Euclidean Reconstruction of Deformable Structure Using a Perspective Camera with Varying Intrinsic Parameters

Xavier Lladó   Alessio Del Bue   Lourdes Agapito
Department of Computer Science. Queen Mary, University of London
Mile End Road, London, E1 4NS, U.K
{llado,alessio,lourdes}@dcs.qmul.ac.uk

Abstract

In this paper we present a novel approach for the 3D Euclidean reconstruction of deformable objects observed by a perspective camera with variable intrinsic parameters. We formulate the non-rigid shape and motion estimation problem as a non-linear optimization where the objective function to be minimised is the image reprojection error. Our approach is based on the observation that often some of the points on the observed object behave rigidly, while others deform from frame to frame. We propose to use the set of rigid points to obtain an initial estimate of the camera's varying internal parameters and the overall rigid motion. The prior information that some of the points in the object are rigid can also be added to the non-linear minimization scheme in order to avoid ambiguous configurations. Results on synthetic and real data prove the performance of our algorithm even when using a minimal set of rigid points and when varying the intrinsic camera paremeters.

1. Introduction

Tomasi and Kanade's factorization algorithm [7] has been one of the most successful approaches to solve the *structure from motion* problem for rigid scenes. During the last years different works have presented extensions of this factorization framework to deal with non-rigid objects [1, 3, 8, 10]. These methods assume weak perspective viewing conditions and are based on the fact that any configuration of the shape can be explained as a linear combination of basis shapes that define the principal modes of deformation of the object.

Several factorization algorithms have also been proposed for rigid scenes viewed under full perspective conditions. These are based on the observation that if the projective depths were known the measurement matrix could be rescaled and decomposed into motion and shape matrices using factorization. Xiao and Kanade [9] have recently proposed a factorization method for reconstruction of 3D non-rigid shapes under the full perspective camera model. Their method uses the subspace constraint to estimate the projective depths and then upgrades the projective reconstruction to a metric one using an extension of their affine closed form solution to the perspective camera case. However, their method still relies on the assumption that there be a set of frames in which the basis shapes are known to be independent. Besides, as pointed out by Brand [2], their affine closed form solution breaks down with noisy data.

In this paper we propose an approach to obtain a metric reconstruction of a deformable object observed by a perspective camera with varying intrinsic parameters. Our main assumption is that the observed object is composed of a mixture of some points which undergo purely rigid motion and others which exhibit deformations. This situation occurs frequently: for instance when observing a human face the points on the nose or temples typically do not deform and only undergo overall rotations and translations. We then formulate the non-rigid shape estimation problem as a non-linear optimization where the objective function to be minimised is the image reprojection error. We propose to use the set of rigid points to obtain an initial estimate of the camera's varying internal parameters and the overall rigid motion — using a standard self-calibration algorithm — which in turn provides the transformation to upgrade the non-rigid structure to metric space. Besides, the knowledge that some of the points on the object are rigid provides a strong prior on the shape parameters which can guide the non-linear minimization and allow to solve for ambiguities.

The experimental evaluation of our algorithm is performed on both synthetic and real sequences. We concentrate on analysing the effects of varying the intrinsic parameters of the camera throughout the sequence on the recovery of 3D structure and motion.

2. Non-rigid factorization

Assuming a perspective projection camera model a 3D point $\mathbf{X}_j$ is projected onto image frame i according to $\mathbf{x}_{ij} =$

$\frac{1}{\lambda_{ij}}$ $\mathtt{P}_i\mathbf{X}_j$, where $\mathbf{x}_{ij}$ and $\mathbf{X}_{ij}$ are both expressed in homogeneous coordinates, $\mathtt{P}_i$ is the projection matrix and λ_{ij} is the projective depth for that point. The projection matrix may be parameterized as $\mathtt{P}_i = \mathtt{K}_i[\mathtt{R}_i|\mathbf{T}_i]$ where $\mathtt{K}_i$ is the calibration matrix, $\mathtt{R}_i$ the rotation matrix and $\mathbf{T}_i$ the translation vector. The collection of the projection of all the scene points $\mathbf{x}_{ij}$ in all image frames is known as the measurement matrix $\mathtt{W}$. For rigid objects when the measurement matrix is rescaled with the correct projective depths its rank is constrained to be at most 4. This rank constraint can then be exploited to factorize $\mathtt{W}$ into its motion and shape components.

When the object is deforming the non-rigid 3D shape can be approximated by a linear combination of a set of D basis shapes $\mathtt{B}_d$ which represent the principal modes of deformation of the object. The 3D vectors are expressed in homogeneous coordinates so the shape at frame i is

$$\mathtt{X}_i = \begin{bmatrix} \sum_{d=1}^{D} l_{id}\mathtt{B}_d \\ \mathbf{1} \end{bmatrix} \qquad \mathtt{X}_i \in \Re^{4\times N} \quad \mathtt{B}_d \in \Re^{3\times N} \qquad (1)$$

where $\mathtt{B}_d$ are the $3\times N$ shape bases (N is the number of points), l_{id} are the corresponding deformation coefficients and $\mathbf{1}$ is an N-vector of ones. The projection of the shape at any frame i onto the image is then governed by:

$$\mathtt{W}_i = \mathtt{P}_i\mathtt{X}_i = \mathtt{P}_i \begin{bmatrix} \sum_{d=1}^{D} l_{id}\mathtt{B}_d \\ \mathbf{1} \end{bmatrix} \qquad (2)$$

Note that if the projective depths are known the rank of the rescaled measurement matrix is now at most $3D+1$.

3. Our approach

The idea of our approach for solving the metric non-rigid shape and motion estimation is to minimize the geometric distance between the measured image points and the estimated reprojected points $\sum_{i,j} \parallel \mathbf{x}_{ij} - \hat{\mathbf{x}}_{ij} \parallel^2 = \sum_{i,j} \parallel \mathbf{x}_{ij} - \mathtt{P}_i\mathbf{X}_{ij} \parallel^2$. Therefore, the cost function being minimized is:

$$\arg \min_{\mathtt{K}_i\mathtt{R}_i\mathbf{T}_i\mathbf{B}_d l_{i,d}} \sum_{i,j} \parallel \mathbf{x}_{ij} - \mathtt{K}_i[\mathtt{R}_i|\mathbf{T}_i] \begin{bmatrix} \sum_{d=1}^{D} l_{id}\mathbf{B}_{dj} \\ 1 \end{bmatrix} \parallel^2 \qquad (3)$$

3.1. Initialization

Non-linear minimization schemes require the initial estimate to be close to the global minimum. To obtain an initial fit, we propose a novel method that, provided an initial segmentation of the scene into rigid and non-rigid points, firstly estimates the rigid metric structure and then models the remaining contributions of the deformations. In this paper we do not cover the segmentation step and instead we assume it has been achieved in a preliminary stage using a similar approach to the work proposed by Del Bue et al. [4, 5].

To upgrade the rigid structure to metric, we apply the well-known self-calibration method proposed by Pollefeys et al. [6]. This provides initial estimates for the 3D metric structure and for the intrinsic parameters, the overall rigid motion and the mean shape. An initial estimate for the first basis shape of the non-rigid points can be directly computed given the original measurement matrix containing the trajectories of the non-rigid points and the projection matrices $\mathtt{P}_i$ obtained by the self-calibration algorithm. The rest of the basis shapes which encode the $D-1$ non-rigid components are initialized to small random values. Finally, the deformation weights associated with the mean shape are initialised to 1 while the rest are initialised to small values. Similar initialisation has previously been used in [4, 5, 8].

The prior information that some of the points in the object are rigid can also be added to the non-linear minimization scheme in order to avoid ambiguous configurations. Our prior expectation is that the rigid points have a zero non-rigid component and can therefore be modelled entirely by the first basis shape. We write these constraints as priors on the coordinates of the basis vectors $\mathbf{B}_{dj}$ and solve the problem as a Maximum A Posteriori (MAP) estimation.

3.1.1. Varying intrinsic camera parameters

The self-calibration algorithm allows to impose different constraints on each of the camera intrinsic parameters (focal length, principal point and aspect ratio). For instance, in the minimization, each of the parameters may be considered to be known, unknown but constant between views or unknown and varying. In the experimental section we will analyse the effect of these different assumptions on the Euclidean reconstructions.

4. Experimental results

4.1. Synthetic data

The 3D data consisted of a set of random points sampled inside a box of size $100 \times 100 \times 100$ units. Different sequences were generated using different ratios of rigid/non-rigid points. In particular, we used a fixed set of 10 rigid points while using 10 and 50 non-rigid points. The deformations for the non-rigid points were generated using random basis shapes and random deformation weights. The first basis shape had the largest weight equal to 1. We also created different sequences varying the number of basis shapes ($D = 3$ and $D = 5$) for both ratios of rigid/non-rigid points. Finally, in order to evaluate different levels of perspective distortion we used 2 different camera setups in which we varied the distance of the object to the camera and the focal length (Setup 1: z=250, f=900; Setup 2: z=200, f=600). The 3D data was then projected onto 50 im-

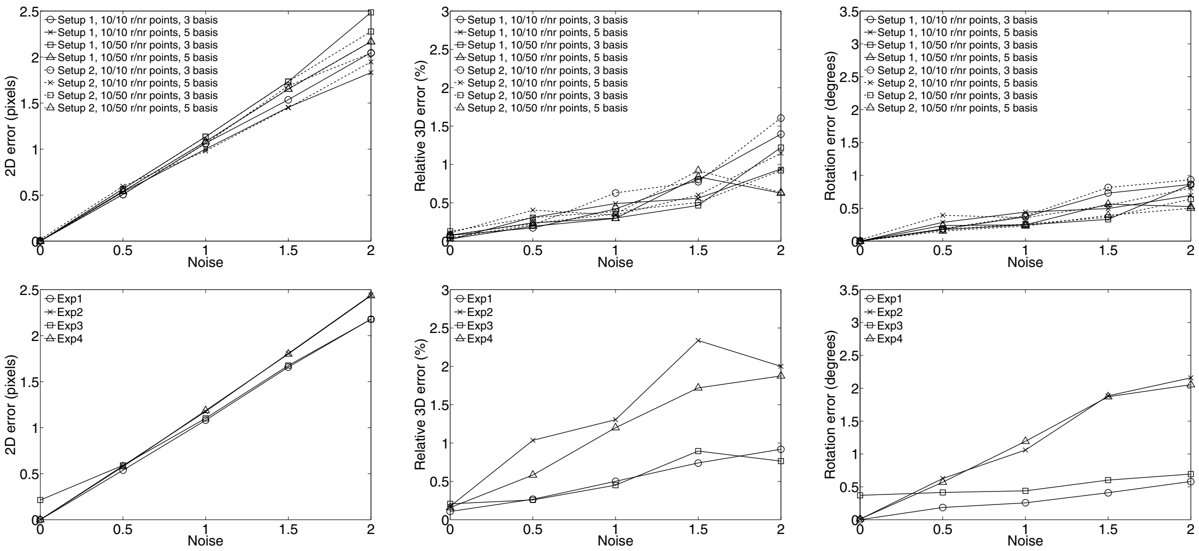

Figure 1. 2D error, 3D error and rotation error curves. First row: results obtained when the focal length was constant. Second row: results obtained for the 4 experiments with varying intrinsics.

ages applying rotations and translations over all the axes. Gaussian noise of increasing levels of variance was added to the image coordinates.

4.1.1. Constant intrinsics

For the first set of experiments we assumed that all the camera parameters remained constant over the sequence. We then applied our 3D reconstruction algorithm to all the experimental setups described before. The results are summarized on the first row of Figure 1 where we show the r.m.s. 2D image reprojection error (pixels), 3D metric reconstruction error (percentage relative to the scene size) and the absolute rotation error (degrees). The plots show the mean values of 5 random trials per level of noise.

Our proposed algorithm appears to perform well in the presence of noise. The 3D reconstruction error is low even for large perspective distortions and for a large proportion of non-rigid versus rigid points. The 2D error is also small and it appears to be of the same order as the image noise. Figure 1 also illustrates that the rotations are correctly estimated. Reliable estimates for the internal camera parameters (focal length and principal point) were obtained even in the presence of noise, although we do not show them here.

4.1.2. Varying intrinsics

We then performed a set of experiments in which some of the internal parameters of the camera were varying throughout the sequence. We designed 4 different experiments using camera setup 2 (z=200,f=600), a ratio of 10 rigid to 50 non-rigid points and 5 basis shapes. For Experiment 1 the focal length of the camera varied linearly throughout the sequence while the rest of the internal parameters remained constant. In the optimization algorithm we considered the focal length unknown and allowed it to vary while the principal point and the aspect ratio were considered to be unknown but fixed throughout the sequence and the skew was considered known and equal to zero. Experiment 2 had the same setup with the focal length being the only parameter that varied but during the optimization process we allowed both the focal length and the principal point to vary. In Experiment 3 the focal length and the principal point both varied throughout the sequence. In the minimization we considered the focal length unknown and allowed it to vary but the principal point was assumed to be fixed but unknown. Finally, in Experiment 4 we used the same setup as in Experiment 3 but allowed both the focal length and the principal point to vary in the minimization.

The results for all 4 experiments are illustrated on the second row of Figure 1. The results obtained for the internal camera parameters are summarised on table 1. Note that for the noisy cases in which the real principal point was varying better estimates were obtained assuming the principal point constant during the minimization.

4.2. Experiments with real data

The real scene consisted of a set of 12 rigid points (9 on two boxes and 3 over a chair) and a set of 20 deformable points situated on a pillow which was deforming during the sequence (see first row of Figure 2). The 3D data was captured using a VICON motion capture system. The 3D points were then projected synthetically onto an image sequence

	Noise	0	0.5	1	1.5	2
Exp1	mean f	0	0.56	1.68	1.69	3.90
	std. dev f	0	0.18	1.26	0.94	1.99
	max. err f	0	0.83	3.49	3.22	7.16
	mean p_u	0	0.59	1.48	1.29	6.03
	mean p_v	0	0.91	2.43	2.50	3.46
Exp2	mean f	0.01	2.93	5.14	10.28	10.97
	std. dev f	0.01	0.79	2.92	6.96	4.33
	max. err f	0.02	3.91	8.36	20.12	14.92
	mean p_u	0.09	11.17	18.01	26.68	27.50
	mean p_v	0.08	6.66	14.80	22.93	28.91
Exp3	mean f	0.69	1.04	1.16	3.10	2.92
	std. dev f	0.27	0.50	0.38	2.58	1.15
	max. err f	1.04	1.75	1.81	5.96	4.47
	mean p_u	2.97	2.96	3.01	3.77	3.97
	mean p_v	3.49	3.34	3.47	5.88	3.79
Exp4	mean f	0.05	2.11	4.93	10.40	10.38
	std. dev f	0.04	1.05	3.51	2.92	4.66
	max. err f	0.09	3.60	8.80	14.27	14.17
	mean p_u	0.10	5.95	12.71	16.01	16.31
	mean p_v	0.07	3.49	10.61	14.34	15.54

Table 1. Mean, standard deviation and maximum value of the relative error in the focal length estimate. Absolute mean error in principal point estimate.

75 frames long using a perspective camera model. Gaussian noise of 0.5 pixels was added to the image coordinates. The size of the scene was $61 \times 82 \times 53$ units, it was at a distance of 150 units from the camera and the focal length was 900 pixels and constant during the sequence. Figure 2 shows the ground truth (squares) and reconstructed shape (crosses) from two different viewpoints. The number of basis shapes was $D = 6$, the 2D reprojection error was 0.95 pixels, the absolute 3D error was 1.34 units, the absolute rotation error was 2.11 degrees and the focal length was estimated to be 899. Note that the deformations are well captured even for the frames in which they were most exaggerated. We repeated the same experiment but varying the focal length from 700 to 1000 pixels during the sequence. In this case, the 2D error was 0.96 pixels, the absolute 3D error was 1.65 units, the rotation error was 2.77 degrees and the mean focal length error was 34.84 pixels.

5. Conclusions

We have proposed a new approach for the estimation of Euclidean non-rigid shape from uncalibrated images. The experiments have shown that even when using a minimal set of rigid points and when varying the intrinsic parameters it is possible to obtain reliable metric information.

6. Acknowledgments

This work has been supported by EPSRC grant GR/S61539/01. ADB holds a Queen Mary Studentship.

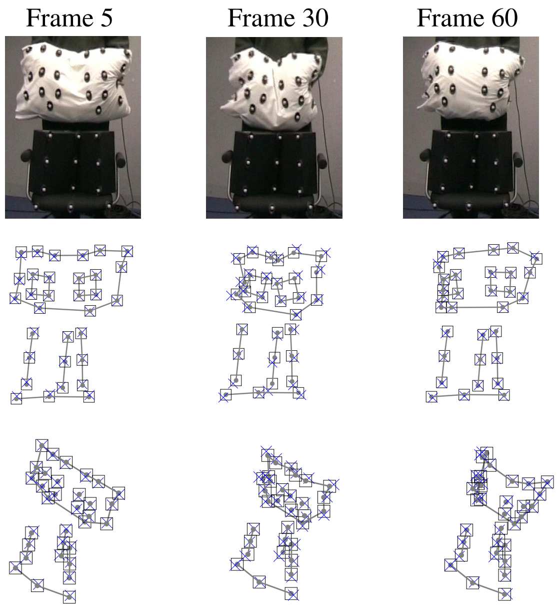

Figure 2. First row shows examples of the analysed scene. Second and third rows show two views of the reconstructed scene.

References

[1] M. Brand. Morphable models from video. In *Proc. IEEE CVPR*, pages 456–463, December 2001.
[2] M. Brand. A direct method for 3d factorization of nonrigid motion observed in 2d. In *Proc. IEEE CVPR*, pages 122–128, 2005.
[3] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3d shape from image streams. In *Proc. IEEE CVPR*, pages 690–696, June 2000.
[4] A. Del Bue, X. Lladó, and L. Agapito. Non-rigid face modelling using shape priors. In *IEEE International Workshop AMFG, held in conjunction with ICCV-05*, volume 3723 of *Lecture Notes in Computer Science*, pages 96–107. 2005.
[5] A. Del Bue, X. Lladó, and L. Agapito. Non-rigid metric shape and motion recovery from uncalibrated images using priors. In *Proc. IEEE CVPR. To appear*, New York, 2006.
[6] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. *International Journal in Computer Vision*, 32(1):7–25, 1999.
[7] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization approach. *International Journal in Computer Vision*, 9(2):137–154, 1992.
[8] L. Torresani, D. Yang, E. Alexander, and C. Bregler. Tracking and modeling non-rigid objects with rank constraints. In *Proc. IEEE CVPR*, pages 493–500, December 2001.
[9] J. Xiao and T. Kanade. Uncalibrated perspective reconstruction of deformable structures. In *Proc. 10th ICCV*, pages 1075–1082, October 2005.
[10] J. Xiao, J. Chai and T. Kanade. A closed-form solution to non-rigid shape and motion recovery. In *Proc. 8th ECCV*, pages 573–587, May 2004.