NON-RIGID 3D SHAPE RECOVERY USING STEREO FACTORIZATION

Alessio Del Bue *Lourdes Agapito*

Department of Computer Science
Queen Mary, University of London
London, E1 4NS, U.K.
{alessio,lourdes}@dcs.qmul.ac.uk

ABSTRACT

In this paper we address the problem of recovering 3D non-rigid structure from a sequence of images taken with a stereo pair. We have extended existing non-rigid factorization algorithms to the stereo camera case and presented an algorithm to decompose the measurement matrix into the motion of the left and right cameras and the 3D shape, represented as a linear combination of basis-shapes. The added constraints in the stereo camera case is that both cameras are viewing the same structure and that the relative orientation between both cameras is fixed. Our focus in this paper is on the recovery of flexible 3D shape rather than on the correspondence problem. We propose a method to compute reliable 3D models of deformable structure from stereo images. Our experiments with real data show that improved reconstructions can be achieved using this method.

1. INTRODUCTION

Recent work in non-rigid factorization [1, 2, 3] has proved that it is possible under weak perspective viewing conditions to infer the principal modes of deformation of an object alongside its 3D shape within a structure from motion estimation framework. These non-rigid factorization methods stem from Tomasi and Kanade's factorization algorithm for rigid structure [4], developed in the early 90's. The key idea of this algorithm is the use of rank-constraints to express the geometric invariants present in the data. This allows the factorization of a measurement matrix – which contains the image coordinates of a set of features tracks – into its shape and motion components.

In the case of non-rigid factorization, the 3D shape recovered by the algorithms is represented as a linear combination of a number of detected modes of deformation. These models can subsequently be used as compact representations of the objects suitable for use in tracking [5], animation or other analysis.

The authors would like to thank the Royal Society European Science Exchange Programme for financial support and Enrique Muñoz, José Miguel Buenaposada and Luis Baumela for useful comments.

There have been other computer vision systems able to build similar morphable 3D models of non-rigid objects. However, most of them rely on having additional information — for instance depth estimates available from 3D scanning devices [6] or multiview reconstruction [7] — or have been specialised to the specific object under observation: for example physically-based human face models [8].

Crucially, the new factorization methods work purely from video in an unconstrained case: a single uncalibrated camera viewing an arbitrary 3D surface which is moving and articulating. The 2D point tracks needed as input data by the algorithm can be obtained initially using a local feature tracker provided the patch around the feature has high texture content (corner features). Alternatively, robust optic flow can also be obtained in areas of the image with low texture by exploiting the rank constraint [3, 2], an approach inspired by its rigid equivalent [9].

In this paper we have extended the non-rigid factorization algorithm to the multiple camera case. More specifically, we have formulated the problem for a stereo rig, where the two cameras remain fixed relative to each other throughout the sequence. In this case the measurement matrix requires not only the temporal tracks of points in the left and right image sequences but also the stereo correspondences between left and right image pairs. Assuming that both cameras are synchronized and that the stereo correspondences are known, the measurement matrix may be factorized into the left and right motion matrices and the 3D non-rigid shape.

Note that our work requires the temporal and spatial correspondences to be known. Previous work on non-rigid factorization for a single camera has mainly concentrated its efforts in solving the temporal tracking problem, exploiting the rank constraints to obtain correspondences between frames [3, 2]. However, our focus here is on the recovery of flexible 3D shape rather than on the correspondence problem. The results of our experiments – where matching is aided by the use of markers – show that using a stereo framework gives improved 3D reconstructions.

The paper is organized as follows. In section 2 we describe the use of rank constraints to compute motion and

3D shape within the factorization framework. We first outline the factorization algorithm in the rigid case and then describe the existing non-rigid factorization algorithms for a single camera. We then formulate non-rigid factorization for the case of multiple cameras in section 3 and describe an algorithm that imposes the extra constraints. Finally in section 4 we present some experimental results on real image sequences showing improved 3D reconstructions.

2. BACKGROUND: FACTORIZATION

2.1. Rigid factorization

Tomasi and Kanade's factorization algorithm [4] for rigid structure provides a maximum likelihood estimate for affine structure and motion under the assumption of isotropic gaussian noise. The key idea is to gather the 2D image coordinates of a set of P points tracked throughout F frames into a measurement matrix $\mathtt{W}_{2F\times P}$. Assuming affine viewing conditions, the measurement matrix can be expressed analytically as a product of two matrices: $\mathtt{W} = \mathtt{MS}$ where $\mathtt{M}$ is a 2F×3 motion matrix which expresses the pose of the camera and $\mathtt{S}$ is the 3×P shape matrix which contains 3D locations of the reconstructed scene points. Therefore the rank of the measurement matrix is constrained to be $r \leq 3$. This constraint can be easily imposed by taking the Singular Value Decomposition of the measurement matrix and truncating it to rank 3: $SVD_3(\mathtt{W}) = U_{2F\times 3}D_{3\times 3}V_{3\times P} = M_{2F\times 3}S_{3\times P}$. In this way the image measurement matrix can be factorized into its motion and shape components.

2.2. Non-rigid motion: the single camera case

Tomasi and Kanade's factorization algorithm has recently been extended to the case of non-rigid deformable 3D structure [1, 2, 3]. Here, a model is needed to express the deformations of the 3D shape in a compact way. The chosen model is a simple linear model where the 3D shape of any specific configuration of a non-rigid object is approximated by a linear combination of a set of K basis-shapes which represent the K principal modes of deformation of the object for P points. A perfectly rigid object would correspond to the situation where K=1. Each basis-shape $(\mathtt{S}_1, \mathtt{S}_2, ..., \mathtt{S}_K)$ is a 3 × P matrix which contains the 3D locations of P object points for that particular mode of deformation. The 3D shape of any configuration can then be expressed as a linear combination of the basis-shapes $\mathtt{S}_i$:

$$\mathtt{S} = \sum_{i=1}^{K} l_i \mathtt{S}_i \qquad \mathtt{S}, \mathtt{S}_i \in \Re^{3\times P} \quad l_i \in \Re$$

where l_i are the deformation weights. If we assume a scaled orthographic projection model for the camera, the coordinates of the 2D image points observed at each frame f are related to the coordinates of the 3D points according to the following equation:

$$\mathtt{W}_f = \begin{bmatrix} u_{f,1} & \dots & u_{f,P} \\ v_{f,1} & \dots & v_{f,P} \end{bmatrix} = \mathtt{R}_f \left(\sum_{i=1}^{K} l_{f,i} \mathtt{S}_i \right) + \mathbf{T}_f \quad (1)$$

where $\mathtt{R}_f$ is a 2×3 matrix which contains the first and second rows of the camera rotation matrix and $\mathbf{T}_f$ contains the first two components of the camera translation vector. Weak perspective is a good approximation when the depth variation within the object is small compared to the distance to the camera. The weak perspective scaling (f/Z_{avg}) is implicitly encoded in the $l_{f,i}$ coefficients. We may eliminate the translation vector $\mathbf{T}_f$ by registering image points to the centroid in each frame. If the same P points can be tracked throughout an image sequence we may stack them into a 2F×P measurement matrix $\mathtt{W}$ and we may write:

$$\mathtt{W} = \begin{bmatrix} \mathtt{W}_1 \\ \vdots \\ \mathtt{W}_F \end{bmatrix} = \begin{bmatrix} l_{1,1}\mathtt{R}_1 & \dots & l_{1,K}\mathtt{R}_1 \\ \vdots & & \vdots \\ l_{F,1}\mathtt{R}_F & \dots & l_{F,K}\mathtt{R}_F \end{bmatrix} \begin{bmatrix} \mathtt{S}_1 \\ \vdots \\ \mathtt{S}_K \end{bmatrix} = \mathtt{MS} \quad (2)$$

Since $\mathtt{M}$ is a 2F×3K matrix and $\mathtt{S}$ is a 3K×P matrix, the rank of $\mathtt{W}$ when no noise is present must be $r \leq 3K$. Note that, in relation to rigid factorization, in the non-rigid case the rank is incremented by three with every new mode of deformation.

2.3. Non-rigid factorization

The rank constraint on the measurement matrix $\mathtt{W}$ can be easily imposed by truncating the SVD of $\mathtt{W}$ to rank 3K. This will factor $\mathtt{W}$ into a motion matrix $\tilde{\mathtt{M}}$ and a shape matrix $\tilde{\mathtt{S}}$. However, the result of the factorization of $\mathtt{W}$ is not unique since any invertible 3K×3K matrix $\mathtt{Q}$ can be inserted in the decomposition leading to the alternative factorization $\mathtt{W} = (\tilde{\mathtt{M}}\mathtt{Q})(\mathtt{Q}^{-1}\tilde{\mathtt{S}})$. The problem is to find a transformation matrix $\mathtt{Q}$ that imposes the replicated block structure on the motion matrix $\tilde{\mathtt{M}}$ shown in (2) and that removes the affine ambiguity upgrading the reconstruction to a metric one. The block structure is not required if we only wish to determine image point motion, but is crucial for the recovery of 3D shape and motion.

2.3.1. Computing the transformation matrix $\mathtt{Q}$

Whereas in the rigid case the problem of computing the transformation matrix $\mathtt{Q}$ to upgrade the reconstruction to a metric one can be solved linearly [4], in the non-rigid case imposing the appropriate repetitive structure to the motion matrix $\tilde{\mathtt{M}}$ results in a non-linear problem. A satisfactory solution is yet to be found for this problem but various methods have been proposed so far in the literature [2, 3].

One of the approaches proposed by Brand [2] consists of correcting each column triple independently applying the rigid metric constraint to each $\tilde{\mathtt{M}}^k_{2F\times 3}$ vertical block in $\tilde{\mathtt{M}}$ shown here:

$$\tilde{\mathtt{M}}=\left[\,\tilde{\mathtt{M}}^1\ \dots\ \tilde{\mathtt{M}}^K\,\right]=\begin{bmatrix}\tilde{\mathtt{M}}_{11} & \dots & \tilde{\mathtt{M}}_{1K}\\ \vdots & & \vdots\\ \tilde{\mathtt{M}}_{F1} & \dots & \tilde{\mathtt{M}}_{FK}\end{bmatrix}=\begin{bmatrix}l_{1,1}\mathtt{R}_1 & \dots & l_{1,K}\mathtt{R}_1\\ \vdots & & \vdots\\ l_{F,1}\mathtt{R}_F & \dots & l_{F,K}\mathtt{R}_F\end{bmatrix}$$

Since each 2×3 $\tilde{\mathtt{M}}_{fk}$ sub-block is a scaled rotation (truncated to dimension 2 for weak perspective projection) a 3×3 matrix $\mathtt{Q}_k$ (with $k = 1...K$) can be computed to correct each vertical block $\tilde{\mathtt{M}}^k$ by imposing orthogonality and equal norm constraints on the rows of each $\tilde{\mathtt{M}}_{fk}$. Each $\tilde{\mathtt{M}}_{fk}$ block will contribute with 1 orthogonality and 1 equal norm constraint to solve for the elements in $\mathtt{Q}_k$.

Each vertical block will then be corrected in the following way: ($\hat{\mathtt{M}}^k \leftarrow \tilde{\mathtt{M}}^k\mathtt{Q}_k$). Unlike the method proposed in [1], this provides a unique corrective transform for each column-triple of $\tilde{\mathtt{M}}$. The overall 3K×3K correction matrix $\mathtt{Q}$ will therefore be a block diagonal matrix. The 3D structure matrix will then be corrected appropriately using the inverse transformation: $\hat{\mathtt{S}} \leftarrow \mathtt{Q}^{-1}\tilde{\mathtt{S}}$. Brand included two further steps to refine the transformation. However, we did not notice improved results when we applied them.

2.3.2. Factorization of the motion matrix $\tilde{\mathtt{M}}$

The final step in the non-rigid factorization algorithm deals with the factorization of the motion matrix $\tilde{\mathtt{M}}$ into the 2×3 rotation matrices $\mathtt{R}_f$ and the deformation weights $l_{f,k}$. Bregler et al. [1] proposed a second factorization round where each each motion matrix 2 row sub-block $\tilde{\mathtt{M}}_{\mathtt{f}} = \mathbf{l}_f^T \otimes \mathtt{R}_f$ (with $f = 1...F$) is rearranged as an outer product of rotation parameters and deformation coefficients and then decomposed using a series of rank-1 SVD's. However, in the presence of noise the second and higher singular values of the sub-blocks do not vanish and this results in bad estimates for the rotation matrices and the deformation weights.

Brand proposed an alternative method in [2] to factorize each motion matrix 2 row sub-block $\tilde{\mathtt{M}}_{\mathtt{f}} = \mathbf{l}_f^T \otimes \mathtt{R}_f$ using orthonormal decomposition, which factors a matrix directly into a rotation and a vector. Summarizing his approach: a matrix $\mathtt{A}_{2\times 3}$ is built by re-arranging and manipulating each motion matrix sub-block $\tilde{\mathtt{M}}_{\mathtt{f}}$ (see [5] for details). The analytic form of $\mathtt{A}$ is:

$$\mathtt{A} = \begin{bmatrix} kr_1 & kr_2 & kr_3\\ kr_4 & kr_5 & kr_6\end{bmatrix} \tag{3}$$

where $k = l_{f,1} + l_{f,2} + ... + l_{f,K}$ (the sum of all the deformation weights for that particular frame f) and $r_1, .., r_6$ are the coefficients of the truncated rotation matrix $\mathtt{R}_f$. Since the rows of $\mathtt{R}_f$ are orthonormal, the equation $\mathtt{A}\mathtt{R}_f^T = \sqrt{\mathtt{A}\mathtt{A}^T}$ is satisfied, leading to $\mathtt{R}_f^T = \sqrt{\mathtt{A}\mathtt{A}^T}/\mathtt{A}$. This allows one to find a linear least-squares fit for $\mathtt{R}_f$ and subsequently for the weights l_{fk} exploiting the orthonormality of $\mathtt{A}$.

3. THE STEREO CAMERA CASE

The main contribution of this paper is to extend the non-rigid factorization methods to the case of a stereo rig, where the two cameras remain fixed relative to each other throughout the sequence. However, the same framework could be used in the case of 3 or more cameras. Torresani et. al. [3] formulated the factorization problem for the multiple camera case but did not provide an implementation or any experimental results.

3.1. The stereo motion model

When two cameras are viewing the same scene, the measurement matrix $\mathtt{W}$ will contain the image measurements from the left and right cameras resulting in a 4F×P size matrix. Assuming that both cameras are synchronized and that not only the single-frame tracks but also the stereo correspondences are known we may write the measurement matrix $\mathtt{W}$ as:

$$\mathtt{W}=\begin{bmatrix}\mathtt{W}^L\\ \mathtt{W}^R\end{bmatrix} \tag{4}$$

where for each frame f the stereo correspondences are:

$$\mathtt{W}_f^L=\begin{bmatrix}u^L_{f,1} & \dots & u^L_{f,P}\\ v^L_{f,1} & \dots & v^L_{f,P}\end{bmatrix}\ \mathtt{W}_f^R=\begin{bmatrix}u^R_{f,1} & \dots & u^R_{f,P}\\ v^R_{f,1} & \dots & v^R_{f,P}\end{bmatrix} \tag{5}$$

Note that, since we assume that the cameras are synchronized, at each time step f the left and right cameras are observing the same 3D structure and an additional constraint is that the the structure matrix $\mathtt{S}$ and the deformation coefficients $l_{f,k}$ are shared by left and right camera. The measurement matrix $\mathtt{W}$ can be factored into a motion matrix $\mathtt{M}$ and a structure matrix $\mathtt{S}$ which take the following form:

$$\mathtt{W}=\begin{bmatrix}l_{1,1}\mathtt{R}_1^L & \dots & l_{1,K}\mathtt{R}_1^L\\ \vdots & & \vdots\\ l_{F,1}\mathtt{R}_F^L & \dots & l_{F,K}\mathtt{R}_F^L\\ \hline l_{1,1}\mathtt{R}_1^R & \dots & l_{1,K}\mathtt{R}_1^R\\ \vdots & & \vdots\\ l_{F,1}\mathtt{R}_F^R & \dots & l_{F,K}\mathtt{R}_F^R\end{bmatrix}\begin{bmatrix}\mathtt{S}_1\\ \vdots\\ \mathtt{S}_K\end{bmatrix} \tag{6}$$

Note that the assumption that the deformation coefficients are the same for the left and right sequences relies on the fact that the weak perspective scaling f/Z_{avg} must be the same for both cameras. However, this assumption is generally true in a symmetric stereo setup where f and Z_{avg} are usually the same for both cameras.

3.2. Non-rigid stereo factorization

Once more the rank of the matrix measurement $\mathtt{W}$ is $r \leq 3K$ since $\mathtt{M}$ is a 4F×3K matrix and $\mathtt{S}$ is a 3K×P matrix, where P is the number of points. Assuming that the single frame tracks and the stereo correspondences are all known the measurement matrix $\mathtt{W}$ may be factorized into the product of a motion matrix $\mathtt{M}$ and a shape matrix $\mathtt{S}$ by truncating the SVD of $\mathtt{W}$ to rank 3K.

$$\mathtt{W}=\begin{bmatrix} \mathtt{W}^L \\ \mathtt{W}^R \end{bmatrix}=\tilde{\mathtt{M}}\tilde{\mathtt{S}}=\begin{bmatrix} \tilde{\mathtt{M}}^L \\ \tilde{\mathtt{M}}^R \end{bmatrix}\tilde{\mathtt{S}} \tag{7}$$

3.2.1. Computing the transformation matrix $\mathtt{Q}$

The result of the factorization is not unique since $(\tilde{\mathtt{M}}\mathtt{Q})(\mathtt{Q}^{-1}\tilde{\mathtt{S}})$ would give an equivalent factorization. We proceed to apply the metric constraint in the same way as we did for the single camera case in section 2.3.1, correcting each 4F×3 vertical block in $\tilde{\mathtt{M}}$ independently. Note that in this case we have four constraints per frame: 1 orthogonality and 1 equal norm constraint from each camera. Each vertical block will then be corrected: $\hat{\mathtt{M}}^k \leftarrow \tilde{\mathtt{M}}^k\mathtt{Q}_k$ and the shape matrix will be corrected with the inverse transformation: $\hat{\mathtt{S}} \leftarrow \mathtt{Q}^{-1}\tilde{\mathtt{S}}$.

3.2.2. Factorization of the motion matrix $\tilde{\mathtt{M}}$

In the stereo case we factorize each 4×3K sub-block of the motion matrix (which contains left and right measurements for each frame f) into its truncated 2×3 rotation matrices $\mathtt{R}_f^L$ and $\mathtt{R}_f^R$ and the deformation weights $l_{f,k}$ using orthonormal decomposition. The structure of the sub-blocks can be expressed as:

$$\begin{bmatrix} \mathtt{M}_{f1}^L & \dots & \mathtt{M}_{fK}^L \\ \mathtt{M}_{f1}^R & \dots & \mathtt{M}_{fK}^R \end{bmatrix}=\begin{bmatrix} \mathtt{l}_{f,1}\begin{bmatrix} \mathtt{R}_f^L \\ \mathtt{R}_f^R \end{bmatrix} & \dots & \mathtt{l}_{f,K}\begin{bmatrix} \mathtt{R}_f^L \\ \mathtt{R}_f^R \end{bmatrix} \end{bmatrix} \tag{8}$$

Since now we have 4 rows per frame, in the stereo case we build a 4×3 matrix $\mathtt{A}$ by re-arranging and manipulating each motion matrix sub-block $\tilde{\mathtt{M}}_\mathtt{f}$. The analytic form of this matrix, built directly from $\tilde{\mathtt{M}}_\mathtt{f}$, is:

$$\mathtt{A}=\begin{bmatrix} kr_1^L & kr_2^L & kr_3^L \\ kr_4^L & kr_5^L & kr_6^L \\ kr_1^R & kr_2^R & kr_3^R \\ kr_4^R & kr_5^R & kr_6^R \end{bmatrix}=\begin{bmatrix} \mathtt{A}_L \\ \mathtt{A}_R \end{bmatrix} \tag{9}$$

where $k = l_{f,1} + \dots + l_{f,K}$. Since the rows of $\mathtt{R}^L$ and $\mathtt{R}^R$ are orthonormal, the following equation is satisfied:

$$\begin{bmatrix} \mathtt{R}_L & 0 \\ 0 & \mathtt{R}_R \end{bmatrix}_{4\times6}\begin{bmatrix} \mathtt{A}_L^T & 0 \\ 0 & \mathtt{A}_R^T \end{bmatrix}_{6\times4}=\sqrt{\begin{bmatrix} \mathtt{A}_L\mathtt{A}_L^T & 0 \\ 0 & \mathtt{A}_R\mathtt{A}_R^T \end{bmatrix}_{4\times4}} \tag{10}$$

Therefore, a linear least-squares fit can be obtained for the rotation matrices $\mathtt{R}_L$ and $\mathtt{R}_R$ and the weights l_{fk} can be subsequently estimated exploiting the orthonormality of $\mathtt{A}$.

3.2.3. Imposing the fixed relative rotation constraint

There is a further constraint that may be imposed if we assume that the left and right cameras remain fixed with respect to each other throughout the sequence: the relative rotation between the left and right cameras is fixed. This can be expressed in the following way:

$$\mathtt{R}_{f,k}^R = \mathtt{R}_{f,k}^L\hat{\mathtt{R}} \quad f = 1...F, k = 1...K \tag{11}$$

Given all the $\mathtt{R}_{f,k}^R$ and $\mathtt{R}_{f,k}^L$ the previous expression can be used to find the linear least squares fit of $\hat{\mathtt{R}}$. However, we have implemented this estimation and found that we did not obtain improved estimates for the rotation matrices and for the modified 3D shape $\mathtt{S}$.

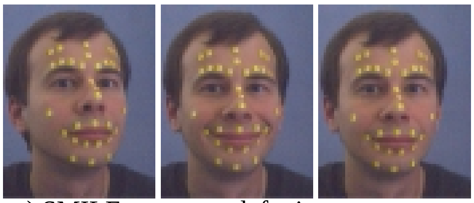
We are currently developing a non-linear method to refine the estimates of all the parameters by minimizing image reprojection error using bundle-adjustment.

4. EXPERIMENTAL RESULTS

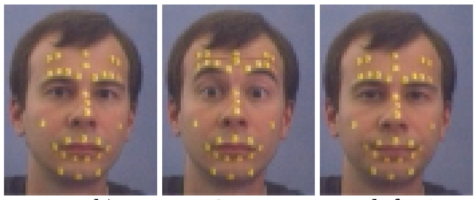
In this section we present some experimental results obtained with real image sequences taken with a pair of synchronized Fire-i digital cameras with 4,65mm built in lenses. The stereo setup was such that the baseline was 20cm and the relative orientation of the cameras was around $15\,\mathrm{deg}$. Two sequences of a human face undergoing rigid motion and flexible deformations were used: the SMILE sequence (82 frames), where the deformation was due to the subject smiling and the EYEBROW (115 frames) sequence where the subject was raising and lowering the eyebrows. Figure (1) shows 3 frames chosen from the sequences taken with the left and right cameras.

In order to simplify the temporal and stereo matching the subject had some markers placed on relevant points of the face such as along the eyebrows, the chin and the lips. A simple colour model of the markers using HSV components was used and this representation was used to track each marker throughout the left and right sequences respectively. The stereo matching was initialized by hand in the first image pair and then the temporal tracks were used to update the stereo matches.

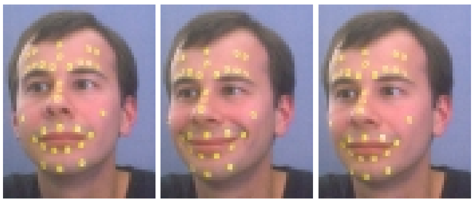
Figure (2) shows front, side and top views of the 3D reconstructions obtained for the SMILE sequence. First we applied the single camera factorization algorithm described in section 2.3 independently to the left and right sequences. We then applied the proposed stereo algorithm to the stereo sequence. In all cases the number of tracked points was $P = 31$ and the chosen number of basis shapes was $K = 5$. Figure (2) shows how the stereo reconstruction clearly provides improved results. The reconstructions obtained from the left and right sequences have worse depth estimates (see top views) and the symmetry of the face is only preserved in the stereo sequence. Figure (3) shows the front views



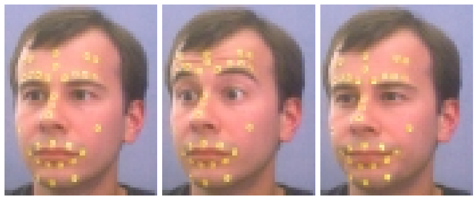
a) SMILE sequence: left view



b) EYEBROW sequence: left view

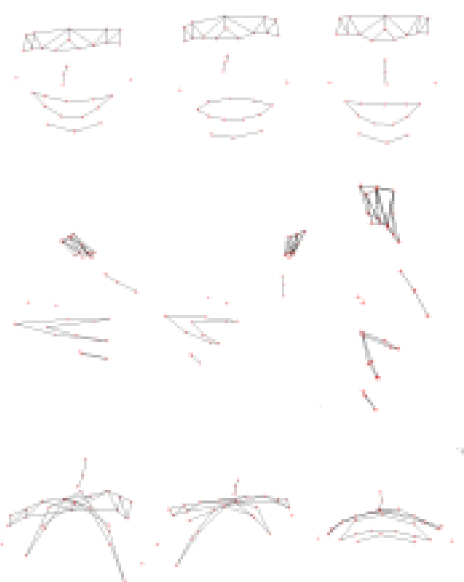


c) SMILE sequence: right view



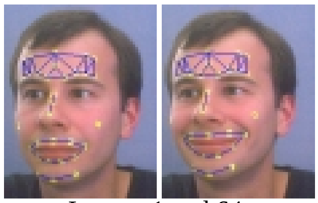
d) EYEBROW sequence: right view

Fig. 1. *Three images from the left (a) and right (c) views of the SMILE sequence and left (b) and right (d) views of the EYEBROW sequence.*

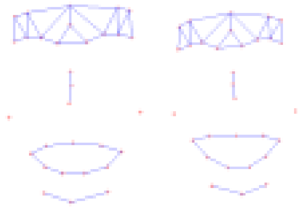


a) Left camera b) Right camera c) Stereo algorithm

Fig. 2. SMILE sequence: Front, side and top views (above, middle, bottom) of the 3D model for the a) left camera, b) right camera and c) stereo setup for K=5.



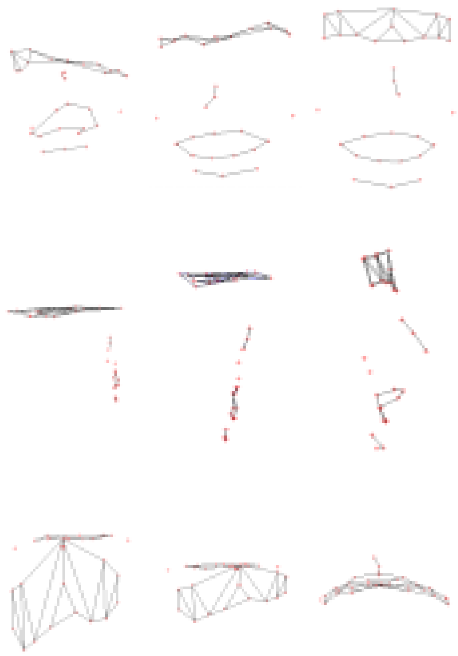
Images 1 and 64



Front view of reconstructed frames

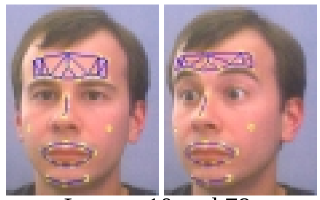
Fig. 3. SMILE sequence: Frames 1 and 64 of the sequence and results of reconstruction with the stereo algorithm

of the 3D reconstructions obtained for frames 1 and 64 of the stereo sequence. We can appreciate how the deformable structure is well captured in the 3D models (notice how the upper lip is curved first and then straightened). Figure (4) shows the 3D reconstructions obtained for the EYEBROW sequence. Once more, the single camera factorization algorithm was applied to the left and right sequences and the stereo algorithm was then applied to the stereo sequence. In this sequence the 3D model obtained using stereo factorization is significantly better than the ones obtained with the left and right sequences. In fact, the left and right reconstructions have very poor quality, particularly the depth estimates. Note that there was less rigid motion in this sequence and therefore the single camera factorization algorithm is not capable of recovering correct 3D information whereas the stereo algorithm provides a good deformable model. Figure (5) shows the 3D reconstructions obtained for frames 10 and 78 of the stereo sequence, showing that the flexibility of the face is well recovered.

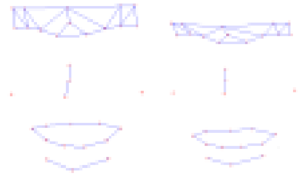


a) Left camera b) Right camera c) Stereo algorithm

Fig. 4. EYEBROW sequence: Front, side and top views (above, middle, bottom) of the 3D model for the a) left camera, b) right camera and c) stereo setup sequences for K=5.



Images 10 and 78



Front view of reconstructed frames

Fig. 5. EYEBROW sequence: Frames 10 and 78 and front view of the reconstruction with the stereo algorithm

5. SUMMARY AND CONCLUSIONS

We have developed a factorization algorithm to obtain non-rigid 3D models from image correspondences obtained from a stereo pair. The algorithm imposes the extra constraints that arise from the fact that both stereo cameras are viewing the same 3D structure. Our focus has been on the recovery of 3D shape. Experiments with real data prove that improved 3D models can be achieved using the stereo factorization method. We have simplified the temporal and spatial correspondence problem by using markers and some manual matching.

We are currently working on a non-linear estimation method to refine the initial estimates of the motion and shape parameters obtained with factorization by minimizing image reprojection error using bundle-adjustment.

6. REFERENCES

[1] C. Bregler, A. Hertzmann, and H Biermann, “Recovering non-rigid 3d shape from image streams,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head, South Carolina*, jun 2000, pp. 690–696.

[2] M. Brand, “Morphable models from video,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii*, December 2001.

[3] L. Torresani, D. Yang, E. Alexander, and C. Bregler, “Tracking and modeling non-rigid objects with rank constraints,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii*, 2001.

[4] C. Tomasi and T. Kanade, “Shape and motion from image streams: a factorization method,” *International Journal in Computer Vision*, vol. 9, no. 2, pp. 137–154, 1991.

[5] M. Brand and Bhotika R, “Flexible flow for 3d nonrigid tracking and shape recovery,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii*, December 2001, pp. 315–22.

[6] T. Vetter and V. Blanz, “A morphable model for the synthesis of 3d faces,” in *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics*, 1999, pp. 187–194.

[7] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D. H. Salesin, “Synthesising realistic facial expressions from photographs,” in *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics*, 1998.

[8] I. Essa and S. Basu, “Modeling, tracking and interactive animation of facial expressions and head movements using input from video,” in *Proceedings of Computer Animation Conference*, Geneva, Switzerland, June 1996.

[9] M. Irani, “Multi-frame optical flow estimation using subspace constraints,” in *Proc. 7th International Conference on Computer Vision, Kerkyra, Greece*, 1999.