

Konzept für Bildanalysen in Hochdurchsatz-Systemen am Beispiel des Zebrabärblings

Zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften
der Fakultät für Maschinenbau
Karlsruher Institut für Technologie (KIT)

genehmigte
Dissertation

von

Dipl. Ing. Rüdiger Alshut
geboren am 14. Dezember 1980 in Mannheim

Tag der mündlichen Prüfung:

12.Juli.2016

Hauptreferent:

Prof. Dr.-Ing. habil. G. Bretthauer

Korreferent:

Prof. Dr.-Ing. Ralf Mikut

Korreferent:

Prof. Dr. Uwe Strähle



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung –
Weitergabe unter gleichen Bedingungen 3.0 Deutschland Lizenz
(CC BY-SA 3.0 DE): <http://creativecommons.org/licenses/by-sa/3.0/de/>

Danksagung

Die vorliegende Dissertation entstand in der Zeit von November 2008 bis Mai 2013 während meiner Tätigkeit am Institut für Angewandte Informatik des Karlsruher Instituts für Technologie. Im Rückblick auf die Entstehungszeit dieser Arbeit habe ich reichlich Grund zur Dankbarkeit.

Herr Prof. Georg Bretthauer hat mir durch die Anstellung diese Arbeit ermöglicht und mich durch Diskussionen und wertvolle Hinweise beim Verfassen unterstützt. Frau Professor Jivka Ovtcharova hat das Korreferat übernommen. Herr Prof. Uwe Strähle ermöglichte durch seine Ideen und Zielsetzungen das Entstehen der Methoden und deren Anwendung für die Biologie. In besonderem Maße bin ich Herrn Prof. Ralf Mikut dankbar. In seiner Arbeitsgruppe und unter seiner Leitung wurden die hier vorgestellten Ergebnisse erarbeitet. Er war für mich sowohl ein großartiger Mentor als auch Ansprechpartner auf professioneller aber auch persönlicher Ebene. Keines unserer Treffen verließ ich nicht ein wenig klüger und motivierter als zuvor.

Herrn Dr. Markus Reischl danke ich für die wertvolle fachliche Unterstützung und die unterhaltsamen Diskussionen, die einen erheblichen Teil zum Gelingen dieser Arbeit beigetragen haben. Sein Wissen und seine Fachkenntnis haben mich manches Mal verblüfft und mir oft weitergeholfen. Auch seinen Humor und seinen sportlichen Ehrgeiz möchte ich nicht missen. Herr Dr. Patrick Waibel hat mit mir, in jener Zeit, durch das gemeinsame Büro mehr Zeit verbracht als jeder andere Mensch und ist seit den ersten Tagen am Institut sowohl in fachlicher als auch in persönlicher Hinsicht wichtig für mich gewesen. Für die weiterhin anhaltende Freundschaft bin ich sehr dankbar. Herrn Dr. Jörg Matthes danke ich für die fachliche und freundschaftliche Unterstützung und die vielen Diskussionen während meiner Anstellung.

Auch die Kollegen und Studenten am Institut für angewandte Informatik waren stets eine große Hilfe bei diversen Problemstellungen. Namentlich erwähnen möchte ich Frau Dr. Jasmin Lampert, Frau Daniela Sanchez, Herrn Prof. Lutz Gröll, Herrn Dr. Alexander Pfriem, Herrn Prof. Christian Pylatiuk.

Für die finanzielle Unterstützung durch das Auslandsstipendium möchte ich dem Karlsruhe House of Young Scientists (KHYS) danken, welches die Zusammenarbeit mit den Kollegen der Harvard Medical School ermöglicht hat. Die kurze aber intensive Kollaboration war außerordentlich gewinnbringend. Von den Kollaborationspartnern gilt mein Dank besonders PhD Randall Peterson, in dessen Lab die Forschung stattfand, PhD Anjali Nath und PhD Xiang Li, die mich sofort kollegial in das Umfeld eingeführt haben und vor allem PhD David Kokel, der mich nicht nur in die Arbeiten integriert hat, sondern mir gleich vom ersten Tag an ein zu Hause in der Fremde geben hat.

Schlussendlich gilt ein großer Dank meiner Familie: Meiner Frau Christine für das fleißige Korrekturlesen, meinen Eltern sowie meiner Schwester Marion für ihre Unterstützung in Worten und Taten.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Bedeutung der Arbeit	1
1.2	Darstellung des Entwicklungsstandes	2
1.2.1	Biologische Grundlagen	2
1.2.2	Grundlagen zur Bildakquise & -verarbeitung für den Hochdurchsatz	12
1.2.3	Existierende Lösungsansätze	15
1.2.4	Offene Probleme	19
1.3	Ziele und Aufgaben	20
2	Neues Konzept zur Versuchsauswertung von Hochdurchsatz-Untersuchungen am Zebrafisch	23
2.1	Übersicht	23
2.2	Neues Konzept zur Versuchsauswertung	23
2.3	Anforderungen an bildbasierte Hochdurchsatz-Untersuchungen	28
2.4	Identifikation der Versuchs- und Auswerteparameter	33
2.4.1	Biologie und Bildakquise	33
2.4.2	Analyse und Interpretation	39
2.4.3	Einfluss der Versuchsparameter auf die Bildqualität	40
2.5	Mathematisches Modell der Hochdurchsatz-Untersuchung	44
2.6	Anforderungsgerechte Versuchsauslegung	52
2.7	Bewertung	56
3	Modulkatalog für die Auswertung und Präsentation von Hochdurchsatz-Untersuchungen am Zebrafisch	57
3.1	Einführung	57
3.2	Bildstrom-Vorverarbeitung	59
3.2.1	Auswahl von Zeitpunkten	59
3.2.2	Neues Verfahren zur Bildfilterung und Normalisierung für inhomogene Datensätze	60
3.2.3	Bildkombination	66
3.3	Segmentierung der Nutzinformation	68
3.3.1	Neue Methode zur Trennung von Einzelobjekt und Hintergrund	68
3.3.2	Neue Methode zur Trennung von mehreren Objekten und Hintergrund	69
3.3.3	Neues Verfahren zum Tracking des Chorions	70
3.3.4	Differenzbilder	72
3.4	Nutzsignalzeitreihen und -merkmale	72
3.4.1	Instantane Merkmale	73

3.4.2	Auf Bewegungen bezogene Merkmale	75
3.4.3	Auf Signalstärke bezogene Merkmale	77
3.5	Merkmalsauswahl	78
3.6	Auswertung und Klassifikation	79
3.6.1	Klassifikation	79
3.6.2	Regression der Dosis-Wirkungs-Kurve	80
3.7	Präsentation	81
3.8	Bewertung	83
4	Implementierung und Skalierung	85
4.1	Übersicht	85
4.2	Umsetzung und Methoden	86
4.3	Skalierung	89
4.4	Werkzeuge	91
4.4.1	Programmpaket Gait-CAD	91
4.4.2	Grafische Oberfläche PIMP	92
4.5	Bewertung	96
5	Anwendungen	97
5.1	Übersicht	97
5.2	Fisch Embryo Test (FET)	98
5.2.1	Auswirkungen der Orientierung der Larve im Ei	99
5.2.2	Fisch Embryo Test (FET) – Koagulation & EC ₅₀ Regression	102
5.2.3	Fisch Embryo Test (FET) – Bewegungen	113
5.3	Photomotor Response (<i>PMR</i>)	124
5.3.1	Übersicht	124
5.3.2	<i>PMR</i> : Hintergrund, Rahmenbedingungen und Schwachpunkte	124
5.3.3	Versuchsplanung: Biologie und Bildakquise	126
5.3.4	Versuchsplanung: Analyse und Interpretation	127
5.4	Bewertung	135
5.4.1	Fisch Embryo Test	135
5.4.2	Photomotor Response	136
6	Zusammenfassung und Ausblick	137
	Symbolverzeichnis	139
	Literaturverzeichnis	147
	Tabellenverzeichnis	165

1 Einleitung

1.1 Bedeutung der Arbeit

Bildbasierte Hochdurchsatz-Systeme ermöglichen das systematische massenweise Testen und Einordnen der Aktivität chemischer Substanzen in diversen Bereichen der Biologie, Toxikologie, Pharmakologie und Genetik. Machbarkeitsstudien zellbasierter Screens mit niedrigen Stückzahlen entwickelten sich in den letzten Jahren schnell zu robusten und umfassenden Industriestandards mit teilweise immensen Stückzahlen von Einzelversuchen sowie einer Größenordnung bis in die Hunderttausende. Allerdings sind Schlussfolgerungen (z.B. Auswirkungen auf den Menschen und die Umwelt) aus den bisher fast ausschließlich zellbasierten Untersuchungen schwierig, Erkenntnisse müssen meist durch langwierige Tierversuche überprüft und schließlich durch diese oft widerlegt werden [189]. Wird das Potenzial des Hochdurchsatzes durch die Ausweitung der Analyse auf einen vollständigen Modellorganismus, wie z.B. die Eier des Zebrafischarbings, ausgeweitet, so sind folgende ökonomische sowie ethische Ziele erreichbar [10, 183, 188]:

- eine höhere Effektivität durch geringere Fehlprognosen,
- die Reduktion von Tierversuchen zur Bestätigung von Ergebnissen,
- die Erhöhung des Durchsatzes bereits bestehender manueller Untersuchungen an Gesamtorganismen wie dem Zebrafischarbling sowie
- die Einsparung von Kosten, Personal und Zeit für die Versuchsdurchführung.

Eine geeignete Hochdurchsatz-Untersuchung erfordert eine auf die speziellen Anforderungen abgestimmte Versuchsauslegung, die schnelle und zuverlässige Erfassung und Analyse der Bildinhalte sowie eine klare zusammenfassende Darstellung der Ergebnisse. Herausfordernd ist, dass sich der auszuwertende Effekt von Versuch zu Versuch stark unterscheidet und durch die hohe Versuchsanzahl die Qualität der aufgezeichneten Daten stark schwankt. Daher beschränken sich bisherige Lösungen auf die (Teil)Automatisierung von einfachen, informationsarmen oder die manuelle Auswertung von komplizierteren, informationsreichen Untersuchungen. In der vorliegenden Arbeit werden ein neues Konzept sowie neue Verfahren zur Gewinnung von Informationen aus Hochdurchsatz-Untersuchungen an Modellorganismen entwickelt und am Beispiel des Zebrafischarbings erprobt, was

für eine schnelle automatisierte Erforschung biologischer Zusammenhänge und der Analyse von Einflüssen innerhalb von Organismen sowie zur Entdeckung neuer Wirkstoffe eine Schlüsseltechnologie darstellt.

1.2 Darstellung des Entwicklungsstandes

1.2.1 Biologische Grundlagen

1.2.1.1 Übersicht Hochdurchsatz-Untersuchung

Als Hochdurchsatz-Untersuchungen werden Batch-Tests oder Siebttests bezeichnet, welchen in der Biologie gewöhnlich genetische, pharmakologische oder toxikologische Fragestellungen zu Grunde liegen und die eine große Anzahl an Einzeluntersuchungen beinhalten. Hochdurchsatz-Untersuchungen sind besonders in der Pharmakologie weit verbreitet und ermöglichen das schnelle Durchführen von Millionen von Tests. Mittels solcher Tests lassen sich beispielsweise rasch Antikörper, (bio)aktive Substanzen oder Gene, die bestimmte biologische Wirkungspfade beeinflussen, identifizieren. Sie ermöglichen, in kurzer Zeit eine große Anzahl von Proben in weiten Teilen automatisiert zu testen. In der Toxikologie wird aus den Ergebnissen der Tests zum einen das Gefahrenpotenzial von Substanzen bestimmt, zum anderen das Verständnis biochemischer Prozesse verbessert. In der Pharmakologie dient die Identifikation bioaktiver Substanzen zudem als Ausgangspunkt zur Erforschung neuer Wirkstoffe, die potenziell als Basis neuartiger Medikamente in Frage kommen, oder wird zur Erforschung biochemischer Prozesse im Organismus herangezogen [18, 64, 127].

Die Abb. 1.1 zeigt den derzeit üblichen Ablauf der Versuchsauslegung für Hochdurchsatz-Untersuchungen. Auf die Parameter der Blöcke „Biologie“, „Bildakquise“ und „Analyse & Interpretation“ wird hierbei Einfluss genommen. Es wird die Art der Bildakquise, die Auswertemethode, Versuchsparameter sowie die Anzahl der durchzuführenden Einzeluntersuchungen manuell festgelegt. Die Auslegung der Untersuchungen wird in der Forschung gemäß der Erfahrung des Experimentators durchgeführt und basiert nicht auf systematischen Entscheidungskriterien.

Unabhängig von Fragestellung, Ziel und Umfang der Hochdurchsatz-Untersuchung wird ein biologisches Signal benötigt, welches sich bei der Datenerfassung beobachten und quantifizieren lässt. Unter dem Begriff *Biosignal* wird meist eine elektrische Spannung als Ergebnis biologischer Aktivitäten verstanden [153] und als Nutzsinal in den folgenden Schritten weiter verarbeitet. Nutzsinalen sind oft mittels Elektroden gewonnene (Änderungen von) Spannungen wie etwa die Elektrokardiografie (EKG), die Elektroenzephalografie (EEG), die Elektromyografie (EMG) oder das Elektretinogramm (ERG). Generalisiert betrachtet umfasst der Begriff jedoch auch alle beobachtbaren und quantifizierbaren nicht-

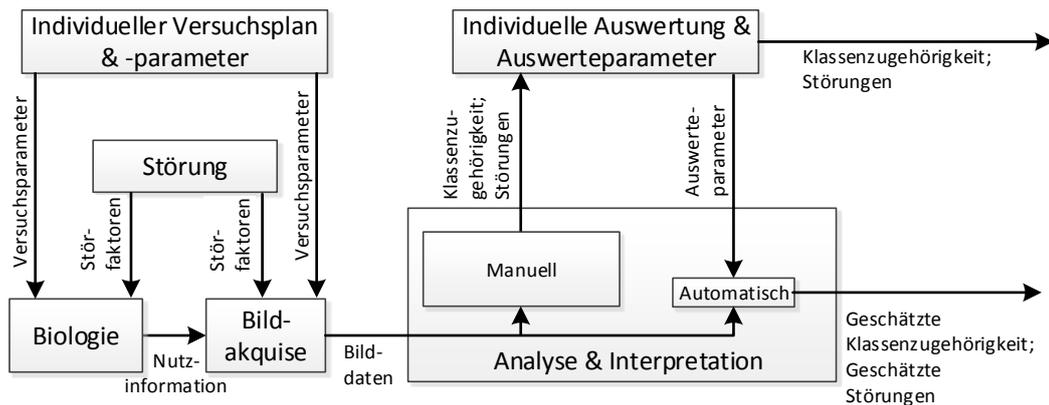


Abbildung 1.1: Bisherige Vorgehensweise bei der Versuchsauslegung bildbasierter Hochdurchsatz-Untersuchungen. Die Größe der Blöcke „Manuell“ und „Automatisch“ lehnt sich an die Verbreitung der Verfahren an (Analyse und Interpretation zurzeit weitgehend manuell).

elektrischen Signale aus z.B. der Mechanik, Akustik, Chemie oder Optik. Für die in der vorliegenden Arbeit betrachteten bildbasierenden Verfahren sind vor allem optische Signale von Interesse, welche durch einen biologischen Effekt in der untersuchten Probe hervorgerufen werden. Die Effekte müssen mit geeigneten Mitteln reproduzierbar erzeugt und sichtbar gemacht werden, beispielsweise durch geeignete Präparation oder Stimulation. Im Anschluss werden die Signale quantifiziert (siehe hierzu Abschnitt 1.2.2), um Rückschlüsse ziehen zu können.

Bei Hochdurchsatz-Untersuchungen hat sich die Anwendung sog. Mikrotiterplatten etabliert [104]. Hierbei handelt es sich um kostengünstige Kunststoffplatten, in welche gitterartig angeordnete Vertiefungen (sog. *Wells*) eingebracht sind. Typische Platten besitzen 96, oder Vielfache von 96 (z.B. 384, 1536, 3456) solcher Vertiefungen und sind kommerziell erhältlich. Je nach durchzuführendem Versuch gibt es unterschiedliche Ausführungen bzgl. der Farbe, beispielsweise mit Seitenwänden aus schwarzem Plastik, oder bzgl. der Form, z.B. mit runden oder flachen Böden der Nöpfchen. Die Mikroskopie ist bei der Mehrheit der Untersuchungen invers, dabei wird das Bild „von unten“ in Durchlicht-Aufnahmetechnik akquiriert. Die Nöpfchen werden in den meisten Fällen mit einer Trägerflüssigkeit, beispielsweise einer wässrigen Lösung von *Dimethylsulfoxid* (DMSO) und, je nach Versuch, einer oder mehreren zu untersuchenden Chemikalien gefüllt. Während die Trägerflüssigkeit in allen Versuchen gleich bleibt, wird die gelöste Chemikalie variiert. Üblicherweise werden zu Kontrollzwecken einige Nöpfchen pro Platte mit lediglich der Trägerflüssigkeit (sog. *Negativkontrollen*) gefüllt und alle Versuche in mehrfacher Ausführung durchgeführt (sog. *Replika*). Zusätzlich werden Nöpfchen mit einem Reaktanten exponiert, dessen Wirkung auf die Organismen bekannt ist (sog. *Positivkontrollen*). Die genannten Maßnahmen dienen bei der Auswertung zur Überprüfung, ob während des Versuchs Ungereimtheiten auftreten und das Testverfahren auch tatsächlich funktioniert. Bei Auffälligkeiten muss die entsprechende Platte oder sogar die gesamte Versuchsreihe ausgeschlossen oder wiederholt werden [64]. Die genauen Anweisungen, welche zur Versuchsdurchführung notwendig sind, werden als „Versuchsprotokoll“ bezeichnet.

Mikrotiterplatten werden nicht nur bei der Versuchsdurchführung angewandt, sondern auch zur Lagerung und Sortierung von chemischen Komponenten verwendet. In einem solchen Fall lagert in jedem Nöpfchen eine Substanz, von der eine detaillierte Dokumentation existiert. Derart befüllte Platten sind als Bibliotheken auch kommerziell erhältlich und werden als „Stock-plates“ (engl.) bezeichnet (vgl. Abb. 1.2). Die Stock-Plates werden oft gefroren gelagert und nicht direkt in der Hochdurchsatz-Untersuchung verwendet, sondern eine definierte Menge der zu untersuchenden Substanzen wird entnommen und der Hochdurchsatz-Untersuchung zugeführt. Meist sind Wirkstoff-Familien auf einer solchen Stock-Plate zusammengefasst, und daher wird die gesamte Stock-Plate auf diese Weise dupliziert und untersucht. In einem typischen Versuch wird in das Gemisch aus Chemikalie und Trägerflüssigkeit eine biologische Einheit gegeben. Bei einer biologischen

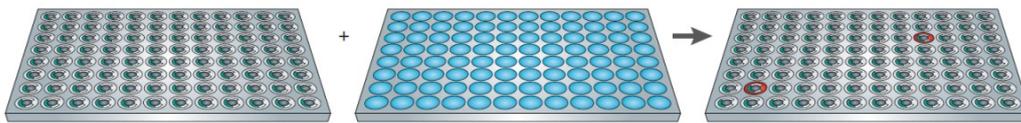


Abbildung 1.2: Mikrotiterplatten organisieren im Hochdurchsatz sowohl die Einzeluntersuchungen, als auch die chemischen Bibliotheken. Im obigen Beispiel aus [91] werden auf die in einer 96 Well-Plate vorbereiteten Larven (links) Substanzen aus einer Bibliothek (Mitte) geben. Im rechten Bild ist das exemplarische Ergebnis dargestellt. Hier wurden zwei Wells markiert, welche sich von den anderen unterscheiden, z.B. durch einen abweichenden Entwicklungsprozess der Larven.

Einheit handelt es sich z.B. um eine Zelle, Zellkultur oder einen biologischen Modellorganismus wie etwa einen tierischen Embryo. Nach einer Inkubationszeit, in welcher der Wirkstoff mit der biologischen Einheit interagieren kann, werden Messungen mittels diverser Detektoren (vgl. Abschnitt 2.4) automatisiert oder manuell von den Einheiten durchgeführt [151]. In bildbasierenden Hochdurchsatz-Untersuchungen erfolgt die Generierung der Messwerte meist durch visuelle Auswertung der Versuche [32]. Die Messwerte werden auf Auffälligkeiten, sog. *hits* (engl. für Treffer), untersucht. Die Möglichkeit, einzelne Zellen z.B. durch fluoreszierende Proteine zu markieren, hat die Robustheit und den Durchsatz der Versuche, zumindest bei zellbasierten Versuchen, in den vergangenen Jahren stark gesteigert [179]. Diese einfach gehaltenen Untersuchungen, bei denen lediglich die Anwesenheit oder Menge des fluoreszierenden Proteins ausgewertet wird, sind selbst ohne fortgeschrittene Bildverarbeitung und Segmentierungstechnik möglich. Kombiniert mit Automatisierungstechnik lassen sich bei derartigen Untersuchungen sehr hohe Zahlen an Einzelversuchen pro Tag erreichen. Der Begriff *Hochdurchsatz* ist indes nicht genau definiert, wird aber weithin bei einem Durchsatz zwischen 10 000 und 100 000 Proben pro Tag verwendet [180].

Manuell durchgeführte Experimente sind meist der Ausgangspunkt, aus dem ein Hochdurchsatz-Experiment abgeleitet wird. Die Versuchsprotokolle der manuellen Untersuchungen im Labor müssen angepasst werden, damit sie sich für den Hochdurchsatz eignen [66]. Wichtige Bedingungen sind:

- Gute Reproduzierbarkeit und Stabilität des Nutzsignals zwischen einzelnen Näpfchen und Platten. Dies gilt sowohl bezüglich des Reaktanten, der biologischen Einheit sowie der verwendeten Geräte (z.B. bei Verwendung mehrerer Mikroskope etc.)
- Hohe Zuverlässigkeit der verwendeten Positiv- und Negativkontrollen
- Ausreichende Sensibilität des Nutzsignals zur Identifikation von Substanzen mit schwachem Effekt auf die biologische Einheit
- Wirtschaftlichkeit der Hochdurchsatz-Untersuchung (standardmäßig gemessen in Kosten pro Näpfchen).

Aus den Abweichungen und Schwankungen zwischen den Platten und Kontrollen lassen sich Kontrollparameter berechnen, welche Auskunft über die Robustheit des Versuchs geben. Zur Validierung, inwieweit sich ein Protokoll für die Hochdurchsatz-Untersuchung eignet, haben sich verschiedene Parameter etabliert (vgl. Tabelle 1.1).

Das Signalfenster und der Z -Faktor dienen zur Abschätzung zwischen dem Minimum und Maximum der Messungen und der Präzision der Werte innerhalb einer Platte und plattenübergreifend. Das Signalfenster bestimmt einen Parameter, welcher die Signifikanz zwischen dem maximalen und minimalen Messwert einer Versuchsreihe quantifiziert. Dieser ist jedoch nicht in gleichem Maße verlässlich wie der Z -Faktor [69]. Der Z -Faktor ermöglicht durch die einheitenlose Skala (von 0 bis 1) den Vergleich verschiedener Versuche und Versuchsdurchläufe miteinander. Hierbei werden auch Kontroll-Näpfchen (Z') und Versuch-Näpfchen (Z'') miteinander verglichen. Niedrige Werte innerhalb der Kontrollen deuten beispielsweise auf eine Kontaminierung der Näpfchen oder ein anderes Problem in der Versuchsdurchführung hin. Die Sensitivität der Untersuchung ist ein Messwert zur Bestimmung der Genauigkeit und Reproduzierbarkeit, welche sich mittels des Minimum-Signifikanz-Verhältnisses (MSR) quantifizieren lässt [9, 42].

1.2.1.2 Zebrabärbling als Modellorganismus für den Hochdurchsatz

Als biologische Einheiten kommen im Hochdurchsatz meist Modellorganismen zum Einsatz. Modellorganismen sind z.B. ausgewählte Pflanzen, Tiere, Bakterien oder Pilze, die mit einfachen Methoden gezüchtet bzw. untersucht werden können und von besonderer Bedeutung für die biologische/biomedizinische Forschung sind. Sie zeichnen sich in der Regel durch eine kostengünstige sowie unkomplizierte Haltung bzw. Nachzucht aus und sind in vielfältiger Hinsicht sehr gut dokumentiert. Darüber hinaus gehören einige dieser Arten zu den ersten Spezies, deren komplettes Genom entschlüsselt werden konnte. Die Wahl des Modellorganismus hängt in der Regel von der biologischen Fragestellung ab. An Einzellern lassen sich beispielsweise gut zellbiologische Prozesse untersuchen. Mehrzellige Lebewesen hingegen werden vor allem für entwicklungsbiologische oder toxikologische Untersuchungen benötigt. Umfassend behandelt werden Modellorganismen in [117].

Parameter	Formel	Beschreibung
Variationskoeffizient	$CV = \frac{\sigma}{\mu} \cdot 100$	Parameter zur Quantifizierung der Messgenauigkeit relativ zum Mittelwert aller Messungen eines Versuchs in Prozent. Akzeptable Werte sind <15%.
Signal-Rausch-Verhältnis	$SNR = \frac{\mu_{\max} - \mu_{\min}}{\sigma_{\min}}$	Parameter zur Quantifizierung der Stärke des Nutzsignals
Signal-Hintergrund-Verhältnis	$SHV = \frac{\mu_{\max}}{\mu_{\min}}$	Parameter, welcher in der Praxis mit Hilfe der Kontrollen ermittelt wird. Erwartungswert der Messwerte eines Versuchs (Nutzsignal) im Verhältnis zum Erwartungswert der Messwerte der Kontrollen (Hintergrund). Akzeptable Werte sind >2.
Signalfenster	$SF = \frac{\mu_{\max} - \mu_{\min} - 3(\sigma_{\max} + \sigma_{\min})}{\sigma_{\max}}$	Parameter zur Quantifizierung der Signifikanz zwischen dem Maximal- und Minimalwert der Kontrollen. Akzeptable Werte sind >2.
Z'-Faktor	$Z' = 1 - \frac{3(\sigma_{\max} + \sigma_{\min})}{ \mu_{\max} - \mu_{\min} }$	Alternative Repräsentation des Signalfensters. Der Wert wird meist sowohl für die Kontrollen als auch für den Versuch selbst ermittelt. Akzeptable Werte sind >0.5.
Minimum-Signifikanz-Verhältnis	$MSR = 10^{2\sigma_d}$	Parameter zur Quantifizierung des minimalen Verhältnisses zwischen zwei Messungen, welches statistisch relevant ist (95% Konfidenz).

Tabelle 1.1: Übersicht ausgewählter Validierungsparameter für Hochdurchsatz-Untersuchungen [66] mit σ Standardabweichung der Messwerte eines Versuchs; $\sigma_{\max/\min}$ Versuch mit größter/kleinster Standardabweichung; σ_d Standardabweichung der ermittelten Wirksamkeiten in logarithmischer Skala (z.B. LC_{50} -Werte, vgl. Abschnitt 3.6.2); μ Erwartungswert der Messwerte eines Versuchs; $\mu_{\max/\min}$ Versuch einer Versuchsreihe mit größtem/kleinstem Erwartungswert.

Bei der Forschung an Modellorganismen wird versucht, allgemein gültige und auf andere Organismen, insbesondere dem des Menschen, übertragbare Erkenntnisse zu erhalten. Ein Grundproblem der etablierten Modellorganismen in der Toxikologie und Entwicklungsbiologie, wie z.B. der Fruchtfliege (*Drosophila melanogaster*), dem Fadenwurm (*Caenorhabditis elegans*) oder dem Krallenfrosch (*Xenopus laevis*), ist die relativ große entwicklungsgeschichtliche Distanz zum Menschen [65]. Säugetiere, wie z.B. die Maus oder die Ratte, besitzen jedoch im Vergleich zu Insekten und Amphibien eine niedrigere Anzahl an Nachkommen und eine ungleich längere Generationszeit. Zudem entwickeln sich die Nachkommen innerhalb des Mutterleibes, wodurch toxikologische Untersuchungen bzw. Manipulationen am Embryo technisch aufwändig sind und sich nur schwer observieren lassen. Säugetierorganismen sind dem menschlichen Organismus am ähnlichsten, so dass an ihnen erzielte Forschungsergebnisse oft aussagekräftige Informationen über Verhältnisse beim Menschen liefern. Jedoch ist eine Hochdurchsatz-Untersuchung aufgrund der oben geschilderten Nachteile und aufgrund der für die Analyse benötigten hohen Anzahl an Messdaten und damit an Individuen in der Praxis häufig nicht wirtschaftlich, technisch schwierig durchführbar und ethisch bedenklich.

Der Zebraquarienfisch ist ein auf dem indischen Subkontinent beheimateter, ausgewachsen ca. 3 bis 5 cm großer Karpfenfisch (*Cyprinidae*) und hat in den wissenschaftlichen Laboratorien der Entwicklungsbiologen eine erstaunliche Karriere gemacht [43]. Auch in der Industrie gewinnt der Zebraquarienfisch an Bedeutung. Eine ständig wachsende Zahl von Laboren versucht, anhand des Fisches biologische Vorgänge zu verstehen, die Ursachen menschlicher Erkrankungen aufzuklären und neue Medikamente zu entwickeln.

Der Zebraquarienfisch verbindet die Vorteile einer kurzen Generationszeit mit einer hohen Anzahl von Nachkommen und technisch leicht zugänglichen Larven, die sich außerhalb des Mutterleibes entwickeln. Als Vertreter der Wirbeltiere (*Vertebraten*) besitzt er alle Organsysteme, die auch im Menschen vorkommen wie z.B. Auge, Hirn, Herz, Drüsen etc. Sein Immunsystem lässt sich im Gegensatz zu dem von Insekten gut mit dem des Menschen vergleichen. Des Weiteren wird davon ausgegangen, dass die überwiegende Zahl aller menschlichen Gene im Fisch nicht nur vorkommt, sondern auch sehr ähnliche oder sogar die gleichen Funktionen besitzen. Zudem sind die Zellen sowie die äußere Fruchthülle um die Larven (das sog. Chorion) transparent, wodurch eine einfache Beobachtung mit dem Lichtmikroskop möglich ist. Die hohe Transparenz der Zellen ermöglicht eine genaue Beobachtung der sich bildenden Organe oder der Herzfrequenz und der Blutströmung am lebenden Embryo. Während von den Zebraquarienfischlarven scharfe Detailaufnahmen die Identifizierung einzelner Zellen ermöglichen, stehen von anderen, lebenden, vertebralem Embryonen oder auch von lebenden menschlichen Embryonen nur weitaus schlechtere Bilder, wie z.B. unscharfe und kontrastarme Ultraschallaufnahmen, zur Verfügung. Laut der EU Richtlinie 2010/63/EU fallen Versuche an frühen Tierembryonen nicht unter das Tierschutzgesetz und müssen somit nicht als Tierversuche angemeldet werden, was den administrativen Aufwand für Versuche mit hohen Stückzahlen verringert. Der Zeitpunkt, ab dem die Embryonen unter das Tierschutzgesetz fallen, ist der Zeitpunkt, ab dem diese eigenständig auf Futtersuche gehen. In [167] wird der Zeitpunkt von 120 Stunden nach der Befruchtung [hpf] empfohlen. Weiterführende Literatur zum Zebraquarienfisch als Mo-

dellorganismus in der Forschung findet sich z.B. in [86, 107, 161, 169].

Die Zebraabärblingslarve hat für Hochdurchsatz-Untersuchungen großes Potenzial. Bereits 1996 wird von [46] auf zahlreiche Literatur zum Potenzial des Zebraabärblings hingewiesen, die Vorteile des Fisches in vielen Forschungsgebieten ausführlich diskutiert und dieser als der Nachfolger der Fruchtfliege bezeichnet. An der Fruchtfliege als Modellorganismus gelangen bahnbrechende Erfolge bei der Identifizierung essentieller Gene in der Embryonalentwicklung. Diese Arbeit von Christiane Nüsslein-Volhard wurde 1995 mit dem Nobelpreis für Medizin und Physiologie ausgezeichnet [120]. So förderte die Europäische Kommission seit Anfang 2004 bis Ende 2009 ein Gemeinschaftsprojekt von 15 europäischen Forschungseinrichtungen zur Forschung am Zebraabärbling mit dem Namen *ZF-MODELS - Zebrafish Models for Human Development and Disease* unter Leitung des Tübinger Max-Planck-Instituts für Entwicklungsbiologie mit 12 Millionen Euro. Ziel des Projektes war es, am Zebraabärbling die Entwicklung von Wirbeltieren sowie menschliche Erbkrankheiten zu erforschen. Während des Projektes wurden die Abläufe und Zusammenhänge im Zebraabärbling als Vertreter der Wirbeltiere untersucht. Die Ergebnisse waren derart vielversprechend, dass seit Juli 2010 das Folgeprojekt *Zebrafish Regulomics for Human Health (ZF-HEALTH)* für einen Zeitraum von fünfeinhalb Jahren mit 11,4 Millionen Euro gefördert wird. Das ausgesprochene Ziel des neuen Programmes ist u.a. die Durchführung von Hochdurchsatz-Phänotypisierung mittels Verhaltens-Untersuchungen am Zebraabärbling in 3D- und 4D-Bildaufnahmen.

Verglichen mit anderen Wirbeltieren verläuft die Entwicklung der Larve des Zebraabärblings sehr schnell und fast jeder Entwicklungszeitpunkt lässt sich verfolgen und in Bildern aufzeichnen. Die Eizelle, die sich oberhalb der Dotterkugel befindet, beginnt sich bereits Minuten nach dem Ablachen und der Befruchtung zu teilen (Abb. 1.3a). Der beschriebene Vorgang wurde bereits mittels Bildverarbeitung geprüft und verfolgt von [74, 108, 109]. Diese Art von Auswertungen sind jedoch bisher aufgrund der großen Datenmengen nicht für den Hochdurchsatz geeignet.

Aus den ersten Zellen bildet sich innerhalb von vier Stunden durch Zellteilungen eine Gruppe aus einigen tausend Zellen (Abb. 1.3b), welche in den folgenden sechs Stunden den Dotter umwachsen. Zehn Stunden nach der Befruchtung zeichnet sich allmählich der Kopf der Larve ab (Abb. 1.3c-e) und nach 18 Stunden sind am Kopf die ovalen Vorläufer der Augen sowie im Rumpf die ersten Muskeln erkennbar. Kurze Zeit später beginnt der Schwanz sich zu strecken, die meisten Organe sind angelegt, das Herz fängt an zu schlagen und die nur knapp 24 Stunden alte Larve bewegt sich bereits im Chorion. Während des dritten Tages nach der Befruchtung ist die Larve fertig entwickelt. Sie schlüpft und schwimmt die ersten kurzen Strecken (Abb. 1.3f-i). Nach fünf Tagen schließlich hat sie ihren Dottervorrat aufgebraucht und sucht selbstständig nach Futter. Trotz der rasanten Entwicklung dauert es dann immerhin noch drei Monate, bis ein Zebraabärbling geschlechtsreif wird. Eine genaue Beschreibung des Entwicklungsprozesses findet sich in [76]. Die beschriebenen Merkmale des Zebraabärblings lassen sich lichtmikroskopisch leicht observieren. Alle Aufnahmen in Abb. 1.3 wurden mit einem herkömmlichen Lichtmikroskop aufgezeichnet. Es existieren bereits automatisierte Systeme zum Erkennen diverser Phä-

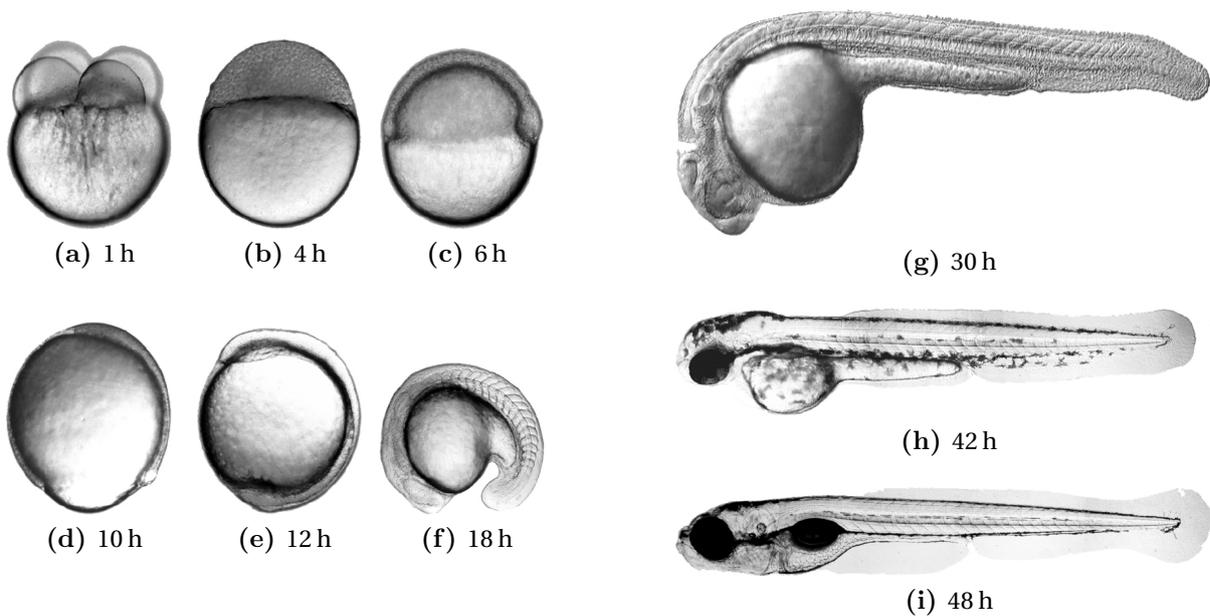


Abbildung 1.3: Ausgewählte Entwicklungsstadien der Zebrafischlarven. Unter den Bildern ist jeweils das ungefähre Alter in Stunden nach der Befruchtung angegeben. [177]

notypen oder Endpunkte [49, 160, 173].

Zum Verständnis eines weiteren Vorteils des Zebrafisch, dem Einsatz sog. Transgene, wird im Folgenden kurz auf das Grün Fluoreszierende Protein eingegangen. 1961 beschrieb und extrahierte erstmals Osamu Shimomura das Grün Fluoreszierende Protein (GFP) einer Tiefseequalle [156, 157], welches bei Anregung mit blauem oder ultraviolettem Licht grün fluoresziert. Von unschätzbarem Wert ist das GFP, da es sich als Marker für die Genexpression nutzen lässt. Eine mit Hilfe des Proteins markierte Zelle erfüllt die gleiche Funktion wie eine nicht markierte Zelle, mit dem Unterschied, dass die markierten Zellen nach Anregung mit einer passenden Wellenlänge fluoreszieren d.h., leuchten. Ein Organismus, in den ein solcher Marker eingebracht wurde, wird als *Transgen* bezeichnet. Auch im Zebrafisch lassen sich auf die beschriebene Art Transgene generieren. Unmittelbar mit der Entstehung einer derart markierten Zelle im Zebrafisch lassen sich räumliche und zeitliche Verteilung und Bewegungen der Zelle studieren. Für die Entdeckung und Extraktion des GFP wurde im Jahr 2008 der Nobelpreis für Chemie an Osamu Shimomura, Martin Chalfie und Roger Tsien verliehen. Es existieren neben dem GFP bereits Varianten in anderen Farben (z.B. CFP (cyan) oder YFP (yellow)), so dass sich auch mehrere Zelltypen des Zebrafisch gleichzeitig und unabhängig beobachten lassen [134, 172]. Die Möglichkeit, die genannte Technik des Markierens von Zellen im Zebrafisch anzuwenden [92], eröffnet für den Modellorganismus alle Möglichkeiten, die die moderne Fluoreszenzmikroskopie bietet [159]. Hervorzuheben sind hier Aufnahmetechniken, die es ermöglichen, Zellbewegungen über die Zeit und/oder in 3D zu akquirieren. Die erzielbare Auflösung reicht von makroskopischen Bildern einfachster Aufbauten [17] über die bereits als Standard geltende Konfokal-Mikroskopie [129], die spezialisierte Hochdurchsatz-Mikroskopie [90] bis hin zur Nanoskopie, welche sich der Fluoreszenz bedient, um

die physikalische Beugungsgrenze zu überwinden [8, 56]. Die einfachste Variante ist das Akquirieren des Fluoreszenzkanals mittels eines herkömmlichen Lichtmikroskops, welches mit einem Fluoreszenzfilter ausgestattet ist. Beim Standard für die 3D-Bildakquise, der Konfokal-Mikroskopie, wird nur ein sehr kleiner Teil des Präparats zu jedem Zeitpunkt zur Fluoreszenz angeregt. Dies ermöglicht das Abrastern der Proben und so das Rekonstruieren von 3D-Modellen im Rechner. Das Abrastern benötigt jedoch Zeit und somit ist das Akquirieren von zeitlichen Veränderungen schwierig bzw. die zeitliche Auflösung oftmals zu begrenzt. Eine weitere wichtige Mikroskopie-Technik, die sich der Fluoreszenz bedient und beim Zebrabärbling Anwendung findet, ist die Single Plane Illumination Microscopy (SPIM) [63]. Hierbei ist das Prinzip ähnlich wie bei der Konfokal-Mikroskopie. Allerdings wird nicht lediglich ein Punkt zur Fluoreszenz angeregt, sondern ein scharf definiertes Lichtvolumen. Durch das hohe Aspektverhältnis des Volumens wird auch von einem Lichtblatt gesprochen. Diese Variante ermöglicht eine wesentlich höhere Akquise-geschwindigkeit bei einem guten Signal-Rausch-Verhältnis.

1.2.1.3 Relevante Forschungsgebiete für eine bildbasierte Hochdurchsatz-Untersuchung im Zebrabärbling

Der Vorteil, den eine Hochdurchsatz-Untersuchung im gesamten Organismus gegenüber der Zelle hat, ist, dass Verknüpfungen und Gene in den Organismen oft redundant vorkommen oder erst durch das Zusammenspiel aller biochemischen Prozesse in der Wirkkette der Einfluss von Chemikalien aussagekräftig erforscht werden kann. Der Zebrabärbling vereint die beschriebenen Möglichkeiten der Bildakquise mit den oben dargelegten Vorteilen in Haltung, gentechnischer Relevanz, kurzen Generationszyklen und klaren Mikroskopieaufnahmen. Somit erscheint das Tier als optimal geeignet für neue Hochdurchsatz-Untersuchungen mit bisher nicht erreichter Aussagekraft [10, 91, 97, 162, 183, 188]. Des Weiteren existieren bereits Firmen, welche spezielle Hochdurchsatz-Untersuchungen auf Auftragsbasis zur Entwicklung von Medikamenten am Zebrabärbling durchführen. Eine Liste von kommerziellen Anbietern von Hochdurchsatz-Untersuchungen findet sich in [19]. Bezüglich der Einsatzgebiete für Hochdurchsatz-Untersuchungen lassen sich drei weitläufige und biologisch relevante Bereiche unterscheiden:

Entwicklungsbiologische oder genetische Forschung: Hier wird das Verständnis über Funktionen und Eigenschaften von Genen und Mutationen erforscht. Die entwicklungsbiologische Forschung oder Ontogenese (aus dem Griechischen für *Wesen Geburt Entstehung*) beschäftigt sich mit den Vorgängen des Wachstums und der Entwicklung einzelner Organismen. Sie hat ihren Ursprung in der Embryologie. Im Hochdurchsatz lassen sich die genetische Kontrolle von Zellwachstum, Zelldifferenzierung und Zellspezialisierung in verschiedenen Zelltypen und Organen erforschen. Hierbei werden diverse Techniken wie *Forward Genetics* und *Reverse Genetics* angewandt [87]. Damit ist es beispielsweise möglich, gezielt die Expression von Genen auszulösen oder zu verhindern. Durch die Techniken lassen sich Einblicke in die Funktion und Wirkungsweise der Gene gewinnen. Das Potenzial von Hochdurchsatz-Untersuchungen im Zebrabärbling wird u.a. in [24, 127, 144] aufgezeigt.

Toxikologische Forschung: Hier wird der Einfluss von Chemikalien auf den Organismus erforscht und Substanzen bezüglich der Stärke ihrer Toxizität untersucht. Es existieren auch bereits Regularien, die den sog. *Fisch Embryo Test* (FET) beschreiben und festlegen, wie eine Hochdurchsatz-Untersuchung für toxikologische Versuche und Frischwasser-Untersuchungen durchgeführt werden soll [22, 68, 105, 118, 121, 122]. Der Test soll dazu dienen, eine Vielzahl von Tierversuchen an ausgewachsenen Fischen zu vermeiden. Die Toxizitäts-Tests müssen laut den gesetzlichen Vorschriften, wie sie in REACH¹ gefordert sind, für nahezu jede Chemikalie auf dem Markt durchgeführt werden [83, 178]. Substanzen, vor allem Kunststoffe, Verbundstoffe und Polymere, die eine Jahresproduktion von einer Tonne erreichen, müssen nach dem neuen EU Gesetz ökotoxikologisch getestet werden. Unter das Gesetz fallen seit 2007 ca. 30.000 Substanzen. Die Toxizitäts-Tests mittels Tierversuchen durchzuführen, würde bis zu 8 Mio. EUR pro Substanz kosten und jeweils bis zu 5 Jahre dauern [131]. Daher investierte die EU seit 1986 über 300 Mio. EUR in alternative Tests, welche die Anzahl an Tierversuchen reduzieren können [1]. Ein weiterer (öko)toxikologischer Anwendungsbereich ist die Bestimmung der Toxizität von Sedimentgestein [73]. Hierfür wird der *Dantox*-Test (von *danio rerio*, der lateinischen Bezeichnung des Zebrafisch) in einem BMBF-Projekt gefördert². Die Aussagekraft und Anwendbarkeit von Hochdurchsatz-Untersuchungen im Zebrafisch wurde allgemein bestätigt, jedoch wird diskutiert, inwieweit die Membranfunktion der Frucht- oder Eihülle um die Zebrafischlarve die Ergebnisse gegenüber den herkömmlichen Testverfahren verfälscht [21, 57, 84, 85, 171, 183]. Das Potenzial der Hochdurchsatz-Untersuchung im Zebrafisch auf diesem Gebiet wird ausführlich beschrieben in [24, 45, 58, 106, 152, 163, 167, 183]. Limitierungen sind dem Verfahren allerdings durch die Bindung an das Wasser gesetzt. So können z.B. nicht wasserlösliche Substanzen oder Gase mit Hilfe des Fisch Tests oder Fisch Embryo Tests nur unzureichend getestet werden.

Pharmakologische Forschung: Hier werden Chemikalien und kleine Moleküle bezüglich ihrer Eignung als Medikamente geprüft. Das Ziel ist es, Wirkstoffe zur Behandlung bisher unheilbarer Krankheiten oder Alternativen zu bereits etablierten Wirkstoffen zu finden [23, 48, 64]. Bei pharmakologischen Hochdurchsatz-Untersuchungen

- wird die Wirkung und das Wirkungsspektrum qualitativ und quantitativ ermittelt,
- wird versucht, den Angriffspunkt sowie den Wirkungsmechanismus zu klären,
- der Einfluss auf verschiedene Organfunktionen festgestellt,

¹Verordnung (EG) Nr. 1907/2006 des Europäischen Parlaments und des Rates vom 18. Dezember 2006 zur Registrierung, Bewertung, Zulassung und Beschränkung chemischer Stoffe (REACH)

²DanTox - Entwicklung und Anwendung eines Verfahrens zur Ermittlung spezifischer Toxizität und molekularer Wirkungsmechanismen sedimentgebundener Umweltschadstoffe mit dem Zebrafisch (*Danio rerio*)

- die lokale und allgemeine Verträglichkeit geprüft und
- auf toxische Effekte geachtet.

Das Potenzial für Hochdurchsatz-Untersuchungen im Zebrafisch auf diesem Gebiet wird ausführlich beschrieben in [12,13,20,33,184]. Eine Übersicht, welche detektierbaren oder vergleichbaren Auswirkungen menschliche Krankheiten bzw. Krankheitserreger auf den Zebrafisch haben sowie eine Tabelle wichtiger Forschungsarbeiten bietet [91].

1.2.2 Grundlagen zur Bildakquise & -verarbeitung für den Hochdurchsatz

Um eine Hochdurchsatz-Untersuchung durchzuführen, müssen, wie bereits in Abschnitt 1.2.1.1 beschrieben ist, Biosignale erfasst werden. Dabei ist ein solches Signal immer notwendige Voraussetzung, unerheblich, auf welchem der Anwendungsgebiete aus Abschnitt 1.2.1.3 die Hochdurchsatz-Untersuchung durchgeführt werden soll. Für die unterschiedlichen Signale bietet der Markt spezielle Detektoren an. Die Dimension und der Aufwand der Akquise müssen dem Signal angepasst werden, um ein für den Hochdurchsatz geeignetes Signal-Rausch-Verhältnis zu erreichen [38]. Wird die Auswahl auf optische Signale beschränkt, so quantifizieren alle Detektoren die Menge an einfallendem Licht. Das auf den Sensor projizierte Signal ist zunächst nichts weiter als eine zweidimensionale, zeitabhängige, kontinuierliche Verteilung von Lichtenergie. Die Lichtenergie kann ihren Ursprung in einer Lichtquelle oder in einer fluoreszierenden Zelle haben. Der Detektor hat die Aufgabe, drei wesentliche Schritte zu erfüllen [70]:

- Eine räumliche Abtastung der kontinuierlichen Lichtverteilung,
- eine zeitliche Abtastung der daraus resultierenden Funktion und
- eine Quantifizierung zur digitalen Darstellbarkeit und Verarbeitung.

Ein typisches Sensorelement ist eine Photodiode, deren Messung in der digitalen Repräsentation als Pixel bezeichnet wird. Je nach Anordnung der Sensorelemente, geschickter Auswahl von Beleuchtungstechniken, zeitlichen Aufnahmen usw. lässt sich die Dimension, also die zeitliche bzw. räumliche Abtastung des akquirierten Datensatzes, beeinflussen. Die räumliche Abtastung (*spatial sampling*) erfolgt in der Regel durch die Geometrie des Detektors, die zeitliche Abtastung (*temporal sampling*) geschieht durch Steuerung der Zeit, über die die Messung der Lichtmenge durch die einzelnen Sensorelemente erfolgt. Werden zu mehreren Zeitpunkten Bilder akquiriert, entsteht eine Bildsequenz. Wie sich mehrdimensionale Repräsentationen von Zebrafischlarven mittels solcher Detektoren akquirieren lassen, wird im Verlauf der vorliegenden Arbeit aufgezeigt.

Wird lediglich ein Sensorelement oder eine Reihe solcher verwendet, lassen sich eindimensionale Signale akquirieren. In [16, 126, 132] werden solche Detektoren beispielsweise eingesetzt, um die Menge an abgeschattetem Licht oder die Stärke an Fluoreszenz zu quantifizieren. Eine kontinuierliche Aufnahme von Messwerten ist mittels fast aller Detektoren möglich und ergibt eine Zeitreihe. Wird eine Zebraabärlingslarve definiert an einem Sensor vorbei geführt, so lässt sich ein Profil über der Zeit aufzeichnen. Ein besonders hoher Durchsatz lässt sich erzielen, wenn Zebraabärlingslarven in einem Schlauch fortwährend geprüft werden. Die Information über den Ort der Quelle des Signals geht hierbei verloren oder ist durch die Anbringung des Sensors fix.

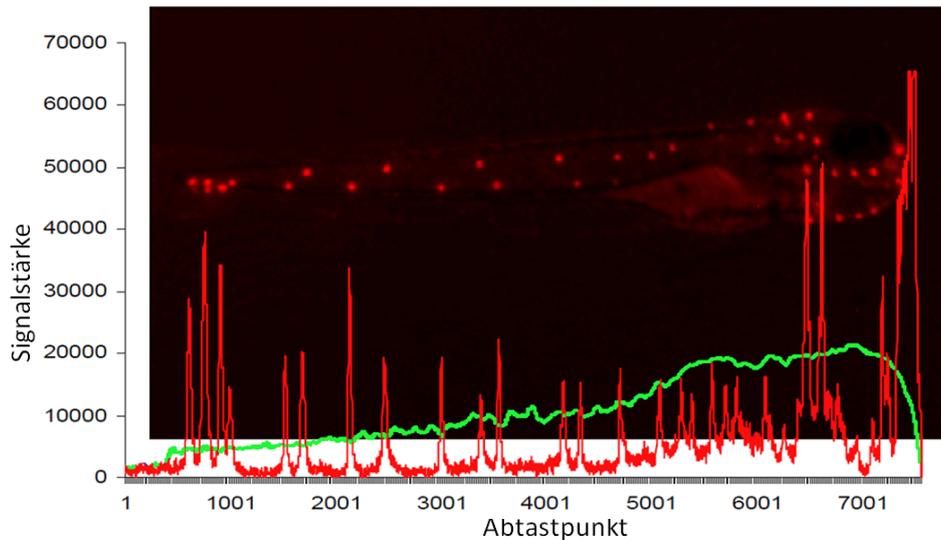


Abbildung 1.4: Darstellung eines Zebraabärlings mittels 1D-Profilen im Vergleich zu einer 2D-Bildaufnahme [16]. Die grüne Linie zeigt die Menge an abgeschattetem Licht, während die rote Linie die Menge des Fluoreszenzsignals an der jeweiligen X-Position akkumuliert. Die jeweiligen Signale wurden getrennt aufgezeichnet und dann zur Visualisierung überlagert.

Abb. 1.4 zeigt eine Überlagerung einer zweidimensionalen Fluoreszenzaufnahme und Profile eindimensionaler Detektoren im Vergleich. Der erste Detektor (grüne Linie) quantifiziert die Menge an abgeschattetem Licht, d.h. je höher der Wert, desto weniger Licht durchdringt die Larve. Der zweite Detektor (rote Linie) quantifiziert die Menge an von der Larve fluoreszierend ausgesendetem Signal, welches aufgrund eines anregenden Lasers ausgesendet wird. Der Vorteil der Methode liegt im erreichbaren Durchsatz und im günstigen Preis, während auf die Information der zweiten Ortskoordinate und eine bessere örtliche Auflösung verzichtet werden muss.

Eine typische Kamera, wie sie auch zur Akquise von Zebraabärlingen verwendet wird, enthält eine regelmäßige und rechtwinklige Anordnung solcher Sensoren bzw. Sensorelemente. Das aufgezeichnete Bild ist somit eine Matrix \mathbf{I} mit den ganzzahligen Koordinaten $i_x = 1 \dots I_x$; $i_y = 1 \dots I_y$ von Pixelwerten $I(i_x, i_y)$ innerhalb der Farbtiefe ϕ :

$$\mathbf{I} \in \mathbb{N}^{I_x \times I_y} \quad \text{mit} \quad i_x, i_y \in \mathbb{N} \quad \text{und} \quad I(i_x, i_y) < \phi \quad (1.1)$$

Für die Farbtiefe ϕ werden üblicherweise positive ganze Zahlen im Bereich $[0 \dots 2^\phi - 1]$ benutzt. Ein typisches Bild kann beispielsweise $\phi = 8$ bit besitzen und die Intensitätswerte $[0 \dots 255]$ annehmen. Kleine Werte stehen hierbei für dunkle und hohe Werte für helle Pixel.

Die 2D-Repräsentation eines dreidimensionalen Zebra**ä**rblings ist je nach verwendetem Mikroskop, Objektiv, numerischer Apertur etc. meist nicht über die gesamte Tiefe des Modellorganismus scharf. Die optischen Limitationen beschränken die Fokusebene auf einen bestimmten Bereich, die sog. Tiefenschärfe [145]. Eine dennoch scharfe Abbildung des gesamten Zebra**ä**rblings kann erfolgen, indem mehrere Bilder unterschiedlicher Fokusebenen fusioniert werden. Die Technik wird als *Extended Focus* bezeichnet [47, 170]. Hierbei wird über alle aufgenommenen Fokusebenen mittels einem Algorithmus (beispielsweise der Wavelet-Transformation) bestimmt, welche Pixel in welcher der Fokusebenen eine scharfe Abbildung des Objektes sind. Lediglich die scharf abgebildeten Pixel werden in einem einzigen resultierenden Bild mit *erweitertem* Fokus gespeichert. Aus der räumlichen Position des Pixels kann weiter eine dreidimensionale Repräsentation rekonstruiert werden. Da keine Information vom Inneren des Volumens und von den verdeckten Flächen vorhanden ist, wird hier von einer 2,5D-Abbildung gesprochen (vgl. Abb. 1.5).

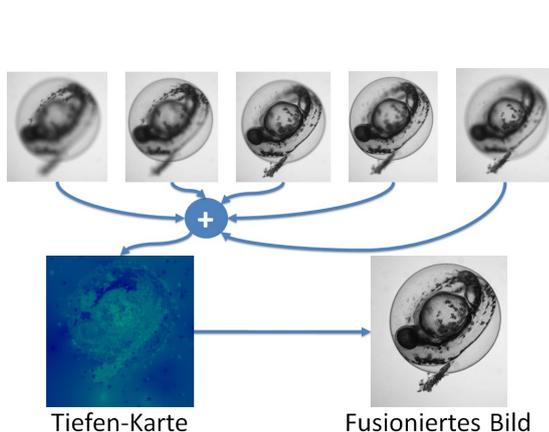


Abbildung 1.5: Mikroskopaufnahmen verschiedener Fokusebenen eines Zebra**ä**rblings; aus der Wavelet-Transformation berechnete Tiefenkarte; fusioniertes Bild mit erweiterter Tiefenschärfe

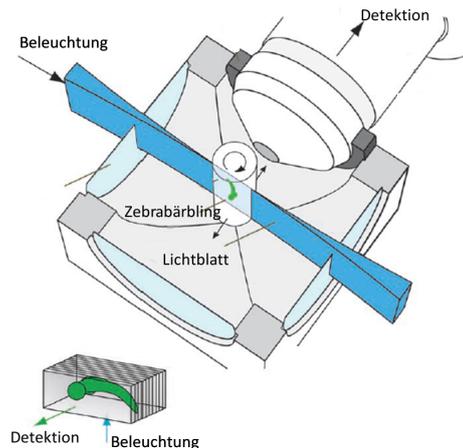


Abbildung 1.6: Prinzipskizze eines Lichtblatt-Mikroskops nach [63] zur Akquise eines 4D-Datensatzes. Der Detektor ist hier am Ende des Strahlenganges des Mikroskops angebracht.

Vollständige Volumenmodelle des Fisches lassen sich zerstörungsfrei aus Schichtaufnahmen, wie sie beispielsweise mittels der Konfokal-Mikroskopie akquiriert werden können, erstellen. Die Schichtaufnahmen werden im Rechner zu einem vollständigen 3D-Modell verbunden. Der Detektor ist auch hier ein herkömmlicher 2D-Detektor. Entscheidend für die Ermöglichung der dreidimensionalen Akquise ist das Vorhandensein und die definierte Anregung von fluoreszierenden Zellen im Zebra**ä**rbling sowie die spätere Zuordnung der örtlichen Information. Werden die Techniken der dreidimensionalen Fluoreszenzmikrosko-

pie mit einer zeitlichen Aufnahme kombiniert wie beispielsweise bei der *SPIM*-Methode (vgl. Abb. 1.6), so lassen sich vierdimensionale Datensätze aufzeichnen [63, 74].

1.2.3 Existierende Lösungsansätze

Trotz den in Wirtschaft und Forschung weit verbreiteten, meist zellbasierten Hochdurchsatz-Untersuchungen existieren nur wenige Publikationen, die, basierend auf Zebrabärblingen, Hochdurchsatz-Untersuchungen realisieren. Aus der Erfahrung der zellbasierten Untersuchungen ist bekannt, dass sowohl die Durchführung der Hochdurchsatz-Untersuchung als auch deren Auslegung strukturiertes Vorgehen verlangt. Die wichtigsten Arbeiten und deren Vorgehensweise werden im Folgenden kurz zusammengefasst.

1.2.3.1 Existierende bildbasierte Hochdurchsatz-Untersuchungen im Zebrabärbling

In der Literatur sind verschiedene Ansätze zu finden, die das große Potenzial des Zebrabärblings für den Hochdurchsatz tatsächlich nutzen. Alle unterscheiden sich in der Art der Umsetzung, wie etwa dem Automatisierungsgrad oder der verwendeten Software. Es zeichnen sich bisher weder standardisierte Vorgehensweisen ab, noch hat sich eine Software etabliert. Zwar existiert in der Literatur eine freie, auf die Verwendung mit Zebrabärblingen spezialisierte Bildverarbeitungs-Software. Die Software befindet sich allerdings in einem sehr frühen Entwicklungsstadium und wurde seit 2009 nicht weiter aktualisiert [94–96]. Alle veröffentlichten, bildbasierten Siebttest-Untersuchungen größeren Umfangs verfolgen ein eigens erstelltes Konzept mit einer eigens erstellten Bildverarbeitungsroutine oder benötigen eine kommerzielle und spezialisierte Plattform [28, 52]. Tabelle 1.2 bietet eine Auswahl an Hochdurchsatz-Untersuchungen, eine weitere Übersicht findet sich in [89].

In den Untersuchungen von [49] werden sowohl Hellfeld-Aufnahmen in unterschiedlichen Fokusebenen als auch Fluoreszenzsignale, nach manueller Vorarbeit, automatisiert akquiriert und die Stärke des Fluoreszenzsignals mittels spezialisierten Algorithmen den Regionen im Fisch zugeordnet und miteinander verglichen. Da es sich beim Ziel der Untersuchung um die Identifikation von Genkombinationen handelt, kommen keine Chemikalien während des Screenings zum Einsatz. Die hochspezialisierte Bildverarbeitung leistet sowohl die Rotation der Fische als auch die vollautomatische Erkennung von Regionen sowie die Messung und Zuordnung von Fluoreszenzsignalen zu einzelnen Bereichen. Schließlich erfolgt die Auswertung und Visualisierung der Ergebnisse in sog. *Fingerprints*, bei denen der Stärke des Signals eine Helligkeit und der Region im Fisch jeweils eine Farbe und ein Teil eines Rechtecks zugeordnet werden. Dies ermöglicht das schnelle Vergleichen und Ordnen der Ergebnisse sowie das Ableiten von Schlussfolgerungen. Die gesamte Programmierung, Bild- und Datenauswertung erfolgt mit der Software *MATLAB*³. Es sei

³The MathWorks, Inc.

Jahr	Art des Screens	Anz. Chem.	Anz. Fische	Methode	Dim. der Bildakquise	Ref.
2012	Mutant-Identifizierung		>1000	Bildauswertung und -akquise automatische Station: Tecan Infinite M1000	2D + Fluoreszenz	[174]
2012	Medikamenten-Untersuchung	3	3000	Automatisches Mikroskop, Hellfeld + Fluoreszenz, automatisches Sortierverfahren, manuelle Bildverarbeitung (<i>ImageJ</i>)	1D+2D + Zoom + Fluoreszenz	[35]
2011	Mutant-Identifizierung	5	1000	Automatisches Mikroskop, Hellfeld, Skriptbasierte Bildverarbeitung	2D	[93]
2011	Transgenetische Reporter	3		Automatisches Mikroskop, Hellfeld + Fluoreszenz, Skriptbasierte Bildverarbeitung (<i>Labview</i>)	2D + Zoom + Fluoreszenz	[130]
2010	Verhaltens-Untersuchung	14 000	70 000	Automatisches Mikroskop, Kamera, <i>Metamorph</i> -Control Software	2D + Zeit	[80]
2010	Verhaltens-Untersuchung	5 648		Megapixel Objektiv, Infrarot Kamera, <i>Viewpoint</i> Tracking Software	2D + Zeit	[143]
2010	Medikamenten-Untersuchung	34	1 700	Automatisches Mikroskop, Hellfeld + Fluoreszenz, Skriptbasierte Bildverarbeitung (<i>Labview</i>)	2,5D + Fluoreszenz	[41]
2010	Entwicklung einer Hochdurchsatz-Methode		450	speziell angefertigte Hochdurchsatz-Lösung, mehrere Detektoren, keine automatisierte Bildauswertung	1D + 2D + Fluoreszenz	[126]
2009	Verhaltens-Untersuchung	3		Infrarot Kamera, <i>Noldus</i> Tracking Software	2D + Zeit	[67]
2009	Mutant-Identifizierung	115		Manuelles Mikroskop, Skriptbasierte Bildverarbeitung	2D	[30]
2009	Identifikation von Genkombinationen		17 730	Automatisiertes Mikroskop, Hellfeld + Fluoreszenz, Skriptbasierte Bildverarbeitung (<i>MATLAB-Software</i>)	2,5D + Fluoreszenz	[49]

Tabelle 1.2: Ausgewählte Zebrafisch-Untersuchungen im Hochdurchsatz oder mit Hochdurchsatz-Potenzial mit Anzahl an Fischen oder Substanzen (falls in der Veröffentlichung angegeben). Eine weitere ausführliche Übersicht und Gegenüberstellung bietet [113].

angemerkt, dass in [82] vorgeschlagen wird, solche *Fingerprints* mittels moderner *cloud-basierter* Techniken zu vernetzen.

In [30] führen die Autoren die Hochdurchsatz-Untersuchung zwar vollständig manuell durch, zeigen jedoch nur das Hochdurchsatz-Potenzial ihrer Methode auf, ohne derartige Versuche durchzuführen. Ziel ist es, einen mutierten Zebrabärbling zu identifizieren, dessen Phänotyp ein gekrümmter Rücken ist. Die Autoren stellen einen sehr einfachen Bildverarbeitungsalgorithmus vor, der in der Lage ist, den Rücken der Tiere zu finden und auf Krümmung zu prüfen. Bei zu dicht beieinander auftretenden Fischen versagt die Methode jedoch. Die Autoren vergleichen die Ergebnisse ihrer automatischen mit einer manuellen Analyse, kommen zu übereinstimmenden Ergebnissen und schließen so auf Hochdurchsatz-*Fähigkeit* des Verfahrens. Zur Steuerung der Mikroskope kommt die Software *Metamorph*⁴ zum Einsatz. Die Autoren machen keine Angabe, in welcher Software die Bildverarbeitung realisiert wurde.

In [67] wird die zurückgelegte Distanz von Zebrabärblingslarven sowohl bei Helligkeit als auch im Dunkeln gemessen. Die Vorbereitung der Larven erfolgt manuell, während die Bildakquise innerhalb einer kommerziell erhältlichen gekapselten Box der Firma *Noldus*⁵ ohne Verwendung eines Mikroskops erfolgt. Die Bilder werden lediglich mittels einer Industriekamera akquiriert. Die Bildverarbeitung erfolgt durch Verwendung einer Tracking-Software der Firma *Noldus*.

Das Ziel der Autoren in [126] ist die Durchführung einer Machbarkeitsstudie, um eine neue Automatisierungstechnik zur Analyse von Zebrabärblingen zur Verfügung zu stellen. Die Autoren stellen einen komplexen Apparat vor, der den Fisch innerhalb eines Schlauchs in einer Glas-Kapillaren transportiert, dort rotiert und währenddessen Bilder akquiriert. Durch die Akquise von allen Seiten wird das Problem der undefinierten Lage der Tiere gelöst, welches bei herkömmlicher Mikroskopie unumgänglich ist. Eine belastbare Anwendung der Konstruktion im Hochdurchsatz wurde bis zum heutigen Zeitpunkt jedoch noch nicht vorgestellt.

In [41] bedienen sich die Autoren eines automatisierten Mikroskops, wie es in [90] vorgestellt ist. Es werden automatisiert mehrere Fluoreszenzkanäle akquiriert und es wird gezeigt, dass die Auswertung mittels eines automatisierten Skriptes erfolgen kann. Anders als in der Arbeit von [49] wird hier nicht der gesamte Fisch in Regionen aufgeteilt, sondern es werden lediglich Regionen von Interesse im Fisch (sog. Neuromasten) erkannt und die Menge an Fluoreszenz gemessen. Ein Vergleich der automatisierten mit der manuellen Auswertung ergibt ähnliche Ergebnisse. Da die Bildverarbeitung nicht in der Lage ist, jede der manuell erkennbaren Regionen zu identifizieren, ist es notwendig, die Menge an untersuchten Fischen zu erhöhen, um zu gleichen Ergebnissen von manueller und automatischer Analyse zu kommen. Die Steuerung der Mikroskope sowie die Bildverarbeitung wurde mit Hilfe der Software *Labview*⁶ realisiert.

⁴Molecular Devices, LLC, USA

⁵Noldus Information Technology

⁶National Instruments Corporation

In [143] werden Schlaf- und Aktivitätsphasen der Fische klassifiziert und sog. Verhaltensprofile erstellt. Die Profile werden für über 5000 Substanzen ermittelt. Das Beobachten und Tracken der Fische erfolgt mittels einer kommerziell erhältlichen Box der Firma *Viewpoint*⁷, welche wie in [67] auf ein Mikroskop verzichtet, lediglich Hellfeld-Aufnahmen durchführt und eine Tracking-Software nutzt. Das Verhaltensprofil der mit Chemikalien exponierten Fische wird mit dem der unbehandelten Fische verglichen. Abweichungen können mittels eines Farbcodes visualisiert und gruppiert werden. Mit der vorgestellten Methode lassen sich ähnlich wirkende Substanzen schnell zusammenfassen.

In [80] werden Fische mittels eines Stimulus zu einer Bewegung animiert. Während des Experiments wird die Bewegung der Fische durch ein automatisches Mikroskop der Firma Nikon und einer Kamera aufgezeichnet. Die Synchronisierung der Hardware sowie die Bildauswertung erfolgt mittels der Software *Metamorph*⁸. Es werden über 14000 Substanzen geprüft und vergleichbar zu [143] wird eine Gruppierung der Substanzen vorgenommen, die auf einem Vergleich der Auswirkungen der Substanzen basiert.

In [130] wird eine Bildanalyse bereits während der Bildakquise eingebunden. Dies hat den Vorteil, dass nicht relevante Strukturen oder Proben aussortiert werden können, wohingegen Bereiche im Fisch von Interesse in hoher Auflösung akquiriert werden. Die Mikroskopsteuerung und Bildverarbeitung erfolgt vollständig mittels einer skriptbasierten speziellen Kombination aus *XML*-Dateien⁹ und *Labview*-Skripten¹⁰.

In [93] werden Larven durch ein automatisches System vorbereitet und mit Nanopartikeln beaufschlagt. Aus einer Reihe von Merkmalen, die aus den Bildern extrahiert werden, lassen sich drei verschiedene Endpunkte automatisch identifizieren. Das System wird an ca. 1000 Embryonen demonstriert. Die Erfolgsquote liegt bei über 90% [23, 48, 64].

In [35] werden Larven automatisiert durch Injektion mit Tuberkulosebakterien infiziert. Es wird ein technisch realisierbarer Durchsatz von 2000 Larven pro Stunde angeführt. Das kommerzielle *COPAS* System wird zur Analyse der Daten verwendet.

In [174] findet ein kommerzielles hochspezialisiertes Auswertesystem für Mikrotiterplatten (Tecan Infinite M1000) Anwendung. Es wird ein theoretisch sehr hoher Durchsatz von über 50000 Einheiten pro Tag angegeben. Das System ist allerdings beschränkt auf Auswertungen in Fluoreszenzaufnahmen.

Die vorgestellten Hochdurchsatz-Untersuchungen am Zebrafisch sind bislang oftmals noch auf manuelle Schritte bei der Versuchsvorbereitung angewiesen. Es existieren jedoch bereits automatische robotergestützte Verfahren, deren Ziel es ist, eine vollständige Automatisierung der Vorbereitung der Untersuchungen zu ermöglichen. Um DNA, RNA oder Proteine in Zebrafischseiern einzubringen, wird beispielsweise die Mikroinjektion

⁷Viewpoint S.A. - Viewpoint Life Sciences, Inc.

⁸Molecular Devices, LLC, USA

⁹XML = Extensible Markup Language (engl. für „erweiterbare Auszeichnungssprache“)

¹⁰National Instruments Corporation

angewandt [182]. Der Vorbereitungsschritt wurde bereits mehrfach durch automatisierte Systeme gelöst [59, 98, 175, 181]. Ebenso existieren Lösungsansätze zur automatischen Bestückung der Mikrotiterplatten sowie der automatischen Entfernung des Chorions, welches die Zebrabärblingseier umgibt [102, 186].

Einige der Veröffentlichungen beschäftigen sich speziell mit der Analyse des Herzschlags im Zebrabärbling. Anders als die in Tabelle 1.2 aufgeführten Publikationen beschränken sich die Untersuchungen auf die methodische Durchführbarkeit etwa durch Extraktion des Signals der roten Blutkörperchen aus Hellfeld-Aufnahmen [14] oder die Detektion des Herzschlags ohne ein Markieren des Herzens bzw. der Zellen [15]. Sie erfordern zumeist ein exaktes Positionieren der Larven [36, 123]. Allen ist gemein, dass bisher keine systematische Analyse in größerem Umfang durchgeführt wurde, was meist auf die komplizierten manuellen Präparationsschritte zurückgeführt werden kann. Somit war ein Einordnen in die Tabelle 1.2 nicht möglich.

Es existieren eine ganze Reihe weiterer Ansätze, von denen hier nur auf eine Auswahl der wichtigsten eingegangen werden konnte. Dabei wurden vor allem die jüngeren Veröffentlichungen aufgeführt und solche, die eine hohe Stückzahl bzw. einen hohen Durchsatz erzielen.

1.2.4 Offene Probleme

Zusammenfassend ergibt sich aus den vorangegangenen Abschnitten die Erkenntnis, dass der Zebrabärbling nach übereinstimmender Meinung vieler Autoren erhebliches Potenzial für den Hochdurchsatz besitzt. Offensichtlich ist jedoch, dass aufgrund einer Reihe offener Probleme bisher nur ein geringer bis mittlerer Durchsatz erzielt werden konnte. Die wesentlichen technischen Probleme, weshalb Hochdurchsatz-Untersuchungen im Zebrabärbling bisher scheitern, sind:

1. *Entwurf der Hochdurchsatz-Prozesskette:*

Bisher existiert kein systematisches Entwurfskonzept einer Hochdurchsatz-Prozesskette. Alle publizierten Ansätze versuchen lediglich, manuelle Schritte teilweise oder ganz zu ersetzen. Die technischen, logistischen und informationstechnischen Voraussetzungen sind bei einer Anzahl von mehreren tausend Versuchen pro Tag jedoch komplex, daher bedarf es einer strukturierten Planung.

2. *Auswahl der geeigneten Art der Datenakquise:*

Die Wahl der Art der Datenakquise aller in der Literatur vorhandenen Lösungsansätze basiert im Wesentlichen auf der Erfahrung und dem Equipment, welches dem Entwickler zur Verfügung steht. Entscheidende Vor- und Nachteile der einzelnen Methoden werden nicht oder nur unzureichend berücksichtigt bzw. ausgeschöpft.

Dies führt zu ineffizienter Datenakquise, oft unzureichenden Daten oder zu Datenvolumina, die ohne entsprechendes Fachwissen nicht zu bewältigen sind.

3. *Entwicklung und Auswahl geeigneter Bildverarbeitungs- und Datenanalysealgorithmen:*

Ein systematischer Überblick über existierende Methoden und deren Anwendbarkeit für die biologischen Hochdurchsatz-Untersuchungen mit Zebrafärblingen existiert nicht. Ähnlich der Auswahl der Datenakquise-Methode werden auch die Bildverarbeitungs-Methoden basierend auf Vorlieben der Entwickler und dessen Kenntnisstand ausgewählt. Gleiches gilt für die gesamte Datenverarbeitung einschließlich Normierungs-, Klassifikations- und Visualisierungsstrategien.

4. *Es existiert keine systematische Abstimmung der Datenakquise und Datenverarbeitung:*

Die Bildverarbeitung muss oft auf unzureichend und nicht auf die Bedürfnisse der automatisierten Auswertung abgestimmte Bilddaten zurückgreifen. Die beiden zuvor genannten Punkte sind hochgradig voneinander abhängig, werden bei der Versuchsauslegung jedoch kaum aufeinander abgestimmt. Es existieren keine Untersuchungen bezüglich der Auswirkung der Auswahl bestimmter Akquise-Methoden auf die Auswertung und umgekehrt.

1.3 Ziele und Aufgaben

Hochdurchsatz-Untersuchungen stellen ein leistungsstarkes Mittel zum Erlangen neuer Erkenntnisse in den Wissensgebieten der Genetik, Toxikologie oder Pharmazie dar. Lösungen für Hochdurchsatz-Untersuchungen existieren bereits, beschränken sich jedoch entweder auf Versuche mit Zellen oder können, wenn sie einen gesamten Modellorganismus verwenden (wie z.B. den Zebrafärbling), den gewünschten hohen Durchsatz nicht erreichen. Ein System zu finden, welches die Vorteile der Hochdurchsatz-Untersuchung mit denen eines gesamten Modellorganismus verbindet, ist demnach von beachtlichem Wert. Ziel der vorliegenden Arbeit ist es daher, ein Konzept zu entwickeln, welches den systematischen Entwurf einer bildbasierten Hochdurchsatz-Untersuchung auf Basis des Zebrafärblings im Hinblick auf die Automatisierbarkeit ermöglicht. Hierfür sind die folgenden wissenschaftlichen Teilziele notwendig:

1. Herleitung eines Konzeptes

Ein Konzept zur Versuchsauslegung ist zu entwickeln. Dieses muss universell einsetzbar und an die jeweilige biologische Fragestellung und den gewünschten Durchsatz adaptierbar sein.

2. Ableitung von Auswertungskriterien

Zur Beurteilung der Durchführbarkeit und zum Entwurf einer Hochdurchsatz-Untersuchung sind Kriterien zu finden, die in der Lage sind, bekannte und neue Lösungsansätze zu analysieren und zu bewerten.

3. Entwicklung neuer Datenverarbeitungsverfahren

Neue Daten- und Bildverarbeitungsverfahren sind im Hinblick auf die Anforderungen der Hochdurchsatz-Untersuchung abzuleiten.

4. Anpassungsfähige Implementierung

Um die einfache und benutzerfreundliche Anwendbarkeit der Verfahren zu sichern, ist eine grafische Benutzeroberfläche zu entwickeln, die die erarbeiteten Lösungen in sich vereint und den direkten Zugriff und die Visualisierung der Ergebnisse ermöglicht.

5. Experimentelle Erprobung am Zebrabärbling

Im Hinblick auf die Besonderheiten, die sich bei der Verwendung des Zebrabärblings bei Hochdurchsatz-Untersuchungen ergeben, ist das entwickelte Verfahren zu konkretisieren. Insbesondere biologische, technische Assistenten müssen ohne Expertenwissen im Bereich der Datenverarbeitung in der Lage sein, Versuchsdaten einzulesen und auszuwerten.

6. Ableitung einer Aussage über die Leistungsfähigkeit

Zum Nachweis der Funktionalität sind Fallstudien durchzuführen. Die Ergebnisse geben Auskunft über den erreichbaren Durchsatz und die erreichbare Qualität des neu entwickelten Verfahrens.

In Kapitel 2 werden zunächst die Anforderungen an das zu entwickelnde Verfahren definiert und das neue Konzept zur Versuchsauslegung vorgestellt. Mit Hilfe des Konzeptes ist es dann möglich, die Versuchsauslegung zu konkretisieren und die Versuchsparameter anforderungsgemäß zu wählen. Die zur technischen Umsetzung des Konzeptes neu entwickelten Methoden werden anschließend in Form von Modulen in Kapitel 3 vorgestellt. Darauf aufbauend vereint Kapitel 4 die vorgestellten Methoden in einer grafischen Benutzeroberfläche, welche Zugriff auf alle wichtigen Parameter der Datenverarbeitung bietet. Anwendung findet das erarbeitete Verfahren in Kapitel 5 anhand konkreter biologischer Problemstellungen, für welche eine hochdurchsatzfähige Lösung vorgestellt wird. Eine Zusammenfassung der wesentlichen Ergebnisse der Arbeit sowie ein Ausblick über weitere mögliche Untersuchungen sind Gegenstand von Kapitel 6.

2 Neues Konzept zur Versuchsauswertung von Hochdurchsatz-Untersuchungen am Zebrabärbling

2.1 Übersicht

In diesem Kapitel wird erstmals ein zielgerichtetes, systematisches Konzept zur Versuchsauswertung bei Hochdurchsatz-Untersuchungen vorgestellt.

Ein solches Konzept kann unter Praxisbedingungen für Hochdurchsatz-Untersuchungen an Zebrabärblingen nur dann erfolgreich sein, wenn es die Schritte

- Definition der Versuchs- und Auswerteparameter und
- anforderungsgerechte Versuchsauslegung

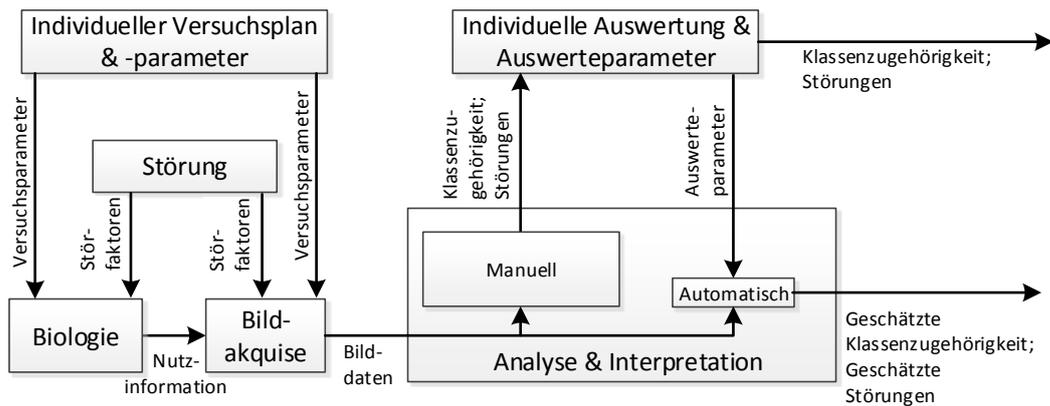
beinhaltet.

In Abschnitt 2.2 wird zunächst das Konzept beschrieben. Darauf aufbauend werden die notwendigen Anforderungen bezüglich der Durchführbarkeit, der Messbarkeit und der Auswertbarkeit von Hochdurchsatz-Untersuchungen zusammengetragen (Abschnitt 2.3). In Abschnitt 2.4 werden Versuchs- und Auswerteparameter identifiziert und daraufhin mathematisch beschrieben (Abschnitt 2.5). Auf den Ergebnissen aufbauend wird abschließend die Vorgehensweise zur anforderungsgerechten Versuchsauslegung vorgestellt (Abschnitt 2.6).

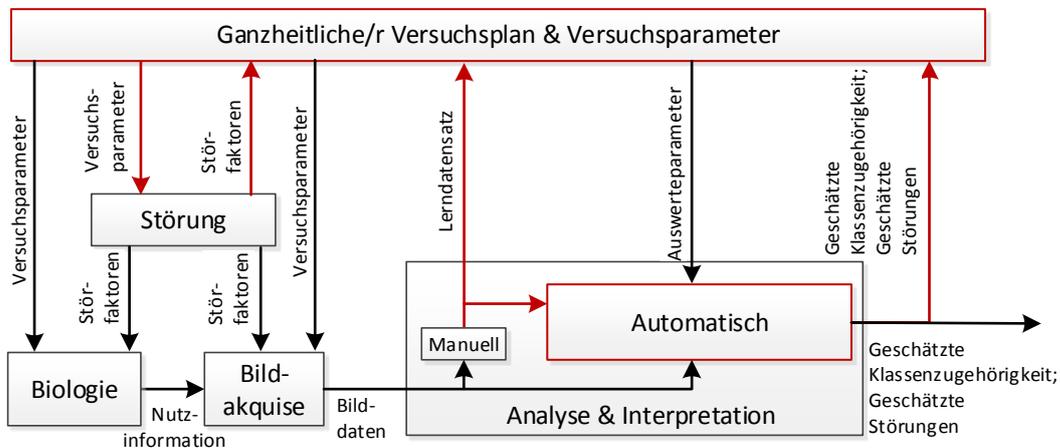
2.2 Neues Konzept zur Versuchsauswertung

Der Auslegung bildbasierter Hochdurchsatz-Untersuchungen kommt aufgrund der hohen Spezifizierung auf das Nutzsignal eine entscheidende Bedeutung für den Erfolg des Ver-

suchs zu. Durch die Biologie muss festgelegt werden, welches das Nutzsignal, d.h. die Information von Interesse, im zu untersuchenden Informationsträger ist. Das Nutzsignal (z.B. eine sich bewegende Zebraabärlingslarve) muss daraufhin mit geeigneten Mitteln (z.B. Vereinzelung der Larven) derart präpariert werden, dass die Bildakquise und deren Verfahren (z.B. Mikroskopie) das Nutzsignal in Bilddaten abbilden kann. Die Daten müssen schließlich innerhalb des Blockes *Analyse und Interpretation* nach neuen Erkenntnissen bezüglich der Fragestellung des Versuchs ausgewertet werden. Die Auswertung erfolgt manuell oder durch eigens entwickelte bzw. entsprechend adaptierte Auswertelgorithmen. Dabei wirken sowohl auf die Biologie als auch auf die Bildakquise Störungen, die Einfluss auf die objektive Realität oder deren Abbildung haben.



(a) Heutige Auslegung



(b) Neue Auslegung

Abbildung 2.1: Bisherige Vorgehensweise (heutige Auslegung) und Sollzustand (neue Auslegung) für die Versuchsauslegung bildbasierter Hochdurchsatz-Untersuchungen.

Die biologischen Schritte zur Bereitstellung der Nutzinformation berücksichtigen bisher lediglich, inwiefern die Information mittels Bildakquise erfasst werden kann, nicht ob sich diese zur (automatisierten) Auswertung eignet und ob Möglichkeiten zur Skalierung bestehen. Abb. 2.1a stellt die heutige Vorgehensweise bei der Auslegung dar. In Biologie und Bildakquise werden durch Versuchsparameter die Randbedingungen festgelegt. Auf den Versuch wirkt eine Störung in Form von Störfaktoren. In den biologischen Schritten

wird eine Nutzsignalinformation (kurz: Nutzinformation) bereitgestellt, die in den Bilddaten abgebildet und durch Analyse und Interpretation einer Klasse zugeordnet wird. Die Klassenzuordnung geschieht bisher manuell. In neueren Arbeiten wird nun versucht, den manuellen Schritt durch eine Nachahmung der manuellen Arbeit zu ersetzen. Hierfür sind Auswerteparameter notwendig. Das Ergebnis sind geschätzte Klassenzugehörigkeiten. Ebenso lassen sich die Störungen bzw. deren Klassenzugehörigkeiten schätzen. Der Anteil an automatisch ausgeführten Auswertungen ist klein, was durch den deutlich kleineren Block im Schaubild visuell verdeutlicht wird. Die Auswahl der Versuchsparameter wie z.B. der Akquise-Methode beruht meist auf einer individuellen, nicht systematischen Entscheidungsgrundlage wie „Verfügbarkeit der Mikroskope“ oder „persönliche Expertise des Laboranten“ etc. Die in der Literatur zu findenden Hochdurchsatz-Untersuchungen führen die Versuchsauslegung oft willkürlich, aus Erfahrungswerten oder anhand der vorhandenen Mikroskop- und Rechnerinfrastruktur durch [127, 147, 168]. Ein methodisches und ganzheitliches Vorgehen ist jedoch unumgänglich, um neuartige drei-, vier- oder auch n-dimensionale (durch Berücksichtigung von Veränderungen im Raum und über die Zeit) Datensätze effizient auswerten zu können. Des Weiteren basieren die Ergebnisse aller sich anschließenden Analyse- und Interpretationsschritte auf den derart akquirierten Bilddaten. Die Qualität der Bilddaten limitiert somit die erreichbare Qualität des Versuchs. Die Auswertung erfolgt bisher in weiten Teilen manuell oder mittels (semi)automatischer Auswerteroutinen (vgl. Literaturübersicht in Tabelle 1.2). Die weitgehend individuelle und ungeordnete Versuchsauslegungsphilosophie hat den Nachteil, dass nachfolgende Verarbeitungsschritte nicht in die Entscheidungsfindung zur Parameterwahl mit einbezogen werden. Dies ist auch ein Grund, weshalb sich zur Zeit nur ein geringer Anteil der Analyse- und Interpretationsschritte automatisiert durchführen lässt. Deshalb führt die vorliegende Arbeit eine Analyse der Parameter durch, um eine Versuchsauslegung anforderungsgerecht durchführen zu können.

Das hier vorgeschlagene neuartige Konzept zur Versuchsauslegung der Hochdurchsatz-Untersuchung bildet gemäß Abb. 2.1b ebenfalls die Nutzinformation mittels Bilddaten ab, jedoch werden alle Entscheidungen, Versuchs- sowie Auswerteparameter zielgerichtet auf die spätere automatisierte Auswertung hin getroffen. Der Prozess der Versuchsauslegung wird ganzheitlich vorgenommen und der Einfluss der jeweils getroffenen Entscheidung nicht nur auf den sich direkt anschließenden Schritt, sondern auf alle Blöcke betrachtet. Die ganzheitliche Betrachtung schließt die Auswahl der Akquise-Methode sowie die Identifikation möglichst vieler Störgrößen ein. Auf Basis der gesammelten Information werden Versuchsparameter gewählt. Des Weiteren wird der Anteil der automatisch vollzogenen Analyse und Interpretation stark erhöht, was durch den großen Block verdeutlicht ist. Der manuelle Anteil ist deutlich kleiner und dient im optimalen Fall lediglich dazu, eine Wissensbasis für die automatische Auswertung zu schaffen. Der ausgewertete oder gelabelte (von engl. *to label* = markieren, beschriften) Datensatz der Wissensbasis wird verwendet, einen Klassifikator anzulernen, mit dessen Hilfe die Klassenzugehörigkeit unbekannter Daten geschätzt wird. Selbst einer vollständigen Automatisierung der Hochdurchsatz-Untersuchung muss indes zwingend eine manuelle Deutung der Ergebnisse folgen, da das Ziel, bzw. die Art der Treffer, welche durch die Untersuchung aufgezeigt werden sollen, im Vorfeld nicht oder nur unzureichend bekannt sind. Eine Ausnahme bilden selten durchgeführte

Hochdurchsatz-Untersuchungen, deren Ziel es ist, eine Hypothese zu prüfen.

Da der bisherige Entwurf der Hochdurchsatz-Untersuchung im Wesentlichen der Versuch ist, die manuellen Schritte der Bildauswertung automatisch nachzubilden, müssen nun für das neue Konzept die Ansprüche aus Biologie und Automatisierungstechnik aufeinander abgestimmt werden. Für diese interdisziplinäre Aufgabe bedarf es einer Schnittstelle, mit deren Hilfe eine erfolgreiche Versuchsauslegung erfolgen kann. Die Schnittstelle sind die Einflussgrößen auf die übergeordneten Blöcke „Biologie“ und „Bildakquise“ sowie „Analyse und Interpretation“. Die entsprechenden Parameter sind demnach die zu bestimmenden Größen bei der Versuchsauslegung. Zwar muss die generelle Struktur der Hochdurchsatz-Untersuchung individuell festgelegt werden, denn die in den Grundlagen der vorliegenden Arbeit bereits aufgezeigte große Bandbreite an möglichen biologischen, toxikologischen und genetischen Untersuchungen schließt eine universell anwendbare Lösung aus. Dennoch lässt sich jede individuell generierte Struktur anhand der Versuchsparameter und Auswerteparameter auslegen und optimieren.

Versuchsparameter fassen hierbei alle Einflussgrößen aus Biologie und Bildakquise zusammen, während Auswerteparameter alle Einflussgrößen beschreiben, die bei der Analyse und Interpretation festzulegen sind. Die Literatur zeigt, dass eine gewissenhafte Auslegung der Parameter den Erfolg der Untersuchung zwar nicht garantiert, wohl aber wahrscheinlich macht und somit notwendige Voraussetzung für das Gelingen der Hochdurchsatz-Untersuchung ist [50,127,155]. Eine mangelhaft ausgeführte Planung kann zum Scheitern der Untersuchung führen oder dazu zwingen, den gesamten Versuch wiederholen zu müssen. Da die Durchführung der Untersuchung, je nach Anzahl und Aufwand der Einzelversuche, bis zu mehreren Monaten dauert, ist es sinnvoll, jeden Schritt sorgfältig in Vorversuchen zu prüfen und zu optimieren.

Die Auswirkung einer Anpassung einzelner Parameter zeigt sich meistens erst nach dem Vollzug der vollständigen Prozesskette, daher ist es notwendig, die Auswahl der Parameter iterativ vorzunehmen und in Vorversuchen die gesamte Kette mehrfach zu durchlaufen. Eine Änderung an einer beliebigen Stelle hat Auswirkungen auf viele, wenn nicht alle Teilprozesse. So kann beispielsweise eine geänderte Behandlung der Fischeier oder die Verwendung einer Mutation der Fischeier (z.B. mit einem fluoreszierenden Marker) Änderungen bei der Wahl der Mikroskopie- bzw. Akquise-Techniken erfordern. Eine solche Änderung führt wiederum zu einer Änderung der Auswerteparameter der Analyse, deren Ergebnisse entsprechend in Interpretation und Evaluation neu dargestellt werden müssen.

Zu Beginn wird die Struktur der Hochdurchsatz-Untersuchung geplant, auf deren Basis Vorversuche durchgeführt werden. Ziel ist es, schrittweise eine robuste Zuordnung des Nutzsignals zu einer Klassen- oder Trefferzuweisung zu erhalten, die Anzahl der Einzelversuche zu höheren Werten hin zu skalieren und gleichzeitig den Aufwand pro Einzelversuch so gering wie möglich zu halten. Nach vollzogener Versuchsauslegung soll in einem Versuchsplan feststehen, welche Versuche mit welchen Behandlungen durchzuführen sind, mit welcher Methode die Nutzinformation akquiriert wird und auf welche Art die ak-

quirierten Daten verarbeitet und ausgewertet werden. Das hier vorgestellte Konzept zur Versuchsauslegung beschränkt sich allerdings auf die *technische* Realisierbarkeit von Hochdurchsatz-Untersuchungen und deren Optimierung. Für weitere Details zur *statistischen* Versuchsplanung, der Erstellung einer Rahmenvorschrift zur praktischen Durchführung der Einzelversuche und der Minimierung und Optimierung der Anzahl der Versuche sei auf die Fachliteratur zur statistischen Versuchsplanung verwiesen [77, 116, 138, 146].

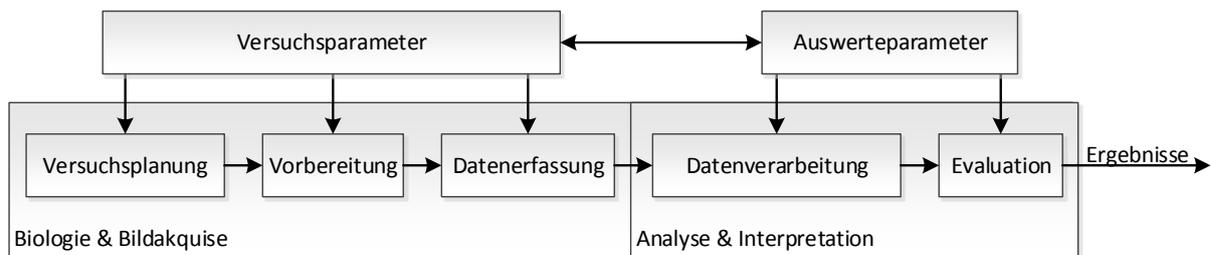


Abbildung 2.2: Neues Konzept zur Optimierung der Versuchsauslegung durch Anpassung der Parameter der Prozesskette.

Eine bildbasierte Hochdurchsatz-Untersuchung besteht üblicherweise aus zwei aufeinander folgenden Teilschritten (vgl. Abb. 2.2): Der erste Teil, die „Biologie & Bildakquise“, setzt sich zusammen aus einer „Versuchsplanung“, gefolgt von der „Vorbereitung“ der Proben und der „Datenerfassung“. Der zweite Teil, die „Analyse & Interpretation“, besteht aus der „Datenverarbeitung“ sowie deren „Evaluation“. Das vorgestellte Konzept ermittelt die Parameter aller in Abb. 2.2 dargestellten Blöcke. Der Informations- bzw. Datenfluss verläuft von links nach rechts, während die Parameter auf alle Blöcke Einfluss nehmen und daher übergeordnet sind. Wichtig hierbei ist die Einflussnahme und Berücksichtigung der Auswerte- und Versuchsparameter aufeinander, was durch den Doppelpfeil zwischen den Parameterblöcken visuell verdeutlicht wird.

Die Versuchsplanung im Block „Biologie & Bildakquise“ steht für die statistische Versuchsplanung. Das Vorgehen bei der statistischen Versuchsplanung sollte sich an der entsprechenden Fachliteratur orientieren und die Prinzipien der Wiederholung, Randomisierung und Blockbildung berücksichtigen [11, 77, 116, 138, 146]. Ergebnis des Blockes „Versuchsplanung“ ist ein statistisch optimaler Versuchsplan, welcher die Anzahl der Einzelversuche und deren jeweilige Behandlung festlegt und optimiert. Für das vorgestellte Konzept sind die innerhalb der statistischen Versuchsplanung identifizierten Störfaktoren von Bedeutung. Der Block „Vorbereitung“ bezieht sich auf alle Präparationen, die dazu beitragen, eine maximale Qualität der Datenakquise zu erzielen. Bei biologischen Versuchen besteht oftmals die Möglichkeit, eine Versuchseinheit derart vorzubereiten, dass die Nutzinformation besser erkennbar wird. Ein Beispiel ist die Betäubung und Ausrichtung der Zebrafärbingslarven. Unterschiedlich aufwändige Präparationsschritte bieten so die Möglichkeit, die Bildauswertung zu vereinfachen. In manchen Versuchen ist das Ausrichten sogar zwingend erforderlich, um das Nutzsignal erfassen zu können. Werden die Larven bei der Präparation z.B. durch speziell geformte Platten immer lateral ausgerichtet, können Bereiche im Zebrafärbing leichter anhand von geometrischen Beziehungen

gefunden werden [49]. Versuchsparameter sind somit alle gewählten Vorbereitungsschritte und Störfaktoren, z.B. die gewählte Position der Versuchseinheit in der Mikrotiterplatte.

Die Datenerfassung zeichnet mittels der gewählten Akquise-Methode einen Bildstrom auf. Je detaillierter der aufgenommene Bildstrom ist, desto mehr (Nutz)Informationen enthält er. Allerdings steigt die Größe des Datensatzes. Der Anteil an „unnötigen“, d.h. Hintergrundinformationen, wächst meist überproportional, was die Gefahr birgt, dass die Daten allein aufgrund der Datenmenge schwer zu verarbeiten und zu archivieren sind. Auch der Aufwand für die Extraktion des Nutzsignals steigt. Vor allem bei neuartigen mehrdimensionalen Mikroskopietechniken wie etwa der SPIM-Mikroskopie (vgl. Abschnitt 1.2.1) fallen immens große Datenmengen von mehreren Gigabyte (GB) oder sogar Terabyte (TB) pro Versuchseinheit an. Mit Abschluss der Datenerfassung ist der übergeordnete Block „Biologie & Datenakquise“ abgeschlossen und alle Versuchsparameter sind bestimmt. In Form des akquirierten Bildstroms steht nun eine Messung des Nutzsignals zur Verfügung. Dem Nutzsignal jeder Versuchseinheit einer Klasse bzw. einer Trefferfunktion zuzuordnen, ist Aufgabe des übergeordneten Blocks „Analyse & Interpretation“, welcher sich wiederum weiter in die Schritte „Datenverarbeitung“ und „Evaluation“ unterteilen lässt.

Die Datenverarbeitung trennt das Nutzsignal von unwichtigen Informationen, d.h. von der Hintergrundinformation. Sie extrahiert Merkmale, welche schließlich der Klassifikation zugeführt werden. Der Prozess schließt die Bildverarbeitung mit ein und wird in der vorliegenden Arbeit detailliert betrachtet (vgl. Kapitel 3). Das Ergebnis der Klassen- oder Trefferzuordnung wird in möglichst übersichtlicher Form dargestellt, sodass biologische Schlüsse gezogen werden können. Dies geschieht im Block „Evaluation“. Aus der automatischen Klassifikation muss der Biologe manuell eine Schlussfolgerung ziehen. Hierbei ist eine übersichtliche Präsentation der Versuchsergebnisse hilfreich. Die anhand der Präsentation getroffene Deutung des Biologen soll die Fragestellung der Hochdurchsatz-Untersuchung beantworten. In den Vorversuchen muss nach der ersten Evaluation anhand der vorläufigen Ergebnisse abgewogen werden, ob die Qualität der Ergebnisse ausreichend ist, um den Versuch auf die finale Größe zu skalieren, d.h. ob mit dem gewählten Verfahren die gewünschten Daten aussagekräftig abgebildet werden können und ob das gesuchte Nutzsignal in den Daten stark genug vorhanden ist. Ist dies der Fall, wird mit der tatsächlichen Durchführung des vollständigen Versuchs begonnen.

Mit dem dargestellten Konzept steht nun eine einheitliche Vorgehensweise und Schnittstelle für die Versuchsauswertung zur Verfügung.

2.3 Anforderungen an bildbasierte Hochdurchsatz-Untersuchungen

Die Hochdurchsatz-Untersuchungen müssen vielschichtigen Anforderungen gerecht werden. Die Anforderungen werden hier in strukturierter Form herausgearbeitet. Anforder-

rungen sind je nach Versuchsvorgaben quantitativ (z.B. Anzahl der zu untersuchenden Substanzen) oder qualitativ (z.B. Sichtbarkeit des biologischen Effekts). Während für quantitative Anforderungen klare Angaben gemacht werden können, lassen sich für qualitative Anforderungen keine genau messbaren Grenzwerte angeben, da z.B. stets ein möglichst gut sichtbarer biologischer Effekt oder eine geringe Fehlerrate anzustreben sind.

Für Hochdurchsatz-Untersuchungen lassen sich Ziele und Vorgaben allgemein bzw. in Abhängigkeit des konkreten biologischen Versuchs formulieren. Der Versuch spielt hierbei für die Anzahl bzw. Komplexität der Anforderungen die maßgebende Rolle, daher sind nicht immer alle Anforderungen für den Einzelfall unbedingt erforderlich. Um den erreichbaren Durchsatz sicherzustellen, ist die Automatisierbarkeit aller notwendigen Versuchsschritte übergeordnet. Zusätzlich wurden in der vorliegenden Arbeit folgende Anforderungen für Hochdurchsatz-Untersuchungen identifiziert:

1. Durchführung

- Legitimität

Die Untersuchung muss den gesetzlichen Vorschriften sowie den ethischen Prinzipien der jeweiligen Forschungseinrichtung entsprechen. Es sind beispielsweise Vorgaben zur Genmanipulation, dem Umgang mit toxikologischen Substanzen und Tierversuchen zu beachten.

- Logistische Machbarkeit

Da es sich bei den Versuchen mit Modellorganismen um sich entwickelnde und damit veränderliche Proben handelt, müssen vergleichbare Daten auch zu vergleichbaren Entwicklungs-Zeitpunkten aufgenommen werden. Die Forderung ist von logistischer und biologischer Bedeutung, da für einen gewünschten Durchsatz der entsprechende Nachschub an Fischeiern gegeben sein muss. Die meisten Arbeitsschritte müssen weitgehend parallel erfolgen. Da die Eier jedoch immer zu einem bestimmten Zeitpunkt, dem Sonnenaufgang, gelegt werden und somit die Entwicklung immer gleich startet, ist das Zeitfenster für die Bildakquise vergleichbarer Daten an einem Tag klein.

- Biologische Realisierbarkeit

Die zu untersuchende Fragestellung muss sich am Zebrabärbling bzw. dem gewählten Modellorganismus durchführen lassen.

- Reproduzierbarkeit

Ein erneutes Durchführen identischer Versuche muss zum qualitativ gleichen Ergebnis führen.

- Finanzielle Realisierbarkeit

Durch die hohe Anzahl an Einzeluntersuchungen entstehen hohe laufende Kosten für die Verbrauchsmaterialien und für die zu untersuchenden Substanzen. Hinzu kommen Kosten für z.T. hochauflösende Mikroskope, robotergestützte Automatisierungstechniken und Computertechnik zur Bildverarbeitung sowie Archivierung. In der Summe übersteigen solche Kosten schnell das Budget kleinerer Forschungseinrichtungen.

2. Messung

- Eindeutige Präsenz der biologisch relevanten Information

Die grundlegende Forderung der eindeutigen Präsenz ist bei vielen Versuchen nicht erfüllt. So können Biologen oftmals aus Erfahrung und anhand von kleinen Hinweisen Dinge (etwa biologische Gewebe, Phänotypen etc.) in Organismen erkennen, welche für eine weniger geübte Person und auch für die automatische Datenverarbeitung in den Daten nicht zu identifizieren sind. In einem solchen Fall ist mit automatischen Verfahren keine robuste Auswertung möglich.

- Rückwirkungsfreiheit der Messung

Die Akquise darf die Ergebnisse nicht wesentlich beeinflussen. Manche Akquise-Techniken erfordern beispielsweise sehr lichtstarke Beleuchtungen oder bei der konfokalen Mikroskopie sogar Laser zur Akquise der Daten. Biologische Proben können z.B. bei Verhaltensuntersuchungen auf Licht reagieren oder fluoreszierende Marker können durch starke Beleuchtung ausgebleichen werden. Daher muss vor Beginn der Untersuchung der Einfluss der Akquise-Methode auf das Nutzsignal berücksichtigt werden.

- Schnelligkeit der Messung

Die Akquise muss in adäquater Zeit erfolgen. Da aus den o.g. logistischen Einschränkungen der Durchsatz deutlich unter der theoretischen Kapazität automatischer Mikroskope liegt, muss geprüft werden, ob die Fragestellung auch in der gewünschten Zeit realisierbar ist. Die Realisierbarkeit ist abhängig von den zur Verfügung stehenden menschlichen und technischen Ressourcen.

3. Auswertung

- Segmentierbarkeit des biologischen Effekts im Bild

Der biologische Effekt muss nach der Messung klar in den aufgezeichneten Bildern oder dem aufgezeichneten Bildstrom vorhanden sein, sodass eine Segmen-

tionierung, d.h. Abgrenzung des biologischen Effekts von anderen Bildinhalten, möglich ist.

- Quantifizierbarkeit des biologischen Effekts im Bild

Der biologische Effekt kann beispielsweise ein Phänotyp sein, welcher nur selten unter Tausenden von Untersuchungen vorkommt, oder eine Statistik, die aus einer hohen Stückzahl an Bildern extrahiert wird. Das Nutzsignal muss sich durch Zahlenwerte, die aus den Bildern ermittelt werden, repräsentieren lassen, d.h. die Bildauswertung muss die relevanten Informationen extrahieren und die Merkmalsextraktion der Bildauswertung muss so konzipiert sein, dass sich das biologische Nutzsignal aus den Merkmalen klassifizieren lässt. Auch in der großen Anzahl an Einzeluntersuchungen müssen selten vorkommende Ereignisse erkennbar sein und auch erkannt werden.

- Niedrige Fehlerrate bei der Detektion des biologischen Effekts

Die Bildauswertung muss robust sein, d.h. Fehlklassifikationen durch Bildfehler aus Akquise, Präparation, Robotik etc. sollen ebenso wie ein kritischer Abbruch der Bildverarbeitungsroutine vermieden werden.

- Schnelligkeit der Auswertung

Die Bildauswertung muss in adäquater Zeit erfolgen, da je nach biologischer Fragestellung u.U. Echtzeitfähigkeit gefordert werden muss. Zumeist ist jedoch eine Rechenzeit von einigen Minuten pro Einzelexperiment vertretbar, solange das Gesamtexperiment in einem angemessenen Zeitraum (üblicherweise zwischen einigen Stunden und Wochen) auswertbar ist.

- Präsentierbarkeit der Auswertung

Die Ergebnisse müssen nach der Durchführung und Auswertung der Hochdurchsatz-Untersuchung dargestellt werden. Die hohe Anzahl von Einzeluntersuchungen macht eine übersichtliche Art und Weise der Darstellung notwendig. Die Klassifikationsergebnisse und die ggf. identifizierten *Treffer* (selten vorkommende Ereignisse, auf welche die Hochdurchsatz-Untersuchung ausgerichtet sein kann) müssen visualisiert werden und den Versuch in übersichtlicher Weise darstellen, sodass von Biologen weitere Schlüsse gezogen werden können.

- Wissenschaftliche Archivierung

Um die Reproduzierbarkeit zu sichern, müssen alle Versuche inkl. Versuchsparmeter, akquirierten Bildern, verwendeter Software, deren Version und der Visualisierung archiviert werden. Die Archivierung schließt sowohl die Bildverarbeitungsroutine als auch den Zugriff auf alle Auswertungs- und Visualisie-

rungsroutinen ein. Bei mehrdimensionalen Untersuchungen im Hochdurchsatz fallen sehr hohe Datenmengen an. Die Daten müssen bei den meisten biologischen Untersuchungen für mindestens 10 Jahre archiviert werden [166]. Die zur Verfügung stehende Archivierungseinrichtung muss daher der Datenmenge und Dauer der Archivierung entsprechend ausgelegt sein.

2.4 Identifikation der Versuchs- und Auswerteparameter

Zur konkreten Anwendung des im vorangegangenen Abschnitt eingeführten Konzeptes müssen in allen Prozessschritten (vgl. Abb. 2.2) Entscheidungen getroffen werden, durch die Versuchs- und Auswerteparameter bestimmt werden. Die Versuchsparameter bilden zusammen mit der Anweisung zur praktischen Durchführung aus der statistischen Versuchsplanung den Versuchsplan der Hochdurchsatz-Untersuchung. Der folgende Abschnitt leistet einen Beitrag zur Entscheidungsfindung für die technische Durchführung und zeigt wichtige Eigenschaften und gegenseitige Abhängigkeiten zwischen den einzelnen Versuchsparametern und Entscheidungen auf. Die gewählten Optionen haben unmittelbare Auswirkungen auf die Aussagen und den Durchsatz, der mittels einer Hochdurchsatz-Untersuchung erzielt werden kann. Zum Teil stehen für ein beabsichtigtes Ergebnis bzw. eine bestimmte biologische Fragestellung und somit ein gewünschtes Nutzsignal eine Reihe an Entscheidungen aus Gründen der Realisierbarkeit bereits fest. Als einführendes Beispiel sei ein Zelltracking einer einzelnen Zelle angeführt. Eine solche Aufgabenstellung ist in einer Hellfeld-Aufnahme schwer realisierbar, da die Nutzsignalinformation (die Zelle von Interesse) schwer von benachbarten Zellen zu unterscheiden ist. Hier bietet sich das Markieren der Zellkerne durch fluoreszierende Proteine an, da sich diese, bei Anregung mit der entsprechenden Wellenlänge separat, ohne Auftreten der nicht markierten Zellen, aufzeichnen lassen. Ein auf morphologischen Formen basierendes Nutzsignal ist dagegen nur mit Hellfeld-Aufnahmen abbildbar. Die beiden Beispiele zeigen die Notwendigkeit auf, die Akquise-Methode so zu wählen, dass das Nutzsignal abgebildet wird. Im Folgenden werden Vor- und Nachteile der Entscheidungsmöglichkeiten diskutiert und anhand von Beispielen die Folgen verdeutlicht. Manche der Entscheidungen schließen sich im Übrigen nicht aus, sondern es können auch zwei oder mehrere Entscheidungen gleichzeitig getroffen werden (im vorigen Beispiel das gleichzeitige Aufzeichnen von Hellfeld- und Fluoreszenzaufnahmen und die spätere Zuordnung zu biologischen Signalen). Ein solches Vorgehen ist für biologische Fragestellungen zum Teil sogar unmittelbar erforderlich und hat erheblichen Einfluss auf den für die Bildakquise bzw. Bildauswertung notwendigen Aufwand und die zu akquirierende Datenmenge.

In Abb. 2.3 ist eine Übersicht der wichtigsten Entscheidungsmöglichkeiten zur Bestimmung der Versuchsparameter dargestellt. Die Übersicht ist nach dem Schema des Konzeptes aus Abschnitt 2.2 gegliedert und die bereits eingeführten, übergeordneten Blöcke „Biologie & Bildakquise“ sowie „Analyse & Interpretation“ lassen sich abgrenzen.

2.4.1 Biologie und Bildakquise

Im Biologie- und Bildakquiseteil müssen die Versuchsparameter bestimmt werden. Für eine gute Qualität bei der Datenerfassung bietet es sich an, während der Versuchsvorbereitung die Larven zu präparieren. Der Aufwand für das Präparieren fällt unterschiedlich hoch aus, ist jedoch in fast allen Fällen manuell durchzuführen, limitiert somit den Durchsatz und setzt ein entsprechendes Personal voraus. Die wichtigsten Präparationsschritte

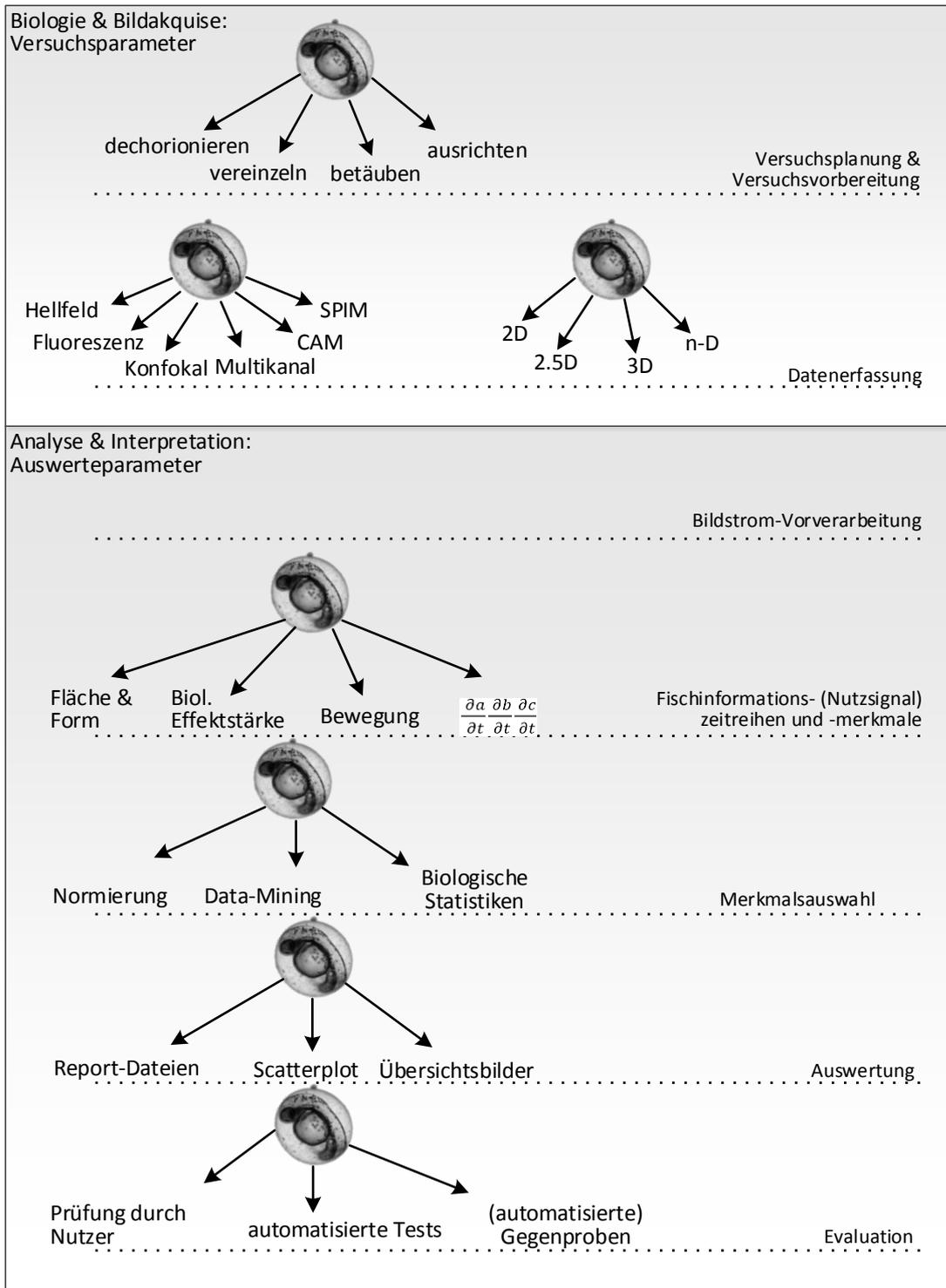


Abbildung 2.3: Entscheidungsmöglichkeiten und Ablauf einer bildbasierten Hochdurchsatz-Untersuchung

sind das Dechorionieren, das Vereinzeln, das Ausrichten und das Betäuben der Larven. Das Dechorionieren ist das Befreien der Larven von der umschließenden Fruchthülle (dem sog. Chorion). Dabei wird die Fruchthülle entweder mit einer Pinzette manuell entfernt oder mittels einer Chemikalie aufgelöst. Bei Verwendung der Chemikalie lassen sich viele Larven gleichzeitig dechorionieren. Abb. 2.4 stellt einer Larve im Chorion eine dechorionierte Larve gegenüber.

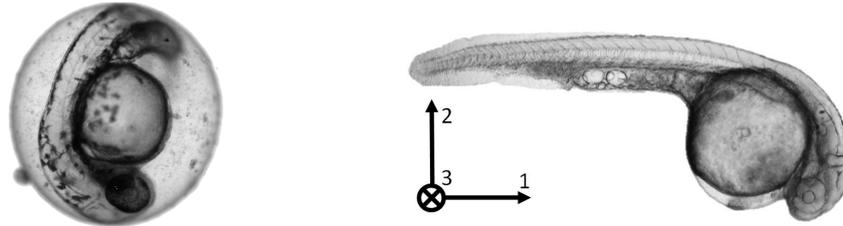


Abbildung 2.4: Larve eines Zebrafischblings im Chorion (links) und nach dem Dechorionieren (rechts), d.h. dem Entfernen der Fruchthülle

Der Vorteil der Präparation ist, dass sich die Larven nach dem Dechorionieren ausstrecken und hierdurch alle seitlich sichtbaren Bereiche akquirierbar sind. Bis zu einem Alter von 72 hpf¹ liegt die Larve zudem definiert entweder in lateraler oder dorsaler Lage². Durch die runde Form des Dottersacks kann die Larve allerdings leicht in eine beliebige Richtung gekippt sein, was zur Folge hat, dass die absolute Höhe von Bereichen im Fisch variiert. Da es technisch nicht möglich ist, die gesamte Larve, auch in der Tiefe, mit einer Aufnahme scharf abzubilden, müssen auch dechorionierte Larven in mehreren Fokusebenen akquiriert werden. Für die Bildauswertung ist die definierte seitliche Position allerdings von großem Vorteil, da sich ein Koordinatensystem definieren lässt und geometrische Beziehungen zum Auffinden von Bereichen im Fisch ableitbar sind [49], d.h. die Larven müssen lediglich um die dritte Achse rotiert und ggf. gespiegelt werden, um eine vergleichbare Position zu erhalten (vgl. Abb. 2.4). Den genannten Vorteilen steht der große zeitliche Aufwand für die Entfernung des Chorions und für das Ausrichten auf der Seite gegenüber. Dies gilt im Besonderen, wenn auf die Dechorionierung mittels Chemikalien aufgrund von möglichen wechselseitigen Abhängigkeiten zwischen den Substanzen, die bei dem Versuch zum Einsatz kommen, verzichtet und somit das Dechorionieren für jede Larve einzeln und manuell durchgeführt werden muss. Zudem ist die Vereinzeln von dechorionierten Zebrafischblings schwieriger, da die Larven bereits „stromlinienförmig“ sind und das exakte Pipettieren und vor allem dessen Automatisierung vor Probleme stellt.

Die Vorteile der Zebrafischlarven im Chorion sind u.a. die leichtere Automatisierbarkeit sowie die kurze Präparationszeit. Bis eine robuste robotergestützte Lösung für die o.g. Präparation entwickelt ist, kann für Hochdurchsatz-Untersuchungen lediglich die Larve im Chorion empfohlen werden, denn bei manueller Präparation lässt sich der geforderte Durchsatz von mehreren tausend Einzelversuchen pro Tag nicht erreichen, was für die Bildverarbeitung jedoch den Nachteil, dass die Larven im Chorion frei beweglich sind. Somit ist deren Lage auf den akquirierten Bildern äußerst unterschiedlich. Als

¹hpf = hours past fertilization = Stunden nach der Befruchtung

²Zu einem späteren Zeitpunkt richtet sich die Larve auf, da sich die Schwimmblase entfaltet hat [177].

Vorteil der Zebra**ä**rblingslarve als Modellorganismus wird zwar vor allem Transparenz angeführt, was jedoch nicht bedeutet, dass alle Seiten der Larve gleich gut auf dem Bild sichtbar sind. Selbst kontraststarke Bereiche wie das Auge sind, wenn sie verdeckt auftreten, zwar durch die Larve hindurch noch zu erahnen, die Konturen werden von den darüber liegenden Körperteilen (z.B. Kopf oder Dottersack) jedoch stark verwischt oder abgeschwächt. Je nach Lage der Larve sind somit verschiedene Details im Zebra**ä**rbling mit Chorion nicht oder nur schwer zu erkennen. Damit lassen sich hier lediglich von der Lage unabhängige Nutzsignale untersuchen.

Die letzte der hier aufgezeigten Präparationsmöglichkeiten ist das Betäuben der Zebra**ä**rblingslarven. Ab ca. 24 hpf beginnen die Larven, sich spontan in ihrer Eihülle zu bewegen. Die Häufigkeit dieser spontanen Bewegungen schwankt während der Entwicklung der Larve. Im Mittel kann jedoch alle drei Sekunden eine Bewegung beobachtet werden, was zur Folge hat, dass, wenn die Bewegung während der Bildakquise auftritt, verwischene Konturen in der Mikroskopaufnahme erscheinen. Solche Aufnahmen sind meist für die weitere Auswertung unbrauchbar. Bei Aufnahmetechniken wie etwa der Konfokal-Mikroskopie oder SPIM sind lange Belichtungszeiten pro Aufnahme ohne Bewegung erforderlich. Soll über eine längere Zeit eine Zuordnung von Bereichen in der Larve möglich sein, müssen solche Bewegungen unterbunden werden. Dies kann durch ein Betäubungsmittel, z.B. Trika**ä**ne, erreicht werden, hat allerdings einen manuellen Präparationsschritt mit o.g. Nachteilen zur Folge. Auch kann ein Zusammenwirken des Anästhetikums mit den Substanzen, deren Einfluss bestimmt werden soll, auftreten. Daher muss bei der Versuchsauslegung geprüft werden, ob die Verwendung von Betäubungsmitteln vertretbar ist. Eine alternative Möglichkeit zum Demobilisieren der Larven ist es, die Larven kurz vor der Bildakquise mit Hilfe von kaltem Wasser zu kühlen [185].

Nachdem nach Abb. 2.3 eventuelle manuelle Präparationsschritte gewählt wurden und die Larven in definierter Anzahl (üblicherweise 1-10) in Mikrotiterplatten platziert und ggf. mit Chemikalien exponiert sind, muss nun für das Nutzsignal eine geeignete Akquise-Technik gewählt werden (vgl. Abschnitt 1.2.2). Für jede Zebra**ä**rblingslarve wird nun eine Messung mittels Detektoren erfolgen. Unabhängig von den gewählten Akquise-Techniken wird bei den aufgezeichneten Daten immer von einem Bildstrom gesprochen. Je nach Protokoll der Aufzeichnung handelt es sich bei den Daten um ein einzelnes Bild, eine Bildsequenz, oder mehrere Bildsequenzen innerhalb einer längeren Zeitspanne. Es müssen also, ebenfalls unabhängig von der gewählten Akquise-Technik und angepasst an das Nutzsignal, die Aufnahmezeitpunkte und deren Frequenz festgelegt werden. Dafür muss geklärt werden, ob die Nutzsignalinformation nur selten auftritt, unter bestimmten Voraussetzungen erkennbar ist oder durch geeignete Mittel provoziert werden kann. Dies hat direkten Einfluss auf Anzahl der Wiederholungen sowie Dauer und Frequenz der Akquise pro Einzelversuch. Im einfachsten Fall ist der Zeitpunkt der Akquise bzw. das Zeitfenster für vergleichbare Ergebnisse groß (mehrere Stunden bis Tage) und der Effekt immer sichtbar. Ist das nicht der Fall, z.B. wenn ein Entwicklungsvorgang der Larven oder von Zellen untersucht wird, so sind Bildsequenzen vonnöten, deren Abtastraten an die Entwicklungsgeschwindigkeit anzupassen sind. Die steigende Akquise-Zeit der Einzelversuche hat Leerlaufzeiten bzw. Wartezeiten der Mikroskope zur Folge. Solche Zeiten des Still-

standes schränken die erreichbare Anzahl an Einzeluntersuchungen pro Tag ein, da die Mikroskope länger durch die Bildakquise je Einzelversuch blockiert werden und durch die geforderte Vergleichbarkeit der Einzelversuche und die schnelle Entwicklung der Larven nur ein kleines Zeitfenster zur Akquise existiert. Die Auftrittshäufigkeit hat daher überproportionalen Einfluss auf den erreichbaren Durchsatz.

Bei der Datenerfassung hat die Wahl der Akquise-Technik, wie z.B. die Präparation, unterschiedliche, weitreichende Auswirkungen auf die Bildauswertung. Entsprechend Abb. 2.3 bieten sich verschiedene Optionen, welche die Modalität des Bildstroms bestimmen. Die Akquise-Methode ist dem Auftreten des Nutzsignals anzupassen. Es muss in jedem Fall sichergestellt werden, dass der Bildstrom das Nutzsignal und damit die Nutzsignalinformation abbildet.

Ist die auszuwertende Beobachtung beispielsweise in Hellfeld-Aufnahmen und in Gewebeflächen der Larve zu finden, so ist eine einzelne Mikroskopaufnahme bereits ausreichend, um die weiteren Schritte erfolgreich abzuschließen. Ist die Aufnahme ein Grauwertbild, so bleibt die vom Detektor aufgezeichnete Matrix zweidimensional. Für Farbbilder sind bereits drei Kanäle (für rot, grün und blau) aufzuzeichnen. Tritt der biologische Effekt lediglich unter Verwendung von Fluoreszenzmikroskopie in Erscheinung, so kann ebenfalls ein zweidimensionales Bild akquiriert werden. Oft ist jedoch die Zuordnung des fluoreszierenden Bereichs im Hellfeld-Kanal gewünscht, was die Dimension der Akquise mindestens um einen Grad erhöht. Zu Beginn der Akquise der Hochdurchsatz-Untersuchung werden somit Modalität bzw. die Anzahl der Kanäle des Bildstroms festgelegt. Hierbei stehen die in der Einleitung (Abschnitt 1.2.2) beschriebenen Möglichkeiten zur Verfügung.

Für eine Hochdurchsatz-Untersuchung bieten sich in erster Linie Hellfeld-Aufnahmen an, da solche traditionell der Standard in der Mikroskopie und daher weit verbreitet sind. Des Weiteren steht eine große Auswahl an (teil)automatisierten Mikroskopen zur Verfügung. Sie können zur oberflächlichen Abbildung fast aller Objekte verwendet werden. Es lassen sich der Zebrafärbung als Ganzes oder auch Bereiche, wie z.B. dessen einsehbare Organe akquirieren. Eine besondere Schwierigkeit ist das (automatische) Auffinden der Fokusebene, welche das Nutzsignal optimal abbildet. Die optimale Schärfe der Zebrafärbungslarven im Chorion ist unter mehreren Gesichtspunkten schwierig zu ermitteln. Für jedes Mikroskop kann der scharf abgebildete Bereich, die sog. Schärfentiefe, aufgrund der physikalischen Gegebenheiten (Auflösung/Numerische Apertur) errechnet werden. Bei hinreichend ebenen Objekten wird daher der gesamte Bereich scharf abgebildet. Da das Chorion jedoch eine relativ große Ausdehnung in der Tiefe hat, lässt sich in jeder Fokusebene nur ein kleiner Teil scharf darstellen. Zwar können z.B. immer die Mitte oder der obere Bereich scharf fokussiert werden, da die Lage der Larven jedoch variiert, ist immer ein anderer morphologischer Bereich der Larve scharf abgebildet. Um das Nutzsignal später vergleichen zu können, muss es jedoch auch möglichst gleich akquiriert sein. Werden nun verschiedene z-Ebenen aufgenommen, sog. Stacks, muss dasjenige ausgewählt werden, welches die interessierende Region möglichst scharf wiedergibt. Eine Vergrößerung der Anzahl an Fokusebenen ergibt jedoch nur so lange Sinn, wie die Summe der Schärfentiefen in den Bildern nicht den Durchmesser des Chorions des Zebrafärbungslarven überschreitet.

Je nach Genauigkeit und Abgrenzbarkeit des Nutzsignals in der Larve muss somit die Anzahl der Fokusebenen festgelegt werden. Es entsteht eine sog. 2.5-dimensionale Aufnahme. Die Modalität ist hierbei Eins (bzw. Drei für Farbaufnahmen) und die Daten sind dreidimensional. Werden Entwicklungsverläufe oder Bewegungsmuster der Zebraabärblinge ausgewertet, müssen eine oder mehrere Bildsequenzen aufgezeichnet werden. Für jeden weiteren Abtastzeitpunkt multipliziert sich die Größe des Bildstroms.

Wird die Fluoreszenzmikroskopie angewandt, lassen sich für eine Hochdurchsatz-Untersuchung die Intensitätswerte eines mittels Markers im Zebraabärbling gekennzeichneten Bereichs akquirieren. Bei der Fluoreszenzmikroskopie ist die Selektivität der Aufzeichnung gleichzeitig Vor- und Nachteil. Ist die Nutzsignalinformation markiert, wird bei der Fluoreszenzmikroskopie lediglich die Nutzsignalinformation akquiriert. Dies hat für die Bildverarbeitung den Vorteil, dass die Region von Interesse nicht vom Hintergrund getrennt werden muss und auch vorliegt, wenn dessen Quelle im Inneren des Zebraabärblings liegen sollte. Der Nachteil jedoch ist, dass in den akquirierten Daten keinerlei Information über die Position der Daten relativ zum Zebraabärbling enthalten ist. Sind beispielsweise nur wenige Zellen markiert, enthält der Bildstrom keine Informationen, ob die Zellen z.B. aus dem Kopf oder dem Rücken stammen. Bildsequenzen lassen sich mit der Konfokal-Mikroskopie nur mit relativ niedrigen Abtastfrequenzen und langen Aufnahmezeiten realisieren. Daher ist eine Betäubung der Larven erforderlich. Um die Fluoreszenzkanäle dennoch einem Bereich im Zebraabärbling zuordnen zu können, besteht die Möglichkeit, sowohl Fluoreszenz- als auch Hellfeld-Aufnahmen von der gleichen Larve zum möglichst gleichen Zeitpunkt unter möglichst gleichen Bedingungen zu akquirieren.

Erweitern lassen sich die Multikanalaufnahmen durch das Hinzufügen weiterer Fluoreszenzkanäle, da bei einer Hochdurchsatz-Untersuchung mehrere Zellen unterschiedlich markiert sein können. Dies können sowohl Autofluoreszenzkanäle als auch markierte Kanäle unterschiedlicher Emissionswellenlängen sein. Je akquirierter Fluoreszenzwellenlänge ist ein gesonderter Kanal aufzuzeichnen und für eine Hellfeld-Aufnahme ein weiterer (bzw. drei für Farbaufnahmen). Bei der Akquise solcher Multikanalaufnahmen entstehen somit schnell große Datenmengen, da sich hier die Datenmenge zusätzlich zu den Abtastzeitpunkten nochmals mit der Kanalanzahl multipliziert. Es sei darauf hingewiesen, dass sich Zellen biologisch oft nicht sofort markieren lassen. Sog. transgene Linien, die sich für diesen Zweck eignen, müssen erst gezüchtet werden. Das Züchten führt zu einer Wartezeit eines oder mehrerer voller Generationszyklen der Fische. Zudem muss beachtet werden, dass für eine Hochdurchsatz-Untersuchung eine entsprechend große Population der transgenen Linie vorhanden sein muss, um den Nachschub an Eiern pro Tag zu gewährleisten, was zu weiteren Wartezeiten führen kann.

Bei der 3. Dimension kann entweder die Hinzunahme der Zeit oder die geometrische Ausdehnung in Richtung der z-Achse gemeint sein. Zur Vermeidung von Missverständnissen werden in der vorliegenden Arbeit nur räumliche Bildströme von Zebraabärblingslarven als dreidimensional bezeichnet, ansonsten wird von Abtastzeitpunkten gesprochen. Wie bei den 2.5-dimensionalen Aufnahmen erweitert sich der Bildstrom bezüglich der z-Achse und alle Pixel werden zu Voxeln. Allerdings ist bei solchen Daten im Gegen-

satz zu 2.5-dimensionalen Bildströmen eine Volumeninformation gegeben, da die verdeckt liegenden Bereiche im Fisch ebenfalls abgebildet werden können. Es steigen jedoch die Anforderungen an die Computer-Hardware sowohl durch den größeren Speicherplatzbedarf als auch in der Auswertung, da die Volumenbilder meist als Ganzes in den Arbeitsspeicher geladen werden müssen.

Die größten Bildströme entstehen bei der Kombination von Multikanalaufnahmen mit dreidimensionalen Aufnahmen. Solche n-dimensionalen Datensätze lassen sich mit dem sog. SPIM Mikroskop aufzeichnen (vgl. Abschnitt 1.2.2). Für die Hochdurchsatz-Untersuchung bleibt die Verarbeitung hoch-dimensionaler Bilddaten jedoch zum Zeitpunkt des Entstehens der vorliegenden Arbeit nur eine theoretische Möglichkeit, da sowohl die Datenakquise, die Rechenleistung und die Speicherkapazitäten solche Datensätze für den Hochdurchsatz ausschließen.

Einen Kompromiss bietet CAM (Computer Aided Microscopy) [130]. Mittels der CAM-Technologie wird ein Übersichtsbild oder ein dreidimensionales Bild in niedriger Auflösung aufgenommen, live ausgewertet und lediglich der Bereich von Interesse akquiriert, beispielsweise mit Hilfe eines Zoomobjektives. Der in sehr hoher Auflösung und ggf. Multikanaltechnik akquirierte Bereich beschränkt sich hierbei möglichst lediglich auf den Bereich im Zebrafisch, der die Nutzsinalinformation trägt.

2.4.2 Analyse und Interpretation

Bei der Analyse und Interpretation (vgl. weiterhin Abb. 2.3) müssen nach der Bildstrom-Vorverarbeitung die Art des biologischen Effekts von Interesse (das Nutzsignal) und die darin enthaltene Information (die Nutzsinalinformation) bestimmt werden. Eine typische Nutzsinalinformation sind entweder eine charakteristische Form oder Fläche eines Bereichs im Zebrafisch oder auch des gesamten Modellorganismus, die biologische Effektstärke, die Bewegung oder die Veränderung einer der zuvor genannten Nutzsinalinformationen über die Zeit.

Die in der Segmentierung gefundenen Bereiche müssen bezüglich des Nutzsignals quantifiziert werden. Je nach Nutzsignal können entweder Merkmale, nach einer Bildvorverarbeitung wie z.B. Verbesserungen des Kontrastes (vgl. Abschnitt 3.2), direkt extrahiert werden oder es müssen erst Zeitreihen ermittelt und die Zeitreihen weiter zu Merkmalen reduziert werden. Hierbei handelt es sich also um Fischinformationszeitreihen und -merkmale oder einfacher Nutzsinalinformationszeitreihen und -merkmale. In Momentaufnahmen wird oft ein Bild in einem Zwischenschritt durch die Segmentierung erzeugt. Solche Bilder enthalten nach Möglichkeit lediglich das Nutzsignal, dessen Information sich durch formbeschreibende Werte quantifizieren lässt und damit Aufschluss über die charakteristischen Bereiche, d.h. der Nutzsinalinformation des Zebrafisches, gibt. Typische solcher Merkmale sind z.B. die Größe, die Rundheit, der Umfang des Bereichs oder die lange/kurze Halbachse einer umschließenden Ellipse. Die biologische Effektstärke zeigt sich

in der Signalstärke im Bild. Das Merkmal der Effektstärke wird vornehmlich bei Fluoreszenzaufnahmen eingesetzt. Die vom Detektor aufgezeichneten Lichtintensitätswerte entsprechen hier der Signalstärke eines bestimmten Bereichs und sind das direkte Maß für die Stärke des biologischen Effekts. Bewegungen lassen sich gut in Bildsequenzen detektieren, von Interesse sind hierbei die Häufigkeit einer Bewegung oder Bewegungsmuster. Quantifiziert werden Bewegungsmuster meist über die Veränderung von Pixelwerten über der Zeit. In Bildsequenzen besteht zusätzlich die Möglichkeit, alle zuvor genannten Merkmale an unterschiedlichen Zeitpunkten zu bestimmen. Die so entstehenden Zeitreihen lassen eine Bewertung der Veränderung über die Zeit zu. Beispielsweise kann die Veränderung der Größe des Dottersacks während der Entwicklung ein solches Maß sein.

Alle extrahierten Merkmale repräsentieren die Nutzsignalinformation und können nun weiter verarbeitet werden. Ziel ist es, die Merkmale der zugehörigen Ausgangsklasse zuzuweisen. Zu Beginn wird der Einfluss von Störfaktoren mittels einer Normierung versucht zu beseitigen. Daraufhin werden möglichst aussagekräftige Merkmale extrahiert und die trennungsstärksten in einer Merkmalsauswahl identifiziert. Für die Identifikation sowie für die spätere Klassifikation ist ein Lerndatensatz erforderlich. Hierfür wird ein Teil der Daten manuell anhand der Bilddaten der entsprechenden Klasse zugeordnet. Mittels der gefundenen signifikanten Merkmale und des Lerndatensatzes wird daraufhin ein Klassifikator angelernt und die Klassenzugehörigkeit der restlichen, nicht manuell zugeordneten Daten geschätzt. Aus den Klassenzugehörigkeiten lassen sich dann in einer Auswertung (vgl. weiter Abb. 2.3) charakteristische bzw. biologisch oder toxikologisch wichtige Kurven und Werte berechnen und interpretieren. Bei einer toxikologischen Untersuchung kann z.B. eine Dosis-Effekt-Kurve dargestellt werden. Dabei wird die geschätzte Klassenzugehörigkeit jeder Einzeluntersuchung gegenüber der Dosis einer Chemikalie aufgetragen. Aus dieser Kurve lassen sich dann charakteristische Werte wie z.B. die höchste Konzentration, bei der kein Effekt auftritt (NOEL; engl. **No Observed Effect Level**), berechnen. Ein weiterer typischer Wert ist die Konzentration, bei der die Hälfte aller Versuchseinheiten nicht überlebensfähig ist (EC₅₀; engl. **Effect Concentration 50**).

Bei der Evaluation muss ein Biologe die präsentierten Daten entsprechend interpretieren. Die Deutung geschieht fast ausschließlich manuell. Für bekannte Effekte lässt sich jedoch eine Plausibilitätsprüfung oder eine automatisierte Gegenprobe einführen, welche eine Sicherheit gegenüber Fehlinterpretationen bietet. Die Ergebnisse müssen schließlich archiviert werden, um zu einem späteren Zeitpunkt eine Reproduktion der Ergebnisse zu ermöglichen.

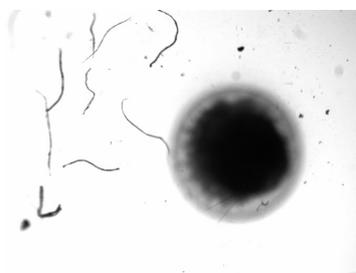
2.4.3 Einfluss der Versuchsparameter auf die Bildqualität

Die für Hochdurchsatz-Untersuchungen geforderten Parameter wie z.B. möglichst kurze Gesamtzeit der Versuchsdurchführung haben zur Folge, dass die Bildakquise automatisiert durchgeführt werden muss. Hierfür finden die im vorigen Abschnitt erwähnten speziellen Hochdurchsatz-Mikroskope Anwendung. Die schnelle Akquise führt jedoch leicht zu Inho-

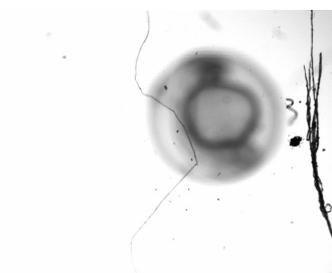
mogenitäten in den Bildströmen, da die Aufnahmen ohne menschliche Überwachung oder Korrektur durchgeführt werden. Solche Inhomogenitäten beeinflussen die Bildqualität und -information. Die Bildverarbeitung und Klassifikation muss entsprechend abgestimmt werden [75, 142]. Zur Betrachtung der Auswirkungen auf die Bildauswertung lassen sich die Inhomogenitäten in zwei Bereiche einteilen:

1. Inhomogenität der *Qualität* durch Fehler bei der Mikroskopie und
2. Inhomogenität der *Information* durch fehlerhafte Objekte.

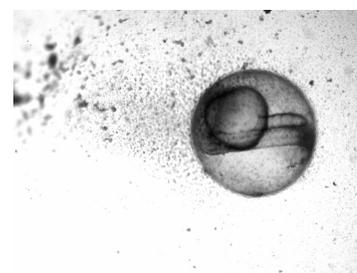
Die Qualität bezieht sich hierbei auf die Güte der Bilder bezüglich Schärfe, Beleuchtung und Reinheit, während mit Information der Inhalt der Bilder gemeint ist, also inwieweit die Akquise den biologischen Effekt von Interesse im Zebraabärbling auf dem Bild wiedergibt. Die Information ist die Forderung, dass der Bildstrom die Nutzsignalinformation enthält. Die Akquise-Methode muss also so gewählt werden, dass der biologische Effekt von Interesse im Bildstrom vorhanden ist. Abb. 2.5 zeigt Beispiele für Beeinträchtigungen, die bei der Bildakquise im Hellfeld entstehen können. In Abb. 2.5a und Abb. 2.5b stellt sich der automatische Fokus des Mikroskops auf Schmutz in der Flüssigkeit oder auf der Linse scharf. Dadurch sind die Larven nur undeutlich zu erkennen. In Abb. 2.5c und Abb. 2.5d sind die Larven von Schmutz umgeben und teilweise überdeckt. In Abb. 2.5e und Abb. 2.5f schließlich treten große Fremdkörper in den Bildern auf, die zum Teil sogar die Kontur um die Eihülle unterbrechen.



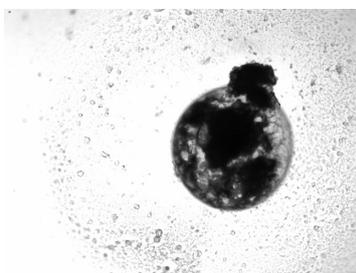
(a) Fehlerhafte Fokussierung



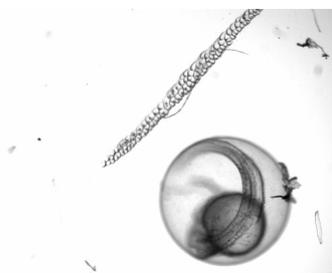
(b) Fehlerhafte Fokussierung



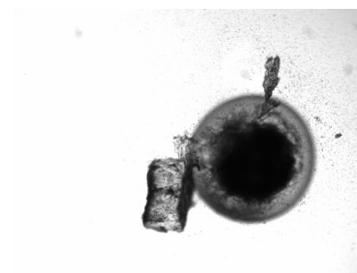
(c) Schmutzpartikel im Wasser



(d) Schmutzpartikel im Wasser



(e) Fremdkörper im Bild



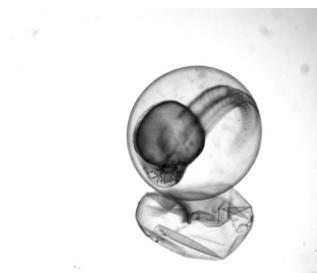
(f) Fremdkörper im Bild

Abbildung 2.5: Qualitative Inhomogenität des Bildstroms. Zur Akquise wurde ein Hochdurchsatz-Mikroskop des Typs Olympus-Scan^R verwendet [4].

Die Qualität des Bildstroms lässt sich maßgeblich durch Sorgfalt bei der Bildakquise und Präparation verbessern. So werden z.B. sämtliche Verunreinigungen, die sich auf dem Objektiv, in der Luft oder im Wasser um die Larven befinden, mit aufgezeichnet und verdecken oder verschlechtern Informationen und damit die Qualität aller weiteren Schritte der Auswertung und können im schlimmsten Fall zu einer Fehlklassifikation führen. Je konstanter die Umgebungsbedingungen wie z.B. die Raumbeleuchtung sind, desto konstanter ist auch die Qualität des Bildstroms. Ebenso kann der automatische Fokus des Mikroskops sich durch übermäßige Verschmutzungen der Probe falsch justieren, sodass irrelevante Bildteile scharf abgebildet werden. In den im Rahmen der vorliegenden Arbeit betrachteten Bildströmen erwiesen sich Schmutzpartikel, Bildunschärfe und eine ungleichmäßige Ausleuchtung als die am häufigsten auftretenden qualitativen Einschränkungen.



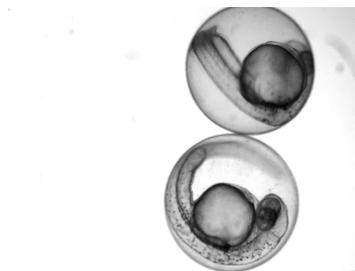
(a) Eihülle ist mechanisch zerstört



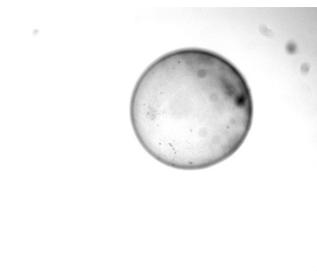
(b) Weitere zerfallene Eihülle im Bild



(c) Eihülle unvollständig



(d) Zwei Larven abgebildet



(e) Ei ist unbefruchtet



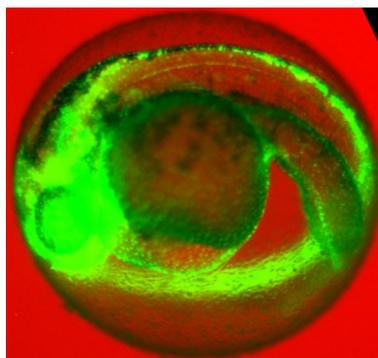
(f) Leeres Näpfchen abgelichtet

Abbildung 2.6: Informelle Inhomogenität der Datensätze [4]

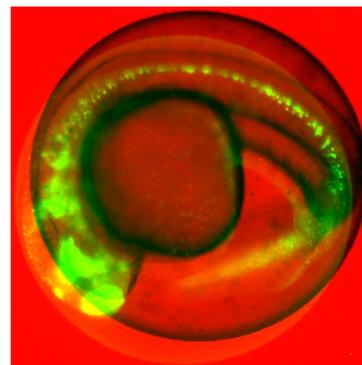
Oft sind die Aufnahmen zwar von guter Qualität bezüglich der Schärfe und der Beleuchtung, jedoch wird die Nutzsignalinformation vom Bildstrom nicht oder nicht komplett wiedergegeben. Larven sind z.B. nicht vollständig auf dem Bild abgelichtet oder auf einem Bild sind mehrere Larven bzw. auch gar keine zu finden (vgl. Abb. 2.6), obwohl im Versuchsprotokoll genau eine Larve gefordert war. In Abb. 2.6a ist das Zebrafärbärlingsei durch mechanische Einwirkung bei der Vereinzelung mit der Pipette oder beim Transport zerstört worden. In Abb. 2.6b befindet sich neben der Eihülle der lebenden Larve eine weitere, leere Eihülle, die möglicherweise von einer bereits geschlüpften Larve stammt. Abb. 2.6c bildet nur ca. 1/3 der Larve ab, während in Abb. 2.6d zwei Larven gleichzeitig bei der Bilderfassung in dem Näpfchen der Mikrotiterplatte waren. In Abb. 2.6e ist das Zebrafärbärlingsei nicht befruchtet, es erscheint auf dem Bild als „leerer Ring“. In Abb. 2.6f schließlich ist ein leeres Näpfchen abgelichtet. Die gezeigten Beispiele können nicht

das gesamte Spektrum von Inhomogenitäten abbilden, sollen jedoch die Vielfältigkeit der Fehlerquellen aufzeigen.

Bei Multikanalaufnahmen ist ein weiterer Aspekt die Problematik des zeitlichen Versatzes, welcher zwischen den Aufnahmen der einzelnen Kanäle entsteht. Z.B. nimmt das Mikroskop Olympus-Scan^R bei mehreren Fokusebenen in einem ersten Durchlauf alle Fokusebenen der Reihe nach auf, beginnend mit den Hellfeld-Aufnahmen. Daraufhin wird das Durchlicht durch eine Blende verschlossen, ein Fluoreszenzfilter automatisch eingesetzt und die entsprechenden Fokusebenen für die Fluoreszenzkanäle akquiriert. Die Bildakquise erfolgt damit nicht simultan [90], was Abweichungen der Bilder zueinander zur Folge hat, da zum einen Toleranzen des Mikroskops den optischen Fluss zwischen beiden Aufnahmeserien verändern und zum anderen sich das Objekt entweder eigenständig (z.B. durch Muskelzucken der Larven) oder durch Störungen bewegt haben kann. Das hat zur Folge, dass gleiche Strukturen im Fisch nicht an der gleichen Stelle im Bild abgebildet werden. Kleine Toleranzen können oder müssen in einem solchen Fall durch eine Registrierung, d.h. eine Zuordnung der einzelnen Kanäle durch Translationen und Rotationen, ausgeglichen werden, während die Registrierung bei großen Toleranzen scheitert oder einen zu großen Aufwand erfordert. Solche Fälle müssen dann ggf. mittels geeigneter Validitätsprüfungen erkannt und ausgeschlossen werden (vgl. Kapitel 3). Ein Beispiel für eine solche Abweichung ist in Abb. 2.7 gegeben. Während in der linken Abbildung 2.7a sich der Fluoreszenzkanal mit dem Hellfeldkanal deckt, sind in Abb. 2.7b die Kanäle nicht deckungsgleich was deutlich am Kopfbereich des Fisches zu sehen ist.



(a) richtig



(b) falsch

Abbildung 2.7: Beispiele für die Registrierung von Multikanalaufnahmen. Überlagerung der Grauwertbilder durch grün eingefärbte Fluoreszenzaufnahmen auf rotem Hintergrund (Falschfarben-Darstellung)

2.5 Mathematisches Modell der Hochdurchsatz-Untersuchung

Bei der Betrachtung der Hochdurchsatz-Untersuchung ist eine allgemein anwendbare Notation, welche sich auf alle Ausprägungen der Untersuchung adaptieren lässt, hilfreich. Daher wird im folgenden Abschnitt für alle identifizierten Parameter der Hochdurchsatz-Untersuchung eine modellhafte Beschreibung eingeführt, welche sich universell auf Hochdurchsatz-Untersuchungen anpassen lässt und ein einheitliches Beschreibungsmodell darstellt.

Werden die identifizierten Parameter aus Sicht der Eingangsdaten der Bildverarbeitung beschrieben, so kann eine formale Betrachtung vorgenommen werden, aus der sich eine modellhafte Beschreibung ableiten lässt. Der biologische Effekt und somit das Nutzsignal werden charakterisiert durch vier Parameter, die sich ausnahmslos auf die Art oder den Ort beziehen, an dem das Nutzsignal auftritt. Die Parameter sind Auftrittsmodalität, -ort, -häufigkeit und -dauer (vgl. Abb. 2.8)



Abbildung 2.8: Charakterisierung des Nutzsignals

Bei der Betrachtung der Hochdurchsatz-Untersuchung muss die **Auftrittsmodalität** des Signals geklärt und in der Versuchsauslegung berücksichtigt werden. Der Begriff Auftrittsmodalität steht für die verschiedenen Kanäle, mit denen ein Nutzsignal aufgezeichnet werden kann, etwa mittels der Aufzeichnung von Grauwertbildern von einem Hellfeld-Mikroskop und z.B. einem weiteren Fluoreszenz-Kanal. Die Auftrittsmodalität hat daher Einfluss auf die Dimension der Bildakquise. Für den Parameter wird der Bezeichner m_ψ eingeführt. Ein weiterer Parameter ist der **Auftrittsort** des Nutzsignals. Der Auftrittsort ist gleichzeitig der Ort bzw. der Raum von Interesse (engl. *(ROI)* Region of Interest) und durch Indizes definiert, welche die Lage des Raums innerhalb des Nutzsignals beschreiben. Für den Auftrittsort wird \mathbf{u}_ψ eingeführt mit $\mathbf{u}_\psi = [x1_\psi, x2_\psi, y1_\psi, y2_\psi, z1_\psi, z2_\psi]^T$. Die **Auftrittshäufigkeit** beschreibt, inwieweit das zu beobachtende Nutzsignal ständig sichtbar ist (z.B. ständig oder nur zu einer bestimmten Wahrscheinlichkeit in einem von 1000 Zeitpunkten). Für die Auftrittshäufigkeit wird der Bezeichner f_ψ eingeführt. Während die Auftrittshäufigkeit des Nutzsignals vornehmlich zu Wartezeiten und niedrigerem Durchsatz führt, wirkt sich die **Auftrittsdauer** auf die Länge der Bildsequenzen aus,

sollte das Nutzsignal über die gesamte Auftrittsdauer beobachtet werden müssen. Für die gewählte Aufnahmedauer des Nutzsignals wird der Bezeichner t_{Ψ} eingeführt.

Eine konkrete bildbasierte Hochdurchsatz-Untersuchung besteht aus einer Anzahl N von divers behandelten Zebrabärblingslarven. Es wird vereinfachend davon ausgegangen, dass eine Larve Träger genau eines biologischen Effekts und damit die Versuchseinheit ist. Die Larven existieren als reale Versuche \mathbf{RV} , d.h. in Form von kontinuierlichen Werten und deren digitaler Abbildung in einem Bildstrom \mathbf{BS} . Die Anzahl an untersuchten Zebrabärblingslarven setzt sich zusammen aus drei Klassen. Die Klasse der Positiv-Kontrollen bestehend aus n_{C^+} Versuchseinheiten, die der Negativ-Kontrollen bestehend aus n_{C^-} Versuchseinheiten und die Klasse der den Rahmenbedingungen ausgesetzten Proben bestehend aus n_s Versuchseinheiten. Die Anzahl an Zebrabärblingslarven in der Hochdurchsatz-Untersuchung ist dann gegeben durch:

$$n = 1 \dots N \quad \text{mit} \quad N = n_{C^+} + n_{C^-} + n_s, \quad N \in \mathbb{N}^+. \quad (2.1)$$

Die Versuchseinheit wird durch Größen beschrieben, die die Versuchsbedingungen bestimmen. Die Größen werden in Anlehnung an [138] als Faktoren bezeichnet. Es erweist sich als zweckmäßig, Plan- und Störfaktoren zu unterscheiden. Die Planfaktoren beschreiben jene Größen, deren Einfluss durch systematische Variation untersucht werden soll. In den Störfaktoren werden alle anderen Größen zusammengefasst, die ebenfalls einen Einfluss auf die Versuchsergebnisse haben. In der Literatur werden die Störfaktoren weiterhin unterteilt in Konstant-, Rest- und Blockfaktoren [138]. Konstantfaktoren sind Faktoren, die ihren Wert während der Versuchsdurchführung nicht ändern, Blockfaktoren sind Faktoren, die sich zu Blöcken oder Stufen zusammenfassen lassen und Restfaktoren sind die übrigen Faktoren. Von der zusätzlichen Unterteilung wird in der vorliegenden Arbeit abgesehen, da die Unterteilung vornehmlich für die statistische Versuchsplanung von Interesse ist. Jeder Faktor besteht aus einer Menge von Klassen, den Ausgangsklassen.

Jede Hochdurchsatz-Untersuchung enthält eine Anzahl $i_{s_y} = 1 \dots s_y$ von Planfaktoren y und eine Anzahl $i_{s_z} = 1 \dots s_z$ von Störfaktoren z . Die Planfaktoren lassen sich anhand der Versuchseinheiten in einer (N, s_y) -dimensionalen Matrix \mathbf{Y} anordnen. Im Folgenden wird meist, zur einfacheren Notation, der Fall eines eindimensionalen Planfaktors ($s_y = 1$) betrachtet. Analog hierzu werden die Störfaktoren in einer (N, s_z) -dimensionalen Matrix \mathbf{Z} notiert. Die Anzahl an Klassen des i_{s_y} -ten Planfaktors werden im Folgenden mit $m_{y, i_{s_y}}$ und die Anzahl an Klassen des i_{s_z} -ten Störfaktors als $m_{z, i_{s_z}}$ bezeichnet. Für jede Versuchseinheit wird bezüglich jedes Faktors eine Klassenzuordnung getroffen. Weiter werden alle Faktoren durch die noch zu definierenden Versuchsparameter $(\mathbf{p}_z)_n, (\mathbf{p}_y)_n$ bestimmt

$$\mathbf{Y} = \begin{pmatrix} y_1[1] & \cdots & y_{s_y}[1] \\ \vdots & \ddots & \vdots \\ y_1[N] & \cdots & y_{s_y}[N] \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} z_1[1] & \cdots & z_{s_z}[1] \\ \vdots & \ddots & \vdots \\ z_1[N] & \cdots & z_{s_z}[N] \end{pmatrix}. \quad (2.2)$$

Zum besseren Verständnis sei als Beispiel eine Hochdurchsatz-Untersuchung angeführt, in welcher Zebrabärblingseier nach Koaguliertheit und Herzschlag untersucht werden

sollen und welche an zwei Mikroskopen und an unterschiedlichen Tagen aufgenommen wurde. In diesem Beispiel sind die Zebrabärblingseier die Versuchseinheiten. Der Herzschlag sowie die Koaguliertheit sind die Planfaktoren, mit jeweils $m_{y,1} = m_{y,2} = 3$ Klassen, $y_1[n] = \{1,2,3\}$, $y_2[n] = \{1,2,3\}$. Hierbei steht bei beiden Planfaktoren die Klasse 1 für *ja*, die Klasse 2 für *vielleicht* und Klasse 3 für *nein*. Die Mikroskope und Aufnahmetage sind Störfaktoren mit jeweils einer Klasse für jedes Mikroskop und jeden Aufnahmetag. Die gewählten Parameter des Mikroskops wie Vergrößerung, Beleuchtungsstärke etc. sind die Versuchsparameter jeder Versuchseinheit $(\mathbf{p}_z)_n, (\mathbf{p}_y)_n$. Die Versuchsparameter sind für die Planfaktoren und Störfaktoren normalerweise identisch ($\mathbf{p}_z = \mathbf{p}_y = \mathbf{p}_{zy}$). Der gesamte Ablauf der Hochdurchsatz-Untersuchung, vom mit \mathbf{RV} bezeichneten realen Versuch bis zur Klassenzuweisung der jeweiligen Klassen aus \mathbf{Y}, \mathbf{Z} ist in dem Schema in Abb. 2.9 dargestellt.

Für die wichtigsten Faktoren der verwendeten Zebrabärblingslarven, nämlich dem Alter t_{hpf} und dem Faktor für die gesamte erforderliche Zeit für die Realisierung t_{RV} , werden Bezeichner eingeführt. Alle weiteren Faktoren werden in \mathbf{z}_{RV} zusammengefasst. Die Realisierung jedes Versuchs \mathbf{RV} ist Element einer Menge, deren Inhalt abhängig von den Randbedingungen des Versuchs ist und mit \mathbb{RV} bezeichnet wird. Es lässt sich nun formulieren:

$$\begin{pmatrix} \mathbf{RV}_1 \\ \vdots \\ \mathbf{RV}_N \end{pmatrix} ; \quad \mathbf{RV} \in \mathbb{RV} \quad (2.3)$$

$$\mathbf{RV}_n = f(\mathbf{p}_{zy})_n \quad (2.4)$$

$$\text{mit } (\mathbf{p}_{zy})_n = (t_{\text{hpf}}, t_{RV}, \mathbf{z}_{RV})_n^T. \quad (2.5)$$

Die notwendige Zeit für einen Versuch (ohne Parallelisierung der Arbeitsabläufe) t_{RV} setzt sich zusammen aus der Summe der Dauer aller notwendigen Präparationsschritte $(t_p)_{i_p}$, mit der Anzahl an Präparationsschritten n_p und der Dauer der Akquise t_{aq}

$$t_{RV} = \sum_{i_p=1}^{n_p} (t_p)_{i_p} + t_{aq}. \quad (2.6)$$

Die Zugehörigkeit der Versuche \mathbf{RV} zu den jeweiligen Klassen \mathbf{y} ist unbekannt und soll in der Hochdurchsatz-Untersuchung durch einen Klassifikator geschätzt werden. Schätzungen werden mit einem Dach gekennzeichnet. Ebenso lässt sich die Klassenzugehörigkeit der Restfaktoren aus \mathbf{z} schätzen. Es ergeben sich die Abbildungen S_y und S_z

$$S_y(\mathbf{p}) : \mathbf{RV} \mapsto \hat{\mathbf{y}} \quad (2.7)$$

$$S_z(\mathbf{p}) : \mathbf{RV} \mapsto \hat{\mathbf{z}}. \quad (2.8)$$

Die Zuordnung zu den Klassen von \mathbf{y} ist Ziel der Hochdurchsatz-Untersuchung und wird in Abhängigkeit des Nutzsymbols und des noch zu definierenden Parametervektors \mathbf{p} erfolgen. Das Nutzsymbols steht nicht unmittelbar zur Verfügung, sondern muss aus dem Bild extrahiert werden. Die Extraktion setzt sich aus den Teilschritten Datenerfassung S_{RV_1} , Bildstrom-Vorverarbeitung S_{RV_2} und Segmentierung S_{RV_3} zusammen.

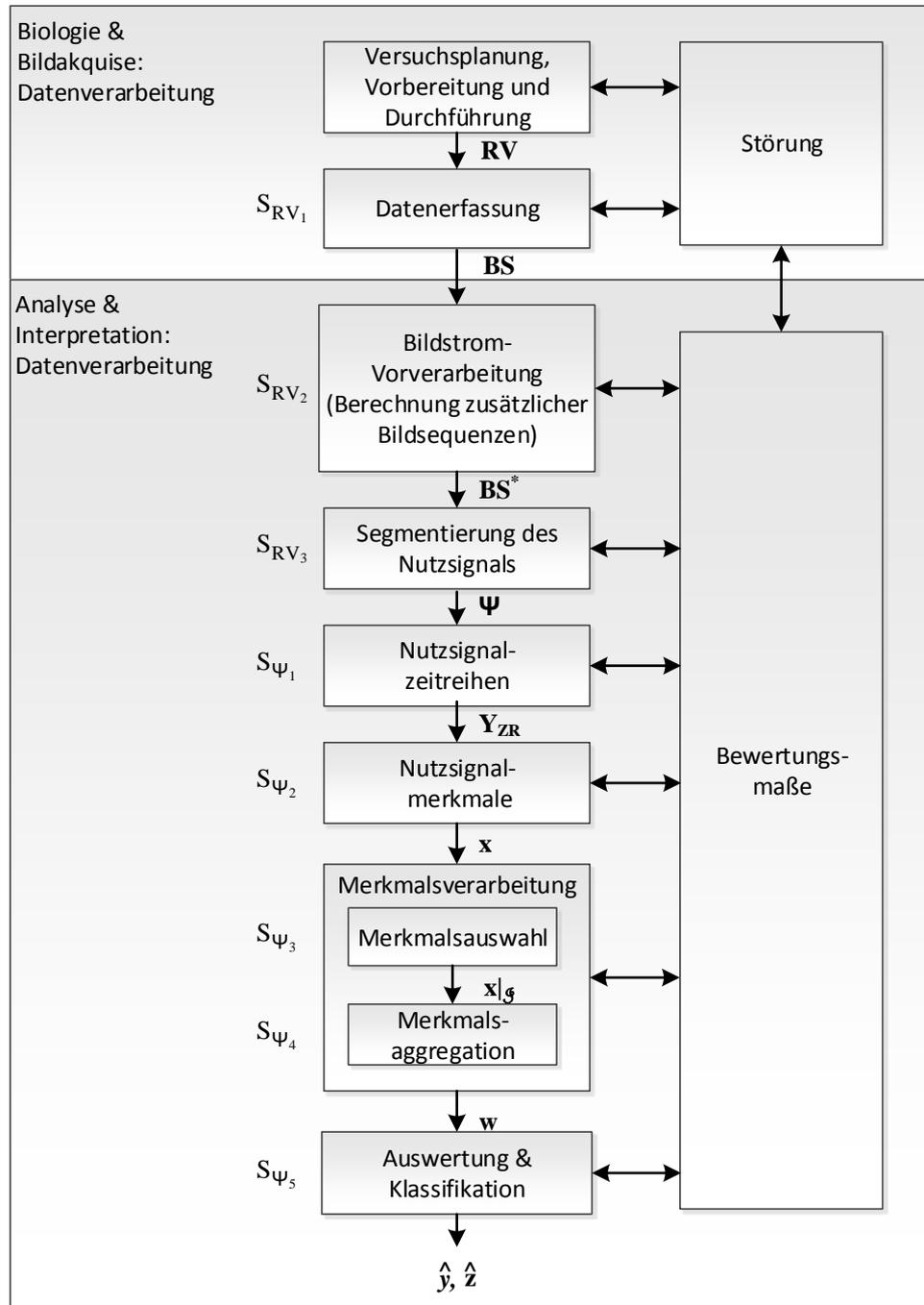


Abbildung 2.9: Schema zum mathematischen Modell für bildbasierte Hochdurchsatz-Untersuchungen aus Sicht der Datenverarbeitung

Das Nutzsinal wird mittels der Matrix Ψ dargestellt. Für eine gute Fragestellung muss das Nutzsinal in jedem realen Versuch \mathbf{RV} enthalten sein. Der Bildstrom \mathbf{BS} ist die digitale Repräsentation des realen Versuchs \mathbf{RV} . Er ist die Abtastung von \mathbf{RV} bezüglich der Zeit und des Ortes und muss für eine erfolgreiche Hochdurchsatz-Untersuchung das Nutzsinal abbilden. Daher muss für die Abtastung auch das Abtasttheorem [154] eingehalten werden. Die Bildakquise S_{RV_1} bildet somit \mathbf{RV} derart auf \mathbf{BS} ab, dass der biologische Effekt und damit das Nutzsinal Ψ aus dem Bildstrom extrahierbar sind:

$$S_{RV_1}(\mathbf{p}_{BS}) : \mathbf{RV} \mapsto \mathbf{BS}. \quad (2.9)$$

Der Bildstrom ist abhängig von den Parametern \mathbf{p}_{BS} . Dabei handelt es sich zum einen um die Art des Detektors und der aufgezeichneten Wellenlänge, d.h. der Akquise-Modalität des Bildstroms $(m_{BS})_n$. Des Weiteren muss der Raum der Akquise als Parameter berücksichtigt werden und wird durch den Vektor $(\mathbf{u}_{BS})_n = [x_{BS_1}, x_{BS_2}, y_{BS_1}, y_{BS_2}, z_{BS_1}, z_{BS_2}]_n^T$ beschrieben. Schließlich müssen noch die Anzahl akquirierter Frames $(t_{BS})_n$ und die Aufnahmefrequenz (engl. Framerate) $(f_{BS})_n$ gewählt werden. Damit lässt sich für den Bildstrom schreiben:

$$\mathbf{BS}_n = f(\mathbf{p}_{BS})_n \quad (2.10)$$

$$\text{mit } (\mathbf{p}_{BS})_n = (m_{BS}; \mathbf{u}_{BS}; t_{BS}; f_{BS})_n^T. \quad (2.11)$$

Die wichtigste Forderung an die Bildakquise ist, dass der Bildstrom \mathbf{BS} ein geeignetes Nutzsinal enthält. Das bedeutet, dass die Parameter des Bildstroms in Abhängigkeit der Parameter des Nutzsignals gewählt werden. Die Parameter des Nutzsignals \mathbf{p}_Ψ lassen sich somit in Übereinstimmung mit Abb. 2.8 analog zu den Parametern des Bildstroms formulieren

$$\Psi_n = f(\mathbf{p}_\Psi)_n \quad (2.12)$$

$$\text{mit } (\mathbf{p}_\Psi)_n = (m_\Psi; \mathbf{u}_\Psi; t_\Psi; f_\Psi)_n^T. \quad (2.13)$$

Gut gewählte Akquise-Parameter des Bildstroms bilden also möglichst ausschließlich das Nutzsinal ab.

Der mit genannten Parametern akquirierte Bildstrom \mathbf{BS} setzt sich aus Einzelbildern \mathbf{I}_E zusammen, welche jeweils durch eine Matrix von Intensitätswerten beschrieben werden:

$$(\mathbf{I}_E[i_f, i_w])_n \in \mathbb{N}^{I_x \times I_y \times I_z}, \quad (2.14)$$

$$i_w, i_f \in \mathbb{N}.$$

Hierbei ist $i_f = 1 \dots r_F$ der Index aller Abtastungen und $i_w = 1 \dots r_w$ der Index der Wiederholungen. Weiter ist I_x die Auflösung des Bildes in x - und I_y die Auflösung des Bildes in y -Richtung, $i_z = 1 \dots I_z$ der Index für Schichtaufnahmen bzw. Fokusebenen und $i_c = 1 \dots I_c$, der Index der Modalitäten. Der gesamte Bildstrom lässt sich somit schreiben als:

$$\mathbf{BS}_n = \begin{pmatrix} \mathbf{I}_E[1, 1] & \cdots & \mathbf{I}_E[1, r_w] \\ \vdots & \ddots & \vdots \\ \mathbf{I}_E[r_F, 1] & \cdots & \mathbf{I}_E[r_F, r_w] \end{pmatrix}_{i_c, n}, \quad (2.15)$$

$$\mathbf{BS}_n \in \mathbb{N}^{I_x \times I_y \times I_z \times I_c \times r_F \times r_w},$$

$$i_c, r_F, r_w \in \mathbb{N}.$$

Es ergeben sich die Bild-Rohdaten des Bildstroms \mathbf{BS} mit insgesamt B Bildern.

$$B = I_c \cdot I_z \cdot r_w \cdot r_f. \quad (2.16)$$

Der Einfluss der Bildstromparameter ist anschaulich in Abb. 2.10 dargestellt.

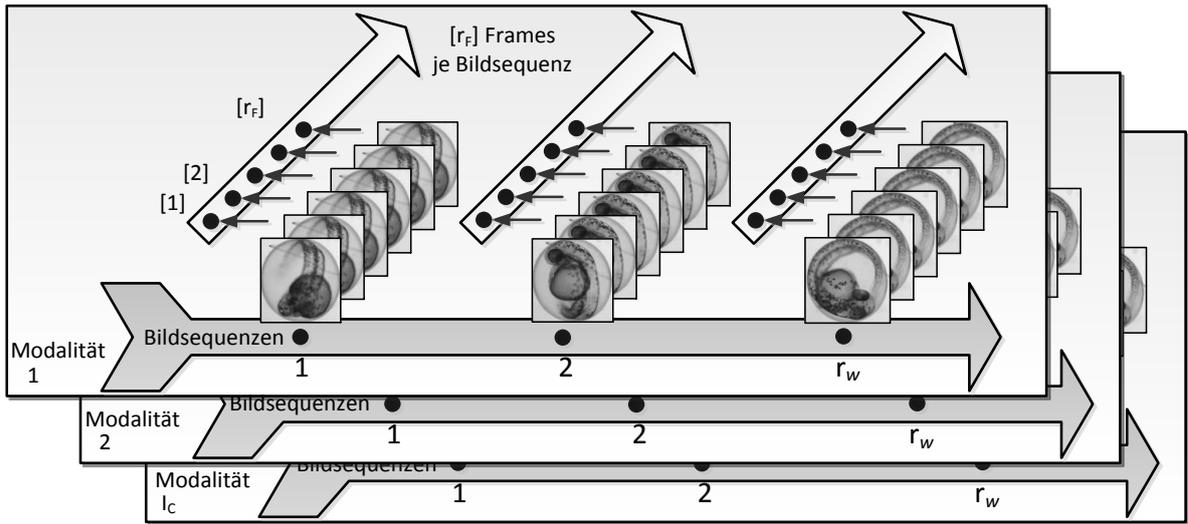


Abbildung 2.10: Grafische Veranschaulichung eines Elementes des Bildstroms \mathbf{BS}_n [7]

In vielen Fällen ist eine Vorverarbeitung des Bildstroms sinnvoll oder notwendig. In einer Bildvorverarbeitung werden z.B. Extended-Fokus-Bilder berechnet (vgl. Abschnitt 1.2.2) oder es werden Punktoperationen wie Histogrammanpassungen oder Korrekturen des Gammawertes vorgenommen. Ebenso fallen alle linearen und nichtlinearen Bildfilter-Operatoren in den genannten Bereich. Die Bildvorverarbeitung wird hier allgemein beschrieben als eine Abbildung des Bildstroms \mathbf{BS} auf einen für die weitere Verarbeitung verbesserten Bildstrom \mathbf{BS}^* , welche von einem oder mehreren Filterparametern \mathbf{p}_{f_n} abhängt. Die Parameter \mathbf{p}_{f_n} sind z.B. die gewählten Filtermasken, Schwellenwerte usw. Der Aufbau von \mathbf{BS}^* ist hierbei analog zu Formel (2.15)

$$S_{RV_2}(\mathbf{p}_{f_n}) : \mathbf{BS} \mapsto \mathbf{BS}^*. \quad (2.17)$$

Im Allgemeinen ist eine Dimensionsreduktion der Daten beabsichtigt. In speziellen Fällen steigt die Dimension jedoch auch an, z.B. wenn es notwendig ist, neue Bilder zu berechnen. Bildfusionen hingegen führen immer zu einer Dimensionsreduktion.

Die Bildverarbeitung hat nun die Aufgabe, das Nutzsignal Ψ in jedem Bild aus dem Bildstrom \mathbf{BS} zu extrahieren oder zu segmentieren. Die Segmentierung S_{RV_3} bildet somit

\mathbf{BS}^* auf Ψ ab

$$S_{RV_3}(\mathbf{p}_\Psi) : \mathbf{BS}^* \mapsto \Psi = \begin{pmatrix} \mathbf{I}_{E\Psi}[1, 1] & \cdots & \mathbf{I}_{E\Psi}[1, r_{w\Psi}] \\ \vdots & \ddots & \vdots \\ \mathbf{I}_{E\Psi}[r_{F\Psi}, 1] & \cdots & \mathbf{I}_{E\Psi}[r_{F\Psi}, r_{w\Psi}] \end{pmatrix}_{i_{c\Psi}, n}, \quad (2.18)$$

$$\Psi_n \in \mathbb{N}^{I_{x\Psi} \times I_{y\Psi} \times I_{z\Psi} \times I_{c\Psi} \times r_{F\Psi} \times r_{w\Psi}}. \quad (2.19)$$

Hierin sind $\mathbf{I}_{E\Psi}$ Ausschnitte jedes Einzelbildes \mathbf{I}_E des originalen Bildstrom, welche nun möglichst nur das Nutzsignal enthalten

$$\begin{aligned} (\mathbf{I}_{E\Psi}[i_f, i_w])_n &\in \mathbb{N}^{I_{x\Psi} \times I_{y\Psi} \times I_{z\Psi}}, \\ i_{w\Psi}, i_{f\Psi} &\in \mathbb{N}. \end{aligned} \quad (2.20)$$

Der segmentierte Bildstrom setzt sich analog zum Bildstrom zusammen. Die Dimension von Ψ ändert sich bei der Segmentierung. Da auf die Region von Interesse Bezug genommen wird, wird die Dimension gegenüber \mathbf{BS}^* meist geringer. Lediglich die Modalität $I_{c\Psi}$ kann durch das Errechnen bzw. Extrahieren von informationstragenden neuen Bildern, wie beispielsweise Differenzbilder, auch größer werden.

Als Klassifikatoren werden Systeme bezeichnet, die in der Lage sind, Eingangsgrößen verschiedenen Klassen zuzuordnen. Die extrahierte Nutzsignalinformation wird nun in Anlehnung an [110, 137, 141] einer Klasse zugewiesen. Der Ablauf der Klassifikation setzt sich aus den Teilschritten Merkmalsextraktion S_{Ψ_1} und S_{Ψ_2} , Merkmalsauswahl S_{Ψ_3} , Merkmalsaggregation S_{Ψ_4} und der Entscheidungsfindung mittels einer Entscheidungsregel S_{Ψ_5} zusammen (vgl. Abb. 2.9 unten). Für jeden Teilschritt sind Parameter zu wählen, für die die Bezeichner $\mathbf{p}_{S\Psi} = [\mathbf{p}_{S\Psi_1}, \mathbf{p}_{S\Psi_2}, \mathcal{S}, \mathbf{p}_{S\Psi_4}, \mathbf{p}_{S\Psi_5}]^T$ eingeführt werden und auf die im Folgenden näher eingegangen wird.

Es soll nun nach Abb. 2.9 mittels der Nutzsignalinformation Ψ jeder Versuchseinheit eine Klasse der Planfaktoren aus \mathbf{Y} zugewiesen werden. Analog lässt sich auch die Klassenzugehörigkeit zu einer der Störgrößen aus \mathbf{Z} schätzen:

$$S_{\psi_y} : \Psi \mapsto \hat{\mathbf{y}} \quad (2.21)$$

$$S_{\psi_z} : \Psi \mapsto \hat{\mathbf{z}} \quad (2.22)$$

Da eine direkte Lösung des Problems anhand der Bilddaten nicht möglich ist, wird die Nutzsignalinformation durch Ersatzgrößen, die Merkmale, charakterisiert. Die Aufgabe ist die Berechnung niederdimensionaler und informationstragender, d.h. das Nutzsignal Ψ beschreibender, Merkmale. Die wichtigsten Merkmale basieren auf Pixeln, Kanten, Texturen, Regionen, Objekten und Szenen (vgl. [88, 110]).

Die Merkmalsextraktion transformiert die Daten der Nutzsignalinformation für jeden Abtastzeitpunkt in einen m_1 -dimensionalen Merkmalsraum in Abhängigkeit der Parameter $\mathbf{p}_{S\Psi_1}$. Für jede Zebrabärblingslarve enthält die Matrix daher Merkmalswerte y_{ZR} zu jedem Abtastzeitpunkt und somit eine Zeitreihe \mathbf{Y}_{ZR} für jedes Merkmal. Diese Zeitreihen werden

für die Hochdurchsatz-Untersuchung Fisch-Informationszeitreihen genannt.

$$S_{\Psi_1}(\mathbf{p}_{S\Psi_1}) : \Psi \mapsto \mathbf{Y}_{ZR} = \begin{pmatrix} y_{ZR}[1, 1] & \cdots & y_{ZR}[1, r_w] \\ \vdots & \ddots & \vdots \\ y_{ZR}[r_F, 1] & \cdots & y_{ZR}[r_F, r_w] \end{pmatrix}_{m_1}, \quad (2.23)$$

$$\mathbf{Y}_{ZR} \in \mathbb{R}^{r_F \times r_w \times m_1}.$$

Die Fisch-Informationszeitreihen werden durch eine Dimensionsreduktion wiederum mittels Merkmalen beschrieben. Es entsteht schließlich ein m_2 -dimensionaler Merkmalsraum in Abhängigkeit der Parameter $\mathbf{p}_{S\Psi_2}$. Solche Merkmale werden für die Hochdurchsatz-Untersuchung Fisch-Informationsmerkmale genannt

$$S_{\Psi_2}(\mathbf{p}_{S\Psi_2}) : \mathbf{Y}_{ZR} \mapsto \mathbf{x}^T = (x_1, \dots, x_{m_2}), \quad (2.24)$$

$$\mathbf{x} \in \mathbb{R}^{1 \times m_2} \quad (2.25)$$

Die extrahierten Merkmale sind häufig redundant oder enthalten z.T. keine Information, die sich für eine Klassifikation eignet. Daher wird eine Merkmalsauswahl angewandt, die sich eines Kriteriums bedient, um die aussagekräftigsten Merkmale zu identifizieren. Eine Übersicht zu solchen Kriterien ist in [3, 115] zu finden. Der Bezeichner \mathcal{S} steht für die Indexmenge der ausgewählten m_3 Merkmale und es ergibt sich die Abbildung

$$S_{\Psi_3} : \mathbf{x} \mapsto \mathbf{x}|_{\mathcal{S}} = (x_1, \dots, x_{m_3}), \quad (2.26)$$

$$\mathbf{x}|_{\mathcal{S}} \in \mathbb{R}^{1 \times m_3}, \quad m_2 \geq m_3.$$

Die Merkmalsaggregation dient zur Reduktion der Dimension des Merkmalsraumes. Hierdurch werden zudem Klassifikationen, die auf Merkmalsräumen mit $m_3 > 3$ basieren, grafisch darstellbar. Es existieren verschiedene Verfahren unterschiedlicher Arbeitsweise, von denen die bekanntesten die Hauptkomponentenanalyse (HKA) und die lineare Diskriminanzanalyse (LDA) sind. Eine Übersicht findet sich in [34, 71, 141, 150]. Beide Verfahren liefern als Ergebnis einen m_d -dimensionalen Merkmalsraum, welcher eine Linearkombination aller m_3 Merkmale ist und von dem Parameter $\mathbf{p}_{S\Psi_4}$ der gewählten Aggregation abhängt.

$$S_{\Psi_4}(\mathbf{p}_{S\Psi_4}) : \mathbf{x}|_{\mathcal{S}} \mapsto \mathbf{w} \quad (2.27)$$

$$\mathbf{w} \in \mathbb{R}^{1 \times m_d}, \quad m_3 \geq m_d.$$

Den abschließenden Schritt des Klassifikatorentwurfs stellt die Konstruktion der Entscheidungsregel dar, die den entstandenen Merkmalsatz einer der Klassen von \mathbf{Y} bzw. \mathbf{Z} , in Abhängigkeit der Klassifikationsparameter $\mathbf{p}_{S\Psi_5}$, zuweist. Da es sich hierbei um eine Schätzung handelt, wird die Zuordnung $\hat{\mathbf{y}}$ bzw. $\hat{\mathbf{z}}$ bezeichnet

$$S_{\Psi_5 y}(\mathbf{p}_{S\Psi_5}) : \mathbf{w} \mapsto \hat{\mathbf{y}}, \quad (2.28)$$

$$S_{\Psi_5 z}(\mathbf{p}_{S\Psi_5}) : \mathbf{w} \mapsto \hat{\mathbf{z}}. \quad i \quad (2.29)$$

Nun sind alle Versuchsparameter des Versuchsplans der Hochdurchsatz-Untersuchung bestimmt und werden in dem Parametervektor \mathbf{p} zusammengefasst

$$\mathbf{p} = (\mathbf{p}_y; \mathbf{p}_z; \mathbf{p}_{BS}; \mathbf{p}_{\Psi}; \mathbf{p}_{fn}; \mathbf{p}_{S\Psi})^T. \quad (2.30)$$

Die gesamte Abbildung ergibt dann:

$$\mathbf{RV} \xrightarrow{S_y} \hat{\mathbf{y}} \hat{=} \mathbf{RV} \xrightarrow{S_{RV_1}} \mathbf{BS} \xrightarrow{S_{RV_2}} \mathbf{BS}^* \xrightarrow{S_{RV_3}} \mathbf{\Psi} \xrightarrow{S_{\Psi_1}} \mathbf{Y}_{ZR} \xrightarrow{S_{\Psi_2}} \mathbf{x} \xrightarrow{S_{\Psi_3}} \mathbf{x}|\mathcal{I} \xrightarrow{S_{\Psi_4}} \mathbf{w} \xrightarrow{S_{\Psi_5}} \hat{\mathbf{y}} \quad (2.31)$$

$$\mathbf{RV} \xrightarrow{S_z} \hat{\mathbf{z}} \hat{=} \underbrace{\mathbf{RV} \xrightarrow{S_{RV_1}} \mathbf{BS} \xrightarrow{S_{RV_2}} \mathbf{BS}^* \xrightarrow{S_{RV_3}} \mathbf{\Psi}}_{\text{Bildverarbeitung}} \xrightarrow{S_{\Psi_1}} \underbrace{\mathbf{Y}_{ZR} \xrightarrow{S_{\Psi_2}} \mathbf{x} \xrightarrow{S_{\Psi_3}} \mathbf{x}|\mathcal{I} \xrightarrow{S_{\Psi_4}} \mathbf{w}}_{\text{Merkmalsextraktion/Klassifikation}} \xrightarrow{S_{\Psi_5}} \hat{\mathbf{z}}. \quad (2.32)$$

2.6 Anforderungsgerechte Versuchsauslegung

Nach der Formulierung der Anforderungen in Abschnitt 2.3 und dem Aufzeigen der zur Verfügung stehenden Versuchs- und Auswerteparameter besteht eine schwierige Aufgabe darin, diejenigen Parameter auszuwählen, welche die Anforderungen auch erfüllen. Daher wird, um die Struktur der Hochdurchsatz-Untersuchung den Anforderungen entsprechend zu erstellen, im folgenden Abschnitt ein Flussdiagramm zur erfolgreichen Gestaltung eines Hochdurchsatzversuchs vorgestellt, welches in Abb. 2.11 dargestellt ist. Beim Durchlaufen des Diagramms werden Schritt für Schritt Parameter festgelegt und jeweils geprüft, ob die Auslegung den Anforderungen gerecht wird.

In diesem Flussdiagramm werden vier Versuchspläne erstellt, die zur erfolgreichen Strukturierung erforderlich sind:

1. Versuchsplan für die Identifikation des biologischen Effekts, d.h. Ermittlung des Nutzsignals,
2. Versuchsplan für die Akquise des Nutzsignals, d.h. Durchführung der Messung,
3. Versuchsplan für die Extraktion und Klassifikation der Merkmale, d.h. Auswertung des Bildstroms,
4. Versuchsplan für die Durchführung der Hochdurchsatz-Untersuchung, d.h. die Erstellung des statistischen Versuchsplans.

Innerhalb der ersten drei Schritte werden, für gewählte Parameter, die in Abschnitt 2.3 formulierten Anforderungen geprüft, bevor der 4. Schritt, die Durchführung der Hochdurchsatz-Untersuchung, vollzogen werden kann. Dabei können nicht alle Anforderungen in der Reihenfolge, in der sie in den Kategorien beschrieben sind, geprüft werden. Beispielsweise ist die Prüfung auf Skalierbarkeit erst möglich, wenn alle anderen Parameter bereits gewählt sind, da jeder Abschnitt sich für die hohe Stückzahl an Einzelversuchen eignen muss. Die zu prüfenden Kategorien sind nach Abschnitt 2.3:

- Anforderungen an die Durchführung,
- Anforderungen an die Messung,

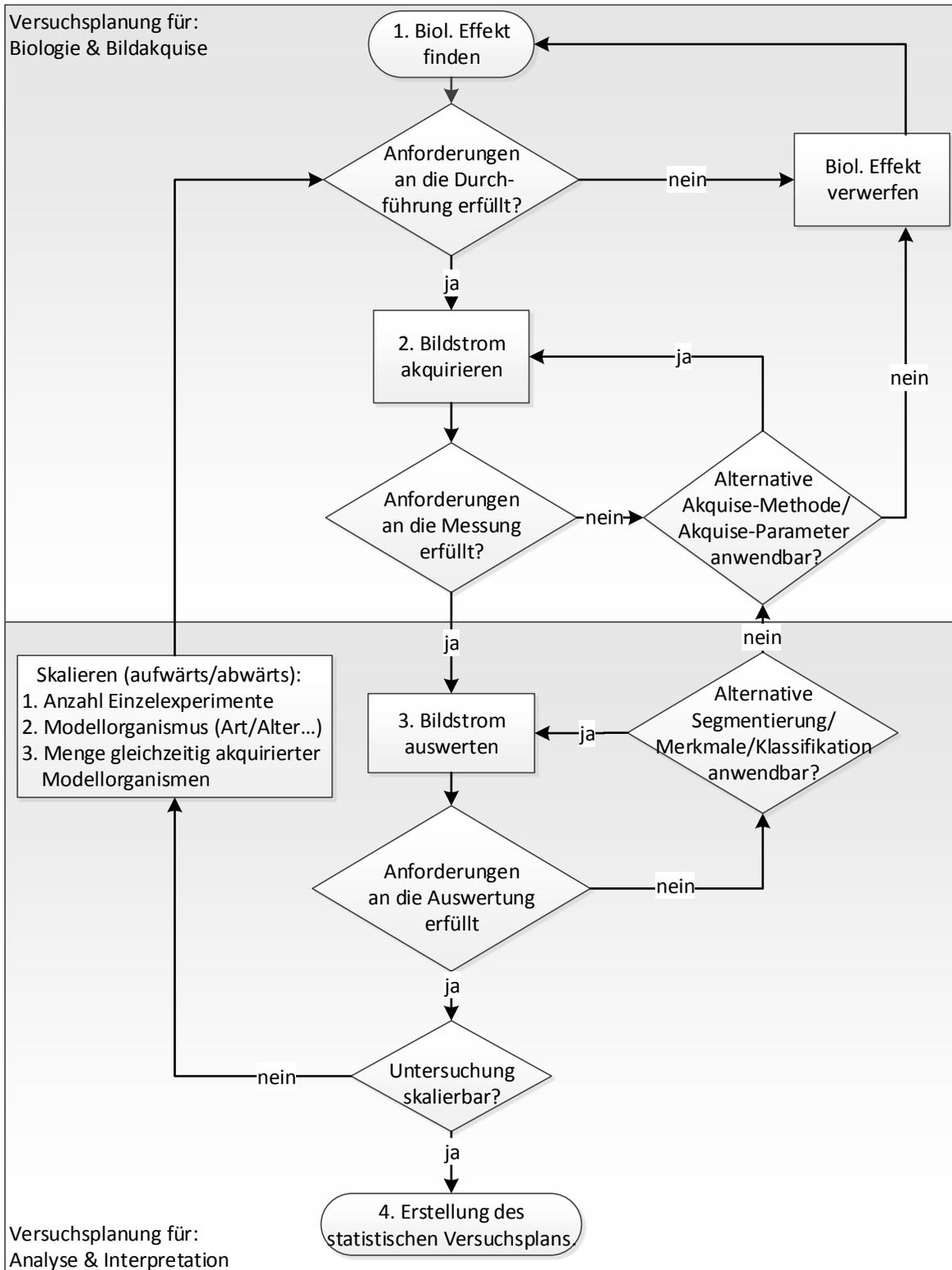


Abbildung 2.11: Flussdiagramm für die anforderungsgerechte Strukturierung des Versuchsplans von Hochdurchsatz-Untersuchungen

- Anforderungen an die Auswertung
- und Skalierbarkeit der Hochdurchsatz-Untersuchung.

Der 1. Schritt, das Finden des biologischen Effekts (des Nutzsignals), ist vollständig biologisch motiviert. Die gesetzlichen Vorschriften müssen beachtet werden, eine permanente Prüfung bei der Auswahl neuer Verfahren ist vorzusehen. Die biologische Realisierbarkeit zeigt sich im Vorversuch, d.h. der Biologe muss verständlicherweise in der Lage sein, den Effekt, der im Hochdurchsatz untersucht werden soll, beispielsweise durch eine Stimulation manuell zu erzeugen. Entscheidend für die technische Durchführbarkeit der Hochdurchsatz-Untersuchung ist, ob das Nutzsignal reproduziert werden kann. Die Reproduzierbarkeit kann durch mehrfaches Wiederholen des biologischen Versuchs geprüft werden. Ist dies nicht gegeben, so ist mit dem jeweiligen Effekt keine Hochdurchsatz-Untersuchung möglich und ein anderer, besser geeigneter Effekt muss gewählt werden. Der Schritt prüft die Parameter der beiden Arbeitsschritte „Versuchsplanung“ und „Vorbereitung“ in Abb. 2.2.

Wurde ein reproduzierbarer biologischer Effekt gefunden, wird versucht, das Nutzsignal zu messen, d.h. mittels einer der erwähnten Methoden zu akquirieren. Um den Anforderungen *Eindeutige Präsenz der biologisch relevanten Information* und *Rückwirkungsfreiheit der Messung* aus Abschnitt 2.3 zu entsprechen, muss die gewählte Methode dabei:

- das Nutzsignal robust abbilden,
- ein gutes Nutzsignal-Rausch-Verhältnis aufweisen,
- die Probe wenig beeinflussen,
- dem Abtasttheorem gerecht werden.

Die Prüfung kann weitgehend durch einfaches Ansehen der Bilder erfolgen. Sind die auszuwertenden Bereiche beispielsweise ein Gewebe in der Larve, so ist auf einen guten Kontrast dieser Bereiche zu achten. Beim Überprüfen der gewählten Abtastfrequenz und bei mit niedriger Frequenz auftretenden Nutzsignalen muss deren Frequenz zuvor mittels entsprechender Messeinrichtungen ermittelt werden. Eine möglichst vollständige Repräsentation des Nutzsignals im akquirierten Bildstrom muss angestrebt werden. Die große Bandbreite an möglichen biologischen Effekten und damit Nutzsignalen lässt eine allgemein gültige Prüfmethode nicht zu. Die Erfahrung zeigt jedoch, dass sich Daten qualitativ für eine automatische Auswertung eignen, wenn es einem ungeübten menschlichen Betrachter möglich ist, das Nutzsignal im Bildstrom leicht zu identifizieren und ggf. zu markieren. Andernfalls muss eine alternative Methode zur Bildakquise angewandt werden oder der biologische Effekt muss verworfen werden (vgl. Abb. 2.11). Die Akquise findet sich in Abb. 2.2 im Block „Mikroskopie“. Das Ergebnis der Akquise ist der Bildstrom.

Ist das Nutzsignal robust im Bildstrom enthalten, besteht der dritte Schritt darin, Methoden der Bildverarbeitung anzuwenden und die Anforderungen an die Auswertung zu prüfen. Hier zeigt sich, ob die Präsenz des Nutzsignals mit der Akquise-Methode bzw. den gewählten Akquise-Parametern des 2. Schritts ausreichend ist. Es ist zu prüfen, ob:

- das Nutzsignal durch die Segmentierung von anderen Informationen im Datensatz getrennt werden kann,
- die Segmentierung ausreichend robust gegenüber Helligkeit, Reflexionen, Lage des Objekts u.a. ist,
- der Berechnungsaufwand vertretbar ist.

Nach der Extraktion von Merkmalen aus dem Bildstrom sind die Merkmale ebenfalls auf das Erreichen der Anforderungen aus Abschnitt 2.3 zu prüfen. Merkmale müssen zwar, je nach Nutzsignal der Hochdurchsatz-Untersuchung, unterschiedlichen Bedingungen gerecht werden, dennoch lassen sich aus der Praxis allgemein gültige Prüfungen formulieren:

1. Vom Nutzsignal unabhängige Größen (wie etwa die Beleuchtung) dürfen die Merkmale nicht oder nur unwesentlich beeinflussen.
2. Der Berechnungsaufwand soll möglichst gering sein.
3. Die multivariate Trenngüte muss gewährleistet sein, d.h. die Kombination der Merkmale muss eine gute Unterscheidung der Klassen des Nutzsignals erlauben.
4. Die Ausprägungen der Merkmale müssen möglichst robust sein und daher eine geringe Empfindlichkeit gegenüber Rauschen, Messfehlern oder zeitvarianten Änderungen aufweisen.
5. Das Ergebnis muss sich übersichtlich präsentieren lassen.

Sollte einer der genannten Punkte nicht erfüllt sein, muss, wie in Abbildung 2.11 ablesbar, geprüft werden, ob mittels einer anderen Akquise-Methode aussagekräftige Merkmale extrahiert werden können. Falls nicht, so muss der biologische Effekt verworfen werden.

Nach erfolgreicher Merkmalsextraktion ist das Nutzsignal durch Zahlenwerte abgebildet und es muss abschließend geprüft werden, ob der Umfang der Hochdurchsatz-Untersuchung mit der ausgearbeiteten Lösung durchgeführt werden kann. Auf der Basis der bis zum jetzigen Zeitpunkt verfügbaren Informationen kann eine Hochrechnung für die wichtigsten Rahmenbedingungen für den Versuch erfolgen. Zu prüfende Werte sind die Akquise-Dauer, der notwendige Speicherplatz, die notwendige Rechenkapazität und zur Verfügung stehende Computer-Hardware, die zur Verfügung stehenden Modellorganismen, der Durchsatz der Mikroskop-Plattformen, die Arbeitskraft für die manuell durchzuführenden Schritte sowie die Präsentierbarkeit der Ergebnisse. Die berechneten Umfänge müssen mit den zur

Verfügung stehenden Mitteln verglichen werden. Ist die Untersuchung nicht durchführbar, müssen Parameter skaliert werden. Eine Möglichkeit bietet z.B. das Anpassen der Anzahl von Einzelexperimenten. Bezüglich des Modellorganismus kann auf einen größeren oder kleineren Organismus ausgewichen werden. Beim ZebraBärbling lässt sich z.B. die Größe durch das Alter skalieren, d.h. durch das Verwenden jüngerer oder älterer Fische bzw. Larven. Auch die Anzahl gleichzeitig akquirierter Modellorganismen kann variiert werden. Ebenso ist das Verwenden von mehreren Mikroskopen zur Erhöhung des Durchsatzes möglich.

Wichtig bei der Skalierung ist, dass sich bei Anwendung einer der aufgezeigten Lösungsvorschläge (oder auch anderer) eine Veränderung auf jeden Schritt der Untersuchung auswirkt, wie im Konzept in Abschnitt 2.2 erarbeitet. Daher muss nach jeder Iteration das *gesamte* Flussdiagramm, mit den genannten Prüfungen, nochmals durchlaufen werden. Erst wenn die Hochdurchsatz-Untersuchung in der gewünschten Weise ohne weitere Anpassungen durchführbar ist, darf mit der tatsächlichen Ausführung des Experiments begonnen werden. Ein typischer Fehler ist es, während der Durchführung den Prozess, z.B. durch Parallelisierung, ohne erneute Prüfung zu beschleunigen. Ein solches Vorgehen führt oftmals zu Störfaktoren, deren Einfluss sich im Nachhinein nur schwer beseitigen lässt oder gar zum Scheitern der Hochdurchsatz-Untersuchung führt.

2.7 Bewertung

Das Ziel des in Kapitel 2 vorgestellten neuen Konzeptes ist es, den Erfolg einer automatisierten Auswertung von Hochdurchsatz-Untersuchungen sicherzustellen, was durch eine gesamtheitliche Betrachtung des Versuchs bei der Auswahl der Versuchs- und Auswerteparameter erreicht wird. Hierfür wurde ein breites Spektrum solcher Parameter identifiziert, eine allgemein anwendbare formale Notation eingeführt und hieraus wurden möglichst allgemein formulierte Anforderungen für eine erfolgreiche Versuchsauslegung abgeleitet. Mittels des abschließend vorgestellten Flussdiagramms wird eine strukturierte Vorgehensweise zur Umsetzung des Konzeptes unter Einhaltung der Anforderungen für die Praxis geliefert. Im folgenden Kapitel wird ein Modulkatalog vorgestellt, der für die Umsetzung des vorgeschlagenen Konzeptes die notwendigen Werkzeuge enthält.

3 Modulkatalog für die Auswertung und Präsentation von Hochdurchsatz-Untersuchungen am Zebrabärbling

3.1 Einführung

Das in der Kapitel 1 aufgezeigte breit gefächerte Anwendungsfeld der Hochdurchsatz-Untersuchungen wird über Jahre hinweg neue Nutzsignale am Zebrabärbling hervorbringen, für die eine bildbasierte Hochdurchsatz-Untersuchung eine geeignete und schnelle Analysemethode ist [99]. In allen Fällen wird nach der Bildakquise ein Bildstrom vorliegen, der analysiert und interpretiert werden muss. Jedes weitere Nutzsignal ist eine neue Herausforderung an jeden Teilbereich der Auswertung, was eine universell anwendbare Lösung für alle bekannten wie zukünftigen Fragestellungen ausschließt. Die Analyse der in der Literatur zu findenden Hochdurchsatz-Untersuchungen sowie die Erfahrung aus den für die vorliegende Arbeit durchgeführten Untersuchungen (vgl. Kapitel 5) zeigen jedoch, dass bei Hochdurchsatz-Untersuchungen von biologischen Datensätzen verschiedene Methoden sowie Möglichkeiten der Vereinfachung aus Sicht der Bildverarbeitung wiederholt vorkommen. Der hohe Durchsatz ermöglicht es, auf aufwändige Auswertungen schlechter oder schwieriger Daten von Einzelversuchen zu verzichten, solche Daten zu verwerfen und mit der Auswertung der übrigen Einzelversuche fortzufahren. Eine kurze Auswertedauer und ein gutes, übersichtliches Datenmanagement sowie die nachvollziehbare Darstellung des Lösungsweges sind von größerer Bedeutung als die perfekte Analyse des Einzelversuches. Durch die massenhafte Präparation und Akquise ist zudem eine hohe Schwankung der Qualität der Daten unumgänglich. Die Schwankung resultiert u.a. aus unterschiedlichen Füllhöhen der Näpfchen mit Flüssigkeit, schwankender Beleuchtung, schwankendem Alter der Larven und Bewegungen der Larven während der Aufnahme (vgl. Abschnitt 2.4.3).

Trotz der genannten Schwierigkeiten ist es für jede Hochdurchsatz-Untersuchung möglich, die Auswertungsschritte, die am Bildstrom vollzogen werden, in wenige Kategorien zusammenzufassen. Für jede der Kategorien schlägt die vorliegende Arbeit eine Reihe von Lösungen vor. Die Lösungen bestehen für jede Kategorie aus zu Modulen zusammengefassten Methoden, welche in den folgenden Abschnitten vorgestellt werden. Alle Module

einer Kategorie haben gemeinsam, dass sie die gleichen Ein- und Ausgangsgrößen besitzen. Für die jeweils vorliegende biologische Fragestellung müssen aus jeder Kategorie ein oder mehrere passende Module gewählt und spezifiziert werden. Die vorkommenden Datenverarbeitungsmethoden wurden entweder in der vorliegenden Arbeit entwickelt, wie z.B. die neuartige Normalisierung für Hochdurchsatz-Untersuchungen (Abschnitt 3.2.2), die Trennung von Einzelobjekt (Fischei) vom Hintergrund (Abschnitt 3.3) und das neue Verfahren zum Tracking des Chorions (Abschnitt 3.3.3) oder es handelt sich um bereits bekannte Methoden, die für die Hochdurchsatz-Untersuchung modifiziert wurden. In Abb. 3.1 ist der Modulkatalog visualisiert und die neu entwickelten Module sind rot hervorgehoben. Das vorliegende Kapitel stellt alle Module jeder Kategorie zu einem Modulkatalog zusammen. Der Vorteil des Kataloges ist, dass es mit dessen Hilfe möglich wird, einen Großteil der bekannten als auch zukünftigen Problemstellungen von bildbasierten Hochdurchsatz-Untersuchungen ohne großen Entwicklungsaufwand zu lösen. Lediglich die Auswahl geeigneter Module und die Anpassung der jeweiligen Auswerteparameter auf die vorliegende Fragestellung sind durch den Anwender durchzuführen.

Der Ablauf der Datenverarbeitung, die Analyse und Interpretation des im Bildstrom enthaltenen Nutzsignals, lässt sich gemäß Abb. 3.1 in sieben Kategorien aufspalten, auf die in den folgenden Abschnitten eingegangen wird. Die Entwurfsphase der Auswertung verläuft nicht schematisch, sondern ist ein iterativer Prozess, wie in Abb. 2.11 vorgestellt wurde. Dabei werden anhand der Zwischenergebnisse Module, Auswerteparameter und Bewertungsmaße variiert. Die Bewertungsmaße sind z.B. die Güte der Klassifikationen oder die visuelle Kontrolle der Ergebnisse der Segmentierung. Sie nehmen Einfluss auf die genannte Variation und basieren auf den Zwischen- und Endergebnissen der Blöcke der Modulkategorien. Sie helfen bei der Findung der am besten geeigneten Struktur und Auswerteparameter sowie bei der Einschätzung der Qualität der Lösung [110].

Zum besseren Verständnis der Module werden einige durch konkrete Beispiele veranschaulicht. Bei der Vorstellung der Module wird zudem bei Beispielen immer versucht, die Beispiele und Datensätze möglichst einfach zu halten. Bildverarbeitungsmodule werden beispielsweise bevorzugt anhand von einzelnen Grauwertbildern entsprechend Formel (1.1) beschrieben, auch wenn die Module in der Praxis auf den wesentlich komplexeren Bildstrom nach Formel (2.15) angewandt werden. Sollte die Anwendung auf einen Datensatz höherer Dimension nicht möglich sein, wird an der entsprechenden Stelle darauf hingewiesen.

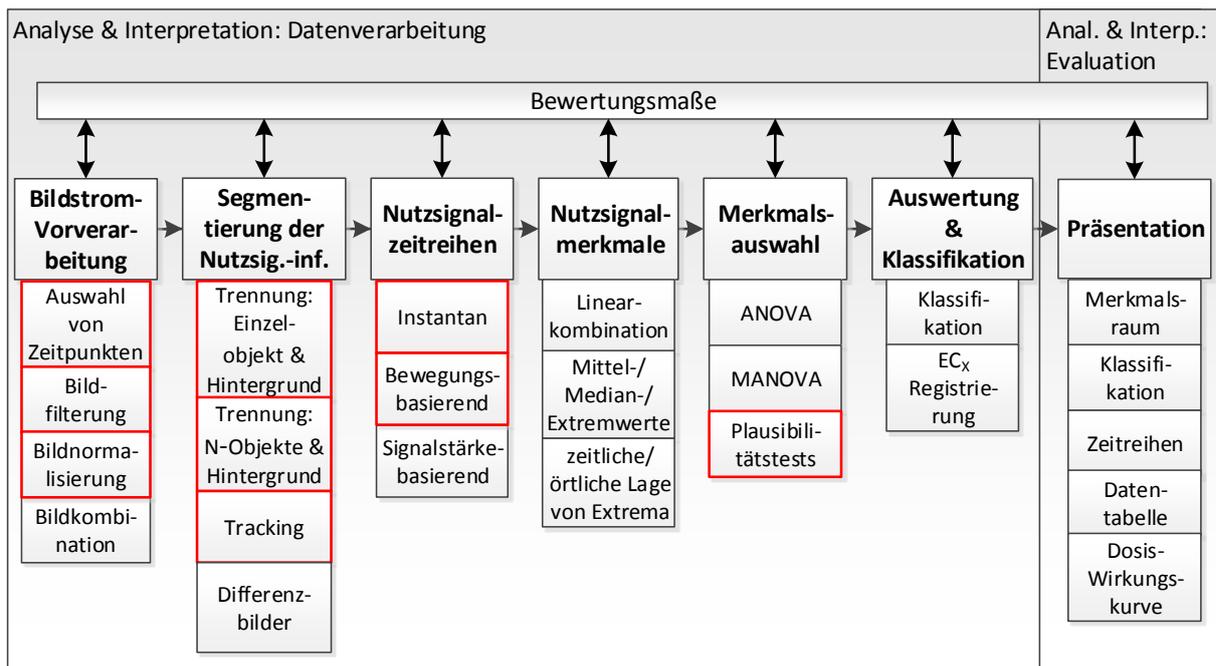


Abbildung 3.1: Modulkatalog mit Kategorien und den übergeordneten Bewertungsmaßen. In der vorliegenden Arbeit neu entwickelte Module sind rot gekennzeichnet.

3.2 Bildstrom-Vorverarbeitung

Notwendige Voraussetzung für eine erfolgreiche Auswertung ist ein Bildstrom, dessen Bild-daten ein konsistentes und ausreichend starkes Nutzsignal enthalten. Durch den Einfluss der vielen bereits in Abschnitt 2.4 diskutierten Störungen ist dies in realen Problemstellungen oft nicht unmittelbar gegeben. Des Weiteren müssen die Rohdaten bei Schicht-aufnahmen erst zu einem verarbeitbaren Datensatz zusammengesetzt werden. Die Bild-Vorverarbeitung hat zum Ziel, den Bildstrom für die weitere Verarbeitung aufzubereiten bzw. zu verbessern.

Eingang der Modulkategorie *Bildstrom-Vorverarbeitung* sind die Rohdaten, d.h. der un-verarbeitete Bildstrom **BS**, wie er bei der Akquise aufgezeichnet wurde. Als Ausgang steht nach erfolgreicher Verarbeitung ein verbesserter Bildstrom **BS*** zur Verfügung, der sich für die Segmentierung eignet und dessen Dimension so groß wie nötig, jedoch so gering wie möglich ist.

3.2.1 Auswahl von Zeitpunkten

Der erste Unterpunkt der Bildstrom-Vorverarbeitung in Abb. 3.1, die Auswahl von Zeitpunkten, steht für das Auswählen nur bestimmter Aufnahmen zu bestimmten Zeitpunkten aus dem Bildstrom. Er stellt eine Dimensionsreduktion des Bildstroms **BS** dar. Hierbei

wird ein großer Bildstrom z.B. in mehrere Teile unterteilt oder auf der Basis von Vorwissen oder durch Sichtprüfung auf einen wesentlichen Teil beschränkt. Die Methode kommt vor allem dann zum Einsatz, wenn der Bildstrom zur Auswertung mehrerer Nutzsignale akquiriert wurde und der auszuwertende Planfaktor im Bildstrom redundant vorhanden ist. Ein Beispiel sind Bewegungssequenzen, die zur Detektion von Herzschlägen aufgezeichnet wurden. Mit einem solchen Datensatz lassen sich neben den Bewegungen auch instantane Nutzsignale wie z.B. die entwickelte Größe bestimmen. Für solche Nutzsignale ist aus der Sequenz bereits ein einziges Bild ausreichend, daher können die anderen zeitlichen Daten bei der Analyse ausgeschlossen werden. Ein Auswerten aller Bilder des Bildstroms würde wenig neue Information zur Klasseneinteilung beitragen. In der Praxis zeigt sich zudem, dass manche Detektoren für die Erfassung von Bildern eine kurze Zeit benötigen, um die Betriebstemperatur zu erreichen, was zur Folge hat, dass die ersten Aufnahmen im Bildstrom z.T. verfälscht dargestellt werden. Daher ist es ratsam, bei der Auswahl nicht intuitiv die erste(n) Aufnahme(n), sondern zeitlich später aufgezeichnete Bilder zu wählen, oder, wenn das Nutzsignal es zulässt, die Auswahl der Bilder nach einer festgesetzten Vorschrift zu vollziehen. Die Auswahl der Bilder erfolgt somit zumeist nach einer manuell festgelegten Liste oder nach einer automatischen Auswahlvorschrift, wie z.B. der Ermittlung der schärfsten Schichtaufnahme mittels Wavelet-Transformation [72].

In jedem Fall wird eine Indexmenge \mathcal{I}_{BV} bestimmt, die aus dem Bildstrom nur bestimmte Bilder auswählt. Das „Weglassen“ der restlichen Bilder ist hier eine einfache Möglichkeit zur Verringerung von Dimension und Größe des Datensatzes. Es sei jedoch darauf hingewiesen, dass der Schritt der Bildauswahl bei nicht sorgfältiger Auswahl die Messgenauigkeit verringert. Die Auswahl muss daher immer mithilfe von Vorversuchen und Bewertungsmaßen überprüft werden.

3.2.2 Neues Verfahren zur Bildfilterung und Normalisierung für inhomogene Datensätze

Einige Einflüsse von Störfaktoren bei Hochdurchsatz-Untersuchungen lassen sich mit vergleichsweise wenig A-priori-Wissen durch Methoden der Bildvorverarbeitung beseitigen oder abschwächen. Die Bildfilterung bezeichnet lineare und nichtlineare Punktoperationen wie Kontrasterhöhung, Registrierung, Korrektur von Beleuchtungseffekten, Rauschunterdrückung und Reduzierung von Artefakten [25,51]. Auch Operationen zur Transformation, Änderung der Farbtiefe oder die Skalierung z.B. mittels sog. Bild- oder Gauß-Pyramiden fallen darunter [27]. Sie transformieren den Bildstrom wie in Formel (2.17) allgemein beschrieben. Das Ergebnis der genannten Methoden sind Bilder, welche sich im Vergleich zu den Rohdaten besser für alle nachfolgenden Schritte der Auswertung des Nutzsignals eignen.

Fast immer ist eine Bildnormalisierung notwendig, die als eine besondere Form von Filterung betrachtet werden kann. Gemäß Abb. 2.1a wirken Störgrößen direkt auf Biologie und Bildakquise. Solche Störungen sind z.B. das Aufzeichnen der Bilddaten zu unter-

schiedlichen Tageszeiten und Temperaturen, durch unterschiedliche Laboranten (z.B. mit Unterschieden in der Vorgehensweise bei nicht genau spezifizierten Versuchsprotokollen), mit geänderter Beleuchtung usw. Die Absolutwerte der für die Auswertung und Klassifikation aus dem Bildstrom extrahierten Merkmalszeitreihen (vgl. Abschnitt 3.4) hängen somit nicht, wie gewünscht, lediglich von der Nutzinformation ab, sondern ebenso von den genannten Störgrößen. Data-Mining Methoden zur Merkmalsauswahl sind somit mit hoher Wahrscheinlichkeit in der Lage, Merkmale zu identifizieren, die auf signifikante Unterschiede im Nutzsignal hindeuten. Die Ursache dieser Signifikanz ist jedoch nicht zwingend im Nutzsignal begründet, was wünschenswert ist, sondern oftmals durch die Störgrößen verursacht. Wird auf eine Normierung gänzlich verzichtet, können Einflüsse der Störgrößen ein Auswerten der Daten verschlechtern oder scheitern lassen.

Als Beispiel sei eine typische Aufnahme des Hochdurchsatzmikroskops Olympus-Scan^R angeführt. Die Intensitätswerte konzentrieren sich lediglich auf einen kleinen Wertebereich (vgl. Abb. 3.2), was zur Folge hat, dass eine Vielzahl üblicher Algorithmen wie etwa adaptive Schwellenwertverfahren [119, 149] oder Regionen-Wachstum-Verfahren [2] aufgrund des schlechten relativen Kontrastes scheitern. In einem solchen Fall wird die Verarbeitungskette aufgrund von Validitätsprüfungen abgebrochen, oder, sollten diese fehlen, werden sogar gänzlich falsche Werte zur Klassifikation herangezogen. Derartig falsche Merkmalswerte führen im ungünstigsten Fall zu systematischen Fehlern und Fehlklassifikationen.

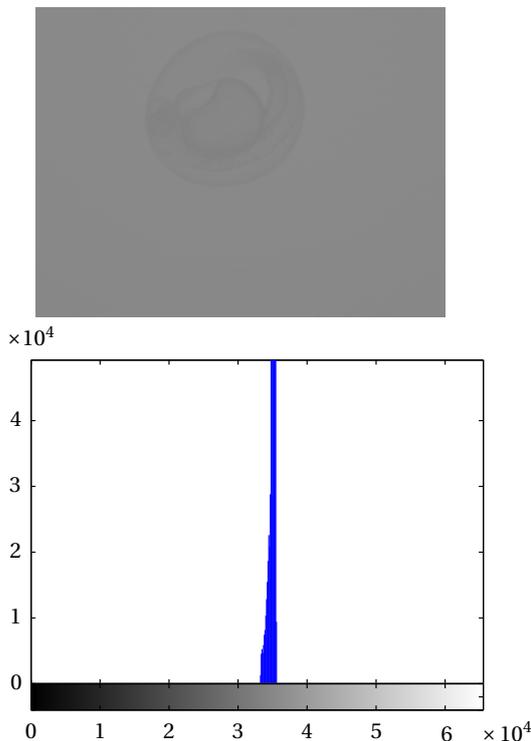


Abbildung 3.2: Bild und Histogramm der Rohdaten des Olympus-Scan^R-Hochdurchsatzmikroskops

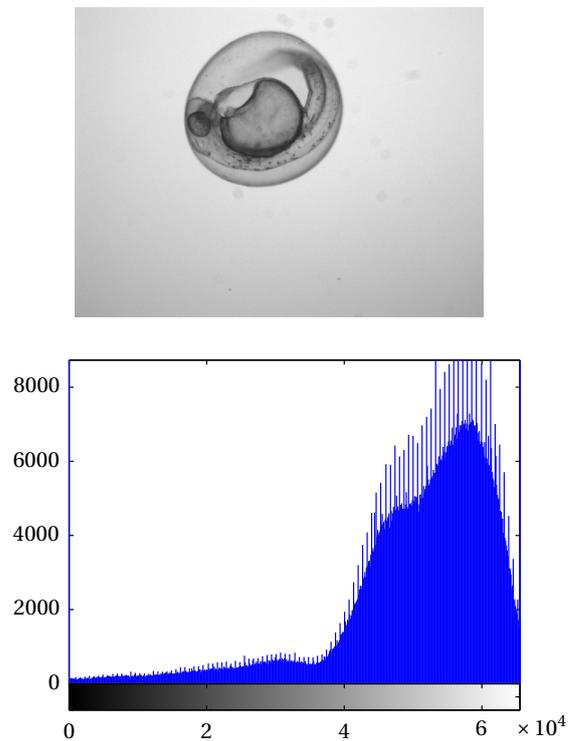


Abbildung 3.3: Bild und Histogramm der normalisierten Rohdaten (gleiche Rohdaten wie Abb. 3.2) gemäß Formel (3.4) mit den Parametern $I_{\text{hoch}}=98\%$ und $I_{\text{tief}}=2\%$.

Die Störgrößen sind hier u.a. das nicht vollständige Ausnutzen des zur Verfügung stehenden Wertebereichs, was zu einer schlechten Bilddynamik führt sowie abweichende Mikroskop-Einstellungen, das Rauschen des Detektors etc. Ein Unterdrücken des Einflusses der Störgrößen kann durch eine Korrektur der Intensitätswerte erfolgen. Hierbei wird jeder Pixel mittels des Parameters q zentriert und durch einen die Verteilung beschreibenden Parameter ν dividiert. Für ein zweidimensionales Bild gilt:

$$I_{i_x i_y}^* = \frac{I_{i_x i_y} - q}{\nu}. \quad (3.1)$$

Für die Parameter q und ν eignen sich z.B. statistische Parameter wie Mittelwert oder Varianz der Intensitätswerte

$$q_1 = \frac{1}{I_x I_y} \sum_{i_x=1}^{I_x} \sum_{i_y=1}^{I_y} I_{i_x i_y}, \quad \nu_1 = \sqrt{\frac{1}{I_x I_y - 1} \sum_{i_x=1}^{I_x} \sum_{i_y=1}^{I_y} (I_{i_x i_y} - q)^2}. \quad (3.2)$$

oder auch Quantile und Extremwerte

$$q_2 = \min_{i_x=1 \dots I_x} \left(\min_{i_y=1 \dots I_y} (I_{i_x i_y}) \right), \quad \nu_2 = \max_{i_x=1 \dots I_x} \left(\max_{i_y=1 \dots I_y} (I_{i_x i_y}) \right) - \min_{i_x=1 \dots I_x} \left(\min_{i_y=1 \dots I_y} (I_{i_x i_y}) \right). \quad (3.3)$$

Durch das Einführen der Forderung einer prozentualen Mindestanzahl an gesättigten Pixeln lässt sich die Robustheit der Normierung mittels Gleichung (3.4) deutlich verbessern. Sinnvolle Werte für I_{tief} sind 2 Prozent und I_{hoch} 98 Prozent:

$$I_{i_x i_y}^* = \begin{cases} 0 & \text{für } I_{i_x i_y} \leq I_{\text{tief}} \\ \frac{I_{i_x i_y} - I_{\text{tief}}}{I_{\text{hoch}} - I_{\text{tief}}} & \text{für } I_{\text{tief}} < I_{i_x i_y} < I_{\text{hoch}} \\ 1 & \text{für } I_{i_x i_y} \geq I_{\text{hoch}} \end{cases} \quad (3.4)$$

Das vorgestellte Verfahren wurde nach Formel (3.4) auf das Bild in Abb. 3.2 angewandt. Das Ergebnis ist in Abb. 3.3 dargestellt. Im Originalbild wird lediglich ein kleiner Bereich der möglichen Grauwerte ausgenutzt, wie im Histogramm zu erkennen ist (Abb. 3.2, unten). Auch ist die Larve im Bild fast nicht zu erkennen. Nach Anwendung des Verfahrens sind die Grauwerte über den gesamten Wertebereich verteilt und die Kontraste gut ausgebildet. Ohne weitere Validitätsprüfung birgt die Methode allerdings die Gefahr, dass lediglich das Bildrauschen auf den gesamten Wertebereich gestreckt wird. Der genannte Fall tritt z.B. ein, wenn bei der Präparation versehentlich ein Näpfchen nicht befüllt wurde. Solche Bilder lieferten in der weiteren Verarbeitung scheinbar sinnvolle Werte und können somit die Versuchsergebnisse stark verfälschen. Abhilfe schaffen hier Validitätsprüfungen, die über den gesamten Bildinhalt Auskunft liefern und fehlerhafte Bilder aussortieren. Eine einfache und robuste Prüfung ist beispielsweise der Mittelwert aller Intensitätswerte. Für ein leeres Well (*Well* = Näpfchen in einer Kunststoffplatte vgl. Abschnitt 1.2.1.1) fällt dieser aufgrund der fehlenden Abschattung durch das Präparat deutlich höher aus als mit Präparat.

Die Methode korrigiert jedoch keine Faktoren, die sich direkt auf den Versuch beziehen, wie z.B. ein unterschiedliches Alter der Fische oder einen unterschiedlichen Mikroskoptyp.

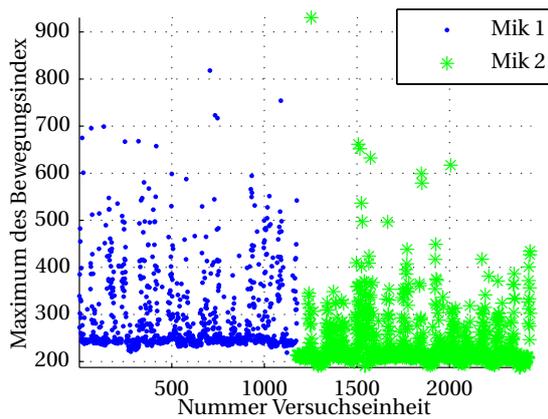


Abbildung 3.4: Verteilung der Merkmalswerte in einer Hochdurchsatz-Untersuchung zur Ermittlung der akuten Toxikologie an zwei baugleichen Mikroskopen vom Typ Olympus-Scan^R

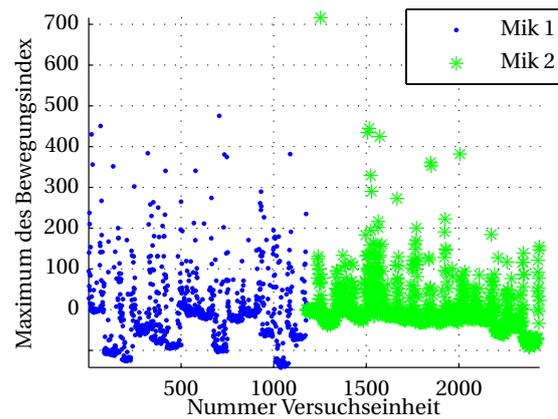


Abbildung 3.5: Scatterplot der mittels der Kontrollen aus Abb. 3.4 normierten Merkmalswerte

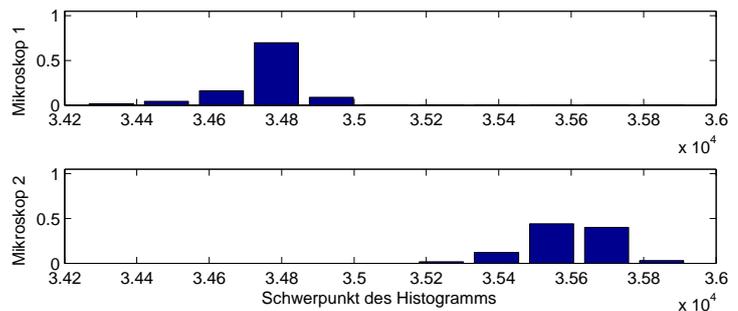
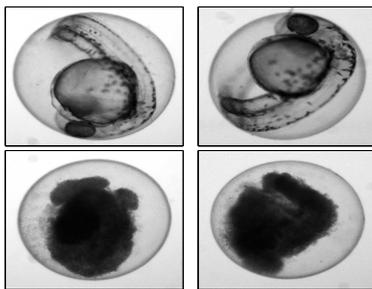


Abbildung 3.6: a (links): Mikroskopbilder mit verschiedenen Planfaktoren (*lebendig* in Zeile 1, *koaguliert* in Zeile 2) und Störfaktoren (*Mikroskop 1* in Spalte 1, *Mikroskop 2* in Spalte 2). b (rechts): Schwerpunkt des Histogramms aller Bilder zweier Mikroskope des in Abbildung 3.4 gezeigten Beispiels [139]. Obwohl die Aufnahmen in 3.6a, spaltenweise verglichen, optisch gleich wirken überlappen die Histogramme des dargestellten Merkmals nicht.

Als Beispiel sei hier die Hochdurchsatz-Untersuchung zur Ermittlung der akuten Toxikologie gewählt [122]. In Abb. 3.4 ist ein Scatterplot einer Hochdurchsatz-Untersuchung von 2436 Versuchseinheiten (hier jeweils ein Zebrafärbling) aufgetragen, welche an zwei unterschiedlichen Mikroskopen aufgezeichnet wurde. Ein typisches Merkmal zur Klassifikation der Bewegung wurde ausgewählt (vgl. hierzu Abschnitt 3.4.2). Der Einfluss der beiden Mikroskope ist deutlich, durch einen Offset der Merkmalswerte zueinander zu erkennen. Solche Einflüsse, wie sie in diesem Beispiel bereits zwischen zwei Mikroskopen des gleichen Typs entstehen, sind zwischen unterschiedlichen Laboren ungleich größer.

Um Versuche auch zwischen unterschiedlichen Laboren, Mikroskopen und Tagen vergleichbar zu machen, muss der Bildstrom durch auf den Versuch bezogene Parameter normalisiert werden. Hier ist meist das Normalisieren mittels der Negativ- oder Positiv-Kontrollen erfolgreich [101]. Im gezeigten Beispiel scheitert die Methode jedoch ebenfalls,

da auch die Merkmalswerte in den Positiv-Kontrollen schwanken. Kontrollen sollten zwar ein vergleichbares Verhalten zeigen, was hier jedoch nicht der Fall ist, wie Abb. 3.5 zeigt. In der Abbildung wurden alle Merkmalswerte mittels der Kontrollen normiert. Insbesondere bei den Merkmalswerten von Mikroskop 1 (blau dargestellt) streuen nach der Normierung die Merkmalswerte stärker als zuvor. Beispielsweise haben die ersten 84 Versuchseinheiten (96 Versuche auf einer Platte davon 12 Kontrollen zur Normalisierung) nach der Normalisierung Merkmalswerte welche minimal etwa bei 0 liegen. Die nächste Platte (Nummer Versuchseinheit 84 bis 168) zeigt Minimalwerte unter -100. Klarer wird die unterschiedliche Verteilung der Merkmalswerte durch den Einfluss der Mikroskope in Abbildung 3.6, in der diese anhand eines Histogramms veranschaulicht wird. Die hier dargestellten Larven sind sich in beiden Aufnahmen (Reihenweise verglichen) sehr ähnlich, die Merkmalswerte (hier Schwerpunkt des Histogramms, x_1 vgl. Abschnitt 3.4.1) zeigen jedoch erhebliche Unterschiede. In einem konsistenten und normalisierten Versuch sollte sich wenig Unterschied bei der Merkmalsverteilung der gleichen Planfaktoren von vergleichbaren Versuchseinheiten zeigen, selbst wenn sie von unterschiedlichen Mikroskopen stammen. Bisherige Lösungen müssen einen solchen Datensatz verwerfen.

Deshalb wurde im Rahmen der vorliegenden Arbeit eine neue Methode erarbeitet, welche in der Lage ist, ein solches Problem zu lösen und auch inhomogene Datensätze vergleichbar macht [139]. Ein Klassifikator ist umso besser, je exakter der Klassifikator in der Lage ist, die Faktoren \hat{Y} oder \hat{Z} der wahren Klasse Y oder Z zuzuordnen. Es lässt sich die Erfolgsrate für Planfaktoren ξ_y und Störfaktoren ξ_z für jede Versuchseinheit n über alle Ausgangsklassen s_y angeben durch:

$$\xi_{y,n} = \frac{1}{s_y} \sum_{i_{s_y}=1}^{s_y} \xi_{y,i_{s_y},n} \quad \text{mit} \quad \xi_{y,i_{s_y},n} = \begin{cases} 1 & \hat{y}_{i_{s_y},n} = y_{i_{s_y},n} \\ 0 & \text{sonst} \end{cases}, \quad (3.5)$$

$$\xi_{z,n} = \frac{1}{s_z} \sum_{i_{s_z}=1}^{s_z} \xi_{z,i_{s_z},n} \quad \text{mit} \quad \xi_{z,i_{s_z},n} = \begin{cases} 1 & \hat{z}_{i_{s_z},n} = z_{i_{s_z},n} \\ 0 & \text{sonst} \end{cases}. \quad (3.6)$$

Die mittlere Klassifikationsgüte $\overline{\xi_y}, \overline{\xi_z}$ wird bestimmt durch:

$$\overline{\xi_y} = \frac{1}{N} \sum_{n=1}^N \xi_{y,n}, \quad \overline{\xi_z} = \frac{1}{N} \sum_{n=1}^N \xi_{z,n}. \quad (3.7)$$

Im Falle einer perfekten Klassifikation werden alle Planfaktoren korrekt der zugehörigen Ausgangsklasse zugeordnet, während sich die Klassen aller anderen Faktoren nicht trennen lassen und daher gleichverteilt nach dem Zufallsprinzip auf die Ausgangsklassen aufgeteilt werden:

$$\overline{\xi_{y,\text{ziel}}} = 1, \quad \overline{\xi_{z,\text{ziel}}} \approx \frac{1}{s_z} \sum_{i_{s_z}=1}^{s_z} \frac{1}{m_{y,sz}}. \quad (3.8)$$

Die letzte Formel setzt allerdings voraus, dass die Merkmale von den Störfaktoren unabhängig sind, was in realen Systemen nicht der Fall ist. Daher müssen alle Parameter in \mathbf{p} so gewählt werden, dass der Einfluss der Störfaktoren klein wird. Dieses Problem

kann mittels Filter- oder Wrapperansätzen gelöst werden [53, 79]. Filter bestimmen die Versuchsparameter ohne Kenntnis der wahren oder geschätzten Klassenzugehörigkeit

$$\mathbf{p}^* = f(\mathbf{p}). \quad (3.9)$$

Wrapper hingegen berücksichtigen die Klassifikationsergebnisse bei der Parameteradaptation:

$$\mathbf{p}^{**} = f(\mathbf{p}, \mathbf{Y}, \mathbf{Z}, \hat{\mathbf{Y}}, \hat{\mathbf{Z}}). \quad (3.10)$$

Wrapper liefern üblicherweise bessere Ergebnisse auf Kosten eines deutlich höheren Rechenaufwands.

Die Optimierung sucht nach einer Parameterkombination, welche einen Kompromiss zwischen $\overline{\xi}_y - \overline{\xi}_{y,\text{ziel}}$ und $\overline{\xi}_z - \overline{\xi}_{z,\text{ziel}}$ sucht.

Ein mögliches Kriterium zur Optimierung ist

$$\min_{\mathbf{p}}(O) \quad \text{mit} \quad O = \alpha \left| \overline{\xi}_y - \overline{\xi}_{y,\text{ziel}} \right| + (1 - \alpha) \left| \overline{\xi}_z - \overline{\xi}_{z,\text{ziel}} \right|, \quad \alpha \in [0.5, 1]. \quad (3.11)$$

Der Parameter α wichtet die Klassifikationsergebnisse zwischen Plan- und Störfaktoren und muss für sinnvolle Ergebnisse mit signifikantem Einfluss der Planfaktoren größer als 0.5 gewählt werden (Bei Werten < 0.5 ist der Einfluss der Störfaktoren auf das Optimierungskriterium größer als der Einfluss der Planfaktoren). Für den Fall $\alpha = 1$ wird die Optimierung zu einem gewöhnlichen Wrapper [140]. Es lassen sich Lösungsansätze sowohl mit Wrappern als auch Filtern realisieren oder auch durch eine Kombination beider. Eine Kombination kann etwa die Optimierung eines normalisierten Bildstroms BS^* mittels eines Wrappers sein.

Für den Fall, dass alle Merkmalswerte zur gleichen Klasse der Störfaktoren gehören, liefern konventionelle Klassifikatoren in einer Kreuzvalidierung gute Ergebnisse, auch für Testdaten [44, 78]. Gehören die Testdaten jedoch einer anderen Klasse der Störfaktoren an, welche bei der Kreuzvalidierung nicht vertreten war, kann das tatsächliche Ergebnis deutlich schlechter ausfallen. Daher ist die Kreuzvalidierung kein geeignetes Mittel, die Validität der Klassifikation zu überprüfen. Die besten Ergebnisse lassen sich erzielen, wenn in den Trainings- und Testdaten alle Klassen der Stör- und Planfaktoren vertreten sind.

Den Erfolg der Methode verdeutlichen die Abbildungen 3.7 und 3.8 am Beispiel der Merkmalswerte aus Abbildung 3.6. Die Datenpunkte in Abbildung 3.7 welche mit 1 gekennzeichnet sind gehören der Klasse *koaguliert* an, Datenpunkte mit 2 gehören der Klasse *lebendig* an. Datenpunkte von Mikroskop 1 sind rot, Datenpunkte von Mikroskop 2 sind grün dargestellt. Der Einfluss des Störfaktors *Mikroskop* ist deutlich zu erkennen. Obwohl die Merkmalswerte beider Mikroskope Versuchseinheiten mit ähnlichem Status abbilden, überschneidet sich die Streuung der Merkmalswerte nicht. In Abbildung 3.8 schließlich sind die gleichen, jedoch mittels des Wrapperansatzes normalisierten Daten dargestellt. Ebenfalls gehören die mit 1 gekennzeichneten Datenpunkte der Klasse *koaguliert* an und

die mit 2 gekennzeichneten Datenpunkte der Klasse *lebendig* an. Datenpunkte von Mikroskop 1 sind wie zuvor rot und Datenpunkte von Mikroskop 2 grün gekennzeichnet. Nach der Normalisierung ist eine Unterscheidung der Mikroskope, wie gewollt, nicht mehr unmittelbar möglich. Es lässt sich lediglich der Planfaktor *Koaguliertheit* klar abgrenzen. Damit sind die zuvor sowohl durch die Plan- als auch durch die Störfaktoren trennbaren Merkmalsräume nach der Normalisierung nur noch durch die Planfaktoren unterscheidbar und somit deutlich robuster für eine Hochdurchsatz-Untersuchung.

3.2.3 Bildkombination

Der letzte Modulblock in der Kategorie Bildstrom-Vorverarbeitung sind bildkombinierende Methoden. Solche kommen zum Einsatz, wenn bei der Bildakquise mehrere Fokusebenen oder Schichtaufnahmen akquiriert wurden, was vor allem bei Nutzsignalen der Fall ist, bei denen bestimmte Bereiche im Fisch unterschieden werden sollen. Die Bildkombination erstellt aus mehreren zweidimensionalen Bildern kombinierte Bilder und reduziert somit die Dimension. Bei Hochdurchsatz-Untersuchungen kommt vor allem die Kombination von Farbkanälen zu einem Grauwertkanal zum Einsatz, da Farbkanäle in bestimmten Fällen keine zusätzliche Information zur Nutzsignalinformation liefern. Eine ebenfalls wichtige Methode ist der Extended Focus [133]. Die Verfahren hierzu basieren auf der Annahme, dass alle zu fusionierenden Bilder den gleichen Inhalt zeigen, die Bildinformation in einer der Aufnahmen jedoch weniger gestört aufgezeichnet wurde als in anderen. Hier wird mittels eines Bewertungsmaßes in jeder Fokusaufnahme versucht, die beste Repräsentation der Szene auszuwählen und in das Ergebnisbild einzufügen. Es lassen sich damit deutliche Steigerungen der Schärfentiefe und damit des auswertbaren Bereichs innerhalb des Zebraabärblings erzielen. Eine ausführliche Beschreibung der Methoden findet sich u.a. in [25, 70, 165].

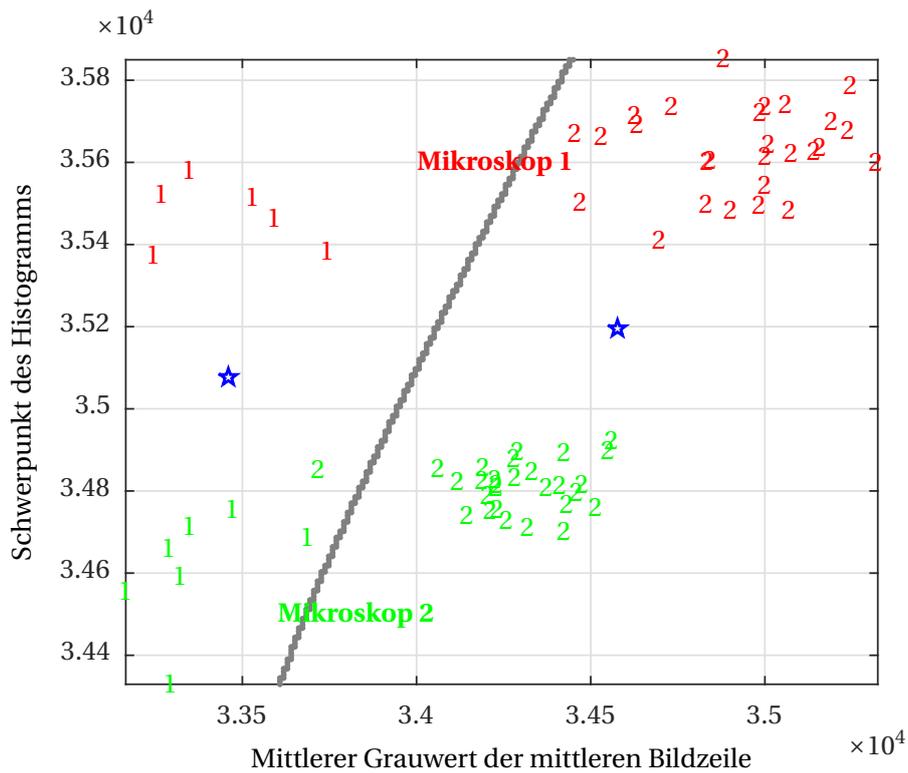


Abbildung 3.7: Merkmalsraum und Diskriminanzfunktion für zwei Platten aus Abbildung 3.4 vor Normalisierung [139].

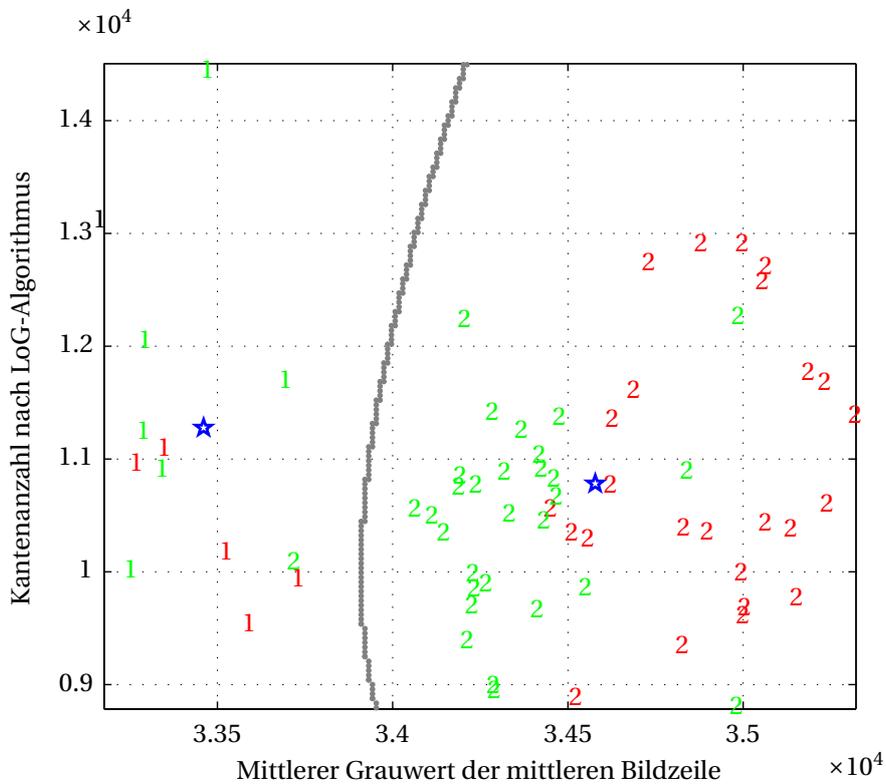


Abbildung 3.8: Merkmalsraum und Diskriminanzfunktion für zwei Platten aus Abbildung 3.4 nach der Normalisierung mittels des Wrapperansatzes [139].

3.3 Segmentierung der Nutzinformation

Um das biologische Nutzsignal einer Hochdurchsatz-Untersuchung quantifizieren zu können, muss es von allen anderen Daten im Bildstrom separiert werden. Alle hier vorgestellten Module haben als Eingang den vorverarbeiteten Bildstrom \mathbf{BS}^* und als Ausgang Merkmalsbilder Ψ . In die Kategorie *Segmentierung der Nutzinformation* gehören die folgenden drei Module.

1. Je Bildstrom ist genau eine Versuchseinheit zu jedem Zeitpunkt aufgezeichnet (Trennung Einzelobjekt und Hintergrund).
2. Im Bildstrom sind mehrere Versuchseinheiten zu jedem Zeitpunkt aufgezeichnet (Trennung mehrere Objekte und Hintergrund).
3. Zusätzlich zu 1. oder 2. ist die Versuchseinheit über einen bestimmten Zeitraum aufgezeichnet (Tracking).

Je Versuchseinheit werden Merkmalsbilder errechnet. Merkmalsbilder sind aus den Rohdaten abgeleitete Bilder wie z.B. Binärbilder (nach Anwendung eines Schwellenwertes) oder Differenzbilder von Bildsequenzen. Die Separation der Information im vorverarbeiteten Bildstrom \mathbf{BS}^* erfolgt durch die sog. Segmentierung. Die Segmentierung ist das Gruppieren zusammengehöriger Bereiche. Im Folgenden werden die Bildregionen, welche die Information des biologischen Effekts enthalten, als Vordergrund und alle anderen Regionen als Hintergrund bezeichnet. Die Segmentierung trennt somit den Vordergrund vom Hintergrund und ermöglicht so die weitere Auswertung der im Vordergrund enthaltenen Nutzsignalinformation. Alle Methoden bewirken eine Reduktion des vorverarbeiteten Bildstroms \mathbf{BS}^* auf die Nutzsignalinformation. Dabei unterscheidet sich das Vorgehen stark für die Fälle, ob mehrere Versuchseinheiten im Bild vorhanden sind bzw. sein können oder ob in der Versuchsauslegung festgelegt ist, dass lediglich eine Versuchseinheit, z.B. der Zebrafisch (inkl. Chorion), im Bildstrom vorliegt.

3.3.1 Neue Methode zur Trennung von Einzelobjekt und Hintergrund

Zur Trennung der Nutzsignalinformation einer einzelnen Versuchseinheit vom Hintergrund sind bei akzeptabler Bildqualität bereits Schwellenwertverfahren in Kombination mit morphologischen Operatoren erfolgreich [4]. Adaptive Schwellenwertverfahren wie die Methode von Otsu [124, 125] oder Fuzzy-basierende Schwellenwertverfahren [55, 61] bieten hier jedoch deutlich robustere Ergebnisse [5, 6]. Der Vorteil der genannten Verfahren ist die kurze Rechenzeit, was gerade bei Roboteranwendungen wie z.B. dem automatischen Pipettieren von Eiern wichtig ist [132]. Ein Schwellenwertverfahren trennt jedoch ohne jegliche Prüfung Vorder- von Hintergrundpixeln. Daher muss für jedes gefundene Objekt nachträg-

lich geprüft werden, ob es sich um eine Versuchseinheit oder ein Artefakt, z.B. Schmutz o.ä., handelt. Die besten Ergebnisse lassen sich unter Einbezug von Vorwissen, beispielsweise der Größe und Kontur der Nutzsignalinformation der Hochdurchsatz-Untersuchung, erzielen. Solche Informationen unterscheiden sich stark für die beiden Fälle:

1. Die Larve befindet sich im Chorion.
2. Die Larve befindet sich nicht mehr im Chorion.

In der vorliegenden Arbeit wurden neue Methoden für den ersten Fall entwickelt, da für den Hochdurchsatz das Dechorionieren, wie in Abschnitt 2.4.1 beschrieben, deutliche Nachteile hat. Unter Einbezug von Vorwissen, wie zum Beispiel der Kontur des Zebrabärblings, kann jedoch auch für den zweiten Punkt eine Segmentierung erfolgreich sein [49].

Im ersten Fall (der Zebrabärbling befindet sich im Chorion), lässt sich das zu findende Objekt gut durch die Größe und Form charakterisieren. Daher wird ein Schwellenwert mittels Intervallschachtelung so lange angepasst, bis das größte Objekt den Suchkriterien der Hochdurchsatz-Untersuchung gerecht wird. Bei der Suche nach dem Ei des Zebrabärblings ist das Ziel der Suche ein einzelnes Objekt, welches ein Mindestmaß für Rundheit und eine Größe innerhalb eines zuvor festgelegten Toleranzbereichs besitzt. Als Maßeinheit empfiehlt es sich, sich von der Bildauflösung unabhängige Maße, wie z.B. die wahre Größe in mm, zu wählen, da sich die Algorithmen somit auch auf andere Mikroskope, Kameras und Vergrößerungen übertragen lassen. Werden nach einer maximalen Anzahl an Iterationen die o.g. Kriterien nicht erreicht, wird die Suche abgebrochen und die Versuchseinheit als fehlerhaft markiert. Das Verfahren setzt voraus, dass die Mehrzahl der Versuche korrekt präpariert ist, d.h. genau ein Ei enthält, da solche Versuchseinheiten sonst als fehlerhaft detektiert, aus der weiteren Verarbeitung ausgeschlossen werden und somit wertlos sind. Die Validitätsprüfung geschieht hier durch Bestimmung von instantanen Merkmalen des größten Objektes. Typische Werte für die Ei-Größe und das Verhältnis der Durchmesser einer umschließenden Ellipse werden zu Beginn für typische Beispiele ermittelt und dann mit den Werten jeder Versuchseinheit verglichen. Alle Bilder außerhalb eines Toleranzbereichs von 20% des ermittelten Wertes werden verworfen. Für nähere Ausführungen zu Merkmalen und deren Extraktion sei auch auf Abschnitt 3.4 verwiesen. Schließlich wird das Bild auf den Bereich, der das größte Objekt enthält, zugeschnitten. Der Zuschnitt erfolgt durch Anpassung der Indizes des Bildstroms (vgl. Formel (2.18)).

3.3.2 Neue Methode zur Trennung von mehreren Objekten und Hintergrund

Für die Trennung mehrerer Eier von Zebrabärblingen vom Hintergrund kann ebenfalls auf eine Kette von Schwellenwerten und morphologischen Operatoren zurückgegriffen werden. Je mehr Versuchseinheiten pro Bild aufgezeichnet werden, desto größer ist jedoch die Wahrscheinlichkeit, dass sich die Versuchseinheiten bei der Aufnahme berühren

und nach der Anwendung eines Schwellenwertes im Binärbild als ein einziges großes Objekt erscheinen. Wenn es nicht möglich ist, solche Cluster zu verwerfen, müssen diese anschließend durch zusätzliche Bildverarbeitungsschritte getrennt werden, z.B. mittels Regionen-Wachstumsverfahren oder Levelset-Algorithmen [39, 128, 187]. Da es sich hierbei um aufwändigere Verfahren handelt, wächst der Rechenaufwand deutlich und macht solche Verfahren für Hochdurchsatz-Untersuchungen weniger geeignet. Durch Ausnutzung von spezifischem Vorwissen lässt sich jedoch eine robuste Segmentierung finden, welche den besten Kompromiss zwischen Rechenaufwand und Robustheit liefert. Bei der Segmentierung mehrerer Zebrabärblingseier lässt sich eine Konstante im Bild ausnutzen, die in jeder Bildsituation und in jeder Position sowie bei sich bewegenden Larven vorkommt: Die Außenkante des Chorions. Die Außenkante bildet zudem auch den Abschluss des zu untersuchenden Objektes zu Nachbarobjekten und dem Hintergrund. Das Chorion kann vereinfacht durch einen Kreis beschrieben werden, welcher, wenn dieser einen akzeptablen Kontrast zum Hintergrund aufweist, durch den weitgehend bekannten Radius eine robuste Detektion zulässt. Zudem verändert sich der Durchmesser des Chorions während der Entwicklung der Larve nur unwesentlich [177]. Damit lässt sich die Trennung von mehreren Versuchseinheiten zumindest für die initiale Detektion der Eier auf eine Kreissuche reduzieren. Für die Kreissuche existieren in der Literatur bereits diverse Ansätze [31, 60]. Die Hough-Kreisdetektion kann erfolgreich zur Detektion des Chorions angewandt werden. Hierfür wird zu Beginn ein Kantenbild beispielsweise mittels des Canny-Filters errechnet [29]. Jedem Pixel im Kantenbild werden daraufhin alle potenziellen Mittelpunkte des Chorions zugeordnet und jeweils in einer Akkumulationsmatrix markiert. Mittels der Maxima der so entstandenen Matrix lässt sich daraufhin die Position der Zebrabärblingseier ermitteln. Limitierungen hat die beschriebene Methode bei niedrigem Kontrast des Chorions zum Hintergrund. In einem solchen Fall wird das Chorion nicht vollständig als Kante erkannt und das Maximum in der Akkumulationsmatrix ist weniger ausgeprägt. Bei sehr schlechtem Kontrast führt dies dazu, dass der Rücken der Zebrabärblinge stärker im Kantenbild repräsentiert ist als der Rand des Chorions, was zu einer fehlerhaften Kreisdetektion führt. Abbildung 5.19 in Kapitel 5 zeigt ein Beispiel der Anwendung der Methode. Allerdings ist der Rechenaufwand zum Erstellen der Akkumulationsmatrix größer als bei den erwähnten reinen Schwellenwertverfahren. Im Gegensatz zu jenen ist die Kreisdetektion jedoch sehr robust und das Ergebnis lässt sich anhand einfacher Validitätskriterien wie z.B. dem mittleren Grauwert des Kreisinhaltes überprüfen.

3.3.3 Neues Verfahren zum Tracking des Chorions

Liegen Bildsequenzen von mehreren Zebrabärblingseiern vor, ist es notwendig, die Position des Nutzsymbols an jedem Zeitpunkt zu bestimmen und einander zuzuordnen. Die in Abschnitt 3.3.2 vorgestellte Methode zur Detektion des Chorions lässt sich zwar auf jedes Frame eines Bildstroms anwenden, allerdings bedeutet dies einen hohen Rechenaufwand. Günstiger ist es, das Tracking mittels Kreuzkorrelation hinzu zu kombinieren [70]. Tabelle 3.1 stellt den Rechenaufwand der beiden Methoden gegenüber.

Method	Rechenaufwand [s]	Rechenaufwand für typische Sequenz [s]
Kreisdetektion	0.131284	131.284
Korrelation	0.001889	15.112

Tabelle 3.1: Gegenüberstellung des Rechenaufwands von Korrelation und Kreisdetektion auf einem handelsüblichen Desktop-PC für eine typische Sequenzlänge von 1000 Frames und 8 Versuchseinheiten

Die Kreisdetektion benötigt gegenüber der Korrelation zweier Bilder signifikant mehr Rechenzeit. Ein besserer Ansatz ist es daher, die in einer initialen Kreisdetektion gewonnene Information der Eiposition auszunutzen und das Chorion mittels Korrelation zu tracken. Allerdings muss die Korrelation für jedes zu trackende Ei durchgeführt werden, während der Aufwand der Kreisdetektion von der Eianzahl unabhängig ist. Dennoch ergibt sich für eine typische Sequenz mit etwa 1000 Frames und 8 Eiern ein deutlicher Geschwindigkeitsvorteil gegenüber der ausschließlichen Kreisdetektion. Bei der Korrelation wird der Suchraum des Ortes im darauf folgenden Frame (das sog. Zielbild) auf einen kleinen Bereich um die bekannte Eiposition verringert. Die Einschränkung des Suchraums kann getroffen werden, da sich das Ei zwischen zwei Frames nur um eine endliche Pixelanzahl verschiebt und zudem die Driftbewegungen der Eier verhältnismäßig langsam bezüglich der Abtastrate sind. Problematisch hierbei sind jedoch kleine Abweichungen jeder gefundenen Position in jedem getrackten Frame, was zu einem „Wegdriften“ der aktuellen Position von der tatsächlichen Eiposition führen kann. Das Problem des Wegdriftens wurde gelöst, indem als Suchbild beim Tracking nicht das gefundene Ei, sondern ein einfacher Kreis mit dem Radius der Eihülle verwendet wurde. Da im Suchbild nur die Eihülle einen Kreis darstellt und im Suchraum nur ein Ei vorhanden sein kann, ist ein Wegdriften des Suchfeldes bei ausreichender Bildqualität ausgeschlossen. Mit dem zu Beginn aus der Hough-Kreisdetektion ermittelten Eidurchmesser wird das Suchbild mit lediglich einem Kreis erstellt. Dieses Bild wird im Suchraum mittels Korrelation so positioniert, dass es das Chorion bzw. das Kantenbild des Chorions möglichst gut abdeckt. Hierfür wird eine sog. Güte-Matrix erstellt, die einen Wert für die Übereinstimmung des Suchbildes im Zielbild enthält. Das Maximum in der Güte-Matrix stellt die wahrscheinlichste neue Position des Chorions dar und wird als Startpunkt für die Suche im nächsten Frame verwendet. Der Vorgang wird wiederholt, bis die gesamte Bildsequenz abgearbeitet ist und in jedem Frame die Position des Chorions bestimmt wurde. Abb. 3.9 veranschaulicht den Ablauf des Trackings.

Als Empfehlung zur Bestimmung der Parameter erweist es sich als zweckmäßig, den mittleren Eidurchmesser in den ersten Frames der ersten Bildsequenzen einer Hochdurchsatz-Untersuchung manuell zu ermitteln. Dieser kann dann mit einer Toleranz von 10% zur weiteren automatischen Kreissuche angewandt werden. Für den Suchraum wurden um das Suchbild problemspezifisch definierte Entfernungen (Standardparameter: vier Pixel) in X- und Y-Richtung hinzugefügt. Dies erweist sich für die Zebrabärblinge bei einer Bildwiederholungsfrequenz von 30 fps als ausreichend.

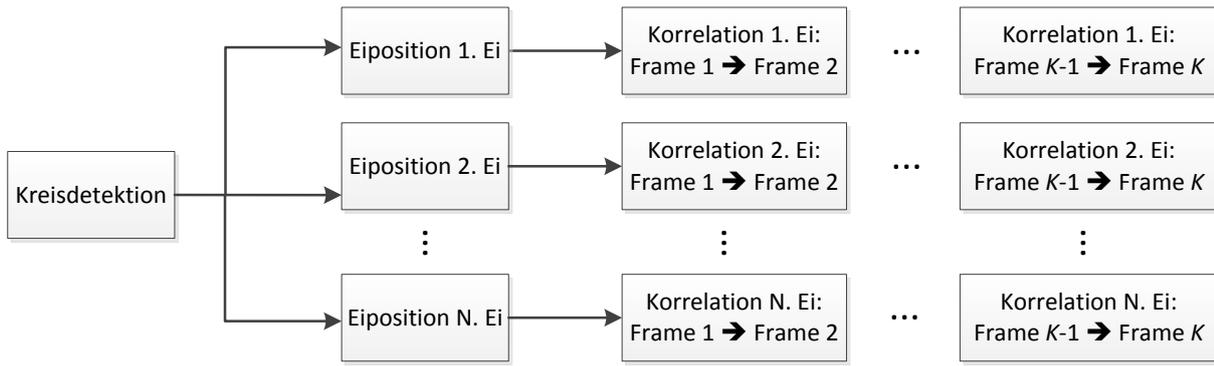


Abbildung 3.9: Ablaufschema des Trackings bei einer Hochdurchsatz-Untersuchung mit noch im Chorion befindlichen Zebraabärblingslarven. K bezeichnet die Anzahl an Frames und N die Anzahl an Eiern in der untersuchten Sequenz.

3.3.4 Differenzbilder

Differenzbilder sind eine einfache Möglichkeit die Veränderung der Pixelwerte und damit die Bewegung der Zebraabärblingslarve in einem Bild darzustellen. Hierzu wird bei einer Bildsequenz das zeitlich später aufgezeichnete Bild vom vorhergehenden abgezogen. Um negative Pixelwerte zu vermeiden, wird der Absolutwert gebildet. Werden alle Bilder des Bildstroms \mathbf{BS} nach dem Zeitpunkt der Akquise geordnet und wird sich zur einfacheren Notation hierbei auf lediglich eine Modalität $I_c = 1$ und eine Schichtaufnahme $I_z = 1$ beschränkt, so ergibt sich eine Zeitreihe $I_{i_x i_y}[k]$ mit $k = 1 \dots K$ Abtastzeitpunkten. Für die Differenzbilder lässt sich dann schreiben:

$$\Psi_d[k] = |I[k] - I[k-1]| \quad \text{mit } k = 2 \dots K \quad (3.12)$$

Die entstandenen Differenz- oder Merkmalsbilder bilden die Basis für eine Vielzahl von Einzelmerkmalen, deren gemeinsames Ziel es ist, die Menge an Bewegung im gesamten Bild oder in Ausschnitten zu quantifizieren. Ein einzelnes Differenzbild aus einer solchen Zeitreihe wird im Folgenden mit \mathbf{I}_d bezeichnet.

3.4 Nutzsinalzeitreihen und -merkmale

Drei Typen von Nutzsinalen werden unterschieden:

1. Nutzsinalen, die sich auf das Aussehen des Zebraabärblings beziehen,
2. Nutzsinalen, die sich auf die Bewegung des Zebraabärblings beziehen und
3. Nutzsinalen, die sich auf die Signalstärke von z.B. fluoreszierenden Markern beziehen.

Ziel der Merkmale ist es, die Nutzsignale zu quantifizieren und somit vergleichbar zu machen. Für jeden Typ der Nutzsignale wurden mehrere Merkmale entwickelt. Zwar ist ein einziges, signifikantes Merkmal zur Klassifikation oft bereits ausreichend, doch kann durch die Berücksichtigung mehrerer Merkmale eine niedrigere Fehlerrate erzielt werden. Gerade bei der starken Beeinträchtigung der Nutzsignale durch Störfaktoren, wie es bei Hochdurchsatz-Untersuchungen der Fall ist, ist die Extraktion von mehreren Merkmalen von Vorteil. Generell hat es sich für Data-Mining Methoden bei Hochdurchsatz-Untersuchungen bewährt, möglichst viele Merkmale zur Auswertung heranzuziehen und signifikante Merkmale später ausfindig zu machen. In den folgenden Abschnitten wird ein spezifischer, beschreibender Merkmalsatz aufgebaut, auf dessen Basis daraufhin eine Bewertung erfolgen kann. Damit ergeben sich als Eingangsgrößen der Modulkategorie von der Segmentierung bereitgestellten Nutzsignale Ψ und als Ausgangsgrößen Zeitreihen \mathbf{Y}_{ZR} oder Merkmale \mathbf{X} (vgl. Formel (2.27) und Abb. 2.9).

3.4.1 Instantane Merkmale

Für den ersten Typ Merkmale, solche, die sich auf das Aussehen des Zebrabärblings beziehen, sind Einzelbilder zur Merkmalsextraktion ausreichend. Sie werden entweder durch Auswahl von Zeitpunkten, nach den Methoden aus Abschnitt 3.2.1 aus dem Bildstrom entnommen, oder es wurden bereits bei der Bildakquise lediglich Einzelbilder je Versuchseinheit akquiriert. Die Merkmale beurteilen ohne Berücksichtigung einer zeitlichen Veränderung den Zustand zum jeweiligen Aufnahmezeitpunkt. Zur Merkmalsextraktion wird ein einzelnes, vorverarbeitetes und segmentiertes Bild herangezogen.

Ein histogramm-basierendes Merkmal ist der Schwerpunkt des Histogramms H über die Pixelwerte im Bild \mathbf{I} und berechnet sich für ein Einzelbild gemäß:

$$x_1 = \frac{\sum_{i_h=0}^{\phi} i_h \cdot H(i_h)}{\sum_{i_h=0}^{\phi} H(i_h)}. \quad (3.13)$$

wobei für das Histogramm H gilt:

$$H(i_h) = \text{card}\{(i_x, i_y) | I(i_x, i_y) = i_h\} \quad \text{mit } i_h = 1 \dots \phi, i_h \in \mathbb{N}. \quad (3.14)$$

Der Operator *card* bezeichnet die Kardinalität oder Mächtigkeit, welche für endliche Mengen gleich der Anzahl der Elemente einer bestimmten Menge ist.

Ein ähnliches, grauwert-basiertes Merkmal ist der Mittelwert über alle Pixel im segmentierten Bild \mathbf{I} . Es hat sich hierbei bewährt, Hintergrundpixel zuvor auf den Wert 0 zu

setzen und bei der Berechnung auszuschließen:

$$x_2 = \frac{\sum_{i_x=1}^{I_x} \sum_{i_y=1}^{I_y} I(i_x, i_y)}{I_x \cdot I_y - \text{card}\{(i_x, i_y) \mid I(i_x, i_y) = 0\}}. \quad (3.15)$$

Beide Werte geben einen allgemeinen Hinweis, ob der Inhalt des segmentierten Bildes vergleichbar ist. Eine ähnlich entwickelte Zebrabärblingslarve hat bei gleichen Akquise-Parametern auch eine vergleichbare Grauwertverteilung und somit einen vergleichbaren Schwerpunkt des o.g. Histogramms bzw. einen ähnlichen Mittelwert der Grauwerte.

Steht etwa, wie bei den Eiern des Zebrabärblings, das Vorwissen zur Verfügung, dass sich die relevante Information in der Mitte des segmentierten Bildes (d.h. im Zentrum des Chorions) befindet, lassen sich weitere Merkmale wie der „mittlere Grauwert der mittleren Bildzeile“ und die „Schwankung der Grauwerte entlang der mittleren Bildzeile“ berechnen. Beide Merkmale sind ein guter Indikator für den Entwicklungsstand der Zebrabärblinge. Die Merkmale lassen sich weiter verbessern, wenn sie auf das Zentrum des segmentierten Bereichs einschränken werden, indem lediglich die mittleren 30% der mittleren Bildzeile ausgewertet werden:

$$x_3 = \frac{\sum_{i_x=n_1}^{n_2} I\left(i_x, \left(\frac{I_y}{2}\right)\right)}{n_2 - n_1 + 1} \quad (3.16)$$

mit gerundeten Werten $n_1 = [0.3 \cdot I_x]$ und $n_2 = [0.7 \cdot I_x]$.

Die Schwankung um die mittlere Bildzeile lässt sich nach Korrektur um die mittleren Grauwerte c berechnen durch

$$\delta(i_x) = \left| I\left(i_x, \left(\frac{I_y}{2}\right)\right) - c \right|. \quad (3.17)$$

Nach Aufsummieren ergibt sich das Merkmal ebenfalls über den mittleren Bereich der mittleren Bildzeile

$$x_4 = \sum_{i_x=n_1}^{n_2} \delta(i_x). \quad (3.18)$$

Die Anzahl an Kanten in Kantenbildern ist ein guter Indikator für die Anzahl an ausgeprägten Details im Zebrabärbling und somit für die Vergleichbarkeit des Entwicklungsstandes. Die Algorithmen zur Kantendetektion „Canny“ I_{Canny} [29] und „Laplacian of Gaussian (LoG)“ I_{LoG} [62] wurden angewandt und ausgewertet:

$$x_5 = \text{card}\{(i_x, i_y) \mid I_{\text{Canny}}(i_x, i_y) = 1\}, \quad (3.19)$$

$$x_6 = \text{card}\{(i_x, i_y) \mid I_{\text{LoG}}(i_x, i_y) = 1\}. \quad (3.20)$$

Es lassen sich weitere Merkmale ermitteln, indem das Aussehen der Objekte im Bild untersucht wird. Beispielsweise lässt sich die „Kugeligkeit“ oder „Rundheit“ des Bildinhalts des segmentierten Bildstroms bestimmen. Dafür wird nach Anwendung des ISO-Data Algorithmus [55] die längste M_{max} mit der kürzesten Halbachse M_{min} der umschließenden

Ellipse verglichen. $(\mu_{11}, \mu_{22}, \mu_{12})$ sind hierfür die notwendigen zentralen Momente des gefilterten Bildes (vgl. [54]).

$$M_{\max} = 2 \cdot \sqrt{2} \cdot \sqrt{\mu_{11} + \mu_{22} + \sqrt{(\mu_{11} - \mu_{22})^2 + 4 \cdot \mu_{12}^2}} \quad (3.21)$$

$$M_{\min} = 2 \cdot \sqrt{2} \cdot \sqrt{\mu_{11} + \mu_{22} - \sqrt{(\mu_{11} - \mu_{22})^2 + 4 \cdot \mu_{12}^2}} \quad (3.22)$$

$$x_7 = M_{\max} - M_{\min}. \quad (3.23)$$

Für einen Validitätstest wird die Größe jedes Bildes als Merkmal extrahiert. So lassen sich bei der Auswertung deutlich größere oder kleinere Bildinhalte ausschließen.

$$x_8 = I_x \cdot I_y. \quad (3.24)$$

3.4.2 Auf Bewegungen bezogene Merkmale

Alle Merkmale, die sich auf die Bewegung der Zebraabürblinge beziehen, wurden aus Differenzbildern extrahiert. Die Merkmale unterscheiden sich vornehmlich dadurch, dass sie unterschiedlichen Wert auf charakteristische Phänomene der Bewegungen der Larven, wie etwa lokal oder global auftretende Änderungen, legen.

Ein geeignetes Merkmal ist die Standardabweichung in den Differenzbildern nach Formel (3.12):

$$x_9[k] = \frac{1}{I_x I_y - 1} \sum_{i_x=1}^{I_x} \sum_{i_y=1}^{I_y} \left(I_{d,i_x i_y}[k] - \bar{I}_d[k] \right)^2 \quad (3.25)$$

mit

$$\bar{I}_d[k] = \frac{1}{I_x I_y} \sum_{i_x=1}^{I_x} \sum_{i_y=1}^{I_y} I_{d,i_x i_y}[k]. \quad (3.26)$$

Das Merkmal quantifiziert die Variation der Pixel im Differenzbild. Bei Bewegung steigt die Standardabweichung an. Eine Variante ist es, das Merkmal über die Anzahl der Pixel und die Farbtiefe ϕ (Die Farbtiefe wird üblicherweise bei der Bildakquise festgelegt) zu normalisieren:

$$x_{10}[k] = \frac{1}{I_x I_y \phi} \sum_{i_x=1}^{I_x} \sum_{i_y=1}^{I_y} I_{d,i_x i_y}[k] \quad (3.27)$$

Das Merkmal x_{10} gibt Auskunft über die mittlere relative Änderung der Pixel.

Um Differenzen innerhalb dunkler und heller Bildbereiche direkt vergleichbar zu machen, wird der Mittelwert über die Intensitätswerte der Pixel normalisiert. Um Divisionen durch Null zu vermeiden, wird eine Konstante x_{off} im Nenner addiert:

$$x_{11}[k] = \frac{1}{I_x I_y} \sum_{i_x=1}^{I_x} \sum_{i_y=1}^{I_y} \frac{I_{d,i_x i_y}[k]}{I_{\max,i_x i_y}[k] + x_{\text{off}}} \quad (3.28)$$

mit

$$I_{\max, i_x i_y}[k] = \max(I_{i_x i_y}[k], I_{i_x i_y}[k-1]). \quad (3.29)$$

Stark bewegte Regionen im Bild lassen sich quantisieren, wenn die Anzahl relevanter Pixeländerungen im Differenzbild bestimmt wird. Dazu werden Pixel gezählt, die Differenzen größer als ein dynamischer Schwellenwert I_{dyn} aufweisen:

$$x_{12}[k] = \text{card}_{i_x, i_y}(I_{d, i_x i_y}[k] | I_{d, i_x i_y}[k] > I_{\text{dyn}}[k]) \quad (3.30)$$

Der dynamische Schwellenwert wird so bestimmt, dass er Differenzwerte übertrifft, die aus Bildrauschen resultieren. Bildrauschen bewirkt eine Abweichung der gemessenen Pixelwerte, ohne tatsächliche Änderung des aufgezeichneten Objektes. Wird ein Pixel eines mit Rauschen überlagerten Bildes mit I_d^* bezeichnet, so liefern der Mittelwert und die Standardabweichung (hier durch das dreifache der Standardabweichung 99,7% aller Pixelwerte berücksichtigt) dieses Pixels I_{dyn} :

$$I_{\text{dyn}}[k] = \bar{I}_d^*[k] + 3s_d^*[k] \quad (3.31)$$

Durch Bildrauschen entstehen Ausreißer im Differenzbild mit Werten oberhalb eines mittels des c -Quantils über den Pixelwerten des Bildes gebildeten Schwellenwertes I_{quant_c} oder unterhalb eines mittels des $(1-c)$ -Quantils über den Pixelwerten des Bildes gebildeten Schwellenwertes $I_{\text{quant}_{1-c}}$. Ihre Pixelwerte lassen sich bestimmen durch:

$$\{(i_x, i_y)\} = \{(i_x, i_y) | I_{\text{quant}_c} < I_{d, i_x i_y} < I_{\text{quant}_{1-c}}\}, \quad 0 \leq c < 0.5. \quad (3.32)$$

c muss hierbei kleiner sein als der Anteil der maximal im Bild auftretenden Bewegung. Ein Wert von $c=0.4$ hat sich in der Praxis bewährt. Damit ergibt sich

$$s_d^*[k] = \frac{1}{(1-c)I_x I_y - 1} \sum_{(i_x, i_y) | c < I_{d, i_x i_y} < 1-c} (I_{d, i_x i_y}[k] - \bar{I}_d^*[k])^2 \quad (3.33)$$

mit

$$\bar{I}_d^*[k] = \frac{1}{(1-c)I_x I_y - 1} \sum_{(i_x, i_y) | c < I_{d, i_x i_y} < 1-c} I_{d, i_x i_y}[k] \quad (3.34)$$

Das Merkmal $x_{13}[k]$ bezieht sich auf die Anzahl relevanter, relativer Pixeländerungen im Differenzbild und berechnet sich wie x_{12} in Formel (3.30), allerdings mit normalisierten Differenzwerten $\frac{I_{d, i_x i_y}[k]}{I_x I_y}$ anstelle von $I_{d, i_x i_y}[k]$.

$$x_{13}[k] = \text{card} \left((i_x, i_y) \left| \frac{I_{d, i_x i_y}[k]}{I_x I_y} > I_{\text{dyn}}[k] \right. \right) \quad (3.35)$$

Die maximale Änderung im Differenzbild ist ein Maß für die stärkste Bewegung der Zebraabärlingslarve:

$$x_{14}[k] = \max_{i_x, i_y}(I_{d, i_x i_y}[k]) \quad (3.36)$$

Eine Variante des Merkmals x_{14} ist es, das Merkmalsbild zuvor mit einem 3x3 Gaußfilter (Maximum im geglätteten Differenzbild) zu falten. Dies reduziert den Einfluss einzelner Ausreißer im Differenzbild. Es hat sich bewährt für σ 0.5 zu wählen:

$$h(i_x, i_y) = \exp -\frac{i_x^2 + i_y^2}{2\sigma^2} : \quad (3.37)$$

$$x_{15}[k] = \max_{I_x, I_y}(I_{d,smoothed,i_x i_y}[k]) \quad (3.38)$$

mit

$$I_{d,smoothed,i_x i_y}[k] = I_{d,i_x i_y}[k] * h(i_x, i_y). \quad (3.39)$$

wobei $*$ für den Faltungsoperator und $h(i_x, i_y)$ für die Filtermaske steht.

Die Summe aller Änderungen im Differenzbild ist ein Maß für die Menge an Bewegung im Bild. Verschiedene Einflüsse bei der Bildakquise können dazu führen, dass im Bild ein Beleuchtungs- bzw. Helligkeitsverlauf auftritt, was die Bilddynamik beeinflusst. Da die Absolutwerte der Differenzbilder von der Dynamik in den Rohdaten begrenzt wird, entstehen so auch Abweichungen in der Messung. Um die Vergleichbarkeit der Merkmale und deren absoluten Werte sicherzustellen, ist eine Normierung sinnvoll. Daher wurden die Merkmalswerte mittels der Standardabweichung über alle Pixelwerte normalisiert. Somit berechnet sich das Merkmal nach:

$$x_{16}[k] = \frac{1}{x_9[k]} \sum_{i_x=1}^{I_x} \sum_{i_y=1}^{I_y} I_{d,i_x i_y}[k] \quad (3.40)$$

Wichtig bei einer solchen Normalisierung ist jedoch ein vergleichbarer Bildinhalt. Ist in den Bildern beispielsweise eine variierende Anzahl an Eiern abgebildet, so scheitert die Methode.

Die einfachste Form zur Ermittlung der Bewegung im Differenzbild ist die Summe aller Änderungen, ohne jegliche Normalisierung nach:

$$x_{17}[k] = \sum_{i_x=1}^{I_x} \sum_{i_y=1}^{I_y} I_{d,i_x i_y}[k] \quad (3.41)$$

3.4.3 Auf Signalstärke bezogene Merkmale

Merkmale zur Bestimmung der Signalstärke finden Anwendung, wenn die Stärke des auszuwertenden Nutzsignals durch den optischen Detektor erfasst wurde. Bei Hochdurchsatz-Untersuchungen mit Zebrabärblingen ist dies vor allem bei mittels Fluoreszenzmikroskopie generierten Bilddaten der Fall. Solche Merkmale gestatten einen Rückschluss auf die Menge an fluoreszent leuchtenden Markern. Die Berechnung erfolgt meist mittels Addition aller Pixelwerte im segmentierten Bildstrom oder es wird auf das arithmetische Mittel bzw. den Median über die Pixelwerte zurückgegriffen. Bei der Auswertung von

auf Basis fluoreszierend leuchtender Marker akquirierter Bilder ist zur richtigen Deutung jedoch eine Berücksichtigung der Messeinrichtung nötig. Eine weitere Schwierigkeit sind fluoreszente Strahlungen des Hintergrundes sowie ein oft hohes Rauschen im aufgezeichneten Bild. Eine genauere Betrachtung sowie eine Checkliste zu den Besonderheiten der Quantifizierung von Fluoreszenzbildern finden sich in [176].

3.5 Merkmalsauswahl

Bei der Merkmalsauswahl werden die extrahierten Merkmale für die Klassifikation vorbereitet bzw. ausgewählt. Eingangsgrößen der Modulkategorie sind alle m_2 extrahierten Merkmale \mathbf{X} . Die Ausgangsgröße ist eine reduzierte Anzahl an Merkmalen m_3 bzw. m_d . Zu Beginn werden Merkmalswerte, welche von fehlerhaften Bilddaten wie leeren Näpfchen, zu vielen Versuchseinheiten, schlechter Beleuchtung etc. stammen, in einem Validitätstest ausgeschlossen. Hierbei kann auf einfache Schwellenwerte zurückgegriffen werden, die im Allgemeinen aus Vorwissen stammen. Wurde beispielsweise die Eiggröße als Merkmal extrahiert, kann ein Bereich sinnvoller Werte bestimmt werden. Merkmale außerhalb des bestimmten Bereichs sollten verworfen werden. Sind solche Pixel ausgeschlossen, kann eine Auswahl der besten Merkmale erfolgen, denn die extrahierten Merkmale sind z.T. redundant und es kommt im konkreten Anwendungsfall auf die wirkende Störung an, welches der Merkmale eine trennstärke Information enthält. Je mehr redundante Merkmale zur Verfügung stehen, desto schwieriger wird die Entscheidungsfindung. Eine zu große Anzahl an hierbei berücksichtigten Merkmalen kann sogar zu einer Fehlklassifikation führen [40]. Wird die Auswertung in Echtzeit zur Verfügung gestellt und auf die Akquise zurückgekoppelt wie z.B. bei Roboteranwendungen, ist es schon aus Gründen der Rechenzeit notwendig, die Berechnung auf notwendige Merkmale zu limitieren.

Zur automatischen Merkmalsauswahl kommen verschiedene bekannte Methoden zum Einsatz. Beispiele sind die univariate Varianzanalyse (ANalysis Of VAriance – ANOVA) und die multivariate Varianzanalyse (Multivariate ANalysis Of VAriance – MANOVA) [3]. Die Verfahren vergleichen die Innerklassenvarianzen mit den Zwischenklassenvarianzen der Merkmale und geben so Aufschluss darüber, welches Merkmal oder welche Merkmalskombination am besten zur Unterscheidung der gewählten Klassen geeignet ist. Mit Hilfe der ANOVA wird das Merkmal identifiziert, dessen Relevanz zur Klassentrennung den höchsten Wert erreicht. Daraufhin wird das identifizierte Merkmal, mit Hilfe der MANOVA, mit allen verbliebenen Merkmalen kombiniert, die Relevanz der Merkmalskombination ermittelt und die beste Zweier-Merkmalskombination bestimmt. Der Schritt entspricht der Formel (2.26) und Formel (2.27). Es hat sich in der Praxis bei Hochdurchsatz-Untersuchungen bewährt, den Merkmalsraum auf eine solche Zweier-Merkmalskombination zu reduzieren und zur Klassifikation heranzuziehen. Eine solche Zweier-Merkmalskombination ist ein guter Kompromiss zwischen Interpretierbarkeit, Rechenzeit und Güte der Klassifikation [3, 110].

3.6 Auswertung und Klassifikation

Nach Auswahl geeigneter Module zur Extraktion von informationstragenden Merkmalen sind mit der formulierten Fragestellung aus Abschnitt 2.6 alle notwendigen Voraussetzungen für eine erfolgreiche Lösung eines Klassifikations- oder Regressionsproblems in Hochdurchsatz-Untersuchungen am Zebraäbrbling gegeben. Zur Klassenzuordnung kann zumeist auf bekannte Verfahren der Klassifikation zurückgegriffen werden [110]. Entscheidend für den Erfolg ist die Bereitstellung eines informationstragenden Merkmalsatzes \mathbf{W} . Wie bei einer solchen Art von Problemstellung üblich, muss ein Kompromiss zwischen einer niedrigen Fehlerrate und einer hohen Interpretierbarkeit der Ergebnisse, sowie einem möglichst geringen Rechenaufwand gefunden werden. Die Eingangsgrößen bei dieser Modulkategorie sind die vorverarbeiteten Merkmale \mathbf{W} und die Ausgangsgrößen sind die Klassenzuweisungen der Planfaktoren $\hat{\mathbf{Y}}$ bzw. der Störfaktoren $\hat{\mathbf{Z}}$ (vgl. auch Formel (2.28)).

3.6.1 Klassifikation

Zur Durchführung der Klassifikation unbekannter Daten muss der Klassifikator zu Beginn mittels eines Lerndatensatzes angelernet werden. Die Frage nach der Anzahl an Versuchseinheiten, die hierfür notwendig ist, kann nicht pauschal beantwortet werden und hängt von den Eigenschaften und Störfaktoren der Hochdurchsatz-Untersuchung ab [110]. In den meisten Fällen sind jedoch 20 bis 50 Versuchseinheiten einer Klasse ausreichend. Für das Labeln, also das manuelle Zuordnen einer Versuchseinheit der Lerndaten zu einer Klasse, wurde eine komfortable Softwarelösung entwickelt und in das Softwarepaket Gait-CAD implementiert (vgl. Abschnitt 4.4.1), welcher dem Anwender ein zufälliges Beispiel aus dem Datensatz zeigt und die Möglichkeit bereitstellt, per Mausklick einer der vorhandenen Ausgangsklassen zuzuweisen. So kann auf schnelle Weise für eine Hochdurchsatz-Untersuchung am Zebraäbrbling der Lerndatensatz kreiert werden.

Zur Klassifikation kommen vornehmlich Bayes-Klassifikatoren zum Einsatz, deren Aufgabe es ist, die Zugehörigkeit der Versuchseinheiten auf Basis der extrahierten, reellwertigen Merkmalswerte \mathbf{W} einer der Klassen aus \mathbf{Y} oder \mathbf{Z} diskret zuzuordnen. Mittels des Lerndatensatzes wird ein hierfür funktioneller Zusammenhang bestimmt. Der Zusammenhang wird als Diskriminanzfunktion bezeichnet und ist entweder eine Entscheidungsfunktion, die den Merkmalswerten eine Klasse explizit zuordnet, oder besteht aus Distanzen zu den Klassen, welche die Zugehörigkeit implizit ausdrücken. Es wird eine Normalverteilung der Merkmale für jede Ausgangsklasse angenommen. Die Diskriminanzfunktion wird mittels der Mahalanobis-Distanz und klassenspezifischen Kovarianzmatrizen bestimmt [100]. Im Anschluss lassen sich mittels Validierungsverfahren die Güte über unbekannte Datentupel abschätzen und die Auswirkungen zufälliger Störeinflüsse ermitteln [110].

3.6.2 Regression der Dosis-Wirkungs-Kurve

Oft ist es das Ziel, den Effekt des in der Hochdurchsatz-Untersuchung systematisch variierten Störfaktors \mathbf{Z} auf einen oder mehrere Planfaktoren \mathbf{Y} zu bestimmen. Bei Toxinen werden hier üblicherweise Verdünnungsreihen angelegt. Es kann auch der Effekt anderer Störfaktoren, wie z.B. Spannungen oder Temperaturen, beobachtet werden. Jede Versuchsreihe wird mit einer unterschiedlichen Konzentration (oder Dosis) ausgeführt. Voraussetzung ist, dass für jede gewählte Stufe eine ausreichende Anzahl an Versuchseinheiten präpariert wurde. Die Klassifikation erfolgt dabei nacheinander für jede der gewählten Konzentrationsstufen und es wird die Klassenzugehörigkeit aller Versuchseinheiten $\hat{\mathbf{Y}}$ geschätzt. Wird die beobachtete Wirkung durch den variierten Störfaktor stimuliert, so wird in den meisten Fällen eine kontinuierliche Zunahme des Effekts erwartet. Die Konzentration wird von einem niedrigen Effekt-Niveau $(p_y)_{\min}$ zu einem hohen Effekt-Niveau $(p_y)_{\max}$ verlaufen (vgl. Abb. 3.10). Die Dosis-Wirkungs-Kurve folgt typischerweise einer Sigmoidfunktion, welche mittels Regression an die Messwerte angepasst wird. Das entstehende Optimierungsproblem ist parameternichtlinear und wird numerisch gelöst.

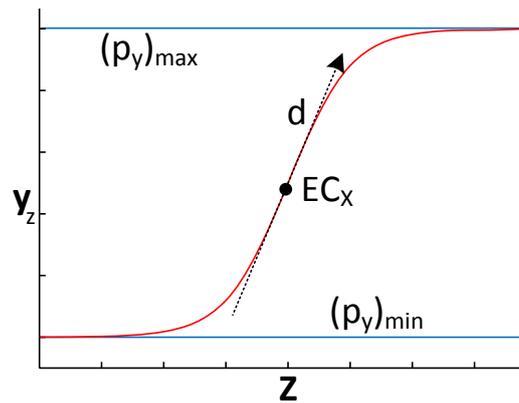


Abbildung 3.10: Sigmoidkurve der Dosis-Wirkungs-Kurve mit den Parametern EC_X , $(p_y)_{\min}$, $(p_y)_{\max}$ und d [37].

$$y_{z_i} = (p_y)_{\min} + \frac{(p_y)_{\max} - (p_y)_{\min}}{1 + \left(\frac{z_i}{EC_X}\right)^d} \quad (3.42)$$

Hier ist y_z der Wert bezüglich des untersuchten Effekts (meist in Prozent), $(p_y)_{\min}$, $(p_y)_{\max}$ der niedrigste bzw. höchste ermittelte Wert des Effekts, der Vektor \mathbf{z}_c enthält die Konzentrationen und d ist der Absolutwert der Steigung der Sigmoidfunktion [37]. EC_X bezeichnet die Wendestelle der Kurve, welcher einer Konzentration/Dosis zugeordnet wird. Der Parameter d wird als Hill-Koeffizient bezeichnet. Für den Fall, dass steigende Konzentrationen die Wirkung hemmen, wird der Parameter negativ und es ergibt sich ein abfallender Verlauf der Kurve in Abb. 3.10. Die Wirksamkeit auf den gesuchten Effekt lässt sich am EC_X -Wert ablesen. Entscheidend für den Erfolg der Methode ist jedoch, dass sich der betrachtete Effekt bei der Variation des Planfaktors auch einstellt. Sollte sich der Effekt nicht einstellen, scheitert die Regression und es muss ggf. eine Anpassung des Versuchsplanes erfolgen.

3.7 Präsentation

Entscheidend für den Erfolg der Hochdurchsatz-Untersuchung ist es, die Präsentation des Ergebnisses derart zu gestalten, dass die eingangs gestellte biologische Frage möglichst direkt beantwortet wird. Ist das nicht unmittelbar möglich, so muss versucht werden, möglichst exakte Hinweise zu liefern. Biologische Wirkzusammenhänge sowie die Komplexität der Auswertungskette machen es notwendig, nicht nur das Endergebnis, sondern auch den Lösungsweg transparent darzustellen. Hierzu ist es wichtig, auch Zwischenergebnisse jedes Moduls nachvollzieh- und interpretierbar darstellen zu können. Eingangsgrößen der Modulkategorie *Präsentation* sind die Klassifikationsergebnisse \hat{Y} bzw. \hat{Z} . Ausgangsgrößen sind die aufbereiteten Daten in diverser Form (vgl. Tabelle 3.2) sowie die Archivierung der Untersuchungsergebnisse.

Ist die Fragestellung beispielsweise die Suche nach der Effekt-Konzentration 50 (EC_{50}), bei der ein gesuchter Effekt bei 50% der Versuchseinheiten auftritt, so ist das Ergebnis der Hochdurchsatz-Untersuchung klar formulierbar. Entweder die Regression der Messwerte konvergieren bezüglich der Dosis-Wirkungs-Kurve und die gesuchte Konzentration EC_{50} lässt sich ermitteln oder die gewählten Konzentrationsreihen müssen angepasst werden. Ein möglichst genauer Hinweis für eine solche Anpassung kann durch die Darstellung der Klassifikationsergebnisse in Form eines Histogramms gegenüber der Konzentration erfolgen. Andere Fragestellungen können nicht direkt beantwortet werden, sondern es kann lediglich ein Hinweis durch die Daten gegeben werden, der von Experten gedeutet werden muss. Um einen guten Überblick über alle Daten zu erhalten, sind automatisch generierte Reportdateien ein geeignetes Mittel. Solche fassen für jede Larve oder jede Platte wichtige Ergebnisse übersichtlich zusammen.

Die beste Präsentationstechnik ist daher abhängig von der durchzuführenden Analyse zu wählen. In Tabelle 3.2 ist eine Auswahl solcher Techniken über bestimmten Problemstellungen aufgelistet, die sich bei Hochdurchsatz-Untersuchungen bewährt haben. Die Aufzählung ist keinesfalls erschöpfend. Eine etwas allgemeinere Aufzählung der genannten und weiterer Präsentationstechniken findet sich in [110]. Die Tabelle ordnet einer Art der Analyse jeweils eine mögliche Präsentationstechnik zu. In der Bildverarbeitung sind beispielsweise Merkmalsbilder, d.h. Bilder, die typische Bildinhalte in Form von Merkmalen repräsentieren und aus den Rohdaten (Rohbildern) berechnet werden, nützlich (vgl. Abschnitt 3.3). Auch eine Überlagerung solcher Merkmalsbilder mit den Rohdaten ist oftmals sehr aussagekräftig (vgl. z.B. Abb. 5.21). Ein weiteres Beispiel ist die Präsentation des Ergebnisses aus dem Anlernen eines Klassifikators. Das Ergebnis kann z.B. eine Diskriminanzfunktion sein. Hier eignet sich ein Scatterplot über den Lerndaten mit der Überlagerung der Diskriminanzfunktion. So kann bereits schon durch Anschauen des Ergebnisses eine Aussage über die Güte des Klassifikators getroffen werden (Ein Beispiel findet sich in Abb. 3.8).

Analyse	Präsentationstechnik
Bildverarbeitungsanalyse	Merkmalsbilder bzw. Sequenzen von Merkmalsbildern. Überlagerung von Rohdaten mit Merkmalsbildern
Diskriminanzfunktion	Scatterplot über Lerndaten mit Einblendung der Trennflächen
Ergebnisinterpretation	Sequenzen und Bilder des Bildstroms evtl. mit Einblendung der trennstärksten Merkmale
Falschfarbenbilder	Darstellung eines Intensitäts- oder Grauwertbilds durch Zuordnung einer Farbe meist als Überlagerung eines anderen Bildes (vgl. Abb. 2.7)
Gesamtübersicht	Reportdatei mit einer Auswahl der hier genannten Präsentationstechniken beziehend auf Plan- und Störfaktoren
Heatmaps	Darstellung der Werte einer Matrix in Form eines Bildes mittels einer Farbkodierung (vgl. Abb. 5.20)
Klassifikationsergebnisse	Klassenspezifische Histogramme, Ergebnistabellen, Konfusionsmatrix, Überlagerung von Bildern und Sequenzen des Bildstroms mit den Klassifikationsergebnissen, Scatterplot mit Klassenzuweisung und Diskriminanzfunktion
Merkmalsselektion	Sortierte Merkmalslisten mit Merkmalsrelevanzen
Merkmalsverteilung	Boxplot, Histogramm eines Merkmals
Merkmalsverteilung (paarweise)	Scatterplot, Falschfarbenbilder, Heatmaps
Rohdatenanalyse	Sequenzen und Bilder der Rohdaten sowie deren Grauwertverteilung, z.B. in Form von Histogrammen
Zeitreihen	2D-Plot, Heatmap, evtl. mit Einblendung von Klassifikationsergebnissen wie z.B. Bewegungsphasen und -ereignissen

Tabelle 3.2: Ausgewählte Präsentationstechniken für Hochdurchsatz-Untersuchungen

3.8 Bewertung

Der in Kapitel 3 vorgestellte Modulkatalog bietet eine Auswahl an neuen (wie z.B. Abschnitt 3.2.2 und Abschnitt 3.3.1) sowie bewährten Methoden (z.B. Abschnitt 3.2.1) für die Bildanalyse in Hochdurchsatzsystemen am Beispiel des Zebrabärblings. Je nach den Gegebenheiten des zu untersuchenden Nutzsignals und akquirierten Bildstroms ist es nun für eine Vielzahl von Versuchen möglich, passende Methoden auszuwählen und zu einer leistungsfähigen Auswertungskette zusammenzufügen. Der Katalog deckt dabei alle Verarbeitungsschritte von den Rohdaten bis hin zur Präsentation und Archivierung ab. Die in Abb. 3.1 übersichtlich nach Reihenfolge der Abarbeitung aufgelisteten Methoden sind vorzugsweise nach dem bereits in Abschnitt 2.6 vorgestellten Ablaufdiagramm auszuwählen und im Hinblick auf die Erfüllung der in Abschnitt 2.3 eingeführten Anforderungen im Zusammenspiel mit den anderen Versuchsparametern zu prüfen. Der modulare Aufbau ermöglicht zudem das leichte Ergänzen weiterer, an bisher unbekannte Problemstellungen angepasste Methoden. Die folgenden Kapitel zeigen die exemplarische Implementierung und Anwendung des erarbeiteten Konzeptes und des Modulkataloges.

4 Implementierung und Skalierung

4.1 Übersicht

Ziel der Implementierung für Hochdurchsatz-Untersuchungen ist es, die vorgestellten Module lauffähig umzusetzen. Die Implementierung soll möglichst einfach und anwenderfreundlich sein und somit allen Beteiligten der interdisziplinären Projekte zur Verfügung stehen. Des Weiteren soll Einfluss auf die Parameter bestehen und sowohl die Bilddaten als auch die zugehörigen Repräsentationen, Faktoren und Klassen müssen einsehbar und anschaulich präsentierbar sein. Ein Filtern der Daten auf Basis der Faktoren ist ebenso unverzichtbar wie die direkte Ausführbarkeit der Bildverarbeitungsalgorithmen. Zum Beispiel kann es von Interesse sein, alle Bildsequenzen einzusehen, die mittels eines bestimmten Mikroskops aufgenommen wurden oder alle Merkmalswerte eines bestimmten Merkmales, welches an einem bestimmten Tag aufgenommen wurde, zu plotten. Die Bilddaten werden meist auf einem zentralen, über eine Netzwerkverbindung angeschlossenen Server gespeichert und sollten für eine schnelle, parallele Berechnung auf mehrere Computer (sog. Clients) verteilbar ausgeführt werden können. Eine Parallelisierung wiederum erfordert eine zusätzliche Möglichkeit, die Teilergebnisse zu fusionieren. Der schnelle direkte Zugriff auf alle Methoden, Daten und Klassenzugehörigkeiten ist wichtig für ein gutes Verständnis des Datensatzes und zur Kontrolle auf Inkonsistenzen. So kann z.B. das Auflösen von Daten nach Positionen in der Mikrotiterplatte neue Störfaktoren aufzeigen. Aufgrund der interdisziplinären Problemstellung ist es zudem wichtig, dass die Auswertung sowie eine Anpassung der Parameter auf neue Datensätze weitgehend auch ohne Expertenwissen auf dem Gebiet der Informatik (z.B. von einem Biologen) ausführbar sind. Die Datentypen sind hierbei sehr heterogen, und es sind jeweils Bilddaten, Merkmale (und Zeitreihen) und Metadaten zuzuordnen und neu generierte Daten zu speichern. Die Bilddaten müssen prozessiert werden, es muss bekannt sein, mit welchem Algorithmus, welche Zwischenergebnisse erstellt wurden und wie die Einzelversuche miteinander zu vergleichen sind. Des Weiteren muss die Berechnung schnell erfolgen und die Ergebnisse müssen präsentierbar sein. Ein weiterer Aspekt ist die Archivierung. Dabei müssen nicht nur die Versuchsdaten und Ergebnisse reproduzierbar abgelegt werden, sondern auch die verwendete Version der Algorithmen und die gewählten Parameter archiviert werden.

Um die genannten Aufgaben zu erfüllen, wurde für die vorliegende Arbeit eine neue grafische Oberfläche (GUI) erstellt und die bereits bestehende MATLAB Toolbox Gait-CAD erweitert. Skriptbasierte Lösungen wurden verworfen, da solche den Nachteil haben, dass der Zugriff auf Parameter und Optionen oft kryptisch ist und eine lange Einarbeitungszeit

erfordert. In einer GUI kann jeder Parameter mit entsprechenden Hinweisen versehen werden. Auch ist eine GUI oft selbsterklärend oder die Auswirkung einzelner Optionen kann durch den Nutzer leicht in Erfahrung gebracht werden. Zudem müssen die Ergebnisse im Merkmalsraum oder die Merkmalsbilder ohnehin in einer grafischen Ausgabe dargestellt werden.

Zusammenfassend kann für die Implementierung gefordert werden, dass sie beziehend auf Abb. 2.1b und die Kategorien des Modulkatalogs gemäß Abb. 3.1

1. die Einzelversuche verknüpft mit
 - den zugehörigen Faktoren,
 - dem Bildstrom,
 - den Zwischenergebnissen,
 - der Repräsentation des Nutzsignals (Nutzsignalzeitreihen bzw. Nutzsignalmerkmale),
 - der Zuordnung zu den Ausgangsklassen.
2. Zugriff auf die Datenverarbeitung und deren Parameter erlaubt,
3. die Berechnung skalierbar macht,
4. die Ergebnisse präsentiert und
5. die Untersuchung und Ergebnisse archiviert.

Im Folgenden wird zu Beginn die Umsetzung der GUI und die Implementierung der Algorithmen beschrieben. Daraufhin wird auf die Möglichkeit zur Skalierung und zum verteilten Rechnen, wie es zum schnellen Berechnen der Daten notwendig ist, eingegangen.

4.2 Umsetzung und Methoden

Jede Hochdurchsatz-Untersuchung wird als ein abgeschlossenes Projekt betrachtet. Ein Projekt enthält für jede Versuchseinheit eine Sammlung an Daten oder Verlinkungen (vgl. Abb. 4.1). Eine Verlinkung ist eine Referenzierung auf den tatsächlichen Speicherort einer Datei. Das bedeutet, dass die Bilddaten nicht physisch im Projekt gespeichert werden, sondern lediglich auf die Datei verwiesen wird. Durch die Verweise bleiben die Projekte

handlich, es entstehen keine mehrfachen Kopien der Bilddaten und es ist dennoch der direkte Zugriff auf die Daten über die Verlinkung gegeben. Eine solche Sammlung wird für jede Versuchseinheit im Folgenden Datentupel genannt. Die Daten in einem Datentupel lassen sich in drei Gruppen einteilen: die Faktoren, die Bilddaten und die zu den Bilddaten gehörige Repräsentation des Nutzsignals. Mit der Repräsentation ist die jeweilige Quantifizierung des Nutzsignals aus dem Bildstrom bezeichnet, wie sie in Kapitel 3 durchgeführt wurde. Ein Projekt enthält zu Beginn nicht alle notwendigen Daten, sondern es wird durch die Methoden eine Reihe von Daten erzeugt, welche schließlich die Klassenzuweisung ermöglicht. Alle Daten lassen sich zu jedem Zeitpunkt präsentieren. Die Faktoren bestehen aus den Stör- und Planfaktoren sowie geschätzten Klassenzugehörigkeiten. Die Bilddaten sind der akquirierte Bildstrom und die Repräsentation deren Quantifizierung.

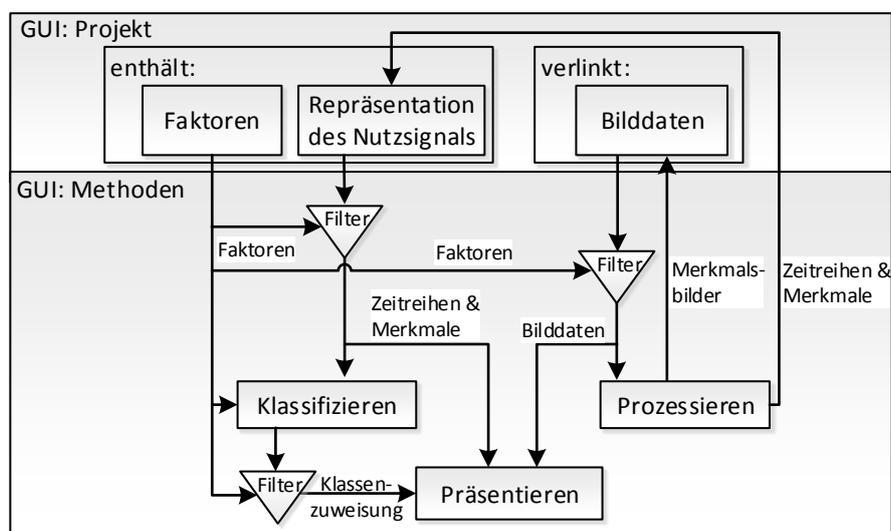


Abbildung 4.1: Prozessschritte und Signalfluss der grafischen Benutzeroberfläche zur Datenverarbeitung bei Hochdurchsatz-Untersuchungen

Bei Projekterstellung werden sowohl die Bilddaten als auch die zugehörigen bekannten Störfaktoren eingelesen. Die Störfaktoren werden in einem solchen Zusammenhang auch als Metadaten bezeichnet und sind üblicherweise entweder im Dateinamen und bzw. oder in einer zugehörigen Metadatei gespeichert. Wie in Abb. 4.1 dargestellt, verlinkt das dargestellte Projekt Bilddaten und enthält Faktoren sowie Repräsentationen der Bilddaten (Zeitreihen und Merkmale). Auf das Projekt lassen sich Methoden anwenden. Die zur Verfügung stehenden Methoden sind entweder bereits in Gait-CAD (vgl. Abschnitt 4.4.1) integrierte Funktionalitäten oder die Implementierung der Module aus Kapitel 3. Die Methoden lassen sich in drei Gruppen einteilen:

1. Methoden zum Prozessieren
2. Methoden zum Klassifizieren
3. Methoden zum Präsentieren.

Dabei bezeichnet der Begriff Prozessieren alle Berechnungen zur Ermittlung der Merkmale, die das Nutzsignal repräsentieren, somit also die Ermittlung von \mathbf{X} aus dem Bildstrom \mathbf{BS} (vgl. Abschnitt 2.5 bzw. Abb. 2.9). Die Methoden müssen dabei nicht zwingend auf die gesamten Daten angewandt werden, sondern es besteht die Möglichkeit, die Daten anhand der Faktoren zu filtern.

Nach dem Einlesen steht somit ein Projekt zur Verfügung, welches bereits zum Erkunden des Datensatzes verwendet werden kann. So lassen sich (je nach aufgezeichneten Metadaten) beispielsweise alle Bilddaten eines bestimmten Aufnahmedatums oder Laboranten filtern und darstellen. Bei der Repräsentation der Bilddaten handelt es sich immer um eine Quantifizierung. Die Quantifizierung ist Ergebnis der Bildverarbeitung und kann mit Hilfe der Benutzeroberfläche auf alle Daten angewandt werden. Ebenso ist es möglich, die Daten innerhalb der Oberfläche mittels der Faktoren zu filtern. Nach dem Prozessieren steht die Quantifizierung in Form von Fisch-Informationsmerkmalen zur Verfügung. Hierbei werden auch weitere Faktoren erzeugt. Beispielsweise wird gespeichert, ob die Bildverarbeitung erfolgreich war oder nicht. Sollte die Bildverarbeitung z.B. aufgrund von Fehlern oder Inhomogenitäten in den Bilddaten gescheitert sein, so können solche Datentupel von der späteren Verarbeitung ausgeschlossen werden. Zwischenschritte des Prozessierens, wie etwa Zeitreihen oder Merkmalsbilder, können gespeichert und präsentiert werden. Bei Bildern als Ergebnis der Berechnung wird wiederum nur ein Link zu den Bildern im Projekt gespeichert und die Datei physisch auf dem Datenspeicher abgelegt, während die Werte der berechneten Fisch-Informationsmerkmale und Fisch-Informationszeitreihen vollständig im Projekt abgelegt werden. Da es sich hierbei pro Bild, Merkmal und Abtastzeitpunkt im Normalfall um nur einen Wert handelt, bleibt die Größe eines Projektes auch bei einer großen Anzahl an Datentupeln überschaubar¹. Jeder Zwischenschritt lässt sich präsentieren und mittels der Faktoren filtern bzw. gegen die Faktoren plotten (beispielsweise in Histogrammen). Das Filtern und Plotten ist sowohl für Merkmalsbilder als auch für Zeitreihen und Merkmale möglich.

Da zur Erstellung eines Klassifikators ein Lerndatensatz benötigt wird, muss für dessen Erstellung zumindest für einen Teil der Datentupel die Klassenzugehörigkeit der Planfaktoren bekannt sein. Da die Klassenzuordnung von Experten anhand der Bilddaten erfolgt, wurde eine Funktionalität implementiert, die es ermöglicht, die zugehörigen Bilder oder Bildsequenzen von Datentupeln nacheinander anzeigen zu lassen und die Klassenzuordnung von Hand vorzunehmen. Der Vorgang wird als Labeln bezeichnet und kann in zufälliger Reihenfolge oder der Reihe nach erfolgen. Es besteht zusätzlich die Möglichkeit, die Klassenzugehörigkeit aus einer Datei einzulesen, sollte die Information über die Daten von einem Experten z.B. bereits in einer Excel-Datei abgelegt worden sein. Die Information wird als weiterer Faktor den Bildern zugeordnet und abgespeichert. Anhand der bereits quantifizierten Bilddaten und der neuen Information durch das Labeln ist es nun möglich, einen Klassifikator zu erstellen, der auf alle unbekanntenen Daten angewandt werden kann. Ist ein solcher Klassifikator z.B. anhand von Kontrollen erstellt worden, lässt sich der Klassifikator exportieren und in andere Projekte importieren. Ein solcher Export ermöglicht, unter der Voraussetzung konsistenter Datensätze, die Klassifikation weiterer

¹Typische Projektgröße: ca. 9 KB pro Datentupel

Projekte ohne die Notwendigkeit eines erneuten Labelns. Nach der Anwendung des Klassifikators sind alle Schritte abgeschlossen und können präsentiert werden. Es wurde zudem eine Funktionalität implementiert, die es gestattet, ein PDF-Dokument zu generieren, das alle wichtigen Daten in Form eines Reports übersichtlich zusammenstellt.

4.3 Skalierung

Die große Anzahl N an Einzelversuchen, welche in hohem Tempo in automatisierten Mikroskopen akquiriert werden, macht eine ebenso schnelle Datenverarbeitung notwendig. Die angewandte Methode zur Skalierung setzt ein schnelles lokales Datennetzwerk (LAN) voraus, ohne welches das einfache Übertragen der Daten mehr Zeit in Anspruch nimmt als sich durch verteiltes Rechnen einsparen lässt. Für die Datenübertragung stand bei Durchführung der vorliegenden Arbeit eine theoretische Übertragungsrate von 1000 Megabit/s (Gigabit-Ethernet) zur Verfügung. Das Prinzip der hier angewandten verteilten Berechnung ist in Abb. 4.2 dargestellt.

Eine Anzahl an Mikroskopen zeichnet parallel die Bildströme auf und kopiert die Rohdaten inkl. der Metadaten auf einen zentralen Datenspeicher. Auf den Datenspeicher greifen einzelne Computer (Clients in Abb. 4.2 als PC 1 bis PC j_p bezeichnet) zu. Auf den Clients läuft jeweils eine Instanz der Bildverarbeitung ab. In der vorliegenden Arbeit ist die Instanz auf den Clients entweder das Werkzeug Gait-CAD oder die Bildverarbeitungsoberfläche *PIMP* (vgl. Abschnitt 4.4). Da beim Einlesen, Durchsehen und Berechnen der Bilddaten mehr als einmal auf die Bilddaten zugegriffen werden muss, hat es sich als sinnvoll erwiesen, auf jedem Client eine lokale Kopie der zu prozessierenden Daten abzuspeichern und nach erfolgreicher Berechnung lediglich die Änderungen zurück auf den Server zu speichern. Hierfür wurde ein Programm der Firma Microsoft namens *Synctoy ver. 2.1* eingesetzt, welches sich automatisiert mit Skripten steuern lässt und für die Synchronisation über lokale Netzwerkstrukturen spezialisiert ist. Jeder Client erstellt in einer eigenen Instanz, wie im vorangegangenen Kapitel beschrieben, eine oder mehrere Projektdateien. In der Praxis hat es sich bewährt, je Mikroskopdatensatz eine Projektdatei anzulegen. Die aufbereiteten Daten (Merkmalsbilder etc.) werden von den Clients auf den Datenspeicher zurückgeschrieben und die Projektdateien werden nacheinander fusioniert. Wichtig ist hierbei, dass die Referenzierung auf die Daten nicht absolut, sondern relativ erfolgt und auf den Clients die gleiche Verzeichnisstruktur beibehalten wird wie auf dem Datenspeicher. Ein anderes Vorgehen führt zu fehlerhaften Referenzierungen der Bilddaten nach der Fusion. Nach Abschluss der Berechnung, Synchronisation und Fusion kann das finale Gesamtprojekt weiterverarbeitet werden. Üblicherweise wird ein Klassifikator geladen oder erstellt, die Klassenzuweisung wird durchgeführt und die Ergebnisse werden visualisiert.

Um sicherzustellen, dass alle Clients die identischen Arbeitsschritte selbstständig durchführen, wurde eine Kombination aus Makro- und Batchdateien erstellt. Eine Makrodatei enthält eine feste Abfolge von Befehlen, welche an einem (Teil)Projekt durchgeführt wer-

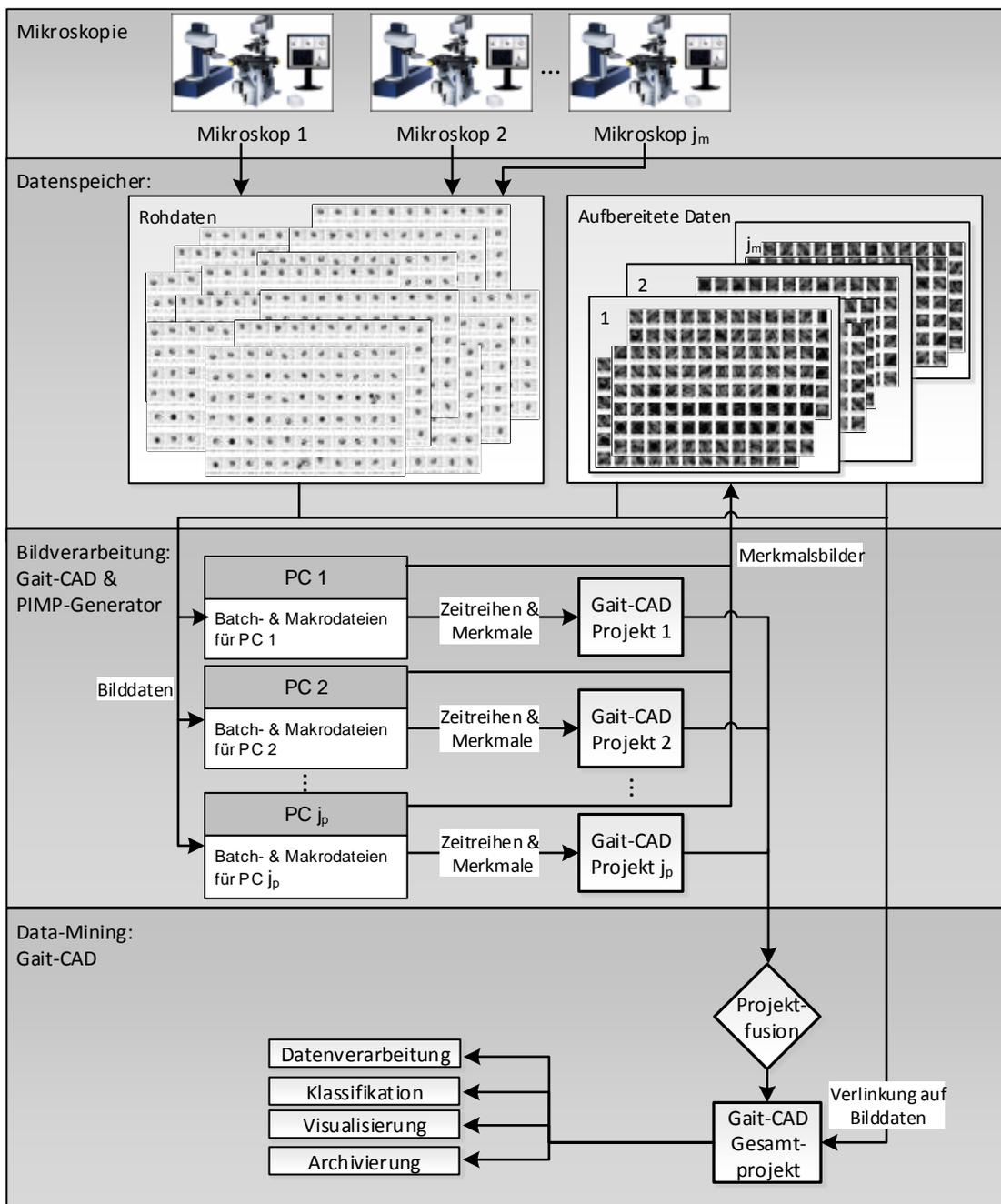


Abbildung 4.2: Skalierung für die Bildverarbeitung im Hochdurchsatzverfahren

den. Eine Batchdatei ist in der Lage, eine definierte Anzahl an Projekten zu öffnen und Makrodateien darauf anzuwenden. Die Makrodateien enthalten für jeden Mikroskopdatensatz die gleiche Vorgehensweise bei der Berechnung der Repräsentationen für die Bilddaten. Der einzige Unterschied von Client zu Client sind die zugeordneten Bilddaten je Client. Jeder Client enthält somit für jeden Mikroskopdatensatz einen Satz Makrodateien, welche nacheinander abgearbeitet werden. Die Aufspaltung in Makro- und Batchdateien ermöglicht eine erleichterte Fehlersuche. Es wird während der Ausführung jeweils eine Ereignisprotokolldatei (Logdatei) erstellt, welche für jeden Makro- und jeden Batchbefehl

einen Eintrag erhält, ob die Ausführung erfolgreich war oder nicht. Durch den vollständig automatisierten Ablauf des o.g. Prozesses muss der Anwender in den Batchdateien nun lediglich die Speicherorte der Bilddateien anpassen. Die Erstellung der Projekte und des Gesamtprojektes erfolgt dann automatisch durch die Abfolge der beschriebenen Batch- und Makrodateien. Das Erstellen der Batchdateien ist im Gegensatz zur Projekterstellung nicht vollständig grafisch implementiert, da in den Dateien lediglich die Referenzierungen zu den Bilddaten angepasst werden müssen und ein solch geringer Aufwand an Anpassung den zusätzlichen Aufwand einer grafischen Umsetzung nicht rechtfertigt. In der Implementierung wurde das parallele Rechnen auf mehreren Kernen eines modernen Prozessors ausgenutzt, welches, unter Einhaltung diverser Restriktionen, direkt von MATLAB mittels des MATLAB-Befehls *parfor* zur Verfügung steht. Zur Archivierung ist es ausreichend, die einzelnen Projektdateien und das Gesamtprojekt inkl. aller verwendeter Batch- und Makrodateien auf den Datenspeicher zu kopieren.

4.4 Werkzeuge

4.4.1 Programmpaket Gait-CAD

Am Forschungszentrum Karlsruhe werden seit 1998 Algorithmen zur Klassifikation entwickelt und aus der Toolbox KAFKA (**K**arlsruher **F**uzzy Modellbildungs-, **K**lassifikations- und datengestützte **A**nalyse-Toolbox) entstand ab dem Jahr 2001 die auf MATLAB basierende Toolbox Gait-CAD (Gait: englisch für Gang und CAD: **C**omputer **A**ided **D**iagnosis) [114, 115]. Ab dem Jahr 2006 wurde Gait-CAD mit einer grafischen Benutzeroberfläche als freie Software im Internet unter der GNU General Public License (GNU GPL) veröffentlicht [26, 111, 112, 164]. Die Anwendungsgebiete der Toolbox Gait-CAD haben sich von anfänglich ausschließlicher Ganganalyse in weitere Felder erstreckt, unter anderem für die vorliegende Arbeit auf das Feld der Bildverarbeitung mit zu Beginn Einzelbildern und schließlich auch Bildsequenzen.

Die Funktionalitäten von Gait-CAD wurden im Rahmen der vorliegenden Arbeit um Bildverarbeitungsroutinen und für die Verarbeitung von Bildströmen erweitert. Es stehen nun eine Reihe der vorgestellten Module in Form von Plugins zur Verfügung, die sich auf die Datensätze anwenden lassen. Die umgesetzten Module sind in Abb. 4.3 hervorgehoben und wurden sowohl für Einzelbilder als auch für Bildsequenzen implementiert. Alle dargestellten Module sind entweder in Form von Plugins verfügbar oder direkt von der GUI zu starten. Die bereits erwähnte Funktionalität zum Labeln wurde ebenfalls ergänzt. Die Auswertung des Anwendungsbeispiels *Fisch Embryo Test* (FET), vgl. Kapitel 5, wurde vollständig in Gait-CAD integriert und steht „auf Knopfdruck“ bis hin zur Erstellung eines automatischen Reports in Form eines PDF-Dokuments zur Verfügung. Die Abb. 4.4 zeigt einige Screenshots der Funktionalitäten. Im Vordergrund rechts ist eine EC50-Registrierung zu sehen. Im Vordergrund links ist ein Scatterplot über zwei Merkma-

le mit errechneter Diskriminanzfunktion dargestellt. Im Hintergrund ist eine beispielhafte Auswahl von Datentupeln sowie die Bedienoberfläche von Gait-CAD zu sehen.

Für das Anwendungsbeispiel des *Photomotor Response Screen* wurden Gait-CAD Projektdateien für alle Einzelversuche erstellt und durch Fusion in ein die gesamte Untersuchung umfassendes Projekt überführt. Innerhalb eines solchen Projektes ist es daraufhin möglich, Klassifikationen, z.B. bezüglich verschiedener Bewegungsmuster, durchzuführen. Die Bildverarbeitung erfolgt hier allerdings aus verschiedenen Gründen außerhalb von Gait-CAD wie im folgenden Abschnitt beschrieben wird (vgl. Abschnitt 4.4.2).

4.4.2 Grafische Oberfläche PIMP

Die ebenfalls in MATLAB entworfene Auswertungseinheit *PIMP* ist als grafische Benutzeroberfläche implementiert und enthält Routinen zum komfortablen Einlesen der Videosequenzen, der Auswahl und der manuellen Adaption von Auswerteparametern sowie dem Ausführen der Berechnung anhand der gewählten Parameter. Die Bildsequenzen lassen sich sowohl einzeln als auch in Gruppen einlesen und prozessieren. Vor der Durchführung der Berechnung wird automatisch eine Vorschau erstellt, bei der das Ergebnis der Segmentierung direkt ersichtlich ist und anhand derer Parameter visuell, auch ohne Expertenwissen, angepasst werden können. Abbildung 4.5 zeigt einige Screenshots der Oberfläche. Im Vordergrund ist ein Einzelversuch mit den Ergebnissen des Trackings und der Bewegungsklassifikation (vgl. Abschnitt 3.3.3) dargestellt. Dahinter ist eine Übersicht mit über 50 solcher Einzelversuche in Form einer Heatmap dargestellt. Darauf folgt ein Screenshot des Report-Generators sowie schließlich ein Screenshot des Hauptteils der Oberfläche, in welchem sich die Daten der Versuche auswählen und alle Funktionen starten lassen.

Eine neue Oberfläche wurde entworfen, da sie außerhalb des Gesamtpakets Gait-CAD eine Reihe von Vorteilen bietet. Die Oberfläche ist übersichtlich und beschränkt sich auf die für den *PMR* notwendigen Parameter. *PIMP* wurde so konzipiert, dass es auch ohne eine vollständige MATLAB-Installation (mit den entsprechenden MATLAB-Runtime-Paketen) lauffähig ist. Zusätzlich dazu ist es plattformunabhängig und sowohl auf WINDOWS wie auch auf MAC/OS und UNIX Betriebssystemen lauffähig. Das ermöglicht die Ausführung auf beliebig vielen Computern, welche lokal, z.B. direkt neben den aufzeichnenden Mikroskopen, platziert sind und spart das „doppelte“ Kopieren wie bei der auf Gait-CAD basierten Lösung, bei welcher vom Mikroskop auf den Datenspeicher, von dem Datenspeicher auf den Client und schließlich wieder zurück kopiert werden muss. Lediglich die fertigen Daten werden zur Archivierung auf den Datenspeicher einmalig abgelegt. Damit wird die neue Softwarelösung auch attraktiv für Labore, die weder über ein schnelles lokales Netzwerk noch über große ständig erreichbare Speicherkapazitäten verfügen.

²www.mathworks.de/matlabcentral/fileexchange/26978-hough-transform-for-circles

³www.mathworks.de/matlabcentral/fileexchange/25500-peakfinder

⁴www.mathworks.de/matlabcentral/fileexchange/28590-template-matching-using-correlation-coefficients

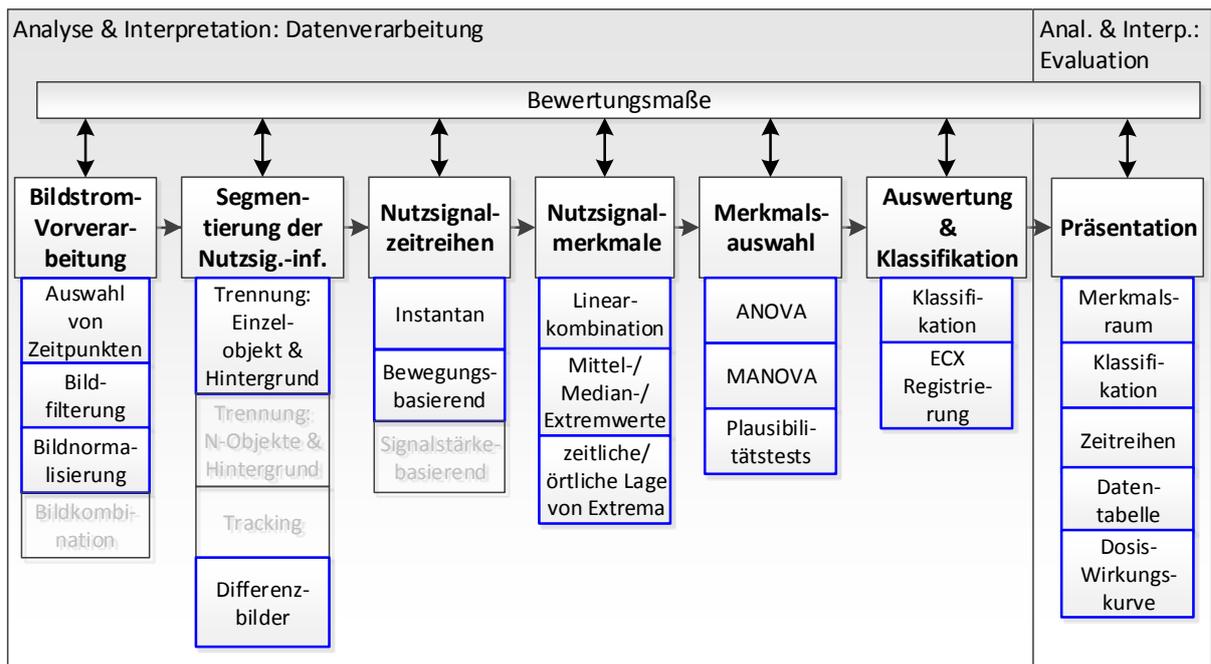


Abbildung 4.3: Die in Gait-CAD implementierten Module des Modulkatalogs sind hervorgehoben.

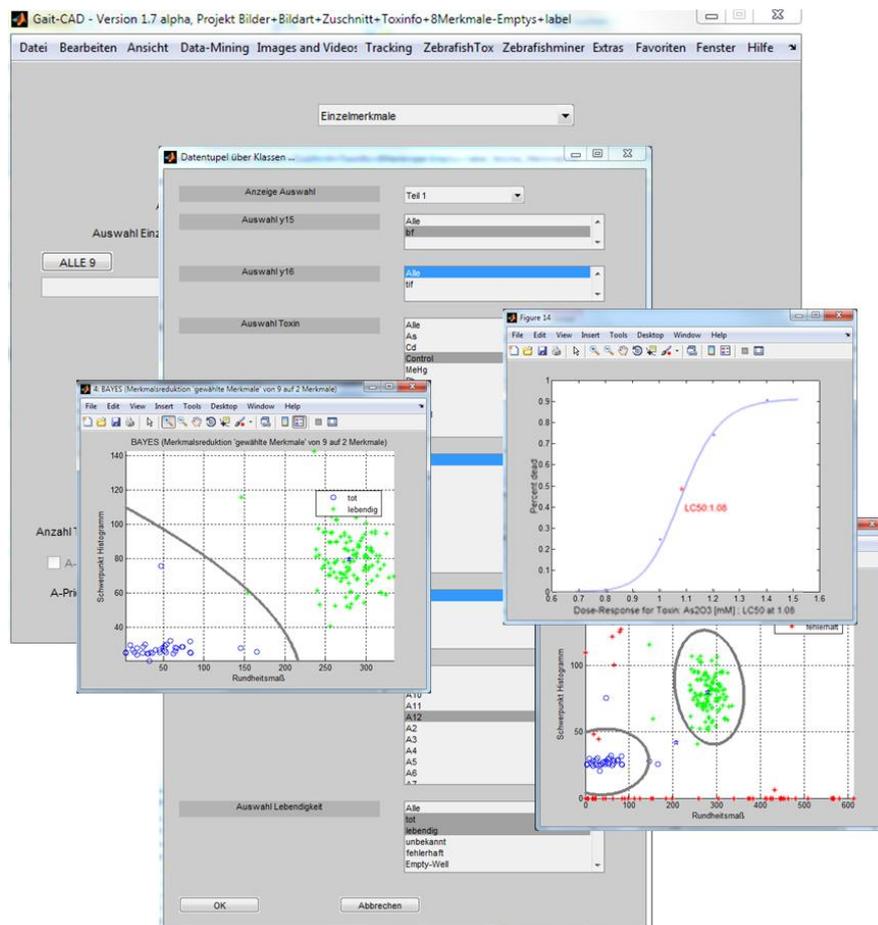


Abbildung 4.4: Gait-CAD: Die TOX-Erweiterung für den FET

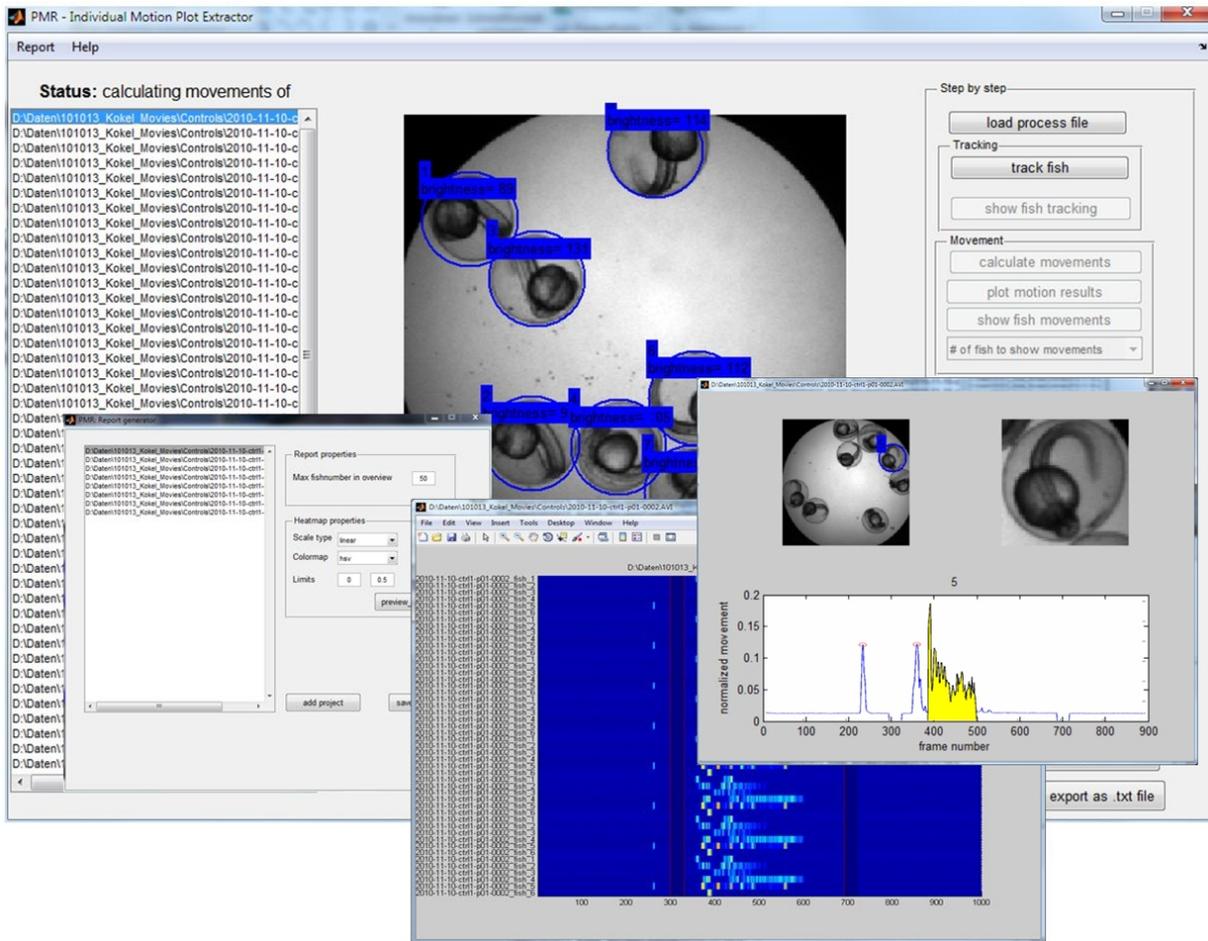


Abbildung 4.5: PIMP: Die grafische Oberfläche zum Einlesen, Auswerten und Visualisieren von PMR-Bildsequenzen

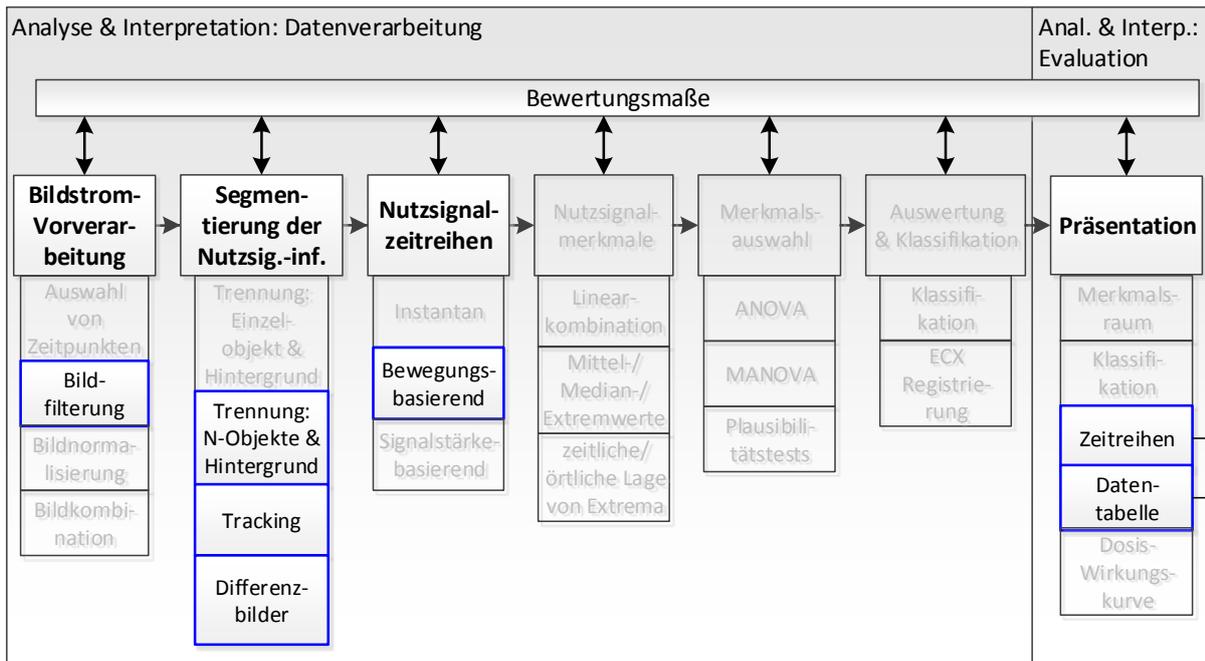


Abbildung 4.6: In PIMP implementierte Module des Modulkatalogs (Nicht implementierte Module sind ausgegraut.)

Programm	Beschreibung	Author	Lizenz
CombineZ	Auf die Kombination mehrerer Fokusaufnahmen spezialisierte, batchfähige Software	Alan Hadley	GPL
Hough transform for circles ²	Implementierung des Hough-Algorithmus zur Kreisdetektion	David Young	BSD
ImageJ	Bildbearbeitungs- und Bildverarbeitungsprogramm	Wayne Rasband	Public Domain
IrfanView	Bildbetrachter mit umfangreichen Funktionen	Irfan Škiljan	Freeware
LabVIEW add-on: Vision/Image Processing	Bildverarbeitungserweiterung für LabVIEW	National Instruments	Proprietär
MATLAB Image Processing Toolbox	Umfangreiche Sammlung von Bildverarbeitungs- und Bildanalysefunktionen	Mathworks	Proprietär
PeakFinder ³	Implementierung zur Minima- und Maximaldetektion innerhalb verrauschter Daten	Nate Yoder	BSD
Sync-Toy	Tool zur Synchronisierung von Dateien und Ordnern	Microsoft	Freeware
Template Matching using Correlation Coefficients ⁴	Implementierung des Template-Matching Verfahrens	Yue Wu	BSD
TortoiseSVN	Windows-Client für den Versionsverwaltungsdienst Subversion-SVN	The TortoiseSVN team	GPL

Tabelle 4.1: Übersicht der Fremdimplementierungen und verschiedener Tools, welche in Kombination mit der in Kapitel 4 vorgestellten Software zum Einsatz kommen.

Zudem wurde die Option implementiert, bereits prozessierte Daten erneut einzulesen und so einzelne Bewegungsmuster einer jeden Versuchseinheit zu untersuchen und gegen andere zu plotten. Die Option ermöglicht das schnelle Inspizieren auffälliger Ergebnisse. Der zusätzlich integrierte Reportgenerator ermöglicht das Erstellen eines ausführlichen Berichtes mit einer Zusammenfassung sowie Zwischenergebnissen im PDF- oder wahlweise im HTML-Format. Die erzeugten Ergebnisse werden zudem für die weitere Verarbeitung und Klassifikation in Form von Gait-CAD Projekten abgespeichert. Die innerhalb von *PIMP* umgesetzten Module des Modulkatalogs sind in Abb. 4.6 hervorgehoben. Tabelle 4.1 bietet eine Übersicht verschiedener, bei der Umsetzung verwendeter Fremdimplementierungen sowie einige Tools, welche sich während der Arbeit bewährt haben. Beispielsweise das auf der Programmiersprache JAVA aufbauende *ImageJ*, welches durch den offengelegten Quellcode leicht erweiterbar ist oder auch die außerordentlich leistungsfähige, wenn auch proprietäre, Bildverarbeitungs-Erweiterung in der Software MATLAB.

PIMP wird im Rahmen der vorliegenden Arbeit zur Auswertung der Datensätze in Abschnitt 5.3 verwendet und ist darüberhinaus derzeit als Auswerteeinheit der am KIT entwickelten Plattform für Hochdurchsatz-Untersuchungen am Zebrafisch im Einsatz [103]. Desweiteren befindet sich die Software derzeit im Labor von *Randall T. Peterson*⁵ sowie im Labor von *David Kokel*⁶ der Harvard Medical School und des Massachusetts General Hospital (Boston, USA) im Routineeinsatz.

4.5 Bewertung

Die vorgestellte Implementierung, die grafische Oberfläche und das Skalierungskonzept lassen sich für die Auswertung beliebiger Daten von Hochdurchsatz-Untersuchungen anwenden und falls notwendig adaptieren. Während das Werkzeug *PIMP* für den Photomotor Response Screen spezialisiert ist, lässt sich das Programmpaket Gait-CAD durch Plugins leicht auf neue Problemstellungen anpassen bzw. erweitern. Der grundlegende Ablauf vom Einlesen der Bilddaten bis hin zur Klassifikation bleibt hierbei identisch. Es muss für weitere Problemstellungen lediglich die Segmentierung ausgetauscht und anhand gelabelter Daten ein individueller Klassifikator entworfen werden. Somit steht nun, mit den vorgestellten Werkzeugen, eine zwar prototypische, jedoch vollständige Implementierung für Hochdurchsatz-Untersuchungen zur Verfügung. Eine erste Anwendung ist in [81] zu finden. Im folgenden Kapitel werden die konzeptionellen und theoretischen Methoden an zwei praktischen Problemstellungen erprobt.

⁵<http://www.rtplab.org/>

⁶<http://kokellab.com/>

5 Anwendungen

5.1 Übersicht

Um die Leistungsfähigkeit des neuen Konzeptes nachzuweisen, wird es im Folgenden auf zwei Probleme, die sich mit Hochdurchsatz-Untersuchungen lösen lassen, angewandt. Dabei werden die Problemstellung und damit der zu untersuchende Bilddatensatz jeweils komplexer. Die vorgestellten Projekte sind:

- eine toxikologische Untersuchung
 - anhand von Einzelaufnahmen von vereinzelt Zebrabärblingen (Abschnitt 5.2.2) und
 - anhand von Bildsequenzen von vereinzelt Zebrabärblingen (Abschnitt 5.2.3).
- eine Hochdurchsatz-Untersuchung anhand von Bildsequenzen von mehreren Zebrabärblingen (Abschnitt 5.3).

iAlle Projekte werden mittels des eingeführten Konzeptes entworfen. Es folgt die Erstellung einer Auswertungskette durch Auswahl geeigneter Module aus Kapitel 3. Die Ergebnisse werden anschließend präsentiert und diskutiert. Weitere, nicht im Rahmen des Kapitels vorgestellte Anwendungen, auf die entweder direkt Einfluss genommen wurde oder die auf die Ergebnisse der vorliegenden Arbeit aufbauen, finden sich in [103,132,135,136,148]. Die Anwendungen wurden aus verschiedenen Gründen nicht in das Kapitel 5 aufgenommen: [132,135] haben ihren Schwerpunkt weniger in der Bildverarbeitung und -analyse als in der Automatisierungstechnik. In der Anwendung von [136,148] werden einige Aspekte des hier vorgestellten Konzeptes umgesetzt und eine Bildverarbeitungsroutine zur Erkennung des Herzschlags und zur Extraktion der Herzfrequenz erarbeitet. Obwohl die Ergebnisse vielversprechend sind, war deren Umfang zum Zeitpunkt des Entstehens der vorliegenden Arbeit noch nicht ausreichend für eine Aufnahme in das folgende Kapitel.

5.2 Fisch Embryo Test (FET)

Der FET stellt ein geeignetes Einsatzszenario für das erarbeitete Konzept dar, da alle Teilschritte ein hohes Potenzial zur Automatisierung bieten und in Teilbereichen wie z.B. der Präparation bereits Automatisierungstechnik zur Verfügung steht. Die bisher manuell erfolgende Klassifikation von sog. *Endpunkten*, wie sie in den OECD-Richtlinien definiert sind [121, 122], zu automatisieren, ist der offensichtlich nächste Schritt zur Vervollständigung der Auswertungskette. Im o.g. Standard sind mehrere dieser *Endpunkte* definiert. Hierbei handelt es sich jeweils um dichotome Prüfungen, d.h. jede Larve erreicht einen der definierten Endpunkte oder nicht. Die Endpunkte und deren Erscheinungsformen sind:

- Koagulation

Die Koagulation ist die einsetzende Gerinnung der Proteine im Ei. Bei unbefruchteten oder sich nicht entwickelnden Organismen tritt Koagulation ein. Sie ist in den Mikroskopaufnahmen durch einen dunklen Fleck im Inneren des Eies erkennbar. Koaguliertheit wird klassifiziert, wenn keine Larve sondern der beschriebene Fleck im Bild vorhanden ist.

- Herzschlag

Der ab ca. 24 hpf einsetzende Herzschlag ist in der gesunden Larve, je nach Lage des Herzens, unterschiedlich deutlich sichtbar. Auch der Blutfluss in der Larve ist durch die Transparenz des Gewebes z.T. erkennbar und ein indirektes Maß für den Herzschlag. Der Herzschlag wird als vorhanden klassifiziert, wenn die Kontraktion des Herzens oder das Fließen der Blutkörperchen in der Bildsequenz erkennbar ist.

- Spontanbewegung

Die ebenfalls nach ca. 24 hpf eintretenden spontanen Bewegungen der Larve innerhalb des Chorions sind in Bildsequenzen durch ein Verwinden der Larve im Ei erkennbar. Die Spontanbewegung wird als vorhanden klassifiziert, wenn die Larve ihre Position im Bild deutlich verändert, d.h. eine klar andere Position eingenommen hat, wie z.B. eine 180° Drehung.

- Ablösung des Schwanzes vom Dottersack

Nach ca. 24 hpf löst sich während der Entwicklung der Larve der Schwanz vom Dottersack. Das Ablösen des Schwanzes wird als vorhanden klassifiziert, wenn ein Freiraum zwischen dem Schwanz und dem Dottersack in den Mikroskopaufnahmen erkennbar ist, d.h. der Schwanz der Krümmung des Dottersacks nicht unmittelbar folgt.

- Somiten

Ab ca. 16 hpf werden im Schwanz und Rücken der Zebrabärblingslarve sog. Somiten sichtbar. Somiten stellen sich auf den Bildern als regelmäßige Linien im Schwanz dar und können je nach Position des Schwanzes u.U. verdeckt oder auch unscharf abgebildet sein. Somiten werden als vorhanden klassifiziert, wenn die Struktur im Rücken der Larve erkennbar ist.

Für jeden der Endpunkte soll gemäß des Standards eine ebenfalls dichotome Klassenzuweisung erfolgen. Nach Einschätzung von Experten ist jedoch sowohl die Ablösung des Schwanzes vom Dottersack als auch die Ausbildung von Somiten im Rücken selbst vom Menschen nicht immer scharf einer der Klassen $\{ja, nein\}$ zuzuordnen und eignet sich somit weniger für eine automatische Klassifikation. Es wird im Folgenden daher nur auf die ersten drei der genannten Endpunkte eingegangen.

5.2.1 Auswirkungen der Orientierung der Larve im Ei

Die Larven können sich ab einem Alter von ca. 48 hpf innerhalb des Eies frei bewegen. Damit ist die Präsenz der biologisch relevanten Information in den Daten lageabhängig. Bei einer entwickelten Fischlarve sind je nach Position mehr oder weniger Organe sichtbar. Der Dottersack, mindestens ein Auge und der Rücken sind fast immer zu identifizieren. Feinere Strukturen wie das Herz, die Leber usw. sind nur schwer auszumachen. Schlägt die Entwicklung komplett fehl, zersetzt sich das Ei und zeigt Koagulation. Um in der Lage zu sein, eine Hochdurchsatz-Untersuchung erfolgreich durchzuführen, muss für einen akquirierten Datensatz untersucht werden, inwieweit sich der Datensatz zur automatisierten Klassifikation der o.g. Endpunkte eignet. Das Ziel der folgenden Analyse ist somit, die akquirierten Daten bezüglich den spezifischen Anforderungen der genannten Endpunkte zu untersuchen.

Die theoretisch mögliche Anzahl von Orientierungen, die eine sich frei bewegende Larve innerhalb des Chorions annehmen kann, ist durch die kontinuierliche Positionierung unendlich groß. Es lässt sich jedoch eine Anzahl von typischen Fällen mit geometrischen Eigenschaften in Gruppen zusammenfassen, welche potenziell für die spezifische Untersuchung bezüglich der vorgestellten Endpunkte herangezogen werden können. Die Gruppen werden im Folgenden herausgearbeitet.

Ein typischer Datensatz wird bei ca. 48 hpf akquiriert, da nach OECD-Norm sich hier alle der o.g. Endpunkte prüfen lassen [121, 122]. Bei einem Entwicklungszeitpunkt von 48 hpf sind verschiedene Bereiche und Details in der Larve bereits unterscheidbar. Je nach Größe sind die Details jedoch zum Teil nur schwer zu erkennen oder werden bei ungünstiger Lage der Larven von verschiedenen Körperteilen verdeckt. Die wichtigsten Bereiche, die sich in nahezu jedem Bild abgrenzen lassen, sind nach aufsteigendem Schwierigkeitsgrad der Auffindbarkeit geordnet: das Auge, der Dottersack, der Rücken und der Schwanz.

Die Vorgehensweise bei der Einteilung in Klassen orientiert sich vornehmlich daran, welche Bereiche der Larven gut zu erkennen sind. Hierfür werden zur leichteren Klassifikation mehrere Merkmale zur Positionsbestimmung definiert. Markante Bereiche liegen üblicherweise im Vordergrund und werden nicht durch mehr oder minder transparente andere Körperteile verdeckt. Das erste Merkmal ist somit das markanteste Merkmal am Kopf: die Augen, welche einen sehr starken Kontrast aufweisen und zumeist als dunkle Flächen oder „Scheiben“ erscheinen. Das zweite Merkmal ist der Dottersack bzw. wie viel Prozent des Dottersacks ohne Überdeckung, d.h. im Vordergrund des Bildes, erscheinen. Das dritte Merkmal zur Klasseneinteilung stellt der Rücken dar. Gemessen wird der Abstand der Konturlinie der Larve, also das Ende des Rückens, bis zum Dottersack. Der Wert gibt an, wie viel Prozent des Durchmessers sichtbar ist, bezogen auf den maximalen Durchmesser einer „normal“ entwickelten Larve bei einem Alter von 48 hpf. Das letzte Merkmal ist der Schwanz. Da keiner der ausgewählten Endpunkte Details innerhalb des Schwanzes ausgewertet, ist das Auftauchen des Schwanzes im Vordergrund eher hinderlich, weil es dann zumeist andere Körperteile verdeckt. Das Merkmal misst die Menge an vom Schwanz verdeckten anderen Körperteilen im Verhältnis zum Gesamtvolumen.

Für die Einordnung der Lage der Larven werden in der vorliegenden Arbeit vier Klassen definiert. Die Einteilung der Klassen erfolgt anhand der oben definierten Merkmale durch Anwendung folgender Regeln, welche aus Erfahrungswerten abgeleitet sind:

seitlich Die weitaus größte Anzahl der Larven liegt auf der Seite. Bezogen auf die eingeführten vier Merkmale liegt eine Larve seitlich, wenn ein Auge deutlich im Vordergrund zu sehen ist, während das andere nicht im Vordergrund ist und sich im optimalen Fall sogar deckungsgleich auf der „Rückseite“ der Larve befindet. Beim Merkmal der Augen wird das Auge, welches einen größeren Anteil an Pixeln im Vordergrund aufweist, als Auge 1 und das andere als Auge 2 bezeichnet. Somit ist für das Merkmal *Auge* bei der Klasse *Seitlich* Auge 1 zu mehr als 90% im Vordergrund festgelegt, während von Auge 2 nicht mehr als 5% im Vordergrund erkennbar sein darf. Der Dottersack ist immer im Bild sichtbar und kann lediglich durch den Schwanz oder den Rücken teilweise überdeckt werden. Da bei seitlicher Lage der Rücken zumeist über der Larve liegt, sollte der Dottersack nicht oder nur unwesentlich verdeckt sein. Als zusätzliches Kriterium ist somit eine Sichtbarkeit des Dottersacks von 80% festgelegt. Eine weitere Voraussetzung für die seitliche Lage ist, dass der Rücken zu mehr als 50% über dem Dottersack sichtbar ist und sich weder vor noch hinter anderen Körperteilen befindet. Da bei einer Überdeckung der Organe der Larve durch deren Schwanz solche wichtigen Bereiche im vorderen Teil der Larve nicht erkennbar werden, wird gefordert, dass die Überdeckung des Schwanzes 5% des Volumens der Larve nicht überschreiten darf.

von unten Die Klasse beinhaltet alle Larven, bei denen ein Teil des Rückens durch den Dottersack verdeckt ist und beide Augen zumindest teilweise im Vordergrund erkennbar sind. Bezogen auf die Klasse wird gefordert, dass Auge 1 zu mehr als 90% sichtbar und Auge 2 zu mehr als 60% sichtbar ist. Da der Dottersack den Rücken verdeckt, ist er fast vollständig im Vordergrund und es werden mehr als 80% im Vor-

dergrund für die Klasseneinteilung gefordert. Der Rücken muss verdeckt sein und befindet sich somit zu höchstens 20% im Vordergrund. Eine Überdeckung durch den Schwanz soll möglichst ausgeschlossen werden. Die Überdeckung des Schwanzes soll daher 5% des Volumens der Larve nicht überschreiten.

von oben Hierzu werden alle Larven gezählt, bei denen der Dottersack durch den Rücken verdeckt bzw. in zwei Bereiche geteilt wird und beide Augen zumindest teilweise im Vordergrund sichtbar sind. Die Augen im Vordergrund sind identisch zur Lage *von unten* definiert mit mehr als 90% für Auge 1 und mindestens 60% für Auge 2. Da jedoch der Dottersack in der Lage *von oben* durch den Rücken verdeckt ist, sind die Voraussetzungen gegenüber *von unten* umgekehrt. Der Dottersack muss zu maximal 60% und der Rücken zu mindestens 50% im Vordergrund sichtbar sein. Die Überdeckung des Schwanzes soll daher 5% des Volumens der Larve nicht überschreiten.

hinten Zur Klasse *hinten* werden die Larven gezählt, bei denen kein Auge im Vordergrund sichtbar ist, da der Kopf vom Dottersack verdeckt wird. Auge 1 soll für die Klasse *hinten* zu maximal 10% und Auge 2 zu maximal 5% sichtbar sein. Der Dottersack befindet sich im Vordergrund zu mehr als 50%. Eine Überdeckung des Rückens oder des Schwanzes ist für die Klasse unerheblich und es werden alle Werte akzeptiert, d.h. 0% - 100%.

undefiniert Larven, die sich keiner der übrigen vier Klassen zuordnen lassen, wurden als *undefiniert* klassifiziert. Die klare Zuordnung scheitert in vielen Fällen dann, wenn die Larve schräg oder verdreht im Ei liegt.

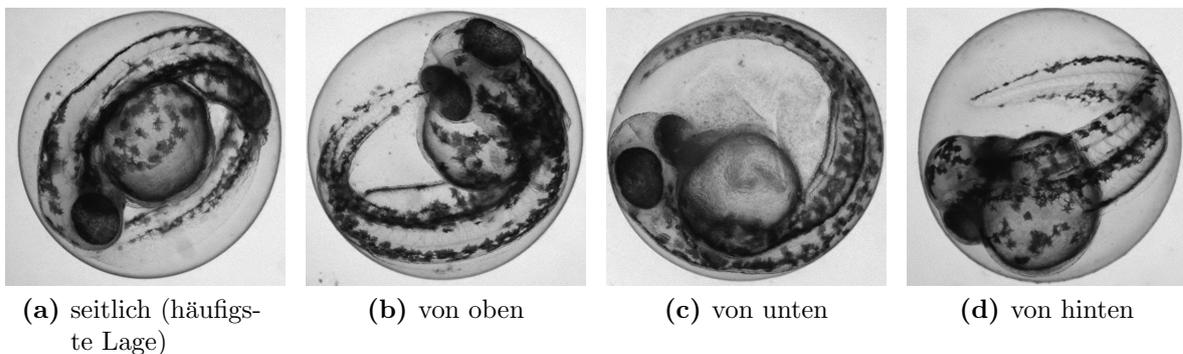


Abbildung 5.1: Beispiele von Zebrafärlingslarven in unterschiedlichen Orientierungen im Ei

In Tabelle 5.1 sind die Regeln der Einteilung zusammengefasst. Ein Beispiel für jede Position ist in Abb. 5.1 gegeben. Ebenso sind hier die Ergebnisse einer manuellen Zuordnung von über 240 Zebrafärlingslarven angegeben. Die Einteilung zeigt, dass nur wenig mehr als die Hälfte aller Larven in einer gleichen Position (der seitlichen Position) liegen. 25% der Larven passen in keine der definierten Klassen. Eine Bildauswertung, die Nutzsignale in der Larve auswertet, muss daher von der Lage der Larven unabhängig sein, wenn keine Larven aufgrund der Lage verworfen werden sollen. Bei einer automatischen Erkennung der vier beschriebenen Klassen können zwar etwa 75% der Larven ausgewertet werden,

allerdings setzt dies voraus, dass das Nutzsignal ebenfalls in jeder Lage sichtbar ist. Des Weiteren muss für jede auszuwertende Klasse eine gesonderte Bildauswertung entwickelt werden. Eine Beschränkung auf die häufigste Klasse hätte einen sehr hohen Ausschuss an Einzeluntersuchungen zur Folge, da etwa die Hälfte aller Versuche nicht ausgewertet werden können, was den Durchsatz halbiert. Beide Optionen eignen sich somit wegen des hohen Entwicklungs- bzw. Kostenaufwands nicht für den Praxiseinsatz.

Klasse	Augen		Dottersack	Rücken	Schwanz	Anzahl
	Auge 1	Auge 2				
seitlich	$\geq 90\%$	$\leq 5\%$	$\geq 80\%$	$\geq 50\%$	$\leq 5\%$	133
von oben	$\geq 90\%$	$\geq 60\%$	$\leq 60\%$	$\geq 50\%$	$\leq 5\%$	15
von unten	$\geq 90\%$	$\geq 60\%$	$\geq 80\%$	$\geq 20\%$	$\leq 5\%$	13
hinten	$\leq 10\%$	$\leq 5\%$	$\geq 50\%$	0-100%	0-100%	18
undefiniert						61
Summe						240

Tabelle 5.1: Zusammenfassung der Regeln zur Klassenzuordnung und manuelle Zuordnung von 240 unbehandelten Zebrabärblingslarven; Augen, Dottersack: Prozent Pixel sichtbar im Vordergrund; Rücken: Prozent Pixel des Rückendurchmessers sichtbar; Schwanz: Verdeckung anderer Teile der Larve in Prozent.

Zusammenfassend zeigt die Untersuchung, dass Larven im Ei sich für Hochdurchsatz-Untersuchungen nur eignen, wenn deren Nutzsignal unabhängig von der Lage der Larve sichtbar bzw. quantifizierbar ist. Für alle anderen Untersuchungen in der Larve muss auf einen zusätzlichen Präparationsschritt zurückgegriffen werden, um eine definierte Position sicherzustellen. Das Ergebnis ist somit, dass unter Verwendung des vorliegenden Datensatzes eine Auswertung nur Erfolg verspricht, wenn lageunabhängige Merkmale für jeden Endpunkt gefunden werden. Ist das nicht möglich, muss die Biologie und Bildakquise erneut mit einem höheren Informationsgehalt, beispielsweise im 3D-Raum, durchgeführt oder unter Verwendung von fluoreszierenden Markern erfolgen, welche weniger empfindlich auf Verdrehungen der Probe reagieren, wie z.B. in [35, 49] gezeigt .

5.2.2 Fisch Embryo Test (FET) – Koagulation & EC₅₀ Regression

Der FET soll in einer Hochdurchsatz-Untersuchung zur Ermittlung der Sterberate bei unterschiedlichen Konzentrationen für mehrere Toxine eingesetzt werden. Zebrabärblingseier sollen hierzu einzeln auf Koagulation untersucht werden und der Einfluss der Toxine bei steigender Konzentration ist zu bestimmen. Im Allgemeinen werden sich bei einer sehr niedrigen Konzentration fast alle Larven normal entwickeln (Sterberate bei wenigen Prozent), während bei einer sehr hohen Konzentration keiner der untersuchten Modellorganismen überlebensfähig ist (Sterberate 100%). Die zu untersuchenden Toxine und die Verdünnungsreihen je Konzentration sind in Tabelle 5.2 gegeben. Die Untersuchungen wurden am KIT durchgeführt.

Toxin	Einheit	Dosis									
		0.80	0.90	1.00	1.20	1.40					
As ₂ O ₃	[mM]	0.80	0.90	1.00	1.20	1.40					
CdCl	[mg/l]	20	25	35	40	50	60	80	100	120	150
Ethanol	[%]	1.5	2.0	2.25	2.5	3					
Methanol	[%]	3.00	4.00	4.50	5.00	6.00					
PbCl ₂	[mg/l]	80	100	120	140	160	180	200	220		

Tabelle 5.2: Toxine und Verdünnungsreihen des untersuchten Datensatzes

Für die erfolgreiche Auswertung der Hochdurchsatz-Untersuchung und zur Anwendung des erarbeiteten Konzeptes wird das in Abb. 2.11 vorgestellte Flussdiagramm zur anforderungskonformen Auswertung durchlaufen. Es werden die übergeordneten Blöcke der Versuchsplanung „Biologie und Bildakquise“ sowie „Analyse und Interpretation“ abgearbeitet. Jede Kategorie des Modulkataloges wird betrachtet und passende Module werden ausgewählt und angewandt. Schließlich wird die Hochdurchsatz-Untersuchung durchgeführt und die Ergebnisse werden präsentiert. Die Anforderungen werden einzeln überprüft. Ist eine Anforderung erfüllt, wird deren Kurzbezeichnung als Verweis auf Abschnitt 2.3 in Klammern angegeben.

5.2.2.1 Versuchsplanung: Biologie und Bildakquise

Die Bestimmung des biologischen Effekts, d.h. des Nutzsignals, ist nach der eingeführten Vorgehensweise der erste Schritt und beim FET durch die OECD-Norm bereits durchgeführt. Ein Beispiel der zu unterscheidenden Fälle ist in Abb. 5.2a gegeben. Der markanteste Unterschied zwischen einer sich entwickelnden Larve und dem Eintreten der Koagulation ist der schwarze Punkt im Zentrum des Eies. Die Anforderungen (vgl. Abschnitt 2.3) für die Durchführung können als erfüllt betrachtet werden, da die Versuche der OECD-Norm den Gesetzmäßigkeiten entsprechen (Legitimität), die Larven in ausreichender Menge zur Verfügung stehen (logistische Machbarkeit), die Endpunkte reproduziert werden können (Reproduzierbarkeit) und auch die notwendigen Ressourcen zur Verfügung stehen (Finanzielle Realisierbarkeit).

Für die Überprüfung der Anforderungen an die Messung müssen deren Randbedingungen abgesteckt werden. Die Messung erfolgt wie folgt: Die Fischlarven werden im Alter von 4 hpf den Toxinen ausgesetzt, und ab 48 hpf werden die Bilddaten akquiriert. Hierfür wird in jedes Nöpfchen von sieben Mikrotiterplatten jeweils eine Larve mittels einer Pipette transportiert. In jede Reihe wird eine höhere Konzentration eines Toxins gegeben. In jeder Mikrotiterplatte werden fünf unterschiedliche Konzentrationen untersucht, wobei eine Reihe von Larven unbehandelt ist und die Negativ-Kontrolle darstellt. Jeweils zwei Reihen der Mikrotiterplatte blieben leer, somit sind 72 Nöpfchen pro Platte besetzt. Jede Untersuchung wird doppelt ausgeführt (Replika vgl. Abschnitt 1.2.1). Insgesamt werden für die Untersuchung somit 1008 Eier von Zebrafärbblingslarven verwendet (zweimal sieben Platten mit je 72 Larven). Das Mikroskop erfasst zu jedem Zeitpunkt immer genau ein

Näpfchen. Da die Steuer-Software des Mikroskops es nur erlaubt, die gesamte Platte zu akquirieren, werden auch Aufnahmen von den leeren Näpfchen aufgezeichnet, was sich auf die Akquise-Dauer niederschlägt. Daher ist es ratsam, die Platten immer möglichst voll zu besetzen.

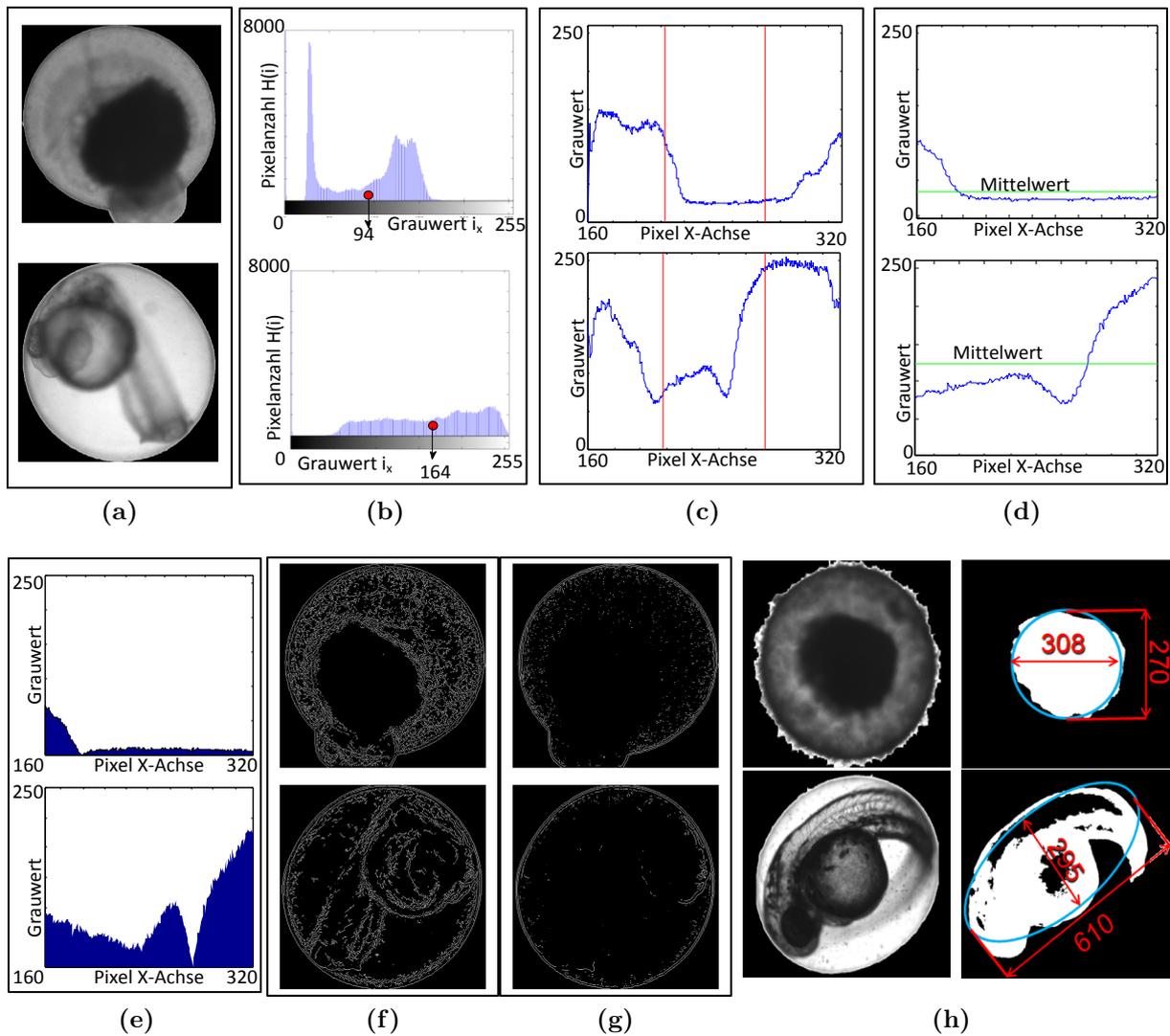


Abbildung 5.2: Beispiel für ein entwickeltes (Abb. 5.2a unten) sowie ein koaguliertes Ei eines Zebrafährblings (Abb. 5.2a oben) und Illustration der extrahierten Merkmale $x_1 - x_7$ (Abb. 5.2b-5.2h; oben für die Klasse *koaguliert*, unten für die Klasse *entwickelt*) [4].

Zur Messung wurde das Hochdurchsatz-Mikroskop Olympus-Scan^R von Olympus Biosystems, ausgestattet mit einem Hamilton Micro-Lab-SWAP-Plattenwechsler, einem 2.5x Objektiv und einer Olympus Biosystems DB-1 (1344x1024 Pixel) CCD-Kamera verwendet. Die Mikroskop-Parameter (Fokus, Belichtungszeit usw.) wurden zu Beginn der Bildakquise einmalig per Hand bestimmt und mittels der mikroskopeigenen Software im XML-Format gespeichert. Es lassen sich jedoch nicht alle Parameter auf automatische Weise speichern. Wichtig ist, dass auch nicht automatisch protokollierte Parameter wie etwa

ein Vorhang oder Störungen wie z.B. das Anstoßen an den Mikroskoptisch die Vergleichbarkeit von unterschiedlichen Durchläufen beeinflussen können und daher möglichst zu vermeiden sind.

Nach der Akquise eines Bildstroms wird überprüft, ob die Anforderungen der Messung erfüllt sind: In Abschnitt 5.2.1 wurde gezeigt, dass sich die akquirierten Daten für die Hochdurchsatz-Untersuchung eignen, unter der Voraussetzung, dass sich signifikante Merkmale extrahieren lassen, welche von der Lage der Zebrabärblinge unabhängig sind (Eindeutige Präsenz der biologischen Information). Da keine fluoreszierenden Marker eingesetzt werden und die Lichtmikroskopie keine bekannten toxischen Auswirkungen auf die Fische hat [135], sind keine Effekte von der Bildakquise auf die Zebrabärblinge zu erwarten (Rückwirkungsfreiheit der Messung). Die Automatisierung der Mikroskopie mittels der Robotertechnik des Olympus-Scan^R soll eine hinreichend schnelle Messung sicherstellen. Die benötigte Zeit zur Messung eines Versuchs ist wie im Konzept beschrieben (vgl. Abschnitt 2.5 Gleichung (2.6)) die Summe aus manuellen Präparationsschritten und der Dauer der Bild-Akquise. Es hat sich als zweckmäßig erwiesen, die Präparationszeit für eine komplette Mikrotiterplatte zu bestimmen, denn typischerweise wird eine bestimmte Anzahl solcher Platten pro Tag verarbeitet und der Tagesablauf bei der Versuchsdurchführung hiernach ausgerichtet. Da bei der durchgeführten Hochdurchsatz-Untersuchung keine manuelle Ausrichtung der Larven erfolgt, sind, bezüglich der Präparationsschritte, lediglich die Dauer für die manuelle Vereinzelnung ($t_{p_1} = \text{ca. } 6\text{s}$ pro Nöpfchen) und für das manuelle Hinzufügen der Chemikalien ($t_{p_2} = \text{ca. } 3\text{s}$ pro Nöpfchen) zu berücksichtigen. Die Dauer der Bildakquise ist abhängig von der Anzahl zu akquirierender Nöpfchen pro Mikrotiterplatte (n_w), der Anzahl von Aufnahmen pro Nöpfchen hintereinander (n_z) und der Anzahl von Wiederholungen¹ der gesamten Platte (n_l). Wird nun noch die Zeit ermittelt, welche das Olympus-Scan^R-Mikroskop zur Aufnahme eines Bildes (t_{m_1}), zum Bewegen zum nächsten Nöpfchen (t_{m_3}) sowie zum Bewegen an den Anfang der Mikrotiterplatte für eine Wiederholungsmessung benötigt (t_{m_2}), so berechnet sich die Akquise-Dauer mittels der oben aufgelisteten Parameter näherungsweise² mittels:

$$t_{aq} = n_w \cdot n_z \cdot n_l \cdot t_{m_1} + (n_w - 1) \cdot n_l \cdot t_{m_2} + (n_l - 1) \cdot t_{m_3} \quad (5.1)$$

In Tabelle 5.3 sind die Parameter und die errechnete Dauer der Akquise für eine Mikrotiterplatte aufgeführt. Damit kann nach der bereits in Abschnitt 2.5 eingeführten Formel (2.6) die Zeit t_{RV} , die für die Durchführung eines Versuchs inkl. Akquise benötigt wird, bestimmt werden. Sie beläuft sich hier auf ca. 11s und somit für eine Mikrotiterplatte auf weniger als 20min. Da das Zeitfenster eines vergleichbaren Alters der Larven bei ca. 4h liegt, können pro Tag und Mikroskop ca. 12 Platten oder 1152 Larven akquiriert werden. Die ermittelten maximal erreichbaren Werte sind für die geplante Hochdurchsatz-Untersuchung ausreichend (Schnelligkeit der Messung), da für sieben Platten theoretisch ca. 2,5 bis 3 Stunden benötigt werden.

¹Mit „Wiederholung der gesamten Platte“ ist hier die Wiederholung der Messung zu einem späteren Zeitpunkt zur Bestimmung von Bewegungen oder Entwicklungen gemeint und nicht zu verwechseln mit den Replika, welche Wiederholungen des gesamten Versuchs (inkl. Präparation und Akquise) sind (vgl. Abschnitt 1.2.1.1).

²Abweichungen ergeben sich z.B. durch Schwankungen in automatisch bestimmten Mikroskop-Parametern oder schwankende Belastung des PCs, z.B. durch Datentransfer.

Parameter	Wert	Bemerkung
n_l	1	Wiederholungen innerhalb der Platte (nicht Replika!)
n_w	96	Standard-Mikrotiterplatte mit 96 Nöpfchen
n_z	3	Anzahl der Fokusebenen
t_{m_1}	0.2	Zeit zur Aufnahme eines Bildes [s]
t_{m_2}	1.2	Zeit für Bewegung an den Anfang der Mikrotiterplatte [s]
t_{m_3}	4	Zeit für Bewegung von Nöpfchen zu Nöpfchen [s]
t_{p_1}	6.2	Zeit zum Vereinzeln einer Larve [s]
t_{p_2}	3.1	Zeit zum Exponieren einer Larve [s]
t_{aq}	1.8	Mittels Formel (5.1) ermittelte Gesamtzeit [s] für eine Larve
$t_{aq_{96}}$	172	Mittels Formel (5.1) ermittelte Gesamtzeit [s] für eine 96er-Platte
t_{RV}	11	Gesamter Zeitaufwand pro Larve [s]
$t_{RV_{96}}$	18	Gesamter Zeitaufwand pro 96er-Platte [min]

Tabelle 5.3: Ermittlung der notwendigen Zeit der Versuchsdurchführung für eine Mikrotiterplatte

5.2.2.2 Versuchsplanung: Analyse und Interpretation

Für die Bildakquise nach o.g. Spezifikationen können nun die Parameter der Hochdurchsatz-Untersuchung gemäß Abschnitt 2.5 bestimmt werden. Der biologische Effekt, die Koagulation, ist der Planfaktor y und soll automatisch, anhand von aus dem Bildstrom **BS** extrahierten Merkmalen \mathbf{x} , geschätzt werden. Der Planfaktor gehört einer der Klassen $\{ja, nein, unbekannt\}$ an. Für einen Teil des Bildstroms wurden die Mikroskopaufnahmen der Eier manuell den Klassen $\{ja, nein\}$ zugeordnet (gelabelt). Alle anderen Aufnahmen gehören der Klasse $\{unbekannt\}$ an, d.h. es ist nicht bekannt, ob die Eier der Zebrabärblinge koaguliert sind oder nicht. Die gelabelten Daten bilden die Lerndaten und werden zur Ermittlung der besten Merkmalskombination sowie zur Berechnung des Klassifikators herangezogen. Mit den Ergebnissen wird die Klassenzugehörigkeit der unbekannt Daten \hat{y} geschätzt und einer der Klassen $\{ja, nein\}$ zugeordnet. Die Konzentration des Toxins ist der Parameter p_y des Planfaktors und soll systematisch variiert werden. Es ist ebenso möglich, den Datensatz nach den bekannten Störfaktoren z zu untersuchen. Diese sind das Toxin, die Position der untersuchten Larve innerhalb einer Konzentration auf der Mikrotiterplatte (hier immer in einer Reihe der Platte, d.h. Position 1 bis 12), das verwendete Mikroskop, die Wiederholung (Replika), das Aufnahmedatum und der Aufnahmezeitpunkt. Das Toxin ist deshalb als bekannter Störfaktor zu betrachten, da zur Bestimmung des LC_{50} -Wertes lediglich die Konzentration des Toxins geändert wird, nicht jedoch die *Art* des Toxins. Eine Klassenzuweisung bezüglich der Störfaktoren ist hier jedoch für die Bestimmung des LC_{50} -Wertes nicht weiter von Interesse. Im Gegenteil soll es bei gleichwertigen Bedingungen der Hochdurchsatz-Untersuchung für einen optimalen Versuch unmöglich sein, eine Unterscheidung anhand der Störfaktoren zu treffen (vgl. Abschnitt 2.2). Als Parameter der bekannten Störfaktoren p_z bleibt lediglich der Typ der verwendeten Mikrotiterplatte, da die Konzentration bereits Parameter des Planfaktors ist.

Der Bildstrom wird in Farbe (Modalität $m_{BS} = 3$) bei einer Auflösung von 1334x1024 auf drei Fokusebenen ($u_{BS} = 1334 \times 1024 \times 3$) zu lediglich einem Zeitpunkt (ein einziger Frame $t_{BS} = 1$) aufgezeichnet. Tabelle 5.4 fasst die Parameter des Versuchs zusammen.

y/\hat{y}	p_y	z/\hat{z}	p_z	p_{BS}		
				m_{BS}	u_{BS}	t_{BS}
Koagulation	Konzentration	Toxin Position Mikroskop Wiederholung Aufnahmedatum Aufnahmezeit	96er-Mikrotiter- platte	3	1334x1024x3	1

Tabelle 5.4: Parameter der Hochdurchsatz-Untersuchung für einen Endpunkt nach Abschnitt 2.5; Mit m_{BS} Anzahl Kanäle, u_{BS} Anzahl Pixel und t_{BS} Anzahl Frames.

Für die Auswertung, den dritten Punkt im Ablaufdiagramm (Abb. 2.11), werden geeignete Module identifiziert und zu einer Auswertungskette zusammenschaltet. Abb. 5.3 hebt die ausgewählten Module im Modulkatalog hervor. Die Beschreibung der Module findet sich in Kapitel 3, daher wird im Folgenden nur auf die Auswahl der Module und die damit gebildete Auswertungskette sowie deren Spezifizierung auf den FET eingegangen.

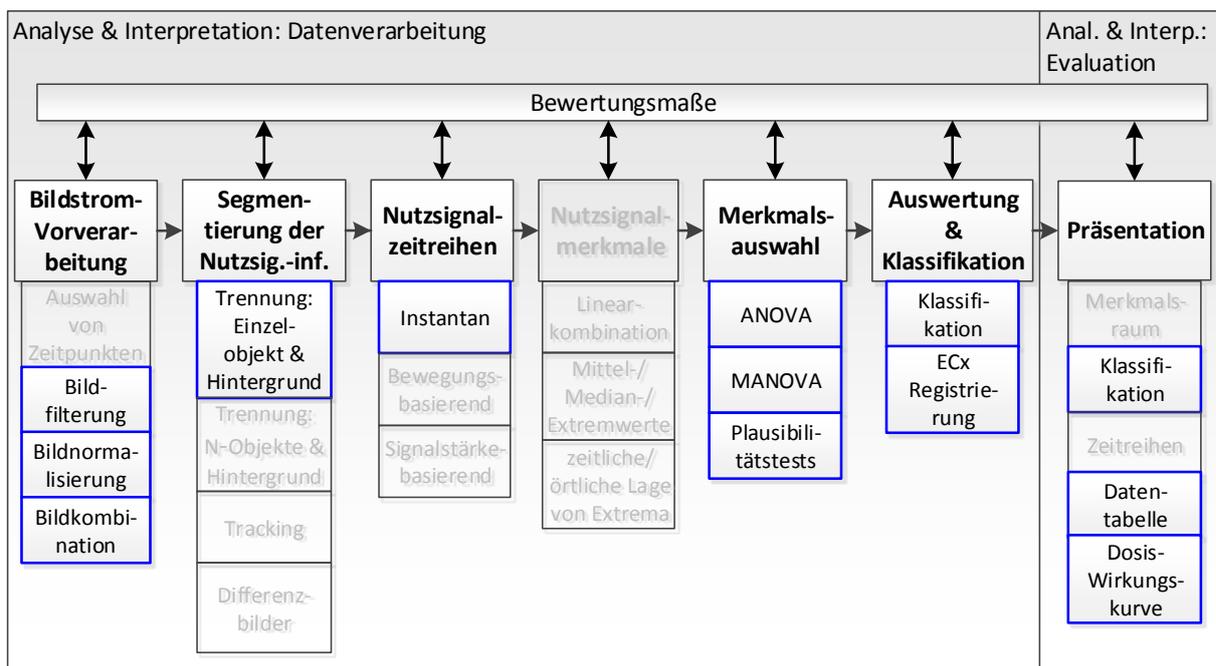


Abbildung 5.3: Module, die zur Auswertung und Bestimmung der EC_{50} Konzentration ausgewählt wurden

Bildstrom-Vorverarbeitung: Für die Bildstrom-Vorverarbeitung wurden die folgenden drei Module eingesetzt: In einer Bildkombination wird aus den Aufnahmen über drei Fo-

kusebenen ein Bild mit erweitertem Fokus erstellt. Hierdurch sind die Details im Fischei deutlicher zu unterscheiden. Die Aufnahmen des Mikroskops Olympus-Scan^R nutzen lediglich einen kleinen Teil des zur Verfügung stehenden Wertebereichs. Daher wird mittels Bildnormalisierung eine Spreizung des Histogramms hinzugefügt, welche sich der Bildfilterung bedient. Als Ergebnis steht ein verbesserter Bildstrom **BS*** zur Verfügung. Da der Bildstrom für das Nutzsignal *Koagulation* lediglich zu einem Zeitpunkt und an einem einzigen Mikroskop aufgezeichnet wurde, können eine aufwendigere Bildnormalisierung und die Auswahl von Zeitpunkten entfallen.

Segmentierung der Nutzsignalinformation: Das Ei des Zebrafärbchens wird bei der Präparation vereinzelt, so dass lediglich ein Ei in jedem Nöpfchen der Mikrotiterplatte und somit in jedem Bildstrom vorhanden ist. Hierfür wird das Modul zum Trennen eines Einzelobjektes vom Hintergrund angewandt. Das Ergebnis ist eine Bildmaske, welche das Ei vom Hintergrund trennt und mit dem Originalbild multipliziert wird. Fehlerhafte Aufnahmen werden wie im Modul beschrieben anhand der Größe und „Rundheit“ des Eies ausgeschlossen. Abb. 5.4 veranschaulicht die Anwendung der Segmentierung.

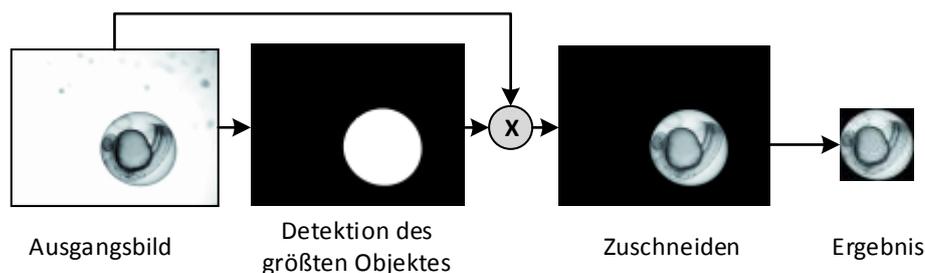


Abbildung 5.4: Veranschaulichung der Anwendung des Moduls *Trennung von Einzelobjekt & Hintergrund*.

Die Segmentierung lässt sich robust auf den akquirierten Datensatz anwenden. Von 1008 Bildern werden sieben automatisch anhand der genannten Kriterien aussortiert. Die niedrige Fehlerrate erlaubt es, auf zusätzliche Validierungsprüfungen zu verzichten. Die Fehler beruhen zudem ausschließlich auf qualitativen Inhomogenitäten in den Bilddaten und können durch genauere Präparation abgestellt werden. Die Bilddaten nach der Segmentierung werden zur späteren Analyse gespeichert. Der vollständige Prozess (inkl. Einlesen und Speichern) benötigt pro Bild 0.46s und für den gesamten Datensatz 7:42 min auf einem normalen PC³ (Segmentierbarkeit des biologischen Effekts im Bild).

Nutzsignalzeitreihen: Der Bildstrom wurde an lediglich einem Zeitpunkt aufgezeichnet. Somit kann auch nur für den Zeitpunkt der Aufzeichnung ein Merkmal extrahiert werden, was zu Einzelwerten bzw. einer Zeitreihe der Länge $k = 1$ führt. Das Fehlen von Bildsequenzen beschränkt die Menge der extrahierbaren Merkmale auf die instantanen Merkmale (vgl. Abschnitt 3.4.1). Voraussetzung für die Merkmale ist, dass sie unabhän-

³CPU: Intel® Core™ i5 750 @ 2.67GHz, Arbeitsspeicher: 4.00GB; OS: Windows 7 Enterprise 64Bit

gig von der Lage der Larve eine Klassifikation erlauben. Ein Beispiel beider Klassen des zu unterscheidenden Nutzsymbols ist in Abb. 5.2a gegeben.

Potenziell eignen sich alle instantanen Merkmale aus Abschnitt 3.4.1 zur quantitativen Beschreibung des Eiinhaltes bezüglich Koagulation. Es wurden daher auch alle Merkmale x_1 bis x_8 extrahiert. Abb. 5.2b-5.2h illustriert Beispiele für einige der Merkmale. Die obere Reihe der Abbildungen ist jeweils für ein koaguliertes Ei, die untere für eine normal entwickelte Larve. In Abb. 5.2b liegt der Wert von x_1 (Schwerpunkt des Histogramms) für koagulierte Eier bei deutlich niedrigeren Werten (hier 94 (koaguliert) zu 164 (entwickelt)), ebenso verhält sich das nicht dargestellte Merkmal x_2 , welches sich auf die mittlere Helligkeit des gesamten Bildes bezieht. In Abb. 5.2c ist die mittlere Bildzeile des Bildes dargestellt. Von der Zeile wird der Mittelwert der mittleren 30% als Merkmal x_3 extrahiert und ist in Abb. 5.2d durch die grüne Linie gekennzeichnet. Der Mittelwert für ein koaguliertes Ei liegt deutlich niedriger. Abb. 5.2e illustriert das Merkmal x_4 , welches die Schwankung der Grauwerte um den Mittelwert von x_3 quantifiziert. Das Merkmal ist die blau eingefärbte Fläche unter der Kurve. Abb. 5.2f-5.2g zeigt ein Beispiel für die Kantenbilder, deren Anzahl durch die Merkmale x_5 und x_6 extrahiert wird. Abb. 5.2h schließlich illustriert die längste und kürzeste Halbachse der umschließenden Ellipse, deren Differenz das Merkmal x_7 darstellt. Das Merkmal x_8 wurde nicht dargestellt, da es lediglich die Bildgröße enthält und als Validitätsprüfung genutzt wird (vgl. hierzu Abschnitt 3.3.1 und Abschnitt 3.4.1).

Dimensionsreduktion: Die Dimensionsreduktion der Zeitreihen zu Einzelmerkmalen entfällt, da lediglich Einzelmerkmale extrahiert wurden.

Merkmalsauswahl: Von einem Experten wurde ein Trainingsdatensatz aus einem Teil des aufgenommenen Datensatzes erstellt. Dabei wurden 35 Larven der Klasse *koaguliert* und 120 Larven der Klasse *entwickelt* zugeordnet. Anhand der Trainingsdaten wurde das trennstärkste Merkmal mittels ANOVA ermittelt und mittels MANOVA die beste 2er-Merkmalsskombination bestimmt. Es zeigt sich, dass das Merkmal x_7 mit einer Güte von 0.872 in der ANOVA das beste Merkmal zur Klassifikation ist. Die beste Zweierkombination ergibt sich unter Hinzuziehung des Merkmals x_3 mit einer erreichten Güte von 0.898 (vgl. Tabelle 5.5). Somit ist eine Merkmalskombination extrahiert, die die Nutzsymbolsinformation beschreibt und eine gute Trennschärfe aufweist (Quantifizierbarkeit des biologischen Effekts im Bild).

Auswertung & Klassifikation: Anhand der identifizierten trennstarken Merkmalskombination und der Trainingsdaten wurde ein Bayes-Klassifikator mit Berücksichtigung der klassenspezifischen Kovarianzmatrizen angelernet [110]. Die Klassifikation zeigt auf den Testdaten in einer 5-fach-Crossvalidierung einen Klassifikationsfehler von $1.2 \pm 0.2\%$ (niedrige Fehlerrate bei der Detektion des biologischen Effekts). In Abb. 5.5 ist ein Scatterplot der Trainingsdaten über die Merkmale x_7 und x_3 sowie die Diskriminanzfunktion des Klassifikators abgebildet. Die der Klasse *koaguliert* zugeordneten Larven zeigen für x_3 und x_7 deutlich geringere Werte als die der Klasse *entwickelt* zugeordneten. Die Diskriminanzfunktion des Bayesklassifikators ist ebenfalls abgebildet. Der Klassifikator kann nun bei

Merkmal	Güte-MANOVA	Güte-ANOVA
x_7	—	0.872
x_3	0.898	0.693
x_4	0.886	0.556
x_6	0.880	0.430
x_5	0.879	0.389
x_1	0.875	0.580
x_8	0.872	0.056
x_2	0.872	0.249

Tabelle 5.5: Merkmalsrelevanzen – Univariate und Multivariate Analyse (bei Auswahl von 2 Merkmalen)

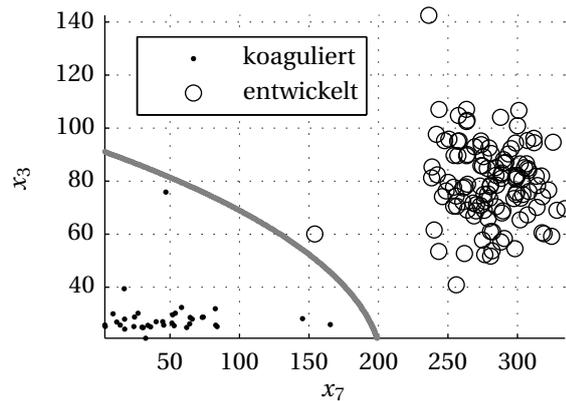


Abbildung 5.5: Auf Basis der Lerndaten berechnete Diskriminanzfunktion

Durchführung der Hochdurchsatz-Untersuchung auf Testdaten angewandt werden. Die Ergebnisse der Klassifikation ermöglichen die Zuordnung einer relativen Sterberate zu jeder Konzentration eines jeden Toxins.

Durch Anwendung des Moduls “EC_x-Registrierung“ zur Ermittlung der Dosis-Wirkungskurve aus den Ergebnissen der Sterberaten der Konzentrationsreihen eines jeden Toxins lässt sich bei geeigneten Werten der EC₅₀-Wert ermitteln. Im vorliegenden Datensatz ließen die erzielten Effekte der Konzentrationen die Ermittlung des Wertes bei drei Toxinen zu, wobei keine Aufnahme so schlecht war, dass sie durch die Validierung ausgeschlossen werden musste. Es konnten somit alle aufgezeichneten Larven klassifiziert werden. Wenn es nicht möglich war, ein EC₅₀-Wert zu ermitteln war die gewählte Verdünnungsreihe nicht passend (vgl. Abschnitt 3.6.2). Die Anzahl der untersuchten Larven, Konzentrationen sowie die getroffenen Klassenzuordnungen sind in Tabelle 5.6 aufgelistet. Tabelle 5.7 zeigt die Aufschlüsselung auf die einzelnen Konzentrationen. Für das Toxin *AS2O3* wurden beispielsweise 120 Larven untersucht (Zusätzlich noch 24 unbehandelte Larven als Kontrolle), wovon sich jeweils 24 auf jede der 5 Konzentrationen verteilen (12 je Replika). Bei Konzentration **K1** von *AS2O3* (0.80 mM vgl. Tabelle 5.2) wurden bei der ersten Wiederholung 0 von 12 Larven als koaguliert eingestuft (d.h. alle 12 Larven sind als *nicht koaguliert* eingestuft), beim zweiten Durchgang wurden 3 von 12 Larven als koaguliert geschätzt. Bei der Konzentration **K2** wurde keine der Larven als koaguliert eingestuft usw. Insgesamt wurden für das Toxin 49 Larven als koaguliert und 71 als entwickelt also nicht koaguliert eingeschätzt. Der ermittelte EC₅₀ liegt bei einer Konzentration von 1.1 mM.

Präsentation: Die Ergebnisse werden zur biologischen Auswertung in Form einer 22 Seiten starken Reportdatei im PDF-Format erstellt. Das Dokument enthält für jedes Toxin eine Reihe von Darstellungen der Ergebnisse, die jeweils auf einer DIN-A4 Seite das Ergebnis bereitstellen. Die Darstellungen sind: eine Überlagerung der Bilddaten mit den Klassifikationsergebnissen (vgl. Abb. 5.6), eine Auswertung der Klassifikationsergebnisse

Toxine	Larven	Konzentrationen	Schätzung koaguliert	Schätzung entwickelt	EC ₅₀
As ₂ O ₃	120	K1-K5	49	71	1.1
CdCl	240	K1-K10	31	209	-
Ethanol	120	K1-K5	27	93	2.5
Methanol	120	K1-K5	74	46	2.5
PbCl	240	K1-K8	47	193	-
Kontrolle	168	-	-	-	-

Tabelle 5.6: Anzahl von untersuchten Larven und Konzentrationen sowie Klassifikationsergebnisse und errechnete EC₅₀-Werte über den Toxinen.

Toxin	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	Kontrolle
AS ₂ O ₃ -1	0	0	3	9	11						1
AS ₂ O ₃ -2	3	0	4	9	10						0
CdCl-1	1	1	5	2	0	0	2	4	2	1	0
CdCl-2	1	0	0	2	1	0	0	3	4	2	1
Ethanol-1	1	0	0	6	7						0
Ethanol-2	0	0	0	4	9						0
Methanol-1	0	6	10	11	12						0
Methanol-2	0	2	10	12	11						0
PbCl-1	3	3	4	0	6	1	5	4			0
PbCl-2	5	6	1	1	3	0	2	4			0

Tabelle 5.7: Als koaguliert eingestufte Larven von jeweils 12 untersuchten Larven je Konzentration über den Toxinen der Mikrotiterplatten aus Tabelle 5.6. Jede Mikrotiterplatte wurde jeweils einmal zur Kontrolle wiederholt. (Bei CdCl-1(Kontrolle), CdCl-2(Kontrolle), PbCl-1(K1), PbCl-2(K1), PbCl-1(K2), PbCl-2(K2), PbCl-1(Kontrolle) und PbCl-2(Kontrolle) sind anstelle von 12 jeweils 24 Larven getestet worden.)

in Tabellenform (vgl. Tabelle 5.6 und Tabelle 5.7), ein klassenspezifisches 2-D-Histogramm über die Konzentration (vgl. Abb. 5.7), die Regression auf die Dosis-Wirkungs-Kurve mit Überlagerung der ermittelten Konzentration EC_{50} (vgl. Abb. 5.8, dargestellt ist das Toxin Methanol), ein Scatterplot über Klassifikationsergebnisse im Merkmalsraum über alle Daten sowie für jede Konzentration. Zusätzlich werden Metainformationen wie etwa der Name des Toxins und der Speicherort der Daten abgelegt. Abb. 5.6 zeigt die automatisch zugeschnittenen und tabellarisch in der Ordnung der Mikrotiterplatten angeordneten Bilder der Zebrafischlarven, wie sie für jedes Toxin in der Reportdatei abgelegt werden. Die Darstellung ergibt somit „einen virtuellen Blick“ in die gesamte Mikrotiterplatte. Die als koaguliert klassifizierte Larven sind rot-gestrichelt umrandet. Abb. 5.7 zeigt das klassenspezifische 2-D-Histogramm. Die Höhe des roten Balkens gibt den Prozentwert an Larven an, die der jeweiligen Klasse koaguliert oder entwickelt (hier englisch *dead* und *alive*) zugeordnet wurden. Die beiden roten Balken ergeben in der Summe somit immer 100%. Mit Hilfe der in der Reportdatei gesammelten Informationen steht dem Anwender das Endergebnis übersichtlich zur Verfügung und es lassen sich anhand der Überlagerung der Ergebnisse aus Segmentierung und Klassifikation alle Verarbeitungsschritte innerhalb weniger Minuten auf grundsätzliche Konsistenz prüfen (Präsentierbarkeit der Auswertung).

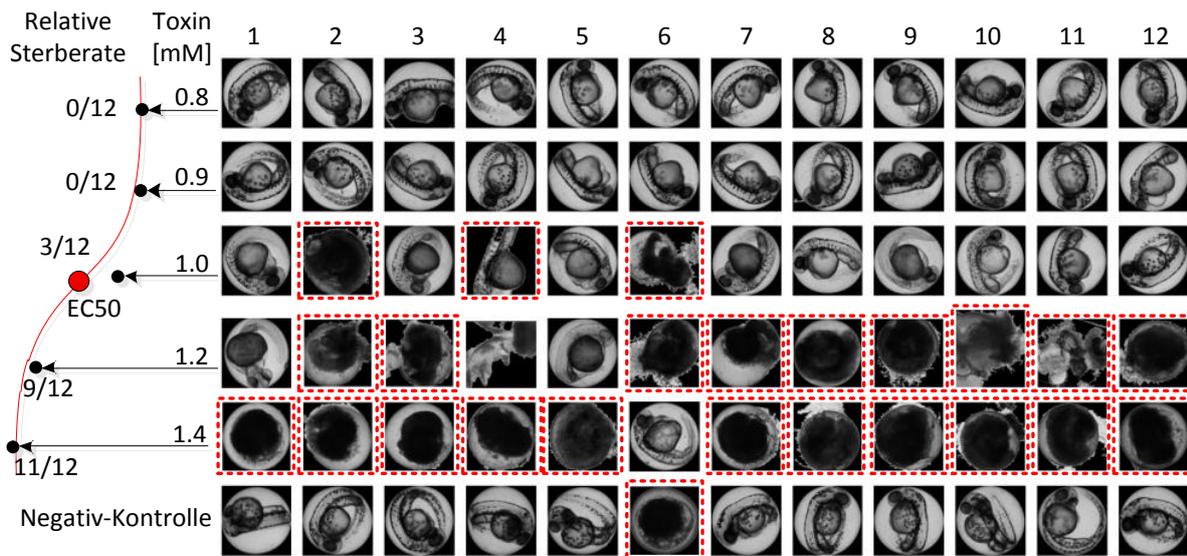


Abbildung 5.6: Visualisierung der Klassifikationsergebnisse der Hochdurchsatz-Untersuchung, hier für das Toxin As_2O_3 ; gestrichelt markierte Larven sind der Klasse *Koaguliert* zugeordnet [5]

5.2.2.3 Durchführung und Resultat

Nach dem Erstellen des Klassifikators lässt sich die erarbeitete Auswertungskette durch die kurze benötigte Zeit zur Auswertung auch auf große Datensätze skalieren. Der gesamte Auswerteprozess benötigt (inkl. Einlesen und Speichern von Bildern, Projektdateien, Ergebnissen) pro Platte 73.5s, d.h. für die gesamte Untersuchung ca. 8:30 min (Schnelligkeit

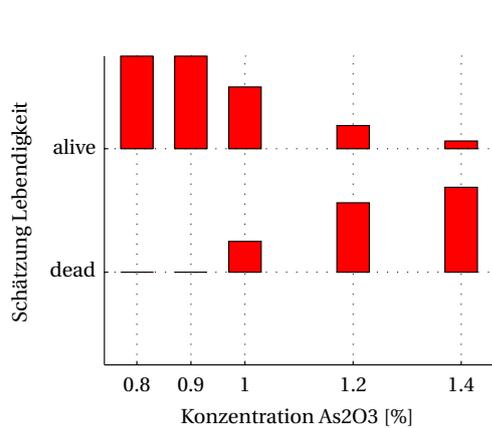


Abbildung 5.7: 2D-Histogramm koagulierter und entwickelter Larven

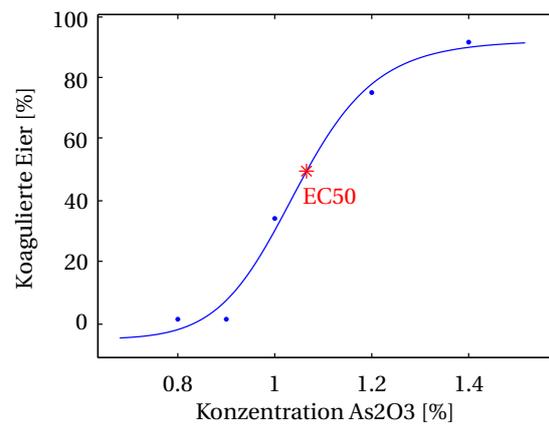


Abbildung 5.8: Automatische Regression und EC₅₀-Berechnung

der Auswertung). Ein erneutes Durchlaufen des Flussdiagramms ist nur notwendig, wenn eine Änderung der Störfaktoren (z.B. Anzahl der Mikroskope, Anzahl der Fischeier pro Nöpfchen) vorgenommen werden muss. Alle Bilddaten, Gait-CAD-Projekte, Zwischenergebnisse und Reportdateien wurden auf einem Archiv des KIT namens LSDF (engl. Large Scale Data Facility) [166], welches für die Speicherung großer Datenmengen über eine lange Zeit erstellt wurde, gespeichert (Wissenschaftliche Archivierung). Das Durchlaufen des vorgeschlagenen Flussdiagramms aus Abschnitt 2.6 führt erwartungsgemäß zu einer aussagekräftigen Hochdurchsatz-Untersuchung, welche die Beantwortung der Fragestellung unter Einhaltung aller Anforderungen erlaubt. Es ist nun erstmals möglich, den FET bildbasierend für den Endpunkt *Koagulation* mit hoher Erfolgsrate automatisch auszuwerten [6, 7].

Eine manuelle Überprüfung der Ergebnisse zeigt jedoch, dass die Klassifikation durch die Beschränkung auf lediglich einen Endpunkt, aus biologischer Sicht, d.h. für die Beurteilung der Lebendigkeit der Zebraäbrlinge, zu einer hohen Falsch-Negativ-Rate kommt. Viele Larven weisen zwar keine Koagulation auf, können jedoch nicht als lebendig eingestuft werden, was sich dadurch zeigt, dass sie weder Herzschlag besitzen noch spontane Bewegungen ausführen. Das Ergebnis legt die Erweiterung der bildbasierten Auswertung bezüglich der beiden Endpunkte *Herzschlag* und *Spontanbewegung* nahe, welche im folgenden Abschnitt 5.2.3 vorgestellt wird.

5.2.3 Fisch Embryo Test (FET) – Bewegungen

Die automatisierte Auswertung des FET aus Abschnitt 5.2.2 soll um die Endpunkte *Herzschlag* und *Spontanbewegung* erweitert werden, da so die Sensitivität der Untersuchung verbessert und eine geringere Falsch-Negativ-Rate ermöglicht wird. Das Ziel der vorliegenden Versuche besteht darin, einen Standardablauf für das systematische Testen mit den genannten drei Endpunkten zu entwickeln. Dazu wurden drei Toxine mit bekannten

Wirkungen gewählt, um typische Effekte sicher abzubilden. Methanol und Ethanol führen zu koagulierten Eiern, während das Arzneimittel Valproat in hohen Konzentrationen Muskeln sowie Nervensystem und folglich Herzschlag sowie Spontanbewegung beeinträchtigt. Für die erfolgreiche Erarbeitung einer Hochdurchsatz-Untersuchung wird wiederum das in Abb. 2.11 vorgestellte Flussdiagramm zur anforderungskonformen Auswertung durchlaufen. Es kann hierbei auf einige Ergebnisse aus Abschnitt 5.2.2 zurückgegriffen werden. Alle Larven wurden von Experten auf die drei Endpunkte untersucht. Pro Toxin wurde eine Mikrotiterplatte mit 96 Nöpfchen verwendet, von denen 72 mit Fischlarven belegt wurden. Fünf Reihen mit je zwölf Fischlarven wurde je einer bestimmten Toxinkonzentration ausgesetzt, während eine Reihe mit zwölf unbehandelten Fischlarven als Negativkontrolle ohne Toxinbehandlung diente. Damit wurde eine Gesamtzahl von 288 Larven untersucht. Der aufgenommene Datensatz und die manuelle Klassifikation bezüglich der drei Endpunkte sind in Tabelle 5.8 gegeben.

Toxin	Einheit	Dosis					präsen- te Koagu- lation	präsen- te Spontan- bewegung	präsen- ter Herzschlag
Methanol	[%]	1.67	2.27	2.5	2.75	3.75	9	63	29
Ethanol	[%]	1.33	1.82	2	2.2	3	16	31	8
Valproat I	[mg/l]	186.7	254.5	280	308	420	6	21	10
Valproat II	[mg/l]	186.7	254.5	280	308	420	17	25	18

Tabelle 5.8: Toxine, Verdünnungsreihen und manuell ermittelte Klassenzugehörigkeiten des untersuchten Datensatzes.

5.2.3.1 Versuchsplanung: Biologie und Bildakquise

Der biologische Effekt, d.h. die Endpunkte, sind bereits in der OECD-Norm definiert [121, 122] und die Anforderungen zur Durchführbarkeit sind wie bei der Klassifikation zur Koagulation erfüllt (Legitimität, logistische Machbarkeit, biologische Realisierbarkeit, Reproduzierbarkeit, finanzielle Realisierbarkeit).

Zur Überprüfung der Anforderungen an die Messung werden lediglich die Unterschiede zum vorangegangenen Abschnitt geprüft. Das Mikroskop sowie die Präparation und die Zeitpunkte, an denen die Larven dem jeweiligen Toxin exponiert werden, sind identisch. Von jedem Ei sind, um die Messung und Detektion von Herzschlag und Spontanbewegung zu realisieren, Bildzeitreihen notwendig. Der Herzschlag der Zebrafärbblingslarve hat eine Frequenz von ca. 120 Schlägen pro Minute. Da der Herzschlag eine schnelle Bewegung darstellt, müssen hierfür Bildzeitreihen mit einer hohen Abtastfrequenz ($f > 4$ Hz) aufgenommen werden. Nach Voruntersuchungen wurde festgestellt, dass die maximale Akquise-Geschwindigkeit des Olympus-Scan^R-Mikroskops von zwei Bildern pro Sekunde (engl. fps - frames per second) ausreicht, um die Präsenz des Herzschlags in einer Bildsequenz über zwei Sekunden zu identifizieren. Hierbei auftretende Aliasing-Effekte können vernachlässigt werden, weil nur eine dichotome Klassifikation, nicht aber die Bestimmung

der Herzfrequenz Ziel der Untersuchung ist. Die Spontanbewegung der Fischlarven erfolgt nicht konstant wie der Herzschlag. Zuweilen bewegt sich die Larve auch für mehrere Minuten nicht. Daher sind zu deren Analyse Bildzeitreihen mit niedrigen Frequenzen ($f \ll 1 \text{ Hz}$) notwendig. Eine Analyse der Klassifikationsergebnisse mit Mikroskopaufnahmen, die sich auf lediglich eine Fokusebene beziehen (vgl. Abb. 1.5), zeigt keine signifikante Verschlechterung der Fehlerrate, wenn die ausgewertete Fokusebene weitgehend in der Mitte des Eies der untersuchten Zebrafärbungs-Larve ist. Der Vorteil ist jedoch, dass sich der Datensatz um den Faktor 3 verringert und auf die Bildkombination verzichtet werden kann.

Für die automatische Bildakquise wurde für die vorliegende Arbeit, um in der Lage zu sein, in nur einem Durchgang sowohl Herzschlag als auch Spontanbewegung aufzuzeichnen, ein zyklischer Akquiseablauf in die Steuerungssoftware des Olympus-Scan^R-Mikroskops programmiert. Damit wird ein Bildstrom nach Formel (2.15) aufgezeichnet. Abb. 5.9 illustriert die jeweils 5 schnell hintereinander aufgenommenen Bilder zur Herzschlagdetektion sowie die längeren Pausen zwischen den einzelnen Sequenzen [7]. Der Bildstrom BS vereinfacht sich hier entsprechend, da nur eine Modalität ($I_c = 1$) und auch nur eine Fokusebene ($I_z = 1$) aufgezeichnet wurde. Dazu nimmt das Mikroskop für jede Larve eine zwei Sekunden lange Bildzeitreihe für den Herzschlag mit maximaler Akquise-Geschwindigkeit auf, was fünf Aufnahmen ($n_z = 5$) entspricht. Daraufhin setzt das Mikroskop die Aufnahme unmittelbar mit der nächsten Larve fort, bis alle Nöpfchen der 96er-Mikrotiterplatte ($n_w = 96$) akquiriert wurden. Das Vorgehen wird 30 mal wiederholt ($n_l = 30$). Die große Anzahl von Wiederholungen ist notwendig, da die Larven nach dem Einsetzen in das Mikroskop, möglicherweise aufgrund der Erschütterung, sich für eine bestimmte Zeitdauer wenig bewegen. Eine nur kurze Akquisitionszeit hat somit falsch-negativ (engl. *false-negative*) Klassifikationen zur Folge. In Abb. 5.10 sind die anfänglich verminderten Bewegungen der Larven anhand der mittleren Bewegungsmenge nach Merkmal x_{15} (vgl. Abschnitt 3.4.2) von 96 gleichzeitig ausgewerteten Larven unmittelbar nach dem Einsetzen in das Mikroskop gezeigt. Erst nach 20–30 min erreicht das Merkmal einen etwa gleichbleibenden Wert, d.h. bis zu diesem Zeitpunkt sind deutlich weniger Bewegungen der Larven vorhanden. In Abb. 5.11 ist die Anzahl von Larven derselben Platte, welche mindestens eine Bewegung vollzogen haben über der Akquisitionszeit dargestellt.

Es lässt sich ablesen, dass erst nach etwa einer Stunde alle Larven mindestens einmal eine Spontanbewegung durchgeführt haben. Aufgrund dieser Ergebnisse wurden die 30 Wiederholungen ($n_l = 30$) festgelegt, was in einer Akquisezeit von ca. zwei Stunden resultiert. Somit ist sichergestellt, dass mindestens eine Bewegung von gesunden Larven erfasst wird. Damit kann die Akquise-Dauer mittels Formel (5.1) ermittelt werden. Abb. 5.9 veranschaulicht den Datensatz anhand der Beschreibung des Bildstroms BS (vgl. Formel (2.15)) aus Abschnitt 2.5. Die Zeit für die Akquise eines Versuchs beläuft sich auf $t_{RV} = 76 \text{ s}$. Dabei handelt es sich jedoch nur um einen theoretischen Wert, da die Wartezeiten zwischen den einzelnen Aufnahmen nicht gekürzt werden können, um sicherzustellen, dass eine entwickelte Larve während der Akquise auch mindestens einmal eine Spontanbewegung durchgeführt hat (Eindeutige Präsenz der biologisch relevanten Information). Auch durch die längere Dauer der Akquise sind keine Auswirkungen auf die Larven zu erwarten (Rückwirkungsfreiheit der Messung). Jede Mikrotiterplatte benötigt bei der Akquise

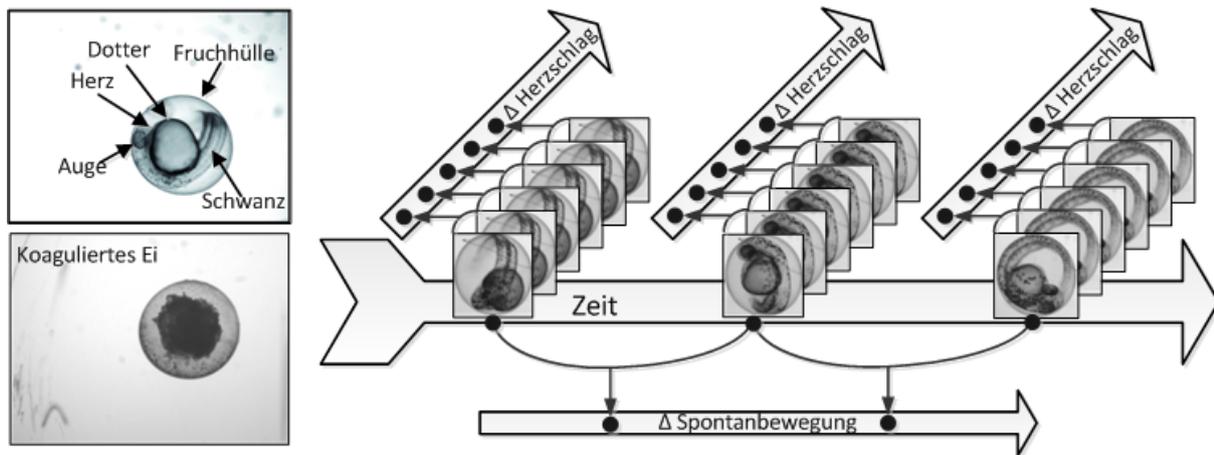


Abbildung 5.9: Links: Typische Zebrafischlarven (oben: entwickelt mit einzelnen Organen, unten: koaguliert), rechts: Bildzeitreihen für den Herzschlag wurden mit 2 Frames pro Sekunde akquiriert; Bildzeitreihen für die Spontanbewegung mit 0.25 Frames pro Minute.

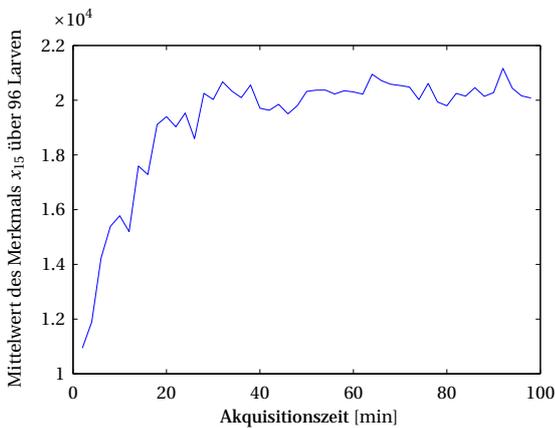


Abbildung 5.10: Mittelwert des Merkmals x_{15} (vgl. Abschnitt 3.4.2) über 96 Larven einer Platte unbehandelter Larven (Kontrollen). Nach etwa 30 min erreichen die Larven eine gleichbleibende Bewegungsmenge.

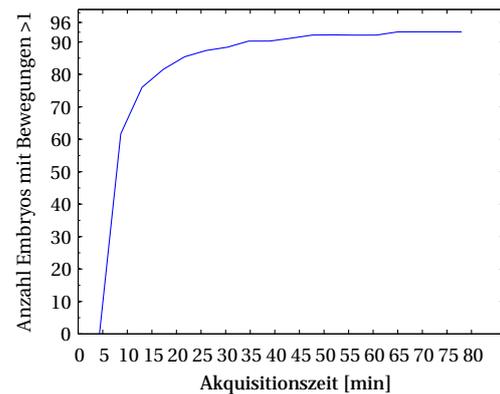


Abbildung 5.11: Anzahl von insgesamt 96 Larven aus der gleichen Platte wie Abb. 5.10, bei welchen mindestens eine Spontanbewegung auf Basis des Merkmals x_{15} (vgl. Abschnitt 3.4.2) identifiziert wurde.

mittels den in Tabelle 5.9 angegebenen Parametern 123 min. Der zu verarbeitende Datensatz wächst auf eine Größe von fast 400 MB Speicherplatz pro Modellorganismus und umfasst für den gesamten Datensatz ca. 1 TB. Der Größenunterschied im Vergleich zum Datensatz zur Ermittlung der Koagulation kommt durch den Einfluss der Bildsequenzen und der erwähnten langen Akquisezeit zustande.

Bei dem Zeitfenster zur Akquise von ca. 4h pro Tag lassen sich somit lediglich 2 Platten oder 192 Larven an einem Mikroskop akquirieren. Bei einer großen Anzahl zu untersuchender Substanzen müssen somit mehrere Mikroskope parallel betrieben werden. Der vorliegende Datensatz wurde auf nur einem Mikroskop aufgezeichnet, da die Erfahrung

Parameter	Wert	Bemerkung
n_l	30	Wiederholungen
n_w	96	Standard Mikrotiterplatte mit 96 Nöpfchen
n_z	5	Anzahl der Frames
t_{m_1}	0.2	Zeit zur Aufnahme eines Bildes [s]
t_{m_2}	1.2	Zeit für Bewegung an den Anfang der Mikrotiterplatte [s]
t_{m_3}	4	Zeit für Bewegung von Nöpfchen zu Nöpfchen [s]
t_{p_1}	6.2	Zeit zum Vereinzeln einer Larve [s] (manuell)
t_{p_2}	3.1	Zeit zum Exponieren einer Larve [s] (manuell)
t_{aq}	67	Mittels Gleichung (5.1) ermittelte Gesamtzeit [s] für eine Larve
$t_{aq_{96}}$	107	Mittels Gleichung (5.1) ermittelte Gesamtzeit [min] für eine 96er-Platte
t_{RV}	76	Gesamter Zeitaufwand pro Larve [s]
$t_{RV_{96}}$	123	Gesamter Zeitaufwand pro 96er-Platte [min]

Tabelle 5.9: Ermittlung der notwendigen Zeit der Versuchsdurchführung für eine Mikrotiterplatte

zeigt, dass die Verwendung mehrerer Mikroskope zu inkonsistenten Datensätzen führen kann. Um die Vergleichbarkeit inhomogener Datensätze sicherzustellen, muss der Einfluss der Mikroskope mittels spezieller Normalisierung beseitigt werden. Für solche Fälle wurde das Modul in Abschnitt 3.2.2 entwickelt, welches auch dazu dienen kann, Datensätze zwischen einzelnen Laboren vergleichbar zu machen. Die Leistungsfähigkeit wurde in [139] gezeigt. Auch wenn das Modul in den Datensätzen der vorliegenden Arbeit nicht zum Einsatz kommt, ist es für die Skalierbarkeit der Untersuchung wichtig und daher in den Modulkatalog aufgenommen. Die vier Mikrotiterplatten des betrachteten Datensatzes lassen sich an einem Mikroskop innerhalb von minimal zwei Tagen akquirieren (Schnelligkeit der Messung).

5.2.3.2 Versuchsplanung: Analyse und Interpretation

Die Parameter der Hochdurchsatz-Untersuchung sind bis auf die Planfaktoren \mathbf{Y} , die Anzahl von akquirierten Frames ($t_{\mathbf{BS}}$) und der Akquise-Frequenz ($f_{\mathbf{BS}}$) identisch mit den Parametern der Hochdurchsatz-Untersuchung für Koagulation. Die beiden neuen Planfaktoren *Herzschlag* und *Spontanbewegung* sollen automatisch, anhand aus dem Bildstrom \mathbf{BS} extrahierter Merkmale \mathbf{x} , geschätzt werden. Zum Nachweis der Leistungsfähigkeit der automatischen Evaluation sind alle Aufnahmen der Larven von Experten einer der Klassen $\{ja, nein, fehlerhaft\}$ zugeordnet worden. Zwei Mikrotiterplatten (Methanol und Valproat I) wurden zum Anlernen der Algorithmen und die verbleibenden (Ethanol und Valproat II) zum Validieren der Algorithmen verwendet. Die Parameter der Hochdurchsatz-Untersuchung sind in Tabelle 5.10 zusammengefasst. Der Unterschied zu Tabelle 5.4 besteht in den Parametern zur Aufnahme der beschriebenen Bildsequenz mit entsprechen-

den Abtastraten und Wiederholungen.

y/\hat{y}	p_y	z/\hat{z}	p_z	p_{BS}			
				m_{BS}	u_{BS}	t_{BS}	f_{BS}
Koagulation Herzschlag Spontanbew.	Konzentration	Toxin Position Mikroskop Wiederholung Aufnahmedatum Aufnahmezeit	96er-Mikro- titerplatte	3	1334x1024x3	5	0.0042

Tabelle 5.10: Parameter der Hochdurchsatz-Untersuchung für drei Endpunkte nach Abschnitt 2.5; mit m_{BS} Anzahl Kanäle, u_{BS} Anzahl Pixel, t_{BS} Anzahl Frames und f_{BS} Akquise-Frequenz [Hz]

Die für die Auswertung der Endpunkte *Herzschlag* und *Spontanbewegung* zusammengesetzten Module sind in Abb. 5.12 hervorgehoben. Die Auswertung des Endpunktes *Koagulation* erfolgt analog zum vorangegangenen Abschnitt, daher wird im Folgenden hierauf nicht weiter eingegangen.

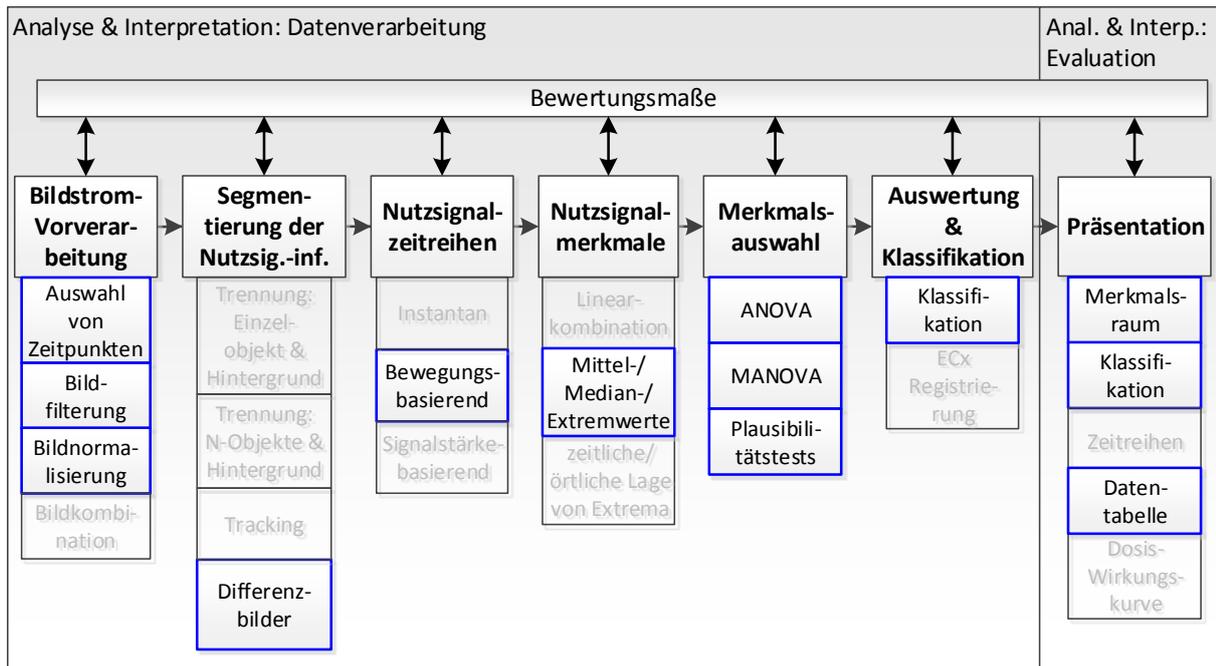


Abbildung 5.12: Ausgewählte Module zur Auswertung von Herzschlag und Spontanbewegung

Bildstrom-Vorverarbeitung: Die Bildstrom-Vorverarbeitung verläuft identisch wie bei der Hochdurchsatz-Untersuchung zur Koagulation, jedoch kann, da nur eine Fokusebene akquiriert wurde, auf die Bildkombination für den Extended Focus verzichtet werden. Für die Auswertung der Spontanbewegung ist es ausreichend, lediglich ein Bild jeder

Kurzzeitreihe zu verwenden. Daher wurde das Modul zur zeitlichen Unterscheidung hinzugefügt, was die notwendige Rechenzeit um den Faktor fünf verringert.

Segmentierung der Nutzsinalinformation: Für beide Endpunkte ist lediglich die Bewegung im Bild entscheidend. Da nur eine Larve pro Bild akquiriert wird und davon ausgegangen werden kann, dass sich Hintergrundobjekte und Fremdkörper nicht bewegen, kann auf eine Trennung von Einzelobjekt und Hintergrund sowie ein Zuschneiden verzichtet werden. Zudem würde beides eine Registrierung der zugeschnittenen Bilder erforderlich machen, was rechenintensiv ist. Weiter kann gezeigt werden, dass sich sowohl die Spontanbewegung als auch der Herzschlag durch die Veränderung der Pixel über die Zeit widerspiegelt, was in Abb. 5.13 dargestellt ist. Die Abbildung illustriert den Einfluss der Bewegungen des Herzschlags und des Blutflusses. Bildbereiche mit einem hohen Anteil an Bewegung erscheinen im Differenzbild (Abb. 5.13b) als helle Regionen, welche sich durch einen Schwellenwert vom Hintergrund trennen lassen (Abb. 5.13c). Überlagert man schließlich die segmentierten Bildbereiche mit dem Originalbild ergibt sich Abb. 5.13d.

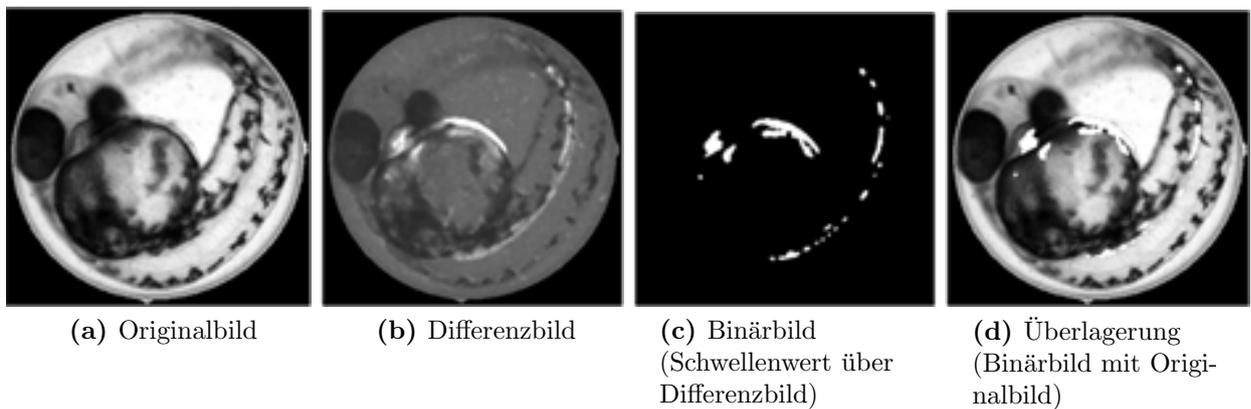


Abbildung 5.13: Quantifizierung der Bewegung im Bild, von links nach rechts: Originalbild; Differenzbild aus der Bildzeitreihe für Herzschlag; Differenzbild nach Schwellenwertoperation; Überlagerung Originalbild und Differenzbild für Herzschlag.

Zur Segmentierung des Bildstroms und für die Merkmalsextraktion von Spontanbewegung und Herzschlag werden lediglich die Merkmalsbilder über absolute Differenzen im Bild errechnet und abgespeichert. Der segmentierte Bildstrom besteht somit aus Kurzzeitreihen von je vier Differenzbildern für den Herzschlag und 29 Differenzbildern für die Spontanbewegung. Die Differenzbilder der Kurzzeitreihen wurden über die mit hoher Frequenz aufgezeichneten Bilder errechnet und die für Spontanbewegung über die mit niedriger Frequenz aufgezeichneten Bilder.

Nutzsignalzeitreihen: Aus jedem Differenzbild werden Merkmale extrahiert. Hierdurch entsteht eine Zeitreihe für den Herzschlag und eine Zeitreihe für die Spontanbewegung. Es wurden die Merkmale $x_9 - x_{15}$ extrahiert, wie sie im Modulkatalog im Abschnitt 3.4.2 beschrieben sind.

Dimensionsreduktion: Der Merkmalsatz wird durch Anwendung von Maximum- (MAX), Median- (MEDIAN) und Mittelwertoperatoren (MEAN) über die Abtastzeitpunkte gebildet. Die Operatoren MAX extrahieren die maximale Bewegung über der Zeit, was starke Bewegungsänderungen der Larve widerspiegelt. Der Median und der Mittelwert geben Auskunft über die Menge der Bewegung in der gesamten Zeitreihe.

Merkmalsauswahl: Anhand der Trainingsdaten wurde das trennstärkste Merkmal mittels ANOVA ermittelt und mittels MANOVA die beste 2er-Kombination bestimmt. In Tabelle 5.11 und Tabelle 5.12 sind die ermittelten Merkmalsrelevanzen für jedes alle Merkmale aufgelistet.

Die Kombination der Merkmale *MEAN ZR* x_{15} und *MAX ZR* x_{12} liefert für die Klassifikation der Spontanbewegung die beste Güte (vgl. Tabelle 5.11). Das erstgenannte Merkmal quantisiert den Durchschnitt der maximalen Bewegung. Das Merkmal weist, für starke Bewegungen, die über die gesamte Zeitreihe auftreten, hohe Werte auf. Besonders bei gesunden, aktiven Larven tendiert das Merkmal daher zu hohen Werten. Das zweite Merkmal quantifiziert das Maximum signifikanter Bewegungen. Da der Maximalwert extrahiert wird, toleriert es somit auch Ruhephasen ohne Bewegungen, wenn die Larven einige Zeit still liegen bleiben. Eine tote oder beeinträchtigte Fischlarve bewegt sich gar nicht oder nur schwach, was in beiden Merkmalen durch niedrige Werte angezeigt wird.

Die Merkmalsrelevanzen der ANOVA und MANOVA, welche den Herzschlag beschreiben, sind hingegen relativ niedrig (vgl. Tabelle 5.12). Das beste Ergebnis ergibt sich bei Merkmalen, welche nicht eine einzelne Bewegung im Gesamtbild, sondern eine Änderung beschreiben, welche über die gesamte Zeitreihe konstant hoch ist. Das ist verständlich, da der Herzschlag im Vergleich zur Spontanbewegung ein eher lokales Phänomen ist, das wenig globale Pixeländerungen hervorruft jedoch ohne Pause auftritt. Die extrahierten Merkmale beschreiben somit die Nutzsignalinformation und weisen eine ausreichende Trennschärfe auf (Quantifizierbarkeit des biologischen Effekts im Bild).

Auswertung & Klassifikation: Der Klassifikator der Spontanbewegung zeigte 1.92% Fehler auf den Lerndaten (156 Beispiele) und 5.52% Fehler auf den Testdaten (145 Beispiele). Wie in Abb. 5.14 zu erkennen ist, überlappen sich die Klassen, was zu weiteren Klassifikationsfehlern führt. Die Überlappung ist auf den Einfluss der Toxine zurückzuführen, welche die Nerven und Muskelaktivitäten nicht schlagartig, sondern stufenweise ausschalten. Eine Diskussion mit dem Experten ergab, dass bereits bei der manuellen Klassifikation viele Zuordnungen nicht scharf vornehmbar waren. Der Grund hierfür ist, dass es im Übergangsbereich stark vom subjektiven Eindruck des Experten abhängig ist, ob eine leichte Bewegung noch als Herzschlag gewertet wird oder nicht.

Die Klassifikation des Herzschlags weist 5.0% Fehler über den Lerndaten (140 Beispiele) und 9.85% Fehler über den Testdaten (132 Beispiele) auf. Die Herzschlagdetektion ist sensibler für Fehlklassifikationen, da die bewegte Stelle sehr klein ist. Ebenso ist der Übergang von einem sichtbaren Herzen zu einem verdeckten fließend, was wiederum zu Fehlklassifikationen führen kann. Des Weiteren bewirken die Toxine oftmals eine stark

Merkmal	Güte- MANOVA	Güte- ANOVA
MEAN ZR x_{15}	—	0.879
MAX ZR x_{12}	0.895	0.789
MAX ZR x_{13}	0.894	0.801
MAX ZR x_9	0.893	0.859
MEAN ZR x_{13}	0.887	0.834
MEAN ZR x_{12}	0.887	0.829
MAX ZR x_{10}	0.886	0.771
MEDIAN ZR x_{13}	0.885	0.819
MEDIAN ZR x_{12}	0.885	0.813
MAX ZR x_{14}	0.883	0.721
MAX ZR x_{15}	0.882	0.730
MEAN ZR x_9	0.882	0.820
MEDIAN ZR x_9	0.881	0.801
MEAN ZR x_{14}	0.881	0.878
MEAN ZR x_{10}	0.880	0.769
MEDIAN ZR x_{10}	0.880	0.746
MEDIAN ZR x_{14}	0.879	0.873
MEDIAN ZR x_{15}	0.879	0.874
MAX ZR x_{11}	0.879	0.777
MEDIAN ZR x_{11}	0.879	0.749
MEAN ZR x_{11}	0.879	0.772

Merkmal	Güte- MANOVA	Güte- ANOVA
MEAN ZR x_{14}	—	0.342
MEDIAN ZR x_9	0.577	0.016
MEAN ZR x_9	0.542	0.032
MEAN ZR x_{13}	0.489	0.015
MEAN ZR x_{12}	0.478	0.033
MEDIAN ZR x_{13}	0.476	0.004
MEDIAN ZR x_{12}	0.471	0.057
MEAN ZR x_{10}	0.464	0.017
MEDIAN ZR x_{10}	0.461	0.007
MAX ZR x_{12}	0.447	0.000
MAX ZR x_{13}	0.440	0.033
MAX ZR x_{10}	0.432	0.031
MAX ZR x_9	0.423	0.036
MAX ZR x_{14}	0.344	0.261
MAX ZR x_{15}	0.344	0.261
MEDIAN ZR x_{14}	0.343	0.297
MEDIAN ZR x_{15}	0.343	0.297
MEAN ZR x_{15}	0.342	0.342
MAX ZR x_{11}	0.342	0.033
MEDIAN ZR x_{11}	0.342	0.009
MEAN ZR x_{11}	0.342	0.020

Tabelle 5.11: Merkmalsrelevanzen für Spontanbewegung – Univariate und Multivariate Analyse (bei Auswahl von 2 Merkmalen)

Tabelle 5.12: Merkmalsrelevanzen für Herzschlag – Univariate und Multivariate Analyse (bei Auswahl von 2 Merkmalen)

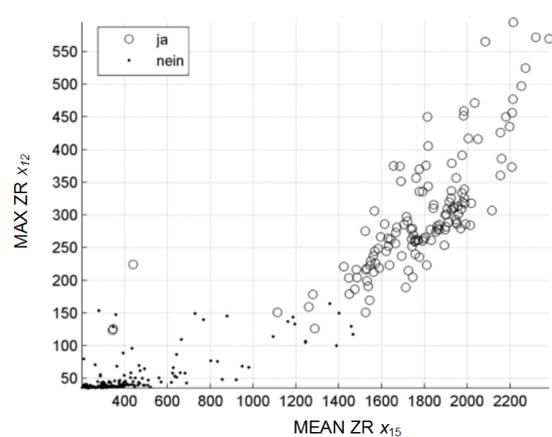
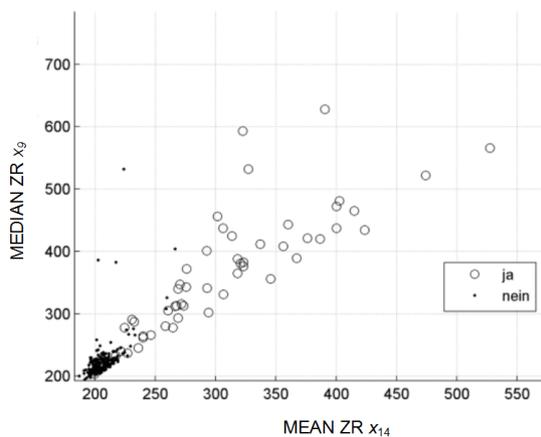


Abbildung 5.14: Scatterplot der besten Merkmale nach MANOVA; Herzschlag (links) und Spontanbewegung (rechts) [7]

herabgesetzte Funktion oder Frequenz des Herzens, was ebenfalls zu Fehlklassifikationen führt, da ein Experte üblicherweise nach einem kontinuierlich schlagenden Herzen sucht, während der Klassifikator selbst einen einzelnen Schlag abbildet. Die Scatterplots in Abb. 5.14 zeigen die Verteilung aller Datenpunkte. Wenige Bewegungen in einem Bild führen zu niedrigen Merkmalswerten; daher befinden sich die Datenpunkte toter oder beeinträchtigter Fischlarven bei niedrigen Werten. Hohe Bewegung, wie es z.B. bei vorhandenem Herzschlag bzw. Spontanbewegung der Fall ist, hat demnach größere Werte und höhere Varianzen zur Folge. Deutlich zu erkennen ist der „fließende“ Übergang zwischen den Einzelversuchen der Klassen (in Abbildung 5.14 durch Punkte bzw. Kreise dargestellt) vor allem beim Herzschlag. Insgesamt liefert die Klassifikationsroutine eine ausreichende Genauigkeit, da sich alle toxikologischen Größen von Interesse aus den Ergebnissen berechnen lassen (niedrige Fehlerrate bei der Detektion des biologischen Effekts).

Alle Klassifikatoren wurden im Anschluss zu einer Kaskade kombiniert (vgl. Abb. 5.15). Im ersten Schritt erfolgt die Bildverarbeitung, wie sie in den vorangegangenen Abschnitten beschrieben ist, bis hin zur Merkmalsauswahl. Mithilfe der Merkmale werden die Klassifikatoren angelern. Jede unbekannte Larve, d.h. nicht gelabelte Versuchseinheit, durchläuft nun die abgebildete Kaskade. Im ersten Schritt der Klassifikation wird, mit Hilfe der ermittelten Merkmalswerte, auf Koaguliertheit geprüft. Alle nicht koagulierten Larven werden dann bezüglich Spontanbewegung und im Anschluss daran auf Herzschlag geprüft. Nur wenn keiner der Endpunkte erreicht wird, wird die Larve als *entwickelt* eingestuft. Alle anderen Larven werden einem biologischen Endpunkt zugeordnet. Aus den Zuordnungen lassen sich daraufhin alle statistischen Werte von Interesse berechnen.

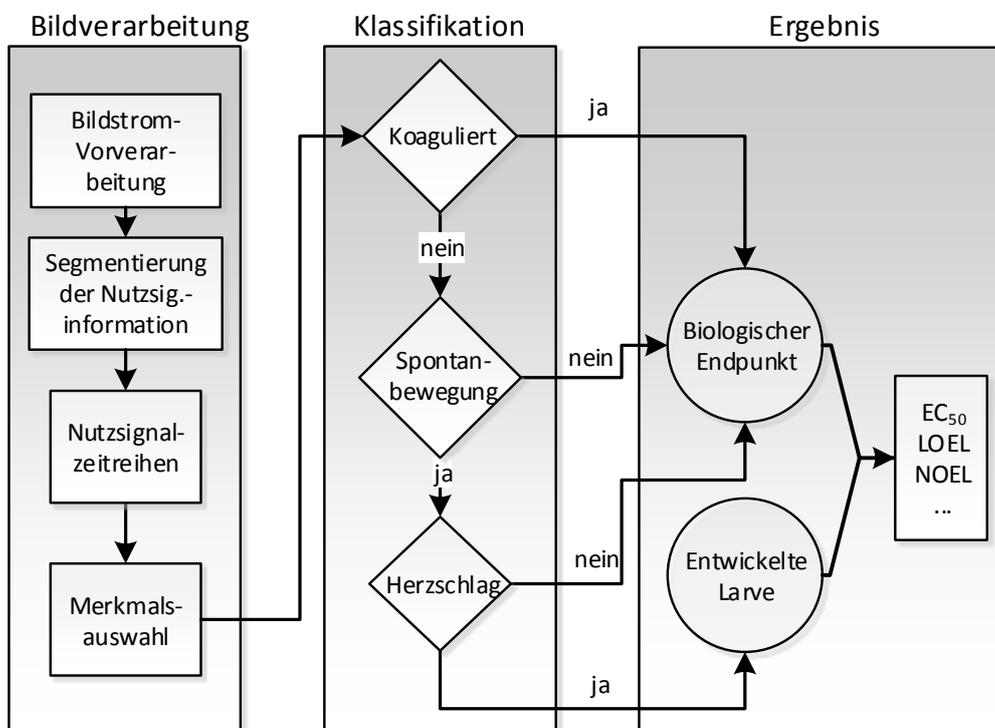


Abbildung 5.15: Klassifikatorkaskade über drei Endpunkte [7]

Präsentation: Die Ergebnisse werden identisch zum Endpunkt *Koagulation* dargestellt (Präsentierbarkeit der Auswertung) und archiviert (Wissenschaftliche Archivierung).

5.2.3.3 Durchführung und Resultat

Die vorgestellte neue Strategie zur Bildakquise gestattet es, alle drei Endpunkte anhand eines einzigen Datensatzes zu prüfen. Die durch den nur sporadisch auftretenden biologischen Effekt verursachte lange Akquise-Dauer und der demzufolge große Datensatz von ca. 1 TB Rohdaten lässt die Auswertezeit pro Platte ansteigen. Um bei Verwendung mehrerer Mikroskope die Hochdurchsatzfähigkeit zu sichern, wurden die Algorithmen für die vorliegende Arbeit so implementiert (vgl. Kapitel 4), dass ein verteiltes Rechnen möglich ist (Schnelligkeit der Auswertung).

Zum Nachweis der Leistungsfähigkeit der Klassifikatorkaskade wurden die drei Endpunkte mit den manuell gelabelten Daten bestimmt und mit den geschätzten Ergebnissen für einen Endpunkt und der Kaskade über alle drei Endpunkte verglichen (vgl. Tabelle 5.13).

	Nur Koaguliertheit	3 Endpunkte
Datensatz I (Tabelle 5.2)	99 % (98,8 %)	—
Datensatz II (Tabelle 5.8)	42,3 % (41,1 %)	95,8 % (82,2 %)

Tabelle 5.13: Vergleich der Klassifikationsergebnisse auf Basis von einem und von drei Endpunkten. Jeweils Ergebnisse über Lerndaten und in Klammern über Testdaten bzw. bei Datensatz I das Ergebnis aus der 5-fach Crossvalidierung. (Die manuell gelabelten Werte in Datensatz II entsprechen dem Ergebnis der gesamten Klassifikatorkaskade)

Wirken die Toxine vor allem auf die Bewegungen bzw. Muskeln der Larven, ist die Falsch-Positiv-Rate, trotz der guten Klassifikationsgüte für den Endpunkt *Koagulation*, hoch, da sich dies im Endpunkt *Koagulation* nicht widerspiegelt. Solche Larven weisen keinerlei Merkmale der *Koagulation* auf. Die kombinierte Klassifikation der drei Endpunkte hingegen zeigt eine gute Übereinstimmung (< 80% auf Testdaten) mit den manuell gelabelten Daten. Die Nachteile sind jedoch der deutlich größere Datensatz sowie die lange Dauer der Akquise und der Auswertung. Die Zeit zur Aufnahme einer einzigen Platte $t_{RV_{96}}$ beispielsweise beträgt 123 min im Vergleich zu lediglich 18 min in Abschnitt 5.2.2 (vgl. auch Tabelle 5.3 und Tabelle 5.9).

5.3 Photomotor Response (PMR)

5.3.1 Übersicht

Der Photomotor Response (*PMR*) ist ein biologischer Versuch mit Zebraabärblingslarven, welcher im Hochdurchsatz zur Anwendung kommt und erstmals von *Kokel et al.* vorgestellt wurde [80]. Die Arbeit stellt in Umfang und Durchsatz die bisher größte veröffentlichte Hochdurchsatz-Untersuchung mit Zebraabärblingen dar. Dennoch wird in der Arbeit erhebliches Potenzial, gerade in Bezug auf die Datenextraktion und Bildverarbeitung, nicht genutzt. Auch ist das Verhältnis von Nutzsignal zu Hintergrundrauschen aufgrund verschiedener Fehlereinflüsse wie Beleuchtung und Kontrast hoch. Die Anwendung bietet somit ein optimales Szenario, die in Kapitel 2 vorgestellte Methodik anzuwenden und deren Potenzial an einem realen Beispiel zu belegen. In einer Projektarbeit zwischen dem Institut für Angewandte Informatik und den o.g. Autoren von der Harvard Medical School wurde das ausgearbeitete Konzept auf die Auswertung des Bildstroms angewandt und implementiert. Die Vorgehensweise und Ergebnisse sind im Folgenden dargestellt. Zu Beginn wird der Hintergrund des Versuchs zum besseren Verständnis skizziert, daraufhin die vorgestellte Methodik angewandt und schließlich deren Ergebnisse vorgestellt.

5.3.2 PMR: Hintergrund, Rahmenbedingungen und Schwachpunkte

Der *PMR* nutzt ein stereotypes Verhalten von gesunden Zebraabärblingslarven, welches durch einen intensiven Lichtstimulus ausgelöst werden kann. Zebraabärblingslarven können sich bereits in einem frühen Entwicklungsstadium, mit einem Alter von ca. 28 hpf, innerhalb der Eihülle bewegen. Werden die Larven nun einem starken Lichtimpuls ausgesetzt, so antworten sie mit einem typischen *Zappeln*. Alle Bewegungen der Larven werden in [80] durch Vergleich von fünf Bildzeilen eines jeden Frames gemessen. Ein repräsentatives Beispiel einer solchen Bewegungsaufzeichnung über alle Larven der Bildsequenz ist in Abb. 5.16 dargestellt. Der gesamte Versuch wird als Photomotor Response (*PMR*) bezeichnet und lässt sich grob in folgende vier Phasen einteilen: Die erste Phase (Ruhezustand) des *PMR* ist das natürliche Verhalten bzw. die sog. *Spontanbewegungen* der Larven ohne jeglichen Einfluss von außen. Entsprechend bewegen sich die Larven mit der typischen Häufigkeit wie es bei ca. 28 hpf üblich ist [177]. Die Bewegungen treten zufällig auf und werden in [80] als Hintergrundsignal aufgefasst. Die erste Phase ist in Abb. 5.16 der Bereich von 0 bis 10 Sekunden des Versuchs. Die typischen Spontanbewegungen sind die Ausschläge im Bewegungsindex. Die Pfeile markieren jeweils den Zeitpunkt eines Lichtstimulus. In einem typischen *PMR*-Versuch werden zwei Lichtstimuli verwendet. Die zweite Phase kann als Reaktionszeit oder Latenzzeit bezeichnet werden und wird durch den Einsatz des Lichtstimulus nach 10 Sekunden sowie dem Beginn starker Bewegungen der Larven bei ca. 11 Sekunden abgegrenzt. Es ist somit die Ruhephase zwischen dem ersten

Stimulus und der Reaktion. Die dritte Phase ist die Bewegungsphase oder Anregungszeit. Die Phase zeichnet sich durch starke Zappelbewegungen der Larven aus und endet mit dem Abklingen der starken Bewegungen bei ca. 20 Sekunden. Es werden zwei typische Bewegungsereignisse unterschieden: Das sog. „Coiling“ und das „Swimming“. Die Begriffe stehen für unterschiedliche charakteristische Bewegungen, welche die Larve typischerweise nach einem Lichtimpuls durchführt. Coilings sind starke Drehungen der Larve und werden als ein instinktives „Umdrehen“ der Fische in eine zufällige Richtung interpretiert. Swimmings sind Bewegungen, bei denen die Larve, ohne große Drehungen auszuführen, sich auf der Stelle mit geringen Amplituden bewegt. Ein Swimming kann auch als eine Art „Zittern“ beschrieben werden und wird als ein fiktives „Flüchten“ oder Wegschwimmen der Larve interpretiert. Nach der Anregungszeit erfolgt ein weiterer Lichtstimulus, auf welchen die Larven typischerweise nicht reagieren und der die vierte und letzte Phase einleitet, die sog. Refractory-Phase oder Resistenzzeit. In der letzten Phase sind die Larven nicht nur unempfindlich gegenüber dem zweiten Lichtstimulus, sondern die Anzahl der natürlichen, zufällig auftretenden Spontanbewegungen ist auch deutlich geringer, verglichen mit dem *Ruhezustand* zwischen 0 und 5 Sekunden.

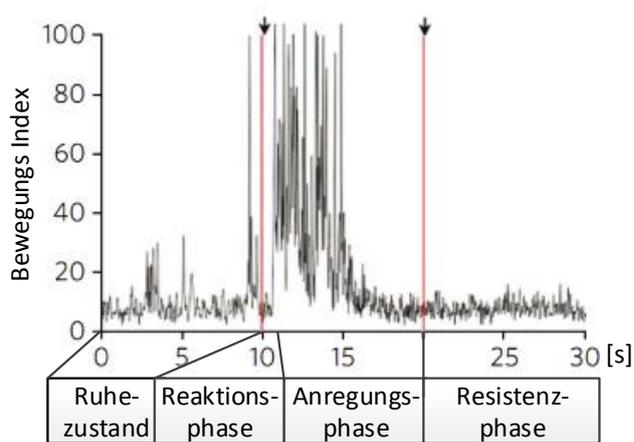


Abbildung 5.16: Repräsentatives Beispiel für die aggregierte Bewegung während eines *PMR*-Versuchs über der Zeit. Die Pfeile markieren die Zeitpunkte der Lichtstimuli (in Anlehnung an [80])

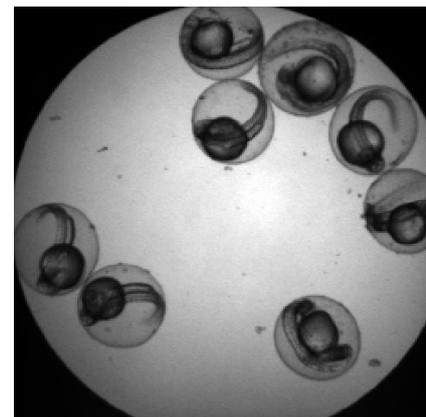


Abbildung 5.17: Einzelframe einer Bildsequenz aus einem *PMR*-Versuch. Acht Larven befinden sich in einem Nöpfchen einer 96er-Mikrotiterplatte [80].

Das Verhalten der Zebraäbrblingslarven ist hochgradig reproduzierbar und der Versuch lässt sich derart skalieren, dass er sich automatisiert in 96er-Mikrotiterplatten durchführen lässt. Beim vorliegenden Hochdurchsatzversuch wird eine große Anzahl von Eiern (bis zu 5000 Larven pro Tag) gewonnen. Die Eier werden bis zu einem Alter von 28 hpf in einem Inkubator gehalten und daraufhin in Mikrotiterplatten in Gruppen zu je 8-10 Larven pipettiert (vgl. Abb. 5.17). Durch den Einfluss von chemischen Substanzen kann das Bewegungsmuster der Larven erheblich von dem typischen oben beschriebenen Muster (vgl. Abb. 5.16) un behandelter Larven abweichen. Die Abweichungen lassen sich systematisch zu verschiedenen Einflussgruppen zusammenfassen, aus denen charakteristische Fingerabdrücke der Chemikalien erstellt werden können [82].

Die in [80] beschriebene Vorgehensweise und Bildverarbeitung hat sich als durchaus robust erwiesen, jedoch sind verschiedene Probleme, gerade im Bereich der Datenanalyse, nicht optimal gelöst. Beispielsweise erfolgt jegliche Datenakquise bzw. Zeitreihenextraktion aus den beschriebenen fünf Linien, welche äquidistant über das Bild verteilt sind, was den Nachteil hat, dass keine Information über das einzelne Individuum und dessen Bewegung generiert werden kann. Es wird lediglich der Mittelwert über alle Bewegungen gebildet und so die aggregierte Bewegung berechnet. Des Weiteren wird in [82] die Anzahl von Larven im Nöpfchen nicht exakt berücksichtigt. Da die untersuchte Anzahl jedoch zwischen 8 und 10 Larven variiert [82], unterliegt somit auch die Auswertung der gleichen Variation. Zudem ist nicht sichergestellt, ob die ausgewerteten Linien die Eier der Larven überhaupt bzw. nahe des Zentrums schneiden, was zur Folge hat, dass die Auswertung je nach Verteilung der Larven im Nöpfchen schwankt. Deshalb ist zu keinem Zeitpunkt klar, wie viel relevante Information ausgewertet wird und wie hoch der Rauschanteil ist. Die als Merkmal zur Charakterisierung der Bewegungsmuster herangezogene Messung unterliegt somit bei gleichen Bedingungen deutlichen Schwankungen.

Durch Anwendung des in der vorliegenden Arbeit vorgestellten Konzeptes lassen sich die genannten Fehlereinflüsse erheblich verringern, z.T. beseitigen und somit die Aussagekraft des Versuchs deutlich verbessern, ohne Einbußen im Durchsatz in Kauf nehmen zu müssen. Die folgenden Untersuchungen beschränken sich auf unbehandelte Larven, da es notwendig ist, die natürliche Reaktion der Larven auf Lichtstimuli vollständig zu untersuchen und zu beschreiben, bevor Vergleiche mit behandelten Larven vorgenommen werden können. Das Ziel der Anwendung des Konzeptes ist es,

- charakteristische Bewegungsereignisse von einzelnen Larven zu klassifizieren sowie
- Auftrittshäufigkeit und Dauer der Bewegungsereignisse einzelner Larven zu bestimmen.

Hierzu ist es notwendig,

- einzelne Larven in den Bildsequenzen zu finden und
- Merkmale zur Quantifizierung der Bewegung zu extrahieren.

5.3.3 Versuchsplanung: Biologie und Bildakquise

Da in [80] die Untersuchung bereits im Hochdurchsatz durchgeführt wurde, sind auch die Anforderungen an die Durchführung und Messung erfüllt. Somit können diese Punkte im Flussdiagramm Abb. 2.11 übersprungen werden und direkt mit der Erstellung der Auswertungskette fortgefahren werden. Um die Vergleichbarkeit der Auswertung mit den ursprünglichen Ergebnissen von [80] zu gewährleisten, wurde nur das Merkmal x_{16} aus Abschnitt 3.4.2 extrahiert, welches identisch mit dem in [80] verwendeten ist. Da die Aus-

sagekräftigkeit des Merkmales von [80] bereits gezeigt wurde, wurde auf einen Vergleich mit den anderen in der vorliegenden Arbeit eingeführten Merkmalen zur Beschreibung von Bewegungen ($x_9 - x_{17}$) verzichtet. Die Klassifikation der Bewegungsereignisse muss zudem direkt an der extrahierten Zeitreihe erfolgen. Aus diesen Gründen kann auf die Kategorien *Dimensionsreduktion* und *Merkmalsauswahl* des Modulkatalogs verzichtet werden.

5.3.4 Versuchsplanung: Analyse und Interpretation

Die Parameter der Hochdurchsatz-Untersuchung für die Auswertung werden bestimmt. Der biologische Effekt, die Bewegungsereignisse Coiling und Swimming, sind die Planfaktoren \mathbf{Y} und deren Auftreten soll automatisch, anhand von aus dem Bildstrom \mathbf{BS} extrahierten Zeitreihen $\mathbf{y}_{\mathbf{ZR}}$ geschätzt werden. Die Planfaktoren gehören den Ereignissen $\{ja, nein, unbekannt\}$ an. Im Hochdurchsatz soll der Versuch zum Ermitteln des Einflusses und zum Kategorisieren von Chemikalien p_y angewandt werden. Als Störfaktor z wurde die Position der Larven auf einer 96er-Mikrotiterplatte aufgezeichnet. Der Typ der Mikrotiterplatte ist der Parameter p_z des Störfaktors. Die Bildsequenzen wurden in Farbe $m_{\mathbf{BS}} = 3$ bei einer Auflösung von 250x250 Pixeln auf einer Fokusebene für 1000 Frames bei 30fps akquiriert. Eine Bildsequenz, eines Wells, belegt einen Speicherplatz von ca. 200MB. Für die Auswertung lag ein Datensatz von 60 Bildsequenzen vor. Tabelle 5.14 fasst die Parameter der Hochdurchsatz-Untersuchung zusammen.

y/\hat{y}	p_y	z/\hat{z}	p_z	$p_{\mathbf{BS}}$		
				$m_{\mathbf{BS}}$	$u_{\mathbf{BS}}$	$t_{\mathbf{BS}}$
Coiling Swimming	Chemikalie	Position	96-Mikro- titerplatte	3	250x250x1	1000

Tabelle 5.14: Parameter der Hochdurchsatz-Untersuchung nach Abschnitt 2.5 für PMR-Versuch; Mit $m_{\mathbf{BS}}$ Anzahl Kanäle, $u_{\mathbf{BS}}$ Anzahl Pixel und $t_{\mathbf{BS}}$ Anzahl Frames.

Die für die Auswertung der Bewegungsereignisse Coiling und Swimming ausgewählten Module sind in Abb. 5.18 hervorgehoben und werden im Folgenden an den jeweiligen Abschnitten erklärt.

Bildstrom-Vorverarbeitung: Da keine Farbinformationen von Interesse sind, können alle Farbebenen durch Mittelung der Farbkanäle zu einem Grauwertbild kombiniert werden, was den Speicherbedarf auf ein Drittel der ursprünglichen Größe reduziert. Zur Verminderung von Bildrauschen wird der Bildstrom mittels eines Gaußfilters geglättet. Die Auswahl von Zeitpunkten kann vorgenommen werden, da während der beiden Lichtstimuli die Kamera lediglich weiße (gesättigte) Bilder zeigt und die ersten drei Aufnahmen meist noch unterbelichtet sind. Alle Bilder, die während der beiden Lichtstimuli (jeweils während der Frames oder Abtastzeitpunkte $k = 250 \dots 300$, $k = 650 \dots 700$) aufgezeichnet

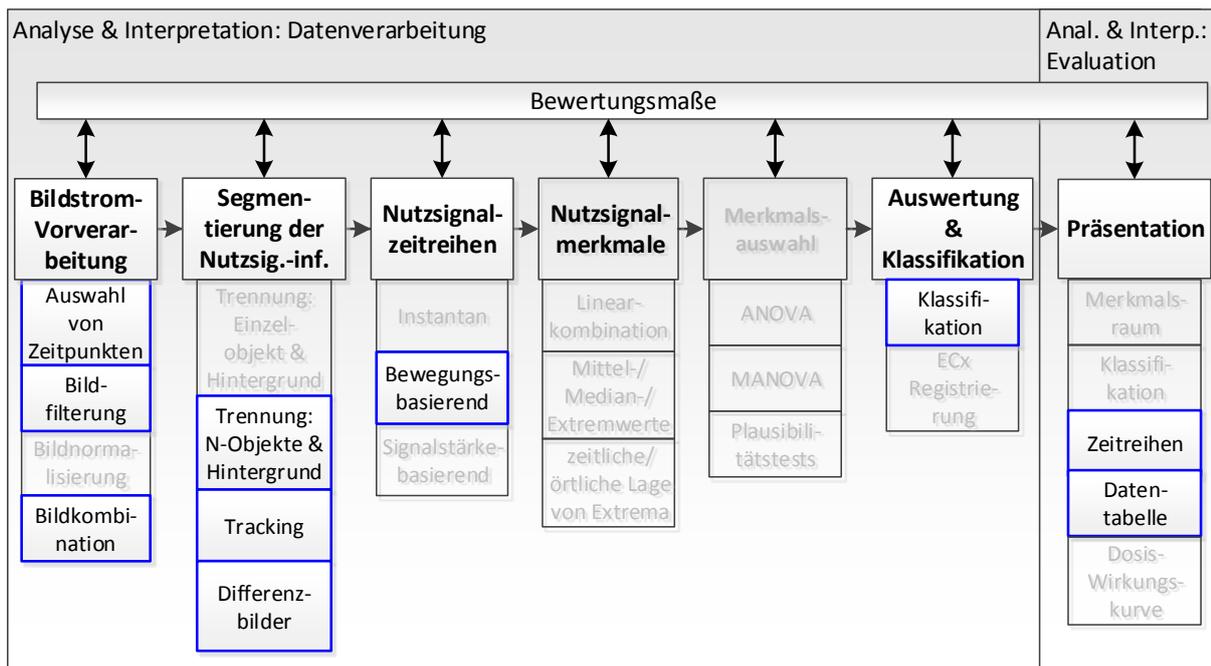


Abbildung 5.18: Module zur Auswertung der Bewegungsereignisse Coiling und Swimming.

werden sowie die ersten drei Aufnahmen des Bildstroms ($k = 1 \dots 3$) werden daher verworfen (Auswahl von Zeitpunkten).

Segmentierung der Nutzsiginformation: Nach der Bildstrom-Vorverarbeitung müssen alle Objekte von Interesse im Bild gefunden, vom Hintergrund getrennt und voneinander unterschieden werden. Dabei handelt es sich bei den Aufnahmen um eine unbekannte Anzahl von Eiern. Die Anzahl schwankt üblicherweise zwischen 8 und 10, kann jedoch bei einem versehentlich nicht gefüllten Nöpfchen auf 0 sinken oder auch deutlich höher liegen, z.B. durch einen Pipettier-Fehler. Zur Anwendung kommt das Modul *Trennung mehrerer Objekte vom Hintergrund*. Es wird auf das erste betrachtete Frame ($k = 4$) angewandt. Im binären Kantenbild, welches bei der Kreissuche initial erstellt wird, sind die zu suchenden Kreise deutlich zu erkennen, wie Abb. 5.19 exemplarisch zeigt.

In der Abbildung ist weiter gut zu erkennen, dass die Kreise nur bei ausreichendem Kontrast durch das Kantenbild vollständig wiedergegeben werden. Es zeigte sich, dass z.B. ein niedriger Wasserstand im Nöpfchen das Licht der Durchlichtbeleuchtung derart streut, dass sich ein starker Helligkeitsverlauf zum Rand hin einstellt. In bestimmten Toleranzgrenzen kann der Helligkeitsverlauf von dem Algorithmus kompensiert werden. Im betrachteten Beispiel ist der Grauwert des Hintergrundes im Zentrum gesättigt, d.h. vollständig weiß bei einem Pixelwert von 255, während der Wert der Pixel zum Rand hin auf 80 abfällt. Je dichter die Eier am Rand liegen und je dunkler der Hintergrund ist, desto geringer ist der Kontrast und desto schlechter können die Eier erkannt werden. Die zur Kreisdetektion im nächsten Schritt berechnete Akkumulationsmatrix der Hough-Kreisdetektion des Moduls ist in Abb. 5.20 dargestellt. Es wurde nach Kreisen mit einem Radius zwischen 20 und 30 Pixel gesucht. Bei der Darstellung wird jedes Pixel im Bild als

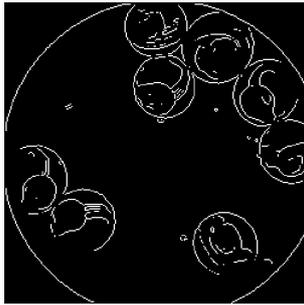


Abbildung 5.19: Kantenbild des Beispiel-Screenshots aus Abb. 5.17. Die Eihüllen sind deutlich als z.T. unterbrochene Kreise zu erkennen.

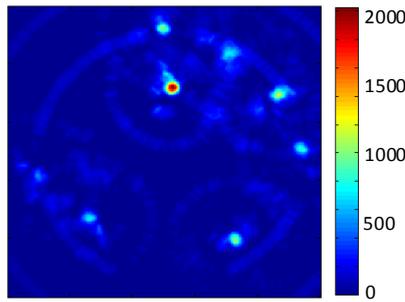


Abbildung 5.20: Akkumulationsmatrix (heatmap) der Hough-Kreisdetektion für Radien zwischen 20 und 30 Pixel

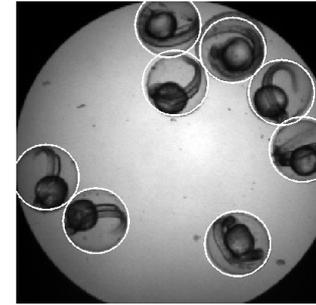


Abbildung 5.21: Identifizierte Kreise aus Mittelpunkten aller Maxima der Akkumulationsmatrix über definiertem Schwellenwert

potentieller Kreismittelpunkt betrachtet. Für jedes weiße Pixel im Kantenbild werden alle theoretisch möglichen Kreismittelpunkte in der Akkumulationsmatrix markiert. Tatsächliche Kreismittelpunkte erhalten so einen hohen Score in der Akkumulationsmatrix. Alle Punkte über einen Score von 500 in der Akkumulationsmatrix werden schließlich als tatsächlicher Kreismittelpunkt aufgefasst. Der Beleuchtungseffekt ist deutlich an schwächeren Maxima in den Randregionen der Akkumulationsmatrix zu erkennen. In Abb. 5.21 sind die segmentierten Eier markiert. Ist der Beleuchtungseffekt zu groß, werden Eier fehlerhaft segmentiert, was eine Validitätsprüfung erfordert. Die tatsächlich vorhandenen Kreise müssen von den fälschlicherweise detektierten unterschieden werden. Es wird der Abstand aller Kreismittelpunkte ermittelt und alle diejenigen identifiziert, welche näher beieinander liegen als 90% des typischen Eidurchmessers. Daraufhin werden nacheinander Kreismittelpunkte verworfen, welche potenziell die Zebraablarve nicht vollständig erfassen, wie es z.B. in Abb. 5.21 oben zu sehen ist. Ein typisches Merkmal für eine vollständig erfasste Larve ist der mittlere Grauwert innerhalb des Kreises. Da Vordergrundpixel, also Pixel, welche die Larve abbilden, dunkel erscheinen, ist der mittlere Grauwert niedriger, wenn die Larve korrekt gefunden wurde und daher mehr Vordergrundpixel enthalten sind. Es werden nacheinander Kreise mit hohen Mittelwerten verworfen, bis alle übrigen einen Abstand von mehr als 90% des typischen Eidurchmessers zueinander haben. Zwei weitere Plausibilitätskriterien schließen gefundene Eier aus, welche entweder einen deutlich zu hohen oder niedrigen Mittelwert aufweisen. Ein solcher Fall kann eintreten, wenn z.B. ein unbefruchtetes Ei, also eine leere Eihülle, in der Bildsequenz vorkommt. Manchmal sind auch Luftbläschen auf den Bildern, welche fälschlicherweise als Eier erkannt werden. Durch die Plausibilitätskriterien können solche Fälle abgefangen werden.

Werden alle Eier im Bild fehlerfrei gefunden, so ist das initiale Segmentieren aller Zebraablarven abgeschlossen und die Position im Bild sowie die Abmaße des umspannenden Rechtecks werden gespeichert. Der nächste Schritt ist das Verfolgen der Eier, hierfür wird das Modul „*Tracking*“ angewandt. Exemplarisch ist die Anwendung des Moduls in Abb. 5.22 abgebildet. Zur besseren Visualisierung wurde in der Abbildung der Suchraum nicht durch Vorwissen eingegrenzt, wie es bei der Implementierung der Fall ist. Bei dem Verfahren wird das in der Abbildung rot eingerahmte Suchbild (engl. Template) über das

Kriterium [#]	Beschreibung
1	Frames, die 25% heller sind als der zuvor bestimmte Mittelwert korrekter Frames, werden verworfen.
2	Teilweise oder vollständig außerhalb der Bildmatrix detektierte Eier werden verworfen.
3	Detektierte Eier mit einer mittleren Helligkeit von über 200 werden verworfen.
4	Detektierte Eier mit einer mittleren Helligkeit unter 50 werden verworfen.
5	Eier mit einem Durchmesser größer oder kleiner als die Suchradien (hier zwischen 20 und 30 Pixel) werden verworfen.
6	Außerhalb des Suchraums für das Tracking (hier 4 Pixel in X- und Y-Richtung) detektierte Eier werden verworfen.
7	Eier, deren Mittelpunkte näher als 90% des zuvor ermittelten typischen Eidurchmessers sind, werden verworfen.
8	Bewegungen außerhalb des detektierten Kreises der detektierten Eier werden verworfen.

Tabelle 5.15: Übersicht der Validitätskriterien beim PMR

Zielbild geschoben und mittels Kreuzkorrelation die Übereinstimmung ermittelt. Es lassen sich sieben Maxima erkennen, von denen sich eines deutlich abhebt. Da sich zwischen einzelnen Frames das Ei nur wenige Pixel bewegt haben kann, ist der Suchraum auf jeweils vier Pixel in X- und Y-Richtung eingeschränkt. Damit ist eine Verwechslung der Maxima ausgeschlossen. Das Modul ermittelt die wahrscheinlichste Position des Suchbildes im Zielbild.

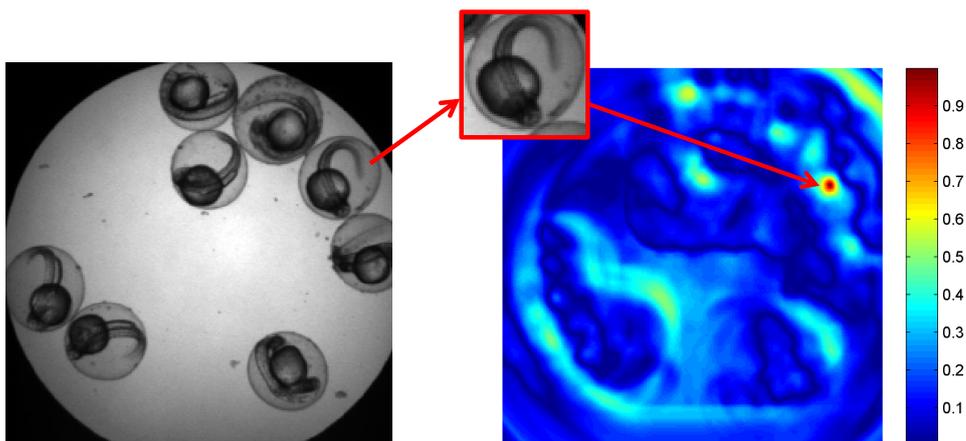


Abbildung 5.22: Kreuzkorrelation des Quell- oder Objektbildes (Templates) mit dem Zielbild oder Referenzbild aus dem darauf folgenden Frame. Links: Originalbild; Mitte: Template; rechts: Erzielte Übereinstimmung des Templates für die jeweilige Position im Zielbild. Hohe Werte bedeuten eine hohe Übereinstimmung.

Die Bilder der Eier zu jedem Abtastzeitpunkt werden nicht zugeschnitten und erneut abgespeichert, sondern es wird lediglich die Position und Ausdehnung des Eies aufgezeichnet. Das Ergebnis ist ein Datensatz, der zu den Rohdaten an jedem Abtastzeitpunkt charakteristische Größen für jedes Ei enthält. Die Größen sind Eigröße und Position sowie der mittlere Grauwert. Das Tracking aller Eier einer Bildsequenz benötigt ca. 20 s (inkl. Einlesen der Bildsequenz und Speichern der Ergebnisse) auf einem zum Zeitpunkt des Entstehens der vorliegenden Arbeit üblichen Desktop-PC (Segmentierbarkeit des biologischen Effekts/Schnelligkeit der Auswertung). In der Benutzeroberfläche (vgl. Kapitel 4) werden als Parameter Eidurchmesser zwischen 20 und 30 Pixel gewählt (bei einer Auflösung der Sequenzen von 250 x 250 Pixel) und detektierte Eier mit einem mittleren Grauwert unter 50 und über 200 werden verworfen. Tabelle 5.15 listet alle Validitätskriterien auf.

Nutzsignalzeitreihen: Ist die Position der Eier in jedem Frame ermittelt, müssen für jeden Zeitpunkt Merkmale für die Bewegung extrahiert werden. Bewegungen im Bild führen direkt zu Veränderungen in Pixelwerten, welche nun individuell für jedes Ei berechnet werden können. Jede Bewegung der Larve ist mit einer Veränderung der Pixelwerte in der zugeordneten Region verbunden, daher wird für die Bewegung die Summe aller Absolutwerte der Pixeldifferenzen der Region gebildet. Das Ergebnis ist eine Zeitreihe, welche die Bewegung der jeweiligen Larve quantifiziert und im Folgenden *Bewegungsindex* genannt wird. Da das Merkmal nicht mehr wie zuvor die Bewegung mittels Bildzeilen über die gesamte Breite der Rohdaten quantifiziert, sondern der Bewegungsindex für jede Larve individuell bestimmt wird, schwankt die Dynamik und damit der Bewegungsindex im Bildausschnitt, je nachdem ob sich die Larve an einer gut oder schlecht ausgeleuchteten Position im Well befindet. Um den Beleuchtungseffekt zu minimieren, wird der Bewegungsindex mit der Standardabweichung des jeweiligen Bildausschnitts normalisiert. Der Bewegungsindex setzt sich aus den Merkmalswerten des Merkmals x_{16} des Modulkatalogs zu jedem Abtastzeitpunkt zusammen.

Das Ermitteln der Eiposition mithilfe des Moduls zum Tracken auf Basis eines einfachen Kreises ist zwar äußerst robust gegenüber dem beschriebenen Wegdriften der Eiposition, jedoch weicht die gefundene Position in manchen Fällen um wenige Pixel von der optimalen Position ab. In solchen Fällen werden Veränderungen der Pixelwerte gemessen, welche nicht auf die Bewegung der Larve, sondern auf die Bewegung der Bildausschnitte zueinander zurückzuführen ist. Um lediglich Bewegungen der Larven zu verfolgen, wurde eine Feinjustierung durchgeführt. Hierzu wird als Suchbild anstatt des einfachen Kreises der Ausschnitt des Originalbildes verwendet, welches auf alle Nachbarpositionen verschoben wird und daraufhin erneut o.g. Merkmal berechnet wird. Die Position, an der der Wert des Merkmals minimal wird, ist identisch mit der optimalen Position des Suchbildes und enthält die reine Bewegungsinformation. Zur weiteren Verbesserung der Robustheit wurde das beschriebene Verfahren für eine Abweichung von bis zu zwei Pixeln durchgeführt. Eine weitere Verbesserung der Merkmalswerte wird durch Hinzufügen einer zusätzlichen Region von Interesse (ROI) innerhalb des segmentierten Bereichs erzielt. Da die Bildausschnitte in Matlab nur als rechteckige Matrizen gespeichert werden können, erscheinen Teile von Nachbarlarven, welche sich diagonal zu dem aktuellen Ei befinden, im ausgewerteten Bildausschnitt. Bewegt sich nun die Nachbarlarve, erzeugt diese Bewegung bei

Berücksichtigung des kompletten Bildausschnitts eine leichte Bewegung in der Zeitreihe, die jedoch nur der Nachbarlarve zuzuordnen ist. Die ROI wurde durch die Multiplikation der Bildausschnitte mit einer kreisrunden Maske durchgeführt, welche den typischen Eisdurchmesser aufweist und alle Pixel außerhalb des auszuwertenden Eies auf den Wert 0 setzt. Das Ergebnis ist eine Zeitreihe der Bewegung einer einzelnen Larve und exemplarisch in Abb. 5.23 dargestellt. Es wurden 425 Bewegungszeitreihen von Larven, also ca. sieben Larven pro Bildsequenz, gefunden und extrahiert (Quantifizierbarkeit des biologischen Effekts im Bild).

Auswertung & Klassifikation: Da für jede Larve eine detaillierte Information über die Reaktion auf den Lichtstimulus extrahiert wurde, kann nun eine genaue Untersuchung der typischen Reaktionsformen und Bewegungsereignisse folgen. Die Zeitreihen werden als Gait-CAD-Projekt gespeichert und dort bezüglich der Bewegungsereignisse klassifiziert. Im Bewegungsprofil sind Coiling Bewegungen als deutlich separierte hohe Ausschläge erkennbar, während das Swimming ein etwas längerer Abschnitt mittlerer Amplitude ist. Zur Ermittlung der Anzahl und zum Unterscheiden zwischen beiden Ereignissen werden zuerst alle hohen Ausschläge im Bewegungsindex identifiziert (vgl. Abb. 5.23).

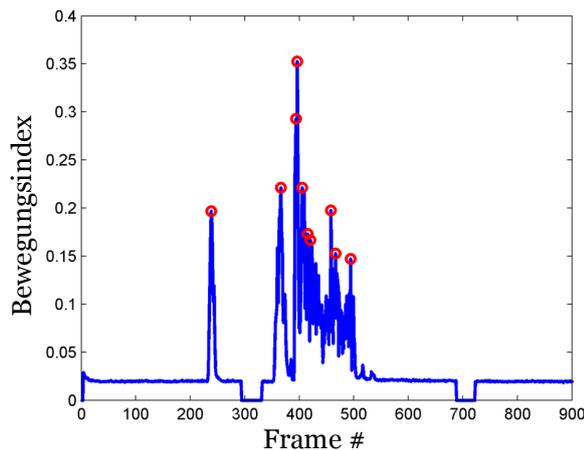


Abbildung 5.23: Klassifikation der Coiling-Ereignisse anhand von Peak-Erkennung in der Bewegungszeitreihe.

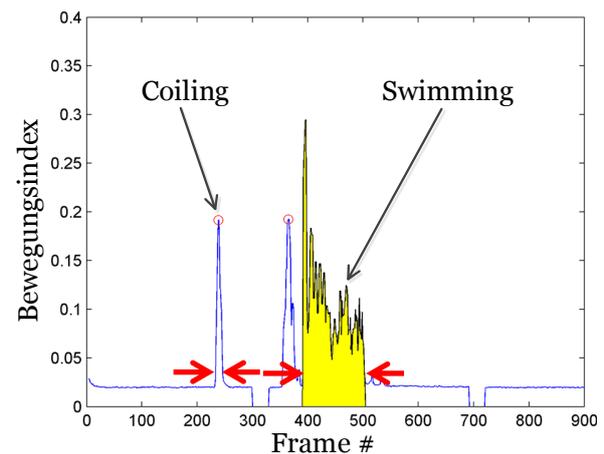


Abbildung 5.24: Klassifikation der Swimming-Ereignisse anhand von zwei Schwellenwerten.

Daraufhin wird die Länge des Intervalls bestimmt, in dem sich die Larve bewegt. Ein manueller Vergleich von Videosequenzen und den Werten des Bewegungsindex ergibt, dass typische Coiling-Ereignisse weniger als 40 Frames andauern (oder 1,33 s bei 30,03 fps). In Abb. 5.24 ist die Länge zweier Intervalle anhand roter Pfeile gekennzeichnet. Das linke Intervall ist ein Coiling-Ereignis, während das rechte, deutlich längere Intervall ein Swimming-Ereignis ist (gelb gekennzeichnet). Mittels der so bestimmten Länge der Intervalle um die detektierten Ausschläge werden alle Ereignisse mit einer Intervalllänge von weniger als 40 Frames als Coiling-Ereignisse und alle anderen als Swimming-Ereignisse klassifiziert.

Die Anwendung der Klassifikation auf alle extrahierten 425 Zeitreihen ermöglicht es, typische Bewegungsmuster und die Zeitdauer verschiedener Phasen zu bestimmen. In Abb. 5.25 ist die Wahrscheinlichkeit für ein Auftreten genau eines Coiling-Events (grün) bzw. eines Swimming-Events (blau) an jedem Frame aufgetragen. Deutlich zu sehen ist der Peak der Wahrscheinlichkeit direkt nach dem ersten Lichtstimulus (Frame 300) sowie die Abnahme der Wahrscheinlichkeit für einen Coil in der Resistenzphase (Frame 700-900) im Vergleich mit dem Ruhezustand (Frame 0-300). Die Wahrscheinlichkeit für ein Swimming hat ihren Peak kurz nach dem Peak Coiling-Wahrscheinlichkeit.

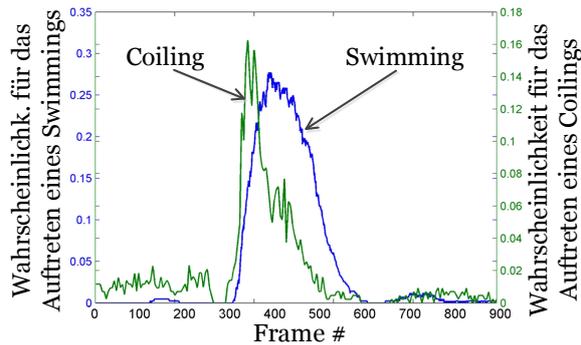


Abbildung 5.25: Wahrscheinlichkeit für das Auftreten von genau einem Swimming- oder Coiling-Ereignis für jedes Frame

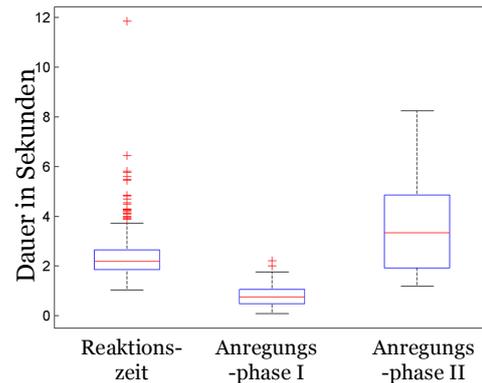


Abbildung 5.26: Dauer der Phasen, ermittelt anhand der Klassifikationen

In Abb. 5.26 ist auf der Grundlage der gemessenen Bewegungen die jeweilige Dauer der Phasen bestimmt und über alle Larven als Boxplot dargestellt. Die Reaktionszeit ist die Dauer in Sekunden vom Lichtstimulus bis zur ersten messbaren Reaktion der Larve (Median bei etwa 2 Sekunden Dauer). Die in Abb. 5.16 eingeführte Anregungsphase lässt sich mit Hilfe der Klassifikationen nun in zwei Teile aufspalten. Anregungsphase I ist die Dauer vom Ende der Reaktionszeit bis zum Einsetzen der Swimming-Ereignisse (Median bei etwa einer Sekunde). Die Dauer der Swimming-Ereignisse ist Anregungsphase II (Median bei etwa drei Sekunden). Die gesamte Auswertung einer Bildsequenz benötigt ca. 2 min. Eine 96er-Mikrotiterplatte mit ca. 770 Larven kann an einem einzigen Computer somit unter 3:20 h ausgewertet werden. Die Art der Implementierung (siehe Kapitel 4) ermöglicht zudem die Beschleunigung der Auswertung durch das verteilte Rechnen auf mehreren Computern (Schnelligkeit der Auswertung).

Präsentation: Ziel der Visualisierung ist es, die große Menge an Bewegungszeitreihen und Klassifikationen derart zu veranschaulichen, dass Bewegungsmuster erkannt werden können. Wird die Bewegungszeitreihe je nach Höhe des Ausschlags mit einem Farbcode eingefärbt und wird die Kurve „von oben“ betrachtet, so ergibt sich ein Farbprofil, in welchem blau für keine Bewegung und rot für starke Bewegungen definiert wurde (vgl. Abb. 5.27). Wird der Farbcode für jede Larve in jeweils einer Zeile aufgetragen, ergibt sich eine Heatmap-Übersicht wie sie in Abb. 5.28 dargestellt ist. Deutlich sind die vereinzelt Ausschläge im Ruhezustand, die Ruhepause während der Reaktionszeit, die Anregungsphase

sowie die Resistenzzeit zu erkennen. Die Übersicht ermöglicht einen schnellen Überblick über alle Larven und lässt sich bis ca. 500 Larven gut visualisieren. Die Anzahl von Ereignissen wiederum lässt sich zu *Datentabellen* zusammenfassen. Die Zeitdauer der Phasen lässt sich entsprechend der Abb. 5.26 z.B. mittels Boxplots darstellen.

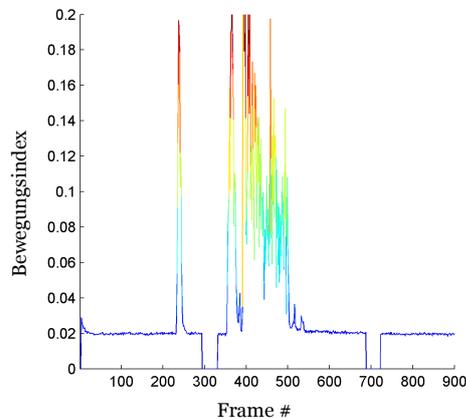


Abbildung 5.27: Extrahierte Bewegungszeitreihe. Hohe Werte sind rot, niedrige blau eingefärbt.

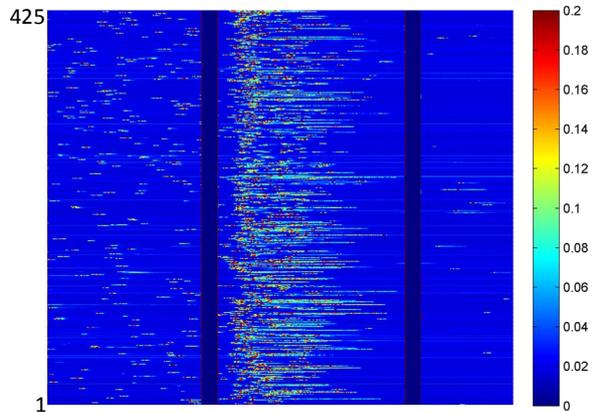


Abbildung 5.28: Zusammenfassung aller 425 Fischlarven als Heatmap. Die einzelnen Bewegungsphasen der Larven sind deutlich zu erkennen.

Zur biologischen Auswertung und Archivierung werden die Ergebnisse in Form eines PDF-Dokumentes automatisch generiert. Das Dokument enthält einen Screenshot des ersten und letzten Frames. Hierin sind die Ergebnisse der Segmentierung überlagert und die Larven durchnummeriert (ähnlich zu Abb. 5.21). So kann mit einem Blick geprüft werden, ob die Segmentierung und das Tracking erfolgreich sind. Ebenso werden die Bewegungszeitreihen der Larven in o.g. Heatmaps dargestellt (Präsentierbarkeit der Auswertung). Alle Daten wurden identisch zu den anderen Untersuchungen archiviert (Wissenschaftliche Archivierung).

5.3.4.1 Durchführung und Resultat

Mit den Zeitreihen ist es nun erstmals möglich, die individuellen Bewegungen zu klassifizieren. Es wurden Frames mit 1.540 Coiling-Ereignissen und 19.775 Frames mit Swimming-Ereignissen identifiziert (Aus einer Gesamtzahl von 382.500 Frames). Zeitlich ereignet sich das Maximum der Swimming-Ereignisse nach dem Maximum der Coiling-Ereignisse. Dies lässt sich auch in Abb. 5.25 ablesen. In der Abbildung ist die Wahrscheinlichkeit für das Auftreten eines Coilings (blaue Kurve) und das Auftreten eines Schwimmings (grüne Kurve) zu jedem Zeitpunkt angegeben. Die Verteilung zeigt ein starkes Ansteigen der Bewegungen nach dem Stimulus. Nach der Reaktionszeit von 2.2 s (66 Frames) auf den Stimulus steigt das Auftreten der Coiling-Frequenz um etwa das 10fache von 0.07 Coilings/s auf 0.82 Coilings/s für die mittlere Dauer von ca. 0.8 s (Abb. 5.26). Während vor dem Stimulus

nur bei 2 Larven ein Swimming klassifiziert wurde, zeigen 161 Larven nach dem Stimulus ein Swimming-Ereignis, mit einer mittleren Dauer von 3.3s. Die Werte legen den Schluss nahe, dass die Anregungsphase sich in eine Anregungs-Phase I und Anregungs-Phase II aufspalten lässt. Nach der Anregungsphase fällt die Rate der Coilings auf einen Wert von 0.02 Coilings/s ab und gesunde Larven reagieren nicht auf einen weiteren Lichtstimulus (Abb. 5.28). Die Ergebnisse trugen schließlich zur Identifikation von biologischen Wirkungsketten und zu einem tieferen Verständnis der Reaktionen und Bewegungsmuster der Larven bei [81]. In [81] findet sich zudem eine genauere biologische Interpretation der o.g. Werte.

5.4 Bewertung

In Kapitel 5 wurde die Funktionsweise des in der vorliegenden Arbeit neu entwickelten Konzeptes aus Kapitel 2 anhand von zwei realen Problemstellungen mit jeweils steigender Komplexität belegt. Für die Untersuchungen steht nun eine automatische und aussagekräftige Auswerterroutine zur Verfügung. Auf die Ergebnisse beider Projekte wird im Folgenden gesondert eingegangen.

5.4.1 Fisch Embryo Test

In Abschnitt 5.2 wurde das neue Konzept auf zwei toxikologische Untersuchungen angewandt. Die Auswertung des Fisch Embryo Tests (FET) wurde für drei verschiedene biologische Endpunkte bildbasierend und automatisiert durchgeführt. Anhand von Einzelaufnahmen gelingt es, den Endpunkt *Koagulation* zu klassifizieren. Mittels Bildsequenzen wird die Auswertung auf die Endpunkte *Herzschlag* und *Spontanbewegung* ausgeweitet.

In einer Voruntersuchung wurden die Versuchs- und Auswerteparameter nach dem neuen Konzept gemäß Abb. 2.1b festgelegt. Zur Sicherstellung der Berücksichtigung aller in Abschnitt 2.6 ermittelten Anforderungen an eine Hochdurchsatz-Untersuchung wurde bei der Auslegung das Flussdiagramm Abb. 2.11 durchlaufen. Ein wesentliches Ergebnis der Voruntersuchung ist, dass für eine erfolgreiche Untersuchung alle extrahierten Merkmale zur Klassifikation lageunabhängig sein müssen.

Für beide Untersuchungen wurde auf Basis der Voruntersuchung eine Auswertungskette durch Auswahl geeigneter Module aus Kapitel 3 erstellt. Durch die neuen Methoden zur Segmentierung und zur Vorbereitung des Bild- und Datenstroms gelingt für beide Fälle die Segmentierung und Merkmalsextraktion. Die für die vorliegende Arbeit entwickelten instantanen Merkmale x_1 bis x_8 des Modulkatalogs (vgl. Abschnitt 3.4.1) ermöglichen die Klassifikation des Endpunktes *Koagulation*. Die Merkmale x_9 bis x_{15} zur Klassifikation von Bewegungen ermöglichen die Erweiterung auf die Endpunkte *Herzschlag* und *Spon-*

tanbewegung. Die Ergebnisse der Klassifikation wurden durch geeignete Methoden des Modulkatalogs präsentiert.

Abschließend werden die Ergebnisse beider Untersuchungen zu einer Klassifikatorkaskade kombiniert (vgl. Abb. 5.15), mit der nun erstmals die bildbasierte automatisierte Auswertung des FET anhand von drei Endpunkten zur Verfügung steht.

5.4.2 Photomoter Response

In Abschnitt 5.3 wurde das in der vorliegenden Arbeit erarbeitete Konzept aus Kapitel 2 erneut erfolgreich angewandt. Nach der Anwendung auf Einzelaufnahmen und der anschließenden Erweiterung auf Bildsequenzen für vereinzelte Zebrabärblingslarven innerhalb der Versuche des FET in Abschnitt 5.2 ist es nun erstmals möglich, einzelne Larven auch ohne vorherige manuelle Vereinzelung zu untersuchen. Für den PMR bedeutet das, bei unverändertem Durchsatz, eine erhebliche Verringerung der Fehlereinflüsse und eine deutliche Verbesserung der Aussagekraft verglichen mit dem in [80] vorgestellten Verfahren. Die verbesserte Aussagekraft des in der vorliegenden Arbeit vorgestellten Verfahrens trug direkt zur Identifikation sog. *Response Cells* im Rautenhirn (engl. Hindbrain) bei, welche für die Auslösung der stereotypen Antwort der Zebrabärblinge auf Lichtimpulse verantwortlich sind [81].

Die Versuchs- und Auswerteparameter wurden wieder nach dem neuen Konzept gemäß Abb. 2.1b festgelegt und die Erfüllung aller in Abschnitt 2.6 ermittelten Anforderungen sichergestellt. Durch Auswahl geeigneter Module aus dem Modulkatalog wurde eine Auswertungskette erstellt. Die in der vorliegenden Arbeit neu entwickelten Methoden zur Segmentierung und zum Tracking der Fischeier ermöglichen, nach Extraktion von Merkmalen, das Ermitteln eines charakteristischen Bewegungsindex für jede Fischlarve einzeln. Auf Basis des Bewegungsindex wird eine Klassifikation der beiden Ereignisse *Coiling* und *Swimming* möglich. Schließlich wird die Leistungsfähigkeit der Auswertungskette auf 425 Zebrabärblingslarven belegt und die Dauer charakteristischer Phasen (Anregungsphase I und II, Reaktionszeit) ermittelt.

In Form der Auswertungskette steht nun ein leistungsfähiges Verfahren zur Analyse und Klassifikation einzelner Zebrabärblinge, ohne die Notwendigkeit einer vorherigen manuellen Vereinzelung, zur Verfügung.

6 Zusammenfassung und Ausblick

Bildbasierte Hochdurchsatz-Untersuchungen am Zebrabärbling setzen hohe Anforderungen sowohl an den Versuchsentwurf als auch die automatische Analyse und Interpretation. Die große Anzahl an Einzeluntersuchungen und die dementsprechend große Datenmenge nach der Bildakquise machen ein umfassendes Konzept zur Versuchsauslegung sowie neue Auswerteverfahren notwendig.

Zur optimierten Versuchsauslegung schlägt die vorliegende Arbeit zum einen ein strukturiertes Verfahren vor, mit dem es möglich wird, die Versuchsparameter, angepasst für die automatisierte Bildauswertung, zu bestimmen. Zum anderen werden neue Bildverarbeitungsmodulare, die sowohl für den Zebrabärbling als auch den Hochdurchsatz entwickelt wurden, vorgestellt. Die Kombination beider Teile bietet dem Anwender, i.A. dem Biologen, einen Lösungsweg für Hochdurchsatz-Untersuchungen am Zebrabärbling, der nicht nur neue Nutzsignale auswertbar macht, sondern auch die Auslegung optimiert und somit Datenmenge, redundante Informationen und Arbeitsaufwand sowie Klassifikationsfehler reduziert.

Die Kernpunkte der vorliegenden Arbeit umfassen:

- Die Entwicklung eines Konzeptes zur strukturierten Versuchsauslegung für die Hochdurchsatz-Untersuchung am Zebrabärbling (Kapitel 2).
- Die Entwicklung neuer robuster, hochdurchsatzfähiger Bildverarbeitungsmodulare für Untersuchungen am Zebrabärbling (Kapitel 3).
- Die Implementierung der Algorithmen zur skalierbaren Anwendung auf Hochdurchsatz-Untersuchungen (Kapitel 4) sowie
- die Anwendung und der Nachweis der Leistungsfähigkeit des Konzeptes und der Module auf zwei exemplarische biologische Problemstellungen (Kapitel 5).

Die wichtigsten Ergebnisse der vorliegenden Arbeit sind:

1. Entwicklung eines neuen Konzeptes zum systematischen Entwurf einer Hochdurchsatz-Prozesskette aus Versuchsplanung, Versuchsvorbereitung, Datenerfassung, Datenverarbeitung und Evaluation.

2. Ableitung einer Entscheidungslogik für die Auslegung einer geeigneten Hochdurchsatz-Datenakquise.
3. Herleitung neuer mathematischer Beschreibungen zur Bewertung von Versuchsauslegungen von Hochdurchsatzversuchen, die speziell auf die Anforderungen zur Automatisierbarkeit des Gesamtprozesses eingehen.
4. Entwicklung einer neuen Methode zur Minimierung des Einflusses von Störgrößen auf die Klassifikation.
5. Ableitung einer neuen dichotomen Klassifikatorkaskade zur Evaluation von Herzschlag und Spontanbewegung von Zebrabärblingslarven für den Hochdurchsatz.
6. Entwicklung eines neuen Verfahrens zur robusten Segmentierung und Tracking von Zebrabärblingseiern.
7. Entwicklung eines neuen Verfahrens zur Extraktion von Bewegungen und Klassifikation der Bewegungsmuster von Zebrabärblingslarven.
8. Ableitung einer Methodik zur Realisierung eines automatischen Reportgenerators sowie Nachweis der Funktionalität.
9. Entwicklung der neuen Benutzeroberfläche *PIMP* zum benutzerfreundlichen Import, zur Adaption, Ausführung, Klassifikation, Visualisierung und Archivierung von Hochdurchsatz-Untersuchungen am Zebrabärbling.
10. Nachweis der Funktionalität des Konzeptes durch Versuchsentwurf und -durchführung zweier hochdurchsatzfähiger Untersuchungen.
11. Nachweis der Leistungsfähigkeit der vorgestellten Module durch Anwendung auf zwei reale Problemstellungen.
12. Implementierung der Algorithmen in die Matlab-Toolbox Gait-CAD in Form eines Plug-ins.

In Bezug auf die automatisierte Bildauswertung müssen sich weiterführende Arbeiten mit der vollständigen Automatisierung der z.T. manuell durchgeführten Vorbereitungsschritte befassen.

Eine zukünftige Weiterentwicklung erfordert, die Bildauswertung, Bildakquise sowie Versuchsdurchführung weiter zu verknüpfen. Hierbei wird der Ansatz der computergestützten Bildakquise weitergeführt zum computergestützten Experiment. Der Verlauf derartiger Experimente ist von der Reaktion der Proben abhängig. Eine solche Rückkopplung der Bildauswertung auf den Versuchsverlauf würde z.B. das Feld des Behaviorismus [158] erstmals für den Hochdurchsatz eröffnen. Weiteres Potenzial für methodische Entwicklungen bietet z.B. die dreidimensionale Registrierung von Zebrabärblingseiern nach Spontanbewegungen im Chorion. Eine solche Methode ermöglicht eine Hochdurchsatz-Untersuchung, bei der sich der Entwicklungsprozess von einzelnen Organen ständig und ohne Präparation verfolgen lässt. Weiteres Potenzial bietet sich im Bereich der Auswertung und Rückführung von Klassifikationsergebnissen. Eine Möglichkeit ist es, die Optimierung der erzielten

Güte der Klassifikation nicht nur, wie in der vorliegenden Arbeit vorgestellt, zur Merkmalsauswahl und Normierung zu verwenden, sondern auf die Auswahl aller Parameter der Verarbeitungskette auszuweiten (siehe erste Ansätze in [75]). Abb. 6.1 zeigt einen möglichen Aufbau einer solchen rückgekoppelten Parameterauswahl wie sie bei Weiterführung der hier vorgestellten Ergebnisse aussehen könnte.

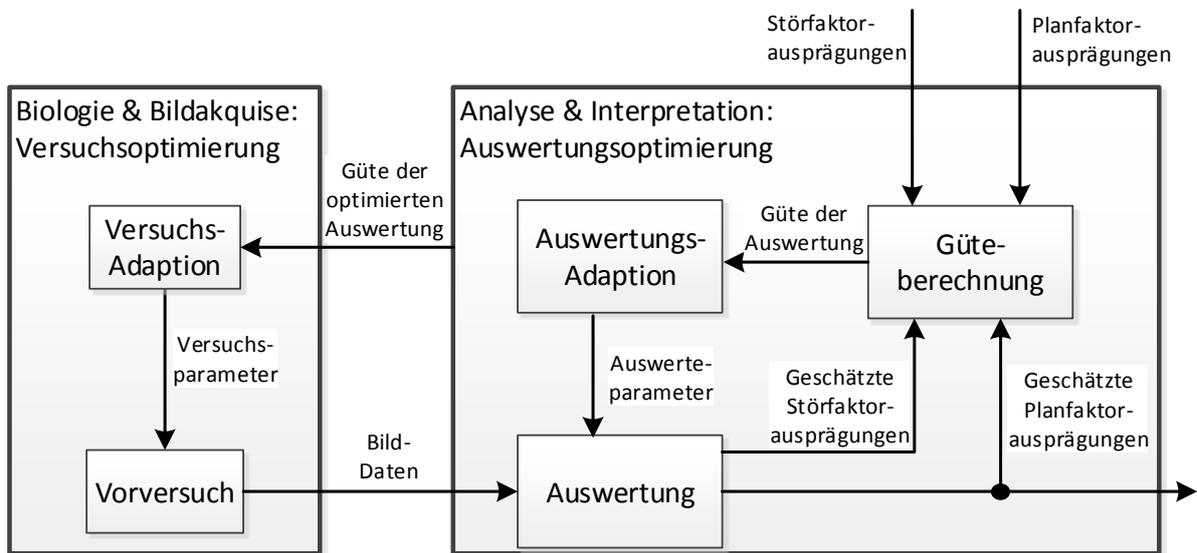


Abbildung 6.1: Ausblick: Anhand der erzielten Güte einer Hochdurchsatz-Untersuchung werden sowohl Auswerte- als auch Versuchsparameter angepasst und verbessert.

Symbolverzeichnis

Notationsvereinbarung

Skalare	kursiv: a, b, c, \dots
Vektoren	fett: $\mathbf{a}, \mathbf{b}, \mathbf{c}, \dots$
Matrizen	fett, groß: $\mathbf{A}, \mathbf{B}, \mathbf{C}, \dots$
Mengen	kalligraphisch, groß: $\mathcal{A}, \mathcal{B}, \mathcal{C}$
Konstanten, Bezeichner	Kleinbuchstaben: a, b, c

Symbole

Formelzeichen	Bedeutung
a	Schwellenwert im Bewegungsindex
B	Bildanzahl
\mathbf{BS}	Bildstrom eines Einzelversuchs
\mathbf{BS}^*	Bildstrom eines Einzelversuchs nach der Bildvorverarbeitung
c	Quantil bei der Merkmalsberechnung
CV	Messgenauigkeitsparameter
d	Hill-Koeffizient; Absolutwert der Steigung der Dosis-Wirkungs-Kurve
EC_x	Wendestelle der Dosis-Wirkungskurve
f_{BS}	Aufnahmehäufigkeit des Bildstroms
f_{Ψ}	Abtastfrequenz des Signals
H	Histogramm eines Bildes
$h(x, y)$	Filtermaske
\mathbf{I}	Bildmatrix
\mathcal{I}	Indexmenge ausgewählter Merkmale
\mathcal{I}^*	Durch Filterung adaptierte Indexmenge \mathcal{I}
\mathcal{I}^{**}	Durch einen Wrapper adaptierte Indexmenge \mathcal{I}
\mathcal{I}_{BV}	Indexmenge zur zeitlichen Unterscheidung im Bildstrom
i_c	Laufindex für Modalitäten eines segmentierten Bildstroms
I_c	Anzahl an Modalitäten eines Bildstroms

$i_{c\Psi}$	Laufindex für Modalitäten eines segmentierten Bildstroms
$I_{c\Psi}$	Anzahl an Modalitäten eines segmentierten Bildstroms
\mathbf{I}_d	Differenzbild
I_d^*	Pixelwert eines mit Rauschen überlagerten Differenzbildes
$\mathbf{I}_{d,\text{smoothed}}$	Geglättetes Differenzbild
I_{dyn}	Dynamischer Schwellenwert bei der Merkmalsberechnung
\mathbf{I}_E	Einzelne Bildmatrix innerhalb eines Bildstroms
i_f	Laufindex der Frames einer Bildsequenz
$i_{f\Psi}$	Laufindex der Frames einer Bildsequenz im segmentierten Bildstrom
i_h	Laufindex im Histogramm
I_{hoch}	Oberer Schwellenwert zur Bildnormalisierung
$I_{i_x i_y}$	Pixel einer Bildmatrix
$I_{i_x i_y}^*$	Pixelwerte nach Bildnormalisierung
$\mathbf{I}_{\text{kante}}$	Binäres Kantenbild
I_{max}	Maximaler Intensitätswert eines Bildes
i_p	Laufindex Präparationsschritte
I_{pmr}	Indikator bei der Klassifikation des Bewegungsindex
I_{quant_c}	Mittels eines Quantils c ermittelter Schwellenwert über Pixelwerte
$I_{\text{quant}_{1-c}}$	Mittels eines Quantils $1 - c$ ermittelter Schwellenwert über Pixelwerte
I_{schwell}	Schwellenwert in einem Bild
i_s	Laufindex für Spitzen im Bewegungsindex
i_{sy}	Laufindex Planfaktoren
i_{sz}	Laufindex Störfaktoren
I_{tief}	Unterer Schwellenwert zur Bildnormalisierung
i_w	Laufindex für Wiederholungen einer Abtastung
$i_{w\Psi}$	Laufindex für Wiederholungen einer Abtastung im segmentierten Bildstrom
i_x	Laufindex für Spalten in einem Bild
I_x	Anzahl Spalten in einem Bild
$i_{x\Psi}$	Laufindex für Spalten im segmentierten Bildstrom
i_y	Laufindex für Zeilen in einem Bild
I_y	Anzahl Zeilen in einem Bild
$i_{y\Psi}$	Laufindex für Zeilen im segmentierten Bildstrom
i_z	Laufindex für Schichten in einem dreidimensionalen Bild
I_z	Anzahl Schichten in einem dreidimensionalen Bild
$i_{z\Psi}$	Laufindex für Schichten im dreidimensionalen und segmentierten Bildstrom
J	Randpunkte des Bewegungsindex
j	Laufvariable
j_m	Anzahl an Mikroskopen
j_p	Anzahl an PCs bzw. Projekten
K	Anzahl Abtastzeitpunkte

k	Laufindex Abtastzeitpunkte
M	Intervall um eine Spitze im Bewegungsindex
m_1	Anzahl der Zeitreihen
m_2	Anzahl der Merkmale
m_3	Anzahl ausgewählter Merkmale
m_{BS}	Modalität des Bildstroms
m_d	Anzahl aggregierter Merkmale
M_{\max}	Längste Halbachse einer umschließenden Ellipse
M_{\min}	Kürzeste Halbachse einer umschließenden Ellipse
MSR	Minimum-Signifikanz-Verhältnis
$m_{y,i_{sy}}$	Anzahl an Klassen des i_{sy} -ten Planfaktors
$m_{z,i_{sz}}$	Anzahl an Klassen des i_{sz} -ten Störfaktors
m_{Ψ}	Modalität des Nutzsignals
n	Laufindex Versuchseinheiten
N	Anzahl Einzelversuche in einer Hochdurchsatz-Untersuchung
\mathbb{N}	Raum der natürlichen Zahlen
n_{C^+}	Anzahl an Positiv-Kontrollen
n_{C^-}	Anzahl an Negativ-Kontrollen
n_F	Anzahl an Bildsequenzen eines Bildstroms
n_l	Anzahl an Wiederholungsdurchläufen innerhalb einer Platte
n_p	Anzahl an Präparationsschritten zur Realisierung eines Versuchs
n_s	Anzahl an präparierten Versuchseinheiten
n_w	Anzahl an Näpfchen (Wells)
n_z	Anzahl an Fokusebenen
O	Optimierungskriterium zur Normalisierung
\mathbf{p}	Vektor aller Versuchsparameter
\mathbb{P}	Sensor- oder Pixelwerte
\mathbf{p}^*	Durch Filterung adaptierte Versuchsparameter \mathbf{p}
\mathbf{p}^{**}	Durch einen Wrapper adaptierte Versuchsparameter \mathbf{p}
\mathbf{p}_{BS}	Versuchsparameter des Bildstroms
\mathbf{p}_{fn}	Versuchsparameter der Bildstrom-Vorverarbeitung
$\mathbf{p}_{S\Psi}$	Versuchsparameter der Klassifikation mit $\mathbf{p}_{S\Psi} = [\mathbf{p}_{S\Psi_1}, \mathbf{p}_{S\Psi_2}, \mathcal{I}, \mathbf{p}_{S\Psi_4}, \mathbf{p}_{S\Psi_5}]^T$
\mathbf{p}_y	Versuchsparameter der Planfaktoren
\mathbf{p}_z	Versuchsparameter der Störfaktoren
\mathbf{p}_{Ψ}	Versuchsparameter des Nutzsignals
q	Parameter zur Zentrierung bei Normalisierung
$\mathbb{R}, \mathbb{R}^s, \dots$	Raum der reellen Zahlen
\mathbf{RV}	Reale Werte des Einzelversuchs
\mathbb{RV}	Raum welcher den realen Einzelversuch beschreibt
r_F	Anzahl an Frames einer Bildsequenz
$r_{F\Psi}$	Anzahl an Frames einer Bildsequenz im segmentierten Bildstrom

r_w	Anzahl an Wiederholungen einer Abtastung
$r_{w\Psi}$	Anzahl an Wiederholungen einer Abtastung im segmentierten Bildstrom
SF	Signal-Fenster
SHV	Signal-Hintergrund-Verhältnis
SNR	Signal-Rausch-Verhältnis
SRV_1	Abbildung bei der Bildakquise
SRV_2	Abbildung zur Bildstrom-Vorverarbeitung
SRV_3	Abbildung bei der Segmentierung
s_y	Anzahl Planfaktoren
S_y	Gesamt-Abbildung zur Schätzung der Klassenzugehörigkeit des realen Versuchs bezüglich der Planfaktoren
s_z	Anzahl Störfaktoren
S_z	Gesamt-Abbildung zur Schätzung der Klassenzugehörigkeit des realen Versuchs bezüglich der Störfaktoren
S_{Ψ_1}	Abbildung zur Bestimmung von Nutzsingalzeitreihen
S_{Ψ_2}	Abbildung zur Bestimmung von Nutzsingalmerkmalen
S_{Ψ_3}	Abbildung zur Merkmalsauswahl
S_{Ψ_4}	Abbildung zur Merkmalsaggregation
$S_{\Psi_{5y}}$	Abbildung zur Entscheidungsfindung bezüglich der Planfaktoren
$S_{\Psi_{5z}}$	Abbildung zur Entscheidungsfindung bezüglich der Störfaktoren
t_{aq}	Erforderliche Zeit für die Akquise einer Standard-Platte mit 96 Nöpfchen
t_{BS}	Anzahl an Frames des Bildstroms
t_{hpf}	Alter der Zebrabärblingslarven in einer Hochdurchsatz-Untersuchung
t_p	Erforderliche Zeit für eine Bewegung des automatischen Mikroskops
t_p	Erforderliche Zeit für einen Präparationsschritt einer Probe in einer Hochdurchsatz-Untersuchung
t_{RV}	Erforderliche Zeit für die Akquise einer Standard-Platte mit 96 Nöpfchen
t_{Ψ}	Aufnahmedauer bei der Bildakquise
\mathbf{u}_{BS}	Indizes, die den Auftrittsort des Nutzsingals innerhalb des Bildstroms beschreiben mit $\mathbf{u}_{BS} = [x_{BS_1}, x_{BS_2}, y_{BS_1}, y_{BS_2}, z_{BS_1}, z_{BS_2}]^T$
\mathbf{u}_{Ψ}	Indizes, die den Auftrittsort des Nutzsingals innerhalb des realen Versuchs beschreiben mit $\mathbf{u}_{\Psi} = [x_{\Psi_1}, x_{\Psi_2}, y_{\Psi_1}, y_{\Psi_2}, z_{\Psi_1}, z_{\Psi_2}]^T$
v	Parameter der Verteilung (z.B. Standardverteilung) zur Normalisierung
VK	Variationskoeffizient
\mathbf{W}	Matrix der aggregierten Merkmale

x	Merkmal (allgemein)
\mathbf{X}	Matrix der Merkmale (N Zeilen, s Spalten)
x^*	Hilfsvariable bei der Merkmalsberechnung
x_{off}	Konstante bei der Merkmalsberechnung
x_I	Abszissenachse eines Bildes
y	Planfaktor
\hat{y}	Geschätzte Klasse eines Planfaktors
$\hat{\mathbf{y}}$	Geschätzte Klassenzugehörigkeit der Planfaktoren
\mathbf{Y}	Matrix der Planfaktoren
\mathbf{y}_c	Erwarteter Effekt nach Regression
y_I	Ordinatenachse eines Bildes
\mathbf{Y}_{ZR}	Zeitreihe von Merkmalswerten
z	Störfaktor
\hat{z}	Geschätzte Klassenzugehörigkeit der Störfaktoren
\mathbf{Z}	Matrix der Störfaktoren
Z'	Parameter zur Quantifizierung der Signifikanz einer Hochdurchsatz-Untersuchung
Z''	Parameter zur Quantifizierung der Signifikanz eines Einzelversuchs einer Hochdurchsatz-Untersuchung
\mathbf{z}_c	Konzentrationen einer Untersuchung
\mathbf{z}_{RV}	Ersatzgröße weiterer Faktoren einer Hochdurchsatz-Untersuchung
α	Parameter zur Wichtung zwischen Plan- und Störfaktoren bei der Optimierung
δ	Bildzeile
μ_{max}	Maximaler Erwartungswert einer Versuchsreihe
μ_{min}	Minimaler Erwartungswert einer Versuchsreihe
ξ_y	Erfolgsrate bei Schätzung der Planfaktoren
$\overline{\xi_y}$	Klassifikationsgüte bei Schätzung der Planfaktoren
$\overline{\xi_{y, \text{ziel}}}$	Optimale Klassifikationsgüte für ξ_y
ξ_z	Erfolgsrate bei Schätzung der Störfaktoren
$\overline{\xi_z}$	Klassifikationsgüte bei Schätzung der Störfaktoren
$\overline{\xi_{z, \text{ziel}}}$	Optimale Klassifikationsgüte für ξ_z
σ	Standardabweichung der Messwerte eines Versuchs
σ_d	Standardabweichung der Differenz von Versuchen
$\sigma_{\text{max/min}}$	Versuch mit größter/kleinster Standardabweichung der Messwerte
ϕ	Farbtiefe eines Bildes
Ψ	Segmentierter Bildstrom

Literaturverzeichnis

- [1] ACKERMAN, F.; MASSEY, R.; MINISTERRÅD, N.; RÅD, N.: *The true costs of REACH*. Nordic Council of Ministers, 2004.
- [2] ADAMS, R.; BISCHOF, L.: Seeded region growing. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 16 (1994) 6, S. 641–647.
- [3] AHRENS, H.; LÄUTER, J.: *Mehrdimensionale Varianzanalyse: Hypothesenprüfung, Dimensionserniedrigung, Diskrimination bei multivariaten Beobachtungen*. Berlin: Akademie-Verlag, 1974.
- [4] ALSHUT, R.: *Entwicklung eines Bildverarbeitungsalgorithmus zur automatisierten Evaluierung eines Hochdurchsatzexperimentes mit toxikologisch behandelten Modellorganismen*. Diplomarbeit, Karlsruher Institut für Technologie (KIT), 2008.
- [5] ALSHUT, R.; LEGRADI, J.; LIEBEL, U.; YANG, L.; VAN WEZEL, J.; STRÄHLE, U.; MIKUT, R.; REISCHL, M.: Methods for automated high-throughput toxicity testing using zebrafish embryos. *Lecture Notes in Artificial Intelligence* 6359 (2010), S. 219–226.
- [6] ALSHUT, R.; LEGRADI, J.; MIKUT, R.; STRÄHLE, U.; REISCHL, M.: Robust identification of coagulated zebrafish eggs using image processing and classification techniques. In: *Proc., 19. Workshop Computational Intelligence*, S. 9–21, 2009.
- [7] ALSHUT, R.; MIKUT, R.; LEGRADI, J.; LIEBEL, U.; STRÄHLE, U.; BRETTHAUER, G.; REISCHL, M.: Automatische Klassifikation von Bildzeitreihen für toxikologische Hochdurchsatz-Untersuchungen. *at-Automatisierungstechnik* 59(5) (2011), S. 259–268.
- [8] AQUINO, D.; SCHÖNLE, A.; GEISLER, C.; v MIDDENDORFF, C.; WURM, C.; OKAMURA, Y.; LANG, T.; HELL, S.; EGNER, A.: Two-color nanoscopy of three-dimensional volumes by 4Pi detection of stochastically switched fluorophores. *Nature Methods* 8 (2011) 4, S. 353–359.
- [9] ASSAY GUIDANCE MANUAL COMMITTEE MARYLAND: Assay guidance manual version. In: *Eli Lilly and Company and NIH Chemical Genomics Center, Bethesda*, 2005.

- [10] BAKER, M.: Screening: the age of fishes. *Nature Methods* 8 (2011) 1, S. 47–51.
- [11] BANDEMER, H.: *Theorie und Anwendung der optimalen Versuchsplanung*, Bd. 1. Akademie-Verlag, 1977.
- [12] BEST, J. D.; ALDERTON, W. K.: Zebrafish: An in vivo model for the study of neurological diseases. *Neuropsychiatric Disease and Treatment* 4 (2008) 3, S. 567.
- [13] BEST, J. D.; BERGHMANS, S.; HUNT, J.; CLARKE, S. C.; FLEMING, A.; GOLDSMITH, P.; ROACH, A. G.: Non-associative learning in larval zebrafish. *Neuropsychopharmacology* 33 (2008), S. 1206–1215.
- [14] BHAT, S.; LIEBLING, M.: Cardiac tissue and erythrocyte separation in bright-field microscopy images of the embryonic zebrafish heart for motion estimation. In: *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, S. 746–749, IEEE, 2009.
- [15] BHAT, S.; OHN, J.; LIEBLING, M.: Motion-based structure separation for label-free high-speed 3-D cardiac microscopy. *Image Processing, IEEE Transactions on* 21 (2012) 8, S. 3638–3647.
- [16] BIOMETRICA, U.: Complex object parametric analyzer and sorter (COPAS). In: <http://www.unionbio.com>, 2008.
- [17] BLACKBURN, J. S.; LIU, S.; RAIMONDI, A. R.; IGNATIUS, M. S.; SALTHOUSE, C. D.; LANGENAU, D. M.: High-throughput imaging of adult fluorescent zebrafish with an LED fluorescence microscope. *Nature Protocols* 6 (2011) 2, S. 229–241.
- [18] BLEICHER, K. H.; BOHM, H.-J.; MULLER, K.; ALANINE, A. I.: Hit and lead generation: beyond high-throughput screening. *Nature Reviews Drug Discovery* 2 (2003) 5, S. 369–378.
- [19] BLOW, N.: High-throughput screening: designer screens. *Nature Methods* 6 (2009) 1, S. 105–108.
- [20] BOWMAN, T. V.; ZON, L. I.: Swimming into the future of drug discovery: In vivo chemical screens in zebrafish. *ACS Chemical Biology* 5 (2010) 2, S. 159–161.
- [21] BRAUNBECK, T.; BÖTTCHER, M.; HOLLERT, H.; KOSMEHL, T.; LAMMER, E.; LEIST, E.; RUDOLF, M.; SEITZ, N.: Towards an alternative for the acute fish LC50 test in chemical assessment: the fish embryo toxicity test goes multi-species – an update. *ALTEX - Alternativen zu Tierexperimenten* 22(2) (2005), S. 87–102.
- [22] BRAUNBECK, T.; HINTON, D. E.; STREIT, B.: *Fish ecotoxicology*. Birkhäuser, 1998.

- [23] BROACH, J. R.; THORNER, J.: High-throughput screening for drug discovery. *Nature* 384 (1996) 6604 Suppl, S. 14–16.
- [24] VAN DEN BULCK, K.; HILL, A.; MESENS, N.; DIEKMAN, H.; DE SCHAEFDRIJVER, L.; LAMMENS, L.: Zebrafish developmental toxicity assay: A fishy solution to reproductive toxicity screening, or just a red herring? *Reproductive Toxicology* 32 (2011) 2, S. 213–219.
- [25] BURGER, W.; BURGE, M. J.: *Digitale Bildverarbeitung*. Springer, 2006.
- [26] BURMEISTER, O.; REISCHL, M.; BRETTHAUER, G.; MIKUT, R.: Data-Mining-Analysen mit der MATLAB-Toolbox Gait-CAD. *at-Automatisierungstechnik* 56(7) (2008), S. 381–389.
- [27] BURT, P.; ADELSON, E.: The laplacian pyramid as a compact image code. *Communications, IEEE Transactions on* 31 (1983) 4, S. 532–540.
- [28] CACHAT, J.; STEWART, A.; UTTERBACK, E.; HART, P.; GAIKWAD, S.; WONG, K.; KYZAR, E.; WU, N.; KALUEFF, A. V.: Three-dimensional neurophenotyping of adult zebrafish behavior. *PLoS One* 6 (2011) 3.
- [29] CANNY, J.: Finding edges and lines in images. Techn. Ber., Massachusetts Institute of Technology, 1983.
- [30] CAO, Y.; SEMANCHIK, N.; LEE, S. H.; SOMLO, S.; BARBANO, P. E.; COIFMAN, R.; SUN, Z.: Chemical modifier screen identifies HDAC inhibitors as suppressors of PKD models. *Proceedings of the National Academy of Sciences* 106 (2009) 51, S. 21819–21824.
- [31] CAROLYN, K.; BALLARD, D.; SKLANSKY, J.: Finding circles by an array of accumulators. *Commun. ACM* 18 (1975) 2, S. 120–122.
- [32] CARPENTER, A. E.: Image-based chemical screening. *Nature Chemical Biology* 3(8) (2007), S. 461.
- [33] CARRADICE, D.; LIESCHKE, G.: Zebrafish in hematology: sushi or science? *Blood* 111 (2008) 7, S. 3331–3342.
- [34] CARREIRA-PERPINÁN, M. A.: A review of dimension reduction techniques. *Department of Computer Science. University of Sheffield. Tech. Rep. CS-96-09* 9 (1997), S. 1–69.
- [35] CARVALHO, R.; DE SONNEVILLE, J.; STOCKHAMMER, O.; SAVAGE, N.; VENEMAN, W.; OTTENHOFF, T.; DIRKS, R.; MEIJER, A.; SPAINK, H.: A high-throughput screen for tuberculosis progression. *PLoS One* 6 (2011) 2.

- [36] CHAN, P.; LIN, C.; CHENG, S.: Noninvasive technique for measurement of heart-beat regularity in zebrafish (*danio rerio*) embryos. *BMC Biotechnology* 9 (2009) 1, S. 11.
- [37] CLELAND, W.: The kinetics of enzyme-catalyzed reactions with two or more substrates or products: I. Nomenclature and rate equations. *Biochimica et Biophysica Acta (BBA)-Specialized Section on Enzymological Subjects* 67 (1963), S. 104–137.
- [38] COLOWICK, S. P.; KAPLAN, N. O.; ABELSON, J. N.; SIMON, M. I.: *Methods in enzymology: Measuring biological responses with automated microscopy*. Academic Press, 2006.
- [39] CREMERS, D.; ROUSSON, M.; DERICHE, R.: A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. *International Journal of Computer Vision* 72 (2007) 2, S. 195–215.
- [40] CRISTIANNI, N.; SHAW-TAYLOR, J.: *Support vector machines*. Cambridge, United Kingdom: Cambridge University Press, 2000.
- [41] D'ALENCON, C.; PENA, O.; WITTMANN, C.; GALLARDO, V.; JONES, R.; LOOSLI, F.; LIEBEL, U.; GRABHER, C.; ALLENDE, M.: A high-throughput chemically induced inflammation assay in zebrafish. *BMC Biology* 8 (2010) 1, S. 151.
- [42] EASTWOOD, B. J.; FARMEN, M. W.; IVERSEN, P. W.; CRAFT, T. J.; SMALLWOOD, J. K.; GARBISON, K. E.; DELAPP, N. W.; SMITH, G. F.: The minimum significant ratio: a statistical parameter to characterize the reproducibility of potency estimates from concentration-response assays and estimation by replicate-experiment studies. *Journal of Biomolecular Screening* 11 (2006) 3, S. 253–261.
- [43] EATON, R. C.; FARLEY, R. D.: Spawning cycle and egg production of zebrafish. *Copeia* (1974), S. 195–204.
- [44] EFRON, B.; TIBSHIRANI, R.: Cross-validation and the bootstrap: estimating the error rate of a prediction rule. Techn. Ber. TR-477, Dept. of Statistics, Stanford University, 1995.
- [45] EIMON, P.; RUBINSTEIN, A.: The use of in vivo zebrafish assays in drug toxicity screening. *Expert Opinion on Drug Metabolism and Toxicology* 5 (2009) 4, S. 393–401.
- [46] EISEN, J. S.: Zebrafish make a big splash. *Cell* 87 (1996) 6, S. 969–977.
- [47] FORSTER, B.; VAN DE VILLE, D.; BERENT, J.; SAGE, D.; UNSER, M.: Complex wavelets for extended depth-of-field: A new method for the fusion of multichannel microscopy images. *Microscopy Research and technique* 65 (2004) 1-2, S. 33–42.

- [48] GAD, S.: *Drug discovery handbook*, Bd. 10. Wiley Online Library, 2005.
- [49] GEHRIG, J.; REISCHL, M.; KALMAR, E.; FERG, M.; HADZHIEV, Y.; ZAUCKER, A.; SONG, C.; SCHINDLER, S.; LIEBEL, U.; MÜLLER, F.: Automated high throughput mapping of promoter-enhancer interactions in zebrafish embryos. *Nature Methods* 6 (2009) 12, S. 911–916.
- [50] GOLDSMITH, P.: Zebrafish as a pharmacological tool: The how, why and when. *Current opinion in pharmacology* 4 (2004) 5, S. 504–512.
- [51] GONZALEZ, R. C.; WOODS, R. E.: *Digital image processing*. Pearson Prentice Hall, 3. Aufl., 2008.
- [52] GREEN, J.; COLLINS, C.; KYZAR, E. J.; PHAM, M.; ROTH, A.; GAIKWAD, S.; CACHAT, J.; STEWART, A. M.; LANDSMAN, S.; GRIECO, F.; ET AL.: Automated high-throughput neurophenotyping of zebrafish social behavior. *Journal of Neuroscience Methods* 210 (2012) 2, S. 266–271.
- [53] HALL, M.: *Correlation-based feature selection for machine learning*. Dissertation, University of Waikato, Hamilton, New Zealand, 1999.
- [54] HARALICK, R. M.; SHAPIRO, L. G.: *Computer and robot vision*. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 1992.
- [55] HATHAWAY, R.; BEZDEK, J. C.; TUCKER, W.: An improved convergence theory for the fuzzy ISODATA clustering algorithms. In: *Analysis of Fuzzy Information* (BEZDEK, J., Hg.), Bd. 3, S. 123–132, Boca Rota: CRC Press, 1987.
- [56] HELL, S. W.; WICHMANN, J.: Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics letters* 19 (1994) 11, S. 780–782.
- [57] HENN, K.; BRAUNBECK, T.: Dechoriation as a tool to improve the fish embryo toxicity test (FET) with the zebrafish (*Danio rerio*). *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* 153 (2011) 1, S. 91–98.
- [58] HILL, A. J.; TERAOKA, H.; HEIDEMAN, W.; PETERSON, R. E.: Zebrafish as a model vertebrate for investigating chemical toxicity. *Toxicological Sciences* 86 (2005) 1, S. 6–19.
- [59] HOGG, R. C.; BANDELIER, F.; BENOIT, A.; DOSCH, R.; BERTRAND, D.: An automated system for intracellular and intranuclear injection. *Journal of Neuroscience Methods* 169 (2008) 1, S. 65–75.
- [60] HOUGH, P. V. C.: Method and means for recognizing complex patterns. 1962.

- [61] HUANG, L.; WANG, M.: Image thresholding by minimizing the measures of fuzziness. *Pattern Recognition* 28 (1995) 1, S. 41–51.
- [62] HUERTAS, A.; MEDIONI, G.: Detection of intensity changes with subpixel accuracy using Laplacian-Gaussian Masks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8(5) (1986), S. 651–664.
- [63] HUISKEN, J.; SWOGER, J.; DEL BENE, F.; WITTBRODT, J.; STELZER, E.: Optical sectioning deep inside live embryos by selective plane illumination microscopy. *Science* 305 (2004) 5686, S. 1007–1009.
- [64] HÜSER, J.: *High-throughput screening in drug discovery*. Wiley-VCH, 2006.
- [65] INGHAM, P.: Zebrafish genetics and its implications for understanding vertebrate development. *Human Molecular Genetics* 6 (1997), S. 1755–1760.
- [66] INGLESE, J.; JOHNSON, R. L.; SIMEONOV, A.; XIA, M.; ZHENG, W.; AUSTIN, C. P.; AULD, D. S.: High-throughput screening assays for the identification of chemical probes. *Nature Chemical Biology* 3 (2007) 8, S. 466–479.
- [67] IRONS, T.; MACPHAIL, R.; HUNTER, D.; PADILLA, S.: Acute neuroactive drug exposures alter locomotor activity in larval zebrafish. *Neurotoxicology and Teratology* 32 (2010) 1, S. 84–90.
- [68] ISO: *ISO 15088:2007 Water quality –determination of the acute toxicity of waste water to zebrafish eggs (Danio rerio)*. 2007.
- [69] IVERSEN, P. W.; EASTWOOD, B. J.; SITTAMPALAM, G. S.; COX, K. L.: A comparison of assay performance measures in screening assays: Signal window, z' factor, and assay variability ratio. *Journal of Biomolecular Screening* 11 (2006) 3, S. 247–252.
- [70] JÄHNE, B.: *Digitale Bildverarbeitung*. Springer, 6., überarb. und erw. Aufl. Aufl., 2005.
- [71] JAIN, A. K.; DUIN, R. P. W.; MAO, J.: Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(1) (2000), S. 4–36.
- [72] KAUTSKY, J.; FLUSSER, J.; ZITOVÁ, B.; ŠIMBEROVÁ, S.: A new wavelet-based measure of image focus. *Pattern Recognition Letters* 23 (2002) 14, S. 1785–1794.
- [73] KEITER, S.; PEDDINGHAUS, S.; HOLLERT, H.; FEILER, U.; REIFFERSCHIED, G.; V. D. GOLTZ, B.; BRAUNBECK, T.; HAFNER, C.; OTTERMANN, R.; HAMMERSWIRTZ, M.; ET AL.: DanTox–ein BMBF-Verbundprojekt zur Ermittlung spezifi-

- scher Toxizität und molekularer Wirkungsmechanismen sedimentgebundener Umweltschadstoffe mit dem Zebrafisch (*Danio rerio*). *Umweltwissenschaften und Schadstoff-Forschung* 22 (2010) 2, S. 94–98.
- [74] KELLER, P.; SCHMIDT, A.; WITTBRODT, J.; STELZER, E.: Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. *Science* 322 (2008) 5904, S. 1065–1069.
- [75] KHAN, A.; REISCHL, M.; SCHWEITZER, B.; WEISS, C.; MIKUT, R.: Automatic tuning of image segmentation routines by means of fuzzy feature evaluation. *Advances in Intelligent Systems and Computing* 190 (2013) 5, S. 459–467.
- [76] KIMMEL, C.; BALLARD, W. W.; KIMMEL, S. R.; ULLMANN, B.; SCHILLING, T. F.: Stages of embryonic development of the zebrafish. *Developmental Dynamics* 203 (1995) 3, S. 253–310.
- [77] KLEIN, B.: *Versuchsplanung-DoE: Einführung in die Taguchi/Shainin-Methodik*. Oldenbourg Wissenschaftsverlag, 2007.
- [78] KOHAVI, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proc., International Joint Conference on Artificial Intelligence*, 1995.
- [79] KOHAVI, R.; JOHN, G. H.: Wrappers for feature subset selection. *Artificial Intelligence* 97 (1-2) (1997), S. 273–324.
- [80] KOKEL, D.; BRYAN, J.; LAGGNER, C.; WHITE, R.; CHEUNG, C. Y. J.; MATEUS, R.; HEALEY, D.; KIM, S.; WERDICH, A. A.; HAGGARTY, S. J.; MACRAE, C. A.; SHOICHET, B.; PETERSON, R. T.: Rapid behavior-based identification of neuroactive small molecules in the zebrafish. *Nature Chemical Biology* 6 (2010) 3, S. 231–237.
- [81] KOKEL, D.; DUNN, T. W.; AHRENS, M. B.; ALSHUT, R.; CHEUNG, C. Y. J.; SAINT-AMANT, L.; BRUNI, G.; MATEUS, R.; VAN HAM, T. J.; SHIRAKI, T.; ET AL.: Identification of nonvisual photomotor response cells in the vertebrate hindbrain. *The Journal of Neuroscience* 33 (2013) 9, S. 3834–3843.
- [82] KOKEL, D.; RENNEKAMP, A.; SHAH, A.; LIEBEL, U.; PETERSON, R.: Behavioral barcoding in the cloud: embracing data-intensive digital phenotyping in neuropharmacology. *Trends in Biotechnology* (2012).
- [83] LAHL, U.; HAWXWELL, K. A.: REACH-The new european chemicals law. *Environmental science & technology* 40 (2006) 23, S. 7115–7121.

- [84] LAMMER, E.: *Refinement of the fish embryo toxicity test (FET) with zebrafish (*Danio rerio*): Is it a real replacement of the acute fish toxicity test?* Dissertation, Combined faculty for mathematics and natural sciences, Ruprecht-Karl-University Heidelberg, 2009.
- [85] LAMMER, E.; CARR, G.; WENDLER, K.; RAWLINGS, J.; BELANGER, S.; BRAUNBECK, T.: Is the fish embryo toxicity test (FET) with the zebrafish (*Danio rerio*) a potential alternative for the fish acute toxicity test? *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* 149 (2009), S. 196–209.
- [86] LANGHEINRICH, U.; HENNEN, E.; STOTT, G.; VACUN, G.: Zebrafish as a model organism for the identification and characterization of drugs and genes affecting p53 signaling. *Current Biology* 12 (2002) 23, S. 2023–2028.
- [87] LAWSON, N. D.; WOLFE, S. A.: Forward and reverse genetic approaches for the analysis of vertebrate development in the zebrafish. *Developmental cell* 21 (2011) 1, S. 48–64.
- [88] LEHMANN, T.; MEYER ZU BEXTEN, E.: *Handbuch der Medizinischen Informatik*. München: Hanser-Verlag, 2002.
- [89] LESSMAN, C. A.: The developing zebrafish (*Danio rerio*): A vertebrate model for high-throughput screening of chemical libraries. *Birth Defects Research Part C: Embryo Today: Reviews* 93 (2011) 3, S. 268–280.
- [90] LIEBEL, U.; STARKUVIENE, V.; ERFLE, H.; SIMPSON, J.; POUSTKA, A.; WIEMANN, S.; PEPPERKOK, R.: A microscope-based screening platform for large-scale functional protein analysis in intact cells. *FEBS Letters* 554 (2003) 3, S. 394–398.
- [91] LIESCHKE, G. J.; CURRIE, P. D.: Animal models of human disease: Zebrafish swim into view. *Nature Reviews Genetics* 8 (2007) 5, S. 353–367.
- [92] LIN, S.: Transgenic zebrafish. In: *Developmental biology protocols: Volume II* (TUAN, R. S.; LO, C. W., Hg.), Bd. 136 von *Methods in Molecular Biology*, S. 375–383, Humana Press, 2000.
- [93] LIU, R.; LIN, S.; RALLO, R.; ZHAO, Y.; DAMOISEAUX, R.; XIA, T.; LIN, S.; NEL, A.; COHEN, Y.: Automated phenotype recognition for zebrafish embryo based in vivo high throughput toxicity screening of engineered nano-materials. *PloS One* 7 (2012) 4.
- [94] LIU, T.: A quantitative zebrafish phenotyping tool for developmental biology and disease modeling. *Signal Processing Magazine, IEEE* 24 (2007) 1, S. 126–129.

- [95] LIU, T.; NIE, J.; LI, G.; GUO, L.; WONG, S.: ZFIQ: Zebrafish image quantitor. In: *Proc., IEEE/NIH Life Science Systems and Applications Workshop*, S. 59–62, 2007.
- [96] LIU, T.; NIE, J.; LI, G.; GUO, L.; WONG, S. T. C.: ZFIQ: a software package for zebrafish biology. *Bioinformatics* 24 (2008) 3, S. 438–439.
- [97] LOVE, D. R.; PICHLER, F. B.; DODD, A.; COPP, B. R.; GREENWOOD, D. R.: Technology for high-throughput screens: the present and future using zebrafish. *Current Opinion in Biotechnology* 15 (2004) 6, S. 564–571.
- [98] LU, Z.; CHEN, P. C.; NAM, J.; GE, R.; LIN, W.: A micromanipulation system with dynamic force-feedback for automatic batch microinjection. *Journal of Micro-mechanics and Microengineering* 17 (2007), S. 314–321.
- [99] MACRAE, C.: Zebrafish in toxicity testing. *Annual Review of Pharmacology and Toxicology* 52 (2012) 1.
- [100] MAHALANOBIS, P. C.: On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta)* 2 (1936), S. 49–55.
- [101] MALO, N.; HANLEY, J. A.; CERQUOZZI, S.; PELLETIER, J.; NADON, R.: Statistical practice in high-throughput screening data analysis. *Nature Biotechnology* 24 (2006) 2, S. 167–175.
- [102] MANDRELL, D.; TRUONG, L.; JEPHSON, C.; SARKER, M. R.; MOORE, A.; LANG, C.; SIMONICH, M. T.; TANGUAY, R. L.: Automated zebrafish chorion removal and single embryo placement. *Journal of Laboratory Automation* 17 (2012) 1, S. 66–74.
- [103] MARCATO, D.; ALSHUT, R.; BREITWIESER, H.; MIKUT, R.; STRÄHLE, U.; PYLATIUK, C.; PERAVALI, R.: An automated and high-throughput Photomotor Response platform for chemical screens. In: *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, S. 7728–7731, IEEE, 2015.
- [104] MARUYAMA, N.; SHIBATA, Y.; SEKINE, T.: *Microtiter plate*. US Patent, 4,735,778, 1988.
- [105] MAYER, F. L.; ELLERSIECK, M. R.; FISH, U. S.: *Manual of acute toxicity: Interpretation and data base for 410 chemicals and 66 species of freshwater animals*. US Dept. of the Interior, Fish and Wildlife Service, 1986.
- [106] MCGRATH, P.; LI, C.: Zebrafish: a predictive model for assessing drug-induced toxicity. *Drug Discovery Today* 13 (2008) 9-10, S. 394–401.

- [107] METSCHER, B.; AHLBERG, P.: Zebrafish in context: uses of a laboratory model in comparative studies. *Developmental Biology* 210 (1999), S. 1–14.
- [108] MIKULA, K.; PEYRIÉRAS, N.; REMEŠÍKOVÁ, M.; SARTI, A.: 3D embryogenesis image segmentation by the generalized subjective surface method using the finite volume technique. *Finite Volumes for Complex Applications V: Problems and Perspectives* (2008), S. 585–592.
- [109] MIKULA, K.; PEYRIÉRAS, N.; REMEŠÍKOVÁ, M.; SMÍŠEK, M.: 4D Numerical schemes for cell image segmentation and tracking. *Finite Volumes for Complex Applications VI: Problems and Perspectives* (2011), S. 693–701.
- [110] MIKUT, R.: *Data Mining in der Medizin und Medizintechnik*. Universitätsverlag Karlsruhe, 2008.
- [111] MIKUT, R.; BURMEISTER, O.; BRAUN, S.; REISCHL, M.: The open source Matlab toolbox Gait-CAD and its application to bioelectric signal processing. In: *Proc., DGBMT-Workshop Biosignalverarbeitung, Potsdam*, S. 109–111, 2008.
- [112] MIKUT, R.; BURMEISTER, O.; GRUBE, M.; REISCHL, M.; BRETTHAUER, G.: Interaktive Auswertung von aufgezeichneten Zeitreihen für Fehlerdiagnosen und Mensch-Maschine-Interfaces. *atp - Automatisierungstechnische Praxis* 49(8) (2007), S. 30–34.
- [113] MIKUT, R.; DICKMEIS, T.; DRIEVER, W.; GEURTS, P.; HAMPRECHT, F. A.; KAUSLER, B. X.; LEDESMA-CARBAYO, M. J.; MARÉE, R.; MIKULA, K.; PANTAZIS, P.; ET AL.: Automated processing of zebrafish imaging data: A survey. *Zebrafish* 10 (2013) 3, S. 401–421.
- [114] MIKUT, R.; LOOSE, T.; BURMEISTER, O.; BRAUN, S.; REISCHL, M.: Gait-CAD. <http://sourceforge.net/projects/gait-cad/>, 2007.
- [115] MIKUT, R.; PETER, N.; MALBERG, H.; JÄKEL, J.; GRÖLL, L.; BRETTHAUER, G.; ABEL, R.; DÖDERLEIN, L.; RUPP, R.; SCHABLOWSKI, M.; GERNER, H.: *Diagnoseunterstützung für die instrumentelle Ganganalyse (Projekt GANDI)*. Forschungszentrum Karlsruhe, 2001.
- [116] MONTGOMERY, D. C.: *Design and analysis of experiments*. Wiley, 6. Aufl., 2005.
- [117] MÜLLER, W. A.; HASSEL, M.: *Entwicklung bedeutsamer Modellorganismen II: Wirbeltiere*, S. 115–170. Springer, Berlin, 2006.
- [118] NAGEL, R.: DarT: The embryo test with the Zebrafish (*Danio rerio*) –a general model in ecotoxicology and toxicology. *ALTEX - Alternativen zu Tierversimenten* 19 Suppl 1 (2002), S. 38–48.

- [119] NIBLACK, W.: *An introduction to digital image processing*. Strandberg Publishing Company, 1985.
- [120] NÜSSLEIN-VOLHARD, C.; WIESCHAUS, E.: Mutations affecting segment number and polarity in *Drosophila*. *Nature* 287 (1980), S. 795–801.
- [121] OECD: *Background paper on fish embryo toxicity assay*. 2006.
- [122] OECD: *Fish embryo toxicity (FET) test. Draft OECD guideline for the testing of chemicals*. 2006.
- [123] OHN, J.; LIEBLING, M.: In vivo, high-throughput imaging for functional characterization of the embryonic zebrafish heart. In: *Proc., IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, S. 1549–1552, IEEE, 2011.
- [124] OTSU, N.: A Threshold Selection Method from Gray-Level Histograms. *Automatica* 11 (1975), S. 285–296.
- [125] OTSU, N.: A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics* 9 (1979), S. 62–66.
- [126] PARDO-MARTIN, C.; CHANG, T.; KOO, B. K.; GILLELAND, C. L.; WASSERMAN, S. C.; YANIK, M. F.: High-throughput in vivo vertebrate screening. *Nature Methods* 7 (2010) 8, S. 634–636.
- [127] PATTON, E.; ZON, L.: The art and design of genetic screens: Zebrafish. *Nature Reviews Genetics* 2(12) (2001), S. 956–966.
- [128] PAVLIDIS, T.; LIOW, Y.: Integrating region growing and edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 12 (1990) 3, S. 225–233.
- [129] PAWLEY, J.; MASTERS, B. R.: Handbook of biological confocal microscopy. *Optical Engineering* 35 (1996) 9, S. 2765–2766.
- [130] PERAVALI, R.; GEHRIG, J.; GISELBRECHT, S.; LÜTJOHANN, D.; HADZHIEV, Y.; MÜLLER, F.; LIEBEL, U.: Automated feature detection and imaging for high-resolution screening of zebrafish embryos. *BioTechniques* 50 (2011) 5, S. 319–324.
- [131] PERKEL, J.: LIFE SCIENCE TECHNOLOGIES: Animal-free toxicology: Sometimes, in vitro is better. *Science* 335 (2012) 6072, S. 1122–1125.
- [132] PFRIEM, A.; SCHULZ, S.; PYLATIUK, C.; ALSHUT, R.; BRETTHAUER, G.: Robotersysteme für Hochdurchsatzverfahren in der Bioanalysetechnik. *at-Automatisierungstechnik* 59 (2011) 2, S. 134–140.

- [133] PIEPER, R. J.; KORPEL, A.: Image processing for extended depth of field. *Applied Optics* 22 (1983) 10, S. 1449–1453.
- [134] PRASHER, D. C.: Using GFP to see the light. *Trends in Genetics* 11 (1995) 8, S. 320–323.
- [135] PYLATIUK, C.; PFRIEM, A.; LIEBEL, U.; SCHULZ, S.; BRETTHAUER, G.: Ingenieurtechnische Besonderheiten bei der automatischen Handhabung von biologischen Organismen. *at-Automatisierungstechnik* 59 (2011) 11, S. 692–698.
- [136] PYLATIUK, C.; SANCHEZ, D.; MIKUT, R.; ALSHUT, R.; REISCHL, M.; HIRTH, S.; ROTTBAUER, W.; JUST, S.: Automatic zebrafish heartbeat detection and analysis for zebrafish embryos. *Zebrafish* 11 (2014) 4, S. 379–383.
- [137] QUINLAN, J. R.: Induction of decision trees. *Machine Learning* 1 (1986), S. 81–106.
- [138] RASCH, D.; VERDOOREN, L. R.; GOWERS, J. I.: *Planung und Auswertung von Versuchen und Erhebungen*. Lehrbuch international, München: Oldenbourg, 2. Aufl., 2007.
- [139] REISCHL, M.; ALSHUT, R.; MIKUT, R.: On robust feature extraction and classification of inhomogeneous data sets. In: *Proc., 20. Workshop Computational Intelligence*, S. 124–143, KIT Scientific Publishing, 2010.
- [140] REISCHL, M.; GRÖLL, L.; MIKUT, R.: Optimierte Klassifikation für Mehrklassenprobleme am Beispiel der Bewegungssteuerung von Handprothesen. In: *Proc., 13. Workshop Fuzzy Systeme*, S. 124–143, Forschungszentrum Karlsruhe, 2003.
- [141] REISCHL, M.; GRÖLL, L.; MIKUT, R.: Optimized classification of multiclass problems applied to EMG-control of hand prostheses. In: *Proc., IEEE International Joint Conference on Neural Networks*, S. 1473–1478, 2004.
- [142] REISCHL, M.; MIKUT, R.: Challenges of Uncertainty Propagation in Image Analysis. In: *Proceedings. 22. Workshop Computational Intelligence*, S. 55, KIT Scientific Publishing, 2014.
- [143] RIHEL, J.; PROBER, D. A.; ARVANITES, A.; LAM, K.; ZIMMERMAN, S.; JANG, S.; HAGGARTY, S. J.; KOKEL, D.; RUBIN, L. L.; PETERSON, R. T.; SCHIER, A. F.: Zebrafish behavioral profiling links drugs to biological targets and rest/wake regulation. *Science* 327 (2010) 5963, S. 348–351.
- [144] RINKWITZ, S.; MOURRAIN, P.; BECKER, T. S.: Zebrafish: An integrative system for neurogenomics and neurosciences. *Progress in Neurobiology* 93 (2011) 2, S. 231–243.

- [145] RUSS, J.: *The image processing handbook*. CRC Press, 2002.
- [146] RUXTON, GRAEME D. ; COLEGRAVE, N.: *Experimental design for the life sciences*. Oxford Univ. Press, 3. Aufl., 2011.
- [147] SABALIAUSKAS, N.; FOUTZ, C.; MEST, J.; BUDGEON, L.; SIDOR, A.; GERSHENSON, J.; JOSHI, S.; CHENG, K.: High-throughput zebrafish histology. *Methods* 39 (2006), S. 246–254.
- [148] SÁNCHEZ, D.: *Entwicklung einer Methode zur automatisierten Extraktion der Herzfrequenz aus Videosequenzen von Zebrafischlarven*. Diplomarbeit, Karlsruher Institut für Technologie (KIT), 2012.
- [149] SAUVOLA, J.; SEPPANEN, T.; HAAPAKOSKI, S.; PIETIKAINEN, M.: Adaptive document binarization. In: *Document Analysis and Recognition, 1997., Proceedings of the Fourth International Conference on*, Bd. 1, S. 147–152, IEEE, 1997.
- [150] SCHÖLKOPF, B.; SMOLA, A.; MÜLLER, K.: Nonlinear component analysis using a kernel eigenvalue problem. *Techn. Ber.* 44, Max-Planck-Institut für biologische Kybernetik, 1996.
- [151] SCHREIBER, S.; KAPOOR, T. M.; WESS, G.: *Chemical biology: From small molecules to systems biology and drug design*. Wiley-VCH, 2007.
- [152] SELDESLAGHS, I. W.; BLUST, R.; WITTERS, H. E.: Feasibility study of the zebrafish assay as an alternative method to screen for developmental toxicity and embryotoxicity using a training set of 27 compounds. *Reproductive Toxicology* 33 (2012) 2, S. 142 – 154.
- [153] SEMMLOW, J. L.: *Biosignal and biomedical image processing: MATLAB-based applications*. CRC Press, 2004.
- [154] SHANNON, C.: Communication in the presence of noise. *Proceedings of the IEEE* 86 (1998) 2, S. 447–457.
- [155] SHARIFF, A.; KANGAS, J.; COELHO, L. P.; QUINN, S.; MURPHY, R. F.: Automated image analysis for High-Content screening and analysis. *Journal of Biomolecular Screening* 15 (2010) 7, S. 726 –734.
- [156] SHIMOMURA, O.: The discovery of aequorin and green fluorescent protein. *Journal of microscopy* 217 (2005) 1, S. 3–15.
- [157] SHIMOMURA, O.; JOHNSON, F. H.; SAIGA, Y.: Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, aequorea. *Journal of Cellular and Comparative Physiology* 59 (1962) 3, S. 223–239.

- [158] SKINNER, B. F.: *Science and human behavior*. Simon and Schuster, 1953.
- [159] SPECTOR, D.; GOLDMAN, R.: *Basic methods in microscopy: Protocols and concepts from cells: A laboratory manual*. CSHL Press, 2006.
- [160] SPOMER, W.; PFRIEM, A.; ALSHUT, R.; CHRISTIAN, P.: High-throughput screening of zebrafish embryos using automated heart detection and imaging. *Journal of Laboratory Automation* (2012).
- [161] SPRAGUE, J.; BAYRAKTAROGLU, L.; CLEMENTS, D.; CONLIN, T.; FASHENA, D.; FRAZER, K.; HAENDEL, M.; HOWE, D. G.; MANI, P.; RAMACHANDRAN, S.; ET AL.: The Zebrafish Information Network: The zebrafish model organism database. *Nucleic acids research* 34 (2006) suppl 1, S. 581–585.
- [162] STARKUVIENE, V.; PEPPERKOK, R.: The potential of high-content high-throughput microscopy in drug discovery. *British journal of pharmacology* 152 (2007) 1, S. 62–71.
- [163] STEGEMAN, J. J.; GOLDSTONE, J. V.; HAHN, M. E.; DR. STEVE F. PERRY, D. M. E.: Perspectives on zebrafish as a model in environmental toxicology. In: *Zebrafish*, Bd. Volume 29, S. 367–439, Academic Press, 2010.
- [164] STEGMAIER, J.; ALSHUT, R.; REISCHL, M.; MIKUT, R.: Information fusion of image analysis, video object tracking, and data mining of biological images using the open source matlab toolbox gait-cad. *Biomedizinische Technik (Biomedical Engineering)* 57 (S1) (2012), S. 458–461.
- [165] STEINBRECHER, R.: *Bildverarbeitung in der Praxis*. Oldenbourg, 1993.
- [166] STOTZKA, R.; HARTMANN, V.; JEJKAL, T.; SUTTER, M.; VAN WEZEL, J.; HARDT, M.; GARCIA, A.; KUPSCH, R.; BOUROV, S.: Perspective of the Large Scale Data Facility (LSDF) supporting nuclear fusion applications. In: *Parallel, Distributed and Network-Based Processing (PDP), 2011 19th Euromicro International Conference on*, S. 373–379, IEEE, 2011.
- [167] STRÄHLE, U.; SCHOLZ, S.; GEISLER, R.; GREINER, P.; HOLLERT, H.; RASTEGAR, S.; SCHUMACHER, A.; SELDERSLAGHS, I.; WEISS, C.; WITTERS, H.; ET AL.: Zebrafish embryos as an alternative to animal experiments – A commentary on the definition of the onset of protected life stages in animal welfare regulations. *Reproductive Toxicology* 33 (2012) 2, S. 128–132.
- [168] TRAN, T. C.; SNEED, B.; HAIDER, J.; BLAVO, D.; WHITE, A.; AIYEJORUN, T.; BARANOWSKI, T. C.; RUBINSTEIN, A. L.; DOAN, T. N.; DINGLEDINE, R.; ET AL.: Automated, quantitative screening assay for antiangiogenic compounds using transgenic zebrafish. *Cancer research* 67 (2007) 23, S. 11386–11392.

- [169] TRAVER, D.; HERBOMEL, P.; PATTON, E. E.; MURPHEY, R. D.; YODER, J. A.; LITMAN, G. W.; CATIC, A.; AMEMIYA, C. T.; ZON, L. I.; TREDE, N. S.: The zebrafish as a model organism to study development of the immune system. *Advances in Immunology* 81 (2003), S. 253–330.
- [170] VALDECASAS, A.; MARSHALL, D.; BECERRA, J.; TERRERO, J.: On the extended depth of focus algorithms for bright field microscopy. *Micron* 32 (2001) 6, S. 559–569.
- [171] VAUGHAN, M.; VAN EGMOND, R.: The use of the zebrafish (*danio rerio*) embryo for the acute toxicity testing of surfactants, as a possible alternative to the acute fish test. *ATLA. Alternatives to laboratory animals* 38 (2010) 3, S. 231–238.
- [172] VEITH, D.; VEITH, M.: Biologie fluoreszierender Proteine: Ein Regenbogen aus dem Ozean. *Biologie in unserer Zeit* 35 (2005) 6, S. 394–404.
- [173] VOGT, A.; CHOLEWINSKI, A.; SHEN, X.; NELSON, S.; LAZO, J.; TSANG, M.; HUKRIEDE, N.: Automated image-based phenotypic analysis in zebrafish embryos. *Developmental Dynamics* 238 (2009) 3, S. 656–663.
- [174] WALKER, S. L.; ARIGA, J.; MATHIAS, J. R.; COOTHANKANDASWAMY, V.; XIE, X.; DISTEL, M.; KÖSTER, R. W.; PARSONS, M. J.; BHALLA, K. N.; SAXENA, M. T.; ET AL.: Automated reporter quantification in vivo: High-throughput screening method for reporter-based assays in zebrafish. *PloS One* 7 (2012) 1.
- [175] WANG, W.; LIU, X.; SUN, Y.: Contact detection in microrobotic manipulation. *The International Journal of Robotics Research* 26 (2007) 8, S. 821–828.
- [176] WATERS, J. C.: Accuracy and precision in quantitative fluorescence microscopy. *The Journal of Cell Biology* 185 (2009) 7, S. 1135–1148.
- [177] WESTERFIELD, M.: *The zebrafish book: a guide for the laboratory use of zebrafish (Brachydanio rerio)*. University of Oregon Press Eugene, OR, 1993.
- [178] WILLIAMS, E. S.; PANKO, J.; PAUSTENBACH, D. J.: The european union's REACH regulation: A review of its history and requirements. *Crit Rev Toxicol* 39 (2009) 7, S. 553–575.
- [179] WITTMANN, C.; REISCHL, M.; SHAH, A.; MIKUT, R.; LIEBEL, U.; GRABHER, C.: Facilitating drug discovery: An automated high-content inflammation assay in zebrafish. *Journal of Visualized Experiments: JoVE* 65 (2012).
- [180] WÖLCKE, J.; ULLMANN, D.: Miniaturized HTS technologies - uHTS. *Drug Discovery Today* 6 (2001) 12, S. 637–646.

- [181] XIE, Y.; SUN, D.; TSE, H. Y.; LIU, C.; CHENG, S. H.: Force sensing and manipulation strategy in robot-assisted microinjection on zebrafish embryos. *IEEE/ASME Transactions on Mechatronics* 16 (2011) 6, S. 1002–1010.
- [182] XU, Q.: Microinjection into zebrafish embryos. *Methods in molecular biology (Clifton, NJ)* 127 (1999), S. 125–132.
- [183] YANG, L.; HO, N.; ALSHUT, R.; LEGRADI, J.; WEISS, C.; REISCHL, M.; MIKUT, R.; LIEBEL, U.; MÜLLER, F.; STRÄHLE, U.: Zebrafish embryos as models for embryotoxic and teratological effects of chemicals. *Reproductive Toxicology* 28 (2009), S. 245–253.
- [184] YEE, N. S.; PACK, M.: Zebrafish as a model for pancreatic cancer research. *Pancreatic Cancer: Methods and Protocols* (2005), S. 273–298.
- [185] ZHANG, T.; RAWSON, D. M.: Studies on chilling sensitivity of zebrafish (*Brachydanio rerio*) embryos. *Cryobiology* 32 (1995) 3, S. 239–246.
- [186] ZHANG, X.; LU, Z.; GELINAS, D.; CIRUNA, B.; SUN, Y.: Batch transfer of zebrafish embryos into multiwell plates. *IEEE Transactions on Automation Science and Engineering* 8 (2011) 3, S. 625–632.
- [187] ZIOU, D.; TABBONE, S.: Edge detection techniques-an overview. In: *International Journal of Pattern Recognition and Image Analysis*, Citeseer, 1998.
- [188] ZON, L. I.; PETERSON, R.: The new age of chemical screening in zebrafish. *Zebrafish* 7 (2010) 1, S. 1–1.
- [189] ZON, L. I.; PETERSON, R. T.: In vivo drug discovery in the zebrafish. *Nature Reviews Drug Discovery* 4(1) (2005), S. 35–44.

Abbildungsverzeichnis

1.1	Bisherige Vorgehensweise bei der Versuchsauslegung bildbasierter Hochdurchsatz-Untersuchungen.	3
1.2	Einsatz von Mikrotiterplatten im Hochdurchsatz	4
1.3	Ausgewählte Entwicklungsstadien der Zebrabärblingslarven.	9
1.4	1D-Profil und 2D-Bildaufnahme einer Zebrabärblingslarve im Vergleich . .	13
1.5	Mikroskopaufnahmen verschiedener Fokusebenen eines Zebrabärblings . . .	14
1.6	Prinzipskizze eines Lichtblatt-Mikroskops	14
2.1	Bisherige und neue Versuchsauslegung für Hochdurchsatz-Untersuchungen .	24
2.2	Neues Konzept zur Optimierung der Versuchsauslegung	27
2.3	Entscheidungsmöglichkeiten und Ablauf einer bildbasierten Hochdurchsatz-Untersuchung	34
2.4	Larve eines Zebrabärblings im Chorion und nach dem Dechorionieren . . .	35
2.5	Qualitative Inhomogenität des Bildstroms	41
2.6	Informelle Inhomogenität der Datensätze [4]	42
2.7	Beispiele für die Registrierung von Multikanalaufnahmen.	43
2.8	Charakterisierung des Nutzsignals	44
2.9	Schema zum mathematischen Modell für bildbasierte Hochdurchsatz-Untersuchungen	47
2.10	Grafische Veranschaulichung eines Elementes des Bildstroms \mathbf{BS}_n [7]	49
2.11	Flussdiagramm für die anforderungsgerechte Strukturierung	53
3.1	Modulkatalog	59
3.2	Bild und Histogramm der Rohdaten	61
3.3	Bild und Histogramm der normalisierten Rohdaten	61
3.4	Verteilung der Merkmalswerte in einer Hochdurchsatz-Untersuchung	63
3.5	Scatterplot der Merkmalswerte der Kontrollen	63
3.6	Mikroskopbilder mit zugehörigen Histogrammen eines Merkmals	63
3.7	Merkmalsraum und Diskriminanzfunktion für zwei Platten	67
3.8	Merkmalsraum und Diskriminanzfunktion für zwei Platten	67
3.9	Ablaufschema des Trackings bei einer Hochdurchsatz-Untersuchung	72
3.10	Sigmoidkurve der Dosis-Wirkungs-Kurve	80
4.1	Prozessschritte und Signalfluss der grafischen Benutzeroberfläche	87
4.2	Skalierung für die Bildverarbeitung im Hochdurchsatzverfahren	90
4.3	Die in Gait-CAD implementierten Module des Modulkatalogs	93
4.4	Gait-CAD: Die TOX-Erweiterung für den FET	93
4.5	PIMP: Die grafische Oberfläche zum Einlesen, Auswerten und Visualisieren	94

4.6	In <i>PIMP</i> implementierte Module des Modulkatalogs	94
5.1	Beispiele von Zebrabärblingslarven in unterschiedlichen Orientierungen . . .	101
5.2	Beispiele und Merkmale eines entwickelten sowie eines koaguliertes Eies eines Zebrabärblings	104
5.3	Module zur Auswertung und Bestimmung der EC_{50} Konzentration	107
5.4	Veranschaulichung der Anwendung des Moduls <i>Trennung von Einzelobjekt & Hintergrund</i>	108
5.5	Auf Basis der Lerndaten berechnete Diskriminanzfunktion	110
5.6	Visualisierung der Klassifikationsergebnisse der Hochdurchsatz-Untersuchung	112
5.7	2D-Histogramm koagulierter und entwickelter Larven	113
5.8	Automatische Regression und EC_{50} -Berechnung	113
5.9	Typische Zebrabärblingslarven und Bildzeitreihen für den Herzschlag	116
5.10	Mittelwert des Merkmals x_{15}	116
5.11	Zeitdauer bis zur ersten Bewegung einer Larve	116
5.12	Module zur Auswertung von Herzschlag und Spontanbewegung	118
5.13	Quantifizierung der Bewegung im Bild	119
5.14	Scatterplot der besten Merkmale nach MANOVA	121
5.15	Klassifikatorkaskade über drei Endpunkte	122
5.16	Repräsentatives Beispiel für die aggregierte Bewegung während eines <i>PMR</i> -Versuchs über der Zeit.	125
5.17	Einzelframe einer Bildsequenz aus einem <i>PMR</i> -Versuch.	125
5.18	Module zur Auswertung der Bewegungsereignisse Coiling und Swimming. .	128
5.19	Kantenbild des Beispiel-Screenshots aus Abb. 5.17.	129
5.20	Akkumulationsmatrix (heatmap) der Hough-Kreisdetektion	129
5.21	Identifizierte Kreise aus Mittelpunkten aller Maxima	129
5.22	Kreuzkorrelation des Quell- oder Objektbildes	130
5.23	Klassifikation der Coiling-Ereignisse	132
5.24	Klassifikation der Swimming-Ereignisse	132
5.25	Wahrscheinlichkeit für das Auftreten von genau einem Swimming- oder Coiling-Ereignis	133
5.26	Dauer der Phasen, ermittelt anhand der Klassifikationen	133
5.27	Extrahierte Bewegungszeitreihe	134
5.28	Zusammenfassung aller 425 Fischlarven als Heatmap.	134
6.1	Optimierung sowohl der Auswerte- als auch der Versuchsparameter	139

Tabellenverzeichnis

1.1	Übersicht ausgewählter Validierungsparameter für Hochdurchsatz-Untersuchungen [66]	6
1.2	Ausgewählte Zebrafisch-Untersuchungen im Hochdurchsatz oder mit Hochdurchsatz-Potenzial mit Anzahl an Fischen oder Substanzen (falls in der Veröffentlichung angegeben). Eine weitere ausführliche Übersicht und Gegenüberstellung bietet [113].	16
3.1	Gegenüberstellung des Rechenaufwands von Korrelation und Kreisdetektion	71
3.2	Ausgewählte Präsentationstechniken für Hochdurchsatz-Untersuchungen	82
4.1	Übersicht der Fremdimplementierungen	95
5.1	Zusammenfassung der Regeln zur Klassenzuordnung	102
5.2	Toxine und Verdünnungsreihen des untersuchten Datensatzes	103
5.3	Ermittlung der notwendigen Zeit der Versuchsdurchführung	106
5.4	Parameter der Hochdurchsatz-Untersuchung für einen Endpunkt	107
5.5	Merkmalsrelevanzen – Univariate und Multivariate Analyse (bei Auswahl von 2 Merkmalen)	110
5.6	Anzahl von untersuchten Larven und Konzentrationen	111
5.7	Als koaguliert eingestufte Larven von jeweils 12 Larven der Konzentration über den Toxinen	111
5.8	Toxine, Verdünnungsreihen und manuell ermittelte Klassenzugehörigkeiten	114
5.9	Ermittlung der notwendigen Zeit der Versuchsdurchführung für eine Mikrotiterplatte	117
5.10	Parameter der Hochdurchsatz-Untersuchung für drei Endpunkte nach Abschnitt 2.5	118
5.11	Merkmalsrelevanzen für Spontanbewegung – Univariate und Multivariate Analyse (bei Auswahl von 2 Merkmalen)	121
5.12	Merkmalsrelevanzen für Herzschlag – Univariate und Multivariate Analyse (bei Auswahl von 2 Merkmalen)	121
5.13	Vergleich der Klassifikationsergebnisse auf Basis von einem und von drei Endpunkten.	123
5.14	Parameter der Hochdurchsatz-Untersuchung nach Abschnitt 2.5 für <i>PMR</i> -Versuch	127
5.15	Übersicht der Validitätskriterien beim <i>PMR</i>	130