ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

Figure-ground Segmentation using Metrics Adaptation in Level Set Methods

A. Denecke[1,2], I. Ayllon Clemente[1,2], H. Wersing[2], J. Eggert[2], J. J. Steil[1]

1- CoR-Lab - Bielefeld University
P.O.-Box 10 01 31, D-33501 Bielefeld - Germany
{adenecke,iayllon}@cor-lab.uni-bielefeld.de

2- Honda Research Institute Europe
Carl-Legien-Str. 30, 63073 Offenbach/Main - Germany

Abstract. We present an approach for hypothesis-based image segmentation basing on the integration of level set methods and discriminative feature clustering techniques. Building up on previous work, we investigate Localized Generalized Matrix Learning Vector Quantization (LGMLVQ) to train a classifier for fore- and background of an image. We extend this concept towards level set segmentation algorithms, where region descriptors are used to adapt the object contour according to the image features. The fusion of both methods outperforms their individual applications and improve the performance compared to other state of the art segmentation methods.

1 Introduction

In computer vision, figure-ground segmentation (FGS) divides an image into two regions containing the object of interest and the background respectively. Hypothesis-driven approaches for FGS rely on an initial hypothesis that provides an a priori assumption (e.g. from user interaction [1] or depth estimation [2]) about a pixelwise affiliation to object or background. Unfortunately they typically include incomplete or partially erroneous cues which can be caused by the user or algorithmic problems. Hypothesis-based FGS consists of two steps: the modeling of the feature-statistics of the hypothetical fore- and background and the consecutive integration of those statistics in energy minimization techniques like Markov random field formulations [1] or level set methods [3]. These algorithms allow for advanced concepts like interactions of neighboring pixels or contour constraints to derive compact regions. For example, Rother et al.[1] use Gaussian mixture models together with the min-cut algorithm to optimize the partition of an image. Similarly in [4], histograms are used as region descriptors and are integrated into a level set energy functional including a smoothness term to derive compact foreground segmentations. The statistical or descriptive modeling of fore- and background does not account for the discriminability of the used features (e.g. in the case of identical colors in fore- and background). In [2], the statistics are modeled with prototypical feature representatives, where an extended learning vector quantization approach [5, 6] is used to train a classifier for fore- and background. There an integrated feature weighting is employed, that discriminates between both regions.

In this paper, we extend the concept of metrics adaptation [6] for FGS towards a level set formulation. On the one hand, the feature weighting mechanism implemented by the metrics adaptation in Generalized Learning Vector Quantization (GLVQ) [5, 6] improves the discrimination capabilities and yields a more precise region modeling. On the other hand, the introduction of an additional region constraint provided by the level set formulation leads to spatially coherent results and reduces the dependence on the initial hypothesis.

2 Metrics Adaptation

Several methods can be used to model the statistics of fore- and background, e.g. descriptive models like histograms [4]. In previous work [2] instead we use a prototype-based classifier to represent homogeneous image regions by prototypical feature representatives. The image data $\mathcal{I}$ consist of $M = 5$ feature maps $\mathcal{F} := \{F_i(\mathbf{x})|i = 1..M\}$ (RGB color and position information) and form the dataset $\mathcal{D} := \{\vec{\xi}|\vec{\xi}(\mathbf{x}) = (F_1(\mathbf{x})..F_M(\mathbf{x}))^T, \mathbf{x} \in \mathcal{I}\}$, comprising a feature vector $\vec{\xi}(\mathbf{x})$ at every image position $\mathbf{x}$. To represent the dataset $\mathcal{D}$ by a set of prototypical representatives, several learning methods e.g. standard vector quantization, can be used. The Generalized Learning Vector Quantization (GLVQ) [5] algorithm is defined by a network of N class-specific prototypical feature representatives $\mathcal{P} := \left\{w_p \in \mathbb{R}^M|p = 1..N\right\}$. Since LVQ is a supervised learning method, a two class setup is used for figure-ground segmentation, where $c(w_p) \in \{0, 1\}$ encodes the a priori (e.g. by the user) assigned class-membership of every prototype. The goal of the GLVQ learning dynamics is to optimize the representatives w_p according to the classification error defined by the functional $E[\mathcal{D}, \mathcal{P}] = \sum_{\vec{\xi}\in\mathcal{D}} \frac{1}{1+e^{-\mu(d)}}$, with $\mu(d) = \frac{d_J - d_K}{d_J + d_K}$. Here the variables $d_J = d(\vec{\xi}, w_J)$ and $d_K = d(\vec{\xi}, w_K)$ are the distances of a randomly selected feature vector $\vec{\xi} \in \mathcal{D}$ to the most similar prototype w_J, $c(\vec{\xi}) = c(w_J)$ from the correct class and w_K from an incorrect class, respectively.

Instead of using the standard Euclidean metrics $d(\vec{\xi}, w_p) = \| \vec{\xi} - w_p \|$, recently several adaptive metrics were proposed [6]. In the most general case, a Mahalanobis-like metrics $d(\vec{\xi}, w_p) = (\vec{\xi} - w_p)^T \Lambda_p (\vec{\xi} - w_p)$ is used, where the distance computation is extended towards a prototype specific $M \times M$ matrix Λ_p of relevance factors (Localized Generalized Matrix LVQ, LGMLVQ). Using metrics adaptation allows a weighting of the features according to the classification task as well as complex non-linear decision boundaries also for a reduced number of prototypes compared to the standard LVQ with multiple prototypes. The prototypes w_J and w_K as well as the corresponding relevance factors Λ_J and Λ_K are optimized by means of a stochastic gradient descent method according to $E[\mathcal{D}, \mathcal{P}]$ on randomly chosen pairs $(\vec{\xi}, c(\vec{\xi}))$ (see [2] for the detailed derivatives of $E[\mathcal{D}, \mathcal{P}]$). Classification relies on nearest neighbor search where the label of the prototype with smallest distance $d(\vec{\xi}, w_p)$ is assigned to a given feature $\vec{\xi}$. The decision boundary or the confidence of the classification is represented by the normalized margin $\mu(d)$, which is small if $\vec{\xi}$ has a similar distance to the

ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

prototypes w_J and w_K.

To apply this method for image segmentation [2], an hypothesis is used in the following way. The hypothesis $\mathcal{H}$ is represented as a binary map indicating which pixels belong to the foreground $\mathcal{H}(\mathbf{x}) = 1$ or the background $\mathcal{H}(\mathbf{x}) = 0$. In the case of metrics adaptation, $\mathcal{H}$ is used as label $c(\vec{\xi}(\mathbf{x})) := \mathcal{H}(\mathbf{x})$ for the image features to allow for an optimization of the prototypes $\mathcal{P}$. To segment an image on the basis of the adapted classifier, the image is partitioned into N segments (binary maps) $V_p \in \{0, 1\}$ by assigning all feature vectors $\vec{\xi}(\mathbf{x})$ (i.e. pixels of a particular image) independently to the best matching prototype. The final segmentation $\mathcal{A}$ is combined by choosing the binary maps from the prototypes assigned to the foreground $\mathcal{A} = \sum_p^N c(w_p)V_p$.

3 Extension towards level set methods

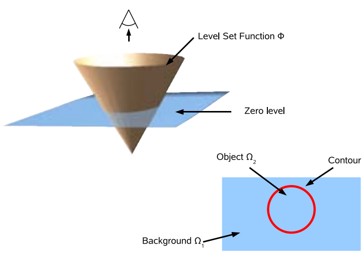

Figure 1: Level set model. The level set function $\phi(\mathbf{x})$ as a function of the image position $\mathbf{x}$ returns a height defining a 2D surface. The cone-shaped surface intersects the X-Y plane at zero height, implicitly representing the contour.

Level set methods [3] are a class of numerical algorithms derived from active contour approaches. They use local information measured around the contour (e.g. the image gradient or global features as color and texture) to align the contour with the object boundary:

$$\phi(\mathbf{x}) = \begin{cases} \phi(\mathbf{x}) < 0 & \text{if} \quad \mathbf{x} \in \Omega_2 \\ \phi(\mathbf{x}) = 0 & \text{if} \quad \mathbf{x} \in \Omega^- \\ \phi(\mathbf{x}) > 0 & \text{if} \quad \mathbf{x} \in \Omega_1 \end{cases} . \quad (1)$$

There are two approaches to represent active contours: explicitly (e.g. by a set of control points changing their position) and implicitly. In implicit representation approaches as level set methods, the contour is defined by the level set function $\phi(\mathbf{x}) \in \Omega \mapsto \Re$ (Eq. 1), which divides the image plane Ω into two disjoint regions, Ω_1 representing the background region, Ω_2 the segmented object, and Ω^- for the contour of the segmented object itself.

For implicit contour representations, prominent formulations of energy functional for image segmentation were given by Mumford and Shah [7], where they use the mean gray value of a region as a simple region descriptor. This concept was adopted by [8] formulating an extended energy functional, where additional constraints on the contour length and region size are imposed.

$$E(\phi(\mathbf{x})) = \sum_{i=1}^{2} \int_\Omega X_i(\phi(\mathbf{x}))\cdot(\vec{\xi}(\mathbf{x})-\rho_i)^2 d\mathbf{x} + \nu \int_\Omega |\nabla H(\phi(\mathbf{x}))| d\mathbf{x} + \gamma \int_\Omega X_2 d\mathbf{x} \quad (2)$$

Here $X_1 = H(\phi(\mathbf{x}))$ only equals '1' when $\phi(\mathbf{x}) > 0$ and $X_2 = 1 - H(\phi(\mathbf{x}))$ when $\phi(\mathbf{x}) < 0$, where $H(\phi(\mathbf{x}))$ is the Heaviside function.

The region descriptors ρ_1 and ρ_2 in Eq. 2 are the average values of both regions, i.e. $\rho_1 = \frac{\int_\Omega X_1 \vec{\xi}(\mathbf{x}) d\mathbf{x}}{\int_\Omega X_1 d\mathbf{x}}$ and $\rho_2 = \frac{\int_\Omega X_2 \vec{\xi}(\mathbf{x}) d\mathbf{x}}{\int_\Omega X_2 d\mathbf{x}}$, were the first term of the energy functional is minimal for a grouping into homogeneous regions. The additional smoothness constraint favors compact regions as well as smooth region boundaries.

The following Halmiton-Jacobi equation results from the minimization of the energy functional with respect to the level set function $\phi(\mathbf{x})$ using gradient descent:

$$\frac{\partial \phi(\mathbf{x})}{\partial t} = \delta(\phi(\mathbf{x}))[\nu \cdot \operatorname{div}\left(\frac{\nabla \phi(\mathbf{x})}{|\nabla \phi(\mathbf{x})|}\right) + \gamma + \lambda_1(\vec{\xi}(\mathbf{x}) - \rho_1)^2 + \lambda_2(\vec{\xi}(\mathbf{x}) - \rho_2)^2]. \quad (3)$$

This method combines the level set evolution by mean curvature [9] (i.e. $\frac{\partial \phi}{\partial t} = |\bigtriangledown \phi| \, div(\frac{\bigtriangledown \phi}{|\bigtriangledown \phi|})$) with an optimization of a single prototype for each region.

Extension of the GLVQ functional. Instead of using only one prototype to represent each region [8, 7] a LVQ network can be regarded as a generalization towards multiple ones, allowing for a heterogeneous appearance of an object and its background. To generalize the concept of metrics adaptation towards a level set formulation, the GLVQ error function (Sec. 2) can be extended by the contour term as:

$$E(\phi(\mathbf{x})) = \int_\Omega \frac{1}{1 + e^{-\mu(d)}} d\mathbf{x} + \nu \cdot \int_\Omega |\bigtriangledown H(\phi(\mathbf{x}))| d\mathbf{x}. \quad (4)$$

The first term corresponds to the classification error where the sum over all pixels is replaced by the integral over the level set function $\phi(\mathbf{x})$. This error term is minimal if both regions can be well represented and discriminated whereas the second term prefers short contours and compact regions as discussed before. To minimize the proposed level set function, the gradient can be approximated as:

$$\frac{\partial \phi}{\partial t} = \delta(\phi(\mathbf{x}))[\nu \cdot \operatorname{div}\left(\frac{\bigtriangledown \phi(\mathbf{x})}{|\bigtriangledown \phi(\mathbf{x})|}\right) - C(\phi(\mathbf{x})) \cdot \mu(d(\xi(\mathbf{x}))) + (1 - C(\phi(\mathbf{x}))) \cdot \mu(d(\xi(\mathbf{x})))]. \quad (5)$$

Informally, the level set function is modified by the confidence of the classification, represented by the margin $\mu(d(\xi(\mathbf{x})))$ (Sec. 2). In regions where the classification is very confident, indicated by a large margin, a strong adaptation occurs in the direction estimated by the classifier (indicated by $C(\phi(\mathbf{x}))$, where $C(\phi(\mathbf{x})) = 1$ if the pixel is classified as foreground and 0 otherwise).

Optimization. The algorithm starts with an initial contour provided by the hypothesis $\mathcal{H}$, i.e. $\phi_{init}(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad H(\mathcal{H}) = 0 \\ -1 & \text{if} \quad H(\mathcal{H}) = 1 \end{cases}$. The iterative optimization of the level set function $\phi(\mathbf{x})$ consists of two steps. The first steps keeps $\phi(\mathbf{x})$ fixed and minimizes the energy with respect to the prototypes $\mathcal{P}$ and relevance matrices Λ by standard LGMLVQ learning (Sec. 2) according to an intermediate hypothesis $\mathcal{H} = (1 - H(\phi(\mathbf{x})))$. In the second step, the level set function is adapted according to Eq. 5 using Heun's method [10], following the general form $y_{i+1} = y_i + \epsilon \cdot h$, with ϵ extrapolating from an old value y_i to a new value y_{i+1} with a step size h. Both steps are iteratively computed along the initial level set function until the function $\phi(\mathbf{x})$ converges or a maximum number of iterations is reached. In general, the level set function is updated close to the zero level set Ω^- determined by the regularized delta function $\delta(\phi, \tau) = \frac{1}{\pi} \cdot \frac{\tau}{\tau^2+\phi^2}$, where $\tau = 2.25$. Further parameters are the weighting of the mean curvature evolution $\nu = 1.3$, the parameters of the metrics adaptation adopted from [2] ($\alpha = 0.05$, $\beta = 0.005$ using 10.000 adaptation steps each iteration) and the number of prototypes (5 for foreground and 3 for background in all experiments).

4 Experiments

The performance of the proposed method is evaluated on public benchmark data [1]. The dataset consists of images of sample objects together with the ground truth segmentation and a Trimap $T = \{T_I = 0, T_B = 64, T_U = 128, T_F = 255\}$ specifying the affiliation of every pixel to foreground T_F or background T_B (unknown status T_U, ignored regions T_I). The initial hypothesis $\mathcal{H}$ is generated by selecting T_I, T_B for background and T_F, T_U for foreground. The quality of the segmentation is evaluated according to the pixelwise similarity to the ground truth segmentation in two different setups (Condition A: single set of parameters for all images, Condition B: individual contour weight ν for every image). The results in Table 1 show that in Condition A the proposed method can successfully improve the segmentation accuracy (it exceeds the baseline similarity of hypothesis $\mathcal{H}$) as well as improve the result compared to the individual application of metrics adaptation and level set methods with histograms [4]. Adapting the curvature weight ν for every image separately, finally yields a significantly improved performance in Condition B. This experiment indicates the effectiveness of the method according to the chosen parameters but is not directly comparable to the other results.

5 Discussion

We propose a new level set formulation, where a discriminative approach is followed to model the statistics of fore- and background instead of descriptive region modeling. We show that the proposed integration of metrics adaptation and level set methods yields a mutual benefit and achieves competitive results on public benchmark data. The introduction of an additional region constraint provided by the level set formulation leads to spatially coherent results while the iterative minimization also reduces the dependence of the initial data labeling. Nevertheless problems can be identified, which are related to the choice of the level set parameters, in particular the weighting for the curvature term, which is a well known problem for level set methods and open for future work. Several extensions to the proposed method are also possible, e.g. the estimation of the parameters on the data, incremental methods to estimate the model complexity

ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

Tab. 1

Method	Error rate
$\mathcal{H}$	07.72% ± 03.41
Condition A	02.41% ± 01.96
Condition B	01.73% ± 01.63
LGMLVQ [1]	04.15% ± 03.15
Level-set	04.98% ± 03.31
Graph-Cut [1]	12.90% ± 12.70

Figure 2: Left: Comparison of the segmentation to the ground truth (mean and std. dev. of pixel-wise error rates for 50 images), showing the similarity of $\mathcal{H}$ as well as the similarity of the foreground segmentation and comparable results from other state-of-the art methods. Right: Four examples (blue outline for ground truth, red outline for the boundary of foreground segmentation). Some problems are visible due to a wrong curvature weight ν as well as systematic errors because of shadows occurring near the object boundary.

(number of prototypes), an integration of user-constraints or a direct extension towards a three dimensional segmentation of video data.

References

[1] C. Rother V. Kolmogorov, A. Blake. "GrabCut": interactive foreground extraction using iterated graph cuts. *ACM Trans. on Graphics*, 23(3):309–314, 2004.

[2] A. Denecke, H. Wersing, J J. Steil, and E. Körner. Online figure-ground segmentation with adaptive metrics in generalized LVQ. *Neurocomputing*, 72(7-9):1470 – 1482, 2009.

[3] S. Osher R. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces.* Springer, October 2002.

[4] D. Weiler and J. Eggert. Multi-dimensional histogram-based image segmentation. In *Procs of the 14th Int. Conf. on Neur. Information Processing*, 2007.

[5] A. Sato K. Yamada. Generalized learning vector quantization. In *Advances in Neural Information Processing Systems*, volume 7, pages 423–429, 1995.

[6] P. Schneider, M. Biehl, and B. Hammer. Relevance matrices in LVQ. In *Proc. of the Europ. Symp. on Art. Neur. Netw. (ESANN)*, pages 37–42, 2007.

[7] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. on Pure and App. Mathematics*, 42(5):577–685, 1989.

[8] T. F. Chan and L. A. Vese. Active contours without edges. *Image Processing, IEEE Transactions on*, 10(2):266–277, 2001.

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

[10] S.C. Chapra and R.P. Canale. *Numerical Methods for Engineers.* McGraw-Hill, 2nd edition, 1988.