# 14

# Non-rigid Stereo-motion

Alessio Del Bue[1] and Lourdes Agapito[2]
[1] *Instituto Superior Técnico*
*Portugal*
[2]*Queen Mary University of London*
*United Kingdom*

## 1. Stereo, motion and structure

Using a calibrated stereo pair is a common and practical solution to obtain reliable 3-D reconstructions. In its simpler formulation, once the stereo rig is calibrated, the depth of points in the image is estimated by applying triangulation (Trucco & Verri, 1998). In order to obtain accurate depth estimates, the cameras are usually separated from each other by a significant baseline thus creating widely spaced observations of the same object. The disadvantage of this configuration though, is that having a wide baseline makes the matching of features between pairs of views a more challenging problem.

On the other hand, the task of computing temporal tracks from single camera sequences is relatively easier since the images are closely spaced in time. As a drawback, disparities may be insufficient to obtain a reliable depth estimation and, as a result, longer sequences are needed to infer the 3-D structure. Particularly, in the case of non-rigid structure, a sufficient overall rigid motion is necessary to allow the algorithms to estimate the reconstruction parameters correctly.

Hence, a question of relevant interest is the feasibility of an approach that efficiently fuses the positive aspects of both methods. The problem of recovering 3-D structure using a stereo-rig moving in time or a stereo rig looking at a moving object has been defined for the rigid case as the stereo-motion problem (Waxman & Duncan, 1998; Dornaika & Chung, 1999; Stein & Shashua 1998; Mandelbaum et al., 1999). Ho and Chung (Ho & Chung, 2000) were the first to formulate this problem within the factorization scenario. Following a similar direction, we introduce a multi-camera motion model that is able to deal with a time-varying shape and we present a linear solution based on the factorization framework that is subsequently optimized with a non-linear procedure.

Schematically, the chapter is structured as follows. The inference of 3-D structure from an image sequence (single camera case) is introduced in Section 2, focusing particularly on the case of a deforming body. The next section will show how the presented framework, based on a factorization solution of the problem, can be consistently extended to the case of multiple cameras viewing a deforming body and a linear solution to the problem will be as well provided. Section 3 introduces a non-linear optimization strategy to refine the linear solution obtained with the previous method and Section 4 will validate the presented approaches with experimental tests on synthetic and real deforming bodies. Finally we present further considerations over the presented framework and its future extensions.



Source: Scene Reconstruction, Pose Estimation and Tracking, Book edited by: Rustam Stolkin,
ISBN 978-3-902613-06-6, pp.530, I-Tech, Vienna, Austria, June 2007

## 2. Background: the monocular case

### 2.1 Rigid factorization

Tomasi and Kanade's factorization algorithm (Tomasi & Kanade, 1992) 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 2-D image coordinates of a set of $P$ points tracked throughout $F$ frames into a measurement matrix $W_{2F \times P}$. Assuming affine viewing conditions, the measurement matrix can be expressed analytically as a product of two matrices: $W = M\,S$ where $M$ is a $2F \times 3$ motion matrix which expresses the pose of the camera and $S$ is the $3 \times P$ shape matrix which contains 3-D 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: $\mathrm{SVD}(W) = U_{2F \times 3}\, D_{3x3}\, V_{3xP} = M_{2F \times 3}\, S_{3xP}$. 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 3-D structure (Bregler et al., 2000). Here, a model is needed to express the deformations of the 3-D shape in a compact way. The chosen model is a simple linear model where the 3-D shape of any specific configuration of a non-rigid object is approximated by a linear combination of a set of $D$ basis-shapes which represent the $D$ principal modes of deformation of the object for $P$ points. A perfectly rigid object would correspond to the situation where $D$=1. Each basis-shape ($S_1\ S_2\ \ldots\ S_D$) is a $3 \times P$ matrix which contains the 3-D locations of $P$ object points for that particular mode of deformation. The 3-$D$ shape of any configuration can then be expressed as a linear combination of the basis-shapes $S_i$:

$$S = \sum_{d=1}^{D} l_d S_d \qquad S, S_d \in \Re^{3 \times P} \quad l_d \in \Re \tag{1}$$

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

$$W_i = \begin{bmatrix} u_{i1} & \ldots & u_{iP} \\ v_{i1} & \ldots & v_{iP} \end{bmatrix} = R_i \left( \sum_{d=1}^{D} l_{id} S_d \right) + T_i \tag{2}$$

where

$$R_i = \begin{bmatrix} r_{i1} & r_{i2} & r_{i3} \\ r_{i4} & r_{i5} & r_{i6} \end{bmatrix} \tag{3}$$

is a 2 x 3 matrix which contains the first and second rows of the camera rotation matrix and $T_i$ 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 $l_i$ coefficients. We may eliminate the translation vector $T_i$ by registering image points to the

centroid in each frame. In this way, the 3-D coordinate system will be centred at the centroid of the shape $S$. If the same $P$ points can be tracked throughout an image sequence we may stack them into a $2F$ x $P$ measurement matrix $W$ and we may write:

$$W = \begin{bmatrix} W_1 \\ \vdots \\ W_F \end{bmatrix} = \begin{bmatrix} l_{11}R_1 & \dots & l_{1D}R_1 \\ \vdots & & \vdots \\ l_{F1}R_F & \dots & l_{FD}R_F \end{bmatrix} \begin{bmatrix} S_1 \\ \vdots \\ S_D \end{bmatrix} = MS \tag{4}$$

Since $M$ is a $2F$ x $3D$ matrix and $S$ is a $3D$ x $P$ matrix, the rank of $W$ when no noise is present must be at most 3D. 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 3-D pose, and shape (basis-shapes and deformation coefficients) of the object from the correspondence points stored in $W$.

### 2.3 Non-rigid factorization

The rank constraint on the measurement matrix $W$ can be easily imposed by truncating the SVD of $W$ to rank $3D$. This will factor $W$ into a motion matrix $\widetilde{M}$ and a shape matrix $\widetilde{S}$. However, the result of the factorization of $W$ is not unique since any invertible $3D$ x $3D$ matrix $Q$ can be inserted in the decomposition leading to the alternative factorization $W = (\widetilde{M}Q)(Q^{-1}\widetilde{S})$. The focal problem is to find a transformation matrix $Q$ that imposes the replicated block structure on the motion matrix $\widetilde{M}$ shown in (4) 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 $Q$ to upgrade the reconstruction to a metric one can be solved linearly (Tomasi & Kanade, 1992), in the non-rigid case imposing the appropriate repetitive structure to the motion matrix $\widetilde{M}$ 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 3-D shape and motion.

Most of the model-free approaches to non-rigid factorization are based either on closed-form solutions (Xiao et al., 2004), assuming prior knowledge over the structure of the basis shapes, or iterative non-linear optimisation techniques (Brand, 2005; Del Bue et al., 2007; Torresani et al., 2001), which require an appropriate initialisation in order to converge.

## 3. The stereo camera case

The main contribution we present here 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. (Torresani et al., 2001) first introduced the factorization problem for the multiple camera case but they did not provide an algorithm or any experimental results.

### 3.1 The stereo motion model

When two cameras are viewing the same scene, the measurement matrix $W$ will contain the image measurements from the left and right cameras resulting in a $4F$ x $P$ matrix where $F$ is the number of frames and $P$ the number of points. Assuming that not only the single-frame

tracks but also the stereo correspondences are known we may write the measurement matrix $W$ as:

$$W = \begin{bmatrix} W^L \\ W^R \end{bmatrix} \tag{5}$$

where for each frame $i$ the stereo correspondences are:

$$W_i^L = \begin{bmatrix} w_{i1}^L & \dots & w_{iP}^L \end{bmatrix} \quad W_i^R = \begin{bmatrix} w_{i1}^R & \dots & w_{iP}^R \end{bmatrix} \tag{6}$$

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

$$W = \begin{bmatrix} l_{11}R_1^L & \dots & l_{1D}R_1^L \\ \vdots & & \vdots \\ l_{F1}R_F^L & \dots & l_{FD}R_F^L \\ l_{11}R_1^R & \dots & l_{1D}R_1^R \\ \vdots & & \vdots \\ l_{F1}R_F^R & \dots & l_{FD}R_F^R \end{bmatrix} \begin{bmatrix} S_1 \\ \vdots \\ S_D \end{bmatrix} \tag{7}$$

where $R^L$ and $R^R$ are the rotation components for the left and right cameras. Once more, we have eliminated the translation for both cameras 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. This assumption is generally true in a symmetric stereo setup where $f$ and $Z_{avg}$ are usually the same for both cameras.

It is also possible to express the stereo motion matrix $M$ 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 in terms of the matrix that encodes their relative orientation matrix $R_{rel}$ such that: $R^R = R_{rel}\, R^L$. The motion matrix $M$ in equation (6) can be consequently expressed as:

$$M = \begin{bmatrix} l_{11}R_1^L & \dots & l_{1D}R_1^L \\ \vdots & & \vdots \\ l_{F1}R_F^L & \dots & l_{FD}R_F^L \\ l_{11}R_{rel}R_1^L & \dots & l_{1D}R_{rel}R_F^L \\ \vdots & & \vdots \\ l_{F1}R_{rel}R_F^L & \dots & l_{FD}R_{rel}R_F^L \end{bmatrix} \tag{8}$$

### 3.2 Non-rigid stereo factorization

Once more the rank of the measurement matrix W is at most 3D since M is a 4F x 3D matrix and S is a 3D x 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 W may be factorized into the product of a motion matrix *M* and a shape matrix *S* by truncating the SVD of *W* to rank 3*D* (see section 2.3):

$$SVD(W) = \widetilde{W} = \begin{bmatrix} \widetilde{M}^L \\ \widetilde{M}^R \end{bmatrix} \widetilde{S} \tag{9}$$

### 3.3 Computing the transformation matrix *Q*

The result of the factorization is not unique since $\widetilde{W} = (\widetilde{M}Q)(Q^{-1}\widetilde{S})$ would give an equivalent factorization. We proceed to apply the metric constraint by correcting each 4*F* x 3 vertical block in $\widetilde{M}$ independently. Note that in this case we have used five constraints per frame: 2 orthogonality constraints (one from each camera) and 3 equal norm constraints (computed from rows 2*i*-1, 2*i*, 2*i*+2*F*-1, 2*i*+2*F* of the motion matrix $\widetilde{M}$ where *i* is a generic frame). Each vertical block will then be corrected as: $\hat{M}_d \leftarrow \widetilde{M}_d Q_d$. The overall transformation *Q* is a block diagonal matrix such that:

$$Q = \begin{bmatrix} Q_1 & 0 & \dots & 0 \\ 0 & Q_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & Q_D \end{bmatrix} \tag{10}$$

The shape matrix will be corrected with the inverse of the block-diagonal transformation: $\hat{S} \leftarrow Q^{-1}\widetilde{S}$.

### 3.4 Factorization of the motion matrix *M*

In the stereo case we factorize each 4 x 3*D* sub-block of the motion matrix (which contains left and right measurements for each frame *i*) into its truncated 2 x 3 rotation matrices $R_i^L$ and $R_i^R$ and the deformation weights $l_{id}$ using an orthonormal decomposition. The structure of the sub-blocks can be expressed as:

$$\mathbf{M_i} = \begin{bmatrix} \mathbf{M}^{\mathbf{L}}_{\mathbf{i}1} & \dots & \mathbf{M}^{\mathbf{L}}_{\mathbf{iD}} \\ \mathbf{M}^{\mathbf{R}}_{\mathbf{i}1} & \dots & \mathbf{M}^{\mathbf{R}}_{\mathbf{iD}} \end{bmatrix} = \begin{bmatrix} \mathbf{l}_{\mathbf{i}1}\begin{bmatrix} \mathbf{R}^{\mathbf{L}}_{\mathbf{i}} \\ \mathbf{R}^{\mathbf{R}}_{\mathbf{i}} \end{bmatrix} & \dots & \mathbf{l}_{\mathbf{iD}}\begin{bmatrix} \mathbf{R}^{\mathbf{L}}_{\mathbf{i}} \\ \mathbf{R}^{\mathbf{R}}_{\mathbf{i}} \end{bmatrix} \end{bmatrix} \tag{11}$$

The approach used to estimate the rotation components for the left and right cameras use the orthogonality constraints on each block of the motion matrix. Since now we have 4 rows per frame, we arrange the motion sub-blocks such that:

$$\widetilde{\mathbf{M}}_{\mathbf{i}} \rightarrow \overline{\mathbf{M}}_{\mathbf{i}} = \begin{bmatrix} \mathbf{l}_{\mathbf{i}1}\begin{bmatrix} \bar{\mathbf{r}}^{\mathbf{L}}_{\mathbf{i}} \\ \bar{\mathbf{r}}^{\mathbf{R}}_{\mathbf{i}} \end{bmatrix} & \mathbf{l}_{\mathbf{i}2}\begin{bmatrix} \bar{\mathbf{r}}^{\mathbf{L}}_{\mathbf{i}} \\ \bar{\mathbf{r}}^{\mathbf{R}}_{\mathbf{i}} \end{bmatrix} & \dots & \mathbf{l}_{\mathbf{iD}}\begin{bmatrix} \bar{\mathbf{r}}^{\mathbf{L}}_{\mathbf{i}} \\ \bar{\mathbf{r}}^{\mathbf{R}}_{\mathbf{i}} \end{bmatrix} \end{bmatrix} \tag{12}$$

where $\bar{r}_i^L = [r_{i1}^L \ldots r_{i6}^L]^T$ is a column vector which contains the coefficients of the left rotation matrix $R_i^L$ and similarly for $\bar{r}_i^R$. Post-multiplying the rearranged matrix $M_i$ by the 2$D$ unity vector $c = [1 \ldots 1]^T$ gives a column vector $a_i$:

$$a_i = \overline{M}_i c \tag{13}$$

which may be rearranged into a 4 x 3 matrix $A_i$ with analytic form:

$$A_i = \begin{bmatrix} kr_{i1}^L & kr_{i2}^L & kr_{i3}^L \\ kr_{i4}^L & kr_{i5}^L & kr_{i6}^L \\ kr_{i1}^R & kr_{i2}^R & kr_{i3}^R \\ kr_{i4}^R & kr_{i5}^R & kr_{i6}^R \end{bmatrix} = \begin{bmatrix} A_i^L \\ A_i^R \end{bmatrix} \tag{14}$$

where $k = l_{i1} + \ldots + l_{iD}$. Since $R^L$ and $R^R$ are orthonormal matrices, the following equation is satisfied:

$$\begin{bmatrix} R_i^L & 0 \\ 0 & R_i^R \end{bmatrix}_{4\times 6} \begin{bmatrix} \left(A_i^L\right)^T & 0 \\ 0 & \left(A_i^R\right)^T \end{bmatrix}_{6\times 4} = \sqrt{\begin{bmatrix} A_i^L \left(A_i^L\right)^T & 0 \\ 0 & A_i^R \left(A_i^R\right)^T \end{bmatrix}_{4\times 4}} \tag{15}$$

Therefore, a linear least-squares fit can be obtained for the rotation matrices $R^L$ and $R^R$ and the weights $l_{id}$ can be subsequently estimated by rearranging the sub-block matrix $M_i$ in a different way from equation (11):

$$\widetilde{M}_f \rightarrow M_i' = \begin{bmatrix} l_{i1}\bar{r}_i^T \\ \ldots \\ l_{iD}\bar{r}_i^T \end{bmatrix} \tag{16}$$

where $\bar{\mathbf{r}}_\mathbf{i} = \left[\left(\bar{\mathbf{r}}_\mathbf{i}^\mathbf{L}\right)^\mathbf{T} \quad \left(\bar{\mathbf{r}}_\mathbf{i}^\mathbf{R}\right)^\mathbf{T}\right]^\mathbf{T}$. The configuration weights for each frame $i$ are then derived exploiting the orthonormality of $R_i$ since:

$$M_i'\bar{r}_i^T = \begin{bmatrix} l_{i1}\bar{r}_i^T\bar{r}_i \\ \ldots \\ l_{iD}\bar{r}_i^T\bar{r}_i \end{bmatrix} = 4\begin{bmatrix} l_{i1} \\ \ldots \\ l_{iD} \end{bmatrix} \tag{17}$$

The linear estimation can be furthermore refined by using a regularization scheme similar to the one used by Brand in his flexible factorization algorithm (Brand, 2001) which enforces the deformations in $\widetilde{S}$ being 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 similar to the shape regularization used by other authors (Torresani et al. 2001; Aanæs & Kahl, 2002).

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 (Brand, 2001). While this new method improves the quality of the 3-D reconstructions with respect to those using a monocular sequence, it still performs a partial upgrade of the motion and 3-D structure matrices since $Q$ is computed initially as a block diagonal matrix and then corrected with Brand's flexible factorization. In order to obtain a solution which completely respects the structure of equation (7), we will now describe a non-linear optimization scheme which renders the appropriate structure to the motion matrix, allowing to disambiguate between the motion and shape parameters.

## 4. Stereo non-linear optimization

### 4.1 The non-rigid cost function

The goal is to estimate the motion parameters $R_i$, the relative orientation between cameras $R_{rel}$, the 3-D basis shapes $S_d$ and the deformation weights $l_{id}$ such that the distance between the measured image points $w_{ij}$ and the reprojection of the estimated 3-D points is minimised. However, the coordinates in $W$ are extracted by a measurement process and, therefore, they are affected by noise or by a certain degree of uncertainty $n_{ij}$. The measured coordinates $w_{ij}$ for the left and right camera at frame $i$ can be expressed in terms of the exact measurements $x_{ij}$ such that:

$$w_{ij} = \begin{bmatrix} w_{ij}^L \\ w_{ij}^R \end{bmatrix} = x_{ij} + n_{ij} \tag{18}$$

The projection equation for a 3-D point $j$ in image frame $i$ is given by:

$$x_{ij} = \begin{bmatrix} R_i^L \sum_d l_{id} S_{dj} \\ R_{rel} R_i^L \sum_d l_{id} S_{dj} \end{bmatrix} \tag{19}$$

where $x_{ij}$ are the image coordinates of the point for the left and right cameras and $S_i$ is the $3D$ x 1 parameterisation of the shape basis for a deformable point $j$ such that:

$$S_j = \begin{bmatrix} S_{1j} \\ S_{2j} \\ \vdots \\ S_{Dj} \end{bmatrix} \tag{20}$$

with the 3-vector $S_{dj}$ defining the $d$ basis component for point $j$.

Following equation (18), the uncertainty over the measurements is obtained from the residual given by $n_{ij} = w_{ij} - x_{ij}$. This residual is generally referred to as the reprojection error of the image coordinates in the literature and it expresses the difference between the image coordinates given the estimated model parameters and the measured data. Hence, it is possible to recast the problem of estimating the non-rigid structure and motion parameters by minimizing the norm of the reprojection error of all the points in all the frames such that:

$$\min_{R_i^L R_{rel} I_{id} S_{dj}} \sum_{i,j}^{F,P} \left\| n_{ij} \right\|^2 = \min_{R_i^L R_{rel} I_{id} S_{dj}} \sum_{i,j}^{F,P} \left\| w_{ij} - x_{ij} \right\|^2 \tag{21}$$

Note that the error is a sum of *FP* quadratic cost functions. Assuming the noise can be modelled with a Gaussian distribution, the minimization of equation (21) provides a true Maximum Likelihood (ML) estimate of the parameters.

The definition of this non-rigid cost function could rise two major criticisms. First, the number of parameters can increase dramatically with the number of frames composing the scene and the complexity of the modelled object. This may render the minimization of equation (21) computationally unfeasible given the size of the parameter space. Second, the high non-linearity of the cost function is likely to produce multiple minima which would result in a difficult convergence to the global minimum of the function. The solution proposed is a reformulation of bundle-adjustment techniques for deformable structure from motion which we describe in the following sections.

### 4.2 A bundle-adjustment approach to deformable modelling

The non-linear optimization of the cost function in (21) is achieved using a Levenberg-Marquardt (Levenberg, 1944; Marquardt 1963; Moré, 1977) iterative minimization scheme modified to take advantage of the sparse block structure of the matrices involved. This method is generically termed bundle-adjustment in the computer vision (Triggs et al. 2000) and photogrammetry (Atkinson, 1996) communities and it is a standard procedure successfully applied to numerous 3-D reconstruction tasks (Hartley & Zisserman, 2000). Our main contribution here is an analysis of its applicability to the non-rigid modelling framework.

In the next section, we will review the concepts involved in bundle-adjustment (Levenberg-Marquardt minimization and sparse computation) and reformulate the factorization framework as a non-linear, large-scale minimization problem.

### 4.3 Levenberg-Marquardt minimization

Levenberg-Marquardt methods use a mixture of Gauss-Newton and gradient descent minimization schemes switching from the first to the second when the estimated Hessian of the cost function is close to being singular. An algorithm with mixed behaviors usually obtains a higher rate of success in finding the correct minimum than other approaches. Other similar second-order or quasi-Newton algorithms may be used to minimize the cost function. However, Levenberg-Marquardt techniques have been studied and tested thoroughly in many Computer Vision applications (Hartley & Zisserman, 2000) and they have been found to deliver satisfactory results. Examples are mostly given for classical inference problems in Computer Vision such as fundamental matrix computation (Bartoli & Sturm, 2004), camera calibration (Pollefeys, 1999), and 3-D sparse reconstruction (Guilbert et al., 2004). However second-order methods have been successfully applied to less conventional geometric problems such as model-based face reconstruction (Fua, 2000), mosaicing (McLauchlan & Jaenicke, 2002) and reconstruction of curves (Berthilsson, 2001).

Most of the computational burden of iterative second-order methods is represented by the Gauss-Newton descent step, each iteration of which requires the calculation of the inverse of the Hessian of the cost function *C*. Specifically to the deformable factorization case, *C* can be

expressed in terms of the *N*-vector $\Theta$ containing the model parameters such that $\Theta = (\Theta_{l1},\ldots,\Theta_{lF},\ \Theta_{R1},\ldots,\Theta_{RF},\Theta_{S1},\ldots,\Theta_{SP})^T$, where $\Theta_{li}$, $\Theta_{Ri}$ and $\Theta_{Sj}$ represent respectively the parameters for the configuration weights, orthographic cameras and 3-D basis shapes for each view and each point. Hence, the cost function *C* can be written as a sum of squared residuals:

$$C(\Theta) = \sum_{i,j}^{F,P} \left\| n_{ij} \right\|^2 \tag{22}$$

where the residual for each frame and each point can be expressed as a 2*FP* x 1 vector *n* such that $n = [n_{11}^T \ldots n_{FP}^T]^T$. At each iteration *t* of the algorithm, an update $\Delta^t$ is computed in order to descend to the minimum of the cost function such that the new set of parameters is given by $\Theta^{t+1} = \Theta^t + \Delta^t$. By dropping the iteration index *t* for notation clarity, it is necessary to express the generic increment $\Delta$ in the model parameters as a second order Taylor expansion assuming local linearities in the cost function such that:

$$C(\Theta + \Delta) \approx C(\Theta) + g^T \Delta + \frac{1}{2}\Delta^T H \Delta \tag{23}$$

where $g = J^T n$ is the *N* x 1 gradient vector and *H* is the *N* x *N* Hessian matrix that can be approximated as $H = J^T J$ (Gauss-Newton approximation of the Hessian matrix; see (Triggs et al. 2000) for details) with $J = \frac{\partial n}{\partial \Theta}$ representing the 2*FP* x *N* Jacobian matrix in the model parameters. In order to find the increment $\Delta$, the minimum of the quadratic function $e = g^T \Delta + \frac{1}{2}\Delta^T H \Delta$ is computed by imposing $\frac{\partial e}{\partial \Delta} = 0$. Thus, the expression of the Gauss-Newton descent step can be finally expressed as:

$$H\Delta = -g \tag{24}$$

Levenberg-Marquardt algorithms differ from a pure Gauss-Newton method since they apply a damping term to equation (24) obtaining:

$$(H + \lambda I)\Delta = -g \tag{25}$$

The added term $\lambda I$ has a twofold effect in the minimization. Firstly, by modifying the parameter $\lambda$, it is possible to control the behavior of the algorithm that can switch between first order (for high values of $\lambda$) and second order (low $\lambda$) iterations. Secondly, $\lambda I$ makes the solution of (17) numerically stable by forcing that $H + \lambda I$ is a full-rank matrix and thus properly invertible.

### 4.4 Sparse structure of the Jacobian

Solving for the normal equations in equation (22) is a problem of complexity $O(N^3)$ and this step has to be repeated at each iteration. In order to render the computation feasible as the number of parameters increases, it is possible to exploit the sparse structure of the Jacobian *J*. Motion components (configuration weights and camera parameters) are unrelated

between different views and, similarly, structure components are unrelated between different point trajectories. As a result, the Jacobian matrix contains a large number of entries for which the partial derivatives are zero.
It is possible to solve for the increment $\Delta$ in (25) efficiently by calculating the inverse of $H$ using the sparse structure of $J$. Standard approaches for sparse computation are described in (Hartley & Zisserman, 2000) and (Triggs et al. 2000). Notice that, again, this property is valid for any rigid and non-rigid factorization model, since the sparseness relation is given by the independency between motion parameters (for each frame) and 3-D structure (for each point) in the multi-view cost function and thus independent of the chosen model.

### 4.5 Proposed implementation

The cost function of a deformable object presents more degrees of freedom than in the rigid case, which could lead to the existence of multiple local minima for the motion, deformation and structure components. It is possible to reduce the chance of falling into local minima by carefully designing the algorithm with respect to the following two aspects: initialisation and model parameterisation.
The camera matrices $R_i$ are parameterised using unit quaternions (Horn 1987) giving a total of 4 x $F$ rotation parameters, where $F$ is the total number of frames. Quaternions ensure that there are no strong singularities and that the orthonormality of the rotation matrices is preserved by merely enforcing the normality of the 4-vector. This would not be the case with the Euler angle or the rotation matrix parameterisations, where orthonormality of the rotations is more complex to preserve. The quaternion normalization is directly enforced in the cost function by dividing the quaternion with its norm. Indeed, in an initial implementation the 3-D pose was parameterised using the 6 entries of the rotation matrices $R_i$ and $R_{rel}$, however the use of quaternions led to improved convergence and to much better results for the rotation parameters and the 3-D pose.
The method proposed by Bar-Itzhack (Bar-Itzhack, 2000) in an attitude control context is used to obtain the quaternions from the set of rotation matrices $R_i$. The algorithm has the main advantage of yieldieng the closest quaternion representation if the constraints of matrix orthonormality are not exactly satisfied. This eventuality usually appears during the initialisation of the non-linear optimization scheme after the first computation of the corrective transform $Q$. Schematically, the method first defines the matrix $B$ given the singular elements $\{r_{mn}\}$ belonging to a generic 3 x 3 rotation matrix $R_i$:

$$B = \frac{1}{3}\begin{bmatrix} r_{11}-r_{22}-r_{33} & r_{21}+r_{12} & r_{31}+r_{13} & r_{23}+r_{32} \\ r_{21}+r_{12} & r_{22}-r_{11}-r_{33} & r_{32}+r_{23} & r_{31}-r_{13} \\ r_{31}+r_{13} & r_{32}+r_{23} & r_{33}-r_{22}-r_{11} & r_{12}-r_{21} \\ r_{23}-r_{32} & r_{31}-r_{13} & r_{12}-r_{21} & r_{11}+r_{22}+r_{33} \end{bmatrix} \tag{26}$$

The algorithm then follows with the following three steps:

- Compute the eigenvalues of $B$.
- Find the largest eigenvalue $\lambda_{max}$.
- Extract the eigenvector of $B$ which corresponds to $\lambda_{max}$.

The given eigenvector is the closest quaternion to the matrix $R$. In the case of an exact orthonormal matrix we would obtain $\lambda_{max} = 1$. Finally, the structure is parameterised with the (3 x $D$) x $P$ coordinates of the $S_d$ shape bases and the $D$ x $F$ deformation weights $l_{id}$.
The linear method proposed in the previous section is used to obtain an initial estimation of the model parameters. The initial estimate for the constant relative orientation $R_{rel}$ between the left and right cameras is estimated from the camera matrices $R^L$ and $R^R$ using a least squares estimation. If the internal and external calibration of the stereo rig were known in advance after a process of calibration or self-calibration, an alternative initialisation could be computed by recovering the 3-D structure and performing Principal Component Analysis (PCA) on the data to obtain an initial estimate for the basis shapes and the coefficients. However, our choice was to use an initialisation that does not require a pre-calibration of the cameras.

## 5. Experimental results

This section shows the performance of the proposed stereo-motion algorithms. Firstly, synthetic stereo sequences are generated under different Gaussian noise and deformation conditions to assess the validity of the method. A further synthetic test using a computer graphic (CG) generated face model will show the behaviour of the configuration weights and motion components when the object in the stereo sequence is static (only deforming). We then carry out some real experiments where the object underwent only a small amount of rigid motion (apart from the deformations) and we will show the improvement of the method by comparing the output of the monocular factorization and the stereo algorithms. Non-linear optimization will follow the computed linear solutions.

### 5.1 Experiments with a synthetic non-rigid cube

A set of deformable points is randomly sampled inside a cube of 50 x 50 x 50 units. A minimal overall rigid motion is introduced to avoid possible ambiguities arising from a completely static object. The 3-D structure computed at each frame is then projected with 2 orthographic cameras displaced by a baseline of 20 units and relatively rotated by 30 degrees about the y-axis. Finally, different levels of Gaussian noise (σ = 0.5, 1, 1.5, 2) are added to the measurements obtained by the stereo pair. Notice that the setup is constructed in such way that the overall rigid motion is not enough to reconstruct the sequences using monocular factorization followed by bundle adjustment. We performed a test and we obtained a relative 3-D reconstruction error of 50% resulting in a meaningless reconstruction.
The results show the plots for the relative 3-D error, rotation error and reprojection error tested over 25 trials with a 3-D shape deforming with different numbers of basis shapes (see Fig. 1) and different degrees of non-rigidity (see Fig. 2) defined as $ratio = \|S_{rigid}\| / \|S_{nonrigid}\|$.
Notice in this case a higher reconstruction error of the relative 3-D structure compared to the monocular case with higher degrees of deformation.

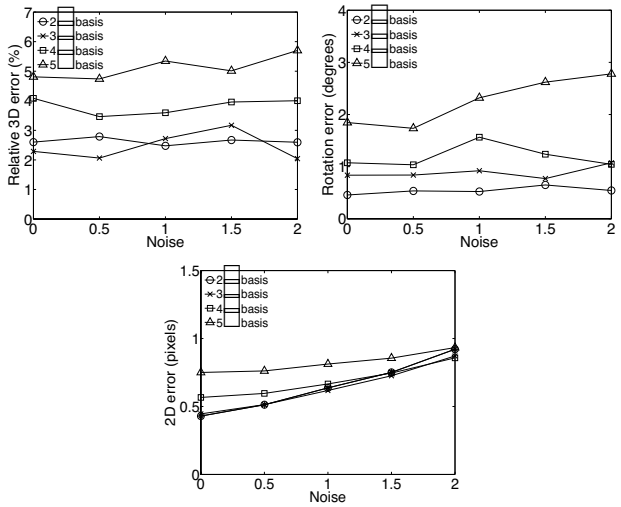



Figure 1. Relative 3-D error (%), r.m.s. rotation error (degrees) and 2-D reprojection error for the synthetic experiments with a stereo pair for different basis shapes $D$ = 2 …5 and increasing levels of Gaussian noise. The ratio of non rigidity is fixed to 40% for all the trials.

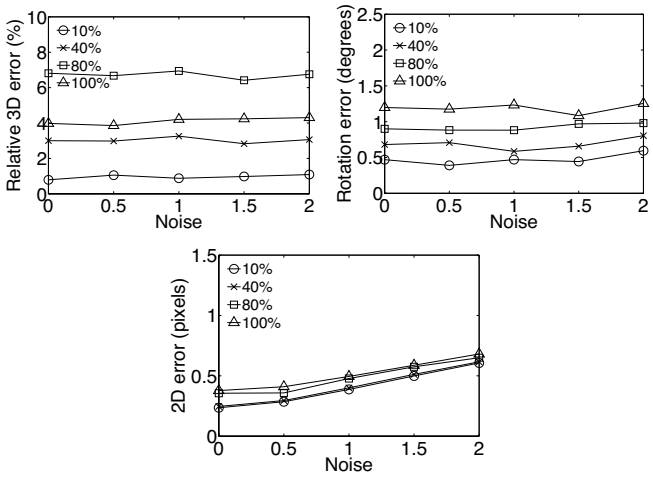



Figure 2. Relative 3-D error (%), r.m.s. rotation error (in degrees) and 2-D reprojection error for the synthetic experiments for different ratios of deformation (10%, 40%, 80%, 100%) and increasing levels of Gaussian noise.

### 5.2 Synthetic experiments with a CG generated face

In this section we have generated a sequence using a synthetic face model originally developed by (Parke & Waters, 1996). This is a 3-D 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 3-D shape and motion. The sequence was 125 frames long. The model deforms between frames 1 and 50, remains static and rigid until frame 100 and deforms once again between frames 100 and 125.

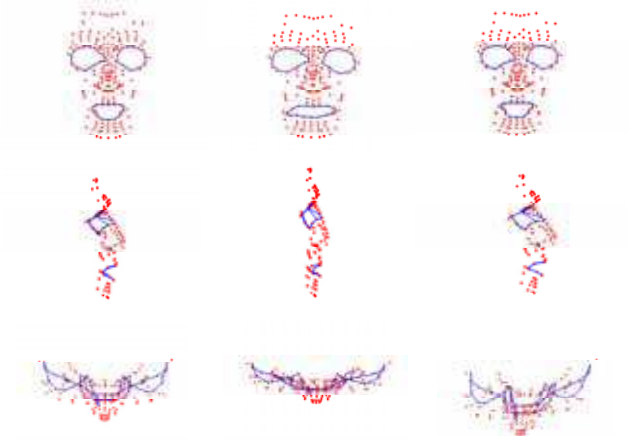


Figure 3. Front, side and top views of the 3-D synthetic face for frame 20. The first column shows the shape ground truth while the following two columns present the 3-D reconstructions for the linear and bundle adjustment algorithms. Deformations are present mainly in the mouth region. Notice that the face does not perform rigid motion for the whole sequence.

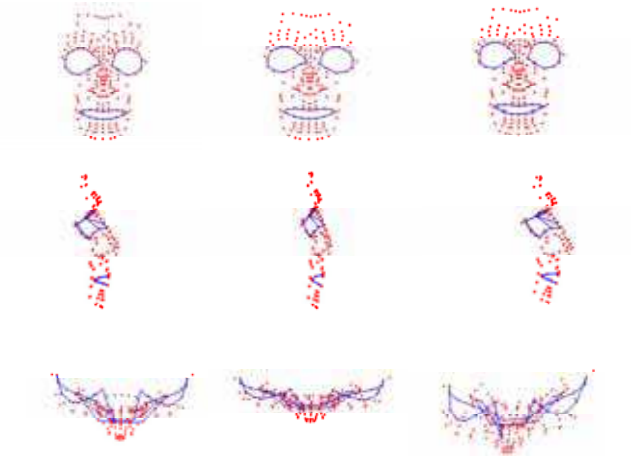


Figure 4. Front, side and top views of the 3-D synthetic face for frame 70. The first column shows the shape ground truth while the following two columns present the 3-D reconstructions for the linear and bundle adjustment algorithms. The shape is completely static in this frame.

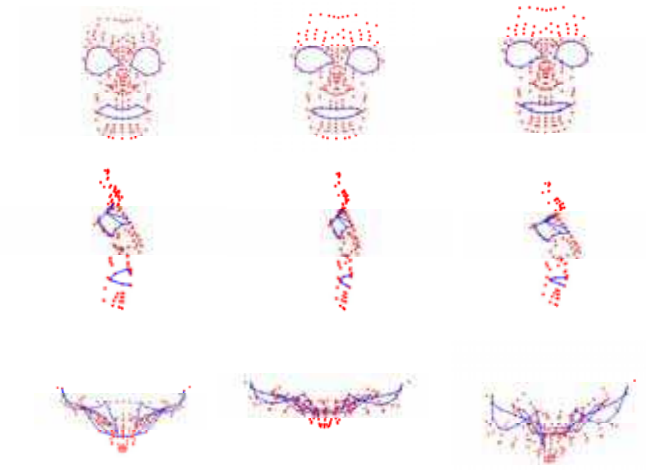


Figure 5. Front, side and top views of the 3-D synthetic face for frame 125. The first column shows the shape ground truth while the following two columns present the 3-D reconstructions for the linear and bundle adjustment algorithms. Deformations are localized in the mouth and cheek regions.

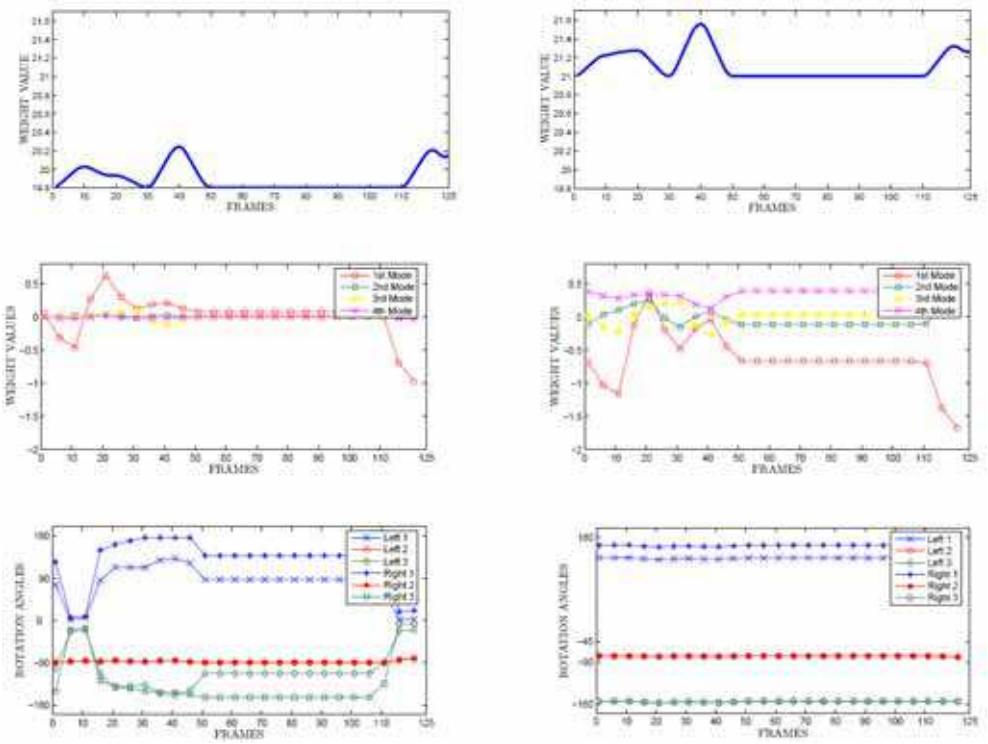
Once the model was generated we projected synthetically 160 points evenly distributed on the face, onto a pair of stereo cameras. The geometry of the cameras was such that both optical axes were lying on the XZ plane and each pointing inwards by 15 degrees. Therefore the relative orientation of the cameras 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

We show in the following figures the comparisons between three key frames of the synthetic sequence providing the 3-D ground truth and the 3-D reconstructions for the linear and bundle adjustment algorithms. Fig. 3 presents a deformation localised in the mouth region at frame 20. A first visual inspection shows that the result obtained by the bundle adjustment has a qualitative advantage over the stereo linear algorithm. While the general mean shape is close to the ground truth, only the optimised solution with bundle adjustment can model properly the deformations. Frame 70 (see Fig. 4) shows the synthetic face (ground truth) with no deformations appearing. The static pose of the shape permits to compare the 3-D depth reconstructed by the algorithms. Compared to the ground truth, the shape obtained by the stereo algorithm shows a good frontal reconstruction but the estimation of the relief is not satisfactory (see side and top views). The non-linear solution obtains a depth estimate qualitatively closer to the ground truth. Finally Fig. 5 presents the reconstruction obtained for frame 125 where the synthetic face shows consistent deformations in the cheeks and mouth area. The stereo algorithm obtains a reasonable mean 3-D shape but it fails in capturing the deformations appearing in the ground truth.

Fig. 6 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 degrees. 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 problems around frames 10 and 115, when the deformations are occurring. After the bundle adjustment step the relative rotation about the Y axis is estimated with a final result of 27 degrees resulting in a 3 degrees error given the ground truth. The relative orientations about the X and Z axes are correctly estimated to 0 degrees -- 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.




(A) STEREO ALGORITHM (B) BUNDLE ADJUSTMENT

Figure 6. 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

### 5.3 Experiments with real data: comparison with the monocular solution

In this section we compare the performance of our stereo factorization algorithm -- before the non-linear optimization -- with Brand's single camera non-rigid factorization method. 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 30 degrees. 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. Fig. 7 shows 3 frames chosen from the sequences taken with the left and right cameras.

a) SMILE sequence: left view

b) EYEBROW sequence: left view

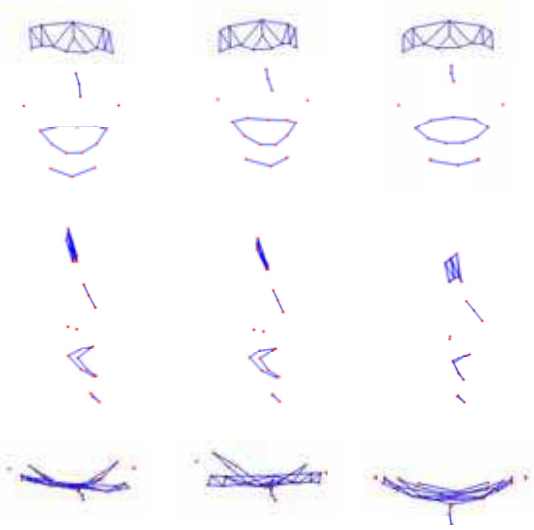
c) SMILE sequence: right view

d) EYEBROW sequence: right view

Figure 7. 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

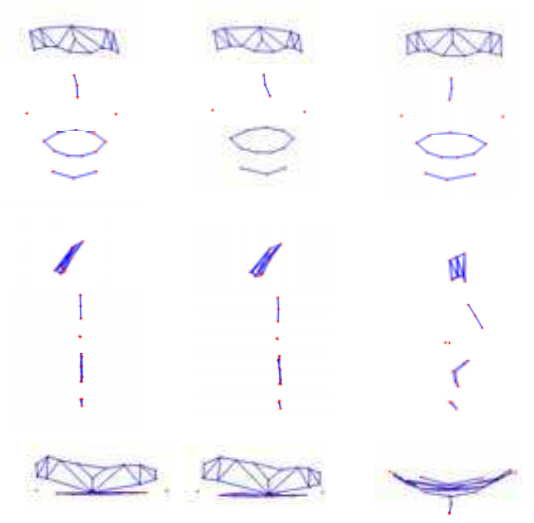
In order to simplify the temporal and stereo matching the subject had some markers placed on relevant points of the face such as along the eyebrows, the chin and the lips. A simple colour model of the markers using HSV components provided the representation used to track each marker throughout the left 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.

Fig. 8 shows front, side and top views of the 3-D reconstructions obtained for the SMILE sequence. First we applied the single camera factorization algorithm developed by Brand 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 heuristically fixed to $D$=5.



a) Left camera b) Right camera c) Stereo

Figure 8. SMILE sequence: Front, side and top views (above, middle, bottom) of the 3-D model for the a) left camera, b) right camera and c) stereo setup for $D$=5



a) Left camera b) Right camera c) Stereo

Figure 9. EYEBROW sequence: Front, side and top views (above, middle, bottom) of the 3-D model for the a) left camera, b) right camera and c) stereo setup sequences for $D$=5

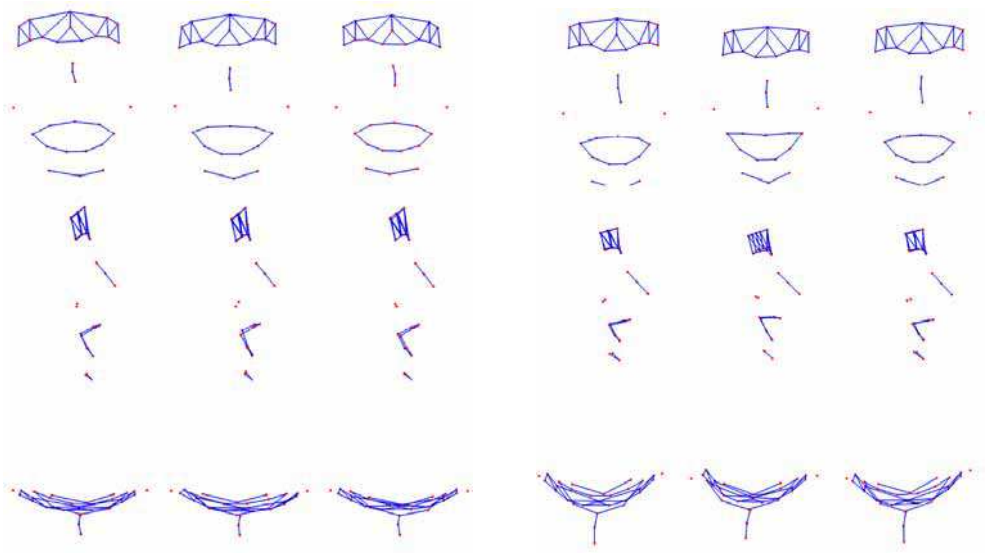



Figure 10. Front, side and top views of the reconstructed face for the SMILE sequence using the stereo algorithm (left) and after bundle adjustment (right). Reconstructions are shown for frames 16, 56 and 81 of the sequence

Fig. 8c shows how the stereo reconstruction provides improved results. The reconstructions obtained using singularly the information from the left and right sequences have worse depth estimates that can be noticed especially in the side and top views. The reconstructed face is strongly asymmetric especially in the mouth region and the points on the forehead are almost belonging to a plane. Differently, after merging the data from both sequences in the stereo algorithm, we obtained a symmetric shape and a satisfactory curvature of the forehead.

Fig. 10(A) shows the front, side and top views of the 3-D reconstructions obtained for frames 16, 58 and 81 of the SMILE sequence. While the 3-D 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 *flexible factorization* step proposed by Brand. 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 paragraphs that the bundle adjustment step resolves the ambiguity between motion and shape parameters and succeeds in modelling the non-rigid deformations.

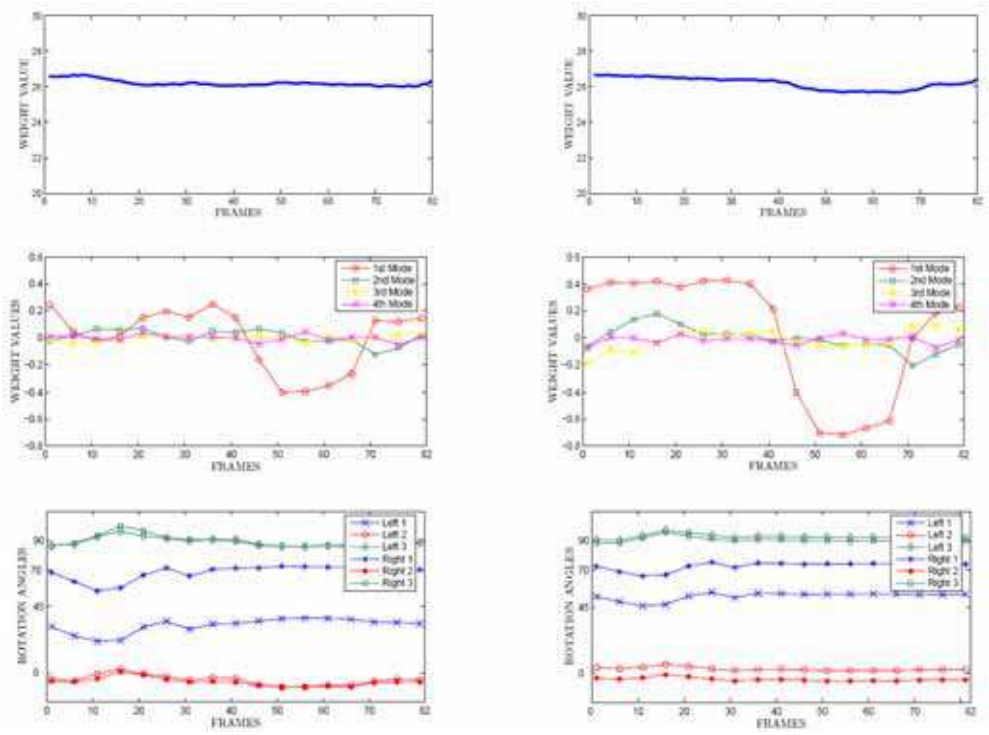
Fig. 9 shows the 3-D reconstructions obtained for the EYEBROW sequence. Once more, the single camera factorization algorithm was applied to the left and right sequences and the stereo algorithm was then applied to the stereo sequence. In this sequence the 3-D model obtained using stereo factorization 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. The points belonging to the nose, mouth and chin are almost planar (see side view) while the ones on the forehead have a particularly wrong depth estimate (see top view). Note that there was less rigid motion in this sequence and therefore the single camera factorization algorithm is not capable of recovering correct 3-D information whereas the stereo algorithm provides a good deformable model.

### 5.4 Experiments with real data: results after non-linear optimization

In this section we show the results obtained after the final non-linear optimization step. Fig. 10(B) shows the front, side and top views of the 3-D 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 3-D 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.

Fig. 11 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 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 15 degrees. 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 $R^R = R_{rel}\, R^L$ where $R_{rel}$ is the estimated relative orientation. Fig. 11(B) 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 15 degrees after bundle adjustment than before.

Fig. 11 also shows the evolution throughout the sequence of the values of the configuration weights associated with the mean 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.




(A) STEREO ALGORITHM (B) BUNDLE ADJUSTMENT

Figure 11. 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

## 6. Summary

A stereo-motion approach has been presented with the aim to reconstruct the 3-D shape of a deformable object using image sequences extracted from a stereo-pair. As a result, the non-rigid factorization framework has been accordingly updated to accommodate the constraint that trajectories in the left and right camera refer to the same 3-D object.

By construction, the method fuses naturally the advantages of motion and stereo approaches. A global solution for the time varying motion and 3-D structure is obtained from the image tracks without any prior calibration of the stereo pairs. Widely separated stereo views allow a more reliable estimation of motion and deformation parameters even in the absence of rigid motion of the object.

Additionally, non-linear optimization, as presented in the previous chapter, is performed to obtain the correct replicated structure in *M*. Results show a relevant improvement in the motion and structure estimates and thus the optimization stage is strongly recommended to obtain a correct solution.

The main assumption of our method is that the cameras must be synchronized and stereo matches be available. Synchronization can be enforced using the method presented in (Tresadern & Reid, 2003) but nowadays it is common to obtain synchronized video from

stereo cameras. Stereo matching could be tackled by extending current techniques (Ho & Chung, 2000; Oliveira et al., 2005} to deal with the non-rigid case.
Finally, notice that the solution for a stereo pair is trivially extendable to the case of multiple cameras both for the linear and non-linear approach. Moreover, the constraint over the fix baseline can be loosened to permit freely moving cameras; in this case the stereo-motion model needs to include parameters for the weak perspective scaling for each camera. This will allow to solve for a general multi camera system modelling non-rigid shapes from uncalibrated data.

## 7. Acknowledgements

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 is supported by Fundação para a Ciência e a Tecnologia (ISR/IST plurianual funding) through the POS_Conhecimento Program that includes FEDER funds.

## 8. References

Aanæs, H. & Kahl, F. (2002). Estimation of deformable structure and motion. *Workshop on Vision and Modelling of Dynamic Scenes*, Copenhagen, Denmark.

Atkinson, K. B. (1996). *Close Range Photogrammetry and Machine Vision. Engineering and Science*. Whittles Publishing.

Bar-Itzhack, I. Y. (2000). New method for extracting the quaternion from a rotation matrix. *Journal of Guidance, Control and Dynamics*, Vol. 23, No. 3, pp. 1085–1087.

Bartoli, A. & Sturm, P. (2004). Nonlinear estimation of the fundamental matrix with minimal parameters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 4, pp. 426– 432, 2004.

Berthilsson, R.; Astrom, K. & Heyden A. (2001). Reconstruction of general curves, using factorization and bundle adjustment. *International Journal of Computer Vision,* Vol . 41, No. 3, pp. 171–182.

Brand, M. (2001) . Morphable models from video. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp 456–463, Vol. 2, Kauai, Hawaii , December 2001.

Brand, M. (2005). A direct method for 3d factorization of nonrigid motion observed in 2d. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, California, pp. 122–128, 2005.

Bregler, C.; Hertzmann, A. & Biermann, H. (2000). Recovering non-rigid 3d shape from image streams. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Hilton Head, South Carolina, pages 690–696, June 2000.

Del Bue, A.; Smeraldi, F. & Agapito, L. (2007). Non-rigid structure from motion using ranklet–based tracking and non-linear optimization. *Image and Vision Computing*, Vol. 25, No. 3, pp. 297-310, March 2007.

Dornaika, F. & Chung, R. (1999). Stereo correspondence from motion correspondence. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pp. 70–75, Fort Collins, Colorado, June 1999.

Fua, P. (2000). Regularized bundle-adjustment to model heads from image sequences without calibration data. *International Journal of Computer Vision*, Vol. 38, No. 2, pp. 153–171, 2000.

Guilbert, N.; Kahl, F.; Astrom, K.; Oskarsson, M.; Johansson, M. & A. Heyden (2004). Constraint enforcement in structure and motion applied to closing an open sequence. *Asian Conference of Computer Vision*, Vol. 1, Jeju, South Korea.

Hartley, R. I. & Zisserman, A. (2000). *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Ho, P. & Chung, R. (2000). Stereo-motion with stereo and motion in complement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 2, pp. 215–220, February 2000.

Horn, B. K. P. (1987). Closed-form solution of absolute orientation using unit quaternions. *Jounal of the Optical Society America*, pp. 629–642, Vol. 4, No. 4, 1987.

Levenberg, K. (1944). A method for the solution of certain problems in least squares. *The Quarterly of Applied Mathematics*, Vol. 2, pp. 164–168, 1944.

Mandelbaum, R. ; Salgian G. & H. Sawhney (1999). Correlation-based estimation of ego-motion and structure from motion and stereo. *Proceedings 7th International Conference on Computer Vision*, pp. 544–550. Kerkyra, Greece, 1999.

Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, Vol. 11, No. 2, June 1963, pp. 431-441.

McLauchlan, P. F. & Jaenicke, A. (2002). Image mosaicing using sequential bundle adjustment. *Image and Vision Computing*, Vol. 20, No. 9, pp. 751–759, 2002.

Moré, J. (1977). The Levenberg–Marquardt algorithm: Implementation and theory. *Numerical Analysis, Lecture Notes in Mathematics*, Vol. 630, pp. 105–116, 1977.

Pollefeys, M. (1999). *Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences*. PhD thesis, Katholieke Universiteit Leuven, Belgium.

Stein, G. & Shashua (1998). Direct estimation of motion and extended scene structure from a moving stereo rig. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pp. 211–218, Santa Barbara, California, 1998.

Tomasi, C. & Kanade, T. (1992). Shape and motion from image streams under orthography: A factorization approach. *International Journal of Computer Vision*, Vol. 9, No. 2, pp. 137–154, 1992.

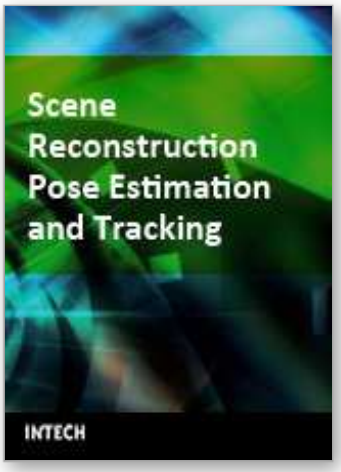
Torresani, L.; Yang, D.; Alexander, E. & Bregler, C. (2001). Tracking and modeling non-rigid objects with rank constraints. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition,* Kauai, Hawaii, 2001.

Triggs, B.; McLauchlan, P.; Hartley, R. I. & Fitzgibbon, A. (2000). Bundle adjustment – A modern synthesis. In: *Vision Algorithms: Theory and Practice*, pp. 298–375. Springer Verlag, 2000.

Trucco, E. & Verri, A. (1998). *Introductory Techniques for 3-D Computer Vision*. Prentice-Hall.

Waxman, A. & Duncan, J. (1986). Binocular image flows: steps toward stereo-motion fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, No. 6, 1986, pp. 715–729.

Xiao, J.; Chai, J. & Kanade, T. (2004). A closed-form solution to non-rigid shape and motion recovery. *Proc. 8th European Conference on Computer Vision*, Copenhagen, Denmark, pp. 573–587, May 2004.




**Scene Reconstruction Pose Estimation and Tracking**

Edited by Rustam Stolkin

ISBN 978-3-902613-06-6

Hard cover, 530 pages

**Publisher** I-Tech Education and Publishing

**Published online** 01, June, 2007

**Published in print edition** June, 2007

This book reports recent advances in the use of pattern recognition techniques for computer and robot vision. The sciences of pattern recognition and computational vision have been inextricably intertwined since their early days, some four decades ago with the emergence of fast digital computing. All computer vision techniques could be regarded as a form of pattern recognition, in the broadest sense of the term. Conversely, if one looks through the contents of a typical international pattern recognition conference proceedings, it appears that the large majority (perhaps 70-80%) of all pattern recognition papers are concerned with the analysis of images. In particular, these sciences overlap in areas of low level vision such as segmentation, edge detection and other kinds of feature extraction and region identification, which are the focus of this book.



**InTech Europe**

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

**InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821