Improving Shape-Based CBIR for Natural Image Content Using a Modified GFD

Yupeng Li[1], Matthew J. Kyan[2], and Ling Guan[1]

[1] Multimedia Research Laboratory, Ryerson University, Toronto, ON Canada
[2] School of Electrical & Information Engineering, University of Sydney, NSW Australia
{yli, mkyan, lguan}@ee.ryerson.ca

Abstract. We present a modified version of the Generic Fourier Descriptor (GFD) that operates on edge information within natural images from the COREL image database for the purpose of shape-based image retrieval. By incorporating an edge-texture characterization (ETC) measure, we reduce the complexity inherent in oversensitive edge maps typical of most gradient-based detectors that otherwise tend to contaminate the shape feature description. We find that the proposed techniques not only improve overall retrieval in terms of shape, but more importantly, provide for a more accurate similarity ranking of retrieved results, demonstrating greater consideration for dominant internal and external shape details.

1 Introduction

In Content Based Image Retrieval (CBIR), shape information is widely considered to play a key role in the characterization of scenes. Shape of itself however, is a difficult property to measure, and much work has been directed toward this effort. Recent descriptors proposed for capturing shape information, generally fall into two categories: contour-based methods (such as the popular fourier descriptor and its variants) and region-based methods (such as Zernike moments, Hu's geometric moments, etc). Due to the difficulty in measuring and assessing shape properties of natural images, the majority of CBIR work reported using shape information alone, tends to focus on simplified shape databases such as those of binary logos [1] and iconified/trademark graphics. In such collections, scenes often involve only a single object with a well defined shape, wherein a single class is often comprised of a set of images with only minor variations to the dominant shape, that whilst altered, remain well defined. In collections involving natural scenes however, such as those found in the well known COREL database, scenes are generally much more complex, involving many combinations of objects, of a variety of shapes and sizes that may or may not be embedded in equally complex backgrounds. Often shape information becomes contaminated by the mixture of content in a scene, rendering shape based retrieval results, relatively poor. In this work, we propose a modified region-based technique to better deal with CBIR applications in the domain of natural image databases.

One of the major problems in attempting to use the contour-based standard Fourier descriptor (FD) as a feature for assessing the similarity between images based on shape, lies in its dependence on the prior knowledge of boundary information. In

M. Kamel and A. Campilho (Eds.): ICIAR 2005, LNCS 3656, pp. 593 – 600, 2005.

particular, it assumes that for each image, we have an ordered description of the points that form the connected path responsible for a particular boundary. As such, when considering the boundary of regions of interest, some form of higher level segmentation becomes necessary. Unfortunately, even if such segmentation is available, it is often the case that multiple boundaries will occur within the image (either due to internal shape content or multiple regions of interest). In previous work [2], the FD has been applied to the description of edge information of natural images found in the COREL database, and has met with limited success in terms of shape identification.

In other work focusing on shape alone, Zernike moment descriptors (ZMD) have been proposed as a preferred technique over other region based techniques such as geometric moments (e.g. Hu's moments) [3]. Derived from a complex set of orthogonal polynomials over the unit disk, a more rotationally invariant description of shape information is achieved, independent of boundary information. Limitations exist in terms of computational complexity and a tendency to capture spatial moments in the radial directions rather than spectral features, thus spectral information is not captured evenly at each order resulting in loss of significant features useful for shape description [4].

As an alternative, a region-based 2D polar Fourier transform (PFT) attempts to better capture the spectral content of angular & radial information by transforming the polar description of an image into a rectangular image of radial vs. angular distribution of image intensities, upon which a standard 2D FT may be applied. This approach has been demonstrated to maintain rotational invariance in that a rotated shape generally yields a similar spectral definition. A generalized version of this technique (Generic Fourier Descriptor - GFD) was proposed in [4],[5]. Translation invariance was achieved by choosing the origin for polar space to be the centroid of the shape in question, thus all radial & angular content is then calculated from this origin. Scale invariance was achieved by normalizing the coefficients of each spectral component in the PFT (i.e. one for each combination of radius & angle, or each point in the rectangular mapped image of radius vs. angle), by its DC component, whilst the DC component itself was normalized by the mass or area over which the polar image was taken. This met with good results in simplified binary shape databases, although no real application to natural image data has been reported.

2 Modified Edge-Based GFD

In this current work, unlike that of [5], we consider the Canny edge description of natural image queries as input to a GFD inspired operator for shape description. Direct application of [5] might see a binary image formed from the original (either by thresholding or similar), such that a 'caricature' of the original could be utilized as input to the GFD. The problem with this is that the computation becomes quite extensive as more pixels need to be considered in the shape image. In addition, achieving a consistent thresholding for a natural image is not trivial as it is very sensitive to contrast, etc. By using an edge description we reduce the computational load (less pixels to consider in the polar mapping).

The GFD operator, as proposed by [5], involves first finding the polar description of the input edge image (mapped into a normal rectangular image format of radius vs. angle) with a 2D FT applied to this transformed image. The modified 2D FT is calculated as follows:

$$PFT(\rho,\phi) = \sum_r \sum_i f(r,\theta_i)\exp[j2\pi(\frac{r}{R}\rho + \frac{2\pi i}{T}\phi)] \ , \tag{1}$$

where (r,θ) denote the polar coordinates in the image plane and (ρ,ϕ) denote the polar coordinates in the frequency plane. $0 \le r = [(x-x_c)^2 + (y-y_c)^2]^{1/2} < R$ and $\theta_i = i(2\pi/T)$; $(x_{c,} y_c)$ define the center of the shape; $0 \le \rho < R$, $0 \le \phi < T$, where R and T are the radial and angular resolutions. To get the GFD, we then normalize the PFT calculation as described previously, for more details see [4].

3 Ignoring Textured Regions Using ETC

Quite often, excessive textured regions contaminate the edge description of dominant objects in a scene. To remove the contaminating effect of over textured regions, we employ a further refinement in the shape image prior to extracting the shape description, by attempting to establish (and thus ignore) Canny responses resulting more from texture rather than more dominant edges. In this way, we attempt to supply a set of edges that better reflect the regional boundaries within an image rather than every intensity variation. The edge-texture characterization (ETC) approach introduced in [6] provides a fuzzy discrimination between edge & textured regions and is adopted in this work.

The principal of ETC is founded in examining the changes in variance occurring in a windowed local region when it is blurred by an averaging filter. In smooth regions the variance does not really change, however the response in textured versus edge images is quite marked. This is exploited in the ETC measure. Based on the size of the local window considered, the ratio $k = \overline{\sigma}/\overline{\sigma}'$ between the standard deviation $\overline{\sigma}$ of original versus $\overline{\sigma}'$ of blurred intensities, yields a measure of deviation due to the underlying nature of the image content in that region. A simple range of values captured by this measure can be attributed to a textured region, thus we can establish a regional mask over the textured regions so that they may be later ignored in the shape descriptor calculation. The equations of $\overline{\sigma}$ and $\overline{\sigma}'$ are defined as:

$$\overline{\sigma}^2 = \frac{1}{|N|}\sum_{(i,j)\in N}(x_{i,j} - \overline{x})^2 \tag{2}$$

$$\overline{\sigma}'^2 = \frac{1}{|N|}\sum_{(i,j)\in N}(x'_{i,j} - \overline{x}')^2 \tag{3}$$

In the equation (2) and (3), N denotes a neighborhood set around the current pixel, $x_{i,j}$ denotes the gray level value of pixel (i, j) in the set, and $x'_{i,j}$ is the corresponding smoothed gray level value under 5 X 5 averaging. $\bar{x}$ and $\bar{x}'$ are the mean of the gray level values of the original and smoothed variables in the neighborhood set. Most of the estimated k values are restricted to the interval [0 5]. Experimental results show that the textured area of most images is located in the interval [2.1 5]. In order to extract the textured area from the image, we apply a morphological operation (erosion and dilation) to deal with the texture map extracted by ETC measurement and form a complete texture area, then we can mask and eliminate the texture area from the edge map extracted by Canny filter.

Fig. 1. (a) Original image; (b) Edge map; (c) Edge map (texture removed)

From fig.1(b), we see edge map of original image (Lenna) fig.1(a), extracted by the Canny filter, the edge map shows us that the feather texture of her hat produces an oversensitive edge response complicating the overall shape, fig.1(c) is the edge map after removing the textured part found with ETC and shows us the clear edge shape information.

4 Experimental Results

In order to test the retrieval performance of our proposed algorithm, we select three different shape descriptors: the standard Fourier Descriptor (FD), Modified Generic Fourier Descriptor (MGFD) and MGFD after removing texture part of edge map. Our simulations were carried out using a subset of the COREL image database consisting of 1000 natural color images (JPEG), from 10 classes that appeared to be more dominated by shape. Each class included 100 conceptually similar images.

Simulations were conducted to compare the retrieval effectiveness on this database when indexed with one of three alternative shape descriptors. The first used the lowest 50 coefficients from a standard FD (denoted FD) as a feature vector for each edge mapped image. The second feature vector is a set of 36 coefficients of the GFD, calculated for an equally distributed set of 4 radii and 9 angles (denoted MGFD). The third feature vector is the same as the second, however the GFD calculation is performed on the texture removed edge map of each image (denoted MGFD1). Each image in the database was then indexed with each of the three different feature vectors. In retrieval, similarity was measured using the Euclidean distance between the feature vector of a query image, and those of all other images in the database.

To measure general retrieval performance, statistical results were calculated by considering 10 different query images from each class (forming 100 queries in total). For each query, the first 16 most similar images were retrieved to evaluate the performance of the retrieval. The table 1 shows us the Retrieval Rate (the percentage of images in the 16 retrieved, belonging to the same class as the query image).

Table 1. Retrieval performance from three different sets of shape features

Class	1	2	3	4	5	6	7	8	9	10	Average
FD(%)	17.9	21.0	22.3	18.6	41.2	23.8	41.2	63.7	32.5	20.0	30.2
MGFD(%)	30.5	19.0	23.0	52.5	31.3	21.6	45.0	51.2	60.0	68.0	40.2
MGFD1(%)	31.6	22.5	22.0	62.5	30.0	22.0	43.0	55.0	68.7	69.5	42.6

Table 1 tells us that the retrieval results of most classes using the proposed MGFD and MGFD1 outperform results using FD. There were, however, a few classes demonstrating similar, but slightly worse performance. Such classes exhibit a much higher variation in shape between images considered of the same class conceptually, thus many images from different classes that have similar shape distribution are often confused. This generally highlights the limitations in using this performance measure to evaluate shape based results. This being said, the proposed MGFD1 method gives much better representation in the classes in which the images shape structure is more consistent across images, a factor especially evident in classes 4, 9 and 10, reflecting sets of rock formation, flag and aircraft images respectively.

Although the retrieval rate improves about 12 percent using MGFD1 over the standard FD approach. In most classes, the retrieval rate is similar to the edge mapped application of GFD (MGFD), with slight overall improvement. To effectively gauge each descriptor more intuitively, we look at some explicit visual results, and offer a more subjective view of their relative success, in terms of the shapes of images retrieved (regardless of class), and their ranking in terms of similarity to the query.

The left top first image in each of the following figures is the query image we selected. The order of similarity ranking is from left to right, top to bottom. In Fig. 2(a), FD only retrieves 3 flag images, whilst in Fig 2(b) The whole boundary shape of the flags (roughly rectangular due to the flags waving) as well as the internal Union Jack feature becomes significant. MGFD not only extracts a more accurate exterior boundary feature than FD, but also considers interior shape features, retrieving 8 flag images with a strong feature in the top left of the flag. In Fig 2(c), some of the more textured details in the flags are eliminated by the ETC consideration, thus only dominant edges (both internal and external) contribute to the search and improve the retrieval result. The proposed MGFD1 approach has the best performance, evident not only in a higher retrieval rate (10 flag images), but more importantly, in the set of Union Jack based flags dominating the *most* similar images, as opposed to the scattered flags found by the MGFD. This reflects a greater accuracy in the ranking of similar result images by the MGFD1.

Fig. 2. Flag query – CLASS 9 (a) FD top; (b) MGFD middle; (c) MGFD1 bottom

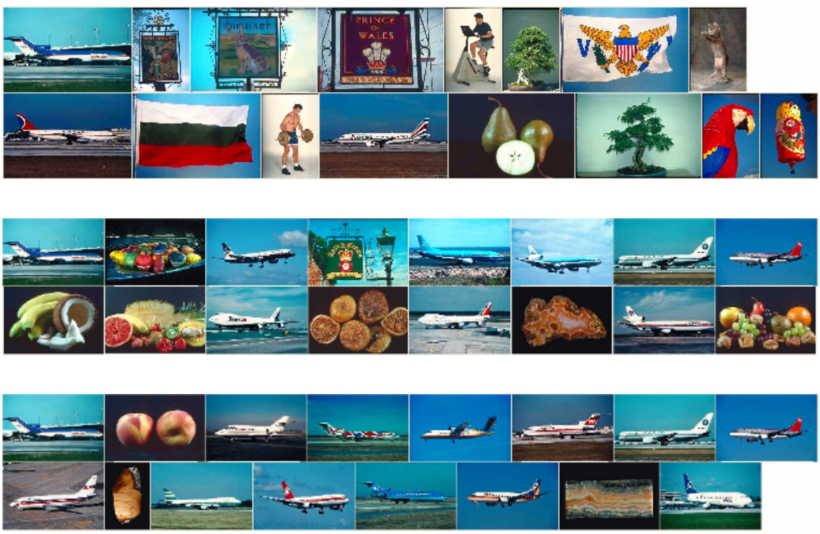

Fig. 3. Aircraft query – CLASS 10 (a) FD top; (b) MGFD middle; (c) MGFD1 bottom

In Fig. 3(a) FD only retrieves 3 aircraft images. Like the flags, the aircraft images exhibit some regularity in terms of shape, although this isn't captured effectively by FD. In Fig. 3(b) MGFD retrieves 9 aircraft images, but due to the influence of internal texture part, the fruit images (the 9th, 10th, 12th and 16th images) are retrieved falsely. Their apparent similarity may be in that the fruit is distributed in an elongated manner, yet internal textures are erratic and confuse the similarity matching.

In Fig. 3(c) MGFD1 removes the texture 'contamination' (the grass and internal part of the plane) and retrieves 13 aircraft images. Note also that the rock contour in the 15th retrieved image, where the shape somehow is similar to the straight body of the plane, with the kinking tail protruding upward at one end of the plane. This is similar to the contour in the rock formation.

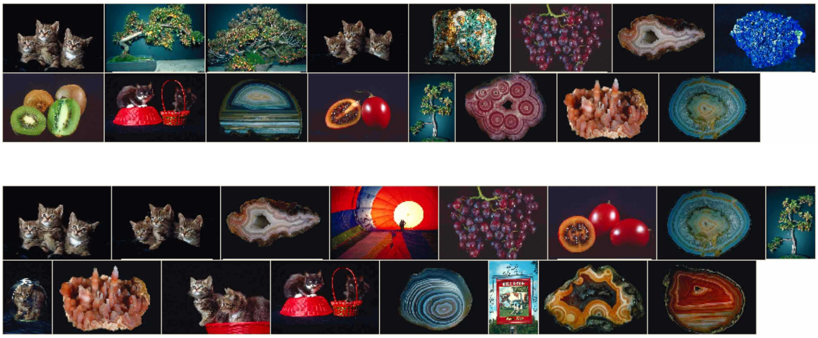

Fig. 4. Cat query – CLASS 3 (a) MGFD top; (b) MGFD1 bottom

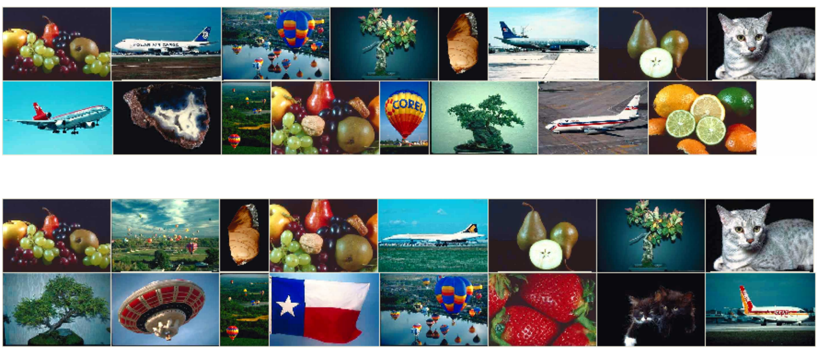

Fig. 5. Fruit query – CLASS 1 (a) MGFD top; (b) MGFD1 bottom

In Fig. 4 the concept of a cat does not necessarily coincide with a consistent silhouette, in fact, in this class different numbers of cats may exist in some images. MGFD retrieves 3 cat images. The 4th result, with 3 kittens is similar to the query. Likewise, the distribution in the tree images (2nd and 3rd), fruit (9th) and rock formations (7th, 15th) are more similar in terms of overall shape than the images with one cat. In the MGFD1 result, not only are 5 cat images retrieved, but the rank of the 3 kittens image is improved from 4th to 2nd, (i.e. it is considered *most* similar to the query), as opposed to the MGFD result. In other queries from class 3 (not shown), MGFD occasionally yields a higher retrieval rate than MGFD1 (hence a slightly higher performance in Table 1). This is misleading however, as closer inspection reveals that images from the correct class may be retrieved by the MGFD method yet have a very different shape from the query. Also, some of the class images with different shape are often ranked as more similar to the query image than those intuitively closer in shape (using MGFD). The experimental results demonstrated that the MGFD1 approach achieves a better result as a shape feature. This effect should mean that if MGFD1 is combined with other features (colour, etc), we expect that the other features will help capture more images from the same class, allowing MGFD1 to sift out and rank the captured set more accurately.

In the same way, the fruit query of Fig. 5 (FD omitted), shows that the clustered balloons are considered to be similar to the clustered fruit objects. The MGFD1 result however, is still more accurate (same fruit ranked 4th for MGFD1, 12th for MGFD).

5 Conclusions

In this paper, we have proposed a modified generic Fourier descriptor (MGFD1) for image retrieval. Comparing with GFD, where the authors extracted the shape feature from the whole shape image in the MPEG-7 region shape database, if GFD is applied on natural image instead of binary trademark images, the computation will be very expensive because of high resolution and complicated shape information in natural images. GFD is not suitable for natural image retrieval. Our proposed MGFD1 overcame this drawback.

The PFT was applied to the Canny edge maps of images, thereby decreasing computational complexity. The MGFD1 shape feature improved average image retrieval rate (except where shape varied dramatically). In such cases however, the similarity rankings were more intuitive. It was argued that in the cases where retrieval rate was lower than that of MGFD, the ultimate performance of the system still showed improvement as images from the correct class, yet with different shape to the query, were rejected or ranked lower in terms of similarity to the query – reflecting an order that more closely fit the notion of shape. In the future, we will apply the MGFD1 method both in combination with other successful low level features to further improve overall retrieval rate for natural image databases, and in retrieval based on the more specific notion of ROI (Region of Interest), where it is expected that overall retrieval rate as well as accuracy will be improved.

References

[1] A. Folkers and H. Samet, "Content-based image retrieval using Fourier descriptors on a logo database". *Proc. the 16th International Conference on Pattern Recognition.* Vol. 3, pp.521-524, Aug. 2002, Quebec City, Canada

[2] K.Jarrah, P. Muneesawang, I.Lee and L.Guan, "Minimizing human-machine interactions in automatic image retrieval". *Proc. Canadian Conference on Electrical and Computer Engineering*. Vol. 3, pp. 1589-1592 May 2004, Niagra Falls, Canada

[3] D. S. Zhang and G. Lu, "A comparative study of three region shape descriptors" DICTA 2002: *Digital Image Computing Techniques and Applications*, Jan. 2002, Melbourne, Australia

[4] D. S. Zhang, "Image retrieval based on shape". Chapter 5, 6 *PhD Thesis*, 2002.

[5] D. S. Zhang and G. Lu "Generic Fourier Descriptor for Shape-based Image Retrieval", *Proceedings of IEEE Int. Conf. On Multimedia and Expo*. Vol.1, pp. 425-428, Aug. 2002.

[6] H. S. Wong and L.Guan, "A neural learning approach for adaptive image restoration using a fuzzy model-based network architecture". *IEEE Transaction on Neural Networks*. Vol. 12, pp. 516-531. May 2001