Image Retrieval Based on P²-Invariant

Ruihang Wu Author1,[†] , Shaozi Lir Author2[††],and Fengmei Zou Author3[††]

Xiamen Univeersity, Department of Computer Science, Xiamen, Fujian, China

Summary

The geometrical description of an object can be separated into two parts: a) the registration information and b) the 'shape' (which is invariant under registration transformation). A common choice of registration is the group of Euclidean similarity transformations, and then the geometrical properties that are invariant under this group of transformations are known as 'similarity shape'. In this paper the class of projective/permutation p^2 -invariant based image retrieval algorithm is given. Computation of the invariants of points set in 2D projective space is given. The correspondence between two images is established making use of the p² –invariant representation of five-tuple of features.

Key words:

p^2 *-invariant, cross ratio, feature correspondence, projective geometry.*

1. Introduction

Feature correspondence is one of the fundamental tasks in computer vision. The two sets of features involved in a correspondence problem will be referred to as the reference set and the transformed set. A full projective transformation is assumed to map features form the reference to the transformed set. To be able to match the two sets projective invariant representations have to be computed.

For real images the explicit correspondence problem usually is solved making use of metrical properties, i.e., the correspondence is obtained from proximity relations. Salient features (most often corners) are detected in both images. The two feature sets are related to identical coordinate systems and the correspondences are sought in windows defined around the reference feature locations. Whenever the two images are taken from very different viewpoints (the two feature sets are connected by a strong affine or projective transformation) the search window based correspondence methods will yield many false matches. In this case the windows have to be large to account for the changes between the two images and numerous candidates are included.

The method described in this paper belongs to the class of invariant representations. Today, invariance is a major

research area of computer vision. Invariant representation of a configuration of k features is a mapping from the high-dimensional space spanned by the independent parameters of the configuration to the low-dimensional space of its invariants.

In this paper we introduce the concept of projective and permutation (p^2) invariants. In the following section, the p²-invariant of five coplanar points is derived. In section 3 the new image correspondence algorithm is described and the experimental results are analyzed in Section 4.

2. P²-Invariant representation

2.1 Projective Invariant

An absolute scalar invariant will be called projective and permutation (p^2) invariant when it remains unchanged under any permutation of the labels of the features from which it is computed. For more detailed description of the p²-invariants see [9]. The p^2 -invariants are obtained exploiting the properties of the fundamental projective invariant, the cross-ratio.

The fundamental projective invariant is the cross ratio of four collinear points A_i , $i = 1,\ldots,4$. Let the points have homogenous coordinates $x = \left(x_1^{(i)}, x_2^{(i)}\right)^t$, then their cross ratio can be defined as

$$I_1(x_1,x_2,x_3,x_4) = \lambda = \frac{(A_1 A_3)(A_2 A_4)}{(A_3 A_2)(A_4 A_1)}$$

$$= \frac{\begin{vmatrix} x_1^{(1)} & x_1^{(3)} \\ x_2^{(1)} & x_2^{(3)} \end{vmatrix} \begin{vmatrix} x_1^{(2)} & x_1^{(4)} \\ x_2^{(2)} & x_2^{(4)} \end{vmatrix}}{\begin{vmatrix} x_1^{(3)} & x_1^{(2)} \\ x_2^{(3)} & x_2^{(2)} \end{vmatrix} \begin{vmatrix} x_1^{(4)} & x_1^{(1)} \\ x_2^{(4)} & x_2^{(1)} \end{vmatrix}} \quad (1)$$

Where $(A_i A_j)$ is the oriented length of the segment delineated by the points A_i and A_j .

A point in the *n*-dimensional projective space, P^n, has the homogeneous coordinates $x_i = (x_1^{(i)},\ldots,x_{n+1}^{(i)})^t$, with at least one $x_j^{(i)}$ nonzero. The projective invariants in P^n are

defined based on $n+3$ points in general position. Similar to notation

$$[x_1,\ldots,x_{n+1}] = \begin{vmatrix} x_1^{(1)} & \cdots & x_1^{(n+1)} \\ x_2^{(1)} & \cdots & x_2^{(n+1)} \\ \vdots & \vdots & \vdots \\ x_{n+1}^{(1)} & \cdots & x_{n+1}^{(n+1)} \end{vmatrix} \quad (2)$$

will be frequently used. The determinant (2) of the homogeneous coordinates of $n+1$ points is proportional to the oriented volume of the parallelepiped defined by those points. In double algebra, it is known as the bracket over the $n+1$-dimensional vector space.

The definition of a cross ratio in P^n can then be written as

$$\begin{aligned} & I_n(x_1,\ldots,x_{n+3}) \\ & = \frac{[x_1,\ldots,x_{n-1},x_n,x_{n+2}][x_1,\ldots,x_{n-1},x_{n+1},x_{n+3}]}{[x_1,\ldots,x_{n-1},x_n,x_{n+3}][x_1,\ldots,x_{n-1},x_{n+1},x_{n+2}]} \quad (3) \end{aligned}$$

From the properties of the bracket follows that $I_n(x_1,\ldots,x_{n+1})$ is invariant to projective transformations of the point set and does not depend on the homogenous representatives chosen for the projective points.

2.2 Projective and Permutation Invariants

The projective invariants we presented were all expressed as cross ratios, which are not permutation invariants. The group of permutations S_4 has 24 elements, but there are six different cross ratios for four collinear points depending on their order. If one of them is noted by $\lambda_1 = \lambda$, the five other values are $\lambda_2 = \dfrac{1}{\lambda}$, $\lambda_3 = 1-\lambda$, $\lambda_4 = \dfrac{1}{\lambda-1}$, $\lambda_5 = \dfrac{\lambda}{\lambda-1}$, $\lambda_6 = \dfrac{\lambda-1}{\lambda}$.

To obtain permutation invariant cross ratios of four elements, it is sufficient to take an arbitrary symmetric function of the $\lambda_i, i=1,\ldots,6$. Any symmetric polynomial in λ_i has already be invariant of the permutation of the indices. These polynomials are also p^2-invariants by the projective invariance of the cross ratios. The simplest two symmetric functions $J_0(\lambda) = \sum_{i=1}^6 \lambda_i = 3$ and $J_0'(\lambda) = \prod_{i=1}^6 \lambda_i = 1$ are not interesting because they give constant values.

For a systematic investigation of all the possible second-order permutation invariants, representations of the permutation group S_4 acting on the four points must be considered. It was shown in [8] that these invariants can be expressed as a linear combination of

$$J_1[\lambda] = \frac{\lambda^6 - 3\lambda^5 + 3\lambda^4 - \lambda^3 + 3\lambda^2 - 3\lambda + 1}{\lambda^2(\lambda-1)^2}$$

$$J_2[\lambda] = \frac{2\lambda^6 - 6\lambda^5 + 9\lambda^4 - 8\lambda^3 + 9\lambda^2 - 6\lambda + 2}{\lambda^2(\lambda-1)^2} \quad (4)$$

$$J_3[\lambda] = 3$$

$$J_4[\lambda] = -3$$

The nontrivial p^2-invariants $J_1[\lambda]$ and $J_2[\lambda]$ are unbounded functions. Their ratio

$$J[\lambda] = \frac{J_2[\lambda]}{J_1[\lambda]} \quad (5)$$

However, is bounded between 2 and 2.8 and for computational convenience it is taken as the p^2-invariant function of the four collinear points. Thus, independent of how the labels of the four points were chosen, when the cross ratio computed from (5) the same value is obtained. So it will be the p^2-invariant used in the correspondence algorithm.

Given a labeling of the $n+3$ points in **P**^{n} it can be proven that

$$\begin{aligned} & J[I_n(x_1,\ldots,x_{n+3})] \\ & = J[I_n(\pi\{x_1,\ldots,x_{n-1}\},\sigma\{x_n,\ldots,x_{n+3}\})] \end{aligned} \quad (6)$$

For all permutations π over the first $n-1$ points, and all permutations σ over the last four points, any exchange of a point x_i, $i=1,\ldots,(n-1)$ with a point x_j, $j=n,\ldots,(n+3)$ violates (6) since the hyperplane defined by the new first $n-1$ points is different. There are only $\binom{n+3}{n-1} = \binom{n+3}{4}$ different arrangements of the labels in (6). Thus the p^2-invariant of $n+3$ points in **P**^{n} is a vector **J** with $\binom{n+3}{4}$ components, each having a different argument of $J[\cdot]$.

2.3 Projective and Permutation Invariants

In P^2, a configuration of five points x_i, $i = 1, \ldots, 5$, no three of them collinear, defines two independent projective invariants :

$$\lambda_1 = I_2(x_1, x_2, x_3, x_4, x_5) = \frac{[x_1 x_2 x_4][x_1 x_3 x_5]}{[x_1 x_3 x_4][x_1 x_2 x_5]}$$

$$\lambda_2 = I_2(x_2, x_1, x_3, x_4, x_5) = \frac{[x_2 x_1 x_4][x_2 x_3 x_5]}{[x_2 x_3 x_4][x_2 x_1 x_5]} \quad (7)$$

Since $n = 2$, the computed components of the p^2-invariant **J** vector are

$$J^{(1)} = J[\lambda_1] \qquad J^{(2)} = J[\lambda_2]$$

$$J^{(3)} = J[\frac{\lambda_1}{\lambda_2}] \qquad J^{(4)} = J[\frac{\lambda_2 - 1}{\lambda_1 - 1}] \qquad (8)$$

$$J^{(5)} = J[\frac{\lambda_1(\lambda_2 - 1)}{\lambda_2(\lambda_1 - 1)}]$$

and are sorted in ascending order, $J^{(i)}$. The vector **J** $= \left(J^{(1)}, J^{(2)}, J^{(3)}, J^{(4)}, J^{(5)}\right)^t$, is the sought p^2-invariant

3. Correspondence Algorithm

Several procedures have to be incorporated into a robust point correspondence algorithm.

1. Compute the locations of points from each image (Section 3.1);
2. Elimination of degenerate configurations (Section 3.2);
3. Definition of configuration dependent bounds on the components of the p^2-invariant vector to compensate for the effect of positional errors (Section 3.3);
4. Constraints for reducing erroneously matches (Section 3.4);
5. Comparisons of invariant values and the similarity between two images (Section 3.5).

3.1 Points Localization

To be able to compute invariants of points in space, it is sufficient to be able to compute the locations of the points from their images. David G. Lowe [2] has shown that the SIFT (Scale Invariant Feature Transform) features derived from the images are invariant to image translation, scaling,

and rotation, and partially invariant to illumination changes and affine or 3D projection. Here we use the approach to identify the point locations.

To achieve rotation invariance and a high level of efficiency, we have chosen to select key locations at maxima and minima of a difference of Gaussian function applied in scale space. This can be computed very efficiently by building an image pyramid with resampling between each level. Furthermore, it locates key points at regions and scales of high variation, making these locations particularly stable for characterizing the image. For more detailed description about the SIFT keys see [2].

3.2 Eliminate Degenerate Cases

The points in a five-tuple must be in a general position with on three of them collinear. Should three points be quasi-collinear, the bracket (2) is close to zero and the computation of the invariants (7) becomes numerically unstable. Let x_1, x_2 and x_3 be three points. The value of their bracket is not a reliable indicator for near singularity, i.e., of quasi-collinearty. Instead, following we use the moment matrix M_{123}

$$M_{123} = \sum_{i=1}^{3} x_i x_i^t \qquad (9)$$

The matrix is also known as scatter or Gram matrix in the literature. The smallest eigenvalue of M_{123} is also the smallest squared singular value of the matrix $X = (x_1, x_2, x_3)$, i.e., the matrix having as columns the vectors of the three points. The closeness to rank deficiency of X (the closeness of its smallest singular value to zero) measures how collinear the three points are (the linear dependency of their vectors).

The correspondence algorithm will not include three points into a five-tuple if the smallest eigenvalue is less than 0.001. Apply the SIFT key localization method which is described in 3.1 to each image. Take the point set derived from the query image as the reference set and the point sets derived from the candidate images as the transformed sets. The collinearity verification procedure is applied to both the reference and transformed sets.

3.3 Positional Uncertainty

The cross-ratio is very sensitive to noise corrupting the point coordinates. In a practical point correspondence algorithm, configuration dependent bounds must be established for each component of the p^2-invariant vector, **J**, to reduce the amount of mismatched five-tuple.

Neglecting the second order noise terms in the expression of the determinants and using the same amplitude, ε, for all the noise processes, the bounds $(\lambda_{1\min}, \lambda_{2\max})$ and $(\lambda_{1\min}, \lambda_{2\max})$ can be computed for the two independent cross-ratios. From these bounds the range of the components of **J** is obtained .

For convenience, all the positional uncertainty is associated with the reference set containing n points. For all the possible five-tuple of points in the reference set, the three vectors **J$_{min}$**, **J** and **J$_{max}$** determine configuration dependent regions in the five-dimensional space. Whenever a **J** vector from the transformed set falls within such a region, the reference and transformed five-tuple are candidates for a match. The vectors should match component-wise:

$$J_{\min}^{(i),ref} < J^{(i),trans} < J_{\max}^{(i),ref} \qquad i = 1,\ldots,5 \quad (10)$$

A matched five-tuple pair uniquely establishes five point correspondences. The value of the parameter ε has only a weak influence on the matching performance.

3.4 Eliminate Degenerate Cases

A projective transformation preserves the convex hull of a point set; Hartley extended the result for the projective transformation which exists between two images of the same coplanar point set. A pair of matched five-tuple puts five points in correspondence. A necessary (but not sufficient) condition for the match to be correct is that the two convex hulls should also be in correspondence.

The convex hull of five points may contain 3, 4, or 5 of the points. The following conditions must be satisfied by the points in a matched reference and transformed five-tuple pair:

1. The number of points on the convex hull must be the same.
2. Corresponding points must both lie on or inside the convex hull.
3. For points lying on the convex hull, neighborhood relations must be preserved.

3.5 Comparison and Similarity

Take the points set derived from the query image as the reference set while each candidate image as the transformed set. The only parameter is the tolerable positional uncertainty associated with the reference set ε.

First the p^2-invariant representations are computed for all Nondegenerate five-tuple of the reference set. Given ε, the bounds **J$_{min}$**, **J** and **J$_{max}$** are also obtained. For randomly chosen a five-tuple from the reference set, its invariant

representation is computed and matched with the reference set. The convex hull constraints eliminate most false matches. If more than one reference five-tuple remains as matching candidate, the one having the smallest Euclidean distance in the space of **J** vector is chosen as the final match. The procedure is repeated for N times.

After a final match was established, the Euclidean distances vote into a contingency table. The similarity of the two images is measured by the sum of the Euclidean distances in the contingency table.

4. Paragraphs and Itemizations

The image database used for performance testing contains 5,000 general-purpose images. These images have been already classified into several different classes with at least 800 images in each class. We take this original classification as the ground truth for judging our retrieval result. The top 10 images found by the p^2-invariant algorithm are shown in Fig. 1, Fig. 2 and Fig. 3 respectively. The query image is the left-up image in the group. And they are sorted by the similarity value in ascending order.

Fig. 1 shows the experiment on the class of battle planes. As we can seen in the figure, no matter the object scales or rotates, it can always be captured by the algorithm, and it has returned a satisfied result.

In the second experiment, we choose some more complicated image class for testing, the class of flowers. The retrieval results are shown in Fig. 2.

In Fig. 3, we test our algorithm on the collection of 2,000 animal images. The experimental result further verifies the effectiveness of the algorithm.

Fig. 1 Retrieval results of battle planes.
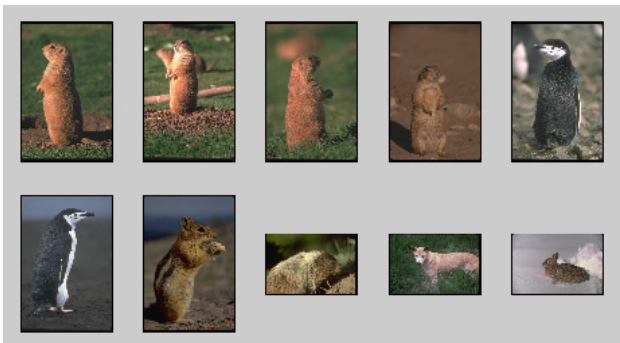
Fig. 2 Retrieval results of flowers.

Fig. 3 Retrieval results of animals.

5. Discussion

A novel image retrieval framework based on p^2-invariant is proposed in this paper. The p^2-invariant representations improve on previous approaches by being largely invariant to changes in affine/projective distortions. The performance depends on the number of features shared by the query image and the candidate images. It greatly improves the drawback of the using only low-level features for the description of image content. Experiments based on an images database with 5,000 general-purpose and randomly selected query images show the effectiveness of the proposed techniques.

Acknowledgments

References

[1]   Csurka. G. and Faugeras. O. Algebraic and Geometric Tools to Compute Projective and Permutation Invariants. In: IEEE transactions on pattern analysis and machine intelligence, Vol. 21, No. 1 January (1999)

[2]   David G. Lowe, "Object Recognition from Local Scale-Invariant Features", In Proc. of the International Conference on Computer Vision, Corfu (1999)

[3]   Faugeras, O. Three Dimensional Computer Vision. MIT Press: Cambridge, MA (1993)

[4]   Faugeras, O. and Mourrain, B. On the geometry and algebra of the point and line correspondences between N images (2004)

[5]   Frederic Cao., Application of the Gestalt principles to the detection of good continuations and corners in image level lines. In: Computing and Visualization in Science. Springer-Verlag, April (2004)

[6]   Hartley, R., Chirality. In: International Journal of Computer Vision, 26 (1), 41–61 (1998).

[7]   James I, Cronley and Alice C. Parker, A Representation for Shape Based on Peaks and Ridges in the Difference of Low-Pass Transform. May (1998)

[8]   Meer, P., Ramakrishna, S., and Lenz, R.: Correspondence of Coplanar Features Though P^2-Invariant Representations. In Application of Invariance in Computer Vision, J.L. Mundy et al. (Eds), Springer-Verlag, (1994) 473-492

[9]   Meer, P., Ramakrishna, S., and Lenz, R.: Efficient Invariant Representations, In: International Journal of Computer Vision 26(2), 137–152 (1998)

[10] Mohr, R and Triggs, B, Projective Geometry for Image Analysis. In : A Tutorial given at ISPRS, Vienna, July (1996)

[11] R. Hartley, Invariants of Points Seen in Multiple Image, preprint submitted for publication

[12] Rothwell, C., Csurka, G. and Faugeras, O. A Comparison of Projective Reconstruction Methods for Pairs of Views.

[13] Yao Yu-rong, Zhang Yu-jin, Shaped Based Image Retrieval Using Wavelet and Moment. In Journal of Image and Graphics. Vol.5(A). No. 3 Mar (2000)

[14] Zhang, Y.J. Content-Based Image Visual Information Retrieval. Science Publisher (2003)

[15] Zhang, Y.J. Image Engineering (1): Image Processing and Analysis, Tsinghua University Press (1999)

[16] Zhang, Y.J. Image Segmentation. Science Publisher (2001).

Ruihang Wu received the B.S. degree in mathematics from Hunan Univ. in 2004. She is now studying in the Department of Computer Science at Xiamen Univ. for M.S. degree. Her interest research interest includes multimedia information retrieval.

Shaozi Li received the B.E. degree from Hunan Univ. in 1983 and M.E. degree from Xi'an Jiaotong Univ. in 1988. Aftering working as a dean of computer teaching-research section, a director of institute of computer in Luoyang technical institute (from 1983), and an assistant professor (from 1998) in the Department of Computer Science, Xiamen Univ., he has been a professor at Xiamen Univ. since 2006. His research interest includes AI, multimedia information retrieval, and image retrieval.

Fengmei Zou received the B.S. degree from Fujian Normal University in 1991, M.S. degree from Beijing Normal University in 1994, M.E. degree from State Univ. of New York at Buffalo in 2000. He received the Dr. Science degree from Chines Academy of Sciences in 1997. After working as a software engineer in AT&T, ENREACH TECH. INC. and VUGENIX INC., he has been an assistant professor at Xiamen Univ.. His research interest includes video/audio frequency compression/decompression, algebraic invariant, pattern recognition and digital image processing.