Spatiotemporally Adaptive Estimation and Segmentation of OF-fields

H.-H. Nagel[1,2] and A. Gehrke[1]

[1] Institut für Algorithmen und Kognitive Systeme
Fakultät für Informatik der Universität Karlsruhe (TH)
Postfach 6980, D-76128 Karlsruhe / Germany
[2] Fraunhofer-Institut für Informations- und Datenverarbeitung (IITB)
Fraunhoferstr. 1, D-76131 Karlsruhe / Germany
Tel. +49-721-6091-210; Fax +49-721-6091-413; email hhn@iitb.fhg.de

Abstract. A grayvalue *structure tensor* provides knowledge about a local grayvalue variation. This knowledge can be used to devise a *spatiotemporally adaptive* optic flow estimation process. Such an adaptive estimation lowers the level at which the resulting optic flow (OF) field is disturbed by noise and estimation artefacts. This in turn substantially simplifies the analysis of remaining – often subtle – effects which easily jeopardize a 'naive' segmentation approach. Appropriate treatment of such effects eventually results in a basically simple, but nevertheless surprisingly robust segmentation approach. Various stages of this approach are illustrated by examples for the extraction of moving vehicle images from a digitized road intersection video-sequence.

1 Introduction

In order to extract a weak, straight line edge segment from a noisy image, it is advantageous in general to employ a gradient filter with a suitably elongated support – see, e.g., [10]. The low-pass contribution to the derivative filter will then act *more along* the edge segment *than across* it. Evidently, such an approach implies a hen-and-egg dilemma: in order to extract the edge segment, one has to know its orientation, and in order to determine its orientation, one has to know the edge segment.

An analogous problem appears in the case where one has to *segment* an optic flow (OF) field to be estimated in the first place: segmentation of an OF-field requires the detection of discontinuities in an estimated vector field, i. e. it implies at first glance some kind of derivative operation, followed by a detection step which decides whether the local change appears significant enough to decide in favor of a discontinuity in the OF-field. Alternatively, one might consider region growing from some 'seed region'. For both alternatives, the hen-and-egg problem pops up in a different disguise: either location *and* structure of a segment boundary element or a seed region *together* with some appropriate stopping criterion for region growing have to be determined *simultaneously* in the OF vector field.

Optic flow – the apparent shift velocity of a grayvalue structure in the image plane – will be considered here as a threedimensional vector $\mathbf{u} = (u_1, u_2, 1)^T$ in an (x,y,t)-space formed by the image plane coordinates $\mathbf{x}' = (x, y)^T$ and time t. A location in this (x,y,t)-space will be given by the threedimensional vector $\mathbf{x} = (x, y, t)^T$. OF is usually *estimated* based on the postulate that the grayvalue $g(x, y, t)$ is locally stationary as a function of x, y, and t: $dg(x, y, t) = 0$. This results in the so-called 'Optic Flow Constraint Equation (OFCE)' [4]:

$$(g_x, g_y, g_t)\,\mathbf{u} = (\nabla g)^T \mathbf{u} = 0 \quad \text{with} \quad g_x = \frac{\partial g(x, y, t)}{\partial x}\,, \quad \text{etc.} \;. \tag{1}$$

Due to space limitations, we refer to [1] for a general review of optic flow estimation. In the sequel, we first concentrate on our approach in order to provide a frame of reference for a discussion of related research in the concluding section.

2 The Grayvalue-Local-Structure-Tensor GLST

The gradient operator ∇ for the computation of $\nabla g(\mathbf{x})$ will be realized by a convolution of g(**x**) with partial derivatives of a trivariate Gaussian G(**x**) given by

$$G(\mathbf{x}) = \frac{1}{(2\pi)^{3/2}\sqrt{|\Sigma|}}\, e^{-\frac{1}{2}\mathbf{x}^T \Sigma^{-1} \mathbf{x}} \;. \tag{2}$$

The covariance matrix Σ is initially set to

$$\Sigma_{\text{init}} = \begin{pmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_y^2 & 0 \\ 0 & 0 & \sigma_t^2 \end{pmatrix} \tag{3}$$

with values for the standard deviations σ_x in the x-direction, σ_y in the y-direction, and σ_t in the t-direction chosen by the experimenter on the basis of a-priori knowledge.

Due to the fact that (1) is underdetermined at a single location **x**, one modifies the estimation postulate into the requirement that the squared magnitude of the OFCE, averaged over some local environment, should be minimized by a suitable choice of **u**:

$$\overline{|(\nabla g)^T \mathbf{u}|^2} = \overline{((\nabla g)^T \mathbf{u})^T\, ((\nabla g)^T \mathbf{u})} = \mathbf{u}^T\, \overline{\nabla g(\mathbf{x})\, (\nabla g(\mathbf{x}))^T}\, \mathbf{u} \stackrel{!}{=} \min_{\mathbf{u}} \;. \tag{4}$$

According to [7], let the spatiotemporal Gaussian introduced by (2) with covariance matrix 2Σ describe the local environment around a location **x**. We then define the location-dependent 'Grayvalue-Local-Structure-Tensor (GLST)' as

$$GLST(\mathbf{x}) = \int_{-\infty}^{+\infty} d\boldsymbol{\xi}\; \frac{\nabla g(\boldsymbol{\xi} - \boldsymbol{x})\, (\nabla g(\boldsymbol{\xi} - \boldsymbol{x}))^T}{(2\pi)^{3/2}\sqrt{2|\Sigma|}}\, e^{-\frac{1}{4}(\boldsymbol{\xi} - \boldsymbol{x})^T \Sigma^{-1} (\boldsymbol{\xi} - \boldsymbol{x})} \;. \tag{5}$$

By definition, the GLST is positive-semidefinite. $\mathrm{GLST}_{\mathrm{init}}$ denotes a Grayvalue-Local-Structure-Tensor computed by using Σ_{init} as given by (3).

Let $\mathbf{e}_{GLST_{\mathrm{init}},i}$ denote the i-th eigenvector of $\mathrm{GLST}_{\mathrm{init}}$ and $\lambda_{\mathrm{init},i}$ the corresponding eigenvalue, with $\lambda_{\mathrm{init},1} \geq \lambda_{\mathrm{init},2} \geq \lambda_{\mathrm{init},3} \geq 0$. $\mathrm{GLST}_{\mathrm{init}}$ can then be written in the form

$$GLST_{\mathrm{init}} = (\mathbf{e}_{GLST_{\mathrm{init}},1},\ \mathbf{e}_{GLST_{\mathrm{init}},2},\ \mathbf{e}_{GLST_{\mathrm{init}},3}) \begin{pmatrix} \lambda_{\mathrm{init},1} & 0 & 0 \\ 0 & \lambda_{\mathrm{init},2} & 0 \\ 0 & 0 & \lambda_{\mathrm{init},3} \end{pmatrix} \begin{pmatrix} \mathbf{e}^T_{GLST_{\mathrm{init}},1} \\ \mathbf{e}^T_{GLST_{\mathrm{init}},2} \\ \mathbf{e}^T_{GLST_{\mathrm{init}},3} \end{pmatrix} . \tag{6}$$

According to (4), optic flow is given as the projection of the GLST-eigenvector $\mathbf{e}_{GLST,3}(\mathbf{x}) = (e_{GLST,31},\ e_{GLST,32},\ e_{GLST,33})^T$, which corresponds to the smallest eigenvalue $\lambda_{GLST,3}(\mathbf{x})$ of $\mathrm{GLST}(\mathbf{x})$, into the image plane. This is equivalent to the hypothesis that the eigenvector related to the smallest GLST-eigenvalue points into the direction of smallest *temporal* change, which in turn is ascribed to a grayvalue structure shifting smoothly in the (x, y, t)-space. The projection of this eigenvector into the image plane will be normalized to a unit value for the third (i. e. temporal) component in order to remain compatible with the definition given in Sect. 1:

$$\mathbf{u}(\mathbf{x}) = \left(\frac{e_{GLST,31}}{e_{GLST,33}} , \frac{e_{GLST,32}}{e_{GLST,33}} , \frac{e_{GLST,33}}{e_{GLST,33}} = 1 \right)^T . \tag{7}$$

3 An Adaptive Filter Based on the GLST

Let us substitute $\mathrm{GLST}_{\mathrm{init}}(\mathbf{x})$ for Σ^{-1} in (2) and recompute the gradient of $\mathrm{g}(\mathbf{x})$. A large eigenvalue of $\mathrm{GLST}_{\mathrm{init}}(\mathbf{x})$ will severely restrict the extent over which grayvalues contribute to the partial derivative in the corresponding direction, whereas the low-pass action implied by the Gaussian will extend much more in the directions corresponding to $\mathbf{e}_{GLST_{\mathrm{init}},2}$ and $\mathbf{e}_{GLST_{\mathrm{init}},3}$. We thus obtain exactly the desired effect of *less* low-pass filtering in the direction of largest local grayvalue change than in the directions perpendicular to it. In general, the amount of low-pass filtering in the direction corresponding to an eigenvalue $\lambda_{\mathrm{init},i}$ of $\mathrm{GLST}_{\mathrm{init}}(\mathbf{x})$ will be determined by the magnitude of this eigenvalue.

We may exploit this information in order to *improve the derivative operation.* Use of $\mathrm{GLST}_{\mathrm{init}}(\mathbf{x})$ instead of a constant Σ^{-1} during a *recomputation* of the gradient at each image location $\mathbf{x}$ automatically adapts the low-pass action of the Gaussian in the partial derivative operators to the local grayvalue structure.

3.1 Delimiting the Extent of an Adaptive Convolution Mask

There occur problems, however, unless we proceed with caution. In rather homogeneous image regions, the area of support can locally grow to a size where a

standard implementation of a Finite Impulse Response (FIR) filter by a digital convolution operation may begin to fail. An uncontrolled adaptation can result, for example, in excessive mask sizes. On the other hand, it may well happen that an eigenvalue becomes very large in image areas with particularly strong grayvalue transitions. As a consequence, the Gaussian in (2) will decrease so sharply that the sampling theorem may be violated upon conversion of the resulting filter to a digital version for a sampling grid given by the - already - digitized image sequence. We are thus forced to restrict the mask size to be used, without jeopardizing the desired adaptation effect whenever the eigenvalues of $\mathrm{GLST}_{\mathrm{init}}(\mathbf{x})$ remain compatible with the minimal and maximal mask size provided by the available implementation of a FIR digital convolution operation.

Let $\sigma^2_{\mathrm{minsize}}$ be a parameter which forces a diagonal element in the covariance matrix for the Gaussian in (2) to remain compatible with the smallest admissible mask size. $\sigma^2_{\mathrm{maxsize}}$ should be defined such that $\sigma^2_{\mathrm{minsize}} + \sigma^2_{\mathrm{maxsize}}$ determines the largest admissible mask size. Let

$$U = (\ \mathbf{e}_{GLST_{\mathrm{init}},1},\quad \mathbf{e}_{GLST_{\mathrm{init}},2},\quad \mathbf{e}_{GLST_{\mathrm{init}},3}\ ) \tag{8}$$

denote the 3D rotation matrix in the (x, y, t)-space which aligns the coordinate axes with the eigenvector directions of $\mathrm{GLST}_{\mathrm{init}}(\mathbf{x})$. Denoting the 3x3 unit matrix by I and using $I = U(\mathbf{x})U^T(\mathbf{x})$, we introduce a *locally adapted* covariance matrix $\Sigma(\mathbf{x})$ as

$$\Sigma(\mathbf{x}) =$$

$$U(\mathbf{x}) \left(\sigma^2_{\mathrm{minsize}}\, I + \begin{pmatrix} \frac{\sigma^2_{\mathrm{maxsize}}}{1+\sigma^2_{\mathrm{maxsize}}\lambda^w_1(\mathbf{x})} & 0 & 0 \\ 0 & \frac{\sigma^2_{\mathrm{maxsize}}}{1+\sigma^2_{\mathrm{maxsize}}\lambda^w_2(\mathbf{x})} & 0 \\ 0 & 0 & \frac{\sigma^2_{\mathrm{maxsize}}}{1+\sigma^2_{\mathrm{maxsize}}\lambda^w_3(\mathbf{x})} \end{pmatrix} \right) U^T(\mathbf{x}) \tag{9}$$

with $\lambda^w_i(\mathbf{x}) = \lambda_{\mathrm{init},i}(\mathbf{x})$, $i \in \{1,2,3\}$. For experiments to be discussed shortly, we have chosen $\sigma^2_{\mathrm{minsize}} = 0.5$ and $\sigma^2_{\mathrm{maxsize}} = 4.0$. Let us assume for the moment that $\mathrm{GLST}_{\mathrm{init}}(\mathbf{x})$ has an eigenvalue $\lambda_{\mathrm{init},3}$ close to zero due to lack of grayvalue variation in the corresponding direction. In the limit of zero for $\lambda_{\mathrm{init},3}$, the third eigenvalue of $\Sigma(\mathbf{x})$ will become $\sigma^2_{\mathrm{minsize}} + \sigma^2_{\mathrm{maxsize}}$, thereby restricting a digitized version of the resulting Gaussian to the chosen maximal mask size. In case of a very *strong* straight line grayvalue transition front in the vicinity of image location $\mathbf{x}$, the first eigenvalue $\lambda_{\mathrm{init},1}(\mathbf{x})$ of $\mathrm{GLST}_{\mathrm{init}}(\mathbf{x})$ will force the second contribution to the corresponding eigenvalue of $\Sigma(\mathbf{x})$ in (9) to be small, delimiting the sum of both terms from below by a suitably chosen $\sigma^2_{\mathrm{minsize}}$.

Equation (9) will only yield acceptable results for the eigenvalues of $\Sigma(\mathbf{x})$, if the eigenvalues of $\mathrm{GLST}_{\mathrm{init}}$ range between $\sigma^2_{\mathrm{minsize}}$ and $\sigma^2_{\mathrm{maxsize}}$. If most of the initial eigenvalues, however, are far outside of this interval - by experience we have seen that most of the eigenvalues from $\mathrm{GLST}_{\mathrm{init}}$ range above 50 - the eigenvalues of $\Sigma(\mathbf{x})$ which are determined according to (9) are all close to $\sigma^2_{\mathrm{minsize}}$

and therefore all are similar. Normalizing the eigenvalues by $\frac{1}{3}trace(GLST(\mathbf{x}))$ – i. e. setting $\lambda_i^w(\mathbf{x}) = \frac{\lambda_{\mathrm{init},i}(\mathbf{x})}{\frac{1}{3}trace(GLST_{\mathrm{init}}(\mathbf{x}))}$, $i \in \{1,2,3\}$ – forces the eigenvalues of $\Sigma(\mathbf{x})$ to vary between reasonable limits.

3.2 Estimation of an Improved GLST

The matrix $\Sigma(\mathbf{x})$ given by (9) describes the local grayvalue variation at location $\mathbf{x}$ in such a manner that the effects of noise on gradient computation will be reduced in comparison with that based on (3), since the low-pass filter action along the directions of smaller gray value variations will be increased. In addition, the influence of strong neighboring grayvalue structures is expected to be reduced due to a limitation of the filter extent in the direction of strong changes, which should lead to a less distorted gradient estimation at location $\mathbf{x}$. The choices for the parameters $\sigma^2_{\mathrm{minsize}}$ and $\sigma^2_{\mathrm{maxsize}}$ facilitate to incorporate a-priori knowledge about the spatiotemporal extent of semantically relevant grayvalue changes in an image sequence.

We may now exploit the knowledge about the local spatiotemporal grayvalue variation at image sequence location $\mathbf{x}$ in order to *recompute* the 'Grayvalue-Local-Structure-Tensor' as defined by (5), but this time using the spatiotemporally adaptive $\Sigma(\mathbf{x})$ given by (9) in the Gaussian of (2) instead of the constant Σ given by (3).

4 Adaptive Estimation of Optic Flow: Special Cases

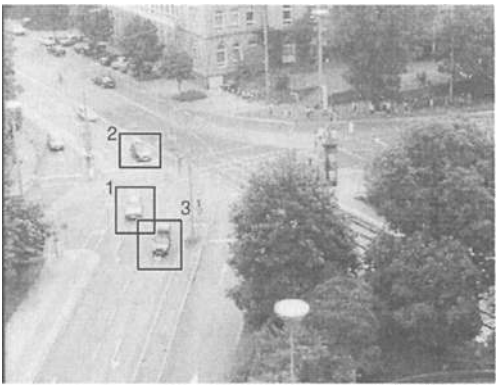

Fig. 1. Image frame No. 424 from the image sequence 'dt_v'. The optical flow fields for the three selected windows will be illustrated subsequently in more detail.

In certain situations, a determination of the OF-vector according to (7) will not yield acceptable results. One condition occurs obviously at a location $\mathbf{x}$

where the local grayvalue is distributed so homogeneously that *all* eigenvalues of GLST practically vanish. This case can simply be detected by comparing trace(GLST(**x**)) – which represents the squared norm of the gradient averaged around **x** – against a minimal size threshold *min_norm_of_GLST*. We shall call pixel positions, where this threshold is *not* surpassed, as *neutral*. At such a location, it will not be possible to reliably detect any motion at all. A forteriori, it thus is not possible to decide at such a location whether the image of a scene surface element depicted at **x** moves relative to the recording camera against a non-contrasting background.

The convention which underlies (7) will lead to counter-intuitive results for a different situation, too. At a location **x** with a locally dominant *straight line* grayvalue transition front, all *spatial* gradients in the neighborhood around **x** point more or less into the same direction. Since $\Sigma(\mathbf{x})$ adapts to this local grayvalue structure, the GLST computed using $\Sigma(\mathbf{x})$ will have a very small component *perpendicular* to the prevailing *spatial* grayvalue gradient direction. It happens occasionally that this spatial variation is smaller than the temporal one, even if the local grayvalue structure shifts smoothly in the image plane with time. In such situations, one is confronted with the fact that the eigenvector corresponding to the smallest eigenvalue of GLST(**x**) *does not reflect the* apparent *shift* of a grayvalue structure *with time.*

We can detect this special condition by inspection of the third, i. e. temporal, component $e_{GLST,33}(\mathbf{x})$ of the eigenvector related to the smallest eigenvalue $\lambda_{GLST,3}(\mathbf{x})$ of GLST(**x**). If $e_{GLST,33}(\mathbf{x}) \leq e_{33_{\mathrm{minflow}}}$, we consider the smallest eigenvalue to represent a *purely spatial* local grayvalue variation. We denote such a pixel as *spatially_tangent* since the eigenvector $\mathbf{e}_{GLST,3}(\mathbf{x})$ corresponding to the smallest eigenvalue is oriented in this case *within the image plane* to be tangential to an edge: it points into a direction with an essentially constant local grayvalue distribution. The characteristics of such a spatially_tangent location differ from those of a neutral one by exhibiting *at least one* eigenvalue significantly different from zero – see Fig. 2(c). Since the three eigenvectors of GLST(**x**) are mutually perpendicular to each other, any temporal variation at a ST-pixel must then be reflected by a vector in the normal plane to $\mathbf{e}_{GLST,3}(\mathbf{x})$. The direction of the smallest grayvalue variation with a temporal component will thus be given by the second eigenvector $\mathbf{e}_{GLST,2}(\mathbf{x})$ which corresponds to the middle eigenvalue.

5 Pixel Assignment to Categories

As a preparatory step for the segmentation of an OF-field, we first categorize each pixel by an hierarchical classification procedure according to its local spatiotemporal grayvalue structure. As will be seen, subtle characteristics of this spatiotemporal structure may substantially influence subsequent clustering, split, and merge steps.

The first test at a pixel location **x** determines if trace(GLST(**x**)) $\leq$ *min_norm_of_GLST* is true: such a pixel is assigned to the category *neutral* (*N*). In the case where trace(GLST(**x**)) $>$ *min_norm_of_GLST*, we can be sure that at

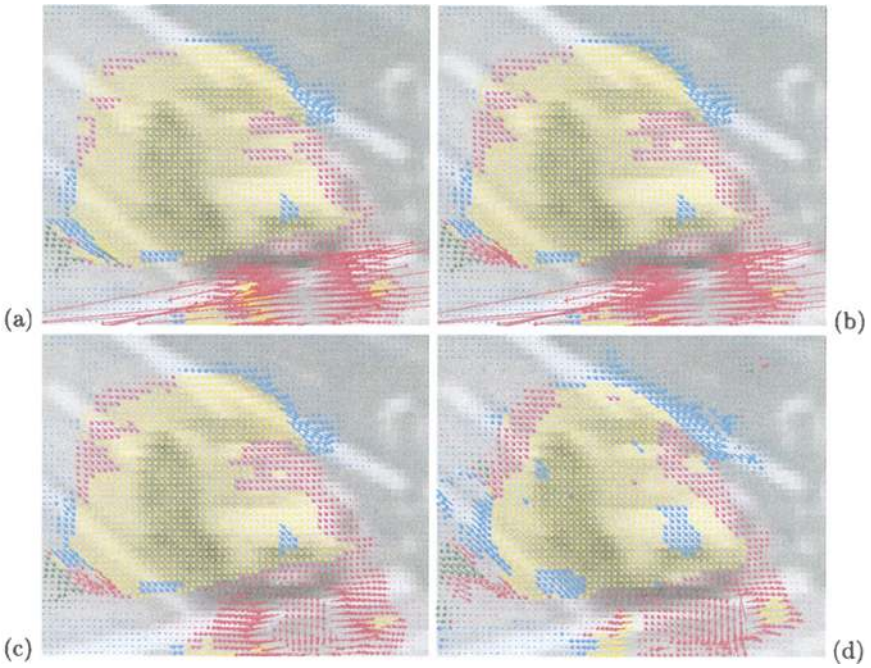

Fig. 2. Enlargement of window 2 in Figure 1. Overlaid are the OF-vectors wich are color-coded as follows: The yellow vectors show *regular_optical_flow (R_OF)*. The vectors at the *OF_discontinuity (OF_D)* locations are color-coded in red and at the *dominant_gradient_direction (DGD)* locations in turquoise. If a pixel could be assigned to the category *OF_discontinuity (OF_D)* and simultaneously complied with the criterion for the category *dominant_gradient_direction (DGD)*, it has nevertheless been assigned to the category *OF_discontinuity (OF_D)* and is, therefore, shown in red. The assignment of pixels to these categories is explained in Sect. 5. (a): The optic flow was estimated based on $\mathrm{GLST}_{\mathrm{init}}$. Note that the pixels, which have been assigned to the category *dominant_gradient_direction (DGD)*, are located at roadway markings with a homogeneous grayvalue structure in spatial direction. (b): Compared to panel (a), we have lowered the discontinuity_threshold. The number of locations assigned to category *OF_D* increases, thereby reducing the areas covered by accepted regular optic flow vectors. (c): Analogous to (b), but treating optical flow vectors at *spatially_tangent (S_T)* locations according to Sect. 4. The concerned pixels are mainly located at the roadway marking in front of the vehicle which exhibit a more or less homogenous grayvalue within the marking. At these locations, therefore, the OF-vector point along the direction of homogenous grayvalue, if we used the eigenvector corresponding to the smallest eigenvalue for the OF estimation – see panel (b). (d): Analogous to panel (c), but the overlaid estimated optical flow vectors are based on $\mathrm{GLST}(x)$ as described in Sect. 3 rather than based on $\mathrm{GLST}_{\mathrm{init}}$. The number of pixels at the depicted vehicle image, which are *incorrectly* assigned to the category *OF_D*, is *smaller* than indicated in panel (c) although we use the same low discontinuity_threshold value as in panel (b) and (c). Simultaneously, the 'discontinuity wall' around the vehicle image in panel (d) contains less holes than the 'discontinuity wall' depicted in panel (c).

least one eigenvalue of GLST(**x**) differs significantly from zero. It thus is justified to speak about the *smallest* eigenvalue $\lambda_{GLST,3}(\mathbf{x})$ of GLST(**x**).

In case $e_{GLST,33}(\mathbf{x}) > e_{33_{\text{minflow}}}$ is true, the pixel at image location **x** is assigned to category *OF_discontinuity (OF_D)*, provided

$$\frac{\lambda_{GLST,3}(\mathbf{x})}{\frac{1}{3}trace\ (GLST(\mathbf{x}))} > discontinuity\_threshold\ . \tag{10}$$

Figure 2 shows the OF-vectors at these locations color-coded in red.

This test captures the observation that the temporal change – represented here by $\lambda_{GLST,3}(\mathbf{x})$ – is larger than deemed acceptable for a smooth shift of a grayvalue structure in the image plane. A large $\lambda_{GLST,3}(\mathbf{x})$ signifies that there is no direction with small grayvalue variation. According to (1) this means that there exist different OF-vectors in the local region considered. Please note that the left hand side of (10) varies by definition between 0 and 1 which allows to restrain the choice of a threshold to this range.

If, however, this location must be treated as *spatially_tangent (S_T)*, because $e_{GLST,33}(\mathbf{x}) \leq e_{33_{\text{minflow}}}$, the second eigenvalue $\lambda_{GLST,2}(\mathbf{x})$ indicates the smallest *temporal* change. Optic flow must then be defined as

$$\mathbf{u}(\mathbf{x}) = \left(\frac{e_{GLST,21}}{e_{GLST,23}}\ ,\ \frac{e_{GLST,22}}{e_{GLST,23}}\ ,\ \frac{e_{GLST,23}}{e_{GLST,23}} = 1 \right)^T\ , \tag{11}$$

see Fig. 2(c). In this case, the test for a discontinuity in the optic flow field must be applied to the second eigenvalue, i. e. location **x** is assigned to category *OF_D* if

$$\frac{\lambda_{GLST,2}(\mathbf{x})}{\frac{1}{3}trace\ (GLST(\mathbf{x}))} > discontinuity\_threshold\ . \tag{12}$$

Among the remaining image plane locations which are neither *neutral (N)* nor *OF_discontinuity (OF_D)*, we still have to detect any potential bias due to the aperture problem which occurs whereever the grayvalue structure is dominated by a straight line grayvalue transition front: the second eigenvalue differs from zero, but only by a small amount which may be insufficient to facilitate a reliable estimation of optic flow. We detect such situations, therefore, by the following test:

$$\frac{\lambda_{GLST,2}(\mathbf{x}) + \lambda_{GLST,3}(\mathbf{x})}{\frac{2}{3}trace\ (GLST(\mathbf{x}))} \leq threshold_{dominant\_gradient\_direction}\ , \tag{13}$$

Again, the left hand side of this equation varies by definition between 0 and 1.

All pixels not assigned in the course of this test sequence to one of the categories *neutral (N)*, *OF_discontinuity (OF_D)*, or *dominant_gradient_direction (DGD)*, will be treated as *regular_optical_flow (R_OF)* locations. These surviving locations are depicted in Fig. 2 through the yellow vectors.

6 Segmentation of Estimated Optic Flow Fields

The basic approach consists in identifying image locations where the temporal grayvalue change does not become small enough to justify a classification of the local spatiotemporal grayvalue variation as a smooth *temporal shift* of a characteristic *spatial* grayvalue structure. Ideally, such OF_discontinuity-locations should form a 'wall of discontinuities' around the image of an object moving relative to the camera with a velocity different from that of its environment. We proceed in four basic steps:

1. Each pixel is assigned – according to Sect. 5 – to one of the following categories:
   - neutral (N),
   - OF_discontinuity (OF_D),
   - dominant_gradient_direction (DGD),
   - regular_optical_flow (R_OF).
2. Pixels which have been assigned to the same category are clustered into 4-connected components, using a standard run-length algorithm as described, e. g., in [3].
3. If the variance of OF-vectors within a connected-component suggests two or more significant clusters, the originally obtained connected-component is split in order to increase the homogeneity of OF-vectors within a region.
4. Certain combinations of the connected components resulting from previous steps are merged in order to improve a mask covering the image of an object which potentially moves in the scene relative to the recording video camera.

6.1 'Discontinuity Walls'

The approach outlined above should result in clusters of R_OF-locations, surrounded by 'walls' of OF_discontinuity-(OF_D-)locations. In particular, a 'discontinuity wall' is expected around the image region corresponding to the image of an object moving in the scene relative to the camera.

As is shown in Fig. 2(a), a number of breaches can be detected in the 'discontinuity wall' around the depicted vehicle image. We may lower the discontinuity_threshold in order to detect more discontinuity locations in the hope to close many, if not all, of these breaches. As Figure 2(b) shows, a reduction of the discontinuity threshold from 0.03 to 0.023 does indeed close some holes, but only at the cost of many additional false alarms, i. e. a lot of spurious OF_D-locations are marked.

Figures 2(a) and (b) also illustrates that choosing the eigenvector associated with the smallest eigenvalue of the GLST may result in OF-vectors which are incompatible with intuition: as can be seen in Figs. 2(a) and (b), OF-vectors in the vicinity of strong, straight line grayvalue transition fronts tend to be oriented tangentially to these transition fronts – even in cases where a transition front (or a shadow cast in that image area) moves more or less in the gradient direction. Figure 2(c) demonstrates the improvement if we subject not the smallest, but

the next larger eigenvalue to the discontinuity test in case of *(S_T)*-locations, using the same threshold as in Fig. 2(b).

So far, all depicted OF-vectors have been computed based on $GLST_{\text{init}}$. Figure 2(d) shows the result analogous to Fig. 2(c), but now computed on the basis of the locally adapted GLST(**x**) rather than on the basis of $\text{GLST}_{\text{init}}(\mathbf{x})$. It is seen that the 'discontinuity wall' contains less holes. Simultaneously, the number of spuriously detected discontinuities could be reduced, too.

6.2 Gaps at 'Dominant_Gradient_Direction'-Locations

The advantage of using this still very simple, but nevertheless more sensitive approach for the detection of discontinuities in the optic flow field consists not only in the fact that the number of 'breaches' in the 'discontinuity wall' is reduced. Even more important is the fact that a specific characteristic of the remaining gaps can be identified: the OF-vectors in gaps, which cause a connected-component determination to 'leak' into the background, do not differ significantly in magnitude and orientation from the acceptable ones covering the image of the moving object. One will notice, however, that the remaining weak points in the 'discontinuity wall' belong to the category *dominant_gradient_direction (DGD)* - see Fig. 2(d). This insight immediately suggests a remedy: in DGD-cases, it can not be decided reliably whether there is a discontinuity in an OF-field due to a straight line segment which delimits the object image in a direction parallel to the image motion, since the optic flow component in the direction of a strong gradient vanishes for locations which either belong to the image of the moving object or to the background occluded by it. The more or less homogeneous nature of the occluded background (or, analogously, of the occluding foreground) results in a dominant gradient direction due to the border between foreground and background. Since there is no motion perpendicular to this border, there is no way to check for a discontinuity of optic flow in this direction.

We thus decided to treat such DGD-situations as 'potential discontinuity' locations and to mark them separately. As a consequence, the connected-component subprocess will be prevented to extend across such a barrier and we obtain surprisingly clean masks covering the images of moving objects – *without* having to introduce any threshold on magnitude or orientation differences.

6.3 Splitting Connected-Components with a Significantly Inhomogeneous OF-Vector distribution

Optic flow in the background area differs significantly from flow vectors associated with a moving vehicle although we avoided so far to exploit a-priori knowledge about the fact that the camera remains stationary with respect to the background. We rather proceed on the *weaker* assumption that optic flow within the image area corresponding to a vehicle image differs significantly from that outside of this area.

The scatter plot in Fig. 4 shows the distribution of u_1 and u_2 components of optic flow within the right connected-component from Fig. 3(b). Two clusters

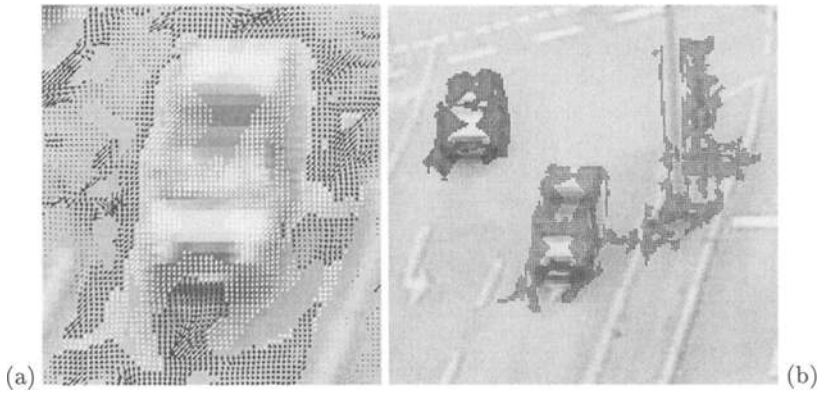

Fig. 3. (a): The same window as window no. 3 from Fig. 1, but recorded 80 msec later, has been enlarged and superimposed with OF-vectors estimated at the pixels which belong to the categories OF_D (dark vectors) and R_OF (bright vectors). One notices a gap at the right border of the depicted vehicle image in the 'discontinuity wall': white *R_OF*-vectors leak through the dark 'discontinuity wall' around the vehicle image into the background region. (b): A greater image region cropped around the area shown in panel (a). Superimposed are the masks for which all pixels at *regular_optical_flow (R_OF)* locations are connected, provided these locations are completely surrounded by pixels at *neutral (N)*, *OF_discontinuity (OF_D)*, or *dominant_gradient_direction (DGD)* locations. Due to the gap in the 'discontinuity wall', the right mask comprises a part of the background.

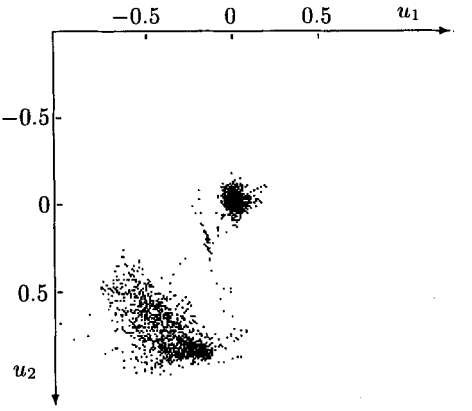

Fig. 4. The distribution of u_1 and u_2 components of the optical flow vectors within the larger connected-component (right hand side) from Fig. 3(b). One can clearly recognize an approximately circular cluster around $(u_1 = 0, u_2 = 0)$ corresponding to the stationary background, and another elliptical cluster centered around $(u_1 = -0.4, u_2 = 0.75)$ which corresponds to regular optic flow vectors associated with the moving vehicle.

pop out immediately which correspond to the two different segments of optic flow vectors contained in this connected-component. A standard clustering algorithm allows to detect and separate these two clusters automatically. It then is a straightforward procedure to separate the original connected-component into two segments, one corresponding to the image of the vehicle and the other corresponding to the background. Pixels, for which the estimated optic flow yields values in the uncertainty area between these two clusters, are simply suppressed: each pixel-cluster will only accept pixels with an OF-vector whose Mahalanobis distance from the nearest OF-cluster-center remains below a threshold.

6.4 Merging Certain Combinations of OF-Connected-Components

Although exclusion of DGD-locations helped to prevent 'leakage' of connected-components for vehicles into the background, it has the disadvantage to generate 'holes' within an object mask whenever the object image comprises marked straight line edge segments with substantial contrast. As can be seen in Fig. 5(b), such DGD-pixel locations can form connected-components of their own which are totally surrounded by R_OF locations. It should be noted that, in this case, the OF-vectors estimated for DGD-locations in the interior of the car image do not differ in an immediately noticeable manner from their surrounding R_OF-vectors.

We may now remove these 'Swiss-Cheese holes' by the simple requirement that a DGD-connected-component, which is completely surrounded by a R_OF-connected-component, can be merged into the surrounding component, provided the Mahalanobis distance between their mutual OF-vector distributions is compatible with the hypothesis that this distance vanishes. We thus first determine the cluster center coordinates $\mathbf{u}_{DGD}$ and $\mathbf{u}_{R\_OF}$, respectively, together with the corresponding covariance matrices Σ_{DGD} and Σ_{R_OF}. Subsequently, these two segments are merged, if

$$(\mathbf{u}_{DGD} - \mathbf{u}_{R\_OF})^T (\Sigma_{DGD} + \Sigma_{R\_OF})^{-1} (\mathbf{u}_{DGD} - \mathbf{u}_{R\_OF}) \leq threshold_{regionmerge-DGD-with-R\_OF} \quad . \quad (14)$$

As a result, we obtain a mask shown in Fig. 5(d).

Figure 6 presents an overlay of all 'object-image-masks' obtained in this manner for the entire frame. In order to simplify the visual detection of these masks by a viewer, we suppressed connected-components which either are smaller than 10 pixels or for which the average optic flow does not exceed a small threshold of 0.17 pixels per frame.

Since we sample video images of the depicted scene at a rate high compared to the temporal change related to scene motion, the apparent shift of object masks from frame to frame is small. The similarity of corresponding object masks from consecutive frames is further increased by the implied extended temporal averaging along the locally prevailing optic flow direction, an immediate consequence of the manner in which optic flow vectors are estimated in this approach.

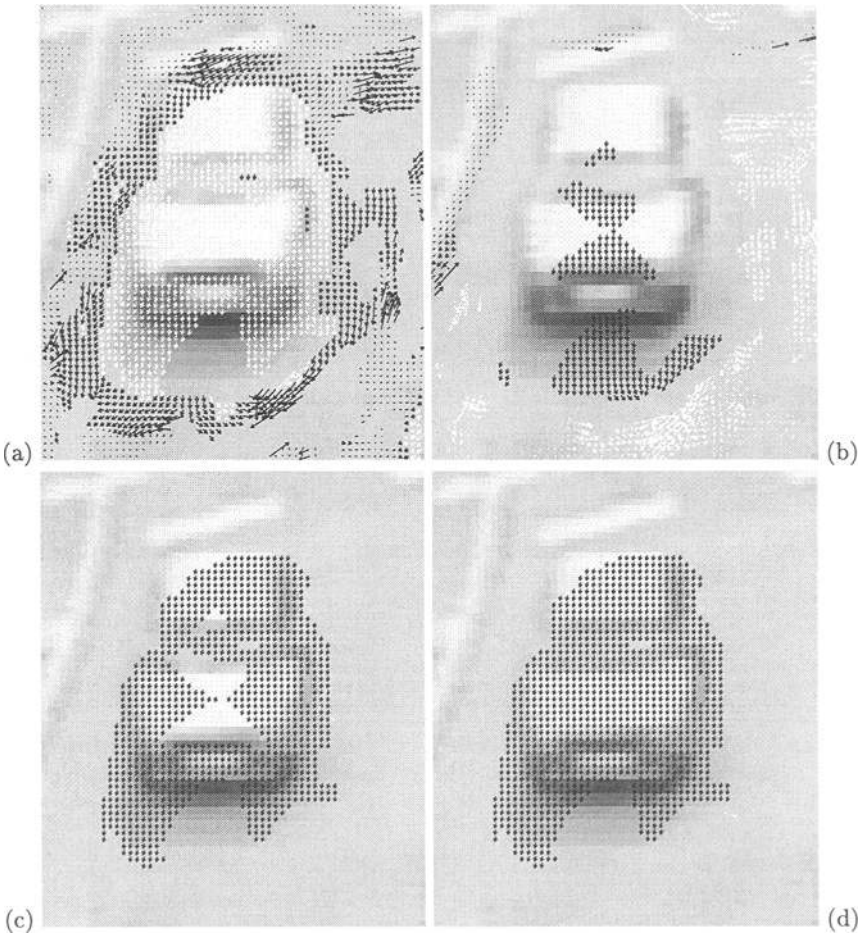

Fig. 5. Enlargement of window No. 1 in Fig. 1. (a): Overlaid are the OF-vectors, which are estimated at pixel positions belonging to the category *regular_optical_flow (R_OF)* (colored with white) and *OF_discontinuity (OF_D)* (colored with black). (b): The OF-vectors colored black in this panel have been estimated at pixels which have been assigned to the category *dominant_gradient_direction (DGD)*. Note that most of these vectors are placed in the interior of the car image or around the high contrast road markings. The white OF-vectors have been estimated at pixels belonging to the category *neutral*. (c): Masks determined in analogy to those from Fig. 3. Due to pixels at DGD-locations, the depicted *regular_optical_flow (R_OF)*-mask exhibits a hole. (d): 'object-image-mask' resulting from the requirement to include DGD-segments, provided these are completely surrounded by R_OF-locations and the OF-vectors estimated at the DGD locations are similar (according to (14)) to the OF-vectors which are estimated at the *regular_optical_flow (R_OF)* -mask.

Fig. 6. Image frame as shown in Fig.1, with superimposed connected regions (indicated by their contour lines) determined according to the algorithm described in Sect. 6. The smaller masks in the top right quadrant as well as the one close to the top left corner correspond to moving people. Note the two masks around the top of a pole in the upper left quadrant: these two masks belong to a vehicle which is partially occluded by this pole.

7 Discussion of Related Research and Conclusion

[11] analysed potential error sources in the usual pseudo-inverse solution for differential OF-estimation and investigated a 'Total Least Squares' approach similar to the one underlying (4). These authors suggested the use of a filter set comprising various orientations, bandwidths, and resolutions. They had to devise means to combine results obtained by different filters – in contradistinction to our adaptive approach which obviates the need to recombine different filter results. [8], too, employ multiple filters based on Hermite polynomials and parameterize their OF-estimation approach directly by 3-D motion parameters. Their algorithm evaluates intermediate results in order to properly diagnose critical grayvalue configurations. These authors do not, however, determine an OF-field with the density and resolution required for the examples used above. No attempt has been made by these authors to actually segment the estimated OF-fields and to evaluate the segmentation results in order to facilitate an assessment of OF-estimation.

Xiong and Shafer [13, 14] investigated hypergeometric filters somewhat similar to Gabor filters (see, e. g., [13, Figure 5]) in order to estimate optic flow, too. Xiong and Shafer do not attempt to segment OF-fields estimated in this manner, nor do they apply their method to image sequences with discontinuities of the optical flow field comparable to those treated in this contribution. It thus remains open whether their approach could cope with such situations.

The idea of a 'structure tensor' had been investigated already way back by [6]. Subsequent generalizations of Knutsson's ideas to spatiotemporal grayvalue structures are discussed, e. g., by [12] who, however, only treated a few (fairly coarse) synthetic image examples. Knutsson also influenced the work of [5]. [2] recently extended the investigations of [5] and applied them to a few short image sequences. Neither [12] nor [5], however, exploited knowledge contained in the GLST about the local spatiotemporal grayvalue structure in order to devise a *matched derivative convolution filter*. These authors show several examples of estimated OF-fields although they do not explicitly generate masks covering images of moving objects. As far as can be inferred from their illustrations, the difficulties diagnosed and overcome by the algorithm described in the preceding sects. 4 through 6 had not yet even been recognized by these authors.

An adaptive spatiotemporal filter somewhat similar to the one described here has already been reported by [9] who assumed that relevant grayvalue structures could be modelled as spatiotemporal Gaussians. The authors exploited this assumption in an attempt to separate the covariance matrix defining a local grayvalue structure in an image sequence from the filter covariance matrix. Difficulties arose in areas where image noise becomes relevant or where the assumption about the underlying grayvalue structure begins to break down.

[7] used an adaptive filter in 2D which is constructed in a manner similar to the one described here. But no attemps have been made by [7] to apply the adaptive filter to the estimation and segmentation of optic flow fields.

Extended own investigations and a judicious combination of experiences reported in the literature suggested a renewed attempt to simultaneously estimate and segment an OF-field. We converted the knowledge provided by a GLST into a *spatiotemporally adaptive* OF-estimation approach. As can be seen from experimental evidence presented above, we thereby obtained a *much cleaner signal about what happens at a particular space-time location in an image sequence.* This improvement, in turn, greatly facilitated to identify *structural* criteria which distinguish various types of failures that may occur in estimation and segmentation steps.

This diagnostic capacity allowed us to design a fairly simple, but nevertheless robust segmentation algorithm for an OF-field which complies with our basic assumptions about motion and distances between objects in the scene relative to the recording video camera: objects, whose images have to be segmented from the background, move only a small distance – in comparison to their distance from the center of projection – during the time between consecutive video-frames, thus yielding essentially parallel and equal-magnitude OF-vectors. Obviously, our approach will have to be modified if the object motion is more complicated than a small translation, since only the latter results in a more or less homogeneous optic flow field.

Please note that we do not base our approach on an *explicit* a-priori characterization of similarity required between neighboring optic flow vectors: we just assume OF-vectors to be sufficiently similar due to the manner in which they have been estimated with a high overlap of their support areas. This 'inherited' similarity quickly breaks down, however, *near the boundary* of an image area corresponding to an object moving rigidly in the depicted scene.

The methodology underlying our approach exploits *entire image regions* with acceptably *homogeneous OF-vectors* as a *tool* for quickly identifying locations where the homogeneity assumption is violated: clustering pixel locations with equal characteristics into connected-components often *kind of magnifies* any estimation or segmentation *deficiency*. Such a deficiency is likely to result in counterintuive region boundaries which can be quickly identified and analysed. This 'methodological leverage' can only be applied, however, since the increased reliability of our *spatiotemporally adaptive OF-estimation* approach removes enough artefacts and noise effects that we obtain a chance to *structurally* analyse failures – as opposed to fiddling around endlessly with attempts to find 'the' optimal parameter combination.

As a result, we obtain 'moving object masks' whose derivation depends on intuitively accessible parameters – so far in an apparently uncritical manner. The resulting mask quality appears sufficient to initialize model-based tracking of vehicles. It may even facilitate a purely data-driven tracking approach in the picture domain. Further research into these directions has been started.

8 Acknowledgements

Partial support of these investigations by the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged.

References

1. J.L. Barron, D.J. Fleet, and S.S. Beauchemin: Performance of Optical Flow Techniques. International Journal of Computer Vision **12** (1994) 43-77
2. H. Haußecker and B. Jähne: A Tensor Approach for Precise Computation of Dense Displacement Vector Fields. E. Paulus and F.M. Wahl (Hrsg.), Mustererkennung 1997, 19. DAGM-Symposium, Braunschweig/Germany, 15.-17. September 1997; Informatik aktuell, Springer-Verlag Berlin Heidelberg 1997, pp. 199–208
3. R.M. Haralick and L.G. Shapiro: Computer and Robot Vision. Addison-Wesley Publishing Company, Reading / MA 1992 (Vol. I)
4. B.K.P. Horn: Robot Vision. The MIT Press, Cambridge/MA and London/UK 1986
5. B. Jähne: Spatio-Temporal Image Processing, Theory and Scientific Applications. Lecture Notes in Computer Science, Vol. 751, Springer-Verlag Berlin Heidelberg 1993.
6. H. Knutsson: Filtering and Reconstruction in Image Processing. Ph.D. Thesis, Department of Electrical Engineering; In: Linköping Studies in Science and Technology, Dissertations No. **88**, Linköping University, S-581 83, Linköping, Sweden, 1982.

7. T. Lindeberg, J. Gårding: Shape-Adapted Smoothing in Estimation of 3-D Depth Cues from Affine Distortions of Local 2-D Brightness Structure. Proc. 3rd European Conference on Computer Vision ECCV '94, 2–6 May 1994, Stockholm/S; J.-O. Eklundh (Ed.), Lecture Notes in Computer Science LNCS **800**, Springer-Verlag Berlin Heidelberg New York/NY 1994, pp. 389–400.
8. H. Liu, T.-H. Hong, M. Herman, and R. Chellappa: A General Motion Model and Spatio-Temporal Filters for Computing Optical Flow. International Journal of Computer Vision **22**:2 (1997) 141–172.
9. H.-H. Nagel, A. Gehrke, M. Haag, and M. Otte: Space- and Time-Variant Estimation Approaches and the Segmentation of the Resulting Optical Flow Field. Proc. Second Asian Conference on Computer Vision, 5-8 December 1995, Singapore, Vol. II, pp. 296–300.
10. V. S. Nalwa, T. O. Binford: On Detecting Edges. IEEE Transactions on Pattern Analysis and Machine Intelligence **PAMI-8** (1986) 699-714.
11. J. Weber and J. Malik: Robust Computation of Optical Flow in a Multi-Scale Differential Framework. Proc. Fourth International Conference on Computer Vision ICCV '93, 11-14 May 1993, Berlin/Germany, pp. 12-20; see, too, Int. Journal of Computer Vision **14**:1 (1995) 67–81.
12. C.-F. Westin: A Tensor Framework for Multidimensional Signal Processing. Ph.D. Thesis, Department of Electrical Engineering; In: *Linköping Studies in Science and Technology, Dissertations No.* **348**, (ISBN 91-7871-421-4) Linköping University, S-581 83, Linköping, Sweden, 1994.
13. Y. Xiong and S.A. Shafer: Moment and Hypergeometric Filters for High Precision Computation of Focus, Stereo and Optical Flow. International Journal of Computer Vision **22**:1 (1997) 25–59.
14. Y. Xiong and S.A. Shafer: Hypergeometric Filters for Optical Flow and Affine Matching. International Journal of Computer Vision **24**:2 (1997) 163–177.