Space-Time Tracking

Lorenzo Torresani and Christoph Bregler

Computer Science Department,
Stanford University,
Stanford, CA 94305, USA
{ltorresa, bregler}@cs.stanford.edu

Abstract. We propose a new tracking technique that is able to capture non-rigid motion by exploiting a space-time rank constraint. Most tracking methods use a prior model in order to deal with challenging local features. The model usually has to be trained on carefully hand-labeled example data before the tracking algorithm can be used. Our new model-free tracking technique can overcome such limitations. This can be achieved in redefining the problem. Instead of first training a model and then tracking the model parameters, we are able to derive trajectory constraints first, and then estimate the model. This reduces the search space significantly and allows for a better feature disambiguation that would not be possible with traditional trackers. We demonstrate that sampling in the trajectory space, instead of in the space of shape configurations, allows us to track challenging footage without use of prior models.

1 Introduction

Most of the tracking techniques that are able to capture non-rigid motion use a prior model. For instance, some human face-trackers use a pre-trained PCA model or parameterized 3D model, and fit the model to 2D image features. Combining these models with advanced sampling techniques (like particle filters or multiple hypothesis approaches) result in algorithms capable of overcoming many local ambiguities. There are many cases where a prior-model is not available. In fact, often the main reason for performing tracking is to estimate data that can be used to build a model. In this case, model-free feature trackers have to be used. Unfortunately many non-rigid domains, such as human motion, contain challenging features that make tracking without a model virtually impossible. Examples of such features are points with degenerate or 1D texture (points along lips and eye contours, cloth and shoe textures). We propose an innovative model-free tracking solution that can overcome such limitations. This can be achieved in redefining the tracking problem.

1. **Traditional Tracking**: *Given* $\mathbf{M} \rightarrow$ *Estimate* α*:*
   Standard model-based approaches assume a known (pre-trained) parameterized model $\mathbf{M}(\alpha)$. The model $\mathbf{M}$ stays constant over the entire time sequence (for example $\mathbf{M}$ might coincide with a set of basis-shapes). The parameters

A. Heyden et al. (Eds.): ECCV 2002, LNCS 2350, pp. 801–812, 2002.

α change from time frame to time frame (for example the interpolation coefficients between the basis shapes). Traditional tracking solves by estimating frame by frame the parameters $\alpha(1)..\alpha(F)$ that would fit $\mathbf{M}(\alpha)$ to the data.

2. **Reverse-Order Tracking**: *Estimate* $\alpha \rightarrow$ *Estimate* $\mathbf{M}$*:*
Our new technique first estimates the $\alpha(1)..\alpha(F)$ without knowing the model. Given the α parameters, it then estimates the model $\mathbf{M}$.

To reverse the order of computations we have to overcome two major hurdles: 1) How can the parameters α be estimated in a model-free fashion? 2) How can a model $\mathbf{M}$ be derived from the parameters α? We will also demonstrate why the reverse order tracking is advantageous over the traditional order.

Based on the assumption that the non-rigid motion can be factorized into a rigid motion and a blend-shape (non-rigid) motion component, we can establish a global low-rank constraint on the measurement-matrix (of the entire image sequence). This low-rank constraint allows us to estimate model-free all deformation parameters α over the entire image sequence. It is inspired by recent work by Irani [8] for the rigid case, and by extensions to non-rigid motion [17].

Since our tracking is based on direct image measurements, we have to minimize over a nonlinear error surface. In traditional tracking the search space grows with the number of time frames and the degrees of freedom of the model, and is therefore prone to many local minima (even with the use of nonlinear sampling techniques). With this new setup, and by fixing α over the entire sequence, we can show that the model estimation becomes a search in a very low-dimensional space. Sampling techniques in this small search space allow us to estimate models with high accuracy and overcome many local ambiguities that would exist in the traditional tracking framework.

We demonstrate this new technique on video recordings of a human face and of shoe deformations. Both domains have challenging (degenerate) features, and we show that the new algorithm can track densely all locations without any problem.

Section 3 describes the general rank constraint, and section 4 details how to exploit it for the two-step tracking. In section 5 we summarize our experiments and we conclude by discussing the results in section 6.

2 Previous Work

Many non-rigid tracking solutions have been proposed previously. As mentioned earlier, most methods use an a-priori model. Examples are [11,3,5,14,1,2,10]. Most of these approaches estimate non-rigid 2D motion, but some of them also recover 3D pose and deformations based on a 3D model.

What is most closely related to our approach, and in part inspired this solution, is work by Irani and Anandan [8,9] as well as methods for non-rigid decompositions [4,17], although these are either in the framework of rigid scenes, based on preexisting point tracks or related to Lucas-Kanade tracking.

3 Low-Rank Constraint for Non-rigid Motion

Our tracking algorithm relies on the assumption that the non-rigid 3D object motion can be approximated by a 3D rigid motion component (rotation and translation) and 3D non-rigid basis shape interpolations.

In this section we describe how we can justify a rank bound on the tracking matrix W without the prior knowledge of a specific object model. This gives an important insight on the roles of two matrices Q and M resulting from the decomposition of the tracking matrix. Section 4 shows how this decomposition is used in the tracking process.

3.1 Matrix Decomposition

The tracking matrix W describes the dense optical flow of P pixel or the tracks of P feature points over a sequence of F video frames:

$$W = \begin{bmatrix} \mathbf{U}^{\mathbf{F}\times\mathbf{P}} \\ \hline \mathbf{V}^{\mathbf{F}\times\mathbf{P}} \end{bmatrix} \tag{1}$$

Each row of $\mathbf{U}$ holds all x-displacements of all P locations for a specific time frame, and each row of $\mathbf{V}$ holds all y-displacements for a specific time frame. It has been shown that if $\mathbf{U}$ and $\mathbf{V}$ describe a 3D rigid motion, the rank of $[\frac{\mathbf{U}}{\mathbf{V}}]$ has an upper bound, which depends on the assumed camera model [16,8]. This rank constraint derives from the fact that $[\frac{\mathbf{U}}{\mathbf{V}}]$ can be factored into two matrices: $Q \times M$. $Q^{2F\times r}$ describes the relative pose between camera and object for each time frame, and $M^{r\times P}$ describes the 3D structure of the scene which is invariant to camera and object motion.

In previous works we have shown that a similar rank constraint holds also for the motion of deforming objects [4,17]. Assuming the non-rigid 3D deformations can be approximated by a set of K modes of variation, the 3D shape of a specific object configuration can be expressed as a linear combination of K basis-shapes $(S_1, S_2, ...S_K)$. Each basis-shape S_i is a $3 \times P$ matrix describing the 3D positions of P points for a specific "key" shape configuration of the object [1]. A deformation can be computed by linearly interpolating between basis-shapes: $S = \sum_k l_k S_k$.

Assuming weak-perspective projection, at a specific time frame t the P points of a non-rigid shape S are projected onto 2D image points $(u_{t,i}, v_{t,i})$:

$$\begin{bmatrix} u_{t,1} \ ... \ u_{t,P} \\ v_{t,1} \ ... \ v_{t,P} \end{bmatrix} = R_t \cdot \left(\sum_{i=1}^{K} l_{t,i} \cdot S_i \right) + T_t \tag{2}$$

where R_t contains the first two rows of the full 3D camera rotation matrix, and T_t is the camera translation. The weak perspective scaling (f/Z_{avg}) of the

[1] We want to emphasize that no prior knowledge of the basis-shapes of the object will be assumed by the method: the K unknown key-shapes will instead be implicitly estimated by the tracking algorithm.

projection is implicitly coded in $l_{t,1}, ...l_{t,K}$. As in [16], we can eliminate T_t by subtracting the mean of all 2D points, and henceforth can assume that S is centered at the origin.

We can rewrite the linear combination in (2) as a matrix multiplication:

$$\begin{bmatrix} u_{t,1} \ ... \ u_{t,P} \\ v_{t,1} \ ... \ v_{t,P} \end{bmatrix} = \begin{bmatrix} l_{t,1}R_t \ ... \ l_{t,K}R_t \end{bmatrix} \cdot \begin{bmatrix} S_1 \\ S_2 \\ ... \\ S_K \end{bmatrix} \tag{3}$$

We stack all point tracks from time frame 1 to F into one large tracking $2F \times P$ matrix W. Using (3) we can write:

$$W = \underbrace{\begin{bmatrix} l_{1,1}R_1 \ ... \ l_{1,K}R_1 \\ l_{2,1}R_2 \ ... \ l_{2,K}R_2 \\ ... \\ l_{F,1}R_F \ ... \ l_{F,K}R_F \end{bmatrix}}_{Q} \cdot \underbrace{\begin{bmatrix} S_1 \\ S_2 \\ ... \\ S_K \end{bmatrix}}_{M} \tag{4}$$

Since Q is a $2F \times 3K$ matrix and M is a $3K \times P$ matrix, in the noise free case W has at most rank $r \leq 3K$.

Beyond the important derivation of the rank constraint, the analysis above allows us to conclude that it is possible to separate the components of non-rigid motion, represented in Q in form of rotation matrices and deformation coefficients, from the structure (basis-shape model) of the object stored in M.

In the following sections we will show how we can take advantage of this decomposition to derive a solution to the tracking problem.

4 Tracking Algorithm

Section 3 described how W can be decomposed into a matrix Q and M. We do not have the tracking matrix W yet, this is the goal of our tracking algorithm. As outlined in section 1, we achieve this in reversing the traditional tracking order: first estimating the motion parameters Q, and then estimating the model M that fits the data.

4.1 Non-rigid Motion Estimation

The motion parameters are the interpolation coefficients $l_{t,k}$ and the rotation matrices R_t that are coded in the $Q^{2F \times r}$ matrix of equation (4).

It is possible to estimate Q without the full availability of W. It is based on the following observation (inspired by [8]): if the rank of $W^{2F \times P}$ is r, a subset of r point tracks (non-degenerate columns in W) will span the remaining tracks of W.

We reorder W into a set of m known "reliable" tracks W_{rel}, and a set of n unknown "unreliable" tracks W_{unrel} ($n = P - m$):

$$W^{2F\times P} = [W_{rel}^{2F\times m}|W_{unrel}^{2F\times n}] = Q^{2F\times r}\cdot[M_{rel}^{r\times m}|M_{unrel}^{r\times n}] \tag{5}$$

Usually r is significantly smaller than P and it is easy to find at least $m > r$ reliable tracks [17] that can be computed for the entire image sequence. Assuming W_{rel} is of rank r, we can estimate a matrix $\hat{Q}^{2F\times r}$ with following factorization:

$$W_{rel}^{2F\times m} = \hat{Q}^{2F\times r}\cdot\hat{M}_{rel}^{r\times m} \tag{6}$$

The factorization is not uniquely defined. Any invertible matrix $G^{r\times r}$ defines another valid solution: $\hat{Q}_2 = \hat{Q}\cdot G$ and $\hat{M}_{rel2} = G^{-1}\hat{M}_{rel}$. Since W_{rel} has rank r, the original matrix Q for the full W matrix (in equation (4) and (5)) is also related to $\hat{Q}$ with an invertible matrix G: $Q = \hat{Q}\cdot G$. In [4,17] several methods are described for how to calculate G (specifically for non-rigid reconstruction). In the context of our tracking problem here, we do not have to know G, since it does not change W. We only need to know that $\hat{Q}$ and Q are related by some (unknown) invertible G. The model M that we obtain in section 4.2 is then just multiplied by G^{-1} to get a correct 3D model (using results from [17]).

We calculate $\hat{Q}$ and $\hat{M}_{rel}$ in the standard way with SVD:

$$svd(W_{rel}) = U\cdot S\cdot V^T = \underbrace{U\cdot C}_{\hat{Q}}\cdot\underbrace{C\cdot V^T}_{\hat{M}_{rel}} \tag{7}$$

where C is the upper $r\times r$ sub-block of $\sqrt{S}$.

This factorization gives us (up to unknown G) the motion parameters $\hat{Q}$ and the 3D shape coordinates of the reliable points $\hat{M}_{rel}$.

4.2 Non-rigid Shapes Estimation

The task we have left is that of estimating the shape elements in the $r\times n$ matrix M_{unrel}.

The corresponding image features for point i has degenerate texture, and no reliable feature extraction and tracking schema for those points is assumed to work. Even probabilistic point tracks (as assumed in [9]) with uncertainty measures are not available. Nonlinear probabilistic trackers (e.g. particle filter-based) fail since the density for each feature location has too much spread or is a-priori uniformly distributed. Obviously in those cases, a known basis-shape model would help dramatically and would constrain the possible feature locations in a reliable way, but we do not have such a model M_{unrel} yet.

Let us consider a single column m_i of M_{unrel} which represents the spatial (unknown) positions of point i for the K main deformations of the object. It turns out that if we multiply $\hat{Q}$ with m_i, we obtain an r-dimensional family of image trajectories $w_i = \hat{Q}\cdot m_i$. Now we have a very strong parameterized model for the unreliable point, but along the time axis instead of the space axis: the low-dimensional variability is expressed as temporal variations of the

trajectory curve of a single point across time. In a sense we have a very accurate high resolution "dynamical model" of the point, without having the "kinematic constraints" in place yet. The probability of each possible trajectory in this constrained (r-dimensional) subspace can be computed much more reliably from the image data. As local texture might be ambiguous in a single frame, it is unique across the entire sequence of F frames, if constrained in the trajectory subspace. For instance, the famous aperture problem of 1D texture vanishes in this framework, as noted by [8] already.

Sampling in trajectory space. Since we only use the image sequence itself as features, the probability density for our trajectory family is nonlinear. Also any possible initialization heuristic for the trajectory of the unknown point track i might be far of. We therefore adopt a stochastic estimation technique based on factored sampling [7] to find the most likely values for m_i.

Factored sampling is a Monte Carlo technique for estimation of conditional probability densities. Let us assume we are trying to estimate the function $p(\mathbf{X}|\mathbf{Z})$ where $\mathbf{X}$ and $\mathbf{Z}$ are continuous random variables statistically related by some unknown dependency. When $p(\mathbf{X}|\mathbf{Z})$ cannot be sampled directly but the function $p(\mathbf{Z}|\mathbf{X} = x)$ can be computed for any x, factored sampling proposes the following recipe. Let $x_1, ..., x_N$ be N samples drawn from the prior $p(\mathbf{X})$ and let us generate a third random variable $\mathbf{Y}$ by choosing samples $y = x_s$ at random with probabilities

$$p_s = \frac{p(\mathbf{Z}|\mathbf{X} = x_s)}{\sum_{k=1}^{N} p(\mathbf{Z}|\mathbf{X} = x_k)} \tag{8}$$

It has been shown that the probability distribution of $\mathbf{Y}$ tends to $p(\mathbf{X}|\mathbf{Z})$ as $N \to \infty$ [7]. An approximation of the mean of the posterior can be computed as

$$E[\mathbf{X}|\mathbf{Z}] = \sum_{s=1}^{N} x_s p_s \tag{9}$$

In our case $p(\mathbf{X}|\mathbf{Z}) = p(\mathbf{m_i}|\mathbf{Z})$ where $\mathbf{Z}$ are measurements derived from the image sequence. We evaluate each hypothesis (sample) $m_i^{(s)}$ for $\mathbf{m_i}$ by computing for each frame f the sum of squared differences between a small window around the point i in the reference frame and the corresponding window in frame f translated according to $m_i^{(s)}$:

$$p(\mathbf{Z}|\mathbf{m_i} = m_i^{(s)}) \propto exp\{-\sum_{f=1}^{F} \sum_{(x,y)\epsilon ROI_i} \frac{(I_0(x,y) - I_f(x + u_f^{(i,s)}, y + v_f^{(i,s)}))^2}{2\sigma^2}\} \tag{10}$$

where $u_f^{(i,s)} = q_u^{(f)} \cdot m_i^{(s)}$, $v_f^{(i,s)} = q_v^{(f)} \cdot m_i^{(s)}$, and $q_u^{(f)}$, $q_v^{(f)}$ are the f-th rows of Q_u and Q_v, with $\hat{Q} = [\frac{\mathbf{Q_u}}{\mathbf{Q_v}}]$.

Alternatively, the outputs of a set of steerable filters tuned to a range of orientations and scales could be used to compare image patches [6,13].

At the end of this density estimation process, each column m_i of M_{unrel} is computed as the expected value of the posterior using equation (9).

Note again how each hypothesis is tested against all the frames of the sequence. This large amount of measurements per sample is the key reason of the robustness of this approach.

The speed of convergence to the posterior density depends on how well the samples $m_i^{(s)}$ are chosen with respect to the unknown distribution $p(\mathbf{m_i}|\mathbf{Z})$. Also, evaluating the likelihood is computationally expensive. Ideally we would like to draw the $m_i^{(s)}$ from areas where the likelihood $p(\mathbf{Z}|\mathbf{m_i})$ is very large instead of wasting computational resources on samples with negligible p_s, that are clearly of little contribution for a first approximation of the unknown density. This is the intuitive idea underlying the theory of importance sampling [15].

Suppose we are given an auxiliary function $g_i(x)$ describing the areas of the random variable space that are believed to better characterize the posterior. Now we can use the importance function $g_i(x)$ to draw the samples $m_i^{(s)}$ and thus achieve faster convergence. We simply need to introduce a correction term in equation (8) to reflect our use of a different sampling distribution:

$$p_s = \frac{p(\mathbf{Z}|\mathbf{X} = x_s)p(x_s)/g(x_s)}{\sum_{k=1}^{N} p(\mathbf{Z}|\mathbf{X} = x_k)p(x_k)/g(x_k)} \tag{11}$$

In this case we define the importance function $g_i(x)$ by assuming the object has smooth surface[2]: we compute μ_{M_i}, for $i = 1, ..., n$, by interpolation from the shapes of the reliable points. $g_i(x)$ is then defined as a Gaussian around μ_{M_i}:

$$g_i(x) = \frac{1}{(2\pi)^{r/2}|\Sigma_i|^{1/2}} exp\{-\frac{1}{2}(x - \mu_{M_i})^T \Sigma_i^{-1}(x - \mu_{M_i})\} \tag{12}$$

We apply this technique for each of the n unreliable m_i separately.

Again, it is important to emphasize the fact that by sampling in the space of deformations as opposed to estimating the inter-frame relationships, we fully exploit the information in the sequence by evaluating each sample on ALL of the images. Whereas for conventional approaches the number of frames represents the number of sub-problems to solve, with this technique images are only measurements. The longer the sequence the more data we have available to constrain the solution.

5 Experiments

The method presented in this article has been tested on two separate video recordings of human motion with different types of deformations. The first sequence was originally employed in [17] to evaluate the algorithm of tracking and

[2] The smoothness assumption is clearly violated at depth discontinuities, but this hypothesis is not critical for the convergence of the method. It is used here only to speed up the process of density estimation.

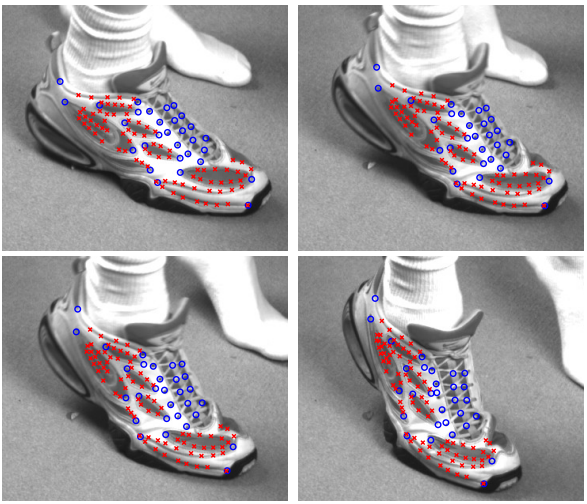

Fig. 1. Example tracks of the shoe sequence. The blue circles are the reliable points utilized to derive Q. The red crosses are edge features: their optical flow is determined by estimating the model M.

non-rigid 3D reconstruction described in that article. Here we show the improved tracking performance that is achievable with this new approach. The video consists of a 500 frames-long sequence of a shoe undergoing very large rotations and non-rigid deformations. While the reliable points are features with 2D texture tracked using the technique of Lucas-Kanade [12], the other 80 points are edges or degenerate features whose optical flow cannot be recovered using local operators. In order to prevent drifting along edges and 1D texture the solution

Fig. 2. Samples distribution. The shape samples are reprojected onto the images according to the occurring non-rigid motion. This is reflected in the sample distributions in the two frames: the principal motion is eyebrow deformation in the first image and head rotation in the second.

in [17], based on an integration of the rank constraint and the Lucas-Kanade linearization, had to be augmented with a regularization term enforcing smooth flow among neighboring features. We have tried to run the tracking solution presented in this article on the shoe sequence making sure that parameters ($r = 9$) and feature points were the same as those used in the original experiment. Although we are now not using any spatial smoothness heuristic, we found the results derived with the new technique even more accurate than those presented in the previous work. Figure 1 shows some examples of the non-rigid motion characterizing the sequence, together with the tracks recovered by our new solution based on sampling trajectories with rank constraints.

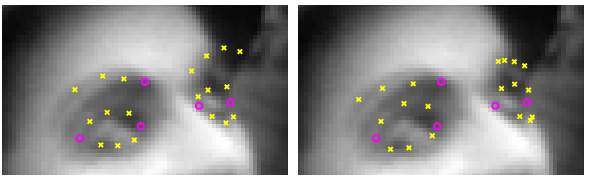

Fig. 3. Eyebrow detail. The left figure shows the initialization of the tracking with shapes extrapolated from the reliable points. The final optical flow recovered by trajectory sampling is shown on the right.

The second group of experiments was carried out on equally challenging data. We extracted 400 frames of digital footage from a Charlie Chaplin film originally recorded in 1918. The old technology used for the recording as well the non-uniform light of the outdoor scene make the sequence very difficult for a tracking task. We focused on a segment of the movie containing a close-up of the actor with the goal of capturing his famous facial expressions. The rigid head motion is very large and causes considerable changes in the features appearance. The non-rigid deformations are similarly extreme and mostly due to eyebrow, lips and jaw motion. We restricted the tracking to important facial features such as eyebrows, eyes and the moustache: most of these points are edges and they are virtually impossible to track with conventional model-free techniques without incurring in features drifting. We could find only 9 features that could be tracked over the entire sequence with the technique of Lucas-Kanade employing affine transformations for the local windows centered at these points. The locations of these reliable points are marked with pink circles in figure 2. These points proved to be anyway sufficient for capturing the main modes of deformations of the face. All the results on this footage reported in the article were obtained with $r = 4$ or $r = 5$. The tracks of these points were used to estimate Q and to consequently recover the main elements of non-rigid motion. Figure 2 shows the reprojections onto the image of the shape samples for different non-reliable features in example frames. It is very apparent how the search for the best feature match takes place along the direction of the occurring non-rigid motion encoded in the Q matrix. The samples in the left frame are spread vertically because

Fig. 4. Tracking of degenerate features on the Chaplin footage. The pink circles are reliable points.

of the large deformation along that direction picked up by the reliable point located on the left eyebrow. Similarly the sample distributions in the right image match the occurring head rotation. This experimentally validates the motion-shape decomposition expressed by equation (4) and shows the reduced size of the search space.

The shape initialization extrapolated from the reliable points provides only a very rough initial approximation, especially in frames where the non-rigid component of the motion is large. The optical flow resulting from the initialization that is shown on the left of figure 3 misses completely to capture deformation and correct 3D position of the right eyebrow. This initialization is not critical for the convergence of the algorithm as exemplified in the right image.

Using our unconventional approach we could track reliably and accurately all of the 29 degenerate features throughout the whole sequence of 400 frames, a task extremely difficult to achieve without the use of prior models. Example frames taken from the footage together with their estimated tracks are shown in figure 4. See `http://movement.stanford.edu/nonrig` for mpeg or quicktime video showing the results of tracking on the entire sequence.

The tracking algorithm described in this article was implemented in interpreted Matlab code and tested on a PC Pentium 3 with a 900MHz CPU and 512M RAM. The number of samples per feature point was chosen to be 500 for all the experiments presented. With this setup the average time to estimate the trajectory of one feature in a video of 400 frames is 2-3 seconds.

6 Discussion

We have presented a new algorithm that is able to track non-rigid object motion and disambiguate challenging local features without the use of a prior model. The key idea of the algorithm is the exploitation of a low-rank constraint that allows us to perform a sampling-based search over trajectory subspaces instead of over the space of shape configurations. From this result we derive a robust solution for tracking without models.

We demonstrated the algorithm on tracking of film footage of Charlie Chaplin's facial expressions, and of sport-shoe deformations. In both cases, model-free point trackers are able to track only a few corner features. With the use of this new technique we were able to track challenging points suffering from the aperture problem and other degenerate features.

Acknowledgements. We would like to thank Hrishikesh Deshpande for providing advice and help with the experiments. This research was funded in part by the National Science Foundation and the Stanford BIO-X program.

References

1. M.J. Black and Y. Yacoob. Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion. In *ICCV*, 1995.

2. M.J. Black, Y.Yacoob, A.D.Jepson, and D.J.Fleet. Learning parameterized models of image motion. In *CVPR*, 1997.
3. A. Blake, M. Isard, and D. Reynard. Learning to track the visual motion of contours. In *J. of Artificial Intelligence*, 1995.
4. C. Bregler, A. Hertzmann, and H. Biermann. Recovering Non-Rigid 3D Shape from Image Streams. In *CVPR*, 2000.
5. D. DeCarlo and D. Metaxas. Deformable model-based shape and motion analysis from images using motion residual error. In *ICCV*, 1998.
6. W. Freeman and E Adelson. The design and use of steerable filters. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1991.
7. U. Grenander, Y. Chow, and D.M. Keenan. *HANDS. A Pattern Theoretical Study of Biological Shapes*. Springer Verlag. New York, 1991.
8. M. Irani. Multi-frame optical flow estimation using subspace constraints. In *ICCV*, 1999.
9. M. Irani and P. Anandan. Factorization with uncertainty. In *ECCV*, 2000.
10. M. Isard and A. Blake. Contour tracking by stochastic propagation of conditional density. In *ECCV*, 1996.
11. A. Lanitis, Taylor C.J., Cootes T.F., and Ahmed T. Automatic interpretation of human faces and hand gestures using flexible models. In *International Workshop on Automatic Face- and Gesture-Recognition*, 1995.
12. B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *Proc. 7th Int. Joint Conf. on Artif. Intell.*, 1981.
13. P. Perona. Deformable kernels for early vision. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1995.
14. F. Pighin, D. H. Salesin, and R. Szeliski. Resynthesizing facial animation through 3d model-based tracking. In *ICCV*, 1999.
15. B.D. Ripley. *Stochastic Simulation*. Wiley. New York, 1987.
16. C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *Int. J. of Computer Vision*, 9(2):137–154, 1992.
17. L. Torresani, D.B. Yang, E.J. Alexander, and C. Bregler. Tracking and Modeling Non-Rigid Objects with Rank Constraints. In *CVPR*, 2001.