# Image Interpretation Based On Similarity Measures of Visual Content Descriptors – An Insight

**Mungamuru Nirmala**
Lecturer, Department of Computer Science,
Eritrea Institute of Technology, Asmara, State of Eritrea.
Email:nirmala.mungamuru@gmail.com

**Kaliyaperumal Karthikeyan,**
Lecturer, Department of Computer Science,
Eritrea Institute of Technology, Asmara, State of Eritrea.
Email:kirithicraj@gmail.com

**Sreedhar Appalabatla**
Faulty, Department of Information and Communication Technology
Zoba Maekel, Ministry of Education, Asmara, Eritrea, North East Africa
Email: appalabatla.s@gmail.com

Raja Adeel Ahmed
Lecturer, Department of Computer Science,
Eritrea Institute of Technology, Asmara, State of Eritrea.
Email: adeelrajj@yahoo.com

***Abstract:*** *Efficient and effective retrieval techniques of images are desired because of the explosive growth of digital images. Content based image retrieval is a promising approach because of its automatic indexing and retrieval based on their semantic features and visual appearance. Interest in the potential of digital images has increased enormously over the last few years. Content-Based Image Retrieval (CBIR) is desirable because most webs based image search engines rely purely on the meta-data which produces a lot of garbage in the result.*

*In the content-based image retrieval, the search is based on the similarity of the content of the images such as color, texture and shape. However, "a picture is worth a thousand words." Image contents are much more versatile compared with text, and the amount of visual data is already enormous and still expanding very rapidly.*

*Most content based image retrieval systems focus on overall and qualitative similarity of scenes. Conventional information retrieval is based solely on text, and these approaches to textual information retrieval have been transplanted into image retrieval in a variety of ways, including the representation of an image as a vector of feature values. By measuring the similarity between image in the database and the query image using a similarity measure, one can retrieve the image. This paper describes Content-Based Image Retrieval methods using visual content descriptors.*

***Keywords:*** *CBIR (Content Based Image Retrieval), visual content descriptors, similarity measure, Digital image, (CBVIR) Content-Based Visual Information Retrieval, TBIR (Text Based Image Retrieval)*

## 1. Introduction

Content-Based Visual Information Retrieval (CBVIR) is an application of the computer vision to the problem of digital pictures retrieval in a large data base. Information contained in an image can be visual information or semantic information. The visual information can be stated in general contexts in form of colors, textures, shapes, spatial relations, or in other specified forms which valid in the domain of certain problems. CBIR is a retrieval technique which uses the visual information by retrieving collections of digital images. The visual information are then extracted and stated as a feature vector which in the sequel then forms a feature database.

The retrieval of data stored in the form of text or documents is carried out by giving the keywords in the search engine. The keywords are compared with the text of the documents in the database. Based on the degree of comparison, the most pertinent documents are retrieved. Since 1990s the content-based image retrieval system has been a fast advancing research area and remarkable progress has been achieved in theoretical research and in developing the retrieval system. The ideal CBIR system from a user perspective would involve what is referred to as *semantic* retrieval. An advance in CBIR was marked by commercial development of image retrieval system for government organization, private institutions and hospitals.

## 2. Content Based Image Retrieval

CBIR has been an interesting topic of research that attracts many researchers since the early of 90's. In the last decade, a lot of progress attained in theoretical research CBIR or in the development of the CBIR system. But, up to now there are still many challenging problems in the field of CBIR which attracts attention of many scientists from various disciplines.

Late of 1970's is the beginning era of research in CBIR. On 1979 there was conference on application of database technique in image, held in Florence. Since then, the application of image database management technique became an area of research that attracted many scientists. The technique used in the beginning of CBIR did not use the visual content yet, but relied on the textual information of each image -Text Based Image Retrieval (TBIR). On the other hand, additional textual information is needed of every image before retrieval. Then the images can be retrieved by using the textual approach DBMS. Text based image retrieval uses the traditional database technique to manage the image database. Through textual description, image can be organized based on topic or level of hierarchies to make the navigation process and browsing easier.

Generally in an image retrieval system, the visual content of images is stored in a multi dimension feature vector. To retrieve an image, users enter inputs of the form image query or sketch. Then the CBIR system computes the feature vector of the query images or sketch. Similarity between the feature vectors of query image/sketch is obtained based on a measure of distance or index scheme.

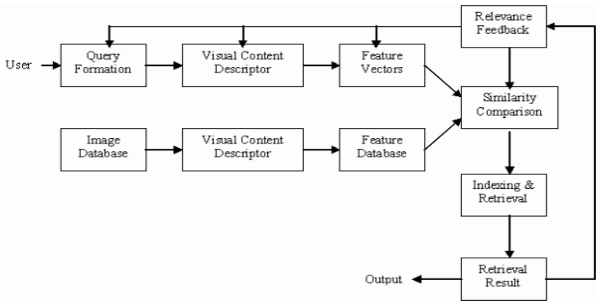

Fig 1: Content based image retrieval System overview

Some of the commercial CBIR systems are IBM's QBIC, Virage's VIR Image Engine, and Excalibur's Image Retrieval Ware. The current CBIR system is limited to effectively operate at the primitive feature level. They are not efficient to search for some semantic queries. For example, say, a photo of a cat. But some semantic queries can be handled by specifying them in terms of primitives.

## 3. Literature Review

An image is called 'the same' with an image in the database if the value of similarity measure is 'small'. This means that a good CBIR retrieval system must be supported by an accurate similarity measure. The Robust Distance for Similarity Measure of Content Based Image Retrieval, Dyah E. Herwindiati and Sani M Isa [3]

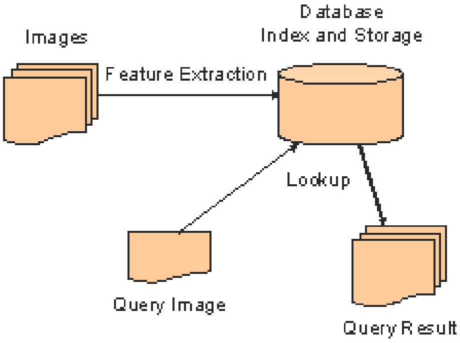

Fig 2: Typical flow of CBIR

Shape matching is an important ingredient in shape retrieval, recognition and classification, alignment and registration, and approximation and simplification. Various aspects that are needed to solve shape matching problems: choosing the precise problem, selecting the properties of the similarity measure that are needed for the problem, choosing the specific similarity measure, and constructing the algorithm to compute the similarity.
Remco C. Veltkamp, "Shape Matching: Similarity Measures and Algorithms," smi, pp.0188, International Conference on Shape Modeling & Applications, 2001[7]

## 4. Visual Contents

Visual contents are divided as primitive features and semantic features. Primitive features are the low-level visual features such as color, texture, shape and spatial relationships - directly related to perceptual aspects of image content. It is usually easy to extract and represent these features and fairly convenient to design similarity measures. System of image retrieval generally stores visual contents of an image in a multi dimension feature vector. It is a very challenging task to extract and manage meaningful semantics and to make use of them to achieve more intelligent and user-friendly retrieval. In an image retrieval process, users enter inputs of the form query image. Depending upon the image a suitable content can be chosen and a descriptor is defined for retrieval.

### 4.1 Content Descriptor

A *descriptor* is the collection of features or attributes of an object. A good visual content descriptor must be invariant to the variance caused by the process of image formation or local. The global descriptor makes use visual information from the whole images, meanwhile the local descriptor makes use visual information of image region to describe the visual content of images. Any object or image should satisfy the criteria of both local and global features. To obtain the local descriptors, the image is partitioned into parts of equal size and shape. A better method is to use some criteria to divide the image into homogeneous regions or a region segmentation algorithm. A complex method of dividing an image is to consider a complete object segmentation to obtain semantically meaningful objects like dog, ball, donkey etc. Depending upon the image, the description of the global features and the local features will differ with respect to the visual content under consideration.

### 4.2 Feature Vectors

CBR system computes the feature vector of the query image. Similarity between the feature vectors of the query images is obtained based on the measurement of distance of index scheme. The features of the visual contents are extracted from the image. The collection of features of the contents is known as a *feature vector*. Therefore, they are multi-dimensional vectors. Feature vectors describe particular characteristics of an image based on the nature of the extraction method. For feature extraction, a number of extraction algorithms have been proposed. Visual feature extraction is the basis of any content-based image retrieval technique. For example, properties of a bounding box, a curvature, spherical functions etc. The feature vectors are used for indexing and image retrieval using a similarity measure. The collection of feature vectors is termed as *feature database* of the images in the database. Feature vector and the different approaches on feature-based similarity search techniques are discussed in [4] and [5].

## 5. Color

Color provides a significant portion of visual information to human beings and enhances their abilities of object detection. Each pixel in an image can be represented as a point in a 3D color space. The common color spaces used for image retrieval are RGB, CIE L*a*b*, CIE L*u*v*, HSV (or HSL, HSB), and opponent color space. The perceptual attributes of color are brightness, hue and saturation. Brightness is the luminance perceived from the color. Hue refers to its redness, greenness etc. Uniformity of a color is the desirable characteristics of an appropriate color space for image retrieval. If two color pairs are equal in similarity distance in a color space then they are perceived as equal by the viewer.

### 5.1 Color Descriptor

Color is widely used for content-based image and video retrieval in multimedia databases. The retrieval method based on the content, *color*, uses the similarity measure of features derived from color. For example, the proportion of the colors in the images is computed for the sample image. Then the images in the database having more or less the same proportion of the colors can be retrieved using the similarity measures. The descriptors are defined based on the attributes of color. The color descriptors are Color Histogram, Color Moments, Color Coherence Vector, and Color Correlogram. The descriptors color histogram and color moments do not include spatial color distribution.

### 5.2 Color Histogram

Color histogram is the most commonly used descriptor in image retrieval. Images added to the collection are analyzed to compute the color histogram, which shows the proportion of pixels of each color within the image. The color histogram is stored in the database. The user has to specify the proportion of each color to search for a desired image. The user can also input an image from which a color histogram is calculated. The matching process retrieves those images whose color histograms match with those of the query images very closely.

Further the retrieval system includes: (i) the definition of adequate color space with respect to a specific application (ii) an appropriate extraction algorithms (iii) evaluation of similarity measures.
The color histogram extraction algorithm can be divided into three steps:

(i) Partition of the color space into cells
(ii) Associate each cell to a histogram bin
(iii) Count the number of image pixels of each cell and store this count in the corresponding histogram bin.

The descriptor is invariant to translation and rotation.

### 5.3 Color Moments

Color moments have been used in many retrieval systems, when the image contains just the object. The first three moments have been proved to be efficient and effective in representing color distributions of images. They are defined as:

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} x_{ij}$$

$$\sigma_i = \left(\frac{1}{N}\sum_{j=1}^{N}(x_{ij}-\mu_i)^2\right)^{1/2}$$

$$s_i = \left(\frac{1}{N}\sum_{j=1}^{N}(x_{ij}-\mu_i)^3\right)^{1/3}$$

Where $x_{ij}$ is the value of the ith color component of the image pixel j and N is the number of the pixels in the image.

### 5.4 Color Coherence Vector (CCV)

This descriptor incorporates spatial information into the color histogram. Each pixel of the image is classified as either coherent or incoherent, depending upon whether or not it is a part of a "large" similarly-colored region. A region is assumed to be large if its size exceeds a fixed user-set value. By counting coherence and incoherent pixels separately, the method offers a finer distinction between images than color histograms.

For each color $c_i$ the number of coherent pixels, $\alpha_{ci}$, and the number of non-coherent pixels, $\beta_{ci}$, are computed; each component in the CCV is a pair ($\alpha_{ci}$, $\beta_{ci}$), called a *coherence pair*. The coherence vector is given by

$$V_c =< (\alpha_{c1},\beta_{c1}),(\alpha_{c2},\beta_{c2}),...,(\alpha_{cN},\beta_{cN}) >$$

The sum $\alpha_{ci} + \beta_{ci}$ is the number of pixels of color $c_i$ present in the image; the set of sums for i = 1,2,…,N represents the color histogram.

## 5.5 Color Correlogram

The color correlogram encodes the spatial correlation of colors. The first and second dimension of the three-dimensional histogram is the colors of any pixel pair and the third dimension is their spatial distance. The color correlogram is a table indexed by color pair, where the k-th entry for <i, j> specifies the probability of finding a pixel of color j at a distance k from a pixel of color i, in the image.

Let I represent the entire set of pixels of the image and $I_{c(i0}$ represent the set of pixels whose colors are *c(i).*

The color correlogram is defined as:

$$r_{i,j}^{(k)} = \Pr_{p_1 \in I_{c(i)}, p_2 \in I} [p_2 \in I_{c(j)} / | p_1 - p_2 |= k]$$

Where i, j $\in$ {1, 2,…, N}, k$\in$ {1,2,…,*d*} and $|p_1-p_2|$ is the distance between pixels $p_1$ and $p_2$. If all possible combinations of color pairs are considered the size of the color correlogram will be very large ($O$ ($N^2$, $d$)). But instead, color auto-correlogram captures the spatial correlation between identical colors, hence the size is reduced to $O(Nd)$. Even though the auto-correlogram is having high dimensionality when compared to color histogram and CCV, its computational cost is very high.

## 6. Texture

Texture is a low-level descriptor for image search and retrieval applications. There are three texture descriptors considered in MPEG-7. It describes spatial relationships among grey-levels in an image. Texture is observed in the structural patterns of surfaces of objects such as wood, grain, sand, grass and cloth. The ability to match on texture similarity can be useful in distinguishing between areas of images with similar color such as sea and sky, grass and leaves. A variety of techniques have been used for measuring texture similarity. A technique named as modeling-from-reality [5] has been proposed for creating geometric models of virtual objects, and is used for texture mapping of color images. The color and texture descriptors [6] of MPEG-7 standard are described and the effectiveness of these descriptors in similarity retrieval is also evaluated.

## 6.1 Texture Descriptors

A basic texture element is known as texels. A Texel contains several pixels. The placement of the texel could be periodic or random. Natural textures are random and artificial textures are periodic or deterministic. Texture can be smooth, regular, irregular, linear, rippled etc. Texture is broadly classified into two main categories, namely *Statistical* and *Structural*. Textures that are random in nature are suitable for statistical categorization. Structural textures are deterministic texels, which repeat according to some placement rules, deterministic or random. A Texel is isolated by identifying a group of pixels having certain invariant properties, which repeat in the given image. The pixel can be defined by its gray level, shape, or homogeneity of some local property like, size, orientation or concurrence matrix. The placement rules are the spatial relationship of the pixels. In deterministic rule, the spatial relationships may be expressed in terms of adjacency, closest distance, etc and the texture is said to be strong.

For randomly placed texels, the associated texture is said to be weak and the placement rules may be expressed in terms of edge density, run length of maximally connected texels, relative extreme density. Apart from these two classifications there is another model called *Mosaic model*, which is a combination of both statistical and structural approaches.

The texture is described based on its features such as, coarseness, contrast, directionality, likeliness, regularity and roughness. These features are designed based on the psychological studies on the human perception of texture.

## 6.2 Coarseness

Coarseness is a measure of the granularity of the texture. Coarseness is calculated using moving averages. Let $M_k$(x, y) denote the moving average at each pixel (x, y) in the window of size $2^k$ x $2^{k,}$ where k = 0, 1, …, 5)

$$M_k(x,y) = \sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} g(i,j)/2^{2k}$$

Where g(i, j) is the pixel intensity at (i, j). The differences between the pairs of non-overlapping moving averages in the horizontal and vertical directions for each pixel are computed. The value of k, which maximizes, the difference calculated in either direction is used to set the best size (say) *S* for each pixel.

$$S_{best}(x, y) = 2^k$$

The coarseness is computed by averaging $S_{best}$ over the entire image.

$$coarseness = \frac{1}{mxn}\sum_{i=1}^{m}\sum_{j=1}^{n} \mathrm{S}_{\mathrm{best}}(i, j)$$

### 6.3 Contrast

Contrast may be defined as the difference in perceived brightness. Detection of light spots depends on the brightness, size of the space and duration as well as the contrast between the spot and the background.

$$contrast = \frac{\sigma}{\alpha_4^{1/4}}$$

Where the kurtosis $\alpha_4 = \mu_4/\sigma^4$, $\mu_4$ is the fourth central moment, and $\sigma^2$ is the variance. The contrast can be computed using this formula for both the entire image and a region of the image.

## 7. Shape Retrieval

The shape of an object refers to its profile and physical structure. These characteristics can be represented by the boundary, region, moment, and structural representations. Shapes are divided into two main categories namely, static shapes and dynamic shapes. Static shapes are rigid shapes. They do not change due to deformation or articulation. For example the shape of a rigid object like a water jug is a static shape. The object like human face is a dynamic shape as the shape of the human face changes with the change in expressions and actions of the human being. A number of features of the object shape can be computed for each stored image. The same set of features is computed for the query image. The images that are having close similarity with the query image are retrieved from the database. The various aspects that are needed to solve shape matching problems like choosing precise problem, selection properties of similarity measure are stated in [7].

### 7.1 Shape Descriptors

Shape descriptors are classified into boundary-based and region-based methods. This classification takes into account whether shape features are extracted from the contour or from the whole shape region. That can be further divided into structural (local) and global descriptors. If the shape is represented by segments or sections, it is structural and if it is from the whole shape region, it is global. Another classification categorizes the shape description into spatial and transform domain techniques, depending on whether direct measurements of the shape are used or a transformation is applied. The literature on content-based retrieval methods are evaluated [8] with respect to several requirements of the retrieval system.

### 7.2 Shape Matching

The retrieval method based on shape can be divided into three broad categories: (1) feature based methods, (2) graph based method and (3) other methods.

Feature Based Methods: The shape feature can be classified into regenerative features and measurement features. Boundaries, regions, moments, structural and syntactic features are identical to the regenerative features. Geometry and Moments are paired with measurement features. In 3D shapes, features denote geometric and topological properties of 3D shapes. Based on the type of shape feature used, the feature based method can be divided into: (1) global features, (2) spatial maps, (3) global feature distributions and (4) local features. The first three represent features of a shape using a single descriptor. The descriptor is a vector of *n*-dimension, and *n* is fixed for all shapes.

The global features are used to characterize the overall shape of the objects. These methods use the global features like that of area; volume, statistical moments, and Fourier transform coefficients. These methods do not discriminate above the object details. These methods support the user feed back. The global feature distribution method is a refinement of the global feature method. Spatial maps are representations that use the spatial location of an object. The entries in the map are locations of the object and are arranged in the relative positions of the features in an object. Local features are derived from the segment or part of the image. At the outset, the image is partitioned using a suitable criterion into equal sizes or meaningful objects. Then based on the features, similarity is measured.

Graph Based Methods: In graph based method, the geometric meaning of a shape is extracted and a graph is constructed. The graph represents the shape components and the links. The graph based method includes (1) model graph, (2) Reeb graph, and (3) Skeletons. Model based graph are useful to 3D solid models created using CAD systems. This method is difficult to find the similarity for models of natural shapes like humans and animals. Because of the fact that the shape should be solid, it is difficult to represent the natural shapes as similar to a sphere, a cylinder, or a place. The skeletal points are connected in an undirected acyclic shape graph, using *Minimum Spanning Tree algorithm*. The shape information of the 3D objects [9] are used to form a skeletal graph and a graph matching technique is used for retrieval Moreover, as an extension of the skeleton method of shape comparison the objects may be segmented as semantically meaningful parts (shapes), the skeleton of the parts can be used for similarity measure. The skeleton of the parts can be converted into graph, and graph matching techniques can be used for comparison.

## 8. Similarity Measure

The measure of similarity between two objects is obtained, based on the distance between pairs of descriptors using a dissimilarity measure. If the distance is small, then it means small dissimilarity and large similarity. The dissimilarity measure can be defined as a non-negative valued function.

Let *d* be the dissimilarity measure on a set S. *d*: S x S → R+ U {0}. The following properties are defined on *d*.

(i) Identity: For all x Є S, *d*(x, x) = 0

(ii) Positivity: For all x, y Є S, *d*(x, y)>0, where x ≠y

(iii) Symmetry: : For all x, y Є S, *d*(x, y) = *d*(y, x)

(iv) Triangle Inequality: For all x, y, z Є S, *d*(x, z) = *d*(x, y) + *d*(y, z)

(v) Transformation Invariance: For any chosen transformation group T,
for all x, y Є S, t Є T, *d*(t(x),t(y)) = *d*(x, y).

The identity property states that the descriptor is completely similar to itself. The positive property implies that different descriptors are never completely similar. This is a very strong property for a high-level descriptor and it is rarely contented. This will not affect the result much if the dissimilarity is on the negligible part of the image.

According to human perception, a visual content of the image, say, shape is not always similar. Human perception does not find a shape x similar to y, as y is similar to x. If partial matching of objects is used, the Triangle Inequality is not satisfied. Since the part of the object is matched the distance between the object would be very small.

Transformation Invariance should be satisfied in all types of descriptors, as the comparison and the extraction process of the descriptors are independent of the place, orientation and scale of the object in the Cartesian coordinate system. If a dissimilarity measure is affected by any transformation, an alternative formulation may be used. For example, (v) can be defined as

(v) Transformation Invariance: For any chosen transformation group T,
for all x,y Є S, t Є T, *d*(t(x),y) = *d*(x,y).

If all the properties (i)-(iv) hold, then dissimilarity is called a *metric*. It is called *pseudo-metric* if (i), (iii) and (iv) hold and *semi-metric* if only (i), (ii) and (iii).

Future Work:

The future study involves in deriving a retrieval method for natural objects in images. It includes comparison of different similarity search methods and the feature extraction methods for the shape of the objects in the image.

Some of the Challenges are:

- ➢ Semantic gap
    - o The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation.
    - o User seeks semantic similarity, but the database can only provide similarity by data processing.
- ➢ Huge amount of objects to search among.
- ➢ Incomplete query specification.
- ➢ Incomplete image description.

## 9. Conclusion

In this paper, the basic concept of Content-Based Image Retrieval methods using visual content is described. The descriptors should be transformation invariance to measure the similarity between any two descriptors. Instead of comparing the query image as a whole with the images in the database, the descriptors are compared for retrieval. With an extension of this comparison the feedback of the retrieved images can be used to further refine the retrieval process. Based on the feedback the descriptor can be redefined and an iterative similarity checking can be implemented to improve the retrieval of proper images with reference to the query image. Another approach in CBIR system is that a priori feature extraction is defined. The features are selected from the predefined set of features. This method uses a set of primitive features and logical features. Based on both primitive and logical features, the query is processed. This may result in the reduction of cost in feature extraction.

## References

1. Eakins, John & Margaret Graham: Content-Based Image Retrieval. *JISC Technology Applications.* Report 39: 1-65 (1999).

2. Benjamin, Bustos, Keim Daniel, Saupe Dietmar & Schreck Tobias: Content-based 3D Object Retrieval. (2007).

3. Isa Dyah E. Herwindiati and Sani M: The Robust Distance for Similarity Measure of Content Based Image Retrieval. Proceedings of the World Congress on Engineering 2009.

4. Vol II WCE 2009, July 1 - 3, 2009, London, U.K.Benjamin, Bustos, Keim Daniel, Saupe Dietmar, Schreck Tobias & Dejan V. Vranic: Feature-Based Similarity Search in 3D Object Databases. *ACM Computing Surveys (CSUR)*, 37: 345-387 (2005).

5. Benjamin, Bustos, Keim Daniel, Saupe Dietmar & Dejan V. Vranic: An Experimental Effectiveness Comparison of Methods for 3D Similarity Search. *International Journal on Digital Libraries, Special Issue on Multimedia Contents and Management*, 6(1): 39-54. (2006).

6. Kurazume Ryo, Ko Nishino, Zhengyou Zhang & Katsushi Ikeuchi: Simultaneous 2D images and 3D geometric model registration for texture mapping utilizing reflectance attribute. Paper presented in the *5th Asian Conference on Computer Vision*, 23–25 Melbourne, Australia (2002).

7. Manjunath, B. S., Jens-Rainer Ohm, Vinod V. Vasudevan, Akio Yamada: Color and Texture Descriptors. *IEEE* Transactions on Circuits and Systems for Video Technology 11/6: 703-714 (2001).

8. Remco C. Veltkamp, "Shape Matching: Similarity Measures and Algorithms," smi, pp.0188, International Conference on Shape Modeling & Applications, 2001 Veltkamp. C. Remco (2001) Shape Matching: Similarity Measures and Algorithms *SMI* (2001).

9. Tangelder, Johan W.H. & Remco C. Veltkamp: A Survey of Content Based 3D Shape Retrieval Methods.
www.cs.princeton.edu/~funk/cacm05.pdf (2005).