International Journal of Computer Vision 66(2), 193–207, 2006

DOI: 10.1007/s11263-005-3958-5

Non-Rigid Stereo Factorization

ALESSIO DEL BUE AND LOURDES AGAPITO
Department of Computer Science, Queen Mary, University of London, London, E1 4NS, UK
alessio@dcs.qmul.ac.uk
lourdes@dcs.qmul.ac.uk

Received May 4, 2004; Revised January 21, 2005; Accepted February 14, 2005

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 are 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. The algorithm includes a non-linear optimization step that minimizes image reprojection error and imposes the correct structure to the motion matrix by choosing an appropriate parameterization. We show that 3D shape and motion estimates can be successfully disambiguated after bundle adjustment and demonstrate this on synthetic and real image sequences. While this optimization step is proposed for the stereo camera case, it can be readily applied to the case of non-rigid structure recovery using a monocular video sequence.

Keywords: non-rigid structure from motion, non-linear optimization, stereo, deformable model

1. Introduction

Much interest has been recently devoted in computer vision to the study of the extent to which 3D information about the world can be inferred directly from image sequences taken from a moving camera, when the specific details of the camera and its motion are all unknown in advance. Such free-form inference can only succeed if certain assumptions are made, the standard one being that the scene observed by the camera is rigid. In this paper we explore the more challenging case of scenes that are not completely rigid, but which have certain degrees of flexibility or deformation.

Recent work in non-rigid factorization (Bregler et al., 2000; Brand and Bhotika, 2001; Torresani et al., 2001) 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 (Tomasi and Kanade, 1991), 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 (Brand and Bhotika, 2001), animation or other analysis. Similar to the rigid case, the underlying geometric constraints are expressed as a rank constraint which is used to factorize the measurement matrix to obtain the 3D pose, configuration coefficients and a pre-specified number of 3D basis shapes.

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 (Vetter and Blanz, 1999) or multi-view reconstruction (Pighin et al., 1998)—or have been specialised to the specific object under observation: for example physically-based human face models (Essa and Basu, 1996).

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 (Torresani et al., 2001; Brand, 2001), an approach inspired by its rigid equivalent (Irani, 1999).

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. We have developed a new method to factorize the measurement matrix into the left and right motion matrices and the 3D non-rigid shape. Note that our work requires both cameras to be synchronized. However, this could be elegantly solved inside a factorization framework with the solution proposed by Tresadern and Reid (2003) for the synchronization of stereo video sequences in an uncalibrated scenario.

1.1. Previous Work

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 (Torresani et al., 2001; Brand, 2001). However, our focus here is on the recovery of flexible 3D shape and motion estimates.

Bregler et al. (2000) were the first to use a factorization-based method for the recovery of non-rigid structure and motion. The decomposition between motion and shape parameters is not unique however, and the motion matrix is only obtained up to a post multiplication by a transformation matrix. While this matrix can be easily computed in the case of rigid structure by enforcing orthonormality constraints on the camera motion, its computation in the non-rigid case is not trivial since the motion matrix has a replicated block structure which must be imposed.

Several methods have been proposed so far to compute the transformation matrix. Bregler et al. (2000) enforced orthonormality constraints on the camera rotations in a similar way to the rigid factorization scheme. Later, (Brand, 2001) proposed an improvement to Bregler et al.'s method using numerically well-behaved heuristics to compute the transformation matrix and adding a final minimization to regularize the shape. Torresani et al. (2001) also extended the method by Bregler et al. (2000) by introducing a final trilinear optimization on the motion and structure parameters. However, none of these approaches are completely satisfactory since they do not impose the full block structure on the motion matrix.

It was not until recently that (Xiao et al., 2004) proved that the orthonormality constraints on the camera rotations are not sufficient to compute the transformation matrix and they proposed a new set of constraints on the shape basis. Their work proves that when both sets of constraints are imposed a closed-form solution exists to the problem of non-rigid structure from motion.

The main contribution of this paper has been to extend non-rigid factorization methods to the stereo camera case. The new algorithm is particularly well suited to sequences in which the overall rigid motion is small, where monocular algorithms will fail to give reliable 3D shape estimates. Here we have used an alternative method to solve for the ambiguity in the recovery of motion and shape parameters. First we obtain an initial estimate of the motion and structure parameters using an extension of Brand's algorithm (Brand, 2001) to the stereo camera case (Del Bue and Agapito, 2004). We then use a bundle adjustment step to refine the initial solution given by the stereo algorithm, by minimizing im-

age reprojection error: a geometrically meaningful error function. The parameterization of the problem takes into account the fixed geometry of the stereo rig. Aanæs and Kahl first proposed the use of bundle-adjustment in the non-rigid case (Aanæs and Kahl, 2002) for a monocular sequence. However, our approach differs in the choice of initialization and in the parameterization of the problem leading to improved results.

Previous work on non-rigid factorization in the multi-view case includes work by Tan and Ishikawa (2004), however their approach limits itself to estimating the 3D coordinates of a set of points on a deformable object at each frame in the sequence without modelling the deformations.

The results of our experiments—where matching is aided by the use of markers—show that the stereo framework succeeds in computing reliable 3D reconstructions in cases where monocular algorithms fail to recover the correct shape (sequences where the overall rigid motion is small). We also present results on synthetic and real data to show that the bundle adjustment refinement step improves the estimates of motion and shape parameters. Note that while we have proposed this optimization step in the stereo camera case, it can be readily applied to the case of non-rigid structure recovery using a monocular video sequence (Del Bue et al., 2004).

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. In Section 3.3 we describe the non-linear optimization scheme and finally in Section 4 we present some experimental results of the stereo algorithm and the bundle adjustment refinement on synthetic and real image sequences showing improved 3D reconstructions.

2. Background: Factorization

2.1. Rigid Factorization

Tomasi and Kanade's factorization algorithm (Tomasi and Kanade, 1991) 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 $2F \times 3$ motion matrix which expresses the pose of the camera and $\mathtt{S}$ is the $3 \times P$ 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: $SVD_3(\mathtt{W}) = \mathtt{U}_{2F\times 3}\mathtt{D}_{3\times 3}\mathtt{V}_{3\times P} = \mathtt{M}_{2F\times 3}\mathtt{S}_{3\times P}$. 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 (Bregler et al., 2000; Brand, 2001; Torresani et al., 2001). 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 $(\mathtt{S}_1, \mathtt{S}_2, \ldots, \mathtt{S}_K)$ is a $3 \times P$ 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 $\mathtt{S}_i$:

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

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:

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

where

$$\mathtt{R}_f = \begin{bmatrix} r_{f,1} & r_{f,2} & r_{f,3} \\ r_{f,4} & r_{f,5} & r_{f,6} \end{bmatrix} \tag{2}$$

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 its 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. In this way, the 3D coordinate system will be centred at the centroid of the shape $\mathtt{S}$. If the same P points can be tracked throughout an image sequence we may stack them into a $2F \times P$ measurement matrix $\mathtt{W}$ and we may write:

$$\mathtt{W} = \begin{bmatrix} \mathtt{W}_1 \\ \vdots \\ \mathtt{W}_F \end{bmatrix} = \begin{bmatrix} l_{1,1}\mathtt{R}_1 & \dots & l_{1,K}\mathtt{R}_1 \\ \vdots & & \vdots \\ l_{F,1}\mathtt{R}_F & \dots & l_{F,K}\mathtt{R}_F \end{bmatrix} \begin{bmatrix} \mathtt{S}_1 \\ \vdots \\ \mathtt{S}_K \end{bmatrix} = \mathtt{MS} \tag{3}$$

Since $\mathtt{M}$ is a $2F \times 3K$ matrix and $\mathtt{S}$ is a $3K \times P$ matrix, the rank of $\mathtt{W}$ when no noise is present must be at most $3K$. Note that, in relation to rigid factorization, in the non-rigid case the rank is incremented by three with every new mode of deformation. The goal of factorization algorithms is to exploit this rank constraint to recover the 3D pose, and shape (basis-shapes and deformation coefficients) of the object from the correspondence points stored in $\mathtt{W}$.

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 $3K \times 3K$ 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 (3) and that removes the affine ambiguity upgrading the reconstruction to a metric one. 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 (Tomasi and Kanade, 1991), in the non-rigid case imposing the appropriate repetitive structure to the motion matrix $\tilde{\mathtt{M}}$ results in a non-linear problem.

It is important to note that while the block structure is not required if we only wish to determine image point motion, it is crucial for the recovery of 3D shape and motion which is the main goal of our work.

2.3.1. *Computing the Transformation Matrix* $\mathtt{Q}$.

The approach proposed by Brand (2001) 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:

$$\tilde{\mathtt{M}} = \begin{bmatrix} \tilde{\mathtt{M}}^1 & \dots & \tilde{\mathtt{M}}^K \end{bmatrix} = \begin{bmatrix} \tilde{\mathtt{M}}_{11} & \dots & \tilde{\mathtt{M}}_{1K} \\ \vdots & & \vdots \\ \tilde{\mathtt{M}}_{F1} & \dots & \tilde{\mathtt{M}}_{FK} \end{bmatrix}$$

$$= \begin{bmatrix} l_{1,1}\mathtt{R}_1 & \dots & l_{1,K}\mathtt{R}_1 \\ \vdots & & \vdots \\ l_{F,1}\mathtt{R}_F & \dots & l_{F,K}\mathtt{R}_F \end{bmatrix}$$

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 $k = 1 \dots K$) 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: $(\hat{\mathtt{M}}^k \leftarrow \tilde{\mathtt{M}}^k\mathtt{Q}_k)$. The overall $3K \times 3K$ correction matrix $\mathtt{Q}$ will therefore be a block diagonal matrix with the following structure:

$$\begin{bmatrix} \mathtt{Q}_1 & 0 & \dots & 0 \\ 0 & \mathtt{Q}_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \mathtt{Q}_K \end{bmatrix} \tag{4}$$

Unlike the method proposed by Bregler Bregler et al. (2000)—where the metric constraint was imposed only to the rigid component so that $\mathtt{Q}_i = \mathtt{Q}_{\text{rigid}}$ for $i = 1 \ldots K$—this provides a corrective transform for each column-triple of $\tilde{\mathtt{M}}$. The 3D structure matrix will then be corrected appropriately using the inverse transformation: $\hat{\mathtt{S}} \leftarrow \mathtt{Q}^{-1}\tilde{\mathtt{S}}$.

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. (2000) proposed a second factorization round where each motion matrix 2 row sub-block $\tilde{\mathtt{M}}_f = \mathbf{l}_f^T \otimes \mathtt{R}_f$—where $\otimes$ indicates the tensor product—(with $f = 1 \ldots F$) 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 (2001) to factorize each motion matrix 2 row sub-block $\tilde{\mathtt{M}}_f = \mathbf{l}_f^T \otimes \mathtt{R}_f$ using orthonormal decomposition, which factors a matrix directly into a rotation and a vector.

Each motion matrix sub-block $\tilde{\mathtt{M}}_f$ (see Brand and Bhotika, 2001 for details) is rearranged such that:

$$\tilde{\mathtt{M}}_f \rightarrow \hat{\mathtt{M}}_f = \begin{bmatrix} l_{f,1}\mathbf{r}_f^T & l_{f,2}\mathbf{r}_f^T & \ldots & l_{f,k}\mathbf{r}_f^T \end{bmatrix} \tag{5}$$

where $\mathbf{r}_f = [r_{f1}, \ldots, r_{f6}]$ are the coefficients of the rotation matrix $\mathtt{R}_f$. The motion matrix $\hat{\mathtt{M}}_f$ of size $6 \times K$ is then post-multiplied by the $K \times 1$ unity vector $\mathbf{c} = [1 \ldots 1]$ thus obtaining:

$$\mathbf{a}_f = k\mathbf{r}_f^T = \hat{\mathtt{M}}_f \mathbf{c} \tag{6}$$

where $k = l_{f,1} + l_{f,2} + \cdots + l_{f,K}$ (the sum of all the deformation weights for that particular frame f). A matrix $\mathtt{A}_f$ of size 2×3 is built by re-arranging the coefficients of the column vector $\mathbf{a}_f$. The analytic form of $\mathtt{A}_f$ is:

$$\mathtt{A}_f = \begin{bmatrix} kr_1 & kr_2 & kr_3 \\ kr_4 & kr_5 & kr_6 \end{bmatrix} \tag{7}$$

Since $\mathtt{R}_f$ is an orthonormal matrix, the equation $\mathtt{A}_f\mathtt{R}_f^T = \sqrt{\mathtt{A}_f\mathtt{A}_f^T}$ is satisfied, leading to $\mathtt{R}_f^T = \sqrt{\mathtt{A}_f\mathtt{A}_f^T}/\mathtt{A}_f$ This allows one to find a linear least-squares fit for the rotation matrix $\mathtt{R}_f$.

In order to estimate the configuration weights the sub-block matrix $\tilde{\mathtt{M}}_f$ is then rearranged in a different way from (5):

$$\tilde{\mathtt{M}}_f \rightarrow \bar{\mathtt{M}}_f = \begin{bmatrix} l_{f,1}\mathbf{r}_f \\ \ldots \\ l_{f,k}\mathbf{r}_f \end{bmatrix} \tag{8}$$

The configuration weights for each frame f are then derived exploiting the orthonormality of $\mathtt{R}_f$ since:

$$\bar{\mathtt{M}}_f \mathbf{r}_f^T = \begin{bmatrix} l_{f,1}\mathbf{r}_f\mathbf{r}_f^T \\ \ldots \\ l_{f,k}\mathbf{r}_f\mathbf{r}_f^T \end{bmatrix} = 2 \begin{bmatrix} l_{f,1} \\ \ldots \\ l_{f,k} \end{bmatrix} \tag{9}$$

Brand (2001) included a final minimization scheme in his *flexible factorization* algorithm: the deformations in $\tilde{\mathtt{S}}$ should be as small as possible relative to the mean shape. The idea here is that most of the image point motion should be explained by the rigid component. This is equivalent to the shape regularization used by other authors (Torresani et al., 2001; Aanæs and Kahl, 2002).

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. (2001) 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 $4F \times P$ matrix. Assuming that not only the single-frame tracks but also the stereo correspondences are

known we may write the measurement matrix $\mathtt{W}$ as:

$$\mathtt{W} = \begin{bmatrix} \mathtt{W}^L \\ \mathtt{W}^R \end{bmatrix} \tag{10}$$

where for each frame f the stereo correspondences are:

$$\mathtt{W}_f^L = \begin{bmatrix} u_{f,1}^L & \dots & u_{f,P}^L \\ v_{f,1}^L & \dots & v_{f,P}^L \end{bmatrix} \quad \mathtt{W}_f^R = \begin{bmatrix} u_{f,1}^R & \dots & u_{f,P}^R \\ v_{f,1}^R & \dots & v_{f,P}^R \end{bmatrix} \tag{11}$$

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 this results in the additional constraint that 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:

$$\mathtt{W} = \begin{bmatrix} l_{1,1}\mathtt{R}_1^L & \dots & l_{1,K}\mathtt{R}_1^L \\ \vdots & & \vdots \\ l_{F,1}\mathtt{R}_F^L & \dots & l_{F,K}\mathtt{R}_F^L \\ \hline l_{1,1}\mathtt{R}_1^R & \dots & l_{1,K}\mathtt{R}_1^R \\ \vdots & & \vdots \\ l_{F,1}\mathtt{R}_F^R & \dots & l_{F,K}\mathtt{R}_F^R \end{bmatrix} \begin{bmatrix} \mathtt{S}_1 \\ \vdots \\ \mathtt{S}_K \end{bmatrix} \tag{12}$$

where $\mathtt{R}^L$ and $\mathtt{R}^R$ are the rotation components for the right and left cameras. Once more, we have eliminated the translation vector **T** by registering image points to the centroid in each frame.

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.

It is possible to express the stereo motion matrix $\mathtt{M}$ by including explicitly the assumption that a fixed stereo rig is being used. In this case the rotation pair for the left and right cameras can be expressed with the introduction of a relative orientation matrix $\mathtt{R}_{rel}$ such that: $\mathtt{R}^R = \mathtt{R}_{\text{rel}}\mathtt{R}^L$. The motion matrix $\mathtt{M}$ in Eq. (12) can be consequently expressed as:

$$\mathtt{M} = \begin{bmatrix} l_{1,1}\mathtt{R}_1^L & \dots & l_{1,K}\mathtt{R}_1^L \\ \vdots & & \vdots \\ l_{F,1}\mathtt{R}_F^L & \dots & l_{F,K}\mathtt{R}_F^L \\ \hline l_{1,1}\mathtt{R}_{\text{rel}}\mathtt{R}_1^L & \dots & l_{1,K}\mathtt{R}_{\text{rel}}\mathtt{R}_F^L \\ \vdots & & \vdots \\ l_{F,1}\mathtt{R}_{\text{rel}}\mathtt{R}_F^L & \dots & l_{F,K}\mathtt{R}_{\text{rel}}\mathtt{R}_F^L \end{bmatrix} \tag{13}$$

3.2. *Non-Rigid Stereo Factorization*

Once more the rank of the matrix measurement $\mathtt{W}$ is at most $3K$ since $\mathtt{M}$ is a $4F \times 3K$ matrix and $\mathtt{S}$ is a $3K \times P$ 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 $3K$.

$$\mathtt{W} = \begin{bmatrix} \mathtt{W}^L \\ \mathtt{W}^R \end{bmatrix} = \tilde{\mathtt{M}}\tilde{\mathtt{S}} = \begin{bmatrix} \tilde{\mathtt{M}}^L \\ \tilde{\mathtt{M}}^R \end{bmatrix} \tilde{\mathtt{S}} \tag{14}$$

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 was described for the single camera case in Section 2.3.1, correcting each $4F \times 3$ vertical block in $\tilde{\mathtt{M}}$ independently. Note that in this case we have used five constraints per frame: 2 orthogonality (one from each camera) and 3 equal norm constraints (the first row has the same norm as each of the other 3). 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 $4 \times 3K$ 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:

$$\begin{bmatrix} \mathtt{M}_{f1}^{L} & \dots & \mathtt{M}_{fK}^{L} \\ \mathtt{M}_{f1}^{R} & \dots & \mathtt{M}_{fK}^{R} \end{bmatrix} = \left[l_{f,1} \begin{bmatrix} \mathtt{R}_{f}^{L} \\ \mathtt{R}_{f}^{R} \end{bmatrix} \dots l_{f,K} \begin{bmatrix} \mathtt{R}_{f}^{L} \\ \mathtt{R}_{f}^{R} \end{bmatrix} \right] \tag{15}$$

The approach used to estimate the rotation components for the left and right cameras is similar to the algorithm described in Section 2.3.2. Since now we have 4 rows per frame, we arrange the motion sub-blocks such that:

$$\tilde{\mathtt{M}}_f \rightarrow \hat{\mathtt{M}}_f = \left[l_{f,1} \begin{bmatrix} \mathbf{r}_{f}^{L} \\ \mathbf{r}_{f}^{R} \end{bmatrix} \; l_{f,2} \begin{bmatrix} \mathbf{r}_{f}^{L} \\ \mathbf{r}_{f}^{R} \end{bmatrix} \; \dots \; l_{f,K} \begin{bmatrix} \mathbf{r}_{f}^{L} \\ \mathbf{r}_{f}^{R} \end{bmatrix} \right] \tag{16}$$

where $\mathbf{r}_f^L = [r_{f1}^L \dots r_{f6}^L]^T$ is a column vector which contains the coefficients of the left rotation matrix $\mathtt{R}_f^L$ and similarly for $\mathbf{r}_f^R$. Post-multiplying the rearranged matrix $\hat{\mathtt{M}}_f$ by the $2K$ unity vector $\mathbf{c} = [1 \dots 1]^T$ gives a column vector $\mathbf{a}_f$:

$$\mathbf{a}_f = \hat{\mathtt{M}}_f \mathbf{c} \tag{17}$$

which may be rearranged into a 4×3 matrix $\mathtt{A}_f$ with analytic form:

$$\mathtt{A}_f = \begin{bmatrix} kr_{f,1}^L & kr_{f,2}^L & kr_{f,3}^L \\ kr_{f,4}^L & kr_{f,5}^L & kr_{f,6}^L \\ kr_{f,1}^R & kr_{f,2}^R & kr_{f,3}^R \\ kr_{f,4}^R & kr_{f,5}^R & kr_{f,6}^R \end{bmatrix} = \begin{bmatrix} \mathtt{A}_L \\ \mathtt{A}_R \end{bmatrix} \tag{18}$$

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

$$\begin{bmatrix} \mathtt{R}_L & 0 \\ 0 & \mathtt{R}_R \end{bmatrix}_{4\times 6} \begin{bmatrix} \mathtt{A}_L^T & 0 \\ 0 & \mathtt{A}_R^T \end{bmatrix}_{6\times 4} = \sqrt{\begin{bmatrix} \mathtt{A}_L \mathtt{A}_L^T & 0 \\ 0 & \mathtt{A}_R \mathtt{A}_R^T \end{bmatrix}_{4\times 4}} \tag{19}$$

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 in a similar way as shown in Section 2.3.2.

Finally a minimization scheme similar to the one used by Brand (2001) in his *flexible factorization* algorithm is applied here. The assumption is that the deformations should be small relative to the mean shape so that most of the image motion is explained by the rigid component.

So far we have presented an extension of non-rigid factorization methods to the case of a stereo camera pair. In particular our algorithm follows the approach by Brand (2001). While this new algorithm improves the quality of the 3D reconstructions obtained from a monocular sequence it still fails to render the appropriate replicated block structure to the motion matrix $\mathtt{M}$.

In the next section we will describe a non-linear optimization scheme which renders the appropriate structure to the motion matrix, allowing to disambiguate between the motion and shape parameters. Note that an alternative approach would have been to extend Xiao et al.'s (2004) method to the stereo case.

3.3. *Non-Linear Optimization*

Our approach is to obtain an initial solution for the non-rigid shape and the 3D pose using the stereo algorithm described in the previous section and then to perform a non-linear optimization step by minimizing image reprojection error.

The goal is to estimate the left and right camera matrices $\mathtt{R}_i^L$ and $\mathtt{R}_i^{\mathtt{R}}$, the configuration weights l_{ik} and the basis-shapes $\mathtt{S}_k$ such that the distance between the measured image points $\mathbf{x}_{ij}$ and the estimated image points $\hat{\mathbf{x}}_{ij}$ is minimized. Since the relative orientation between the left and right cameras is fixed we have expressed the rotation of the right camera for each frame in terms of the rotation matrices of the left camera and the relative rotation $\mathtt{R}_{\text{rel}}$ such that $\mathtt{R}_i^{\mathtt{R}} = \mathtt{R}_{\text{rel}} \mathtt{R}_i^L$. The cost function to be minimized is expressed below:

$$\begin{aligned} &\min_{\mathtt{R}_{\text{rel}} \mathbf{R}_i^L \mathtt{S}_k l_{i,k}} \sum_{i,j} \| \mathbf{x}_{ij} - \hat{\mathbf{x}}_{ij} \|^2 \\ &= \min_{\mathtt{R}_{\text{rel}} \mathbf{R}_i^L \mathtt{S}_k l_{i,k}} \sum_{i,j} \left\| \mathbf{x}_{ij}^L - \left(\mathtt{R}_i^L \sum_k l_{i,k} \mathtt{S}_k \right) \right\|^2 \\ &\quad + \left\| \mathbf{x}_{ij}^{\mathtt{R}} - \left(\mathtt{R}_{\text{rel}} \mathtt{R}_i^L \sum_k l_{i,k} \mathtt{S}_k \right) \right\|^2 \end{aligned} \tag{20}$$

This method is generically termed bundle-adjustment in the computer vision and photogrammetry communities and it provides a Maximum Likelihood estimate

provided that the noise can be modelled with a Gaussian distribution. The non-linear optimization of the cost function was achieved using a Levenberg Marquadt minimization scheme modified to take advantage of the sparse block structure of the matrices involved in the process (Triggs et al., 2000).

The work presented here is most closely related to the work by Aanæs and Kahl (2002). However, their approach differs in some fundamental aspects. Firstly their method uses a monocular sequence and requires a calibrated camera while our approach is uncalibrated. Our unique assumption is that the camera has zero skew and unity aspect ratio but the focal length and the extrinsic calibration are both unknown. In this way our method allows for scenarios in which a fully calibrated approach may not be applicable such as when there is auto-focus, zoom, or when the configuration of the cameras might change during action. Secondly our approach uses a different parameterization and initialization.

In terms of their experimental evaluation, Aanæs and Kahl do not provide a quantitative analysis of the recovered parameters, only some qualitative results. In contrast, our experimental analysis shows numerical results for the shape and motion parameters (see Section 4).

3.3.1. Implementation. We have chosen to parameterize the camera matrices $\mathtt{R}_i^L$ and $\mathtt{R}_{\text{rel}}$ using quaternions (Horn, 1987; Bar-Itzhack, 2000) giving a total of $(4 \times F) + 4$ rotation parameters, where F is the total number of frames. Quaternions ensure that there are no singularities and that the orthonormality of the rotation vectors is preserved. The structure was parameterized with the $3 \times K \times P$ coordinates of the $\mathtt{S}_k$ basis shapes (where K is the number of basis shapes and P is the total number of structure points) and the $K \times F$ deformation weights l_{ik}.

Given the large number of parameters involved in the non-linear minimization the objective function is highly non-linear and so it is crucial to provide an initial estimate that is sufficiently close to the global minimum. The output given by our stereo non-rigid factorization algorithm—described in the previous section—was used for initialization of the model parameters $\mathtt{R}_L$, $\mathtt{R}_R$, $\mathtt{S}_k$ and l_{ik}.

The constant relative orientation $\mathtt{R}_{\text{rel}}$ between the left and right cameras is estimated from the camera matrices $\mathtt{R}_L$ and $\mathtt{R}_R$ using a least squares estimation. Unit quaternions were used as the parameterization and the orthogonality constraint was enforced by fixing the 4-vector norm to unity such that the solution space is constrained to lie on a hypersphere of dimension 4.

An alternative initialization which we have observed also gives a good starting point for the non-linear minimization—although we do not show results here—is the one used by Torresani et al. (2001) in their final tri-linear minimization scheme. The idea is to initialize the rotation matrices with the motion corresponding to the rigid component, since it encodes the most significant part of the motion. This assumption works well in the scenario of human facial motion analysis, but would not be valid for highly deformable objects such as a hand or the human body. The basis shapes were initialized with the values obtained using the stereo non-rigid factorization method as were the weights associated with the rigid component. However, the weights associated with the basis shapes that account for the non-rigid motion were initialized to a very small value.

If the internal and external calibration of the stereo rig were known in advance after a process of calibration or self-calibration, an alternative initialization could be computed by recovering the 3D structure and performing Principal Component Analysis on the data to obtain an initial estimate for the shape bases and the coefficients. However, we have chosen to use an initialization that does not require a pre-calibration of the cameras.

Occasionally we have found that the 3D points tend to lie on a plane as a result of the minimization. To overcome this situation, a prior on the 3D shape has been added to the cost function. Our prior states that the depth of the points on the object surface will not change significantly from one frame to the next since the images are closely spaced in time adding the term $\sum_{i=2,j=1}^{i=F,j=P} \| S_z^{i-1,j} - S_z^{i,j} \|^2$ to the cost function. In this way we can preserve the relief present in the 3D data. Similar regularization terms have also been reported in Torresani et al. (2001), Brand (2001) and Aanæs and Kahl (2002).

4. Experimental Results

4.1. Results of the Stereo Factorization Algorithm

In this section we compare the performance of our stereo factorization algorithm—before the non-linear optimization—with Brand's single camera non-rigid

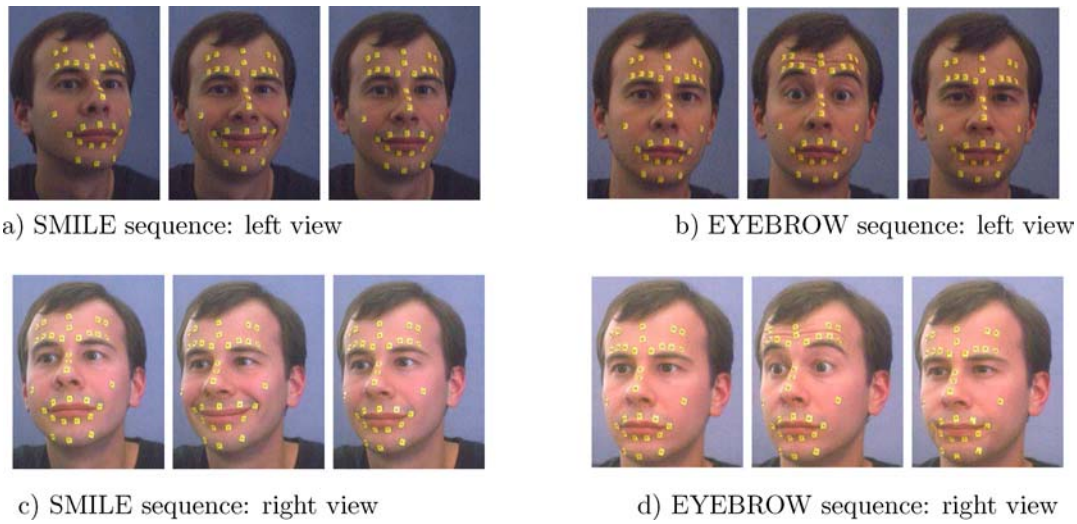

Figure 1. Three images from the left (a) and right (c) views of the SMILE sequence and left (b) and right (d) views of the EYEBROW sequence.

factorization method. We present some experimental results obtained with real image sequences taken with a pair of synchronized Fire-i digital cameras with 4,65 mm built in lenses. The stereo setup was such that the baseline was 20 cm and the relative orientation of the cameras was around 30deg. 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 developed by Brand—described in Section 2.3—to the left and right monocular 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.

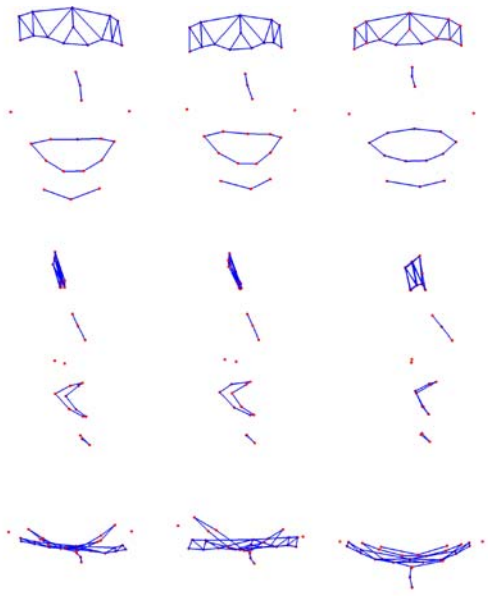

Figure 2. SMILE sequence: Front, side and top views (above, middle, bottom) of the 3D model for the (a) left camera, (b) right camera and (c) stereo setup for $K = 5$.

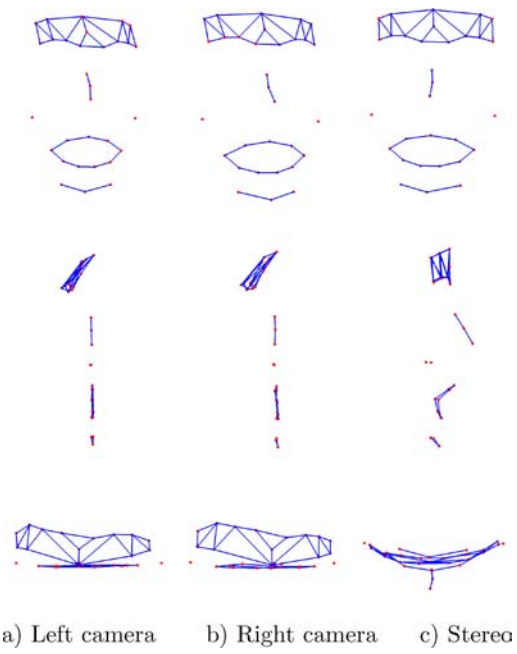

Figure 3. EYEBROW sequence: Front, side and top views (above, middle, bottom) of the 3D model for the (a) left camera, (b) right camera and (c) stereo setup sequences for $K = 5$.

Figure 4(A) shows the front, side and top views of the 3D reconstructions obtained for frames 16, 58 and 81 of the SMILE sequence. While the 3D shape appears to be well reconstructed, the deformations are not entirely well modelled. Note how the smile on frame 58 is not well captured. This was caused by the final regularization step proposed by Brand described in Section 3.2.2. We found that while this regularization step is essential to obtain good estimates for the rotation parameters it fails to capture the full deformations in the model. This is due to the fact that the assumption is that the deformations should be small relative to the mean shape so that most of the image motion is explained by the rigid component which results in a poor description of the deformations. However, we will see in the following section that the bundle adjustment step resolves the ambiguity between motion and shape parameters and succeeds in modelling the non-rigid deformations.

Figure 3 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

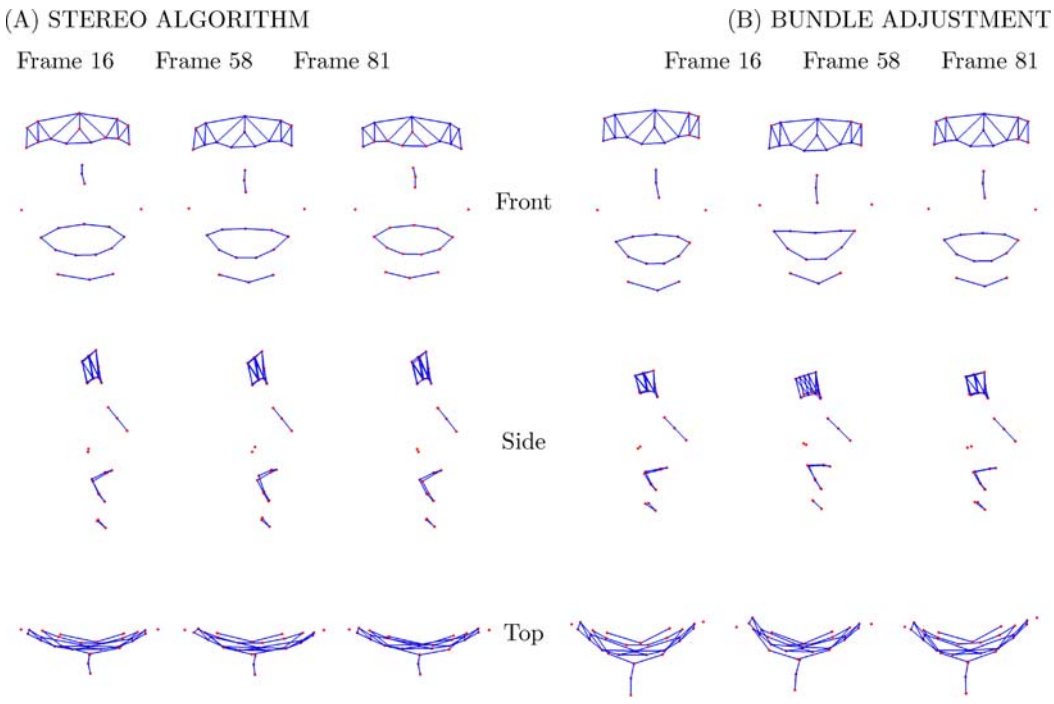

Figure 4. Front, side and top views of the reconstructed face for the SMILE sequence using the stereo algorithm (*left*) and after bundle adjustment (*right*). Reconstructions are shown for frames 16, 56 and 81 of the sequence.

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.

4.2. Results after Non-Linear Optimization

4.2.1. Real Sequences. In this section we show the results obtained after the final non-linear optimization step. The same video sequences that were used in the previous section were used here.

Figure 4 shows the front, side and top views of the 3D reconstructions before and after the bundle adjustment step for three frames of the SMILE sequence. The initial estimate is shown on the left and the results after bundle adjustment are shown on the right. While the initial estimate recovers the correct 3D shape, the deformations on the face are not well modelled. However, bundle adjustment succeeds to capture the flexible structure—notice how the upper lip is curved first and then straightened.

Figure 5 shows the results obtained for the estimated motion parameters and configuration weights using the initial stereo factorization method and the improved results after bundle adjustment. The bottom graphs show the rotation angles about the X, Y and Z axes recovered for each frame of the sequence for the left and right cameras (up to an overall rotation). The recovered angles for the left and right camera after bundle adjustment reflect very well the geometry of the stereo

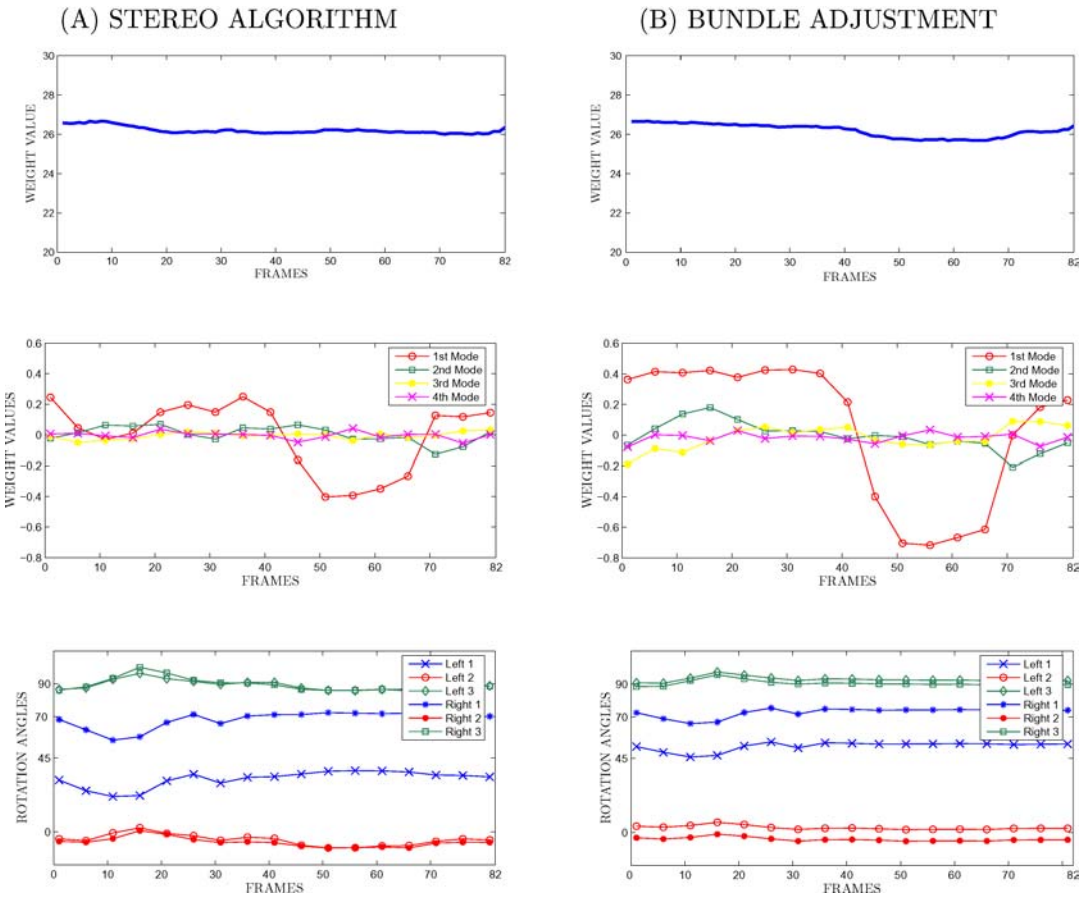

Figure 5. Values obtained for the rigid component (*top*), deformation weights (*middle*) and rotation angles (*bottom*) before (A) and after bundle adjustment (B) for the SMILE sequence

camera setup. This was such that both optical axes lay approximately on the XZ plane—therefore there was no relative rotation between the cameras about the X and Z axes—and the relative rotation about the Y axis was about 15deg. Note that these values are not ground truth and only approximate as they were not measured accurately. Also note that the rotation matrices for the right camera are calculated as $\mathtt{R}^{\mathtt{R}} = \mathtt{R}_{\mathrm{rel}}\, \mathtt{R}^{L}$ where $\mathtt{R}_{\mathrm{rel}}$ is the estimated relative orientation. Figure (5B) shows how the estimates of the rotations about the X and Z axes (in blue and green) for the left and right views are close to being zero. The relative rotation between left and right cameras about the Y axis (in red) is closer to 15deg after bundle adjustment than before.

Figure 5 also shows the evolution throughout the sequence of the values of the morph weights associated with the rigid component (top) and the 4 modes of deformation (middle). The values appear to be larger after bundle adjustment confirming that the non-linear optimization step has achieved to model the deformations of the face. It is also interesting to note how the first mode of deformation experiences a big change starting around frame 40 until frame 75. This coincides with the moment where the subject started and finished the smile expression. Similar results were obtained for the EYEBROW sequence although we have decided not to show them here.

4.2.2. Synthetic Data. In this section we have generated a sequence using a synthetic face model originally developed by Parke and Waters (1996). This is a 3D model which encodes 18 different muscles of the face. Animating the face model to generate facial expressions is achieved by actuating on the different facial muscles. In particular we have used a sequence where the head did not perform any rigid motion, only deformations a situation where, clearly, monocular algorithms would fail to compute the correct 3D shape and motion. The sequence was 125 frames long. The model deforms while rotating between frames 1 and 50, remains static and rigid until frame 100 and deforms once again between frames 100 and 125.

Once the model was generated we projected 160 points evenly distributed on the face, onto a pair of stereo cameras. The geometry of the cameras was such that both optical axes were lying on the XZ plane and each pointing inwards by 15 degrees. Therefore the relative orientation of the cameras about the Y axis was 30 degrees and 0 about the X and Z axes. The camera model used to project the points was a projective model however, the viewing conditions were such that the relief of the scene was small compared to the overall depth. Figure 8 shows the front, side and top views of the 3D model—ground truth—for three different frames in the sequence.

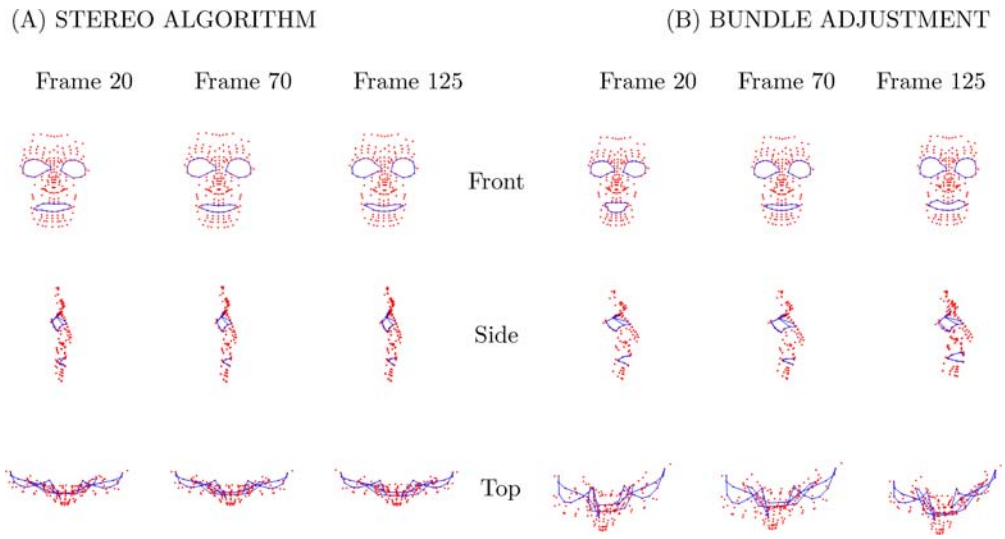

Figure 6. Front, side and top views of the reconstructed face for the synthetic sequence using the stereo algorithm (*left*) and after bundle adjustment (*right*). Reconstructions are shown for frames 20, 70 and 125 of the sequence.

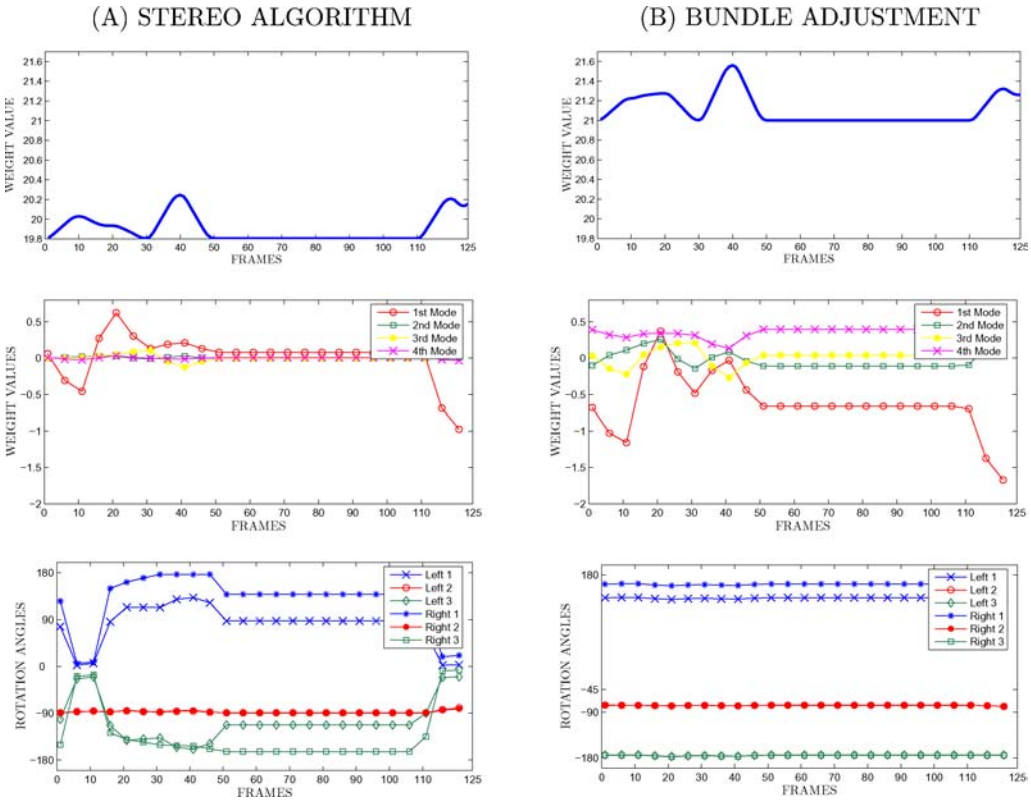

Figure 7. Values obtained for the rigid component (*top*), deformation weights (*middle*) and rotation angles (*bottom*) before (A) and after bundle adjustment (B) for the synthetic sequence.

Figure 6 shows the results obtained for the 3D reconstruction of the face before and after bundle adjustment. Not only is the overall rigid shape of the face best recovered after bundle adjustment—notice the depth estimates are not entirely satisfactory before bundle adjustment—but also the deformations.

Figure 7 shows the results for the estimated rotation angles and configuration weights before and after the non-linear optimization step. The results after bundle adjustment describe fairly accurately the geometry of the cameras and the deformation of the face. In particular, the stereo setup was such that there was no rigid motion of the face (only deformation), the optical axes of the left and right cameras lay on the XZ plane and the relative rotation of the cameras about the Y axis was constant and equal to 30 deg. In this case we have ground truth values for the relative orientation of the cameras since the sequence was generated synthetically. Notice how the values obtained for the rotation angles before bundle adjustment—left—exhibit some

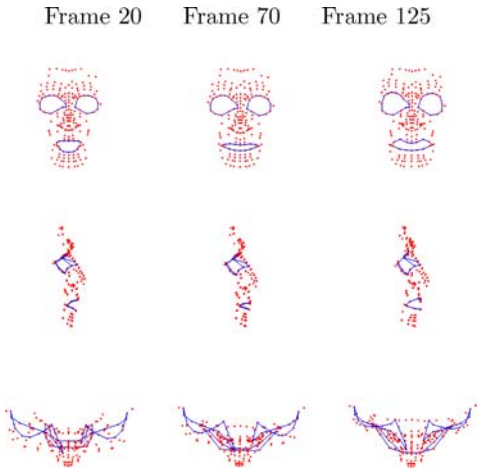

Figure 8. Front, side and top views of the 3D synthetic face used in the experiments. Frames 20, 70 and 125 of the sequence are shown. Note that there is no rigid motion in this sequence. The points which exhibit most deformation have been joined with a wire-frame for clarity.

problems around frames 10 and 115, when the deformations are occurring. After the bundle adjustment step the rotation angles are correctly estimated with a final result of 27 deg for the relative rotation about the Y axis and 0 deg about the X and Z axes—notice that the graphs for the left and right angles are superimposed.

Once more the estimated values for the deformation weights after bundle adjustment have larger values than before the optimization. This explains the fact that the model succeeds to explain the non-rigid deformations accurately. Interestingly, the coefficients remain constant between frames 50 and 110, when no deformations were occurring.

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. Furthermore, a novel non-linear optimization technique based on the bundle adjustment framework is used to refine both motion and shape components. The parameters obtained from the stereo factorization are used as the initial estimate in the minimization process. Experiments with real and synthetic data show that accurate 3D models can be achieved using the stereo factorization method and further improved applying the non-linear optimization scheme.

While our solution can provide a description of a nonrigid object in terms of motion, shape and deformations, some issues still require a proper solution. We plan to extend our framework to deal with missing measurements. This problem can be solved directly by including the uncertainty of the measurements in the non-linear minimization scheme. In this sense, we plan to extend the approach developed in Sugaya and Kanatani (2004) for the rigid case to the case of non-rigid deformable structure.

Our focus here has been on the recovery of 3D shape: we have simplified the temporal and spatial correspondence problem by using markers and some manual matching. However, a future avenue which we plan to explore is to exploit the rank constraint to obtain further stereo matches given a small set of correspondences.

Acknowledgments

The authors would like to thank the Royal Society European Science Exchange Programme, EPSRC Grant GR/S61539/01 and the Spanish Ministry of Science project TIC2002-00591 for financial support. Enrique Muñoz, Jose Miguel Buenaposada and Luis Baumela provided the code for the synthetic 3D face model. Thanks to Andrew Fitzgibbon for matlab functions for bundle adjustment. Alessio Del Bue holds a Queen Mary Studentship award.

References

Aanaes, H. and Kahl, F. 2002, Estimation of deformable structure and motion. In *Workshop on Vision and Modelling of Dynamic Scenes, ECCV'02*, Copenhagen, Denmark.

Bar-Itzhack, I.Y. 2000. New method for extracting the quaternion from a rotation matrix. *Journal of Guidance, Control and Dynamics*, 23(3):1085–1087.

Brand, M. 2001. Morphable models from video. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* Kauai, Hawaii.

Brand, M. and Bhotika, R. 2001. Flexible flow for 3D nonrigid tracking and shape recovery. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* Kauai, Hawaii, pp. 315–322.

Bregler, C., Hertzmann, A., and Biermann, H. 2000. Recovering non-rigid 3D shape from image streams. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* Hilton Head, South Carolina, pp. 690–696.

Del Bue, A. and Agapito, L. 2004. Non-rigid 3D shape recovery using stereo factorization. In *Asian Conference on Computer Vision (ACCV2004)*, vol. 1. Jeju, South Korea.

Del Bue, A., Smeraldi, F., and Agapito, L. 2004, Non-rigid structure from motion using nonparametric tracking and non-linear optimization. In *Workshop in Articulated and Nonrigid Motion ANM04, held in Conjunction with CVPR2004*. Washington.

Essa, I. and Basu, S. 1996. Modeling, tracking and interactive animation of facial expressions and head movements using input from video. In *Proceedings of Computer Animation Conference*. Geneva, Switzerland.

Horn, B. 1987, Closed form solutions of absolute orientation using unit quaternions. *J. Optical Soc. of America A*. 4(4): 629–642.

Irani, M. 1999, Multi-frame optical flow estimation using subspace constraints. In *Proc. 7th International Conference on Computer Vision,* Kerkyra, Greece.

Parke, F.I. and Waters, K. 1996. *Computer Facial Animation*. A.K. Peters, Ltd.

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

Sugaya, Y. and Kanatani, K. 2004. Extending interrupted feature point tracking for 3-D affine reconstruction. *IEICE Transactions on Information and System*, E87-D(4):1031–1039.

Tan, J. and Ishikawa, S. 2001. Deformable shape recovery by factorization based on a spatiotemporal measurement matrix. *Computer Vision and Image Understanding*, 82:101–109.

Tomasi, C. and Kanade, T. 1991. Shape and motion from image streams: A factorization method. *International Journal in Computer Vision*, 9(2):137–154.

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

Tresadern, P. and Reid, I. 2003. Synchronizing image sequences of non-rigid objects. In *Proc. British Machine Vision Conference*, Norwich.

Triggs, B., McLauchlan, P., Hartley, R., and Fitzgibbon., A. 2000. Bundle adjustment—A modern synthesis. In W. Triggs, A. Zisserman, and R. Szeliski (Eds.), *Vision Algorithms: Theory and Practice*, LNCS. Springer Verlag, pp. 298–375.

Vetter, T. and Blanz, V. 1999. A morphable model for the synthesis of 3D faces. In *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics*, pp. 187–194.

Xiao, J., Chai, J., and Kanade, T. 2004. A closed-form solution to non-rigid shape and motion recovery. In *The 8th European Conference on Computer Vision (ECCV 2004)*.