Multi-Dimensional Histogram-Based Image Segmentation

Daniel Weiler, Julian Eggert

2007

Preprint:

This is an accepted article published in Proceedings of the 14th International Conference on Neural Information Processing (ICONIP). The final authenticated version is available online at: https://doi.org/[DOI not available]

Multi-Dimensional Histogram-Based Image Segmentation

Daniel Weiler[1], Julian Eggert[2]

[1] Darmstadt University of Technology, Darmstadt D-64283, Germany
[2] Honda Research Institute Europe GmbH, Offenbach D-63073, Germany

Abstract. In this paper we present an approach for multi-dimensional histogram-based image segmentation. We combine level-set methods for image segmentation with probabilistic region descriptors based on multi-dimensional histograms. Unlike stated by other authors we show that colour space histograms provide a reasonable and efficient description of image regions. In contrast to Gaussian Mixture Model based algorithms no parameter learning and estimation of the number of mixture components is required. Compared to recent level-set based segmentation methods satisfying segmentation results are achieved without specific features (e.g. texture). In a comparison with state-of-the-art image segmentation methods it is shown that the proposed approach yields competitive results.

1 Introduction

In the field of image segmentation, two major approaches can be distinguished: *multi region segmentation* and *figure-background segregation*. While the former tries to group similar (by their image features $\mathbf{f}$) and related (by their spatial properties like location, etc.) pixels of an image into separate regions, the latter attempts to find a salient region of an image considering it as a foreground "figure", labelling all the reminder without any further differentiation as background. In this paper we address the problem of figure-background segregation based on multi-dimensional histogram-based region descriptors.

In state-of-the-art figure-background segregation algorithms (see "GrabCut" [1], "Graph cut" [2], "Knockout 2" [3] and "Bayes Matte" [4]) probabilistic colour distribution models are commonly used. In recent years also level-set methods [5–9] became a powerful tool for image segmentation. The former algorithms model colour distributions in a three dimensional colour space, whereas state-of-the-art level-set methods are able to work on arbitrary feature maps [10]. These feature maps may incorporate the three colour components but might be extended by any other characteristic property of a region (e.g. texture and motion [11]). So far level-set methods assume the feature maps to be independent, which constitutes a major difference to the algorithm proposed here.

The method presented in this paper combines the multi-dimensional approach of colour distributions of state-of-the-art figure-background segregation algorithms with the feature maps used by level-set methods. The combined algorithm is formulated in a two-region level-set framework. Whereas state-of-the-art image segmentation methods commonly model the colour distribution by means of Gaussian Mixture Models, we use colour space histograms that do not require parameter learning and the estimation of the number of mixture components and thus are more efficient to implement. In contrast to state-of-the-art level-set methods it is shown that competitive segmentation results are achieved without any additional specific feature maps, like texture.

Level-set methods [5] separate all image pixels into two disjoint regions by favouring homogeneous image properties for pixels within the same region and distinct image properties for pixels belonging to different regions. The level-set formalism describes the region properties using an energy functional that implicitly contains the region description and that has to be minimised. The formulation of the energy functional dates back to e.g. Mumford and Shah [6] and to Zhu and Yuille [7]. Later on, the functionals were reformulated and minimised using the level-set framework by e.g. [8] and [9].

Among all segmentation algorithms from computer vision (see Sect. 2), level-set methods provide perhaps the closest link with the biologically motivated, connectionist models as e.g. represented by [12]. Similar to neural models, level-set methods work on a grid of nodes located in image/retinotopic space, interpreting the grid as having local connectivity, and using local rules for the propagation of activity in the grid. Time is included explicitly into the model by a formulation of the dynamics of the nodes activity. Furthermore, the external influence from other sources (larger network effects, feedback from other areas, inclusion of prior knowledge) can be readily integrated on a node-per-node basis, which makes level-sets appealing for the integration into biologically motivated system frameworks.

In this paper, we apply an extended level-set formalism to compare the representation of region characteristics by several independent features and by features located in a common feature space and show the advantages of the latter. In Sect. 2 state-of-the-art figure-background segregation algorithms are briefly described. Section 3 introduces the level-set method we use for image segmentation and its extension to multi-dimensional histogram-based region descriptors. In Sect. 4 we present the results of the proposed algorithm. A short discussion finalises the paper.

2 State-of-the-Art Figure-Background Segregation

In [1] a comprehensive summary of recent figure-background segregation methods is given. The reminder of this section compares two major approaches: "trimap"-based algorithms, introduced in Sect. 2.1 and level-set methods, described in Sect. 2.2. Inspired by these two methods, we introduce an extension to standard level-set methods for image segmentation in Sect. 3.

2.1 "Trimap"-Based Methods

A number of state-of-the-art figure-background segregation algorithms (e.g.: "GrabCut" [1], "Graph cut" [2], "Knockout 2" [3] and "Bayes Matte" [4]) perform the image segmentation task based on "trimaps". Starting with an initial "trimap" $T = \{T_B, T_U, T_F\}$ – that specifies known background T_B, known foreground T_F and unknown T_U regions of the image – the pixels of the unknown region are assigned to the foreground and background regions. The assignment is commonly based on probabilistic colour distribution models. Depending on the algorithm, the assignment is in a binary or probabilistic manner and the probabilistic colour distribution models are computed based only on the initial "trimap" or iteratively updated using the previous assignments within the region T_U. To represent the probabilistic colour distribution models, different approaches are proposed. For grey values histograms are often used, whereas a common choice for the RGB colour space are Gaussian Mixture Models. According to [1] it is impractical to construct adequate colour space histograms, which will be disproved in this paper.

In addition to the "trimap", a smoothness term is used to control the granularity of the segmentation. The smoothness term acts in a way that encourages coherence of the assignments of neighbouring, unknown pixels within the region T_U. Therefore adjacent pixels are forced to similar assignments depending on the difference of their corresponding colour and grey values, respectively. The more similar the pixel values are, the higher is the force to assign them to the same region T_F and T_B, respectively.

2.2 Level-Set Methods

Level-set methods are front propagation methods. Starting with an initial contour, the figure-background segregation task is solved by iteratively moving the contour according to the solution of a partial differential equation (PDE). The PDE is often originated from the minimisation of an energy functional. Famous representatives of energy functionals for image segmentation problems are those by Mumford and Shah [6] and by Zhu and Yuille [7]. While the former work in its original version on grey value images (i.e. on scalar data), utilise the mean grey value of a region as a simple region descriptor and were only later extended to vector valued data [10] (e.g. colour images), the latter use more advanced probabilistic region descriptors that are based on the distributions of each feature channel inside and outside the contour. In many cases it is sufficient to model these distributions by unimodal Gaussian distributions. In some rare cases the distributions are approximated in a multimodal way [9] e.g. by Gaussian Mixture Models or Nonparametric Parzen Density Estimates [13]. Regardless of the way the distributions are modeled, the features are in all approaches assumed to be independent. Thus, they are not located in a common feature space which leads to a separate model for each feature. Within a region the models of all features together add up to the region descriptor.

Similar to the "trimap"-based approaches, level-set methods use a smoothness term to control the granularity of the segmentation. A common way is to penalise the length of the contour, that can be formulated in the energy functional by simply adding the length of the contour to the energy that is to be minimised. In doing so, few large objects are favoured over many small objects as well as smooth object boundaries over ragged object boundaries.

Compared to "active contours" (snakes) [14], that also constitute front propagation methods and explicitly represent a contour by supporting points, level-set methods represent contours implicitly by a level-set function that is defined over the complete image plane. The contour is defined as an iso-level in the level-set function, i.e. the contour is the set of all locations, where the level-set function has a specific value. This value is commonly chosen to be zero, thus the inside and outside regions can easily be determined by the Heaviside function $H(x)$[1].

3 Multi-Dimensional Histogram-Based Image Segmentation

3.1 Standard Level-Set based Region Segmentation

The proposed multi-dimensional histogram-based image segmentation framework is based on a standard two-region level-set method [9, 15]. In a level-set framework, a level-set function $\phi \in \Omega \mapsto \mathbb{R}$ is used to divide the image plane Ω into two disjoint regions, Ω_1 and Ω_2, where $\phi(x) > 0$ if $x \in \Omega_1$ and $\phi(x) < 0$ if $x \in \Omega_2$. Here we adopt the convention that Ω_1 indicates the background and Ω_2 the segmented object. A functional of the level-set function ϕ can be formulated that incorporates the following constraints:

– Segmentation constraint: the data within each region Ω_i should be as similar as possible to the corresponding region descriptor ρ_i.
– Smoothness constraint: the length of the contour separating the regions Ω_i should be as short as possible.

This leads to the expression[2]

$$E(\phi) = \nu \int_{\Omega} |\nabla H(\phi)| dx - \sum_{i=1}^{2} \int_{\Omega} \chi_i(\phi) \log p_i \, dx \qquad (1)$$

with the Heaviside function $H(\phi)$ and $\chi_1 = H(\phi)$ and $\chi_2 = 1 - H(\phi)$. That is, the χ_i's act as region masks, since $\chi_i = 1$ for $x \in \Omega_i$ and 0 otherwise. The first term acts as a smoothness term, that favours few large regions as well as smooth regions boundaries, whereas the second term contains assignment probabilities $p_1(x)$ and $p_2(x)$ that a pixel at position x belongs to the inner and outer regions Ω_1 and Ω_2, respectively, favouring a unique region assignment.

[1] $H(x) = 1$ for $X > 0$ and $H(x) = 0$ for $X \leq 0$.
[2] Remark that ϕ, χ_i and p_i are functions over the image position x.

Minimisation of this functional with respect to the level-set function ϕ using gradient descent leads to

$$\frac{\partial \phi}{\partial t} = \delta(\phi)\left[\nu \operatorname{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right) + \log\frac{p_1}{p_2}\right] \quad . \tag{2}$$

A region descriptor $\rho_i(\mathbf{f})$ that depends on the image feature vector $\mathbf{f}$ serves to describe the characteristic properties of the outer vs. the inner regions. The assignment probabilities $p_i(x)$ for each image position are calculated based on an image feature vector via $p_i(x) := \rho_i(\mathbf{f}(x))$. The parameters of the region descriptor $\rho_i(\mathbf{f})$ are gained in a separate step using the measured feature vectors $\mathbf{f}(x)$ at all positions $x \in \Omega_i$ of a region i.

For standard images, there may be only a single feature vector component like the pixel grey values. The case with several image features is – in standard level-set based region segmentation – covered by assuming independent contributions from each feature vector channel f_j using assignment probabilities $p_1 = \prod_j p_{1j}$ and $p_2 = \prod_j p_{2j}$. In many cases, the p_{ij}'s are modeled by unimodal Gaussian region descriptor distributions so that $p_{ij}(x) = \mathcal{N}_{f_j}(\mu_{ij}, \sigma_{ij})$ [10], with mean μ_{ij} and variance σ_{ij}. Furthermore, μ_{ij} and σ_{ij} may act as locally calculated parameters that depend on the pixel position x. Remark that if we assume a single μ_{ij} and σ_{ij} for the entire region, (1) reduces to the standard Mumford-Shah functional as used in [8]. There are also approaches where the distributions are approximated in a multimodal way [9] e.g. by Gaussian Mixture Models or Nonparametric Parzen Density Estimates [13].

3.2 A Multi-Dimensional Histogram-Based Level-Set Method for Image Segmentation

For the multi-dimensional histogram-based level-set method presented in this paper, we propose to use multi-dimensional nonparametric region descriptor functions. In comparison to the commonly used Gaussian Mixture Models, we present an approach that represents the region descriptors extensively in a multi-dimensional grid-based way. Thus, the feature vector channels f_j are no longer assumed to contribute independently from each other to the assignment probabilities p_i via the p_{ij}'s, but span a *single* multi-dimensional feature space $\rho_i(\mathbf{f})$. To this end, we calculate for the entire feature space $\mathbf{f}$ inside a region i a normalised histogram-vector $\mathbf{h}_i$ with single entries indexed by $\boldsymbol{k} = (k_1, k_2, \cdots, k_j, \cdots, k_J)^T$ where

$$h_{i\boldsymbol{k}} = \frac{\int\limits_{\Omega} \chi_i(\phi)\hat{h}_{\boldsymbol{k}}(x)dx}{\int\limits_{\Omega} \chi_i(\phi)dx} \tag{3}$$

and

$$\hat{h}_{\boldsymbol{k}}(x) = \prod_j \left(H(f_j(x) - b_{k_j}) - H(f_j(x) - b_{k_j+1})\right) \tag{4}$$

with hyper-bins indexed by vector $\boldsymbol{k}$ and borders of the histogram hyper-bins defined by b_k[3]. For equally spaced b_k's, the hyper-bins become hyper-cubes in the feature space of $\mathbf{f}$. Smoothed versions of the multi-dimensional histogram $\mathbf{h}_i$ can be gained by convolving it with a multi-dimensional Gaussian kernel of the same dimensionality, but in our applications smoothing the histogram did not change the results substantially.

The standard level-set method as described in the above section is extended by using the normalised multi-dimensional histogram $\mathbf{h}_i$ as the feature-dependent region descriptor $\rho_i(\mathbf{f})$. The region assignment probability is then calculated by

$$p_i(x) = \sum_{\boldsymbol{k}} \hat{h}_{\boldsymbol{k}}(x) h_{i\boldsymbol{k}} := \sum_{k_1} \sum_{k_2} \cdots \sum_{k_j} \cdots \sum_{k_J} \hat{h}_{\boldsymbol{k}}(x) h_{i\boldsymbol{k}} \tag{5}$$

i.e., by extracting the histogram entry of $\mathbf{h}_i$ that corresponds to the hyper-bin indicated by $\mathbf{f}(x)$. In this way, both the region descriptor function as well as the computation of the region assignment become computationally inexpensive, since they amount to calculating and extracting single entries from normalised multi-dimensional histograms.

4 Main Results

In order to show the performance and some internal details of the proposed algorithm two exemplary source images were chosen. Both images are coloured, given in the RGB colour space and used without further preprocessing, thus the segmentation is based on three feature channels, namely the red, green and blue colour channel. The method proposed in this paper is not constrained to these specific features or to exactly three features, since other features, e.g. texture, might be utilised as well. The usage of other features was deliberately omitted to show the capability of the algorithm even in the elementary and commonly used RGB colour space.

The first image shows a zebra standing in its natural environment, the steppe. The image consits of the black and white and shades of grey of the zebra, which constitutes the object to segment and the green and beige colouring of the surrounding steppe. Zebra images are common test images for texture based segmentation algorithms. Here we show that even without a description of texture the segmentation task can be successfully accomplished. Figure 1 shows the image overlaid by the initial and final level-set contours of the segmentation process. On the left, the initial level-set contour, a circle centred in the middle of the image and featuring a radius of one fourth of the smallest image dimension, is displayed. This initial level-set contour is commonly used to express the expectation of an object, e.g. gained by a preprocessing stage previous to the segmentation framework that focuses on salient points, like in autonomous mobile robotics. Figure 1, right, displays the final level-set contour that is obtained

[3] Assuming for simplicity same bin spacing for all feature dimensions j.

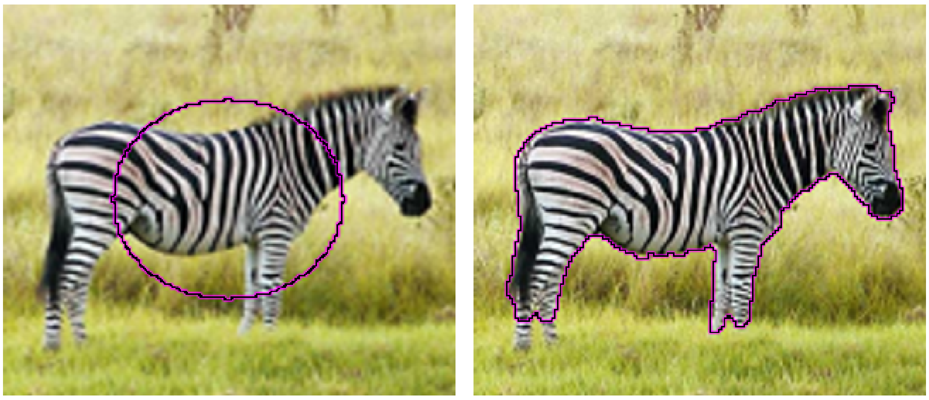

Fig. 1. Initial (left) and final (right) level-set contour of the zebra test image. The segmentation result was achieved after 37 iterations with the multi-dimensional, histogram-based RGB region-descriptor and without any further specific feature channel (e.g. texture).

after 37 iterations of (2). The evolution of the level-set function is stopped according to the development of the value of the energy-functional (1). Figure 2 displays the progress of the values of the energy-functional over iterations. For convenience, the values are normalised to the interval $[0, 1]$. After 29 iterations, the energy has converged to its minimum. The algorithm needs eight consecutive iterations to detect the convergence and stop the segmentation process. Figure 3 displays the region descriptors for the inside and outside regions of the final level-set contour, $\rho_1(\mathbf{f})$ and $\rho_2(\mathbf{f})$, respectively. In the case of using the RGB colour space as the only features, the region descriptors equal the colour distribution of the object and its surrounding. In Fig. 3, left, the distribution of the colours belonging to the zebra, which is mainly composed of black and white and shades of grey, can be observed as the colours are grouped along the diagonal from black to white. The colour distribution of the outside, that mainly consists of a green and beige colouring, can be noticed in Fig. 3, right, where the colours stay in the "greenish" corner of the colour space.

The second image is used in [1] to compare different state-of-the-art image segmentation methods. It was chosen to show the competitive results of the approach proposed in this paper. Figure 4 displays the final level-set contour of the segmentation process, as described in the preceding paragraph. With the ground-truth data provided in [1] and the error measurement introduced by [1] we achieve an error rate of 1.28% of misclassified pixels w.r.t. the number of initially unclassified pixels. This errore rate is comparable to the average error rate of the best performing state-of-the-art image segmentation method, which is specified by 1.36% in [1].

In Fig. 5 we show segmentation results of additional exemplary test images from the database provided in [1]. The segmentation results show an error rate of 1.63%, 0.72% and 1.43% misclassified pixels.

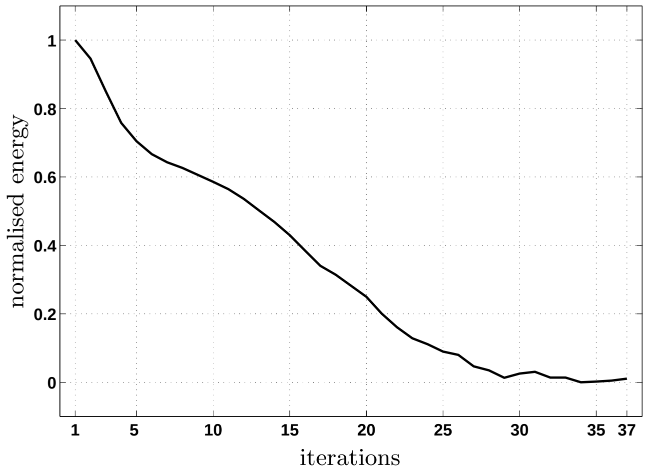

Fig. 2. Progress of the (normalised) energy over iterations. The energy converges after 29 iterations. The algorithm requires eight consecutive iterations to detect the convergence and stop the segmentation process.

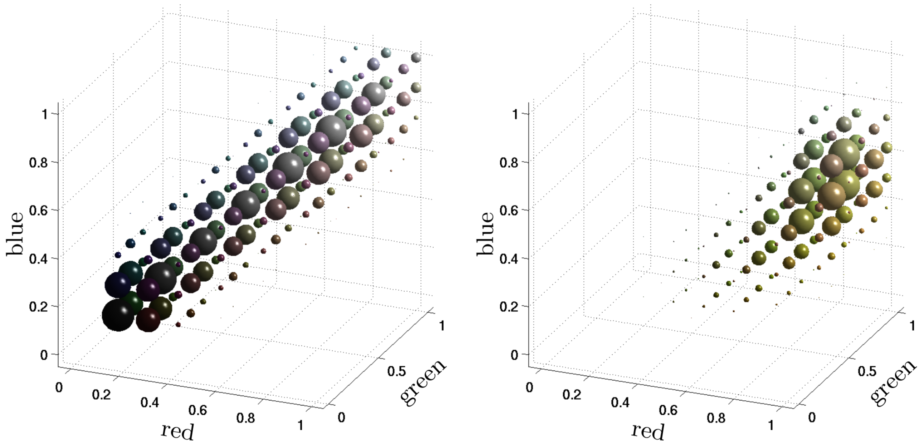

Fig. 3. Distribution (multi-dimensional colour histograms) inside (left) and outside (right) of the final level-set contour of the zebra test image, shown in the three-dimensional feature space spanned by the three colours red, green and blue. Larger and smaller blobs indicate larger and smaller histogram values, respectively. Only colours with a contribution greater than 1% are displayed.

Fig. 4. Final contour of the llama test image from [1] achieved with the segmentation method proposed in this paper. The segmentation result shows an error rate of 1.28% misclassified pixels based on the error measurement and ground-truth data provided in [1].

Fig. 5. Final contour of exemplary test images from the database provided in [1]. The segmentation results show an error rate of 1.63%, 0.72% and 1.43% misclassified pixels based on the error measurement and ground-truth data provided in [1] (from left to right). A preliminary evaluation of the proposed method with all 50 benchmark images (without special tuning to the database) resulted in an average error rate of 2.25%.

5 Conclusion

We have presented an approach for multi-dimensional histogram-based image segmentation that is embedded in a level-set framework for two-region segmentation. Contrary to standard level-set methods for image segmentation we assumed that the features on which the segmentation is based on are part of a *single* feature space. In contrast to recent state-of-the-art image segmentation methods, we did not model the feature distributions based on Gaussian Mixture Models, but applied multi-dimensional histogram-based feature models and showed that the proposed approach yields competitive results. Furthermore no specific features (e.g. texture) were needed to achieve the presented results.

A number of state-of-the-art image segmentation methods provide an alpha mask as segmentation result, that assigns each pixel in a probabilistic manner to the inside and outside region, respectively. In a level-set framework, an alpha mask is not explicitly incorporated but can be easily extraceted as a by-product by evaluating the $p_i(x)$ of (5) as $\alpha(x) = p_2(x)/\left(p_1(x) + p_2(x)\right)$.

References

1. Rother, C., Kolmogorov, V., Blake, A.: "GrabCut": Interactive foreground extraction using iterated graph cuts. ACM Trans. Graph. **23**(3) (2004) 309–314
2. Boykov, Y.Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on Computer Vision **1** (2001) 105–112
3. Corel Corperation: Knockout User Guide. (2002)
4. Chuang, Y.Y., Curless, B., Salesin, D., Szeliski, R.: A bayesian approach to digital matting. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume 2. (2001) 264–271
5. Osher, S., Sethian, J.A.: Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. J. Comput. Phys. **79** (1988) 12–49
6. Mumford, D., Shah, J.: Optimal approximation by piecewise smooth functions and associated variational problems. Commun. Pure Appl. Math **42** (1989) 577–685
7. Zhu, S.C., Yuille, A.L.: Region competition: Unifying snakes, region growing, and bayes/MDL for multiband image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **18**(9) (1996) 884–900
8. Chan, T., Vese, L.: Active contours without edges. IEEE Trans. Image Process. **10**(2) (2001) 266–277
9. Kim, J., Fisher, J.W., Yezzi, A.J., Çetin, M., Willsky, A.S.: Nonparametric methods for image segmentation using information theory and curve evolution. In: International Conference on Image Processing, Rochester, New York. Volume 3. (2002) 797–800
10. Rousson, M., Deriche, R.: A variational framework for active and adaptative segmentation of vector valued images. IEEE Workshop on Motion and Video Computing, Orlando, Florida (2002)
11. Brox, T., Rousson, M., Deriche, R., Weickert, J.: Unsupervised segmentation incorporating colour, texture, and motion. Computer Analysis of Images and Patterns **2756** (2003) 353–360
12. Grossberg, Stephen, Hong, Simon: A neural model of surface perception: Lightness, anchoring, and filling-in. Spatial Vision **19**(2-4) (2006) 263–321
13. Parzen, E.: On the estimation of a probability density function and mode. Annals of Mathematical Statistics **33** (1962) 1065–1076
14. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International Journal for Computer Vision **1**(4) (1988) 321–331
15. Chan, T., Sandberg, B., Vese, L.: Active contours without edges for vector-valued images. J. Visual Communication Image Representation **11**(2) (2000) 130–141