Occlusion Tracking with Logic Models*

James H von Brecht**, Sheshadri R Thiruvenkadam***, and Tony F Chan

UCLA Department of Mathematics
405 Hilgard Avenue, Los Angeles, CA 90095

Abstract. We present a variational PDE based model for tracking objects under occlusion. Here, prior shape information is used within a *logic-based* framework as a means for detecting and segmenting objects under partial occlusions. The model was tested on real and synthetic image sequences with promising results.

1 Introduction

Occlusion tracking [6–8, 13] presents a difficult problem since most, if not all, of the information which characterizes a particular object becomes unreliable under occlusions. A successful algorithm for occlusion tracking, then, must have some means of identifying the object of interest when such information is lacking, or inaccurate. Typically, this is accomplished by using information from other frames of a video sequence to aid in the identification of the object in an occluded frame.

In this work, we assume that the boundary of the object of interest is available, and use this *prior shape* information to track the object in occluded frames. There have been previous works[10–12] that use prior knowledge of the shape of objects to facilitate segmentation specially under low contrasts, occlusions and other undesirable noisy conditions. Most of these works incorporate the shape term additively within the segmentation energy which results in locally optimal solutions. The novelty in this work is that the shape prior is combined with the image term using *logical operations* pertaining to a unique occlusion scenario, thus leading to "meaningful" solutions. Our work is based on Sandberg et. al.[1] algorithm for logical segmentation of multi channel images, and related to the joint segmentation and registration framework of Moelich et. al. [2].

2 Description of the Model

Our model is a forward tracking algorithm; we rely solely upon data from previous frames in order to identify and segment the object in the current frame.

* This research supported by NSF grants DMS-0601395, DMS-0610079 and ARO MURI grant 50363-MA-MUR.

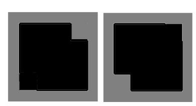
** jub@ucla.edu

*** sheshad@math.ucla.edu

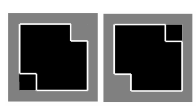
Throughout the remainder of the discussion, we make three assumptions: the first frame of the video sequence, which we refer to as the "template" frame, contains the entire, un-occluded boundary of the object (*shape prior*) we wish to track; second, the object boundary does not deform; and third, the object undergoes only affine movement between frames. The first assumption will allow us to not have to use any *a priori* about the particular occlusion scenario for the algorithm to succeed. The second assumption results from the desire to incorporate prior shape information into the algorithm. We make the last assumption for simplicity, in that more complicated motions amount to a more sophisticated registration model than we use here.

2.1 Logic Models

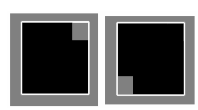
Before introducing our model, we briefly describe the region-based logic models [1, 2] based upon the Chan-Vese segmentation energy [4]. While dealing with multi-channel images, often there are disparities in the appearance of an object in each of the channels. In such cases, there is more than one valid interpretation of the actual object. These alternate interpretations correspond to different logical interpretations of the images. The *logic models*, developed by Sandberg et. al. [1], are designed to segment multi-channel images according to such logical interpretations. For example, given two images f_1 and f_2 in Figure 1(a) which contain two different instances of a particular object of interest, logic models allow us to interpret the actual object (white curve in Figure 1 (b) and (c)) by combining the segmentation in each frame according to a pre-selected logical operation.



(a) Images



(b) $f_1 \cap f_2$ (AND)



(c) $f_1 \cup f_2$ (OR)

Fig. 1. Logical Segmentation on Two Frames

In particular, we consider two logic segmentation models, denoted as $f_1 \cap f_2$(AND) and $f_1 \cup f_2$(OR). The AND model interprets the actual object as the intersection of the object regions that appear in the two frames. Similarly, the OR model, then, is the union of the object regions which appear in the frames. As discussed in [1], a segmentation energy for a single channel can be easily recast in to a logical framework when dealing with multi channels. We briefly review the discussion from [1] here.

Given two frames f_1 and f_2, each which contain an object of interest (that might appear different in each frame), we first define the functions

$$\begin{aligned} z_1^{in} &= \frac{\left(f_1 - c_1^{in}\right)^2}{M_1} \quad z_1^{out} = \frac{\left(f_1 - c_1^{out}\right)^2}{N_1} \\ z_2^{in} &= \frac{\left(f_2 - c_2^{in}\right)^2}{M_2} \quad z_2^{out} = \frac{\left(f_2 - c_2^{out}\right)^2}{N_2}. \end{aligned} \tag{1}$$

As in the standard C-V model, c_i^{in} $(i = 1, 2)$ represents the average intensity inside the object in frame f_i, respectively. Similarly, c_i^{out} represents the average intensities of the background. The constants M_i and N_i ensure that each z function takes values only between 0 and 1. The segmentation energy takes the familiar form

$$E = \int_\Omega f_{in} H(\phi) + \int_\Omega f_{out} (1 - H(\phi)) + \int_\Omega (|\nabla \phi| - 1)^2 , \tag{2}$$

where the last term is a regularization term which prevents the level-sets of ϕ from becoming too flat. The functions f_{in} and f_{out} vary depending upon which logical combination is desired.

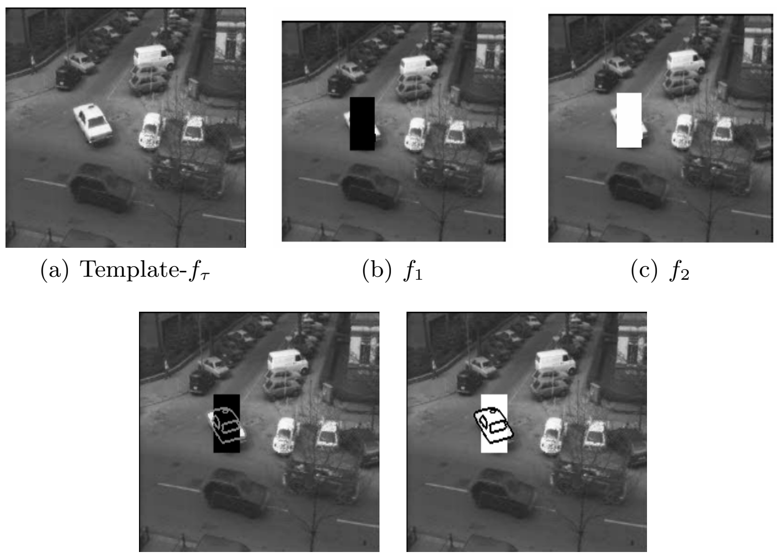
Based upon the definitions of the z-functions (1), we have $z_1^{in} \approx 0/1$ inside/outside the object in f_1 and similarly $z_1^{out} \approx 0/1$ outside/inside the object in f_1. The case is similar for the functions z_2. When taking the OR model, we desire $f_{in} = 0$ for all points inside the object in *at least* one of the frames (see Figure 1(c)), and $f_{out} = 0$ for all points that lie outside the object in *both* frames, which we achieve by defining f_{in} and f_{out} as

$$\begin{aligned} f_{in}^{\cup} &= \sqrt{z_1^{in} z_2^{in}} \\ f_{out}^{\cup} &= 1 - \sqrt{\left(1 - z_1^{out}\right)\left(1 - z_2^{out}\right)}. \end{aligned} \tag{3}$$

For the AND model (Figure 1(b)), the case is reversed. We desire $f_{in} = 0$ for all points inside the object in both frames, and $f_{out} = 0$ for points outside the object in either frame. The resulting definitions are

$$\begin{aligned} f_{in}^{\cap} &= 1 - \sqrt{\left(1 - z_1^{in}\right)\left(1 - z_2^{in}\right)} \\ f_{out}^{\cap} &= \sqrt{z_1^{out} z_2^{out}}. \end{aligned} \tag{4}$$

Before we discuss our model, we first motivate using the logic models in a tracking algorithm. In Figure 2, we provide a simple demonstration how the logic models as discussed above can be used to recover the boundary of an artificially occluded object (the white turning car) . The first image (a), or template f_τ, is used to segment the other two occluded frames (b and c). In the second image (b), a region of different intensity occludes the object, and hence we take the model $f_\tau \cup f_1$ to recover the boundary of the object (curve shown in (d)). In the image (c), the occlusion is of similar intensity to the car, so $f_\tau \cap f_2$ yields the desired result (curve shown in (e)). Note how the appropriate logic model depends upon the intensity of the occlusion relative to the object. When no occlusion is present, either logic model will give the desired result, since the object appears identical in each frame. Thus we see that the application of the shape prior (through the template image f_τ) depends on the occlusion type, which allows our algorithm to *avoid local minima* problems of models that just additively introduce the shape term.



(a) Template-f_τ (b) f_1 (c) f_2

(d) $f_\tau \cup f_1$ (e) $f_\tau \cap f_2$

Fig. 2. Occlusion Segmentation

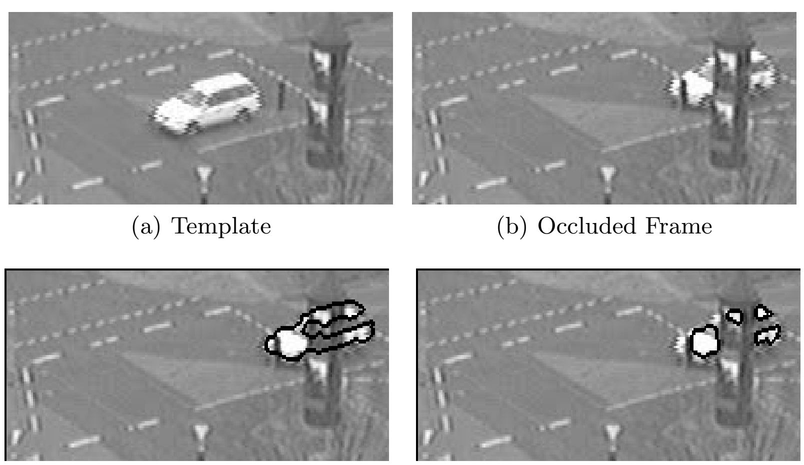
To summarize, in this paper, we deal with two types of occlusion scenario, depending on the intensity of the occlusion relative (similar/different) to the object being tracked. We use an appropriate logic segmentation model (AND/OR) for the occlusion scenario, to correctly segment the object from the current frame, using the template image. Finally, we discuss a technique which we use to automatically switch between the logic models to deal with changing occlusion scenario across frames.

2.2 Joint Registration and Segmentation

The logic models as presented in [1] assume pre-registered images. Consequently, for use in a tracking algorithm, we must incorporate a registration model into our algorithm to reflect object motion. For our purposes, we employ a Joint Registration and Segmentation algorithm similar to [3] and extended to the logic models by Moelich and Chan in [2]. Given a template frame $f_\tau : \Omega_1 \rightarrow \Re^+$ and an occluded frame $f_i : \Omega \rightarrow \Re^+$, we register the two images by introducing a spatial correspondence between the domains Ω_1 and Ω, denoted by g, with the parameters $\{\Delta x \ \Delta y \ \theta\}$. Here, Δx and Δy represent translation and θ represents rotation. The selection of the transformation g is arbitrary; many other valid choices exist which allow for more general object motions. For further details, see [3] or [2].

2.3 Automation of Logic Models

As discussed earlier, while segmenting the boundary of an occluded object (e.g. Figure 2), the correct logical model to be used depends upon the similarity of the intensities of the object and the occlusion. The application of the incorrect logic model will lead to not only an error when segmenting the images (see Fig. 3 (d)), but can also conceivably cause an error in the registration of the images as well. Consequently, in order to employ the logic models in a tracking algorithm, we introduce a method by which we can *automatically* determine the appropriate choice of logical segmentation. To determine the appropriate logic



(a) Template (b) Occluded Frame

(c) Correct Model (OR) (d) Incorrect Model (AND)

Fig. 3. Need for Automation

model, we make use of the prior shape of the object, given by the contour C_τ in the template frame f_τ. Also, we denote the contour given by logical segmentation of the current frame by C. Regardless of the intensity of the occluding object, the correct logic model is that which gives the least shape dissimilarity between C_τ

and C. In this work, we use area difference as the shape dissimilarity measure. Therefore, the correct logic model is that which minimizes the quantity

$$\left(AREA\left(inside\left(C_{\tau}\right)\right)-AREA\left(inside\left(C\right)\right)\right)^{2}. \tag{5}$$

In the level-set framework, we denote by ψ the function used to implicitly represent the contour C_{τ}, and ϕ represents C. The variational form of (5) then becomes

$$\int_{\Omega}\left(H\left(\psi_{g}\right)-H\left(\phi\right)\right)^{2}, \tag{6}$$

where $\psi_{g}=\psi\left(g^{-1}\right)$, and H is the Heaviside function. To enforce this constraint in practice, we compute two functions, $\phi^{\cap}$ using the AND model, and $\phi^{\cup}$ using the OR model, check the quantity (6) in each case, and select as ϕ that which produces a minimum.

2.4 Variational Framework

We now describe the level-set formulation of our model. Given a template frame f_{τ} and the i^{th} frame from a video sequence f_i, define $F_{\tau}=f_{\tau}\left(g^{-1}\right)$. Extending the logic models to include a registration component is then straightforward. The z-functions (1) become

$$\begin{aligned} z_{\tau}^{in} &= \frac{\left(F_{\tau}-c_{\tau}^{in}\right)^{2}}{M_{\tau}} \quad z_{\tau}^{out}=\frac{\left(F_{\tau}-c_{\tau}^{out}\right)^{2}}{N_{\tau}} \\ z_{i}^{in} &= \frac{\left(f_{i}-c_{i}^{in}\right)^{2}}{M_{i}} \quad z_{i}^{in}=\frac{\left(f_{i}-c_{i}^{out}\right)^{2}}{N_{i}}. \end{aligned} \tag{7}$$

Note that in practice, to target the correct object in the i^{th} frame, we fix $c_i^{in}=c_{\tau}^{in}$ and only update c_i^{out} along with the contour. The functions $f_{in,out}$ then become

$$\begin{aligned} f_{in}^{\cap} &= 1-\sqrt{\left(1-z_{\tau}^{in}\right)\left(1-z_{i}^{in}\right)} \\ f_{out}^{\cap} &= \sqrt{z_{\tau}^{out}z_{i}^{out}}. \end{aligned} \tag{8}$$

$$\begin{aligned} f_{in}^{\cup} &= \sqrt{z_{\tau}^{in}z_{i}^{in}} \\ f_{out}^{\cup} &= 1-\sqrt{\left(1-z_{\tau}^{out}\right)\left(1-z_{i}^{out}\right)}. \end{aligned} \tag{9}$$

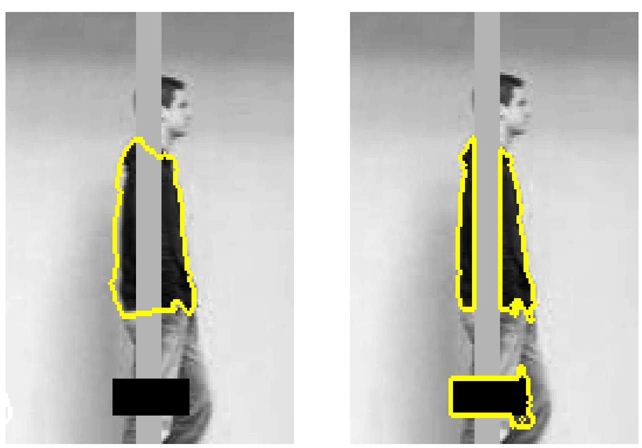
And finally, in our formulation, for each of the logic models (AND/OR), we add the shape term (5) to the segmentation energy. The addition of the shape term has several benefits. Foremost, it helps to ensure a correct registration between frames. Also, it helps prevent unwanted portions (usually similar to the object's intensity) of the image from being included in the final segmentation. Thus, we may introduce the variational form of our model,

$$E\left(\phi,\Delta x,\Delta y,\theta\right)=\int_{\Omega}f_{in}H\left(\phi\right)+f_{out}\left(1-H\left(\phi\right)\right)\ dx$$
$$+\beta\int_{\Omega}\left(H\left(\psi_g\right)-H\left(\phi\right)\right)^2\ dx+\lambda\int_{\Omega}\left(\left|\nabla\phi\right|-1\right)^2\ dx. \qquad (10)$$

Again, H is the Heaviside function, and ψ represents the boundary of the object in the (unregistered) template frame. The function ψ_g then, is defined as $\psi_g = \psi\left(g^{-1}\right)$. λ and β are parameters to balance the terms.

In this context, our algorithm reduces to finding sequentially, for each frame f_i of the sequence, the function ϕ_i and the parameters $p_i = \{\Delta x_i\ \Delta y_i\ \theta_i\}$ which minimize the energy (10). Since the functions $f_{in,out}$ differ depending upon the logic model, we minimize the energy separately for each case, to produce two sets of functions with coupled parameters, $\{\phi_i^{\cap},\ p_i^{\cap}\}$ and $\{\phi_i^{\cup},\ p_i^{\cup}\}$, then select as $\{\phi_i, p_i\}$ that which minimizes (6). This selection gives the desired segmentation; the segmentation closest to the shape prior.

In summary, we combine both shape information and the correct logical interpretation of images to achieve the desired result. We can thus avoid many local minima that other models, which just additively introduce shape, may encounter. In Fig. 4, the first frame demonstrates the result of our algorithm, which combines prior shape and the automated choice of logic model to achieve the desired segmentation. The final frame demonstrates that prior shape alone is not sufficient, and might result in a local minimum.



(a) Logic and Shape (b) Shape without Logic

Fig. 4. Various Combinations of Logic Models and Prior Shape

3 Numerical Implementation

When implementing the algorithm, we begin with an initial ϕ_0 and an initial set of parameters $\{\Delta x_0\ \Delta y_0\ \theta_0\}$, and evolve them according to the Euler-Lagrange equations of (10) until a minimum is reached. The equations for gradient descent are given by

$$\frac{\partial \phi}{\partial t} = \left[f_{out} - f_{in} + 2\beta \left(H\left(\psi_g\right) - H\left(\phi\right)\right)\right] \delta\left(\phi\right) + 2\lambda \left(\nabla^2 \phi - div\left(\frac{\nabla \phi}{|\nabla \phi|}\right)\right) \quad (11)$$

and

$$\begin{aligned}
\frac{\partial \Delta x}{\partial t} &= -\int_{\Omega} \left(H\left(\phi\right) \frac{\partial f_{in}}{\partial F_\tau} \frac{\partial F_\tau}{\partial \Delta x} + 2\beta \left(H\left(\psi_g\right) - H\left(\phi\right)\right) \delta\left(\psi_g\right) \frac{\partial \psi_g}{\partial \Delta x}\right) dx. \\
\frac{\partial \Delta y}{\partial t} &= -\int_{\Omega} \left(H\left(\phi\right) \frac{\partial f_{in}}{\partial F_\tau} \frac{\partial F_\tau}{\partial \Delta y} + 2\beta \left(H\left(\psi_g\right) - H\left(\phi\right)\right) \delta\left(\psi_g\right) \frac{\partial \psi_g}{\partial \Delta y}\right) dx. \\
\frac{\partial \theta}{\partial t} &= -\int_{\Omega} \left(H\left(\phi\right) \frac{\partial f_{in}}{\partial F_\tau} \frac{\partial F_\tau}{\partial \theta} + 2\beta \left(H\left(\psi_g\right) - H\left(\phi\right)\right) \delta\left(\psi_g\right) \frac{\partial \psi_g}{\partial \theta}\right) dx. \quad (12)
\end{aligned}$$

To minimize (10), we follow the procedure outlined in [2, 3]. Beginning from the initial set $\{\Delta x_0\ \Delta y_0\ \theta_0\ \phi_0\}$, we first hold ϕ fixed and perform one iteration of (12), and update the functions F_τ and ψ_g with the new parameters. The registration parameters are then fixed and ϕ is updated via one iteration of (11). This process is continued until convergence.

In practice, selecting the initial set of registration parameters for the i^{th} frame presents the largest barrier for our algorithm to track successfully. That is, how to select $p_0^i = \left\{\Delta x_0^i\ \Delta y_0^i\ \theta_0^i\right\}$. In simpler cases, it suffices simply to take $p_0^i = p_{final}^{i-1}$. However, when the sequence has moderate to severe occlusions or low contrast (see Fig. 7), the algorithm becomes more sensitive to local minima, and hence we must select the parameters more carefully. From our experience, we have developed several techniques to combat this sensitivity. Based upon the definitions of $f_{in}^{\cup}$ and $f_{in}^{\cap}$, we see that the OR model allows for more possible registration local minima than the AND model, and hence the AND model is less sensitive to the prediction of p_0^i. Consequently we can first compute $\phi^{\cap}$ and use those final parameters as the initial parameters for the computation of $\phi^{\cup}$. Alternatively, we might hypothesize a motion trajectory for the tracked object (in the case of Fig. 7-a linear trajectory), and use this to generate each p_0^i. Finally, to reduce the computational burden we run the algorithm only locally around the object of interest in each frame. That is, we simply crop from each frame a small region around the object of interest (this process is automated via use of the shape prior and the prediction of p_0 for each frame) and run the algorithm on the reduced frame. Once we have the desired contour, we simply bring the result back to the original image. Due to such difficulties, and to try and develop a more robust and computationally efficient algorithm, we plan to incorporate the described logical framework into a particle filtering algorithm, such as that described in [13].

4 Experimental Results

We now give results of our algorithm on both synthetic and real examples. In all cases, the first frame f_1 was used as the template, thus the sequences begin with frame f_2. The first example (Figure 5) demonstrates the need to automate the choice of logic model via (6). Without some means of determining the appropriate logical interpretation of the images, an undesirable segmentation can result. In this example, the arbitrary choice was made to use the OR model across all frames. While the algorithm tracks successfully through the first occlusion, when the person reaches the second occlusion, the OR model is no longer correct, and so the algorithm fails. In the later frames of the sequence, when the person has passed completely through the second occlusion, the correct segmentation is once again realized since in such frames, when no occlusion is present, the object appears identical in each channel and hence either logic model gives the desired result. However, in more severe cases, since the registration prediction also depends upon the accuracy of the final segmentation in the previous frame, if the incorrect model is used, the algorithm can completely lose track of the object.

The second example (Figure 6) shows the full algorithm on the same sequence as in (Figure 5). It demonstrates the capability of the algorithm to handle occlusions of both types when the automation method is utilized. The intersection model was taken automatically as the person passes through the black-line occlusion, and union automatically through the second, gray-line occlusion. The current model is unable to cope with the intermediate case, in which the object of interest is simultaneously occluded by regions of both similar intensity and different intensity to the object itself. Such a scenario requires a combination AND/OR model, and we are currently experimenting with a multi-phase level-set method to handle this final case.

The final example, (Figure 7) demonstrates the algorithm on a real video sequence, and was the most challenging. We employed the full algorithm as described, which selected only the union model across each frame. As the sequence progresses, poor image contrast and more severe occlusions make the tracking more difficult, but with a careful choice of the target intensity c_τ^{in} and a careful prediction of the initial parameters at each step, our algorithm succeeded. Video clips of the last two examples are available for viewing at `http://www.math.ucla.edu/~sheshad/mansequence.avi` and `http://www.math.ucla.edu/~sheshad/carsequence.avi`.

4.1 Acknowledgments

The images in Figures 2 and 7 were taken from `http://i21www.ira.uka.de/image_sequences`. The images in Figure 6 were provided by Tasha N. Carey.

(a) f_2

(b) f_{17}

(c) f_{21}

(d) f_{23}

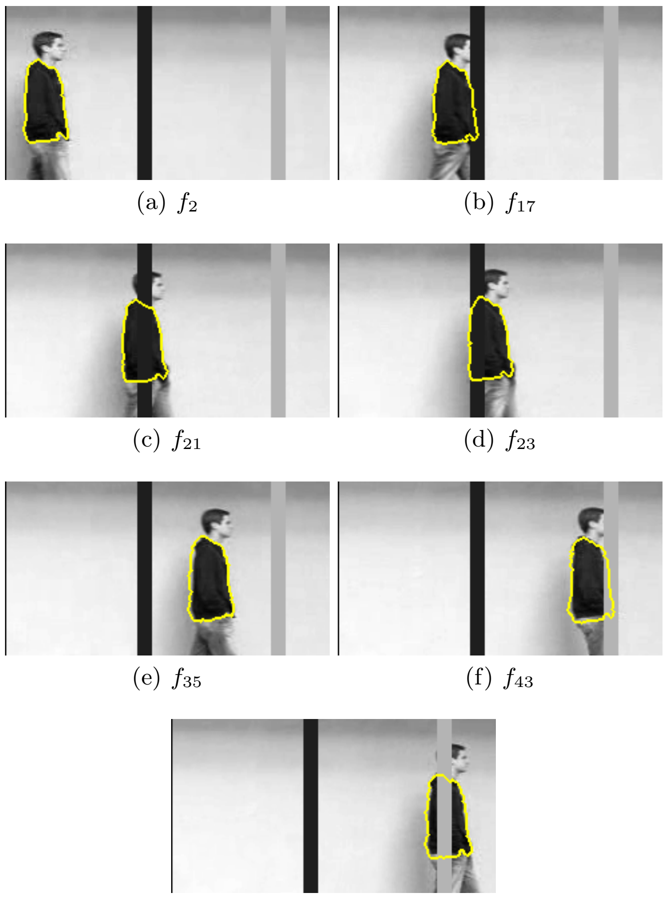
(e) f_{35}

(f) f_{43}

(g) f_{48}

(h) f_{77}

Fig. 5. Result without using (6)



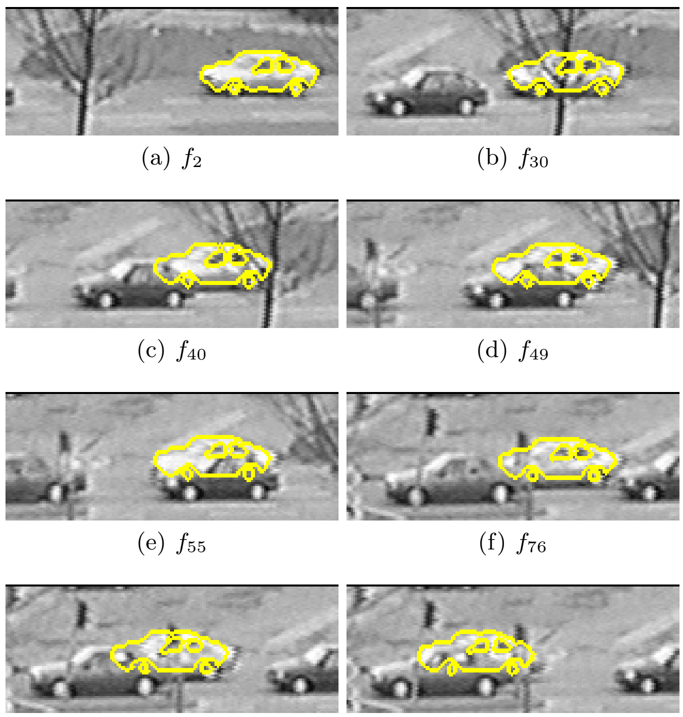
(a) f_2 (b) f_{17}

(c) f_{21} (d) f_{23}

(e) f_{35} (f) f_{43}

(g) f_{48}

Fig. 6. Tracking through both types of occlusions



(a) f_2 (b) f_{30}

(c) f_{40} (d) f_{49}

(e) f_{55} (f) f_{76}

(g) f_{83} (h) f_{90}

Fig. 7. Real sequence with poor image contrast

References

1. B. Sandberg and T. Chan. *Logic Operators for Active Contours on Multi-Channel Images*. UCLA CAM Report 02-12, 2002.
2. M. Moelich and T. Chan. *Joint Segmentation and Registration Using Logic Models*. UCLA CAM Report 03-06, 2003.
3. A. Yezzi, L. Zollei, and T. Kapur. *A variational approach to joint segmentation and registration*. In Proc. IEEE Conf. on Comp. Vision and Pattern Recogn., 2001.
4. T. Chan and L. Vese. *Active Contours Without Edges*. IEEE Transactions on Image Processing, vol. 10, no. 8, pp. 266-277, 2001.
5. T. Chan, B. Sandberg, and L. Vese. *Active contours without edges for vector-valued images.* Journal of Visual Communication and Image Representation, 11:130-141, 1999.
6. Y. Shi, J. Konrad, and W. Karl. *Multiple Motion and Occlusion Segmentation with a Multiphase Level Set Method*. 2004.
7. A. Bartesaghi and G. Sapiro. *Tracking of Moving Objects under Severe and Total Occlusions*. IMA Preprint Series 2015, 2005.
8. J. Jackson, A. Yezzi, and S. Soatto. *Tracking Deformable Moving Objects Under Severe Occlusions*. IEEE Conf. on Decision and Control, 2004.
9. M. Moelich and T. Chan. *Tracking Objects with the Chan-Vese Algorithm*. UCLA CAM Report 03-14, 2003.
10. Y. Chen, H. Tagare, S. Thiruvenkadam, F. Huang, D. C. Wilson, K.S. Gopinath, R W. Briggs, and E. A. Geiser. *Using prior shapes in geometric active contours in a variational framework.* IJCV, 50(3):315-328, 2002.
11. Timothy F. Cootes, Christopher J. Taylor, David H. Cooper, and Jim Graham. *Active shape models-their training and application*. CVPR, 61(1):38-59, 1995.
12. M. Leventon, W. L. Grimson, and O. Faugeras. *Statistical shape influence in geodesic active contours.* In CVPR, volume 1, pages 316-323, 2000.
13. Y. Rathi, N. Vaswani, A. Tannenbaum, and A. Yezzi. *Particle Filtering for Geometric Active Contours with Application to Tracking Moving and Deforming Objects.*