Near-Duplicate Image Recognition and Content-based Image Retrieval using Adaptive Hierarchical Geometric Centroids

Mai Yang[1], Guoping Qiu[1], Jiwu Huang[2] and Dave Elliman[1]

[1]School of Computer Science and Information Technology, The University of Nottingham, UK
[2]School of Information Science and Technology, Sun Yat-sun University, Guangzhou, China

Abstract

In this paper, we present a new feature extraction method that simultaneously captures the global and local characteristics of an image by adaptively computing hierarchical geometric centroids of the image. We show that these hierarchical centroids have some very interesting properties such as illumination invariant and insensitive to scaling. We have applied the method for near-duplicate image recognition and for content-based image retrieval. We present experimental results to show that our method works effectively in both applications.

1. Introduction

Image matching is important and has broad applications in pattern recognition and machine intelligence. In this paper, we present a new image matching method for content-based image retrieval (CBIR) and for near duplicate image recognition. CBIR has been a popular research topic for many years and can have important application in managing large image and video databases [3, 4, 5]. With the ease of image distribution made possible by the success of the Internet, issues associated with copy-right infringement and content pirating have become increasingly important, near-duplicate image recognition can be used as an alternative to traditional watermarking for copy-righted image content protection [1, 2].

An essential step in image matching is feature extraction. To a large extend, the quality of the features used for image matching determines the matching performance. In this paper, we present a method for computing hierarchical image features suitable for content-based image retrieval and for near-duplicate image recognition.

2. Related Work

Since feature extraction is such an important aspect in many applications of pattern recognition and computer vision, there is a very large body of literature on this topic. Here we give a very brief review of the most related literature. In the computer vision community, researchers have developed various local image features that are invariant to rotation, scaling and various transformations; see e.g. [6, 7]. In image indexing and retrieval literature, researchers have developed various image content descriptors for comparing image contents, including local and global color and texture features, see e.g. [3, 4, 5].

Both local features and global feature have their weaknesses. Using local features for image matching necessarily requires more complicated feature matching methods such as voting [8] to rank the database images similarity, where the query image's features are first individually compared (based on some distance, e.g., L_2) with features from the database images and then followed by a verification step to account for spatial or geometric relationships between the features. Matching local feature individually not only fails to capture the co-occurrences of the local features, which can convey much useful information of the images, it is also computationally very expensive. Global features, such as histogram-based image descriptors [3], only capture the overall distribution of the image features but not their spatial locations. For example, images can have the same color histogram but have completely different spatial color distribution patterns. Desirable image features should simultaneously capture both the local and global characteristics of the image contents so that simple distance measures such as L_1 or L_2 can be directly used to match the images. In this paper, we present a method for extracting hierarchical invariant features that captures global and local invariant image characteristics for image retrieval and for near duplicate image recognition.

3. Geometric Centroid of Image

An image $I(x,y)$ can be regarded as a two-dimensional planar lamina, i.e., a two-dimensional planar closed surface Ω which has a surface density $I(x,y)$. The "mass" of the image M can be defined as

$$M = \iint_{\Omega} I(x,y)dxdy \tag{1}$$

The coordinates of the centroid (also called the center of gravity) are

$$x_c = \frac{\iint_{\Omega} xI(x,y)dxdy}{M} \quad y_c = \frac{\iint_{\Omega} yI(x,y)dxdy}{M} \tag{2}$$

For a color image, $I(x,y) = \{R(x,y), G(x,y), B(x,y)\}$, we define the normalized centroids as (in discrete form)

$$x_{rc} = \frac{\sum_{x=1}^{m}\sum_{y=1}^{n} xR(x,y)}{m\sum_{x=1}^{m}\sum_{y=1}^{n} R(x,y)} \quad y_{rc} = \frac{\sum_{x=1}^{m}\sum_{y=1}^{n} yR(x,y)}{n\sum_{x=1}^{m}\sum_{y=1}^{n} R(x,y)} \tag{3}$$

$$x_{gc} = \frac{\sum_{x=1}^{m}\sum_{y=1}^{n} xG(x,y)}{m\sum_{x=1}^{m}\sum_{y=1}^{n} G(x,y)} \quad y_{gc} = \frac{\sum_{x=1}^{m}\sum_{y=1}^{n} yG(x,y)}{n\sum_{x=1}^{m}\sum_{y=1}^{n} G(x,y)} \tag{4}$$

$$x_{bc} = \frac{\sum_{x=1}^{m}\sum_{y=1}^{n} xB(x,y)}{m\sum_{x=1}^{m}\sum_{y=1}^{n} B(x,y)} \quad y_{bc} = \frac{\sum_{x=1}^{m}\sum_{y=1}^{n} yB(x,y)}{n\sum_{x=1}^{m}\sum_{y=1}^{n} B(x,y)} \qquad (5)$$

A centroid vector, C = (x_{rc} y_{rc} x_{gc} y_{gc} x_{bc} y_{bc}), can then be formed to characterize the image content. Before presenting our adaptive hierarchical centroid vector, we first look at some of the properties of the centroid vector. In particular, we show that the centroids are invariant to illuminations and are insensitive to scaling.

Illuminant Invariant: The color of a pixel is determined by the surface reflectance, illuminant source and the image sensor characteristics. Suppose under illuminant A, a surface's color is (R_a G_a B_a), under illuminant B, the same surface's color imaging by the same sensor is (R_b G_b B_b), then, according to the coefficient model or von Kries model [9], the two colors have the following relation: R_a=αR_b; G_a=βG_b; B_a=γB_b, where α, β, and γ are unknown coefficients. Based on this model, it is not difficult to verify that the centroid vectors of an image under illuminant A, C_A=(x_{rca} y_{rca} x_{gca} y_{gca} x_{bca} y_{bca}) and that of the same image under illuminant B, C_B=(x_{rcb} y_{rcb} x_{gcb} y_{gcb} x_{bcb} y_{bcb}) are identical, i.e., C_A = C_B. Therefore, the centroid vector of the image as defined here is invariant to illumination changes (also called color constancy in computer vision literature). This is an important and useful property for matching the same object/scene imaged in different lighting conditions. For example, in content-based image retrieval, we may want to retrieval the same scene imaged at different times of the day, and in content protection, we may want to find a pirated image that has been subjected to re-colorization by unauthorized users. The illumination (color) invariant property of the centroids can be exploited to help these tasks.

Insensitive to Scaling: Suppose that an image $I(x, y)$ is scaled by factors of Δ_x and Δ_x along x and y dimensions respectively, then we have

$$x_c(\Delta_x,\Delta_y) = \frac{\sum_{x=1}^{m/\Delta_x}\sum_{y=1}^{n/\Delta_y} xR(x\Delta_x, y\Delta_y)}{(m/\Delta_x)\sum_{x=1}^{m/\Delta_x}\sum_{y=1}^{n/\Delta_y} R(x\Delta_x, y\Delta_y)} \approx x_c = \frac{\sum_{x=1}^{m}\sum_{y=1}^{n} xR(x,y)}{m\sum_{x=1}^{m}\sum_{y=1}^{n} R(x,y)} \qquad (6)$$

This shows that although not exactly invariant to scaling, the centroid vectors are *insensitive* to scaling. Again, insensitive to scaling will be useful in CBIR and near-duplicate image recognition.

4. Adaptive Computation of Hierarchical Image Geometric Centroids

To capture local as well as global image content characteristics by exploiting the useful properties of the image geometric centroids as discussed above, we have developed a method to adaptively compute multi-level image geometric centroids in a hierarchical manner as illustrated in Fig. 1, and is described as follows. For a given (color) image, we compute the geometric centroids of different color channels independently. For each color channel, we first compute the first level geometric centroid, the center of gravity of the whole image and denote it as C_0=($x_{c(0)}$,$y_{c(0)}$). A horizontal line and a vertical line that pass through C_0 partition the image into quadrants. We then compute the centroid of each of these 4 sub-images, which forms the second level geometric centroids and denote them as C_1=($x_{c(1,0)}$, $y_{c(1,0)}$, $x_{c(1,1)}$, $y_{c(1,1)}$, $x_{c(1,2)}$, $y_{c(1,2)}$, $x_{c(1,3)}$, $y_{c(1,3)}$). We again use a horizontal line and a vertical line that pass through each of these four 2nd level centroids to partition the 4 sub-images into quadrants thus partitioning the whole image into 16 sub-images. We then compute the geometric centroid of each of these 16 sub-images to form the 3rd level image geometric centroids. The procedure can be performed recursively using vertical and horizontal lines to pass through the 16 3rd level centroids to partition the image into 64 sub-images, and then compute the 64 4th level image geometric centroids.

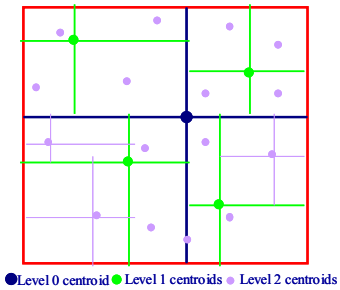

Fig. 1 Schematic diagram illustrates the process of adaptively computing of hierarchical image geometric centroids.

The method can be seen as crude top-down adaptive image segmentation using a quadtree data structure. What is different here is that, firstly, our goal is not image segmentation per se, but rather we seek to characterize the content distributions of the image; secondly, unlike in traditional image segmentation where some region homogeneity measures are used to divide the image, we partition the image and their sub-regions using the centers of gravity.

From the definition of the center of gravity, (1) – (5), it is clear that it is a function of the spatial distributions of pixel values. The adaptively computed hierarchical centroids of all color channels of an image should be able to capture the characteristics of an image's content, which in turn can be used for image matching. In the next two sections, we will present experimental results of using the adaptive hierarchical image geometric centroids for near duplicate image recognition and for content-based image retrieval.

5. Near-Duplicate Image Recognition

The success of the Internet has made it extremely easy to distribute digital data including text, music, image and video. One problem associated with it is content pirating. One of the ways to safeguard copyrighted images and trademarks is to detect near-replicas of registered contents on the Internet and return a list of suspected URLs and then the owners of the contents can check to see if suspected images are indeed unauthorized copies. Recent work in this area appeared in the literature include [1, 2].

The ways in which an image can be pirated include, format conversion (e.g., from JPEG to GIF), compression (e.g. using JPEG), scaling, cropping, re-sampling, re-

coloring, adding small amount of noise, filtering, etc. A pirated copy would be unlikely changed substantially such that it would loss the perceptual similarity to the original. In other word, a pirated copy would be a "near-duplicate" of the original.

In order to detect near-duplicates from a large collection of images, we first represent each image using 3 levels of adaptive hierarchical image geometric centroids. To form the feature vector of the image, we concatenate the geometric centroids of all color channels and all hierarchical levels. For 3 levels, the feature vector of an image has a dimensionality of 3x(2x1 + 2x4 +2 x 16) = 126. We use the following simple distance measure to compute the similarity of two images represented by a feature vector U and V

$$D(U,V)=\sum_i \frac{|u_i - v_i|}{1+u_i+v_i} \quad (7)$$

Our experimental database consists of 5000 color photographs from the Corel photo CD collection. We randomly pick 4 images (shown in Fig. 2) and for each image, we perform following transformations

(1) Re-coloring the Red channel by +40% (RCR)
(2) Re-coloring the Green channel by +40% (RCG)
(3) Re-coloring the Blue channel by +40% (RCB)
(4) Contrast enhancement +60% (CEI)
(5) Contrast enhancement -60% (CED)
(6) Cropping, remove the outer borders of the image to reduce its size by 20% and then re-scale it back to the original size (CRO)
(7) Adding 20% random noise to the image (NOI)
(8) Pinch special effect (PIN)
(9) Scale down 4 times and then up to the original size (SDU)
(10) Scale up 4 times and then down to the original size (SUD)
(11) Change the hue of the image by +50% (HUE)
(12) Change the luminance of the image by +50% (LUM)

In total, we have created 4 x 12 = 48 images. In our experiment, we embed each of these images in the 5000-image database and use the original image as the query image. For each query, we compute the distance between the query image's feature vector and those in the database using (7) and rank the database images in increasing order based on their distances from the query image. Ideally, we want the transformed versions of the query image returned in the first rank. Results are shown in table 1.

It is seen that apart from the CEI and CED transforms of the Stonehenge image and the CEI transform of the Football image (these images were ranked in 847/5000, 186/5000 and 24/5000 positions in the returned list respectively), the method achieves excellent results. Inspecting the images of these transforms on the Stonehenge image (Fig. 3), it is clear that the distortions are so severe that the whole image structure has been badly destroyed, hence such a result is understandable. Overall, it is shown that our new feature works very effectively in recognizing near duplicate images.

Fig. 2 Four query images. 12 different transforms are performed on each image to produce 48 target images.

6. Content-based Image Retrieval

Content-based image retrieval (CBIR) has been extensively studied in the past decade [3-5]. Using our adaptive hierarchical image geometric centroid technique of Section 4, we can form image content descriptors for CBIR. One possible approach is to using the centroids and the spectral image features such as color distribution and texture characteristics of each partitioned sub-image together to form the image descriptor, which should capture the overall content characteristics of the image.

Table 1. The numbers in each cell are the ranks of the transformed image in the returned image list when using the original image to query a 5000-image database.

	Image			
Transforms				
RCR	1	1	1	1
RCG	1	1	1	1
RCB	1	1	1	1
CEI	1	847	24	1
CED	1	186	1	1
CRO	1	1	1	1
NOI	1	1	1	1
PIN	1	1	1	1
SDU	1	1	1	1
SUD	1	1	1	1
HUE	1	1	1	1
LUM	1	1	1	3

Fig. 3, left to right: original, CEI, CED and RCG. These distortions are pretty severe. It is interesting that while CEI and CED returned with a very low rank, RCG and other equally severely distorted images return with a first rank.

In this experiment, we form an image's content descriptor vector as two parts (we again used a three-level partition as in previous section). The first part of the content descriptor vector is the centroids, which is a 126-d vector, and the second part of the descriptor is the Red, Green and Blue values of the centroid regions (note that other possible spectral image features of the sub images such as texture can also be used, we only use the simplest features in this experiment), this part is a 3x(1 + 4 + 16) = 63 dimensional vector. Let C(A) be the 126-d centroid part and F(A) be the 63-d color part of the content descriptor of image A, and similarly, C(B) be the 126-d centroid part and F(B) be the 63-d color part of the content descriptor of image B, then the similarity of image A and B is measured as

$$S(A,B)=D(C(A),C(B))+\lambda D(F(A),F(B)) \quad (8)$$

where D is a distance measure according to (7), λ determines a weighting between the two parts of the image content descriptor. By changing λ, we can adjust the relative importance of the two terms. For example, if we want to make the retrieval illumination invariant, e.g., we would like to find images of the same scene imaged

under different lighting conditions (different times of day for example), we can set λ to a smaller value to de-emphasis this illuminant dependant term and give more weight to the centroids which are illumination invariant. In other cases, we may be more interested in the actual color contents of the image and we can set λ to a larger value to achieve this objective.

In this experiment, we again use the 5000-image database. For query, we have collected two sets of images. The first set consists of 300 pairs of similar image (3 examples are shown in Fig. 4). The second set consists of 4 groups and each consists of 100 similar images (two example groups are shown in Fig. 5). As a comparison, we have also used a 256-bin MPEG-7 color structure descriptor (CS) [4] to perform the same experiments. In all results presented here, we set λ = 1 for our new features (not necessarily optimal).

Ours =1, CS = 1017 Ours =1, CS = 79 Ours =1, CS = 808

Fig. 4 Examples of the 1st set query/target image pairs. The numbers are the retrieval ranking of the target images for our new feature and that of MPEG-7 color structure (CS).

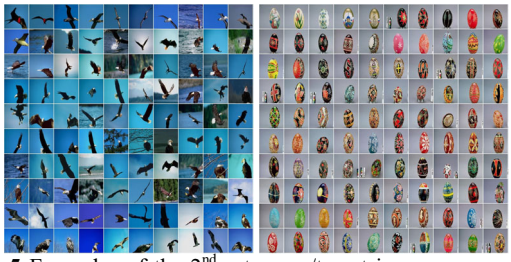

Fig. 5 Examples of the 2nd set query/target image groups. Use one of the images as query, the target is to retrieve all other images in the same group as the query.

The first set of image is used to evaluate retrieval precision. Use one of the pair as query we want to retrieve the other image of the same pair to as high a rank as possible. We use following measure for precision [5]

$$P(k) = \sum_{\forall i} \left| \{ Q_i | Rank(I_i) \le k \} \right| \qquad (9)$$

where Q_i is the ith query image and I_i is the unique retrieval target for the query. $P(k)$ measures how many correct targets are retrieved within the first kth returns, a larger $P(k)$ indicates a better performance. Results of our method and that of the MPEG-7 CS are shown in Table 2. It is seen that our new features outperforms a well-established image content descriptor. Example images for which our method significantly performed better are shown in Fig. 4.

Table 2: The $P(k)$ performance of our new method and MPEG-7 CS performed using 300 query/target pairs and a database of 5000 images.

	k = 50		k = 100		k = 300	
	ours	CS	ours	CS	ours	CS
$P(k)$	**147**	**134**	**168**	**153**	**211**	**179**

We use the second dataset to evaluate the recall ability of our new feature. We use each one of the 400 images as query, and the objective is to retrieve all those 99 images in the same group as the query image. We use following average recall measure [5]

$$AR(l) = \sum_i \left(\frac{\left| \{ Q_i | Rank(Q_i(j)) < l \} \right|}{N_i} \right) \qquad (10)$$

where Q_i is the ith querying image and $Q_i(j)$, j = 1, 2, …, 99, are the images in the same group as Q_i, N_i = 99. $AR(l)$ is a weighted score of how many correct answers are returned in the first l positions, accumulated over all queries. It is therefore a measure of recall performance. A higher value of $AR(l)$ indicates a better performance. Results accumulated over all four groups are shown in Fig. 6. It is seen that our new method performs better also in this test.

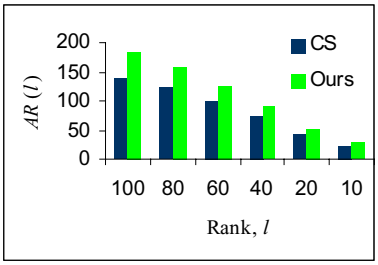

Fig. 6 Average recall performance of our new feature and that of MPEG-7 CS content descriptor performed over 400 queries to a 5000-image database.

7. Concluding Remarks

In this paper, a new method that adaptively drives hierarchical image features has been presented. The new features are shown to be illumination invariant and are insensitive to scaling. We have successfully applied the new features to near-duplicate image recognition and content-based image retrieval.

References

1. A. Qamra, Y. Meng, and E. Y. Chang, "Enhanced Perceptual Distance Functions and Indexing for Image Replica Recognition", IEEE Trans. PAMI, Vol. 27, No. 3, MARCH 2005, pp. 279 – 391
2. D Zhang and S-F Chang, "Detecting Image Near-Duplicate by Stochastic Attributed Relational Graph Matching with Learning", ACM conference of Multimedia 2004
3. Smeulders, A.W.M. Worring, M. Santini, S. Gupta, A. Jain, R., "Content-based image retrieval at the end of the early years", IEEE Trans. PAMI, Vol. 22, pp. 1349-1380, Dec 2000
4. MPEG7 FCD, ISO/IEC JTC1/SC29/WG11, March 2001, Singapore.
5. G Qiu and K-M Lam, "Frequency layered color indexing for content-based image retrieval", IEEE Transactions on Image Processing, vol. 12, no.1 pp. 102 -113, 2003
6. D. G. Lowe, "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, 60(2): 91–110, 2004
7. M-K. Hu, "Visual pattern recognition by moment invariants", IRE Trans. on Information Theory, IT-8: pp. 179-187, 1962
8. K. Mikolajczyk and C. Schmid. Indexing Based on Scale Invariant InterestPoints. In Proceedings of the International Conference on Computer Vision, pages 525–531, 2001
9. J. von Kries, "Chromatic Adaptation", In MacAdam, D.L. Ed. Sources of Color Vision, MIT Press, Cambridge, 1970