2009 Seventh International Conference on Advances in Pattern Recognition

Color Image Retrieval Using M-Band Wavelet Transform Based Color-Texture Feature

Malay K. Kundu and Priyank Bagrecha
Machine Intelligence Unit
Indian Statistical Institute
Kolkata - 700108, India

***Abstract*—Feature Extraction algorithm is a very important component of any retrieval scheme. We propose M-band Wavelet Transform based feature extraction algorithm in this paper. The MxM sub-bands are used as primitive features, over which energies computed in a neighborhood are taken as the features for each pixel of the image. These features are clustered using FCM to obtain image signature for similarity matching using the Earth Mover's Distance. The results obtained were compared with MPEG-7 content descriptor based system and found to be superior.**

I. Introduction

With the advent of internet, there has been an explosion in the amount of visual data available to us. Over the years it has become increasingly difficult to manage ever increasing database of multimedia. Due to the difficulty of manual annotation of images, it is imperative to index images automatically, for efficient retrieval depending on its content. Several solutions to this effect have been proposed to the problem and it is still an active area of research. Several color, texture, shape etc. based approaches have been proposed independently [7], [16], [2]. Texture of an image is a very useful characteristic for image retrieval, so Texture based retrieval using wavelet transform is an active topic of research, owing to many beautiful properties of the wavelet transform. The ability of wavelet transform to capture textural characteristics is comparable to Gabor Transforms, and the availability of fast algorithms for computation of wavelet transform has facilitated to its increasing use over other methods.

Manjunath et al extensively studies the use of Gabor filter banks for texture based retrieval. [8], [16] MPEG-7 or Multimedia Content Description Interface, an ISO standard for multimedia retrieval, includes two Gabor filter banks based texture descriptors [7], [8], [12]. Acharyya et. al. [1] have used M-band wavelet representation and used wavelet filters for computation of texture features.In the present work, we have used similar kind of approach for the computation of color texture feature.

Wang et. al. [14], [15] have used a 2-step algorithm using wavelet transform for developing a CBIR system for retrieval of color images. They transform the image to a color space similar to opponent color space prior to wavelet decomposition. At first, sub-band variances as representative features are used for crude matching, the output of which are used for a finer matching.

We propose the use of M-band Wavelet Transform for Content Based Retrieval of images. The motivation for the use of 2D M-band Wavelet Transform is based on the alignment of its properties to that of the Human Visual System in dividing the spatial frequency domain into bands. Wavelet coefficients in the intensity and chromaticity planes are used as primitive features for computation of local energy. Clustering of the pixels over the local energy measurement, divides the image into perceptually similar regions. Use of a perceptually inclined metric like Earth Mover's Distance has improved the retrieval result remarkably.

II. M-band Wavelet Transform

Orthogonal M-band wavelet transform is a direct generalization of dyadic orthogonal wavelet transform [3]. Dyadic wavelet transform is not suitable for analysis of high frequency signals, as it decomposes the frequency channel logarithmically but M-band wavelet transform divides the time-scale space both logarithmically as well as linearly thereby giving better resolution at high frequencies[3]. We use M-channel filters decomposing the time-scale space into M×M sub-bands.

III. M-Band Wavelet Transform based Color-Texture Feature

Human eye shows varying sensitivity response to different spatial frequencies. A Human Visual system [10] divides an image into several bands, than actually visualizing the complete image as a whole. This fact motivated us to use the M-band filters which are essentially frequency and direction oriented band pass filters. We use a 1-D, 16 tap 4 band orthogonal filters with linear phase and perfect reconstruction for the multi-resolution analysis. The 1-D, M-band filter transfer functions are denoted byH_i, $1 \leq i \leq 4$. The image, prior to M-band wavelet decomposition, is transformed to Y-Cb-Cr color space. This ensures that the textural characterization of the image is independent of the color characterization. Wavelet decomposition over the intensity plane characterizes the texture information, while the wavelet decomposition over chromaticity planes characterizes color. Wavelet transform is applied to Y, Cb and Cr planes. An over-complete decomposition resulting in the same size of the sub-bands as the image is important here, to obtain the features for each pixel of the

978-0-7695-3520-3/09 $25.00 © 2009 IEEE
DOI 10.1109/ICAPR.2009.34

image, to be clustered further. The 16 sub-bands coefficients obtained are used as the primitive features.

Natural images exhibit spatial variation of the texture. As a result, texture based retrieval of images cannot assume the textures to be homogeneous. A localized characterization of textures thus becomes necessary. Hence we estimate local energy for each of the 16 sub-band images. The Absolute Gaussian energy, for each pixel, is computed over a neighborhood, the size of which is determined using a spectral flatness measure (SFM).

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

$$1 \leq m_1 \leq M, 1 \leq m_2 \leq M$$

where$N$is the neighborhood size 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 filterH_{m_2}.

$$G(x,y) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(1/2\sigma^2)(x^2+y^2)}$$

SFM gives a measure of the global frequency content of the image. It is defined as the ratio of arithmetic mean and the geometric mean of the Fourier coefficients of the image. It has been reported in literature that the size of the neighborhood for computation of localized energy range from 11×11 to 31×31, while SFM varies from 1 to 0 [1]. 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. Since images generally are formed by regions, as in from objects and their surroundings, clustering or quantizing the wavelet energy reduces the feature size retaining maximum information. Based on this basic assumption, the energy values, for each sub-band and for each plane of the color image are used as the feature for a pixel and clustered using Fuzzy C-Means.

It has been shown [17] that Earth Mover's Distance(EMD)is a very is very useful distance metric while measuring perceptual distance between two color texture We have used Earth Mover's Distance(EMD) as the metric for similarity matching. Earth Mover's distance uses a signature over traditional histogram for similarity matching [11] has successfully reported the use of EMD as an efficient metric for content based image retrieval with several advantages over other similarity and dissimilarity measures. The Earth Mover's Distance is formally discussed in the next section. The feature vector comprising of the cluster centers of the energy measurement over subbands, with the number of pixels of the image in each cluster comprises the image signature. To keep the computations minimum, FCM was preferred keeping the number of clusters as 3.

IV. Earth Mover's Distance(EMD)

For an image retrieval system to find images which are visually similar to query image, it should have an efficient representation of the system as well as an efficient measure that can determine the extent of similarity between the images in the database and the query image. Various measures like Minkowski Distance, Histogram Intersection, Kullback-Leibler Divergence, Jeffrey Divergence or χ^2 Statistics have been used by the researchers in texture retrieval [11].

A signature is defined as $\{cl_i = (feature_i, weight_{feature_i})\}$ characterizes each image independently and efficiently. A complex image will have a larger signature while a simple image will have a small signature, by adapting the number of clusters depending on the complexity of the image. Computation of EMD is based on a solution to Monge-Kantorovitch mass transfer problem. The problem can be represented in terms of flow of goods between suppliers and consumers [11]. Assuming supply of goods from several suppliers each having a capacity to supply to several consumers each having a consumption capacity, the transportation problem is then to find the least expensive flow of goods which satisfies the consumer's demands. Signature matching then becomes analogous to the transportation problem by defining one signature as supplier and the other as the consumer, and the ground distance between an element in the first signature to an element in the second as the cost for a supplier-consumer pair. The EMD [11] thus measures the minimum amount of work required to transform one signature into other. It is formally defined as follows

Let $P = \{(p_1, w_{p_1}), ..., (p_m, w_{p_m})\}$be the first signature with m clusters where p_i is a cluster representative and w_{p_i} is the weight of the cluster. $Q = \{(q_1, w_{q_1}), ..., (q_n, w_{q_n})\}$ is the second signature with n clusters. Let $D = [d_{ij}]$the ground distance matrix where $d_{ij} = d(p_i, q_j)$is the ground distance between clusters p_i andq_j, chosen according to the task at hand. Computing EMD thus becomes finding a flow $F = [f_{ij}]$with $f_{ij}$the flow between p_i and $q_j$which minimizes the overall cost

$\mathbf{WORK}(P,Q,F) = \sum_{i=1}^{m}\sum_{j=1}^{n} d(p_i,q_j)f_{ij}$subject to the constraints:

$$f_{ij} \geq 0, 1 \leq i \leq m, 1 \leq j \leq n \tag{1}$$

$$\sum_{j=1}^{n} f_{ij} \leq w_{p_i}, 1 \leq i \leq m, \tag{2}$$

$$\sum_{i=1}^{m} f_{ij} \leq w_{q_j}, 1 \leq j \leq n, \tag{3}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = \min(\sum_{i=1}^{m} w_{p_i}, \sum_{j=1}^{n} w_{q_j}) \tag{4}$$

Constraint Eq.1 ensures movement of goods from suppliers to consumers and not the other way. Constraint Eq.2 defines the upper bound on the capacity of the suppliers while Eq.3

defines the upper bound on the capacity of the consumers. Constraint Eq.4 ensures that maximum possible supplies to be moved from suppliers (P) to consumers (Q), called the total flow. Once the solution to optimal flow is obtained EMD is defined as the work normalized by the total flow:

$$EMD(P,Q) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} d(p_i,q_j)f_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}}$$

EMD by its definition extends to distance between sets or distributions of elements, thereby facilitating partial matches. The cost of moving earth from sand piles to holes, to fill them defines the nearness property properly as compared to histogram, information theoretic or statistics based approach. It can be shown that, EMD is a metric, if the ground distance is a metric and the total weights of two signatures are equal.

V. MPEG-7 AND EXISTING METHODOLOGIES:

MPEG-7 or Multimedia Content Description Interface is an ISO standard focusing on multimedia retrieval. It includes descriptors and description schemes for efficient retrieval of images, videos, audio and graphic files based on the content. [7], [9] Several low level feature extraction algorithms using color, texture, motion, and shape are facilitated for image and video retrieval and benchmarking of new schemes. [8], [9] MPEG-7 provides Scalable Color Descriptor (SCD), Color Structure Descriptor (CSD), Dominant Color Descriptor and Color Layout Descriptor for color based retrieval and Texture Browsing Descriptor, Homogeneous Texture Descriptor (HTD) and Edge Histogram Descriptor (EHD) for texture based retrieval. [7]

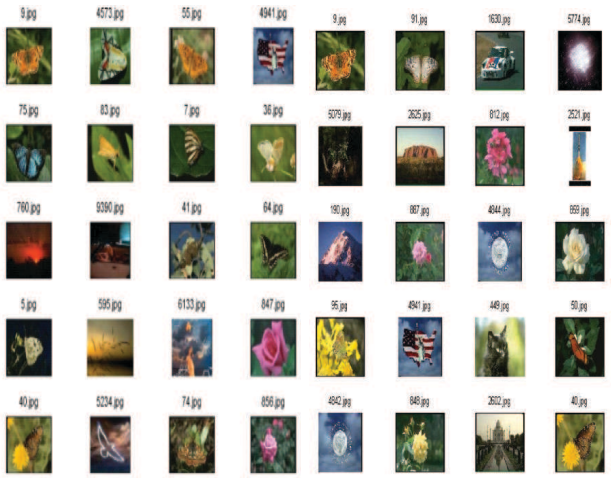
Texture Browsing Descriptor is a compact descriptor based on Gabor Wavelets. [16] It characterizes a texture's regularity, directionality and coarseness. [7] HTD characterizes image texture by filtering the image with a bank of scale and orientation sensitive Gabor filters. It computes the energy and standard deviation of the energy of the output of the filter banks in the frequency bands as the features. [7] The underlying algorithm for Texture Browsing Descriptor and HTD assumes that the images comprises of homogeneous textures. EHD on the other captures spatial distribution of edges which gives a better texture measurement even if it is not homogeneous.

VI. EXPERIMENTAL RESULTS

The experiments were performed on Dell Precision T7400 PC. The performance of the image retrieval system is tested upon two databases: (a) SIMPLIcity images [14] and (b) Corel 10000 miscellaneous database [15]. SIMPLIcity database has 1000 images in 10 categories (People, Beach, Buildings, Bus, Dinosaur, Elephant, Flower, Horses, Mountains and Food). The Corel database has 9908 images belonging to 79 semantic categories. It should be noted that the images in the Corel database are of lower resolution and do not belong to one semantic category alone. The lower resolution of the images makes it difficult to obtain very fine texture features, which are widely used in internet. Extensive experiments were performed and the results are compared with the Edge Histogram Descriptor included in MPEG-7 standard which becomes almost a standard norms [18] for the evaluation of newly proposed image features for CBIR system.

Fig. 1(a) and Fig. 1(b) shows the retrieval result for a query image of class butterfly in the Corel Miscellaneous database respectively. M-band Wavelet Transform definitely performs better than EHD. Retrieval result is obtained for each image in the database as a query. Fig. 2(a) and Fig. 2(b) shows the retrieval result for query of class animal (animal in grassland). EHD is computed by using a block based algorithm, which divides the image into 4×4 sub-images. These sub-images are further sub-divided into a constant number of image blocks. Each macro-block is treated as a 2×2 pixel image. Five edge operators including 4 directional and one isotropic edge operators are used to measure edge strength. Thus for each macro-block 5 edge strengths are computed and if found greater than a threshold, then the image block is considered to be an edge block and used for computation of edge histogram [7]

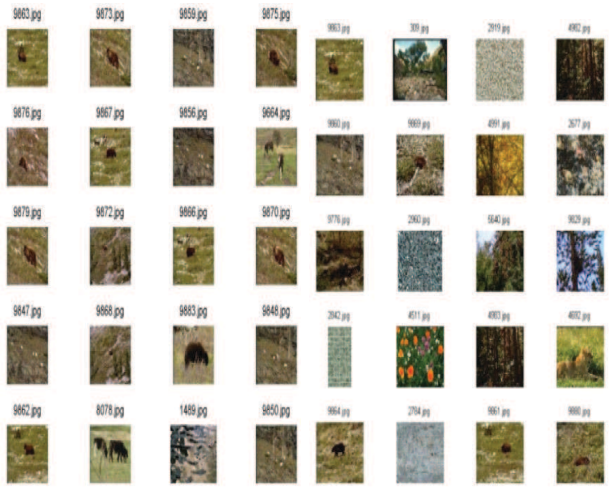
Retrieval results were first obtained for Earth Mover's Distance directly. A better retrieval is obtained by a weighted distance obtained over the three EMD's computed over each color plane separately.



(a) using M-band Wavelet Transform (b) using EHD

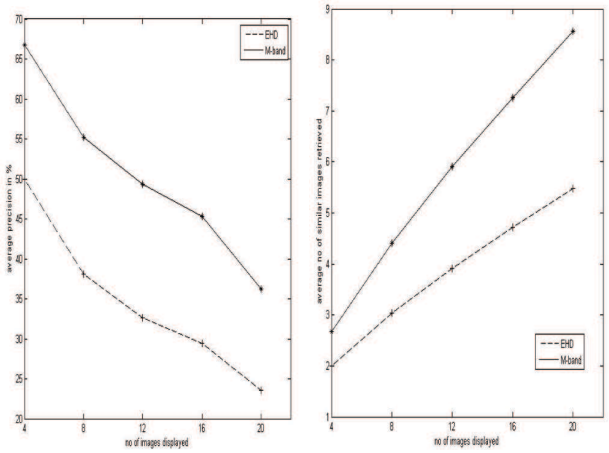
Fig. 1. Retrieval Results

The human visual system is less sensitive to chrominance than luminance [10]. Accordingly, the weights for computing the weighted distance is based on the convention relative perceptual importance as used in the JPEG 2000 that is Y: Cb: Cr = 4:2:1. Here the weights are chosen though seemingly are heuristic; but they are based on the fact that a human eye cannot detect more than a certain number of colors. Also, a color, or intensity detected by a human eye, is dependent on not just its value, but also the intensities and colors in the region surrounding it. As a result we have chosen wavelet analysis in the chromaticity planes as opposed to chromaticity



(a) using M-band Wavelet Transform (b) using EHD

Fig. 2. Retrieval Results



(a) Average precision curve (b) Average retrieval result

Fig. 3. Average precision curve and average retrieval curve for M-band and EHD

histograms for color information. Although we have chosen a fixed set of weights, 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. Nevertheless, the set of weights used by us gives good performance in almost all semantic classes of the database.

As can be seen from the average Precision curve in Fig.3, the proposed methodology using M-band Wavelet Transform performs a better retrieval than the EHD based system for which ranking of the semantically similar images are lower compared to the proposed methodology as displayed in 20 images resulted on retrieval.

VII. Conclusion

Several tree based hierarchical clustering algorithms for huge database are available in literature. the use of which would facilitate an adaptive clustering of the images depending on its complexity and has already been reported [13] its effectiveness in region based retrieval of gray images. The same methodology could be adapted over a color-space closer to human visual system, such as HSI or HMMD. The adaptive choose of weights for computing weighted distance, depending on the query image should facilitate to achieve the performance of the CBIR system. As mentioned the proposed methodology performs better than EHD proposed in MPEG-7 standard, and hence may also be useful for retrieval of video shots.

References

[1] M. Acharyya and M. K. Kundu, An adaptive approach to unsupervised texture segmentation using M-band wavelet transform, Signal Processing 81, pp.1337-1356, 2001.

[2] M. Bober, MPEG-7 Visual Shape Descriptors, IEEE Transactions on Circuits and Systems for Video Technology, 11(6), pp.716-719, 2001.

[3] C. S. Burrus, A. Gopinath and H. Guo, Introduction to Wavelets and Wavelet Transform: A Primer, Prentice Hall International Editions, placeCityEnglewood Cliffs, NJ, 1998.

[4] I. Daubechies, Orthogonal bases for compactly supported wavelets, Commun. Pure Appl. Math. 41, pp.773-789, 1988.

[5] M. N. Do and M. Vetterli, Wavelet-based Texture Retrieval using Generalized Gaussian Density and Kullback-Leibler Distance, IEEE Transactions on Image Processing, 11(2), pp. 146-158, 2002.

[6] M. Lamard, G. Cazuguel, G. Quellec, L. Bekri, C. Roux and B. Cochener, "Content Based Image Retrieval based on Wavelet Transform coefficients distribution", Conf. Proc. IEEE Eng Med Biol Soc. 2007; 1: 4532-4535.

[7] B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and Texture Descriptors", IEEE Transactions on Circuits and Systems for Video Technology, 11(6), pp. 70-715, 2001.

[8] B. S. Manjunath, P. Salembier, and T. Sikora (editors), Introduction to MPEG-7: Multimedia Content Description Interface, John Wiley and Sons, Inc, USA 2002.

[9] J. M. Martinez, MPEG-7 Overview (version 10), ISO/IEC JTC1/SC29/WG11N6828, 2004.

[10] K. N. Plataniotis and A. N. Venetsanopoulos, Color Image Processing and Applications, Springer Verlag, Heidelberg, 2000.

[11] Y. Rubner and C. Tomasi, Perceptual Metrices for Image Database Navigation, Kluwer Academic Publishers, 2001.

[12] T. Sikora, The MPEG-7 visual Standard for Content Description-An Overview, IEEE Transactions on Circuits and Systems for Video Technology, vol. 11(6), 2001.

[13] N. Suematsu, Y. Ishida, A. Hayashi and T. Kanbara, "Region based Image Retrieval using Wavelet Transform", Proc. 15th International Conference on Vision Interface, Canada, pp. 9-16, 2002.

[14] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: Semantics-sensitive integrated matching for picture libraries", IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(9), pp. 947-963, 2001.

[15] J. Z. Wang, G. Wiederhold, O. Firschein and S. X. Wei, Content-based image indexing and searching using Daubechies' wavelets, Int J Digit Libr, 1, pp.311-328, 1997.

[16] P. Wu, B. S. Manjunath, S. Newsam and H. D. Shin, A texture descriptor for browsing and similarity retrieval, Signal Processing: Image Communication 16, 33-43, 2000.

[17] Y. Rubner, C. Tomasi and L.J. Guibas, Earth Mover's distance as a metric for image retrieval,International Journal of Computer Vision 40(2), pp. 99-121, 2000.

[18] M. Banerjee and M. K. Kundu, Image Retrieval Using Fuzzy Relevance Feedback and Validation with MPEG-7 Content Descriptors, Proc.2nd International Conference on Pattern recognition and Machine Intelligence(PREMI-07), LNCS 4815, pp.144-152, 2007.