This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2014.2308415, IEEE Transactions on Multimedia

Non-Rigid Structure-from-Motion with Uniqueness Constraint and Low Rank Matrix Fitting Factorization

Imran Khan

Abstract

Non-rigid structure-from-motion is one of the difficult and challenging problems in computer vision, especially when the only input available is 2D correspondences in monocular video sequence. This paper proposed a new constraint based framework for underconstrained non-rigid structure-from-motion problem to constrain the space of solution. The proposed method is based on a point trajectory approach with an additional uniqueness constraint applied to shape coefficients to reduce the basis required to construct the non-rigid 3D shape. A framework for occluded and incomplete measured data is also proposed using low rank matrix fitting which is a robust factorization scheme for the matrix completion problem. This method offers not only new theoretical insight, but also a practical, everyday solution, to non-rigid structure-from-motion. The proposed method is positively compared to the state-of-the-art in non-rigid structure-from-motion, providing improved results on high-frequency deformations of both articulated and simpler deformable shapes.

Index Terms

Non-rigid structure-from-motion, uniqueness constraint, low rank matrix fitting, least squares estimation.

Imran Khan is with the Department of Computer Engineering, Center for Advanced Studies In Engineering, Islamabad, Pakistan, e-mail: imi_case@yahoo.com, tel: +92 51 5507256

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2014.2308415, IEEE Transactions on Multimedia

I. Introduction

The display of images in 3D provides important benefits for everyone interacting with the display. The added depth information in 3D images allows comprehending images more quickly and precisely. In computer vision non-rigid structure-from-motion (NR-SFM) is a way of getting time varying 3-D information of deformable object in a scene from two or more 2-D images with interest points. The NR-SFM is proved to be a promising framework in many computer vision applications [1–5] ranging from tracking, human computer interaction, motion-capture, 3D face reconstruction to augmented reality [6–8]. It is also extensively used in medical applications where 3D information is used for better examining the body organs and helps to diagnose various diseases that otherwise are not possible if the only 2D information is available. The entertainment industry also uses this technology in making animations close to reality, majority of which are based on 3D technology.

Matrix factorization is widely applied for solving SFM and NR-SFM problems [9–12]. The input to this procedure is a tracking or measurement matrix W that is a result of a feature tracking algorithm and is formed by collecting all features tracked point correspondences. In the presence of missing data in measurement matrix W of rank r formulates NR-SFM problem more complicated and exigent. In factorization process, the measurement matrix W of size m $\times$ n is decomposed into low-rank matrices M and S of dimensions m $\times$ r and r $\times$ n, such that the error between measurement matrix W and product of two matrices M and S is minimized [10]. Also the rapidly changing shape of object in NR-SFM makes the process even more difficult to model because of infinite deformation patterns.

Structure-from-motion term is first coined by Tomasi and Kanade [10]. The authors used factorization approach to recover 3D shape and motion from monocular image sequence under orthographic projection model. Later this work is extended for non-rigid structure-from-motion by Bregler *et al.* [9] in their seminal work. They proposed a framework for approximating the structure of 3D object as a linear combination of basis shapes under orthonormality constraints. This model pioneered new computational and theoretical challenges in this field. Xiao *et al.* [13] introduced additional constraints to remove ambiguity in the solution that was due to insufficient of orthonormality constraints. To improve reconstruction process a rank deficient basis are used. But due to the presence of missing and noisy data, this closed form solution fails to deliver improved results. Torresani *et al.* [14] selected a framework based on priors to constrain the solution. They introduced a Gaussian prior to reduce the coefficients. They also imposed a metric constraint for camera matrices, but the update of the camera matrix is only an approximation and is not an optimal. Brand [15] improved the numerical stability of the estimation process in NR-SFM by

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2014.2308415, IEEE Transactions on Multimedia

directly minimizing deviation from the required orthogonal structure of the projection matrix. All these advances are founded on an assumption of prior knowledge of the measured data, also results are shown using objects having a significant number of points moving rigidly.

Akhter *et al.* [16] proposed trajectories based model for non-rigid 3D object. They proposed a model based on DCT basis trajectories instead of shape basis, the 3D point trajectories are modeled compactly in the domain of the Discrete Cosine Transform (DCT) basis vectors. The trajectory based approach provides improved results on articulated and complex shapes. However, this model fails to present high-quality results in the case of high-frequency deformation without relaxing the rank-3K constraint, also obtained results are over-smoothed. Paladini *et al.* [17] described an alternating least-squares approach associated with a globally optimal projection step onto the manifold of metric constraints, this approach is especially useful in the case of missing data in input measurement matrix and also equally applicable on real video sequences with occluded input data. Constraints are extremely important as illustrated in literatures to attain computationally efficient and numerically stable results for NR-SFM. In [18–22], the authors used priors to better constrain the solution for NR-SFM problem, however unsatisfactory results are reported in [15] when the basic low-rank shape model is used to find 3D shape, so priors are essential to better model the deformations that otherwise difficult specially for NR-SFM.

The paper proposes a novel framework for NR-SFM by modeling non-rigid 3D shape deformations as a point moving in shape space in certain trajectory, by considering only unique points in this trajectory makes the representation of non-rigid shape in shape space possible. The method exploits the repetition of non-rigid shape by considering only those trajectory points that uniquely specify shape at some time instant. Thus, it reduces the possible repetitive trajectory points as shown in Fig. 1. The approach restricts the number of basis and their coefficients required to represent the non-rigid shape by introducing a uniqueness constraint. The proposed system also encounters the partially filled measurement matrix by using low rank matrix fitting which is a robust factorization scheme for the matrix completion problem. This approach modified the standard matrix factorization technique and provides improved reconstruction results for high-frequency deformable and articulated shapes.

The outline of this paper is as follows. The classical factorization model for NR-SFM is described in Section 2. The proposed model for 3D-shape computation is outlined in Section 3. Optimization process is detailed in Section 4. A scheme for handling occluded data is presented in Section 5. Experimental results are discussed in Section 6. Finally, conclusions are drawn in Section 7.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2014.2308415, IEEE Transactions on Multimedia

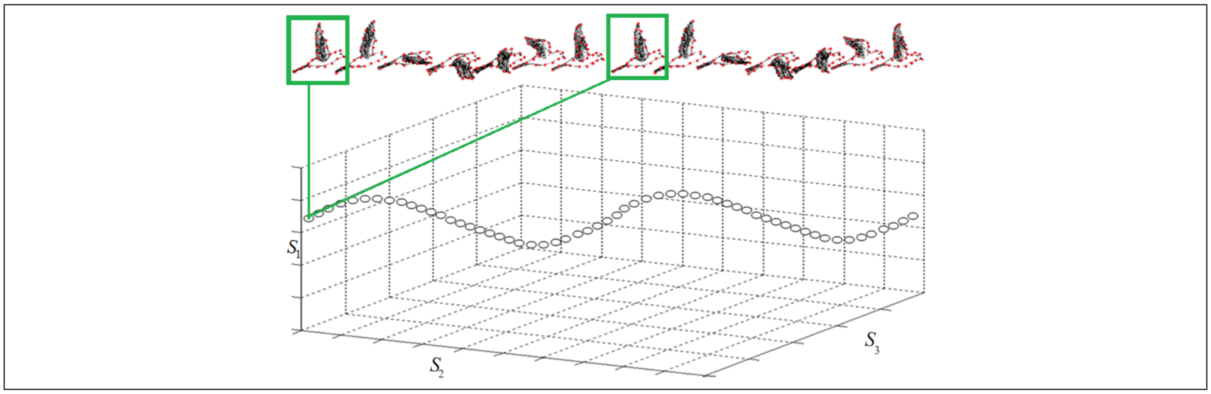

Figure 1. The 3D shape trajectory model where the non-rigid 3D shape is represented as a single point moving in certain trajectory in a shape space. The first frame and eighth frame are identical and represented by unique basis with their coefficients in shape space.

II. Classical Matrix Factorization Model in NR-SFM

Factorization technique is most widely used for obtaining camera matrices and 3D structure using image feature correspondences as an input, especially in the case of rigid structure-from-motion [23–25]. The NR-SFM problem is modeled by considering F frames and P points, the orthographic 2D projection at time instant t of the jth 3D point is denoted by $\left[u_P^{(F)}, v_P^{(F)}\right]^T$ and by arranging all these input points in a stack formed a tracking matrix W defined as,

$$\mathrm{W}^{2F\times P} = \begin{bmatrix} u_1^{(1)} & \cdots & u_P^{(1)} \\ v_1^{(1)} & \cdots & v_P^{(1)} \\ & \vdots & \\ u_1^{(F)} & \cdots & u_P^{(F)} \\ v_1^{(F)} & \cdots & v_P^{(F)} \end{bmatrix} \epsilon\, \mathbb{R}^{2F\times P}.$$

The complete tracking matrix with no missing entries is shown above, here it is assumed that the contents of the measurement matrix W are registered to the centroid of the object by subtracting mean column vector t $\epsilon\, \mathbb{R}^{2F}$ from W making it zero-mean also all the frames are in sequence. The first two rows of W define the complete contents of the frame at time instant $t = 1$ in superscript, and P define the number of points in ith frame in subscript. Bregler *et al.* [9] describes the non-rigid 3D shape at time instant t as a linear combination of K key-frame basis set S_1,$S_2 \cdots S_K$ where (K = 1,2$\cdots$K) define the number of shape basis as,

$$\text{S} = \sum_{i=1}^{K} \; l_i \;\; S_i, \tag{1}$$

here, each key frame S_i is a 3×P matrix describing P points in frame F, the tracking matrix W containing feature correspondence is decomposed using singular value decomposition (SVD) into two low-rank matrices $\hat{M}^{2F\times 3K}$, $\hat{S}^{3K\times P}$and is defined as,

$$\text{W} = \text{U}\Sigma V^T = (\text{U}\Sigma^{\frac{1}{2}})(\text{U}\Sigma^{\frac{1}{2}}) = \text{dot}(\; \hat{M}^{2F\times 3K} \;\; \hat{S}^{3K\times P}),$$

so this model eventually decomposes W as the product of two low-rank matrices $\hat{\text{M}}$ and $\hat{\text{S}}$ and considering the 3K columns of these matrices.

$$\text{W} \; = \; \overbrace{\text{R}(\text{G}\otimes I_3)}^{\text{M}} \overbrace{\begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_K \end{bmatrix}}^{\text{S}}, \tag{2}$$

I_3 represents identity matrix and $\otimes$ denotes the Kronecker product in (2), $\text{S}\,\epsilon\,\mathbb{R}^{3K\times P}$ define the 3D shape structure, the coefficients of factor M are separated in a block-diagonal rotation matrix $\text{R}\,\epsilon\,\mathbb{R}^{2F\times 3F}$ and a shape coefficient matrix $\text{G}\,\epsilon\,\mathbb{R}^{F\times K}$ defined as,

$$\text{R} = \begin{bmatrix} r_1 & & & \\ & r_2 & & \\ & & \ddots & \\ & & & r_F \end{bmatrix}, \quad \text{G} = \begin{bmatrix} g_{1,1} & \cdots & g_{1,K} \\ g_{2,1} & \cdots & g_{2,K} \\ \vdots & \ddots & \vdots \\ g_{F,1} & \cdots & g_{F,K} \end{bmatrix},$$

$\text{r}_t\epsilon\,\mathbb{R}^{2\times 3}$ represents the orientation of the camera at each frame t and each row of matrix G represents the contribution of shape basis for each frame.

III. Proposed Model for 3D-Shape Computation

This section introduces a proposed framework for NR-SFM by modeling 3D shape deformations as a single point moving in shape space in certain trajectory and considering only unique points in this trajectory. The concept described here can be easily grasped by considering Fig. 1, presenting the bird's flight. In this figure shape space is formed by the combination of shape basis $\text{S}_1 \ldots .\text{S}_K$. The small

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2014.2308415, IEEE Transactions on Multimedia

References

[1] J. K. Aggarwal and Q. Cai, "Human motion analysis: A review," in *Nonrigid and Articulated Motion Workshop, 1997. Proceedings., IEEE*. IEEE, 1997, pp. 90–102.

[2] C. Nastar and N. Ayache, "Frequency-based nonrigid motion analysis: Application to four dimensional medical images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, no. 11, pp. 1067–1079, 1996.

[3] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 231–268, 2001.

[4] C. Studholme, D. L. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3d medical image alignment," *Pattern recognition*, vol. 32, no. 1, pp. 71–86, 1999.

[5] F. Rengier, A. Mehndiratta, H. von Tengg-Kobligk, C. M. Zechmann, R. Unterhinninghofen, H.-U. Kauczor, and F. L. Giesel, "3d printing based on imaging data: review of medical applications," *International journal of computer assisted radiology and surgery*, vol. 5, no. 4, pp. 335–341, 2010.

[6] J. Zhu, S. C. Hoi, and M. R. Lyu, "Real-time non-rigid shape recovery via active appearance models for augmented reality," in *Computer Vision–ECCV 2006*. Springer, 2006, pp. 186–197.

[7] M. J. Black and Y. Yacoob, "Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion," in *Computer Vision, 1995. Proceedings., Fifth International Conference on*. IEEE, 1995, pp. 374–381.

[8] D. M. Gavrila, "The visual analysis of human movement: A survey," *Computer vision and image understanding*, vol. 73, no. 1, pp. 82–98, 1999.

[9] C. Bregler, A. Hertzmann, and H. Biermann, "Recovering non-rigid 3d shape from image streams," in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, vol. 2. IEEE, 2000, pp. 690–696.

[10] C. Tomasi and T. 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.

[11] M. Brand, "A direct method for 3d factorization of nonrigid motion observed in 2d," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2. IEEE, 2005, pp. 122–128.

[12] L. Torresani, D. B. Yang, E. J. Alexander, and C. Bregler, "Tracking and modeling non-rigid objects with rank constraints," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. I–493.

[13] J. Xiao, J.-x. Chai, and T. Kanade, "A closed-form solution to non-rigid shape and motion recovery," in *Computer Vision-ECCV 2004*. Springer, 2004, pp. 573–587.

[14] L. Torresani, A. Hertzmann, and C. Bregler, "Learning non-rigid 3d shape from 2d motion," *Advances in Neural Information Processing Systems*, vol. 16, 2003.

[15] M. Brand, "A direct method for 3d factorization of nonrigid motion observed in 2d," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2. IEEE, 2005, pp. 122–128.

[16] I. Akhter, Y. Sheikh, S. Khan, T. Kanade *et al.*, "Nonrigid structure from motion in trajectory space," in *Neural information processing systems*. Citeseer, 2008, pp. 41–48.

[17] M. Paladini, A. Del Bue, M. Stosic, M. Dodig, J. Xavier, and L. Agapito, "Factorization for non-rigid and articulated structure using metric projections," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 2898–2905.

[18] A. Del Bue, X. Llad, and L. Agapito, "Non-rigid metric shape and motion recovery from uncalibrated images using priors," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1. IEEE, 2006,

pp. 1191–1198.

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

[20] A. Bartoli, V. Gay-Bellile, U. Castellani, J. Peyras, S. Olsen, and P. Sayd, "Coarse-to-fine low-rank structure-from-motion," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[21] S. Olsen, A. Bartoli *et al.*, "Using priors for improving generalization in non-rigid structure-from-motion," in *British Machine Vision Conference*, vol. 2, 2007, p. 3.

[22] H. AanÙĹs and F. Kahl, "Estimation of deformable structure and motion," Citeseer, Tech. Rep., 2002.

[23] A. M. Buchanan and A. W. Fitzgibbon, "Damped newton algorithms for matrix factorization with missing data," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2. IEEE, 2005, pp. 316–322.

[24] S. Chen, Y. Wang, and C. Cattani, "Key issues in modeling of complex 3d structures from video sequences," *Mathematical Problems in Engineering*, vol. 2012, 2011.

[25] A. Zaheer, I. Akhter, M. H. Baig, S. Marzban, and S. Khan, "Multiview structure from motion in trajectory space," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2447–2453.

[26] Z. Wen, W. Yin, and Y. Zhang, "Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm," *Mathematical Programming Computation*, vol. 4, no. 4, pp. 333–361, 2012.

[27] G. H. Golub and C. F. van Van Loan, "Matrix computations (johns hopkins studies in mathematical sciences)," 1996.

[28] M. Paladini, A. Del Bue, M. Stosic, M. Dodig, J. Xavier, and L. Agapito, "Factorization for non-rigid and articulated structure using metric projections," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 2898–2905.

[29] P. F. Gotardo and A. M. Martinez, "Kernel non-rigid structure from motion," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 802–809.

[30] A. Del Bue, J. Xavier, L. Agapito, and M. Paladini, "Bilinear modeling via augmented lagrange multipliers (balm)," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 8, pp. 1496–1508, 2012.

[31] A. Datta, Y. Sheikh, and T. Kanade, "Linear motion estimation for systems of articulated planes," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

Imran Khan received the B.S degree in computer science from Barani Institute of Information Technology (UAAR), MS(CE) from Center for Advance Studies in Engineering (CASE), Islamabad, Pakistan. His current research interests focus on image enhancement and reconstruction, super-resolution, 3D face reconstruction, scene change detection, non-rigid structure from motion.