Invariant Based Recognition of Scenes with Multiple Translationally Repeated Components

Ragini Choudhury and Santanu Chaudhury
Dept. of Electrical Engineering
IIT Delhi
Hauz Khas, New Delhi 110016
India
ragini@cse.iitd.ernet.in
santanuc@cse.iitd.ernet.in

J. B. Srivastava
Dept. of Mathematics
IIT Delhi
Hauz Khas, New Delhi 110016
India
jbsrivas@maths.iitd.ernet.in

Abstract:
This paper proposes an invariant based recognition scheme for scenes with multiple translationally repeated components, using a single view of such a scene. In order to do this, three component subsets are identified which characterize the scene completely. We have developed a mathematical framework, for the projective reconstruction, based on relative affine structure, of each such three component building block, to serve as an input to the recognition strategy. This is extended to the case when each of the components is a quadric. A set of projective invariants of three quadrics have also been obtained by us. These can be computed for each three quadric component subset and their values used for recognition purposes. As a recognition strategy for scenes with multiple translationally repeated quadric components, we propose to compute and store invariant values for each such three component subset. Experiments on real data have shown the applicability of this approach for recognizing a scene with multiple quadric repetitions. The discriminatory power and stability of these invariants have also been established.

1 Introduction
It is difficult to recognize images of man-made structures like power plants, factories and historical monuments using only image based features because the images of these scenes can differ substantially due to changes in viewing positions and consequent perspective distortions. This problem can be circumvented by exploiting invariants based on 3D structure of such scenes. In this paper, we present a recognition scheme for scenes with multiple translationally repeated components, using a single view of the scene. This is achieved via a reconstruction framework and use of projective invariants. We have characterized such scenes by considering three components at a time. We give a mathematical framework for the projective reconstruction of three times translationally repeated objects. This is then extended to the specific case of quadrics. Quadrics are 3D shapes like ellipsoids, spheres and hyperboloids. A set of joint projective invariants are computed with these reconstructed quadrics, and their values are used for the recognition of such scenes.

Work on reconstruction of repeated objects has concentrated on objects which repeat two times [6, 7]. Recognition issues have been limited to translationally repeated objects [4]. In case of quadrics, only reconstruction has been done [10, 2]. Recognition issues have not been dealt with in these papers.

The most significant contribution of our work is the reconstruction based recognition strategy. The three element repetition based reconstruction strategy is general and utilizes
the trinocular stereo framework. The use of appropriate invariants can characterize such clusters in a multi-component scene. In this paper, although we have only used invariants of quadrics, our approach can be applied with invariants computed using other repeated objects. We have considered subsets of three because these can characterize multiple repetition comprehensively, compared to pairs which may lead to large number of constraints for the multiply repeated scenario.

2 Reconstruction scheme for a three component subset

In a scene with multiple translationally repeated components, we pick sets, of three components which serve as building blocks to characterize the object. If we can reconstruct each of these three component subsets, then they would serve as inputs to our reconstruction based recognition strategy for scenes with multiple repetition. In this section we give a framework for the projective (relative affine) reconstruction of a three component building block and then take the specific case when each component is a quadric.

Each building block consists of three components \( S, S', S'' \) where \( S' = T_{13}(S) \) and \( S'' = T_{12}(S) \) and \( T_{13} = \begin{pmatrix} I_3 & t_{12} \\ 0 & 1 \end{pmatrix} \) and \( T_{12} = \begin{pmatrix} I_3 & t_{13} \\ 0 & 1 \end{pmatrix} \) for some \( t_{12} = (a \ b \ c)' \) and \( t_{13} = (a' \ b' \ c')' \) are the translations. The problem is to reconstruct such an object from a single image of it. This is best handled by converting it to the equivalent trinocular stereo framework.

Let the three cameras in the trinocular stereo setup be \( \hat{P}_i = [P_i \; p_i] \), \( \det(P_i) \neq 0, i = 1, 2, 3 \). Here \( COP_{k} = \begin{pmatrix} -P_{k-1}^{-1}p_k \\ 1 \end{pmatrix} \) for \( k = 1, 2, 3 \). Let \( M = (X \ Y \ Z \ 1)' \) represent an arbitrary point in 3D. Define \( [P_1 \; p_1]M = \tilde{m} = \lambda \tilde{m}, \quad m = (x \ y \ 1)' \; [P_2 \; p_2]M = \tilde{m}' = \lambda \tilde{m}', \quad m' = (x' \ y' \ 1)' \) and \( [P_3 \; p_3]M = \tilde{m}'' = \lambda \tilde{m}'' \), \( m'' = (x'' \ y'' \ 1)' \). The epipoles are defined by \( \tilde{e}_{ij} = [P_1 \; p_i] \begin{pmatrix} -P_{j-1}^{-1}p_j \\ 1 \end{pmatrix} = \lambda e_{ij}, e_{ij} = (e_{ij}^1 \ e_{ij}^2 \ e_{ij}^3)' \) for \( i \neq j, i, j = 1, 2, 3 \). Let \( \Pi \) be a plane which does not contain \( COP_i \) and \( COP_j \) and \( H_{\Pi} \) be the homography with respect to the plane \( \Pi \) \([5]\). When \( \Pi = \Pi_{\infty} \) is the plane at infinity, the homography will be denoted by \( H_{\infty_{ij}} \). It is easily seen \([5]\) that \( H_{\infty_{ij}} = P_j P_i^{-1} \).

Finally we define the projective reconstruction matrix as \( \mathcal{H} = \begin{pmatrix} P_1 & p_1 \\ L_N & \nu_N \end{pmatrix} \), where \( L_N = -\nu_{12N}P_1, \quad \nu_N = -\nu_{12N} + \lambda_{21}[\nu_1], \quad H_{12N}^I = \epsilon_{21N}(H_{12} - H_{\infty_{12}}, \quad \epsilon_{jjN} = \epsilon_{jN}^2. \) By using the above properties and definitions, the following camera transformations hold \((i)[P_1 \; p_1] = [I_3 \; 0]H, (ii)[P_2 \; p_2] = [H_{12} \; \epsilon_{21H}H, (iii)[P_3 \; p_3] = [A \; b]H, \) where \( A = H_{\infty_{12}} + \alpha_1\tilde{c}_3 \tilde{N}_{12N}, \quad b = \alpha_1 \tilde{c}_3 \tilde{N}, \quad \alpha_1 = \frac{\lambda_{21}\tilde{c}_3}{\tilde{c}_2 \tilde{N}_2} = \frac{(\epsilon_{21H}A_{12N})\tilde{c}_2 \tilde{N}_2}{\|\tilde{c}_2 \tilde{N}_2\|^2}. \)

In the next theorem (proved in Appendix), we obtain a projective reconstruction of 3D configurations containing three times translationally repeated objects, based on relative affine structure \([5, 9]\). In this, all the three components are reconstructed with respect to the same frame, which was not possible by a direct application of any method available. The reconstruction will be extended to the case when each of the multiple components is a quadric. The \( k, k' \) and \( k'' \) computed in the next theorem characterize the relative affine structure.

**Theorem 2.1 [Projective Reconstruction]**

Let \( M_i, \quad i = 1, 2, 3, 4 \) be four points in general position on the component \( S \) and let \( M'_i, M''_i, \quad i = 1, 2, 3, 4 \) be the corresponding points on the translated components \( S' \) and \( S'' \) respectively. A single uncalibrated perspective image of a scene containing \( S, S' \) and \( S'' \) is given. Also given are the image point correspondences \( m_i, \quad m'_i, \quad m''_i, \quad i = 1, 2, 3, 4 \) of the above points. If \( m = (x \ y \ 1)' \), \( m' = (x' \ y' \ 1)' \) and \( m'' = (x'' \ y'' \ 1)' \) is a point correspondence of a 3D point \( M \in S \) and the corresponding translated points \( M' \in S' \) and \( M'' \in S'' \), then there exists a projective transformation \( \mathcal{H} \) such that the reconstructed points are given by \( \mathcal{H}M \approx (x \ y \ 1)' \), \( \mathcal{H}M' \approx (x' \ y' \ 1)' \) and \( \mathcal{H}M'' \approx (x'' \ y'' \ 1)' \) respect-
where \( k = \frac{(m' \times r_2) \cdot (H_{23} - m \cdot m')}{\| (m' \times r_2) \|} \), \( k' = \frac{(m' \times r_2) \cdot (H_{23} - m' \cdot m')}{\| (m' \times r_2) \|} \), and \( k'' = \frac{(m' \times r_2) \cdot (H_{23} - m'' \cdot m')}{\| (m' \times r_2) \|} \). Here \( \alpha_1 = \frac{(r_2 \times m'' \cdot (H_{23} - m'' \cdot m'))}{\| (r_2 \times m'' \cdot (H_{23} - m'' \cdot m')) \|} \) and the reference plane \( \Pi = \langle M_1, M_2, M_3 \rangle \).

Now we consider the specific case when each of the components \( S, S' \) and \( S'' \) are quadrics \( Q, Q' \) and \( Q'' \) respectively where \( Q' = T_1 Q \) and \( Q'' = T_2 Q \). A quadric surface is an algebraic polynomial of degree 2 in four variables \( x_1, x_2, x_3, x_4 \) over the field \( \mathbb{R} \) of real numbers. General equation of the quadric surface is

\[
S = S(x_1, x_2, x_3, x_4) = \sum_{i=1}^{4} \sum_{j=1}^{4} q_{ij} x_i x_j = 0
\]

i.e., \( S(X) = X^T Q X = 0 \), where \( Q = (q_{ij}) \) is a 4 \( \times \) 4 real symmetric matrix and \( X = (x_1, x_2, x_3, x_4)^T \) is a homogeneous 4 \( \times \) 1 vector. The quadric surface defined above is said to be proper if \( \det(Q) \neq 0 \), i.e., \( Q \) is non-singular. It can be seen that the outline conic \( C \) of the original quadric \( Q \) is the same as the outline conic \( \tilde{C} \) of the reconstructed quadric \( \tilde{Q} = \mathcal{H}(Q) = \{ HM | M \in Q \} \).

Let \( Q = \begin{pmatrix} Q_{33} & q \\ q^t & q_{44} \end{pmatrix} \), \( Q_{33} \) a 3 \( \times \) 3 matrix, \( q \) a 3 \( \times \) 1 vector, \( q_{44} \) a scalar, be a quadric and let \( \tilde{Q} = \mathcal{H}(Q) = \begin{pmatrix} \tilde{Q}_{33} & \tilde{q} \\ \tilde{q}^t & \tilde{q}_{44} \end{pmatrix} \) be the quadric to be reconstructed. Then \( \tilde{C} = \tilde{Q}_{33} - \tilde{q}_{44} \tilde{Q}_{33} \).

In the next theorem (proved in Appendix), we reconstruct the three translationally repeated quadrics. For this we have modified the method of quadric reconstruction of \([10]\) so that all three quadrics are reconstructed with respect to the same frame.

**Theorem 2.2** [Reconstruction of \( \tilde{Q}, \tilde{Q}' \) and \( \tilde{Q}'' \): A single uncalibrated perspective image of a set of three proper quadrics \( Q, Q' \) and \( Q'' \) is given when \( Q' \) and \( Q'' \) are translational repetitions of \( Q \). Further image point correspondences of \( 4 \) non-coplanar points \( M_i, i = 1, 2, 3, 4 \), on \( Q \) and corresponding points \( M'_i, M''_i, i = 1, 2, 3, 4 \), on \( Q' \) and \( Q'' \) respectively are given. Also known in the image are corresponding outline conics. Then the quadrics \( Q, Q' \) and \( Q'' \) can be reconstructed projectively.

These reconstructed quadrics are used to compute the invariants for recognition purposes.

**3 Recognition of a three quadric component subset**

In this section, we compute the joint projective invariants of a set of three proper quadrics for the purpose of recognition of quadric configurations. Invariants are computed for each three component building block using its three component quadrics, which have been reconstructed. These invariant values are then used to identify the scene. The nature of the invariants computed is such that the repeated quadrics with different translations as well as different nature of quadrics, will be recognized as different configurations.

Let \( Q_1, Q_2 \) and \( Q_3 \) be three proper quadrics defined up to a scale by 4 \( \times \) 4 non-singular symmetric matrices \( A, B \) and \( C \) respectively. Let \( T \in PGL_4(\mathbb{R}) \) be any projective transformation. Define transformed quadrics \( \bar{Q}_1 = T(Q_1) = \{ TM | M = (X Y Z 1)^t \in Q_1 \} = \{ TM | M^t AM = 0 \}, \bar{Q}_2 = T(Q_2) = \{ TM | M^t BM = 0 \} \) and \( \bar{Q}_3 = T(Q_3) = \{ TM | M^t CM = 0 \} \).
By the Counting Argument [3], there are 12 independent projective invariants of the configuration space consisting of three quadrics $Q_1$, $Q_2$ and $Q_3$ determined by matrices $A, B$ and $C$ up to a scale. So if the projective invariants are computed with the help of matrices $A, B$ and $C$, then these should be independent of all scalar multiples and all $T \in PGL_4(\mathbb{R})$, which is shown in the following theorem, proved in the Appendix.

**Theorem 3.1:** If $Q_1$, $Q_2$ and $Q_3$ are three proper quadrics defined by symmetric non-singular matrices $A$, $B$ and $C$ respectively up to a scale, then

\[
\begin{align*}
I_1, I_2, I_3, I_4, I_5, I_6, I_7, I_8, I_9, I_{10}, I_{11}, I_{12} \end{align*}
\]

are 12 independent projective invariants of the configuration space consisting of three quadrics $Q_1$, $Q_2$ and $Q_3$ determined by matrices $A, B$ and $C$ up to a scale. So if the projective invariants are computed with the help of matrices $A, B$ and $C$, then these should be independent of all scalar multiples and all $T \in PGL_4(\mathbb{R})$, which is shown in the following theorem, proved in the Appendix.

The conics were fitted interactively by using the Bookstein’s algorithm [4]. The fitted conics for some objects is shown in Fig. 6. The corresponding points on the repeated objects were also picked interactively. The processing routines were developed in MATLAB.

**Discriminatory ability:**
In order to study this, the framework was first
applied to images (Fig. 1.2) of scenes which have only three quadric components repeated. By picking the images of the four distinguished points together with the outline conics, quadrics are reconstructed. These are used to compute invariants. In order to compare two sets of invariant values between two images, a distance measure is required. For all pairs of different images \( im_i, im_j \), we define the distance used by Long Quan [8], \( \Sigma^{15}_{k=1}(V_k[im_i] - V_k[im_j])^2 \) where \( V_k[im_i] \) is the value of the \( k \)th invariant of image \( i \). The distances (scaled) between set of invariant values of Fig. 1 and 2 are given in Table 1. It is seen that (scaled) distances between invariant values of projectively equivalent views (two views of the same scene) as in Fig. 1(a), 1(b) is much less than the distance between Fig. 1(a),2(b) which are not projectively equivalent.

The framework is then applied to real life objects as in an egg carton in Fig. 3, and a vessel in Fig. 4, in which multiple repetitions of quadrics occur. The distances between the invariants obtained from a three quadric building block are shown in Table 2. They confirm the discriminatory ability of the invariants. Finally we have the three images (Fig. 5) of the cooling towers of power plants. Our framework is applied to a subset of three towers. The (scaled) distance between the invariant values of Fig. 5(a),(b) is 1.3649 and that of Fig. {5(a), (c)} and {5(b), (c)} are 6.6033e4 and 6.6635e4 respectively. This supports the similarity in shape between towers of Fig. 5(a) and (b).

\textbf{Stability of Invariants :}

In order to study the stability of these invariants, we consider Fig.6 (a) - (e), which are five views of the same scene containing three quadric components. The values of the invariants are shown in the Table 3. The last row of the table gives the coefficient of variation (\( \gamma = (100 \times (\text{standard deviation/mean})) \)) of these invariant values. The low values of \( \gamma \) indicates stability.

\section{5 Conclusions}

In this paper we have proposed an invariant based recognition scheme for scenes which have multiple translationally repeated components, via a reconstruction framework. The information in such a scene can be characterized by a basic subset of three such translational repetitions. We propose a framework for the projective reconstruction of each such building block. When each of the components is a quadric, they can be recognized by using the set of projective invariants for three proper quadrics which have been computed by us. This is sufficient to recognize the scene with multiple repetitions.
Table 3 : Values of Invariants for Stability Analysis

References


6