A survey on content-based image retrieval/browsing systems exploiting semantic

Paolo Parisen Toldin

2010-09-13

Abstract

In this survey we will present the basic idea behind the content-based image retrieval systems. We will present some low-level feature, techiques for matching similarity and we will conclude by giving basic ideas on semantic.

Contents

1 Introduction

Nowadays, we are surrounded by multimedia devices. Many of us have a camera, a cellular phone, an audio/video player, and so on ...

The explosion of digital data in all of these forms (audio, video, text, image, ...) has arisen a question about retrieve relevant information from repository that could be very large.

Indeed, many areas of our world are affected by this problem. Consider for example hospitals, where a large number of images is created every day. Usually, the only way to searchig an image in those collections was by keyword indexing or, simply, by browsing.

The ability to retrieve stored items from memory, based on similarity, has become a key property that underlies much of what we associate with human intelligence, including reasoning, classification and prediction. What a human mind could solve effortlessly, remains a very difficul task for a machine; these kind of task are often not computable, in general.

The text approach can be tracked back to the '70 years and they provide two main disavantages. One is about human labour; how many hours are needed to manual annotate every image? The other is about the accuracy; How can we be sure that the annotation is correct? it depends on human perception [8].

Study about multimedia databases however have open the possibility to content-based search. The word "Content-based" means that the search will analyze the actual contents of the image rather than the metadata such as keywords, tags, and/or descriptions associated with the image. We use the term "content" to refer to those feature, such as colors, shapes, textures, ..., that can be derived from the image itself.

Nowadays, there are some web services that use the content-based for image retrieval (CBIR). The most famous, maybe, is the tinyeye search engine, whicth allow to search through the web for similar images. Other image search engines, like google images, instead, do not take advantage of these new possibility yet.

Since internet has become increasingly important and a lot of web services were born, this topic has grown very fast in its importance. There are a lot of people that still works [15] on this topic. For this reason, it is not so hard to find general overview in this topic as eg. [18], [2], [12], [11], and so on....

2 Background

One of the first study in this sector is [6]. The authors explained new methodology for image indexing and abstraction in pictorical databases in order to facilitate image retrieval. The pictorical databases consisted of picture as object and picture relations.

The main difference between a text-based and a content-based retrieval system is that the human interaction is an indispensable part of the first system. Human tends to use high-level feature, such as concept, to describe a picture, while a calculating machine is able to use only low level feature, such as color, texture, shape, ...

In general there is no direct link between low-level feature and high-level feature [17].

2.1 Extending standard DBMS

Typically, DBMS are not designed to deal with multimedia information. Indeed multimedia files are often very large and their structures, very complex, are very far away from the normal structures that a DBMS is used to deal with. Furthermore, the structures of multimedia files, often, are not build to handle content-based search.

What we have to do, so, is to exstend standard DBMS in order to be able to deal with multimedia object. Many of DBMS known permits to add user-defined data types and provides ad-hoc function in order to work on them.

3 Low level image feature

Content-based image retrieval is based principally on low level image feature. As it has been found that users are usually more interested in specific region rather then entire image, most current content-based image retrieval systems are region based. This means that tha image is divided in regions on which the other operation are performed. To carry out this first step we need to perform a *segmentation* of the original image.

We will consider several classes of features that are used to specify queries: color, texture, shape.

3.1 Image segmentation

Automatic image segmentation is a difficult task to compute. A lot of techniques have been proposed in the past, such as curve evolution [9] and graph partitioning [13]. A lot of known techniques works well for images that have homogeneous color regions. Let's see an example of segmentation on one picture in figure 3.1.

However, picture from the real world are richer of color and shades. In literature are know many segmentation algorithm, like JSEG [7], blobworld [4] or KMCC [14].

Figure 1: Example of segmentation of a picture. (the red lines divided the regions)

3.2 Color feature

Color features are easy to obtain, often directly from the pixels intensities like color histogram over the whole image, over a fixed subimage, or over a segmented region. It is one of the most used feature in image retrieval. The colors are described by their color space: RGB, LAB, LUV, HSV, ...

RGB is the best known space color and is it commmonly used for visualization. The acronym stands for *Red Green Blue*. This space color can be seen as a cube where the horizontal x-axis as red values increasing to the left, y-axis as blue increasing to the lower right and the vertical z-axis as green increasing towards the top, as in figure 2.

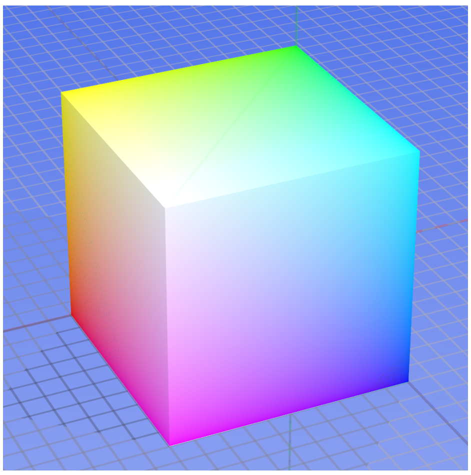

Figure 2: The RGB color model mapped to a cube. The origin, black, is hidden behind the cube.

RGB is a convenient color model for computer graphics because the human visual system works in a way that is similar, though not quite identical, to an RGB color space.

Another "famous" space color is the HSV. The acronym stads for *Hue Saturation Value*. Refering to the image 3 we can see the color space as a cylinder, where the angle around the central vertical axis corresponds to hue, the distance from the axis corresponds to saturation, and the distance along the axis corresponds to value (also called brightness).

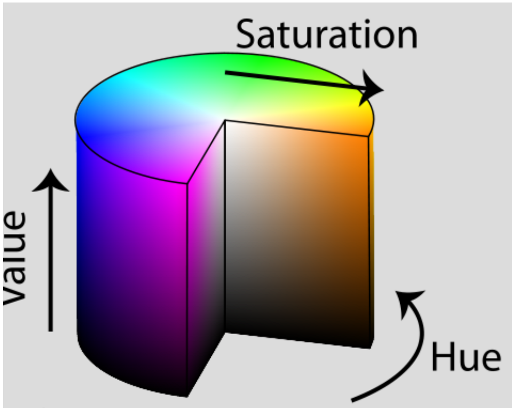

Figure 3: The HSV space color.

3.3 Shape feature

There's no universal definition of what a shape is. Impressions of shape can be conveyed by color or intensity patterns, or texture, from which a geometrical representation can be derived, like Plato's Meno, where a socratic dialog is made around the word "figure".

Figure is the only existing thing that is found always following color - Socrate

Shape feature (like aspect ratio, circularity, Fourier descriptor, consecutive boundary segments, ...) are very important image feture, iven if they are not so commmonly used in Region-Based Image Retrieval Systems. Due to the inaccuracy of segmentation step, it is more difficult to apply shape features instead of color or texture feature.

However in literature are know some Content-Based Image Retrieval Systems that use this feature, like [21], [14] e [20].

3.4 Texture feature

A precise definition of texture is untraceable. The notion of texture generally refers to the presence of a spatial pattern that has some properties of homogeneity. In particular, homogeneity cannot result from the presence of only a single color in the regions, but requires interaction of various colors.

However, texture information is important because it describes the content of many real world images suach as bricks, clouds, fog, etc. . . .

But, what is exactely a texture? let's try to answer this question by suggesting some examples, in figure 3.4.

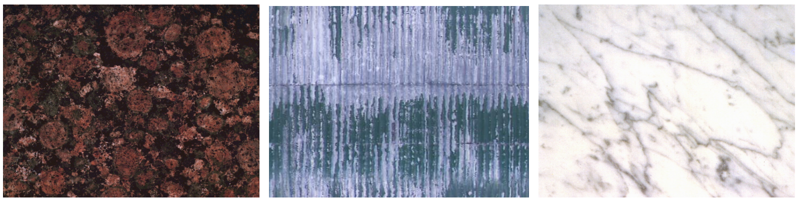

Figure 4: Example of textures

Textures can provide particular information. They are called Tamura feture [19]: Coarseness (granularity of the pattern), Contrast (amount of local changes in brightness), Regularity, Directionality, Line-likeness and Roughness. The first three are considered the most important feature. The other three are related to the first three and do not add much more information on texture description.

4 Similarity

The problem of estimating the relative significance of different features pertains to information retrieval in general. Indeed the human perception cannot be simulated exactly by a computational machine.

In order to approximate the expect result we will compare the low level features of the selected images.

Basically, there are two ways to measure similarity of images:

- One-One Match: Each region of the query image is allowed to match with only one region in the target image and, obviously, vice-versa. Then the overall similarity is calculated as the weighted sum of the similarity of each matched region. In [1] they suggest a way to have a *best match*.

- Many-Many Match: In this case, a region of the query image is allowed to match with one or more regions in the target image. In this case, the most used algorithm is Earth Mover's Distance [16]. It matches perceptual similarity well.

4.1 Color

Every electronic image is composed by a finite number of pixels, described by using a certain type of color model. We could, simply, compare every pixels of two candidate images in order to understand if they look similar or not. By the way, this solution is not practically feasible.

eg: supose to use RGB color space and a depth of 24 bits per image. The number of possibile, distinct, colors are $2^{24} = 16777216$: too much!

Indeed, there are some colors that are indistinguishable from each other. eg: shade of red that our eye cannot distinguish, and so on. . . .

For all of those reasons, it is consider more practically create approximate histograms of desired images, as in figure 4.1.

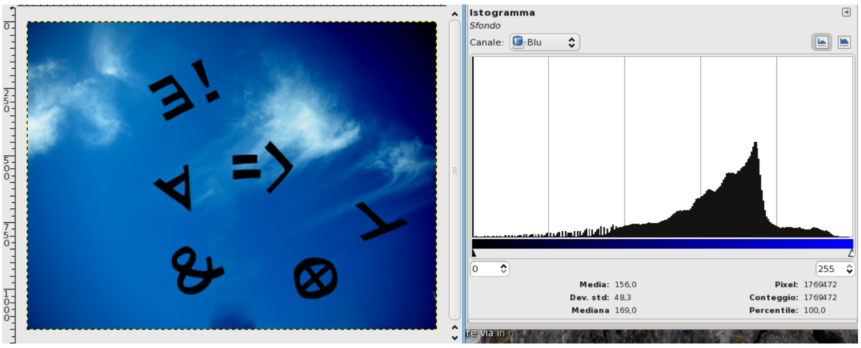

Figure 5: Example of histogram

By using an approximate histogram of desired image, we can, practically, compare images each other. Indeed, histograms are vector and, as a vector, we know how to measure the distance between two different vector.

4.2 Shape

In literature there are a lot of edge detector algorithm. After detected the necessary edges of our figure is important to know how to encode all of them.

We could, simply, encode, after all the points have been ordered, the shape by using a parametric curve. The main problem that arise from this solution is that our shape will be encoded in vector of different length. How can we compare vector of different length?

The idea is to consider only the n most interesting points.

So, given two different shapes, rappresented in vector of same length, a common way to compare them is to use euclidean distance. By the way, this solution is sensitive to alignament of values. Let's see an example, in figure 4.2.

Figure 6: Example of two similar shapes not alligned

What we need, so, is a technique that make us able to match points that lie in near time slices. This is called Dinamic Time Warping (DTW).

4.3 Texture

It is not so easy to compare two different texture. As we saw previously, the Tamura features, texture can express the idea of directional, contrast and coarseness. We can use a Gabor Filter (of Dennis Gabor, nobel prize in 1971) in order to understand presence of a particular pattern along a direction.

Indeed it has been found that Gabor Filter are particularly appropriate for texture representation and discrimination.

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi\frac{x'}{\lambda} + \psi\right)$$

where

$$x' = x\cos\theta + y\sin\theta$$

and

$$y' = -x\sin\theta + y\cos\theta$$

In this equation, λ represents the wavelength of the cosine factor, θ represents the orientation of the normal to the parallel stripes of a Gabor function, ψ is the phase offset, σ is the sigma of the Gaussian envelope and γ is the spatial aspect ratio, and specifies the ellipticity of the support of the Gabor function.

By changing frequencies and orientations it has been found that Gabro filters are very useful in order to extract features from an image.

5 Semantic

As we have seen from previous sections, content-based image retrieval is based principally on low level features.

The main question that may arise could be: “is it possible to elevate the image retrieval to semantic meaning?”. The state of the art for reducing the semantic gap can be classified in different way. By considering application domain, they can be classified as those who works on artwork retrieval, scenery image retrieval, WWW image retrieval, etc ....

In some cases, semantic can be easily derived from our daily language. We can describe the “sky” as an uniform blue region in the upper part of an image. Starting from low level image features we can easly build up a simple vocabulary, as in previous example. This vocabulary is also called *object-ontology*, which provides high level concept for querying.

Quantization of color and texture feature is the main idea on which this concept is based. The most used way is to quantize the colors information by its color name (eg: “red”, “blue”, etc...).

The well known color naming system is “CNS” (Color Name System) proposed by Berk, Brownston and Kaufman [3]. It quantizes the HSL (Hue, Saturation, Luminance) space into 627 distinct colors. Hue value is quantized into a set of basic colors. Saturation and Luminance are quantized into different bins as adjectives signifying the richness and brightness of the color. The complete set of generic hue names in CNS is “red”, “orange”, ”brown”, “yellow”, “green”, “blue”, “purple”, “black”, “gray” and “white”: 10 base colors.

There is also a parallel need for texture naming system which would standardize the description and representation of textures. However, texture naming is found to be more difficult than color naming.

5.1 Image browsing/retrieval

In literature there are a lot of works on system for searching images in databases. As we have seen, there are system that uses search by text annotation, other by using feature similarity. A few systems combine text and image data.

Searching using a simple conjunction of keywords and image features is provided in Blobworld [4]. Here the image segment color is translated, through a pre-processing step, in a series of color categories. Thus, searching over all the image features become a easiest search over text annotation. This solution seems to be efficient but potential for more sophisticate use it is very limited.

Another image retrieval system is Webseer [10]. It uses similar ideas of Blobworld in order to evaluate queries over the web. It also indexes the results of some automatically estimated image features. An example of one of these features is to know whether the image is a photograph or a sketch, specially for face finder.

Cascia [5] integrates text and histograms data for image indexing. Other works have tried to use image feature to refine the query process.

Research in content-based image retrieval is a topic on which a lot of industries are investing their money. Indeed there are a lot interests around this subject and is it probable that in the near future we will see a lot of innovations.

In the past, CBIR it has been focused on low level features and image processing. The experience has established that low-level image feature cannot always describe high level semantic concepts. In the future, CBIR will try to narrow down the semantic gap between user's mind (and its richness semantic) and computational power of our computers.

References

[1] Stefania Ardizzoni, Ilaria Bartolini, and Marco Patella. Windsurf: Region-based image retrieval using wavelets. In *DEXA '99: Proceedings of the 10th International Workshop on Database & Expert Systems Applications*, page 167, Washington, DC, USA, 1999. IEEE Computer Society.

[2] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. Modern information retrieval, 1999.

[3] T. Berk, L. Brownston, and A. Kaufman. A new color-naming system for graphics language. *IEEE Computer Graphics Application*, 2(3):37–44, 1982.

[4] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: image segmentation using expectation-maximization and its application to image querying. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(8):1026–1038, 2002.

[5] M.L. Cascia, S. Sethi, and S. Sclaroff. Combining textual and visual cues for content-based image retrieval on the world wide web. In *IEE workshop on content-based access of image and video libraries*, 1998.

[6] Shi-kuo Chang and Shaou-hung Liu. Picture indexing and abstraction techniques for pictorical databases. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(4):475–484, 1984.

[7] Y. Deng and B. S. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(8):800–810, 2001.

[8] John Eakins and Margaret Graham. Content-based image retrieval. *Library and Information Briefings*, 85:1–15, 1999.

[9] Haihua Feng, D.A. Castanon, and W.C. Karl. A curve evolution approach for image segmentation using adaptive flows. In *ICCV 2001. Proceedings. Eighth IEEE International Conference on computer vision*, volume 2, 2001.

[10] C. Frankel, M.J. Swain, and V. Athitsos. Webseer: an image search engine for the world wide web. 1996.

[11] Daniel Heesch. A survey of browsing models for content based image retrieval. *Multimedia Tools and Applications*, 40:261–284, 2008. 10.1007/s11042-008-0207-2.

[12] Ying Liu, Dengsheng Zhang, Guojun Lu, and Wei-Ying Ma. A survey of content-based image retrieval with high-level semantics. *Pattern Recogn.*, 40(1):262–282, 2007.

[13] W.Y. Ma and B.S. Majunath. Edge flow: a framework of boundary detection and image segmentation. In *ICCV 2001. Proceedings. Eighth IEEE International Conference on computer vision*, 1997.

[14] V. Mezaris, I. Kompatsiaris, and M. Strintzis. An ontology approach to object-based image retrieval, 2003.

[15] N. Rasiwasia, J.C. Pereira, E. Coviello, G. Doyle, G.R.G. Lanckriet, R. Levy, and N. Vasconcelos. A New Approach to Cross-Modal Multimedia Retrieval. 2010.

[16] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. A metric for distributions with applications to image databases. In *ICCV '98: Proceedings of the Sixth International Conference on Computer Vision*, page 59, Washington, DC, USA, 1998. IEEE Computer Society.

[17] I.K. Sethi and I.L. Coman. Mining association rules between low-level image features and high-level concepts. In *Proceding of the SPIE Data Mining and Knowledge Discovery*, volume 3, 2001.

[18] Arnold W. M. Smeulders, Senior Member, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1349–1380, 2000.

[19] Hideyuki Tamura, Shunji Mori, and Takashi Yamawaki. Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics*, 8(6), 1978.

[20] Christopher Town and David Sinclair. Content based image retrieval using semantic visual categories. Technical report, 2001.

[21] J. Wang, J. Li, D. Chan, and G. Wiederhold. Semantics-sensitive retrieval for digital picture libraries. *D-LIB Magazine*, 5(11), November 1999.