Reconstructing a 3D deformable surface from a set of 2D correspondences across an image sequence is the problem known as Non-Rigid Structure from Motion (NRSfM). In this scenario, it is usually assumed that the 3D deformable shape can be represented by a combination of rigid basis shapes with time varying coefficients. Due to its simplicity, linear basis shape models have provided good reconstructions under certain assumptions. However, NRSfM is inherently underconstrained and it remains an open problem. In particular these models successfully reconstruct objects which exhibit small deformations, such as faces, but have difficulties in modelling objects with strong ones.

Inspired by previous work in the fields of computer graphics [1, 2], we propose a quadratic model for non-rigid deformations based on geometric constraints. The shape coordinates are augmented with quadratic and cross terms allowing the model to represent non-linear local deformations. We show that our new model can deal with challenging sequences and difficult motions with strong ones.

In the case of a rigid object, the 3D coordinates of a world point are defined as

\[ \mathbf{w}_{ij} = \begin{bmatrix} u_{ij} \\ v_{ij} \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_3 \\ s_4 & s_5 & s_6 \end{bmatrix} \begin{bmatrix} X_j \\ Y_j \\ Z_j \end{bmatrix} + \mathbf{T}_i, \]

(1)

where \( \mathbf{w}_{ij} = (u_{ij} \ v_{ij})^T \) are the non-homogeneous coordinates of point \( S_j \) in frame \( i \); \( R_i \) is a \( 2 \times 3 \) orthogonal camera matrix that contains the first two rows of a rotation matrix and \( T_i \) is the \( 2 \times 1 \) translation vector. Our quadratic deformation model for non-rigid bodies is based on geometric constraints, where the 3D coordinates of the deforming body at each frame \( i \) is defined as

\[ S = \begin{bmatrix} X_1 & X_2 & \ldots & X_p \\ Y_1 & Y_2 & \ldots & Y_p \\ Z_1 & Z_2 & \ldots & Z_p \end{bmatrix} = \begin{bmatrix} \mathbf{S}^{(1)} \\ \mathbf{S}^{(2)} \end{bmatrix}, \]

(2)

where \( \mathbf{S}^{(i)} \) is the \( 3 \times p \) linear shape matrix which contains the 3D coordinates of the shape at rest and \( \mathbf{S}^{(1)} \) and \( \mathbf{S}^{(2)} \) are simply the \( 3 \times p \) matrices that contain the quadratic and cross terms respectively. To each of these three components of the new shape matrix is associated a deformation matrix that contains the 3D coordinates of the deforming body at each frame \( i \) are defined as

\[ S_i = A_i S = \begin{bmatrix} \Gamma_i & \Omega_i & \Lambda_i \end{bmatrix} S \]

(3)

where \( \Gamma_i, \Omega_i \) and \( \Lambda_i \) are the \( 3 \times 3 \) transformation matrices associated respectively with the linear, quadratic and cross-term deformations at frame \( i \). Notice that \( S \) is fixed for all the frames while \( \Lambda_i \) varies frame-wise. By analogy with equation 1, the 2D image coordinates of the \( i \)-th world point in the \( p \)-th image will be given by \( w_{ij} = R_i A_i S_j + T_i \). Registering the 2D coordinates to the centroid we can form the bi-linear model by stacking all the sub-block matrices for each frame obtaining:

\[ \hat{w} = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_f \end{bmatrix} \begin{bmatrix} \Gamma_1 & \Omega_1 & \Lambda_1 \\ \Gamma_2 & \Omega_2 & \Lambda_2 \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \mathbf{g}^{(1)} \\ \mathbf{g}^{(2)} \\ \vdots \\ \mathbf{g}^{(A)} \end{bmatrix} = \mathbf{K} S, \]

(4)

where \( \mathbf{K} \) is the \( 2f \times p \) measurement matrix.

Due to the non-linear nature of the model we have used a non-linear optimization to minimize image reprojection error, generically known as bundle adjustment [4].

Defining the 2D coordinates of point \( j \) at frame \( i \) given by our quadratic model as \( \hat{w}_{ij} \), the parameters of the model are then estimated by minimizing the following cost function:

\[ \arg \min_{\Gamma, \Omega, \Lambda, r_i, f_i, \Omega, \Lambda, s_i} \sum_{i,j} ||w_{ij} - \hat{w}_{ij}||^2 = \arg \min_{\Gamma, \Omega, \Lambda, r_i, f_i, \Omega, \Lambda, s_i} \sum_{i,j} ||w_{ij} - R_i \Gamma_i \Omega_i \Lambda_i S_j - T_i||^2, \]

(5)

Figure 1: (a) 3D reconstruction of a cylindrical object performing a bending motion. This sequence was obtained with a VICON Motion Capture system and artificially projected to 2D. Ground truth data is represented as green circles, while the reconstructions are presented as black dots; (b) 3D reconstruction of a cushion performing a stretching motion.


