N. Petkov, M. A. Westenberg (Eds.): Computer Analysis of Images and Patterns, LNCS 2756, pp. 353-360, August 2003.

Unsupervised Segmentation Incorporating Colour, Texture, and Motion*

Thomas Brox[1], Mikaël Rousson[2], Rachid Deriche[2], and Joachim Weickert[1]

[1] 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

[2] 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

Abstract. In this paper we integrate colour, texture, and motion into a segmentation process. The segmentation consists of two steps, which both combine the given information: a pre-segmentation step based on nonlinear diffusion for improving the quality of the features, and 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. For the nonlinear diffusion we apply a novel diffusivity closely related to the total variation diffusivity, but being strictly edge enhancing. A multi-scale implementation is used in order to obtain more robust results. In several experiments we demonstrate the usefulness of integrating many kinds of information. Good results are obtained for both object segmentation and tracking of multiple objects.

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 very popular, because in such a framework it is possible to integrate many different cues and models. One can integrate, for instance, boundary information, shape priors as well as region information. Level set theory [12] provides an efficient possibility to find a minimizer of such an energy.

In our paper an unsupervised approach will be proposed that does not depend on previously acquired information. The objective of such an unsupervised approach is 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

* Our research is partly funded by the projects IMAVIS HPMT-CT-2000-00040 within the framework of the *Marie Curie Fellowship Training Sites Programme* as well as the European project *Cogvisys* numbered 3E010361, and the projects WE 2602/1-1 and SO 363/9-1 of the *Deutsche Forschungsgemeinschaft (DFG)*. This is gratefully acknowledged.

as possible. This importance to combine different cues in a segmentation algorithm has also been stressed in the work of Malik et al. [10]. Consequently, 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. The proposed framework based on the work in [19] allows to integrate all this information.

However, the possibility to integrate different kinds of information is only one step. Another question is how to acquire the information from the image. There is no such problem when using only primary features like grey value or colour, but as soon as secondary features like texture or motion are included, it is not obvious how to extract them the best way. Recently a nonlinear version of the linear structure tensor from [6] has been proposed [22]. Its suitability for texture discrimination is demonstrated in [18]. Considering motion, the optic flow is the principal method to integrate this information. Due to the fact that structure tensor techniques are also useful for optic flow analysis [2, 11], the nonlinear structure tensor can be applied here as well [4].

Since the features are often perturbed by noise or details that are useless for segmentation, a pre-processing step is very useful to obtain better results. Such a pre-processing should meet the following requirements: It must remove the perturbations while not loosing any important information. Moreover, like the segmentation, it should combine all the given information. Finally, it should yield results that are already close to a segmentation. Nonlinear diffusion is a well-suited technique to meet these requirements [15]. In this context, we propose to use a new diffusivity that can especially meet the last item.

As soon as motion information is used, it becomes obvious to perform not only segmentation but also tracking. For our segmentation technique it is rather easy to track an object once it has been detected. It becomes even possible to drop the assumption of having only one object, and to perform simultaneous tracking of multiple objects.

The remainder of this paper is organized as follows. In the next section the acquisition of the texture and motion features is briefly specified. Section 3 then describes how the information is employed in the two parts of our technique. In Section 4 the method is extended to tracking of multiple objects. In the succeeding section we show results of our experiments. The paper is concluded by a summary as well as an outlook on future work. A more detailed description and more experiments can be found in a research report [3] available from the internet.

2 Information Extraction

Information extraction is only interesting for texture and motion, since the grey level or colour information is already given by the image itself. For the acquisition of good texture features the nonlinear structure tensor has been shown to be very powerful [18]. It will also be used here.

For optic flow estimation, we use a modification of the method from [4]. This technique has two advantages: first, it induces only one smoothness parameter, and second it also applies the nonlinear structure tensor, so it is in best accordance with the texture feature acquisition. Thus, instead of the nonlinear structure tensor from [4], the slightly modified scheme described in [18] will be applied.

Provided all information is used, a feature vector with 8 components is considered, 3 colour channels (R, G, B), the optic flow components u and v computed with the above-mentioned method, and 3 texture channels ($\sum_i (I_i)_x^2, \sum_i (I_i)_y^2, \sum_i (I_i)_x (I_i)_y$), where i denotes the colour channel and the other subscripts denote partial derivatives.

3 Integration of Cues

3.1 Integrating Cues for Joint Smoothing

For the joint smoothing of the extracted features we apply nonlinear vector-valued diffusion. Nonlinear diffusion was introduced by Perona and Malik [15]. It was extended to vector-valued data in [7] using

$$\partial_t u_i = \operatorname{div}\Big(g\Big(\sum_{k=1}^{N} |\nabla u_k|^2\Big)\nabla u_i\Big) \qquad i = 1, ..., N \tag{1}$$

where u is the evolving feature vector initialized by the previously extracted data, and N is the number of feature channels. The decreasing *diffusivity function* g steers the reduction of smoothing in the presence of discontinuities. Note that g is the same for all channels, so there is a joint smoothing taking the edge information of all channels into account.

The choice of the diffusivity function is a critical point and mainly defines the behaviour of the diffusion process. Since there are first derivatives in the feature vector, the frequently used diffusivities with additional contrast parameters cause problems: Often it is impossible to choose a good global contrast parameter, since the derivatives may have responses of very different magnitude. A diffusivity without a contrast parameter is used in the total variational (TV) flow [1], the diffusion filter corresponding to TV regularization [20]. It leads to piecewise constant results removing oscillations and closing structures. However, TV flow is only one special representative of an entire family of diffusivities having these properties:

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

where ϵ is a small positive constant avoiding the diffusivity to become unbounded. These diffusivities include TV flow for $p = 1$ and so-called *balanced forward backward diffusion* [8] for $p = 2$. While TV flow is exactly the limit between forward and backward diffusion, the diffusivities are strictly edge enhancing for $p > 1$. In the continuous case, well-posedness questions for forward–backward diffusion are still unsolved, but discretization has been shown to resolve this problem [21]. The results for $p > 1$ appear not only to be piecewise constant, they also have steep edges due to the edge enhancement. This is very useful for our application. The exact choice of p is uncritical. 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. In our experiments $p = 1.6$ turned out to be a good compromise.

3.2 Integrating Cues for Partitioning

Two-Region Partitioning. 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 minimization problem following the idea of *geodesic active regions* [14, 19]:

$$E(\Omega_i, p_{ij}) = -\sum_{j=1}^{N} \left(\int_{\Omega_1} \log p_{1j}(u_j(x))dx + \int_{\Omega_2} \log p_{2j}(u_j(x))dx \right) \qquad i = 1, 2. \tag{3}$$

For minimizing 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 regularized 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 $\chi_2(s) = 1 - H(s)$. Moreover, we add a regularization term on the length of the interface $\partial\Omega$ between the two regions Ω_1 and Ω_2. Such a regularization can be expressed using the level set representation; see [23] for details. This allows to formulate a continuous form of the above-mentioned energy functional:

$$E(\Phi, p_{ij}) = -\sum_{i=1}^{2}\sum_{j=1}^{N} \left(\int_{\Omega} \log p_{ij}(u_j)\chi_i(\Phi)dx \right) + \alpha \int_{\Omega} |\nabla H(\Phi)|dx \tag{4}$$

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

$$\partial_t \Phi = H'(\Phi)\Big(\sum_{j=1}^{N} \log \frac{p_{1j}(u_j)}{p_{2j}(u_j)} + \alpha \operatorname{div} \frac{\nabla\Phi}{|\nabla\Phi|} \Big) \tag{5}$$

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

PDF Approximations. The variational framework 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: the mean μ_{ij} and the standard deviation σ_{ij}. Although reasonable, choosing a Gaussian function as PDF is not the only possible solution. Kim et al. [9] 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.

Given the probability densities, the energy can be minimized with respect to Φ using the gradient descent in Eq. 5. Thus the segmentation process works according to the *expectation-maximation* principle [5] with some initial partitioning (Ω_1, Ω_2). The non-parametric 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, it results in more local minima in the objective function and makes it more dependent on the initialization. This problem can be addressed by applying the basic idea of deterministic annealing [16, 17] using a Gaussian function in the first run to get close to the global minimum of the objective function. Then a second run of the minimization process will finally result in this global minimum or a local minimum that is very close to this global minimum. Although there exist counter-examples where this approach will fail, the heuristic works very well in most cases.
In order to further increase the robustness of our approach, we used a multi-scale implementation: the data from a finer scale is downsampled and serves as input for a segmentation at a coarser scale. This segmentation is then used to initialize the segmentation of the finer scale. This eases the problem of local minima. Two levels were used for our experiments.
In the variational formulation, we did not mention which information each channel contained. This general framework permits to combine any kind of information as we will see in the experiments.

4 Extension to 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 optic flow. In [13] and [19] it has already been demonstrated that it is possible to apply segmentation to tracking. In this section that approach will be combined with the classic idea of tracking using optic flow. Both vector components of the optic flow are used as features.
To allow the tracking of multiple objects, the variational formulation must be slightly modified. First, we suppose the positions of each of the moving objects to be known in the first frame. To each of these moving objects a level set function Φ_k is assigned, with $k = 1, ..., M$ and M being the total number of detected objects. We denote by B the static part of the image (which corresponds to the background of the scene) and by p_B the corresponding probability density function. This region is defined as the region where all the level set functions are negative. The global energy is defined as follows:

$$E = \int_\Omega \sum_{k=1}^{M} \Bigg(\underbrace{-\sum_{j=1}^{N} \log p_{kj}\, H(\Phi_k) + \alpha |\nabla H(\Phi_k)|}_{e_k} \Bigg) \underbrace{- \sum_{j=1}^{N} \log p_{Bj}\, \chi_B}_{e_B} \qquad (6)$$

Note that the χ function for the background $\chi_B = \prod_{l=1}^{M}(1 - H(\Phi_l))$ also affects the estimation of the PDFs. The minimization of the new energy leads to the revised gradient descent

$$\partial_t \Phi_k = -H'(\Phi_k)\Big(e_k - e_B \prod_{l \neq k} (1 - H(\Phi_l)) - \alpha \operatorname{div} \frac{\nabla \Phi_k}{|\nabla \Phi_k|} \Big). \qquad (7)$$

5 Results

The performance of our approach was tested with a number of real-world images. For more examples we refer to [3].
First we used static images without motion information to combine texture and colour. Two results are shown in Fig. 1. In Fig. 1a the level set initialization used for all our experiments is depicted. Fig. 2 demonstrates the importance to use all available information for some images. The correct result can only be obtained by using both texture and colour information.
Colour information and optic flow magnitude were integrated on a sequence where a hand is moving in front of a complicated background (Fig. 3). Despite camera noise we obtain a good detection of the hand. Only a small region corresponding to the shadow is merged with the moving object in some frames.
To illustrate the capacities of the tracking method in Section 4, we applied it on the tracking of three players in a soccer sequence with moving camera (Fig. 4). Note that the players are relatively small and close to each other. The tracking initialization is done by clicking on the players we want to track. The results look very promising.

6 Conclusions

We have presented an unsupervised segmentation framework that can incorporate many different kinds of information. It has been possible to integrate colour, texture, and motion. The way to compute the features, the coupled nonlinear diffusion with a novel diffusivity, as well as the statistical region model and a multiscale implementation are responsible for the good results. Our approach uses jointly different cues in both parts of the method. Like humans do when analysing a scene, we tried to extract many kinds of information and integrated them in a general framework.
In several experiments it has been shown that our method works very well with all images that are in accordance with our model assumptions. In natural images such assumptions can sometimes be violated. In order to be able to deal also with such images, the assumption of having only two regions has to be dropped. A good solution for this problem will be a very challenging topic for future research. We also think that it could be advantageous to combine our unsupervised technique with learning techniques known from supervised approaches.

References

1. F. Andreu, C. Ballester, V. Caselles, and J. M. Mazón. Minimizing total variation flow. *Differential and Integral Equations*, 14(3):321–360, Mar. 2001.
2. 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.
3. T. Brox, M. Rousson, R. Deriche, and J. Weickert. Unsupervised segmentation incorporating colour, texture, and motion. Research Report 4760, INRIA Sophia-Antipolis, France, 2003.

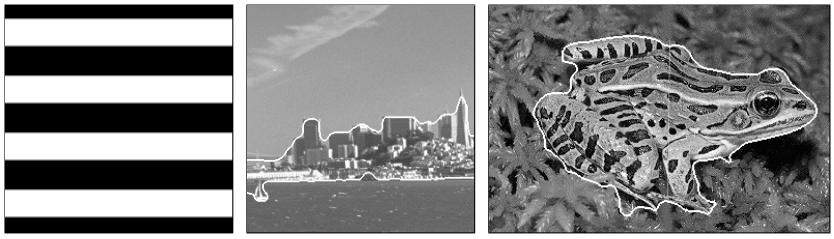

Fig. 1. LEFT: (a) Level set initialization. CENTER: (b) Using colour and texture. RIGHT: (c) Using colour and texture.

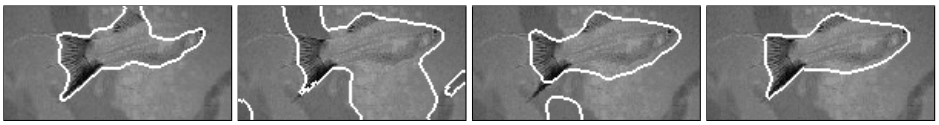

Fig. 2. Segmentation results using different kinds of information. FROM LEFT TO RIGHT: (a) Grey value and texture. (b) Colour (RGB). (c) Colour (CIELAB). (d) Colour and texture.

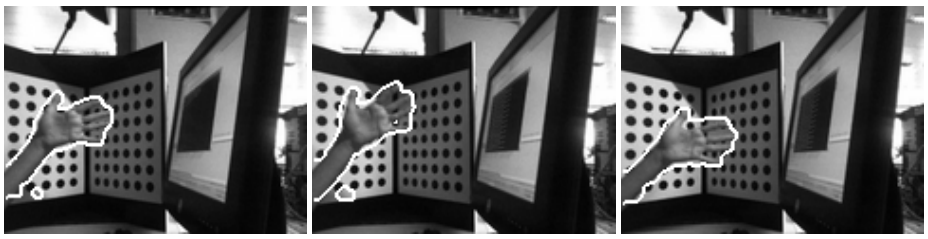

Fig. 3. Tracking result for 3 out of 30 images of the hand sequence.

Fig. 4. Tracking result for 4 out of 27 images of the soccer sequence.

4. 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.
5. A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B*, 39:1–38, 1977.
6. 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.
7. 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.
8. S. Keeling and R. Stollberger. Nonlinear anisotropic diffusion filters for wide range edge sharpening. *Inverse Problems*, 18:175–190, Jan. 2002.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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), Mar. 2000.
14. 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*, pages 249–268, March/June 2002.
15. 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.
16. 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.
17. 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.
18. M. Rousson, T. Brox, and R. Deriche. Active unsupervised texture segmentation on a diffusion based feature space. In *Proc. 2003 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, Madison, WI, June 2003. IEEE Computer Society Press. To appear.
19. 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.
20. L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
21. 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.
22. 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*. AMS, Providence, 2002.
23. 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.