A Probabilistic Prediction Method for Object Contour Tracking

Daniel Weiler, Irene Clemente, Volker Willert, Julian Eggert

2009

Preprint:

This is an accepted article published in Neural Network World. The final authenticated version is available online at: https://doi.org/[DOI not available]

A Probabilistic Prediction Method for Object Contour Tracking

Daniel Weiler[1], Irene Ayllón Clemente[2], Volker Willert[3], Julian Eggert[3]

[1] Darmstadt University of Technology, Darmstadt D-64283, Germany
[2] Research Institute for Cognition and Robotics, Bielefeld D-33615, Germany
[3] Honda Research Institute Europe GmbH, Offenbach D-63073, Germany

Abstract. We present an approach for probabilistic contour prediction within the framework of 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. Utilizing 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 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.

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", labeling the remainder without any further differentiation as background. In this paper we address the problem of figure-background segregation in a sequence of consecutive images (object contour tracking) based on optical flow estimations resulting in a probabilistic prediction method for the contour of an object.

The segmentation occurs by means of level-set methods [1–9], that separate all image pixels into two disjoint regions [1] by favoring 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. Minimizing 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 minimized using the level-set framework e.g. by [4] and [5]. In recent years level-set methods became a powerful tool for image segmentation. State-of-the-art level-set methods are able to work on arbitrary feature maps [6]. These feature maps may incorporate the three colour components of an image but might be extended by any other characteristic property of a region (e.g. texture and motion). Most level-set methods assume the feature maps to be independent and commonly utilize a feature vector $\mathbf{f}$ composed of three colour and three texture components to perform the segmentation [7]. In [9] an approach was introduced that yields competitive results by employing only the three colour components but considering these image features $\mathbf{f}$ as located in a common, multi-dimensional feature space comparable to the probabilistic colour distributions modeled by means of Gaussian Mixture Models in state-of-the-art figure-background segregation algorithms [10–12].

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 [13]. 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, that 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 [14, 15] and correlation-based patch-matching [16–20] algorithms. The former is based on the *gradient constraint equation* that utilizes spatiotemporal derivatives of the image intensity resulting in velocities smaller than one pixel per frame and a high frame rate of the camera. To the contrary, the latter uses similarity (e.g. normalized cross-correlation) or distance (e.g. sum of squared differences) measures between a small patch of an image and its shifted counterpart that leads to discrete velocities and comparatively high computational costs. In recent years several approaches to improve optical flow methods were introduced (e.g. coarse-to-fine strategies and resolution pyramids). Whereas both classes of optical flow estimation provide a confidence measure for the estimated velocity, only the second class of correlation-based patch-matching methods is able to deliver a probability distribution function within a given velocity space that exhibits the capability to reflect ambiguous and multimodal motion.

Object tracking, i.e. locating an object in a sequence of consecutive images, constitutes an elementary task in high level video analysis. In [21] a comprehensive survey of object tracking algorithms is given. Depending on the vision task, object tracking algorithms are based on several object representations (e.g. single point; rectangular, elliptical and part-based multiple patches; object contour and silhouette), object detection strategies (e.g. point detectors, background subtraction, image segmentation) and prediction methods for the object location (e.g. probabilistic or deterministic, parametric or non-parametric models). 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.

Here we propose an approach that combines level-set segmentation algorithms and optical flow estimation methods to form a tracking system. The paper is organized 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 probabilistic 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 finalizes the paper.

2 Level-Set Segmentation and Probabilistic 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 minimization of an energy functional [2, 3]. The solution to the PDE constitutes an initial value problem that is solved by gradient descent. Depending on the initialization of the problem (i.e. on the initial contour) the gradient descent will, in cases of reliable initialization, succeed in finding the global minimum of the energy functional or, in cases of unreliable initialization, fail in doing so and be stuck in a local minimum. Famous representatives of energy functionals for image segmentation problems are those by Mumford and Shah [2] and by Zhu and Yuille [3]. While the former work in its original version on grey value images (i.e. on scalar data), utilizes the mean grey value of a region as a simple region descriptor and were later extended to vector valued data [6] (e.g. colour images), the latter uses 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 [5] e.g. by Gaussian Mixture Models or Nonparametric Parzen Density Estimates [22]. Within a region the models of all features together add up to the region descriptor.

Similar to state-of-the-art figure-background segregation algorithms [10–12], level-set methods use a smoothness term to control the granularity of the segmentation. A common way is to penalize 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 minimized. In doing so, few large objects are favored over many small objects as well as smooth object boundaries over ragged object boundaries.

Compared to "active contours" (snakes) [23], 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 [5, 9, 24]. 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 $x \in \Omega_2$. Here we adopt the convention that Ω_1 indicates the

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

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]

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

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 favors 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 outer and inner regions Ω_1 and Ω_2, respectively, favoring a unique region assignment.

Minimization of this functional [1, 25, 26] with respect to the level-set function ϕ using gradient descent leads to

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

with the smeared-out delta-function $\delta(\phi)$ e.g.:

$$\delta(\phi) = \frac{1}{\pi} \cdot \frac{\tau}{\tau^2 + \phi^2} \ . \tag{3}$$

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 [9]. 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

[2] Note that ϕ, χ_i and p_i are functions over the image position x.

to different kinds of motion-specific correspondence problems (e.g. the aperture problem [15]) and noisy data the measurement is based on. Especially for the case of transparent moving objects that overlap or partly occlude each other several motion hypotheses are needed to fully describe the image movement within the overlapping regions.

As has been suggested and discussed by several authors [18–20], velocity probability density functions (pdf's) are well suited to handle several kinds of motion ambiguities. Following these ideas we [20] model the uncertainty of the optical flow $\mathbf{V}^t$ as follows:

$$P(\mathbf{V}^t|Y^t) = \prod_x P(\mathbf{v}_x^t|Y^t) \qquad \text{with} \quad Y^t = \{\mathbf{I}^t, \mathbf{I}^{t+1}\}\,, \tag{4}$$

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 [19, 20]. 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 Object 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 [3]. 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. This leads to simpler implementations of many level-set algorithms requiring the computation of the gradient and furthermore, the value of the level-set function corresponds to the distance from the contour.

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 and accuracy a maximum time-step value is required [1]. Furthermore, a closer look at Eq. 2 and Eq. 3 shows that the temporal derivation of the level-set function ϕ directly depends on the value of the level-set function via the smeared-out

[3] In [25] Fast Marching Methods that code the minimum time of arrival of the contour by the value of the Level Set Function are introduced

delta-function $\delta(\phi)$. Thus, sections of the contour[4] exhibiting large values of the level-set function in their neighborhood generally move slower than those with smaller values of the level-set function in their neighborhood. Following this 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 the standard level-set framework for image segmentation. Certainly the steering of the velocity of the contour might also be introduced by a local modulation parameter overlaid with the segmentation evolution, but this would cause higher computational effort and require changes in the segmentation algorithm.

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 minimization process one might also use a smoothed (using a filter K_σ) version of the level-set function:

$$\hat{\phi}^t = K_\sigma * \phi^{t-1} \tag{5}$$

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 initialization of the image segmentation algorithm on the basis of the measured transformation. A parametric approach, based on a similarity[5] transformation $\mathbf{A}$ [27], requires the estimation $\boldsymbol{F}$ of four parameters, namely the translation vector $\boldsymbol{t} = (t_x, t_y)^T$, the rotation angle ω and scale factor 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 center of gravity of the inside masks χ_2^{t-1} and χ_2^{t-2}, the rotation by the evaluation of the principal component[6] of the two masks and the scale by the square root of the mask area ratio.

$$\hat{\phi}^t = \mathbf{A}(\phi^{t-1}, \boldsymbol{s}^{t-1}) \qquad \text{with} \quad \boldsymbol{s}^{t-1} = \boldsymbol{F}(\chi_2^{t-1}, \chi_2^{t-2}) \tag{6}$$

[4] Note that the contour is defined as those locations where the level-set function equals zero.

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

[6] 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}.

In contrast to the previous 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 $\boldsymbol{R}$.

$$\hat{\phi}^t = \mathbf{A}(\phi^{t-1}, \boldsymbol{s}^t) \qquad \text{with} \quad \boldsymbol{s}^t = \boldsymbol{R}(\mathbf{V}^{t-1}, \chi_2^{t-1}) \tag{7}$$

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 initialization 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. Utilizing an image processing warp algorithm [27] $\mathbf{W}_v$ 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 $\mathbf{I}^t$.

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

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 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.

$$\hat{\phi}^t = \mathbf{W}_v(\phi^{t-1}, \mathbf{V}^{t-1}, \mathbf{C}) \tag{9}$$

Equation (8) and (9) require an image warping algorithm [27]. Backward warping yielded best results. For the backward warping, also the optical flow need to be estimated backward in time.

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 $\hat{\phi}^t$. Utilizing 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 $\hat{\phi}^t$.

$$\hat{\phi}^t = \mathbf{W}_p(\phi^{t-1}, P(\mathbf{V}^{t-1}|Y^{t-1})) \tag{10}$$

that leads to

$$\hat{\phi}^t(x) = \sum_{v_{x'}^t} P(v_{x'}^{t-1}|Y^{t-1}) \cdot \phi^{t-1}(x - \Delta t \cdot v_{x'}^{t-1}) \tag{11}$$

with the image position x and its neighborhood $x - \Delta t \cdot v_{x'}^{t-1}$ with the sampling time Δt between two consecutive images $\mathbf{I}^t$ and $\mathbf{I}^{t+1}$.

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 reduced and dominated by the prediction algorithm.

4 Main Results

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

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 result (bottom row, left). The only difference of the initial level-set functions

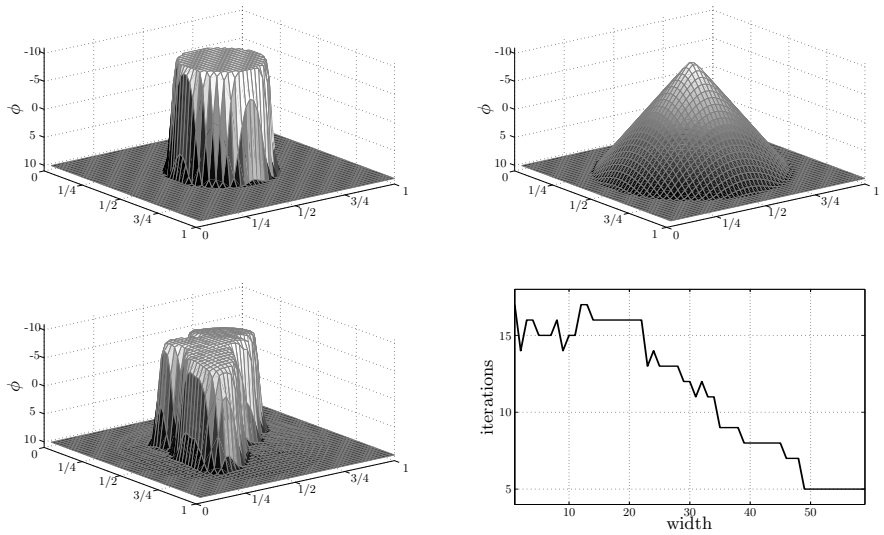

Fig. 1. Top row: Two different initial level-set functions ϕ for the same figure-background regions with steep (left) and flat (right) gradient. Bottom row: Final level-set function, after segmentation, that is independent of the chosen initial level-set function (left) and number of iterations until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

is 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.

4.2 Comparison of selected Algorithms

In order to obtain the predicted initial level-set function, different prediction methods have been implemented. An evaluation of the different algorithms was realized taking into account how they improved the computational cost of the segmentation process and in the case of affine movement models, how accurate they estimate the affine parameters. The algorithms are divided into two groups. The algorithms in the first group are based on affine prediction models, whereas the algorithms in the second group are based on non-parametric prediction models. For demonstrating the performance of the proposed algorithms four 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,

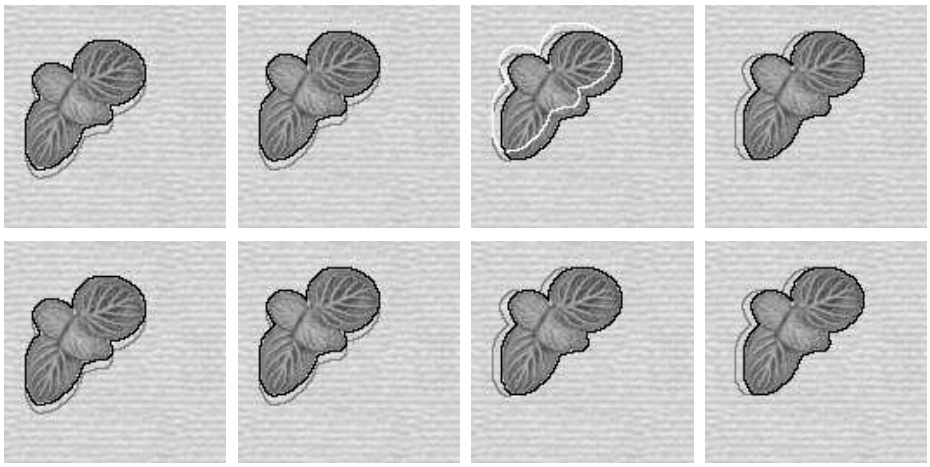

Fig. 2. Part of an artificial test sequence, overlaid with segmentation results: previous segmentation result (grey), current segmentation initial (white) and current segmentation result (black). Identical image sequences in top and bottom row, but different prediction approaches. Top row: First order prediction based on last two segmentation results (Eq. 6). Bottom row: Prediction based on optical flow measurement (Eq. 7).

three real world examples were chosen: one outdoor scene with a driving car and two indoor scenes with moving objects (a duck and a pencil case).

Affine Prediction Models

In this section, the two prediction methods from Sect. 3.2 are compared. Both are based on a similarity transformation, obtained by the estimation of the transformation of the previous level-set inside masks (Eq. 6) and the evaluation of the current optical flow (OF) (Eq. 7), repsectively. Additionally, a state-of-the-art tracking algorithm, the Kanade Lucas Tomasi (KLT) tracker (see [28] for our specific implementation), is considered in the following comparison. The aim of the last algorithm, proposed originally in 1981 by Kanade and Lucas, is to align a template image to an input image constrained to an affine transformation. The sequences used here only include affine transformations.

Figure 2 shows a part of the artificial test image sequence, overlaid with segmentation results: previous segmentation result (grey), current segmentation initial conditions (white) and current segmentation result (black). The image sequences in the top and bottom row are identical, but different prediction approaches were used. In Fig. 2 (top row) a first order prediction based on the last two segmentation results (Eq. 6) was used, whereas in Fig. 2 (bottom row) the prediction is based on the optical flow measurement (Eq. 7). The leaf in the image sequence moves in the first two images from bottom to top and in the last two images from left to right. While the first order prediction is not able to "foresee" the change in the movement of the object, the optical flow based

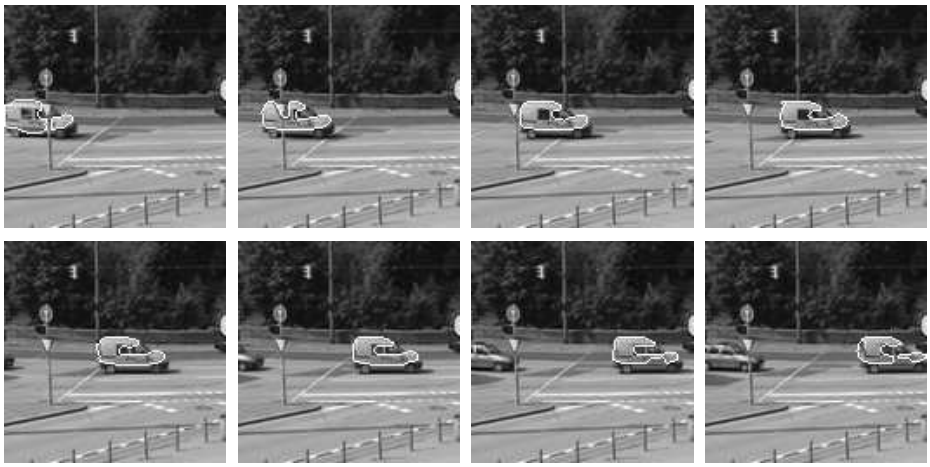

Fig. 3. Real-world sequence of a car driving along the street. At the beginning of the sequence, the car is occluded by a traffic signal. This signal is not segmented with the car. For this sequence we get a better performance and clear segmentation of the car by using the optical flow based prediction model (Eq. 7) (see table 1, last column).

method succeeds in this example. The reason is that the top-row prediction is based on the last two contours, i.e., it is assumed that the past transformation continues in the current frame.

Table 1 shows the Mean Square Error (MSE) values for different affine tracking algorithms in two sequences. The first sequence is the artificial sequence from Fig. 2 and the second sequence is a real world sequence represented by the car of the Fig. 3. The ground truth data is known, for the car sequence the car is translating three pixels to the right each frame. All three affine parameter estimation based prediction models adapt perfectly to the artificial sequence of the leaf (Fig. 2), providing MSE values near to zero. However, in the case of real-world sequences with partial occlusion as in the car sequence, the affine parameterized optical flow (Eq. 7) exhibits advantages. The reason is that we used a special spatiotemporally integrating optical flow estimation method that favors the estimation of movement patterns that exhibit a coherent translation [20].

Table 1. Comparison of different MSE values (pixels for translation, degrees for rotation) of the similarity transformation based prediction and the KLT tracking algorithm.

MSE	Leaf			Car		
	Masks (Eq. 6)	KLT	OF (Eq. 7)	Masks (Eq. 6)	KLT	OF (Eq. 7)
translation t_x	0.222	0.1874	0.0316	0.8276	0.513	0.6256
translation t_y	0.278	0.3676	0.0823	0.1373	0.008	0.0459
scale s	~ 0	0.012	0.001	0.011	~ 0	~ 0
rotation ω	0.0013	0.6912	2.5295	18.3521	24.49	0

Fig. 4. Comparison by means of the artificial leaf sequence between the level set predicted by the parametric model from Eq. 7 (left) and the deterministic (Eq. 8, middle) and probabilistic (Eq. 10, right) non-parametric models. Fig. 2 shows the original sequence.

Non-parametric Prediction Models

Three optical flow based prediction methods (Eq. 7, 8 and 10) were introduced in the previous Sect. 3.2 and 3.3. Whereas Eq. 7 constitutes a parametric method, Eq. 8 and Eq. 10 are a non-parametric deterministic and a non-parametric probabilistic method, respectively. In the following, we concentrate on these non-parametric motion models for level set prediction.

The advantage of the non-parametric models is that non-affine transformations and deformations of the object of interest can be tracked reliably. The disadvantage is that due to the fact of employing patches, that are sets of close pixels with a common translation (coherency constraint), the displacement of some of them is not correctly estimated and then, the warped images are blurred in contrast to those obtained by an affine model. This can be seen in Fig. 4, comparing the left with the other two prediction models (middle, right).

The difference between the two non-parametric models is that the probabilistic model takes into consideration all the possible movement hypotheses and hence, deals with ambiguous motion, leading to a more robust performance as shown in Fig. 6. In the shown example, the moving pencil case is not lost in any moment even though it undergoes arbitrary movements and deformations that could not have been covered by an affine model restricted to translation, rotation, shear and change of scale.

In Fig. 5 one frame of a 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 5 shows the same frame, processed with two different prediction approaches. Whereas Fig. 5 (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} (Eq. 6). This yields an initial condition for the segmentation that leads to unreliable tracking. The reason is that the segmentation algorithm gets stuck in an unfavorable local minimum as the initial

condition is already too far away from the desired final contour because it covers some parts of the fingers. Figure 5 (right) shows an approach based on the probabilistic optical flow measurement (Eq. 10) that is able to track the moving object reliably.

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. Using a 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. Our results suggest that this method is the most promising for tracking applications working on real-world sequences.

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 that have a low initial confidence.

Our results suggest that the non-parametric probabilistic method is the most promising for real-world sequences.

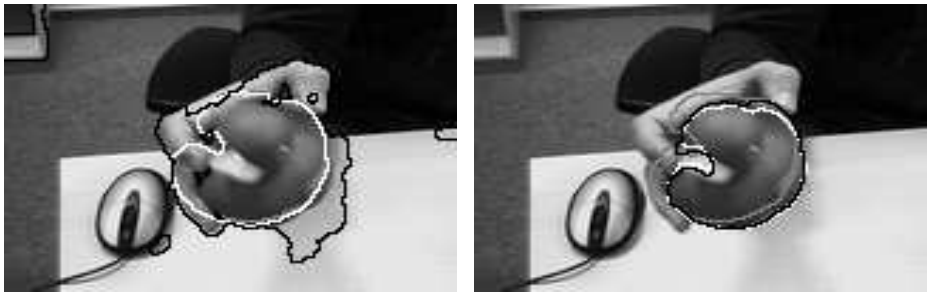

Fig. 5. Detailed view of a real-world image sequence with high dynamic motion, overlaid with segmentation results: previous segmentation result χ_2^{t-1} (grey), current segmentation initial condition $\hat{\phi}^t$ (white) and current segmentation result χ_2^t (black). Shown are identical frames left 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} (Eq. 6), 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 (Eq. 10), being able to track the high dynamically moving object (right).

Fig. 6. In this case, the real world sequence is a moving pencil case. The pencil case is a non-rigid body, that translates, rotates and changes its scale but nevertheless is successfully tracked by the algorithm. This sequence was evaluated using the non-parameterized probabilistic model from Eq. 8.

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. 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 (December 2002)
7. 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
8. Paragios, N., Faugeras, O.D., Chan, T., Schnörr, C., eds.: Variational, Geometric, and Level Set Methods in Computer Vision, Third International Workshop, VLSM 2005, Beijing, China, October 16, 2005, Proceedings. In Paragios, N., Faugeras, O.D., Chan, T., Schnörr, C., eds.: VLSM. Volume 3752 of Lecture Notes in Computer Science., Springer (2005)
9. Weiler, D., Eggert, J.: Multi-dimensional histogram-based image segmentation. In: International Conference on Neural Information Processing (ICONIP), Kitakyushu, Japan (2007)
10. 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
11. 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
12. Rother, C., Kolmogorov, V., Blake, A.: "GrabCut": Interactive foreground extraction using iterated graph cuts. ACM Trans. Graph. **23**(3) (2004) 309–314
13. Grossberg, Stephen, Hong, Simon: A neural model of surface perception: Lightness, anchoring, and filling-in. Spatial Vision **19**(2-4) (2006) 263–321
14. Horn, B.K.P., Schunck, B.G.: Determining optical flow. Artif. Intell. **17**(1-3) (1981) 185–203
15. Beauchemin, S.S., Barron, J.L.: The computation of optical flow. ACM Comp. Surv. **27**(3) (1995) 433–467
16. Singh, A.: An estimation-theoretic framework for image-flow computation. In: 3rd IEEE ICCV. (1990) 168–177
17. Wu, Q.X.: A correlation-relaxation-labeling framework for computing optical flow - template matching from a new perspective. IEEE Trans. Pattern Anal. Mach. Intell. **17**(9) (September 1995) 843–853
18. 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
19. Weiss, Y., Fleet, D.: Velocity likelihoods in biological and machine vision. In: Probabilistic Models of the Brain: Perception and Neural Function. (2002) 77–96
20. Willert, V., Eggert, J., Adamy, J., Körner, E.: Non-gaussian velocity distributions integrated over space, time, and scales. In: IEEE Trans. Syst., Man, Cybern. B. Volume 36. (June 2006) 482–493
21. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. ACM Comput. Surv. **38**(4) (2006) 13
22. Parzen, E.: On the estimation of a probability density function and mode. Annals of Mathematical Statistics **33** (1962) 1065–1076

23. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International Journal for Computer Vision **1**(4) (January 1988) 321–331
24. Chan, T., Sandberg, B., Vese, L.: Active contours without edges for vector-valued images. J. Visual Communication Image Representation **11**(2) (2000) 130–141
25. Sethian, J.: 8. Applied Mathematical Sciences Vol.153. In: Level Set Methods and Fast Marching Methods. Cambridge University Press (1999) 378
26. Osher, S., Fedkiw, R.: 2 & 7. Applied Mathematical Sciences Vol.153. In: Level Set Methods and Dynamic Implicit Surfaces. Springer, Berlin (2002) 295
27. Jähne, B.: Digital Image Processing. 6 edn. Springer, New-York (2005)
28. Eggert, J., Zhang, C., Körner, E.: Template matching for large transformations. In: Artificial Neural Networks, 17. International Conference (ICANN), Springer Verlag (2007) 169–179