Robust Invariant Descriptor for Symbol-Based Image Recognition and Retrieval

Alexander Wong and William Bishop
Department of Electrical and Computer Engineering
University of Waterloo
Waterloo, Ontario, Canada, N2L 3G1
{a28wong,wdbishop}@uwaterloo.ca

Abstract

This paper presents a robust invariant descriptor for symbol-based image recognition and retrieval. A modified Hough-based Transform is used to extract parameter space information (i.e., position data and angular data) from a symbol image to derive an invariant descriptor. The proposed descriptor provides a compact representation of a symbol image that can be evaluated efficiently. The extracted descriptor is highly robust against geometric transformations such as translation, rotation, reflection, and scaling, and image degradation. A series of experiments were conducted using a set of architectural and engineering symbols subjected to geometric transformations and image degradation. The experimental results clearly show that the proposed descriptor can be used effectively for symbol recognition and retrieval.

1 Introduction

Symbol-based recognition and retrieval is an important area of content-based image and video retrieval (CBIVR) where a symbol in an image or video is used to retrieve related media content containing the symbol. Symbol-based recognition and retrieval has been used in a diverse range of applications. These range from semantic drawing intepretation for electronic conversion [1, 2] and content-based trademark retrieval [3] to shape-based image database querying [4] and optical character recognition [5].

Symbol-based image recognition and retrieval is often difficult to accomplish for several reasons. First, symbols in an image or video may be subject to geometric transformations such as scaling, translation, and rotation. This is further complicated by the fact that symbols within an image or video may experience image degradation due to noise. Such geometric transformations and image degradation make it very difficult to use direct pixel-to-pixel similarity comparison techniques. Second, symbols often contain little or no color or texture information that can be used to distinguish one symbol from another. For example, circuit diagrams are typically represented as black and white line drawings. Finally, due to the large volume of images that need to be evaluated for CBIVR systems, the similarity evaluation process should be fast and use memory efficiently. The proposed algorithm aims to address these issues through the introduction of a new invariant descriptor for the purpose of symbol-based image recognition and retrieval.

The main contribution of this paper is a new Hough-based invariant descriptor for the purpose of symbol-based image recognition and retrieval. Related work on descriptors for symbol-based image recognition and retrieval is presented in Section 2. The theory underlying the proposed descriptor is described in detail in Section 3. Testing methods and test data are outlined in Section 4. Experimental results are presented and discussed in Section 5 and conclusions are drawn in Section 6.

2 Related Work

Various descriptors have been proposed for symbol-based image recognition and retrieval. Fourier descriptors are widely used for determining shape similarity [6, 7]. There are a number of issues associated with using Fourier descriptors for symbol-based image recognition. First, a closed contour is required for Fourier descriptors to work. This makes it difficult to use Fourier descriptors in situations where discontinuities occur due to noise. Second, details and shape information contained within a symbol are not represented by Fourier descriptors. This is problematic in situations where a symbol may only be distinguishable from another based on interior details. This issue is illustrated in Fig. 1. Finally, since symbols are represented by a single contour, Fourier descriptors cannot be used to effectively represent symbols consisting of multiple disjointed components. Another popular type of descriptors are those based on moment invariants [8, 9]. One issue with moment

0-7695-2997-6/07 $25.00 © 2007 IEEE
DOI 10.1109/ICSC.2007.24

invariants is that they are region-based descriptors and are therefore mainly applicable to situations where objects with filled regions exist. This is often not the case for symbols, particularly those used in engineering drawings.

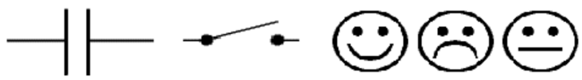

Figure 1. Examples of symbols that cannot be represented by a single contour

The proposed descriptor is closely related to Hough-based descriptors [10, 4]. These methods attempt to extract a descriptor based on angular information obtained from the Hough-space. However, there are issues with these descriptors. The first method [10] requires the selection of a fixed threshold value that depends on the actual image content. This makes it poorly suited for situations where the quality of symbol images varies greatly as is often the case for image databases. Furthermore, the descriptor is sensitive to geometric transformations such as rotation. The second method [4] improves upon the first method by making the descriptor invariant to rotation as well as removing the need for fixed thresholds. However, there are also several issues with this method. First, the descriptor is not reflection invariant. Second, the construction of the descriptor does not account for the effects of noise and degradation. The proposed descriptor addresses all of these issues.

3 Theory

This section describes the underlying theory behind the proposed descriptor. First, the theory behind the relative $|\rho|\theta$ transform used to bring the symbol image into the proposed parameter space is described, along with a discussion on its implications on translation and reflection invariance. Second, the techniques used to extract a descriptor that is invariant to rotation and scaling is explained. An analysis of the robustness of the proposed descriptor to noise and degradation is also presented.

3.1 Relative $|\rho|\theta$ Transform

The Hough transform is a common method used to provide a global description on curve features of an image such as lines and circles. In the classical Hough transform, points in a binary image specified in the xy coordinate space are mapped onto a coordinate space defined by the parameters of the curve feature being represented. For the case of line features, points in the xy coordinate space are mapped onto a $\rho\theta$ parameter space, where ρ and θ correspond to the parameters of a line as defined by the line equation:

$$x\cos(\theta) + y\sin(\theta) = p \tag{1}$$

where ρ represents the length of the normal vector going from the origin of the image (typically defined as the top-left corner of the image) to the line and θ represents the angular orientation of the normal vector with respect to the x-axis. For implementation, the $\rho\theta$ parameter space is discretized into an accumulator array. Each point in the xy coordinate space corresponds to a curve in the $\rho\theta$ parameter space while each point in the $\rho\theta$ parameter space corresponds to a line in the xy coordinate space. Therefore, the value at location (ρ, θ) in the accumulator array represents the number of points in the binary image that lie on the line defined by the parameters ρ and θ. Hence, line features can be recognized by detecting peaks in the accumulator array.

One major advantage of using the Hough transform for representing and identifying curve features is that it is highly robust against image noise and degradation. Furthermore, unlike boundary-based descriptors such as Fourier descriptors, a Hough-based descriptor allows for the representation of symbols with multiple parts and interior details.

In the proposed descriptor extraction method, a modification is made to the classical Hough transform for line representation to allow for the construction of a global structural descriptor that is invariant to both translation and reflection. In the classical Hough transform for line representation, the parameters ρ and θ of a given line are defined relative to the normal vector between the line and the origin at $(x, y) = (0, 0)$. A translation of the symbol with respect to the origin results in a different normal vector and thus a different value of ρ. One approach to making the parameter space invariant to translation is to define the parameters ρ and θ of a given line relative to the centroid of the symbol object in the image rather than the origin. Therefore, the modified line equation becomes:

$$(x - x_c)\cos(\theta) + (y - y_c)\sin(\theta) = \rho \tag{2}$$

where x_c and y_c are the x and y-coordinates of the centroid of the symbol object, respectively.

While the modified Hough transform is invariant to translation, it remains sensitive to reflection transforms. There are two factors that contribute to this sensitivity to reflection. First, the value of parameter ρ may take on either a positive and negative value depending upon the orientation of the normal vector. Therefore, a positive value of ρ may become a negative value after reflection. Second, the orientation of the normal vector may change due to the reflection, leading to a change in the value of θ.

The approach taken in the proposed algorithm is to first remove the effect of reflection on ρ by taking its absolute value. For the purpose of this paper, the resulting parameter space is referred to as the relative $|\rho|\theta$ parameter space. This parameter space is invariant to the effects of reflection and translation on ρ.

3.2 Rotation Invariance

To extract a descriptor that is invariant to rotation, the proposed algorithm projects the information from the relative $|\rho|\theta$ space onto the $|\rho|$-axis. Using this approach, the descriptor is computed as:

$$v(|\rho|) = \frac{\sum_{\theta=0}^{\pi} I(|\rho|\,,\theta)^2}{m} \tag{3}$$

where $I(|\rho|, \theta)$ is the value at the coordinate $(|\rho|, \theta)$ in the accumulator array and m is the sample mean of $v(|\rho|)$ for all values of $|\rho|$. The value of $|\rho|$ lies between 0 and the maximum distance from the centroid of the symbol object to the boundaries of the image. This compacts the $|\rho|\theta$ space in such a way that positional information is explicitly retained while angular information is implicitly represented.

An explicit representation of relative positional information was chosen for an important reason. Angular information does not provide a high level of discrimination between symbols with similar angular characteristics. Technical drawings and certain written languages may contain symbols with very similar angular properties that have different semantic interpretations. An example of this is illustrated in Fig. 2 which shows various word symbols constructed from the Korean Hangul alphabet. The symbols are very distinct from each other. However, from a structural perspective, all of the symbols consist of straight lines running in a mostly parallel or perpendicular manner to each other. This is due to the fact that the Korean Hangul alphabet is composed largely from straight lines that run mostly parallel to the x and y axes. Since there are few angular differences between the word symbols, angular-based descriptors are insufficient. On the other hand, the relative position distribution of the lines is very distinct between the word symbols and so positional-based descriptors provide good discrimination between the word symbols.

민 문 마 표 는 무 도

Figure 2. Examples of word symbols composed from the Korean Hangul alphabet

The projection of information onto the $|\rho|$-axis addresses a number of important issues. First, since angular orientation is not explicitly represented, the descriptor is invariant to the effects of rotation and reflection. The influence of angular distribution is implicitly represented by projecting the data onto the $|\rho|$-axis using a sum of squares. Finally, by compacting the 2D parameter space information into a 1D vector representation, the descriptor is more compact and can be used for fast symbol similarity evaluation.

3.3 Scale Invariance

After a 1D descriptor is obtained using projection, the descriptor is modified to be: i) scale-invariant, and ii) robust against noise and degradation. First, the upper bound of the descriptor is determined as the highest value of $|\rho|$ that contains 1% of the maximum magnitude of the descriptor. The descriptor is then truncated to the range of $|\rho|$ between 0 and the upper bound. Finally, the truncated descriptor is normalized and quantized into a fixed number of bins. After this transformation, the final descriptor is invariant to translation, scaling, rotation, and reflection. The entire transformation process is illustrated in Fig. 3.

Bin size is an important factor that influences the robustness of the descriptor. The bin size should be selected such that it is small enough to provide a high level of discrimination between symbols and large enough to be robust to the effects of noise and degradation. For the various symbol sets used during testing, an initial bin size of 50 bins achieves good results. Good symbol recognition and robustness against noise and minor geometric distortions can be achieved using this approach.

4 Testing Methods

To evaluate the effectiveness of the proposed descriptor for the purpose of symbol-based image recognition, two sets of tests were performed. The test sets are based on the content provided in the 2005 International Symbol Recognition Contest (ISRC 2005) held during GREC 2005. The set of model symbols used at ISRC 2005 consists of symbols used in architectural and engineering drawings. For the purpose of testing, the set of 25 model symbols was used. Some of the symbols from the data set are shown in Fig. 4. To evaluate the similarity between descriptors created using the proposed algorithm, a Sum of Absolute Differences (SAD) metric was used.

Using the model symbols, two 50 image test sets were generated by subjecting the model symbols to random rotation and scaling transformations, as well as various types

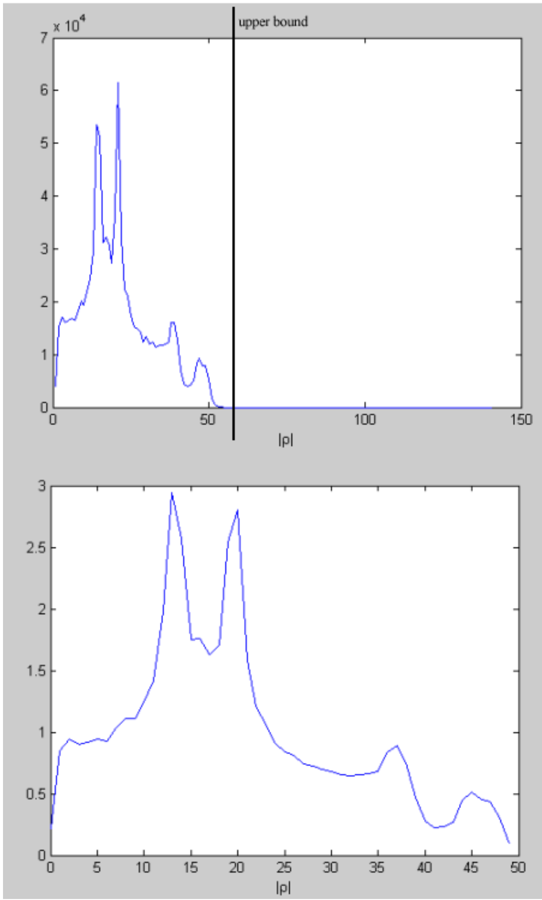

Figure 3. Transformation process: Top: descriptor truncation; Bottom: descriptor normalization and compression

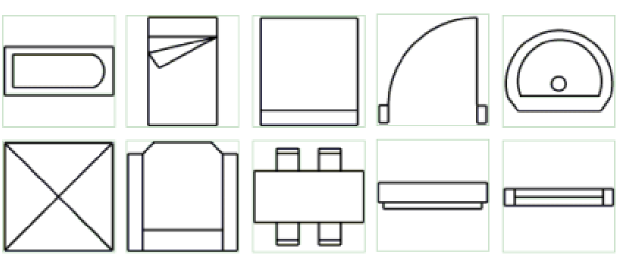

Figure 4. Example of symbols from the ISRC 2005 data set

and levels of image noise and degradation. Examples from each test set are shown in Fig. 5. The document degradation model proposed by Kanungo et al. [11] was used to introduce image noise and degradation. This model is defined by a set of 6 parameters $(\alpha_0, \alpha, \beta_0, \beta, \eta, \kappa)$. A summary of the parameters for each test set is provided below:

TEST1: $(\alpha_0, \alpha, \beta_0, \beta, \eta, \kappa)$ = (0.8, 0.5, 0.02, 0.001, 0, 1)

TEST2: $(\alpha_0, \alpha, \beta_0, \beta, \eta, \kappa)$ = (0.5, 0.5, 0.5, 0.03, 0.1, 0)

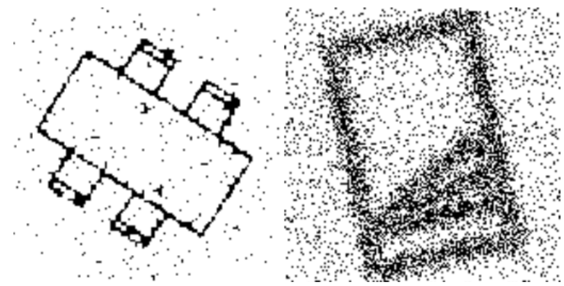

Figure 5. Test images: Left: TEST1; Right: TEST2

To evaluate the robustness of the proposed method against different levels of noise, parametric tests were conducted to evaluate the recognition accuracy of the proposed method with respect to increasing levels of replacement noise. The test data set consists of word symbols constructed from the Korean Hangul alphabet, which are randomly rotated and distorted. Examples of symbol images degraded using the different levels of replacement noise are shown in Fig. 6.

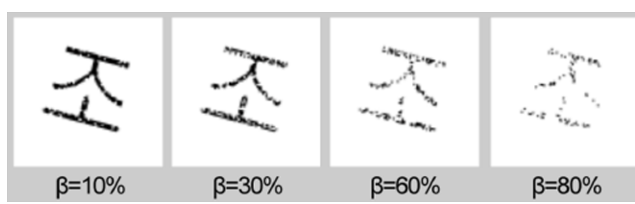

Figure 6. Example symbol under various levels of replacement noise

The proposed algorithm was evaluated against the state-of-the-art algorithm described by Vlachos et al. [4], which was demonstrated to be competitive with popular similarity metrics such as Chamfer distances and Hausdorff distances while providing significant performance improvements in the matching process. Both the proposed method and the method proposed by Vlachos et al. produce descriptors that can be matched for similarity with a computational complexity of $O(n)$.

5 Experimental Results

The symbol recognition accuracies are shown in Table 1. The proposed algorithm achieves a high level of recognition accuracy and outperforms the algorithm described by Vlachos et al. [4] on both test sets. Thus, the the proposed descriptor is robust against geometric transformations such as rotation and scaling as well as image noise and degradation.

Table 1. Symbol recognition accuracy

Test	Recognition Accuracy (%)	
	Vlachos	Proposed
TEST1	64	96
TEST2	64	92

A plot of the recognition accuracy of the proposed algorithm at different levels of replacement noise is shown in Fig. 7. The recognition accuracy of the Vlachos method decreased significantly faster than that of the proposed method as the level of replacement noise increased. These results illustrate the robustness of the proposed descriptor under non-ideal conditions where noise is prominent.

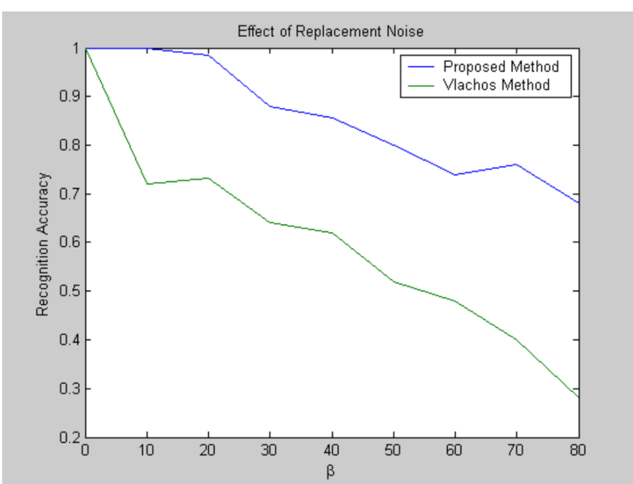

Figure 7. Effect of Replacement Noise on Recognition Accuracy

6 Conclusions

A Hough-based descriptor for symbol-based image recognition and retrieval has been introduced. The proposed descriptor is invariant to geometric transformations such as translation, rotation, scaling, and reflection. It is also robust against image noise and degradation. Good symbol recognition accuracy was achieved using the proposed descriptor on images subjected to various levels of noise and geometric transformations. It is believed that the proposed descriptor can be used effectively for high-performance symbol recognition applications such as technical drawing interpretation and content-based image and video retrieval.

References

[1] A. Filipski and R. Flandrena, "Automated conversion of engineering drawings to cad form," *Proceedings of the IEEE*, vol. 80, no. 7, pp. 1195–1209, July 1992.

[2] L. Wendling and S. Tabbone, "A new way to detect arrows in line drawings," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, , no. 7, pp. 935–941, July 2004.

[3] W. Leung and T. Chen, "Trademark retrieval using contour-skeleton stroke classification," in *IEEE International Conference on Multimedia and Expo*, Lausanne, Switzerland, August 2002.

[4] M. Vlachos, Z. Vagena, P. Yu, and V. Athitsos, "Rotation invariant indexing of shapes and line drawings," in *14th ACM Conference on Information and Knowledge Management*, Bremen, Germany, 2005.

[5] D. Shi, R. Damper, and S. Gunn, "Offline handwritten chinese character recognition by radical decomposition," *ACM Transactions on Asian Language Information Processing*, vol. 2, no. 1, pp. 27–48, 2003.

[6] I. Bartolini, P. Ciaccia, and M. Patella, "Warp: Accurate retrieval of shapes using phase of fourier descriptors and time warping distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 1, pp. 142–147, 2005.

[7] J. Cai and Z. Liu, "Hidden markov models with spectral features for 2d shape recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 12, pp. 1454–1458, 2001.

[8] D. Shen and H. Ip, "Discriminative wavelet shape descriptors for recognition of 2d patterns," *Pattern Recognition*, vol. 32, no. 2, pp. 151–166, 1999.

[9] F. Mokhtarian and S. Abbasi, "Robust recognition based on fusion of multiple shape descriptors," in *IASTED International Conference on Signal and Image Processing*, Honolulu, Hawaii, 2003, pp. 473–478.

[10] P. Franti, A. Mednonogov, V. Kyrki, and H. Klviinen, "Content-based matching of line-drawing images using the hough transform," *International Journal of Document Analysis and Recognition*, vol. 3, no. 2, pp. 117–124, 2000.

[11] T. Kanungo, R. Haralick, H. Baird, W. Stuezle, and D. Madigan, "Statistical, nonparametric methodology for document degradation model validation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1209–1223, 2000.