

Information Selective Image Retrieval

Guoping Qiu and Mai Yang

School of Computer Science and Information Technology
The University of Nottingham, United Kingdom

Jiwu Huang

School of Information Science and Technology
Sun Yat-sen University, Guangzhou 510275, China

Abstract: In this paper, we introduce an information theoretic concept to image retrieval by first developing a novel method to compute the amount of information contained in image parts, and then using the informative-ness of the image parts to develop a flexible information selective image retrieval (ISIR) technology. We present experimental results which show that the new method is effective and is especially useful for retrieving similar objects from a variety of different backgrounds.

Keywords—Information theory, entropy, image database, content-based image retrieval

I. INTRODUCTION

Content-based image retrieval (CBIR) has been studied extensively for sometime now [1]. Although much progressive has been made, it remains to be a challenging subject. The difficulty comes from several fronts, including the extraction of reliable and meaningful visual representation features, intelligent content recognition and user-friendly human computer interaction.

As a starting point for building an automatic image retrieval system, image contents have to be represented by robust feature vectors in such a way that distinctive contents should be represented by distinctive numerical values. The feature vectors should be compact, discriminative and can be easily exploited by machines for automatically recognizing the contents of images. What is more, these features should make “semantic”¹ sense such that they can be easily used to help users to interpret the visual meaning of the image contents at a cognitive level. All these are very challenging tasks and many researchers in the past have developed various solutions with different degrees of success. In the CBIR literature², image content descriptors used by researchers include various forms of color features, texture features and others [1].

For any given image, different parts of the image may have different degree of importance or usefulness for recognizing the image content. There seems to have evidence to suggest that biological visual systems have selective attention mechanisms that process features of different saliencies differently in the process of scene analysis and object recognition. In the computational vision field, researchers have been developing computational models of visual attention and computational methods for salient visual

features, see, e.g. [2]. In the context of CBIR, if we can grade various parts of an image in terms of their importance in conveying information about the whole image, then it would enable the development of more flexible and effective image retrieval technologies.

In this paper, we present a novel method for CBIR. We introduce an information theoretic concept to image retrieval by first developing a novel method to compute the amount of information contained in image parts, and then using the informative-ness of the image parts to develop a flexible information selective image retrieval (ISIR) technology. The organization of the rest of the paper is as follows. In section 2, we present a method for computing the information content of image parts. In section 3, we present a flexible scheme that uses the information content selectively to retrieve images. Section 4 presents experimental results and section 5 concludes the paper.

II. COMPUTING INFORMATION OF IMAGE PARTS

According to Shannon information theory [3], the entropy of a random variable measures the amount of information contained in the random variable. Let X be a random variable, x some real world realization of X , $p(x)$ the probability density function, then the amount of information or entropy of x is

$$E(x) = -\int p(x) \log(p(x)) dx \quad (1)$$

Given that our variable is image parts, which means that the dimensionality of the variable is very high. As a consequence, it is very difficult to estimate the probability density function of the variable and hence difficult to use equation (1) to measure the information content of image parts. Although it is possible to use techniques such as Gaussian mixture model (GMM) or independent component analysis (ICA) to estimate $p(x)$, in this paper we opt for a more intuitive solution and will report results of using GMM and ICA in the future.

The key to our current solution is the following intuitive interpretation of information content: If an event is happening frequently and with regularity, that means it is more predictable, which in turn means that there is not much surprising elements in predicting the event. In the terminology of information theory, the amount of information of such an event is small. On the other hand, if an event happens very rarely and without any predictability, then there will be large amount of surprising elements in trying to predict its happening, in other words, the amount of information

¹ We use this term loosely.

² Although many technologies equally applicable to CBIR have been developed in other aligned fields such as pattern recognition, computer vision, machine learning etc., CBIR and these fields are intrinsically linked and heavily burrows technologies developed in those fields.

contained in such an event is high. We want to mathematically quantify such an intuition.

For simplicity and easy presentation, we first divide an image into non-overlapping image patches as illustrated in Figure 1 and compute the information content of each patch. It is worth noting that the scheme presented here should be regarded as an illustration of principle and there is significant scope for different practical implementations. Assume that the m th patch $x(m)$, which is represented by low-level features such as color, texture or a combination of color and texture plus other features. Our task is to determine the amount of information contained in $x(m)$, in the context of a single image. We can also compute the information of $x(m)$ in the context of the whole image database. We call the formal local information and the later global information. Both local and global information will be useful in developing flexible image retrieval methods, but we will only concentrate on the use of local information in this current paper.

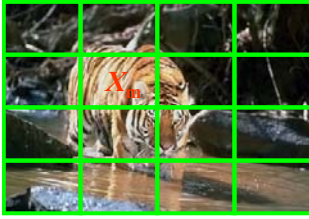


Figure 1: Divide an image into patches and compute the information content of each patch.

Our method is again build on following intuitions. In the context of a given image, for a given patch, if there are many more patches in the image that are very similar, this means the particular given patch is not rare, in other words, the information contained in the patch is low. On the other hand, if a given patch is rare and there are no other patches that are similar to the given one, then this patch contains high amount of information. In the following we present one possible way to quantitatively describe such an intuition.

We first form a similarity matrix $\mathbf{D} = \{d(m, n)\}$, where $d(m, n)$ is some similarity measure between patch $x(m)$ and $x(n)$, which can be measured by some metric norms, such as the Euclidean distance or normalized correlation. We then sum each row of \mathbf{D} , to form an N -dimensional vector, \mathbf{S} , where N is the total number of patches in the image

$$s(m) = \frac{1}{N} \sum_{n=1}^N d(m, n) \quad (2)$$

Note that a larger $s(m)$ means that there are fewer patches that are similar to patch m , while a smaller $s(m)$ means that there are more patches that are similar to patch m . In other words, a larger $s(m)$ indicates patch m contains more information and a smaller $s(m)$ indicates that patch m contains less information. Let $S_{\max} = \max\{s(m)\}$ be the largest of all $s(m)$'s, we quantitatively measure the amount of information contained in patch m as follows:

$$I(m) = \left(\frac{S(m)}{S_{\max}} \right)^{\gamma} \quad (3)$$

In (3), the exponent γ serves as an adjustable variable which adjusts the relative amount of information assigned to the blocks. It serves to make the patches more or less distinctive. Figure 2 shows an example, where the image is divide into 24 patches and each patch is represented by a color histogram for computing its information using (2) and (3).

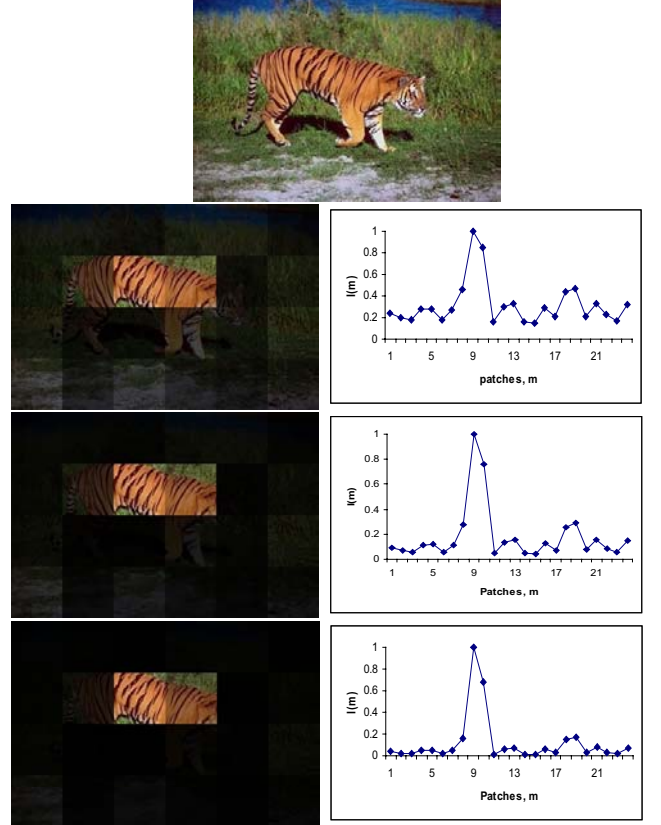


Figure 2: The amount of information contained in image parts for different γ 's. 2nd row $\gamma=3$, 3rd row $\gamma=5$ and 4th row $\gamma=7$. Left column shows the image patches. A brighter patch indicates higher information content. Right column shows the numerical values of the information in each block.

III. INFORMATION SELECTIVE IMAGE RETRIEVAL

Based on the information measure computed from previous section, we develop an information selective image retrieval technique. The rationale is as follows. For each image in the database, we first compute the amount of information contained in their parts. It is reasonable to assume that image parts with higher information content would be more important to the image or image parts with different information content will have different importance and parts with similar information content will have similar importance in content recognition. In the context of example-based image retrieval, we can compare the similarity of two images by comparing their parts with similar amount of information. This approach makes sense. For example, for images with a small object placed against a large (somewhat homogeneous) background, then the information content of the parts containing the foreground object will be higher than that of those parts in the background. If such an image is used as the querying image and the desire is to find similar images in the

database, it is reasonable that we should compare the foreground of the querying image with the foreground of the database images, and do the same for their respective backgrounds. Such operation can in fact be automatically accomplished using the information content computed from last section, and we call such approach information selective image retrieval (ISIR).

To perform ISIR, we compare the parts of two images with similar information; this can be achieved using a technique similar to the frequency layered color indexing (LCI) [4], but instead of using the frequency content of the image parts, we here use the information content. Again, for illustration purpose, we assume two image X and Y are decomposed into their constituent parts, $X = [x(1), x(2), \dots, x(n)]$, $Y = [y(1), y(2), \dots, y(n)]$, each part can be a non-overlapping block like those in section 2 or divided by image segmentation and each parts with complex shape. In this paper, we will use the simple approach of section 2 to divide the image into parts and compute each part's information content. Note that the two images can have different number of parts. For $x(i)$ part of X , there is an information content score $I_x(i)$, and similarly for $y(i)$ part of Y there is an information content score $I_y(i)$. Obviously, both $I_x(i)$ and $I_y(i)$ are continuous values ranging from 0 to 1. We therefore first quantize the information score to a few discrete values. Although it is possible to design the quantizer according to the statistics of these information scores, but in this paper, we take a simple approach by uniformly dividing the range into several intervals (in fact we use 4). If two parts of the images have the same quantized information score, then we regard them as equally important.

Let Q denote the information score quantizer, and $Q(I_x(i)) = m$, $Q(I_y(i)) = n$, then all parts of X with a quantized score of m form the set X_m , and all parts of Y with a quantized score of n form the set Y_n , i.e.,

$$\begin{aligned} X_m &\leftarrow x(i) \quad \text{if} \quad Q(I_x(i)) = m \\ Y_n &\leftarrow y(i) \quad \text{if} \quad Q(I_y(i)) = n \end{aligned} \quad (4)$$

From the set X_m we extract image representation features $F[X_m]$, and from the set Y_n we extract the same type of image features $F[Y_n]$. The features to be used can be those commonly used in the CBIR literature, such as color histogram, various texture descriptors etc. To compute the similarity of the images X and Y , we use following ISIR image similarity measure

$$D(X, Y) = \sum_k \lambda_k \|F[X_k] - F[Y_k]\| \quad (5)$$

where λ_k is a weighting constant which selectively uses the information. A larger λ_k places more importance on the subset of parts with a quantized information score of k . In words, ISIR first divides the images into constituent parts, and computes each part's information content. The information score is then quantized. All parts within the same image that have the same quantized information score are grouped together. For each group of patches, image features are computed. To compare the similarity of two images, we compare the image features of each group of patches from one

image with that of its corresponding group of patches in another image with the same information score.

IV. EXPERIMENTAL RESULTS

In the experiment, we have used an image database containing about 4000 color photographs. The images all have a size of 256 x 384 pixels. Each image is divided into 24 64x64 blocks. We used the method of section 2 to compute the information content of each block. We used a 4-level quantizer to quantize the information score of each block. Therefore image blocks in each image are divided into 4 groups. For representing each group of patches, we construct a 64-bin color histogram using the HMMD color space and the quantization table suggested in MPEG-7 document. As a comparison, we have also implemented MPEG-7 color structure descriptor [5].

In the first experiment, we tested the precision performance. We have collected many image pairs containing similar objects. We then put one of the pair in the database and used the other as querying image. Some results are shown in Figure 3 and Figure 4.

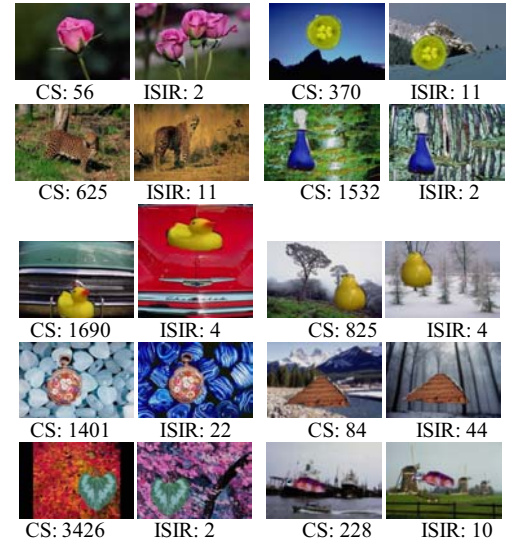


Figure 3: Examples of retrieval precision results. The numbers are the ranks of the target image in the retrieval list. CS = MPEG-7 CS method. ISIR – our new method, for $\gamma = 3$, $\lambda_1 = 0.4$, $\lambda_2 = 0.2$, $\lambda_3 = 0.4$, $\lambda_4 = 0$, and λ_1 corresponds with the highest and λ_4 corresponds with the lowest information scores.

In Figure 3, it is seen that if the querying image and the target image have very different backgrounds, then ISIR can significantly outperform the MPEG-7 CS by appropriately weighting the features according to the information content of the patches. However, if both the background and the foreground objects of the querying image and target image are very similar, then MPEG-7 CS worked as well as ISIR and in some case can outperform ISIR, as illustrated in Figure 4.

In a second experiment, we tested the recall performances of ISIR using the same 4000-image database. Again, we used MPEG-7 CS as a comparison benchmark. The querying images are the same objects against different backgrounds. Two example objects are shown in Figure 5. For each object, there are 100 images on various backgrounds. We used each of the 100 image as querying image and measure how many

images containing the same object returned with high ranks. We use following recall measure [4]

$$AR(L) = \frac{1}{100} \sum_{j=1}^{100} (|\{q(j) | Rank(q(j)) < L\}|) \quad (6)$$

where $q(j)$ is the j th querying image. $AR(L)$ measures on average how many of the images containing the same object as the querying image are returned within the first L images. A higher value of $AR(L)$ indicates a better performance. Figure 6 shows results of retrieving two objects. It is seen that compared with MPEG-7 CS, our method gives better performances. Figure 7 shows an example of retrieved images.

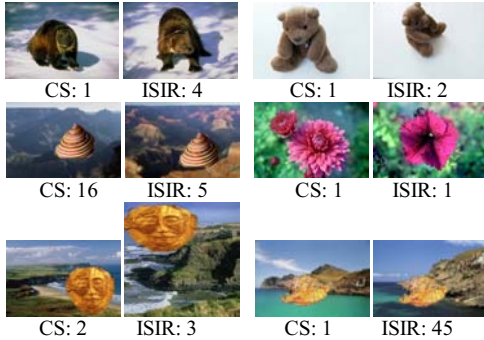


Figure 4: Examples of retrieval precision results. The numbers are the ranks of the target image in the retrieval list. CS – MPEG-7 CS method. ISIR – our new method, for $\gamma = 3$, $\lambda_1 = 0.2$, $\lambda_2 = 0.4$, $\lambda_3 = 0.4$, $\lambda_4 = 0$, and λ_1 corresponds with the highest and λ_4 corresponds with the lowest information scores.



Figure 5: Examples of recall querying images. Each object is placed in 100 different backgrounds. The goal is to use one of the images as querying and retrieve the other 99 images containing the same object.

V. CONCLUDING REMARKS

In this paper, we have presented a novel approach to content-based image retrieval. We introduced the concept of information selective image retrieval by first computing the information contents of image parts and then selectively utilized the image features according to their informative-ness. We have presented experimental results which shown that the technique is effective, and is especially useful for retrieving similar objects in cluttered backgrounds. Our future work will include alternative techniques for computing the information measures, the use of global information and new ways to use the information measures for image retrieval.

REFERENCES

- [1] A. W. M. Smeulders et al, "Content-based image retrieval at the end of the early years", IEEE Trans PAMI, vol. 22, pp. 1349 - 1380, 2000
- [2] L. Itti, C. Koch, E. Niebur, A Model of Saliency-Based Visual Attention for Rapid Scene Analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 11, pp. 1254-1259, Nov 1998 q
- [3] T. M. Cover, Elements of Information Theory, Wiley, 1991
- [4] G Qiu and K-M Lam, "Frequency layered color indexing for content-based image retrieval", IEEE Transactions on Image Processing, vol. 12, no.1 pp. 102 -113, 2003
- [5] MPEG-7, ISO/IEC JTC1/SC29/WG11/N4062, March 2001

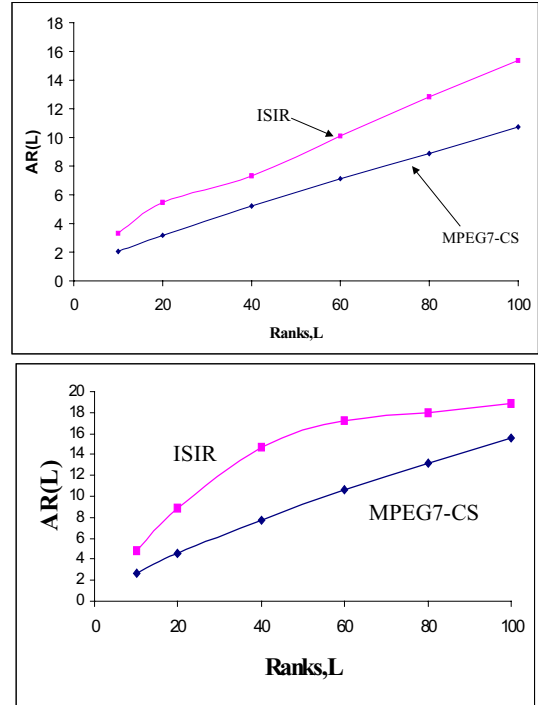


Figure 6: Recall performances of MPEG-7 CS and our new method, for $\gamma = 7$, $\lambda_1 = 1$, $\lambda_2 = 0$, $\lambda_3 = 0$, $\lambda_4 = 0$, i.e., we only used those patches containing the highest information.



Figure7: Top: Querying image and its patches' information content. Bottom: Returned images (order of similarity from left to right top to bottom).