Indexing Images of Buildings Based on Geometrical Invariant Hough Descriptors

Chang-Tsun Li and *Xiang Yuan*

Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK

c-t.li@warwick.ac.uk, x.yuan@dcs.warwick.ac.uk

ABSTRACT

A CBIR system for retrieving images with specific buildings from databases is proposed in this paper. We exploit the *parallelism invariance* property of the line features of buildings in images in the derivation of two new geometrically invariant linear feature descriptors. We call these descriptors *Hough descriptors* as they are extracted from the Hough transform domain. The underlying concept is to utilise the invariance of parallel line features such that the individual edges (local property) can be used in a collective manner to embed their global relationship in the feature descriptors. Upon receiving a query image, the CBIR system transforms the edges detected from the query image into the Hough transform domain. The transform domain is divided into 180 degrees/bins in order to reveal the linear edge distribution. From each bin, the *peak percentage profile* and *distance ratio profile* are calculated to serve as the descriptors of images. That is to say that an image descriptor consists of two components (*peak percentage profile* and *distance ratio profile*), each with 180 elements. The circular correlations between the *peak percentage profile* and *distance ratio profile* of the query image and those of the database images are then taken as the similarity measure for ranking the relevance of the database images to the query.

Index Terms— content based image retrieval (CBIR), feature extraction, matching algorithm, Hough descriptor, multimedia retrieval, image indexing

I. INTRODUCTION

With enormous efforts invested by various sectors of our societies in the last two decades in digitising historical documents, archiving national heritages and multimedia content productions, a wealth of valuable multimedia data in digital forms has been accrued around the world. These collections of multimedia content provide tremendous application opportunities in many walks of our everyday life. Many sectors of our societies such as education, broadcasting, entertainment, tourism, just to name a few, can benefit from this wealth of multimedia if easy access is made available. Among so many possibilities, building image indexing [7, 12, 25] is useful in helping retrieve multimedia contents with specific buildings to serve purposes such as finding news reports of events occurring around a specific building, finding information about a specific building, collecting images of buildings of a specific architecture style, etc. An exciting technological advance in the past decade is the development of content-based image retrieval (CBIR) [3, 5, 11, 23, 24], which allows the indexing systems to rely on multimedia content, rather than text-based annotation, to retrieve required content. For example, by search for a specific building in a large GPS-tagged image archive, the location where a particular painting of the building was painted can be determined. Another example is the so called computational rephotography [1], which allows the creation of a new photo from the same viewpoint of an existing photo that was taken from the same scene a long time ago. With a CBIR system, this can be done by matching the historic image with a modern-day image that matches its perspective. In CBIR systems, a set of features, such as colour [2, 4, 14, 26], texture [2, 9, 11] and shape [5, 7, 8, 10, 13, 23, 24, 25], are used to represent a digital image. Retrieval is performed by comparing the feature set of the query image against those of the database images using a matching algorithm [23]. However, the colours and texture of a building image is highly variable because of varying illumination conditions. Moreover, most digitised versions of historical documents are in gray level. Therefore, colour is not suitable for building

image retrieval in most cases. Perceptually, humans tend to be able to recognise buildings solely by their shapes. These factors have attributed to the popularity of shape features as descriptors for building identification [7, 12, 25]. Shapes are formed by linear features such as lines and edges. Therefore, shape-based methods tend to exploit lines and edges in their search for relevant content [5, 10, 24, 27]. One of the advantages of the approaches that exploit linear features is that they do not require sophisticated segmentation [15] and representation of objects. As a result, the decision as to whether or not particular man-made structures are present in the image can be made without the need of recognising specific instances of the structures of interest through the utilisation of *a priori* knowledge about their properties [7, 12, 13, 25].

The rest of the paper is organised as follows. Section II provides a brief review on related shape description methods. In Section III, we describe the details of our proposed CBIR system. Experimental results are demonstrated in Section IV. Section V concludes the work and points out future work.

II. RELATED WORK: SHAPE-BASED FEATURE DESCRIPTORS

Human can recognise objects solely from their shapes. Therefore, shape features provide a powerful clue for object identification. A good shape descriptor can capture characteristic shape features as far as image retrieval is concerned. Also, an effective shape descriptor should be invariant to geometrical operations such as scaling, rotation, and translation [19]. There have been a great number of CBIR researches based on shape descriptors. Shape descriptors can be broadly divided into two categories, namely, region-based descriptors [19, 21, 23, 24, 26] and contour-based descriptors [5, 10, 24, 27]. Region-based descriptors, such as Zernike moments [18, 22, 24], exploit only shape interior (regional) information. On the other hand, contour-based descriptors such as Fourier descriptors (FD) [26] and curvature scale space (CSS) [5] exploit shape boundary information.

A. Region-based Descriptors

Region-based descriptions of shape specify the object's "body" within the closed boundary, all the pixels within a shape boundary are taken into account to obtain the shape representation, rather than only using boundary information. Region-based descriptors often use moments to describe shapes. Zernike moments are derived from Zernike polynomials which are a sequence of polynomials orthogonal on a unit disk. It is Teague [21] who first introduced Zernike moments to the image retrieval field. His work has inspired many works on image retrieval applications [18, 17, 20, 24]. Zernike polynomials form a complete orthogonal set over the interior of a unit circle and Zernike moments are a series of numerical signal expanded from orthogonal bases. The precision of shape representation depends on the number of Zernike moments extracted from the expansion. Since Zernike basis functions take the unit disk as their domain, this disk must be specified before moments can be calculated. To make the extracted moments scale and translation invariant, all the shapes are normalised to a unit disk, and then the disk is centred on the shape centroid. Using the magnitudes of the moments can make the Zernike moments rotation invariant. The advantage of Zernike moments is that, because higher order moments can be easily constructed, it is capable of representing more complex shapes. However, high order Zernike moments are more sensitive to image noise than the lower ones. Moreover, traditional ways of devising Zernike descriptors only take into account the magnitudes of the moments and ignore the phase information [24]. A few approaches [18] that incorporate both have been proposed. However, this improved performance has to be gained at the expense of higher computational complexity. Another major shortcoming of these continuous moments is that the implementation requires numerical approximation.

Mukundan *et al.* introduced a new set of orthogonal moment functions based on the discrete Tchebichef polynomials [16], which does not involve any numerical approximation in the implementation of moments because the basis set is orthogonal in the discrete domain of the image coordinate space. This property makes Tchebichef moments superior to the conventional orthogonal moments such as Zernike moments, in terms of preserving the analytical properties needed to ensure information redundancy in a moment set. The radial Tchebichef moments are invariant with respect to image rotation. However, like Zernike moments, in order to achieve scale invariance, the original shape has to be normalised to a predetermined size. This shape normalisation process is time consuming and tends to give rise to the loss of some characteristics of a shape. Consequently, distorted moments may be derived. To overcome this limitation, El-ghazal *et al.* have proposed to use the area and the maximum radial distance of a shape to normalise the radial Tchebichef moments in order make them scale invariants [6]. By normalising the moments without normalizing the shape to a predetermined size, characteristics of the shape can be preserved.

B. Contour-based Descriptors

Contour-based descriptors can extract shape features from boundary information. There are mainly two approaches for contour modelling: global approaches and structural approaches [26]. Usually, a feature vector derived from the integral boundary is used to describe the shape in global approaches such as Fourier descriptors, shape signature, etc. While in the structural approaches such as CSS, chain code, polygon, etc, the shape boundary is broken into segments based on some particular criterion. Shape boundary is a set of coordinates identified by contour tracing. Various types of information, such as centroid distance, chain code and cumulative angles, can be derived from shape boundary to form *boundary signatures*. For example, the centroid of a shape is the average coordinates of the normalised

shape boundary. As a common contour-based approach, Fourier descriptors (FD) of shapes are formed by applying the Fourier transform on the boundary signatures [26]. It is has been shown [26] that Fourier descriptors transformed from the centroid distances are invariant to translation and rotation. The advantage of Fourier descriptors is that they are based on the well-developed theory of Fourier analysis, which makes FD relatively easy to implement. Global shape features are captured by lower order Fourier coefficients and the finer shape features are captured by higher order coefficients. Noise, which only appears in very high frequency bands, can be truncated out by low-pass filtering. However, after the Fourier transform, local shape information is distributed to all coefficients. Therefore, local information could be lost. Moreover, Fourier descriptors use only the magnitude components and ignore the useful shape information contained in the phase components of the Fourier transform.

Unlike Fourier descriptors, curvature scale space (CSS) descriptors are used for representing key local shape features [5]. Curvature, as an important local measure of how fast a planar contour is turning, is exploited in the scale space to generate CSS descriptors. In the scale space, both the locations and the degrees of convexities in shape boundaries are detected. The first step to obtaining CSS descriptors is to calculate the CSS contour, which is a multi-scale organisation of curvature points. Curvature points are located in the shape boundary. Then the shape is propagated to next scale by applying Gaussian smooth, and in the new scale space all the curvature points are extracted and located in the shape boundary, which will be evolved into next scale by applying Gaussian smooth again. This process stops when no more curvature points can be found. One of the advantages of CSS descriptors is its ability in representing key local features, such as the locations and the degrees of convexity on the shape boundary. Also, the dimension of CSS descriptors is relatively lower, thus making matching easier. However, the disadvantage of CSS descriptors is that they cannot effectively capture global features, which are also important for shape representation. Also, comparing to Fourier descriptors,

deriving CSS descriptors is usually more computationally expensive.

The aforementioned methods, region-based and contour-based, are useful for object recognition and can certainly provide insight into the derivation of building feature descriptors. However, buildings are complicated structure. The shapes of the same building may look very different when viewed from different angles. Therefore, alternatives have to be found. As reported in [25], we observed that the distribution of points in the Hough transform domain can effectively represent the distribution of linear signals in building images. In [25], the edge map of a building image is extracted first. Secondly, Hough transform is applied to transform the edge map from the spatial domain into the Hough transform domain. By partitioning the Hough transform domain into a number of *orientation* bands (each covering the entire *distance* range), the *centroid* of the points in each band is calculated. Then a band-wise matching (BWM) algorithm is employed to measure the similarity between the query image and the images in the database by taking the centroid set as the feature vector. Experiments showed that the proposed CBIR system is effective in retrieving building images with strong linear features. However, the improperly predefined bandwidth can make the bands of two Hough diagrams miss-matched. Also, the use of *distance centroids* as one of the feature descriptors is later proved to be problematic. For example, in Figure 1, l_2 is obtained by rotating l_1 by $\Delta\theta$ anti-clockwise. We can see that changes in phase are consistent in both the Hough transform domain and spatial domain. However, after the rotation, the distance information in the Hough transform domain changes (i.e., $\rho_1 \neq \rho_2$). Therefore, *distance centroid* cannot exactly represent the distance information in the Hough transform domain. To overcome the limitations of this CBIR system, an improved CBIR system based on a new set of geometrically invariant feature descriptors, called *Hough descriptors*, extracted from the Hough transform is proposed in this work.

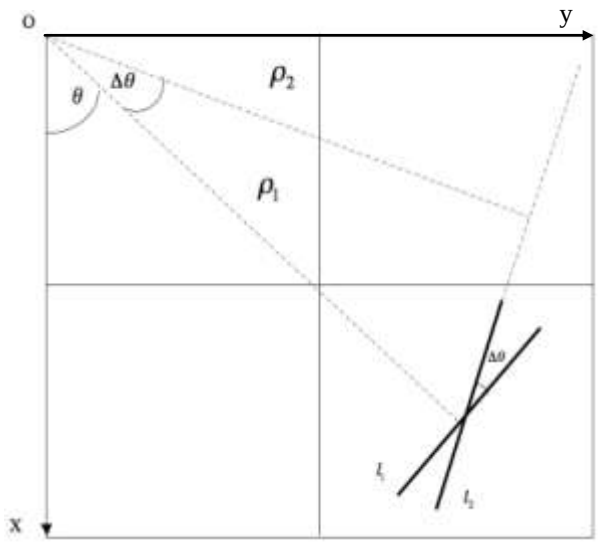


Figure 1. Hough analysis of lines l_1 and l_2 in the spatial domain

III. PROPOSED CBIR SYSTEM

Buildings usually have prominent parallel features. Without taking the perspective factor into account, parallel lines remain parallel under rotation, translation, scaling and shearing operations. For example, when scaled or sheared, the distances between parallel lines change and the lengths of the lines become different, but they remain parallel. When rotated, the orientation of the lines changed, but they still oriented in the same direction. In real imaging situations, the perspective factor can distort parallelism to various extents depending on the camera's relative location to the object. However, in most cases when a building is photographed, the camera is positioned reasonably distant from the building. In such a distance, the parallelism distortion becomes negligible. This "parallelism invariance" forms the basis of our derivation of building feature descriptors proposed in this work.

A. Feature Extraction

Most buildings have linear edges. However, linear edges are low-level local property of objects.

When used alone, it is unlikely that edges can capture the high-level semantics the user is looking for or convey information about parallelism we want to exploit. Therefore, edges and line segments need to be used *collectively* in some way in order to describe objects at a higher level in a wider or even the global context [12, 25]. Hough transform is usually used to detect collinear points and lines in the spatial domain. The points coexist in the spatial domain on the same line oriented θ° from the downwards pointing axis and with a distance ρ to the origin (upper-left corner) contribute to the count / value in the accumulator cell (θ, ρ). Therefore, points on parallel lines in the spatial domain will distribute across various cells in the same orientation (i.e., along the vertical band of a certain orientation in the Hough transform domain (θ-ρ space).

In this work, the first step towards feature extraction is to reduce the interference of noise by applying a 3 × 3 Gaussian kernel with a standard derivation of 0.5. The Canny edge detector is then applied to the Gaussian smoothed image to create an edge map. The edges of the map are then thinned to remove redundancy. Subsequently, the Hough transform is applied to the map of thinned edges. Throughout the rest of this work, we will use the term *Hough peaks* to represent the accumulator cells in the Hough transform domain with high counts. Illustrated in Figure 2(a) is the image to be analysed while Figure 2(b) is the map of the thinned edges and Figure 2(c) is the Hough transform of Figure 2(b). In Figure 2(c), each Hough peak is marked with a square. To show that Hough peaks can reveal strong signature of the linear features in images, we enhanced all the edge pixels in Figure 2(b) that contribute to the formation of the peaks in the Hough transform domain and demonstrated the new edge map in Figure 2(d). Apparently, most of the salient linear features such as the vertical edges of the monument and the horizontal edges of the pool have been enhanced in Figure 2(d) while the irrelevant edge information, such as the edges of the clouds and the reflection of the monument in the water is left out.

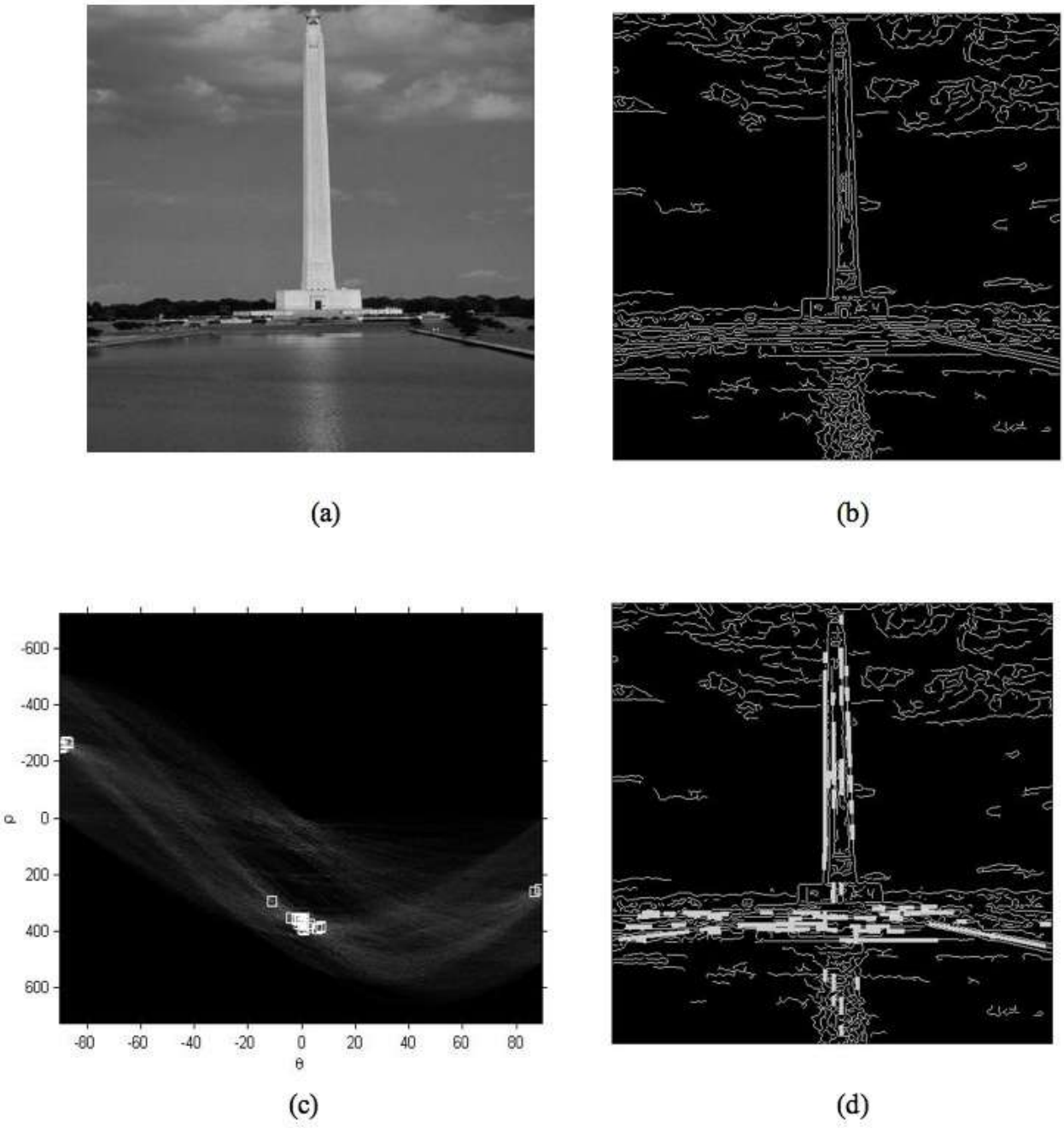
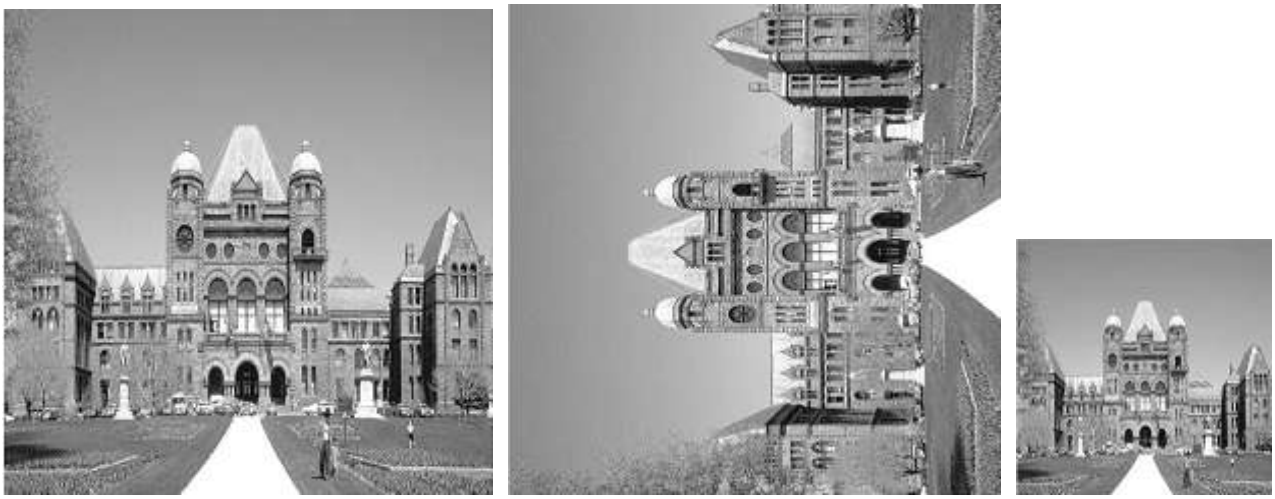


Figure 2. a) An image containing a building, b) the edge map of (a), c) the Hough transform of (b), d) is the enhanced edge map with the points contributing to the formation the peaks in (c).

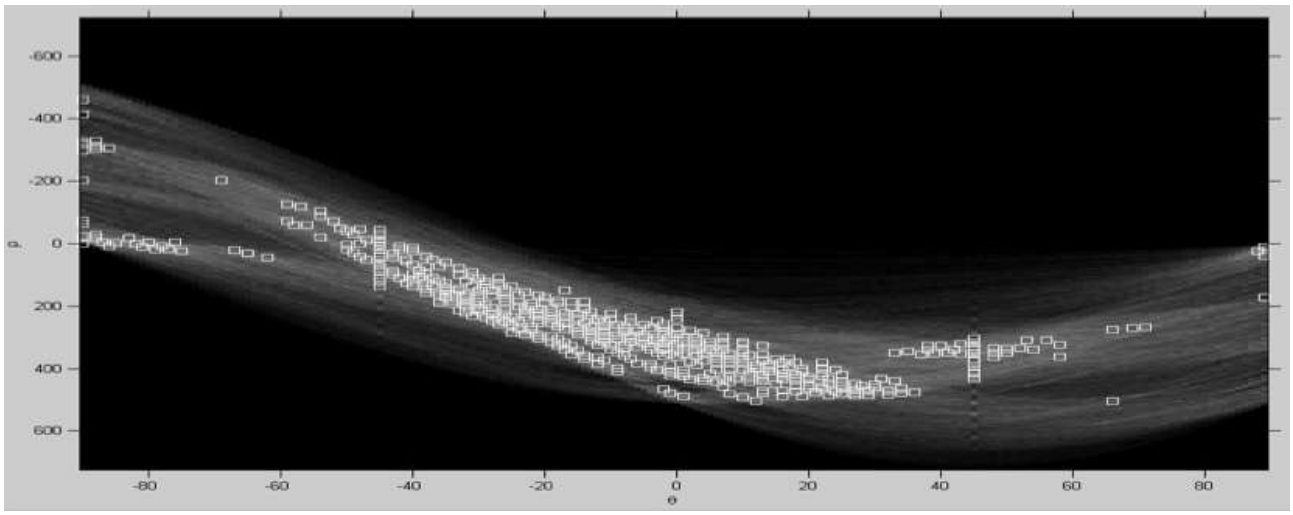
The next step is to make use of the Hough peaks' distribution so that the information about the edges can be exploited in a collective manner to describe buildings. Figure 3(a) and (b) show a building image and its 90˚-rotated version, respectively. By comparing the corresponding Hough transforms of

their edge maps as shown in Figure 4(a) and (b), we can see that the distribution of the Hough peaks from -90˚ to 0˚ in Figure 4(a) are similar to the distribution of the Hough peaks from 0˚ to 90˚ in Figure 3(b) and the distribution of the Hough peaks from 0˚ to 90˚ in Figure 3(a) are similar to the distribution of the Hough peaks from -90˚ to 0˚ in Figure 3(b). From this observation, we can see that the distributions of Hough peaks in Figure 3(a) and (b) are 90˚ different, which is the same as the phase-difference (rotation) in the spatial domain, whereas the distance (magnitude) information of the corresponding Hough peaks in the diagrams have not changed. Figure 3(c) is an image of the same building in Figure 3(a) at a smaller scale. We can see that the distributions of the Hough peaks in their Hough transform domain (see Figure 4(a) and 4(c)) are quite similar too. These two observations and Figure 4 indicate that the Hough transform is invariant to rotation and scaling if the Hough diagram is seen as a circular ring, with -90˚ overlaps with +90˚.

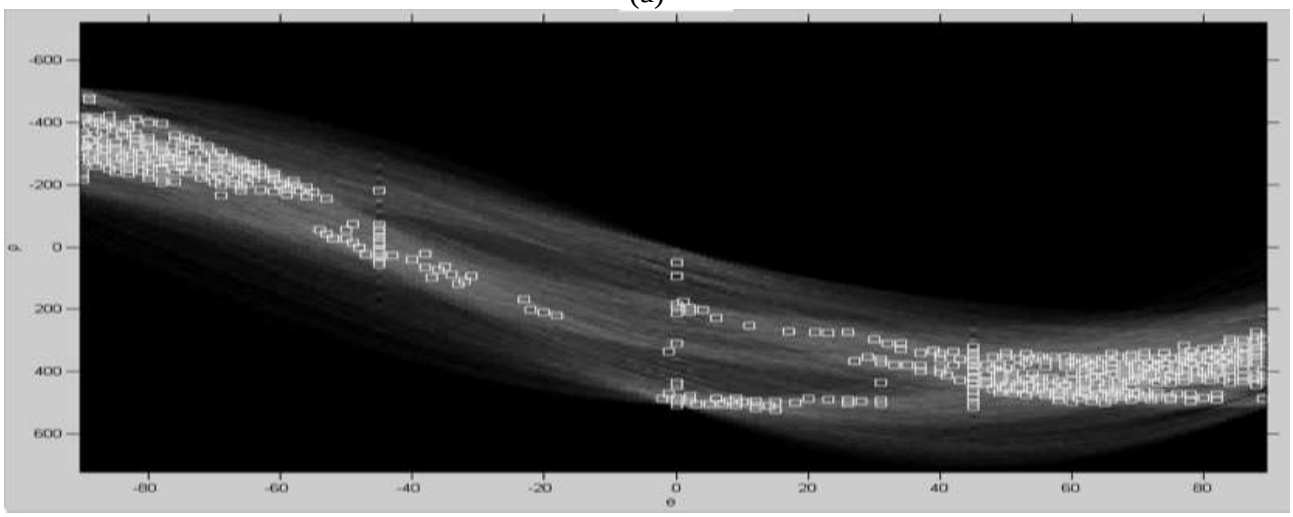


(a) (b) (c)

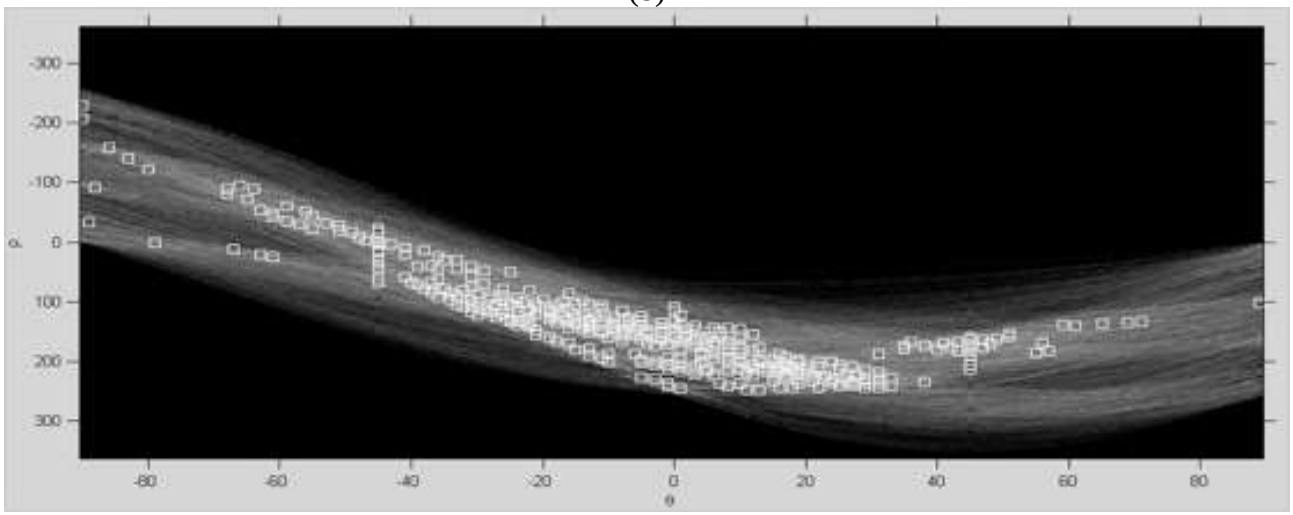
Figure 3. (a) is a grey scale image with a size of 512 × 512 pixels. (b) is a rotated version of (a). (c) is a down scaled version of (a) with a size 256 × 256 pixels.



(a)



(b)



(c)

Figure 4. (a), (b) and (c) are the Hough diagrams of Figure 3(a), (b) and (c), respectively.

In the Hough transform domain, instead of using *distance centroid* as we did in [25], we calculate the *percentage of Hough peaks* distributed in each 1˚-wide bin. The *peak percentage* of the i^{th} orientation bin is defined as:

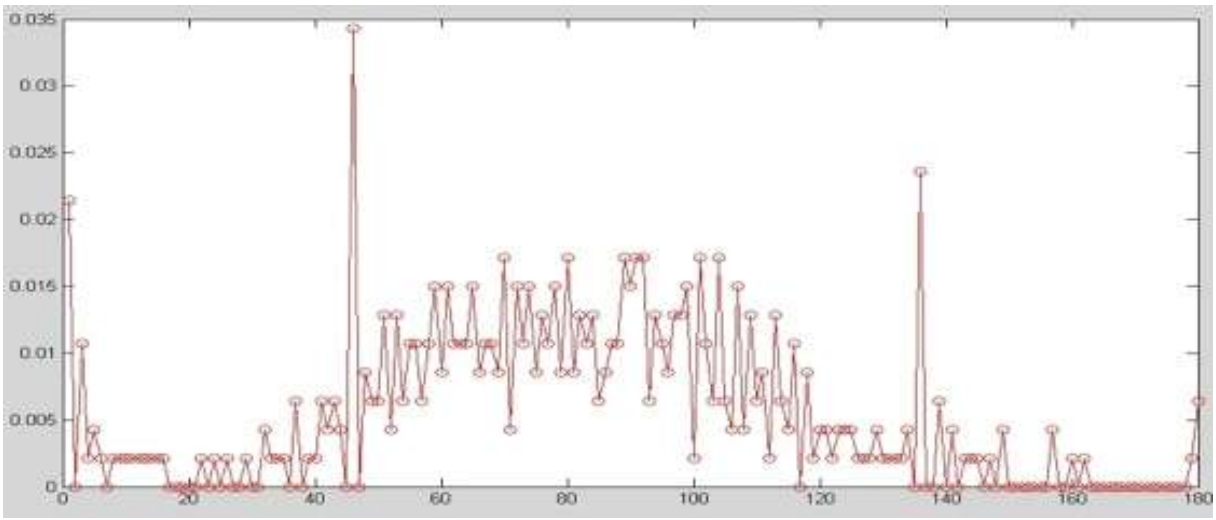
$$p_i = n_i / n \tag{1}$$

where n_i is the number of peaks of the ith bin and n is the total number of peaks in the entire Hough transform domain, i.e.,

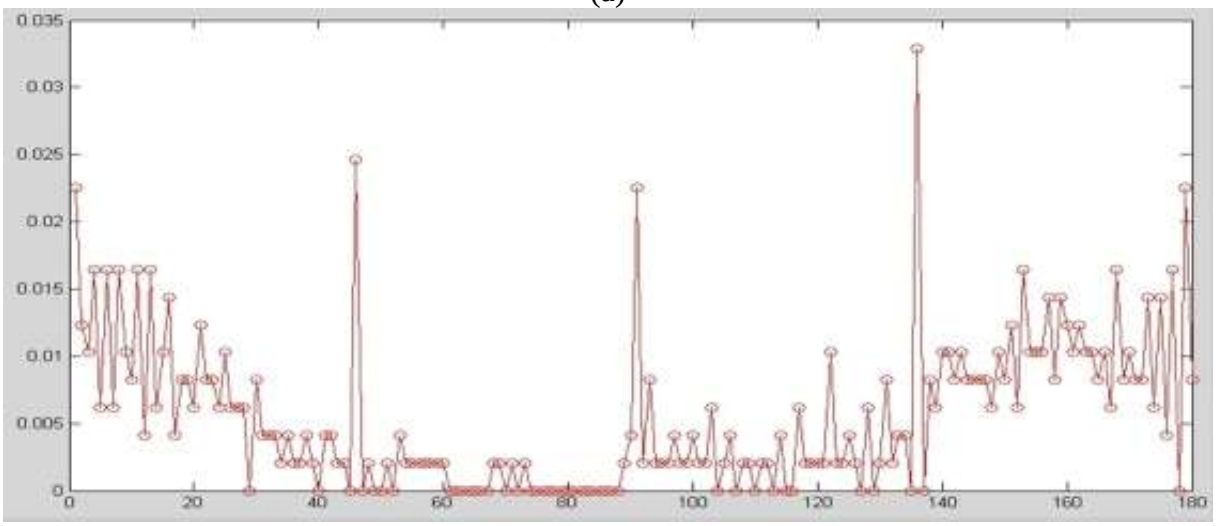
$$n = \sum_{i=1}^{180} n_i \tag{2}$$

The set / profile of all 180 peak percentages, $p = \{p_1, p_2, ..., p_{180}\}$, is used as the first feature descriptor of our work. Figure 5 shows the corresponding *peak percentage profiles* of the three Hough transform diagrams in Figure 4. From Figure 5, we can see that Figure 5(b) is basically a circularly shifted version of Figure 5(a) while Figure 5(a) and Figure 5(c) have very similar *peak percentage profiles*. Although Figure 5(c) represents the down-scaled image comparing to the image represented by Figure 5(a), the scaling operation seems to have no significant impact on these *peak percentage profiles*. From Eq. (1) we can see that p_i conveys the relationship between local (n_i) and global (n) information. The reason we use *peak percentage profile* as a feature descriptor is because in an ideal case:

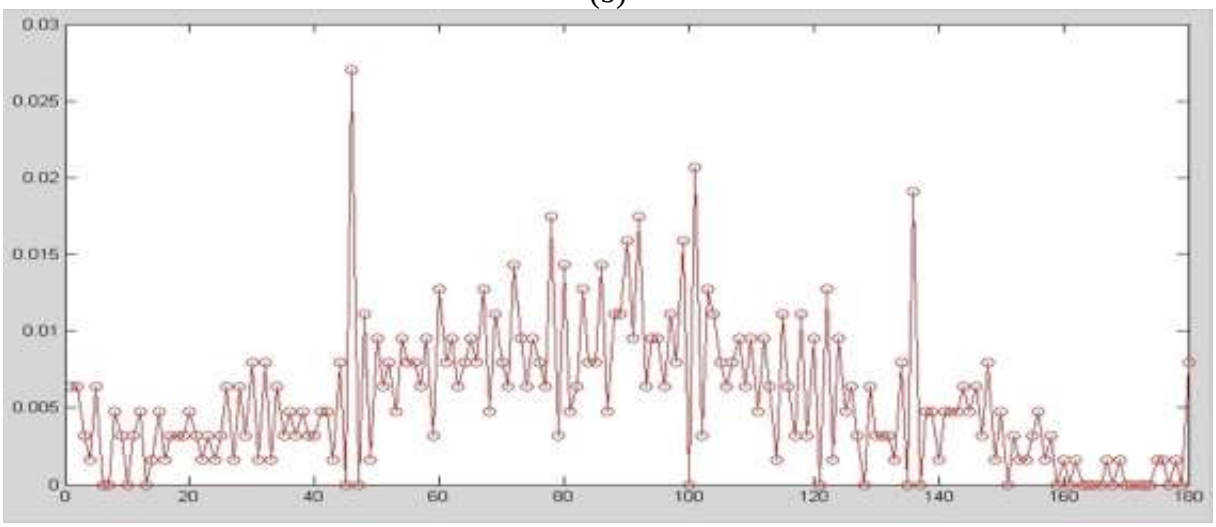
- Rotation operation does not change the value of n and the number of peaks n_i within each orientation bin apart from circularly shifting the *peak percentage profile* by the degree of rotation.
- Scaling changes the number of edge points on each line and the distances between the parallel lines; as a result, the number of points fall in each accumulator cell is different and the distribution of the points among the accumulator cell within the same orientation bin will also be changed. However, globally, the ratio of the number of points within the same bin (n_i) to the total number of point in the entire image (n) remains the same.



(a)



(b)



(c)

Figure 5. (a), (b) and (c) are the *peak percentage profiles* of Figure 4(a), (b), and (c), respectively.

We also observed that the ratio r_i of the *sum of distances* to the *centroid of distances* within each orientation bin, as defined in Eq. (3), exhibits similar invariant characteristics as peak percentage. We call this ratio *distance ratio* for short, The *distance ratio* of the *i*th bin is defined as:

$$r_i = \frac{\sqrt{\sum_{j=1}^{n_i} d_j^2}}{|C_i|} \tag{3}$$

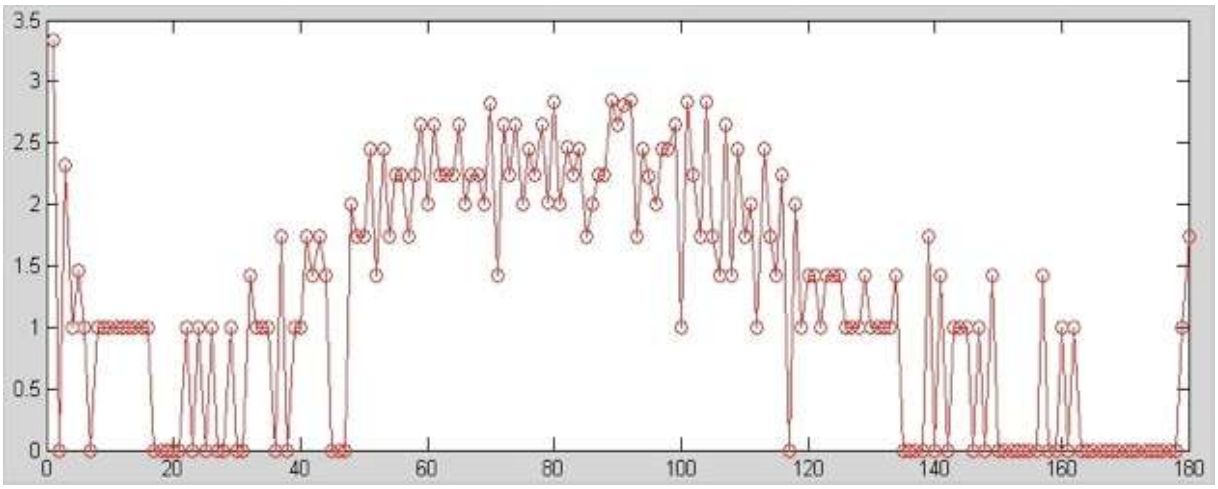
where d_j is the distance of the *j*th Hough peaks and the distance centroid C_i of the *i*th bin is defined as:

$$C_i = \frac{\sum_{j=1}^{n_i} \omega_j \cdot d_j}{\sum_{j=1}^{n_i} \omega_j} \tag{4}$$

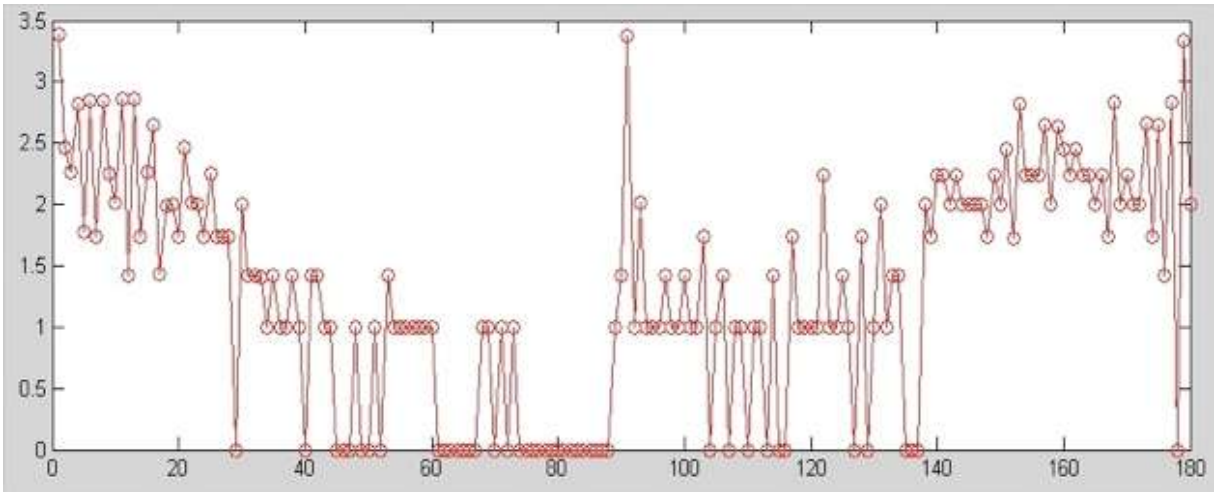
where ω_j is the number of points in the edge map that contribute to the formation of the *j*th Hough peak in the *i*th orientation bin. The set / profile of all 180 distance ratios, $r = \{r_1, r_2, ..., r_{180}\}$, is used as the second feature descriptor of our work. The *distance ratio profiles* of Figure 4(a), (b), and (c) are illustrated in Figures 6(a), (b), and (c), respectively. It is clear from Figures 6(a) and (b) that the latter is a -90˚ rotated version of the former. The high similarity between Figures 6(a) and (c) also indicates the feasibility of *distance ratio profile* in dealing with scaling operation. *Distance ratio* r_i is intended to capture the relationship between the parallel lines oriented along i˚. The algorithm of feature extraction is summarised in Table 1.

Table 1. The proposed feature extraction algorithm

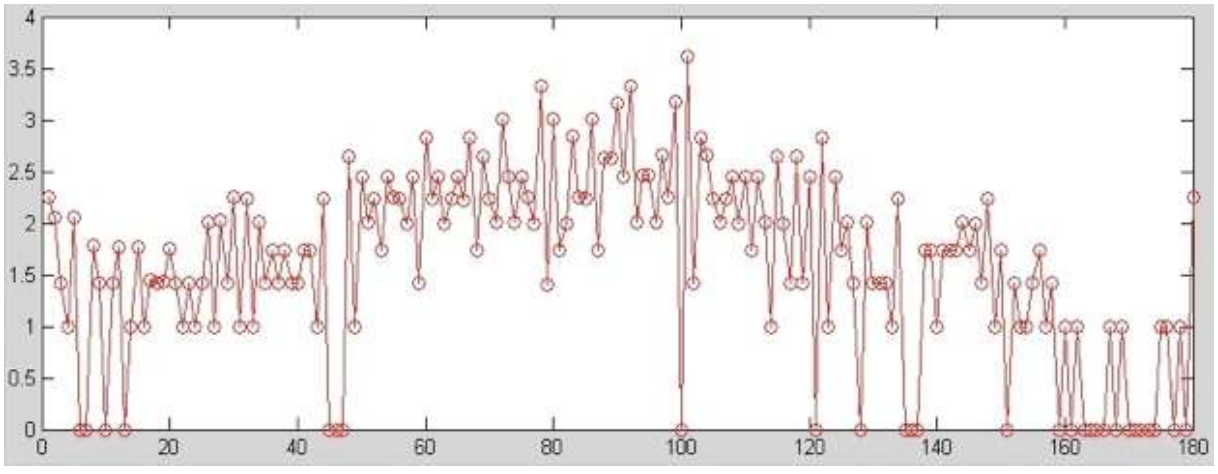
1. Smooth the input image with a 3 × 3 Gaussian kernel with a standard derivation of 0.5
2. Detect edges with the Canny operator to create an edge map
3. Thin the edges in the edge map to remove redundancy
4. Apply the Hough transform on the map of thinned edges
5. Divide the Hough transform domain into 180 bins along the θ-axis, with each bin covers 1 degree
6. Calculate the *peak percentage set p* and *distance ratio r*



(a)



(b)



(c)

Figure 6. (a), (b) and (c) are the *distance ratio profiles* of Figure 4(a), (b), and (c), respectively.

B. Image Retrieval

In the image retrieval process, the *circular correlation* is employed to measure the similarity between the peak percentage profile, p^q, of the query image, I^q, and that (p^k) of a database image, I^k. It is also the similarity measuring algorithm for the distance ratio profile, r^q, of the query image and that (r^k) of a database image. The *m*th circular correlation between p^q and p^k is defined as:

$$u^k(m) = \frac{\sum_{i=1}^{180-m}(p_i^q \cdot p_{i+m}^k) + \sum_{i=180-m+1}^{180}(p_i^q \cdot p_{i+m-180}^k)}{\sqrt{\sum_{i=1}^{180}(p_i^q)^2} \cdot \sqrt{\sum_{i=1}^{180}(p_i^k)^2}} \tag{5}$$

The maximum circular correlation $\hat{u}^k$ between p^q and p^k is:

$$\hat{u}^k = \arg\max_i \{u^k(i)\}. \tag{6}$$

The *m*th circular correlation between r^q and r^k is defined as:

$$v^k(m) = \frac{\sum_{i=1}^{180-m}(r_i^q \cdot r_{i+m}^k) + \sum_{i=180-m+1}^{180}(r_i^q \cdot r_{i+m-180}^k)}{\sqrt{\sum_{i=1}^{180}(r_i^q)^2} \cdot \sqrt{\sum_{i=1}^{180}(r_i^k)^2}} \tag{7}$$

The maximum circular correlation $\hat{v}^k$ between r^q and r^k is:

$$\hat{v}^k = \arg\max_j \{v^k(j)\}. \tag{8}$$

For a database with z images, let $\hat{u} = \{\hat{u}^1, \hat{u}^2, ..., \hat{u}^z\}$ and $\hat{v} = \{\hat{v}^1, \hat{v}^2, ..., \hat{v}^z\}$ be sets of circular correlations of the peak percentage profile and distance ratio profile between the query image and the z database images, and $rank(\hat{u}^k)$ and $rank(\hat{v}^k)$ be the ranking of $\hat{u}^k$ and $\hat{v}^k$ in $\hat{u}$ and $\hat{v}$, respectively. The final ranking $rank^k$ of the database image, I^k, is given as

$$rank(I^k) = rank(\hat{u}^k) + rank(\hat{v}^k) \quad (9)$$

The database images are then displayed / retrieved according to their rankings, with the one having the highest ranking being displayed first. The algorithm of the image retrieval process is summarised in Table 2.

Table 2. The proposed image retrieving algorithm

1. Extract feature descriptors p^q and r^q from the query image I^q using the feature extraction algorithm described in Table 1.
2. For each database image I^k, calculate the maximum circular correlations $\hat{u}^k$ and $\hat{v}^k$ as similarity measurements between I^q and I^k according to Eq. (6) and Eq. (8).
3. Calculate the ranking $rank^k$ for each database image I^k according to Eq. (9).
4. Display database images with the top-ranked first

IV. EXPERIMENTS

We have carried out a series of retrieval tests of our algorithms on a database of 953 building images captured by the authors or downloaded from the Google gallery (*http://images.google.com/*), which contains different types of buildings such as towers, churches, cathedrals, castles, pyramids, residential buildings, university departments, etc. We have also implemented an interface for the user to provide a query image and for the system to display 9 retrieved images on each page, with the most similar one displayed at the upper-left corner of the page according the raster scan order (i.e., row first). Figure 7 demonstrates one of the experiments we have carried out on our proposed system. We can see that 4 relevant images are among the 9 most similar images retrieved from the database. It is noteworthy that the highly rotated images (the first and fifth retrieved images in the raster scan order) are successfully retrieved by the proposed system. A comparative view of the performance of 6 different

shape descriptors is illustrated in Figure 8. The horizontal axis is the number of images retrieved from the database given the same query image as in Figure 7. The most similar one is the first one in the horizontal axis. The vertical axis indicates the number of relevant images among the retrieved images. For example, coordinates (10, 3) mean 3 out of the 10 retrieved images are relevant. We can see that, given this particular query image, the proposed Hough descriptors perform significantly better than moment invariants, CSS and Fourier descriptor, but marginally poorer than Zernike moments. Note that there are only 9 other images in the database that are relevant to this query image.

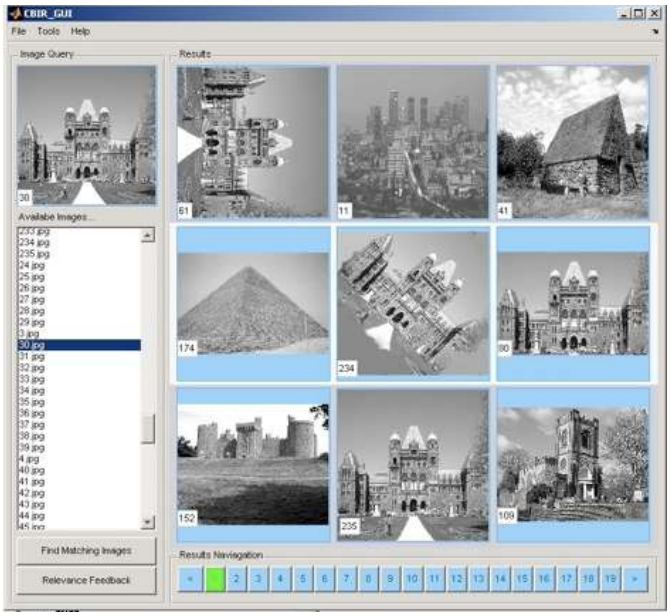


Figure 7. Retrieval results when the proposed methodology is used.

Figure 9 demonstrates the retrieval result of our system given a different query image while Figure 10 shows the performance of 5 different shape descriptors given the same new query image. We can see from Figure 10 that the proposed Hough descriptors outperform all other descriptors, including Zernike moments. Note that there are only 5 other images in the database that are relevant to this query image. The aforementioned demonstrated examples have shown the potential of the two proposed Hough descriptors.

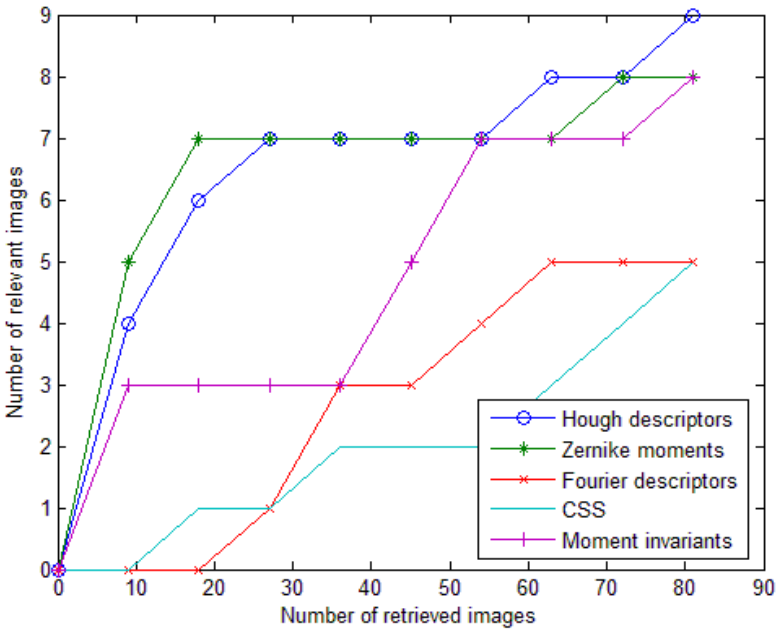


Figure 8. Retrieval performance of 5 different shape descriptors given the same query image as in Figure 7. The horizontal axis represents the number of retrieved images while the vertical axis represents the number the relevant images among the retrieved images.

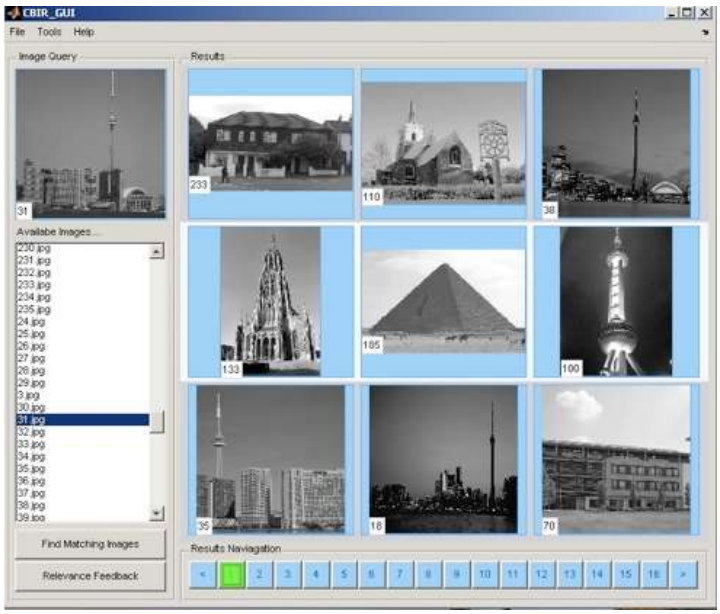


Figure 9. Retrieval results when the proposed methodology is used

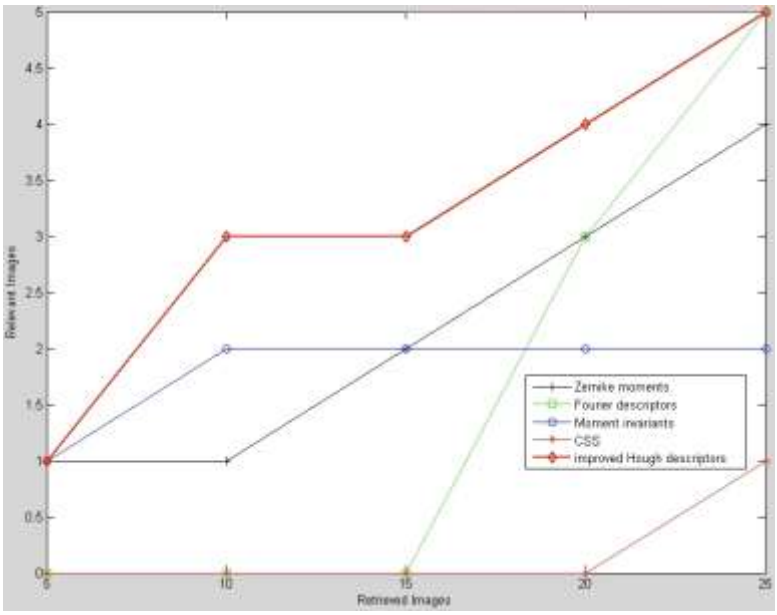


Figure 10. Retrieval performance of 5 different shape descriptors given the same query image as in Figure 9. The horizontal axis represents the number of retrieved images while the vertical axis represents the number the relevant images among the retrieved images.

From Figure 5, we observed that the *peak percentage profile* of an image is a phase-shifted version of the *peak percentage profile* of the rotated version of the same image. The same observation on the *distance ratio profiles* is also valid, as demonstrated in Figure 6. The phase displacement of the *peak percentage profile* is very close or equal to the phase displacement the *distance ratio profile*. As formulated in Eq. (6) and Eq. (8), i and j are the phase displacement needed to achieve the maximum circular correlations between the query image and a database image. Therefore, the difference between i and j should be under some threshold λ (see Eq. (10)) if a database image is to be deemed as the rotated version of the query image. Otherwise what Eq. (6) and (8) say would be inconsistent.

$$|i - j| \leq \lambda \tag{10}$$

The performance of the proposed system with the constraint of Eq. (10) applied ($\lambda = 5$ in this case) is illustrated in Figure 11. Comparing to Figure 7, wherein the constraint of Eq. (10) is not imposed, we

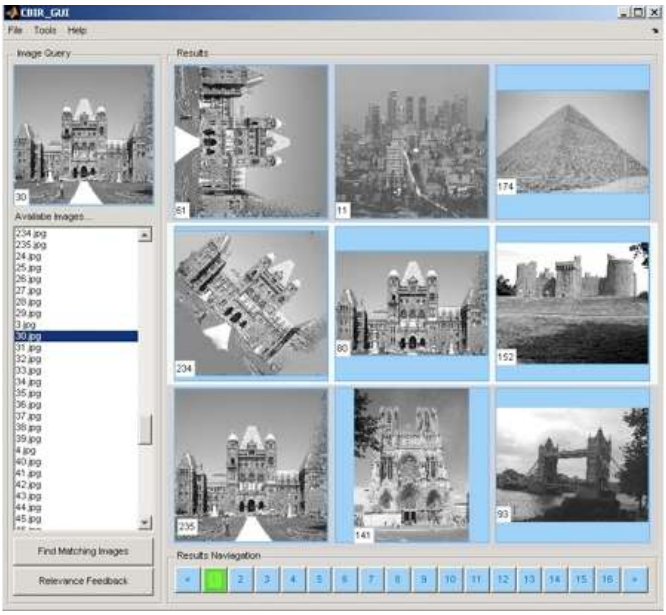


Figure 11 Retrieved results by the proposed methodology when the constraint in Eq. (10) with $\lambda = 5$ is imposed.

can see that the rankings of the four relevant images in Figure 11 have been moved up because irrelevant images with high ranking in Figure 7 have been left out.

V. CONCLUSION

In this work, two *Hough descriptors*, namely, *peak percentage profile* and *distance ratio profile* have been proposed for describing building images. The derivation of these two descriptors is based on the observation that rotation and scaling operation preserve the line parallelism and the profiles of both descriptors in the Hough transform domain, although the two profiles may be phase-shifted due to the rotation operation. Taking the phase displacement of the two descriptors into account, we have also proposed a *circular correlation matching* algorithm for measuring the similarity between the Hough descriptors of images. From our observations and experiments, we can see that the proposed Hough descriptors, combined with the circular correlation matching algorithm, outperform existing shape

descriptors in the context of building image retrieval. We have demonstrated their robustness against rotation and scaling. Currently, the robustness of the proposed CBIR system against different viewing aspects is under investigation.

REFERENCES

[1] S. Bae, A. Agarwala and F. Durand, "Computational Rephotography," *ACM Transactions on Graphics*, vol. 29, no. 3, 2010

[2] Y. D. Chun, N. C. Kim and I. H. Jang, "Content-Based Image Retrieval Using Multiresolution Color and Texture Features," *IEEE Transactions on Multimedia*, vol. 10 , no. 6, pp. 1073 – 1084, 2008

[3] I. J. Cox, M. L. Miller, T. P. Minka, T.V. Papathoman and P.V. Yianilos, "The Bayesian Image Retrieval System: Pichunter Theory, Implementation and Psychological Experiments," *IEEE Transactions Image Processing*, vol. 9, no. 1, pp. 20–37, 2000

[4]Y. Deng, B.S. Manjunath, C. Kenney, M.S. Moore and H. Shin, "An Efficient Color Representation for Image Retrieval," *IEEE Transactions on Image Processing*, vol. 10, no. 1, pp.140–147, 2001

[5] A. Dyana and S. Das, "MST-CSS (Multi-Spectro-Temporal Curvature Scale Space), a Novel Spatio-Temporal Representation for Content-Based Video Retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 8, pp. 1080 - 1094, 2010

[6] A. El-ghazal, O. Basir and S. Belkasim, "Scale Invariants of Radial Tchebichef Moments for Shape-Based Image Retrieval," in *Proc. IEEE International Symposium on Multimedia*, San Diego, USA 2009, pp. 318 - 323

[7]G. Fritz, C. Seifert, M. Kumar and L. Paletta, "Building Detection from Mobile Imagery Using Informative SIFT Descriptors", in *Proc. Scandinavian Conference on Image Analysis*, Joensuu, Finland, 2005, pp.629-638

[8]T. Geodeme and T. Tuytelaars, "Fast Wide Baseline Matching for Visual Navigation," in *Proc. International Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, 2004, pp. 24-29

[9]S. Grigorescu, N. Petkov and P. Kruizinga, "Comparison of Texture Features Based on Gabor Filters," *IEEE Trans. On Image Processing*, vol. 11, no. 10, pp.1160-1167, 2002

[10]J. W. Han and L. Guo, "A Shape-Based Image Retrieval Method Using Salient Edges," *Signal Processing: Image Communication*, vol. 18, no. 2, pp.141-156, 2003

[11] S. C. H. Hoi and M.R. Lyu, "A Multimodal and Multilevel Ranking Scheme for Large-Scale Video Retrieval," *IEEE Transactions on Multimedia*, vol. 10, no. 4, pp. 607-719, 2008

[12]Q. Iqbal and J. K. Aggarwal, "Applying Perceptual Grouping to Content-Based Image Retrieval: Building Images," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, Ft. Collins, USA 1999, pp. 42–48

[13] A. Jain, A. Vailaya and H. J. Zhang, "On Image Classification: City Image vs. Landscapes," *Pattern Recognition*, vol. 31, no. 12, pp.1921-1935, 1998

[14] W.-H. Lai and C.-T. Li, "Skin color-based face detection in colour images," in Proc. IEEE International Conference on Advanced Video and Signal based Surveillance, Sydney, Australia, 22 - 24 November, 2006

[15] C.-T. Li, "Multiresolution Image Segmentation Integrating Gibbs Sampler and Region Merging Algorithm," Signal Processing, vol. 83, no. 1, pp. 67-78, 2003

[16] R. Mukundan, S. H. Ong and P. A. Lee, "Image Analysis by Tchebichef Moments," *IEEE Transactions on Image Processing*, vol. 10, no. 9, pp. 1357-1364, 2001

[17] G. A. Papakostas, Y.S. Boutalis, D.A. Karras and B.G. Mertzios, "A New Class of Zernike Moments for Computer Vision Applications," *Information Sciences*, vol. 177, no. 13, pp. 2802-2819, 2007

[18] J. Revaud, G. Lavoue and A. Baskurt, "Improving Zernike Moments Comparison for Optimal Similarity and Rotation Angle Retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 4, pp. 627 – 636, 2009

[19] T. Sikora, "The MPEG-7 Visual Standard for Content Description - an Overview," *IEEE Transactions on Circuits Systems Video Technology*, vol. 11, no. 6, pp. 702–796, 2001

[20] D. G. Sim, H. K. Kim and R. H. Park, "Invariant Texture Retrieval Using Modified Zernike Moments," *Image and Vision Computing*, vol. 22, no. 4, pp. 331-342, 2004

[21] M. R. Teague, "Image Analysis via the General Theory of Moments," *Journal of Optical Society of America*, vol. 70, no. 8, pp. 920-930, 1980

[22] C.-Y. Wee and R. Paramesran, "On the Computational Aspects of Zernike Moments," *Image and Vision Computing*, vol. 25, no. 6, pp. 967-980, 2007.

[23] C.-H. Wei, C.-T. Li, and R. Wilson, "A Content-based Approach to Medical Image Database Retrieval," in Database Modeling for Industrial Data Management: Emerging Technologies and Applications, ed. by Z. Ma, pp. 258 - 291, Idea Group Publishing, 2006

[24] C.-H. Wei, Y. Li, W. Y. Chau and C.-T. Li, "Trademark Image Retrieval Using Synthetic Features for Describing Global Shape and Interior Structure," *Pattern Recognition*, vol. 42, no. 3, pp. 386-394, 2009

[25] X. Yuan and C.-T. Li, "CBIR Approach to Building Image Retrieval Based on Linear Edge Distribution", in *Proc. IEEE International Conference on Advanced Video and Signal based Surveillance*, Sydney, Australia, 2006, pp. 95-101

[26] D. Zhang and G. Lu, "Review of Shape Representation and Description Techniques," *Pattern Recognition*, vol. 37, no. 1, pp. 1 – 19, 2004.

[27] X. S. Zhou and T. S. Huang, "Edge-Based Structural Features for Content-Based Image Retrieval," *Pattern Recognition Letter*, vol. 22, pp. 457–468, 2001