

Universität des Saarlandes



Fachrichtung 6.1 – Mathematik

Preprint Nr. 147

**Colour, Texture, and Motion in Level Set Based  
Segmentation and Tracking**

Thomas Brox, Mikaël Rousson,  
Rachid Deriche and Joachim Weickert

Saarbrücken 2005



# Colour, Texture, and Motion in Level Set Based Segmentation and Tracking

**Thomas Brox**

Mathematical Image Analysis Group,  
Faculty of Mathematics and Computer Science,  
Saarland University, Building 27.1,  
66041 Saarbrücken, Germany  
brox@mia.uni-saarland.de

**Mikaël Rousson**

Siemens Corporate Research,  
Princeton, NJ 08540, USA  
mikael.rousson@siemens.com

**Rachid Deriche**

I.N.R.I.A, Projet ROBOTVIS/ODYSSEE,  
2004 Rte des Lucioles, BP 93,  
06902 Sophia Antipolis, France  
Rachid.Deriche@sophia.inria.fr

**Joachim Weickert**

Mathematical Image Analysis Group,  
Faculty of Mathematics and Computer Science,  
Saarland University, Building 27.1,  
66041 Saarbrücken, Germany  
weickert@mia.uni-saarland.de

Edited by  
FR 6.1 – Mathematik  
Universität des Saarlandes  
Postfach 15 11 50  
66041 Saarbrücken  
Germany

Fax: + 49 681 302 4443  
e-Mail: [preprint@math.uni-sb.de](mailto:preprint@math.uni-sb.de)  
WWW: <http://www.math.uni-sb.de/>

## Abstract

This paper introduces an approach for the extraction and combination of different cues in a level set based image segmentation framework. Apart from the image grey value or colour, we suggest to add its spatial and temporal variations, which may provide important further characteristics. It often turns out that the combination of colour, texture, and motion permits to distinguish object regions that cannot be separated by one cue alone. We propose a two-step approach. In the first stage, the input features are extracted and enhanced by applying coupled nonlinear diffusion. This ensures coherence between the channels and deals with outliers. We use a nonlinear diffusion technique, closely related to total variation flow, but being strictly edge enhancing. The resulting features are then employed for a vector-valued front propagation based on level sets and statistical region models that approximate the distributions of each feature. The application of this approach to two-phase segmentation is followed by an extension to the tracking of multiple objects in image sequences.

*Key Words:* Image segmentation; tracking; level set methods; nonlinear diffusion; texture; motion

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Feature Extraction and Enhancement</b>	<b>4</b>
2.1	Texture Features from a Nonlinear Structure Tensor . . . . .	4
2.2	Motion Features: Optic Flow . . . . .	6
2.3	Feature Combination . . . . .	7
<b>3</b>	<b>Two-Region Segmentation</b>	<b>8</b>
3.1	Gaussian Density Estimate . . . . .	8
3.2	Nonparametric Density Estimate . . . . .	9
3.3	Smoothness Term . . . . .	9
3.4	Implicit Weighting of Features . . . . .	10
3.5	Multiple Scales . . . . .	11
3.6	Implementation Aspects . . . . .	11
<b>4</b>	<b>Experiments on Segmentation</b>	<b>12</b>
4.1	Colour and Texture Segmentation . . . . .	12
4.2	Motion Segmentation . . . . .	13
<b>5</b>	<b>Simultaneous Tracking of Multiple Objects</b>	<b>14</b>
5.1	Coupled Curve Evolution . . . . .	14
5.2	Experiments . . . . .	18
<b>6</b>	<b>Conclusion</b>	<b>18</b>

# 1 Introduction

The segmentation and tracking of objects in image sequences is a challenging problem in image analysis. Robust algorithms exist for many industrial applications, yet efficiency and robustness are most of the time obtained by adding strong constraints issued from prior knowledge and the controlling of the environmental conditions. Numerous approaches for tracking, for instance, propose to use models of the moving objects or of the background. Although such approaches usually turn out to be quite robust, they often lack generality. Relying on sharp constraints, they are designed to tackle precisely defined problems. Moreover, they need a supervised learning stage, and their performance of the method highly depends on the quality of the learning samples.

In this paper, we want to use as little prior knowledge as possible. In order to nevertheless succeed in the challenging task of image segmentation, it is therefore necessary to use as much information of an image as possible. Thereby, the combination of different cues is essential for the robustness of segmentation and tracking. The combination is to be designed such that if one cue is corrupted, the method can still rely on the remaining information.

Before one can combine different types of information, this information has first to be extracted. There is no such problem in case of primary features like the grey value or colour, which are already given as input data. However, as soon as secondary features like texture or motion play a role, there are many possibilities how to extract them from the image. In the ideal case, the features should be highly discriminative while inducing as little data load as possible.

In the field of texture analysis the most frequently used methods are based on Gabor filters [31, 67, 68, 57]. Neurobiology indicates Gabor filters to be important in human vision [45]. However, they 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) [24]. 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. An alternative has been proposed in [6] where the structure tensor, also known as second moment matrix, is used for this task. It yields only three feature channels and thus provides a significantly larger information content per channel than Gabor features. Recently, the structure tensor has been extended to so-called nonlinear structure tensors based on nonlinear diffusion [13, 10], which are data-adaptive and therefore more accurate in the vicinity of discontinuities. In the present paper, we make use of this concept in order to extract texture features and to combine them with other cues.

Considering motion, the optic flow is the principal method to extract this information. Optic flow estimation is a complete research area on its own, and there exist plenty of different techniques; see e.g. [4, 47, 71] for overviews. Interestingly, the nonlinear structure tensor, already applied for deriving texture features, can also be used here [11, 13].

The combination of features is realised at two different stages: first, nonlinear diffusion is applied to the extracted feature vector, thus inducing a coupled smoothing of all feature channels. This permits to deal with possible outliers, and the spatial coherence between various information is highly improved. Afterwards, these smoothed versions of the features are combined in joint probability density functions that describe the interior of regions. These densities are approximated by Gaussian densities and nonparametric Parzen estimates [59, 40]. We will show that by choosing appropriate statistical models one obtains an implicit weighting of the features by their discriminative power.

Maximising the total a-posteriori probability finally leads to a partitioning of the image domain. For this optimisation problem, we employ level set methods [26, 53, 69, 51, 52]. In

recent years, level sets have become a very popular tool for image segmentation. Starting with edge based active contour models [16, 44, 17, 39], they have later been extended to region based models [55, 19, 18, 20, 57], which can be related to the energy functionals proposed by Mumford and Shah [48] and Zhu and Yuille [81]. Formulating these functionals in the level set framework has several advantages. First of all, the embedding of the 1-D curve into the 2-D level set function as its zero-level line allows for connecting constraints on the curve to constraints on the regions. Furthermore, topological changes can naturally be handled with level set representations. In particular, regions need not necessarily be connected and parts of a region can split or merge. Finally, the optimisation of the partitioning problem by evolving curves is perfectly suited for applying the method to tracking problems: once the object contour in one image is found, only few iterations are necessary to adapt the curve to the shape and position of the object in the next frame.

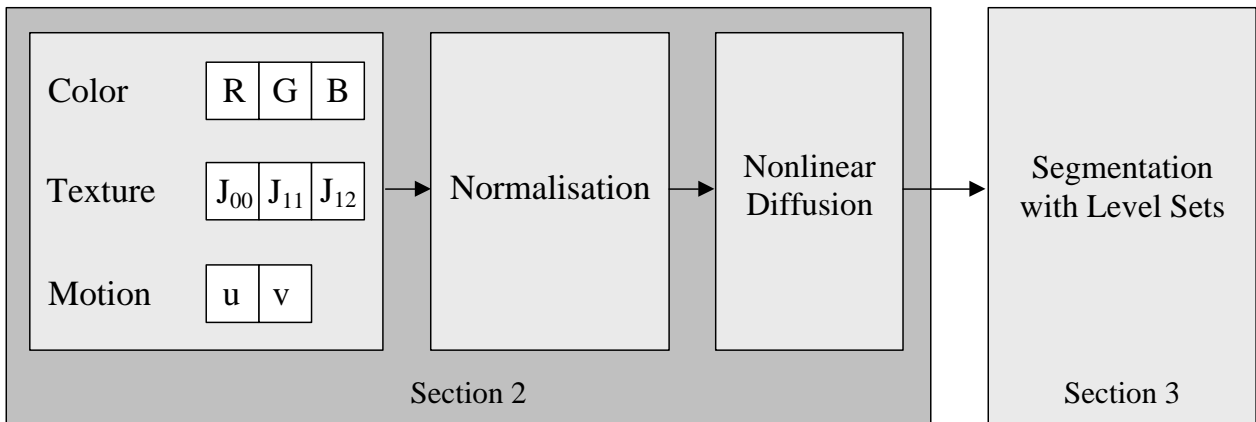


Figure 1: Summary of our approach.

**Related work.** The integration of different types of information has been extensively studied in various fields in the past. We can refer to studies in biology [49], neuroscience [33], statistics [70], and computer vision [5, 70]. Here, image cues will stand for the information to integrate, but it may also be the fusion of data obtained from different sensors, which is a whole domain in itself.

One possibility is to create a feature vector from all the cues and run classical clustering algorithms in the resulting feature space [5, 54]. In [5], a feature space for image segmentation is obtained from the extraction of texture and colour cues. Then, an EM algorithm is employed to estimate the parameters of a Gaussian mixture in the corresponding 6-D feature space. Similarly, it has been suggested to maximise the a-posteriori probability, or equivalently, to minimise the negative log-likelihood [81, 58]. With the assumption of no correlation between cues, a global energy is defined by summing the log-likelihoods for each single cue. Auto-adjustment of these weights has been proposed in the case of image segmentation in [35, 38] and for visual tracking in [70].

Alternatively, one can consider different cues successively instead of in parallel [41, 25]. In [41], e.g., first a coarse segmentation of the moving objects is obtained from using motion information, and then the segmentation is refined by considering the local image gradient. In [25], the authors deal with colour and texture images by first quantizing the image in the colour space, and then introducing a spatial criterion to obtain the segmentation. We refer to Clark and Yuille [22] for a classification of the different approaches in cues integration.

The present article comprises and extends work that has been presented on conferences. The variational framework based on level sets has been introduced in [64]. In [63] it has been applied to texture segmentation based on the nonlinear structure tensor. Finally, in [9], colour, texture, and motion information have been integrated. In particular, the present paper extends these concepts by: (i) a detailed investigation of feature coupling, including the way of normalisation, (ii) by showing that with a Gaussian region model, the partitioning becomes independent from the contrast of the features and comprises an implicit weighting of features by their discriminative power, (iii) by embedding our method in a multi-scale setting, and finally (iv) by a significantly extended experimental evaluation.

**Paper organisation.** In the following section, the extraction of the texture and motion features will be presented. This also comprises the combination of all features in a single feature vector. After that, in Section 3, the variational formulation for image partitioning based on level set methods will be introduced. Section 4 shows experiments on image segmentation using different feature combinations. Finally, Section 5 introduces an extension that allows the tracking of multiple objects. Experimental results are shown for this application as well. The paper is concluded by a summary.

## 2 Feature Extraction and Enhancement

### 2.1 Texture Features from a Nonlinear Structure Tensor

The nonlinear structure tensor introduced in [78, 13] is based on the classic linear structure tensor [30, 6, 62, 42, 34]

$$J_\rho = K_\rho * (\nabla I \nabla I^\top) = \begin{pmatrix} K_\rho * I_x^2 & K_\rho * I_x I_y \\ K_\rho * I_x I_y & K_\rho * I_y^2 \end{pmatrix} \quad (1)$$

where  $K_\rho$  is a Gaussian kernel with standard deviation  $\rho$ ,  $I$  is a grey value image, and subscripts denote partial derivatives. In the case of a colour image  $I = (I_1, I_2, I_3)$ , all channels are taken into account by summing the tensor products of the particular channels [27]:

$$J_\rho = K_\rho * \left( \frac{1}{3} \sum_{i=1}^3 \nabla I_i \nabla I_i^\top \right). \quad (2)$$

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 to address this problem is the replacement of the Gaussian smoothing by nonlinear diffusion. This is achieved by applying the scheme of nonlinear matrix-valued diffusion introduced in [72, 73], 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 [61]. For a review we refer to [75, 76]. The main idea is to reduce the smoothing in the presence of edges. The resulting diffusion equation is

$$\partial_t u = \operatorname{div} (g(|\nabla u|) \nabla u) \quad (3)$$

with  $u(t = 0)$  being the initial image and  $g$  a decreasing *diffusivity function* that is specified further below.



Eq.3 can only be used with scalar-valued data like a grey value image. Gerig et al. [32] introduced a version of nonlinear diffusion for vector-valued data

$$\partial_t u_i = \operatorname{div} \left( g \left( \sum_{k=1}^N |\nabla u_k|^2 \right) \nabla u_i \right) \quad \forall i = 1, \dots, N \quad (4)$$

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. A recent survey on vector-valued diffusion can be found in [74].

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 [72].

An important issue is the appropriate choice of the diffusivity function  $g$ . In [13], diffusivities that lead to total variation flow [3] or related types of diffusion have been suggested for the nonlinear structure tensor

$$g(|\nabla u|) = \frac{1}{|\nabla u|^p} \quad (5)$$

with  $p \in \mathbb{R}$  and  $p \geq 1$ . These diffusivity functions include for  $p = 1$  *total variation (TV) flow* [3, 28], a diffusion filter that is strongly related, and in the space-discrete, 1-D setting even equivalent, to TV regularisation [66, 14]. For  $p = 2$ , one obtains the so-called *balanced forward backward diffusion* introduced in [37]. These diffusivities fit the requirements of the structure tensor very well: there is no contrast parameter that needs to be optimised, the diffusion process removes oscillations, and experiments show that one obtains piecewise constant results. Furthermore, for  $p > 1$ , the underlying diffusion process not only preserves but even enhances edges. While such edge-enhancing diffusion processes are ill-posed in the continuous setting, spatial discretisations have been shown to be well-posed [77]. They are especially well suited for improving the coherence of the feature channels.

Unfortunately, the diffusivity functions in Eq.5 are singular in 0 and will hence lead to numerical problems when the gradient gets close to zero. This problem can be avoided by adding a small positive constant  $\epsilon$  to the denominator [1, 29]

$$g(|\nabla u|) = \frac{1}{(|\nabla u|^2 + \epsilon^2)^{\frac{p}{2}}}. \quad (6)$$

For implementation of the diffusion equation resulting from Eq.4 and Eq.6, the AOS scheme proposed in [79] can be applied. It allows efficient computation of such flows also for small  $\epsilon$ . For  $\epsilon$  in the area of 0.01, where the approximation of such kinds of flow is much better than for larger  $\epsilon$ , causing less blurring effects, the AOS scheme is around 3 orders of magnitude faster than a simple explicit scheme.

Our texture features consist of the three different components of the spatial structure tensor  $J_0$ , i.e.,  $F := (I_x^2, I_y^2, 2I_x I_y)$  with  $I$  being the image grey value. By applying Eq.4 with the diffusivity function of Eq.6 we obtain the smoothed components, i.e., the nonlinear structure tensor. In the case of colour images, the initial condition is extended to the sums  $F := (\frac{1}{3} \sum_i (I_i)_x^2, \frac{1}{3} \sum_i (I_i)_y^2, \frac{2}{3} \sum_i (I_i)_x (I_i)_y)$ . Note that in both cases the third vector component has to be weighted twice, since it appears twice in the matrix.

In the following, we will extend this feature vector by adding further cues, *before* nonlinear diffusion is applied to the vector. In particular, the vector will be extended by two motion components and the image grey value or colour.

## 2.2 Motion Features: Optic Flow

In optic flow estimation, one seeks at each point in the image the motion vector  $(u, v)$  that describes the shift of a pixel  $(x, y)$  in a frame at time  $z$  to another frame at time  $z + 1$ . To estimate this vector, one assumes that image structures do not alter their grey values during their motion. This is expressed by the optic flow constraint [36]

$$I_x u + I_y v + I_z = 0. \quad (7)$$

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 is necessary. Lucas and Kanade proposed to assume the optic flow vector to be constant within some neighbourhood  $B_\rho$  of size  $\rho$  [43]. The optic flow in some point  $(x_0, y_0)$  can then be estimated by the minimiser of the local energy function

$$E(u, v) = \frac{1}{2} \int_{B_\rho(x_0, y_0)} (I_x u + I_y v + I_z)^2 dx dy. \quad (8)$$

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

$$\begin{pmatrix} \int_{B_\rho} I_x^2 dx dy & \int_{B_\rho} I_x I_y dx dy \\ \int_{B_\rho} I_x I_y dx dy & \int_{B_\rho} I_y^2 dx dy \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} -\int_{B_\rho} I_x I_z dx dy \\ -\int_{B_\rho} I_y I_z dx dy \end{pmatrix}. \quad (9)$$

Instead of the sharp window  $B_\rho$ , often a convolution with a Gaussian kernel  $K_\rho$  is used yielding

$$\begin{pmatrix} K_\rho * I_x^2 & K_\rho * I_x I_y \\ K_\rho * I_x I_y & K_\rho * I_y^2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} -K_\rho * I_x I_z \\ -K_\rho * I_y I_z \end{pmatrix}. \quad (10)$$

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 or the aperture problem remains present. In such situations the smaller eigenvalue of the system matrix is close to 0, and 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 dense results are obtained in most cases. A technique that combines the Lucas-Kanade method with dense flow fields can be found in [15]. Obviously the entries of the linear system are five of the components of the spatio-temporal structure tensor

$$J_\rho = K_\rho * \begin{pmatrix} I_x^2 & I_x I_y & I_x I_z \\ I_x I_y & I_y^2 & I_y I_z \\ I_x I_z & I_y I_z & I_z^2 \end{pmatrix} \quad (11)$$

where  $z$  describes the temporal axis and  $K_\rho$  the Gaussian kernel with standard deviation  $\rho$ . In the case of colour, these will be extended by using the corresponding sums as described in the last section for the spatial structure tensor.

Like in case of the spatial structure tensor, also here the Gaussian smoothing can be replaced by matrix-valued nonlinear diffusion resulting in the nonlinear structure tensor as proposed in [11, 13]. For the diffusion, TV flow is applied, i.e.,  $p = 1$ . Employing the components of this nonlinear structure tensor in the system in Eq.10 yields optic flow estimates which are more precise at motion boundaries than with the original structure tensor. Related techniques have been suggested in [50, 46].

The resulting optic flow vector  $(u, v)$  can be added to our feature vector  $F$ . However, since different types of information are coupled here, one needs an appropriate way of normalizing the different components. This will be discussed in the following section.

## 2.3 Feature Combination

In the previous two sections it has been shown how texture and motion features can be extracted from the raw image data. Supplementing the colour information, one hence obtains a feature vector consisting of 3 colour channels (R, G, B), 3 texture channels (the components of the spatial structure tensor  $J_0$ , respecting its symmetry) and 2 optic flow channels (components  $u$  and  $v$  from the optic flow vector).

At this stage it becomes important to combine this information in a way that allows for the simplification of the data, for dealing with outliers, and for the closing of structures, using the cues of *all* channels. For this purpose, vector-valued diffusion according to Eq.4 is very well suited, as it couples all vector channels by a joint diffusivity. This way, the information of all channels is used to decide whether an edge is worth to be enhanced or not.

However, for a balanced coupling, Eq.4 assumes the values of all channels to have approximately the same range. Unfortunately, a simple normalisation of all channels to the same range, which would solve this problem immediately, is not a good approach. If a channel contains no information, e.g. there is no motion, the data in this channel should be constant. A normalisation, however, amplifies the noise in such a situation. Therefore, the problem of how all channels get approximately the same range has to be solved in a different manner.

Since the potential range of the grey value or colour channels is fixed (usually values are between 0 and 255), it is reasonable to normalise the other features to the same potential range. For the texture channels this can be done by replacing the structure tensor by its square root. Given the eigenvalue decomposition of the structure tensor  $J_0 = T(\lambda_i)T^\top$ , the square root can be computed by

$$\tilde{J}_0 := \sqrt{J_0} = T(\sqrt{\lambda_i})T^\top. \quad (12)$$

In the grey value case, this comes down to

$$\tilde{J}_0 = \begin{pmatrix} \frac{I_x}{|\nabla I|} & -\frac{I_y}{|\nabla I|} \\ \frac{I_y}{|\nabla I|} & \frac{I_x}{|\nabla I|} \end{pmatrix} \begin{pmatrix} |\nabla I| & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{I_x}{|\nabla I|} & \frac{I_y}{|\nabla I|} \\ -\frac{I_y}{|\nabla I|} & \frac{I_x}{|\nabla I|} \end{pmatrix} = \begin{pmatrix} \frac{I_x^2}{|\nabla I|} & \frac{I_x I_y}{|\nabla I|} \\ \frac{I_x I_y}{|\nabla I|} & \frac{I_y^2}{|\nabla I|} \end{pmatrix} = \frac{J_0}{|\nabla I|}. \quad (13)$$

With colour images, this simplification is not possible, due to the sum of matrices for each colour channel, and one has to compute  $\tilde{J}_0$  by means of a principal axis transformation according to Eq.12. Multiplying the components of  $\tilde{J}_0$  with the factor that is determined by the discretisation of the derivatives - for central differences and grid size  $h = 1$  this factor is 2 - ensures a potential range of the structure tensor components equivalent to that of the input image.

In the case of the optic flow channels, a comparable potential range can only be achieved by some fixed weighting. We weighted  $u$  and  $v$  with factor 64, so a displacement of 4 pixels per frame corresponds to the possible maximum value in the other channels. This is a reasonable choice, since above a velocity of 4 pixels per frame optic flow estimation with the mentioned method is unreliable, so image sequences with such large displacements should not be used for input. The weight may be adapted to the frame rate of the camera. For larger frame rates, it can be reasonable to choose larger weights.

With these considerations one can assemble the normalised feature vector

$$F(0) := (F_1, \dots, F_8) := \left( R, G, B, 2\tilde{J}_{0,11}, 2\tilde{J}_{0,22}, 4\tilde{J}_{0,12}, 64u, 64v \right). \quad (14)$$

It serves as initial condition for the nonlinear diffusion process according to Eq. 4 with a diffusivity from the family in Eq. 6. For our experiments we have chosen  $p = 1.6$ , though

other values of  $p > 1$  will also work. The larger  $p$ , the more important is the influence of the edge enhancement compared to the simplification effect. After diffusion time  $t$ , one obtains the smoothed feature vector  $F(t)$  that will be used in the following section for partitioning. In order to be able to keep the appearing parameter fixed for arbitrary images, the diffusion time should be set dependent on the image size  $|\Omega|$ . In our experiments we set  $t = 0.002|\Omega|$ .

### 3 Two-Region Segmentation

After the feature vector has been assembled, we will now discuss the partitioning of the image by means of this feature vector. To this end, we first assume the image to consist of only two regions: the object region and the background region. In Section 5 this assumption will be dropped for tracking.

A two-region segmentation splits the image domain  $\Omega$  into two disjoint regions  $\Omega_1$  and  $\Omega_2$  where the elements of  $\Omega_1$  and  $\Omega_2$  respectively are not necessarily connected. Let  $u : \Omega \rightarrow \mathbb{R}^N$  be the computed features of the image and  $p_{ij}(x)$  the conditional probability density function of a value  $u_j(x)$  to appear in region  $\Omega_i$ . 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* [57, 64]:

$$E(\Omega_i, p_{ij}) = - \sum_{j=1}^N \left( \int_{\Omega_1} \log p_{1j}(u_j(x)) dx + \int_{\Omega_2} \log p_{2j}(u_j(x)) dx \right) \rightarrow \min. \quad (15)$$

For minimizing this energy, a *level set function* is introduced. Let  $\Phi : \Omega \rightarrow \mathbb{R}$  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  $\Phi$  is the searched boundary between the two regions. We also introduce a regularised Heaviside function  $H(s)$  with  $\lim_{s \rightarrow -\infty} H(s) = 0$ ,  $\lim_{s \rightarrow \infty} H(s) = 1$ , and  $H(0) = 0.5$ . In particular, we choose the error function with  $\sigma = 1$ . Furthermore, let  $\chi_1(s) = H(s)$  and  $\chi_2(s) = 1 - H(s)$ . This allows to reformulate the functional in Eq.15

$$E(\Phi, p_{ij}) = - \sum_{i=1}^2 \sum_{j=1}^N \left( \int_{\Omega} \log p_{ij}(u_j) \chi_i(\Phi(x)) dx \right). \quad (16)$$

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

$$\partial_t \Phi = \sum_{j=1}^N \left( \log \frac{p_{1j}(u_j)}{p_{2j}(u_j)} H'(\Phi) \right) \quad (17)$$

where  $H'(s)$  is the derivative of  $H(s)$  with respect to its argument. The likelihood ratio involved here implies some implicit weighting of the channels. If there is a channel with a very high likelihood ratio, i.e., the regions can be well distinguished by this feature, it contributes much more to the level set function than a channel with a low likelihood ratio.

#### 3.1 Gaussian Density Estimate

The variational framework is not completely defined so far, since it still lacks the definition of the probability density functions (PDF)  $p_{ij}$ . A reasonable choice is a Gaussian function. Neglecting the correlation between the feature channels, this yields two parameters for the

PDF of each region  $i$  and channel  $j$ : mean  $\mu_{ij}$  and standard deviation  $\sigma_{ij}$ . Given the level set function  $\Phi$ , minimisation of Eq.16 with respect to these parameters (see [64]) leads to the update equations:

$$\mu_{ij} = \frac{\int_{\Omega} u_j(x)\chi_i(x)dx}{\int_{\Omega} \chi_i(x)dx}, \quad (18)$$

$$\sigma_{ij} = \sqrt{\frac{\int_{\Omega} (u_j(x) - \mu_{ij})^2 \chi_i(x)dx}{\int_{\Omega} \chi_i(x)dx}}. \quad (19)$$

Given the probability densities, the energy can be minimised with respect to  $\Phi$  using the gradient descent in Eq.17. Thus the segmentation process works according to the *expectation-maximisation* 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. This iterative process converges to the next local minimum depending on the initialisation.

### 3.2 Nonparametric Density Estimate

Although reasonable, choosing a Gaussian function as PDF is not the only possible solution. Kim et al. [40] proposed nonparametric Parzen density estimates for segmentation. Using discrete histograms, this approach comes down to smoothing the histograms computed for each region  $i$  and channel  $j$  by a Gaussian kernel  $K_{\sigma_h}$ :

$$p_{ij}(g) = K_{\sigma_h} * \frac{\int_{\Omega} \delta_{u_j(x)g} \chi_i(x)dx}{\int_{\Omega} \chi_i(x)dx} \quad (20)$$

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

This nonparametric density 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 generalise 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 starting with smoother versions of the objective function by embedding the minimisation in a multi-scale framework that is described in Section 3.5.

### 3.3 Smoothness Term

So far there has not been any smoothness constraint employed to the partitioning, 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 [20]

$$\alpha \int_{\Omega} |\nabla H(\Phi)| dx \quad (21)$$

where  $\alpha$  is a smoothness parameter. The steepest descent of this energy results in an additional evolution term:

$$\alpha H'(\Phi) \operatorname{div} \left( \frac{\nabla \Phi}{|\nabla \Phi|} \right) \quad (22)$$

It is closely related to mean curvature motion and automatically includes a narrow band. For the relation between this smoothness term and mean curvature motion we refer to [80]. Adding the smoothness term to the optimisation problem in Eq.16 yields the following energy:

$$E(\Phi, p_{ij}) = - \sum_{i=1}^2 \sum_{j=1}^N \left( \int_{\Omega} \log p_{ij}(u_j) \chi_i dx \right) + \alpha \int_{\Omega} |\nabla H(\Phi)| dx \rightarrow \min. \quad (23)$$

The corresponding gradient descent reads

$$\partial_t \Phi = H'(\Phi) \left( \sum_{j=1}^N \log \frac{p_{1j}(u_j)}{p_{2j}(u_j)} + \alpha \operatorname{div} \frac{\nabla \Phi}{|\nabla \Phi|} \right). \quad (24)$$

To get a geometric interpretation of this evolution equation, it can be linked to an explicit evolution of the curve  $\mathbf{C}$  subject to a force  $\beta$  that acts only in normal direction to the curve:  $\mathbf{C}_t \cdot \mathbf{N} = \beta$ . The link is made by replacing  $H'(\Phi)$  by  $|\nabla \Phi|$ , which is only a time rescaling (cf. [80]). Then the gradient descent (Eq.24) becomes

$$\partial_t \Phi = \beta |\nabla \Phi| \quad \text{with} \quad \beta = \sum_{j=1}^N \log \frac{p_{1j}(u_j)}{p_{2j}(u_j)} + \alpha \operatorname{div} \frac{\nabla \Phi}{|\nabla \Phi|}. \quad (25)$$

This is exactly the evolution obtained from  $\mathbf{C}_t \cdot \mathbf{N} = \beta$  using the level set representation (see [44] for details). So the evolution of  $\Phi$  can be interpreted as the evolution of  $\mathbf{C}$  subject to the force  $\beta$  defined above and acting only in normal direction to the contour.

### 3.4 Implicit Weighting of Features

Analyzing the influence of feature weighting on the segmentation leads to an interesting observation. While comparable ranges of all features are important for coupled nonlinear diffusion to work properly, the weighting of each channel has only little or even no influence on the discrimination properties of the curve evolution, if appropriate statistical models are employed. Consider the Gaussian probability density estimate. By multiplying a feature with a factor  $a$ , i.e., increasing its contrast by this factor, both the standard deviation  $\sigma_i$  and the difference of each value to the mean  $\mu_i$  are scaled by  $a$ . For each summand in Eq.17 this yields:

$$\begin{aligned} & -\log(\sqrt{2\pi} a \sigma_1) - \frac{(a(s-\mu_1))^2}{2(a\sigma_1)^2} + \log(\sqrt{2\pi} a \sigma_2) + \frac{(a(s-\mu_2))^2}{2(a\sigma_2)^2} \\ = & -\log(\sigma_1) - \frac{(s-\mu_1)^2}{2\sigma_1^2} + \log(\sigma_2) + \frac{(s-\mu_2)^2}{2\sigma_2^2}. \end{aligned} \quad (26)$$

This means, the region force is independent from  $a$ , and therefore contrast independent. Instead, the influence of a feature depends on how well it distinguishes the two regions, i.e. its discriminative power.

One can easily assure oneself that, apart from a linear scaling of the smoothing parameter  $\sigma_h$ , also the Parzen density estimate is contrast invariant. This shows that it is not necessary to try to automatically determine weights for the features, since this is already achieved by choosing a reasonable statistical model for the regions.

### 3.5 Multiple Scales

The energy defined above is in general highly non-convex. Thus, although the gradient descent according to Eq.24 converges to a local minimum, it is not ensured that this minimum coincides with the global minimum.

Usually, the number of local optima increases with a growing model complexity. This leads to a dilemma: on one hand, the model complexity is sought to be increased in order to accurately incorporate the image contents, but on the other hand a higher model complexity increases also the dependency of the result from a good initialisation.

As a remedy, one can consider continuation methods, also known under the term of graduated non-convexity (GNC) [7]. They provide a strategy to heuristically prevent a method from getting stuck in a local optimum. Continuation methods are based on the idea to start the optimisation with smoothed, simplified versions of the energy functional and to successively increase the level of detail until finally a solution for the original problem is obtained.

For image segmentation, one possibility to create a smoothed version from the original functional is to regard the input image at coarser resolution levels. This removes details from the image which may disturb the method in finding the global optimum. Additionally, the complexity of the region model can be reduced by replacing, for instance, the nonparametric density by a Gaussian density.

Both approaches can be combined. First a pyramid of successively downsampled versions of the input image is considered. Note that the multi-resolution optimisation strategy has a positive side effect apart from the reduced dependency of the method from the initialisation: it is considerably faster. At coarser resolution levels, the distance that the curve has to cover to reach its optimum position is much smaller than at the finest resolution level. Thus fewer iterations are necessary to converge to the steady state. Moreover, iterations at coarser levels are much faster, since there are fewer pixels. In the experiments, we consider two scales of the pyramid. The downsampling factor from one level to the next is 0.5.

Additionally to the pyramidal approach, we reduce the complexity of the region statistics at the coarse level. We start with a global Gaussian density estimate at the coarse scale while at the finer scale, when the curve is already close to the object boundary, we switch to the nonparametric Parzen density estimate, since it can describe the region statistics more accurately than the Gaussian model.

The pyramidal approach further opens the possibility to consider multiple scales for texture discrimination. In the single-scale case it is only possible to consider one scale of the texture without losing accuracy. Now it is feasible to add also the structure tensor components 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.2.

### 3.6 Implementation Aspects

The gradient descent of the energy functional in Eq.23 automatically leads to a narrow band approach, like it has been proposed, e.g., in [21, 2, 60, 56] 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

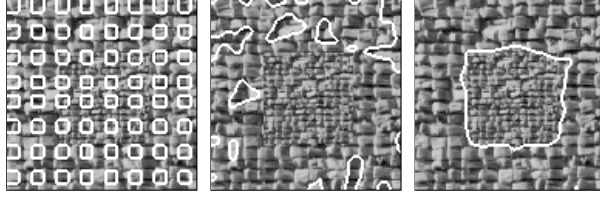


Figure 2: Segmentation of a multi-scale texture. LEFT: Initialisation of the curve. CENTER: Result considering only one scale. RIGHT: Result considering two scales.

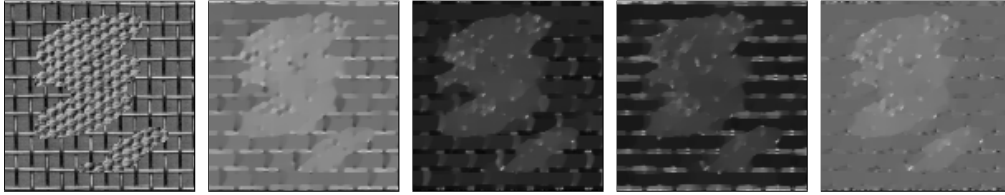


Figure 3: FROM LEFT TO RIGHT: (a) Texture image  $I$ . (b)  $F_0$  based on  $I$ . (c)  $F_1$  based on  $I_x^2$ . (d)  $F_2$  based on  $I_y^2$ . (e)  $F_3$  based on  $I_x I_y$ .

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, one can estimate the PDFs and implement the gradient descent given in Eq.24. Respecting the narrow band idea to speed up computation, we only consider pixels for the gradient descent where  $H'(\Phi(x)) > 0.1$ . Finally, the smoothness term is implemented as mean curvature motion restricted to the narrow band. Together with the multi-scale approach, 100 iterations at each scale are generally sufficient for the curve to become stationary. For a  $200 \times 200$  image this leads to computation times of around 5 seconds on contemporary hardware.

## 4 Experiments on Segmentation

### 4.1 Colour and Texture Segmentation

First the method is applied to texture segmentation. In this case there are four feature channels, one for the image grey value  $I$ , and three for the different components of the spatial structure tensor. An example of what the features look like after the coupled diffusion process is depicted in Fig.3.

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 some of the well-known Brodatz textures [8]. The results depicted in Fig.4 are very satisfactory. The same curve initialisation with 64 small circles distributed uniformly across the image domain was used for all segmentation examples; see Fig.2. 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. An alternative way to consider the scale of texture is to measure the local scale in the image and to supplement



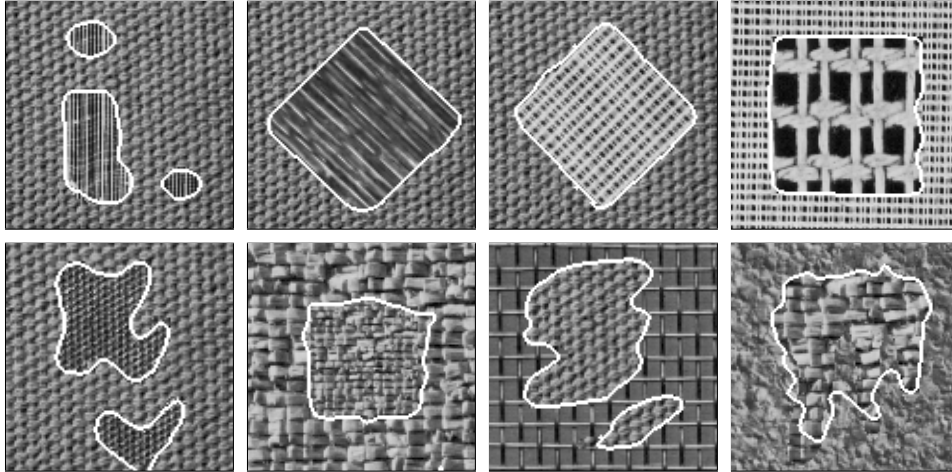


Figure 4: (a)-(h) FROM LEFT TO RIGHT, TOP TO BOTTOM: Segmentation results for synthetic texture images.

it as an additional feature as suggested in [12].

Although synthetic test images are best for verifying the possibilities and limits of a method, results achieved for real images are more interesting. Thus we applied our method also to several 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. Consequently, the feature vector consists of 6 entries, 3 for the colour channels and again 3 for the components of the structure tensor. The results in Fig.6 reveal that the usage of both colour and texture leads to very good results. The combination of texture and colour information can make the method still successful, when it would fail, if only colour *or* texture was considered, as shown in Fig.7.

## 4.2 Motion Segmentation

With image sequences, motion information becomes available and the optic flow components can be added to the feature vector. For a grey value sequence the feature vector therefore consists of 6 components, namely the grey value, the 3 components of the spatial structure tensor, and the two components of the optic flow.

For the special case of motion *detection*, it is often useful to weight the optic flow channels higher than the other channels. Otherwise, the resulting segmentation might not split the moving objects from the non-moving background, but some bright, textured parts from dark parts.

In Fig.9 we show results for different combinations of feature channels on the same test images (Fig.8) from a traffic scene <sup>1</sup>. While the optic flow yields reliable information for the coarse shape and location of the objects, adding intensity and texture information permits to increase the accuracy of the detection. The complete feature vector including all information is the only one that permits to separate the cars on the left, as shown in Fig.9(iii). Similarly, in Fig.10, a coloured and textured moving box is segmented with a very high accuracy when using all the features.

<sup>1</sup>The sequence was kindly provided by the *Institut für Algorithmen und kognitive Systeme* at the University of Karlsruhe, Germany. The size of the whole sequence is  $(700 \times 566 \times 1034)$

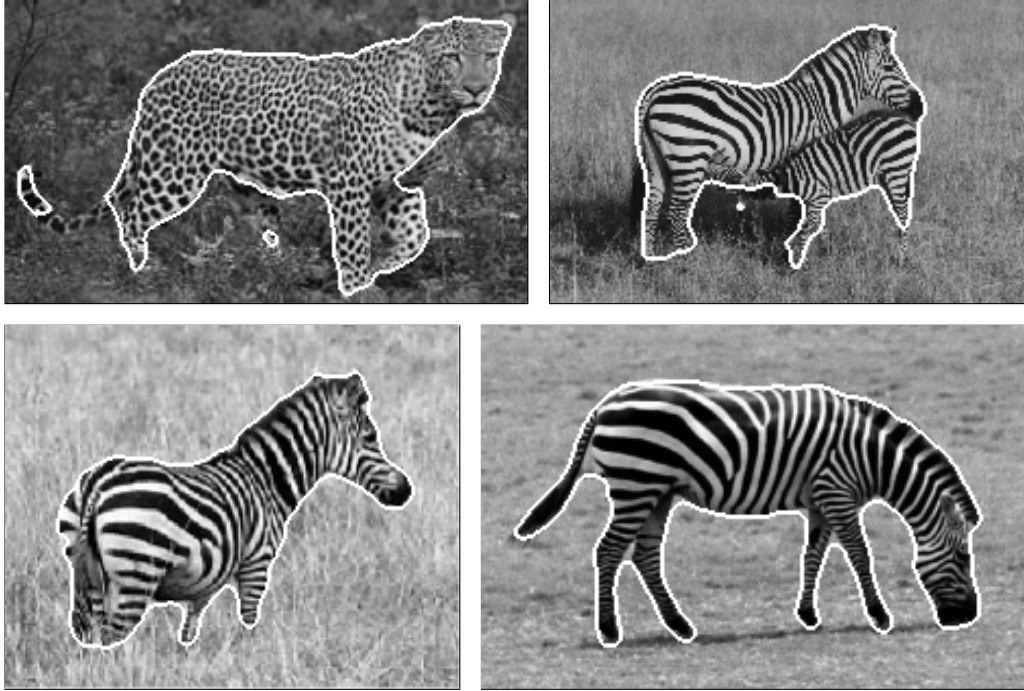


Figure 5: Segmentation results with some real-world images using texture and grey value information.

The complete feature vector permits also to capture small moving objects, as demonstrated in Fig.11. Pedestrians and bicycles can hardly be detected if only the optic flow is considered. When grey value and texture are supplemented, two more objects are detected (these objects are really moving in the sequence).

## 5 Simultaneous Tracking of Multiple Objects

When switching from segmentation to tracking, only few things change. The main issue is that for tracking the initial position and number of objects are known by definition, so local optima and the choice of the initialisation are not that problematic anymore. This permits to drop the assumption of a scene with only one object and the background, and it becomes possible to simultaneously track multiple objects. For this purpose, a new energy functional has to be introduced that allows for the coupled evolution of an arbitrary number of level sets, one for each object. The coupling is important in order to avoid the overlapping of objects.

### 5.1 Coupled Curve Evolution

Before the general case of  $M$  objects is introduced, we first consider the simplified case of only two object regions  $\Omega_1$  and  $\Omega_2$  with the background region  $B$ . One level set function is assigned to each object such that  $\Phi_i(x) > 0$  if and only if  $x \in \Omega_i$ . Adding the non-overlapping constraint, yields the characteristic functions:

$$\begin{aligned}\chi_1 &= H(\Phi_1)(1 - H(\Phi_2)) \\ \chi_2 &= H(\Phi_2)(1 - H(\Phi_1))\end{aligned}\tag{27}$$

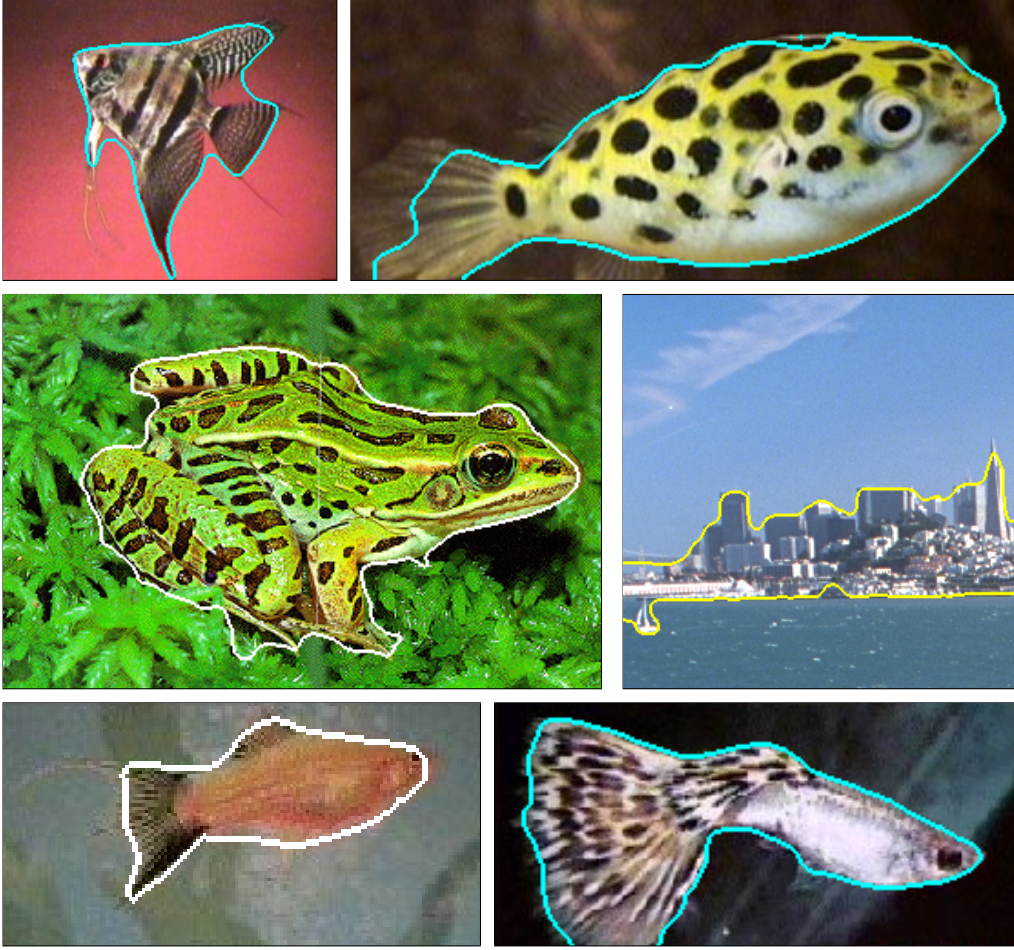


Figure 6: Segmentation results with some real-world images using texture and colour information.

The characteristic function assigned to the background is the remaining part of the image domain:

$$\begin{aligned}\chi_B &= H(\Phi_1)H(\Phi_2) + (1 - H(\Phi_1))(1 - H(\Phi_2)) \\ &= 1 - (\chi_1 + \chi_2)\end{aligned}\quad (28)$$

Let  $e_1$ ,  $e_2$ , and  $e_B$  be the log-likelihoods of a pixel to be in  $\Omega_1$ ,  $\Omega_2$ , and  $\Omega_B$ .

$$e_i = \sum_{j=1}^N \log p_{ij}(u_j) \quad (29)$$

Then the tracking of  $\Omega_1$  and  $\Omega_2$  at frame  $t$  is obtained by minimizing

$$\begin{aligned}E(\Phi_1, \Phi_2) &= - \int_{\Omega} e_1 \chi_1 + e_2 \chi_2 + e_B \chi_B \\ &= - \int_{\Omega} (e_1 - e_B) \chi_1 + (e_2 - e_B) \chi_2 + e_B\end{aligned}\quad (30)$$



Figure 7: Benefits of using joint 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 successive cutout images ( $229 \times 208$ ) of the traffic sequence.

By means of the Euler-Lagrange equations a system of evolution equations for the level set functions is obtained.

$$\begin{aligned}\partial_t \Phi_1 &= \delta(\Phi_1) ((e_1 - e_B)(1 - H(\Phi_2)) - (e_2 - e_B)H(\Phi_2)) \\ \partial_t \Phi_2 &= \delta(\Phi_2) ((e_2 - e_B)(1 - H(\Phi_1)) - (e_1 - e_B)H(\Phi_1))\end{aligned}\quad (31)$$

This model is slightly different from the model we proposed in [9] where the coupling between the level set functions was weaker. Overlapping has been avoided there, yet there has been no competition between objects when they are next to each other. Now in the case of overlapping, the system of evolution equations can be simplified to:

$$\begin{aligned}\partial_t \Phi_1 &= \delta(\Phi_1) (e_1 - e_2) / 2 \\ \partial_t \Phi_2 &= \delta(\Phi_2) (e_2 - e_1) / 2\end{aligned}\quad (32)$$

which effectively represents a competition between the two objects.

The extension to the general case of  $M$  objects is straightforward. Including the smoothness constraint for each level set leads to the energy functional

$$E = - \int_{\Omega} e_B + \sum_{i=1}^M (e_i - e_B) H(\Phi_i) \prod_{j \neq i} (1 - H(\Phi_j)) + \alpha |\nabla H(\Phi_i)| dx \quad (33)$$

and the evolution equations

$$\begin{aligned}\partial_t \Phi_k &= \delta(\Phi_k) (e_k - e_B) \prod_{j \neq k} (1 - H(\Phi_j)) \\ &\quad - \delta(\Phi_k) \sum_{j \neq k} (e_j - e_B) H(\Phi_j) \prod_{l \neq j, l \neq k} (1 - H(\Phi_l)) \\ &\quad + \alpha \delta(\Phi_k) \operatorname{div} \frac{\nabla \Phi_k}{|\nabla \Phi_k|} \quad k = 1, \dots, M.\end{aligned}\quad (34)$$

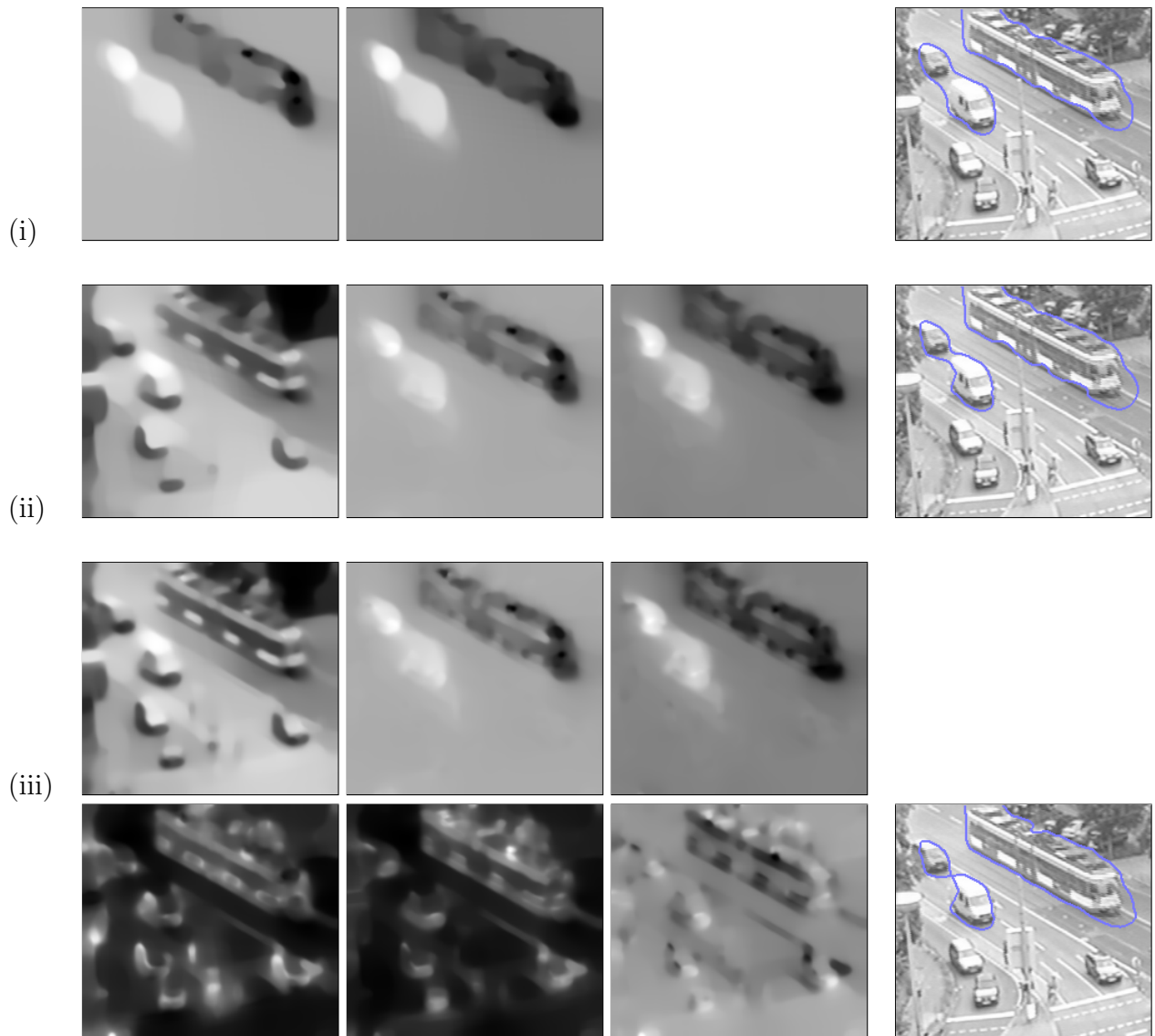


Figure 9: Motion segmentation integrating different cues. LEFT: Feature channels. RIGHT: Final Segmentation. (i) Optic flow. (ii) Intensity and optic flow. (iii) Intensity, optic flow, and texture.

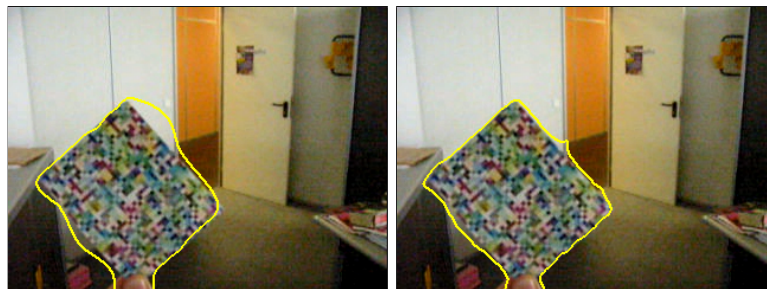


Figure 10: Object detection in a box sequence. LEFT: Using optic flow. RIGHT: Using colour, optic flow, and texture.



Figure 11: Moving object detection between frame 130 and 131 of the traffic sequence. **LEFT:** Using only optic flow. **RIGHT:** Using intensity, optic flow, and texture.

## 5.2 Experiments

The tracking algorithm was first tested with two grey level sequences: the famous Hamburg Taxi sequence<sup>2</sup> (Fig.12) as well as the traffic sequence already considered in the motion detection part (Fig.13).

In the first sequence, the objects are relatively large, yet its quality caused by the camera hardware available at that time is rather challenging. Since the grey value and texture information is unreliable for the van on the right, only the motion information can provide for its rough tracking. Note that also in Fig.13 the small cyclist can only be tracked because of its clear motion.

In Fig.14, a small cutout of the traffic sequence is considered. In this example, very small objects (one pedestrian and two bikes) are tracked along the sequence. In particular, the bike encircled in yellow is tracked despite the very low gradients.

In order to demonstrate the effect of the coupled evolution equations, we have tested our approach also in critical cases where objects overlap each other. First a synthetic example of three balls having crossing trajectories is shown in Fig.15. The coupling between the level set functions permits to deal with this situation. In the second example depicted in Fig.16, we work on a colour sequence of a soccer game where we track three players that are very close to each other. Moreover, the camera is not static here.

## 6 Conclusion

In this paper it has been shown how pertinent information from an image or a sequence of images can be extracted and combined. Ignoring prior knowledge, we concentrated on extracting and incorporating as much information from the given data as possible. It has further been shown that a suitable region model includes implicit weighting of features by their discriminative power. The proposed method allows for segmenting textured images or to detect moving objects even under difficult circumstances. The way we compute the features, the coupled nonlinear diffusion with a special degenerated diffusivity, the statistical region model, as well as the multi-scale approach are the basis for the good results. Moreover, by extending the method to tracking, it has been demonstrated that the approach is not only useful for segmentation but can also be adapted to similar problems. Even though we

<sup>2</sup>The sequence was created at the University of Hamburg and can be obtained from <ftp://csd.uwo.ca/pub/vision>

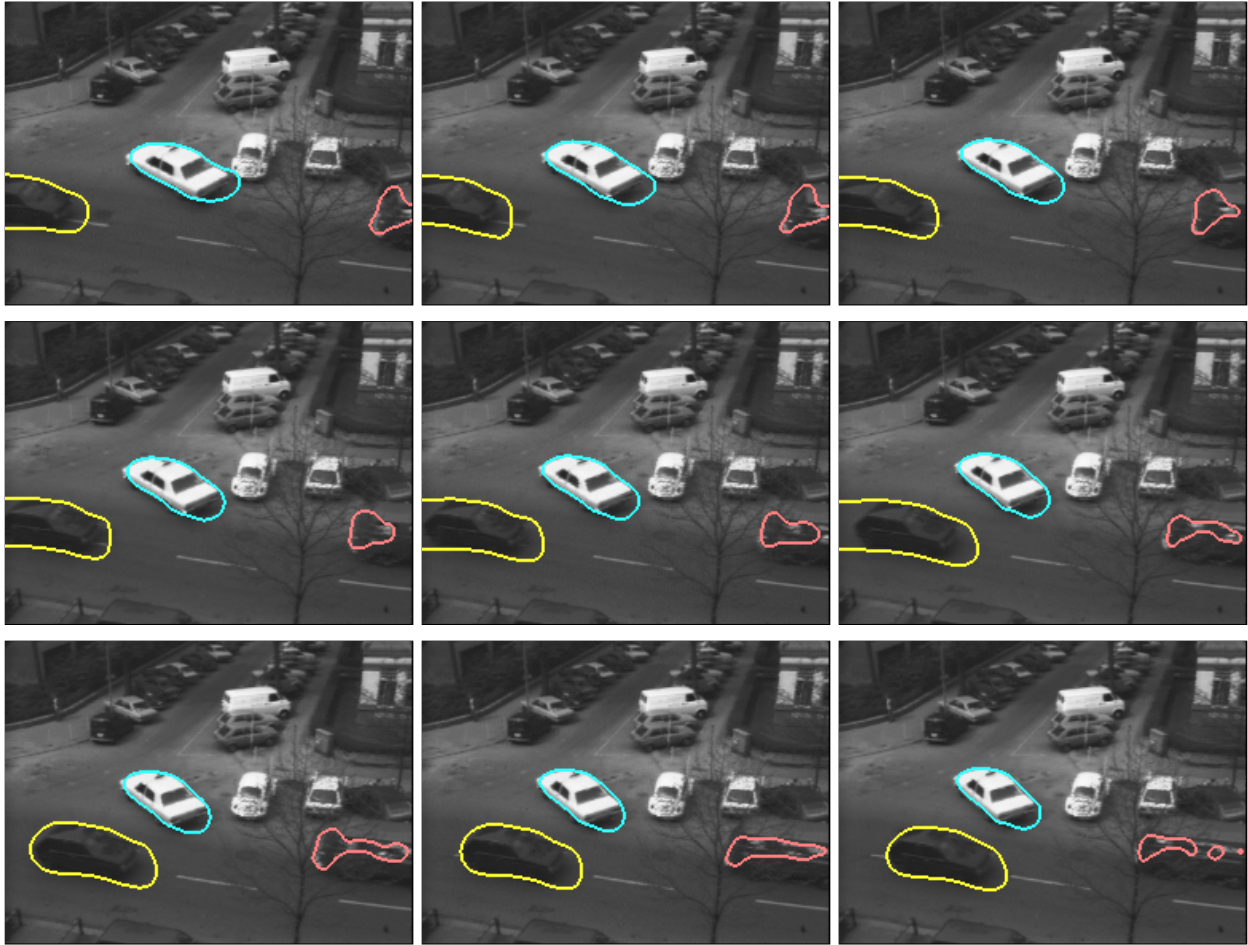


Figure 12: FROM LEFT TO RIGHT, TOP TO BOTTOM: Tracking result for the Hamburg Taxi sequence ( $256 \times 190 \times 20$ ).

have shown that we can achieve a lot with such a bottom-up approach, prior knowledge may certainly be mandatory in more difficult cases with, e.g., partial occlusions, shadows, or non-uniformly textured objects. The presented method is a good basis for extensions that include such knowledge. For instance, shape priors defined for level set representations in [65, 23] could be used without any modification.

## References

- [1] R. Acar and C. R. Vogel. Analysis of bounded variation penalty methods for ill-posed problems. *Inverse Problems*, 10:1217–1229, 1994.
- [2] D. Adalsteinsson and J. A. Sethian. A fast level set method for propagating interfaces. *Journal of Computational Physics*, 118(2):269–277, 1995.
- [3] F. Andreu, C. Ballester, V. Caselles, and J. M. Mazón. Minimizing total variation flow. *Differential and Integral Equations*, 14(3):321–360, March 2001.
- [4] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, February 1994.

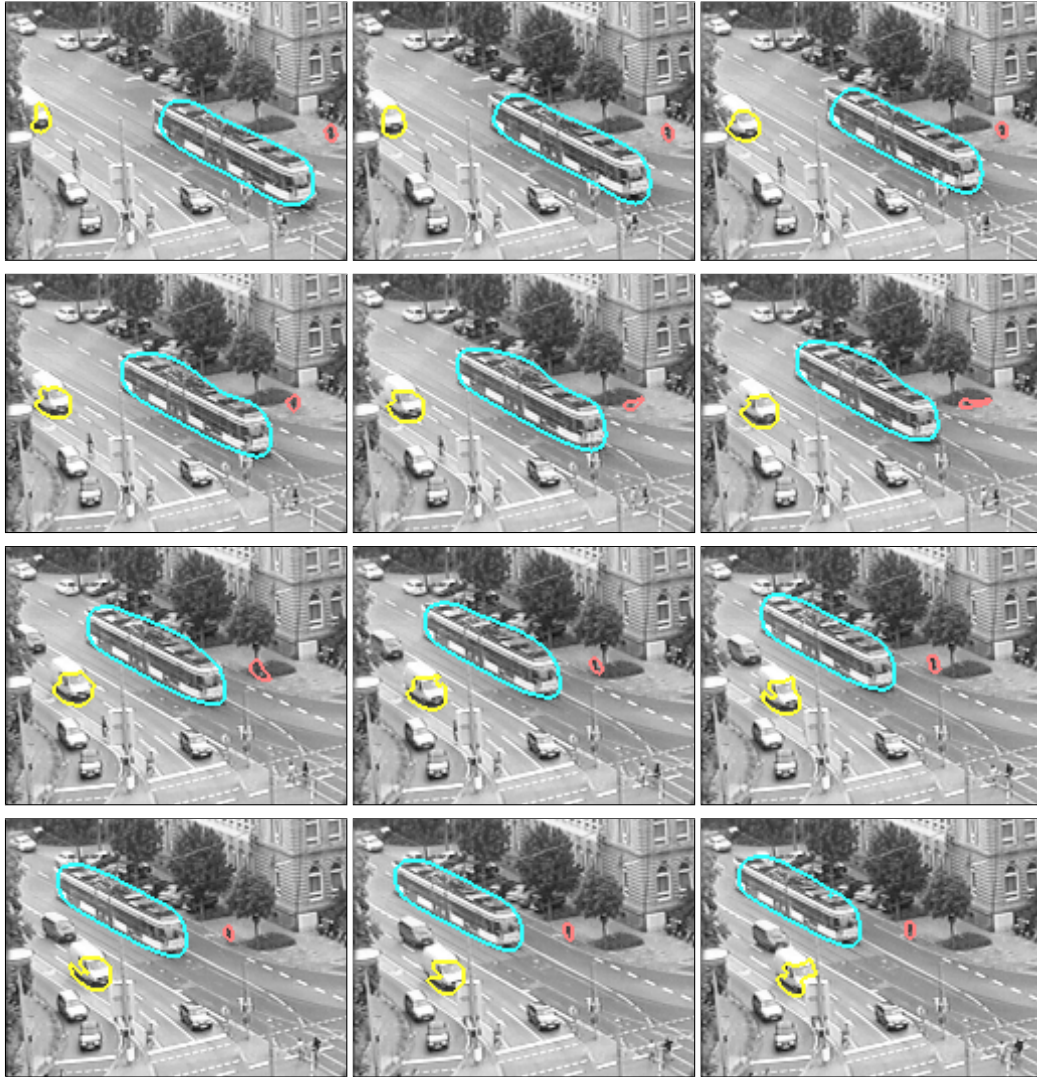


Figure 13: FROM LEFT TO RIGHT, TOP TO BOTTOM: Tracking result for a downsampled and cropped version of the traffic sequence ( $170 \times 130 \times 76$ ).

- [5] S. Belongie, C. Carson, H. Greenspan, and J. Malik. Color- and texture-based image segmentation using the expectation-maximization algorithm and its application to content-based image retrieval. In *Sixth International Conference on Computer Vision*, pages 675–682, Bombay, India, January 1998.
- [6] 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, August 1991.
- [7] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, Cambridge, MA, 1987.
- [8] P. Brodatz. *Textures: a Photographic Album for Artists and Designers*. Dover, New York, 1966.
- [9] T. Brox, M. Rousson, R. Deriche, and J. Weickert. Unsupervised segmentation incorporating colour, texture, and motion. In N. Petkov and M. A. Westenberg, editors, *Computer Analysis of Images and Patterns*, volume 2756 of *Lecture Notes in Computer Science*, pages 353–360. Springer, Berlin, 2003.





Figure 14: FROM LEFT TO RIGHT: Tracking of small objects in a cutout of the traffic sequence ( $110 \times 220 \times 90$ ).

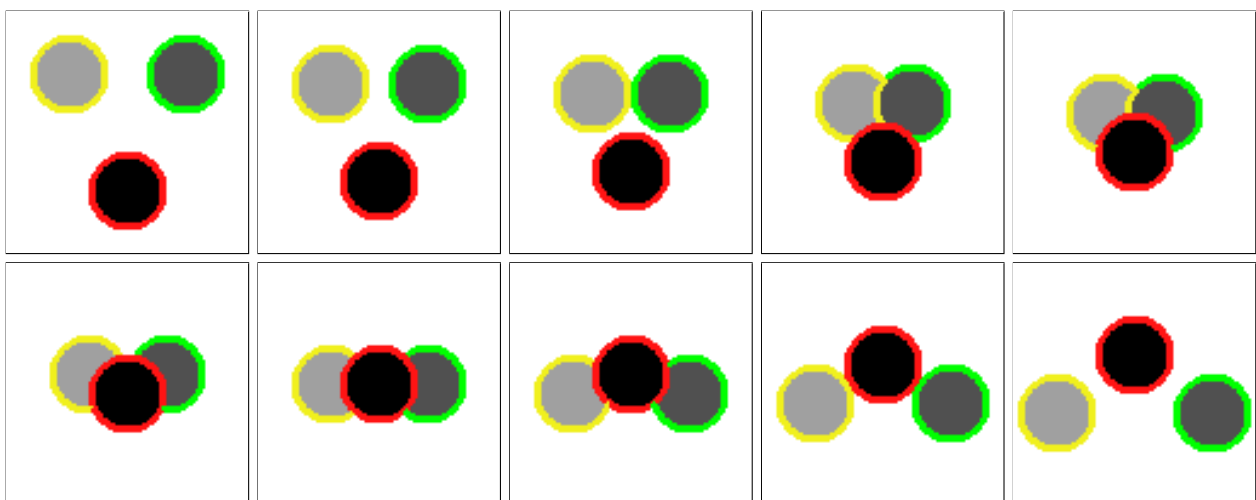


Figure 15: FROM LEFT TO RIGHT, TOP TO BOTTOM: Tracking of 3 overlapping balls ( $100 \times 100 \times 45$ ).

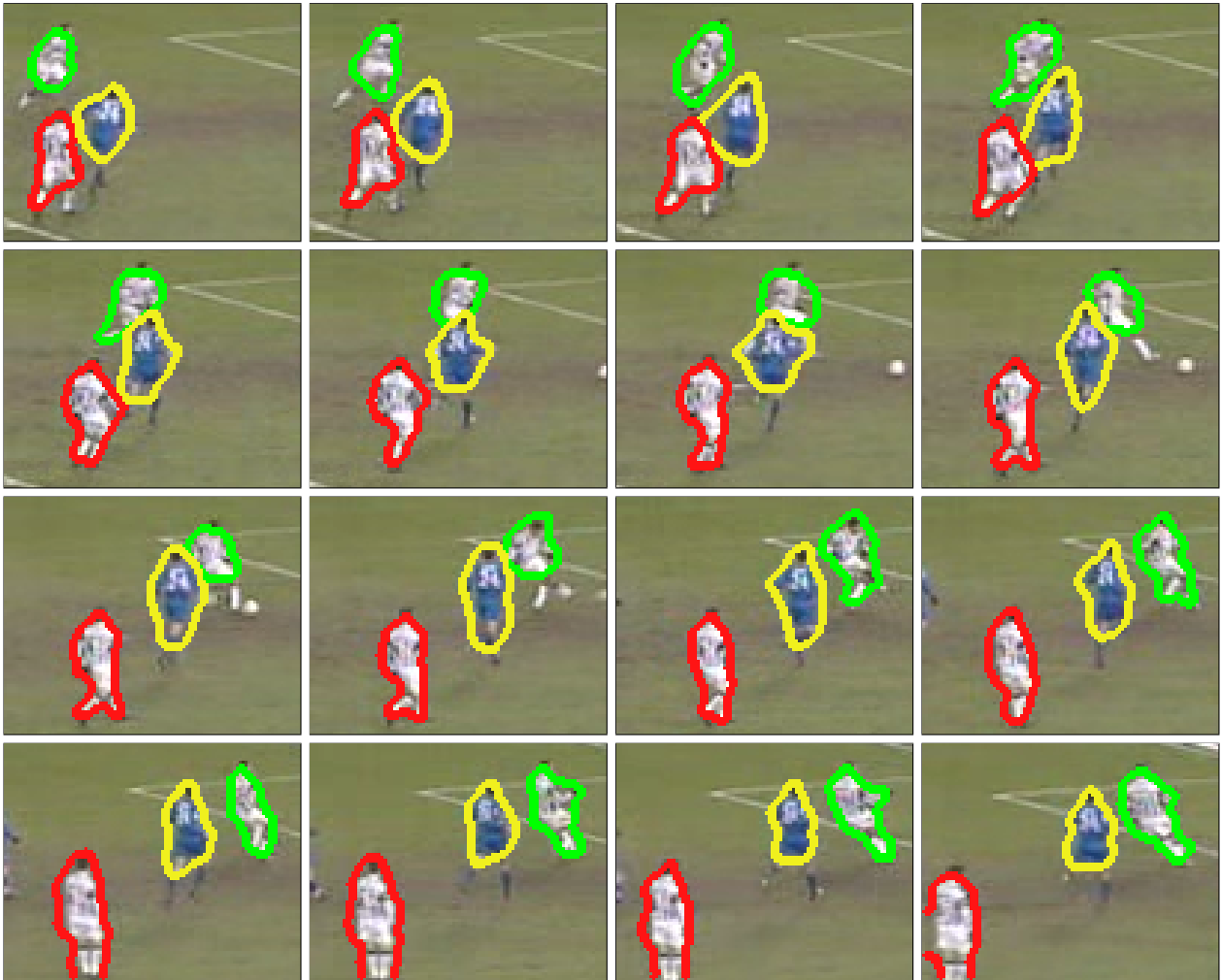


Figure 16: FROM LEFT TO RIGHT, TOP TO BOTTOM: Tracking of 3 players in a soccer sequence ( $180 \times 130 \times 40$ ).

- [10] T. Brox, R. van den Boomgaard, F. B. Lauze, J. van de Weijer, J. Weickert, P. Mrázek, and P. Kornprobst. Adaptive structure tensors and their applications. In J. Weickert and H. Hagen, editors, *Visualization and Processing of Tensor Fields*. Springer, Berlin, 2005. To appear.
- [11] 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, September 2002.
- [12] T. Brox and J. Weickert. A TV flow based local scale estimate and its application to texture discrimination. *Journal of Visual Communication and Image Representation*, 2005. To appear.
- [13] T. Brox, J. Weickert, B. Burgeth, and P. Mrázek. Nonlinear structure tensors. Technical Report 113, Dept. of Mathematics, Saarland University, Saarbrücken, Germany, October 2004.
- [14] T. Brox, M. Welk, G. Steidl, and J. Weickert. Equivalence results for TV diffusion and TV regularisation. In L. D. Griffin and M. Lillholm, editors, *Scale-Space Methods in Computer Vision*, volume 2695 of *Lecture Notes in Computer Science*, pages 86–100. Springer, Berlin, June 2003.
- [15] A. Bruhn, J. Weickert, and C. Schnörr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61(3):211–231, 2005.
- [16] V. Caselles, F. Catté, T. Coll, and F. Dibos. A geometric model for active contours in image processing. *Numerische Mathematik*, 66:1–31, 1993.
- [17] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In *Proc. Fifth International Conference on Computer Vision*, pages 694–699, Cambridge, MA, June 1995. IEEE Computer Society Press.
- [18] T. Chan, B.Y. Sandberg, and L. Vese. Active contours without edges for vector-valued images. *Journal of Visual Communication and Image Representation*, 11:130–141, 2000.
- [19] T. Chan and L. Vese. An active contour model without edges. In M. Nielsen, P. Johansen, O. F. Olsen, and J. Weickert, editors, *Scale-Space Theories in Computer Vision*, volume 1682 of *Lecture Notes in Computer Science*, pages 141–151. Springer, 1999.
- [20] T. Chan and L. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, February 2001.
- [21] D. Chopp. Computing minimal surfaces via level set curvature flow. *Journal of Computational Physics*, 106(1):77–91, May 1993.
- [22] J.J. Clark and A.L. Yuille. *Data fusion for sensory information processing*. Kluwer Academic, 1990.
- [23] D. Cremers, S. Osher, and S. Soatto. A multi-modal translation-invariant shape prior for level set segmentation. In C.-E. Rasmussen, H. Bülthoff, M. Giese, and B. Schölkopf, editors, *Pattern Recognition*, volume 3175 of *Lecture Notes in Computer Science*, pages 36–44. Springer, Berlin, August 2004.
- [24] G. R. Cross and A. K. Jain. Markov Random Field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5:25–39, 1983.
- [25] Y. Deng and B. S. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(8):800–810, 2001.

- [26] A. Dervieux and F. Thomasset. A finite element method for the simulation of Rayleigh–Taylor instability. In R. Rautman, editor, *Approximation Methods for Navier–Stokes Problems*, volume 771 of *Lecture Notes in Mathematics*, pages 145–158. Springer, Berlin, 1979.
- [27] S. Di Zenzo. A note on the gradient of a multi-image. *Computer Vision, Graphics and Image Processing*, 33:116–125, 1986.
- [28] F. Dibos and G. Koepfler. Global total variation minimization. *SIAM Journal on Numerical Analysis*, 37(2):646–664, 2000.
- [29] X. Feng and A. Prohl. Analysis of total variation flow and its finite element approximations. *Mathematical Modelling and Numerical Analysis*, 37(3):533–556, 2003.
- [30] 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.
- [31] D. Gabor. Theory of communication. *The Journal of the Institution of Electrical Engineers*, 93:429–457, 1946.
- [32] 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.
- [33] E. Granger, M. Rubin, S. Grossberg, and P. Lavoie. A what-and-where fusion neural network for recognition and tracking of multiple radar emitters. *Neural Networks*, 14:325–344, 2001.
- [34] G. H. Granlund and H. Knutsson. *Signal Processing for Computer Vision*. Kluwer, Dordrecht, 1995.
- [35] E. Hayman and J.-O. Eklundh. Probabilistic and voting approaches to cue integration for figure-ground segmentation. In *7th European Conference on Computer Vision*, volume 3, pages 469–486, 2002.
- [36] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.
- [37] S. L. Keeling and R. Stollberger. Nonlinear anisotropic diffusion filters for wide range edge sharpening. *Inverse Problems*, 18:175–190, January 2002.
- [38] S. Khan and M. Shah. Object based segmentation of video using color, motion and spatial information. In *Computer Vision and Pattern Recognition, CVPR 2001*, pages 11–13, December 2001.
- [39] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Gradient flows and geometric active contour models. In *Proc. Fifth International Conference on Computer Vision*, pages 810–815, Cambridge, MA, June 1995. IEEE Computer Society Press.
- [40] 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*, volume 3, pages 797–800, Rochester, NY, June 2002.
- [41] G. Kühne, J. Weickert, O. Schuster, and S. Richter. A tensor-driven active contour model for moving object segmentation. In *Proc. 2001 IEEE International Conference on Image Processing*, volume 2, pages 73–76, Thessaloniki, Greece, October 2001.
- [42] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer, Boston, 1994.

- [43] 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, August 1981.
- [44] R. Malladi, J. A. Sethian, and B. C. Vemuri. Shape modeling with front propagation: a level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):158–175, February 1995.
- [45] S. Marcelja. Mathematical description of the response of simple cortical cells. *Journal of Optical Society of America*, 70:1297–1300, 1980.
- [46] M. Middendorf and H.-H. Nagel. Empirically convergent adaptive estimation of grayvalue structure tensors. In L. van Gool, editor, *Proc. 24th DAGM Symposium*, volume 2449 of *Lecture Notes in Computer Science*, pages 66–74, Zürich, Switzerland, September 2002. Springer.
- [47] 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.
- [48] D. Mumford and J. Shah. Boundary detection by minimizing functionals, I. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 22–26, San Francisco, CA, June 1985. IEEE Computer Society Press.
- [49] R. Murphy. Biological and cognitive foundations of intelligent sensor fusion. *IEEE Transactions on Systems, Man, and Cybernetics*, 26(1), 1996.
- [50] H.-H. Nagel and A. Gehrke. Spatiotemporally adaptive estimation and segmentation of OF-fields. In H. Burkhardt and B. Neumann, editors, *Computer Vision – ECCV ’98*, volume 1407 of *Lecture Notes in Computer Science*, pages 86–102. Springer, Berlin, 1998.
- [51] S. Osher and R. P. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*, volume 153 of *Applied Mathematical Sciences*. Springer, New York, 2002.
- [52] S. Osher and N. Paragios, editors. *Geometric Level Set Methods in Imaging, Vision and Graphics*. Springer, New York, 2003.
- [53] 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.
- [54] E. Ozyildiz, N. Krahnstoever, and R. Sharma. Adaptive texture and color segmentation for tracking moving objects. *Pattern Recognition*, 35(10), 2002.
- [55] N. Paragios and R. Deriche. Unifying boundary and region-based information for geodesic active tracking. In *Proc. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 300–305, Forth Collins, Colorado, June 1999.
- [56] 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, March 2000.
- [57] N. Paragios and R. Deriche. 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, 2002.
- [58] N. Paragios and R. Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *International Journal of Computer Vision*, 46(3):223–247, March 2002.

- [59] E. Parzen. On the estimation of a probability density function and the mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.
- [60] 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.
- [61] 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.
- [62] A. R. Rao and B. G. Schunck. Computing oriented texture fields. *CVGIP: Graphical Models and Image Processing*, 53:157–185, 1991.
- [63] 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*, pages 699–704, Madison, WI, June 2003.
- [64] 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*, pages 56–62, Orlando, Florida, December 2002.
- [65] 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.
- [66] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
- [67] 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.
- [68] 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.
- [69] J. A. Sethian. *Level Set Methods and Fast Marching Methods*. Cambridge University Press, Cambridge, UK, second edition, 1999.
- [70] M. Spengler and B. Schiele. Towards robust multi-cue integration for visual tracking. *Lecture Notes in Computer Science*, 2095:93–106, 2001.
- [71] C. Stiller and J. Konrad. Estimating motion in image sequences. *IEEE Signal Processing Magazine*, 16:70–91, 1999.
- [72] D. Tschumperlé and R. Deriche. Diffusion tensor regularization with constraints preservation. In *Proc. 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 948–953, Kauai, HI, December 2001. IEEE Computer Society Press.
- [73] 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.
- [74] D. Tschumperlé and R. Deriche. Vector-valued image regularization with PDE’s: a common framework for different applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):506–517, April 2005.

- [75] 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.
- [76] J. Weickert. Design of nonlinear diffusion filters. In B. Jähne and H. Haußecker, editors, *Computer Vision and Applications*, pages 439–458. Academic Press, San Diego, 2000.
- [77] 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.
- [78] 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.
- [79] 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, March 1998.
- [80] H. K. Zhao, T. Chan, B. Merriman, and S. Osher. A variational level set approach to multiphase motion. *Journal of Computational Physics*, 127:179–195, 1996.
- [81] S.-C. 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, September 1996.