Contour-Based Object Tracking with Occlusion Handling in Video Acquired Using Mobile Cameras

Alper Yilmaz, *Member*, *IEEE*, Xin Li, and Mubarak Shah, *Fellow*, *IEEE*

Abstract—We propose a tracking method which tracks the complete object regions, adapts to changing visual features, and handles occlusions. Tracking is achieved by evolving the contour from frame to frame by minimizing some energy functional evaluated in the contour vicinity defined by a band. Our approach has two major components related to the visual features and the object shape. Visual features (color, texture) are modeled by semiparametric models and are fused using independent opinion polling. Shape priors consist of shape level sets and are used to recover the missing object regions during occlusion. We demonstrate the performance of our method on real sequences with and without object occlusions.

Index Terms—Contour tracking, shape priors, occlusion handling, level sets.

1 Introduction

THE most common approach to track objects is to first detect them using background subtraction and, then, establish correspondence from frame to frame to find the tracks of the objects [1], [2]. Despite its popularity, background subtraction can only be applied to imagery acquired by "stationary cameras" and it provides "coarse object silhouettes" which are not suitable for high level vision tasks, such as fine level action recognition, where detailed analysis of the shape deformation during an action is required. An alternative approach to background subtraction is to find the transformation of the object from frame to frame which is modeled using simple geometric models, e.g., ellipse or rectangle. In [3], Comaniciu et al. use the mean-shift approach to compute the translation of a circular region. They model the object appearance by weighed histograms. Similarly, Jepson et al. [4] compute the affine motion of the object using a probabilistic appearance model that captures the stable object features, object shape, and deals with outliers. Although good tracking performances are achieved, these trackers only track the centroid or the orientation of the object.

Tracking of the complete object can be achieved by employing the active contours, which were introduced by Kass et al. [5]. The objective of active contours is to get a tight contour enclosing the object by minimizing an energy functional:

$$E(\Gamma) = \int_0^1 E_{internal}(\mathbf{v}) + E_{image}(\mathbf{v}) ds,$$

where s is the arc length of contour Γ, E_{image} signifies the energy based on the image observations, and $E_{internal}$ prevents gaps and rapid bending. In practice, E_{image} is commonly defined in terms of the image gradient, ∇I, [5], [6], [7] [8]. Specifically, Caselles et al. [6] set E_{image} to $g(|\nabla I|)$, where g was a sigmoid function. Later, the same functional was adopted by Paragios and Deriche [7]. They initialized the object contour in every frame using the background subtraction. Tracking using image gradient is not suitable for textured images, where boundary between the object and background becomes ambiguous. To overcome this limitation, Zhu and Yuille [9] propose a region-based energy which is minimized using gradient descent. An offline merging step is applied to reduce oversegmentation. The same energy is reformulated using the region descriptors by Jehan-Besson and Barland [10]. In the same context, Paragios and Deriche [11] combine energy terms used in [6] and [9].

Rigid body motion models can also be coupled with the contour energy functionals. In [12], Yezzi et al. combine the Mumford-Shah distance with 2D transformation and compute the contour transformation between two different views. Similarly, Rittscher and Blake [13], use affine motion for contour tracking. In contrast to [12], they train the tracking algorithm with possible affine deformations. In general, rigid-body motion models are not suitable for tracking nonrigid objects, e.g., humans and animals.

Nonrigid object motion can be modeled in terms of optical flow (u, v), which is derived from the brightness constraint. In [8], Bertalmio et al. compute u and v iteratively using two energy functionals: one for contour evolution and the other for intensity morphing. At each iteration, the contour is evolved with the speed computed by projecting the temporal gradient onto the contour normal. Mansouri [14] uses a probabilistic form of the brightness constraint, where the color prior is defined by a Gaussian distribution. He computes optical flow by maximizing $P(I^{t+1}|I^t)$. Brightness constraint requires very small variation in the intensities and is not suitable for images with high dynamic intensity ranges.

In this paper, we present a Bayesian framework for contour tracking formulated as a variational calculus problem. Proposed contour energy functional contains two energy terms: the image energy E_{image} and the shape energy E_{shape}. Image energy, which is partly motivated by [9] (use of the conditional probabilities) and [14] (relation to the brightness constraint), is based on color and texture observations and is evaluated in a band around the contour. The shape energy is based on the past contour observations and preserves the shape of the object during partial and full occlusions. Recently, Cremers et al. [15] proposed a shape energy, which requires training by modeling a set of contours using principal component analysis. Here, we propose an online shape model which is learned from nonrigid contour deformations during the course of tracking. Tracking is achieved by evolving the contour, which is represented using level sets, by minimizing energy in the gradient descent direction.

The paper is organized as follows: In Section 2, we give details of the appearance models used, derive the proposed energy functional, discuss the contour representation and energy minimization, and propose an occlusion handling mechanism. Experimental results, a comparative discussion on the proposed method, and conclusions are sketched in Sections 3, 4, and 5, respectively.

2 Proposed Method

Object tracking can be treated as two-class discriminant analysis of pixels, where the classes correspond to the object, R_{obj}, and the background, R_{bck}, regions. The performance of discriminant analysis depends on the object features, energy functional, energy minimization technique, and the contour representation chosen. In

- *A. Yilmaz and M. Shah are with the School of Computer Science, University of Central Florida, 4000 Central Florida Blvd., Orlando, FL 32816. E-mail: {yilmaz, shah}@cs.ucf.edu.*
- *X. Li is with the Department of Mathematics, University of Central Florida, Physics and Mathematics Building, Rm. 235, Orlando, FL 32816. E-mail: xli@math.ucf.edu.*

Manuscript received 19 May 2003; revised 3 May 2004; accepted 13 May 2004.
Recommended for acceptance by C. Taylor.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0085-0503.

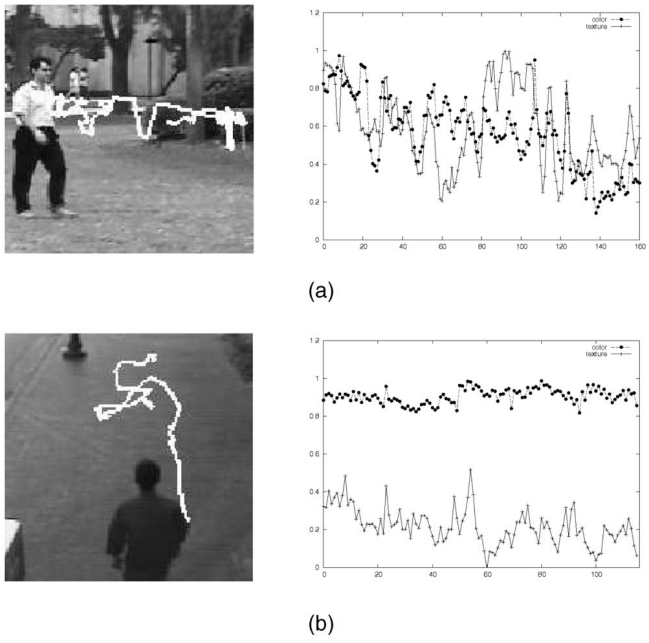

Fig. 1. Left image: trajectory of a contour point, right image: plots of weights as a function of frame number. (a) Both color and texture weights are high at different times and contribute to tracking. (b) The weight of color is higher than texture and contributes the most.

this section, we will address these issues and additionally propose a solution to handle occlusions.

2.1 Appearance Features

During the last two decades, two classes of features have been widely considered for tracking and segmentation purposes: color and texture. We believe an ideal tracking approach should use both of these features. This is evident from Fig. 1, where we show the importance of color and texture for two sequences along the trajectory of a contour point. In the first sequence (Fig. 1a, left), both features contribute to tracking (Fig. 1a, right), whereas in the second sequence (Fig. 1b, left), color contributes more than the texture (Fig. 1b, right).

Therefore, in our approach, we use both of these features. In particular, for color, since in our experiments, we achieved the same qualitative tracking results with RGB, HSV, and YIQ color spaces, we chose the RGB space. Color prior is defined by multivariate kernel density estimation using the Epanechnikov kernel, which is chosen for its property to provide minimum error between the data and its estimate. The texture features, which are obtained from the sub-bands in the steerable pyramid representation [16], are modeled using a mixture of two Gaussians.

Fusion of color and texture models produces a semiparametric statistical model. Using this model, pixels can be clustered as the object or the background by the "independent opinion polling" strategy, which evaluates pixel probabilities prior to membership assignment:

$$P(\alpha|\mathbf{x}) = \frac{\prod_\beta P_\beta(\mathbf{x}|R_\alpha)P(\alpha)}{\sum_\gamma \prod_\beta P_\beta(\mathbf{x}|R_\gamma)P(\gamma)},$$

where $\gamma, \alpha \in \{obj, bck\}$ and $\beta \in \{\text{color}, \{\text{steerable subbands}\}\}$. It can be observed that the discriminant features will be emphasized, otherwise, they will be suppressed (see Figs. 1a and 1b).

2.2 Tracking Energy Functional

Let the image be $I = R_{obj} \cup R_{bck}$. The likelihood of observing the boundary (contour), Γ, is equal to the likelihood of partitioning the space $P(\Gamma) = P(\varphi(I) = \{R_{obj}, R_{bck}\})$, where φ is the partitioning operator [9]. Thus, posteriori contour probability $P(\Gamma)$ can be used interchangeably with a posteriori partitioning probability. Constraining $P(\Gamma)$ with the current image, I^t, and the previous boundaries, $\Gamma^{\tilde{t}} = \Gamma^1 \dots \Gamma^{t-1}$, the tracking energy becomes $P_\Gamma = P\left(\varphi(R^t)|I^t, \Gamma^{\tilde{t}}\right)$. Using Bayes' rule, probability of the contour is approximated as:

$$P_\Gamma \approx \underbrace{P\left(I^t|R^t_{obj}, \Gamma^{\tilde{t}}\right)}_{\text{object}\to P_{R_{obj}}(I^t)} \underbrace{P\left(I^t|R^t_{bck}, \Gamma^{\tilde{t}}\right)}_{\text{background}\to P_{R_{bck}}(I^t)} \underbrace{P\left(\varphi(R^t)|\Gamma^{\tilde{t}}\right)}_{\text{shape}\to P^t_S}. \quad (1)$$

The first two terms in (1) can be computed using the observed features and the object and background priors. The last term in (1) represents the object shape learned over time.

Let there be two subregions R^Γ_{obj} and R^Γ_{bck} defined in the contour neighborhood, such that $R^\Gamma_{obj} \subset R^t_{obj}$ and $R^\Gamma_{bck} \subset R^t_{bck}$. Due to the artifacts (holes inside the object) and noise (quantization errors), contour probability in (1) can be defined in terms of the subregions R^Γ_{obj} and R^Γ_{bck} using $P_\Gamma \le P'_\Gamma = P_{R^\Gamma_{obj}}(I^t)P_{R^\Gamma_{bck}}(I^t)P^t_S$. In the remainder of the paper, we replace P_Γ with P'_Γ. The maximum a posteriori (MAP) estimate of the object contour in the tth frame, $\widehat{\Gamma^t}$, is found by maximizing the probability P'_Γ over the subsets $\Gamma \subset \Omega$, where Ω is the space of all object contours. The MAP estimate can be written in terms of the subregions:

$$\widehat{\Gamma^t} = \arg\max_{\Gamma\subset\Omega} \prod_{\mathbf{x}_1} \left[\prod_{\mathbf{x}_2} P_{R^\Gamma_{obj}}(I^t(x_2)) \prod_{\mathbf{x}_3} P_{R^\Gamma_{bck}}(I^t(\mathbf{x}_3)) P^t_S \right],$$

where $\mathbf{x}_1 \in \Gamma$, $\mathbf{x}_2 \in R_{obj}$ and $\mathbf{x}_3 \in R_{bck}$. Converting this to energy minimization by considering the negative log-likelihood of the probabilities, we have:

$$E = \int_{\mathbf{x}_1} \left[\underbrace{\iint_{\mathbf{x}_2} \Psi_{obj}(\mathbf{x}_2) d\mathbf{x}_2}_{E_A} + \underbrace{\iint_{\mathbf{x}_3} \Psi_{bck}(\mathbf{x}_3) d\mathbf{x}_3}_{E_B} - \log P^t_S \right] d\mathbf{x}_1, \quad (2)$$

where $\mathbf{x}_1 \in \Gamma$, $\mathbf{x}_2 \in R_{obj}(\mathbf{x}_1)$, $\mathbf{x}_3 \in R_{bck}(\mathbf{x}_1)$, and

$$\Psi_\alpha(\mathbf{x}) = -\log P_{R_\alpha}(I^t(\mathbf{x})) : \alpha \in \{obj, bck\}.$$

For the sake of implementation, we choose the subregion $R^\Gamma(\mathbf{x}_i)$ as a set of square regions of size $2m \times 2m$ centered on $\mathbf{x}_i$. Let s be the contour arc length. For each contour position $(f(s), g(s))$, the pixels inside the square region are defined using the parametric curve functions f and g by $x = \tilde{x} + f(s)$ and $y = \tilde{y} + g(s)$. The region membership using the new notation is defined through an indicator function, $1^\Gamma_\alpha(\mathbf{x}) = \frac{1}{1+\exp(-M(\mathbf{x}))}$, where M is the region mask. Changing the contour variables (x, y) to arc length s, we have the following energy: $E_A(s) = \iint_{-m}^{m} \Psi_A(x, y) 1^\Gamma_{obj}\{x, y\} J dx dy$, where Jacobian J is introduced due to the change of variables, and because of the translation $J = 1$. Once E_B is written similarly, (2) results in the following functional:

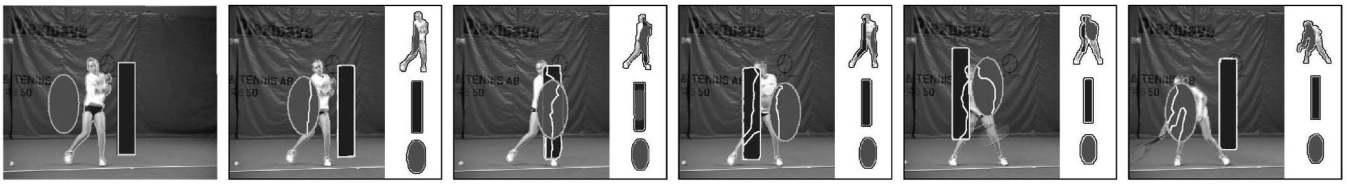

Fig. 2. A sequence with two synthetic objects occluding a player. Left image is the first image of the sequence. The white boxes next to each frame shows extracted objects. Note that multiple object occlusions in the second frame are correctly handled.

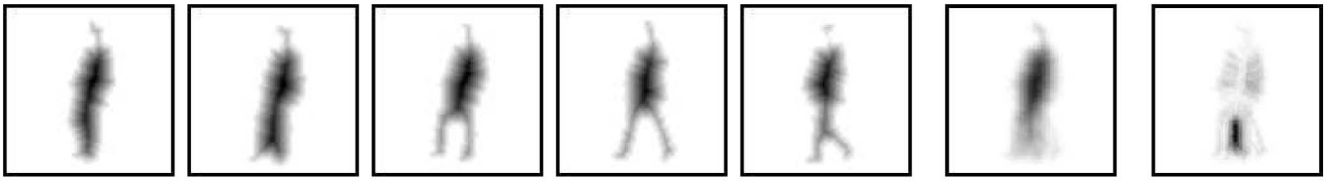

Fig. 3. (a) Shape level sets for the walking sequence. Shape model: (b) mean and (c) standard deviation.

$$E = \int_0^l \Bigg[-\overbrace{\iint_{-m}^{m} \log P_{R_{obj}}(I(\mathbf{x}))1^{\Gamma}_{obj} d\tilde{\mathbf{x}} d\tilde{\mathbf{y}}}^{\Phi_{obj} \Rightarrow \text{posteriori object log likelihood}} - \overbrace{\iint_{-m}^{m} \log P_{R_{bck}}(I(\mathbf{x}))1^{\Gamma}_{bck} d\tilde{\mathbf{x}} d\tilde{\mathbf{y}}}^{\Phi_{bck} \Rightarrow \text{posteriori background log likelihood}} - \overbrace{\log P^{t}_{S}}^{S \Rightarrow shape} \Bigg] ds, \tag{3}$$

where l is the contour length, $\mathbf{x} = (x, y)^T$ and $1^{\Gamma}_{bck} = 1 - 1^{\Gamma}_{obj}$. Note that $-\Phi_{obj} - \Phi_{bck}$ is related to the color observations and is called the "image energy," E_{image} and S is related to the object shape and is called the "shape energy," E_{shape}.

2.3 Energy Minimization and Contour Representation

Object tracking is achieved by evolving the contour in each frame, such that the final energy given in (3) is minimized. The first order necessary condition in this regard is to find the derivative of the parametric energy functional, which is associated with the Euler-Lagrange equations of the functional given in (3):

$$\frac{\delta E}{\delta x} = -\Bigg[\iint_{-m}^{m} \big(-\Psi_{obj}(\mathbf{x}) - \Psi_{bck}(\mathbf{x}) \big) d\mathbf{x} - S \Bigg] \dot{\boldsymbol{y}}, \tag{4}$$

$$\frac{\delta E}{\delta y} = \Bigg[\iint_{-m}^{m} \big(-\Psi_{obj}(\mathbf{x}) - \Psi_{bck}(\mathbf{x}) \big) d\mathbf{x} - S \Bigg] \dot{\boldsymbol{x}}, \tag{5}$$

where $\dot{\boldsymbol{y}} = \frac{\partial y}{\partial s}$ and $\dot{\boldsymbol{x}} = \frac{\partial x}{\partial s}$ for contour parameter s.

Contour evolution is directly related with the representation chosen. There are several contour representations, such as the marker-string, the volume fluid and the level set. Among these, we chose the level set due to its numerical stability and flexibility to split and merge regions [17]. In level sets, the contour is implicitly represented on a fixed grid $\phi : R^2 \times R \rightarrow R^1$, whose values are the distances from the contour, and the inside and outside the contour are defined by $\phi(\mathbf{x}) < 0$ and $\phi(\mathbf{x}) > 0$, respectively. Evolution is obtained by updating the grid, or formally, $\phi^{\xi+1} = \phi^{\xi} + F(\mathbf{x})|\nabla\phi^{\xi}|$, where ξ is the iteration number and F is the speed in the contour normal direction $\vec{n}$. After combining (4) and (5) by setting $\vec{\mathbf{v}} = (x, y)$ and $\vec{n} = [-\dot{\boldsymbol{y}} \;\; \dot{\boldsymbol{x}}]^T$, level set update becomes $\frac{\delta E}{\delta \vec{\mathbf{v}}} = -(\Phi_{obj} + \Phi_{bck} + S)\vec{n}$, which results in the following speed:

$$F_{x,y} = -\Phi_{obj} + \Phi_{bck} - S, \tag{6}$$

where $\mathbf{x}' = (x + i, y + j)$. In our implementation, an object is considered to be either occluded or unoccluded during tracking. Unless an occlusion occurs, where the object shape is dramatically distorted, we set $S = 0$ in (6), and do not compute the shape-based speed. Thus, the negative and the positive terms in (6) correspond to shrinking and expansion forces, respectively, such that when the contour hypothesis is correct, the motion of the contour will be 0. Otherwise, the background (object) likelihood will be higher and the speed will become negative (positive).

2.4 Occlusion Handling

During occlusion, visual features of the occluded objects are not observed and the objects cannot be tracked. We propose a two step approach to handle the occlusions. The first step detects the occlusion, and the second step recovers the shape of the occluded objects.

Occlusion can be detected based on both the distance between the objects and the change of the object size. Let A and B be two objects. For instance, if the distance between A and B is zero and the size of A reduces dramatically, we label A as the *occludee*. If the distance between A and B is high and the size of A is changing slowly, then we infer that the camera is zooming in or out A. Let there be N such objects with level sets ϕ_i. The Euclidean distance $D_{i,j}$ from the object O_i to the object O_j can be obtained using $D_{i,j} = \arg\min \phi_i\big(\Gamma_j(x)\big), i \neq j$. Next, the average object size, A_i^{avg}, and current object size, A_i^t, are computed. Occlusion detection is performed by evaluating $Occ_{i,j} = \frac{1}{\exp(-|D_{i,j}|)+1} \times \frac{A_i^t}{A_i^{avg}}$, such that when $Occ_{i,j} < \rho$, we conclude that an occlusion has occurred.

Fig. 2 shows an occlusion example, where two objects occlude a tennis player. Before the occlusion (Fig. 2, leftmost image), the contours are complete. During occlusion, evolution using only visual features for the tennis player results in broken contours. At this point, distances between the player and the other objects are $D_{\text{player,ellipse}} = D_{\text{player,rectangle}} = 0$. The ratio of the player area and its average area is $\frac{A_{\text{player}}}{A^{avg}_{\text{player}}} < 0.5$. Thus, $Occ_{\text{player,rectangle}} = Occ_{\text{player,ellipse}} < .25$, which implies the player is occluded.

Note that, during occlusion, the missing object parts need to be recovered. For nonrigid objects, the shape is constantly varying. We propose to model the nonrigid changes in object shape using a modified level set representation, which encodes the statistics of the object motion. Each object is first scaled and then a dense level set ϕ_i' is maintained with the outside region set to zero (Fig. 3a). Since ϕ_i' is generated in the scaled space, zooming in or out the object does not create ambiguities.

Ideally, the object shape changes gradually, resulting in a small variation in ϕ' over time. For each grid (k', l') in ϕ', the shape variation is modeled by a single Gaussian, $G_{\phi'}(k', l')$, and the model parameters (Figs. 3b and 3c) are updated until an occlusion is detected. We initialize the shape model with the initial contour with outside regions having zero means. For every frame, if no occlusion is observed, the Gaussian parameters in ϕ' are updated accordingly. Lower probabilities in this model indicate the presence of the object region, whereas higher probabilities indicate the object boundary. Setting $P_S^t = G_{\phi'}$ in (2) relates to maximizing the likelihood of observing the object boundary. After removing constant terms, we evolve the contour to recover the object shape by setting Φ_{obj} and Φ_{bck} to 0 in (6), such that:

$$F(k,l) = \frac{(\phi'_{k',l'} - \mu_{k',l'})^2}{\sigma^2_{k,l}} + \log \sigma_{k,l}, \tag{7}$$

where $\mu_{k,l}$, $\sigma_{k,l}$ are the Gaussian parameters. As can be observed, $F(k,l)$ in (7) is an expansion force which recovers the missing object parts. In Fig. 2, we demonstrate the robustness of the approach on a synthetic sequence, where a player is occluded by two objects.

3 Experiments

We have tested our algorithm on various sequences, some of which are standard sequences used by other researchers. Most of the time, the contours of the tracked objects are very tight. During tracking, object priors are computed online by reevaluating the object and the background changes. The tracking algorithm is initialized with the boundaries of the objects in the first frame and, for each frame, subband analysis of the steerable pyramids is performed using the Gabor wavelets in four directions. We chose a 10×10 analysis patch for each filter. The selection of band size in (6), m, is not sequence dependent and is fixed to 6 for all sequences. For the video sequences and more results, we refer the reader to http://www.cs.ucf.edu/~vision/projects/contour_tracking.

3.1 Single Object Sequences

Fig. 4a demonstrates the performance of tracking on the standard tennis sequence, in which the camera pans and tilts as the player performs strokes. The visual features do not change drastically throughout the sequence. The contour is perfectly tracked (even the pony tail of the player is tracked!). Note that the racket is not tracked due to the fast racket motion. These results show improved tracking performance compared to other work.

In another experiment, we tested the adaptivity of the approach to color and texture changes. Throughout the sequence the camera zooms, pans, and tilts as shown in Fig. 4b. Note that sometimes the object color and texture is similar to the background. For color and texture weighting, please see Fig. 1a. Nevertheless, the object contour is perfectly tracked. In Fig. 4c, we present tracking of a walking person in a low quality surveillance sequence. Both the object and the background textures are similar, and the color feature is sufficient to track the person (see Fig. 1b).

3.2 Occluding Object Sequences

We demonstrate the tracking of the occluded objects in Fig. 4d. Notice the white regions inside the person wearing dark clothes. Occluded person's contour is correctly recovered and both persons are tracked before, during and after the occlusion. Ambiguous head regions of both persons are correctly located. It should be noted that many state-of-the-art tracking methods either only estimate the centroid of the object during the occlusion or do not deal with it at all [7], [14]. In our case, we are able to track the complete object contour during occlusion.

3.3 Infrared Sequences

Aerial infrared (IR) imagery has a very low quality, which is affected by the atmospheric conditions. This results in frequent variation in image intensities, requiring an adaptive feature modeling mechanism. In Figs. 4e, 4f, and 4g, we present the performance of our method on IR sequences from the AMCOM dataset. The proposed method successfully models very small objects (10 to 15 pixels in area) and robustly tracks them regardless of the blurred object boundaries. In Fig. 4e, note that the window of the vehicle (dark region) toward the end of the sequence is tracked as a part of the vehicle (despite its similarity to the background). In Fig. 4f, background priors are updated throughout the sequence. Fig. 4g presents robust tracking performance for two vehicles with similar appearances.

4 Discussion

The proposed method contributes to the tracking field in several aspects. For instance, it works for mobile cameras and does not require camera motion estimation. It adapts its priors to changing color and texture features. It tracks the complete region of the nonrigid objects. It can recover occluded object parts. Despite its merits, the algorithm has limitations. For instance, complete occlusion of similar looking objects may cause ambiguities. In such cases, motion-based terms or terms encoding spatial information may be required. Disregarding the shape term, when there is no occlusion, may have disadvantages, such that for similar objects with different shapes, contour may evolve to the wrong one. Occlusion detection uses area and distance heuristics and can be ambiguous at times when the object size changes due to swift zooming, while it is close to the other objects.

In the following discussion, we will provide comparative discussion of the proposed contour energy functional (3) which consists of two major components the image energy and the shape energy.

4.1 Image Energy

There is an analogy between the image energy proposed here and energy functional proposed in [9], such that both functionals are similar in appearance. However, they have different interpretations. The averaging operation in [9], which resembles our region-based energy term, is only used for noise reduction (similar to a low-pass filter) and can be removed, whereas, in our functional, it is directly related to the contour likelihood and cannot be removed. In addition, the proposed image energy is very general, and the functionals used by Mansouri [14], Caselles et al. [6], and Paragios and Deriche [11] are special cases of it. For instance, considering only the inside of the object, dropping the plane integrals (due to the $\max$ operation) and setting the probabilities to

$$P_\alpha(\mathbf{x}) = \max_{\mathbf{z}:\|\mathbf{z}\| \le m} \exp\left(-\frac{(I^t(\mathbf{x}+\mathbf{z}) - I^{t-1}(\mathbf{x}))^2}{2\sigma^2} \right),$$

E_{image} results in the functional proposed in [14]. Similarly, using the Gaussian of the image gradient, $e^{-|\Delta I|}$, for the pixel probabilities and setting m=1, E_{image} reduces to the functional in [6]. E_{image} unifies the convex combination of boundary and the region-based functionals used in [11] through the use of the

Fig. 4. Tracking results. (a) Tennis player and (b) walking person sequence both captured by a mobile camera. (c) Surveillance sequence captured by a stationary camera. (d) Tracking two objects during occlusion; note that the full-occlusion is correctly handled. Tracking in IR sequences, sequence (e) 14_15, (f) 16_18, and (g) 16_08 from AMCOM data set.

subregional terms defined by the contour. Due to evaluation of the nonlinear image energy in the locality of the contour, the stability of the solution is also increased.

The localization introduced by the band increases the stability of the solution, and naturally generalizes *boundary-based* [5], [6] and *region-based* [9], [11], [14] contour methods, such that, if the band

size is set to 1, it becomes a boundary-based method or, if the band covers the complete object, then it becomes a region based-method.

4.2 Shape Energy

The shape energy is not available in most of the contour tracking or segmentation methods. A general trend among researchers is to introduce the shape-based terms as an external energy in the contour energy. In contrast, the proposed functional (3) naturally introduces a shape term, which is obtained during the derivation of the functional. Proposed shape energy is derived using Bayesian framework by modeling the nonrigid deformation of the contour. Compared to other shape-based contour methods which require training [15], our algorithm learns the object shape online. Note that offline training of the object shape models is not an easy task and is not suitable for tracking all kinds of objects that may appear in the scene.

5 Conclusions

We proposed a contour-based nonrigid object tracking method. Along with color and texture models generated for the object and the background regions, our method maintains a shape prior for recovering occluded object parts during the occlusion. The shape priors encode the motion of the object and are built online. The energy functional is derived using a Bayesian framework and is evaluated around the contour to suppress visual artifacts and to increase numerical stability. We minimized the energy in the gradient descent direction, which in turn maximizes the posteriori contour probability. The results presented show the robust tracking performance with occlusion in video acquired from moving cameras.

References

[1] C.R. Wren, A. Azarbayejani, T. Darrell, and A.P. Pentland, "Pfinder: Real-Time Tracking of the Human Body," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, July 1997.

[2] C. Stauffer and W. Grimson, "Learning Patterns of Activity Using Real Time Tracking," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 8, pp. 747-767, Aug. 2000.

[3] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-Based Object Tracking," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 5, pp. 564-575, May 2003.

[4] A.D. Jepson, D.J. Fleet, and T.F. El-Maraghi, "Robust Online Appearance Models for Visual Tracking," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 10, Oct. 2003.

[5] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *Int'l J. Computer Vision,* vol. 1, 1988.

[6] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic Active Contours," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 694-699, 1995.

[7] N. Paragios and R. Deriche, "Geodesic Active Contours and Level Sets for the Detection and Tracking of Moving Objects," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 3, pp. 266-280, 2000.

[8] M. Bertalmio, G. Sapiro, and G. Randall, "Morphing Active Contours," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 7, July 2000.

[9] S. Zhu and A. Yuille, "Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 9, pp. 884-900, Sept. 1996.

[10] S. Jehan-Besson and M. Barlaud, "Video Object Segmentation Using Eulerian Region-Based Active Contours," *Proc. IEEE Int'l Conf. Computer Vision,* 2001.

[11] N. Paragios and R. Deriche, "Geodesic Active Regions and Level Set Methods for Supervised Texture Segmentation," *Int'l J. Computer Vision,* vol. 46, no. 3, pp. 223-247, 2002.

[12] A. Yezzi, L. Zollei, and T. Kapur, "A Variational Framework for Joint Segmentation and Registration," *Proc. Workshop Math. Methods in Biomedical Image Analysis,* 2001.

[13] J. Rittscher and A. Blake, "A Probabilistic Background Model for Tracking," *Proc. Europen Conf. Computer Vision,* vol. 2, 2000.

[14] A. Mansouri, "Region Tracking via Level Set PDEs Without Motion Computation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 7, pp. 947-961, July 2002.

[15] D. Cremers, T. Kohlberger, and C. Schnörr, "Nonlinear Shape Statistics in Mumford-Shah Based Segmentation," *Proc. Europen Conf. Computer Vision,* 2002.

[16] W. Freeman and E. Adelson, "The Design and Use of Steerable Filters," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 13, no. 9, pp. 891-906, Sept. 1991.

[17] J. Sethian, *Level Set Methods: Evolving Interfaces in Geometry, Fluid Mechanics Computer Vision and Material Sciences.* Cambridge Univ. Press, 1999.