Segmentation of Rigid Motion from Non-rigid 2D Trajectories

Alessio Del Bue[1], Xavier Lladó[2], and Lourdes Agapito[3]

[1] Instituto de Sistemas e Robótica, Instituo Superior Técnico, Lisboa, Portugal
[2] Dept. of Electronics, Computer Eng. and Automatics, Universtiy of Girona, Spain
[3] Department of Computer Science, Queen Mary, University of London, UK

Abstract. In this paper we evaluate an automatic segmentation algorithm able to identify the set of rigidly moving points within a deformable object given the 2D measurements acquired by a perspective camera. The method is based on a RANSAC algorithm with guided sampling and an estimation of the fundamental matrices from pairwise frames in the sequence. Once the segmentation of rigid and non-rigid points is available, the set of rigid points could be used to estimate the internal camera calibration parameters, the overall rigid motion and the non-rigid 3D structure.

1 Introduction

In this paper we evaluate a method that performs the automatic segmentation of a set of rigidly moving points within a deformable object given a set of 2D image measurements. Several works have previously proposed the use of the dimensionality of the subspace in which the image trajectories lie to perform motion segmentation [1,2,3,4]. However, all these methods focus either on the segmentation of independently moving objects or on the segmentation of objects of different nature (rigid, articulated or non-rigid), but none of them can deal efficiently with the segmentation of rigid and non-rigid points on a single deformable object. Moreover, most of these methods assume an affine camera model but in this paper we are interested in the full perspective camera case. To our knowledge, the work of Del Bue et al. [5] is the first attempt to obtain a reliable segmentation of rigid points from non-rigid bodies, however, the method is restricted to affine camera models.

Our segmentation approach in the perspective case is based on a RANSAC [6] algorithm with guided sampling, similarly to the one proposed by Tordoff and Murray [7]. The RANSAC algorithm is used to estimate the fundamental matrices from pairwise frames in the sequence and to segment the scene into rigid and non-rigid points. We perform experiments of the segmentation algorithm on synthetic and real data which show the validity of our proposed method. The result of the segmentation algorithm could then be used to recover the 3D non-rigid structure and motion using the method described in [8,9].

The paper is organised as follows. Section 2 introduces the general idea of the segmentation algorithm, while sections 3−4 describe in detail the different steps

J. Martí et al. (Eds.): IbPRIA 2007, Part I, LNCS 4477, pp. 491–498, 2007.

of the approach. In section 5 we show segmentation results on different synthetic and real data sets.

2 Segmentation of Rigid and Non-rigid Motion Under Perspective Viewing

The approach is based on the fact that rigid points satisfy the epipolar geometry while the non-rigid points will give a high residual in the estimation of the fundamental matrix between pairs of views. We use a RANSAC algorithm [6] to estimate the fundamental matrices from pairwise frames in the sequence and to segment the scene into rigid and non-rigid points. Therefore, in this case we consider the dominant motion to be the rigid one and the non-rigid points to be the outliers.

However, a well known drawback of random sampling and consensus techniques is the computational cost required to obtain a valid set of points when the percentage of outliers is high, due to the large number of samples needed to be drawn from the data. Unfortunately, this is the most likely scenario in non-rigid structure from motion where we normally deal with a small proportion of completely rigid points. Here we exploit a measure of the degree of non-rigidity of a point to infer a prior distribution of the probability of a trajectory being rigid or non-rigid given that measure. These distributions are then used as priors to perform guided sampling over the set of trajectories in a similar approach to the one proposed by Tordoff and Murray [7] for the stereo matching problem.

2.1 Degree of Non-rigidity

In order to increase the likelihood of selecting rigidly moving points in the sampling stage, we associate a measure of non-rigidity to each trajectory. Recently, Kim and Hong [10] introduced the notion of Degree of Non-rigidity (*DoN*) of a point viewed by an orthographic camera as an effective measure of the deviation of the point from the average shape. If the average 3D shape of a time varying shape $\mathtt{X}_i = [\mathbf{X}_{i1} \dots \mathbf{X}_{iP}]$ (in non-homogeneous coordinates) is given by $\check{\mathtt{X}} = [\check{\mathbf{X}}_1 \dots \check{\mathbf{X}}_P]$ the Degree of Non-rigidity for point j is defined as:

$$DoN_j = \sum_{i=1}^{F} (\mathbf{X}_{ij} - \check{\mathbf{X}}_j)(\mathbf{X}_{ij} - \check{\mathbf{X}}_j)^T \tag{1}$$

The 2D projection $\mathtt{C}_j$ of the *DoN* will be thus given by:

$$\mathtt{C}_j = \sum_{i=1}^{F} \mathtt{R}_i (\mathbf{X}_{ij} - \check{\mathbf{X}}_j)(\mathbf{X}_{ij} - \check{\mathbf{X}}_j)^T \mathtt{R}_i^T = \sum_{i=1}^{F} (\mathbf{w}_{ij} - \check{\mathbf{w}}_{ij})(\mathbf{w}_{ij} - \check{\mathbf{w}}_{ij})^T \tag{2}$$

where $\mathtt{R}_i$ are the 2×3 orthographic camera for F frames, $\mathbf{w}_{ij}$ are the non-homogeneous image coordinates of point j in frame i and $\check{\mathbf{w}}_{ij}$ are the coordinates

of its projected mean 3D value over the F frames in the sequence. While the DoN cannot be computed without an estimation of the mean 3D shape (and this implies finding a 3D deformable reconstruction), the value of its projection can be estimated directly from image measurements.

An approximate estimate of the average projected shape $\check{\mathbf{w}}_{ij}$ can be given simply by the rank-3 approximation of the measurement matrix $\mathtt{W}$ (the matrix that contains the 2D coordinates $\mathbf{w}_{ij}$ of all P points viewed in all F frames) computed using singular value decomposition and given by $SVD_3(\mathtt{W}) = \check{\mathtt{M}}\check{\mathtt{S}}$. The projected deviation from the mean for all the points would then be defined by $\{\mathbf{w}_{ij} - \check{\mathbf{w}}_{ij}\} = \mathtt{W} - \check{\mathtt{M}}\check{\mathtt{S}}$. Kim and Hong computed a more sophisticated estimate of the average shape, but for simplicity we have used the above description which has shown to give a reasonable measure of the degree of deformability.

Notice that the previous definitions all assume affine viewing conditions. However, our trajectories reside in a projective space so a redefinition of the measure of non-rigidity is required. First, the original measurement matrix must be re-scaled by the estimated projective weights λ_{ij}. We calculate the projective weights λ_{ij} using subspace constraints [11] and express the rescaled measurement matrix (in homogeneous coordinates) as $\bar{\mathtt{W}} = \{\lambda_{ij}\ [\mathbf{w}_{ij}^T\ 1]^T\}$. Then, we estimate the mean shape as the rank-4 approximation of the rescaled measurement matrix computed using singular value decomposition and given by $SVD_4(\bar{\mathtt{W}}) = \check{\mathtt{M}}\check{\mathtt{S}}$. The projected deviation from the mean would then be defined similarly as before but computing the mean projected shape $\check{\mathbf{w}}_j$ using the normalized non-homogenous coordinates. Therefefore, the projection of the DoN can finally be computed as:

$$\mathtt{C}_j = \sum_{i=1}^{F} (\mathbf{w}_{ij} - \check{\mathbf{w}}_{ij}) (\mathbf{w}_{ij} - \check{\mathbf{w}}_{ij})^T \tag{3}$$

in the form of a 2×2 covariance matrix. Instead of using the full information of $\mathtt{C}_j$, we approximate the score s as the sum of the diagonal values of $\mathtt{C}_j$.

3 Computation of the Prior

Tordoff and Murray [7] showed that guided sampling based on knowledge extracted from the images can greatly improve the performance of a random sampling method, especially in the presence of noise or high number of outliers. In these cases standard RANSAC becomes computationally prohibitive given the large number of random samples that must be drawn from the data. Here we use the 2D projection of the DoN defined in the previous section to provide the score s for each point trajectory which in turn will be used to build a prior distribution of the conditional probability of each point in the object being rigid or non-rigid given this score.

We have inferred the conditional probability density functions for the score s given that a point is rigid $p(s|r)$ (see figure 1(a)) or non-rigid $p(s|\bar{r})$ (see

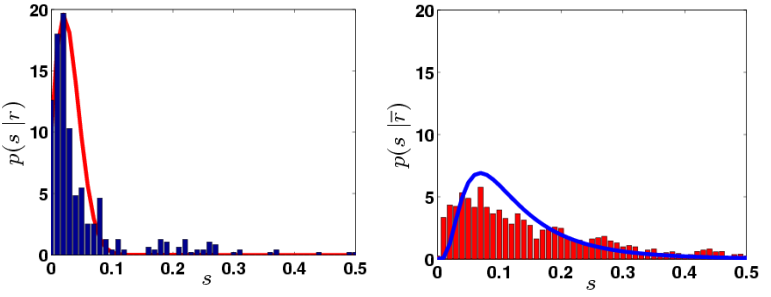


Fig. 1. Conditional densities for the score given: (a) that a point is rigid $p(s|r)$ or (b) non-rigid $p(s|\bar{r})$ approximated from the normalised frequency histograms for different synthetic and real sequences with different degrees of perspective distortion, deformation and ratio of rigid/non-rigid points

figure 1(b)) by computing the normalised frequency histograms over many experimental trials with synthetic and real sequences with different perspective distortions, degrees of deformation and ratios of rigid/non-rigid points. We have then approximated the histograms by fitting appropriate analytical functions (a Gamma distribution for $p(s|r)$ and a Lognormal for $p(s|\bar{r})$).

To derive the prior conditional density function of a point being rigid given the non-rigidity score, $p(r|s)$, we use Bayes theorem:

$$p(r|s) = \frac{p(s|r)p(r)}{p(s)} \propto \frac{p(s|r)}{p(s|r) + p(s|\bar{r})} \tag{4}$$

Figure 2 shows an example of a prior obtained from our experiments. Note that, although the computation of the score is specific to each method, the derivation of the prior given the distribution of the score is general.

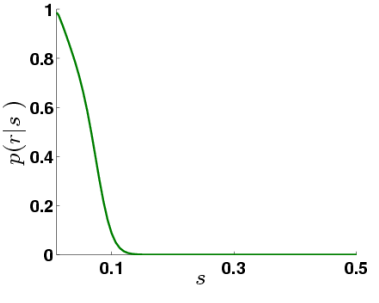


Fig. 2. Estimated prior given by the estimated densities $p(s|r)$ and $p(s|\bar{r})$

4 Guided RANSAC

We use guided RANSAC to estimate the fundamental matrices between pairs of consecutive views for all the F frames composing the sequence. This process will be used to provide a segmentation of the image trajectories into rigid and non-rigid ones since the non-rigid trajectories will not satisfy the epipolar geometry and will therefore give a high residual in the computation of the pairwise fundamental matrices. In order to speed up the process, we use the prior derived in the previous section to draw the point samples: points with the highest conditional probability of being rigid will be chosen more frequently. The RANSAC with priors procedure is outlined as follows:

Algorithm 1. RANSAC with priors

1. Compute the score s for each trajectory in $\bar{\mathtt{W}}$.
2. Sample b trajectories given the prior $p(r)$ and the score s.
3. For each sample, estimate $(F-1)$ fundamental matrices from each pair of consecutive frames.
4. Calculate the distance of the points from the $F-1$ instantiated models and find the trajectories that are within a threshold t.
5. Repeat N times and determine the largest consensus given a set of trajectories.

The method employed to estimate the fundamental matrix is the standard 8-point algorithm [12]. The distance threshold t which decides whether a point is an inlier or an outlier (rigid or non-rigid in this case) was set empirically to be $t = 4.12$. It was fixed by taking into account the sum of the residuals given by the estimation of $F-1$ fundamental matrices using normalised coordinates. Notice that we do not consider outliers in the point matching from frame to frame. We show results of the guided sampling RANSAC algorithm applied to the segmentation of rigid and non-rigid points in the experimental section.

Once the scene has been segmented into the rigid and non-rigid point sets we may compute metric non-rigid shape in two further steps as described in [8,9]. First the rigid points are used to estimate the intrinsic parameters of the camera – which provide the necessary information to upgrade the structure to metric – and the overall rotations and translations. Secondly, the estimation of metric non-rigid shape is formulated as a global non-linear minimization with shape priors over the rigid trajectories.

5 Experimental Results

This experimental section validates the rigid/non-rigid segmentation with synthetic and real experiments. The synthetic tests are designed in such a way as to verify the performance of the method in case of different ratios of rigid/non-rigid points and with two different setups of perspective distortions.

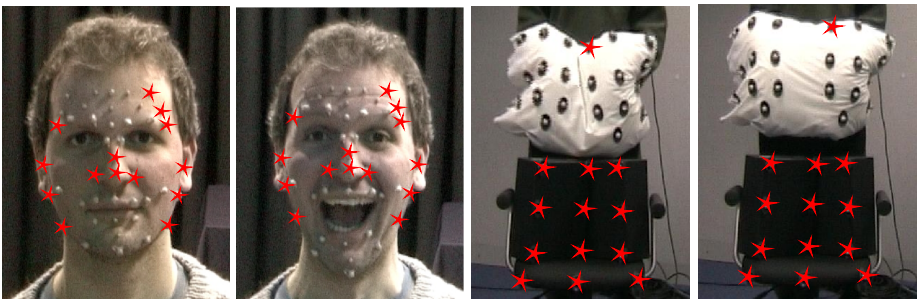
5.1 Synthetic Data

The 3D data consists of a set of random points sampled inside a cube of size $100 \times 100 \times 100$ units. Several 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 as well as 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 images applying random rotations and translations over all the axes. Gaussian noise of increasing levels of variance was added to the image coordinates.

Table 1. Mean error of number of non-rigid points being classified as rigid over 100 trials. Results for different levels of Gaussian noise with variance $\sigma = 0.5$, 1, 1.5, 2 pixels and different experimental setups.

Experiments	Noise				
	0	0.5	1	1.5	2
Exp1: $D = 5$, 10/10, setup 1	0.28	0.48	0.55	0.72	0.77
Exp2: $D = 5$, 10/50, setup 1	0.31	0.38	0.46	0.55	0.72
Exp3: $D = 3$, 10/10, setup 1	0.95	1.36	1.53	1.60	1.54
Exp4: $D = 3$, 10/50, setup 1	2.19	2.38	2.33	2.78	2.51
Exp5: $D = 5$, 10/10, setup 2	0.24	0.27	0.32	0.48	0.62
Exp6: $D = 5$, 10/50, setup 2	0.3	0.34	0.39	0.51	0.58
Exp7: $D = 3$, 10/10, setup 2	0.65	0.94	1.27	1.42	1.45
Exp8: $D = 3$, 10/50, setup 2	2.09	2.37	2.28	2.31	2.27

The RANSAC procedure was tested over 100 trials for each setup and for each level of noise. The number of samples randomly chosen over the prior distribution was fixed to 2500. At each new trial the motion components (rotation and translation) of the objects are randomly generated obtaining a 50 frames long sequence. The results in table 1 show the mean error (over the 100 trials) of the number of non-rigid points being classified as rigid for the different setups. Better performances are obtained for more complex deformations (i.e., more basis shapes, $D = 5$) and for stronger perspective effects (Setup 2) since the effects of perspective distortions and deformations are less ambiguous in such cases (Experiments 5 and 6). Experiments 4 and 8 obtain the worse results achieving a mean error of more than 2 points given the smaller deformations occuring in these data and the higher number of non-rigid points (50).



(a) (b)

Fig. 3. (a) The image frame show the position of the markers used to capture the face deformations. (b) The image shows an example of the setup used for the box and pillow experiment. The segmented rigid points are highlighted with a red star.

5.2 Real Experiments

We tested our segmentation algorithm over two sequences with real deforming objects. The 3D structure was obtained using a VICON motion capture system and was then projected synthetically onto the image plane using a perspective camera model. Gaussian noise of 0.5 pixels was finally added to the image coordinates. The first experiment captured a human face undergoing rigid motion while performing different facial expressions. The subject was wearing 37 markers on the face. Figure 3 (a) shows two frames with the positions of the markers and the selected rigid points. The segmented points mostly belong to the temple and nose areas which udergo a predominantly rigid motion in the sequence. Applying the non-rigid structure from motion algorithm proposed in [9] we obtained a 2D reprojection error (root mean squared) of 0.62 pixels, a 3D error relative to the scene size of 1.81 units (the size of the face model was $168 \times 193 \times 101$ units) and a rotation error of 0.27 degrees.

The second 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 Figure 3 (b)). The segmentation algorithm provided 13 rigid points, including one non-rigid trajectory corresponding to one of the pillow points. However, the inclusion of this non-rigid point did not affect the 3D reconstruction results. We obtained a 2D reprojection error of 0.9 pixels, a 3D relative error of 1.49 units (the size of the scene was $61 \times 82 \times 53$) and a rotation error of 2.82 degrees.

6 Conclusions

We have presented an approach to segment rigid trajectories embedded in a non-rigidly moving shape. The extracted rigid trajectories may be used to obtain prior-based 3D reconstructions as in [9] or to aid non-rigid shape registration tasks. The segmentation stage obtains reasonable results for the configuration of

basis, cameras and points tested. However we noticed a higher misclassification ratio with weak perspective effects and higher proportion of non-rigid points. Also, points that are rigid only for a part of the sequence may appear undetected since they only conform with the epipolar geometry for a subset of frames.

Acknowledgments. This work was partly funded by EPSRC grant GR/S61539/ 01. ADB is supported by Fundação para a Ciência e a Tecnologia (ISR/IST funding) through the POS_Conhecimento Program that includes FEDER funds.

References

1. Costeira, J.P., Kanade, T.: A multibody factorization method for independently moving objects. International Journal of Computer Vision 29(3), 159–179 (1998)
2. Kanatani, K.: Motion segmentation by subspace separation: Model selection and reliability evaluation. International Journal of Image and Graphics 2(2), 179–197 (2002)
3. Vidal, R., Hartley, R.I.: Motion segmentation with missing data using powerfactorization and gpca. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, San Diego, California. vol. 2, pp. 310–316 (2004)
4. Yan, J., Pollefeys, M.: A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate. In: Proc. 9th European Conference on Computer Vision, Graz, Austria. vol. 4. pp. 94–106 (2006)
5. Del Bue, A., Lladó, X., Agapito, L.: Non-rigid face modelling using shape priors. In: Zhao, W., Gong, S., Tang, X. (eds.) AMFG 2005. LNCS, vol. 3723, pp. 96–107. Springer, Heidelberg (2005)
6. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. In: Fischler, M.A., Firschein, O. (eds.) Readings in Computer Vision: Issues, Problems, Principles, and Paradigms, Los Altos, CA. pp. 726–740 (1987)
7. Tordoff, B.J., Murray, D.W.: Guided-MLESAC: Faster image transform estimation by using matching priors. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(10), 1523–1535 (2005)
8. Del Bue, A., Lladó, X., Agapito, L.: Non-rigid metric shape and motion recovery from uncalibrated images using priors. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, New York, NY (2006)
9. Lladó, X., Del Bue, A., Agapito, L.: Euclidean reconstruction of deformable structure using a perspective camera with varying intrinsic parameters. In: Proc. International Conference on Pattern Recognition, Hong Kong (2006)
10. Kim, T., Hong, K.S.: Estimating approximate average shape and motion of deforming objects with a monocular view. International Journal of Pattern Recognition and Artificial Intelligence 19(4), 585–601 (2005)
11. Xiao, J., Kanade, T.: Uncalibrated perspective reconstruction of deformable structures. In: Proc. 10th International Conference on Computer Vision, Beijing, China (2005)
12. Hartley, R.I.: In defense of the eight-point algorithm. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico vol. 19(6), pp. 580–593 (1997)