A Simple Method to Recover 3-D Rigid Structure from Motion using SIFT, RANSAC and the Tomasi-Kanade Factorization

Samir H. Abdul-Jauwad[1], Rehmat Ullah[2] and Farman Ullah[3]

[1]Department of Electrical Engineering, King Fahd University of Petroleum & Minerals, Dhahran, 31261, Saudi Arabia

[2]Department of Computer Systems Engineering, University of Engineering & Technology, Peshawar, 25000, Pakistan

[3]Department of Electrical Engineering, COMSATS Institute of Information Technology, Attock, 43600, Pakistan

Abstract

Traditionally two frames are used to estimate the 3-D structure, while recent approaches have made use of a long sequence of frames. The latter gives a better recovery of a structure because it amasses temporal information over time. Tomasi-Kanade factorization also assumes all features to be visible throughout the entire image stream. This results in a dense 2-D cloud and therefore allows full recovery of the entire 3-D structure.

This paper addresses the problem of 3-D structure reconstruction from motion by using the Tomasi-Kanade factorization method applied to a sequence of frames. Orthographic projection and rigidity is assumed and the singular value decomposition technique is used to factor the measurement matrix (W) into two matrices which correspond to the object's 3-D structure (S) and camera rotation (R) respectively. To construct W, feature correspondences are established by applying a SIFT tracker following which RANSAC is used to discard the false matches detected by the SIFT tracker. The 3-D point-cloud (S) is converted to a mesh of triangles by connecting the nearest three neighboring points. Finally, the mesh of triangles is rendered by means of the Plyview from CyberWare to achieve the final 3-D structure.

Keywords: *3-D Structure, Tomsi-Kanade Factorization Method, Singular Value Decomposition (SVD), SIFT Tracker, RANSAC.*

1. Introduction

In a problem of 3-D reconstruction from motion, both the object structure and camera motion need to be estimated. The following aspects should be taken into account while attempting to solve this problems.

Projection model: Two popular models, namely orthographic and perspective, are widely discussed in the literature. Let $[X, Y, Z]^T$ represent an object point in 3-D space and $[x, y]$ represent its projection on the image plane. Orthographic and perspective projections can then be expressed as follows:

$$\textit{orthographic}: \quad x = X; y = Y$$

$$\textit{perspective}: \quad x = fX / Z; y = fY / Z$$

where f is focal length of camera. The perspective projection model is more realistic than the orthographic one. But the Z in the denominator gives rise to non-linearity which is more difficult to manipulate.

Rigid motion model: Let $[X', Y', Z']^T$ designate the new position of the object after a rigid motion. This rigid motion can be expressed mathematically as follows:

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + T$$

Here R is a rotation matrix and T is a translation vector. The objective is to estimate R and $[X, Y, Z]^T$, given a series of features points $[x, y]$'s.

Number of frames used: In the earlier literature, mostly two frames were used [1] [2] [3], while long-sequence based approaches [4-7] later became popular. The long-sequence based approaches give a better estimation of structure of interest. The structure must remain unchanged while accumulating temporal information over time instants.

Feature correspondence: Feature correspondence plays a vital role in the success of a structure from motion algorithm. Most algorithms [2-7] described in the literature employ sparse feature correspondence. In the case of long sequence, how best to track features over a long time both stably and efficiently remains an open question in the literature.

Considering the above, the proposed 3-D reconstruction scheme is shown in Fig. 1. Given an input image sequence, feature points are detected and tracked first, following which the Tomasi-Kanade algorithm is employed to estimate both the camera motion and the object's shape. However, this structure is not unique and further orthogonality constraints are applied to obtain a unique structure.

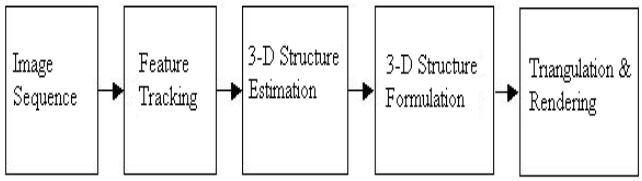

Fig. 1: A general scheme of 3-D structure from motion.

In this research, feature correspondences is achieved using the well-known SIFT tracker [8], whose robustness has already been proved by researchers, and implemented the reconstruction algorithm proposed by Tomasi and Kanade [4]: the so-called factorization method over a long sequence. Next, triangulation is performed on the structure (3-D point cloud), filtering out incorrect polygons and the output is a mesh of triangles. Finally, the mesh is converted into a .ply file which is then rendered by Plyview; a tool developed by CyberWare.

This paper is organized as follows. Section 2 discusses the Tomasi-Kanade factorization method. Section 3 provides experiments and results and Section 4 concludes the research.

2. Methodology: the Tomasi-Kanade Factorization Method

Suppose that F frames of a video sequence are available to us and P features points over F frames have been detected and tracked. In other words, the set $\{(x_{fp}, y_{fp}), f = 1,..., F; p = 1,..., P\}$ is known to us. We

form a measurement matrix W with its dimension being $2F \times P$ by subtracting the mean of each row from each point in that row, as follows:

$$W = \begin{bmatrix} x_{11} & \cdots & x_{1P} \\ y_{11} & & y_{1P} \\ \cdot & \cdots & \cdot \\ x_{F1} & & x_{FP} \\ y_{F1} & & y_{FP} \end{bmatrix} - \begin{bmatrix} x'_1 \\ y'_1 \\ \cdot \\ x'_F \\ y'_F \end{bmatrix};$$

$$x'_f = \frac{1}{P}\sum_{p=1}^{P} x_{fp}\,;\; y'_f = \frac{1}{P}\sum_{p=1}^{P} y_{fp}$$

This subtraction is performed for each frame to recenter the camera reference point at the object centroid by constraining the feature locations to have zero mean.

Assuming orthographic projection and noise-free features correspondences, the measurement matrix W is factored by using the factorization method into two matrices $\bar{R}$ ($2F \times 3$ by dimension) and $\bar{S}$ ($3 \times P$ by dimension), representing camera rotation and object shape/structure respectively. This can be expressed as $W = \bar{R}\bar{S}$. In the presence of noise, a singular value decomposition (SVD) technique is used to provide an approximation to $\bar{S}$ and $\bar{R}$.

However, this decomposition is not unique up to any invertible 3×3 matrix Q. We therefore impose constraints on the orthogonality of rotation matrix $\bar{R}$ to

yield the matrix Q. Once Q is known, we estimate the rotation matrix $R = \bar{R}Q$ and the shape matrix $S = Q^{-1}\bar{S}$. This workflow is shown in Fig. 2.

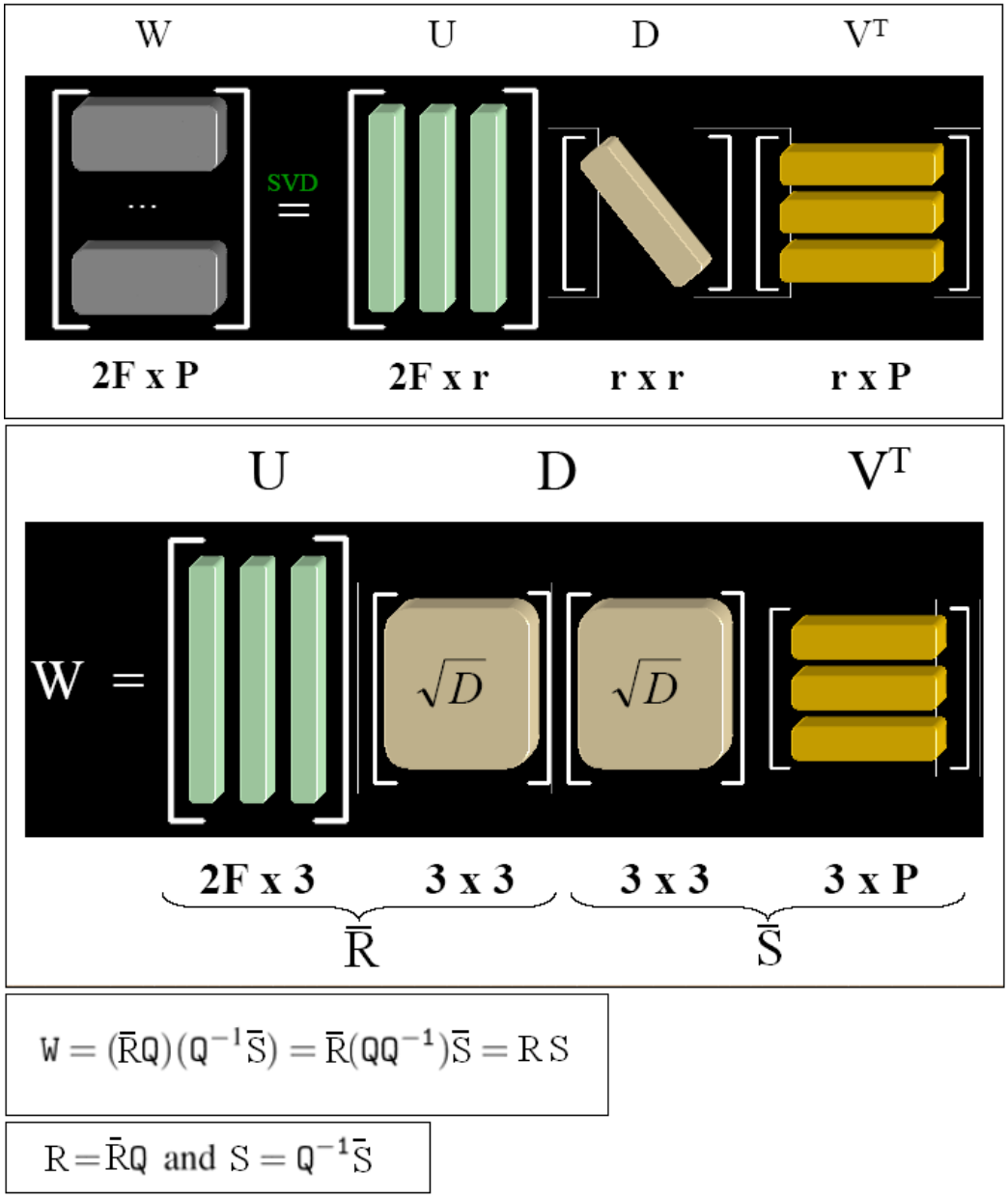

Fig. 2: The Tomasi-Kanade factorization method.

3. Experiments and Results

The code was written in MATLAB and experiments were performed using three common datasets from the literature namely *coin*, *cow* and *dino*. The results are shown below. As evidence by the below figures showing the reconstructed 3-D structures, the Tomasi-Kanade factorization method gives pretty good approximation. Looking at the reconstructed 3-D structures, two important observations are (1) that, some parts are not visible in the three datasets; (2) that the feature points are dense in some parts while sparse in other parts. This is due to these reasons that we can only partially, not fully, recover the structure. We recover what is visible. This is where we need robust and efficient feature correspondences in order to recover a complete structure.

Experiment 1: Dataset: coin

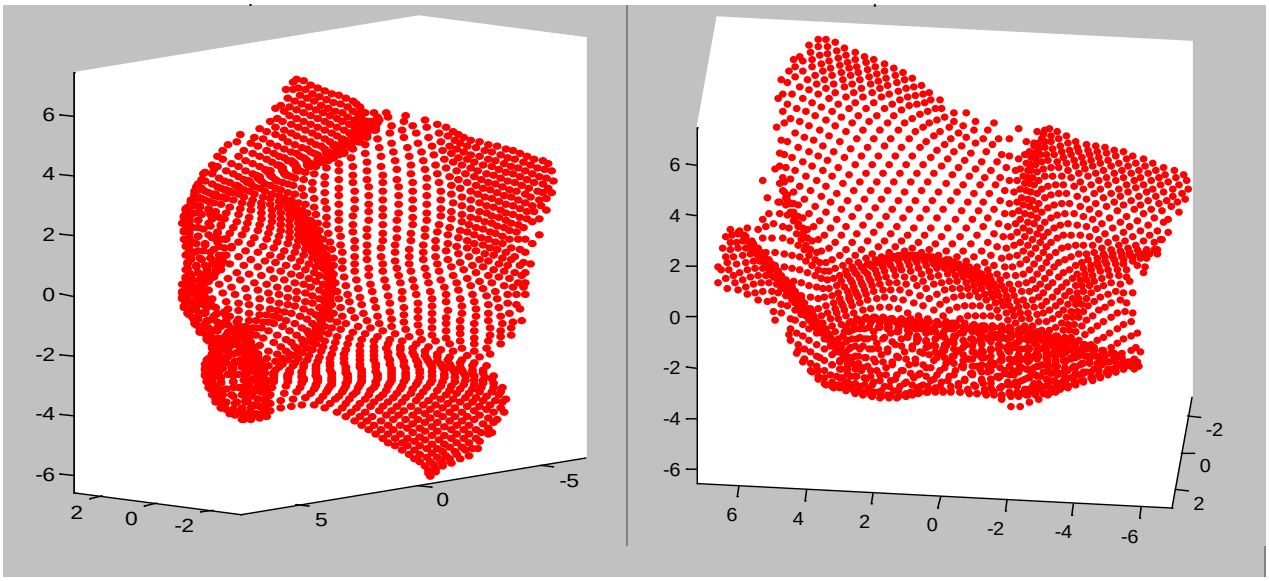

Fig. 3: Two rotated versions of 3-D structure.

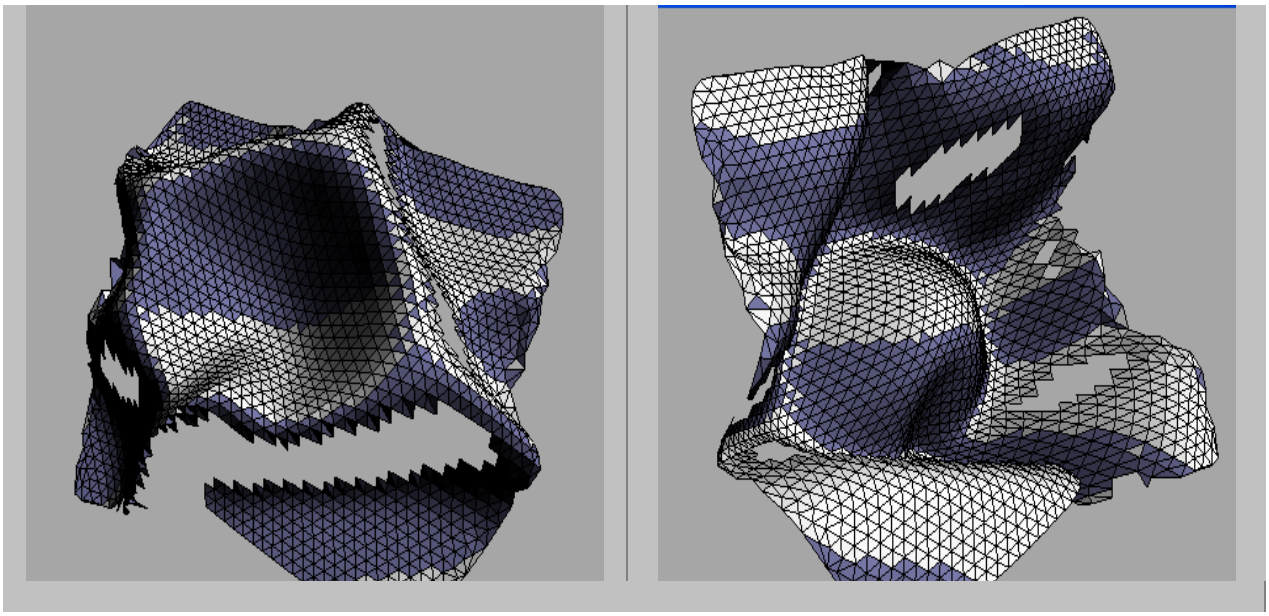

Fig. 4: Two rotated versions of 3-D structure after triangulation and rendering.

Experiment 2: Dataset: cow

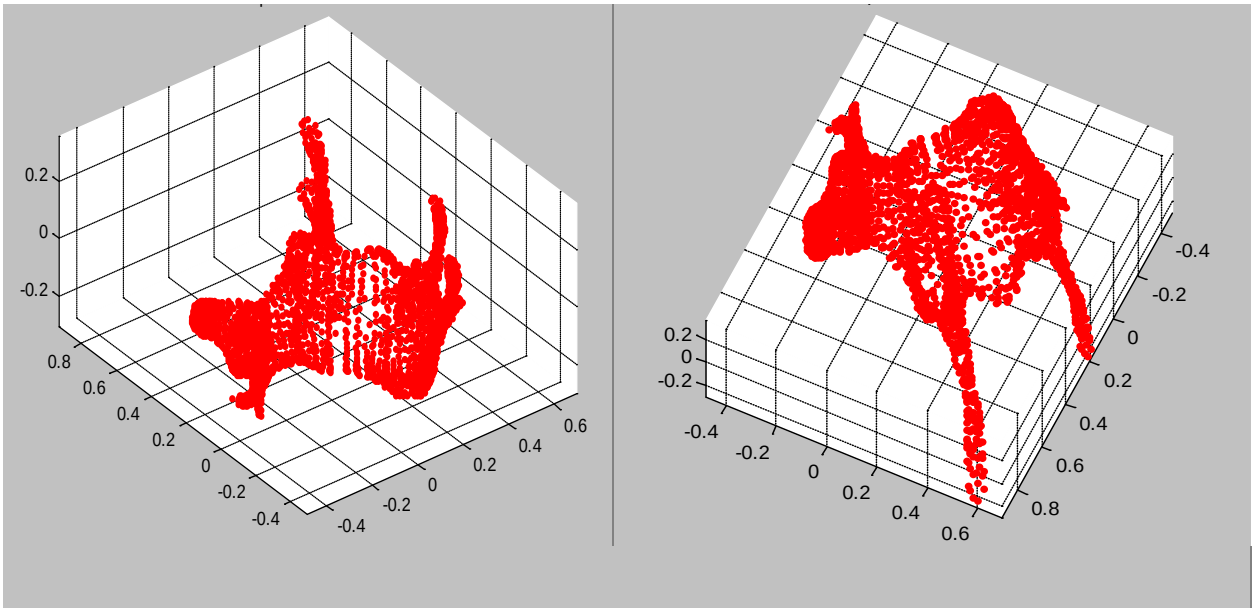

Fig. 5: Two rotated versions of 3-D structure.

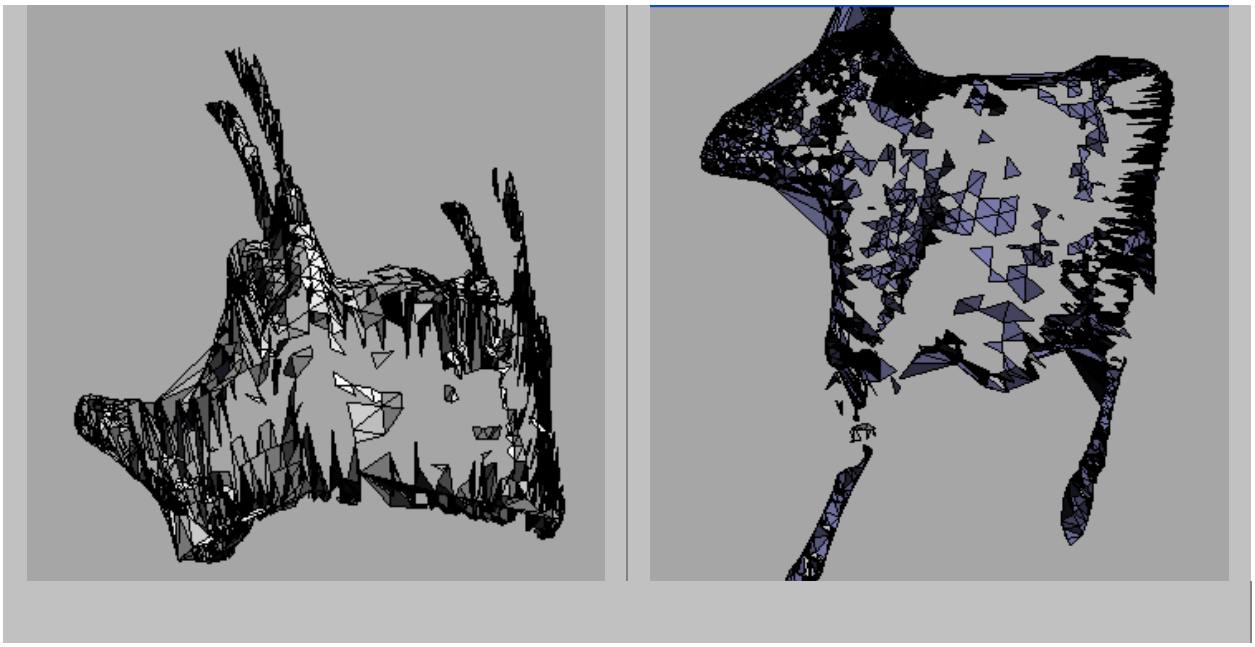

Fig. 6: Two rotated versions of 3-D structure after triangulation and rendering.

Experiment 3: Dataset: dino

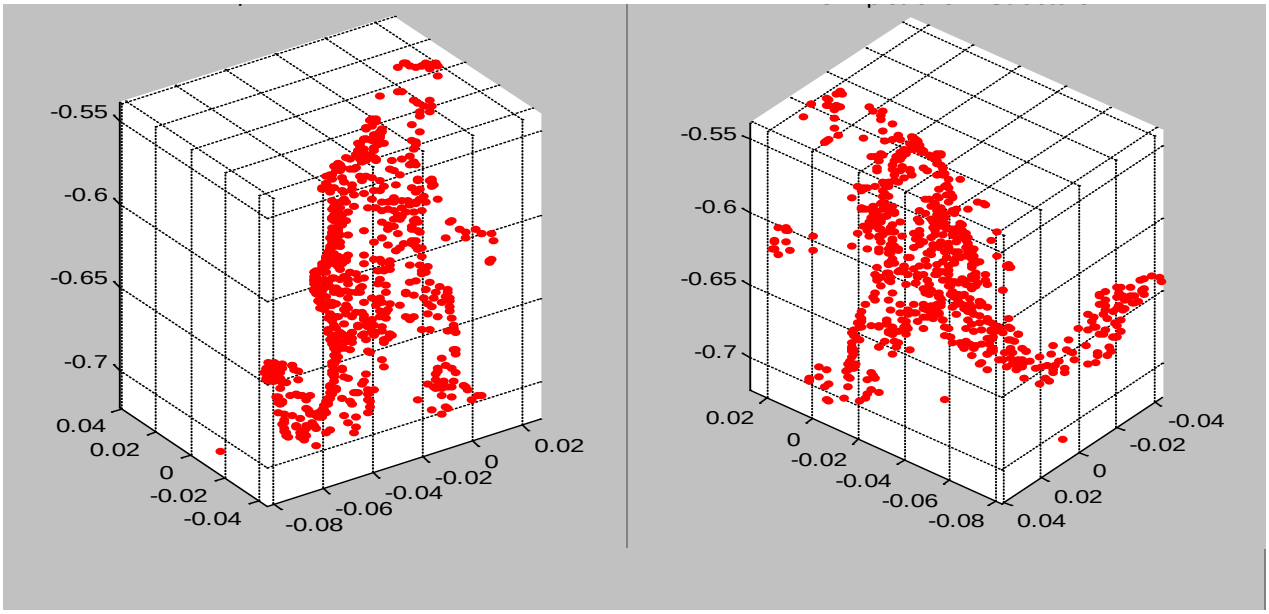

Fig. 7: Two rotated versions of 3-D structure.

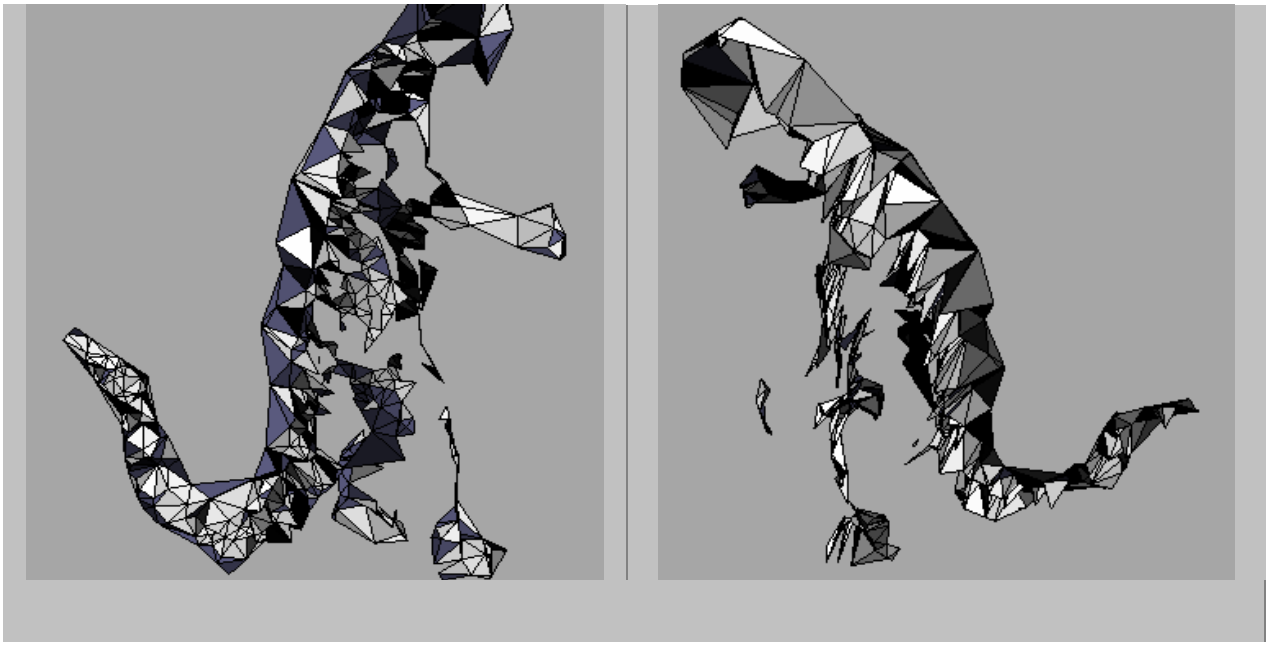

Fig. 8: Two rotated versions of 3-D structure after triangulation and rendering.

4. Conclusion

The Tomasi-Kanade factorization method has traditionally been used for the reconstruction of an objects' 3-D structure. However, in this paper, 3-D visualization or reconstruction from motion problem is solved using the factorization method in the orthographic projection. The robustness of feature correspondence essentially determines the success of the reconstruction algorithms. Especially, in the case of a long sequence, how to track features over a long time both stably and efficiently remains an unresolved challenge in the literature. This work employs the famous SIFT tracker for feature correspondence, whose robustness has already been proved by other researchers. This research further uses RANSAC to eliminate noise by discarding the false matches detected by the SIFT tracker. Finally, Tomasi-Kanade factorization is used to factor the noise-free measurement matrix into two matrices which correspond to the object's 3-D structure and camera rotation.

Acknowledgment

We wish to gratefully acknowledge *King Fahd University of Petroleum & Minerals Dhahran, Saudi Arabia,* for providing the funds to undertake this work. We further thank Ajmal S. Mian of the University of Western Australia, for making available to us the code to convert a point-cloud to a mesh of triangles.

References

[1] R. Ullah, S.H. Abdul-Jauwad, and K.M. Yahya "Structure from Motion using Tomasi-Kanade Factorization, SIFT and RANSAC" EGU General Assembly Conference 2012, Vol. 14, EGU2012-700-3.

[2] R. Y. Tsai and T. S. Huang, "Estimating 3-d motion parameters of a rigid planar patch", *IEEE Transaction. on Acoustic, Speech and Signal Processing*, Vol. 29, pp. 1147 - 1152, 1981.

[3] R. Tsai, T. Huang, and Wei-Le Zhu, "Estimating Three-dimensional motion parameters of a rigid planar patch, II: singular value decomposition", *IEEE Transaction. on Acoustic, Speech and Signal Processing*, Vol. 30, pp. 525-534, 1982.

[4] C. Tomasi, and T. Kanade, "Shape and motion from image streams: a factorization method", *International Journal of Computer Vision*, Vol. 9, pp.137-154, 1991.

[5] C. Poleman, and T Kanade, "A Paraperspective Factorization Method for Shape and Motion Recovery", *CMU Techincal Report*, CMU-CS-93-219, 1993.

[6] T. J. Broida, S. Chandrashekhar, and R. Chellappa, "Recursive estimation of 3D motion from a monocular image sequence", *IEEE Transactions on Aerospace and. Electronuic Systens.*, Vol. 26, pp. 639-656, 1990.

[7] T. J. Broida, and Rama Chellappa, "Estimating the kinematics and structure of a rigid object from a sequence of monocular images", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 13, pp. 497-513, 1991.

[8] C. Tomasi, and T. Kanade, "Detection and tracking of point features", *CMU Technical Report*, CMU-CS-91-132, 1991.

Samir H. Abdul-Jauwad holds a PhD (1985) in Electrical Engineering from the University of Sheffield, England, U.K. He got his Bachelor (1973) and Master degree (1976) in Electrical Engineering from King Fahd University of Petroleum & Minerals (KFUPM), Dhahran, Saudi Arabia. He is currently a Professor of Electrical Engineering at King Fahd University of Petroleum & Minerals (KFUPM), where his teaching and research interests include Signal and Image Compression, Image Processing, Wavelets, Satellite Communications. He is a member of KFUPM Scientific Council, and also a Senior Member of IEEE.

Rehmat Ullah graduated from University of Engineering and Technology Peshawar, Pakistan with a BSc (2004) and MSc (2008) degrees in Computer Systems Engineering. He also obtained his second MSc (2011) from Lahore University of Management Sciences, Lahore, Pakistan, in Computer Engineering with concentration in Image and Video Coding. He is currently a lecturer in Department of Computer Systems Engineering at University of Engineering and Technology, Peshawar, Pakistan. His research interests include Image and Video Coding, Image Analysis, and Computer Vision.

Farman Ullah did his BSc (2006) in Computer Systems Engineering from University of Engineering and Technology, Peshawar, Pakistan and his MSc (2011) in Computer Engineering from Center for Advanced Studies in Engineering, Islamabad, Pakistan. He is pursuing his PhD from Korea Aerospace University, South Korea. He worked as Telemetry Engineer at Advanced Engineering Research Organization, Pakistan for four years. Currently he is working as a lecturer in Department of Electrical Engineering at COMSATS Institute of Information Technology, Attock, Pakistan.