15th IAPR International Conference on Machine Vision Applications (MVA)
Nagoya University, Nagoya, Japan, May 8-12, 2017.

Improving the performance of non-rigid 3D shape recovery by points classification

Junjie Hu[1], Terumasa Aoki[1,2]
[1]Graduate School of Information Sciences (GSIS), Tohoku University,
Aramaki Aza Aoba 6-3-9, Aoba-ku, Sendai, Japan
[2] New Industry Creation Hatchery Center (NICHe), Tohoku University,
Aramaki Aza Aoba 6-6-10, Aoba-ku, Sendai, Japan
{hujunjie, aoki}@riec.tohoku.ac.jp

Abstract

The goal of Non-Rigid Structure from Motion (NRSfM) is to recover 3D shapes of a deformable object from a monocular video sequence. Procrustean Normal Distribution (PND) is one of the best algorithms for NRSfM. It uses Generalized Procrustes Analysis (GPA) model to accomplish this task. But the biggest problem of this method is that just a few non-rigid points in 2D observations can largely affect the reconstruction performance. We believe that PND can achieve better reconstruction performance by eliminating the affection of these points. In this paper, we present a novel reconstruction method to solve this problem. We present two solutions to simply classify the points into non-rigid and nearly rigid points. After that, we use EM algorithm of PND to recover 3D structure again for nearly rigid points. Experimental results show that the proposed method outperforms the existing state-of-the-art algorithms.

1. Introduction

Structure from motion (SfM) is the process to estimate 3D structures and corresponding camera motions of a rigid object from 2D point tracks. Although its theory has been well established over the past two decades [8], recovering time varying 3D shapes of a deformable object is still difficult. This task is called Non-Rigid Structure from Motion (NRSfM). The difficulty is mainly due to the inherently high number of degrees of freedom. For every time varying observed 2D point, there exactly exists one corresponding 3D point, which makes the problem to be ill posed.

Recent research works have been attempted to solve NRSfM by using additional constraints. For instance, some approaches assumed the time varying 3D structure of non-rigid object can be represented as a linear combination of several bases of shapes [1, 9, 10]. However, because the shape basis is highly object-specific, it is difficult to apply it in practical problems. Other approaches represented a 3D point trajectory by using a set of pre-defined trajectory bases [3, 4, 5, 11]. The advantage of this representation is that it is object free and the trajectory basis can be pre-defined, thus achieving a significant reduction in number of unknowns. Besides, Dai *et al.* proposed a well-known algorithm based on the nuclear minimization method that called Block Matrix Method (BMM) [2]. BMM achieved one of the most remarkable reconstruction performances. The similar approaches based on the nuclear minimization were used to reconstruct 3D structures from realistic videos [6, 12]. The advantage of these methods is that it doesn't need any priors such as pre-defined shape basis or trajectory basis.

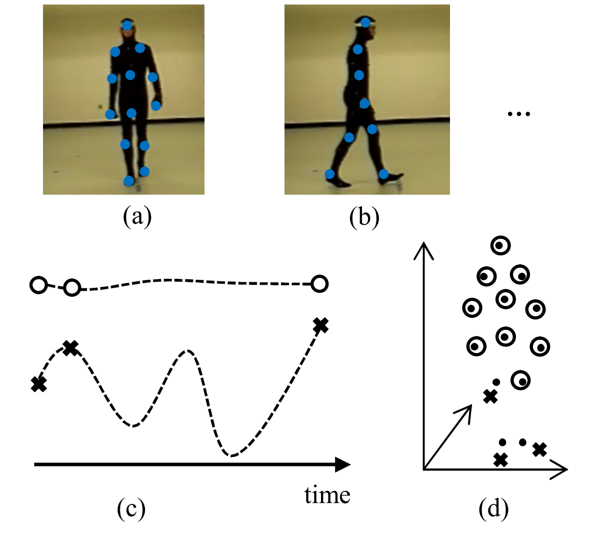


Fig. 1. Basic idea of our classification methods for nearly rigid and non-rigid points. (a) (b) ⋯ are T frames input images; (c) 2D trajectory of nearly rigid and non-rigid point; (d) Aligned 3D shape and its mean shape.

However, as long as the authors know, all of these existing methods are based on the low rank assumption. The rank needs to be set correctly and it is still difficult to know the proper number. Recently, Lee *et al.* proposed a novel algorithm called Procrustean Normal Distribution (PND) [7] that outperforms the existing methods and does not require any rank constraint. It determines rigid motions based on the Generalized Procrustes Analysis (GPA) [13], which improves the accuracy of rotation calculation for NRSfM, but there exists one big problem in PND that only a few points with strong deformation can cause poor reconstruction performance.

In this paper, we present a novel NRSfM method to solve this problem. Here, we define the points with strong deformation as "non-rigid points". On the other hand, the other points are called "nearly rigid points" in this paper. The proposed method is based on PND. The basic idea is that: we classify all points of the object into non-rigid and nearly rigid points, Fig. 1 (c) and (d) show our two approaches to separate these two types of points correctly. They compare the time varying 2D displacement (Fig.1 (c)) and 3D distance between aligned shape and PND's recovered mean shape (Fig.1 (d)). And after that, we use EM algorithm again to achieve more accurate reconstruction result for the nearly rigid points.

The remainder of this paper is organized as follows: Section 2 briefly reviews PND algorithm and introduces our approach to solve the problem of PND. Experiments results and conclusions are presented in section 3 and section 4, respectively.

2. Non-Rigid Structure from Motion

2.1 Procrustean Normal Distribution

PND [7] is a special normal distribution of shape deformation that constrains the non-rigid motions, so that 3D shapes are aligned as closely as possible in its distribution. It determines rigid motions by using modified GPA constraints to minimize non-rigid variation, which improve the accuracy of rotations in NRSfM. GPA can be used to superimpose the multiple landmark shapes to a common reference using rigid transformations. The GPA problem is defined as:

$$\min_{s_i, R_i, t_i, \overline{X}} \sum_{i=1}^{T} \left\| s_i R_i X_i + t_i - \overline{X} \right\|_F^2 \qquad (1)$$
$$\text{subject to } R_i^T R_i = 1, \ \left\| s_i X_i \right\|_F = 1$$

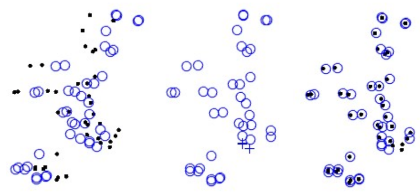
where $X_i \in \mathbb{R}^{3\times n}$, $s_i \in \mathbb{R}^{1\times 1}$, $R_i \in \mathbb{R}^{3\times 3}$, $t_i \in \mathbb{R}^{3\times 1}$ are the 3D shape, scale, rotation, and translation, respectively, for the *i*-th input frame. T is the total number of frames and n is the number of points. $\|A\|_F$ denotes the Frobenius norm of matrix A, *i.e* $\|A\|_F^2 = tr(A^T A) = \|vec\|_2^2$ with a vectorization operator $vec(\cdot)$. Also $\overline{X}$ is the mean shape, represented as:

$$\overline{X} = \frac{1}{T} \Sigma s_i R_i X_i \qquad (2)$$

where T is the number of frames. PND introduces an additional constraint that the norm of the mean shape is one, *i.e.* $\|\overline{X}\|_F^2 = 1$. This constraint can also be rewritten as:

$$vec(s_i R_i X_i)^T vec(\overline{X}) = s_i tr(R_i X_i \overline{X}) = 1 \qquad (3)$$

The modified GPA constraints can be represented as:



(a) (b) (c)

Fig. 2. Reconstructed 3D shape (blue circle) of *Pickup* against ground truth (black dots). a) Result of PND; b) Classification of nearly rigid points (circle) and non-rigid points (plus); c) Result of proposed method.

$$\|\overline{X}\|_F^2 = 1, \ R_i^T R_i = 1, \ s_i tr(R_i X_i \overline{X}^T) = 1, \ R_i X_i \overline{X}^T \in S_+^3 \qquad (4)$$

where S_+^3 is a set of three dimensional positive semi-definite matrices (PSDs), which is convex. The last two constraints in (4) are the convex constraints of aligned shapes. They define the distribution of PND as a special normal distribution. The next step is to use EM algorithm to calculate motion parameters and 3D shapes by maximizing the following log-likelihood function:

$$\begin{aligned} &\sum \log(p(W_i, X_i \mid s_i, R_i, \overline{X}, \Sigma)) \\ &= \sum \log(p(W_i \mid X_i)) + \log(p(X_i \mid s_i, R_i, \overline{X}, \Sigma)) \end{aligned} \qquad (5)$$

where W_i is the 2D observed points of *i*-th 3D shape.

PND is the most accurate algorithm at present [7]. It outperforms other existing algorithms such as Dai's method [2] which is known as state-of-the-art for NRSfM, the performance is inferior in several datasets. Since PND use the modified GPA to calculate rigid motions, a few non-rigid points of an object can cause inaccurate calculation. As seen in Fig. 2. (a), just two non-rigid points can cause poor reconstruction performance.

2.2. Proposed Reconstruction Method

The basic idea of our reconstruction algorithm is: firstly, we use PND to recover 3D shapes as first estimation; then, we classify all points of the object into nearly rigid points and non-rigid points; after that, we use EM algorithm to reconstruct nearly rigid points again. The details of the proposed method are described in Algorithm1.

We present two novel methods to classify two types of points. Since the deformation of non-rigid points is stronger than rigid points, we consider the time varying 2D displacement of these points is also bigger than rigid points. Thus, our first method simply classifies the points just from the 2D observed points. Let $x_{i,j}$ be the *j*-st 2D observed point in frame *i*, then the 2D displacement of *j*-st point in all frames can be calculated by:

$$dis_j = \sum_{i=1}^{T} (x_{i+1,j} - x_{i,j}) \qquad (6)$$

We call the above method as Method1 in this paper. Owing to the fact that PND also calculates the rigid motion parameters (s, R, t) that are used to align 3D shapes to its mean shape, the other approach is to measure the Euclidean distance between 3D aligned shapes and mean shape. The distance of non-rigid point to its corresponding point in mean shape is greater than the distance of nearly rigid point to its corresponding point in mean shape. This method is also very easy to implement. s_i, R_i, t_i are the scale, rotation, and translation, respectively, for the *i*-th input frame. $X_{i,j}$ is the *j*-st 3D point recovered by PND in frame *i*. For the *j*-st point in each aligned shape, we calculate the sum of Euclidean distance to the corresponding point in mean shape by:

$$\mathrm{dis}_j=\sum_{i=1}^{T}\left\|s_i R_i X_{i,j}+t_i-\overline{X}_j\right\|_F^2 \tag{7}$$

We call this method as Method2 in this paper.

Algorithm1

Input: 2D observations of T frames, n tracking points.

Output: recovered 3D shapes.

1. Use PND to recover 3D shapes as first estimation.
2. For every observed 2D point x_j, calculate its 2D time varying displacement dis_j in all frames by equation (6). Or for every recovered 3D point , calculate it's 3D distance dis_j to the corresponding point in mean shape by using equation (7).
3. Normalize dis_j to 0~1.
   if($\mathrm{dis}_j<\sigma$)
      X_j is nearly rigid point.
4. Use EM algorithm of PND to reconstruct the nearly rigid points again.
5. Combine the new recovered 3D points with non-rigid points calculated in step 1 as final recovered 3D shapes.

3. Experiments

In this section, we compared our proposed method against PND and Dai's method (BMM) [2]. Since we proposed two approaches Method1 and Method2 to classify the nearly rigid points and non-rigid points, it's necessary to evaluate the reconstruction performance for each method. We chose the most familiar datasets in this field which contains *Shark* (240/91), *Walking* (260/55) and *Face* (316/40) of [1]; *Drink* (1102/41), *Pickup* (357/41), *Yoga* (307/41), *Stretch* (370/41), *Dance*(264/75) of [4], where (T/n) denotes the number of frames (T) and points (n).The low rank parameter of BMM was set in accordance with their original paper. Following the parameter setting methodology in [2, 3, 4], we set the threshold parameter σ in Method1 and Method2 with different values from 0.2 to 1 ($\sigma\in\{0.2, 0.25, 0.3,\cdots,1\}$) for different data and reported best result. The average 3D reconstruction error of T frames was measured using the same equation as PND:

$$\mathrm{error}=\frac{1}{T}\sum_{i=1}^{T}\left\|X_i-X0_i\right\|_F/\left\|X0_i\right\|_F \tag{8}$$

where X_i and $X0_i$ is recovered 3D shape and ground truth respectively for the *i*-th input frame.

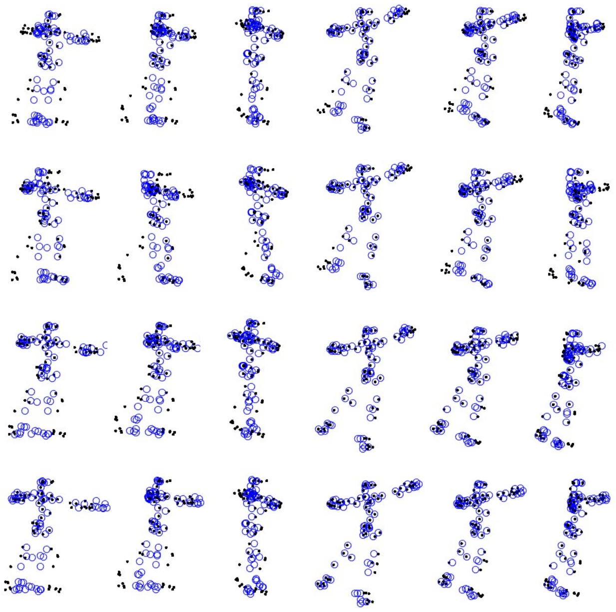


Fig. 3. From first to last row: 3D shapes of *Dance* at frame 1, 11, 30, 130, 141, 150 recovered by BMM, PND, proposed method Method1 and Method2 respectively. Recovered shapes are blue circles and ground truth is dark dots.

Table 1. 3D reconstruction error of BMM, PND and our proposed method: Method1 and Method2.

Method	BMM	PND	Proposed method	
			Method1	Method2
Walking	0.0861	0.0465	0.0462	**0.0435**
Face	0.0223	0.0165	0.0170	**0.0158**
Yoga	0.0224	0.014	0.0130	**0.0119**
Stretch	0.0288	0.0156	**0.0146**	**0.0146**
Pickup	0.0356	0.0372	**0.0164**	**0.0164**
Drink	0.0216	0.0037	**0.0036**	**0.0036**
Dance	0.1451	0.1834	**0.1438**	0.1512
Shark	0.5475	0.0134	**0.0122**	0.0135

We show the experimental result for 3D reconstruction errors. As shown in Table 1, the best reconstruction performance was achieved by the proposed method for all of these datasets. Method1 is better than PND except *Face* and Method2 is better than PND except *Shark*. Fig.3 shows the front view of shapes of *Dance* at frame 1, 11, 30, 130, 141, 150 recovered by different methods. It is seen that although PND was inferior to BMM, the performance was largely improved by our proposed method. Also, Fig.4 shows the side view of reconstruction results of Walking, Face, Yoga by PND and Method2. Our experiments showed that the proposed method can achieve more accurate reconstruction performance than

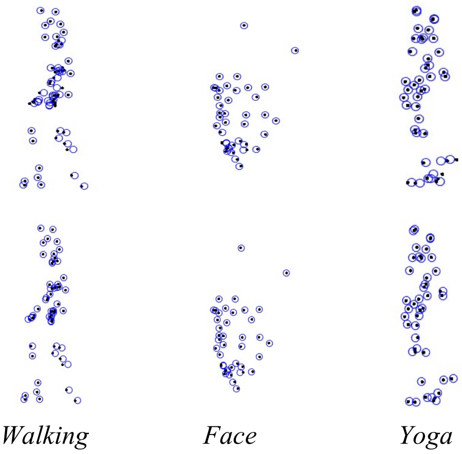


Fig. 4. From first to second row: side views of 3D shapes recovered by PND, Method2, respectively.

Fig. 5. The 3D reconstruction error of Method1 and Method2 for *Pickup* with different value of threshold.

PND. Also, Fig.5 shows the 3D reconstruction error of *Pickup* with different thresholds for classifying non-rigid points. *Pickup* is typical data that the quality of the reconstruction largely suffers in the presence of few non-rigid points. As is clear in Fig.5, the classification performance of Method 2 is more similar than Method1 for different thresholds.

Our method can improve the reconstruction quality for the data that suffers in the presence of few non-rigid points, such as *Pickup*. For data like this, we believe that both Method 1 and Method 2 can achieve good reconstruction performance. Also, we consider that Method 1 is better for data with strong deformation such as *Dance*, because PND would cause large 3D reconstruction error for strong deformation and it would be more difficult to classify the points in 3D space calculated by PND.

We explain clearly that our method can't be used to improve the reconstruction performance for any non-rigid objects. Also, its performance is guaranteed only when suitable threshold is given. However, there is still no automatic method to determine this threshold.

4. Conclusion

In this paper, we solved the problem of PND is that in some cases a few non-rigid points can largely affect the reconstruction performance. We demonstrated that nearly rigid point or non-rigid point can be simply classified by calculating its 2D displacement or 3D distance to the corresponding point in mean shape, thus the affection of non-rigid points can be easily eliminated. The proposed method achieved more accurate performance than BMM and PND. However, now we still don't know the automatic selection way of the value of σ for different data. It is left for our future work.

5. Acknowledgements

The authors would like to thank Y. Dai and M. Lee for the distribution of their code.

References

[1] L. Torresani, A. Hertzmann, and C. Bregler, "Nonrigid structure-from-motion: estimating shape and motion with hierarchical priors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 5, pp. 878–892, 2008.

[2] Y. Dai, H. Li, and M. He. "A simple prior-free method for non-rigid structure from motion factorization," in *Computer Vision and Pattern Recognition (CVPR)*, pp. 2018 – 2025, 2012.

[3] P.F.U. Gotardo and A.M. Martinez. "Non-rigid structure from motion with complementary rank-3 spaces," In *Computer Vision and Pattern Recognition (CVPR)*, pp. 3065-3072, 2011.

[4] I. Akhter, Y. Sheikh, S. Khan, and T. Kanade, "Nonrigid structure from motion in trajectory space," in *Advances in Neural Information Processing Systems (NIPS)*, pp. 42-48, December 2008.

[5] Li K, Yang J, Jiang J. "Non-rigid structure from motion via sparse representation," *International Conference on Multimedia & Expo (ICME)*, pp. 1-6, 2014.

[6] K. Fragkiadaki, M. Salas, P. Arbelaez, and J. Malik."Grouping -based low-rank trajectory completion and 3D reconstruction," In *Advances in Neural Information Processing Systems (NIPS),* pp. 55-63, 2014.

[7] M. Lee, J. Cho, C. H. Choi, and S. Oh. "Procrustean normal distribution for non-rigid structure from motion," In *Computer Vision and Pattern Recognition (CVPR)*, pp. 1280-1287, 2013.

[8] Tomasi Carlo and Takeo Kanade. "Shape and motion from image streams under orthography: a factorization method," *International Journal of Computer Vision*, vol. 9, no. 2, pp.137–154, 1992.

[9] J. Xiao, J. Chai, and T. Kanade. "A closed form solution to non-rigid shape and motion recovery," *International Journal of Computer Vision*, vol. 67, no .2, pp. 33–246, 2006.

[10] C. Bregler, A. Hertzmann, and H. Biermann. "Recovering non-rigid 3D shape from image streams," In *Computer Vision and Pattern Recognition (CVPR)*, vol.2, pp. 690-696, 2000.

[11] H. S. Park, T. Shiratori, I. Matthews, and Y. Sheikh. "3D reconstruction of a moving point from a series of 2D projections," In *European Conference on Computer Vision (ECCV)*, pp. 158-171, 2010.

[12] R. Garg, A. Roussos, and L. Agapito. "Dense variational reconstruction of non-rigid surfaces from monocular video." In *Computer Vision and Pattern Recognition (CVPR)*, pp. 1272-1279, 2013.

[13]Gower, J.C., "Generalized Procrustes analysis." *Psychometrika*, vol. 40, no. 1, pp. 33-51, 1975