Object Tracking and Segmentation in a Closed Loop

Konstantinos E. Papoutsakis and Antonis A. Argyros

Institute of Computer Science, FORTH
and
Computer Science Department, University of Crete
{papoutsa,argyros}@ics.forth.gr
http://www.ics.forth.gr/cvrl/

Abstract. We introduce a new method for integrated tracking and segmentation of a single non-rigid object in an monocular video, captured by a possibly moving camera. A closed-loop interaction between EM-like color-histogram-based tracking and Random Walker-based image segmentation is proposed, which results in reduced tracking drifts and in fine object segmentation. More specifically, pixel-wise spatial and color image cues are fused using Bayesian inference to guide object segmentation. The spatial properties and the appearance of the segmented objects are exploited to initialize the tracking algorithm in the next step, closing the loop between tracking and segmentation. As confirmed by experimental results on a variety of image sequences, the proposed approach efficiently tracks and segments previously unseen objects of varying appearance and shape, under challenging environmental conditions.

1 Introduction

The vision-based tracking and the segmentation of an object of interest in an image sequence are two challenging computer vision problems. Each of them has its own importance and challenges and can be considered as “chicken-and-egg” problems. By solving the segmentation problem, a solution to the tracking problem can easily be obtained. At the same time, tracking provides important input to segmentation.

In a recent and thorough review on the state-of-the-art tracking techniques [1], tracking methods are divided into three categories: *point tracking*, *silhouette tracking* and *kernel tracking*. Silhouette-based tracking methods usually evolve an initial contour to its new position in the current frame. This can be done using a state space model [2] defined in terms of shape and motion parameters [3] of the contour or by the minimization of a contour-based energy function [4, 5], providing an accurate representation of the tracked object. Point-tracking algorithms [6, 7] can also combine tracking and fine object segmentation using multiple image cues. Towards a more reliable and drift-free tracking, some point tracking algorithms utilize energy minimization techniques, such as Graph-Cuts or Belief Propagation on a Markov Random Field (MRF) [8] or on a Conditional

G. Bebis et al. (Eds.): ISVC 2010, Part I, LNCS 6453, pp. 405–416, 2010.

Random Field (CRF) [9, 10]. Most of the kernel-based tracking algorithms [11–13] provide a coarse representation of the tracked object based on a bounding box or an ellipsoid region.

Despite the many important research efforts devoted to the problem, the development of algorithms for tracking objects in unconstrained videos constitutes an open research problem. Moving cameras, appearance and shape variability of the tracked objects, varying illumination conditions and cluttered backgrounds constitute some of the challenges that a robust tracking algorithm needs to cope with. To this end, in this work we consider the combined tracking and segmentation of previously unseen objects in monocular videos captured by a possibly moving camera. No strong constraints are imposed regarding the appearance and the texture of the target object or the rigidity of its shape. All of the above may dynamically vary over time under challenging illumination conditions and changing background appearance. The basic aim of this work is to preclude tracking failures by enhancing its target localization performance through fine object segmentation that is appropriately integrated with tracking in a closed-loop algorithmic scheme. A kernel-based tracking algorithm [14], a natural extension of the popular mean-shift tracker [11, 15], is efficiently combined with Random Walker-based image segmentation [16, 17]. Explicit segmentation of the target region of interest in an image sequence enables reliable tracking and reduces drifting by exploiting static image cues and temporal coherence.

The key benefits of the proposed method are (i) close-loop interaction between tracking and segmentation (ii) enhanced tracking performance under challenging conditions (iii) fine object segmentation (iv) capability to track objects from a moving camera (v) increased tolerance to extensive changes of object's appearance and shape and, (vi) continual refinement of both the object and the background appearance models.

The rest of the paper is organized as follows. The proposed method is presented in Sec. 2. Experimental results are presented in Sec. 3. Finally, Sec. 4 summarizes the main conclusions from this work and future work perspectives.

2 Proposed Method

For each input video frame, the proposed framework encompasses a number of algorithmic steps, tightly interconnected in a closed-loop which is illustrated schematically in Fig.1. To further ease understanding, Fig.2 provides sample intermediate results of the most important algorithmic steps.

The method assumes that at a certain moment t in time, a new image frame I_t becomes available and that a fine object segmentation mask M_{t-1} is available as the result of the previous time step $t-1$. For time $t=0$, M_{t-1} should be provided for initialization purposes. Essentially, M_{t-1} is a binary image where foreground/background pixels have a value of 1/0, respectively (see Fig.2). The goal of the method is to produce the current object segmentation mask M_t. Towards this end, the spatial mean and covariance matrix of the foreground region of M_{t-1} is computed, thus defining an ellipsoid region coarsely representing

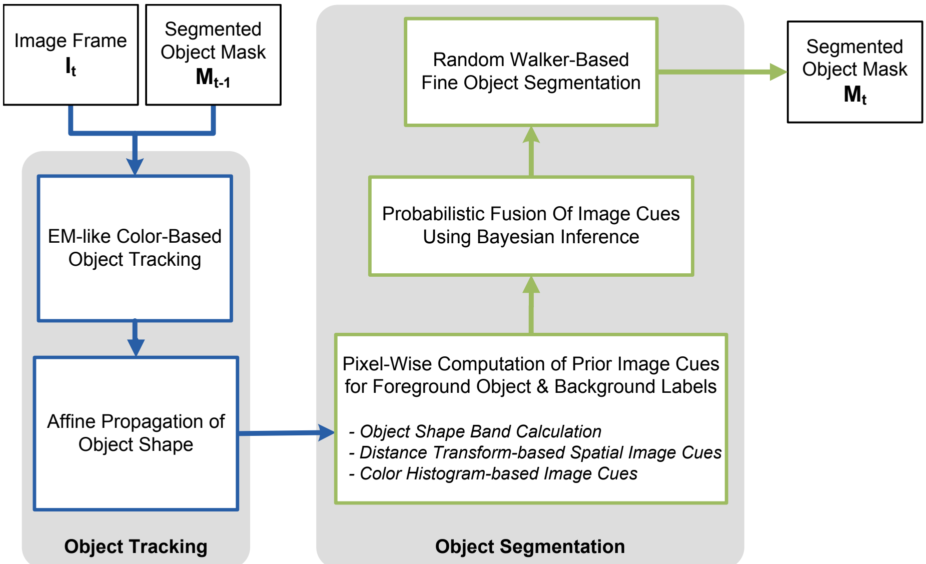

Fig. 1. Outline of the proposed method

the object at $t-1$. Additionally, a color-histogram-based appearance model of the segmented object (i.e., the one corresponding to the foreground of M_{t-1}) is computed using a Gaussian weighting kernel function. The iterative (EM-like) tracking algorithm in [14] is initialized based on the computed ellipsoid and appearance models. The tracking thus performed, results in a prediction of the position and covariance of the ellipsoid representing the tracked object. Based on the transformation parameters of the ellipsoid between $t-1$ and t, a 2D spatial affine transformation of the foreground object mask M_{t-1} is performed. The propagated object mask $M_t^{'}$ indicates the predicted position and shape of the object in the new frame I_t. The Hausdorff distance between the contour points of M_{t-1} and $M_t^{'}$ masks is then computed and a *shape band*, as in [4, 9], around the $M_t^{'}$ contour points is determined, denoted as B_t. The width of B_t is equal to the computed Hausdorff distance of the two contours. This is performed to guarantee that the shape band contains the actual contour pixels of the tracked object in the new frame. Additionally, the pixel-wise Distance Transform likelihoods for the object and background areas are computed together with the pixel-wise color likelihoods based on region-based color histograms. Pixel-wise Bayesian inference is applied to fuse spatial and color image cues, in order to compute the probability distribution for the object and the background regions. Given the estimated Probability Density Functions (PDFs) for each region, a Random Walker-based segmentation algorithm is finally employed to obtain M_t in I_t.

In the following sections, the components of the proposed method are described in more detail.

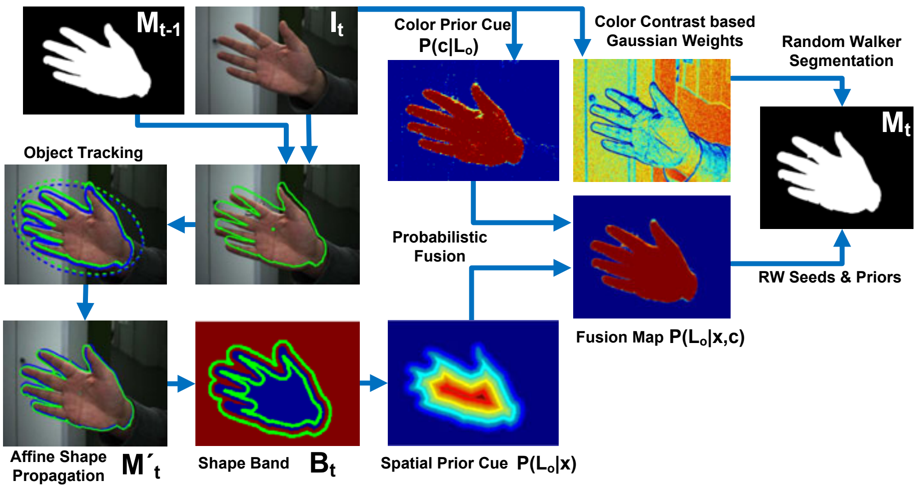

Fig. 2. Sample intermediate results of the proposed method. To avoid clutter, results related to the processing of the scene background are omitted.

2.1 Object Tracking

This section presents the tracking part of the proposed combined tracking and segmentation method (see the bottom-left part of Fig.1).

EM-Like Color Based Object Tracking: The tracking method [14] used in this work is closely related to the widely-used mean-shift tracking method [11, 15]. More specifically, this algorithm coarsely represents the objects' shape by a 2D ellipsoid region, modeled by its center $\boldsymbol{\theta}$ and covariance matrix Σ. The appearance model of the tracked object is represented by the color histogram of the image pixels under the 2D ellipsoid region corresponding to $\boldsymbol{\theta}$ and Σ, and is computed using a Gaussian weighting kernel function. Provided M_{t-1} and I_{t-1}, $\boldsymbol{\theta}_{t-1}$, Σ_{t-1} the object appearance model can be computed for time $t-1$. Given a new image frame I_t where the tracked object is to be localized, the tracking algorithm evolves the ellipsoid region in order to determine the image area in I_t that best matches the appearance of the tracked object in terms of a Bhattacharrya coefficient-based color similarity measure. This gives rise to the parameters $\boldsymbol{\theta_t}$ and Σ_t that represent the predicted object position and covariance in I_t.

Affine Propagation of Object Shape: The tracking algorithm presented above assumes that the shape of an object can be accurately represented as an ellipse. In the general case, this is a quite limiting assumption, therefore the objects' appearance model is forced to include background pixels, causing tracking to drift. The goal of this work is to prevent tracking drifts by integrating tracking with fine object segmentation.

To accomplish that, the contour C_{t-1} of the object mask in M_{t-1} is propagated to the current frame I_t based on the transformation suggested by the parameters $\boldsymbol{\theta_{t-1}}$, $\boldsymbol{\theta_t}$, Σ_{t-1} and Σ_t. A 2D spatial, affine transformation is defined between the corresponding ellipses. Exploiting the obtained Σ_{t-1} and Σ_t covariance matrices, a linear 2×2 affine transformation matrix A_t can be computed based on the square root ($\Sigma^{1/2}$) of each of these matrices. It is known that a covariance matrix is a square, symmetric and positive semidefinite matrix. The square root of any 2×2 covariance matrix Σ can be calculated by diagonalization as

$$\Sigma^{1/2} = Q\Lambda^{1/2}Q^{-1}, \tag{1}$$

where Q is the square 2×2 matrix whose i^{th} column is the eigenvector q_i of Σ and $\Lambda^{1/2}$ is the diagonal matrix whose diagonal elements are the square values of the corresponding eigenvalues. Since Σ is a covariance matrix, the inverse of its Q matrix is equal to the transposed matrix Q^T, therefore $\Sigma^{1/2} = Q\Lambda^{1/2}Q^T$. Accordingly, we compute the transformation matrix A_t by:

$$A_t = Q_t\Lambda_t^{1/2}\Lambda_{t-1}^{-1/2}Q_{t-1}^T. \tag{2}$$

Finally, C_t' is derived from C_t based on the following transformation

$$C_t' = A_t(C_t - \boldsymbol{\theta}_{t-1}) + \boldsymbol{\theta}_t. \tag{3}$$

The result indicates a propagated contour C_t', practically a propagated object mask $M_t^{'}$ that serves as a prediction of the position and the shape of the tracked object in the new frame I_t.

2.2 Object Segmentation

This section presents how the pixel-wise posterior values on spatial and color image cues are computed and fused using Bayesian inference in order to guide the segmentation of the tracked foreground object (see the right part of Fig.1).

Object Shape Band: An object shape band B_t is determined around the predicted object contour C_t'. Our notion of shape band is similar to those used in [4, 9]. B_t can be regarded as an area of uncertainty, where the true object contour lies in image I_t. The width of B_t is determined by the Euclidean, 2D Hausdorff distance between contours C_{t-1} and $C_t^{'}$ regarded as two point sets.

Spatial Image Cues: We use the Euclidean 2D Distance Transform to compute the probability of a pixel $\boldsymbol{x_i}$ in image I_t to belong to either the object L_o or the background L_b region, based on its 2D location $\boldsymbol{x_i} = (x, y)$ on the image plane. As a first step, the shape band B_t of the propagated object contour C_t' is considered and its inner and outer contours are extracted. The Distance Transform is then computed starting from the outer contour of B_t towards the inner part of the object. The probability $P(L_o|x_i)$ of a pixel to belong to the object given its image location is set proportional to its normalized distance from the outer contour of the shape band. For pixels that lie outside the outer contour of B_t, it holds that $P(L_o|x_i) = \epsilon$, where ϵ is a small constant.

Similarly, we compute the Euclidean Distance Transform measure starting from the inner contour of B_t towards the exterior part of the object. The probability $P(L_b|x_i)$ of a pixel to belong to the background given its image location is set proportional to its normalized distance from the inner contour of the shape band. For pixels that lie inside the inner contour of B_t, it holds that $P(L_b|x_i) = \epsilon$.

Color Based Image Cues: Based on the segmentation M_{t-1} of the image frame I_{t-1}, we define a partition of image pixels Ω into sets Ω_o and Ω_b indicating the object and background image pixels, respectively. The appearance model of the tracked object is represented by the color histogram H_o computed on the Ω_o set of pixels. The normalized value in a histogram bin c encodes the conditional probability $P(c|L_o)$. Similarly, the appearance model of the background region is represented by the color histogram H_b, computed over pixels in Ω_b and encoding the conditional probability $P(c|L_b)$.

Probabilistic Fusion of Image Cues: Image segmentation can be considered as a pixel-wise classification problem for a number of classes/labels. Our goal is to generate the posterior probability distribution for each of the labels L_o and L_b, which will be further utilized to guide the Random Walker-based image segmentation. Using Bayesian inference, we formulate a probabilistic framework to efficiently fuse the available prior image cues, based on the pixel color and position information, as described earlier. Considering the pixel color c as the evidence and conditioning on pixel position x_i in image frame I_t, the posterior probability distribution for label L_l is given by

$$P(L_l \mid c, x_i) = \frac{P(c \mid L_l, x_i)P(L_l \mid x_i)}{\sum_{l=0}^{N} P(c \mid L_l, x_i)P(L_l \mid x_i)}, \tag{4}$$

where $N = 2$ in our case. The probability distribution $P(c \mid L_l, x_i)$ encodes the conditional probability of color c taking the pixel label L_l as the evidence and conditioning on its location x_i. We assume that knowing the pixel position x_i, does not affect our belief about its color c. Thus, the probability of color c is only conditioned on the prior knowledge of its class L_l following that $P(c \mid L_l, x_i) = P(c \mid L_l)$. Given this, Eq.(4) transforms to

$$P(L_l \mid c, x_i) = \frac{P(c \mid L_l)P(L_l \mid x_i)}{\sum_{l=0}^{N} P(c \mid L_l)P(L_l \mid x_i)}. \tag{5}$$

The conditional color probability $P(c \mid L_l)$ for the class L_l is obtained by the color histogram H_l. The conditional spatial probability $P(L_l \mid x_i)$ is obtained by the Distance-Transform measure calculation. Both of these calculations have been presented earlier in this section.

Random Walker Based Object Fine Segmentation: The resulting posterior distribution $P(L_l \mid c, x_i)$ for each of the two labels L_o and L_b on pixels x_i guides the Random Walker-based image segmentation towards an explicit and accurate segmentation of the tracked object in I_t.

Random Walks for image segmentation was introduced in [18] as a method to perform K-way graph-based image segmentation given a number of pixels with user (or automatically) defined labels, indicating the K disjoint regions in a new image that is to be segmented. The principal idea behind the method is that one can analytically determine the real-valued probability that a random walker starting at each unlabeled image pixel will first reach one of the pre-labeled pixels. The random walker-based framework bears some resemblance to the popular graph-cuts framework for image segmentation, as they are both related to the spectral clustering family of algorithms [19], but they also exhibit significant differences concerning their properties, as described in [17].

The algorithm is formulated on a discrete weighted undirected graph $G = (V, E)$, where nodes $u \in V$ represent the image pixels and the positive-weighted edges $e \in E \subseteq V \times V$ indicate their local connectivity. The solution is calculated analytically by solving *K-1* sparse, symmetric, positive-definite linear systems of equations, for K labels. For each graph node, the resulting probabilities of the potential labels sum up to 1.

In order to represent the image structure by random walker biases, we map the edge weights to positive weighting scores computed by the Gaussian weighting function on the normalized Euclidean distance of the color intensities between two adjacent pixels, practically the color contrast. The Gaussian weighting function is

$$w_{i,j} = e^{-\frac{\beta}{\rho}(\|c_i - c_j\|)^2} + \epsilon, \tag{6}$$

where c_i stands for the vector containing the color channel values of pixel/node i, ϵ is a small constant (i.e $\epsilon = 10^{-6}$) and ρ is a normalizing scalar $\rho = \max(\|c_i - c_j\|), \forall i, j \in E$. The parameter β is user-defined and modulates the spatial random walker biases, in terms of image edgeness. The posterior probability distribution $P(L_l \mid c, x_i)$ computed over the pixels x_i of the current image I_t suggest the probability of the pixels to be assigned to the label L_l. Therefore, we consider the pixels of highest posterior probability values for the label L_l as pre-labeled/seeds nodes of that label in the formulated graph.

An alternative formulation of the Random Walker-based image segmentation method is presented in [16]. This method incorporates non-parametric probability models, that is, prior belief on label assignments. In [16], the sparse linear systems of equations that need to be solved to obtain a real-valued density-based multilabel image segmentation are also presented. The two modalities of this alternative formulation suggest for using only prior knowledge on the belief of a graph node toward each of the potential labels, or using prior knowledge in conjunction with pre-labeled/seed graph nodes. The γ scalar weight parameter is introduced in these formulations, controlling the degree of effectiveness of the prior belief values towards the belief information obtained by the random walks. This extended formulation of using both seeds and prior beliefs on graph nodes is compatible with our approach considering the obtained posterior probability distributions $P(L_l \mid c, x_i)$ for the two segmentation labels. The two Random Walker formulations that use prior models, suggest for a graph construction similar to the graph-cut algorithm [20], where the edge weights of the constructed graph can be seen as the *N-links* or *link-terms* and the prior belief values of the graph nodes for any of the potential labels can be considered as the *T-links* or the *data-terms*, in graph cuts terminology.

Regardless of the exact formulation used, the primary output of the algorithm consists of K probability maps, that is a soft image segmentation per label. By assigning each pixel to the label for which the greatest probability is calculated, a K-way segmentation is obtained. This process gives rise to object mask M_t for image frame I_t.

3 Experimental Results and Implementation Issues

The proposed method was extensively tested on a variety of image sequences. Due to space limitations, results on eight representative image sequences are presented in this paper. The objects tracked in these sequences go through extensive appearance, shape and pose changes. Additionally, these sequences differ with respect to the camera motion and to the lighting conditions during image acquisition which affects the appearance of the tracked objects.

We compare the proposed joint tracking and segmentation method with the tracking-only approach of [14]. The parameters of this algorithm were kept identical in the stand-alone run and in the run within the proposed framework. It is important to note that stand-alone tracking based on [14] is initialized with the appearance model extracted in the first frame of the sequence and that this appearance model is not updated over time. This is done because in all the challenging sequences we used as the basis of our evaluation, updating the appearance model based on the results of tracking, soon causes tracking drifts and total loss of the tracked object.

Figure 3 illustrates representative tracking results (i.e., five frames for each of the eight sequences). In the first sequence, a human hand undergoes complex articulations, whereas the lighting conditions significantly affect its skin color tone. In the second sequence, a human head is tracked despite its abrupt scale changes and the lighting variations. In the third sequence the articulations of a human hand are observed by a moving camera in the context of a continuously varying cluttered background. The green book tracked in the fourth sequence undergoes significant changes regarding its pose and shape, whereas light reflections on its glossy surface significantly affect its appearance. The fifth sequence is an example of a low quality video captured by a moving camera, illustrating the inherently deformable body of a caterpillar in motion. The sixth and seventh sequences show a human head and hand, respectively, which both go through extended pose variations in front of a complex background. Finally, the last, low resolution sequence has been captured by a medical endoscope. In this sequence, a target object is successfully tracked within a low-contrast background.

Each of the image sequences in Fig.(3) illustrating human hands or faces as well as the green book sequence consists of 400 frames of resolution 640×480 pixels, captured at a frame rate of 5-10 fps. The resolution of each frame of the image sequences illustrated in the second and the fifth row is 320×240 pixels. The last image sequence depicted in the Fig.(3), captured by a medical endoscope consists of 20 image frames of size 256×256 pixels each.

The reported experiments were generated based on a Matlab implementation, running on a PC equipped with an Intel i7 CPU and 4 GB of RAM memory. The runtime performance of the current implementation varies between 4 to 6 seconds per frame for 640×480 images. A near real-time runtime performance is feasible by optimizing both the EM-like component of the tracking method and the solution of the large sparse linear system of equations of the Random Walker formulation in the segmentation procedure.

Each frame shown in Fig.(3) is annotated with the results of the proposed algorithm and the results of the tracking method proposed in [14]. More specifically, the blue solid ellipse shows the expected position and coarse orientation of the tracked object as this results from the tracking part of the proposed methodology. The green solid object contour is the main result of the proposed algorithm which shows the fine object segmentation. Finally, the result of [14] is shown for comparison as a red dotted ellipse. Experimental results on the full video datasets are available online[1].

In all sequences, the appearance models of the tracked objects have been built based on the RGB color space. The object and background appearance models used to compute the prior color cues are color histograms with 32 bins per histogram for both the object and the background. Preserving the parameter configuration of the object tracking algorithm as described in [14], the target appearance model of the tracker is implemented by a color histogram of 8 bins per dimension.

The Random Walker segmentation method involves three different formulations to obtain the probabilities of each pixel to belong to each of the labels of the segmentation problem, as described in Sec. 2.2. The three options refer to the usage of seed pixels (pre-labeled graph nodes), prior values (probabilities/beliefs on label assignments for some graph nodes), or a combination of them. For the last option, the edge weights of the graph are computed by the Eq.(6), where the β scalar parameter controls the scale of the edgeness (color contrast) between adjacent graph nodes. The pixel-wise posterior values are computed using Bayesian inference as described in Sec. 2.2 and are exploited to guide segmentation as seed and prior values in terms of Random Walker terminology. Each pixel x_i of posterior value $P(L_l \mid x_i)$ greater or equal to 0.9 is considered as a seed pixel for the label L_l, thus as a seed node on the graph G. Any other pixel of posterior value $P(L_l \mid x_i)$ less than 0.9 is considered as a prior value for label L_l. In the case of prior values, the γ parameter is introduced to adjust the degree of authority of the prior beliefs towards the definite label-assignments expressed by

[1] `http://www.ics.forth.gr/~argyros/research/trackingsegmentation.htm`

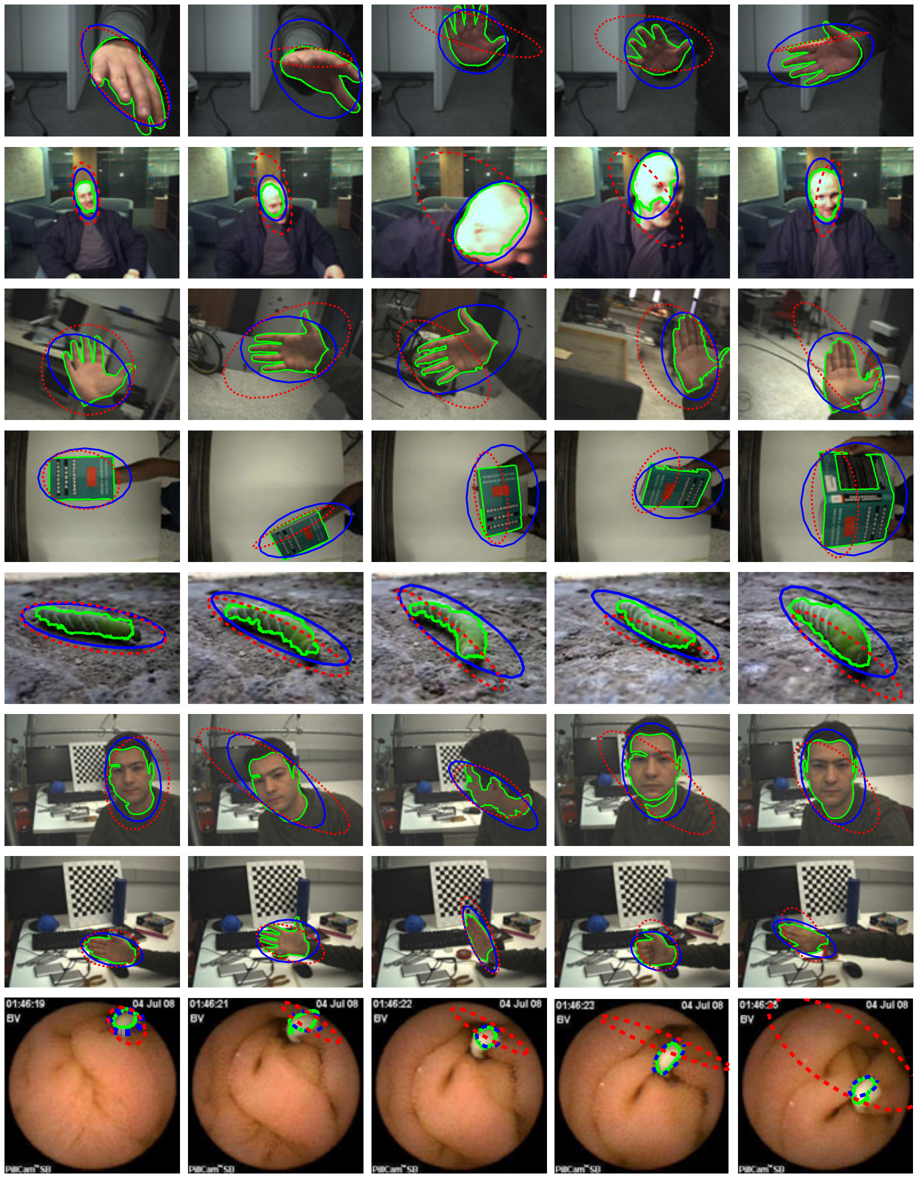

Fig. 3. Experimental results and qualitative comparison between the proposed framework providing tracking and segmentation results (blue solid ellipse and green solid object contour, respectively) and the tracking algorithm of [14] (red dotted ellipse). See text for details.

Table 1. Quantitative assessment of segmentation accuracy. See text for details.

Segmentation option	Precision	Recall	F-measure
Priors	93,5%	92,9%	93,1%
Seeds	97,5%	99,1%	98,3%
Priors and Seeds	97,5%	99,1%	98,3%

the seed nodes of the graph. In our experiments, the β parameter was selected within the interval of $[10 - 50]$, whereas the γ ranges within $[0.05 - 0.5]$.

In order to assess quantitatively the influence of the three different options regarding the operation of the Random Walker on the quality of segmentation results, the three different variants have been tested independently on an image sequence consisting of $1,000$ video frames. For each and every of these frames ground truth information is available in the form of a manually segmented foreground object mask. Table 1 summarizes the average (per frame) precision, recall and F-measure performance of the proposed algorithm compared to the ground truth. As it can be verified, although all three options perform satisfactorily, the use of seeds improves the segmentation performance.

4 Summary

In this paper we presented a method for online, joint tracking and segmentation of an non-rigid object in a monocular video, captured by a possibly moving camera. The proposed approach aspires to relax several limiting assumptions regarding the appearance and shape of the tracked object, the motion of the camera and the lighting conditions. The key contribution of the proposed framework is the efficient combination of an appearance-based tracking with Random Walker-based segmentation that jointly enables enhanced tracking performance and fine segmentation of the target object. A 2D affine transformation is computed to propagate the segmented object shape of the previous frame to the new frame exploiting the information provided by the ellipse region capturing the segmented object and the ellipse region predicted by the tracker in the new frame. A shape-band area is computed indicating an area of uncertainty where the true object boundaries lie in the new frame. Static image cues including pixel-wise color and spatial likelihoods are fused using Bayesian inference to guide the Random Walker-based object segmentation in conjunction with the color-contrast (edgeness) likelihoods between neighboring pixels. The performance of the proposed method is demonstrated in a series of challenging videos and in comparison with the results of the tracking method presented in [14].

Acknowledgments

This work was partially supported by the IST-FP7-IP-215821 project GRASP. The contributions of FORTH-ICS members I. Oikonomidis and N. Kyriazis to the development of the proposed method are gratefully acknowledged.

References

1. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. ACM Comput. Surv. 38, 13 (2006)
2. Isard, M., Blake, A.: Condensation: Conditional density propagation for visual tracking. International Journal of Computer Vision 29, 5–28 (1998)
3. Paragios, N., Deriche, R.: Geodesic active contours and level sets for the detection and tracking of moving objects. IEEE Transactions on PAMI 22, 266–280 (2000)
4. Yilmaz, A., Li, X., Shah, M.: Contour-based object tracking with occlusion handling in video acquired using mobile cameras. IEEE Transactions on PAMI 26, 1531–1536 (2004)
5. Bibby, C., Reid, I.: Robust real-time visual tracking using pixel-wise posteriors. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 831–844. Springer, Heidelberg (2008)
6. Khan, S., Shah, M.: Object based segmentation of video using color, motion and spatial information. In: IEEE Computer Society Conference on CVPR, vol. 2, p. 746 (2001)
7. Baltzakis, H., Argyros, A.A.: Propagation of pixel hypotheses for multiple objects tracking. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Kuno, Y., Wang, J., Pajarola, R., Lindstrom, P., Hinkenjann, A., Encarnação, M.L., Silva, C.T., Coming, D. (eds.) ISVC 2009. LNCS, vol. 5876, pp. 140–149. Springer, Heidelberg (2009)
8. Yu, T., Zhang, C., Cohen, M., Rui, Y., Wu, Y.: Monocular video foreground/background segmentation by tracking spatial-color gaussian mixture models. In: IEEE Workshop on Motion and Video Computing (2007)
9. Yin, Z., Collins, R.T.: Shape constrained figure-ground segmentation and tracking. In: IEEE Computer Society Conference on CVPR, pp. 731–738 (2009)
10. Ren, X., Malik, J.: Tracking as repeated figure/ground segmentation. In: IEEE Computer Society Conference on CVPR, pp. 1–8 (2007)
11. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-based object tracking. IEEE Transactions on PAMI 25, 564–577 (2003)
12. Tao, H., Sawhney, H., Kumar, R.: Object tracking with bayesian estimation of dynamic layer representations. IEEE Transactions on PAMI 24, 75–89 (2002)
13. Jepson, A.D., Fleet, D.J., El-Maraghi, T.F.: Robust online appearance models for visual tracking. IEEE Transactions on PAMI 25, 1296–1311 (2003)
14. Zivkovic, Z., Krose, B.: An em-like algorithm for color-histogram-based object tracking. In: IEEE Computer Society Conference on CVPR, vol. 1, pp. 798–803 (2004)
15. Comaniciu, D., Ramesh, V., Meer, P.: Real-time tracking of non-rigid objects using mean shift. In: IEEE Computer Society Conference on CVPR, vol. 2, p. 2142 (2000)
16. Grady, L.: Multilabel random walker image segmentation using prior models. In: Proceedings of the 2005 IEEE Computer Society Conference on CVPR, vol. 1, pp. 763–770 (2005)
17. Grady, L.: Random walks for image segmentation. IEEE Transactions on PAMI 28, 1768–1783 (2006)
18. Grady, L., Funka-Lea, G.: Multi-label image segmentation for medical applications based on graph-theoretic electrical potentials. In: Sonka, M., Kakadiaris, I.A., Kybic, J. (eds.) CVAMIA/MMBIA 2004. LNCS, vol. 3117, pp. 230–245. Springer, Heidelberg (2004)
19. von Luxburg, U.: A tutorial on spectral clustering. Statistics and Computing 17, 395–416 (2007)
20. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient n-d image segmentation. International Journal of Computer Vision 70, 109–131 (2006)