Interactive Image Retrieval with Wavelet Features

Malay K. Kundu[1], Manish Chowdhury[1], and Minakshi Banerjee[2]

[1] Machine Intelligence Unit, Indian Statistical Institute, Kolkata 700 108, India
[2] RCC Institute of Information Technology, Kolkata 700015, India
{malay,manish_t,minakshi_r}@isical.ac.in

Abstract. This paper presents an iterative Content Based Image Re-trival(CBIR) system with Relevance Feedback (RF), in which M-band wavelet features are used as representation of images. The pixels are clustered using Fuzzy C-Means (FCM) clustering algorithm to obtain an image signature and Earth Mover's Distance (EMD) is used as a distance measure. Fuzzy entropy based feature evaluation mechanism is used for automatic computation of revised feature importance and similarity distance at the end of each iteration. The performance of the algorithm is tested on standard large multi-class image databases and compared with MPEG-7 visual features.

Keywords: Relevance Feedback, Edge Histogram Descriptor, Color Structure Descriptor, Feature Evaluation Index.

1 Introduction

Content Based Image Retrieval (CBIR)[1] which aims at retrieving similar images from a large database by measuring visual similarities between the query image and database images based on automatically derived features like color, texture etc. has become a potential area of research. The performance of a CBIR system strongly depends upon the availability of suitable features for proper representation of semantic aspects automatically and suitable similarity distance measures.

The major problem in the conventional CBIR, is the semantic gap between visual perception of a scene and feature based representation of the scene image. To bridge this gap, user's feedback has come into picture in an interactive manner which is popularly known as Relevance Feedback (RF)[2]. Although several solutions for suitable feature extraction and RF have been proposed, but results achieved so far are not fully upto the user's expectation. Efficient feature extraction mechanism which keeps computation minimum and suitable RF mechanism is still an important research issue.

Wavelet transform based low level features provide an unique multiresolution analysis and highly suitable for characterizing textures of images[3]. Wang *et al.*[4] have used a 2-step algorithm using subband variances of wavelet coefficients for image retrieval applications. Cheng *et al.*[5] have used M-band wavelet transform based histogram for content based retrieval of aerial images.

The relevance feedback which is a post-retrieval step for enhancing the retrieval results[1] mainly uses two approaches (a) "weighing approach" where higher weight is given to emphasize important features (b) "probability approach" where the information representing the query image is modified to make it more similar to the positive images than negative images, according to user feedback. Zin *et al.*[6] have proposed a feature re-weighting technique by using both the relevant and the irrelevant information, to obtain more effective results.

We propose a fuzzy entropy based RF method using EMD by considering both relevant and irrelevant images to adjust automatically the weights of the M-band wavelet features. The results are compared with MPEG-7 visual features which has become almost standard norms for the evaluation of newly proposed image features for CBIR system.

2   M-band Wavelet and Distance Measure

1. **M-band Wavelet:** An M-band wavelet system forms a tight frame for functions $f(x) \in L^2(\Re)$ with (M-1) unitary wavelet filters[7] is given by

$$f(x) = \sum_k c(k)\varphi_k(x) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} \sum_{i=1}^{M-1} M^{j/2} d_{i,j}(k) \psi(M^j x - k) \qquad (1)$$

where φ_k is the scaling function and ψ are the wavelet functions respectively and are associated with the analyzing (or synthesizing) filters. The wavelet coefficient is $c(k) = \int f(x)\varphi(x-k)dx$ and the expansion coefficient of coarser signal approximation is $d_{i,j}(k) = \int f(x)M^{j/2}\psi(M^j x - k)dx$. The scaling function and (M-1) wavelet function also define a Multiresolution Analysis(MRA)and the subspaces form an orthogonal decomposition of functional space for $L^2(\Re)$ defined by the space spanned by the translate of $\psi_i(x)$ for fixed j and $k\epsilon\mathbf{Z}$ as $W_{i,j} = Span\{\psi_{i,j,k}\}$.

2. Earth Mover's Distance: Earth Mover's Distance (EMD) by its definition extends to distance between sets or distributions of elements, thereby facilitating partial matches[8]. If p_i be the centroid of each cluster of a signature $P_i = \{(p_1, w_{p_1}), ..., (p_m, w_{p_m})\}$ with m clusters and w_{p_i} the weight of each cluster, similarly q_i be the centroid of each cluster of a signature $Q = \{(q_1, w_{q_1}), ..., (q_n, w_{q_n})\}$ with n clusters and $D = [d_{ij}]$ is the ground distance matrix, where $d_{ij} = d(p_i, q_j)$. Computing EMD thus becomes finding a flow between p_i and q_j which minimizes the overall cost. A flow between P and Q is any matrix $F = (f_{i\ j}) \epsilon R^{m \times n}$. Thus, EMD is defined as the work normalized by the total flow i.e. $EMD(P,Q) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} d(p_i,q_j)f_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}}$ subject to the constraints (a) $f_{ij} \geq 0, 1 \leq i \leq m, 1 \leq j \leq n$, (b) $\sum_{j=1}^{n} f_{ij} \leq w_{p_i}, 1 \leq i \leq m$, (c) $\sum_{i=1}^{m} f_{ij} \leq w_{q_j}, 1 \leq j \leq n$, and (d) $\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} = \min(\sum_{i=1}^{m} w_{p_i}, \sum_{j=1}^{n} w_{q_j})$.

3   Proposed Technique

(1). 1D, 16 tap 4 band orthogonal filters with linear phase and perfect reconstruction for the MRA are used as a kernel for wavelet filter. (2). Before

decomposing the image, RGB color plane of the image is converted into YCbCr color plane. (3). After color plane tranformation, each image plane is decomposed into M×M channels without downsampling and then energy feature at each pixel is computed. (4). FCM clustering algorithm is used in each pixel to obtain the signature. (5). EMD distance is computed between the query image and the stored image signature in the database and the images are displayed in the first pass. (6). Feature Evaluation Index (FEI) are calculated from the marked relevant and irrelevant set of images and recomputes the weighted EMD over each iteration.

3.1 Color-Texture Feature as Signature

Motivated from the idea that human visual system divides an image into several bands, than actually visualizing the complete image as a whole. M-band wavelet filters are used which are essentially frequency and direction oriented band pass filters. Wavelet transform is applied to Y, Cb and Cr planes where decomposition over the intensity plane characterizes texture and over chromaticity planes characterizes color. For each of 16 subbands, absolute Gaussian energy for each pixel is computed over a neighborhood and the size of which is determined using a spectral flatness measure(SFM) given by the ratio of arithmetic mean and the geometric mean of the Fourier coefficients of the image.

$$energy_{m_1,m_2}(i,j) = \sum_{a=1}^{N} \sum_{b=1}^{N} |Wf_{m_1,m_2}(a,b)| \, G(i-a, j-b), \qquad (2)$$

Here, N is the neighborhood size and $1 \leq m_1 \leq M, 1 \leq m_2 \leq M$ while Wf_{m_1,m_2} is the wavelet transform coefficient obtained by row-wise convolution using the filter H_{m_1} and column-wise convolution with the filter H_{m_2}. The nonlinear transform is succeeded by a Gaussian low-pass (smoothing) filter of the form $G(x,y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(1/2\sigma^2)(x^2+y^2)}$, where σ defines the spatial extent of the averaging filter. We use a neighborhood size of 11×11 for SFM between 1 and 0.65, 21×21 for SFM between 0.65 and 0.35 while 31×31 for 0 to 0.35.

3.2 Iterative Computation of Weighted Based on Feature Relevance

Perceptual importance as used in the JPEG 2000 is Y : Cb : Cr = 4 : 2 : 1. Here, the weights are chosen heuristically. However, an automatic scheme which chooses the weights depending on the color-texture complexity of the image will certainly boost the performance of the CBIR system.

Combining information of relevant and irrelevant images marked by the users, the fuzzy features evaluation index (FEI) is computed considering a pattern classification problem. If $C_1, C_2, ..., C_m$ are m pattern classes in N dimensional features space where class C_j contains, n_j number of samples. The features values along the qth co-ordinate along classes C_j are assigned as standard S-type membership function between 0 and 1[9]. Entropy(H) of C_j gives the measure of intraset ambiguity is given as $H(A) = (\frac{1}{n_j \ln 2}) \sum_i S_n(\mu(f_{iqj})); \, i = 1, 2...n_j$, where $S_n(\mu(f_{iqj})) = -\mu(f_{iqj})\ln\mu(f_{iqj}) - \{1-\mu(f_{iqj})\}\ln\{1-\mu(f_{iqj})\}$ is the Shannon's function (μ). $H_{min} = 0$ for μ=0 or 1, $H_{max} = 1$ for μ=0.5.

The FEI_q for the qth component is defined as $FEI_q = \frac{H_{qjk}}{H_{qj}+H_{qk}}$. H_{qk} is the entropy of class C_k along q^{th} dimension over n_k number of samples. H_{qjk}(interset ambiguity) is entropy along q^{th} dimension combining classes C_j and C_k with $(n_j + n_k)$ number of samples. Lower value of FEI_q, indicates better quality of importance of the qth feature because FEI_q should be decreasing after combining C_j and C_k as the goodness of the qth features in descriminating pattern classes C_j and C_k increases[9]. The weight w_q is a function of the evaluated (FEI_q) is $w_q = F_q(FEI_q)$.

H_{qj} is computed from the set of relevant images $I_r^{(q)}= \{I_{r1}^{(q)}, I_{r2}^{(q)}, I_{r3}^{(q)},...,I_{rk}^{(q)}\}$. Similarly, H_{qk} is computed from the set of irrelevant images where, $I_{ir}^{(q)} = \{I_{ir1}^{(q)}, I_{ir2}^{(q)}, I_{ir3}^{(q)},...,I_{irk}^{(q)}\}$. H_{qkj} is computed combining both the sets. Images are ranked according to EMD distance. The user marks the relevant and irrelevant set from 20 returned images, for automatic evaluation of FEI.

Presently an image I is represented as a signature $P = \{(p_1, w_{p_1}), (p_2, w_{p_2}), ..., (p_m, w_{p_m})\} = \{(p_i, w_{p_i})\}_{i=1}^{m}$with m clusters. The cluster centroid p_i constitutes the wavelet features over 16 subbands of each Y, Cb and Cr plane and obtained with $p_i = [p_{i_Y}, p_{i_{Cb}}, p_{i_{Cr}}]$ which may be further represented as $[f_{1Y}, ..., f_{nY}, f_{1Cb}, ..., f_{nCb}, f_{1Cr}, ..., f_{nCb}]$. Here p_{i_Y}, $p_{i_{Cb}}$ and $p_{i_{Cr}}$ are the local energy values computed overall subbands of each Y, Cb and Cr planes respectively, for e.g $p_{i_Y} = [f_{1Y}, ..., f_{nY}]$ where $p_i \epsilon R^N$ and $N = 3n$ is the feature dimension. Here $n = 16$ and $w_i \geq 0$.

The features considered to compute the feature evaluation index along each component plane for e.g plane Y are $F_Y = \{f_{i_{1Y}}, ..., f_{i_{qY}}\}$, where $i = 1, 2, ..., m$ (m clusters) and $q = 1, 2, ..., n$ (n subband features). Similarly features are considered for Cb and Cr plane. The FEI_{qY} for the component feature q of Y plane are $FEI_{qY} = \frac{H_{qtotal}}{H_{qRel}+H_{qIrrel}}$. Here, H_{qRel}, H_{qIrrel} and H_{qtotal} are the entropies along the qth dimension of relevant, irrelevant and total returned images respectively. And the $(FEI)_{qCb}$ and $(FEI)_{qCr}$ of other two planes are computed similarly. The overall weight factor for the Y plane is given by $W_Y' = \sum_{i=1}^{m} \sum_{q=1}^{n} (FEI)_{qY}$, similarly W_{Cb}' and W_{Cr}' are computed and the weight for the Y plane is $W_Y = \frac{W_Y'}{W_Y'+W_{Cb}'+W_{Cr}'}$.

Similarly W_{Cb} and W_{Cr} are computed. Multiplying with the weight actually transforms the representative points p_i but leave its distribution fixed. The EMD is now computed between the transformed signature maps. $P' = \{g(p_i), w_{pi}\}_{i=1}^{m}$ and $Q' = \{g(q_j), w_{qj}\}_{i=1}^{n}$, where p_i and q_i are the centroids of the query and database images and g is the weight updating factor computed after each iteration. The $EMD_g(P,Q)$, where $g \epsilon (W_Y, W_{Cb}, W_{Cr})$ is computed from the work flow i.e. $WORK(F, P', Q') = \sum_{i=1}^{m} \sum f_{i,j} d(g(p_i), g(q_j))$.

The transformed p_i and q_j is represented as $p_i' = [W_Y p_{i_Y}, W_{Cb} p_{i_{Cb}}, W_{Cr} p_{i_{Cr}}]$ and $q_j' = [W_Y q_{j_Y}, W_{Cb} q_{j_{Cb}}, W_{Cr} q_{j_{Cr}}]$ respectively. The EMD is computed upto k^{th} iteration till it converges i.e $W(F^{(K+1)}, P_K'^{(K+1)}, Q_K'^{(K+1)}) \leq W(F^{(K)}, P_K'^{(K)}, Q_K'^{(K)})$. The signature maps for similar images are expected to be similar. After multiplying with weights, the ranks of the relevant images are not affected much.
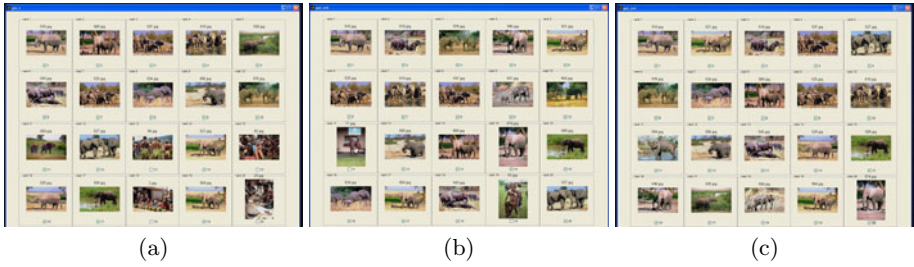
Fig. 1. Database (A) (a) First Pass of the Retrieval Results using M-band Wavelet Features with the top left most image as a query image (b) First Iteration (c) Second Iteration

4   Experimental Results

Considering each image as a query, results are tested upon (A) SIMPLIcity images consisting of 1000 images from 10 different categories (B) Corel 10000 miscellaneous database which is labeled into 79 semantic categories. The results are also compared with ISO standard MPEG-7 visual features i.e. Edge Histogram Descriptor (EHD) and Color Structure Descriptor (CSD)[10].

The retrieval results using different features and distance norms can be best compared when tested upon same database. We grossly partition each image of the database into three meaningful clusters for computing EMD distance. For the query image of elephant as shown in Fig.1(a) on database (A), there are four irrelevant images. Such images are having quite similar color distribution but of different distribution weight. By suitable weighting of the feature planes the average precision ($Precision = \frac{No.\ of\ relevant\ images\ retrieved}{Total\ No.\ of\ images\ retrieved}$)is increased upto 100% after 2nd iteration which are shown in Fig. 1(b) and Fig.1(c). The results on database(B) are also satisfactory, one example is shown in Fig.2(a). Fig.2(b) shows results using M-band, EHD and CSD.
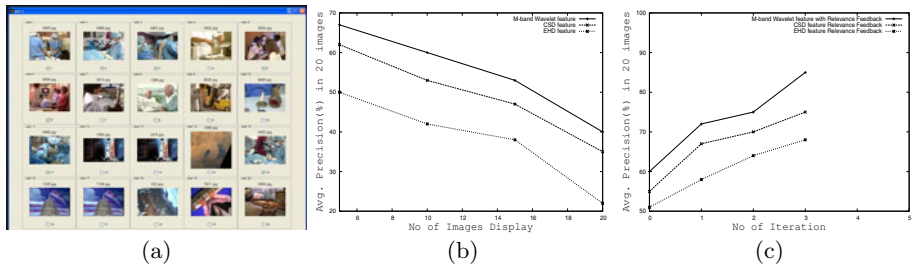

Fig. 2. (a) First Pass of the Retrieval Results using M-band Wavelet Features on Database (B) (b) Average Precision of CSD, EHD and M-band Wavelet Vs No. of Image on Database (A) without RF (c) Successive improvement(Average Precision) Vs No. of Iteration with RF

Fig. 2(c) shows an improvement in average precision (considering all database images) upto 30% (maximum) at 3rd iteration using our RF with EMD on M-band wavelet features. Our results are found better than CSD and EHD which show an improvement of about 10% to 15% after 3rd iteration, where FEI is used with Euclidean distance.

MPEG-7 CSD visual features and M-band wavelet features approximately take 1-2 secs whereas EHD takes 500 ms. The CPU-time taken for each iteration is approximately 50 ms for database (A) and 3 sec for database (B) using MATLAB R2008a on a Dell(T7400, 4GB, RAM) machine.

5   Conclusions

The proposed image retrieval system based on M-band wavelet features is able to improve the retrieval performance satisfactorily within 2 to 3 iterations of relevance feedback. We intend to incorporate partial query using EMD and relevance feedback in partial matching for videos in conjunction with motion information as future scope of research.

References

1. Heesch, D.: A survey of browsing models for content based image retrieval. Multimedia Tools Application 40, 1380–7501 (2008)
2. Chang, F.C., Hang, H.M.: A relevance feedback image retrieval scheme using multi-instance and pseudo image concepts. IEICE Transaction Information System E89-D, 1720–1731 (2006)
3. Acharyya, M., Kundu, M.K.: An adaptive approach to unsupervised texture segmentation using M-band wavelet tranform. Signal Processing 81, 1337–1356 (2001)
4. Wang, J.Z., Li, J., Wiederhold, G.: Simplicity: Semantics-sensitive integrated matching for picture libraries. IEEE Transactions on Pattern Analysis and Machine Intelligence 23, 947–963 (2001)
5. Cheng, Q., Yang, C., Chen, F., Shao, Z.: Application of M-band wavelet theory to texture analysis in content-based aerial image retrieval. International Geoscience and Remote Sensing Symposium 3, 2163–2165 (2004)
6. Jin, Z., King, I., Li, X.: Content-based image retrieval by relevance feedback. In: Laurini, R. (ed.) VISUAL 2000. LNCS, vol. 1929, pp. 521–529. Springer, Heidelberg (2000)
7. Burrus, C.S., Gopinath, A., Guo, H.: Introduction to Wavelets and Wavelet Transform: A Primer. Prentice Hall International Editions, Englewood Cliffs (1998)
8. Rubner, Y., Tomasi, C.: Perceptual Metrices for Image Database Navigation. Kluwer Academic Publishers, Boston (2001)
9. Pal, S.K., Majumder, D.D.: Fuzzy Mathematical Approach To Pattern Recognition. Willey Eastern Limited, New York (1985)
10. Manjunath, B.S., Salembier, P., Sikora, T.: Introduction to MPEG-7: Multimedia Content Description Interface. John Wiley and Sons Inc., USA (2002)