Comparison of Retrieval Effectiveness of Different Region Based Image Representations

Md Monirul Islam, Dengsheng Zhang and Guojun Lu
Gippsland School of Information Technology
Monash University, Gippsland Campus
Churchill, VIC 3842, Australia
e-mail: {md.monirul.islam, dengsheng.zhang, guojun.lu}@infotech.monash.edu.au

***Abstract*—The choice for low level representation of images makes a significant difference in image retrieval performance regardless whether the retrieval engine works on low level or high level features. The rich diversity of natural images makes this decision crucial. This paper reports some initial result of our research in image semantics. The main objective of this part is to find an effective region based image representation that has highly potential clue for image semantics. Therefore, this paper compares retrieval effectiveness of different region based representations for images. Our study shows that complex techniques like automatic segmentation based methods do not guarantee improved image retrieval performance compared to simple approaches. This paper critically evaluates the merits of these approaches.**

I. INTRODUCTION

Although many research activities in content based image retrieval (CBIR) have been carried out over the past decade, the retrieval precision is still unsatisfactory. One of the main reasons behind this poor effectiveness is the semantic gap - the difference between the low level image representation and the high level human perception of image. We now realize that this gap becomes the critical barrier against achieving the high retrieval effectiveness. One approach to reduce semantic gap is to map the low level image features to high level concepts [1]. In order to improve accuracy of the mapping, we need to know which low level image feature is most effective and simple to use. This paper is part of our research in image semantics and will focus on finding the most appropriate low level features.

Early CBIR systems use global features, like color histogram [2], color correlogram [3], color coherence vector [4], color moments [5], Tamura texture features [6], etc. to represent images. These global features are extracted from entire images and often fail to capture local details that exist in most natural images. As a consequence, dissimilar images varying in local details have similar feature vector representations and vice versa. This anomaly creates difficulties for global representations to have direct relations with image semantics. To address this problem, some techniques divide images into regions with a hope that region level representation would capture local details and have improved clues to image semantics. These techniques are formally known as region based image retrieval (RBIR) techniques. Important issues in RBIR include how images are divided into regions, what features are extracted for regions and how images are compared to each other. Many RBIR techniques such as NeTra [7], Blobworld [8], SIMPLIcity [9] use complex segmentation algorithms to divide images into regions and these are known as segmentation based systems. On the other hand many techniques, classified as block based systems [10], [11], simply divide images into rectangular grid of blocks. Regardless how regions are formed, database images are represented and indexed either directly with feature vectors of constituent regions [7]-[10], or indirectly with indices from a codebook generated from a set of training images using complex vector quantization (VQ) algorithm [11]. As RBIR techniques have shown some success to capture image semantics [1], this paper compares retrieval effectiveness of different region based image representations, especially VQ based representation with others. Although there exist few works that compare VQ based representation with different global representations [12], there is no work that compares VQ based representation with other region based representations. Other than VQ based technique, this paper implements and compares one block based technique and two segmentation based techniques – both use same image representation but differ in query formulation. VQ is usually used in image compression. Images, represented through VQ, are available in compressed form. So, if it is proved that VQ based representation is more effective in retrieval than other region based representations, we can directly use compressed images without going through complex segmentation process. Furthermore, VQ-indexed images have stronger clues for image semantics compared to other region based representations. This also motivates us to compare VQ based representation with others.

The rest of this paper is organized as follows. Section II introduces different RBIR techniques. Section III describes the experimental results. Section IV discusses potential impacts of initial findings on our future research. The paper concludes in Section V.

II. REGION BASED TECHNIQUES

Almost all region based techniques work in the same way – divide images into regions and represent them in terms region features. This section describes different region based techniques which we are going to compare.

1-4244-0983-7/07/$25.00 ©2007 IEEE

A. Segmentation based technique

Segmentation based techniques use complex and automatic segmentation algorithms to segment images into regions. However no known segmentation algorithm can guarantee accurate segmentation of natural images. Thus derived regions do not retain natural shapes of objects. In this experiment, we use one of the state-of-the-art segmentation algorithms, JSEG [13], to segment images into regions. We represent each region with MPEG-7 dominant color descriptor (DCD), $\{(c_i, p_i)\}$, i=1, 2,…, 8, where p_i is the area percentage of the ith most dominant color, c_i, in the region. We define a color map with 64 colors by uniformly partitioning each of three HSV color subspace into 4 equal intervals. We select this quantization level because this level shows the best retrieval performance in our experiment. We calculate histogram of each region using the color map and take the highest eight bins to form its DCD. Once features are extracted, we design two different segmentation based systems to check the retrieval effectiveness. These two systems differ in query formulation and image distance calculation. The first one uses query by region of interest same as NeTra [7] or Blobworld [8] do. We name this 'single region query based segmentation technique'. The distance between the query and a database image is the minimum of distances between the query region and all regions of the database image. To calculate the distance between two regions, we use the distance measure recommended by MPEG-7. The second approach uses global query, like SIMPLIcity [9], – an entire query image is matched with an entire database image. We call this 'global query based segmentation technique'. The individual distances between regions of two images are combined into one distance measure using earth mover's distance (EMD) calculation [14]. Note that segmentation approaches in NeTra [7], Blobworld [8] or SIMPLIcity [9] are different. Our objective is to check the retrieval effectiveness of different representations but not the segmentation algorithms.

B. Block based technique

Block based technique divides each image into a rectangular grid of blocks. The regions constructed in this approach are worse in quality than the regions constructed in segmentation based approaches. But this approach is much simpler than segmentation based technique and suitable for online feature extraction for query image. So, if the retrieval effectiveness of block based representation is proved to be same or at least comparable to segmentation based representation, it would be wise to use simple block based representation instead of complex segmentation based representation. In our present experiment, we divide each image into 4 by 4 blocks because this division shows the best retrieval performance in the experiment. Each block is then considered as a region and is represented with a DCD of dimension 8. Feature extraction for blocks is same as feature extraction for regions of segmentation based techniques. During retrieval, a single block is given as query. Distance calculation is also same as segmentation technique that uses single region query.

C. VQ based technique

VQ is basically used for image compression. The main idea is to form a dictionary of image blocks from a set of training images using a clustering algorithm and encode all images with indices from the constructed dictionary. In VQ, this dictionary is known as codebook and the entries in the codebook are known as codewords. The motivation behind VQ compression is that although there are millions of image blocks, only few of them repeat themselves in different images. So, instead of storing millions of image blocks, these few repeating blocks can be used to represent images compactly with acceptable visual quality. Encoded images actually contain indices from the codebook instead of actual codewords. Thus these images can be regarded as indexed images. The main complexity in the VQ based system is to generate appropriate codebook. However it can be constructed offline. Once the codebook is constructed and we get the indexed images, different types of features can be extracted for image retrieval purpose. In this experiment, we divide images into blocks of size 4 by 4 pixels. From a set 60 training images we constructed a codebook of size 1024 codewords. Blocks of size 4 by 4 pixels from each database image are then replaced with indices of the closest blocks from the codebook. We then extract VQ histogram of size 1024 for each database image. Each histogram is then normalized so that the sum of all bins in it becomes 1. During retrieval we use global query. The distance between two images is the L1 distance between two corresponding histograms.

III. EXPERIMENT

A. Experimental setup

Although research in CBIR is being done for few decades, still there is no bench mark database in this community. However, most of the existing CBIR techniques use different subsets of images from Corel Photo Library to judge their systems. For this project, we have also collected 5100 images from 51 categories of Corel image dataset. Each category consists of 100 images. To have a fair comparison between different systems, we have randomly selected a query set consisting of 160 images – 10 images from each of 16 selected categories.

B. Experimental result

We use conventional precision-recall curves to evaluate different systems. Precision is the ratio of the number of retrieved relevant images to the total number of retrieved images. Recall is the ratio of the number of retrieved relevant images to the total number of relevant images in the database. Fig. 1 shows retrieval results of different systems. Among different region based systems, block based representation shows quite good result at the lowest recall value only. On the other hand VQ based histogram outperforms all other region based representations for all recall values. Two segmentation based systems performs worst among all. Segmentation based systems requires query image to go through complex segmentation process at online. Poor performance of these systems is totally discouraging to use them in real time retrieval systems. On the other hand retrieval effectiveness of VQ based representation is quite promising. Once codebook is generated, they can be used in many ways. This paper shows the simplest use, VQ based histogram. Code words, generated by VQ, retain inter pixel relations. Thus, in some extent, codewords inherently capture spatial information which leads towards

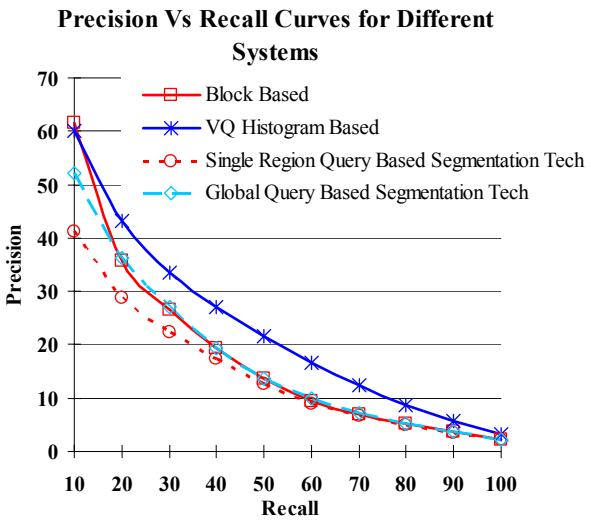

Fig. 1.: Average Precision Recall Curves

higher retrieval precision compared to other representations. In block based representation, dissimilar images often have similar blocks which are represented by quite similar feature vectors with minor differences. Thus query by block often retrieves irrelevant images. A worse problem happens in segmentation based systems. Segments created by segmentation algorithms are more homogeneous in color than blocks resulting from evenly division of images. Thus segments with different shapes and sizes may have also similar color feature vectors. Addressing this problem with shape features will produce misleading result, because it has been proved that no segmentation algorithm can retain natural shapes of objects. Furthermore, Jing et al. [15] found that size information is not also a good feature for segment.

Yet, there is another problem in the opposite direction in block based and segmentation based representations which also lead to low retrieval precision. The problem is that regions having similar perceptual look do not guarantee to have similar feature vector representation and vice versa. To explain both these problems, we have to check what happens at pixel levels. A group of pixels, locating together, individually may have different color values for different pixels. However when we look at the group as a whole, they together make a contrast to our eyes and we often see a single color instead of individual pixel colors. Even two regions having same perceptual look may not have same color composition at pixel levels. How a group of pixel looks depends both on their color composition and spatial organization which is basically their inter pixel relation. Because of these two factors, perceptually different regions may have similar color composition or similar color feature vectors and vice versa. This happens because these feature extraction techniques do not capture inter pixel relation. On the other hand VQ can capture this relation at some level. VQ quantizes groups of pixels instead of individual pixels. During indexing of images with codebook, we may think that groups of pixels from database images are replaced by groups of pixels from the codebook. Thus inter pixel relation is retained in indexed images and thereby any extracted feature from any VQ-indexed image also retain this relation.

IV. DISCUSSION AND IMPACTS ON OUR FUTURE RESEARCH

We have found that VQ based representation that we have implemented has quite promising retrieval effectiveness instead of its simplicity in use. Furthermore, VQ-indexed images in our implementation bear highly potential clues to image semantics. We will explain how it comes true. In our implementation, we have quantized blocks of pixels and derived a codebook of 1024 codewords or quantized blocks. All database images are indexed using these codewords. To understand how VQ-indexed images are related to image semantics, consider the following scenario. Suppose that we have a set of tiger images. Although all of them may not have same background, certain parts of all of them must contain the tiger. Those certain parts must be indexed with same subset of codewords or indices from the codebook. This subset of indices will be definitely a great clue for tiger semantics. This unique and promising advantage can not be achieved with other region based methods; even color indexed images can not achieve this. To understand this, let us have another close look into these representations. To get basic color indexed images, individual pixel values, instead of blocks of pixels, are quantized to form a codebook of colors. Thus pixels in color indexed images are discrete and because of this discreteness, they can not capture any semantic clue as a group. On the other hand, neither of region based representations has a fixed set of feature vectors. Thus images, having same semantic concept, often have regions with different representations. Thus it becomes difficult to correlate them. If we quantize the feature vectors to form a dictionary or codebook of feature vectors and index images with this dictionary of feature vectors, information will be lost in two stages: once when we calculate the feature vectors and again when we quantize them. Thus, this approach will not bring retrieval effectiveness as much as our implementation of VQ does. Direct quantization of regions resulting from segmentation algorithm has another problem. To illustrate the problem, let we have two tiger images. Segmentation algorithm does not guarantee that segmented regions of these two images will be same in number, shapes or sizes. Fig. 2 illustrates this. Therefore, quantization of feature vectors is the only option in such case. But we have already argued that this cannot bring better result than quantization of blocks of pixels.

V. CONCLUSION

In this paper we have shown that VQ based representation performs better than other region based representations. We have also critically analyzed why VQ based representation works better than others. Furthermore, we have also strongly argued that VQ representation has better capability to capture image semantics that other region based representations.

REFERENCES

[1] Y. Liu, D. Zhang, G. Lu and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition,* Vol. 40, Issue 1, pp. 262-282, , January 2007.

[2] M. Swain and D. Ballard, "Color Indexing," *Proceedings of the third International Conference on Computer Vision*, pp. 11–32, 1990.

[3] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu and R. Zabih, "Image Indexing Using Color Correlograms," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 762–768, 1997.

(a) (b) (c) (d)

Fig. 2.: Variations in segmented regions of images with same semantics (a-b: original, c-d: after segmentation)

[4] G. Pass, R. Zabih and J. Miller, "Comparing Images Using Color Coherence Vectors," *Proceedings of the fourth ACM international conference on Multimedia*, Boston, US, pp. 65-73, 1997.

[5] M. Stricker and M. Orengo, "Similarity of Color Images," *Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Databases*, pp. 381–392, 1995.

[6] H. Tamura, S. Mori and T. Yamawaki, "Texture features corresponding to visual perception," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. -8, No. 6, 1978.

[7] W. Y. Ma and B. S. Manjunath, "NETRA: A toolbox for navigating large image databases", *Proc. IEEE International Conference on Image Processing*, Santa Barbara, California, vol. 1, pp. 568-571, Oct. 1997.

[8] C. Carson, M. Thomas, S. Belongie, J. M. Hellerstein, and J. Malik, "Blobworld: A system for region-based image indexing and retrieval," *Proceedings of the Third International Conference VISUAL'99*, Amsterdam, The Netherlands, June 1999, *Lecture Notes in Computer Science*, Vol. 1614, Springer, 1999.

[9] J. Z. Wang, J. Li and G. Wiederhold. "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 9, September 2001.

[10] V. Takala, T. Ahonen and M. Pietikäinen, "*Block-based methods for image retrieval using local binary patterns*," Lecture Notes in Computer Science Vol. 3540, Springer-Verlag, pp. 882-891, 2005.

[11] L. Zhu, C. Tang, A. Rao and A. Zhang, "Using Thesaurus to Model Keyblock-Based Image Retrieval," *IEEE International Conference on Multimedia and Expo*, pp. 170 – 173, 2001.

[12] G. Lu and S. Teng, "A Novel Image Retrieval Technique Based on Vector Quantization", *Proceedings of International Conference on Computational Intelligence for Modeling, Control and Automation*, pp. 36-41, 1999.

[13] Y. Deng and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Transaction on Pattern Analysis and Machine Learning*, Vol. 23, No. 8, pp. 800-810, 2001.

[14] Y. Rubner, C. Tomasi, and L. Guibas, "A metric for distributions with applications to image databases," *Proceedings of IEEE International Conference on Computer Vision*, pp. 59–66, Jan, 1998.

[15] F. Jing, M. Li, H. J. Zhang, and B. Zhang. "Relevance Feedback in Region-Based Image Retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 14, No. 5, May 2004.