

Complete Visual Metrology using Relative Affine Structure

Adersh Miglani
Adersh.Miglani@gmail.com

Sumantra Dutta Roy
sumantra@ee.iitd.ac.in

Santanu Chaudhury
santanuc@ee.iitd.ac.in

J B Srivastava
jbsrivas@gmail.com

Abstract—We propose a framework for retrieving metric information for repeated objects from single perspective image. Relative affine structure, which is an invariant, is directly proportional to the Euclidean distance of a three dimensional point from a reference plane. The proposed method is based on this fundamental concept. The first object undergoes 4×4 transformation and results in a repeated object. We represent this transformation in terms of three relative affine structures along X , Y and Z axes. Additionally, we propose the possible extension of this framework for motion analysis - structure from motion and motion segmentation.

I. INTRODUCTION

In computer vision, invariants are widely used for recognition and classification of objects and three dimensional reconstruction of a scene from one or more uncalibrated images [1][2][3][4]. Broadly interpreted, all these vision tasks use invariants for retrieving geometric properties of objects from images. Here, we use a view-point invariant for retrieving metric measurements of multiple objects with translational and affine repetition from single perspective image. No prior knowledge of camera's internal and external parameters is required in this setup. The camera model used in this work is the central projection.

Repetition of two and three dimensional objects are frequently used in multiple vision tasks. The relative affine structure is one of the widely used tools in the context of repeated objects [4]. This is a projective invariant for repeated objects and turns into an affine invariant under special cases such as parallel projection. Prior work used relative affine structure for reconstruction and recognition of three dimensional objects, retrieving structure from motion and synthesizing new views from multiple prespective images [2][3][5]. Here, we explored its fundamental property that it is directly proportional to the Euclidean distance of a point from the reference plane. This led us to create another visual metrology framework beyond traditional usages of relative affine structure.

In this paper, we blend and develop previous results on relative affine structure and single view metrology [4][6][7][8][9]. Earlier, relative affine structure was used for retrieving three dimensional projective structure of one object from multiple images taken from different view points [4][2]. Later, it was used for projective reconstruction of repeated objects with translational and affine repetition from single perspective image [3]. And, homology, a plane projective transformation, was used for vanishing points based visual metrology techniques and camera calibration [6][7][10][11][12]. Here,

we use relative affine structure and homology, together in a different manner, for computing metric measurements of repeated objects with minimal scene information and reference metric measurements.

In single view metrology, we can compute the distance between two parallel planes when the corresponding points on the planes are along the direction normal to the planes [7]. This is a requirement to establish homology between two parallel planes [10]. But, our framework considers the general configuration of corresponding points on the repeated objects and is not restricted to such point correspondences.

Given reference measurements of first object, measurements of repeated objects, irrespective of translation and affine transformation, are computed. In case of affine repetition, computed measurements are upto a uniform scale along the reference direction that is normal to reference plane. Furthermore, proposed framework can be extended for vision tasks involving multiple views such as motion analysis - structure from motion, motion segmentation and tracking.

Section II describes relative affine structure and its properties suited for the proposed framework. The method to compute measurements of repeated objects, individually, without using relative affine structures for corresponding points is developed in section III. Section IV has details on retrieving metric information for affinely repeated objects up to uniform scale along X , Y and Z directions, respectively. Section V describes relationships between different object transformation matrices and relative affine structures. The results and conclusion with future work are discussed in sections VI and VII, respectively.

II. OVERVIEW

A. Relative Affine Structure

In three dimensional space, a plane π and a point $M_1 \notin \pi$ are chosen. Given two views ψ and ψ' with projection centers O and O' , H_π is a homography that transfers image point m to m' , where m and m' are image of point $M \in \pi$ [10]. Image points of M_1 in two views ψ and ψ' are m_1 and m'_1 . These are related by the following relationship which is derived in [2]:

$$m'_1 \cong H_\pi m_1 + ke' \quad (1)$$

where, e' is epipole and k is relative affine structure. Above relation has a scale factor which can be resolved by appropri-

ately scaling H_π or e' such that [2],

$$m'_0 \cong H_\pi m_0 + e' \quad (2)$$

where m'_0 and m_0 are images of a fixed point $M_0 \notin \pi$. This configuration is shown in figure (1). Geometrically, the relative affine structure is defined as [2],

$$k = \frac{X_1 \lambda_0}{\lambda_1 X_0} \quad (3)$$

where λ_1 and λ_0 are depth of point M_1 and M_0 and X_1 and X_0 are perpendicular distance of these points from reference plane π which is YZ plane in this case. Thus, it is concluded that:

The relative affine structure is proportional to the perpendicular distance from a reference plane.

The scale factor for Eq. (1) is computed in [3]. It is ratio of depths of M_1 with respect to projection centers O and O' .

$$\lambda' m'_1 = \lambda(H_\pi m_1 + ke'_N) \quad (4)$$

where e'_N is normalized epipole. After simplifying it further, the expression for k is written as [3],

$$k = \frac{(m'_1 \times e'_N)^T ((H_\pi m_1) \times m'_1)}{\|(m'_1 \times e'_N)\|^2} \quad (5)$$

Let us consider the dual configuration of what is shown in figure (1). An object S undergoes a transformation T , affine or translation, and results in object S' . In three dimensional space, a point $M \in S$ is related by its corresponding point $M' \in S'$ such as $M' = TM$. The image of M and M' in view ψ with projection center O are m and m' . This configuration can be considered as single image of two repeated objects with transformation T or two different images of single object when two cameras undergo same transformation T . This is called *isometry property*. The relative affine structures for corresponding points m and m' are k and k' . The expression for k is given by Eq. (5). The expression for k' under translational and affine repetition is given by Eq. (6) and (7), respectively.

$$k' = \frac{(m \times e'_N)^T ((H_\pi m' - 2m') \times m)}{\|(m \times e'_N)\|^2} \quad (6)$$

$$k' = \frac{(H_\infty m \times e'_N)^T ((e'_N \nu_{\pi N}^T m' - m') \times m)}{\|(H_\infty m \times e'_N)\|^2} \quad (7)$$

where $\nu_{\pi N}^T = e'^T_N (H_\pi - H_\infty)$ and H_∞ is infinite homography between two views [3].

B. Apparatus for Proposed Framework

The chosen world coordinate system for repeated objects is shown in figure (2). The three orthogonal planes π_{YZ}, π_{ZX} and π_{XY} are *reference planes*. For every point $M_1 \notin \{\pi_{YZ}, \pi_{ZX}, \pi_{XY}\}$ will have three relative affine structures, k_x, k_y and k_z , respectively.

$$k_x = X_1 \frac{1}{\lambda} \frac{\lambda_0}{X_0}, \quad k_y = Y_1 \frac{1}{\lambda} \frac{\lambda_0}{X_0}, \quad \text{and} \quad k_z = Z_1 \frac{1}{\lambda} \frac{\lambda_0}{X_0} \quad (8)$$

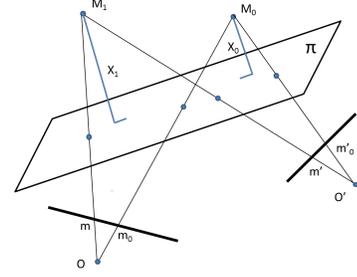


Fig. 1. Geometry: Relative Affine Structure

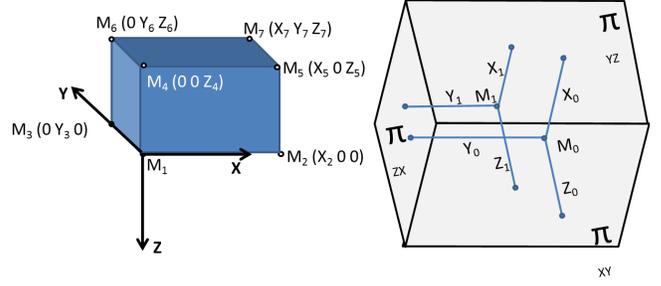


Fig. 2. Geometry: Three Orthogonal Relative Affine Structures

where, X_0 and λ_0 are X coordinate and depth of a fixed point $M_0 \notin \{\pi_{XY}, \pi_{YZ}, \pi_{ZX}\}$. The ratio $\frac{\lambda_0}{X_0}$ is a constant and denoted by μ_{x0} . Similarly, we can write,

$$k_x = X_1 \frac{1}{\lambda} \mu_{x0}, \quad k_y = Y_1 \frac{1}{\lambda} \mu_{y0}, \quad \text{and} \quad k_z = Z_1 \frac{1}{\lambda} \mu_{z0} \quad (9)$$

Therefore, each expression for k is proportional to the perpendicular distance from the chosen reference plane, e.g. $k_x \propto X_1$, $k_y \propto Y_1$ and $k_z \propto Z_1$. There are two constants of proportionality. First constant is inverse of depth of the point $\frac{1}{\lambda}$ which will vary for every point. Second constant (say μ_{x0}) is fixed for every point with respect to the reference plane (say π_{YZ}).

The proposed framework use this fundamental concept behind relative affine structure to determine 3D measurements of translationally and affinely repeated objects. We experiment our framework on a perspective image of repeated cuboids, as shown in figure (3). Any one cuboid is considered as principal object and rest as auxiliary objects.

III. MEASUREMENTS OF INDIVIDUAL OBJECT

As described in section II-B, for a point M_i , a relative affine structure k_x has a fixed constant μ_{x0} and a variable constant $\frac{1}{\lambda_i}$. The fixed constant will be eliminated by taking ratio of two relative affine structures for two different points with respect to same reference plane, π_{YZ} .

Considering points $M_2 = (X_2 \ 0 \ 0)$ and $M_5 = (X_5 \ 0 \ Z_5)$ as shown in figure (2), ratio of their relative affine structures is reduced to the following expression,

$$\frac{k_{x2}}{k_{x5}} = \frac{X_2 \lambda_5 \lambda_0 X_0}{X_5 \lambda_2 X_0 \lambda_0} = \frac{X_2 \lambda_5}{X_5 \lambda_2} \quad (10)$$

The values of k_{x2} and k_{x5} are computed by Eq. (5). The expression for X_5 can be written as

$$X_5 = X_2 \frac{k_{x5} \lambda_5}{k_{x2} \lambda_2} \quad (11)$$

The depth of points M_2 and M_5 , λ_2 and λ_5 , are computed using vanishing points based method given in [6]. Given metric value of X_2 (X_{ref}) and metric value of X_5 is computed. Thus, for any arbitrary point's X coordinate can be computed using relative affine structure and projective depth. Similarly, we can compute the Y and Z coordinates of every point, given reference metric measurements Y_{ref} and Z_{ref} along Y and Z direction, respectively.

$$Y = Y_{ref} \frac{k_y}{k_{yref}} \frac{\lambda}{\lambda_{ref}} \quad \text{and} \quad Z = Z_{ref} \frac{k_z}{k_{zref}} \frac{\lambda}{\lambda_{ref}} \quad (12)$$

IV. PRINCIPAL OBJECT AS REFERENCE

Consider a pair of corresponding points M_i and M'_i on affinely repeated objects S and S' . Alternately, we can arbitrarily choose a pair of points on two repeated objects. The respective relative affine structures k_{xi} and k'_{xi} for m_i and m'_i can be computed by Eq. (5) and Eq. (7). The ratio of k_{xi} and k'_{xi} can be expressed as follows

$$\frac{k_{xi}}{k'_{xi}} = \frac{X_i \lambda'_i \lambda_{i0} X'_{i0}}{X'_i \lambda_i \lambda'_{i0} X_{i0}} \quad (13)$$

$$\frac{k_{xi}}{k'_{xi}} = \frac{X_i \lambda'_i}{X'_i \lambda_i} \psi_{x0} \cong \frac{X_i \lambda'_i}{X'_i \lambda_i} \quad (14)$$

Since $\psi_{x0} = \frac{\lambda_{i0} X'_{i0}}{\lambda'_{i0} X_{i0}}$ is fixed for all points, the ratio $\frac{k_{xi}}{k'_{xi}}$ can be computed up to uniform scale along X -axis. The ratio $\frac{\lambda'_i}{\lambda_i}$ can be computed by solving equation (4).

$$\frac{\lambda'_i}{\lambda_i} = \frac{\|((H_\pi m_i) \times e'_N)\|}{\|m'_i \times m_i\|} \quad (15)$$

Equation (14) can be written as

$$X'_i \cong X_i \frac{k'_{xi} \lambda'_i}{k_{xi} \lambda_i} = X_i \alpha_{xi} \quad (16)$$

Similarly, we can write expressions for Y and Z directions as below,

$$Y'_i \cong Y_i \frac{k'_{yi} \lambda'_i}{k_{yi} \lambda_i} = Y_i \alpha_{yi}, \quad Z'_i \cong Z_i \frac{k'_{zi} \lambda'_i}{k_{zi} \lambda_i} = Z_i \alpha_{zi} \quad (17)$$

Given reference measurements on the principal object along X , Y and Z axes, measurements of affinely repeated object can be computed up to a respective scale by Eq. (16) and (17).

V. TRANSFORMATION OF REPEATED OBJECT

Under specific configurations, relative affine structure, which is projective structure, turns into affine structure. If the reference plane is at infinity or in case of parallel projection, relative affine structure approaches to affine structure [2]. Ratio of two relative affine structures of a point with respect to different reference planes does not depends on the depth. Thus, that ratio is a projective structure. We have seen that relative affine structure is proportional to Euclidean distance of a point

from the reference plane. Therefore, relative affine structure subsumes projective, affine and Euclidean structures [2]. Here, this statement is analyzed mathematically.

Suppose S and S' are repeated objects and are related by $S' = T(S)$, where T is a 4×4 general transformation matrix. A point $M' \in S'$ is corresponding to $M \in S$. By using Eq. (16) and (17), the relation between corresponding points can be written as follows,

$$\begin{pmatrix} X' \\ Y' \\ Z' \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_x & 0 & 0 & 0 \\ 0 & \alpha_y & 0 & 0 \\ 0 & 0 & \alpha_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \mathcal{K} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (18)$$

where \mathcal{K} is the transformation matrix in 3D Euclidean space.

A. Translational Repetition

If object S and S' are related by pure translation, the transformation is represented as,

$$\begin{pmatrix} X' \\ Y' \\ Z' \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & T_x \\ 0 & 1 & 0 & T_y \\ 0 & 0 & 1 & T_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (19)$$

From Eq. (18) and (19),

$$X' = X + T_x = X \alpha_x \Rightarrow T_x = X(\alpha_x - 1) \quad (20)$$

Similarly,

$$T_y = Y(\alpha_y - 1), \quad T_z = Z(\alpha_z - 1) \quad (21)$$

Therefore, translation vector can be represented in terms of three relative affine structures with respect to uniform scale along X , Y and Z axes, respectively.

B. Pure Rotational Repetition

If object S undergoes pure rotation and results in S' , the transformation is represented as,

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = \begin{pmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} \alpha_x & 0 & 0 \\ 0 & \alpha_y & 0 \\ 0 & 0 & \alpha_z \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

Therefore, the rotation between two points is equivalent to the following diagonal matrix

$$R = \begin{pmatrix} \alpha_x & 0 & 0 \\ 0 & \alpha_y & 0 \\ 0 & 0 & \alpha_z \end{pmatrix}$$

where three columns are scaled uniformly along X , Y and Z directions, respectively.

C. Affine Repetition

The most general case is affine repetition that encapsulates rotation, translation, scaling and shearing [10]. This transformation in 3D Euclidean space is equivalent to \mathcal{K} . Once constants ψ_{x0} , ψ_{y0} and ψ_{z0} along X , Y and Z directions are computed from image (coordinates of M_0), affine repetition can be computed uniquely.

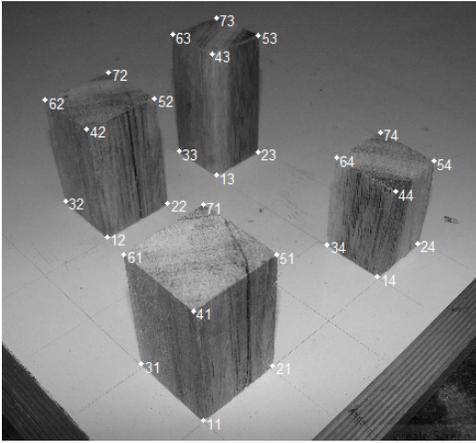


Fig. 3. Repeated Objects

VI. RESULTS

In our experiments, we consider a real image, as shown in figure (3). It has objects with affine repetition. Table I and II display the measurements (centimeter) of objects computed using methods discussed in sections III and IV, respectively. Based on the precision required for an application, the errors can be further reduced by employing efficient techniques for computing point correspondences and vanishing points from image. Additionally, proper uncertainty analysis will also improve the results [6].

TABLE I
MEASUREMENTS - INDIVIDUAL OBJECT

| ID | Source | X5 | Z5 | Y6 | Z6 | X7 | Y7 | Z7 |
|----|--------|------|------|------|-----|------|------|-----|
| 1 | Scene | 3.1 | 5.1 | 3.1 | 5.1 | 3.1 | 3.1 | 5.1 |
| | Image | 3.3 | 5.0 | 3.3 | 4.8 | 3.4 | 3.5 | 4.6 |
| | Error | -0.2 | 0.1 | -0.2 | 0.3 | -0.3 | -0.4 | 0.5 |
| 2 | Scene | 3.1 | 5.1 | 3.1 | 5.1 | 3.1 | 3.1 | 5.1 |
| | Image | 2.3 | 5.3 | 2.1 | 4.2 | 2.1 | 2.1 | 4.3 |
| | Error | 0.8 | -0.2 | 1.0 | 0.9 | 1.0 | 1.0 | 0.8 |
| 3 | Scene | 2.6 | 6.0 | 2.6 | 6.0 | 2.6 | 2.6 | 6.0 |
| | Image | 2.4 | 6.1 | 1.5 | 5.7 | 2.5 | 1.3 | 5.8 |
| | Error | 0.2 | -0.1 | 1.1 | 0.3 | 0.2 | 1.3 | 0.2 |

TABLE II
MEASUREMENTS - PRINCIPAL (I_{st}) OBJECT AS REFERENCE

| ID | Source | X5 | Z5 | Y6 | Z6 | X7 | Y7 | Z7 |
|----|--------|-----|------|------|-----|-----|------|-----|
| 2 | Scene | 3.1 | 5.1 | 3.1 | 5.1 | 3.1 | 3.1 | 5.1 |
| | Image | 2.3 | 5.3 | 3.5 | 4.2 | 2.1 | 3.5 | 4.3 |
| | Error | 0.8 | -0.2 | -0.4 | 0.9 | 1.0 | -0.4 | 0.8 |
| 3 | Scene | 2.6 | 6.0 | 2.6 | 6.0 | 2.6 | 2.6 | 6.0 |
| | Image | 2.4 | 5.9 | 3.2 | 5.2 | 2.6 | 2.9 | 5.1 |
| | Error | 0.2 | -0.1 | -0.8 | 0.8 | 0.0 | -0.3 | 0.9 |

VII. CONCLUSION

We extended prior work on relative affine structure for computing three dimensional measurement from a single perspective image of repeated objects. The transformation

between repeated objects can be represented in terms of relative affine structures along three orthogonal directions. Therefore, one invariant is used to analyze projective, affine and Euclidean space for vision tasks. Camera transformation for repeated object can also be expressed in terms of relative affine structures.

Furthermore, three dimensional motion of an object or a camera can be parameterized in terms of relative affine structure. So, motion analysis related tasks such as motion segmentation and tracking can use relative affine structure, an invariant.

ACKNOWLEDGMENT

This work has been supported under the research grant, towards the setting up of the ‘*Program on Autonomous Robotics*’, from the *Board of Research in Nuclear Sciences (BRNS), India*.

REFERENCES

- [1] Y. Cao and J. McDonald, “Viewpoint invariant features from single images using 3d geometry,” in *Applications of Computer Vision (WACV), 2009 Workshop on*, 2009, pp. 1–6.
- [2] A. Shashua and N. Navab, “Relative affine structure: canonical model for 3d from 2d geometry and applications,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, no. 9, pp. 873–883, 1996.
- [3] R. Choudhury, J. B. Srivastava, and S. Chaudhury, “Reconstruction-based recognition of scenes with translationally repeated quadrics,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 6, pp. 617–632, 2001.
- [4] A. Shashua and N. Navab, “Relative affine structure: theory and application to 3d reconstruction from perspective views,” in *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR ’94., 1994 IEEE Computer Society Conference on*, 1994, pp. 483–489.
- [5] S. Tebaldini, M. Marcon, A. Sarti, and S. Tubaro, “Uncalibrated view synthesis from relative affine structure based on planes parallelism,” in *Proceedings of the International Conference on Image Processing, ICIP 2008, October 12-15, 2008, San Diego, California, USA.* IEEE, 2008, pp. 317–320.
- [6] A. Criminisi, “Accurate visual metrology from single and multiple uncalibrated images,” Ph.D. dissertation, University of Oxford, Dept. Engineering Science, 1999, d.Phil. thesis.
- [7] A. Criminisi, I. Reid, and A. Zisserman, “Single view metrology,” *Int. J. Comput. Vision*, vol. 40, no. 2, pp. 123–148, Nov. 2000.
- [8] K. Peng, L. Hou, R. Ren, X. Ying, and H. Zha, “Single view metrology along orthogonal directions,” in *Proceedings of the 2010 20th International Conference on Pattern Recognition*, ser. ICPR ’10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 1658–1661. [Online]. Available: <http://dx.doi.org/10.1109/ICPR.2010.410>
- [9] G. Wang, Z. Hu, F. Wu, and H.-T. Tsui, “Single view metrology from scene constraints,” *Image Vision Comput.*, vol. 23, no. 9, pp. 831–840, Sep. 2005. [Online]. Available: <http://dx.doi.org/10.1016/j.imavis.2005.04.002>
- [10] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.
- [11] R. Cipolla, T. Drummond, and D. Robertson, “Camera Calibration from Vanishing Points in Images of Architectural Scenes,” in *Proc. British Machine Vision Conference*. Nottingham, UK: British Machine Vision Association, 1999, pp. 382–391.
- [12] E. Guillou, D. Meneveaux, E. Maisel, and K. Bouatouch, “Using vanishing points for camera calibration and coarse 3d reconstruction from a single image,” *The Visual Computer*, vol. 16, no. 7, pp. 396–410, 2000. [Online]. Available: <http://dx.doi.org/10.1007/PL00013394>