Intelligent Systems & Advanced Manufacturing, Pittsburgh, Pennsylvania USA, 13–17 October 1997, Proc. SPIE 3208

Active vision and sensor fusion for inspection of metallic surfaces

F. Puente León and J. Beyerer

Institut für Meß- und Regelungstechnik
Universität Karlsruhe (TH), Postfach 6980
D-76128 Karlsruhe, Germany

ABSTRACT

This paper deals with strategies for reliably obtaining the edges and the surface texture of metallic objects. Since illumination is a critical aspect regarding robustness and image quality, it is considered here as an active component of the image acquisition system.

The performance of the methods presented is demonstrated – among other examples – with images of needles for blood sugar tests. Such objects show an optimized form consisting of several planar grinded surfaces delimited by sharp edges. To allow a reliable assessment of the quality of each surface, and a measurement of their edges, methods for fusing data obtained with different illumination constellations were developed. The fusion strategy is based on the minimization of suitable energy functions.

First, an illumination-based segmentation of the object is performed. To obtain the boundaries of each surface, directional light-field illumination is used. By formulating suitable criteria, nearly binary images are selected by variation of the illumination direction. Hereafter, the surface edges are obtained by fusing the contours of the areas obtained before. Following, an optimally illuminated image is acquired for each surface of the object by varying the illumination direction. For this purpose, a criterion describing the quality of the surface texture has to be maximized. Finally, the images of all textured surfaces of the object are fused to an improved result, in which the whole object is contained with high contrast.

Although the methods presented were designed for inspection of needles, they also perform robustly in other computer vision tasks where metallic objects have to be inspected.

Keywords: surface image acquisition, image contrast, sensor fusion, illumination series, active vision

1 INTRODUCTION

To illustrate the vision problems discussed in this paper, Fig. 1 shows an image of a firing pin print on a cartridge case illuminated by parallel, directed light. For the purpose of forensic examinations, it is important to acquire high quality images of such objects containing all features caused by the fire arm used. In Fig. 1 a small object area can be observed with high contrast; however, most of the relevant surface areas are dark and deliver no useful information. In comparison, Fig. 2 shows the same object with diffuse lighting. The surface is illuminated more homogeneously, but the contrast is lower than in the small, high contrast region of Fig. 1. Especially, a distinct feature resulting from surface topography – indicated by an arrow – is visible only if directed lighting is applied.

It would be desirable to get an image incorporating both advantages, the homogeneity of diffuse illumination, and the high contrast of directed lighting. This can be achieved by acquiring a series of images with systematically varying illumination, so that every relevant surface area is captured with high quality within at least one image. Subsequently, the regions with optimal contrast are cut out and are combined to an image of high quality everywhere. In Fig. 3, four of twenty images of an illumination series are shown. The elevation angle θ was kept constant, whereas the azimuth was increased stepwise ($\Delta\varphi = 18°$). The fusion result can be seen in Fig. 4. Compared with Fig. 2, many details within the firing pin print area which are clearly visible in Fig. 4 become invisible, if diffuse lighting is used.

Further author information –
F.P.L.: Email: f.puente@ieee.org; WWW: http://www-mrt.mach.uni-karlsruhe.de/~puente; Telephone: +49-721-608 3604; Fax: +49-721-661874
J.B.: Email: beyerer@mrt.mach.uni-karlsruhe.de; WWW: http://www-mrt.mach.uni-karlsruhe.de/~beyerer; Telephone: +49-721-608 2209; Fax: +49-721-661874

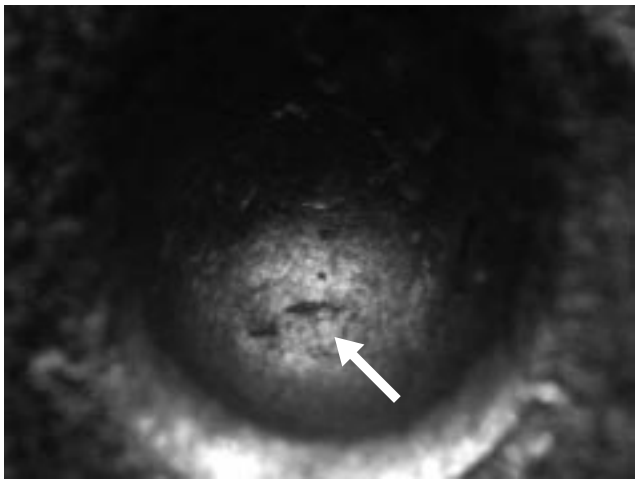


Figure 1. Firing pin print: directed lighting.

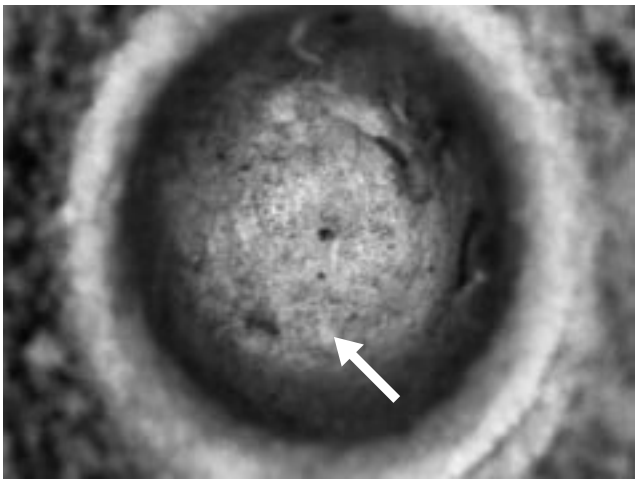


Figure 2. Firing pin print: diffuse lighting.

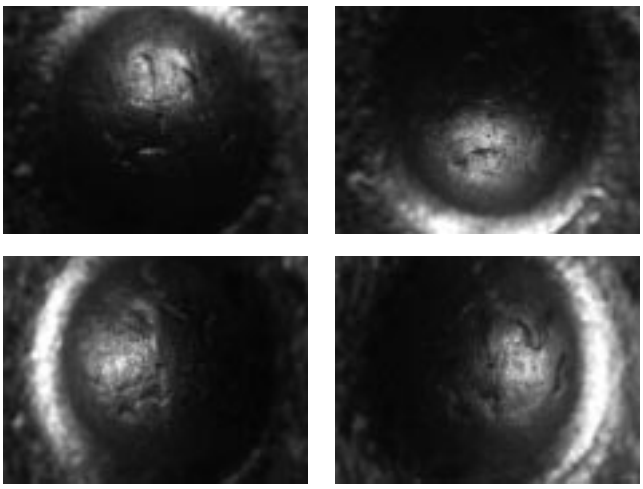


Figure 3. Firing pin print: illumination series.

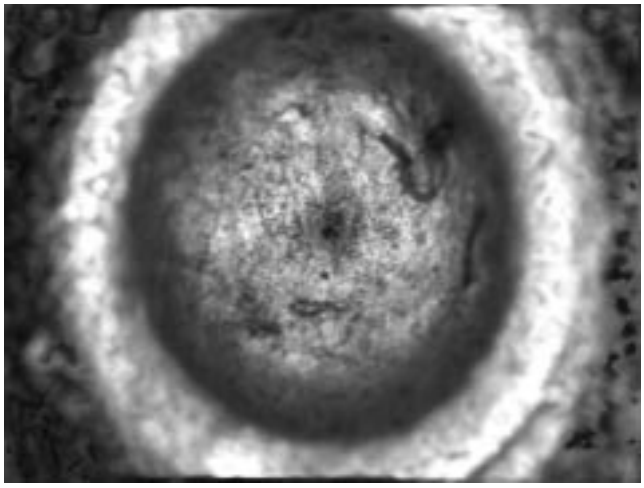


Figure 4. Firing pin print: fusion result.

2 PRINCIPLES OF SENSOR FUSION

Sensor fusion within the scope of this paper means to extract and unite the pertinent contents of a series of M images to N results, which can be images, feature sets or symbolic image descriptions; see Fig. 5.

In *data fusion*, image data are combined to obtain new images. Two different modes of data fusion must be distinguished. In the case that the relevant information is concentrated locally in the images of the series, it is reasonable to cut out these areas and combine them like in a patch work. This mode of data fusion is called *complementary* data fusion. An example is the fusion of the illumination series performed in section 1. On the other hand, if the same relevant information is spread out spatially all over each of the images of the series, it is necessary to perform a kind of averaging over the series. Therefore, we propose to call this mode *averaging* data fusion. In literature, it is sometimes called *competitive* fusion.[11] An example is linear averaging of images of the same scene to reduce noise. Often, it is expedient to perform a mixture of both modes. For example, to avoid artefacts, in Fig. 4, at the boundaries of the patches from different images of the series, a local averaging was done.

Fusion can also be reasonable for more abstract image descriptions than image data itself. *Feature fusion* may be performed to obtain improved numerical estimates of certain image properties.

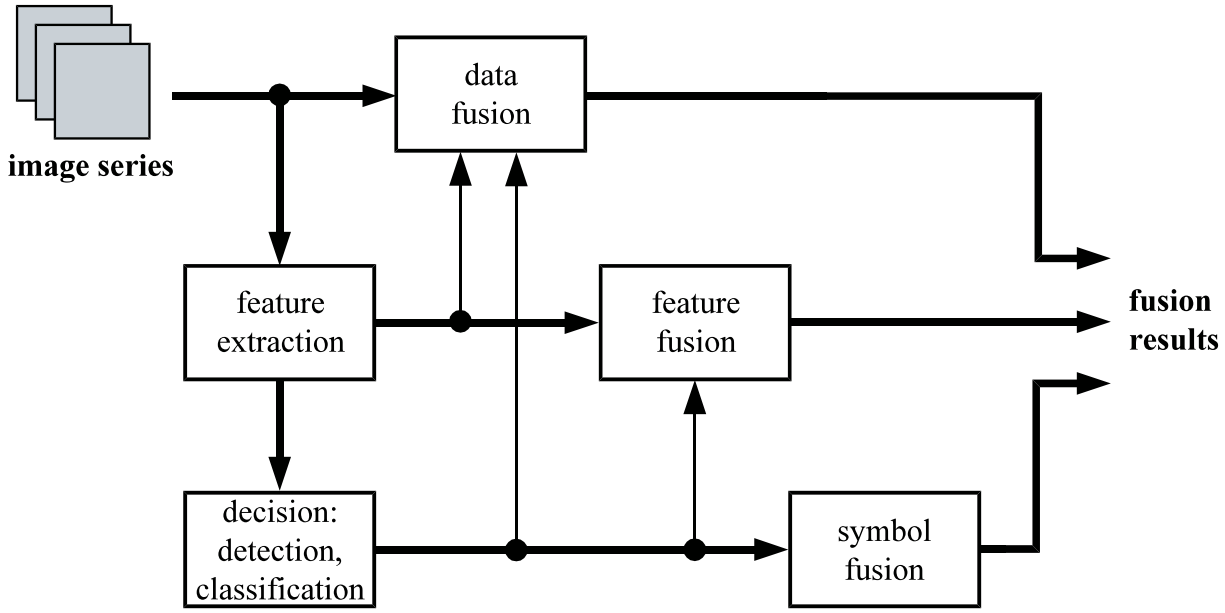


Figure 5. Fusion of image series.

At the highest level of abstraction, symbolic image descriptions, like detection or classification results with regard to individual images of the series, are combined. This is commonly called *decision fusion*, althought, within the scope of this paper, *symbol fusion* is a more appropriate term. For a more detailed taxonomy, we refer to [5].

In data fusion as well as in feature fusion it can be favorable to consider results from higher abstraction levels, as indicated in Fig. 5. For instance, it could be advantageous to use segmentation results (i.e. a symbolic description) of individual series images to facilitate the fusion at data level, because the segmentation delivers several regions with different data properties.

A universal methodology for formalizing fusion problems consists in describing all knowledge available, assumptions, and wishes about data, features, and symbols as well as connections between them with generalized energy terms E_k.[6] These energies E_k should describe the given knowledge and requirements in a monotonous way, that is, the lower the energies are, the better the fusion result should be. The terms E_k are summarized by weighted summation to the so-called energy functional E:

$$E = \sum_k \lambda_k E_k. \tag{1}$$

E represents an implicit, compact description of the fusion problem. Because of the monotony of E, the fusion is accomplished by minimization of the energy functional with respect to the fusion result. Thus, the fusion task is led back to an optimization problem.

There is an interesting connection with Bayesian statistics. According to statistical physics, a Gibbs probability density function (PDF) can be defined for the energy functional:

$$\mathrm{PDF} \propto \mathrm{e}^{-\frac{E}{T}} = \prod_k \mathrm{e}^{-\frac{-\lambda_k E_k}{T}}. \tag{2}$$

T can be thought of as a generalized temperature. Since the energy functional is a sum, the PDF can be decomposed into factors. We will show later that these factors can be interpreted either as a likelihood function or as a priori PDFs. By means of an appropriate normalization of eq. (2), the a posteriori PDF for the fusion

result given the image series is obtained. The monotony of the exponential function ensures that minimizing E is equivalent to maximizing eq. (2). Therefore, the optimization delivers the maximum a posteriori estimate of the fusion result. It should be emphasized that – since the energy terms do not only embody objective knowledge, but also subjective wishes and requirements – the PDF eq. (2) is a subjective probability description. However, the advantage of describing fusion from a probabilistic point of view is that there exists a powerful set of mathematical tools for treating Gibbs PDFs. An important example is the simulated annealing optimization method.[9]

3 FUSION OF ILLUMINATION SERIES

The abstract fusion strategy presented in section 2 will be applied now concretely to fusion of illumination series. For better understanding, fusion of images of edgeless objects – i.e. objects in which neither profile nor textural edges are contained – will be discussed first. In section 3.2, an extension of the proposed approach will be shown that allows to fuse illumination series of more complicated objects.

3.1 Fusion of illumination series of edgeless objects

In this section, only one-dimensional series – i.e. series in which only one acquisition parameter is varied – will be treated. However, the approach presented can be easily extended to enable fusion of multi-dimensional image series.

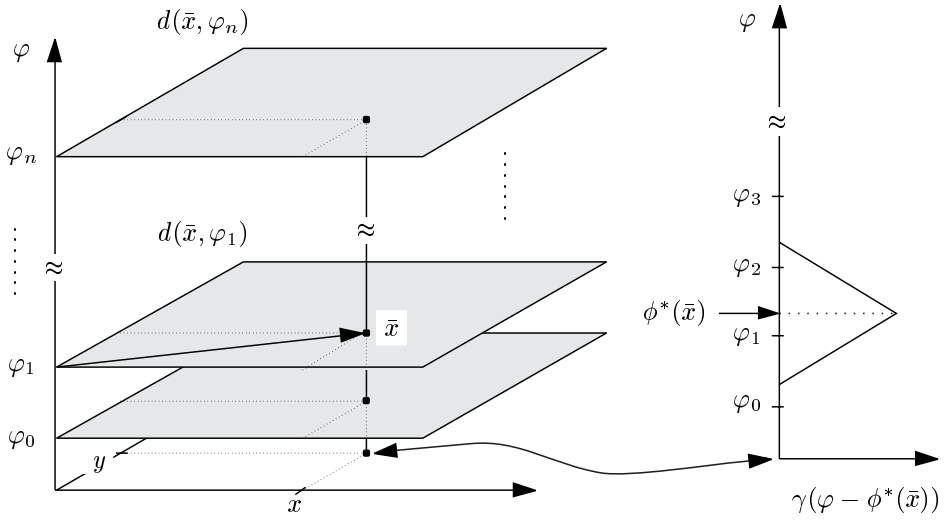


Figure 6. Selection of the best illuminated image segments for fusion.

Before describing the fusion problem mathematically, the fundamental idea will be explained exemplarily for an image series in which the azimuth φ is varied; see Fig. 6. The series consists of $n+1$ images $d(\bar{x}, \varphi_i)$, where φ_i denote the discrete illumination angles. For each location $\bar{x}$, the "best" of the series has to be extracted, and combined to a final result $r^*(\bar{x})$. The optimal illumination azimuth =: $\phi^*(\bar{x})$ specifies from which images of the series image areas are to be "cut out". This optimal parameter function $\phi^*(\bar{x})$ must be a spatially "slowly" varying signal compared with the signal of interest to avoid interferences and artifacts caused thereby. In general, $\phi^*(\bar{x})$ will also adopt values between the discrete angles φ_i. Thus, cutting out has to be understood in a general way in the sense that more than one image of the φ-neighbourhood of $\phi^*(\bar{x})$ can contribute to the fusion result. This is enabled by an interpolator $\gamma(\varphi - \phi^*(\bar{x})) \geq 0$, which is generally unimodal with respect to φ, and narrowly concentrated around ϕ^*. Particularly, in Fig. 6 a linear interpolator is depicted, which provides for a weighted averaging of the two best images of the series at the location $\bar{x}$.

A general energy functional suitable in the sense of eq. (1) for performing data fusion is:

$$E = E_{\mathrm{D}}(D, r) + \lambda E_{\mathrm{C}}(r), \qquad \lambda > 0, \tag{3}$$

$E_{\mathrm{D}}(D, r)$ models the relationship between the given image data (i.e. the image series)

$$D = \{d(\bar{x}, \varphi_i), i = 0, \ldots, n\}, \tag{4}$$

and the fusion result $r^*(\bar{x})$.

$E_{\mathrm{C}}(r)$ models desired or a priori known characteristics of the fusion result $r^*(\bar{x})$ itself. The regularization parameter λ serves to weight both components.

The "energy terms" $E_{\mathrm{D}}(D, r)$ and $E_{\mathrm{C}}(r)$ are to be defined in such a way that the result is the more desirable, the lower the energy is. Consequently, E has to be minimized to obtain the optimum result $r^*(\bar{x})$.

However, for the present fusion task, the energy functional eq. (3) has to be extended accordingly:

$$\begin{aligned} E &= \sum_i \sum_{\bar{x}} \left(r(\bar{x}) - d(\bar{x}, \varphi_i)\right)^2 \gamma(\varphi_i - \phi(\bar{x})) + \lambda_1 \sum_{\bar{x}} \left(\mathrm{HP}\{\phi(\bar{x})\}\right)^2 + \lambda_2 \sum_{\bar{x}} (-1) \cdot C\{r(\bar{x})\} \\ &= E_{\mathrm{D}}(D, r, \phi) + \lambda_1 E_{\mathrm{S}}(\phi) + \lambda_2 E_{\mathrm{C}}(r) \end{aligned} \tag{5}$$

with

$$\varphi_i = \varphi_0 + i\Delta\varphi, \qquad i = 0, \ldots, n, \qquad \varphi_0 \le \phi(\bar{x}) \le \varphi_n, \qquad \lambda_1, \lambda_2 > 0, \quad \mathrm{HP}\{\}\text{: high-pass.}$$

Eq. (5) represents a compact, implicit formulation of the fusion problem, in which all known and desirable characteristics of the quantities involved in the fusion process as well as their mutual relations are given. For the solution of the fusion task, E has to be minimized simultaneously with respect to $r(\bar{x})$ and the azimuth function $\phi(\bar{x})$:

$$E(r^*, \phi^*, D) := \min_{r,\phi}\{E\}. \tag{6}$$

So far, eq. (5) does not provide any information about explicit estimation strategies for $r^*(\bar{x})$ and $\phi^*(\bar{x})$. However, such strategies are given in section 3.1.2.

Formally, compared with eq. (3), an additional energy term $E_{\mathrm{S}}(\phi)$ concerning the illumination azimuth $\phi(\bar{x})$ is included. From the point of view of estimation theory, it represents a *nuisance parameter*,[13] to which reasonable requirements can be imposed which should influence the estimation of the interesting quantity $r^*(\bar{x})$. In our case, it is reasonable to postulate that the optimal illumination azimuth changes locally more slowly than the surface texture to be imaged. Typically, the texture $t(\bar{x})$ has a significant band-pass character.[4] To avoid interferences with the texture $t(\bar{x})$ due to local changes of the illumination direction, $\phi^*(\bar{x})$ is assumed to have a low-pass character. In the ideal case, the support sets of the Fourier spectra $T(\bar{f})$ and $\Phi^*(\bar{f})$ of the signals $t(\bar{x})$ and $\phi^*(\bar{x})$ should be disjointed:

$$\mathrm{supp}\left\{T(\bar{f})\right\} \cap \mathrm{supp}\left\{\Phi^*(\bar{f})\right\} = \emptyset. \tag{7}$$

Hence, the second addend $E_{\mathrm{S}}(\phi)$ penalizes inadmissible high frequency components of $\phi(\bar{x})$ by measuring the energy of the high-pass filtered signal $\mathrm{HP}\{\phi(\bar{x})\}$. Consequently, this addend represents a smoothness constraint for the optimal illumination angle $\phi^*(\bar{x})$.

The first addend $E_{\mathrm{D}}(D, r, \phi)$ in eq. (5) provides for data proximity to $r^*(\bar{x})$. Locally, $r^*(\bar{x})$ has to resemble that image $d(\bar{x}, \varphi_i)$ of the series which shows optimal illumination at the location $\bar{x}$. In the first energy term, the function $\gamma(\varphi_i - \phi(\bar{x}))$ only allows a contribution of the best-suited $d(\bar{x}, \varphi_i)$ and its neighbours.

The third addend $E_{\mathrm{C}}(r)$ evaluates whether the local criterion C (e.g. a local contrast measure), which grows monotonously with image quality, promotes high values of C in the fusion result $r^*(\bar{x})$ globally.

With regard to the assumptions met here, the simultaneous minimization of E with respect to $r(\bar{x})$ and $\phi(\bar{x})$ leads to the optimal fusion result $r^*(\bar{x})$ (and, of course, also to the optimal azimuth $\phi^*(\bar{x})$).

3.1.1 Connection with Bayesian statistics

If we now insert eq. (5) into the Gibbs PDF eq. (2), we obtain:

$$\begin{aligned} p(D,r,\phi) &\propto \mathrm{e}^{-\frac{E_\mathrm{D}}{T}} \cdot \underbrace{\mathrm{e}^{-\frac{\lambda_1 E_\mathrm{S}}{T}} \cdot \mathrm{e}^{-\frac{\lambda_2 E_\mathrm{C}}{T}}}_{} \\ p(D,r,\phi) &= \underbrace{\underbrace{p(D|r,\phi)}_{\text{likelihood}} \cdot \underbrace{p(r,\phi)}_{\text{a priori PDF}}}_{} \\ &\propto \underbrace{p(r,\phi|D)}_{\text{a posteriori PDF}} . \end{aligned} \tag{8}$$

Since there are three energy terms, the PDF also consists of three probability factors. The first one depends on the observed data D as well as on the signal of interest $r(\bar{x})$, and on the azimuth function $\phi(\bar{x})$. Therefore, from a statistical point of view, it can be interpreted as the likelihood function, that is the conditional PDF of the data D given the unknowns $r(\bar{x})$ and $\phi(\bar{x})$. Consequently, since the second and third term are independent of D, they play the role of an a priori PDF $p(r,\phi)$, which describes the properties of $r(\bar{x})$ and $\phi(\bar{x})$. The fact that $p(r,\phi)$ can be factored into $p(\phi) \cdot p(r)$ reflects that the knowledge about $r(\bar{x})$ and $\phi(\bar{x})$ was implicitely assumed to be independent when the energy terms were established.

3.1.2 Implementation strategy

Because of the very high computation time required for performing the fusion task, an *efficient approximation* is proposed, which is based on a separate minimization of the addends of the energy function (5) without considering neither their coupling nor the weighting factors λ_i; see Fig. 7.

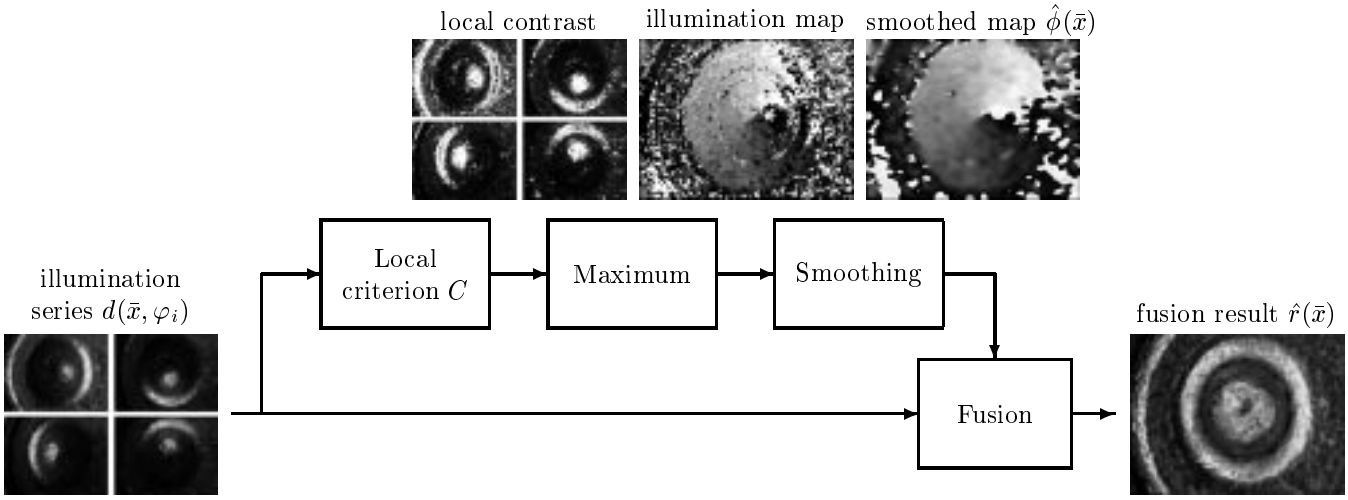


Figure 7. Structure of the minimization strategy for E.

In the first step, the signal

$$\tilde{\varphi}(\bar{x}) = \arg\max_{\varphi_i} C\{d(\bar{x},\varphi_i)\} \tag{9}$$

is calculated, which provides information about the azimuth which maximizes the local criterion C. This essentially leads to a minimization of the third addend $\lambda_2 E_\mathrm{C}(r)$. To achieve a contrast as high as possible, criteria C like the

local grey level variance, the local average gradient magnitude or the local entropy are suitable choices. In other kinds of inspection tasks, where more complex textures consisting of several different components have to be imaged, it may be useful to combine the above-mentioned criteria with feature or direction selective filters; see e.g. [2].

In the next step, the signal $\tilde{\varphi}(\bar{x})$ is smoothed with a low-pass filter:

$$\hat{\phi}(\bar{x}) = \angle \mathrm{TP}\{\mathrm{e}^{j\tilde{\varphi}(\bar{x})}\}. \tag{10}$$

Thereby, essentially the second addend $\lambda_1 E_\mathrm{S}(\phi)$ is minimized. The low-pass filter is chosen to be a binomial filter, because it can be implemented very efficiently in the spatial domain,[7] and it represents an approximation of a continuous Gaussian filter. In the one-dimensional case, Gaussian filters have the favourable property of not generating new extrema. Though this property cannot be extended to multi-dimensional Gaussian filters, these filters are especially benign with respect to generation of artifacts.[10] Moreover, the cyclicity of φ has to be taken into account, because $\varphi = \varphi + 2\pi k$, $k \in \mathbb{Z}$ holds. Thus, not $\tilde{\varphi}(\bar{x})$ itself, but the complex pointer $\exp(j\tilde{\varphi}(\bar{x}))$ has to be smoothed. The resulting function $\hat{\phi}(\bar{x})$ is the angle of the complex result.[12]

The actual fusion is performed by weighted superposition of at most two adjacent images with a linear interpolator γ. This leads to a minimization of the first addend above all. The fusion result is:

$$\hat{r}(\bar{x}) = \frac{\hat{\phi}(\bar{x}) - \varphi_l}{\varphi_{l+1} - \varphi_l} d(\bar{x}, \varphi_l) + \frac{\varphi_{l+1} - \hat{\phi}(\bar{x})}{\varphi_{l+1} - \varphi_l} d(\bar{x}, \varphi_{l+1}), \qquad \varphi_l \le \hat{\phi}(\bar{x}) \le \varphi_{l+1}. \tag{11}$$

The interpolation takes care of a smooth transition between φ-neighbouring images. The narrow extent of $\gamma(.,.)$ provides for an averaging of grey levels of only similarly illuminated images. Thus, an undesirable contrast loss due to destructive interferences of light and shadow in different images of the series is avoided.

3.2 Fusion of illumination series of objects with edges

In the case that images of objects containing textural or profile edges have to be obtained, the smoothness postulate for the illumination map is only partially reasonable. In general, the object areas beside such an edge will require a different illumination direction to image them with high quality. However, the smoothness term $E_\mathrm{S}(\phi)$ formulated in section 3.1 will enforce a slow transition between two optimal illumination directions, leading thus to poor illumination conditions near to the edges. Since fusion of images of objects with edges according to section 3.1 will complicate the information extraction concerning the edge areas from the fusion result, an extension of the methodology presented is needed which allows to fuse images to a high-quality result.

The extension is straight forward, if we combine the method of section 3.1 with a segmentation $S = \{S_1, \ldots, S_J\}$ of the image area into J edgeless regions S_j, $j = 1 \ldots J$ within which it is reasonable to assume a smooth azimuth function ϕ. Then, each region S_j can be treated as before. Thus, the images are modelled hierarchically in the sense that at an upper level of abstraction the images are regarded as a set of regions, and at a lower level the signal contents filling these regions are of interest.

It is easy to extend the energy functional in such a way that it also describes the segmentation process. For this purpose, three energy terms are added:

$$\begin{aligned} E &= \sum_{S_j} \sum_i \sum_{\bar{x} \in S_j} (r(\bar{x}) - d(\bar{x}, \varphi_i))^2 \gamma(\varphi_i - \phi(\bar{x})) + \lambda_1 \sum_{S_j} \sum_{\bar{x} \in S_j} (\mathrm{HP}\{\phi(\bar{x})\})^2 + \lambda_2 \sum_{S_j} \sum_{\bar{x} \in S_j} (-1) \cdot C\{r(\bar{x})\} \\ &\quad + \lambda_3 E_\mathrm{compact}(S) + \lambda_4 E_\mathrm{disjointed}(S) + \lambda_5 E_\mathrm{complete}(S) \\ &= E_\mathrm{D}(D, r, \phi, S) + \lambda_1 E_\mathrm{S}(\phi, S) + \lambda_2 E_\mathrm{C}(r, S) + \lambda_3' E_\mathrm{segmentation}(S). \end{aligned} \tag{12}$$

$E_\mathrm{compact}(S)$ measures the compactness of the regions S_j of S. For example, this could be done by measuring the border length of all regions divided by their area. $E_\mathrm{disjointed}(S)$ formalizes the requirement that the S_j shold be mutually disjointed, so that S is a partition in the ideal case. This can be realized by defining it as $E_\mathrm{disjointed}(S) \propto \sum_{i \neq j} |S_i \cap S_j|$. The term $E_\mathrm{complete}(S)$ serves to ensure that the whole image area is covered by S. An appropriate term would be: *Image area* $- |\cup_{j=1}^{J} S_j|$.

In this case, the desired fusion result $r^*(\bar{x})$ is obtained by minimizing the energy functional E with respect to $r(\bar{x})$, $\phi(\bar{x})$, and S:

$$E(r^*, \phi^*, S^*, D) := \min_{r,\phi,S}\{E\}. \tag{13}$$

In most practical cases, additional information about the object to be imaged is available, e.g. the number J of regions S_j, their approximate shape and size, their arrangement, and material qualities like reflection properties or texture. This information should also be taken into account by adding adequate terms to eq. (12).

3.2.1 Connection with Bayesian statistics

In analogy to section 3.1.1, the joint PDF $p(D, r, \phi, S)$ can be obtained by inserting eq. (12) into eq. (2), and an appropriate normalization leads to the a posteriori PDF $p(r, \phi, S|D)$ for the unknown given observed data:

$$\begin{aligned}
p(D, r, \phi, S) &\propto \mathrm{e}^{-\frac{E_\mathrm{D}}{T}} \cdot \underbrace{\mathrm{e}^{-\frac{\lambda_1 E_\mathrm{S}}{T}} \cdot \mathrm{e}^{-\frac{\lambda_2 E_\mathrm{C}}{T}}}_{\text{a priori PDF}} \cdot \underbrace{\mathrm{e}^{-\frac{\lambda_3' E_{\text{segmentation}}}{T}}}_{\text{a priori PDF of segmentation}} \\
p(D, r, \phi, S) &= \underbrace{p(D|r, \phi, S)}_{\text{likelihood}} \cdot p(r, \phi|S) \cdot \underbrace{p(S)}_{\text{a priori PDF of segmentation}} \\
&\propto \underbrace{p(r, \phi, S|D)}_{\text{a posteriori PDF}} .
\end{aligned} \tag{14}$$

Now, the textural fusion result r as well as the segmentation S are signals of interest, and ϕ is again a meaningful nuisance parameter. The structure of the energy functional gives rise to interpret the Gibbs PDF according to the factorization of the second row of eq. (14). The factor $p(S)$ is the a priori PDF of the segmentation S. It conveys knowledge about the properties that a "good" segmentation should have. The other two factors mean the same as in eq. (8), but now they are conditional on the segmentation S.

3.2.2 Implementation strategy

Unfortunately, the minimization of eq. (12) is computationally even more expensive than the minimization of eq. (5). Consequently, a more efficient implementation strategy is needed in this case as well.

The factorization of eq. (14) suggests to optimize the energy functional eq. (12) by firstly performing a separate minimization of the terms $E_{\text{segmentation}}(S)$ concerning the segmentation task, and subsequently minimizing the other energy terms within each region S_j with the approach for edgeless objects given in subsection 3.1.2. The first stage of this efficient implementation consists in performing a segmentation of the image series. The result of the segmentation step is an image partition which provides regions containing no edges. Thus, within each of these regions it is resonable to apply the fusion methods proposed in section 3.1.

The choice of a suitable strategy for performing the segmentation task highly depends on the object geometry as well as the surface texture and cannot be dealt with in detail here. However, for an important class of metallic objects consisting of plane faces, an illumination-based segmentation method will be presented in section 3.2.3. For a more general discussion on image segmentation, we refer to [8].

Next, a fusion of the image series according to section 3.1 has to be performed for each of the regions obtained in the segmentation step.

Finally, the provisional fusion results have to be merged to one or more final fusion results. To obtain a high-quality image of the object texture, the images resulting from fusion within each of the regions are combined taking the segmentation result into account. In our case, an additional symbolic fusion result is obtained by extracting the edges of the partition resulting from the segmentation step. Thus, this higher level fusion result contains the edges of the object.

3.2.3 Illumination-based segmentation

In this section, an illumination-based segmentation method will be presented which is based on the maximization of a goodness function by means of varying the illumination direction of a parallel lighting device. This method is especially suitable for segmenting images of metallic objects showing several homogeneously textured plane faces.

In contrast to the former discussion, now we are faced with an *active vision* problem dealing with optimization of *each single image* during the image acquisition stage by varying the illumination direction. To distinguish this consideration from the problem of image fusion, the image quality of single images is specified by a goodness function G, which could also be interpreted as a kind of inverse energy function E.

To obtain the boundaries of each surface, directional light-field illumination is used. By formulating a goodness function, the usefulness of an image for being used in the segmentation process can be assessed. In this case the aim of image acquisition is to obtain a series of nearly binary images so that each face of the object is contained with high intensities in just one image of the series, whereas the other faces should show dark intensity levels.

A suitable goodness function for many practical applications is:

$$G(d) = G_{\text{binary}}(d) + \lambda_1 \cdot G_{\text{size}}(T\{d\}) + \lambda_2 \cdot G_{\text{compactness}}(T\{d\}), \tag{15}$$

where $T\{\}$ represents a threshold operator, and the λ_i denote weighting factors. The term G_{binary} measures how far the image d contains binary information by penalizing grey levels close to the average grey level μ_g. An appropriate expression for evaluating the binarity of an image is: $\sum_i (d(\bar{x}_i) - \mu_g)^2$. The second term G_{size} serves for rating the size of the regions obtained by segmenting the image d by means of the threshold operator $T\{\}$. In most inspection tasks, a priori knowledge is available concerning the approximate size of the regions of the object being imaged. In such cases, the difference between the expected region size and the actual region size of the image d can be used to define a suitable distance measure. The third term provides for a high compactness of the regions in the segmentation result. To implement this requirement, the area of all regions obtained by segmentation of d by the operator $T\{\}$ can be divided by the sum of their border lengths.

By maximizing the goodness function eq. (15) with respect to the direction (φ, θ) of a parallel lighting device, a nearly binary image is obtained showing one face of the object to be segmented with high intensities. To segment the remaining image area, pixels belonging to a surface region which has already been detected are excluded from the further segmentation process. This procedure has to be repeated until all regions have been detected or until all image areas have been assigned to a region.

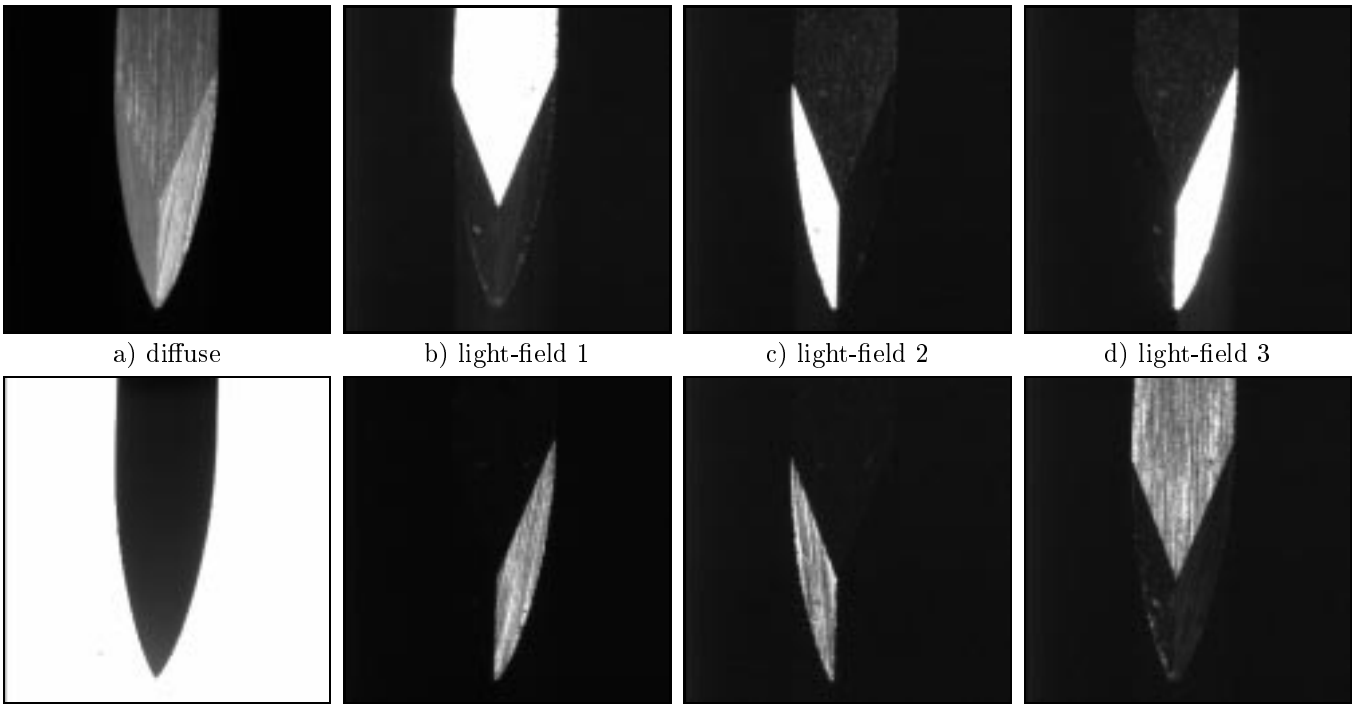
The performance of the illumination-based segmentation method is demonstrated with images of a needle for blood sugar tests. Such objects show an optimized form consisting of several planar grinded surfaces delimited by sharp edges. In Fig. 8a, an image of such a needle obtained with diffuse lighting is shown. To allow a reliable quality assessment, high-quality images of their surfaces have to be obtained, and their edges have to be measured. However, this information can hardly be extracted by using diffuse illumination; see Fig. 8a.

By maximizing the goodness function eq. (15) as explained above, the images Figs. 8b–e were acquired. It can be clearly seen that the information contained in these images is practically binary. Moreover, since the regions obtained by binarization of the images coincide with the faces of the needle, a segmentation of the object, and the extraction of its edges based on these images is straight forward.

In a further step, the illumination direction was varied so that a high contrast was obtained within each of the regions obtained after thresholding the images Figs. 8b–d. The resulting images are depicted in Figs. 8f–h. It can be clearly seen that each of the surfaces shown in these images shows a higher contrast than the same faces in the diffusely illuminated image Fig. 8a.

4 EXPERIMENTAL RESULTS

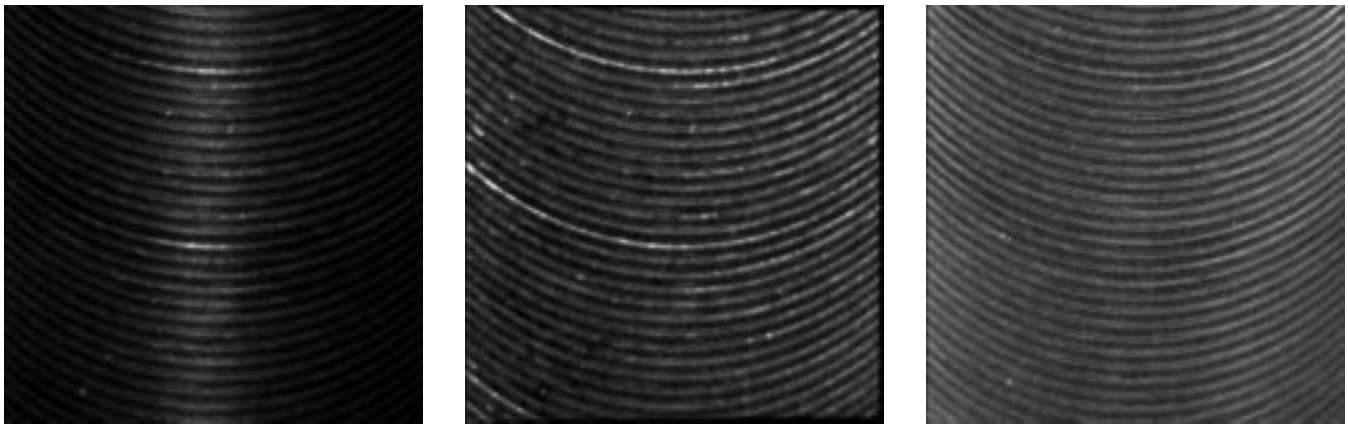
In this section, experimental results of the image fusion methods discussed in section 3 are presented and compared with images which can be obtained without data fusion. To obtain image series, an automated system was set up which consists of a flexible illumination module, and a commercial macroscope. All images throughout this paper were digitized with 512×512 pixels, and 8 bit grey levels.



a) diffuse b) light-field 1 c) light-field 2 d) light-field 3

e) back-lighting f) light-field 4 g) light-field 5 h) light-field 6

Figure 8. Needle for blood sugar tests: a) diffuse illumination; b) back-lighting; c–e) light-field illumination.



a) parallel lighting b) fusion result c) diffuse lighting

Figure 9. End-milling surface: a) image no. 8 of the illumination series; b) fusion result (criterion C: variance in a 5×5-neighbourhood; smoothing of $\tilde{\varphi}(\bar{x})$ with a binomial filter of size 49×49); c) diffuse lighting.

In Fig. 9a, an image of an illumination series consisting of 20 images ($\Delta\varphi \approx 5.6°$) of an end-milling texture can be seen. Due to illumination, the texture shows a bright stripe.[4] Since the object contains neither profile nor textural edges, the method described in section 3.1 was used to fuse the image series. C was chosen as the local grey-level variance in a neighbourhood of size 5×5. The illumination map $\tilde{\varphi}(\bar{x})$ was smoothed with a circular binomial filter with an impulse response of 49×49 pixels. In the fusion result Fig. 9b, the whole surface is illuminated much better, hitherto hidden details become visible, and the stripe-like inhomogeneity can no longer be recognized. By comparison of the fusion result with the same surface illuminated diffusely (Fig. 9c), it can be stated that in the fusion result the grooves are contained with much higher contrast. Thus, the surface quality can be assessed more accurately.[3]

a) diffuse lighting b) symbolic fusion result: edges c) data fusion result: texture

Figure 10. Needle for blood sugar tests: a) diffuse lighting; b) symbolic fusion result (object edges); c) data fusion result (grinding texture).

Next, the results of fusing images of objects containing edges according to section 3.2 will be demonstrated. Fig. 10b shows the segmentation result of an image series of a needle for blood sugar tests; see also section 3.2.3. To obtain this result, the images shown in Figs. 8b–e were binarized using a constant threshold. The resulting images were eroded morphologically by a disc-shaped structuring element of radius $r = 3$ pixels, and subsequently the images were combined additively. The region edges were obtained by computing the morphological skeleton of the background. Hereafter, a component labelling[1] was done, and for each region obtained in this process the image of the illumination series showing the best contrast within this area was faded into the second fusion result Fig. 10c, which shows the surface texture with significantly higher contrast than in the case of diffuse lighting; see Fig. 10a.

5 SUMMARY

In this paper, a general methodology for describing and performing fusion of image series is discussed. The fusion problem is formalized by defining an energy functional which characterizes all knowledge and wishes concerning involved quantities in a monotonous way so that minimizing the energy functional delivers the optimal fusion result with respect to the assumptions met. It is shown how this energy formulation can be interpreted from the point of view of Bayesian statistics by introducing Gibbs probabilities.

This abstract theory is applied to image series of metallic objects with varying illumination. Besides other results, high-quality images are obtained which could not have been acquired physically with only one image. Depending on the kind of objects to be imaged, two cases are distinguished, namely objects without and with edges. In the latter case, additionally the fusion aims at detecting the boundaries of different object areas.

Since the direct minimization of the energy functionals is prohibitively computationally expensive, efficient methods which lead to convincing results are derived for both cases. The usefulness of the approaches is demonstrated by means of several practically relevant examples.

REFERENCES

1. D. H. Ballard and C. M. Brown, *Computer Vision*, Prentice-Hall, Englewood Cliffs, 1982.
2. J. Beyerer, "Suppression of stochastically placed, straight toolmarks to enhance objects and defects", Technisches Messen **59** (10), pp. 389–397, 1992.
3. J. Beyerer, *Analyse von Riefentexturen*, VDI-Verlag, Reihe 8, Nr. 390, Düsseldorf, 1994.
4. J. Beyerer and F. Puente León, "Suppression of inhomogeneities in images of textured surfaces", *Optical Engineering* **36** (1), pp. 85–93, 1997.

5. B. V. Dasarathy, "Sensor Fusion Potential Exploitation – Innovative Architectures and Illustrative Applications", Proc. of the IEEE **85** (1), pp. 24–38, 1997.
6. J. J. Clark and A. L. Yuille, *Data Fusion for Sensory Information Processing Systems*, Kluwer, Boston, 1990.
7. B. Jähne, *Digitale Bildverarbeitung*, Springer, Berlin, 1993.
8. A. Kaup, *Modelle zur regionenorientierten Bildbeschreibung*, VDI-Verlag, Düsseldorf, 1995.
9. S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by simulated annealing". *Science* **220**, pp. 671–680, 1983.
10. T. Lindeberg, *Scale-Space Theory in Computer Vision*, Kluwer Academic Publishers, Boston, 1994.
11. J. R. Maheshkumar, Vijay Veeranna, S. Sitharama Iyengar, Richard S. Brooks, "New computational technique for complementary sensor integration in detection-localization systems", *Optical Engineering* **35** (3), pp. 674–684, 1996.
12. K. V. Mardia, *Statistics of Directional Data*, Academic Press, London, 1972.
13. L. L. Scharf, *Statistical Signal Processing, Detection, Estimation, and Time Series Analysis*, Addison-Wesley Publishing Company, 1991.