Performance Characterization of Shape Descriptors for Symbol Representation

Ernest Valveny[1], Salvatore Tabbone[2], Oriol Ramos[1], and Emilie Philippot[2]

[1] Computer Vision Center, Dep. Ciències de la Computació
Universitat Autònoma de Barcelona, Bellaterra, Spain
{oriolrt,ernest}@cvc.uab.es
[2] LORIA, Université Nancy 2, Nancy, France
{Antoine.Tabbone}@loria.fr

Abstract. In this paper we propose a general framework for the characterization of shape descriptors and show its application to graphic symbols. The framework is based on the combination of several performance measures independent of the application. We have applied this framework using a standard set of descriptors and databases. We show how it can be used to characterize the properties of each descriptor for a given database.

1 Introduction

There has been an increasing interest in research in performance evaluation in Graphics Recognition during the last years. Several contests have been organized in past editions of GREC Workshop concerning raster-to-vector conversion [1,2,3], arc segmentation[4] and symbol recognition[5,6]. In the particular domain of symbol recognition, a general framework of evaluation has been proposed[7] with the goal of getting a deeper understanding of the characteristics, pros and cons of various approaches to symbol recognition. The contests aim to analyze the performance of symbol recognition methods with several types of test data, including different number of symbols and several kinds of transformations and degradations. The results have been very positive as they permit to determine the robustness of participant methods under the different kinds of noise included in the test set. However, we cannot get a global understanding of different approaches to symbol recognition as only few methods (those used by the participants in the competition) were evaluated. In addition, as remarked in the conclusions of the last contest (cf. [6]), not always a detailed information about the techniques employed by each participant method is available and, therefore, we cannot have a good understanding of recognition rates according to the different types of methods.

In this paper we propose a different and more general approach for performance evaluation of methods for symbol recognition. If we take a look at them, we can observe that most of them are based on some kind of shape descriptor, as shape is the most characteristic visual feature of symbols. Indeed, the selection of a suitable shape description and representation that permits

W. Liu, J. Lladós, and J.-M. Ogier (Eds.): GREC 2007, LNCS 5046, pp. 278–287, 2008.

to capture the most relevant features of symbols is a key issue in order to obtain good recognition rates. Actually, a large number of shape descriptors have been proposed in the literature [8,9] and most of them have been applied to the problem of symbol recognition.

Then, our main goal is to propose a framework to evaluate the general performance of several shape descriptors and apply this framework to the particular case of symbol representation. This is the first difference with previous approaches to performance evaluation of symbol recognition. Instead of evaluating specific methods for symbol recognition we will evaluate general shape descriptors that can also be used for other problems in pattern recognition. The second difference is in the final goal of the evaluation framework. We do not want to focus on recognition, but we aim at characterizing the behavior of shape descriptors under several circumstances. Then, our evaluation is not only based on the recognition rate, but on the combination of several measures: recognition rate, homogeneity, separability, precision and recall.

In the experiments we have used two databases the symbol database defined for previous contests on symbol recognition, and the standard MPEG shape database. We have taken several standard and well-known shape descriptors grouped in three categories: pixel-based descriptors (Fourier-Mellin, Generic Fourier Descriptor and Zernike moments), contour-based descriptors (shape context, pixel-level constraint and string matching) and structural descriptors (graph-based).

The paper is organized as follows: first, in section 2 we explain the framework for performance characterization, mainly the evaluation measures. Then, in section 3 we describe the experiments using the selected set of descriptors and shapes. Finally, in section 4 we draw the main conclusions of this work.

2 Performance Characterization

As we have said before, our main goal is the characterization of shape descriptors, i.e, to define a kind of *genetic map* of a number of descriptors, i.e, a list of relevant and intrinsic properties for each family of descriptors. Such list of properties can help to choose the most appropriate family of descriptors given a practical pattern recognition problem.

For such a protocol to be of general use, it must be independent of datasets and must evaluate several properties of each descriptor such as the complexity, the robustness to different kinds of transformations and degradations, the power of discrimination as the number of classes and the variability of shapes grows, the genericity with very different datasets or the influence of the setting of the intrinsic parameters of the descriptor.

In this context, we cannot rely only on recognition rate as evaluation measure. Recognition rates can be very dependant on the type of classifier used. In addition, they are largely linked to one kind of practical problem, the classification of unknown shapes and cannot be the best choice to evaluate other kind of applications, such as shape retrieval.

Therefore, we need more general evaluation criteria in order to get a deep understanding of the properties of descriptors. Thus, we have decided to use 5 different measures for that: two of them (separability and homogeneity) try to be independent of the application as they intend to evaluate how well distributed are the shapes in the space of representation provided by the shape descriptor. The other three measures evaluate the performance of the descriptors in two of the most common real problems: recognition rate for the problem of classifying unknown shapes and precision/recall for shape retrieval.

All these measures rely on the computation of the matrix of distances among the representation of all the images in the shape database obtained using a given descriptor. The definition of the distance will assure that all the distances are normalized between 0 and 1 (for instance, using the Pearson correlation coefficient or the normalized euclidean distance).

- **Homogeneity:** A good description of a class of shapes should yield an homogeneous representation in the sense that the representation of all the shapes should be concentrated in a small area of the feature space. In this sense, we have defined a measure so that values close to zero mean that all the feature vectors are close (the descriptor is more homogeneous). This measure, $\mathcal{H}$ is based on the distance between elements belonging to the same class and is defined in the following way:

$$\mathcal{H} = \sum_{c=1}^{N} \frac{H(c)}{N} \tag{1}$$

$$H(c) = \frac{2h(c)}{M_c(M_c - 1)} \tag{2}$$

$$h(c) = \sum_{i=1}^{M_c} \sum_{j=1, j>i}^{M_c} \delta(v_i, v_j) \tag{3}$$

  where N is the number of classes, M_c is the number of elements in class c, v_i is the representation of element i using a particular descriptor and $\delta(v_i, v_j)$ is the distance between two elements in the feature space normalized between 0 and 1.
- **Separability:** Another property of a good shape description is that elements belonging to different classes have a dis-similar representation. Thus, we have defined a measure of separability that permits to assess this property. The farther feature vectors of elements belonging to different classes are the more separability of the descriptor. This measure, S, is based on the distance between elements belonging to different classes and defined as:

$$\mathcal{S} = \sum_{c=1}^{N} \frac{S(c)}{N} \tag{4}$$

$$S(c) = \frac{s(c)}{M_c \sum_{k=1,k\neq c}^{N}(M_k - 1)} \tag{5}$$

$$s(c) = \sum_{i=1}^{M_c} \sum_{k=1,k\neq c}^{N} \sum_{j=1}^{M_k} \delta(v_i, v_j) \tag{6}$$

- **Recognition rate:** Using the well-known 1-NN classifier, we evaluate the performance of each descriptor for recognition. This is a standard measure that can be used as a benchmark for the performance of the descriptor in recognition tasks.
- **Precision/Recall:** These measures are commonly used in the context of image retrieval and are useful to evaluate the ability of the descriptor to retrieve shapes similar to a given query shape. They can be used to evaluate the performance of the descriptor in retrieval tasks. Precision, P, measures how many retrieved shapes really correspond to the class of the query shape, while recall, R, measures the percentage of the total number of shapes belonging to the query shape actually retrieved by the descriptor. They are defined in the usual way:

$$P = \frac{N_c}{N} \tag{7}$$

$$R = \frac{N_c}{M_c} \tag{8}$$

where c is the class of the query shape, N is the total number of retrieved shapes, N_c is the number of retrieved shapes belonging to class c and M_c is the total number of shapes belonging to class c.

3 Experiments

3.1 Shape Descriptors

All these measures have been applied to evaluate a set of standard and well-known shape descriptors. As it is usually done in the literature we have distinguished between pixel-based, contour-based and structural descriptors.

Pixel-based descriptors are computed directly from the pixels of the whole image. We have used the following descriptors in this category:

- Fourier-Mellin[10]: based on the application of the Mellin and the Fourier transforms to the polar representation of the image. It is invariant to rotation and scaling.
- General Fourier Descriptor (GFD)[11]: based on the Modified Polar Fourier Transform, that applies a 2-D Fourier Transform to the polar representation of the image. The coefficients are conveniently normalized in order to achieve invariance to rotation and scaling.

– Zernike moments[12]: based on computing the projection of the image onto the Zernike polynomials and have been widely used in pattern recognition. They are also invariant to rotation and scaling.

Contour-based descriptors are obtained after extracting the outer contour of the shape. In this category we have used:

– Shape Context[13]: this descriptor is based on taking a sample of points from the contour of the shape and computing the histogram of spatial relations between a reference point and all other sample points in the contour. It is invariant to translation and scaling.
– Pixel-level constraint (PLC)[14]: it is based on the points of the skeleton of the shape. Then, taking any of these points as reference the ratio of angle and length between any other pair of points can be computed and then, the histogram of these ratios is obtained. The histograms obtained taking every point in the skeleton as reference point can be grouped in two matrices, one for angular information and the other one for the length information, that are processed to obtain the final descriptor, that is rotation and scale invariant.
– String matching[15]: based on the representation of the contour as a chain code and applying and edit distance to compute the similarity between chain code of two different shapes.

Structural descriptors are based on representing relationships between components of the shape, normally using graphs or grammars. In our case, we have used a graph representation where nodes correspond to junction points or end points and edges correspond to the lines joining these points. From this graph representation a signature is computed assigning to each node a value based on the number of incident edges and the angle and relative length between them.

3.2 Shape Databases

All these descriptors have been applied to two shape databases: the database of graphic symbols defined for the first contest on symbol recognition at GREC' 2003 [5] and the MPEG-7 contour database.

The GREC database is composed of 50 graphic symbols composed of straight lines and arcs of circumference. The original database generated for the contest contained images with 3 kinds of transformations: geometric transformations (rotation and scaling), binary degradations and vectorial distortions. In our experiments we have used two subsets of images from the original database:

– GREC-50: in this set we have images of the 50 original symbols with rotation, scaling and slight binary degradations (see figure 1(a))
– GREC-Vec: in this set we have included images with vectorial distortion generated by randomly moving junction and end points, but keeping line connectivity as required by the graph-based descriptor. In this case we have only used the 26 symbols composed only of straight lines. In figure 1(b) we can see some examples of the kind of distortions that have been generated.

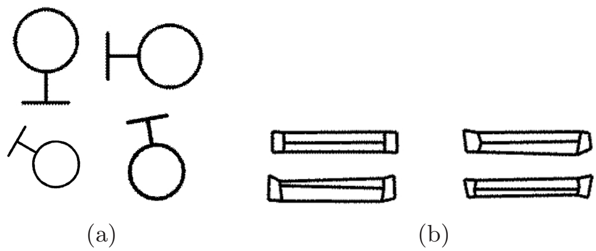

Fig. 1. Example of symbols of GREC database. (a) symbols in the set GREC-50. (b) Vectorial distortion applied to symbols included in the set GREC-Vec.

However, in this kind of images, contour-based descriptors do not perform well. In addition, one of the goals of the proposed protocol for the characterization of descriptors was to test the genericity with different datasets. Then, in order to obtain more general results and to be able to better compare pixel-based and contour-based descriptors we have also used images of the MPEG-7 shape database (some examples can be seen in figure 3.2). For this database we have defined 4 subsets:

- MPEG-99: composed of 99 images belonging to 9 different classes.
- MPEG-216: contains images of 18 classes, 12 images per each class.
- MPEG-1045: 1045 images belonging to 42 classes. There are between 3 and 60 images per class.
- MPEG-Occ: this set is the same as MPEG-99, but we have applied a method to generate random partial occlusions of the contour as can be seen in figure 3.2

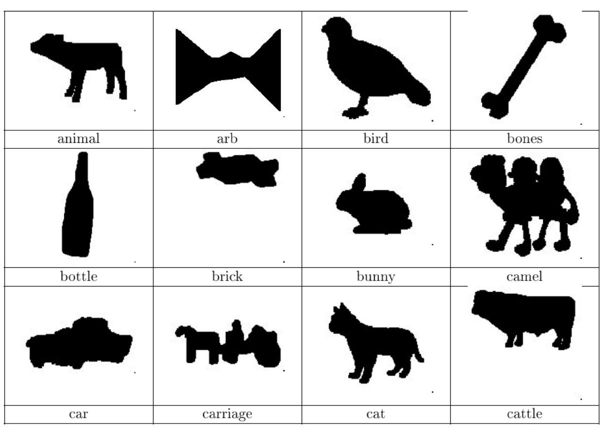

Fig. 2. Examples of the shapes included in the MPEG-7 shape database

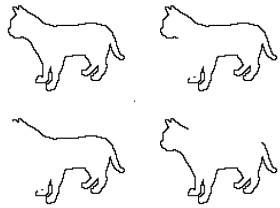

Fig. 3. Examples of the shapes with partial occlusions of the contour

3.3 Analysis of Results

Not all the set of descriptors described in section 3.1 could be applied to all the sets of images explained in section 3.2 due to the properties of each descriptor. Then, in the summary table 1 we can see which descriptors have been applied to each database. Mainly, pixel-based descriptors were applied to both databases, contour-based descriptors only to the MPEG-7 database and structural descriptors only to the GREC database.

Then, once obtained the representation of all the shapes with every descriptor, we computed the distance matrix among all elements and all the evaluation measures: homogeneity, separability, recognition rate and precision/recall. Due to space availability we cannot show the detailed results for all descriptors and databases, but the analysis of these results permit to state some conclusions about the performance of the descriptors, both from a global point of view and from the particular point of view of every descriptor.

If we analyze the results globally, considering all descriptors and databases we can say:

– GFD and Zernike moments have always obtained the best recognition rates. In addition, recognition rates are better for the GREC database than for the MPEG database. This seems logical as shapes in the GREC database have less shape variability.
– In general GFD also gives the best value for homogeneity. For contour-based descriptors, PLC is the descriptor with the best homogeneity. The graph-based descriptor has also a very good homogeneity in the only case where it is used.

Table 1. Summary of the performance of descriptors

	GREC-50	GREC-Vec	MPEG-99	MPEG-216	MPEG-1045	MPEG-Occ
Fourier-Mellin	- - -	+	+	+	-	- - -
GFD	++	++	+++	++	++	+
Zernike	++	++	++	++	+	++
Shape Context			- -	-	- - -	+
PLC			+	++	++	+
String matching			-	-	- -	
Graph		+				

– Zernike moments have always the best separability measure, although they present one of the worst homogeneity values.
– There are some descriptors (GFD, Zernike) with better precision than recall. That means that they prioritize retrieving exact shapes than all meaningful shapes relevant to a given query. On the contrary, other descriptors (Fourier-Mellin, PLC and Shape context) with better recall than precision are better in order to retrieve all relevant shapes although some of them do not correspond to the query. On the other hand, graph-based descriptors and string matching reach a good compromise between precision and recall in the sense that they are able to retrieve a good number of relevant shapes while keeping low the number of non-relevant answers.

Beside these global conclusions we can also state some interesting conclusions about every particular descriptor:

– GFD has a good separability in long and thick shapes and in shapes with curves.
– Fourier-Melllin gives better separability in closed shapes and better homogeneity in long and thick shapes.
– Shape context has better separability in occluded shapes. The recognition rates are, in general, low.
– Pixel-level constraint obtains good recognition rates and good separability in long and thick shapes.
– String matching gives better separability when there are significant changes of direction in the contour of the shape.
– Graph-based descriptors do not really have a good recognition rate. They have difficulty in separating objects with the same structure, but they have good separability for shapes with large number of lines.

We have summarized these conclusions in Table 1 where for every set of images we show how positive (+,++,+++) or negative (-,- -,- - -) are the results obtained for each descriptor. The evaluation of each descriptor is based on the analysis of the results of the recognition rate, the homogeneity and the separability. The precision and recall have not been taken into account. This analysis does not intend to be a rigorous, formal and exact evaluation of the descriptors. However it permits to establish a kind of tendency for each descriptor and can help to choose a descriptor for a given application. For instance, it can be observed that GFD, Zernike and PLC always obtain a positive evaluation, although the "best" descriptor varies depending on the dataset.

4 Conclusions

In this paper we have proposed to use several performance measures for the characterization of shape descriptors. The combination of these measures permits to have a better understanding of the behavior of each descriptor than using single classical indices such as the recognition rate, precision or recall, that are more oriented to specific tasks. We illustrate the usefulness of this approach by analyzing a set of standard shape descriptors using a database of graphic symbols and a database of contour shapes. From the analysis of the results obtained with the proposed evaluation measures we are able to state several conclusions that characterize the performance of the descriptors for each database permitting to summarize it in a table.

References

1. Phillips, I., Chhabra, A.: Empirical performance evaluation of graphics recognition systems. IEEE Transactions on Pattern Analyisis and Machine Intelligence 21(9), 849–870 (1999)
2. Chhabra, A., Phillips, I.: The second international graphics recognition contest - raster to vector conversion: A report. In: Tombre, K., Chhabra, A.K. (eds.) GREC 1997. LNCS, vol. 1389, pp. 390–410. Springer, Heidelberg (1998)
3. Chhabra, A., Philips, I.: Performance evaluation of line drawing recognition systems. In: Proceedings of 15th. International Conference on Pattern Recognition, Barcelona, Spain, September 2000, vol. 4 (2000)
4. Wenyin, L., Zhai, J., Dori, D.: Extended summary of the arc segmentation contest. In: Blostein, D., Kwon, Y. (eds.) GREC 2001. LNCS, vol. 2390, pp. 343–349. Springer, Heidelberg (2002)
5. Valveny, E., Dosch, P.: Symbol recognition contest: A synthesis. In: Lladós, J., Kwon, Y.-B. (eds.) GREC 2003. LNCS, vol. 3088, pp. 368–385. Springer, Heidelberg (2004)
6. Dosch, P., Valveny, E.: Report on the second symbol recognition contest. In: Liu, W., Lladós, J. (eds.) GREC 2005. LNCS, vol. 3926, pp. 381–397. Springer, Heidelberg (2006)
7. Valveny, E., Dosch, P., Winstanley, A., Zhou, Y., Yang, S., Yan, L., Wenyin, L., Elliman, D., Delalandre, M., Trupin, E., Adam, S., Ogier, J.: A general framework for the evaluation of symbol recognition methods. International Journal on Document Analysis and Recognition 9(1), 59–74 (2007)
8. Zhang, D., Lu, G.: Review of shape representation and description techniques. Pattern Recognition 37, 1–19 (2004)
9. da F.R. Costa, L., Cesar Jr., R.M.: Shape analysis and classification. CRC Press, Boca Raton (2001)
10. Adam, S., Ogier, J.M., Cariou, C., Mullot, R., Gardes, J., Lecourturier, Y.: Utilisation de la transformé de fourier-mellin pour la reconnaissance de formes multiorientées et multi-échelles: application l'analyse automatique de documents technique. Traitement du signal 18(1), 17–33 (2001)
11. Zhang, D., Lu, G.: Shape-based image retrieval using general fourier descriptor. Image Communication 17
12. Kim, H., Kim, J., Sim, D., Oh, D.: A modified zernike moment shape descriptor invariant to translation, rotation and scale for similarity-based image retrieval. In: IEEE International Conference On Multimedia and Expo.,

13. Mori, G., Belongie, S., Malik, J.: Eficient shape matching using shape context. IEEE Transactions on Pattern Analyisis and Machine Intelligence 27(11), 1832–1837 (2005)
14. Yang, S.: Symbol recognition via statistical integration of pixel-level constraint histograms: A new descriptor. IEEE Transactions on Pattern Analyisis and Machine Intelligence 27(2), 278–281 (2005)
15. Tsay, Y., Tsai, W.: Model-guided attributed string matching by split-and-merge for shape recognition. International Journal of Pattern Recognition and Artificial Intelligence (1988)