Journal of Theoretical and Applied Information Technology
10th April 2013. Vol. 50 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

IMAGE RETRIEVAL USING MUTUAL INFORMATION

[1]HOU YUANYUAN [2]FAN XUNLI [3]LI JIANGHONG

[1] School of Power and Energy, Northwestern Polytechnical University, Xi'an 710072, China
[2] School of Information Science & Technology, Northwest University, Xi'an 710127, China
[3] School of Power and Energy, Northwestern Polytechnical University, Xi'an 710072, China

ABSTRACT

In this paper, we study an information theoretic approach to image similarity measurement for content-base image retrieval. In this novel scheme, similarities are measured by the amount of information the images contained about one another – mutual information (MI). The given approach is based on the premise that two similar images should have high mutual information, or equivalently, the querying image should convey high information about those similar to it. The method first generates a set of statistically representative visual patterns and uses the distributions of these patterns as images content descriptors. To measure the similarity of two images, we develop a method to compute the mutual information between their content descriptors. Two images with larger descriptor mutual information are regarded as more similar. We present experimental results, which demonstrate that mutual information is a more effective image similarity measure than those have been used in the literature such as Kullback-Leibler divergence and L_2 norms.

Keywords: *Mutual Information(MI), Image Retrieval, Kullback-Leibler Divergence(KLD), Entropy, Content-based Image Indexing and Retrieval(CBIR)*

1. INTRODUCTION

Content-based image indexing and retrieval(CBIR) has been a topic of considerable interest in various applications [1,2]. One of the key technological challenges in CBIR is image similarity measurement. A general CBIR system makes use of different type of queries such as query by example image, sketch or region and provides relevant images from a given database, based not exclusively on textual annotation but on a similarity function using low-level features. CBIR represent image content with representative feature vectors employing color, texture and shape characteristics [10,11,12]. Since image contents invariantly uses low-level visual features, computational image similarity is measured via their low-level similarity, often using some notions of distance measure. It is well known that image retrieval is (in most cases) a high level activity, and low-level similarity does not necessarily correspond to high level similarity perceived by human observers. In recent years, various techniques that uses machine learning to incorporate high level knowledge into image similarity measures have been introduced in the literature, yet a simple, effective and straightforward image similarity measure is still desirable in many applications. In the literature, many approaches have been used to measure the similarity of low level visual features for image retrieval. In a comprehensive empirical evaluation work, [2] studied 9 different similarity measures.

Although Kullback-Leibler divergence (KLD) has been used to measure image similarity [2], a related information theoretic measure, the mutual information (MI) [3], seems to have been ignored by researchers. One of the reasons may be that MI

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

is much more difficult to compute. However, we believe MI should have many advantages as a similarity measure. Firstly, MI measure general statistical relations between variables. Secondly, MI is invariant to monotonic linear transforms performed on the variables. Thirdly, MI has an intuitive similarity explanation. MI measures the average reduction in uncertainty about x that results from learning the value of y, or vice versa. Equivalently, MI measures the amount of information that *x* conveys about *y*.

The organization of the paper is as follows. In section 2, we introduce the general idea of image representation as image similarity measurement. In section 3, we present a method for the estimation of mutual information between images based on their content descriptors. Section 4 presents experimental results, which demonstrate that information theoretic image similarity measures are superior to other related similarity measures. Section 5 concludes the paper.

2. IMAGE REPRESENTATION USING DISTRIBUTIONS OF STATISTICAL VISUAL PATTERN

How to characterize the visual content of images is a nontrivial problem. Past methods mostly use color and texture features [1]. It was shown that how image is represented has a direct bearing on the performances of different similarity measures [2]. In the absence of a standard approach to image content representation, we use an approach that learns statistically representative visual patterns that jointly characterize the texture and color distributions of images.

The method first de-composes a given image into a multilevel Gaussian pyramid [4]. At each level, the image is represented in an opponent color space. Let $\{I_l(x, y)\}=\{rl(x, y), gl(x, y), b_l(x, y)\}$ be the *l-th l*evel image in an image pyramid. These images are then transformed into an opponent space. We use the YC_bC_r space. At each level, images patches (blocks) of $m\times n$ pixels, are formed. Let $\{B_l(i, j)\} = \{(Y_l(i, j), C_{bl}(i, j), C_{rl}(i, j)) \mid i = 1, 2, \ldots, m, j = 1, 2, \ldots, n\}$ be an image patch at level l, for each block, we form two appearance vectors as follows:

$$\begin{cases} a_l = \left\{ \left. \dfrac{Y_l(i,j)}{M_{Bl}} \right|_{\forall i,j} \right\} \\ c_l = \left\{ \left. \dfrac{c_{bl}(2i,2j)}{M_{Bl}}, \dfrac{c_{rl}(2i,2j)}{M_{Bl}} \right|_{i=1,2,\ldots,\frac{m}{2},\, j=1,2,\ldots,\frac{n}{2}} \right\} \\ M_{Bl} = \dfrac{1}{m\times n}\displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n} Y_l(i,j) \end{cases} \quad (1)$$

Where a_l is the achromatic appearance vector, and c_l is the chromatic vector of block $B_l(i, j)$. For computational convenience, we use a single uniform block size *(4×4 pixels)* for all levels of the pyramid, which covers an area of *4×4, 8×8, 16×16, 32×32, …* pixels in the original image depending on which level the vector comes from. Vector quantization [5] is then used to design one codebook for the achromatic vectors (of all resolutions) and one codebook for the chromatic vectors. Clearly the achromatic vectors are of *16* dimensional and the chromatic vectors are of 8 dimensional.

To design the codebook, we used over *200 512×512* pixels true color images (note these images are not part of the testing images reported in section 4). To generate training samples, these images are represented in multilevel Gaussian pyramids and training vectors (over 15 millions) come from all levels of the pyramid.

As expected the achromatic codewords contain much stronger spatial patterns. Since these patterns (codewords) have been designed using a statistical means [5], they should form statistically representative visual patterns, which can in turn be used to characterize image contents.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Once the codebooks have been designed, we can then use them to represent the image content. An image is indexed using the following procedure:

Step 1: Decompose the images into an L-level Guassian pyramid [4].

Step 2: For each level l *(l = 1, 2, …, L)*, divide the image into *4×4* blocks (which can overlap each other). For each block, calculate the achromatic and chromatic vectors according to (1), and encode these vectors according to their respective codebooks.

Step 3: For each level l *(l = 1, 2, …, L)*, construct an achromatic visual pattern histogram and a chromatic visual pattern histogram. These histograms record the frequency each codeword has been used to encode the blocks in the image.

Step 4: Construct the final image descriptor by concatenating all levels' achromatic and chromatic histograms.

3. COMPUTING MUTUAL INFORMATION BETWEEN HISTOGRAMS

Once the images have been represented in visual pattern histograms as described in section 2, we can compare their similarity by comparing their histograms. The straightforward comparison is to compute their L1 or L2 distance, or their KLD. In this section, we present a method to compute the mutual information between two histograms as constructed in section 2.

3.1 Mutual Information Definition For Histogram

We first briefly review some key concepts in Shannon information theory. Instead of describing general definitions, we give our definition specifically for our problem. Let $X = (x_1, x_2, \ldots, x_n)$, $Y = (y_1, y_2, \ldots, y_n)$, be two n-bin histograms, x_i and y_i are the *i-th* bin count of their respective histograms. The mutual information[3] between *X* and *Y* can be defined as:

$$\begin{aligned} I(X;Y) &= H(X)+H(Y)-H(X,Y) \\ &= H(X)-H(X\mid Y) \\ &= H(Y)-H(Y\mid X) \end{aligned} \qquad (2)$$

Where *H(X)* is the Shannon entropy of histogram *X*, computed on the probability distribution of the bin counts. This is not to be confused with the image's entropy, which can be calculated directly from the bin counts. Let the values of the bin counts of *X* be a, $0 \le a \le 1$. The probability distribution of the bin count is defined as follows:

$$\begin{aligned} P(X=a) &= \frac{1}{n}\sum_{i=1}^{n}\delta(x_i-a) \\ H(X) &= -\sum_{x\in A_X} P(x)\log(P(x)) \\ &= -\int P(X=a)\log(P(X=a))da \end{aligned} \qquad (3)$$

Where δ is the Dirac delta function. In discrete implementations, a takes non-continuous values and the integral is replaced by summation.

The entropy and probability distribution of *Y* can be similarly defined. It is also possible to replace the Dirac delta with a Gaussian which becomes a Parzen method [7]. *H(X|Y)* is the conditional entropy, which is based on the conditional probabilities *P(X = a | Y = b)*, the probability of *X* has a bin count value a given that the corresponding bin in Y has a bin count value *b*.

The entropy can be interpreted as an uncertainty measure. Equation (2) can therefore be translated to mean the amount of uncertainty about histogram *X* minus the uncertainty about *X* when histogram *Y* (the content of *Y*) is known. Therefore, the mutual information *I(X; Y)* is the amount by which the uncertainty about *X* decreases when *Y* is known, or equivalently, the amount of information *Y* contains about *X*. Mutual information is symmetric, that is *I(X; Y) = I(Y; X)*, therefore, mutual information is also the amount of information *X* contains about *Y*.

Mutual information can also be defined as joint probability distribution of the histograms as:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$\mathrm{I(X;Y)} = \iint P(x=a, y=b) \log\left(\frac{P(x=a, y=b)}{P(x=a)P(y=b)}\right) dadb \quad (4)$$

3.2 Joint Probability Estimation For Histogram

To estimate the joint probability *P(a, b)=P(X=a, Y=b)* for histograms *X* and *Y*, the most straightforward approach is to compute the co-occurrence matrix of the corresponding bin count values. The entries of the co-occurrence matrix, *CM(a, b)*, records the number of times the bin counts in *X* having a value a coincide with the corresponding bin counts in *Y* having a value of *b*. This is similar to the co-occurrence matrix used in texture characterization [6]. Mutual information based images registration used a similar method to estimate the joint probability of two images [7]. The joint probability is obtained by dividing the entries by the total number of entries in the co-occurrence matrix. The marginal distribution of *P(X=a)* and *P(Y=b)* can be obtained by summation over the rows or columns of the co-occurrence matrix.

The joint probability can also be estimated using Parzen window technique[8]. Let $v_i = (x_i, y_i)$ and $u_j = (x_j, y_j)$ are pairs of corresponding bin values in X and Y, the joint probability of pair $v_i = (x_i, y_i)$ can be defined as:

$$P(v_i) = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{2\pi\sigma^2} e^{\frac{1}{2\pi\sigma^2}\left\|v_i - v_j\right\|^2}\right) \quad (5)$$

4. EXPERIMENTAL RESULTS

Because mutual information is about how much information one image (histogram) contains about the other, it can be used to measure the image similarity in content-based image retrieval. Content-based image retrieval is about using example image to find similar images from the database.

If an image in the database conveys high information about the example image, or vice versa, then they must be similar images in some sense. In other words, similar images should have high mutual information. Therefore use mutual information for images retrieval amounts to finding images in the database having high mutual information with the querying example image.

We have performed various experiments to study the performance of mutual information measure for image retrieval. Our database is a subset of the Corel stock photo data. We use 3 levels of Guanssian pyramid for each image, and the codebook sizes for the chromatic and achromatic patterns are both 64.

In our experiment, we further define two similarity measures based on mutual information. A normalized mutual information (NMI) is defined as:

$$NMI(X,Y) = 1 - \frac{I(X;Y)}{\max\{H(X), H(Y)\}} \quad (6)$$

The information distance measure (MID) [9] is defined as:

$$MID(X,Y) = H(X,Y) - I(X;Y) \quad (7)$$

Where *H(X, Y)* is the joint entropy, MID satisfies the axioms for a distance:

$$\begin{cases} MID(X,\ Y) \ \geq\ 0 \\ MID(X,\ X) \ =\ 0 \\ MID(X,Y) = MID(Y,\ X) \\ MID(X,Y) + MID(Y,\ Z) \geq MID(X,\ Z) \end{cases}$$

As a comparison, we also implemented the KLD measure, which is defined as follows:

$$KLD(X,Y) = \sum_{i=1}^{n} x(i)\log\left(\frac{x(i)}{y(i)}\right) \quad (8)$$

We implemented a normalized correlation measure defined as:

Journal of Theoretical and Applied Information Technology
10th April 2013. Vol. 50 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$\begin{cases} NC(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(\frac{x_i - m_x}{\sigma_x})(\frac{y_i - m_y}{\sigma_y}) \\ \sigma_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - m_x)^2} \\ m_x = \frac{1}{n}\sum_{i=1}^{n} x_i \end{cases} \quad (9)$$

We also use Euclidean distance as a measure of the difference between two histograms:

$$ED(X,Y) = \sqrt{\frac{1}{n}\sum_{i=1}^{N}(x_i - y_i)^2} \quad (10)$$

Let Q_i be the *i-th* query image, $i = 1, 2, \ldots, K$, and let $Q_i(1), Q_i(2),\ldots,Q_i(N_i)$ be N_i "correct" answers to the query *Qi*.

The metrics that were employed to evaluate the performances, were the commonly used metrics of precision and recall[13]. We define the following average accumulated recall measures:

$$RC(l) = \frac{1}{K}\sum_{i}\left(\frac{|\{Q_i(j) \mid rank(Q_i(j)) < l\}|}{N_i}\right) \quad (11)$$

RC(l) measures the fraction of correct answers returned within the first l returned images. Obviously the larger RC(l) is, the better the performance.

We also defined following precision measure:

$$PR(l) = \frac{1}{K}\sum_{i}\left(\frac{|\{Q_i(j) \mid rank(Q_i(j)) < l\}|}{l}\right) \quad (12)$$

Again, the larger PR(l) is, the better the performance.

Ground truth data for image retrieval is still something to be desired in the research community. In the absence of a universal testing ground truth database, we used the categories in the Corel data as ground truth. We regard images within a single subdirectory (100 images each) belong to the same class. In other words, using one of these images as a query example, the correct answers (ground truth) for this query are other 99 images from the same subdirectory.

Although the validity of such data is debatable for many categories in the database, for example, many images belonging to the same category can vary significantly whilst images belong to different categories can have much in common, however, this is a problem beyond the scope of this paper.

We choose 3 categories of images as query example and target images, which are which are horses, dinosaurs, and flowers. The total size of the image database is 10,000, and the retrieve results are shown in Fig.1, Fig.2 and Fig.3.

Fig. 1. Retrieved Results Using The Top Left Image As The Query Image

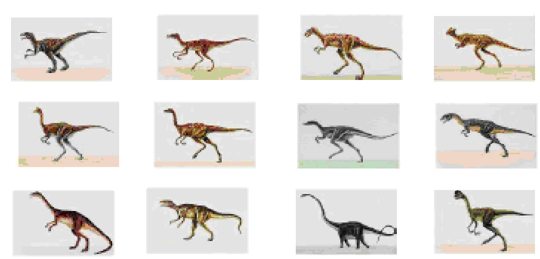

Fig. 2. Retrieved Results, With The Top Left Image As The Query Image.

Journal of Theoretical and Applied Information Technology
10th April 2013. Vol. 50 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Fig. 3. Retrieved Results With The Top Left Image As The Query Image.

For each of the similarity measures, mutual information (MI), normalized mutual information (NMI), mutual information distance (MID), Kullback-Leibler divergence (KLD), normalized correlation (NC), and Euclidean distance (ED), we performed 300 queries, that is each images in the given categories has been used as a query. The recall and precision performance of these queries are shown in Fig. 4 and 5.

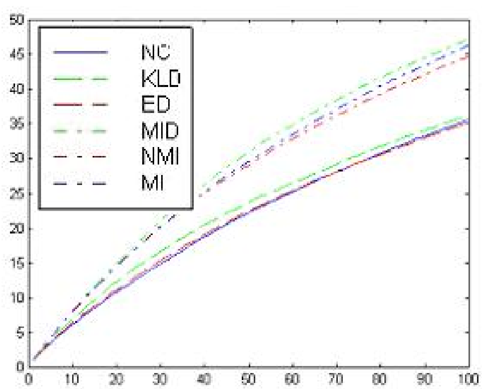

Fig. 4. Recall Performance Of Various Measures

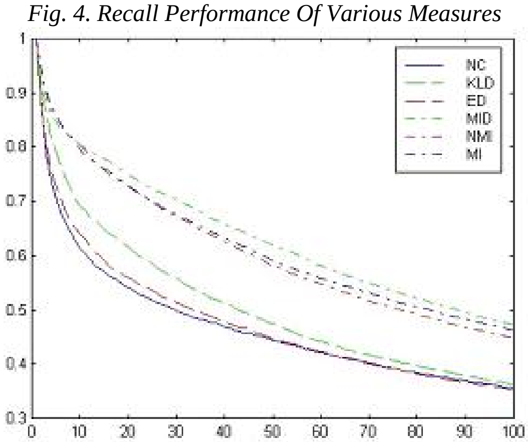

Fig. 5. Precision Performance Of Various Measures

From Fig. 3 and 4, it is seen that mutual information measures all performs similarly and all outperforms other measures. It is interesting to note that the mutual information definition *X* and *Y* is equal to the KLD (4) and the KLD (8) are closely related. In fact, the mutual information *I(X; Y)* between the joint probability function *P(X, Y)* and the product of the probability function *P(X)* and *P(Y)*. However, our results clearly show that mutual information is a more effective similarity measure than KLD.

5. SUMMARY

In this paper, we have introduced a method for image retrieval using mutual information. We justify the use of mutual information as a similarity measure for images based on the argument that mutual information captures higher orders of statistical relations between images. Two images having higher mutual information means that knowing one image conveys more information about the other, hence mutual information is a natural measurement of image similarity. We have developed a method that computes the mutual information between the visual pattern histograms of images. Our experimental results further demonstrate the superiority of mutual information measures over other similarity measures widely used in the literature. This method can be easily extended to other image content descriptors, such as color correlogram, MPEG-7 color descriptors and other widely used image descriptors.

REFRENCES

[1] Minakshi Banerjee, Malay K. Kundu, Pradipta Maji, “Content-based image retrieval using visually significant point features”, *Fuzzy Sets and Systems*, Vol.160, No.23, 2009, pp3323–3341

[2] Panagiotis Sidiropoulos, StefanosVrochidis, Ioannis Kompatsiaris, “Content-based binary image retrieval using the adaptive hierarchical density histogram”, *Pattern Recognition*, Vol..44, No.4, 2011, pp739–750

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[3] Maciej A. Mazurowski, Joseph Y. Lo, Brian P. Harrawood, "Mutual information-based template matching scheme for detection of breast masses: From mammography to digital breast tomosynthesis", *Journal of Biomedical Informatics*, Vol.44, No.5, 2011, pp815–823

[4] Thomas Hurtut, Yann Gousseau, Francis Schmitt, "Adaptive image retrieval based on the spatial organization of colors", *Computer Vision and Image Understanding*, Vol.112, No.2, 2008, pp101–113

[5] Feng Wanga, Cheng Yang, Zhiyi Lin, Yuanxiang Li, Yuan Yuan, "Hybrid sampling on mutual information entropy-based clustering ensembles for optimizations", *Neurocomputing*, Vol.73, No.1, 2010, pp1457–1464

[6] Guoping Qiu, Jeremy Morris, Xunli Fan, "Visual guided navigation for image retrieval", *Pattern Recognition*, Vol.40, No.6, 2007, pp1711-1721

[7] Umarani Jayaraman, Surya Prakash, Phalguni Gupta, "An efficient color and texture based iris image retrieval technique", *Expert Systems with Applications*, Vol.39, 2012, pp4915–4926

[8] Vretos N., Solachidis V., Pitas I., "A mutual information based face clustering algorithm for movie content analysis", *Image and Vision Computing*, Vol.29, 2011, pp 693–705

[9] Wu Sitao, Rahman M.K.M., Tommy W.S. Chow, "Content-based image retrieval using growing hierarchical self-organizing quad tree map", *Pattern Recognition*, Vol.38, No.5, 2005, pp707–722

[10] Tanzila Saba, Amjad Rehman and Ghazali Sulong, "An Intelligent Approach To Image Denoising", *Journal of Theoretical and Applied Information Technology*, Vol 17. No. 1, 2010, pp32-36

[11] Suresh G., Epsiba P., Rajaram M., Sivanandam S.N., "Image And Video Coding With A New Wash Tree Algorithm For Multimedia Services", *Journal of Theoretical and Applied Information Technology*, Vol 12. No. 1, 2010, pp53-59

[12] Stefanos Vrochidis, Anastasia Moumtzidou, Ioannis Kompatsiaris, "Content-based binary image retrieval using the adaptive hierarchical density histogram", *Pattern Recognition*, Vol. 44, No.4, 2011, pp 739–750

[13] Goutte C., Gaussier E., "Aprobabilistic interpretation of precision, recall and f-score, with implication for evaluation", *Proceedings of European Conference on IR Research(ECIR*), Santiago de Compostela, Spain, 2005, 345–359