R INRIA

Unsupervised Segmentation Incorporating Colour, Texture, and Motion

Thomas Brox , Mikaël Rousson, Rachid Deriche , Joachim Weickert

N° 4760

March 2003

THÈME 3

Rapport de recherche

INRIA
SOPHIA ANTIPOLIS

Unsupervised Segmentation Incorporating Colour, Texture, and Motion

Thomas Brox *, Mikaël Rousson†, Rachid Deriche †, Joachim Weickert*

Abstract: In this paper we incorporate different kinds of information, namely colour, texture, and motion, into a segmentation process. The segmentation is based on a variational framework for vector-valued data using a level set approach and a statistical model to describe the interior and the complement of a region. It is assumed here that there are only two regions in the image, basically one object and the background. In order to obtain appropriate texture features, we apply the idea of the nonlinear structure tensor, which turns out to have very good discrimination properties while inducing only three feature channels. The spatio-temporal version of the nonlinear structure tensor is also used to estimate the optic flow as the feature for motion. Before the actual segmentation process, the features are presmoothed by a novel nonlinear diffusion technique closely related to TV flow, but being strictly edge enhancing. The coupling between the channels hereby ensures the enhancement of edges at joint positions in all channels. This kind of presmoothing works well together with the statistical model we use, namely nonparametric Parzen density estimates. We also consider multi-scale segmentation in order to speed up the process and to obtain more robust results.
Furthermore, it is shown that our method can not only be applied to segment images but also to track moving objects. The tracking of several moving objects is taken into account,

 * Mathematical Image Analysis Group, Faculty of Mathematics and Computer Science, Saarland University, Building 27, 66041 Saarbrücken, Germany, {brox,weickert}@mia.uni-saarland.de, www.mia.uni-saarland.de

 † Projet Odyssée, INRIA Sophia-Antipolis, 2004, route des Lucioles, BP 93, 06902 Sophia-Antipolis, France, {Mikael.Rousson,Rachid.Deriche}@sophia.inria.fr, www-sop.inria.fr/odyssee/presentation/index.en.html

where a coupling between the object regions allows to track also partly occluded objects. Our method has been verified in several experiments using synthetic as well as real image data. There we checked the importance of the different kinds of information for obtaining good results. Finally, also the limitations of our approach are described and shown in some examples.

Key-words: Level Set Theory, Segmentation, Texture, Optic Flow, Structure Tensor, Non-linear Diffusion, Tracking.

Segmentation non-supervisée incluant couleur, texture et mouvement

Résumé : Nous proposons dans ce rapport un processus de segmentation incorporant différents types d'information tels que la couleur, la texture et le mouvement. L'approche de segmentation considérée est basée sur une formulation variationnelle pour des images multi-valuées et utilise une modélisation statistique pour décrire l'information région. Nous étudions ici le problème simplifié de segmentation en deux régions, correspondant à un objet et au fond de l'image. Afin d'obtenir des caractérisitiques de textures appropriées, nous utilisons l'idée du tenseur de structure non-linéaire. Celui-ci a la propriété de fournir une information discriminante entre différentes textures en introduisant seulement trois composantes. La partie spatio-temporelle du tenseur de structure est aussi utilisée pour estimer le flot optique comme carctérisation du mouvement dans une séquence d'images. Les différentes images caractéristiques sont lissées grâce à une nouvelle technique de diffusion non-linéaire proche du flot TV. Le couplage entre les composantes assure une prise en compte des contours de toutes les composantes. Ce type de pré-lissage est bien adapté à la modélisation non-paramétrique (fenêtre de Parzen) de l'information statisique qui est utilisée lors de la segmentation. Nous présentons aussi une formulation multi-échelle afin d'améliorer la vitesse et la robustesse de la méthode.

Notre méthode est ensuite étendue pour le suivi d'objets dans une séquence d'images. Afin d'améliorer les résultats dans le cas de cette extension, le résultat de l'image précédente déplacé grâce à l'information fournie par le flot optique est utilisé comme initialisation pour segmenter l'image courante. Cela revient à prédire la position de l'objet dans l'image courante et permet ainsi de rendre notre méthode de suivi d'objets plus robuste. De plus, le suivi de plusieurs objets est aussi pris en compte. La formulation variationnelle proposée induit naturellement un couplage entre les contours de chaque objet, permettant ainsi de traiter les cas d'occlusions partielles entre objets.

Nous avons testé notre méthode sur de multiple exemples avec des images synthétiques et réelles. Lors de ces essais expérimentaux, nous étudions l'importance à donner aux différents types d'information afin d'obtenir de bons résultats. Finalement, les limites de cette approche sont discutées autour d'exemples applicatifs.

Mots-clés : Théorie des courbes de niveaux, segmentation, texture, flot optique, tenseur de structure, diffusion non-linéaire, suivi d'objets.

1   Introduction

Image segmentation is one of the principal problems in computer vision and has been studied for decades. From recent approaches those using a variational framework are quite popular (see e.g. [61, 39]), because in such a framework it is possible to integrate many different cues and models. So it is possible to integrate, for instance, boundary information [30], shape priors [26, 47, 12, 10] as well as region information [59, 8, 22, 39]. Level set theory [36] provides for an efficient possibility to find a minimum solution of such an established energy.

Especially the use of prior knowledge has become very popular in recent time, since it is very useful or even necessary to obtain good segmentation results in difficult image scenes. However, in this paper an unsupervised approach will be proposed that will not depend on previously acquired information. The objective of such an unsupervised approach can be to find good segmentations in less difficult image scenes in order to serve as a knowledge acquisition method for a segmentation based on prior knowledge.

In order to succeed in this task it is necessary to use as much information of an image as possible. This importance to combine different cues in a segmentation algorithm has also been recognised earlier in the work of Malik et al. [29]. Therefore, we propose to use not only the grey value of an image but also colour, texture as well as motion information, if they are available. In the proposed framework based on the work in [46] it is possible to integrate all this information.

However, the possibility to integrate different kinds of information is only one step. The next question appearing is how to acquire the information from the image. There is no such problem when using only primary features like grey value or colour, yet as soon as secondary features like texture or motion are included, it is no longer sure how to extract them from the image the best way in order to ensure a high discrimination power between regions while inducing as few channels as possible.

In the field of texture analysis the most frequently used methods are based on Gabor filters [17, 49, 50, 38]. In the sum Gabor filters can discriminate all kinds of textures, and neurobiology indicates them to be important in human vision [31]. However, Gabor filters have the decisive drawback to induce a lot of redundancy and thus many feature channels. Similar problems appear with the usage of Markov Random Fields (MRFs) [13]. As soon as a MRF of reasonable order is used, there arise many parameters not only causing lots of feature channels but also problems in estimating them. A very interesting solution for the problem of texture discrimination has been proposed by Bigün et al. [4] where they use the structure tensor for this task. In contrast to all the other methods the structure tensor only yields three different feature channels per scale. In this paper it will be demonstrated that these three channels together with a nonparametric statistical model, as proposed by Kim et

al. [24], are sufficient for texture discrimination. The only problem of the original work of Bigün et al. [4] is the Gaussian smoothing used for the structure tensor. It dislocates edges in feature space which leads to inaccurate segmentation results. To overcome this problem of the classical structure tensor it was proposed in [57] to develop a nonlinear structure tensor based on nonlinear matrix-valued diffusion. In [45] we demonstrated that this nonlinear structure tensor in fact performs very well when being applied to texture discrimination.

Considering motion, the optic flow is the principal method to integrate this information. Optic flow estimation is a complete research area on its own, and there exist plenty of different techniques; see e.g. [3, 34, 51] for overviews. Interestingly, the nonlinear structure tensor already applied for the texture can also be used here [6]. Two different kinds of features for motion can be derived from the optic flow: first the optic flow vector as such, second only its magnitude. While the first case is more general and also allows the detection of moving objects in a scene where the camera is moving, too, the second case will be more attractive, if just the motion needs to be detected.

Applications of a segmentation technique integrating all these cues are numerous. We will show here the example of tracking objects. Although the segmentation technique can be used for tracking without changes, as demonstrated in [46], an extension is added, as we introduce the simultaneous tracking of multiple objects. A coupling between the objects has to be taken into account in order to deal with partial occlusions. Like for segmentation it is possible to use arbitrary kinds of information also for tracking. It is self-evident to apply motion information in this case, but also grey value, colour and texture can be added.

The remainder of this paper is organized as follows. In the next section the segmentation framework will be described. This includes the level set formulation, the statistical model as well as some implementation aspects. Section 3 then deals with the acquisition and the treatment of the features. Especially the acquisition of the texture and optic flow features using the nonlinear structure tensor will be specified. Section 4 describes a multi-scale implementation of the technique, followed by Section 5 where the method is applied to tracking. In the succeeding section we show the results of our experiments and specify the limitations of the technique. The paper is concluded by a brief summary as well as an outlook on future work.

2   Two-Region Segmentation

Assume the image to consist of only two regions: the object region and the background region. Then a segmentation splits the image domain Ω into two disjoint regions Ω_1 and Ω_2 where the elements of Ω_1 and Ω_2 respectively are not necessarily connected. Let $u : \Omega \to \mathbb{R}^N$ be the computed features of the image and $p_{ij}(x)$ the conditional probability

density function of a value $u_j(x)$ to be in region Ω_i. Assuming all partitions to be equally probable and the pixels within each region to be independent, the segmentation problem can be formulated as an energy minimisation problem following the idea of *geodesic active regions* [39, 46]:

$$E(\Omega_i, p_{ij}) = -\sum_{j=1}^{N} \left(\int_{\Omega_1} \log p_{1j}(u_j(x))dx + \int_{\Omega_2} \log p_{2j}(u_j(x))dx \right) \to \min \qquad i = 1, 2.$$

(1)

For minimising this energy, a *level set function* is introduced. Let $\Phi : \Omega \to \mathbb{R}$ be the level set function with $\Phi(x) > 0$ if $x \in \Omega_1$ and $\Phi(x) < 0$ if $x \in \Omega_2$. The zero-level line of Φ is the searched boundary between the two regions. We also introduce the regularised heaviside function $H(s)$ with $\lim_{s \to -\infty} H(s) = 0$, $\lim_{s \to \infty} H(s) = 1$, and $H(0) = 0.5$. Furthermore, let $\chi_1(s) = H(s)$ and $\chi_2(s) = 1 - H(s)$. This allows to formulate a continuous form of the above-mentioned energy functional:

$$E(\Phi, p_{ij}) = -\sum_{i=1}^{2} \sum_{j=1}^{N} \left(\int_{\Omega} \log p_{ij}(u_j)\chi_i(\Phi)dx \right).$$

(2)

The minimisation of this energy can be performed using the following gradient descent:

$$\partial_t \Phi = \sum_{j=1}^{N} \left(\log \frac{p_{1j}(u_j)}{p_{2j}(u_j)} H'(\Phi) \right)$$

(3)

where $H'(s)$ is the derivative of $H(s)$ with respect to its argument.

2.1 Gaussian PDF

The variational framework is not completely defined so far, since it still lacks the definition of the probability density function (PDF). A reasonable choice is a Gaussian function. Assumed there is no useful correlation between the feature channels, this yields two parameters for the PDF of each region i and channel j: mean μ_{ij} and standard deviation σ_{ij}. Given the level set function Φ, the minimisation of Eq.2 with respect to these parameters (see [46]) leads to the updating equations:

$$\mu_{ij} = \frac{\int_{\Omega} u_j(x)\chi_i(x)dx}{\int_{\Omega} \chi_i(x)dx},$$

(4)

$$\sigma_{ij} = \sqrt{\frac{\int_{\Omega} (u_j(x) - \mu_{ij})^2 \chi_i(x)dx}{\int_{\Omega} \chi_i(x)dx}}.$$

(5)

Given the probability densities, the energy can be minimised with respect to Φ using the gradient descent in Eq.3. Thus the segmentation process works according to the *expectation-maximation* principle. Initialised with some level set function with both negative and positive values the parameters of the PDFs can be computed. At the same time step the gradient descent adapts the level set function to the new parameters. For a Gaussian PDF this iterative process can be proven to converge to the next local minimum depending on the initialisation.

2.2 Nonparametric PDF

Although reasonable, choosing a Gaussian function as PDF is not the only possible solution. Kim et al. [24] proposed nonparametric Parzen density estimates instead. Using discrete histograms this approach comes down to smoothing the histograms computed for each region i and channel j by a Gaussian kernel K_{σ_h}:

$$p_{ij}(g) = K_{\sigma_h} * \frac{\int_\Omega \delta_{u_j(x)g} \chi_i(x) dx}{\int_\Omega \chi_i(x) dx} \tag{6}$$

with $\delta_{kl} = 1$ if $k = l$ and 0 otherwise, and σ_h being a regularisation parameter that we set to a fixed value ($\sigma_h = 8$).

This nonparametric PDF estimate is much more powerful in describing the statistics within the regions than the Gaussian approximation. Although this yields best usage of the given information, there is also a drawback. Compared to the Gaussian approximation this approach is less able to generalize but adapts more specifically to the given data. Thus it results in more local minima in the objective function and makes it more dependent on the initialisation.

This problem can be addressed by applying the basic idea of deterministic annealing [42, 43]. Using a Gaussian function in the first run can be considered as a smoothed version of the objective function with less local minima. The result of this minimisation process is supposed to be close to the global minimum of the objective function obtained by using the nonparametric approach. A second run of the minimisation process will finally result in this global minimum or a local minimum that is very close to this optimum solution. Although there exist counter-examples where this approach will fail, the heuristic works very well in most cases.

2.3 Smoothness term

So far there has not been a smoothness constraint employed, so the segmentation inclines to encircle small noisy regions. In order to avoid this, it is common to penalise the length of

the boundary between the regions, thus adding the following smoothness term to the energy functional

$$\alpha \int_\Omega |\nabla H(\Phi)| dx \tag{7}$$

where α is a smoothness parameter. The steepest descent of this energy results in an additional evolution term closely related to mean curvature motion

$$\alpha H'(\Phi) \, \text{div} \left(\frac{\nabla \Phi}{|\nabla \Phi|} \right). \tag{8}$$

which automatically includes a narrow band. For the relation between this smoothness term and mean curvature motion we refer to [60].

Adding the smoothness term to the primary energy in Eq.2 the following expression is obtained:

$$E(\Phi, p_{ij}) = - \sum_{i=1}^{2} \sum_{j=1}^{N} \left(\int_\Omega \log p_{ij}(u_j) \chi_i dx \right) + \alpha \int_\Omega |\nabla H(\Phi)| dx \to \min \tag{9}$$

and the resulting gradient descent will be

$$\partial_t \Phi = H'(\Phi) \left(\sum_{j=1}^{N} \log \frac{p_{1j}(u_j)}{p_{2j}(u_j)} + \alpha \, \text{div} \, \frac{\nabla \Phi}{|\nabla \Phi|} \right). \tag{10}$$

To get the geometric interpretation of this equation of evolution, we can link it to an explicit evolution of the curve $\mathbf{C}$ subject to a force β, acting only in the normal direction to the curve: $\mathbf{C_t} \cdot \mathbf{N} = \beta$. The link is made by replacing $H'(\Phi)$ by $|\nabla \Phi|$ as in [60] (this is only a time rescaling). Then the gradient descent (Eq.10) becomes

$$\partial_t \Phi = \beta |\nabla \Phi| \quad \text{with} \quad \beta = \sum_{j=1}^{N} \log \frac{p_{1j}(u_j)}{p_{2j}(u_j)} + \alpha \, \text{div} \, \frac{\nabla \Phi}{|\nabla \Phi|}. \tag{11}$$

This is exactly the evolution obtained from $\mathbf{C_t} \cdot \mathbf{N} = \beta$ using the level set representation (see [30] for details). So, our evolution for Φ can be interpreted as the evolution of $\mathbf{C}$ subject to the force β defined above, acting only in the normal direction to the contour.

2.4 Implementation

The gradient descent of our energy functional automatically leads to a narrow band approach, like it has been proposed for instance in [11, 1, 40, 37] in order to speed up computations. To compute the narrow band, the zero-level line of the level set function is determined in each iteration. Smoothing this line with a Gaussian kernel yields $H'(\Phi(x))$. The corresponding regularised Heaviside function $H(s)$ is consequently the integral of the Gaussian and can be implemented efficiently using a lookup table for the discretised values of the level set function. With $H(\Phi(x))$ and $H'(\Phi(x))$ available, it is easy to estimate the PDFs and to implement the gradient descent given in Eq.10. Respecting the narrow band idea to speed up computation, we only consider pixels for the gradient descent where $H'(\Phi(x))$ is larger than a certain threshold. Finally, the smoothness term is implemented as mean curvature motion restricted to the narrow band.

3 Computation of the Features

With a framework capable to incorporate different kinds of information available, the question of how to acquire this information becomes significant. The segmentation approach of the last section is already powerful enough to work with primary information, like the grey level image or the colour image, respectively, without any further preprocessing steps. This has been demonstrated in [46]. However, for secondary information, like texture or motion, some additional processing is necessary in order to derive useful features from the available primary information. In our approach both texture and motion features will be based on the concept of the *nonlinear structure tensor* and nonlinear diffusion that works solely in backward direction, so it is strictly edge-enhancing.

3.1 Nonlinear structure tensor

The nonlinear structure tensor introduced in [57] is based on the classic linear structure tensor [16, 4, 19, 27, 44]

$$J_\rho = K_\rho * (\nabla I \nabla I^\top) = \begin{pmatrix} K_\rho * I_x^2 & K_\rho * I_x I_y \\ K_\rho * I_x I_y & K_\rho * I_y^2 \end{pmatrix} \tag{12}$$

where K_ρ is a Gaussian kernel with standard deviation ρ, I the image, and subscripts denote partial derivatives. In the case of a colour image all channels are taken into account by

summing the tensor products of the particular channels [14]:

$$J_\rho = K_\rho * \left(\sum_{i=1}^{3} \nabla I_i \nabla I_i^\top \right). \tag{13}$$

The major problem of the classic structure tensor is the dislocation of edges due to the smoothing with Gaussian kernels. This leads to inaccurate results near discontinuities in the data. The basic idea in [57] to address this problem is the replacement of the Gaussian smoothing by nonlinear diffusion. However, they use only the magnitude of the evolving structure tensor to drive the diffusion. Our version of the nonlinear structure tensor uses the whole information by applying the scheme of nonlinear matrix-valued diffusion introduced by Tschumperlé and Deriche [52, 53] and diffusivity functions without a contrast parameter. In the following the details of this procedure will be described.

Nonlinear diffusion is based on the early work of Perona and Malik [41]. For a review we refer to [55]. The main idea is to reduce the smoothing in the presence of edges. The resulting diffusion equation is

$$\partial_t u = \mathrm{div}\,(g(|\nabla u|)\nabla u) \tag{14}$$

with $u(t = 0)$ being the initial image and g a decreasing *diffusivity function*. Perona and Malik proposed two different diffusivity functions

$$g(|\nabla u|) \;\;=\;\; \frac{1}{1 + |\nabla u|^2/\lambda^2} \tag{15}$$

$$g(|\nabla u|) \;\;=\;\; \exp\left(-\frac{|\nabla u|^2}{2\lambda^2} \right) \tag{16}$$

where λ is a contrast parameter steering the transition from forward diffusion to backward diffusion.

Eq.14 can only be used with scalar-valued data like a grey value image. Gerig et al. [18] introduced a version of nonlinear diffusion for vector-valued data

$$\partial_t u_i = \mathrm{div}\left(g\left(\sum_{k=1}^{N} |\nabla u_k|^2 \right) \nabla u_i \right) \qquad \forall i = 1, \ldots, N \tag{17}$$

where u_i is an evolving vector channel and N the total number of vector channels. Note that in this approach all channels are coupled by a joint diffusivity, so an edge in one channel also inhibits smoothing in the others.

When regarding the components of a matrix as components of a vector, what is reasonable,

since the Frobenius norm of a matrix equals the Euclidean norm of the resulting vector, it is possible to diffuse a matrix, such as the structure tensor, with the above-mentioned scheme. In fact, this complies with the scheme proposed in [52].

A rather critical issue is the appropriate choice of the diffusivity function g. Applied to the structure tensor, the diffusivity functions proposed by Perona and Malik, for example, will not work properly. This is because the structure tensor contains first derivatives, which have very local responses of different magnitude. Thus it is mostly impossible to choose an appropriate contrast parameter. Instead we use diffusivity functions of type [54]

$$g(|\nabla u|) = \frac{1}{|\nabla u|^p} \tag{18}$$

with $p \in \mathbb{R}$ and $p \geq 1$. These diffusivity functions include for $p = 1$ *total variation (TV) flow* [2, 15], a diffusion filter that is equivalent to TV regularisation [48, 7]. For $p = 2$, one obtains the so-called *balanced forward backward diffusion* introduced in [23]. Their properties fit the requirements of the structure tensor very well: There is no contrast parameter, they remove oscillations, and experiments show that they yield piecewise constant results. Therefore they can preserve, or for $p > 1$ even enhance, important edges [25].

Unfortunately, these diffusivity functions will lead to numerical problems when the gradient gets close to zero. This problem can be avoided by adding a small positive constant ϵ to the denominator.

$$g(|\nabla u|) = \frac{1}{|\nabla u|^p + \epsilon} \tag{19}$$

For implementation, we apply the AOS scheme [58] that allows efficient computation of such flows also for small ϵ. For ϵ in the area of 0.01, where the approximation of such kinds of flow is much better than for larger ϵ causing less blurring effects, the AOS scheme is around 3 orders of magnitude faster than a simple explicit scheme.

In conclusion, the nonlinear structure tensor is obtained by applying Eq.17 with the diffusivity function of Eq.19. For the spatial version of the structure tensor the data vector will consist of three components initialised by $u(t = 0) = (I_x^2, I_y^2, 2I_x I_y)$ with I being the image grey value. For the colour case the initial conditions are extended to the sums $u(t = 0) = (\sum_i (I_i)_x^2, \sum_i (I_i)_y^2, 2\sum_i (I_i)_x (I_i)_y)$. Note that the third vector component has to be weighted twice, since it appears twice in the matrix.

The same way a spatio-temporal nonlinear structure tensor can be obtained by using corresponding initial conditions $u(t = 0) = (I_x^2, I_y^2, I_z^2, 2I_x I_y, 2I_x I_z, 2I_y I_z)$ where z describes the time axis. In the case of colour these values will again be extended by using the corresponding sums.

3.2   Texture

As proposed by Bigün et al. [4], the components of the structure tensor can directly be used for texture discrimination. In Section 6 it will be demonstrated that this yields very good results. As the components of the structure tensor are not redundant, they are able to collect the whole texture information of one scale with only three channels. In fact, it will be shown in Section 6 that one scale of the structure tensor will even be sufficient to cope with most cases of multi-scale textures, if it is used together with the nonparametric statistical model described in the previous section. Also note the discrimination power of the structure tensor to be fully rotation invariant. A formal proof for this rotation invariance can be found in Appendix A.

3.3   Optic flow

As mentioned earlier, the structure tensor can be used for optic flow estimation, too. This has also been proposed by Bigün et al. [4], though we use the early approach of Lucas and Kanade [28] here, which is almost the same.
The assumption that image structures do not alter their grey values during their motion can be expressed by the optic flow constraint [21]

$$I_x u + I_y v + I_z = 0. \tag{20}$$

Again subscripts denote partial derivatives, and u and v are the searched components of the optic flow vector. As this is only one equation for two flow components, the optic flow is not uniquely determined by this constraint *(aperture problem)*. A second assumption has to be made. Lucas and Kanade proposed to assume the optic flow vector to be constant within some neighborhood B_ρ of size ρ. The optic flow in some point (x_0, y_0) can then be estimated by the minimiser of the local energy function

$$E(u,v) = \frac{1}{2} \int_{B_\rho(x_0,y_0)} (I_x u + I_y v + I_z)^2 dx dy. \tag{21}$$

A minimum (u,v) of E satisfies $\partial_u E = 0$ and $\partial_v E = 0$, leading to the linear system

$$\begin{pmatrix} \int_{B_\rho} I_x^2 dx dy & \int_{B_\rho} I_x I_y dx dy \\ \int_{B_\rho} I_x I_y dx dy & \int_{B_\rho} I_y^2 dx dy \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} -\int_{B_\rho} I_x I_z dx dy \\ -\int_{B_\rho} I_y I_z dx dy \end{pmatrix}. \tag{22}$$

Instead of the sharp window B_ρ often a convolution with a Gaussian kernel K_ρ is used yielding

$$\begin{pmatrix} K_\rho * I_x^2 & K_\rho * I_x I_y \\ K_\rho * I_x I_y & K_\rho * I_y^2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} -K_\rho * I_x I_z \\ -K_\rho * I_y I_z \end{pmatrix}. \tag{23}$$

The linear system can only be solved, if the system matrix is not singular. Such singular matrices appear in regions where the image gradient vanishes. They also appear in regions where the aperture problem remains present, leading to the smaller eigenvalue of the system matrix to be close to 0. In this case one may only compute the so-called normal flow (the optic flow component parallel to the image gradient). Using a sufficient amount of smoothing for the structure tensor, however, will greatly reduce such singular situations and one will obtain dense results in most cases.

Obviously the entries of the linear system are five of the components of the spatio-temporal structure tensor J_ρ. Of course, instead of the classic linear structure tensor, the nonlinear structure tensor will be employed here, as proposed in [6]. For the diffusion of the structure tensor TV flow will be applied, i.e. $p = 1$.

This optic flow estimation method can be related to the work in [32]. The main difference between the approaches in [6] and [32] is that in [6] the smoothing is adapted at each iteration to the new, smoothed data, thus being a strictly nonlinear process.

3.4   Coupled smoothing of all features

Putting all features together yields a feature vector consisting of 3 colour channels (R, G, B), 3 texture channels (components of the spatial structure tensor, respecting its symmetry) and 2 optic flow channels (components u and v of the optic flow vector). Although nonlinear diffusion has been introduced with the nonlinear structure tensor, it is more reasonable to smooth all channels together in a coupled manner. Consequently, at least for the texture features, the nonlinear structure tensor is not applied directly. Instead, the components of the unsmoothed structure tensor are smoothed together with the other features. The advantage of such a procedure is the usage of additional information from the other channels, namely the edges of these channels, for driving the diffusion. For computing the optic flow, however, the structure tensor must be smoothed explicitly with $t > 0$ in order to avoid the aperture problem. Nevertheless, the resulting optic flow components will also be smoothed together with the other features in a sort of post-processing step. The coupled smoothing of all features is in accordance with the segmentation, where all features are employed in one single process, as well.

For the coupled smoothing we propose to use the diffusivity function given in Eq.19 with $p > 1$. In our experiments $p = 1.6$ appeared to be a good choice, though other values of p will also work. Diffusivity functions of this type have not been used in any application so far. They have the very special property to work always in backward direction, so edges are not only preserved but even enhanced. In the continuous case, backward diffusion is ill-posed, however, discretisation has been shown to resolve this problem [56]. This is the reason why the diffusivities given by Perona and Malik, which also include backward dif-

fusion, could be used successfully in many applications. In our case, the larger p, the more important is the influence of the edge enhancement compared to the smoothing effect. We want to stress that p has nothing to do with the contrast parameter λ in the Perona-Malik diffusivities. It specifies the ratio between edge enhancement and smoothing. As edge enhancement has basically a positive effect for our application, it would be best to use large p. However, this will considerably increase diffusion time necessary to obtain also an appropriate smoothing effect. So in order to ensure efficiency and accuracy in practice, p should be chosen in the proposed area.

It should also be noted that the values of all channels are supposed to be approximately in the same range in order to ensure a balanced coupling between the channels. This is not the case for the optic flow channels. Therefore, all channels are normalized to the same range before the diffusion process starts.

Summarizing the feature extraction, the following steps are performed. First the spatio-temporal nonlinear structure tensor is computed with diffusion according to TV flow ($p = 1$). From this nonlinear structure tensor, the optic flow components are determined. Afterwards all available information, i.e. the colour channels, the components of the unsmoothed spatial structure tensor, and the optic flow vector, are assembled to a joined feature vector that is smoothed with nonlinear diffusion and $p > 1$. For all smoothing operations the vector-valued diffusion scheme of Eq.17 can be employed. It is implemented using the AOS scheme.

4   Multi-Scale Extension

There are two reasons why it is beneficial to consider multiple scales for the segmentation. The first reason is the reduction of computation time. The second, more important reason is the improvement in robustness. The initialisation of the level set function can sometimes be the critical point of our method. If an inappropriate initialisation is chosen, the final solution might be a local minimum of the objective function that is far away from the global minimum. On the other hand, it can be assumed that the downsampled version of an image, or its features, yields less local minima than the original data. Of course, this advantage has to be payed with a reduced accuracy.

The idea to combine both robustness and accuracy is simple: The data from a finer scale is downsampled and serves as input for a more robust preliminary segmentation at a coarser scale. The result of this segmentation will then be used to initialise the segmentation of the finer scale which improves accuracy.

Beside the improvements in robustness and computation speed, this multi-scale implemen-

tation opens the possibility to consider multiple scales in texture discrimination. In the single-scale case it was only possible to consider one scale of the texture without loosing accuracy. Now it is feasible to add also the structure tensor of the downsampled image to the feature vector. Consequently, each coarser scale has three additional texture channels compared to its next finer scale. Although this additional information is not available for the finer scale anymore, otherwise accuracy would be lost, it helps in finding the coarse shape of an object. Note that due to the statistical model the finest scale is able to distinguish textures of different scales, but gets trapped in a local minimum very easily. An example is shown in Fig.1.
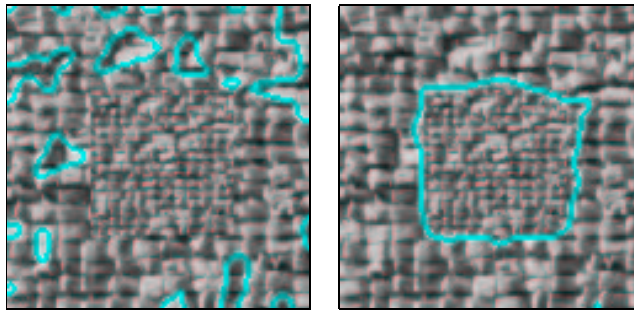
Figure 1: Segmentation of a multi-scale texture. LEFT: Result considering only one scale. RIGHT: Result considering two scales.

5 Tracking

One of several applications for the segmentation approach described in the preceding sections is the tracking of moving objects. Since it becomes possible to employ not only the optic flow to follow the objects but also other information, the tracking is expected to be more reliable than with techniques based only on the optic flow. In [37] and [46] it has already been demonstrated that it is possible to apply segmentation to tracking. In this section we not only supplement the optic flow as an additional feature, but also take multiple objects of a scene into account. A general formulation is presented without adding artificial coupling terms.

The tracking algorithm splits into two parts: a detection part, extracting all moving objects in the first image, and a tracking part, trying to follow these objects. For the detection part our segmentation approach is applied to the first image of the sequence using the magni-

tude of the optic flow estimated between the first and the second image. While the optic flow magnitude will focus the segmentation on the moving objects, the grey value or colour will improve the accuracy of the result, since edges in the primary features of an image are located much more precisely than in the optic flow. However, including the grey value or colour can also be critical, since there is the possibility of certain objects to be considered as important, though they are not moving, just because their grey value or colour is much different from the rest of the image. As we are not interested in the detection of the most dominant object in the image but in the moving parts, we impose a higher weight on the optic flow magnitude than on the other features.

Once the moving parts of the image are separated from the non-moving background, the connected components are extracted from this region. Each connected component is considered as a separately moving object. To each of these objects a level set function Φ_k is assigned, with $k = 1,...,M$ and M being the total number of detected objects. In order to ease further computation, the level set function is ensured to be positive for the moving object and negative for the background. The resulting level sets are the basis for the tracking part.

In the tracking part the level sets Φ_k are updated simultaneously using the data of the next image. In this part both vector components of the optic flow are used as features. Since the object regions are assumed to be disjoint, a coupling between the level sets should be taken into account to prevent a pixel of being set for two different objects. But as we will see in the case of two moving objects, a simple variational formulation can easily avoid such cases without introducing additional terms.

5.1 Tracking of two objects

Before we introduce the general case, we study the tracking of two moving objects O_1 and O_2. We note B the remaining part of the image, corresponding to the background of the scene. As we said before, we use two level set functions Φ_1 and Φ_2 to represent the objects. The image must be composed by three regions given by

$$
\begin{aligned}
R_1 &= \{x \in \Omega, \Phi_1(x) \geq 0\}, \\
R_2 &= \{x \in \Omega, \Phi_2(x) \geq 0\}, \\
R_B &= \{x \in \Omega, \Phi_1(x) < 0 \text{ and } \Phi_2(x) < 0\}.
\end{aligned}
\tag{24}
$$

By e_1, e_2 and e_B we denote positive functions measuring the energy inside each region R_1, R_2 and R_B. The global energy is defined as follows:

$$
E = \int_\Omega \big(e_1 H(\Phi_1) + e_2 H(\Phi_2) + e_B(1 - H(\Phi_1))(1 - H(\Phi_2))\big) \, dx,
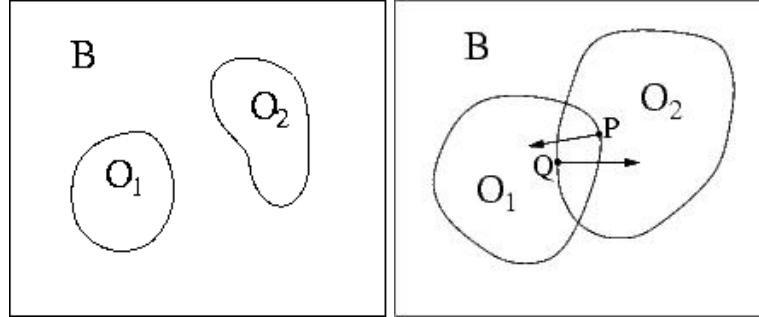\tag{25}
$$

Figure 2: LEFT: (a) Tracking of two objects. RIGHT: (b) Critical case: Overlapping between two objects, $\partial_t \Phi_1(P) = -e_1(P)H'(\Phi_1)$ and $\partial_t \Phi_2(Q) = -e_2(Q)H'(\Phi_2)$

Differentiation with respect to Φ_1 and Φ_2 gives a coupled system of two equations:

$$\begin{cases} \partial_t \Phi_1 = -H'(\Phi_1)\big(e_1 - e_B(1 - H(\Phi_2))\big), \\ \partial_t \Phi_2 = -H'(\Phi_2)\big(e_2 - e_B(1 - H(\Phi_1))\big). \end{cases} \quad (26)$$

These evolution equations naturally forbid any overlapping between objects since in such cases, the system is simplified to

$$\begin{cases} \partial_t \Phi_1 = -H'(\Phi_1)e_1, \\ \partial_t \Phi_2 = -H'(\Phi_2)e_2 \end{cases} \quad (27)$$

and both contours are shrinking with a speed proportional to their energy in the critical region as shown in Fig.2b.

5.2 The general case

The extension to the tracking of M objects is straightforward.

$$E = \int_\Omega \sum_{k=1}^{M} \bigg(\underbrace{-\sum_{j=1}^{N} \log p_{kj}\, H(\Phi_k) - \alpha |\nabla H(\Phi_k)|}_{e_k} \bigg) - \underbrace{\sum_{j=1}^{N} \log p_{Bj}\, \chi_B}_{e_B}, \quad (28)$$

$$\chi_B = \prod_{l=1}^{M} (1 - H(\Phi_l)). \quad (29)$$

Note that the χ function for the background has been redefined affecting also the estimation of the PDFs. The minimisation of the new energy leads to the revised gradient descent

$$\partial_t \Phi_k = -H'(\Phi_k)\Big(e_k - \alpha \,\mathrm{div}\, \frac{\nabla \Phi_k}{|\nabla \Phi_k|} - e_B \chi'_B\Big). \tag{30}$$

6   Experiments

6.1   Non-parametric model

In a first test we verified the ability of the non-parametric model to cope with Gaussian distributions. Fig.3 reveals the test to be successful. In Fig.3a the regions are Gaussian distributions with standard deviation $\sigma = 25$ and mean $\mu_1 = 113$ as well as $\mu_2 = 142$. In Fig.3b the mean of both regions is identical ($\mu = 128$) and the standard deviations are $\sigma_1 = 25$ and $\sigma_2 = 50$. For this test only the grey value information has been used and no presmoothing has been applied. In Fig.3c the initialisation of the level set function is depicted. We used the same kind of initialisation for all the following experiments. The black regions indicate the level set to be negative, while the white regions indicate it to be positive. The blue line denotes the zero-level line.
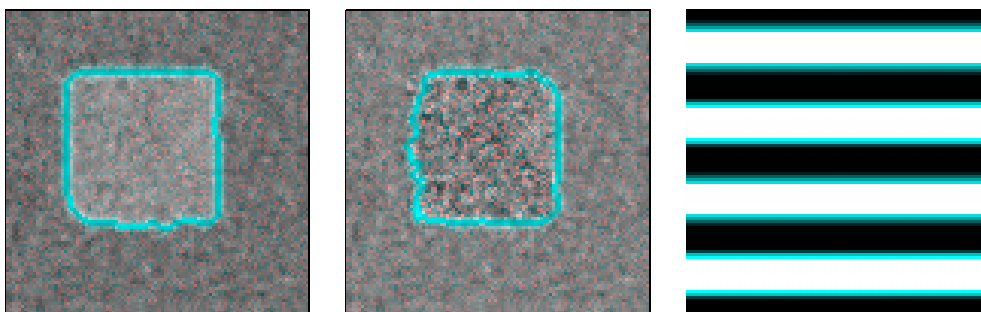
Figure 3: Non-parametric model with Gaussian distributed data. LEFT: (a) $\mu_1 = 113$, $\mu_2 = 142$, $\sigma_1 = \sigma_2 = 25$. CENTER: (b) $\mu_1 = \mu_2 = 128$, $\sigma_1 = 25$, $\sigma_2 = 50$. RIGHT: (c) Initialisation of the level set.

6.2   Segmentation

In order to verify the discrimination power of our segmentation approach with respect to texture, it was first applied to several synthetic test images composed of the well-known

Brodatz textures [5]. The results depicted in Fig.4 are very satisfactory. Even textures that are difficult to distinguish, like in Fig.4f and Fig.4h, can be handled by the nonlinear structure tensor combined with the statistical model. Note that the multi-scale approach also allows to discriminate textures of different scale, like those in Fig.4d and Fig.4f. In fact, the textures in Fig.4f are only different in scale, as the inner texture is the downsampled version of the one outside. We considered two scales for all our experiments.
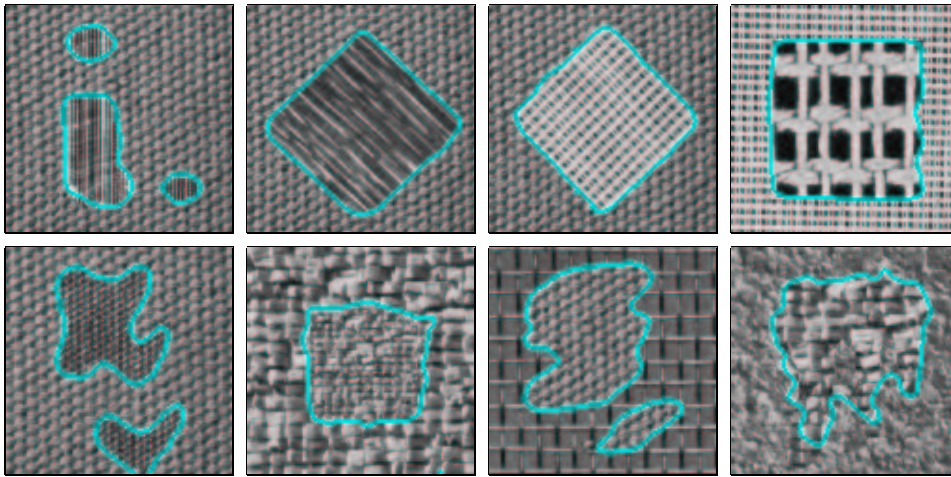
Figure 4: (a)-(h) FROM LEFT TO RIGHT, TOP TO BOTTOM: Segmentation results for synthetic texture images.

Although synthetic test images are best for verifying the possibilities and limits of a method, results achieved for real images are more interesting. So we applied our method also to natural images. Some results with grey level images are depicted in Fig.5.
In some further tests the texture information has been combined with colour. The results in Fig.6 reveal that the usage of both colour and texture leads to very good results as long as both information is coherent. Hence the combination of texture and colour information can make the method still successful, when it would have failed, if only colour *or* texture was considered, as shown in Fig.7. However, as soon as there is, for example, a textured object in front of a background containing different colours, it is not clear anymore whether the texture or the colour is more important. The result will depend on which feature has the larger discrimination power. Two examples where the colour information does not help to improve the result are shown in Fig.8. If colour and texture information contradict each other too much, the segmentation can even fail completely, as depicted in Fig.9.
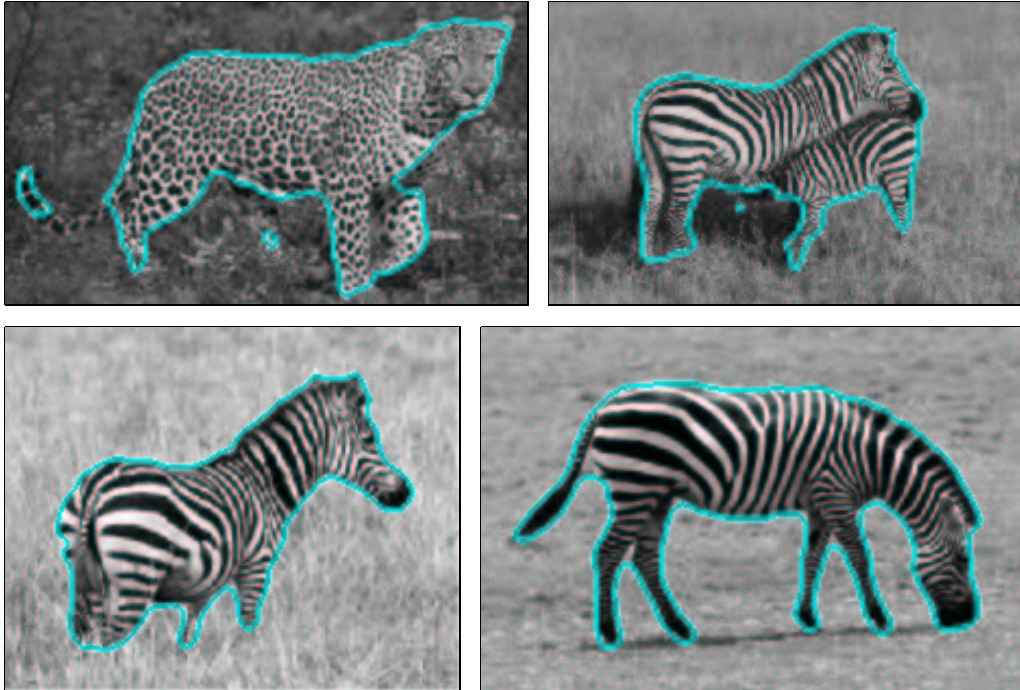
Figure 5: Segmentation results with some real-world images using texture and grey value information.

Note that two assumptions made in our approach are responsible for this failure. First, due to the unsupervised approach, it is assumed that there is no useful prior knowledge available. Of course, this is not the case for an example such as that in Fig.9. Even if somebody has never seen a giraffe, there might be prior knowledge about other animals available, proposing the result in Fig.9 to be impossible. The second assumption is the existence of only two regions in the image, the object and the background. This is also not true for this example. In fact, the background is composed of different regions. Because of the necessity to assign each region either to the background or to the object, some parts of the image that are more similar to the object than to the background are assigned to the object, such as the sky in Fig.9. This assumption of only two regions is in fact mainly responsible for the contradictions between the different feature channels, and explains why the results in Fig.8 are not perfect, as well. Obviously it is necessary to drop both assumptions to obtain good segmentation results with arbitrary images. We will return to this issue in the concluding

section of this paper.
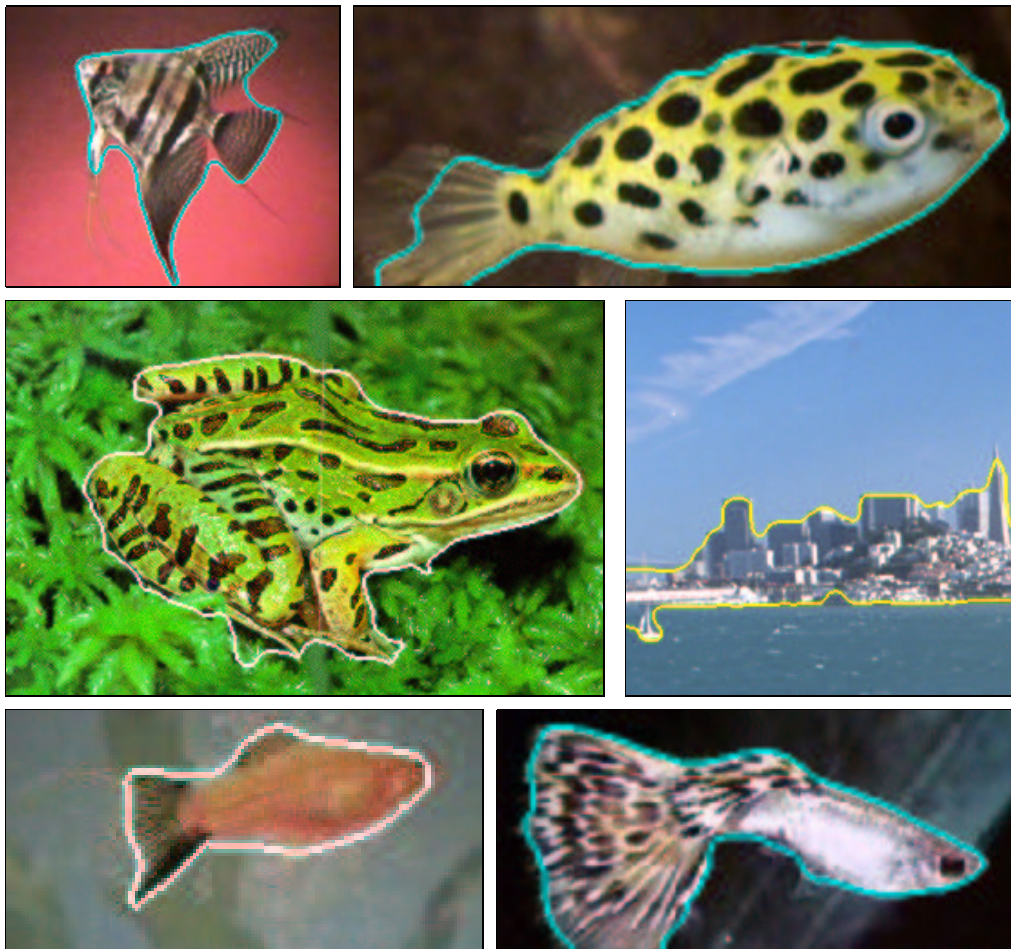
Figure 6: Segmentation results with some real-world images using texture and colour information.
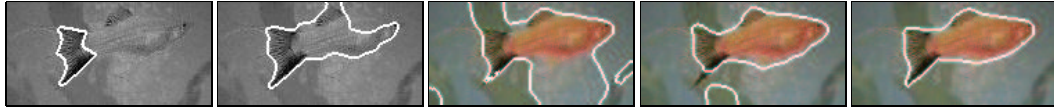
Figure 7: Benefits of using jointly colour and texture information. FROM LEFT TO RIGHT: (a) Grey value. (b) Grey value and texture. (c) Colour (RGB). (d) Colour (CIELAB). (e) Colour and texture.
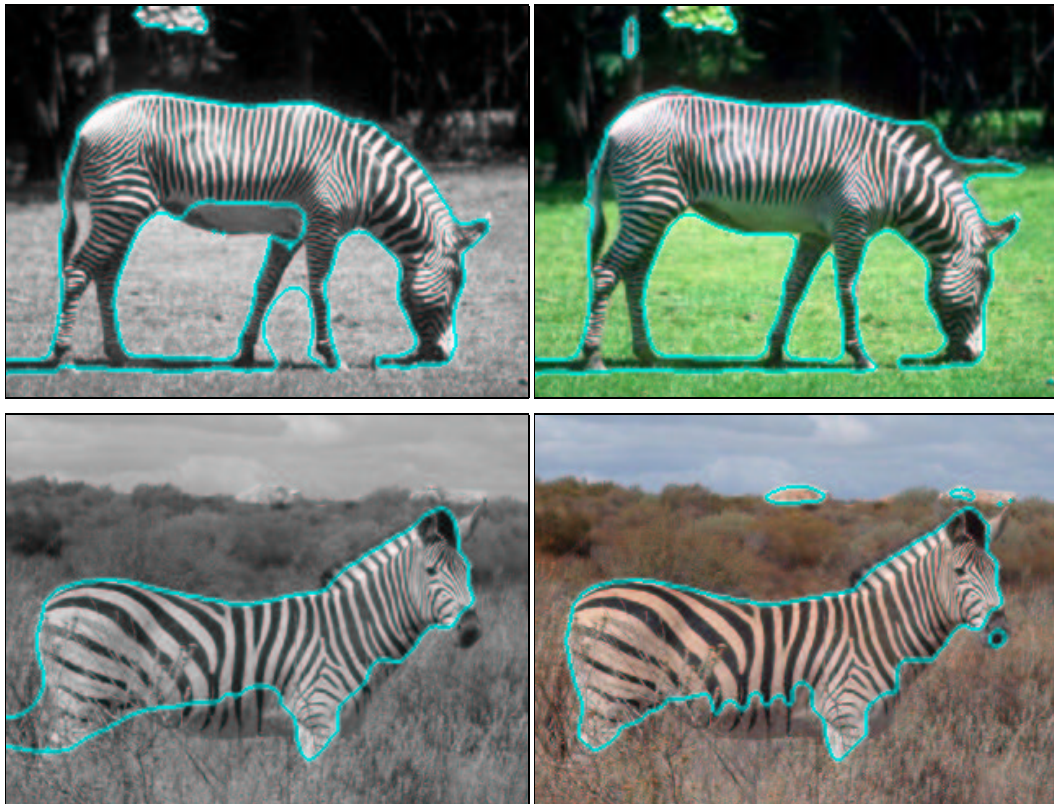

Figure 8: Two examples where additional information cannot improve the result. LEFT: Result when using grey level and texture. RIGHT: Result when using colour and texture.

When the segmentation approach was described, the objective function has been mentioned to have often more than one minimum. Thus the segmentation result can depend on the initialisation. For the examples we presented so far, we always used an initialisation consisting
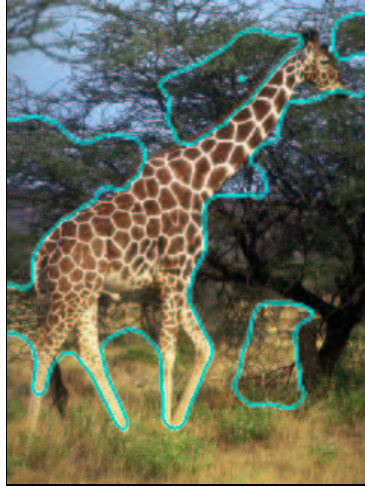
Figure 9: Example where the segmentation fails, because the assumptions of the approach (two regions) are not satisfied.

of several horizontal stripes (Fig.3c), which does not prefer a certain solution. But it is also possible to force a specific segmentation by using an initialisation that is very close to the wanted solution. An example is given in Fig.10. While the unbiased initialisation results in the region that is most different from the rest of the image (the bright background), the initialisation with a rectangle in the middle of the image leads to the fish (and its shadow). This suggests a possibility to steer the segmentation by some additional information, for example that the searched object is always somewhere close to the center of the image. However, we want to stress again that beside the example in Fig.10 all results were obtained with the unbiased initialisation.

Experiments on 4-phase segmentation. By now, our approach was restricted to the 2-phase case. Despite its numerous advantages, the level set representation can only be used to represent the interface between two regions. However, different ways of coupling multiple level set functions have been proposed in the literature, allowing multi-phase segmentation of a higher number of regions. Hence, these methods propose the segmentation of N regions using a different number of level set functions; N level set functions are used in [60, 39] while only $\log N$ are necessary in the approach described in [9]. We consider the last one, since it needs less level set functions and no artificial coupling is introduced unlike in other methods. This method was tested with two level set functions, which cor-
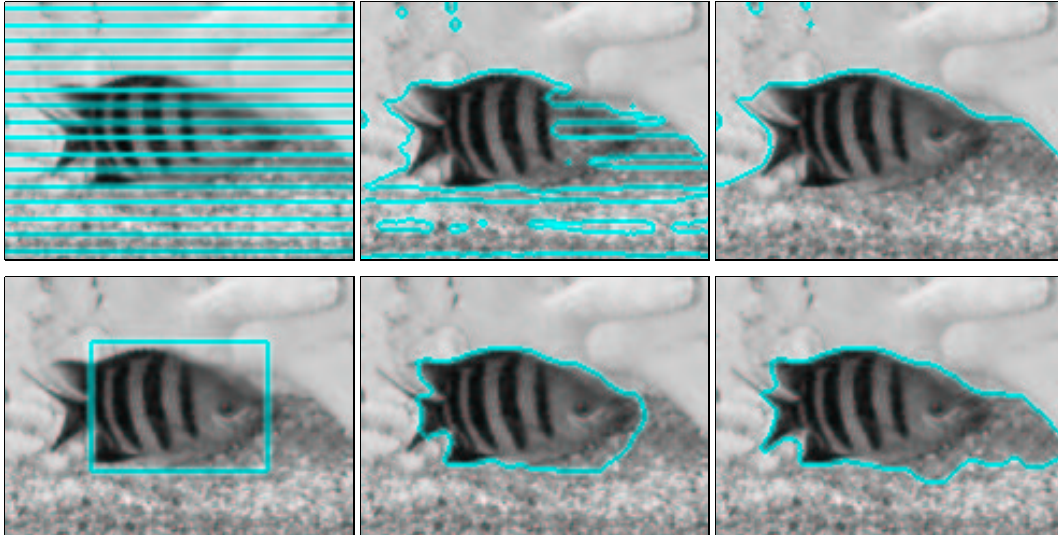
Figure 10: Curve evolution with different initialisations. FROM LEFT TO RIGHT: (a) Initialisation. (b) Intermediate result. (c) Final result.

responds to 4-phase segmentation. Zero crossings of the two level set functions and the corresponding segmentation are shown in Fig.11 on a synthetic image from the Brodatz database. However, the method seems to be more sensitive to the initialisation than the two-phase case.

6.3 Tracking

The tracking algorithm was first tested with two grey level sequences: the famous *Hamburg taxi* sequence[1] (Fig.12) as well as another traffic scene [2] (Fig.13). While in the first sequence the objects are relatively large, its quality caused by the camera hardware available at that time is rather challenging. The quality of the second sequence is better, but there are also very small objects consisting of only a few pixels. The results are depicted in Fig.12 and Fig.13. The usage of both optic flow and grey value turns out to be beneficial for the results. While the optic flow provides for the coarse shape and location of the objects, improving the robustness of tracking, the grey value is responsible for more accurate results. Note that

---

[1]The sequence was created at the University of Hamburg and can be obtained from `ftp://csd.uwo.ca/pub/vision`

[2]The sequence was kindly provided by the *Institut für Algorithmen und kognitive Systeme* at the University of Karlsruhe, Germany
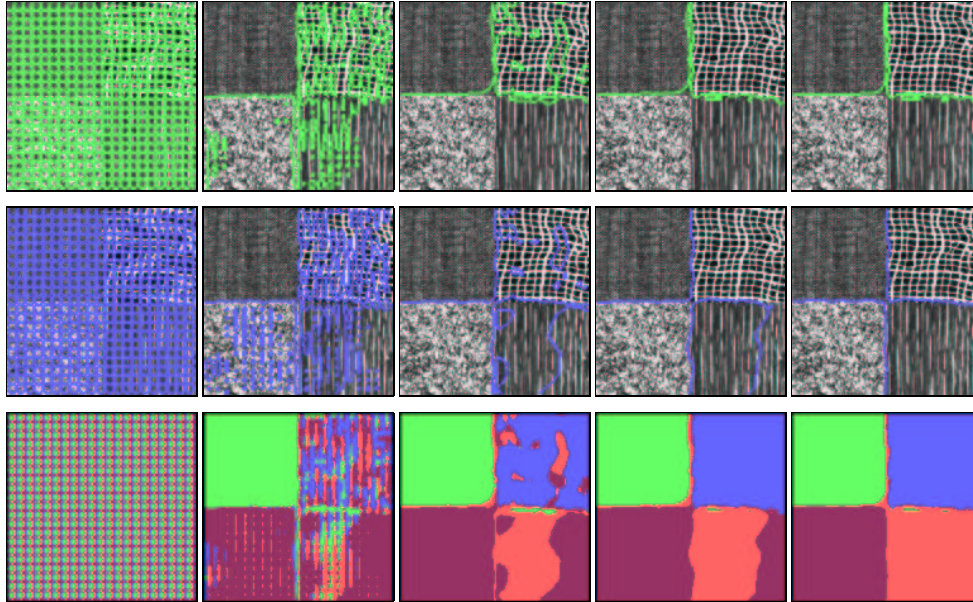
Figure 11: Segmentation in 4 regions of a synthetical image composed by 4 Brodatz textures (classification rate is 97%). FIRST ROW: Zero crossings of the first level set. SECOND ROW: Zero crossings of the second level set. LAST ROW: Corresponding segmentation.

even rather small objects like the pedestrians or the cyclist in Fig.13 can be tracked by the algorithm. The objects are lost in a natural way when they disappear behind an obstacle (magenta object) or the information used for the segmentation becomes very unreliable (green object).

Note that the newly appearing car in Fig.13 is merged with the existing one. This is because it has not been present in the detection part of the algorithm and would actually be ignored. However, since the car gets very close to the existing one and also has approximately the same speed and direction of motion, it is considered as the existing car changing its shape.

Colour information has been integrated on the tracking of the *Hand* sequence where one hand is moving in front of a complicated background. Despite noise due to the camera, we obtain a good tracking of the hand, results are shown in Fig.14. Only a small region corresponding to the shadow gets into the moving object in some images.

We also tested the performance of our approach in the presence of occluding objects. Our variational formulation introduces a coupling between the level sets and permits to avoid the overlapping of regions. It makes the level sets to compete when two objects are very close
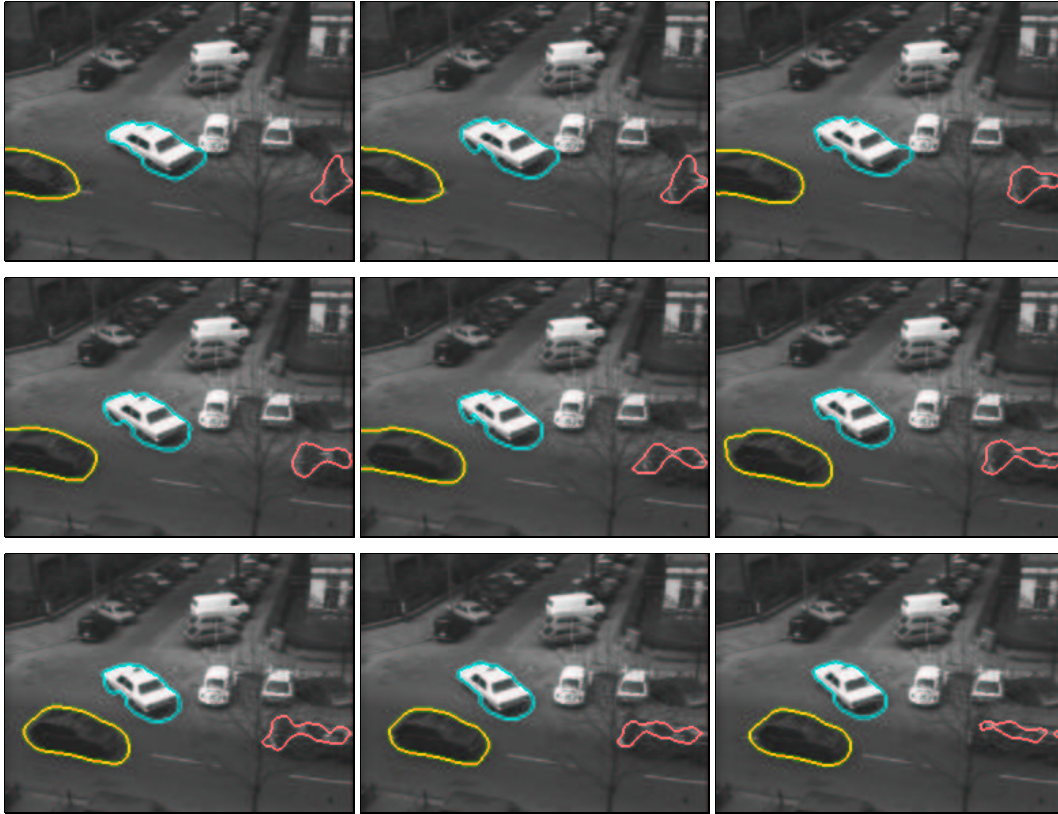
Figure 12: FROM LEFT TO RIGHT, TOP TO BOTTOM: Tracking result shown for 9 out of 20 images from the *Hamburg Taxi* sequence.

or when an occlusion occurs. Fig.15 demonstrates that this competition between objects works very well.

Occlusions also appear in the more complicated soccer sequence in Fig.16. Again the technique with the coupled level sets is able to handle this situation. There is an additional challenge compared to the sequences in Fig.12 and Fig.13. In contrast to those sequences the camera is no longer static here. However, using both optic flow components this causes no problems. Moreover, also the colour information helps to track the objects.

7   Conclusions and Perspectives

In this paper we have presented an unsupervised segmentation framework that is able to incorporate many different kinds of information. It has been possible to integrate colour, texture, and motion. Furthermore, it has been shown that both texture and motion can be obtained from the nonlinear structure tensor. Our way to compute the features, the coupled nonlinear diffusion with a special degenerated diffusivity, as well as the statistical region model are the basis for the good results. The multi-scale approach can further improve the robustness. Moreover, by applying the method to tracking it has been demonstrated that the approach is not only useful for pure segmentation but can also be adapted to similar problems.

In several experiments it has been shown that our method works very well with all images that are in accordance to our model assumptions. In natural images such assumptions may sometimes be violated. In order to be able to deal also with images where our approach fails so far, the assumption of only two regions has to be dropped, or at least the statistical model has to be extended. To find a good solution for this problem is a very challenging topic for future research. The same way one could think of another way to select the best features from the available information. Our approach prefers the information with the largest discrimination power. We plan to investigate whether there are better solutions. We also believe that it can be advantageous to combine our unsupervised technique with learning techniques from supervised approaches. An unsupervised method need not inevitably cope without learned data, it must only learn the data in an unsupervised manner.

Our approach consists of using jointly different cues in both parts of our method: in the feature extraction process by accounting for joint edges in all channels, and in the segmentation process by using the joint density probability of all features. Like humans do when analysing a scene (static or in motion), we extracted many kinds of information and integrated all these different cues in a general framework. Along all our study we tried to present a general and coherent framework to incorporate any kind of information using a unique approach. It takes both boundary and region properties into account, using two techniques: diffusion, constrained by boundary information, and segmentation, based on a region homogeneity criterion. The first technique appears to be very close to the methods proposed from studies on visual processing in the cerebral cortex [33, 20, 35]. For example, Grossberg *et al.* in [33] propose a coupled boundary detection/enhancement and filling-in method which may first look really different from the tools we used, but if we take a deeper look, the idea of a diffusion/filling-in process constrained by edge information/barriers is really close. Starting from this statement, our future research will focus on establishing links between geometric-based and neural-based image processing.

Acknowledgements

A Proof of rotation invariance

Consider an image consisting of two oriented textures T_1 and T_2 with an orientation difference of angle θ. Then the structure tensors of T_1 and T_2 will be

$$
J_0(T_1) = \begin{pmatrix} \sin^2 \phi & \sin \phi \cos \phi \\ \sin \phi \cos \phi & \cos^2 \phi \end{pmatrix}
$$

$$
J_0(T_2) = \begin{pmatrix} \sin^2(\phi+\theta) & \sin(\phi+\theta)\cos(\phi+\theta) \\ \sin(\phi+\theta)\cos(\phi+\theta) & \cos^2(\phi+\theta) \end{pmatrix}
$$

for a rotation angle ϕ of the image. The Euclidean distance between both will be

$$
\|J_0(T_1) - J_0(T_2))\|_2 = \sqrt{(\sin^2\phi - \sin^2(\phi+\theta))^2 + 2(\sin\phi\cos\phi - \sin(\phi+\theta)cos(\phi+\theta))^2 + (\cos^2\phi - \cos^2(\phi+\theta))^2}.
$$

For abbreviation we define

$$
\begin{aligned}
a &= \sin \phi \\
b &= \cos \phi \\
c &= \sin \theta \\
d &= \cos \theta.
\end{aligned}
$$

With some basic theorems from trigonometry, the Euclidean distance can be simplified to

$$
\begin{aligned}
\|J_0(T_1)-J_0(T_2)\|_2 &= \sqrt{(a^2-(ad+bc)^2)^2+2(ab-(ad+bc)(bd-ac))^2+(b^2-(bd-ac)^2)^2} \\
&= \sqrt{(a^2-a^2d^2-2abcd-b^2c^2)^2+2(ab-abd^2+a^2cd-b^2cd+abc^2)^2+(b^2-b^2d^2+2abcd-a^2c^2)^2} \\
&= \sqrt{(a^2c^2-2abcd-b^2c^2)^2+2(abc^2+a^2cd-b^2cd+abc^2)^2+(b^2c^2+2abcd-a^2c^2)^2} \\
&= \sqrt{(c^2(a^2-b^2)-2abcd)^2+2(2abc^2+cd(a^2-b^2))^2+(c^2(a^2-b^2)-2abcd)^2} \\
&= \sqrt{2(c^2(a^2-b^2)-2abcd)^2+2(2abc^2+cd(a^2-b^2))^2} \\
&= \sqrt{2(c^4(a^2-b^2)^2-4abc^3d(a^2-b^2)+4a^2b^2c^2d^2+4a^2b^2c^4+4abc^3d(a^2-b^2)+c^2d^2(a^2-b^2)^2)} \\
&= \sqrt{2(c^4(a^2-b^2)^2+4a^2b^2c^2d^2+4a^2b^2c^4+c^2d^2(a^2-b^2)^2)} \\
&= \sqrt{2((a^2-b^2)^2(c^4+c^2d^2)+4a^2b^2(c^4+c^2d^2))} \\
&= \sqrt{2((c^4+c^2d^2)(1+4(a^4-a^2+a^2b^2)))} \\
&= \sqrt{2((c^4+c^2d^2)(1+4a^2(a^2+b^2-1)))} \\
&= \sqrt{2(c^4+c^2d^2)} \\
&= \sqrt{2c^2(c^2+d^2)} \\
&= \sqrt{2c^2} \\
&= \sqrt{2}\sin\theta
\end{aligned}
$$

what is independent from ϕ.

References

[1] D. Adelseinsson and J. Sethian. A fast level set method for propagating interfaces. *Journal of Computational Physics*, 118:269–277, 1995.

[2] F. Andreu, C. Ballester, V. Caselles, and J. M. Mazón. Minimizing total variation flow. *Differential and Integral Equations*, 14(3):321–360, Mar. 2001.

[3] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, Feb. 1994.

[4] J. Bigün, G. H. Granlund, and J. Wiklund. Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8):775–790, Aug. 1991.

[5] P. Brodatz. *Textures: a Photographic Album for Artists and Designers*. Dover, New York, 1966.

[6] T. Brox and J. Weickert. Nonlinear matrix diffusion for optic flow estimation. In L. Van Gool, editor, *Pattern Recognition*, volume 2449 of *Lecture Notes in Computer Science*, pages 446–453. Springer, Berlin, 2002.

[7] T. Brox, M. Welk, G. Steidl, and J. Weickert. Equivalence results for TV diffusion and TV regularisation. In L. D. Griffin, editor, *Scale-Space Theories in Computer Vision*, Lecture Notes in Computer Science. Springer, Berlin, 2003. To appear.

[8] T. Chan, B. Sandberg, and L. Vese. Active contours without edges for vector-valued images. *Journal of Visual Communication and Image Representation*, 11:130–141, 2000.

[9] T. F. Chan and L. A. Vese. A level set algorithm for minimizing the Mumford–Shah functional in image processing. In *Proc. First IEEE Workshop on Variational and Level Set Methods in Computer Vision*, pages 161–168, Vancouver, Canada, July 2001. IEEE Computer Society Press.

[10] Y. Chen, H. D. Tagare, S. Thiruvenkadam, F. Huang, K. S. Gopinath, R. W. Briggs, and E. A. Geisler. Using prior shapes in geometric active contours in a variational framework. *International Journal of Computer Vision*, 50(3):315–328, Dec. 2002.

[11] D. Chopp. Computing minimal surfaces via level set curvature-flow. *Journal of Computational Physics*, 106:77–91, 1993.

[12] D. Cremers, F. Tischhäuser, J. Weickert, and C. Schnörr. Diffusion snakes: introducing statistical shape knowledge into the Mumford-Shah functional. *International Journal of Computer Vision*, 50(3):295–313, Dec. 2002.

[13] G. R. Cross and A. K. Jain. Markov Random Field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5:25–39, 1983.

[14] S. Di Zenzo. A note on the gradient of a multi-image. *Computer Vision, Graphics and Image Processing*, 33:116–125, 1986.

[15] F. Dibos and G. Koepfler. Global total variation minimization. *SIAM Journal on Numerical Analysis*, 37(2):646–664, 2000.

[16] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Proc. ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, pages 281–305, Interlaken, Switzerland, June 1987.

[17] D. Gabor. Theory of communication. *J. IEEE*, 93:429–459, 1946.

[18] G. Gerig, O. Kübler, R. Kikinis, and F. A. Jolesz. Nonlinear anisotropic filtering of MRI data. *IEEE Transactions on Medical Imaging*, 11:221–232, 1992.

[19] G. H. Granlund and H. Knutsson. *Signal Processing for Computer Vision*. Kluwer, Dordrecht, 1995.

[20] T. Hansen and H. Neumann. Neural mechanisms for representing surface and contour features. In S. Wermter, J. Austin, and D. Willshaw, editors, *Emergent Neural Computational Architectures Based on Neuroscience*, volume 2036 of *Lecture Notes in Computer Science*, pages 139–153. Springer, Berlin, 2001.

[21] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[22] S. Jehan-Besson, M. Barlaud, and G. Aubert. Constrained flows of matrix-valued functions: application to diffusion tensor regularization. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Computer Vision – ECCV 2002*, volume 2352 of *Lecture Notes in Computer Science*, pages 365–380. Springer, Berlin, 2002.

[23] S. Keeling and R. Stollberger. Nonlinear anisotropic diffusion filters for wide range edge sharpening. *Inverse Problems*, 18:175–190, Jan. 2002.

[24] J. Kim, J. Fisher, A. Yezzi, M. Cetin, and A. Willsky. Nonparametric methods for image segmentation using information theory and curve evolution. In *IEEE International Conference on Image Processing*, Rochester, NY, Sept. 2002.

[25] D. V. Kovkov and V. I. Tsurkov. Singular solutions to the problem with initial conditions for the equation of anisotropic diffusion. *Journal of Computer and Systems Sciences International*, 39(4):497–502, 2000.

[26] M. E. Leventon, W. E. L. Grimson, and O. Faugeras. Statistical shape infuence in geodesic active contours. In *Proc. 2000 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 316–323, Hilton Head, SC, June 2000. IEEE Computer Society Press.

[27] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer, Boston, 1994.

[28] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. Seventh International Joint Conference on Artificial Intelligence*, pages 674–679, Vancouver, Canada, Aug. 1981.

[29] J. Malik, S. Belongie, T. Leung, and J. Shi. Contour and texture analysis for image segmentation. *International Journal of Computer Vision*, 43(1):7–27, 2001.

[30] R. Malladi, J. Sethian, and B. Vermuri. Shape modeling with front propagation: A level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):158–175, Feb. 1995.

[31] S. Marcelja. Mathematical description of the response of simple cortical cells. *J. Opt. Soc. Amer.*, 70:1297–1300, 1980.

[32] M. Middendorf and H.-H. Nagel. Empirically convergent adaptive estimation of grayvalue structure tensors. In L. Van Gool, editor, *Pattern Recognition*, volume 2449 of *Lecture Notes in Computer Science*, pages 66–74. Springer, Berlin, 2002.

[33] E. Mingolla, W. Ross, and S. Grossberg. A neural network for enhancing boundaries and surfaces in synthetic aperture radar images. *Neural Networks*, 12(3):499–511, 1999.

[34] A. Mitiche and P. Bouthemy. Computation and analysis of image motion: a synopsis of current problems and methods. *International Journal of Computer Vision*, 19(1):29–55, July 1996.

[35] H. Neumann, L. Pessoa, and T. Hansen. Visual filling-in for computing perceptual surface properties. *Biological Cybernetics*, 85(5):355–369, 2001.

[36] S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton–Jacobi formulations. *Journal of Computational Physics*, 79:12–49, 1988.

[37] N. Paragios and R. Deriche. Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(3):266–280, Mar. 2000.

[38] N. Paragios and R. Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *International Journal of Computer Vision*, 46(3):223, 2002.

[39] N. Paragios and D. R. Geodesic active regions: A new paradigm to deal with frame partition problems in computer vision. *Journal of Visual Communication and Image Representation*, 13(1/2):249–268, March/June 2002.

[40] D. Peng, B. Merriman, S. Osher, H. Zhao, and M. Kang. A PDE-based fast local level set method. *Journal of Computational Physics*, 155(2):410–438, 1999.

[41] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:629–639, 1990.

[42] C. Peterson and B. Söderberg. A new method for mapping optimization problems onto neural networks. *International Journal of Neural Systems*, 1(1):3–22, 1989.

[43] J. Puzicha, T. Hofmann, and J. Buhmann. Deterministic annealing: fast physical heuristics for real-time optimization of large systems. In *Proc. 15th IMACS World Conference on Scientific Computation, Modelling and Applied Mathematics*, Berlin, 1997.

[44] A. R. Rao and B. G. Schunck. Computing oriented texture fields. *CVGIP: Graphical Models and Image Processing*, 53:157–185, 1991.

[45] M. Rousson, T. Brox, and R. Deriche. Active unsupervised texture segmentation on a diffusion based feature space. In *Proc. 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Madison, WI, June 2003. IEEE Computer Society Press. To appear.

[46] M. Rousson and R. Deriche. A variational framework for active and adaptive segmentation of vector valued images. In *Proc. IEEE Workshop on Motion and Video Computing*, Orlando, Florida, Dec. 2002.

[47] M. Rousson and N. Paragios. Shape priors for level set representations. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Computer Vision – ECCV 2002*, volume 2351 of *Lecture Notes in Computer Science*, pages 78–92. Springer, Berlin, 2002.

[48] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[49] C. Sagiv, N. A. Sochen, and Y. Y. Zeevi. Texture segmentation via a diffusion-segmentation scheme in the gabor feature space. In *Proc. Texture 2002, 2nd International Workshop on Texture Analysis and Synthesis*, Copenhagen, June 2002.

[50] B. Sandberg, T. Chan, and L. Vese. A level-set and Gabor-based active contour algorithm for segmenting textured images. Technical Report 39, Math. Dept. UCLA, Los Angeles, USA, July 2002.

[51] C. Stiller and J. Konrad. Estimating motion in image sequences. *IEEE Signal Processing Magazine*, 16:70–91, 1999.

[52] D. Tschumperlé and R. Deriche. Diffusion tensor regularization with contraints preservation. In *Proc. 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 948–953, Kauai, HI, Dec. 2001. IEEE Computer Society Press.

[53] D. Tschumperlé and R. Deriche. Orthonormal vector sets regularization with PDE's and applications. *International Journal of Computer Vision (IJCV, Special Issue VLSM)*, 50:237–252, 2002.

[54] V. I. Tsurkov. An analytical model of edge protection under noise suppression by anisotropic diffusion. *Journal of Computer and Systems Sciences International*, 39(3):437–440, 2000.

[55] J. Weickert. A review of nonlinear diffusion filtering. In B. ter Haar Romeny, L. Florack, J. Koenderink, and M. Viergever, editors, *Scale-Space Theory in Computer Vision*, volume 1252 of *Lecture Notes in Computer Science*, pages 3–28. Springer, Berlin, 1997.

[56] J. Weickert and B. Benhamouda. A semidiscrete nonlinear scale-space theory and its relation to the Perona–Malik paradox. In F. Solina, W. G. Kropatsch, R. Klette, and R. Bajcsy, editors, *Advances in Computer Vision*, pages 1–10. Springer, Wien, 1997.

[57] J. Weickert and T. Brox. Diffusion and regularization of vector- and matrix-valued images. In M. Z. Nashed and O. Scherzer, editors, *Inverse Problems, Image Analysis, and Medical Imaging*, volume 313 of *Contemporary Mathematics*, pages 251–268. AMS, Providence, 2002.

[58] J. Weickert, B. M. ter Haar Romeny, and M. A. Viergever. Efficient and reliable schemes for nonlinear diffusion filtering. *IEEE Transactions on Image Processing*, 7(3):398–410, Mar. 1998.

[59] A. Yezzi, A. Tsai, and A. Willsky. A statistical approach to snakes for bimodal and trimodal imagery. In *Proc. 7th International Conference on Computer Vision*, volume 2, pages 898–903, Kerkyra, Greece, Sept. 1999.

[60] H. Zhao, T. Chan, B. Merriman, and S. Osher. A variational level set approach to multiphase motion. *Journal of Computational Physics*, 127:179–195, 1996.

[61] S. Zhu and A. Yuille. Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900, Sept. 1996.
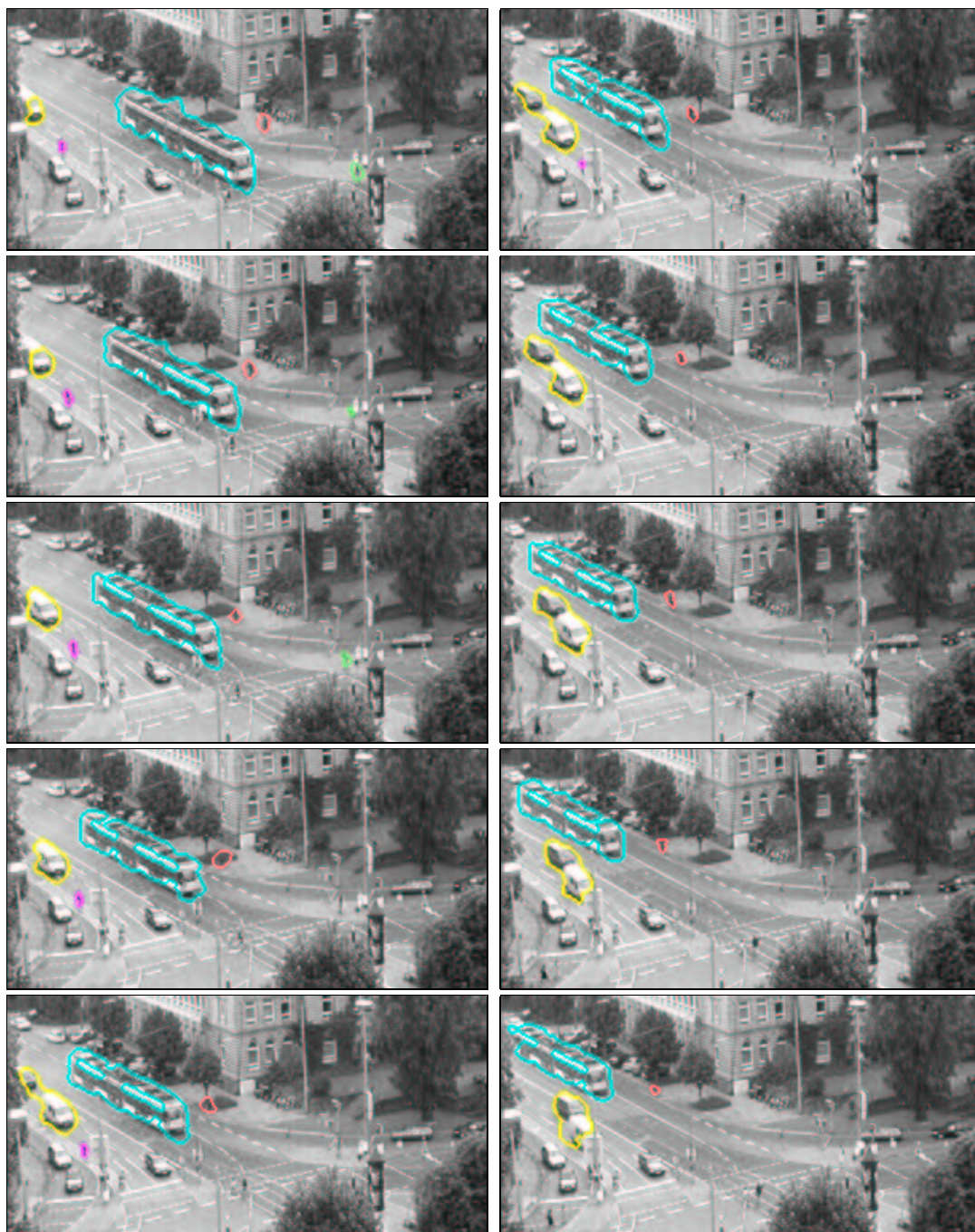
Figure 13: FROM TOP TO BOTTOM, LEFT TO RIGHT: Tracking result shown for 10 out of 100 images from the traffic sequence.
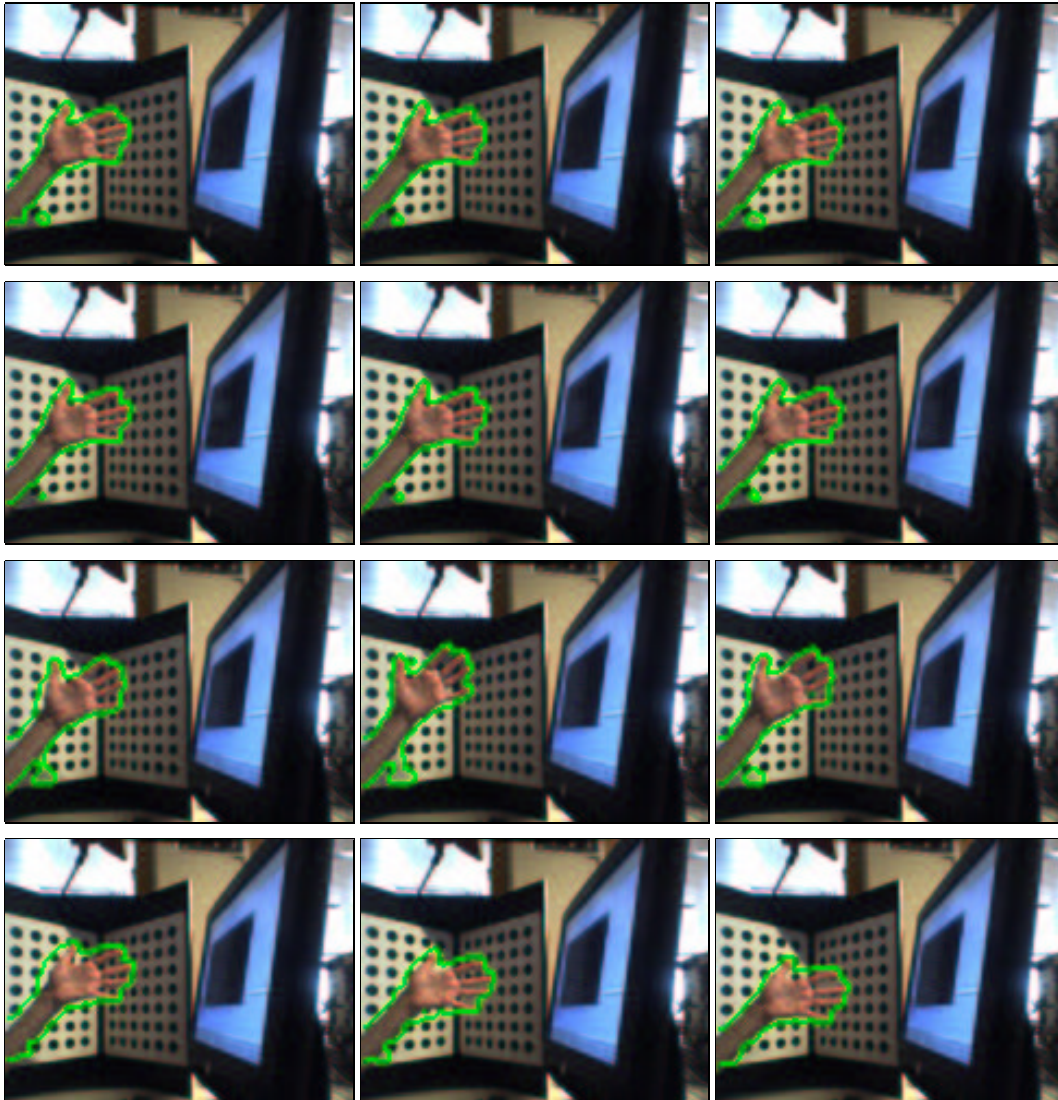
Figure 14: FROM LEFT TO RIGHT, TOP TO BOTTOM: Tracking result shown for 12 out of 30 images of the *Hand* sequence.
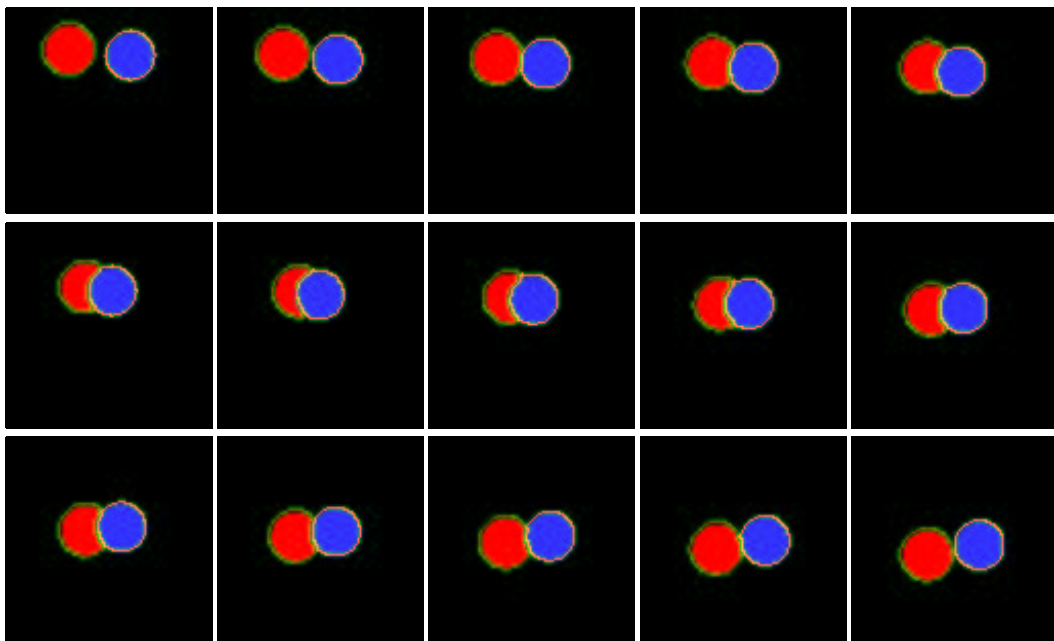
Figure 15: FROM LEFT TO RIGHT, TOP TO BOTTOM: Tracking result for a synthetic sequence. Occlusions can be handled correctly.

Figure 16: FROM LEFT TO RIGHT, TOP TO BOTTOM: Tracking result shown for 12 out of 27 images of the *Soccer* sequence.