Improving the Robustness of Parametric Shape Tracking with Switched Multiple Models

Jacinto Nascimento and Jorge S. Marques

Torre Norte, IST/ISR, Av. Rovisco Pais, 1049-001 Lisboa, Portugal,
{pcjan@pop.,jsm@}isr.ist.utl.pt

Abstract. *This paper addresses the problem of tracking objects with complex motion dynamics or shape changes. It is assumed that some of the visual features detected in the image (e.g. edge strokes) are outliers i.e., they do not belong to the object boundary. A robust tracking algorithm is proposed which allows to efficiently track an object with complex shape or motion changes in clutter environments. The algorithm relies on the use of switched models, i.e., a bank of stochastic motion models switched according to a probabilistic mechanism. Robust filtering methods are used to estimate the label of the active model as well as the state trajectory.*

1 Introduction

Object tracking has various applications in the scope of medical diagnosis, surveillance and human-machine interface. It is usually assumed that the object shape and motion slowly vary during the observation interval, being described by a finite set of parameters.

Typically a stochastic difference equation is used to described the evolution of the shape and motion parameters. These assumptions are not valid in the presence of abrupt changes or complex parameter trajectories which can only be described by nonlinear dynamic systems. Two examples concern the estimation of the lips boundaries e.g., for speech recognition or face animation and the estimation or the human motion [4].

To deal with these difficulties, it was recently advocated the use of switched dynamical models i.e., a set of models switched according to a probabilistic rule, each of them being tailored to a specific motion regime or shape evolution [5–7]. Two problems have to be addressed if we want to use switched dynamic models for shape tracking. First, given a video sequence we have to determine which model is active at each instant of time. This is the labeling problem. Second, we have to estimate the state of the active model using the available data. This amounts to estimate the shape and motion parameters of the object to be tracked. These problems have been addressed either by non parametric techniques [5] or by parametric ones, based on the propagation of Gaussian mixtures [6, 7].

Although switched models are able to describe complex motion and shape evolution, they fail in the presence of outliers i.e., if the image measurements contain invalid data. Typically, the tracker loses the object boundary when wrong edge points are detected in the image, e.g., edge points belonging to the background or to inner regions of the object to be tracked. This is a major drawback which prevents the application of such models in many tracking problems. This difficulty is addressed here. A robust tracking algorithm is presented which extends the method described in [6, 7]. The proposed tracker is based on two key concepts. First, middle level features (strokes) are used instead of low level ones (edge points). Second, two labels (valid/invalid) are considered for each stroke. These techniques have been used with success when a single model is adopted to track the object [8]. The proposed algorithm allows a robust performance of the switched multiple model tracker in the presence of outliers.

2 Switched Dynamical Models

In order to estimate the object position and deformation, three steps are considered [2]: contour prediction, image measurement and contour update. The first step predicts the position of the object boundary in the next image. The second step computes image features in the vicinity of the predicted contour e.g., by sampling the predicted contour at equally spaced points. The third step uses the image measurements to update the contour estimate. It is assumed that image features (edges points) either belong to the boundary of the object to be tracked or they are produced by the background or inner edges (outliers). The main difficulty lies in the presence of false alarms or detection failures which produce undesirable effects. One way to deal with this situation is by considering that each feature is either valid or invalid. This approach is not practical since it involves 2^N hypothesis (data interpretations), N being the number of detected features (sometimes hundreds). We adopt a different approach to reduce the number of hypothesis. The edge points are linked in M strokes, and each stroke is classified either as valid or invalid. This reduces the number of hypothesis to 2^M, with $M \ll N$.

Let $x(t) \in \Re^n$ be a vector containing the shape parameters of the object to be tracked (e.g. control points of a spline curve). We assume that the state vector is generated by a set of a stochastic difference equations [9]

$$x(t) = A_{k(t-1),k(t)} x(t-1) + w(t) \tag{1}$$

where $w(t) \sim \mathcal{N}(0, Q_{k(t-1),k(t)})$ is a white Gaussian noise, $k(t) \in \{1, \ldots, m\}$ is the label of the active model at instant t and m is the number of steady state models (see Fig. 1). Additional models are considered at transitions (matrices A, Q depend on $k(t)$ and $k(t-1)$). It is assumed that the label sequence $k(t)$ is a first order Markov process with the transition probability

$$T_{rq} = p(k(t) = q \mid k(t-1) = r) \tag{2}$$

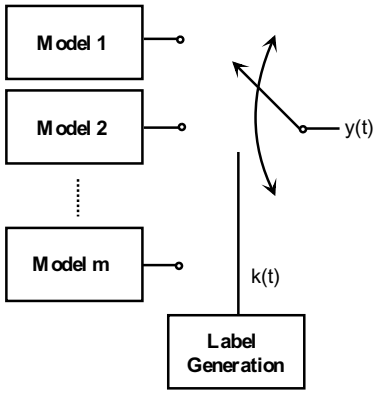

Fig. 1. Bank of switched models.

where $r, q \in \{1, \ldots, m\}$, and m the number of steady state models. Switched dynamic models were studied in control theory and aeronautics to deal with abrupt changes in dynamic systems (e.g., see [9], [3]). The application of these models in object tracking has been considered in [6], assuming that all the data points are valid. In this paper the available observations are the strokes detected in the image. However, we do not know which strokes belong to the object boundary and should therefore be considered as valid. Since this information is not available a label (valid/invalid) is assigned to each stroke and all the label sequences are considered. Each label sequence is denoted as a *data interpretation*. An interpretation I_i is defined as a binary sequence $I_i^1, \ldots, I_i^M$, where $I_i^j \in \{0, 1\}$ is the label of the j-th stroke in the i-th interpretation.

Let $y(t)$ be the vector of all image features detected at instant t and let $y_i(t)$ be a vector with all valid features according in the i-th interpretation, $(y_i(t) \subset y(t))$. It will be assumed that the sensor model for the i-th interpretation is given by

$$y_i(t) = C_i x(t) + \eta(t) \tag{3}$$

where $\eta(t) \sim \mathcal{N}(0, R_i)$ is a white Gaussian noise and C_i is the shape matrix associated to the i-th interpretation.

Fig.2 shows an example in which there are 2^3 interpretations. A possible interpretation is $I_i = (110)$. In this case, matrix C_i includes the rows associated with the indexes $\{b^1, ..., e^1, b^2, ..., e^2\}$ of the image features considered as true. The observation matrices C_i, C_j associated with two interpretations I_i,I_j are different since the observation vectors y_i, y_j contain different data features and they usually have different dimensions.

The state of a switched multiple model is characterized by the transition density $p(x(t), k(t) \mid x(t-1), k(t-1))$, which can be split as follows

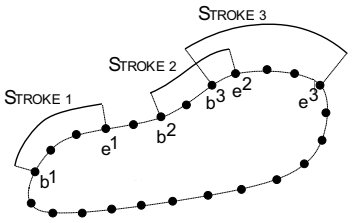

Fig. 2. Initial shape estimate and image strokes.

$$p(x(t),k(t) \mid x(t-1),k(t-1)) = p\big(x(t) \mid k(t),x(t-1),k(t-1)\big)\, p\big(k(t) \mid x(t-1),k(t-1)\big) \tag{4}$$

The first factor depends on the dynamic equation (1) while the second is an element of the transition matrix of the Markov chain $T_{k(t-1),k(t)}$.

3 Density Propagation

The problem to be solved can be formulated as follows: given a set of observations $Y^t = \{y(1), \ldots y(t)\}$ which may contain outliers, what is the best estimate of the state and model label $\hat{x}(t)$, $\hat{k}(t)$.

This is a nonlinear filtering problem. If the joint probability density function, conditioned on the observations is evaluated $p(x(t),k(t) \mid Y^t)$, estimates of the unknown parameters $(\hat{x}(t),\hat{k}(t))$ can be obtained by using the maximum *a posteriori* (MAP) method

$$(\hat{x}(t),\hat{k}(t)) = \arg \max_{x(t),k(t)} p(x(t),k(t) \mid Y^t) \tag{5}$$

Using the law of total probabilities, the *a posteriori* density becomes

$$p(x(t),k(t) \mid Y^t) = \sum_{K^{t-1}} c_{K^t} p(x(t) \mid K^t, Y^t) \tag{6}$$

where $c_{K^t} = p(K^t \mid Y^t)$ and $K^t = \{k(1), \ldots, k(t)\}$ is the model label sequence up to instant t. Since $p(x(t) \mid K^t, Y^t)$ is a Gaussian density, the joint density $p(x(t),k(t) \mid Y^t)$ defined in (6) is a mixture of Gaussians, each of them being associated to different label sequence K^t.

The computation of the mixture modes depends on the method being used. If all the observations are valid, each $p(x(t) \mid K^t, Y^t)$ (Gaussian component) can be updated by Kalman filtering and this is the optimal solution [6]. However, when $y(t)$ is contaminated with outliers, robust filtering methods must be adopted. In fact, assuming that the model sequence K^t is known, the mean and covariance

matrix can be computed using the S-PDAF method, recently proposed in [8], inspired in the work of Bar-Shalom and Fortmann [1] in the context of target tracking. To update the coefficients c_{K^t} in the switched model case, a new update law is required. After straightforward manipulation,

$$c_{K^t} = \gamma \, c_{K^{t-1}} \, T_{k(t-1)k(t)} \sum_i {}^k\alpha_i(t) \prod_{j=1}^{M} \prod_{n=b^j}^{e^j} {}^k\mathcal{E}_i^j(s_n, t) \tag{7}$$

where γ is the normalization constant; $c_{k(t-1)}$ is the predicted mixture coefficient; $T_{k(t-1)k(t)}$ is an element of the transition matrix of the Markov chain; $\alpha_i(t) = P(I_i(t) \mid Y^t)$ is the association probability assigned to the data interpretation $I_i(t)$; M is the number of strokes; b^j, e^j are the indexes of the j-th stroke; $\mathcal{E}$ is a normal or uniform distribution, depending on the stroke j being considered as valid/invalid in the interpretation $I_i(t)$.

The Kalman filter is a particular case of S-PDAF since a single model is used and all the data is considered as valid. Therefore, $\mathcal{E}$ becomes independent of j and i. In this case, equation (7) can be written in this case as

$$c_{K^t} = \gamma \, c_{K^{t-1}} \, T_{k(t-1)k(t)} \prod_{n=1}^{L} {}^k\mathcal{E}(s_n, t) \tag{8}$$

The mean and covariance of the state estimates are updated by the S-PDAF, (see [8] for details)

$$\hat{x}_{K^t} = \hat{x}(t \mid t-1) + \sum_{i=1}^{m_t} \alpha_i(t) K_i(t) \nu_i(t) \tag{9}$$

$$\begin{aligned} P_{K^t} = & \left[I - \sum_{i=1}^{m_t} \alpha_i(t) K_i(t) C_i \right] P(t \mid t-1) \\ & + \sum_{i=0}^{m_t} \alpha_i(t) x_i(t) x_i(t)^T - \hat{x}(t \mid t) \hat{x}(t \mid t)^T \end{aligned} \tag{10}$$

where $K_i(k)$, $\nu_i(k)$ are the Kalman gain and innovation associated to the interpretation $I_i(k)$. The filter defined in (6-10) will be denoted as *RMM robust multi-model tracker*.

The computation of (7,9,10) is organized in a tree structure, each branch being characterized by (see Fig. 3), x_{K^t}, P_{K^t} and c_{K^t}. The structure illustrated in Fig. 3 suggests that the number of leaves (Gaussian mixtures) increases as time goes by. Assuming that we have m label values, the mixture will have m^t modes at time t. It is crucial to limit the growth, in order to obtain a practical solution. Several strategies can be applied to achieve this goal, e.g., by using mode merging and elimination [6]. In this paper the second method is adopted by discarding the mixture components with small enough coefficients.

Let us now consider the estimation of the unknown variables $x(t)$, $k(t)$. The model label $K(t)$ is estimated by the MAP method as follows

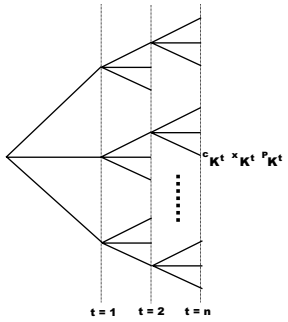

Fig. 3. Tree structure of RMM ($m = 3$).

$$\hat{k}(t) = \arg\max_q P\{k(t) = q \mid Y^t\} \tag{11}$$

$$= \arg\max_q \sum_{K^t : k(t)=q} c_{K^t} \int p(x(t) \mid K^t, Y^t) dx(t) \tag{12}$$

To estimate the state vector, the mean square error method was used instead for the sake of simplicity, leading to

$${}^q\hat{x}_{K^t} = \gamma \sum_{K^t : k(t)=q} c_{K^t} \sum_i {}^k\alpha_i(t) \ {}^k x_i(t \mid t) \tag{13}$$

The state estimate is a weighted sum of the estimates associated to the tree paths ending with the q label.

Shape representation and feature detection are performed as described in [8] and will not be discussed here.

4 Experimental Results

The RMM tracker was used in the estimation of the boundaries of objects with significant shape changes. An example of lip tracking is presented. A comparison between the proposed method and the Kalman multi-model (KMM) filter described in [7] is given.

Two dynamic models were used to describe lip motion. The first model (model 1) performs a vertical contraction of the object boundary using the boundary estimate computed at the previous frame. The second model (model 2) expands the object contour (see Fig. 4). The first model is tailored to describe the evolution of the lip contour while the mouth is closing while the second model is useful when the mouth is opening.

In Fig. 5, we can see the performance of the KMM, RMM trackers using the same input data. It is clear that the outliers produced by the nose and teeth introduce strong distortions in the shape estimates obtained with the KMM tracker,

Fig. 4. Contour prediction using models 1 (dots) and 2 (dashed line).

leading to useless results after a few number of frames. The robust tracker described in this paper overcomes this difficulties and exhibits good tracking performance in this experiment.

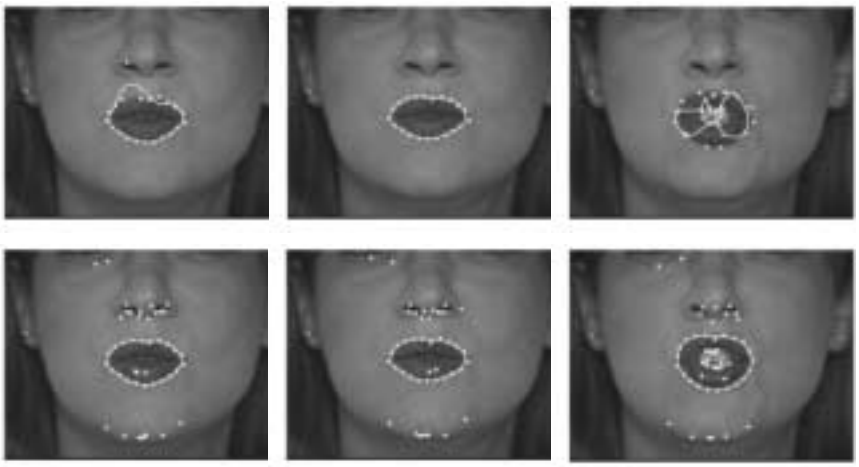

Fig. 5. Lip tracking with KMM tracker (first row, active model: 2 1 1) and RMM tracker (second row, active model: 2 1 2), (frames 8, 9, 13).

A more difficult situation is presented in Fig. 6. In this case, the KMM tracker loses the boundary of the lips and fails to estimate the correct dynamic model. Fig. 6 shows the results given by RMM filter (second row) showing a remarkable robustness with respect to outliers. We have even increased the search area during the feature detection phase, therefore allowing more outliers. The algorithm selects the expansion model in these frames since it is the one which

describes best the opening of the mouth. It is shown the robustness of the RMM even in the presence of a large number of clutter features.

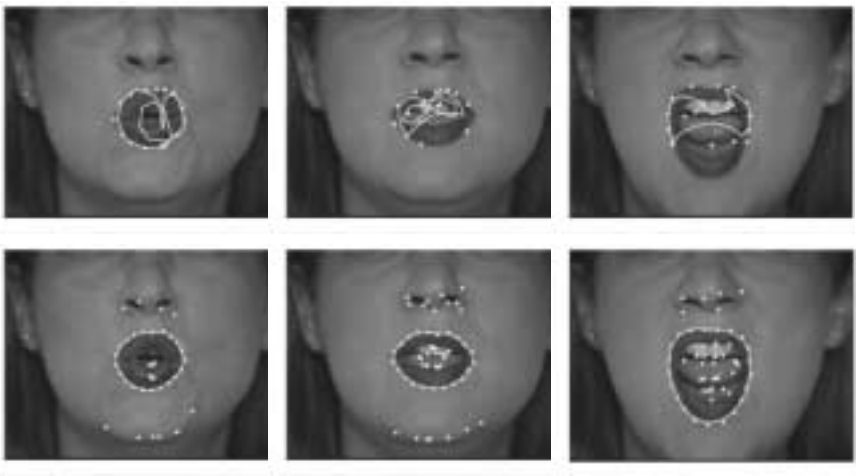

Fig. 6. Lip tracking with KMM tracker (first row, active model: 2 1 1) and RMM tracker (second row, active model: 2 2 2), (frames 16, 27, 46).

Figure 7 show the performance of the RMM algorithm in the presence of sudden shape changes. Three consecutive frames are shown in this figure. The use of multiple models allows to track sudden changes of motion or shape deformation. The expansion model is selected in this example to track the opening of the mouth. It is also displayed the predicted contours obtained by both models showing that the expansion model performs better in these three frames.

5 Conclusions

A new algorithm has been described for object tracking in video sequences. This algorithm allows the use of switched dynamic models, modeled by a bank of stochastic difference equations. Furthermore, it is assumed that the visual features detected in the image contain outliers, i.e., invalid features which do not belong to the object boundary. A robust filtering algorithm is proposed which is able to deal with multiple dynamics and invalid observations. This is accomplished by computing the propagation of the *a posteriori* density using Gaussian mixtures. Experimental results presented in the paper show that significant improvements are achieved, comparing to the results obtained by the Kalman MM filter which was recently proposed in [6, 7]. The algorithm was tested in lip tracking operations. It was experimentally observed that the proposed method

Fig. 7. Lip tracking with RMM: predicted contours (first row) and estimated contours (second row), (frames 45, 46, 47), active model: 2 2 2).

efficiently copes with the presence of abrupt shape changes and noisy measurements corrupted by outliers. This is clearly seen in some test sequences in which the mouth suddenly opens or closes. The RMM tracker still manages to estimate the lips contours well in these cases.

References

1. Bar-Shalom,T. Fortmann, "Tracking and Data Association" Academic Press, 1988.
2. A. Blake and M. Isard, "Active Contours". Springer, 1998.
3. C. Chang, M. Athans, "State Estimation for Discrete Systems with Switching Parameters", in *IEEE Trans. Aerospace Electr. Syst.* 14 (1978) 418-425.
4. A. Baumberg and D. Hogg, "Learning Deformable Models for Tracking the Human Body", in *Motion Based Recognition*, R. Jain, M. Sha Ed., pp. 39-60. Kluwer, 1997.
5. M. Isard and A. Blake, "A Mixed-State Condensation Tracker with Automatic Model-Switching", in *Int. Conference on Computer Vision*, pp. 107-112, 1998.
6. J. S. Marques, J. M. Lemos, "Shape tracking Based on Switched Dynamical Models", in *Proc. IEEE Int. Conf. on Image Processing*, pp. 954-958, Kobe, 1999.
7. J. S. Marques, J. M. Lemos, "Optimal and Suboptimal Shape Tracking Based on Switched Dynamic Models". Image and Vision Computing, to be published, 2001.
8. J. Nascimento, J. S. Marques, "Robust Shape Tracking in the Presence of Cluttered Background", in *Proc. IEEE Int. Conf. on Image Processing* vol. 3, pp. 82-85, Vancouver, 2000.
9. J. Tugnait, "Detection and Estimation for Abruptly Changing Systems", in *Automatica*, vol. 18, pp. 607-615, 1982.