

LEVEL SET TRACKING WITH DYNAMICAL SHAPE PRIORS

Xue Zhou, Xi Li and Weiming Hu

National Laboratory of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences, Beijing, China
{xzhou, lixi, wmhu}@nlpr.ia.ac.cn

ABSTRACT

Dynamical shape priors are curical for level set-based non-rigid object tracking with noise, occlusions or background clutter. In this paper, we propose a level set tracking framework using dynamical shape priors to capture contours changes of an object in a periodic action sequence. The framework consists of two stages— off-line training and on-line tracking. During the off-line training stage, a graph-based dominant set clustering (DSC) method is applied to learn a shape codebook with each codeword representing a certain shape mode. Then a codeword transition matrix is learnt to characterize the temporal correlations of contours of an object. During the on-line tracking stage, we fuse the knowledge of shape priors and current observations, and adopt maximum a posteriori (MAP) estimation to predict the current shape mode. The experimental results on synthetic and real video sequences demonstrate the effectiveness of our method.

Index Terms— Tracking, level set, markov model, dynamical shape priors

1. INTRODUCTION

Level set-based methods are very popular in dealing with many computer vision tasks such as object detection, image segmentation and tracking [1] [2] [3]. In the case of noise, partial occlusions and background clutter, low-level appearance features (color, texture, etc.) are inadequate to perform the above tasks. Thus, some high-level prior knowledge about the shape of objects is necessary to be integrated into the level set evolution framework.

In recent years, there are lots of work on modeling shape priors in a level set framework [2] [4] [5]. Leventon et al. [4] incorporate the shape information into the image segmentation process. A Gaussian model is constructed in the low-dimensional shape subspace. Due to the little variation in shape when dealing with medical imagery segmentation task, the static shape model tackles this task well. Paragios and Rousson [2] propose another pixel-wise static shape model in which a Gaussian density function is described at each grid location. However, the aforementioned static shape models

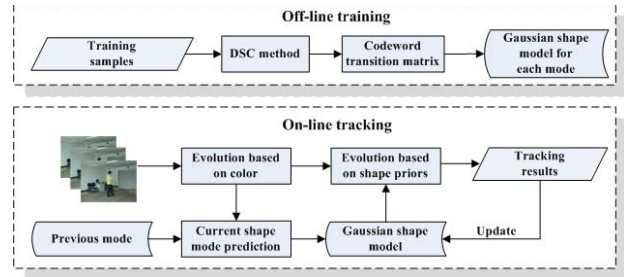


Fig. 1. Flowchart of our method

can't be updated on-line, resulting in the failures of adaptation to new shapes. Active shape models (ASMs) proposed by Cootes et al. [6] are often used to model different aspects of rigid objects in a shape prior formalism. Fussenegger et al. [7] employ a partially learned ASM (only finite samples are used in the training step) to perform segmentation, and then the segmentation results are utilized to update the ASM with an incremental Principal Component Analysis (IPCA) algorithm. With respect to the tracking problem, especially for the nonrigid objects with continuous and large shape changes, a dynamical shape model considering the temporal correlations is imperative. Cremers [5] proposes a linear dynamical shape model for level set-based tracking. The temporal evolution of the eigenmodes of the level set function is approximated by an autogressive model. In the model, a particular shape at the current time instance is reconstructed by the shapes observed at previous time instances.

In our method, we propose a markov model-based dynamical shape model. Unlike the dynamical shape model [5] based on data fitting process and without data understanding, our shape model is a high level understanding of shape knowledge. Our tracking framework consists of two stages: off-line training and on-line tracking. During the off-line shape model training stage, a dominant set-based clustering (DSC) method [8] is used to obtain the dominant shape modes of a periodic action sequence. Each mode is viewed as a codeword and the codeword transition matrix encoding the markov property is computed by counting the times for any codeword transition pair between adjacent frames. Then, for each shape mode, a Gaussian model is constructed at each grid lo-

cation. During the on-line tracking stage, three main steps are needed. Firstly, the initial contour is evolved only with the color feature; secondly, the obtained result and the constructed shape priors are fused to predict the current shape mode; thirdly, the contour is continuously evolved with the shape constraint and the final result is returned to update the shape model. Fig.1 shows the flowchart of our method.

The remainder of this paper is organized as follows: Section 2 describes the off-line training process of the shape prior model. The on-line tracking process is introduced in Section 3. Experimental results are given in Section 4. The last section concludes the paper.

2. OFF-LINE SHAPE MODEL TRAINING

Within this stage, the shape model is constructed using a set of aligned training contours. In our method, contours are represented by the level set method [9] which embeds the shape information in the signed distance map represented by Φ . We define the sign of the interior point as negative and vice versa. The similarity between any two contour samples (Φ_i, Φ_j) is computed as:

$$Sim_{(\Phi_i, \Phi_j)} = e^{-\frac{d(\Phi_i, \Phi_j)^2}{\beta^2}} \quad (1)$$

$$d(\Phi_i, \Phi_j) = \min \left\{ \sum_{x,y} |H_a(\Phi_i(x, y)) - H_a(\Phi_j(A(x, y)))|; \sum_{x,y} |H_a(\Phi_j(x, y)) - H_a(\Phi_i(\tilde{A}(x, y)))| \right\}$$

where $|\cdot|$ is the absolute value operator, A and \tilde{A} are the optimal affine transformation parameters for shape registration [10]: $\Phi_i \xrightarrow{A} \Phi_j$, $\Phi_j \xrightarrow{\tilde{A}} \Phi_i$, β is a constant, and $H_a(\Phi(x, y))$ is a Heaviside function:

$$H_a(\Phi(x, y)) = \begin{cases} 0 & \Phi(x, y) \geq 0 \\ 1 & \Phi(x, y) < 0 \end{cases} \quad (2)$$

After computing the similarity matrix of contour samples, a DSC method [8] (a novel graph-based clustering method adopting an iterative bipartition strategy) is used to discover the typical shape modes. Each mode represents a certain kind of similar shapes. For convenience, we call each mode a shape codeword. Thus, all samples are mapped by DSC method into their corresponding shape codewords which form a shape codebook. Then, a codeword transition matrix is used for modeling the temporal correlations of shape changes in a periodic action sequence. Each element of the codeword transition matrix is computed by counting the times for any codeword transition pair between adjacent frames.

For each shape mode, a Gaussian model is constructed at each grid location [2]. Compared with the global shape subspace model [4], this shape model accounts for local variations. The Gaussian probability density function (pdf) at pixel (x, y) in the shape model is formulated as:

$$p_{x,y}^M(\Phi) = \frac{1}{\sqrt{2\pi}\sigma_M(x, y)} e^{-\frac{(\Phi - \Phi_M(x, y))^2}{2\sigma_M^2(x, y)}} \quad (3)$$

where $\Phi_M(x, y)$ and $\sigma_M(x, y)$ are the mean shape map and the variance of shape deformations respectively.

3. ON-LINE TRACKING

The on-line tracking consists of three main steps: contour evolution based on the color feature, the current shape mode prediction and contour evolution with the shape constraint.

3.1. Contour evolution with the color feature

The method we adopt in this step is a region-based active contours method, modeling the features of both object and background regions in the level set speed model. In our method, we train a color Gaussian Mixture Model (GMM) in object and background regions respectively. The HSV color space is chosen in this model.

The data energy function is formulated based on the segmentation idea similar to [3]. The key factor in this process is to find the optimal partition operator represented by a contour between the object region and the background region. The data energy function E_{data} is defined as follows:

$$E_{data} \approx - \iint_{x_i \in R_{in}} \log P(x_i | \theta_{in}) dx_i - \iint_{x_j \in R_{out}} \log P(x_j | \theta_{out}) dx_j \quad (4)$$

where R_{in} and R_{out} denote respectively the regions inside and outside the object contour, θ_{in} is the parameters of the object GMM and $P(x_i | \theta_{in})$ is the Gaussian likelihood function of pixel x_i . $P(x_i | \theta_{out})$ and θ_{out} are defined by analogy.

Minimizing the above energy function by solving the correlated Euler-Lagrange equations [3], we obtain the level set speed model in which a $(2l+1) \times (2l+1)$ square neighboring subregion around the center pixel is defined. The object and the background posterior probabilities which we denote by $P_{R_{in}}(I_{\tilde{x}})$ and $P_{R_{out}}(I_{\tilde{x}})$ are also calculated in the speed model with the assumption that they have the same prior probabilities:

$$P_{R_{in}}(I_{\tilde{x}}) = P(\tilde{x} | \theta_{in}) / [P(\tilde{x} | \theta_{in}) + P(\tilde{x} | \theta_{out})] \quad (5)$$

$$P_{R_{out}}(I_{\tilde{x}}) = P(\tilde{x} | \theta_{out}) / [P(\tilde{x} | \theta_{in}) + P(\tilde{x} | \theta_{out})] \quad (6)$$

The level set speed model of each pixel (x, y) is obtained by:

$$F_{x,y} = - \sum_{i=-l}^l \sum_{j=-l}^l \log P_{R_{in}}(I_{\tilde{x}}) H_a(\Phi(\tilde{x}, t)) + \sum_{i=-l}^l \sum_{j=-l}^l \log P_{R_{out}}(I_{\tilde{x}}) (1 - H_a(\Phi(\tilde{x}, t))) \quad (7)$$

where \tilde{x} is the neighboring pixels of the center pixel (x, y) : $\tilde{x} = (x+i, y+j)$ and $H_a(\Phi)$ is a Heaviside function (2).

The initial contour is evolved to the desired boundary by modifying Φ iteratively with the speed $F_{x,y}$ in the normal direction:

$$\frac{\partial \Phi}{\partial t} + F_{x,y} \cdot |\nabla \Phi| = 0 \quad (8)$$

We let Φ_c denote the evolution result obtained only with the color feature.

3.2. Shape mode prediction

In this step, the result obtained only considering the color feature and the constructed shape priors are fused to predict the current shape mode. This prediction problem can be modeled as a MAP one. We let the posterior probability of prediction of current shape mode S_t given the current observations O_t and previous shape modes $S_{1:t-1}$ be represented by $P(S_t|O_t, S_{1:t-1})$. Thereby, the optimal current shape mode S_t^* should satisfy:

$$S_t^* = \arg \max_{S_t} P(S_t|O_t, S_{1:t-1}) \quad (9)$$

In accordance with the Bayesian formula, this posterior probability is equal to:

$$P(S_t|O_t, S_{1:t-1}) = \frac{P(O_t|S_t, S_{1:t-1})P(S_t|S_{1:t-1})}{P(O_t|S_{1:t-1})} \quad (10)$$

We assume O_t and $S_{1:t-1}$ are uncorrelated and $P(S_t|S_{1:t-1})$ follows the first order markov property, (10) is further simplified as:

$$P(S_t|O_t, S_{1:t-1}) \propto P(S_t|S_{t-1})P(O_t|S_t) \quad (11)$$

where $P(S_t|S_{t-1})$ reflecting the temporal relations of shape modes is obtained from the codeword transition matrix in Section 2, and $P(O_t|S_t)$ is the likelihood function which estimates the similarity between the predicted shape mode and the current observations. We use $Sim_{(\Phi_c, \Phi_M)}(1)$ to approximate $P(O_t|S_t)$. Φ_c is the result obtained in the last step and Φ_M is the mean shape map associated with S_t .

3.3. Contour evolution with the shape constraint

After the optimal current shape mode S_t^* (9) is obtained, the shape Gaussian model of this shape mode is chosen as the shape constraint with which the contour is evolved continuously. The level set speed function with shape priors is formulated:

$$F_{shape} = \frac{(\Phi(x, y) - \Phi_M(A(x, y)))^2}{\sigma_M(A(x, y))^2} + \log \sigma_M(A(x, y)) \quad (12)$$

where A is the affine parameter for shape registration, Φ_M and σ_M are the Gaussian parameters of the shape model.

The contour is evolved to the final boundary with the overall speed F ($F = F_{curv} + F_{shape}$) in the normal direction:

$$\frac{\partial \Phi}{\partial t} = (F_{curv} + F_{shape}) \cdot |\nabla \Phi| \quad (13)$$

where $F_{curv} = \varepsilon \kappa(x, y)$ is the internal force proportional to the curvature $\kappa(x, y)$ of the contour, it has a smoothness effect on the contour. The initial value of Φ for this iteration equation is Φ_c obtained in Section 3.1. After a number of iterations, the final evolution result is returned to update the shape model of its shape mode.

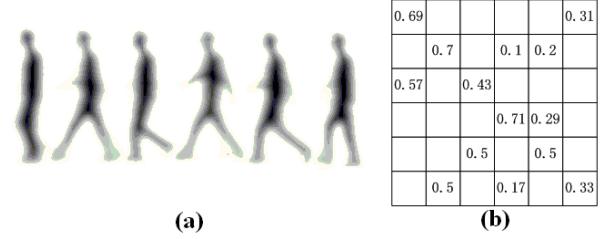


Fig. 2. Illustration of the shape model training: (a)The mean shapes of discovered six shape modes; (b) the codeword transition matrix

4. EXPERIMENTS

To verify the effectiveness of our method, we have performed experiments on synthetic and real video sequences. In the experiments, we track the contours of a walking person under some difficult cases, such as noise, partial occlusions and background clutter. A tracked object is represented with a white contour. At the shape model training stage, for each sequence, a set of hand-segmented contours of each walking person are used as the training samples. Through a DSC method, several dominant shape modes are discovered. For each shape mode, a Gaussian model is constructed at each grid location. The parameter β in (1) is set to be 8000. l in (7) is independent of sequences and is fixed to 2.

The first sequence is an outdoor human walking sequence. The forefront of this sequence is used to train the shape model and the latter part is the testing sequence. During the off-line training stage, the mean shape maps of discovered six modes with respect to this sequence are shown in Fig.2(a). The six modes are typical and representative during a periodic walking sequence. The codeword transition matrix is shown in Fig.2(b). We test our method under two conditions. The first condition is the image is disturbed by the salt&pepper noise. The second condition is the person is partially occluded by a red ellipse whose position is generated randomly around the tracking person. The final tracking results are illustrated in Fig.3. We can find for both the noise condition (Fig.3(a)) and partial occlusions condition (Fig.3(b)), our method keeps good track of the contours of the walking person. Therefore, the dynamical shape model is effective for recovering the missing parts of the tracked objects.

The second sequence is an indoor human walking sequence with mobile camera. The background is cluttered up with some stuffs which have the similar color as the object. Nine dominant shape modes are discovered by the DSC method and the codeword transition matrix is calculated similar to our first experiment. The tracking results are shown in Fig.4. Through the results only considering the color feature (Fig.4(a)), we can find the color feature is sensitive to the background disturbance. Compared with this condition, the tracking results fusing the color feature and the shape priors are robust to background disturbance (Fig.4(b)). Therefore,

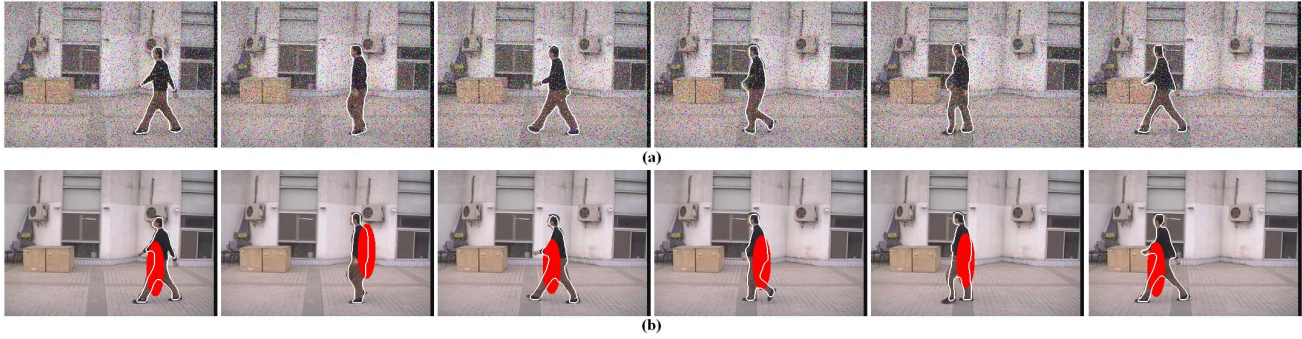


Fig. 3. Tracking results of synthetic sequence: (a) noise condition; (b) partial occlusions condition

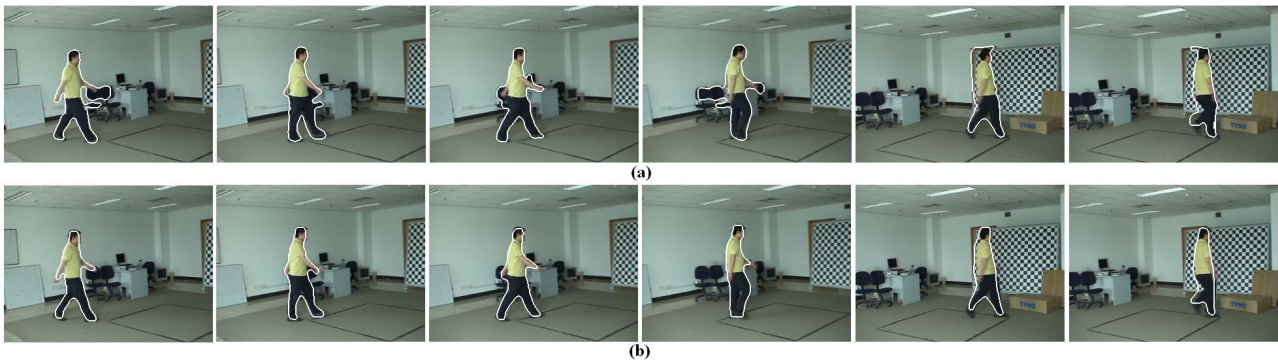


Fig. 4. Tracking results of real sequence: (a) tracking results only considering the color feature; (b) tracking results fusing color feature and dynamical shape priors

the consecutive shape changes are well modeled by our dynamical shape model.

5. CONCLUSIONS

We have proposed a markov model-based dynamical shape priors model in the level set tracking framework. The shape model encodes the temporal correlations of contours in a periodic action sequence and is a high-level understanding of shape knowledge. During the shape model construction process, a DSC method is used to discover the typical shape modes, the transition matrix among them is calculated to characterize the temporal correlations. During the tracking process, shape priors and the result only evolved with the color feature are fused to predict the current shape mode. With the help of shape priors, the contour is evolved to enclose the object correctly. The results presented have shown the effective tracking performance with noise, occlusions and background clutter.

6. ACKNOWLEDGMENTS

This work is partly supported by NSFC (Grant No. 60520120099 and 60672040) and the National 863 High-Tech R&D Program of China (Grant No. 2006AA01Z453).

7. REFERENCES

- [1] N. Paragios and R. Deriche, "Geodesic active contours and level sets for the detection and tracking of moving objects" in *IEEE Trans. PAMI*, vol.22, pp. 266–280, Mar. 2000.
- [2] N. Paragios and M. Rousson, "Shape priors for level set representations" in *Proc. of ECCV*, 2002, pp. 78–92.
- [3] A. Yilmaz, X. Li, and M. Shah, "Contour-based object tracking with occlusion handling in video acquired using mobile cameras" in *IEEE Trans. PAMI*, vol.26, pp. 1531–1536, Nov. 2004.
- [4] M. Leventon, E. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours" in *Proc. of CVPR*, 2000, vol.1, pp. 316–323.
- [5] D. Cremers, "Dynamical statistical shape priors for level set based tracking" in *IEEE Trans. PAMI*, vol.28, pp. 1262–1273, Aug. 2006.
- [6] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active shape models-their training and application" in *CVIU*, vol.61, pp. 38–59, Jan. 1995.
- [7] M. Fussenegger, P.M. Roth, H. Bischof, and A. Pinz, "On-line, incremental learning of a robust active shape model" *DAGM-Symposium*, pp. 122–131, 2006.
- [8] M. Pavan and M. Pelillo, "A new graph-theoretic approach to clustering and segmentation" in *Proc. of CVPR*, 2003, vol.1, pp. 18–20.
- [9] S. Osher and J. Sethian, "Fronts propagation with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulation" in *Journal of Comput. Phys.*, vol.79, pp. 12–49, 1988.
- [10] N. Paragios, M. Rousson, and V. Ramesh, "Matching distance functions: a shape-to-area variational approach for global-to-local registration" in *Proc. of ECCV*, 2002, pp. 775–789.