TEXTURE SIMILARITY MEASUREMENT USING KULLBACK-LEIBLER DISTANCE ON WAVELET SUBBANDS

Minh N. Do[†] *and Martin Vetterli*[†§]

[†]Laboratory for Audio-Visual Communications

Swiss Federal Institute of Technology Lausanne (EPFL), CH-1015 Lausanne, Switzerland

[§]Department of EECS, University of California at Berkeley, Berkeley CA 94720, USA

Email: {Minh.Do, Martin.Vetterli}@epfl.ch

ABSTRACT

The focus of this work is on using texture information for searching, browsing and retrieving images from a large database. In the wavelet approaches, texture is characterized by its energy distribution in the decomposed subbands. However it is unclear on how to define similarity functions on extracted features; usually simple norm-based distances together with heuristic normalization are employed. In this paper, we develop a novel wavelet-based texture retrieval method that is based on the modeling of the marginal distribution of wavelet coefficients using generalized Gaussian density (GGD) and a closed form Kullback-Leibler distance between GGD's. The proposed method provides greater accuracy and flexibility in capturing texture information while its simplified form has close resemblance with existing methods. Experimental results indicate that the new method significantly improves retrieval rates, e.g. from 65% to 77%, against traditional approaches while it has comparable levels of computational complexity.

1. INTRODUCTION

Digital image and video libraries are becoming more widely used as more visual information is put in digital form as well as on-line. To improve human access, however, there must be an effective and precise method for users to search, browse, and interact with these collections and do so in a timely manner.

Typically in a content-based image retrieval (CBIR) system, there are two major tasks. The first one is *Feature Extraction* (FE), where a set of features – called image signatures, is generated to accurately represent the content of each image in the database. A signature is much smaller in size than the original image, typically on the order of hundreds of elements (rather than millions). The second task is *Similarity Measurement* (SM), where some sort of similarity functions between the query image and each image in the database using their signatures are computed so that the top N matched images can be retrieved.

For the texture retrieval application, some of the most popular feature extraction methods are filtering or wavelet-like approaches [2, 8, 6, 10, 7]. Essentially, those methods measure energy (possible weighted) at the output of filter banks as extracted features for texture discrimination. The basic assumption of these approaches is that the energy distribution in the frequency domain identifies a texture. However it remains unclear on how to define similarity measures on extracted features. Many current similarity functions, typically Euclidean-like distances, have little justification in this context.

In a statistical framework, we can view the task of searching for the top N similar images to a given query image from a database of total M images ($N \ll M$) as a selection from multiple hypothesis. Here each candidate image in the database $\mathcal{I}_i, i = 1, 2, \ldots, M$ is assigned with an hypothesis $\mathcal{H}_i$. The goal is to select among the M possible hypotheses the N best ones (with a ranking order) that describe the data from the query image. In this setting, by jointly considering the two problems of FE and SM while keeping in mind the complexity constraint of CBIR applications, we can show [4] that the optimum maximum likelihood (ML) selection rule can be *asymptotically* realized by:

Feature Extraction: Given the data from each image, extracted features are probability model parameters estimated from the data using a ML estimator.

Similarity Measurement: To select the top N matches to a query image, the database is ranked based on the Kullback-Leibler distances between the query model and models of each candidate images.

The Kullback Leibler distance (KLD) or *relative entropy* between two densities f and g is defined as [3]:

$$D(f||g) = \int f \log \frac{f}{g}. \tag{1}$$

Note that the KLD is not symmetric. For our case, the density f represents the query image where g represents a candidate image.

In this paper, we will demonstrate the application of the statistical framework in the wavelet-based texture retrieval problem. The statistical approach fits nicely into this problem since a texture image is often regarded as a realization of an underlying stochastic process.

2. WAVELET-BASED TEXTURE FEATURES

Using the assumption that the energy distribution in frequency domain identifies texture, traditional approaches computed energies of wavelet subband as texture features. Commonly, $\mathcal{L}^1$ and $\mathcal{L}^2$ norms are used as measures.[1]

This work was supported by a Department of Communication Systems, EPFL PhD Fellowship and the Swiss National Science Foundation grant number 21-52439.97.

[1]This is an abuse of terminology since strictly speaking $\mathcal{L}^1$ norm is not an energy function.

On the other hand, statistical approaches treat texture analysis as a probability inference problem. In this work we simply characterize texture images via marginal distributions of their wavelet subband coefficients. This is supported by psychological research on human texture perception which suggests that two homogeneous textures are often difficult to discriminate if they produce similar marginal distributions of responses from a bank of filters [1].

Experiments show that a good probability density function (PDF) approximation for the marginal density of coefficients at a particular subband produced by various type of wavelet transforms may be achieved by adaptively varying two parameters of the *generalized Gaussian* density (GGD) [5, 10], which is defined as:

$$p(x;\alpha,\beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-(|x|/\alpha)^{\beta}}, \qquad (2)$$

where $\Gamma(.)$ is the Gamma function, i.e. $\Gamma(z) = \int_0^{\infty} e^{-t}t^{z-1}dt,\ z>0$.

Here α models the width of the PDF peak (variance), while β is inversely proportional with the decreasing rate of the peak. The GGD model contains the Gaussian and Laplacian PDF's as special cases, using $\beta = 2$ and $\beta = 1$, respectively.

Figure 1 shows a typical example of histogram of wavelet subband coefficients together with a plot of the fitted GGD using the ML estimator that is described in [4]. The fits are generally quite good. Thus it can be seen that the marginal distribution parameters provide a more precise representation of texture information than the wavelet subband energies. But more importantly, the statistical modeling leads to a more natural way of defining the similarity measurement between images.

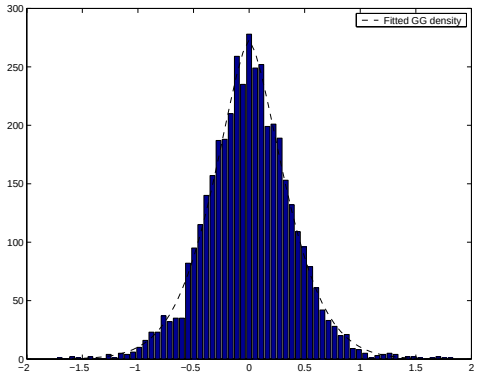


Fig. 1. Example of wavelet subband coefficient histogram fitted with a GGD. The estimated parameters are: $\alpha = 0.46$ and $\beta = 1.52$.

3. SIMILARITY MEASUREMENT ON WAVELET SUBBANDS

3.1. Kullback-Leibler Distance between GGD's

Given the GGD model, the PDF of wavelet coefficients at a subband can be completely specified via two parameters α and β. Substitute (2) into (1) and after some manipulations we obtain the following closed form for the Kullback-Leibler distance (KLD) between two GGD's as:

$$D(p(.;\alpha_1,\beta_1)||p(.;\alpha_2,\beta_2)) = \log\left(\frac{\beta_1\alpha_2\Gamma(1/\beta_2)}{\beta_2\alpha_1\Gamma(1/\beta_1)}\right) + \left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2}\frac{\Gamma((\beta_2+1)/\beta_1)}{\Gamma(1/\beta_1)} - \frac{1}{\beta_1}. \qquad (3)$$

Therefore the similarity measurement between two wavelet subbands can be computed very effectively using the model parameters. Furthermore, using the chain rule of KLD [3] with the reasonable assumption that wavelet coefficients in different subbands are independent, the overall similarity distance between two images is the sum of KLD's given in (3) between corresponding pairs of subbands. That is if we denote $\alpha_i^{(j)}$ and $\beta_i^{(j)}$ as the extracted texture features from the wavelet subband j of the image $\mathcal{I}_i$ then the overall distance between two images $\mathcal{I}_1$ and $\mathcal{I}_2$ (where $\mathcal{I}_1$ is the query image) is:

$$D(\mathcal{I}_1,\mathcal{I}_2) = \sum_j D(p(.;\alpha_1^{(j)},\beta_1^{(j)})||p(.;\alpha_2^{(j)},\beta_2^{(j)})). \qquad (4)$$

3.2. Relation to Energy-based Methods in the Laplacian Case

Consider the case when the parameter β is fixed and equal to 1. That is we are modeling the wavelet coefficients using the Laplacian distribution. The extracted feature from wavelet coefficients $\mathbf{x} = (x_1, x_2, \ldots, x_L)$ at a particular subband using the ML estimator is:

$$\hat{\alpha}_{\mathbf{x}} = \frac{\sum_{i=1}^{L}|x_i|}{L}. \qquad (5)$$

This is the same as the $\mathcal{L}^1$-norm features of wavelet coefficients.

From (3), the KLD between two Laplacian distributions is:

$$D(p(.;\alpha_1,1)||p(.;\alpha_2,1)) = \log\left(\frac{\alpha_2}{\alpha_1}\right) + \frac{\alpha_1}{\alpha_2} - 1$$

This is a convex function of α_2/α_1 and is minimum when $\alpha_2/\alpha_1 = 1$. Therefore in term of selecting the most similar images, we are only interested in the situation when the ratio α_2/α_1 is near the vicinity of 1. Using first-order Taylor approximation of $\log x$ around 1, $\log x \approx x - 1$ when $x \approx 1$, we have,

$$D(p(.;\alpha_1,1)||p(.;\alpha_2,1)) \approx \frac{\alpha_2}{\alpha_1} - 1 + \frac{\alpha_1}{\alpha_2} - 1 = \frac{(\alpha_2-\alpha_1)^2}{\alpha_1\alpha_2}.$$

Summing those distances across wavelet subbands we obtain the overall distance between two images $\mathcal{I}_1$ and $\mathcal{I}_2$ as

$$D(\mathcal{I}_1,\mathcal{I}_2) \approx \sum_j \frac{(\alpha_2^{(j)}-\alpha_1^{(j)})^2}{\alpha_2^{(j)}\alpha_1^{(j)}}. \qquad (6)$$

This distance is essentially the same as the popular weighted Euclidean distance between extracted features $\alpha_i^{(j)}$ where "global" normalization factors $w^{(j)} = \text{var}\{\alpha_i^{(j)} : i = 1, 2, \ldots, M\}$ are replaced by "local" normalization factors $w_{1,2}^{(j)} = \alpha_2^{(j)}\alpha_1^{(j)}$.

So we just demonstrated that the general statistical framework with GGD's on the wavelet coefficients can be simplified to closely resemble and thus provide a justification for the weighted Euclidean distance between $\mathcal{L}^1$-energies of wavelet subbands. This

is an interesting fact since the two approaches are based on totally different assumptions. The former one relies on an underlying stochastic process of the texture image while the later one is based on the energy distribution in the frequency domain.

4. EXPERIMENTAL RESULTS

We used 40 textures obtained from the MIT Vision Texture (VisTex) database [9]. These are real world 512 $\times$ 512 images from different natural scenes. Only gray-scale levels of the images (computed from the luminance component) were used in the experiments. Since we define similar textures as subimages from a single original one, we selected texture images whose visual properties do not change too much over the image. The names of those selected textures can be found in Table 2.

Each of the 512 $\times$ 512 images was divided into sixteen 128 $\times$ 128 non-overlapping subimages, thus creating a test database of 640 texture images. Furthermore, to eliminate the effect of common range in the gray level of subimages from a same original image and to make the retrieval task less biased, each subimage was individually normalized to zero mean and unit variance before the processing.

4.1. Computational Complexity

The proposed texture retrieval system has been implemented in the Matlab environment. The Feature Extraction (FE) step involves talking a wavelet transform of the input image and estimating the GGD model parameters at each subband using ML estimator. It was found that roughly the same amount of time is spent on wavelet transformation and parameter estimation, giving a total of less than 1 second of CPU time on a Sun's Ultra 5 workstation for extracting features from one image.

Thank to the closed form of distance in (3), the Similarity Measurement (SM) between two images involves simple computation using a small number of model parameters. Optimized implementation using lookup tables yields comparable computation time as normalized Euclidean distance.

4.2. Retrieval Effectiveness

In retrieval experiments, a simulated query image is any one of 640 images in our database. The relevant images for each query are defined as the other 15 subimages from the same original VisTex image. Following [6] we evaluated the performance in terms of the average rate of retrieving relevant images as a function of the number of top retrieved images. The new approach (which uses the combination of GGD & KLD) is compared with the traditional methods using the energy-based features in the wavelet domain (both $\mathcal{L}^1$ and $\mathcal{L}^2$ features) together with normalized Euclidean distance as the similarity measurement. We also reported the results when the normalized Euclidean distance (ED) was used on GGD's model parameters, which is similar to the method used in [10]. Note that all compared methods yield the same number of features per images (2 features per wavelet subband). The employed wavelet transforms are the traditional wavelet pyramids (DWT) and the non-subsampled discrete wavelet frames (DWF) using the 8-tap Daubechies orthogonal wavelets. Table 1 summaries the comparison in performance in average percentages of retrieving relevant images in the top 15 matches.

Type of decomposition	Methods		
	$\mathcal{L}^1 + \mathcal{L}^2$	**GGD & KLD**	GGD & ED
1 scale (6 features)			
DWT	55.47	**66.36**	57.47
DWF	56.92	**67.09**	61.34
2 scales (12 features)			
DWT	62.82	**73.35**	65.87
DWF	63.32	**74.01**	69.78
3 scales (18 features)			
DWT	64.83	**76.57**	61.18
DWF	68.48	**78.12**	71.13

Table 1. Average retrieval rate (%) in the top 15 matches using pyramid wavelet transform (DWT) and wavelet frames (DWF).

The first observation is that the statistical approach (GGD & KLD) always outperforms all other tested traditional methods. This is consistent with our expectation since the GGD parameters are more expressive in characterizing textures than the energy-based ones. Furthermore, the inferior results of the GGD & ED method (where the same features with the statistical method were used but similarity measurement uses normalized Euclidean distance) shows that good performance in retrieval comes not just from a good set of extracted features but also together with a suitable similarity measurement. Hence this supports our approach of considering the two problems FE and SM jointly.

Table 2 details the comparison between the $\mathbf{L}^1 + \mathbf{L}^2$ and GGD & KLD methods with 3 levels of pyramid wavelet decomposition. Again we can see that the new method consistently gives superior performance for almost all texture classes, especially for the ones that have structural patterns.

Figure 2 shows a graph illustrating this comparison in retrieval performances as functions of number of top matches considered. As can be seen, it requires almost double the number of retrieved images in the traditional method in comparison with the new method in order to retrieve the same number of relevant images.

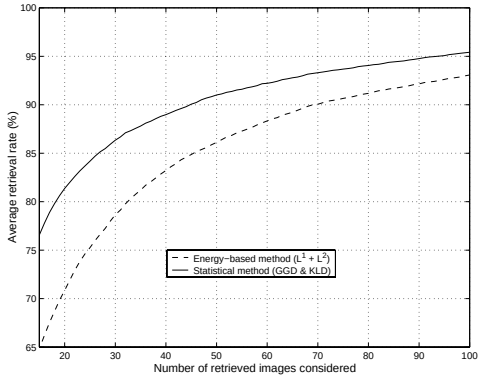
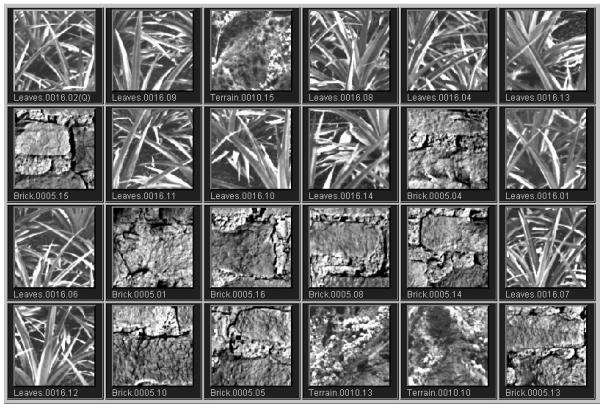


Fig. 2. Retrieval performance according to the number of top matches considered.

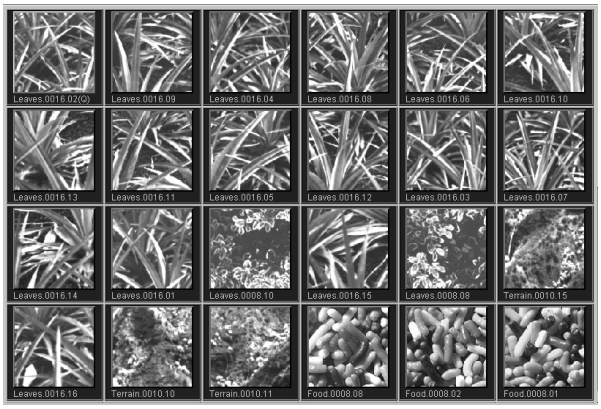
Finally, Figure 3 shows examples of retrieval results using the two compared methods. In this case the energy-based features with normalized Euclidean distance has poor performance in contrast with the new statistical method.

Texture Class	Methods I	Methods II	Texture Class	Methods I	Methods II
Bark0	46.48	53.12	Food8	82.03	97.66
Bark6	39.06	50.39	Grass1	58.98	69.14
Bark8	60.16	73.44	Leaves8	63.67	68.36
Bark9	49.61	61.33	Leaves10	22.27	34.38
Brick1	62.50	71.88	Leaves11	58.98	71.48
Brick4	40.23	66.41	Leaves12	68.75	74.61
Brick5	58.59	83.20	Leaves16	51.56	84.77
Buildings9	71.48	86.72	Metal0	73.83	73.05
Fabric0	69.92	87.50	Metal2	100.00	100.00
Fabric4	66.41	64.84	Misc2	57.42	78.12
Fabric7	98.44	100.00	Sand0	72.27	80.08
Fabric9	70.31	87.89	Stone1	50.00	53.52
Fabric11	66.41	81.25	Stone4	74.22	79.30
Fabric14	89.06	100.00	Terrain10	43.75	52.73
Fabric15	71.88	94.53	Tile1	33.20	53.12
Fabric17	94.92	90.23	Tile4	80.47	99.22
Fabric18	97.66	98.83	Tile7	80.86	100.00
Flowers5	36.33	58.20	Water5	86.72	96.48
Food0	58.59	83.59	Wood1	21.88	35.55
Food5	83.59	89.45	Wood2	80.86	78.52

Table 2. Average retrieval rates for individual texture class using wavelet pyramid transform with 3 decomposition levels. Here I denotes the $\mathcal{L}^1 + \mathcal{L}^2$ method where II denotes the GGD & KLD method.



(a) Method I: $\mathcal{L}^1 + \mathcal{L}^2$



(b) Method II: GGD & KLD

Fig. 3. Examples of retrieval results from 640 texture images based on the VisTex collection. The query image, a subimage of Leaves16, is on the top left corner; all other images are ranked in the order of similarity with the query image from left to right, top to bottom.

5. CONCLUSION

The statistical framework has been applied successfully in the wavelet-based texture retrieval application where simple models on wavelet coefficients that can capture important texture information exist. This results in a new texture similarity measurement in wavelet domain which has a sound theoretical justification with no need for normalization steps. Furthermore, by restricting to simpler models, the new similarity distance becomes closely related to the popular variance-normalized Euclidean distance. Hence the statistical approach can be used as a common framework for other existing methods. Experimental results on 640 texture images of 40 classes from the VisTex collection indicated that the new method has a significant improvement in retrieval rate, e.g. from 65% to 77%, over the traditional approaches using both the pyramid wavelet transform and wavelet frames while requiring comparable computational time.

Acknowledgments: The authors would like to thank Professor Michael Unser and Michael Gastpar for many simulating discussions, and Zoran Pečenović for developing the image retrieval interface used in our experiments.

6. REFERENCES

[1] J. R. Bergen and E. H. Adelson. Theories of visual texture perception. In D. Regan, editor, *Spatial Vision*. CRC press, 1991.

[2] T. Chang and C.-C. J. Kuo. Texture analysis and classification with tree-structure wavelet transform. *IEEE Trans. Image Proc.*, 2(4):429–441, 1993.

[3] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley Interscience, New York, NY, 1991.

[4] M. N. Do and M. Vetterli. Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. submitted to IEEE Trans. on Image Proc., December 1999.

[5] S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Patt. Recog. and Mach. Intell.*, 11(7):674–693, July 1989.

[6] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Trans. Patt. Recog. and Mach. Intell.*, 18(8):837–842, August 1996.

[7] T. Randen and J. H. Husoy. Filtering for texture classification: a comparative study. *IEEE Trans. Patt. Anal. Machine Intell.*, 21:291–310, 1999.

[8] M. Unser. Texture classification and segmentation using wavelet frames. *IEEE. Trans. Image Proc.*, 4:1549–1560, 1995.

[9] MIT Vision and Modelling Group. Vision texture. `http://vismod.www.media.mit.edu`, 1995.

[10] G. V. Wouwer, P. Scheunders, and D. V. Dyck. Statistical texture characterization from discrete wavelet representations. *IEEE Trans. Image Proc.*, 8(4):592–598, April 1999.