Displays 29 (2008) 451–457

Contents lists available at ScienceDirect

Displays

journal homepage: www.elsevier.com/locate/displa

Perceived image similarity and quantization resolution

Hock Chuan Chan *

National University of Singapore, 3 Science Drive 2, Singapore 117543, Singapore

ARTICLE INFO

Article history:
Received 9 November 2007
Accepted 3 March 2008
Available online 12 March 2008

Keywords:
Contest-based image retrieval
Color histogram
Color quantization
Resolution
Image similarity

ABSTRACT

Color quantization is a key step in content-based image retrieval based on color histograms and is critical to the retrieval performance. An important factor related to color quantization is the quantization resolution. It is important to empirically evaluate how resolution levels affect human perceptions of image similarity. A laboratory study was conducted to analyze the effect of different resolution levels on human judgment of image similarity. The results show that the impact of quantization resolutions on perceived image similarity is not linear. In fact, a logarithm relationship fits the data very well. Furthermore, there is a surprising result that the objective measure of colorloss can predict perceived image quite well. The study provides accurate data for content-based image retrieval researchers to decide on the tradeoff between processing speed (which is affected by the choice of resolution level) and perceived image similarity.

1. Introduction

Advances in hardware technology and retrieval methods have contributed to an increase in the use of digital images. Many image retrieval systems have been developed [16,26,46,55]. A common retrieval approach is content-based image retrieval, which captures visual features automatically or semi-automatically so as create image annotation and index [1,46]. Ideally, the extracted features could identify different objects that correspond to human perception. However, most content-based image retrieval systems are pseudo-object-based. That is, low-level visual features are extracted from images, hoping to capture the high-level semantic information through these features. These features in most cases may not have semantic meanings. However, these features limit the search space and greatly increase retrieval efficiency. Typically, retrieval systems implement some similarity measure between the query image and images in the database based on these low features.

Two processes – feature extraction and similarity measurement – are considered to be crucial for content-based image retrieval systems. A commonly used feature is color, which is used for discriminating between relevant and irrelevant images [46]. Many color-related features can be extracted from an image, such as global and local histograms, and prominent and salient colors. For example, color histogram is used in image retrieval systems by Brunelli and Mich [3], Cinque et al. [10], Faloutsos et al. [17], Lu et al. [27], Ogle and Stonebraker [34], Yoo et al. [55]. Color histogram records the colors present in the image, and their quantities. The reason for its popularity is that it is "invariant to translation, rotation about an axis perpendicular to the image, and change only slowly with rotation about other axes, occlusion, and change of the distance to the object" [50]. Color histogram is also found to show good retrieval performance [14,18].

Digital images are commonly represented as intensity values in the RGB color space. For 24-bit images, there are 16.8 million possible values. To maintain a color histogram on 16 million colors is very large and unnecessary for image retrieval. Typically, the color space is quantized to a much lower resolution. Therefore, the resolution level is critical for histogram retrieval method, as it will affect the efficiency and effectiveness of image retrieval systems [52]. For example, use of 16 bins each for red, green and blue will result in 4096 color bins. For each bin, one color is selected as the representative color of that bin. Two colors are the same if they fall into the same bin. Different quantization resolutions have been used, ranging from 64 to 4096 [9,19,32,49,52,51].

It is important to empirically evaluate the effect of different resolution levels (with different color spaces) on human perception of image similarity of the quantized image to the original image, as well as to investigate if there is any objective measure that can be used to predict human perception. Accurate data will allow system developers to know how their decisions on resolution levels are going to affect user perception of the images. This will allow them to make an informed decision on the tradeoff between system processing speed and user perception.

Section 2 of this paper presents a general model of content-based image retrieval, followed by a brief review of a few human factors study in this area. It highlights the paucity and the need for human factors studies. Section 3 clarifies the main research question, and provides a concise review of the relevant literature

* Tel.: +65 65163393.
E-mail address: chanhc@comp.nus.edu.sg

0141-9382/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.displa.2008.03.002

on color histogram and quantization resolution. The laboratory study is detailed in Section 4. The experiment chooses the most commonly used color space, and a very basic quantization method. It is not an aim in this study to identify the best quantization method or color space. Data analyses and results are presented in Section 5. It was found that the relationship between perceived image similarity and resolution levels can be represented by a simple equation. The conclusion is in Section 6.

2. Human factors in content-based image retrieval

This section provides a brief review of content-based image retrieval, followed by a thorough review of relevant empirical studies involving human subjects. It establishes the need for more empirical studies based on actual human judgment of images, leading to the research question in the next section. A literature review of previous studies indicates that several levels of information can be captured from images. Researchers have proposed models ranging from two levels to several levels [8,23,24,26,38,45,46]. Systems can also be classified into two categories: syntactical and pseudo-object levels [8]. At the syntactical level, a single feature is captured to characterize the images. At the pseudo-object level, systems use combinations of these single features to index the images, hoping to approximately identify the objects by clever use of these features [2].

Typically feature-based systems operate in two distinct phases: the first phase is feature extraction, where the features of each image in the database are extracted, usually represented by a vector, and stored. The second phase is retrieval phase. Typically, a user presents a query. Based on a similarity measure in feature space, a system returns some images. These images could be ranked on some similarity values.

One important issue for researchers of content-based image retrieval systems is finding good models and similarity measures which best mirror human perception for comparing images, and which can be integrated into the systems [29,33]. Many factors influence the way people measure similarity. Picard and Minka [37] identified four such factors: visual features, viewpoint, semantics, and culture. Human factors research in this area is gaining increasing attention. Some examples include Chan and Wang [6], Cox et al. [11–13], Han et al. [21], Minka and Picard [30], Mojsilovic et al. [31], Papathomas et al. [35], Picard [39], Rao and Lohse [40], Rui et al. [44], Scassellati et al. [47]. Generally, research works in this area can be classified in three directions: refining query, changing the traditional display techniques, and understanding human perceptions. A few human factor studies on image similarity are briefly reviewed in the following paragraphs.

A study of human perceptions on shape similarity was conducted by Scassellati et al. [47], who tested seven computational shape matching methods against human perceptions. The methods include algebraic moments, parametric curve distance, turning angle, sign of curvature, and modified Hausdorff distance. The database used in the experiment has over 1400 images. Twenty original query shapes were drawn by the researchers. The shapes varied in complexity, number of angles, perimeter, etc. All query shapes were drawn to be perceptually similar to at least one object in the database. The 40 subjects in the experiment were told to choose up to ten matching objects for each query object, without ranking them. No definition of similarity was offered to subjects, but they were told that the relative size and orientation of the pictures were unimportant. The results showed that none of the methods matches the human selections well, although turning angle method provided the best overall results.

Papathomas et al. [35] studied the importance of using semantic information, query feedback memory and relative judgments of image similarity in an image retrieval system, PicHunter. The experiment involved six system versions which differed from each other along the above three factors. The first dimension has three options – purely pictorial features, purely semantic one, or both. The second dimension differentiates whether the system employs the entire history of user response or only the last response. The last dimension uses relative vs. absolute distance for users to give feedback. In the relative distance mode, a user could select several similar images as feedback. In absolute distance mode, only one image (the most similar) can be selected by user as feedback. Six subjects took part in the experiment. The database included 1500 images. Fifteen query images were employed in the experiment. Subjects were required to find the identical target image. Performance was computed as the average iterations that subjects took to complete each search successfully across the 15 targets. The results showed that the use of long-term memory improves performance significantly when relative distance is employed, but there is no improvement for the absolute-distance versions. One surprising finding of the experiment is that the best performance is achieved when using only semantic cues, with memory and relative similarity judgement. The combination of semantic and visual cues under the same condition only achieves the second best result.

Rogowitz et al. [42] conducted a psychophysical experiment aimed at uncovering the dimensions human observers use in rating the similarity of photographic images. They also compared the results with two algorithmic image similarity metrics, one based on similarities in color histograms and the other based on more sophisticated perceptually-relevant features. Ninety-seven JPEG images were selected from 5000 photographic images. Two psychophysical scaling methods were used to measure the perceived similarity of each image with every other image in the set. In the "table scaling" experiment, 97 test images were printed and placed randomly on a large round table. Subjects were required to arrange these images so that the physical distances between them were inversely proportional to their perceived similarity. No definition of similarity was given to the subjects. Nine subjects served in this experiment. In "computing scaling" experiment, the experiment was conducted on a display monitor. The reference stimulus was presented along the left edge of the display, accompanied by two rows of four test stimuli running horizontally along the display. On each trial, the subjects viewed a randomly-selected reference stimulus and eight test stimuli, selected randomly from the set of 97, and judged which of the eight appeared most similar. Again, no definition of similarity was given to the subjects. Fifteen subjects took part in this experiment. A multidimensional scaling analysis in two and three dimensions was performed for the data. Both table and computing scaling methods produced similar results. This similarity suggests that monitor display and paper prints do not have significant effect on perception similarity. When compared with the algorithmic image methods, the conclusion is that even though visual features do not capture the whole semantic meaning of the images, they do correlate a lot with the semantics.

Overall, more and more research in content-based image retrieval systems have focused to human perception. The empirical experiments not only give us more insight on human perception of some basic features, such as texture [40,31] and shape [47], but also cover the dimensions human observers use in rating the similarity of photographic images [42].

3. Research question

The main research question that we are investigating is: how do human perceptions of image similarity vary with the resolution levels? In addition, other objectives measures will be compared