Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems

Ryszard S. Choraś

Abstract—In CBIR (Content-Based Image Retrieval), visual features such as shape, color and texture are extracted to characterize images. Each of the features is represented using one or more feature descriptors. During the retrieval, features and descriptors of the query are compared to those of the images in the database in order to rank each indexed image according to its distance to the query. In biometrics systems images used as patterns (e.g. fingerprint, iris, hand etc.) are also represented by feature vectors. The candidates patterns are then retrieved from database by comparing the distance of their feature vectors. The feature extraction methods for this applications are discussed.

Keywords—CBIR, Biometrics, Feature extraction

Fig. 1. Diagram of the image retrieval process.

I. INTRODUCTION

In various computer vision applications widely used is the process of retrieving desired images from a large collection on the basis of features that can be automatically extracted from the images themselves. These systems called CBIR (Content-Based Image Retrieval) have received intensive attention in the literature of image information retrieval since this area was started years ago, and consequently a broad range of techniques has been proposed.

The algorithms used in these systems are commonly divided into three tasks:

- extraction,
- selection, and
- classification.

The extraction task transforms rich content of images into various content features. Feature extraction is the process of generating features to be used in the selection and classification tasks. Feature selection reduces the number of features provided to the classification task. Those features which are likely to assist in discrimination are selected and used in the classification task. Features which are not selected are discarded [10].

Of these three activities, feature extraction is most critical because the particular features made available for discrimination directly influence the efficacy of the classification task. The end result of the extraction task is a set of features, commonly called a feature vector, which constitutes a representation of the image.

In the last few years, a number of above mentioned systems using image content feature extraction technologies proved reliable enough for professional applications in industrial
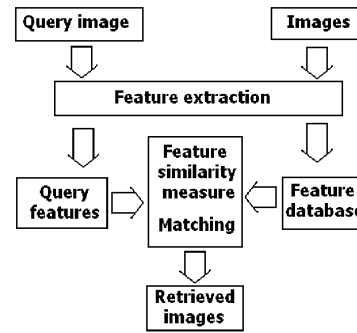
automation, biomedicine, social security, biometric authentication and crime prevention [5].

Fig. 1 shows the architecture of a typical CBIR system. For each image in the image database, its features are extracted and the obtained feature space (or vector) is stored in the feature database. When a query image comes in, its feature space will be compared with those in the feature database one by one and the similar images with the smallest feature distance will be retrieved.

CBIR can be divided in the following stages:

- *Preprocessing*: The image is first processed in order to extract the features, which describe its contents. The processing involves filtering, normalization, segmentation, and object identification. The output of this stage is a set of significant regions and objects.
- *Feature extraction*: Features such as shape, texture, color, etc. are used to describe the content of the image. Image features can be classified into primitives.

CBIR combines high-tech elements such as:
- multimedia, signal and image processing,
- pattern recognition,
- human-computer interaction,
- human perception information sciences.

In Pattern Recognition we extract "relevant" information about an object via experiments and use these measurements (= features) to classify an object. CBIR and direct object recognition, although similar in principle, using many of the same image analysis/statistical tools, are very different operations. Object recognition works with an existing database of objects and is primarily a statistical matching problem. One can argue that object recognition is a particularly nicely defined sub-set of CBIR.

In CBIR, human factors play the fundamental role. Another distinction between recognition and retrieval is evident in less specialized domains - these applications are inherently concerned with ranking (i.e., re-ordering database images according to their measured similarity to a query example) rather than classification (i.e., deciding process whether or not an observed object matches a model), as the result of similarity-based retrieval.

First generation CBIR systems were based on manual textual annotation to represent image content. This technique can only be applied to small data volumes and, to be truly effective, annotation must be limited to very narrow visual domains.

In content-based image retrieval, images are automatically indexed by generating a feature vector (stored as an index in feature databases) describing the content of the image. The similarity of the feature vectors of the query and database images is measured to retrieve the image.

Let $\{F(x,y); x = 1, 2, \ldots, X, y = 1, 2, \ldots, Y\}$ be a two-dimensional image pixel array. For color images $F(x,y)$ denotes the color value at pixel (x,y) i.e., $F(x,y) = \{F_R(x,y), F_G(x,y), F_B(x,y)\}$. For black and white images, $F(x,y)$ denotes the grayscale intensity value of pixel (x,y).

The problem of retrieval is following: For a query image Q, we find image T from the image database, such that distance between corresponding feature vectors is less than specified threshold, i.e.,

$$D(Feature(Q), Feature(T)) \leq t \qquad (1)$$

II. FEATURE EXTRACTION

The feature is defined as a function of one or more measurements, each of which specifies some quantifiable property of an object, and is computed such that it quantifies some significant characteristics of the object.

We classify the various features currently employed as follows:

- General features: Application independent features such as color, texture, and shape. According to the abstraction level, they can be further divided into:
    - Pixel-level features: Features calculated at each pixel, e.g. color, location.
    - Local features: Features calculated over the results of subdivision of the image band on image segmentation or edge detection.
    - Global features: Features calculated over the entire image or just regular sub-area of an image.
- Domain-specific features: Application dependent features such as human faces, fingerprints, and conceptual features. These features are often a synthesis of low-level features for a specific domain.

On the other hand, all features can be coarsely classified into low-level features and high-level features. Low-level features can be extracted directed from the original images, whereas high-level feature extraction must be based on low-level features [8].

A. Color

The color feature is one of the most widely used visual features in image retrieval. Images characterized by color features have many advantages:

- Robustness. The color histogram is invariant to rotation of the image on the view axis, and changes in small steps when rotated otherwise or scaled [15]. It is also insensitive to changes in image and histogram resolution and occlusion.
- Effectiveness. There is high percentage of relevance between the query image and the extracted matching images.
- Implementation simplicity. The construction of the color histogram is a straightforward process, including scanning the image, assigning color values to the resolution of the histogram, and building the histogram using color components as indices.
- Computational simplicity. The histogram computation has $O(X,Y)$ complexity for images of size $X \times Y$. The complexity for a single image match is linear, $O(n)$, where n represents the number of different colors, or resolution of the histogram.
- Low storage requirements. The color histogram size is significantly smaller than the image itself, assuming color quantisation.

Typically, the color of an image is represented through some color model. There exist various color model to describe color information. A color model is specified in terms of 3-D coordinate system and a subspace within that system where each color is represented by a single point. The more commonly used color models are *RGB* (red, green, blue), *HSV* (hue, saturation, value) and Y, C_b, C_r (luminance and chrominance). Thus the color content is characterized by 3-channels from some color model. One representation of color content of the image is by using color histogram. Statistically, it denotes the joint probability of the intensities of the three color channels.

Color is perceived by humans as a combination of three color stimuli: Red, Green, Blue, which forms a color space (Fig. 2). This model has both a physiological foundation and a hardware related one. *RGB* colors are called primary colors and are additive. By varying their combinations, other colors can be obtained. The representation of the *HSV* space (Fig. 2) is derived from the *RGB* space cube, with the main diagonal of the *RGB* model, as the vertical axis in *HSV*. As saturation varies form 0.0 to 1.0, the colors vary from unsaturated (gray) to saturated (no white component). Hue ranges from 0 to 360 degrees, with variation beginning with red, going through yellow, green, cyan, blue and magenta and back to red. These color spaces are intuitively corresponding to the *RGB* model from which they can be derived through linear or non-linear transformations.
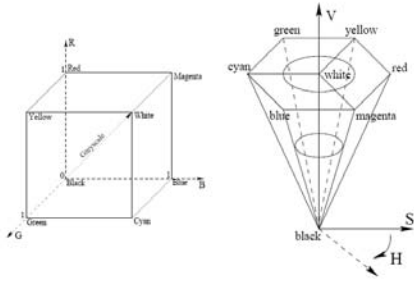
Fig. 2.   The *RGB* color space and the *HSV* color space

Fig. 4.   The *RGB* color space

Fig. 3.   Original image

Fig. 5.   The *HSV* color space

$$H = \cos^{-1}\left\{ \frac{\frac{1}{2}\left[(R-G)+(R-B)\right]}{\sqrt{(R-G)^2 + (R-B)(G-B)}} \right\}$$
$$S = 1 - \frac{3[\min(R,G,B)]}{V} \qquad (2)$$
$$V = \frac{1}{3}(R+G+B)$$

The YC_bC_r color space is used in the JPEG and MPEG international coding standards. In MPEG-7 the YC_bC_r color space is defined as

$$Y = 0.299R + 0.587G + 0.114B$$
$$C_b = -0.169R - 0.331G + 0.500B \qquad (3)$$
$$C_r = 0.500R - 0.419G - 0.081B$$

Fig. 6.   The YC_bC_r

For a three-channel image, we will have three of such histograms. The histograms are normally divided into bins in an effort to coarsely represent the content and reduce dimensionality of subsequent matching phase. A feature vector is then formed by concatenating the three channel histograms into one vector. For image retrieval, histogram of query image is then matched against histogram of all images in the database using some similarity metric.

Color descriptors of images can be global or local and consist of a number of histogram descriptors and color descriptors represented by color moments, color coherence vectors or color correlograms [9].

Color histogram describes the distribution of colors within a whole or within a interest region of image. The histogram is invariant to rotation, translation and scaling of an object but the histogram does not contain semantic information, and two images with similar color histograms can possess different contents.
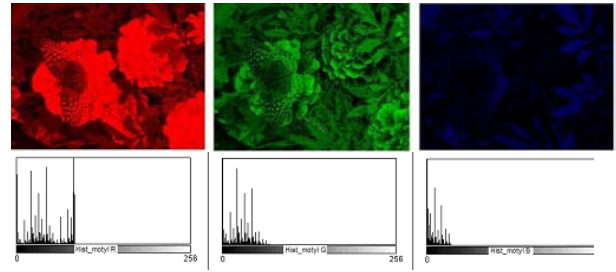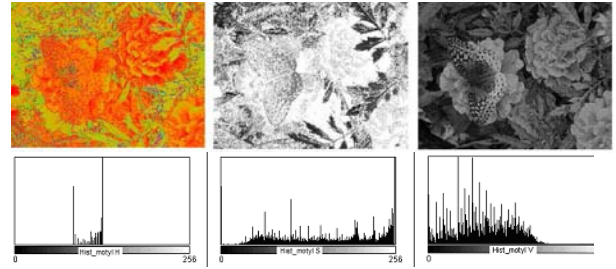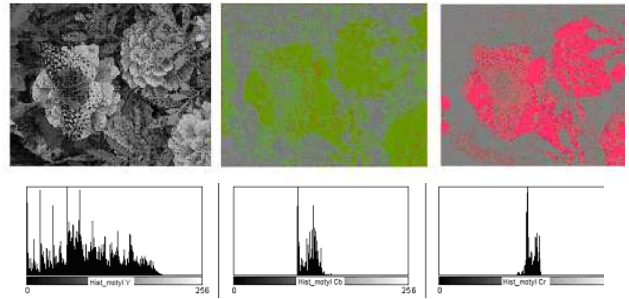
A color histogram H for a given image is defined as a vector $H = \{h[1], h[2], \ldots h[i], \ldots, h[N]\}$ where i represents a color in the color histogram, $h[i]$ is the number of pixels in color i in that image, and N is the number of bins in the color histogram, i.e., the number of colors in the adopted color model.

In order to compare images of different sizes, color histograms should be normalized. The normalized color histogram H' is defined for $h'[i] = \frac{h[i]}{XY}$ where XY is the total number of pixels in an image (the remaining variables are defined as before).

The standard measure of similarity used for color histograms:

- A color histogram $H(i)$ is generated for each image h in the database (feature vector),
- The histogram is *normalized* so that its sum equals unity (removes the size of the image),
- The histogram is then stored in the database,
- Now suppose we select a *model* image (the new image to match against all possible targets in the database).

TABLE I
COLOR MOMENTS

	R	G	B	H	S	V	Y	C_b	C_r
M_k^1	46.672	25.274	7.932	117.868	193.574	78.878	88.192	110.463	134.1
M_k^2	28.088	16.889	9.79	15.414	76.626	41.812	48.427	11.43	7.52
M_k^3	-0.065	0.114	1.059	-1.321	-0.913	0.016	0.215	-0.161	-0.28

We tried 3 kinds of histogram distance measures for a histogram $H(i)$, $i = 1, 2, \ldots, N$.

Color moments have been successfully used in many retrieval systems. The first order (mean), the second (variance) and the third order (skewness) color moments have been proved to be efficient and effective in representing color distributions of images.

The first color moment of the k-th color component ($k = 1, 2, 3$) is defined by

$$M_k^1 = \frac{1}{XY} \sum_{x=1}^{X} \sum_{y=1}^{Y} f_k(x, y) \qquad (4)$$

where $f_k(x, y)$ is the color value of the k-th color component of the image pixel (x, y) and XY is the total number of pixels in the image.

The h-th moment, $h = 2, 3, \ldots$ of k-th color component is then defined as

$$M_k^h = \left(\frac{1}{XY} \sum_{x=1}^{X} \sum_{y=1}^{Y} \left(f_k(x, y) - M_k^1 \right)^h \right)^{\frac{1}{h}} \qquad (5)$$

Since only 9 (three moments for each of the three color components) numbers are used to represent the color content of each image, color moments are a very compact representation compared to other color features.

The similarity function used for retrieval is a weighted sum of the absolute differences between the suitable moments.

Let H and G represent two color histograms. The intersection of histograms is given by:

$$d(H, G) = \sum_{k} \min(H_k, G_k) \qquad (6)$$

Color correlogram characterize color distributions of pixels and spatial correlation of pairs of colors. Let I be an image that comprises of pixels $f(i, j)$. Each pixel has certain color or gray level. Let $[G]$ be a set of G levels $g_1, g_2, \ldots, g_G$ that can occur in the image. For a pixel f let $I(f)$ denote its level g, and let I_g correspond to a pixel f, for which $I(f) = g$. Histogram for level g_x is defined as:

$$h_{g_x}(I) \equiv \Pr_{f \in I} |f \in I_{g_i}| \qquad (7)$$

Second order statistical measures are correlogram and autocorrelogram. Let $[D]$ denote a set of D fixed distances $d_1, d_2, \ldots, d_D$. Then the correlogram of the image I is defined for level pair (g_x, g_y) at a distance d

$$\gamma_{g_x, g_y}^{(d)}(I) \equiv \Pr_{f_1 \in I_{g_x}, f_2 \in I_{g_y}} \lfloor f_2 \in I_{g_x} || f_1 - f_2 = d| \rfloor \qquad (8)$$

which gives the probability that given any pixel f_1 of level g_x, a pixel f_2 at a distance d in certain direction from the given pixel f_1 is of level g_x.

Autocorrelogram captures the spatial correlation of identical levels only $\alpha_g^{(d)}(I) = \gamma_{g,g}^{(d)}(I)$.

B. Texture

Texture is another important property of images. Texture is a powerful regional descriptor that helps in the retrieval process. Texture, on its own does not have the capability of finding similar images, but it can be used to classify textured images from non-textured ones and then be combined with another visual attribute like color to make the retrieval more effective.

Texture has been one of the most important characteristic which has been used to classify and recognize objects and have been used in finding similarities between images in multimedia databases.

Basically, texture representation methods can be classified into two categories: structural; and statistical. Statistical methods, including Fourier power spectra, co-occurrence matrices, shift-invariant principal component analysis (SPCA), Tamura features, Wold decomposition, Markov random field, fractal model, and multi-resolution filtering techniques such as Gabor and wavelet transform, characterize texture by the statistical distribution of the image intensity.

The co-occurrence matrix $C(i, j)$ counts the co-occurrence of pixels with gray values i and j at a given distance d. The distance d is defined in polar coordinates (d, θ), with discrete length and orientation. In practice, θ takes the values $0°; 45°; 90°; 135°; 180°; 225°; 270°; \ and \ 315°$. The co-occurrence matrix $C(i, j)$ can now be defined as follows:

$$
\begin{aligned}
C(i, j) = & \\
= card & \left\{
\begin{array}{l}
((x_1, y_1), (x_2, y_2)) \in (XY) \times (XY) \\
for \quad f(x_1, y_1) = i, f(x_2, y_2 = j \\
\\
(x_2, y_2) = (x_1, y_1) + (d \cos \theta, d \sin \theta); \\
for \quad 0 < i, j < N
\end{array}
\right\} \qquad (9)
\end{aligned}
$$

where $card\{.\}$ denotes the number of elements in the set.

Let G be the number of gray-values in the image, then the dimension of the co-occurrence matrix $C(i, j)$ will be $N \times N$.

So, the computational complexity of the co-occurrence matrix depends quadratically on the number of gray-scales used for quantization.

Features can be extracted from the co-occurrence matrix to reduce feature space dimensionality and the formal definitions of five features from the co-occurrence matrix are done

$$Energy = \sum_i \sum_j C(i,j)^2 \tag{10}$$

$$Inertia = \sum_i \sum_j (i-j)^2 C(i,j) \tag{11}$$

$$Correlation = \frac{\sum_i \sum_j (ij)C(i,j) - \mu_i \mu_j}{\sigma_i \sigma_j} \tag{12}$$

$$DifferenceMoment = \sum_i \sum_j \frac{1}{1+(i-j)^2} C(i,j) \tag{13}$$

$$Entropy = -\sum_i \sum_j C(i,j) \log C(i,j) \tag{14}$$

where

$$\mu_i = \sum_i i \sum_j C(i,j)$$

$$\mu_j = \sum_j j \sum_i C(i,j)$$

σ_i defined as:

$$\sigma_i = \sum_i (i-\mu_i)^2 \sum_j C(i,j)$$

σ_j defined as:

$$\sigma_j = \sum_i (j-\mu_j)^2 \sum_j C(i,j)$$

Motivated by biological findings on the similarity of two-dimensional (2D) Gabor filters there has been increased interest in deploying Gabor filters in various computer vision applications and to texture analysis and image retrieval. The general functionality of the 2D Gabor filter family can be represented as a Gaussian function modulated by a complex sinusoidal signal [4].

In our work we use a bank of filters built from these Gabor functions for texture feature extraction. Before filtration, we normalize an image to remove the effects of sensor noise and gray level deformation.

The two-dimensional Gabor filter is defined as

$$Gab(x,y,W,\theta,\sigma_x,\sigma_y) =$$
$$= \frac{1}{2\pi\sigma_x\sigma_y} e^{\left[-\frac{1}{2}\left(\left(\frac{x}{\sigma_x}\right)^2 + \left(\frac{y}{\sigma_y}\right)^2\right) + jW(x\cos\theta + y\sin\theta)\right]} \tag{15}$$

where $j = \sqrt{-1}$ and σ_x and σ_y are the scaling parameters of the filter, W is the radial frequency of the sinusoid and $\theta \in [0,\pi]$ specifies the orientation of the Gabor filters.
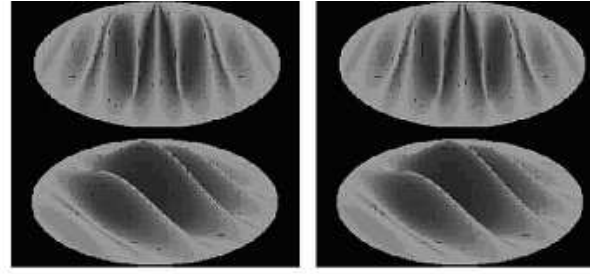

Fig. 7. Gabor filters evaluated in a single location at $0°, 45°, 90°$ and $135°$.

Gabor filtered output of the image is obtained by the convolution of the image with Gabor function for each of the orientation/spatial frequency (scale) orientation (Fig. 8). Given an image $F(x,y)$, we filter this image with $Gab(x,y,W,\theta,\sigma_x,\sigma_y)$

$$FGab(x,y,W,\theta,\sigma_x,\sigma_y) =$$
$$= \sum_k \sum_l F(x-k,y-l) * Gab(x,y,W,\theta,\sigma_x,\sigma_y) \tag{16}$$

The magnitudes of the Gabor filters responses are represented by three moments

$$\mu(W,\theta,\sigma_x,\sigma_y) = \frac{1}{XY} \sum_{x=1}^{X} \sum_{y=1}^{Y} FGab(x,y,W,\theta,\sigma_x,\sigma_y) \tag{17}$$

$$std(W,\theta,\sigma_x,\sigma_y) =$$
$$= \sqrt{\sum_{x=1}^{X} \sum_{y=1}^{Y} ||FGab(x,y,W,\theta,\sigma_x,\sigma_y)| - \mu(W,\theta,\sigma_x,\sigma_y)|^2} \tag{18}$$

$$Skew = \frac{1}{XY} \times$$
$$\times \sum_{x=1}^{X} \sum_{y=1}^{Y} \left(\frac{FGab(x,y,W,\theta,\sigma_x,\sigma_y) - \mu(W,\theta,\sigma_x,\sigma_y)}{std(W,\theta,\sigma_x,\sigma_y)} \right)^3 \tag{19}$$

The feature vector is constructed using $\mu(W,\theta,\sigma_x,\sigma_y)$, $std(W,\theta,\sigma_x,\sigma_y)$ and $Skew$ as feature components.

C. Shape

Shape based image retrieval is the measuring of similarity between shapes represented by their features. Shape is an important visual feature and it is one of the primitive features for image content description. Shape content description is difficult to define because measuring the similarity between shapes is difficult. Therefore, two steps are essential in shape based image retrieval, they are: feature extraction and similarity measurement between the extracted features. Shape descriptors can be divided into two main categories: region-based and contour-based methods. Region-based methods use the whole area of an object for shape description, while

TABLE II
TEXTURE FEATURES FOR LUMINANCE COMPONENTS OF IMAGE "MOTYL"

	$d = 1\ \alpha = 0°$	$d = 1\ \alpha = 90°$	$d = 10\ \alpha = 0°$	$d = 10\ \alpha = 90°$
Energy	872330.0007	864935.0010	946010.0004	1387267.00649
Inertia	1.531547E7	1.049544E7	5.6304255E7	4.1802732E7
Correlation	-2.078915E8	-2.007472E8	-7.664052E8	-8.716418E8
Inverse Difference Moment	788.930555	742.053910	435.616177	438.009592
Entropy	-24419.08612	-24815.09885	-37280.65796	-44291.20651

TABLE III
COLOR MOMENTS

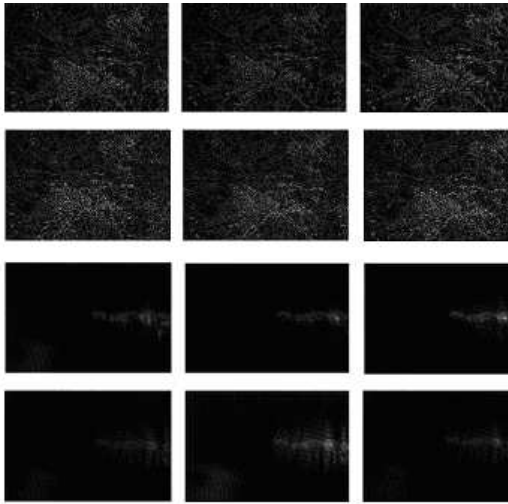|		Coke			Motyl		
θ		$\sigma = 0.3$			$\sigma = 0.3$		
	$\mu(W,\theta,\sigma)$	$std(W,\theta,\sigma)$	$Skew$	$\mu(W,\theta,\sigma)$	$std(W,\theta,\sigma)$	$Skew$	
0°	6,769	15,478	5,887	24,167	25,083	1,776	
45°	5,521	14,888	6,681	20,167	21,549	1,799	
90°	6,782	17,189	6,180	25,018	26,605	1,820	
θ		$\sigma = 3$			$\sigma = 3$		
	$\mu(W,\theta,\sigma)$	$std(W,\theta,\sigma)$	$Skew$	$\mu(W,\theta,\sigma)$	$std(W,\theta,\sigma)$	$Skew$	
0°	6,784	10,518	3,368	29,299	27,357	1,580	
45°	14,175	22,026	3,431	27,758	26,405	1,558	
90°	7,698	12,118	3,578	29,995	28,358	1,586	

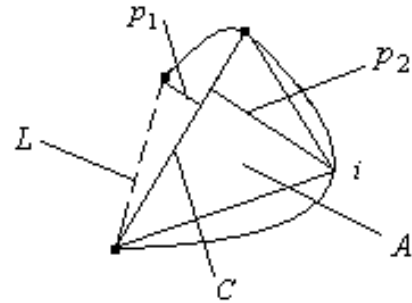Fig. 8.   Gabor filtered output of the image

Fig. 9.   Shape and measures used to compute features.

contour-based methods use only the information present in the contour of an object.

The shape descriptors described here are:

- shape descriptors - features calculated from objects contour: circularity, aspect ratio, discontinuity angle irregularity, length irregularity, complexity, right-angleness, sharpness, directedness. Those are translation, rotation (except angle), and scale invariant shape descriptors. It is possible to extract image contours from the detected edges. From the object contour the shape information is derived. We extract and store a set of shape features from the contour image and for each individual contour. These features are (Fig. 9):

1) Circularity $cir = \frac{4pA}{P^2}$
2) Aspect Ratio $ar = \frac{p_1+p_2}{C}$
3) Discontinuity Angle Irregularity $dar = \sqrt{\frac{(\sum |\theta_i - \theta_{i+1}|)}{2\pi(n-2)}}$
   A normalized measure of the average absolute difference between the discontinuity angles of polygon segments made with its adjoining segments.
4) Length Irregularity $lir = \frac{\sum |L_i - L_{i+1}|}{K}$, where $K = 2P$ for $n > 3$ and $K = P$ for $n = 3$.
   A normalized measure of the average absolute difference between the length of a polygon segment and that of its preceding segment.
5) Complexity $com = 10^{\frac{-3}{n}}$. A measure of the number of segments in a boundary group weighted such that small changes in the number of segments have more effect in low complexity shapes than in high complexity shapes.

6) Right-Angleness $ra = \frac{r}{n}$. A measure of the proportion discontinuity angles which are approximately right-angled.

7) Sharpness $sh = \sum \frac{\max(0, 1 - (\frac{2|\theta - \pi|}{\pi})^2)}{n}$. A measure of the pro-portion of sharp discontinuities (over $90°$).

8) Directedness $dir = \frac{M}{\sum P_i}$. A measure of the proportion of straight-line segments parallel to the mode segment direction.

where: n - number of sides of polygon enclosed by segment boundary, A - area of polygon enclosed by segment boundary, P - perimeter of polygon enclosed by segment boundary, C - length of longest boundary chord, p_1, p_2 - greatest perpendicular distances from longest chord to boundary, in each half-space either side of line through longest chord, θ_i - discontinuity angle between $(i-1)$-th and i-th boundary segment, r - number of discontinuity angles equal to a right-angle within a specified tolerance, and M - total length of straight-line segments parallel to mode direction of straight-line segments within a specified tolerance.

- region-based shape descriptor utilizes a set of Zernike moments calculated within a disk centered at the center of the image.

In retrieval applications, a small set of lower order moments is used to discriminate among different images [15], [12], [16]. The most common moments are:

-     the geometrical moments [15],
-     the central moments and the normalized central moments,
-     the moment invariants [20],
-     the Zernike moments and the Legendre moments (which are based on the theory of orthogonal polynomials) [17],
-     the complex moments.

A object can be represented by the spatial moments of its intensity function

$$m_{pq} = \int \int F_{pq}(x, y) f(x, y) dx dy \qquad (20)$$

where $f(x, y)$ is the intensity function representing the image, the integration is over the entire image and the $F(x, y)$ is same function of x and y for example $x^p y^q$, or a $\sin(xp)$ and $\cos(yq)$.

In the spatial case

$$m_{pq} = \sum_{x=1}^{X} \sum_{y=1}^{Y} x^p y^q f(x, y) \qquad (21)$$

The central moments are given by

$$m_{pq} = \sum_{x=1}^{X} \sum_{y=1}^{Y} (x - I)^p (y - J)^q f(x, y) \qquad (22)$$

where(I, J) are $I = \frac{m_{10}}{m_{00}}$ and $J = \frac{m_{01}}{m_{00}}$.
Normalized central moment μ_{pq}

$$\mu_{pq} = \frac{m_{pq}}{(m_{00})^\alpha}, \quad \alpha = \frac{p + q}{2} + 1 \qquad (23)$$

Hu [9] employed seven moment invariants, that are invariant under rotation as well as translation and scale change, to recognize characters independent of their position size and orientation.

$$\phi_1 = \mu_{20} + \mu_{02}$$

$$\phi_2 = [\mu_{20} - \mu_{02}]^2 + 4\mu_{11}^2$$

$$\phi_3 = [\mu_{30} - 3\mu_{02}]^2 + [3\mu_{21} - \mu_{03}]^2$$

$$\phi_4 = [\mu_{30} + \mu_{12}]^2 + [\mu_{21} + \mu_{03}]^2 \qquad (24)$$

$$\phi_5 = [\mu_{30} - 3\mu_{12}][\mu_{30} + \mu_{12}] \times$$
$$\times [(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] +$$
$$+ [3\mu_{21} - \mu_{03}][\mu_{21} + \mu_{03}] \times$$
$$\times [3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2]$$

$$\phi_6 = [\mu_{20} - \mu_{02}][(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] +$$
$$+ 4\mu_{11}[\mu_{30} + \mu_{12}][\mu_{21} + \mu_{03}]$$

$$\phi_7 = [3\mu_{21} - \mu_{03}][\mu_{30} + \mu_{12}] \times$$
$$\times [(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] -$$
$$- [\mu_{03} - 3\mu_{12}][\mu_{21} + \mu_{03}] \times$$
$$\times [3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2]$$

The kernel of Zernike moments is a set of orthogonal Zernike polynomials defined over the polar coordinate space inside a unit circle. The Zernike moment descriptor is the most suitable for shape similar-based retrieval in terms of computation complexity, compact representation, robustness, and retrieval performance. Shape is a primary image feature and is useful for image analysis, object identification and image filtering applications [11], [13]. In image retrieval, it is important for some applications in which shape representation is invariant under translation, rotation, and scaling. Orthogonal moments have additional properties of being more robust in the presence of image noise.

Zernike moments have the following advantages:

-     Rotation invariance: the magnitude of Zernike moments has rotational invariant property,
-     Robustness: they are robust to noise and minor variations in shape,
-     Expressiveness: Since the basis is orthogonal, they have minimum information redundancy,
-     Effectiveness: an image can be better described by a small set of its Zernike moments than any other types of moments such as geometric moments.
-     Multilevel representation: a relatively small set of Zernike moments can characterize the global shape of pattern. Lower order moments represent the global shape of pattern and higher order moments represent the detail.

Therefore, we choose Zernike moments as our shape descriptor in recognition and/or retrieval systems.

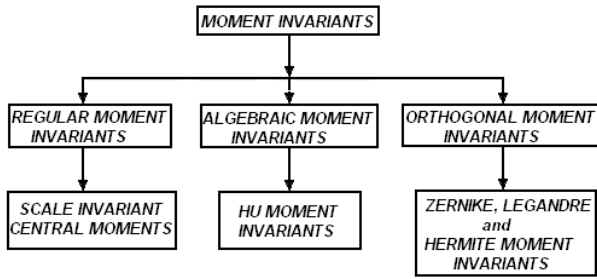Block diagram of computing Zernike moments is presented in Fig. 11 [17].

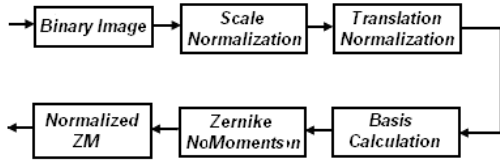Fig. 10.   The categorization of the moment invariants.

Fig. 11.   Block diagram of computing Zernike moments.

Zernike polynomials are an orthogonal series of basis functions normalized over a unit circle. These polynomials increase in complexity with increasing polynomial order [14].
To calculate the Zernike moments, the image (or region of interest) is first mapped to the unit disc using polar coordinates, where the centre of the image is the origin of the unit disc (Fig. 12). Those pixels falling outside the unit disc are not used in the calculation. The coordinates are then described by the length of the vector from the origin to the coordinate point. The mapping from Cartesian to polar coordinates is:

$$x = r\cos\theta, \quad y = r\sin\theta \qquad (25)$$

where $r = \sqrt{x^2 + y^2}, \quad \theta = \tan^{-1}(\frac{y}{x})$.

An important attribute of the geometric representations of Zernike polynomials is that lower order polynomials approximate the global features of the shape/surface, while the higher ordered polynomial terms capture local shape/surface features. Zernike moments are a class of orthogonal moments and have been shown effective in terms of image representation.

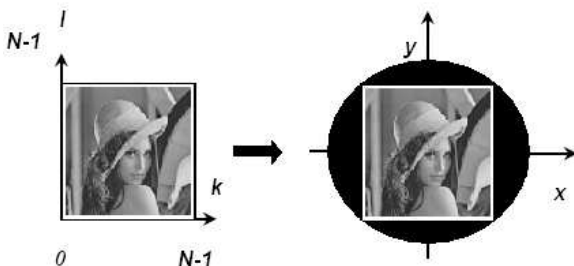The Zernike polynomials are a set of complex, orthogonal polynomials defined over the interior of

Fig. 12.   The square to circular transformation.

a unit circle $x^2 + y^2 = 1$ [13], [21],

$$V_{mn}(r,\theta) = R_{mn}(r)\exp(jn\theta) \qquad (26)$$

where $r,\theta)$ are defined over the unit disc, $j = \sqrt{-1}$ and $R_{mn}(r)$ is the orthogonal radial polynomial, defined as:

$$R_{mn}(r) = \sum_{s=0}^{\frac{m-|n|}{2}} (-1)^s F(m,n,s,r) \qquad (27)$$

where

$$F(m,n,s,r) = \frac{(m-s)!}{s!(\frac{m+|n|}{2}-s)!(\frac{m-|n|}{2}-s)!} r^{m-2s} \qquad (28)$$

where n is a non-negative integer, m is an integer such that $n - |m|$ is even and $|m| \le n$.

We have $R_{mn}(r) = R_{m,-n}(r)$, and $R_{mn}(r) = 0$ if the above conditions depictes and are not true.

So for a discrete image, if $f(x,y)$ is the current pixel then:

$$A_{mn} = \frac{m+1}{\pi}\sum_x\sum_y f(x,y)[V_{mn}(x,y)]^* \qquad (29)$$

where $x^2 + y^2 \le 1$.

It is easy to verify that

$$\begin{aligned}
V_{11}(r,\theta) &= re^{j\theta}\\
V_{20}(r,\theta) &= (2r^2-1)\\
V_{22}(r,\theta) &= r^2 e^{j2\theta}\\
V_{31}(r,\theta) &= (3r^3-2r)e^{j3\theta}
\end{aligned} \qquad (30)$$

and in the Cartesian coordinates

$$\begin{aligned}
V_{11}(x,y) &= x+jy\\
V_{20}(x,y) &= 2x^2+2y^2-1)\\
V_{22}(x,y) &= (x^2-y^2)+j(2xy)\\
V_{31}(x,y) &= (3x^3+3x^2y-2x)+j(3y^3+3xy^2-2y)
\end{aligned} \qquad (31)$$

Zernike moments are rotationally invariant and orthogonal. If A'_{nm} is the moment of order n and repetition m, associated with $f^r(x,y)$ obtained by rotating the original image by an angle φ then

$$f^r(r,\varphi) = f(r,\theta-\varphi) \qquad (32)$$

$$A^r_{nm} = \frac{m+1}{\pi}\sum_x\sum_y f^r(x,y)[V_{mn}(x,y)]^* = A_{nm}e^{-jm\varphi} \qquad (33)$$

If $m = 0$, $A^r_{nm} = A_{nm}$, there is no phase difference between them. If $m \ne 0$, we have $\arg(A^r_{nm}) = \arg(A_{nm}) + m\varphi$ that is $\varphi = \frac{\arg(A^r_{nm})-\arg(A_{nm})}{m}$ which means that if an image has been rotated, we can compute the rotation degree φ.

If $f(\frac{x}{a}, \frac{y}{a})$ represent a scaled version of the image function $f(x, y)$, then the Zernike moment A_{00} of $f(x, y)$ and A'_{nm} of $f(\frac{x}{a}, \frac{y}{a})$ are related by:

$$|A'_{00}| = a^2 |A_{00}| \quad where \quad a = \sqrt{\frac{|A'_{00}|}{|A_{00}|}} \qquad (34)$$

Therefore $|A_{nm}|$ can be used as a rotation invariant feature of the image function. Since $A_{n,-m} = A_{nm}$, and therefore $|A_{n,-m}| = |A_{nm}|$, we will use only $|A_{nm}|$ for features. Since $|A_{00}|$ and $|A_{11}|$ are the same for all of the normalized symbols, they will not be used in the feature set. Therefore the extracted features of the order n start from the second order moments up to the nth order moments.

The first two true invariants are A_{00} and $A_{11}A_{1,-1} = |A_{11}|^2$, but these are trivial where they have the same value for all images, these will not be counted.

There are two second order true Zernike moment invariants

$$A_{20} \quad and \quad A_{22}A_{2,-2} = |A_{22}|^2 \qquad (35)$$

which are unchanged under any orthogonal transformations.

There are have four third order moments A_{33}, A_{31}, $A_{3,-1}$, $A_{3,-3}$. The true invariants are written as the following:

$$A_{33}A_{3,-3} = |A_{33}|^2 \quad and \quad A_{31}A_{3,-1} = |A_{31}|^2 \qquad (36)$$

Teague [16] suggested a new term as

$$A_{33}(A_{3,-1})^3 = A_{33}[(A_{31})^*]^3 \qquad (37)$$

which is an additional invariant. This technique of forming the invariants is tedious and relies on trial and errors to guarantee its function independence.

To characterize the shape we used a feature vector consisting of the principal axis ratio, compactness, circular variance descriptors and invariant Zernike moments. This vector is used to index each shape in the database. The distance between two feature vectors is determined by city block distance measure.

III. CLASSIFIERS

As the feature are extracted, a suitable classifier must be chosen. A number of classifiers are used and each classifier is found suitable to classify a particular kind of feature vectors depending upon their characteristics. The classifiers used commonly is Nearest Neighbor classifier. The nearest neighbor classifier is used to compare the feature vector of the prototype with image feature vectors stored in the database. It is obtained by finding the distance between the prototype image and the database.

IV. APPLICATIONS

The CBIR technology has been used in several applications such as fingerprint identification, biodiversity information systems, crime prevention, medicine, among others. Some of these applications are presented in this section.

A. Medical applications

Queries based on image content descriptors can help in the diagnostic process. Visual features can be used to find images of interest and to retrieve relevant information for a clinical case. One example is a content-based medical image retrieval that supports mammographical image retrieval. The main aim of the diagnostic method in this case is to find the best features and get the high classification rate for microcalcification and mass detection in mammograms [6], [7].

The microcalcifications are grouped into clusters based on their proximity. A set of the features was initially calculated for each cluster:

- Number of calcifications in a cluster
- Total calcification area / cluster area
- Average of calcification areas
- Standard deviation of calcification areas
- Average of calcification compactness
- Standard deviation of calcification compactness
- Average of calcification mean grey level
- Standard deviation of calcification mean grey level
- Average of calcification standard deviation of grey level
- Standard deviation of calcification standard deviation of grey level.

Mass detection in mammography is based on shape and texture based features.

The features are listed below:

- Mass area. The mass area, $A = |R|$, where R is the set of pixels inside the region of mass, and $|.|$ is set cardinal.
- Mass perimeter length. The perimeter length P is the total length of the mass edge. The mass perimeter length was computed by finding the boundary of the mass, then counting the number of pixels around the boundary.
- Compactness. The compactness C is a measure of contour complexity versus enclosed area, defined as: $C = \frac{P^2}{4\pi A}$ where P and A are the mass perimeter and area respectively. A mass with a rough contour will have a higher compactness than a mass with smooth boundary.
- Normalized radial length. The normalized radial length is sum of the Euclidean distances from the mass center to each of the boundary co-ordinates, normalized by dividing by the maximum radial length.
- Minimum and maximum axis. The minimum axis of a mass is the smallest distance connecting one point along the border to another point on the border going through the center of the mass. The maximum axis of the mass is the largest distance connecting one point along the border to another point on the border going through the center of the mass.
- Average boundary roughness.
- Mean and standard deviation of the normalized radial length.
- Eccentricity. The eccentricity characterizes the

lengthiness of a *Region Of Interest*. An eccentricity close to 1 denotes a *ROI* like a circle, while values close to zero mean more stretched *ROIs*.

- Roughness. The roughness index was calculated for each boundary segment (equal length) as

$$R(j) = \sum_{k=j}^{L+j} |R_k - R_{k+1}| \qquad (38)$$

for $j = 1, 2, \ldots, \frac{n}{L}$ where $R(j)$ is the roughness index for the jth fixed length interval.

- Average mass boundary. The average mass boundary calculated as averaging the roughness index over the entire mass boundary

$$R_{ave} \frac{L}{n} \sum_{j=1}^{\frac{L}{n}} R(j) \qquad (39)$$

where n is the number of mass boundary points and L is the number of segments.

B. Iris recognition

A typical iris recognition system often includes iris capturing, preprocessing, feature extraction and feature matching. In iris recognition algorithm, pre-processing and feature extraction are two key processes. Iris preprocessing, including localization, segmentation, normalization and enhancement, is a basic step in iris identification algorithm. Iris feature extraction is the most important step in iris recognition, which determines directly the value of iris characteristics in actual application. Typical iris recognition system is illustrated in Fig. 13.
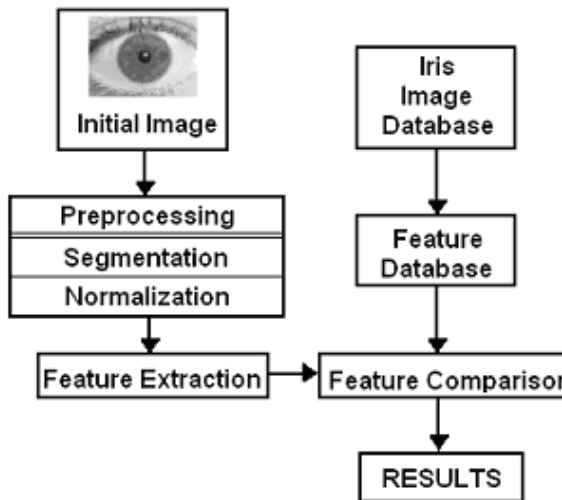

Fig. 13.   Typical iris recognition stages

Robust representations for iris recognition must be invariant to changes in the size, position and orientation of the patterns. Irises from different people may be captured in different sizes and, even for irises from the same eye, the size may change due to illumination variations and other factors. In order to compensate the varying size of the captured iris it is common to translate the segmented iris region, represented in the cartesian coordinate system, to a fixed length and dimensionless polar coordinate system. The next stage is the feature extraction [1], [2].

The remapping is done so that the transformed image is rectangle with dimension 512×32 (Fig. 14).
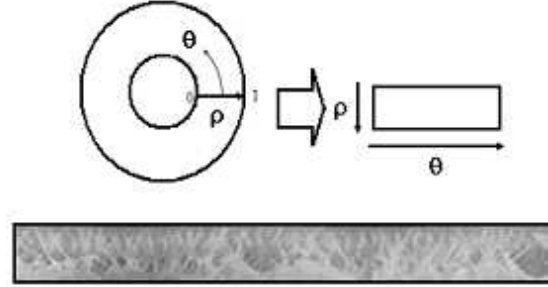

Fig. 14.   Transformed region.

Most of iris recognition systems are based on Gabor functions analysis in order to extract iris image features [3]. It consists of convolution of image with complex Gabor filters which is used to extract iris feature. As a product of this operation, complex coefficients are computed. In order to obtain iris signature, complex coefficients are evaluated and coded.

The normalized iris images (Fig. 14) are divided into two stripes, and each stripe into $K \times L$ blocks. The size of each block is $k \times l$. Each block is filtered according to (16) with orientation angles $\theta = 0°, 45°, 90°, 135°$ (Fig. 15).
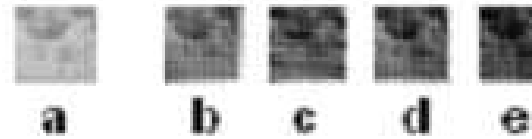

Fig. 15.   Original block iris image (a) and real part of (bcde) for $\theta = 0°, 45°, 90°, 135°$

To encode the iris we used the real part of (bcde) and the iris binary *Code* can be stored as personal identify feature.

V. CONCLUSIONS

The main contributions of this work are the identification of the problems existing in CBIR and Biometrics systems - describing image content and image feature extraction. We have described a possible approach to mapping image content onto low-level features. This paper investigated the use of a number of different color, texture and shape features for image retrieval in CBIR and Biometrics systems.

REFERENCES

[1] W.W. Boles and B. Boashash, "A human identification technique using images of the iris and wavelet transform," *IEEE Transactions on Signal Processing*, 46, pp. 1185–1188, 1998.

[2] J.G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25, pp. 1148–1161, 1993.

[3] J.G. Daugman, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression," *IEEE Trans. Acoust., Speech, Signal Processing*, 36, pp. 1169–1179, 1988.

[4] D. Gabor, "Theory of communication," *J. Inst. Elect. Eng.*, 93, pp. 429–459, 1946.

[5] A.K. Jain, R.M. Bolle and S. Pankanti (eds.), (1999) *Biometrics: Personal Identification in Networked Society*, Norwell, MA: Kluwer, 1999.

[6] J.K. Kim, H.W. Park, "Statistical textural features for detection of microcalcifications in digitized mammograms," *IEEE Transactions on Medical Imaging*, 18, pp. 231–238, 1999.

[7] S. Olson, P. Winter, "Breast calcifications: Analysis of imaging properties," *Radiology*, 169, pp. 329–332, 1998.

[8] E. Saber, A.M. Tekalp, "Integration of color, edge and texture features for automatic region-based image annotation and retrieval," *Electronic Imaging*, 7, pp. 684–700, 1998.

[9] C. Schmid, R Mohr, "Local grey value invariants for image retrieval," *IEEE Trans Pattern Anal Machine Intell*, 19, pp. 530–534, 1997.

[10] *IEEE Computer*, Special issue on Content Based Image Retrieval, 28, 9, 1995.

[11] W.Y. Kim, Y.S. Kim, "A region-based shape descriptor using Zernike moments," *Signal Processing: Image Communications*, 16, pp. 95–102, 2000.

[12] T.H. Reiss, "The revised fundamental theorem of moment invariants," *IEEE Trans. Pattern Analysis and Machine Intelligence* 13, pp. 830–834, 1991.

[13] A. Khotanzad, Y.H. Hong, "Invariant image recognition by Zernike moments," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12, pp. 489–497, 1990.

[14] S.O. Belkasim, M. Ahmadi, M. Shridhar, "Efficient algorithm for fast computation of Zernike moments," in *IEEE 39th Midwest Symposium on Circuits and Systems*, 3, pp. 1401–1404, 1996.

[15] M.K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. on Information Theory*, 8, pp. 179–187, 1962.

[16] M.R. Teague, "Image analysis via the general theory of moments," *Journal of Optical Society of America*, 70(8), pp. 920–930, 1980.

[17] A. Khotanzad, "Rotation invariant pattern recognition using Zernike moments," in *Proceedings of the International Conference on Pattern Recognition*, pp. 326–328, 1988.

[18] R. Mukundan, S.H. Ong and P.A. Lee, "Image analysis by Tchebichef moments," *IEEE Transactions on Image Processing*, 10(9), pp. 1357–1364, 2001.

[19] R. Mukundan, "A new class of rotational invariants using discrete orthogonal moments," in *Proceedings of the 6th IASTED Conference on Signal and Image Processing*, pp. 80–84, 2004.

[20] R. Mukundan, K.R. Ramakrishnan, *Moment Functions in Image Analysis: Theory and Applications*, World Scientific Publication Co., Singapore, 1998.

[21] O.D. Trier, A.K. Jain and T. Taxt, "Feature extraction methods for character recognition - a survey," *Pattern Recognition*, 29(4), pp. 641–662, 1996.

Ryszard S. Choraś is currently Full Professor in the Institute of Telecommunications of the University of Technology & Life Sciences, Bydgoszcz, Poland. His research experience covers image processing and analysis, image coding, feature extraction and computer vision. At present, he is working in the field of image retrieval and indexing, mainly in low- and high-level features extraction and knowledge extraction in CBIR systems. He is the author of Computer Vision. Methods of Image Interpretation and Identification (2005) and more than 143 articles in journals and conference proceedings. He is the member of the Polish Cybernetical Society, Polish Neural Networks Society, IASTED, and the Polish Image Processing Association.