Image Featuring for Retrieval of Multimedia Documents

Nazmul Haque
Department of CS, RMIT-Australia
Email: haque@mds.rmit.edu.au

and

Morshed U.Chowdhury
School of Computing & Mathematics
Deakin University-Australia
Email: muc@deakin.edu.au

Abstract:

The utilization of massive multimedia documents collections, such as multimedia documents in the global Internet, needs search engines which can rank using both text and image evidence. Massive size and (dynamic) nature of collection can make manual indexing prohibitively expensive in such situations automatic indexing essential. Traditional search engines utilize only text components of multimedia documents. But there are information needs, which require the utilization of image evidence. In this paper, we investigate image-feature for large and heterogeneous collections. Both the nature and complexities of information needs are key elements for an effective retrieval. Retrieval needs that dependent on perceptual similarities (as found in art galleries, building architecture) require the utilization of visual cues. In such situations, the retrieval of multimedia document based on image ranking can provide higher effectiveness. Experimental results show that effectiveness of ranking based on image feature can be higher where perceptual similarities are key elements for retrieval than the retrieval effectiveness of algorithms based on text ranking algorithms

Keywords: Multimedia, Image Featuring, Multimedia document and image retrieval

1. Introduction

One of the main objectives of retrieval systems is the development of algorithms providing fast and effective information about the existence and whereabouts of documents that meet users' information needs [4]. The utilization of multimedia document collections requires effective methods of multimedia documents retrieval [1]. Needs for multimedia document retrieval can be visualized from the availability of multimedia documents on the Internet and cheap distribution media such as CDs and DVDs [2],[3]. Multimedia documents} are computer browse able objects, which integrate multiple (thematically-cohesive) components of a presentation using various information channels or media. Currently text and images are dominant media.

Traditional search engines retrieve multimedia documents based on only text evidence and ignore image evidence. Owing to this constraint search engines fails to meet information need that require visual matching. This is paper we present how visual cues can be utilized for multimedia document retrieval. Visual query processing needs identification of image featuring technique that can be used with the existing computing facilities. In general image processing is computationally expensive process and hence we analyze the effectiveness of low cost image retrieval algorithms suitable for uncontrolled image sources.

2. Retrieval Paradigms

Multimedia documents can be retrieved using (a) query-free (exploration) retrieval paradigm [5] or using a query. In the latter case queries and documents can be matched at a syntactic level or a complex semantic (using an inference engine) level. We confine our scope to syntactic level matching paradigm. A query is an expression representing users' information need. Various types of objects can be matched to retrieve

0-7803-7010-4/01/$10.00©2001 IEEE.

multimedia documents. These objects are known as query terms. The use of textual objects may fail to encode users' information needs when the theme of query is highly related to perceptual similarity of image or photographs. There are two main types of multimedia retrieval algorithms: probabilistic and non-probabilistic. A probabilistic retrieval algorithm is based on the relevance feedback from previous retrieval sessions, whereas a non-probabilistic algorithm makes no assumptions about the previous retrieval. In this paper we adopt a non-probabilistic retrieval algorithm.

2.1 Content based Image ranking algorithms

Over the last few years, content-based image retrieval has received a good inception in retrieval research community. Content-based image ranking is based on matching of image features that are abstractions of image contents. In content-based image retrieval one or more images or images regions are used as image query terms and images are matched indirectly by matching features extracted from both query images and images in multimedia documents. The feature extraction process is automatic; hence it is suitable for a large and dynamic collection.

Content-based image retrieval techniques have been used for the retrieval of homogeneous image collections such is TRADEMARK database or GRAPHICS database [7], [8]. Owing to its inherent advantage over keyword based image retrieval, we use content-based retrieval for collecting image evidence used for multimedia document ranking. The image featuring of heterogeneous collection is however different from homogeneous collections in several dimensions:

- Image-contents are different;
- The image acquisition systems are uncontrolled and hence related image attributes are different. This may include image size, contrast resolution, image storage format, lighting condition and color biasness.

Two digital image files can contain different information while they produce perceptually same or nearly same picture after rendering. In content-based image retrieval we want to retrieve pictures on the basis of visual similarity rather than what is stored in the image file. In viewing this we want to minimize differences among files that contain similar image contents. As shown in the Figure 1, two histograms taken from two different files producing the same image.

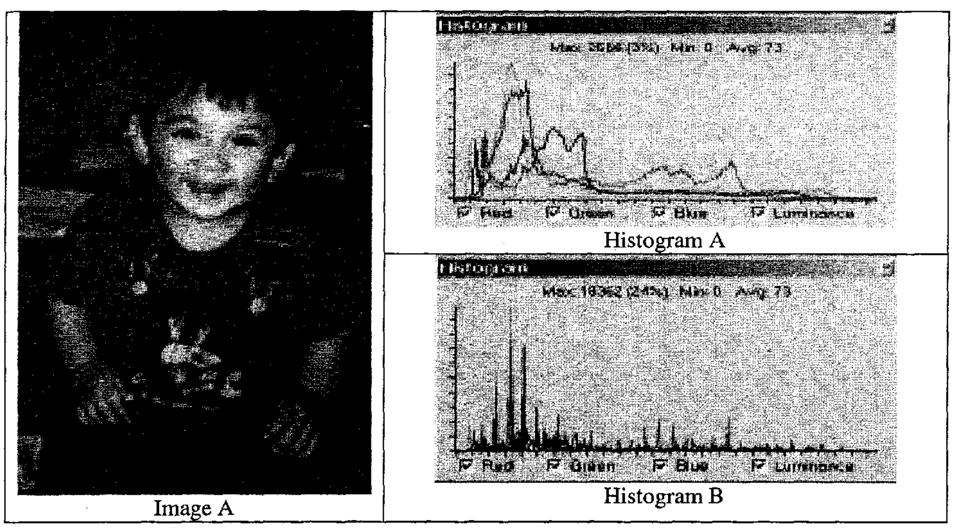

Figure 1: Histogram A and B are taken from two different files producing the same Image A

Image format can be standardized to reduce variations in image files among perceptually similar pictures. As JPEG provide more continuous histograms it is more suitable for image featuring and ranking when simple similarity assessment functions are used. Image dimensions are also important for image featuring. With the increase of image size pixels that represent an element of an object in a scene become more dispersed. Also some details features are captured which are not essential to recognize the objects. For example, pores and hairs of human skin are not essential element for the recognition of a picture of a human. Hence it has been found that image features collected from thumbnail image can become more consistent than those collected from large size images. Thumbnail pictures are also efficient as it contain image data within a smaller number of pixels. Experimental results show that thumbnail of 50x50 resolution can provide more effective ranking. Some image features are dependent on the image dimensions or aspect ratio. In many cases thumbnails with moderate distortions can cope with the variation of aspect ratio.

For an uncontrolled picture sources, color consistency and contrast resolution can introduce variations among different version of the same picture. Human visual systems can perceive a limited amount of variations in gray scale or color wheels. Based on the feature, reduction of contrast resolution can also produce more consistent image feature. Some color spaces are perceptually non-uniform. Hence the optimal contrast solution can be determined by continuously change contrast resolutions. The optimal resolution is also dependent on the color quantization algorithm being used to change colour depth. Some features are less dependent on the original colour variations and contrast resolution. For example, image features extracted by the differential image are not dependent on the colour being dependent but it is dependent on the colour changes. Hence image-ranking algorithms based on differential images have found to produce higher effectiveness. Image contrast resolution is also an important factor for image feature. Too high contrast resolution may put unnecessary details. Hence a higher contrast is not necessarily the optimal one. As shown in the Figure 2, the image is too low contrast can again produce artifacts such as false contours. By keeping the contrast moderate image feature can be improved.

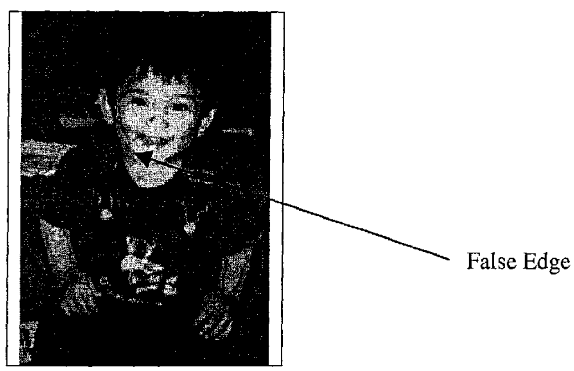

Figure 2: Low contrast image produce artifacts such as false contours

2.2 Image Extraction Algorithms

An image feature is a piece of semantic information extracted from an image. Image features are local, meaningful, extractable part of the images. In content-based image retrievals (CBIR), invariant statistics that can be correlated to the semantics of the image data are algorithmically estimated. An extracted image-feature may represent several different aspects of a perceived feature. Image features can be categorized as either primitive or logical [6], [9]. A primitive feature is a low-level or statistical attribute of an image such as an object boundary or color histogram. Primitive features are automatically extracted directly from the image. A logical feature represents an abstract attribute such as the label 'grass' assigned to a region of an image. Logical features rely on information beyond that contained in the image. Any image processing including image featuring is a computationally expensive process, from both storage and execution point of view, so image featuring with low computational expense is more practicable for image ranking. In this paper we have identified a novel set of primitive image features, which can be extracted easily but still give good retrieval effectiveness. Three kinds of primitive image features are commonly used for image ranking. This includes color, texture and shape. Each of the above class can contain several different physical aspects visuals cues. In the following section we discuss these feature in details. An image feature can be collection from the whole image considering the whole scene as a blob. It can also be collected from regions of regular or irregular boundary to preserve location information. In many situations this can be important. For example, red tints at the top part of a scene may represent a subset or a volcano scene. Finding boundary of an object in a scene it self is a complex research topic. Hence utilization of irregular boundaries is not very common in image featuring. Global histogram capturing feature from the whole image, as blob tends to loose region information.

Orientation and organization of objects can play an important role in image featuring. Certain orientations are natural and hence more likely to come such as organization of body parts along a vertical axis.

Feature collection algorithms should be invariant to image format, especially in a ranking algorithm which ranks images in a heterogeneous collection, where images come from uncontrolled environment. Photographic images collected from an uncontrolled photographing environment can incorporate various kinds of noises, which discourage direct comparison of image data. Two files containing different image data can visually represent the same scene. Image ranking algorithms are sensitive to such variations. A consistent and relatively more continuous color variation in images is important for a color histogram based algorithm as this produces a relatively continuous histogram. Owing to its limit on color variations, a GIF file tends to produce a saw-tooth type histogram while a JPEG produces a relatively continuous histogram. We convert our raw images into JPEG to reduce variations of image data across file formats.

Color Histograms: Color histograms have been used by several researchers in image ranking. It is attractive for it's simplicity but it may loose it's appeal in ranking heterogeneous collections came from uncontrolled image acquisitions. Several types of color histogram are commonly used in image featuring. Various kinds of image processing such as image equalization and contrast enhancement can produce different image files (and consequently different histograms) from the same scene. Hence, a color histogram loose it appeal at once as the image acquisition systems as well as processing can provide different images in multimedia production. The color histograms from a differential image can be free from such problem. An image processed with a differentiation filter (such as a the Robert or Prewitt filters) can be used to record contents' boundary and edge characteristics, which is an important visual are for object recognition. It can also be used to record color variations rather than original color of a scene.

The color component histogram collected can be expressed as

$$h_i = \frac{n_i}{H.W} \quad \forall i = 0 \ldots (C_{Nbit} - 1) \tag{1}$$

where n_i is given by

$$n_{red}(i) = \sum 1 \quad \forall I_{red}(x, y) = i \tag{2}$$

Similarly, N_{blue} and N_{red}-can also be computed. A color histogram of a differential image actually provides light reflectance property of edges of objects in a scene.

Texture Feature: Texture can characterize objects in a scene. It is very difficult to express texture in linguistic terms or a mathematical function. Image texture has been studied by several researchers but there is no consensus that which texture is most suitable for ranking purpose. Texture has been represented by Fourier descriptor, Wavelets and polynomial. Texture can also be represented by run length, which is computed as

$$HRL = \max \| x_2 - x_1 + 1 \| \tag{3}$$

Run length histogram on a differential image can provide an ***edge gradient***, which is another important method defining of the edges of an object.

Shape: Shape is an intuitive feature of objects in a scene though it may not be important to characterize images scenes such as scene of a landscape. Edge detection involves complex image processing algorithms and it is still a research topic. Many people have matched shape indirectly using various kinds of moments such as Zermanic moment, which is independent of various image translation and rotations (that is free body movements).

2.3 Image matching

The similarity of objects in images can be expressed as similarities of corresponding image features. Similarity can be expressed using a mathematical distance function. The retrieval effectiveness of a ranking algorithm mainly depends on the associated distance function and the features. Image features can be in two broad classes – one which is dependent on the feature elements of feature vector which can assume feature elements are orthogonal such that some form ordering feature context are independent. Simple distance functions such as Eucledian, Mantahhan, Squate distance functions are examples of them. Another form distance functions assume that feature context are not independent and tends to associate with them and a quadratic distance function is an example of this.

$$Modified\ Manhattan\ distance = \sum_{i=0}^{M-1} \frac{|Y_i - X_i|}{\max(Y_i, X_i)} \quad \forall X_i \cap Y_i > 0 \qquad (4)$$

Several well known distance functions have been used by CBIR researchers, including modified Manhattan, Monkowski metric, Tonimoto, Correlation, and Canberra distance functions. Investigation of suitable distance functions for image ranking by itself is a separate research topic. Hence, we confine ourselves and try only couple of simple distance functions.

3. Evaluation of Retrieval Effectiveness

The Effectiveness of an image ranking algorithms is an index to reflect the levels of satisfactions of users. It can be assessed qualitatively and quantitatively, but there is no standard way to correlate quantitative and qualitative assessments. Quantitative assessment methods have been used by several researchers to reduce subjectivity. Precision and recall pair at macro level (that is across a series of queries) is usually used to measure ranking effectiveness. Precision is defined as the ratio between the number of relevant items retrieved and the total number of relevant items in the database. Hence, it provides the level of completeness of retrieval. Precision measures the retrieval accuracy and is defined as the ratio between the number of relevant items retrieved and the number of total items retrieved. Hence, precision provides the level of purity of retrieval at a given recall level. Recall measures the ability to retrieve relevant items from the database. It is defined as the ratio between the number of relevant items retrieved and the total number of relevant items in the database. Precision can be represented as

$$Precision = \frac{Number\ of\ relevant\ documents}{Number\ of\ retrieved\ documents}$$

Recall is given by the following expression-

$$Recall = \frac{Number\ of\ relevant\ documents}{Total\ number\ of\ retrieved\ documents}$$

The summary precision-recall curves across all queries were estimated and tested using paired t-test.

4. Experimental Results

A multimedia database of 1880 documents containing (English) natural language text with images was used in the underlying experiment. These documents were sampled randomly from the [2] digital Encyclopedia using a robot program, so the collection was heterogeneous by content. Documents in the collection had well framed pictures collected from uncontrolled sources. 60 image and text queries were made to meet 60 different types of information needs. These queries are collection from several different uses that also browsed gave their relevance judgments. The precision-recall curves presented in Figure 3 and Figure 4 are assessed from the relevance judgments. The experiment results shows (Figure 3 and 4) that a higher precision recall curves is obtainable from multimedia ranking using image evidence than text evidence when the demand for visual information is critical to meet users information needs. As shown in the Figure 3, a little improvement of effectiveness was found with the retrieval using stems, where Loving's stemming algorithms was used. This improvement was statistically insignificant. Several different distance functions were tried for multimedia document ranking using image evidence but modified Manhattan algorithm provided significantly higher effectiveness and it was also consistently across all features. Modified Manhattan distance function was also computationally less expensive then other comparable distance functions. Long horizontal and vertical run lengths imply sharp horizontal and vertical edges, which are commonly found in man, made objects such as building, car, etc. (natural objects have more round or complex edges). As shown in Figure 4, ranking with run-length histograms on differential images provided better recall-precision curves than color histograms, which encode color transitions.

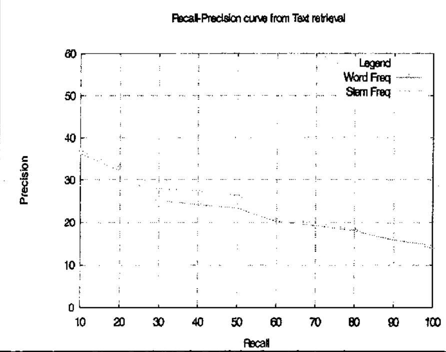

Figure 3 Multimedia document retrieval using text-ranking algorithms from the corpuses of documents.

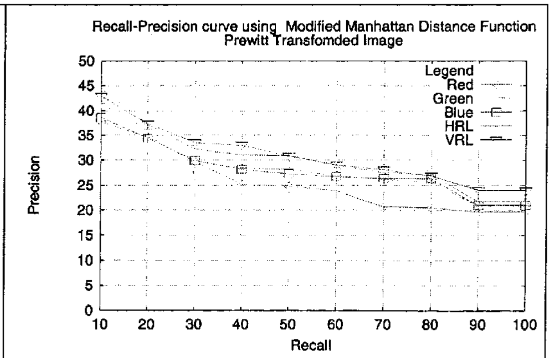

Figure 4 Multimedia document retrieval using image-ranking algorithms.

5 Conclusion

Image evidence can be used to retrieve multimedia documents when a visual aspect of the document is important for retrieval. A significant improvement can be obtained multimedia documents retrieval using image evidence, even with low-level image features applied on differential image that are computationally less expensive. The modified Manhattan distance function is computationally less expensive but it still out performed other distance functions in ranking. Image featuring using differential image on thumbnail picture with JPEG format provide more consistent feature that can provide higher effectiveness. Also differential images obtained from Prewit's transformation provided higher effectiveness than features like raw color component histogram and run-length. The ranking algorithms at six bits contrast resolution provided maximum effectiveness. The collection size used in this study was small, so a conclusive judgment would require bigger database size and larger query sets. The image evidence of multimedia documents may be combined with the text evidence for further improvement in ranking effectiveness. The multimedia ranking using image evidence can be used to rank documents in art gallery database, architectural database. It can also be used for child education and image queries are language neutral.

6 Reference

[1] Martijn Jurjen Hoogeveen, *The viability of Multimedia Retrieval Systems for Marketing and Sales*, Ph.D. Thesis", University of Delft, Beboren te Amsterdam, ISBN 9072125-47-9, 1994

[2] Encarta 98 Encyclopedia, Microsoft, Microsoft Corporation, 1999

[3] Webstar Encyclopedia, 1999, http://www.webstart.com

[4] C J Van Rijsbergen , Information Retrieval, Butterworths, London, 1979, http://www.dcs.gla.ac.uk/Keith/Preface.html

[5] K Parsaye and M Chignell and S Khoshafian and H Wong, *Intelligent Databases: Object-Oriented*, Deductive, Hypermedia Technologies. John Wiley and Sons, Inc., New York, 1989
W.B. Croft and D.J. Harper, Using Probabilistic Models of Document Information, Journal of Documentation, Retrieval without Relevance 35:4, December 1979, 285-295.

[6] Michael Day, *Metadata Image retrieval: combining content-based and metadata-based approaches*, Reports from the Second UK Conference on Image Retrieval held in Newcastle on the 25-26 February 1999,url={http://www.ariadne.ac.uk/issue19/metadata/\#2}

[7] Kato Toshikazu and T Kurita and H Shimogaki, *Intelligent visual interaction with Image Database system - Towards the multimedia Personal Interface*, Journal of information Processing, 1991, Vol14 (2), pp134-143

[8] Kato Toshikazu, Database architecture for content-based image retrieval, 112/SPIE-vol 1662 Image Storage and Retrieval System, 1992

[9] Venkat N Gudivada and Vijay V. Raghavan, *Content-Based Image Retrieval Systems*,IEEE Computer, September,1995