MA T AL BAS D N J CT N A ANC

Rajashekhar, Subhasis Chaudhuri and Vinay. P. Namboodiri

Department of Electrical Engineering
Indian Institute of Technology-Bombay
Mumbai 400076. India
{raja, sc, vinaypn}@ee.iitb.ac.in

ABSTRACT

In this paper we propose an image retrieval scheme based on projectively invariant features. Since cross-ratio is the fundamental invariant feature under projective transformations for points, we use that as the basic feature parameter. We compute the cross-ratios of point sets in quadruplets and a discrete representation of the distribution of the cross-ratio is obtained from the computed values. The distribution is used as the feature for retrieval purposes. The method is very effective in retrieving images having similar planar 3D structures like buildings.

1. INTRODUCTION

Content based image retrieval (CBIR) has been an active area of research for many years. Most of the present technologies rely on annotations attached to the data or visual features such as color, texture or shape along with the spatial relationships between the regions in an image [1, 2]. Among all the features, color is probably the most straightforward one which enables human to recognize different images. Therefore the color histogram [3] is a well known technique used to represent the color of an image. However, it yields many false retrievals. Some researchers have explored the utility of the textural features in retrieving images. Ma and Manjunath [4] evaluated the use of texture features using Gabor wavelet transform representations. In order to capture the texture, Jhanwar *et al.* presented in [5] a translation, and an illumination invariant retrieval scheme using a motif cooccurence matrix. Shape is another important visual attribute that describes the presence of specific types of objects in a scene. A region based shape descriptor invariant to rotation, scale and translation has been presented in [6]. A vector space based approach is used to detect the shape of an object. Texture and shape attributes of an image drastically change under perspective transformations. If one looks at the same scene from a different view angle, the corresponding images may look very different in terms of textural and shape properties. Therefore under perspective transformations image retrieval techniques based on only the low level features cannot effectively capture the characteristics of an image. In other words, the low level visual features are not invariant under the perspective changes. This inspired us to propose a feature which is invariant under projective transformations.

The idea of projective invariance has been used for object recognition [7, 8]. However, to the best of our knowledge there has been no prior work in CBIR that makes use of the 3D properties of a scene. The motivation for developing a CBIR system invariant to perspective changes is as follows. We are currently building an architectural database consisting of many images of buildings, their frontal as well as side views. A CBIR scheme is required to be developed that would retrieve building images very similar to the query in terms of the 3-D structure, irrespective of paints and view angle. None of the currently available techniques was found to be useful. The only way to represent the 3D structure of an object with view point invariance in 2D space is using projective geometry based principles. In this paper we propose an image retrieval technique using cross-ratio which is an invariant feature under perspective transformations. Structural entities like buildings and tracks have similar linear features sets like facades, windows, fishplates, etc. which result in real or virtual cross-ratio sets. These values are very similar across the same structural entity set. The cross ratio is a feature defined for point sets lying on a line. For example cross-ratio values computed for quadruplets in a building image do not change under perspective transformations. We compute the cross-ratio (CR) values of all point sets in collinear quadruplets. We obtain a feature vector of cross-ratio histogram (CRH) based on the CR values. This drastically decreases the length of the feature vector and the time complexity involved in similarity measurement. Due to sparsity of the CRH and the sensitivity to noise in CR calculation we obtain an estimate of the CR distribution by using a smoothing spline on the computed CRH. We assess the retrieval performance by computing the precision and recall rates. Other than buildings, this retrieval system can also be used for retrieving other planar entities which are structurally similar to each other like rail-

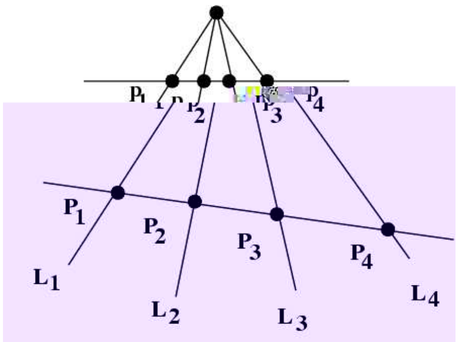


Fig. 1. Illustration of cross-ratio for the point sets.

way tracks captured from different perspective views. Experimental results show one can obtain good retrieval rates.

2. PROPOSED METHOD

Given a point set quadruplet lying on a straight line, the cross-ratio is defined [9] as

$$\mathrm{CR}(P_1, P_2, P_3, P_4) \quad = \quad \frac{(X^3 - X^1)(X^4 - X^2)}{(X^3 - X^2)(X^4 - X^1)}, \quad (1)$$

where P_1, P_2, P_3 and P_4 are the point sets in a real world scene and p_1, p_2, p_3 and p_4 are the points obtained after projecting the world points on the image plane. Cross ratio of the real world points is computed in terms of the cross-ratio of the image points using the equation 1. X^1, X^2, X^3, X^4 represent the corresponding positions of each point along the line, e.g.$(X^3 - X^1)$ is the distance between points P_3 and P_1 as shown in figure 1. Similarly one can compute the cross-ratio of the concurrent lines $L1$, $L2$, $L3$ and $L4$ by computing the angles between the lines (see [9]). Cross-ratio evaluated using the point sets and the lines are identical. It may be noted that the cross-ratios are not permutation invariant, and hence some authors have proposed the use of j-invariants [9] computed from the cross-ratios. However, this is not important in most computer vision problems as generally the left to right ordering is maintained when the view angle of the point set is changed.

We use the following algorithm to compute the cross-ratio values. We obtain all possible feature points in an image by applying the Harris corner detector (see fig 2) (b). We then detect the prominent line segments passing through these corner points using Hough transform operation 2 (c). Individual line segments thus obtained are dilated by using a small circular structuring element. We then perform the logical AND operation on point sets of fig 2 (b) with each dilated line segment to extract the feature points. The feature points thus obtained are used in computing the CR values. We compute the CR values for collinear quadruplets using the equation 1. The number of cross-ratios derived in an image may be quite large. Therefore the time complexity involved in matching the cross ratios directly for retrieval purposes is high. Also indexing becomes inefficient because of a large number of cross-ratio values. Further, the calculation of cross-ratios are sensitive to noise. Therefore we derive a feature vector based on distribution of the cross-ratio values rather than individually matching them. We obtain the distribution (histogram) of the cross-ratios choosing an appropriate bin size. Usually the cross-ratio histogram is sparse in nature. Due to sparsity and sensitivity to noise, we smoothen this feature vector by using a smoothing spline followed by normalization so that the resulting curve actually represents a distribution. The smoothed histogram is then used to obtain similarity measurement. We use Euclidean distance to find the similarity between the smoothed CRHs of the query image and the database images.

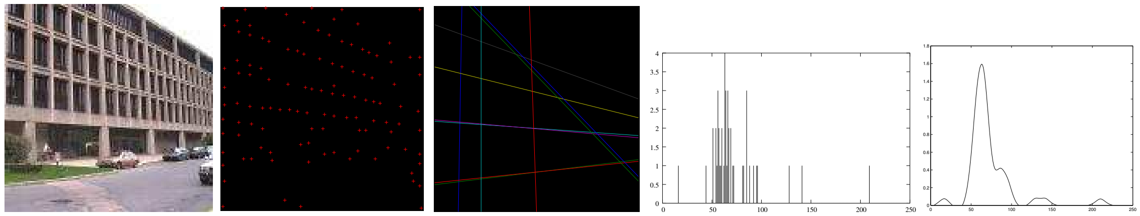
3. EXPERIMENTAL RESULTS

We tested our algorithm considering different types of structures on an image database of size 4000. Our image database collection is mainly from COREL's website. However, it lacks in having a suitable representation of building facades or their side views. We append the VISTEX database and further, added a number of predominantly building images in the combined database. We evaluated the performance of our approach by computing the precision and the recall rates. Precision rate is defined as the fraction of the retrieved images which are relevant. Recall is the fraction of the relevant images which have been retrieved. In all the experimental results image displayed first is the query and the ranking goes from left to right and top to bottom.

One typical building image from the image database is shown in fig 2 (a). Fig 2 (b) describes the result after the corner detection. Fig 2 (c) shows the detected straight lines having at least four collinear corner points. Its CRH and smoothed CRH are shown in Figures 2 (d) and 2 (e), respectively.

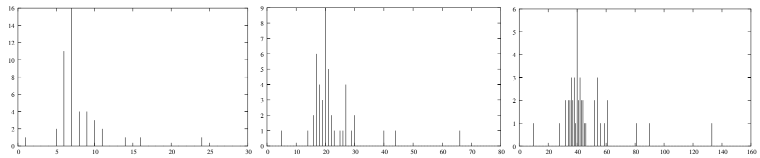
We now show the effect of bin size for quantization of the estimated cross-ratios. Typically the minimum and maximum cross-ratio values for a building image are found to be 0 and 5, respectively. This dynamic range of the cross-ratio values is quantized uniformly by choosing N number of bins. We choose the bin sizes 30, 80, 160, and 250 for quantization. For a small N (30-80), one can observe from Figures 3 (a-b) that the distribution of CR values differs considerably from the distribution shown in fig 2 (d). Beyond 250 bins, experimentally we notice that this distribution does not differ much. Also we find that CRH using a small number of bins affects the retrieval efficiency of our scheme.

We initially analyze the performance of our scheme using CRH feature vector for smaller number of bins. One can see from fig 4 that some of retrieved road, track, tree images are irrelevant. Usually natural scene images do not



(a) (b) (c) (d) (e)

Fig. 2. (a) A building image, (b) points set after Harris corner detection, (c) detected line segments after Hough transform, (d) its CRH using 250 bins, (e) its smoothed CRH.



(a) (b) (c)

Fig. 3. (a-c) The CRHs for N=30, 80, 160 bins, respectively for the fig 2(a).

have a definite structure. For such images one detects some spurious feature point sets. Due to a coarse quantization of the cross-ratios, the CRH of the tree images look similar to the building images. Therefore false retrievals are mainly due to an inappropriate number of bins used to quantize the CR values. Hence we quantize the dynamic range of the CR values of all the images in the database by choosing an appropriate number of bins (N=250). Experimental results for retrieval using the CRH feature vector with 250 bins are shown in fig 5. Due to proper quantization, CRH of the tree images differs much from that of building images. Therefore one can see that such tree images disappear from the list of relevant retrievals. However, we notice some of the railway track images in the retrieval set. Perspectively one can group the building and track images together. Since fish-plates in the rail track give rise to a number of collinear points that may yield a similar set of cross-ratios. As a result we retrieved a few track images along with the building images (see fig 5).

Experimentally we noticed that due to the sparsity of the cross-ratio distribution and its sensitivity to noise, we were not able to obtain a few building images in the database as the relevant retrievals. We therefore performed experiments using the smoothed CRH as the feature vector. Experimental results shown in fig 6 indicate that all the top 25 retrieved images are now relevant to the query. When track image is used as a query image, we not only retrieved track images, but also a few building images as shown in fig 7. Only top 15 retrievals are shown in all the results.

Figure 8 shows the precision-recall diagram for the different features. One can see from the diagram, for small N (30, 80) both the precision and the recall rates are very low. For the top 30 retrievals we obtained a precision rate of 40% and a recall rate of 48% (N=30). Subsequently we notice an improvement in the recall and precision rates with N=250 for CRH feature vector and smooth CRH feature vector. We obtain 73.33% precision rate using CRH feature vector and 83.33% using smoothed CRH feature vector for the top 30 retrievals.

Fig. 4. Retrieved images based on the CRH using 30 bins.

4. CONCLUSION

In this paper we presented an image retrieval scheme using a projective invariant feature. We explained the procedure of extracting the feature points in an image and also the calculation of the cross-ratio values using these feature points. We discussed the issue of bin size in quantizing the cross-ratios. Here we also discussed the the need for smoothing the cross-ratio histogram. The retrieval effectiveness of our scheme is assessed by computing the precision and recall rates. The precision-recall diagram using different features is described. We also showed the images retrieved using CRHs and smoothed CRH feature vectors. Currently we are looking into the problem of incorporating projective invariants for other entities like curves.

Fig. 5. Retrieved images based on the CRH using 250 bins.

Fig. 6. Retrieved images using the smoothed CRH.

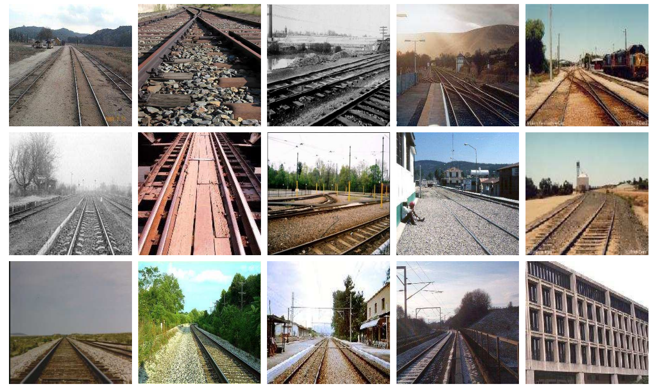


Fig. 7. Retrieved images using the smoothed CRH for rail track query.

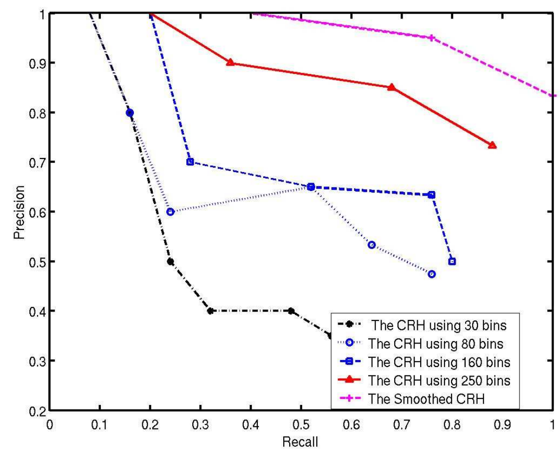


Fig. 8. Precision/Recall diagram for different features.

5. REFERENCES

[1] Y. Rui, T.S. Huang, and S. F. Chang, "Image Retrieval:Current Techniques Promising Directions and Open Issues," *Journal of Visual Communication and Image Representation*, vol. 10, pp. 39–62, 1999.

[2] R. C. Veltkamp and M. Tanse, "A Survey of Content-Based Image Retrieval Systems," in *Content-Based Image and Video Retrieval*, O. Marques and B. Furht, Eds. Kluwer, 2002.

[3] M.J.Swain and D.H.Ballard, "Color Indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, September 1991.

[4] W.Y. Ma and B.S. Manjunath, "A Comparison of Wavelet Transform Features for Texture Image Annotation," in *Proc.IEEE Conference on Image Processing*, Washington D.C, USA, October 1995.

[5] N. Jhanwar, S. Chaudhuri, G. Seetharaman, and B. Zavidovique, "Content Based Image Retrieval Using Motif Coocurence Matrix," in *Proc.ICVGIP*, Ahmedabad, India, December 2002.

[6] H. Kim and J. Kim, "Region-based Shape Descriptor Invariant to Rotation Scale and Translation," *Signal Processing:Image Communication*, vol. 16, no. 1-2, pp. 87–93, September 2000.

[7] A. Zisserman, D. Forsyth, J. Mundy, C. Rothwell, J. Liu, , and N. Pillow, "3D Object Recognition using Invariance," *Artificial Intelligence*, vol. 78, no. 1-2, pp. 239–288, 1995.

[8] C. A. Rothwell, A. Zisserman, D. A. Forsyth, and J. L. Mundy, "Planar Object Recognition using Projective Shape Representation," *IJCV*, vol. 78, no. 1-2, pp. 239–288, 1995.

[9] J. L. Mundy and A. Zisserman, "Projective Geometry for Machine Vision," in *Geometric Invariance in Computer Vision*, J. L. Mundy and A. Zisserman, Eds. MIT press, Cambridge, Massachusetts London, England, 1992.