COMBINE USER DEFINED REGION-OF-INTEREST AND SPATIAL LAYOUT FOR IMAGE RETRIEVAL

Qi Tian, Ying Wu, Thomas S. Huang

Beckman Institute
University of Illinois at Urbana-Champaign, Urbana, IL 61801
{qitian, yingwu, huang}@ifp.uiuc.edu

ABSTRACT

Content-Based Image Retrieval (CBIR) is one of the most active research areas in recent years. Many visual feature representations have been explored and many systems built. However, in most of current systems, only the global features such as overall color histogram and texture moments are used which ignore the actual composition of the image in terms of internal objects. Although relevance feedback was proposed [1] to incrementally supply more information, they may fail due to the lack of higher-level information about what exactly was of interest. Since automatic segmentation of Region-of-Interest (ROI) is not always reliable, human assistance is necessary. In this paper, a novel approach combining user defined Region-of-Interest and spatial layout is proposed for CBIR. Better capture of image object is achieved by the user rather than the computer. Therefore, more accurate relevance feedback is achieved and thus leads to a more powerful search engine.

1. INTRODUCTION

Two main challenges exist for content-based image retrieval (CBIR): (1) the gap between high-level concepts and low-level features (2) subjectivity of human perception of visual content. Relevance Feedback based on interactive retrieval approach was proposed [1] to take into account the above two characteristics in CBIR. During the retrieval process, the user's high-level query and perceptual subjectivity are captured by dynamically updated low-level feature weights based on the user's feedback.

Incorporating relevance feedback in image retrieval solves the abovementioned problem to some extent. However, only *global* feature of the image such as color histogram and wavelet moments are considered in most of the current systems. While these simple global descriptors are fast and often do succeed in partially capturing the essence of the user's query, they more often fail due to the lack of higher-level information about what exactly was of interest to the user in the query image [2]. The observation that spatial information is a critical component of image description has gradually being noticed by researchers [2, 3, 4, 5, 6, 7].

Moreover, what the user typically thinks of as the "object" is seldom captured by the whole image or its global features. Therefore, image object segmentation is very important to image retrieval. However, in an unconstrained domain, for non-preconditioned images, the automatic segmentation of image object or region-of-interest (ROI) is not always reliable. Although many algorithms for segmentation exist [8, 9, 10, 11], what an algorithm can segment is only regions, but not objects. Therefore, to obtain high-level object, which is desirable in Image Retrieval, human assistance is necessary.

In this paper, a novel approach combining user defined Region-of-Interest and spatial layout is proposed for Content-Based Image Retrieval. Spatial layout is first performed and compared to the conventional approach using global features in the relevance feedback process and then user defined ROI is applied as refined spatial layout approach.

The rest of paper is organized as follows. Our approach is described in Section 2. Section 3 presents the similarity measurement. Experimental results are presented in Section 4. Discussions are given in Section 5.

2. OUR APPROACH

Since it is the user who is most qualified to specify the "content" of the image rather than the computer, it is better to let the user identify the regions in the image that he or she is interested in (the "content"). In our retrieval systems, user is asked to define the ROI with the spatial layout as a search constraint.

The image is first cut as $n \times n$ non-overlapping image blocks. For each image block, feature extraction, i.e., color, texture, is pre-processed off-line for each image block. Selections of 2×2, 3×3, 4×4 and 5×5 layouts are available depending on complexity of the internal structures of the query image. The user defined ROI is then applied. Based on how much percentage of overlapping between the user-defined ROI and the image block, the similarity distance for each image is calculated by linearly combining the individual image block similarity distance.

For convenience, in the rest of this paper the conventional approach that relies on global features of the image is denoted as global approach, our spatial layout approach is denoted as layout approach and user defined ROI combined with the spatial layout is denoted as user defined ROI approach.

3. SIMILARITY MEASYREMENT

Relevance feedback is implemented in our systems. By using relevance feedback, user interacts with the system, indicating which returns, i.e., retrieved images he or she thinks are relevant. Based on the user's feedback, query weights are dynamically updated. In this approach, the high-level concept implied in the relevant images is expected to automatically get refined.

3.1. Spatial Layout Approach

The overall similarity distance D_j is obtained by linearly combining individual similarity distance of the i-th feature in the n-th block:

$$D_j = \sum_n \sum_i W_{n,i} S_j(n,i) \ , \ j=1,\ldots, N \tag{1}$$

where D_j is the overall similarity distance of the j-th image in the database to the query Q, $S_j(n,i)$ and $W_{n,i}$ are the similarity distance and its corresponding weight of the i-th feature in the n-th block of the j-th image in the database, respectively. N is the total number of images in the database. $S_j(n,i)$ is a *Mahalanobis* distance defined as:

$$S_j(n,i) = (\mathbf{x}_{n,i} - \mathbf{q}_{n,i})^T \Sigma_{n,i}^{-1} (\mathbf{x}_{n,i} - \mathbf{q}_{n,i}) \tag{2}$$

where $\mathbf{x}_{n,i}$ and $\mathbf{q}_{n,i}$ are the i-th feature vectors of the n-th block for the j-th image in the database and for the query, respectively. The query is the weighted average of the relevant images in the feature space. $\Sigma_{n,i}$ is the covariance matrix of the i-th feature components in the n-th block of the relevant images. Each element of matrix $\Sigma_{n,i}$ is calculated as:

$$\Sigma_{n,i}(l,m) = \frac{\sum_{k=1}^{NR} V(k)[r_{n,i}(k,l)-q_{n,i}(l)][r_{n,i}(k,m)-q_{n,i}(m)]}{\sum_{k=1}^{NR} V(k)} \tag{3}$$

where $V(k)$ is the preference weight for the k-th relevant image provided by the user in the relevance feedback; $r_{n,i}(k,l)$, $r_{n,i}(k,m)$ are the l-th and m-th component values of the i-th feature vector in the n-th block for the k-th relevant image, respectively. $q_{n,i}(l)$ and $q_{n,i}(m)$ are the m-th and n-th component values of the i-th feature for the query, respectively. NR is the total number of the relevant images and $NR > 1$. $\Sigma_{n,i}$ is an identity matrix if $NR = 1$.

The low-level feature weight $W_{n,i}$ in Eq. (1) is automatically updated by:

$$d_{n,i} = \frac{\sum_{k=1}^{NR} V(k) S_k(n,i)}{\sum_{k=1}^{NR} V(k)} \tag{4}$$

$$W_{n,i} = \frac{1}{d_{n,i}} \tag{5}$$

The higher weight is given to the feature that has the smaller average distance $d_{n,i}$ based on the relevant images. This is because these relevant images are more similar, i.e., have smaller distance in this feature than in other features. The index used in Eq. (4) for S_k is different from the index used for S_j in Eq. (1). The former is the k-th relevant image in the relevance feedback stage, and the latter is the j-th image in the whole image database.

3.2 User Defined ROI Approach

The overall similarity distance D_j of the j-th image in the database for user defined ROI approach is calculated as:

$$D_j = \sum_n \sum_i W'_{n,i} S_j(n,i) \ , \ j=1,\ldots, N \tag{10}$$

This is similar as Eq. (1) except that the weight of the i-th feature in the n-th block is updated by:

$$W'_{n,i} = \lambda W_{n,i} \tag{11}$$

where λ is the ratio of the area that user defined ROI overlaps with the image blocks below it to the area of the each image block. If there is no overlap, λ is set to zero.

4. EXPERIMENTAL RESULTS

4.1 Layout vs. Global Approach

The layout approach (3×3) is first tested over a small database of 142 images containing 7 categories. An example is shown in Figure 1. In Fig. 1(b), the middle block has the largest weight among the 9 image blocks. This indicates that the query images (top 2 in Fig. 1(a)) are most similar in the middle part of the image. This fact can be easily verified in Fig. 1(a).

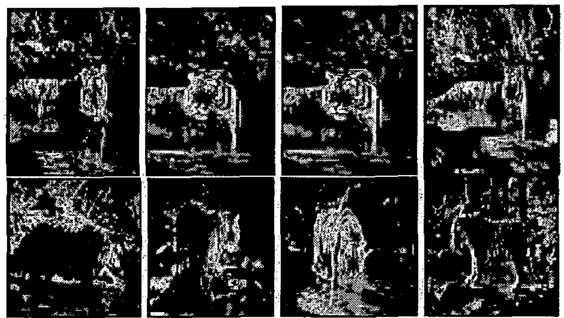


Figure 1(a) Top 8 retrieval results (rank from left to right and from top to bottom, the top 2 are the query images)

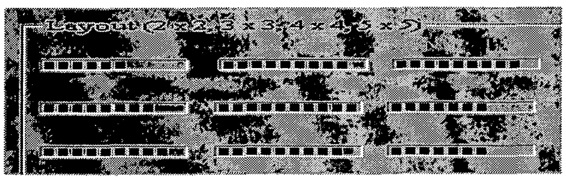
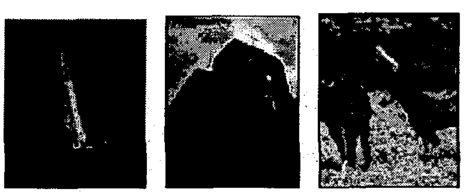


Figure 1(b) Layout (3×3 weights for each image block

The following results show the performances of the layout approach and global approach in the relevance feedback. The layout is 3×3.

The test is over COREL dataset that contains more 17,000 images. Three example query images are shown in Figure 2. Table 1 shows the retrieval results. The layout approach performs better for boat (Fig. 2(a)) than the global approach while the global approach performs better for building (Fig. 2(b)) than the layout approach. This result indicates that the layout approach might have better capability to capture the local details than the global approach, and vice versa. This complementary property is further verified by the experiments over bark, brick, church painting, car, flower and airplane images that are shown in Figure 3. The layout approach performs better for car, flower and airplane while the global approach performs better for bark, brick and church painting.

Since the layout approach shows complementary property to the global approach, a unified approach combining the global and layout approach will certainly be desirable. Table 2 shows an example of retrieval results for Fig. 2(c) using the global, layout and unified approaches. As we expected, the unified approach achieves the best results.

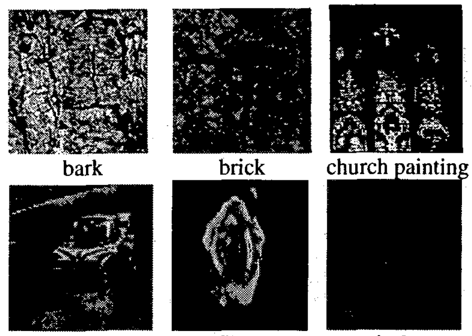


(a) boat (b) building (c) horse

Figure 2 Query images (COREL dataset)

Table 1 Retrieval results (rf – Relevance Feedback)

# of hits in top 20		0 rf	1 rf	2 rf	3 rf	4 rf
Boat Fig. 2(a)	Global	8	10	12	14	14
	Layout	14	18	18	18	18
Building Fig. 2(b)	Global	9	17	18	19	19
	Layout	7	12	13	14	14



bark brick church painting

car flower airplane

Figure 3 Sample images

Table 2 Retrieval results (rf – Relevance Feedback)

# of hits in top 20		0 rf	1 rf	2 rf	3 rf	4 rf
Horse Fig 2(c)	Global	9	17	18	18	19
	Layout	7	12	13	14	14
	Unified	13	17	18	20	20

4.2 User Defined ROI vs. Global Approach

Figure 4(a) shows an example of user defined ROI approach. A rectangular window defines the ROI, the car in this example. Figure 4(b) shows the corresponding image block weights. The layout is 5×5.

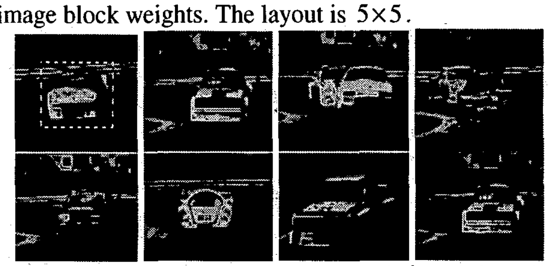


Figure 4(a) Top 8 retrieval results (rank from left to right and from top to bottom, top 1 is the query image with user defined ROI window)

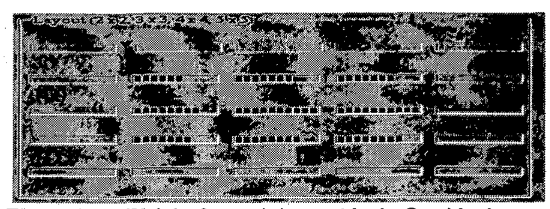


Figure 4 (b) Weight for each image block. Outside the user defined ROI window (shown in the 1st image in Fig. 4(a)), the image block weight is zero.

Table 3 shows the retrieval performance of the global, layout and user defined ROI approaches. For the first 5 query images, the user defined approach performs best and the layout approach ranks 2nd. The global approach has the worst performance. Therefore when decomposing the image information into local details, i.e., focus on the ROI rather than the whole image, better results are obtained by the user defined approach and the layout approach. For the last three uniform images, the global approach performs best. Therefore. When the global information is most important, the global approach performs best.

Table 3 Number of hits in top 15 for image object retrieval

Query	Global	Layout	User Def.
Tiger	7	7	13
Car	2	6	8
Airplane	5	12	12
Flower	9	14	12
Bird	15	15	15
Bark	8	5	4
Brick	13	9	9
Straw	8	6	5

5. DISCUSSIONS

In this paper, a novel approach by combining user defined ROI and spatial layout is proposed for Content-Based Image Retrieval (CBIR). Compared to the global approach that extracts the features from the whole image, the layout approach extracts features from each image block and automatically combines the individual image block similarity based on the relevant images. Experimental results show that the layout approach is more capable of capturing the image details than the global approach while the global approach shows the complementary property in capturing the global information. Therefore, the unified approach combining the global approach and layout approach achieves more satisfactory results than the global approach and layout approach alone.

In user defined ROI approach larger weights are given to the image block that contains the ROI and thus better retrieval results can be obtained than the global approach. User defined ROI leads to a more *user-centric* and a more powerful search engine. User defined ROI approach can be regarded as a refined spatial layout approach since more emphasis is focused on the ROI. However, this approach has some disadvantage due to its spatial dependency. For example, if the user is searching for a tiger in the left corner of the image, the tigers located at the right corner of the images are hard to retrieve. The region-matching algorithm can be employed to solve this problem. The user defined ROI is moving over the whole image block by block. For every block, a similarity distance is recorded. The minimum similarity distance is indexed as the output similarity distance for the image.

The computation complexity of this approach is greatly increased as $O(n^2)$ with the increasing dimension n of layout. Normally that 3×3 layout has a good trade-off between image details and computation complexity from our experience.

In our current spatial layout, there is no overlap among image blocks. However, this work can be easily extended to overlapped layout as in [6]. Moreover, arbitrary shape segmentation instead of rectangular window needs to be investigated in our future work.

6. ACKNOWLEDGEMENTS

This work was supported in part by National Science Foundation Grant CDA 96-24396, and in part by NSF EIA-99-75019

7. REFERENCES

[1] Y. Rui, T. S. Huang, et al., "Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval", *IEEE Trans. on Circuits and Video Tech.*, Special Issue on Segmentation Description, and Retrieval of Video Content, 8(5), Sep. 1998.

[2] B. Moghaddam, H. Biermann and D. Margaritis, "Defining Image Content With Multiple Regions-of-Interest", pp. 89-93, *Proceedings of IEEE Workshop on CBALVL*, 1999.

[3] J. R. Smith, *Integrated Spatial and Feature Image Systems: Retrieval, Compression and Analysis.* PhD Thesis, Columbia University, 1997.

[4] M. Striker, A. Dimai, "Color Indexing with Weak Spatial Constraints", *Proc. of SPIE 2670*, 1996.

[5] I. K. Sethi, I. Coman, et al., "Color-WISE: A System for Image Similarity Retrieval Using Color", *Proc. of SPIE Storage and Retrieval for Image and Video Databases*, vol. 3312, Feb. 1998.

[6] M. Striker, A. Dimai, "Color Indexing with Weak Spatial Constraints", *Proc. of SPIE 2670*, 1996

[7] I. K. Sethi, I. Coman, et al., "Color-WISE: A System for Image Similarity Retrieval Using Color", *Proc. of SPIE Storage and Retrieval for Image and Video Databases*, vol. 3312, Feb. 1998.

[8] M. Lybanon, S, Lea, and S. Himes, "Segmentation of Diverse Image Types Using Opening and Closing", *Proc. of IEEE ICIP'1994.*

[9] X. Q. Li, Z. W. Zhao, et al., "A Fuzzy Logic Approach to Image Segmentation", *Proc. of Int. Conf. on Image Proc.*, 1994.

[10] W. Y. Ma and B. S. Manjunath, "Edge Flow: A Framework of Boundary Detection and Image Segmentation", *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1997.

[11] A. C. She and Thomas S. Huang, "Segmentation of Road Scenes Using Color and Fractal-Based Texture Classification", *Proc. ICIP*, Austin, Nov. 1994.