A Region-based Object Recognition Algorithm

Paulo Sérgio Rodrigues Arnaldo de Albuquerque Araújo

Departamento de Ciência da Computação
DCC - ICEx - UFMG. Caixa Postal 702.
30161-970. Belo Horizonte, MG - BRAZIL
{pssr,arnaldo}@dcc.ufmg.br

Abstract. This paper presents a new algorithm for recognition of objects in a given image. This algorithm transforms the input image into a directed graph which is constructed through several defined rules. The graph characteristics represent the global shape information of the object inside the input image, and are extracted during the graph contruction. This strategy prevents the postprocessing traversing to the graph, with consequent improvement of computational time. The algorithm was tested over a specific data base, and the experiments were conducted to show its performance in the light of two types of problem: object class recognition and similar image retrieval.

1 Introduction

The word "recognition" produces controversy itself. What can be recognized as being one determined object by someone can be interpreted as being something completely different for another one [22]. Moreover, according to the context, one person can give a lot of interpretations to the same image [2, 10, 16]. Nevertheless, there are many works which deal with the task of automatic recognition of objects and scenes [1, 5, 9, 13, 23]. Since the 60's, this word has been widely used in the literature to mean interpretation, classification, cognition and another jobs which are inherently tasks of a human being. As areas like pysichology, neuroscience, analysis of algorithms and computational vision improve, it is known that the task of automatic object recognition, similar to the biological visual system, is still far to be reached, and so much work must be demanded.

Currently, there are only specialist systems with specific objectives and, generaly, in most of the cases, fail in recognizing objects in a general way, even if they are considering only their proper universe of performance [4, 7]. For this task, diverse techniques have appeared over the last three decades [21, 20, 15, 17, 18, 11, 8, 14, 19]. However, the majority of them is a combination of some of few others [22]. Certainly, the most explored are those which use information from the shape for recognition. This fact is also due to the belief that, even so the biological visual system can use different parameters (e.g: color, shape, texture, movement, context, sound, etc.) to reach its objectives, the most used parameter is the shape. Shape can be defined simply as being the external edges, internal ones, disposal of ones in relation to the others, or some set of regions that compose the scene.

Many algorihms works over a unique region, considering only external edges and, rarely the internal edges. This means that objects which are composed by far than one region are not considered. On the other hand, the majority of the recognition methods needs a module of preprocessing before the final step, the recognition itself. Many of these modules include a segmentation phase; and it is known that to segment an image is a cognitive process, too. For this reason, the most popular algorithms currently still produce a deficient output in many cases. A typical example happens when the output of the segmentation module is, not only a unique region which represents the object, but a set of small regions which should be conected. However, these regions keep a global structure which can be used for the later step of recognition.

In light of the above, this work introduces a new algorithm for object recognition, based on the global structure of the objects. In a general way, the algorithm tries to capture global structures of objects and scenes that, normaly, are not captured by traditional algorithms. This algorithm, called *GRAS (Graph Region Arrow Shot)*, gets the characteristics from a set of regions generated in a preliminar phase of preprocessing. This name was chosen because the regions are reached in a way where the neighboring regions are linked (shot) by directed arcs (arrows).

To show the validity of the algorithm, the concept of object view is presented, where the algorithm searchs for similar images that are in different perspectives of the query.

This paper is organized as follows. In Sec. 2 we define some basic concepts which will be used along all this paper. In Sec. 3 we describe the basic preprocessing steps which will be used over the image before it is inputted in the algorithm; in Sec. 4 we describe the proposed algorithm; in Sec. 5 we present the features used by the algorithm; in Sec 6 are presented the experimental results; and finally, some conclusions are given in Sec. 7.

2 Basic concepts

This section presents some definitions and concepts which are used All over the text. Some of them have been already used in the literature.

After the preprocessing image, all pixel with 0 value is set to backgroung. A set of connected pixels with the same no zero value is an individual region; a set of regions is an object; and a set of objects is a scene.

An important concept used in this work is the notion of view. This is the same idea as the perspective in the field of geometry. Then, in this work, each different perspective of an object is called a "view" of the object. Informally speaking, the set of different views of an object represents it in different perspectives.

Fig. 1 shows this idea: 16 images are distributed in four lines of four columns. Each line represents four different views of the same object. For example, in the first line, it is found four views of a pen; the second line shows four views of an eraser; the third line shows four views of a scissor; and the last line shows four views of the object vosg.

Figure 1: An example of 4 different objects, each one with 4 distinct views showed at the same line. The first line presents 4 views of a *pen*, the second line, an *eraser*, the third line, a *scissor* and the fourth line, a *vosg*.

Finally, the last concept to be defined is "recognition". Although this word has been frequently used, in this paper, it is necessary to estabilish its meaning for our purposes. A set of regions, which represents an object, is transformed into a *distance matrix* for the centroids of the regions, and is the input to the GRAS algorithm, which produces, as output, a feature vector based on shape characteristics (see Section 5). Then, for objects which belong to the same class, the word recognition means to produce similar feature vectors. Also, it should generate feature vectors as different as possible, if the objects are from different classes.

3 Preprocessing

First, we can across how the image is inputted to the algorithm. The input data (which represents the input image) to GRAS is an $n \ x \ n$ matrix, constructed from the object regions, which is obtained segmenting the grayscale image. Although the goal of this article is the presentation of the GRAS, we suggest, as an early stage, an edge detection algorithm such as the Sobel operator [6] followed by simple algorithm to reduce noise, and morphologic operations to define salient regions. The preprocessing simplicity is to show, later, that the algorithm supplies satisfactory results, even with a simple preprocessing of the images.

Thus, the used sequence over the images with an unique object and with homogeneous background is the following: Sobel operator application; small regions elimination; and "dilation" morphologic operation with a cross-structured element.

These operations aim to eliminate small regions which may have been originated by factors as aquisition noise, shade, reflectance, etc. Also, they aim to enhance important regions, basic requirement of the algorithm.

Fig. 2 presents an original image of the object sciss, and Fig. 3 presents the result after the operations above have been applied. The resultant regions are labeled with different integer values and their centroids, as well as their distances, are calculated. The *distance matrix*, which represents the distances of the centroids between themselves, is the GRAS's data structure inputted.

4 The GRAS Algorithm

Before the execution of the GRAS, for each object region, the centroids are calculated, and each one of them is labeled with a distinct integer value in an up to down sequence until the buttom region is reached. Fig. 4 shows an atificial example. Thus, the distances between all the centroids are calculated to compose the *distance matrix.* This matrix is the input data to the GRAS. As output, the GRAS generates a feature vector, which complains the characteristics defined in

Figure 2: Original image of the object (*sciss*).

Section 5.

The global idea is traversing the regions linking their centroids, in an up to down way, constructing a directed graph. Having traversed these regions, several shape characteristics are extracted. These characteristics represent global information about the shape of the object, and can not be extracted if the processing is based only locally, as internal edges.

Let i, j $(i < j)$ represent two values of some labeled region, from the set of objec regions. The algorithm uses as input the *distances matrix* for the centroids, and runs based on several rules, which are enumerated below.

The idea of proximity is subjective. So, in this work, the parameter d, used down, is a threshold for it. Then, it was taken as the average number of all distances, and everytime the distance between the centroid on analysis and one of its neighbors exceed d, the latter is not considered by the algorithm; otherwise, it taken in account.

1. if $i = 1$ and there is a centroid j at a distance less than or equal to d, apply rule 2, otherwise, apply rule 3;

2. if a centroid $i \neq j$ is at a distance less than or equal to d from another centroid j, a directed arc is created from i to j. It is said that i shot j by

Figure 3: Obtained regions by the suggested preprocesing chain applied to the object *sciss*) (Fig. 2).

threshold;

3. if a centroid i does not have any other centroid at a distance less than or equal to d, and j is the nearest centroid to i, then, a directed arc from i to j is created. It is said that i shot j by proximity.
4. if a centroid $i \neq 1$ was not shot yet, it is shot by the nearest centroid of the set of centroids which were already shot by another one, according to rules 2 and 3;
5. a region j can not be shot by a region i if i has already shot j.
6. if a centroid i has the greatest label, it does not shoot any one.

The details of the GRAS algorithm can be best described by using a practical example. Consider the set of regions presented in Fig. 4a. Each region was labeled with an integer value from 1 to 15, from left to right and from up to down. Also, let the set of points of in Fig. 4b represent the centroids of each region from Fig. 4a.

GRAS starts from the centroid with the smallest value, 1. In Fig. 4, to simulate that a centroid is at a distance less than d from another centroid, they are linked by light dotted lines; and, to simulate that this

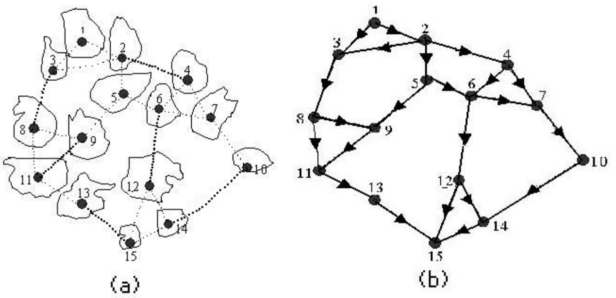

Figure 4: A set of regions of an object (a); connected graph after applying the GRAS algorithm (b).

distance is greater than d, they are linked by dark dotted lines. Thus, in the example above, and according to the specified rules, 1 shot 2 and 3 based on threshold (rules 1 and 2). After that, GRAS continues from centroid 2. The centroids which are at a distance less than or equal to (d) from 2 are 1, 3 and 5. According to rule 5, 2 can not shoot 1, but by rule 2, it can shot 3 and 5 by threshold. For the same rules, 3 can not shoot 1 and 2, but can shoot 8 by proximity. Continuing with the given example: centroid 4, due to its position is shot by 2 (rule 4) and shot 6 and 7 by threshold (rule 2); centroid 5 shot 6 and 9 by threshold (rule 2); 6 shot 7 by threshold (rule 2); 7 shot 10 for proximity (rule 3); 11 shot 13 by threshold (rule 2); centroid 12, as in the case of centroid 4, is shot by proximity by 6 (rule 4) and shot 14 and 15 by threshold (rule 2); centroid 13 shot 15 by proximity (rule 3); centroid 14 shot 15 by threshold (rule 2); and, finally, 15 can not shoot any region (rule 6).

5 Used features

This section presents the characteristics which are used by the proposed algorithm. These characteristics were chosen due to the fact that they capture the global object shape. However, these characteristics are not unique, and it is possible to define and extract others. Despite of this, for the purpose of this work, they are enough.

- **Number of Regions (NR)**: The majority of the most popular algorithms image segmentation, usually, generates as output, not a unique region, which delimits the target object, but a set of small regions that may be separeted one of the others. Fig 3 shows an example. The number of regions can be taken as being a global characteristic of the object.

- **Greater/Minor path (GP/MP)**: According with [3], the longest path between two nodes p and q of a graph is the major number of the nodes which are walked in this path. Similarly, the minor path between these nodes is the minimum number of them.

  For the proposed work, the major/minor path is a characteristic taken from the object regions. The nodes p and q are taken as being the extremes of this path, the major and minor values, respectively.

- **Number of Forks (NF)**: it is the number of nodes of a directed graph which has input degree of at least two.

- **Númber of Splashes (NS)**: it is the number of nodes of a directed graph which has output degree of at least two.

- **Númber of cicles (NC)**: it is the number of cicles of a graph according to the definition of [3].

- **Global Average Axis (GAA)**: each region of an object has a major axis, Em_i, and a minor axis, En_i, transversal to Em_i. Global Average Axis (GAA) is the sum of all Em_i's divided by all En_i's.

  Number of edges (NE): is the number of edges of a graph.

During the GRAS's traversing, some characteristics, that represent globally the shape of the object, can be extracted. However, for the goals of this work, the use of the characteristics just defined is enough. In the example presented in Fig. 4, these characteristics are represented by the following values: NR = 15; GP = 8 (1,2,4,6,7,10,14,15); MP = 5 (1,2,5,6,12,15); NF = 7 (1,2,4,5,6,8,12); NS = 8 (3,6,7,9,11,14,15); NC = 8; NE = 22; GAA = $(E_{m1} + E_{m2} + ... + E_{m15})/(E_{n1} + E_{n2} + .. + E_{n15})$.

6 Experiments

Consider some object i. The proposed algorithm can be used, when receiving as input data a *distance matrix* that represents i, to answer two types of specific questions: (a) what is the object i? and (b) what images, from the data base, are the most similar to i? These two questions can be applied to systems with different goals. Type (a) is necessary when it is desired to know what is a specific object that appears in a given image, as in an intelligence artificial system. Type (b) is a typical task for an image retrieval system.

To evaluate the algorithm under these two points of view, two types of experiments, one for each type of question, were taken in account. The first one measures the degree of certainty to answer the question (a) when the number of queries varies; and the second measures the precision of retrieval when the amount of relevante images is varied. The latter tries to measure the GRAS performance under question (b).

In all experiments, a unique set of images was used. This set is composed of 8 subsets, each one with circa 300 to 600 views of the following objects: eraser, pen, scissor, sharpener, staple, tape, triangle and vosg, at a total of 2000 distinct images approximately. This is the *Prima Web Server* data base, available in [12].

In this paper, the set of 2000 images is called S, and each one of the 8 subsets are defined as follows: E = eraser, P = pen, Sc = sciss, Sh = sharp, St = staple, Tp = tape, Tr = triangle, e V = vosg.

To speed up the process, the queries were done, not directly in S, but in a set B, constructed from S. Then, to construct B, the maximum number of views was fixed for each object. In the case of these experiments, 10 views of each one of the 8 cited subsets were used, in a total of 80 images.

The goal in comparing only 80 images from B with 2000 images from S is to show that, even under restricted condictions, the proposed algorithm worked well. Thus, given an image i of some object, gotten

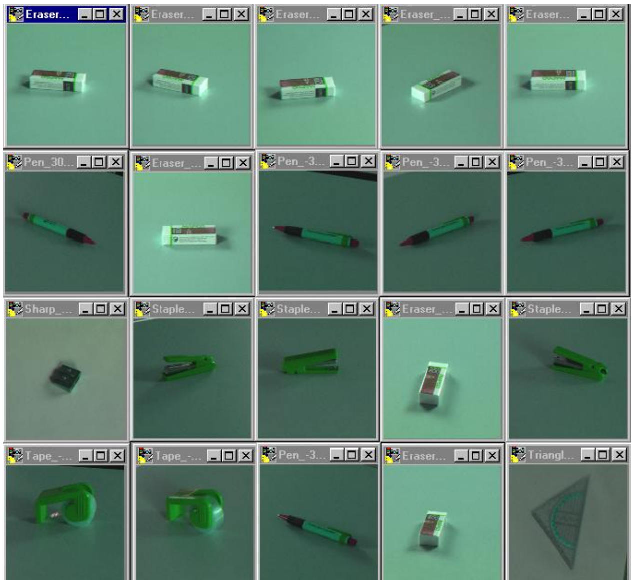

Figure 5: Example of query for an object of the type eraser. The upper lefter image represents the query. In this case, another four types of objects (pen, staple, tape and triangle) were retrieved, too, eventhough with a low similarity value.

randomlly from some subset from S, a query of the type (a) is taken by classifying i as being an image from the the subset B, cited above; and a query of the type (b), ranks a subset of images from B which are most similar to the image of the object i. Na example for the search of the type (b) can be seen in Fig. 5. In this figure, the upper and lefter image is the query, and the remainder are the responses listed and ranked by similarity,in an up-down and left-right manner. Note that the biggest number of possible views of the same type of object, which is used as query, have higher similarity.

The similarity measure used in all the experiments was the Euclidian distance between the feature vectors of the two images to be compared.

To evaluate the algorithm in the light of question (a), graphics *number of searches X % success* were contructed; in Fig. 6 an example for an object of the type E is shown. This specific graph was constructed repeating the query for groups of 1,2,3,...94 random images from E, the subset of S which represents eraser;

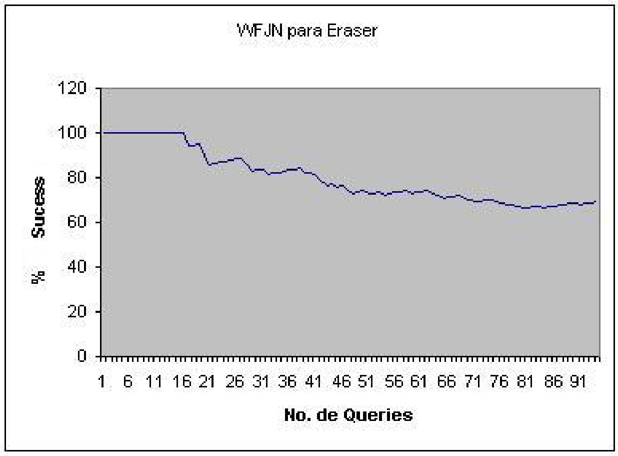

Figure 6: GRAS performance when the object to be searched is of type E, and the question is of type (a).

and, for each group, it was taken the total percentage of the queries, where the most similar images were some view of eraser.

As an example of interpretation of the graph, let take the point which the number of queries is 28, which corresponds to 86% of succes. This means that, when a query was repeated for 28 different images from E, in about 24 times (86%), the object recovered with the higher similarity was classified, accurately, as being an object of the type E.

Thus, the worse performance for the object of the type E occured when the query was repetead for 94 images.

To evaluate the GRAS in the light of question (b), graphics of the type *precision x recall* were constructed as in the example shown in Fig. 7. This type of graph represents, in the horizontal axis, the recall (the ratio between the number of images which the algorihtm must retrieve and the number of images of objects really retrieved); and in the vertical axis, precision (the degree of the similarity achieved). The biggest the number of relevante images (in this case, erasers) with high similarity achieved, the greatest the precision.

As an example of interpretation of this graphic, let the point whose recall is 50%. This means that, when the algorithm achieved a recall of 50% of the relevante images of eraser, they appeared with similarity degree

Figure 7: GRAS performance when the object to be searched is of type E, and the question is of type (b).

(precision) of 75%, in average.

To construct the graphic in Fig. 7, the experiment was repeated 94 times, and the arithmetic average was taken from the results.

The same experiment presented in Fig. 6 was repeated for each one of the 8 subsets of images of the mentioned objects. For comparison, the results were all ploted in the same graph, as can be seen in Fig. 8.

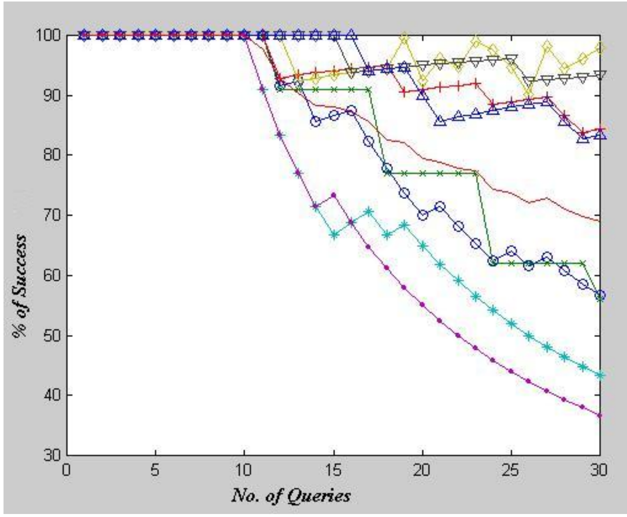

Figure 8: Performance of GRAS for each object: (△) = eraser; (○) = pen; (×)= sciss; (+) = sharp; (∗) = staple; (•) = tape; (⋄) = triangle; (∇) = vosg. The solide line represents the average of the curves, then, it gives an idea of the general behavior of the algorithm.

7 Conclusions

We proposed a new algorithm which can be used to retrieve images of objects in a large image data base, shuch as Content Based Image Retrieval, as well as, to recognize specific objects, such as in intelligence artificial systems. This algotithm, which we called GRAS (Graph Region Arrow Shot), obtained good results even using a simple preprocessing of the images. Due to this GRAS has a low amount of computational time. This is an important characteristic of GRAS. This algorithm traverse the object regions and extract characteristics that represent the object shape global features. This strategy can extract shape information that may not be extracted by algorithms which use local processing only. In the *precision* x *recall* graphic it is possible to evaluate GRAS performance if it is requested to recover all possible images from the data base which are similar to the input query; and in the *numer of queries* x *%success* graphic it is possible to evaluate GRAS performance if it is requested to recover the most similar images from the data base in regard to the input query. For both types of experiments, GRAS was runned comparing a small number of images, as queries, with the whole data base, as target images. Although this is an ongoing work, the results were encouraging.

Acknowledgements

The authors are grateful to CNPq, CAPES and SIAM DCC/PRONEX Project for the financial support of this work.

References

[1] Vlatko Becanovic. Image object classification using saccadic search, spatio-temporal patterns encoding and self-organization. *Pttern Recognition Letters*, 21:453–263, 2000.

[2] M. M. Chun. Contextual cueing of visual attention. *Trends in Cognitive Sciences*, 4(5):170–177, 2000.

[3] T. H. Cormem, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms*. McGraw-Hill, 1997.

[4] J. P. Eakins and M. E. Graham. Content-based image retrieval. Available online at http://www.unn.ac.uk/iidr/CBIR/report.html, October 1999. JISC Technology Applications Programme Report 39.

[5] J. M. Fuertes, M. Lucena, N. P. de la Banca, and J. Chamorro-Martnez. A scheme of colour image retrieval from databases. *Pttern Recognition Letters*, 22:323–337, 2001.

[6] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison Wesley, 1992.

[7] V. N. Gudivada and V. V. Raghavan. Content-based image retrieval. *IEEE Computer*, 28(9):18–22, 1995.

[8] V. N. Gudivada and V. V. Raghavan. Design and evaluation of algorithms for image retrieval by spatial similarity. *ACM Transaction on Information Systems*, 13(2):115–144, 1995.

[9] A. Guri-Dugu and A. Oliva. Classification of scene photographs from local orientation features. *Pttern Recognition Letters*, 21:1135–1140, 2000.

[10] J. M. Henderson and A. Hollingworth. High-level scene perception. *Annual Reviwer of Psychology*, 50:243–271, 1999.

[11] A. K. Jain and A. Vailaya. Image retrieval using color and shape. *Pattern Recognition*, 29(8):1233–1244, 1996.

[12] Gravi laboratories. Prima Web Server Project. Technical report, INRIA - Rhene-Alpes, 2001. htp://www-prima.inrialpes.fr/index.php.

[13] D. Lee and H. Kim. A fast content-based indexing and retrieval technique by the shape information in large image database. *The Journal of Systems and Software*, 56:165–182, 2001.

[14] K. C. Liang and C. C. J. Kuo. Implementation and performance evaluation of a progressive image retrieval system. In I. K. Sethi and R. C. Jain, editors, *Storage and Retrieval for Image and Video Databases*, volume VI, pages 37–48. SPIE 3312, 1998.

[15] F. Liu and R. W. Picard. Periodicity directionality and randommess: Wold features for images modelling and retrieval. *IEEE Transaction on Pattern Analises and Machine Intelligence*, 18(7):722–733, 1996.

[16] N. H. Mackworth and A. J. Morandi. The gaze selects informative details within pictures. *Perceptual Psychophys*, 2:547–552, 1967.

[17] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of large image data. *IEEE Transactions on Pattern Analises and Machine Intelligence*, 18:837–842, 1996.

[18] R. Mehrotra and J. E. Gary. Similar shape retrieval in shape data menagement. *IEEE Computer*, 28(9):57–62, semtember 1995.

[19] S. Ravela and R. Manmatha. Retrieving images by appearance. In *Proceedings of IEEE International Conference on Computer Vision (ICCV98)*, pages 608–613, Bombay, India, 1998.

[20] J. R. Smith and S. F. Chang. *Querying by color regions using the VisualSEEk content-based visual query system*, pages 23–41. AAAI Press, Menlo Park, CA, 1997. Intelligent Multimedia Information Retrieval.

[21] M. J. Swain and D. H. Ballard. Color indexing. *International Journal on Computer Vision*, 7(1):11–32, 1991.

[22] S. Ullman. *High-level vision*. Massachusetts Institute of Technology, 1996.

[23] X. S. Zhou. Edge-based structural features for content-based image retrieval. *Pttern Recognition Letters*, 22:457–468, 2001.