Markus Philipp, Neal Bacher, Jonas Nienhaus, Lars Hauptmann, Laura Lang, Anna Alperovich, Marielena Gutt-Will, Andrea Mathis, Stefan Saur, Andreas Raabe, Franziska Mathis-Ullrich*

Synthetic data generation for optical flow evaluation in the neurosurgical domain

Abstract: Towards computer-assisted neurosurgery, scene understanding algorithms for microscope video data are required. Previous work utilizes optical flow to extract spatio-temporal context from neurosurgical video sequences. However, to select an appropriate optical flow method, we need to analyze which algorithm yields the highest accuracy for the neurosurgical domain. Currently, there are no benchmark datasets available for neurosurgery. In our work, we present an approach to generate synthetic data for optical flow evaluation on the neurosurgical domain. We simulate image sequences and thereby take into account domain-specific visual conditions such as surgical instrument motion. Then, we evaluate two optical flow algorithms, Farneback and PWC-Net, on our synthetic data. Qualitative and quantitative assessments confirm that our data can be used to evaluate optical flow for the neurosurgical domain. Future work will concentrate on extending the method by modeling additional effects in neurosurgery such as elastic background motion.

Keywords: Neurosurgery, surgical microscope, optical flow, evaluation

https://doi.org/10.1515/cdbme-2021-1015

1 Introduction

Worldwide, more than 13 million neurosurgical interventions are conducted annually [1]. Thereby, neurosurgeons need a surgical microscope, which provides them a well-illuminated, stereoscopic view on the surgical site (see Fig. 1). To reduce the surgeon's enormous intraoperative workload, there is a need for computational assistance in neurosurgery. This can be achieved by scene understanding algorithms for microscope video data.

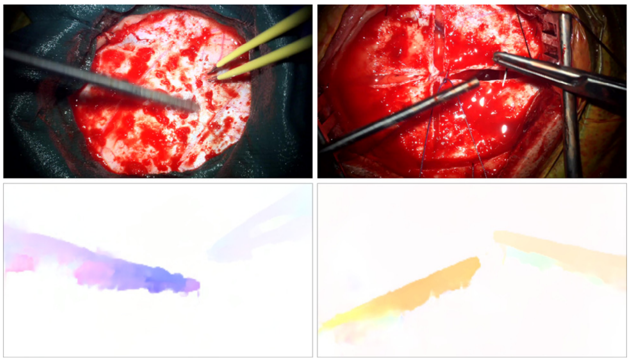


Figure 1: Top row: Intraoperative scenes from our neurosurgical microscope data. Bottom row: Optical flow for the scenes shown above, estimated by PWC-Net [2]. Optical flow is displayed using HSV space. Hue denotes optical flow direction. Saturation decodes the normalized flow magnitude.

Spatio-temporal variations contain important information for automated video understanding, which can be extracted by optical flow. The optical flow captures the apparent motion of objects in a video sequence. For the neurosurgical domain, we show that optical flow is a powerful tool for domain-agnostic instrument localization [3]. Although we prove the benefits of optical flow, we did not yet investigate the choice of the algorithms due to lack of evaluation data.

The question which optical flow algorithm yields highest accuracy on certain video data is addressed by benchmarking. Benchmark datasets consist of video sequences with corresponding ground truth. The common metric for optical flow accuracy is endpoint error [4]. Existing datasets such as FlyingChairs [5] and MPI-Sintel [6] are arguably different from neurosurgical data. To achieve high accuracy on the neurosurgical domain, optical flow algorithms must perform

***Corresponding author: Franziska Mathis-Ullrich:** Karlsruhe Institute of Technology, Karlsruhe, Germany, franziska.ullrich@kit.edu
Markus Philipp: Karlsruhe Institute of Technology, Karlsruhe, Germany
Markus Philipp, Neal Bacher, Jonas Nienhaus, Lars Hauptmann, Laura Lang, Stefan Saur: Carl Zeiss Meditec AG, Oberkochen, Germany
Marielena Gutt-Will, Andrea Mathis, Andreas Raabe: Inselspital Bern, Bern, Switzerland

well given challenging visual conditions: low texture tissue, blur, specularly reflective surfaces, and large amplitude motion. While in other medical domains methods exist to generate synthetic data for optical flow evaluation [7], there are presently no such methods for neurosurgery.

Contributions. Here, we present a method to generate synthetic neurosurgical microscope data for optical flow evaluation. This data takes into account the variety of surgical instruments and challenging visual conditions present in neurosurgery. In our experiments, we compare the accuracy of two state-of-the-art optical flow algorithms on the generated data.

2 Method

We present the approach for synthetic data generation through simulation of the microscope field of view (FOV).

2.1 Synthetic data generation approach

Evaluation of optical flow requires a sequence of at least two images and the corresponding optical flow ground truth. To evaluate optical flow algorithms, we generate short sequences of abstract surgical actions in the microscope FOV. From clinical recordings, we concluded that the majority of motion is caused by surgical instruments. Therefore, our approach focuses on simulating instrument motion.

Our simulation is inspired by the geometric constraints during neurosurgical interventions. Typically, the surgeon operates under a microscope with an instrument in each hand (see Fig. 2 (a)). We model this situation in our simulation (see Fig. 2 (b)). According to discussions with surgical experts, the surgeon's instruments point towards the location of surgical action. We refer to this location as *activity center*, p_{ac} (see Fig. 2 (c)). We model this by simulating instrument tip motion around a (non-visible) activity center. To achieve plausible instrument body motion, we constrain each instrument body i by a prismatic-spherical joint at the random location $p_{h,i}$ (possible surgeon's hand locations).

The data generation process consists of three steps.

(1) Motion modelling. Since optical flow evaluation requires only short temporal motion sequences, we abstain from simulating the complete surgery. Instead, we describe instrument motion by a simple randomized model (see Fig. 2

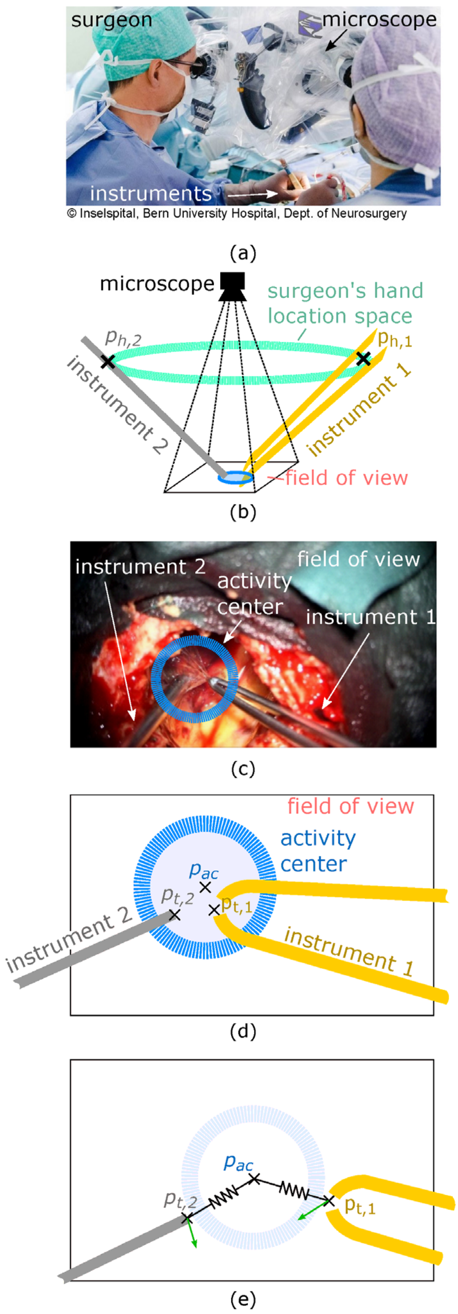
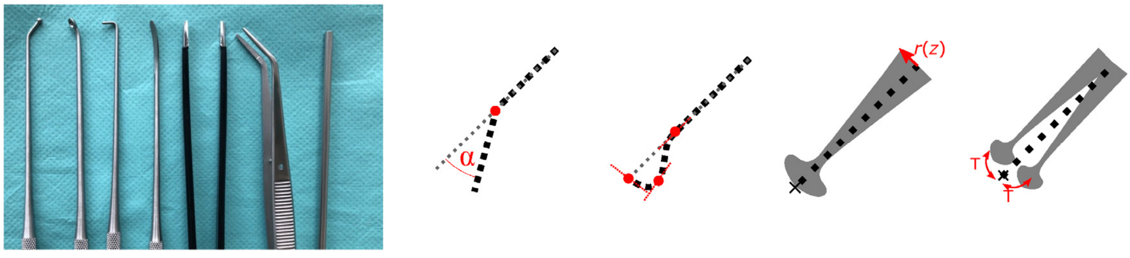


Figure 2: (a) Clinical set-up in neurosurgery: the neurosurgeon operates with two instruments under the microscope (© Inselspital, Bern University Hospital, Dept. of Neurosurgery). (b) Simulation set-up: Two instruments act inside the field of view of a microscope. The instrument's pose is defined by a spherical-prismatic joint at $\boldsymbol{p_{h,i}}$, which represents the surgeon's hand. (c) Clinical recording from our data: scene with marked activity center and instruments. (d) Modeling in our simulation: activity center $\boldsymbol{p_{ac}}$ and instruments with tip locations $\boldsymbol{p_{t,i}}$. (e) Motion model in our simulation: The instruments are linked to the activity center by a spring. Additionally, a randomly oriented force (green) is applied to the tips.



(a) (b) (c) (d) (e)

Figure 3: (a) shows commonly used neurosurgical instruments that we take as motivation for our shape variation methods. (b) - (e) depict our shape variation methods: (b) and (c) display our approach for instrument axis manipulation. (d) shows the variation of instrument cross section. (e) depicts our method to generate two-part instruments.

(e)). For each new sequence, the activity center p_{ac} is set randomly. Its acceleration $\frac{d}{dt}\boldsymbol{v_{ac}}$ is modeled by a Gaussian distribution in eq. 1,

$$\frac{\mathrm{d}}{\mathrm{dt}}\boldsymbol{v_{ac}} = \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{ac}^2\right), \tag{1}$$

where $\boldsymbol{\Sigma_{ac}}$ is tuned manually to match the clinical data. Possible motions are limited inside the FOV.

The instrument tip motion $\boldsymbol{v_{t,i}}$ is modeled by an attraction force towards the activity center. Additionally, we introduce a random acceleration term. The velocity of the instrument tip $\boldsymbol{v_{t,i}}$ is described by eq. 2,

$$\frac{\mathrm{d}}{\mathrm{dt}}\boldsymbol{v_{t,i}} = k_i \boldsymbol{d_{i \to ac}} + \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_i^2\right), \tag{2}$$

whereas $\boldsymbol{d_{i \to ac}}$ describes the vector to p_{ac}. Attraction of activity center k_i and random motion component $\boldsymbol{\Sigma_i}$ are also adjusted manually. We enforce the instruments to stay within the FOV by increasing k_i upon leaving the FOV. Collisions of instrument bodies are avoided by collision detection routines.

(2) Instrument shape variation. Neurosurgical interventions are characterized by a high variability of instrument types and instrument shapes, [8] lists more than 100 different instruments. Thus, accurate modeling of *all* existing neurosurgical instruments is impossible. Instead, we simulate the shape variety by generating arbitrary instrument shapes using structured domain randomization.

Our approach consists of three main operations. Each operation is inspired by properties of existing instruments (see Fig. 3 (a)). First, we create the instrument axis as straight line in cylinder coordinates (z, ϕ, r). Then, we manipulate the instrument axis by inserting a random kink (see Fig. 3 (b)) and/or bending the axis by inserting Beziére curves (see Fig. 3 (c)). Third, we extrude a cross-section along the instrument axis. All instruments generated in our approach have rotational symmetric cross-sections, which depend on the radius $r(z)$ along the z-axis. The radius function $r(z)$ is determined by random combination of elementary functions, including low-order polynomials, exponential and trigonometric functions.

Moreover, some neurosurgical instruments such as bi-polar coagulator or forceps consist of two body parts. Therefore, to model instruments with two body parts, we clone the extruded instrument bodies with a certain probability. We apply a random transformation $\boldsymbol{T}$ to the two instrument body parts relative to the instrument tip $p_{t,i}$. By translation and slight rotation, the intersection point of the bodies is moved along the instrument axis.

(3) Background augmentation and visual property simulation. To increase the degree of realism in our simulation, we include still images from neurosurgical video recordings as background. Including these images prior to rendering allows simulation of characteristic mirroring effects on the instruments. To simulate realistic instrument reflection, we incorporate specular textures. To achieve a realistic lighting set-up of the microscopes, we model a co-axial light source. Furthermore, to simulate the usually limited depth of field in neurosurgical video data we apply artificial blur.

Implementation. The simulation approach was implemented in the open-source framework Blender. First, we generate instruments according to the described model. Next, we create instrument motion sequences of arbitrary length. Optical flow output is directly obtained from Blender.

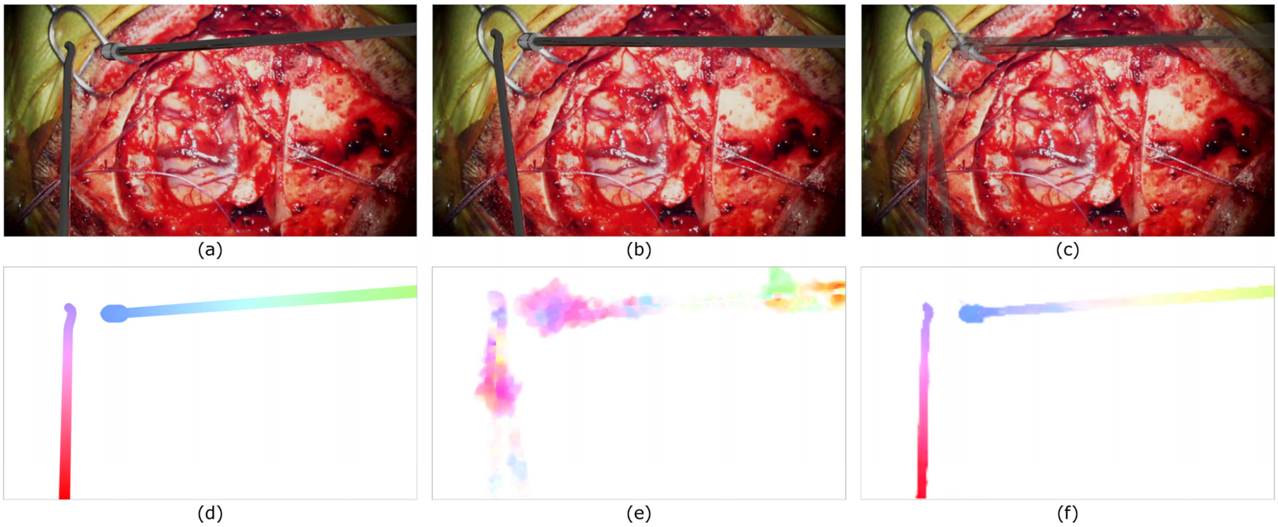


Figure 4: Example scene generated by our method. (a) shows the generated scene at timestep $\boldsymbol{T} = \mathbf{0}$, (b) depicts timestep $\boldsymbol{T} = \mathbf{1}$. (c) displays the overlay of (a) and (b) to highlight the motion. (d) shows the optical flow ground truth data, while (e) is the estimated optical flow using Farneback and (f) the flow estimated by PWC-Net. Optical flow is displayed using HSV space. Hue denotes optical flow direction. Saturation decodes the normalized flow magnitude.

2.2 Experimental setup for optical flow evaluation

We evaluate our synthetic data w.r.t. accuracy of two state-of-the-art optical flow algorithms, Farneback[1] [9] and PWC-Net[2] [2]. For our analysis, we generate a dataset *NeurOF* comprising 1087 sequences, each with 10 images at a resolution of 960 x 540. Background images were inserted from video recordings of 2 tumor, 2 vascular and 2 spine cases, which were conducted at Inselspital Bern, Switzerland with a ZEISS KINEVO 900 surgical microscope. For comparison, we evaluate two public datasets, namely FlyingChairs [5] and MPI-Sintel [6]. We evaluate the accuracy of estimated optical flow $\left(v_{i,j}\right)_{pred}$ versus ground truth $\left(v_{i,j}\right)_{gt}$ by means of endpoint error in eq. 3,

$$AEPE \; = \frac{1}{n} \sum_{i,j} \left(\left(v_{i,j}\right)_{gt} - \left(v_{i,j}\right)_{pred} \right)^2 . \qquad (3)$$

3 Results

First, we present example images as generated by our method.

We evaluate them qualitatively w.r.t. realism and image quality. Additionally, we present the generated optical flow ground truth. The synthetic scene shown in Fig. 4, 5 (a)-(d) prove that our method generates image data that resembles clinical neurosurgical data. Comparing clinical data (Fig. 1) and synthetic data (Fig. 4, 5) we observe similar visual appearances. The computed optical flow from the two algorithms in our evaluation, Farneback and PWC-Net (see Fig. 4, 5 (e)-(f)), indicate our data can be used to assess optical flow accuracy.

Table 1: Numeric results for AEPE for Farneback and PWC-Net on different datasets. PWC-Net performs significantly better than Farneback on all three datasets. The AEPE for *NeurOF* is lower than for FlyingChairs and MPI-Sintel for both, Farneback and PWC-Net.

Dataset	Farneback	PWC-Net (large)
FlyingChairs	8.18	1.44
MPI-Sintel (final)	11.35	3.7
NeurOF	2.48	0.81

Next, we quantitatively assess the accuracy of these two optical flow algorithms on our synthesized data. According to the numerical results (see Table 1), PWC-Net performs at lower AEPE on *NeurOF* than Farneback. Qualitative results in Fig. 4, 5 support our numerical findings. The lower AEPE for PWC-Net than Farneback are in line with numeric results on the public datasets FlyingChairs and MPI-Sintel (see Table 1).

[1] OpenCV implementation. Parameters: pyramid scale = 0.5, levels = 3, window size = 15, iterations = 3, size pixel neighborhood = 5, STD for smoothing derivatives = 1.2

[2] PWC-Net large, implementation and weights (FlyingChairs/FlyingThings3D cycle) by github.com/philferriere/tfoptflow

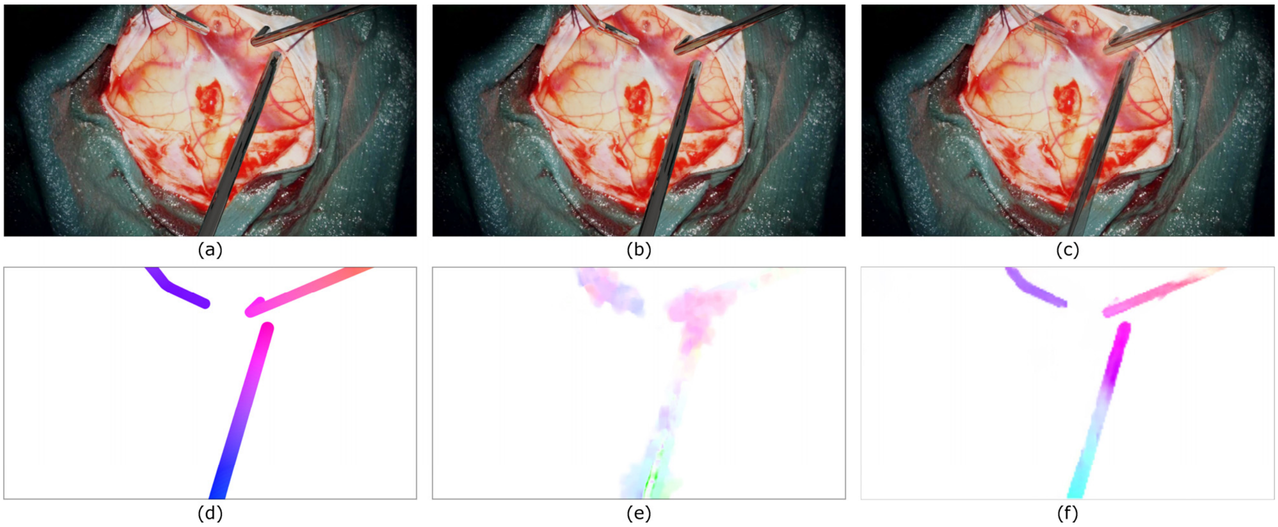


Figure 5: Further example scene generated by our method; captions see Fig. 4.

In direct comparison, the AEPE on *NeurOF* is lower than for FlyingChairs and MPI-Sintel. This observation is expected since *NeurOF* models only foreground instrument motion, while background motion is zero.

4 Conclusions

We demonstrate a synthetic data generation approach for the evaluation of optical flow on the neurosurgical domain. Our approach captures various effects, which are relevant for this domain such as large instrument motion amplitude and specular reflections. We verify the quality of the generated data through visual inspection w.r.t. optical flow calculation. Numerical results show that PWC-Net performs better than Farneback. However, our data models only instrument motions while the background is fixed. Future work will address simulation of a non-rigid background to further improve realism. Then, the benchmark can be extended by evaluating more state-of-the-art optical flow algorithms. Although our experiments focus on using the data generation approach for optical flow evaluation, our method potentially can also improve other computer-assisted surgery applications.

Author Statement

Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

References

[1] Dewan, M.; Rattani, A. et al. (2018): Global neurosurgery: the current capacity and deficit in the provision of essential neurosurgical care. J Neurosurg, 130, 2018.

[2] Sun, D.; Xiaodong, Y. et al. (2018): PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume. IEEE Conference on Computer Vision and Pattern

[3] Philipp, M.; Alperovich, A. et al. (2021): Localizing neurosurgical instruments across domains and in the wild. Medical Imaging with Deep Learning, 2021.

[4] Baker, Simon; Scharstein, D. et al. (2010): A Database and Evaluation Methodology for Optical Flow. International Journal of Computer Vision, 92, 1-31.

[5] Dosovitskiy, A.; Fischer, P. et al. (2015): FlowNet: Learning Optical Flow with Convolutional Networks. IEEE International Conference on Computer Vision, 2015.

[6] Butler, D. J.; Wulff, J. et al. (2012): A naturalistic open source movie for optical flow evaluation. European Conf. on Computer Vision (ECCV), 2012.

[7] Ravasio, C.; Pissas, T. et al. (2020): Learned optical flow for intra-operative tracking of the retinal fundus. International Journal of Computer Assisted Radiology and Surgery, 15, 827–836.

[8] Eddleman, C. (2011): The Neurosurgical Instrument Guide. Thieme Medical Publishers.

[9] Farneback, G. (2003): Two-Frame Motion Estimation Based on Polynomial Expansion. Scandinavian Conference on Image Analysis, 2003.