A Probabilistic Prediction Method for Object Contour Tracking

Daniel Weiler, Volker Willert, Julian Eggert

2008

A Probabilistic Prediction Method for Object Contour Tracking

Daniel Weiler[1], Volker Willert[2], and 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 probabilistic contour prediction in an object tracking system. We combine level-set methods for image segmentation with optical flow estimations based on probability distribution functions (pdf's) calculated at each image position. Unlike most recent level-set methods that consider exclusively the sign of the level-set function to determine an object and its background, we introduce a novel interpretation of the value of the level-set function that reflects the confidence in the contour. To this end, in a sequence of consecutive images, the contour of an object is transformed according to the optical flow estimation and used as the initial object hypothesis in the following image. The values of the initial level-set function are set according to the optical flow pdf's and thus provide an opportunity to incorporate the uncertainties of the optical flow estimation in the object contour prediction.

1 Introduction

In this paper we propose an object contour tracking approach based on level-set methods for image segmentation and correlation-based patch-matching methods for optical flow estimation. Using level-set methods for object detection enables us to overcome the problems imposed by nonrigid object deformations and object appearance changes. In tracking applications with dynamic template adaption these changes lead to template drift and in applications without template adaption to a decreased robustness. Utilising probabilistic optical flow for the prediction of the object contour constitutes a non-parametric prediction model that is capable of representing nonrigid object deformation as well as complex and rapid object movements and thus providing a segmentation method with a reliable initial contour that leads to a robust and quick convergence of the level-set method even in the presence of a comparably low camera frame rate. Furthermore, we introduce a novel interpretation of the value of the level-set function. Unlike most recent level-set methods that consider exclusively the sign of the level-set function to determine an object and its surroundings, we use the value of the level-set function to reflect the confidence in the predicted initial contour. This yields a robust and quick convergence of the level-set method in those sections of the contour with a high initial confidence and a flexible and

mostly unconstrained (and thus also quick) convergence in those sections with a low initial confidence.

The segmentation occurs by means of level-set methods [1–5], which separate all image pixels into two disjoint regions [1] by favouring homogeneous image properties for pixels within the same region and dissimilar 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. Minimising the energy functional leads to the segmentation of the image. The formulation of the energy functional dates back to e.g. Mumford and Shah [2] and to Zhu and Yuille [3]. Later on, the functionals were reformulated and minimised using the level-set framework e.g. by [4]. Among all segmentation algorithms from computer vision, level-set methods provide perhaps the closest link with the biologically motivated, connectionist models as represented e.g. by [6]. 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.

Optical flow estimation, i.e. the evaluation of the pixel-motion in a sequence of consecutive images, yielded two prominent solution classes. Namely gradient-based differential [7, 8] and correlation-based patch-matching [9, 10] algorithms. While the former is based on the *gradient constraint equation* that utilises spatiotemporal derivatives of the image intensity and thus requires nearly linear image intensity resulting in velocities smaller than one pixel per frame and a high frame rate of the camera, the latter uses similarity or distance measures between a small patch of an image and its shifted counterpart that leads to discrete velocities and comparatively high computational costs.

In [11] a comprehensive survey of object tracking algorithms is given. Depending on the vision task, object tracking algorithms are based on several object representations , object detection strategies and prediction methods for the object location . Nonrigid object deformation (e.g. walking person), complex and rapid object movements (e.g. playing children), entire object appearance changes (e.g. front side vs. back side) and object occlusions form some of the numerous challenges in the field of object tracking.

In this paper we propose an approach that combines level-set segmentation algorithms and optical flow estimation methods to form a tracking system. With that combination we are able to overcome some of the principle problems the approaches exhibit, when utilised separately (e.g.: initial level-set function, local optima of the energy functional, aperture problem). The paper is organised as follows. In Sect. 2.1 and 2.2 we describe the level-set method applied for image segmentation and the probabilistic optical flow estimation used for the prediction of the initial object contour, respectively. Section 3 introduces the proposed prob-

abilistic prediction method for object contour tracking. In Sect. 3.1 we suggest a novel interpretation of the value of the initial level-set function. An approach for level-set based object contour tracking based on a *parametric* prediction model is introduced in Sect. 3.2, and extended by a *non-parametric* prediction model in Sect. 3.3. The results of the proposed algorithms are presented in Sect. 4. A short discussion finalises the paper.

2 Level-Set Segmentation and Optical Flow Estimation

2.1 Standard Level-Set based Region Segmentation

Level-set methods are front propagation methods. Starting with an initial contour, a 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 [2, 3].

Compared to "active contours" (snakes) [12], 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].

The proposed object contour tracking framework is based on a standard two-region level-set method for image segmentation. 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 (background) and Ω_2 (object), where $\phi(x) > 0$ if $x \in \Omega_1$ and $\phi(x) < 0$ if $x \in \Omega_2$. 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

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

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 region 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] $H(x) = 1 \,\text{for}\, x > 0$ and $H(x) = 0 \,\text{for}\, x \leq 0$.

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.

2.2 Probabilistic Optical Flow Estimation

The characteristic motion pattern of an object in an image sequence $\mathbf{I}^{1:t}$ at time t is given by the optical flow $\mathbf{V}^t$ within the region that constitutes the object. The optical flow $\mathbf{V}^t = \{\mathbf{v}_x^t\}$ is the set of velocity vectors $\mathbf{v}_x^t$ of all pixels at every location x in the image $\mathbf{I}^t$ at time t, meaning that the movement of each pixel is represented with one velocity hypothesis. This representation neglects the fact that in most cases the pixel movement cannot be unambiguously estimated due to different kinds of motion-specific correspondence problems (e.g. the aperture problem) and noisy data the measurement is based on.

As has been suggested and discussed by several authors [10], velocity probability density functions (pdf's) are well suited to handle several kinds of motion ambiguities. Following these ideas we model the uncertainty of the optical flow $\mathbf{V}^t$ as follows:

$$P(\mathbf{V}^t | Y^t) = \prod_x P(\mathbf{v}_x^t | Y^t) \qquad \text{with} \quad Y^t = \{\mathbf{I}^t, \mathbf{I}^{t+1}\}\,, \tag{3}$$

where the probability for the optical flow $P(\mathbf{V}^t|Y^t)$ is composed of locally independent velocity pdf's $P(\mathbf{v}_x^t|Y^t)$ for every image location x. $P(\mathbf{v}_x^t|Y^t)$ can be calculated using several standard methods, for details refer e.g. to [10]. These pdf's fully describe the motion estimations available for each position x, taking along (un)certainties and serving as a basis for the probabilistic prediction method for object contour tracking as proposed in Sect. 3.3.

3 Probabilistic Prediction Method for Contour Tracking

3.1 Interpretation of the Value of the Initial Level-Set Function

In general, level-set methods evaluate exclusively the sign of the level-set function to determine an object and its surroundings. The exact value of the level-set function is not considered by most approaches. Signed-distance functions are a common means of regulating the value of the level-set function, as they enforce

the absolute value of the gradient of the level-set function to be one. For the approach we propose in this paper (explained in detail in the next section), it is required to extend the common understanding of the values of the level-set function. Considering the front propagation and gradient descent nature of the applied level-set method for image segmentation, the height of the level-set function influences the time (number of iterations) until the occurrence of a zero crossing (change of region assignment). In particular for numerical stability a maximum time step value is required. Thus, sections of the contour exhibiting large values of the level-set function in their neighbourhood generally move slower than those with smaller values. Following that idea, a steep gradient of the *initial* level-set function for a segmentation algorithm yields a slow deformation of the contour, whereas a flat gradient leads to a mostly unconstrained and quick deformation. Altogether this results in the possibility to control the velocity of the propagated front, embedded entirely and without any algorithmic changes in a standard level-set framework for image segmentation.

3.2 Level-Set based Segmentation in Image Sequences

Building an iterative level-set based object tracker, a trivial approach would be the usage of the final level-set function of the preceding image ϕ^{t-1} as the initial level-set function $\hat{\phi}^t$ of the current image. To accelerate the convergence of the minimisation process one might also use a smoothed version of the level-set function:

$$\hat{\phi}^t = K_\sigma * \phi^{t-1} \tag{4}$$

The performance of this approach depends on the velocity and deformations of the tracked object. While the approach will succeed in tracking the object in the presence of small movements and deformations, it is likely to fail under huge deformations or large object movements.

To circumvent the above mentioned problem, tracking algorithms include a prediction stage that estimates the object position in the next frame. Introducing a first order prediction method in our level-set based framework would consider the last two segmentation results χ_2^{t-1} and χ_2^{t-2}, measure the transformation between them and predict the current initialisation of the image segmentation algorithm on the basis of the measured transformation. A parametric approach, based on a similarity[2] transformation $\mathbf{A}$, requires the estimation $\boldsymbol{F}$ of four parameters, namely the translation vector $\boldsymbol{t} = (t_x, t_y)^T$, the rotation ω and scale s, comprised in a state vector $\boldsymbol{s} = (t_x, t_y, \omega, s)^T$. In a level-set framework the object translation might be estimated by the translation of the centre of gravity of the inside masks χ_2^{t-1} and χ_2^{t-2}, the rotation by the evaluation of the principal component[3] of the two masks and the scale by the square root of the mask area

[2] Similarity transformations constitute a subgroup of affine transformations where the transformation matrix A is a scalar times an orthogonal matrix.

[3] Here the principal component is the eigenvector to the largest eigenvalue of the covariance matrix of the positions of the points within the masks χ_2^{t-1} and χ_2^{t-2}.

ratio.

$$\hat{\phi}^t = \mathbf{A}(\phi^{t-1}, \boldsymbol{s}^{t-1}) \qquad \text{with} \quad \boldsymbol{s}^{t-1} = \boldsymbol{F}(\chi_2^{t-1}, \chi_2^{t-2}) \tag{5}$$

In contrast to the above approach with "zero order" prediction, even objects with high velocities can be tracked, as long as they move to some extent in accordance with the assumed similarity transformation model. Object movements that violate the prediction model, in particular high dynamic movements, again lead to failure.

To cope with high dynamic movements, higher order prediction models might be exploited, but they still underlie the limitation to movements that approximately follow the assumed model. Another approach includes the measurement of the real motion of all pixels (optical flow), belonging to the object, thus providing a means to accurately estimate the object position in the next frame, even in the presence of high dynamic movements. In this way, the prediction is not based on previous frames $Y^{t-2} = \{\mathbf{I}^{t-2}, \mathbf{I}^{t-1}\}$ only, but also on the current frame $Y^{t-1} = \{\mathbf{I}^{t-1}, \mathbf{I}^{t}\}$. Extending the above approach by the measurement of optical flow leads to the estimation of the state vector $\boldsymbol{s} = (t_x, t_y, \omega, s)^T$ from the flow field $\mathbf{V}^{t-1}$, that might be achieved by a regression analysis $\boldsymbol{R}$.

$$\hat{\phi}^t = \mathbf{A}(\phi^{t-1}, \boldsymbol{s}^{t}) \qquad \text{with} \quad \boldsymbol{s}^{t} = \boldsymbol{R}(\mathbf{V}^{t-1}, \chi_2^{t-1}) \tag{6}$$

Although the actual pixel velocities within the object are measured and used for an accurate prediction of the object position, a similarity transformation model is used for the prediction of the contour of the object. Strong deformations of the object will still lead to an imprecise initialisation of the image segmentation algorithm that might decrease the speed of convergence and the robustness of the segmentation. In the next section a purely non-parametric approach is introduced that comprises both a non-parametric estimation of the object position and a non-parametric estimation of the object deformation.

3.3 Probabilistic Prediction Method

In the following we propose an extension of the object tracking algorithm, developed in the previous section, that incorporates the optical flow measurement not only in the estimation of the object position, but also in determining the accurate deformation of the object. The optical flow $\mathbf{V}^t$ already contains all information required. Utilising an image processing warp[4] algorithm $\mathbf{W}_v$ that moves each pixel within an image according to a given vector field, enables us to purely non-parametrically predict an initial level-set function $\hat{\phi}^t$ for the segmentation of the current image $\mathbf{I}^t$.

$$\hat{\phi}^t = \mathbf{W}_v(\phi^{t-1}, \mathbf{V}^{t-1}) \tag{7}$$

If the optical flow estimation provides an additional confidence measure $\mathbf{C}$ a modulation of the prediction will lead to large values of the initial level-set function at locations with high confidence and to small values at locations with

[4] Backward warping yielded best results. For the backward warping, also the velocity vectors need to be measured back in time.

low confidence. Thus the flexibility of the moving contour, as introduced in Sect. 3.1 is adapted by the confidence of the optical flow estimation.

$$\hat{\phi}^t = \mathbf{W}_v(\phi^{t-1}, \mathbf{V}^{t-1}, \mathbf{C}) \tag{8}$$

In a last step, to introduce an even more robust and faster convergence of the proposed algorithm, the entire velocity pdf $P(\mathbf{V}^t|Y^t)$ is exploited in the prediction stage to determine not only an accurate initial region $\hat{\chi}_2^t$, but also provide an optimal slope (see Sect. 3.1) of the initial level-set function $\hat{\phi}^t$. Utilising a weighted warping algorithm $\mathbf{W}_p$ that moves each pixel within an image not only in one direction, but in all possible directions and overlays all moved pixels weighted by the probability $P(\mathbf{V}^t|Y^t)$ for the given pixel and direction, enables us to determine both the optimal initial region and the optimal slope of the initial level-set function $\hat{\phi}^t$.

$$\hat{\phi}^t = \mathbf{W}_p(\phi^{t-1}, P(\mathbf{V}^{t-1}|Y^{t-1})) \quad \text{with} \quad \hat{\phi}^t(x) = \sum_{v_{x'}^t} P(v_{x'}^{t-1}|Y^{t-1}) \cdot \phi^{t-1}(x - v_{x'}^{t-1}) \tag{9}$$

Altogether the proposed approach keeps the motion ambiguities of the optical flow estimation and yields a flat gradient of the initial level-set function at those sections of the contour where the information from the optical flow is ambiguous and offers only low confidence, leading to a mostly unconstrained and quick convergence. To the contrary, in regions of the contour where the optical flow has a high confidence, the predicted initial level-set function exhibits a steep gradient, enforcing only little change to the contour. The proposed approach enables a smooth transition between the prediction algorithm and the level-set image segmentation method. Thus, the deformation of the contour is locally controlled depending on which algorithm is superior. In sections of the contour with little structure and thus only small confidence in the optical flow measurement, the segmentation method will drive the contour evolution, whereas in sections, where the optical flow estimation is very accurate, the impact of the segmentation method on the contour deformation is suppressed and dominated by the prediction algorithm.

4 Main Results

In order to show the performance of the proposed approach three exemplary test image sequences were chosen. First, a sequence was artificially created with known ground truth by moving an object in front of a background. The movement was strictly based on similarity transformations, i.e. the transformations of the object exhibit exclusively translation, rotation and scale. Second, two real world examples were chosen. One outdoor scene with a driving car and an indoor scene with a high dynamically moving object.

Figure 1 (top row) shows two initial level-set functions, indicating the same initial figure-background condition of a circle in the middle of the image for the image segmentation algorithm and thus leading to the same segmentation

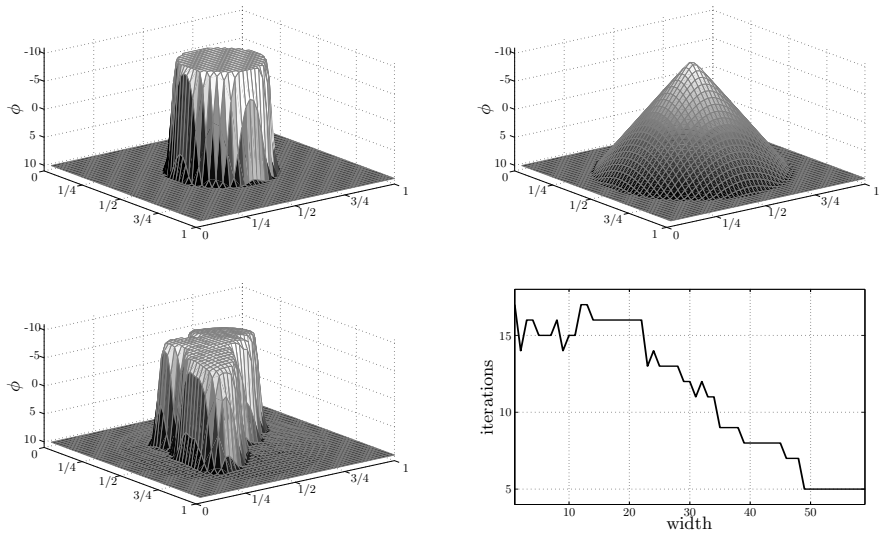

Fig. 1. Top row: Two different initial level-set functions ϕ for the same figure-background regions with steep (left) and flat (right) gradient. Bottom row: Final level-set function, after segmentation (left) and number of iterations until convergence of the segmentation algorithm, plotted for different widths of the contour of the initial level-set function (right).

result (bottom row, left). The only difference of the initial level-set functions are the steepness of their gradients at the contour, yielding different numbers of iterations until convergence of the segmentation algorithm. Figure 1 (bottom row, right) shows the number of iterations until convergence of the segmentation algorithm, depending on the width (steepness of the gradient) of the contour of the initial level-set function. In Fig. 2 one frame of the real-world test image sequence with high dynamic motion is shown in detail. The images are overlaid with segmentation results: previous segmentation result χ_2^{t-1} (grey), current segmentation initial condition $\hat{\phi}^t$ (white) and current segmentation result χ_2^t (black). Figure 2 shows two times the same frame, but processed with different prediction approaches. Whereas Fig. 2 (left) shows the results of a method with first order prediction based on the last two segmentation results χ_2^{t-2} and χ_2^{t-1} (5) yielding a segmentation initial condition that leads to unreliable tracking (the segmentation algorithm gets stuck in an unfavourable local minimum as the initial condition is already to far away from the desired final contour), Fig. 2 (right) shows an approach based on the probabilistic optical flow measurement (9), that is able to track the high dynamically moving object. Figure 3 shows an overview of the used test sequences: an artificial test sequence (top), an outdoor car sequence (middle) and an indoor sequence with high dynamic motion (bottom). The sequences are overlaid with the segmentation result χ_2^t (black) using the prediction method introduced in Sect. 3.3 (9).

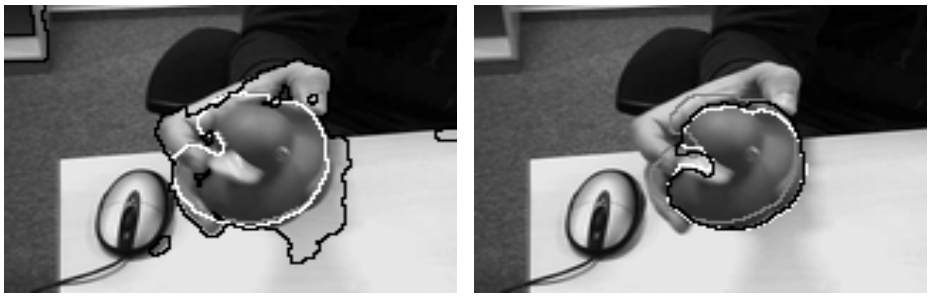

Fig. 2. Detailed view of the real-world image sequence with high dynamic motion, overlaid with segmentation results: previous segmentation result χ_2^{t-1} (grey), current segmentation initial condition $\hat{\phi}^t$ (white) and current segmentation result χ_2^t (black). Shown are identical frames left and right, but different prediction approaches: first order prediction based on the last two segmentation results χ_2^{t-2} and χ_2^{t-1} (5), yielding a segmentation initial that leads to unrobust tracking, as the segmentation algorithm is stuck in a local minimum (left) and probabilistic optical flow based measurement (9), being able to track the high dynamically moving object (right).

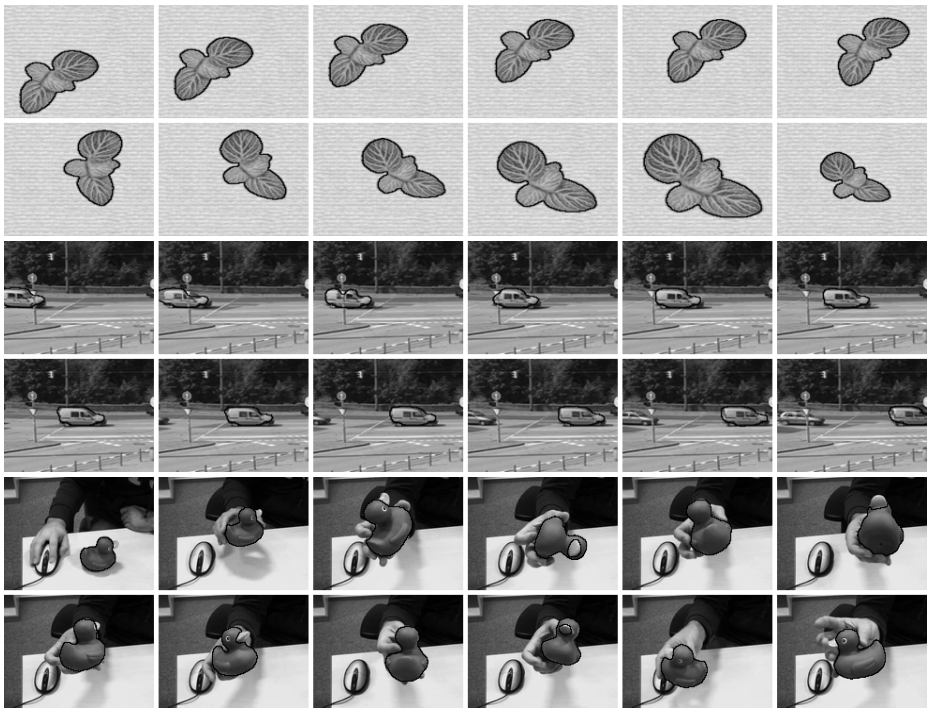

Fig. 3. Overview of used test sequences. Artificial test sequence (top), outdoor car sequence (middle), indoor sequence with high dynamic motion (bottom). The images are overlaid with the segmentation result χ_2^t (black). Shown are the first (upper left), last (lower right) and ten intermidiate frames of the sequence.

5 Conclusions

We presented an approach for object contour tracking, based on a level-set method for image segmentation and a correlation-based patch-matching method for probabilistic optical flow estimation. Utilising the probabilistic optical flow for the prediction of the object contour constitutes a non-parametric prediction model that is capable of representing nonrigid object deformation as well as complex and rapid object movements, thus providing the segmentation method with a reliable initial contour that leads to a robust and quick convergence of the level-set method.

Furthermore we introduced a novel interpretation of the value of the level-set function. Unlike most recent level-set methods that consider exclusively the sign of the level-set function to determine an object and its surroundings, we use the value of the level-set function to reflect the confidence in the predicted initial contour. This yields a robust and quick convergence of the level-set method in those sections of the contour with a high initial confidence and a flexible, mostly unconstrained and quick convergence in those sections with a low initial confidence.

References

1. Osher, S., Sethian, J.A.: Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations. J. Cmpt. Phys. **79** (1988) 12–49
2. Mumford, D., Shah, J.: Optimal approximation by piecewise smooth functions and associated variational problems. Commun. Pure Appl. Math **42** (1989) 577–685
3. Zhu, S.C., Yuille, A.L.: Region competition: Unifying snakes, region growing, and bayes/MDL for multiband image segmentation. PAMI **18**(9) (1996) 884–900
4. Chan, T., Vese, L.: Active contours without edges. IEEE Trans. Image Process. **10**(2) (February 2001) 266–277
5. 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. (September 2002) 797–800
6. Grossberg, Stephen, Hong, Simon: A neural model of surface perception: Lightness, anchoring, and filling-in. Spatial Vision **19**(2-4) (2006) 263–321
7. Horn, B.K.P., Schunck, B.G.: Determining optical flow. Artif. Intell. **17**(1-3) (1981) 185–203
8. Beauchemin, S.S., Barron, J.L.: The computation of optical flow. ACM Comp. Surv. **27**(3) (1995) 433–467
9. Simoncelli, E.P., Adelson, E.H., Heeger, D.J.: Probability distributions of optical flow. In: Proc Conf on Computer Vision and Pattern Recognition, Mauii, Hawaii, IEEE Computer Vision and Pattern Recognition (CVPR) (1991) 310–315
10. Weiss, Y., Fleet, D.: Velocity likelihoods in biological and machine vision. In: Probabilistic Models of the Brain: Perception and Neural Function. (2002) 77–96
11. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. ACM Comput. Surv. **38**(4) (2006) 13
12. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International Journal for Computer Vision **1**(4) (January 1988) 321–331