ON THE RAYLEIGH NATURE OF GABOR FILTER OUTPUTS

Sitaram Bhagavathy, Jelena Tešić, and *B. S. Manjunath*

Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106.
{sitaram, jelena, manj}@ece.ucsb.edu

Abstract

Texture has been recognized as an important visual primitive in image analysis. A widely used texture descriptor, which is part of the MPEG-7 standard, is that computed using multiscale Gabor filters. The high dimensionality and computational complexity of this descriptor adversely affect the efficiency of content-based retrieval systems. We propose a modified texture descriptor that has comparable performance, but with nearly half the dimensionality and less computational expense. This gain is based on a claim that the distribution of (absolute values of) filter outputs have a strong tendency to be Rayleigh. Experimental results show that the dimensionality can be reduced by almost 50%, with a tradeoff of less than 3% on the error rate. Furthermore, it is easy to compute the new feature using the old one, without having to repeat the computationally expensive filtering step. We also propose a new normalization method that improves similarity retrieval and indexing efficiency.

1. INTRODUCTION

Image texture has found wide application in remote sensing, medical diagnosis, quality control, etc. Texture-based features have been proven to be effective for tasks such as segmentation and similarity retrieval. The use of texture for content-based access has been explored by several researchers ([1], for example). An image can be considered as a mosaic of textures and texture features associated with the regions can be used to index the image data for searching and browsing.

Several methods have been used thus far to extract texture features. The prominent ones are based on a) random field texture models [2, 3], and b) multiscale filtering methods [4, 5]. Among these, a widely used feature [5] is that computed using multiscale Gabor filters. This feature is part of the MPEG-7 standard [6] for multimedia content descriptors. This choice is motivated by several factors. The Gabor representation has been shown to be optimal in the sense of minimizing the joint two-dimensional uncertainty in space and frequency [7], and thus are well suited for texture segmentation problems. Furthermore, Gabor filters approximate the characteristics of certain cells in the visual cortex of some mammals [8].

The high dimensionality and computational complexity of the MPEG-7 descriptor adversely affect the performance as well as the storage, computation, and indexing requirements of a content-based retrieval system. In this paper, we study in detail the statistical properties of Gabor filter outputs when the inputs are texture images. Based on the results, we propose a modified texture descriptor that has comparable performance, but with nearly half the dimensionality and less computational expense. Furthermore, it is easy to compute the new feature using the old one, without having to repeat the computationally expensive filtering step. We also propose a new normalization method that improves similarity retrieval and indexing efficiency.

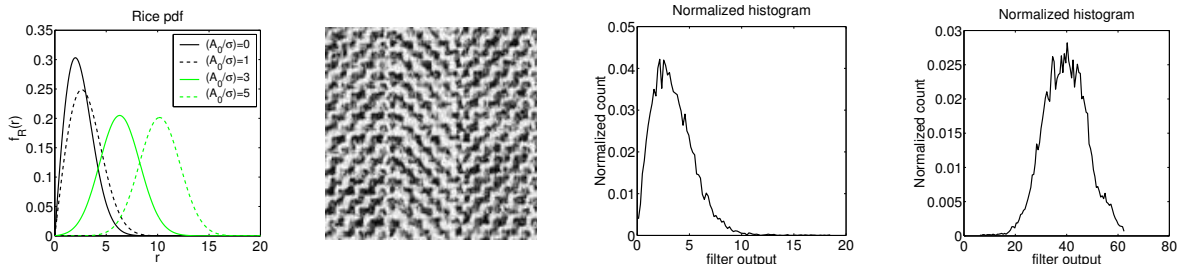
The paper is organized as follows. In Sec. 2, we introduce the Gabor filter bank and the MPEG-7 texture descriptor. In Sec. 3, we study the statistical properties of the Gabor filter outputs, and propose a modified descriptor. In Sec. 4, we compare the performance of the new descriptor and normalization scheme with that of the existing ones. We conclude with a discussion in Sec. 5.

2. A GABOR TEXTURE DESCRIPTOR

A Gabor filter can be viewed as a sinusoidal plane of a particular frequency and orientation, modulated by a Gaussian envelope. A two-dimensional Gabor function is defined as

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)+2\pi jWx\right] \quad (1)$$

This research was supported in part by the following grants and awards: NSF-DLI #IIS-49817432; ISCR-LLNL #W-7405-ENG-48; ONR #N00014-01-0391; NSF Instrumentation #EIA-9986057; and NSF Infrastructure #EIA-0080134.



(a) (b) (c) (d)

Fig. 1. (a) The Rice pdf in (5); (b) a texture image input; (c) and (d) Histograms (with 100 bins) of Gabor filter outputs for two different center frequencies.

where $j = \sqrt{-1}$, and W is the frequency of the modulated sinusoid. A class of self-similar functions, referred to as the Gabor wavelets, is constructed by appropriate dilations and translations of the mother wavelet $g(x, y)$, as follows:

$$\begin{aligned} g_{mn}(x,y) &= a^{-m} g(x', y'), \quad a > 1, \quad m, n \in \mathcal{Z}, \\ x' &= a^{-m}(x\cos\theta + y\sin\theta), \\ y' &= a^{-m}(-x\sin\theta + y\cos\theta), \end{aligned} \tag{2}$$

where $\theta = n\pi/K$, and K is the total number of orientations. The scale factor a^{-m} in (2) ensures that the energy is independent of m. This set of functions forms a non-orthogonal basis for the multi-resolution decomposition.

Gabor filters can be considered to be orientation and scale tunable edge and line detectors, and the statistics of these micro-features can be used to characterize the underlying texture. A Gabor-based homogeneous texture descriptor has been adopted by the MPEG-7 standard for its effectiveness and efficiency. The texture descriptor for s scales and k orientations is given by

$$f_{\mu\sigma} = [\mu_{00}, \sigma_{00}, \mu_{01}, \sigma_{01}, ..., \mu_{s-1,k-1}, \sigma_{s-1,k-1}, \mu_I, \sigma_I], \tag{3}$$

where μ_{mn} and σ_{mn} are the mean and standard deviation of the filter outputs $m(x, y) = |g_{mn}(x, y) * i(x, y)|$, for an input texture image i ($*$ denotes convolution). μ_I and σ_I are the mean and standard deviation of the pixel intensities of the image. Note that the dimensionality of $f_{\mu\sigma}$ is $2sk+2$. Visual dissimilarity between two textures is quantified by computing a distance (usually L_1 or L_2) between their descriptors.

3. STATISTICS OF THE FILTER OUTPUTS

The format of (3) for the texture descriptor is driven by the implicit assumption that the filter outputs have Gaussian-like distributions. Therefore, each of these distributions is taken to be described completely by its mean and standard deviation. In the following, we argue analytically against this approach, and show that the feature vector dimension can be nearly halved.

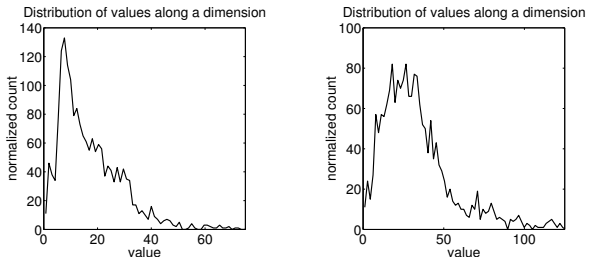
Dunn and Higgins [9] claim (for the 1-D case) that the Gabor filter outputs have a Rice distribution. Their proof is based on modeling the input texture as a periodic lattice of *texels* with random perturbations. Texels are similar, but not necessarily identical, geometric primitives that constitute a texture. We first outline their proof and then draw some inferences.

3.1. Rician model for Gabor filter outputs [9]

Consider the 1-D case for simplicity. The Gabor filter outputs are given by $m(x) = |i(x) * h(x)|$, where $i(x)$ is a real image, and $h(x)$ is a 1-D Gabor function. The 1-D Gabor function can be written as $h(x) = g(x)e^{j\omega_0 x}$, where $g(x)$ is a Gaussian and ω_0 is the frequency. The real and imaginary parts of h are $h_r = g(x)\cos(\omega_0 x)$ and $h_i = g(x)\sin(\omega_0 x)$. The Gabor filter output can then be expressed as $m(x) = |A(x) + jB(x)|$, where $A(x) = i(x) * h_r(x)$ and $B(x) = i(x) * h_i(x)$. Let $H_r(\omega)$ and $I(\omega)$ be the fourier transforms of h_r and i, respectively. If i is a periodic lattice of texels, then $I(\omega)$ is a periodic collection of impulses. $H_r(\omega)$ is a pair of Gaussians centered at $\pm\omega_0$. Supposing that h_r is approximately narrowband and ω_0 coincides with one of the impulses comprising $I(\omega)$, applying h_r to i would result in a pair of impulses at frequencies $\pm\omega_0$. This means $A(x) \approx A_0\cos(\omega_0 x + \psi)$, where ψ is an arbitrary phase angle.

For a nonuniform texture (a periodic texel lattice with random perturbations), it can be shown that (details in [9]) $A(x) = \hat{A}(x) + N(x)$, where $\hat{A}(x) = A_0\cos(\omega_0 x + \psi)$ is the output without perturbation, and $N(x)$ is a zero-mean, wide-sense stationary Gaussian random process with power $E[N^2(x)] = \sigma^2$. $N(x)$ can be modeled as a narrowband process [10], i.e. $N(x) = X(x)\cos(\omega_0 x) - Y(x)\sin(\omega_0 x)$, where $X(x)$ and $Y(x)$ are independent, zero-mean, Gaussian random processes. Thus,

$$\begin{aligned} A(x) &= [A_0\cos(\psi) + X(x)]\cos(\omega_0 x) \\ &\quad -[A_0\sin(\psi) + Y(x)]\sin(\omega_0 x) \\ &= R(x)\cos(\omega_0 x + \Phi(x)), \end{aligned} \tag{4}$$



(a) (b)

Fig. 2. Histograms of the values along an arbitrary dimension of (a) $f_{\mu\sigma}$, and (b) f_γ.

where $R(x)$ is the amplitude, and $\Phi(x)$ is the phase angle of $A(x)$. Since $A(x)$ is the superposition of a sinusoid with narrowband Gaussian noise, it is well known that $R(x)$ has the Rice pdf [10],

$$f_R(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2 + A_0^2}{2\sigma^2}\right) I_0\left(\frac{A_0 r}{\sigma^2}\right), \qquad (5)$$

where $I_0(x)$ is the zero-order modified Bessel function of the first kind. It can then be shown that $m(x) = |A(x) + jB(x)| = \sqrt{A(x)^2 + B(x)^2} = R(x)$. Therefore, $m(x)$ is Rician.

3.2. Practical considerations

The Rice pdf (see Fig. 1a) in (5) can vary from a Rayleigh pdf for small A_0 ($A_0 \approx 0$) to an approximate Gaussian pdf for large A_0 ($A_0 \gg \sigma$). The latter case occurs when the texture is well-defined and periodic, with a highly peaked frequency component at ω_0. The filter outputs tend to have a Rayleigh pdf when the frequency components of the texture are weak in the vicinity of ω_0. The filter bank in (2) used for computing the texture descriptor has predefined center frequencies. Over a wide range of textures, the probability that a given texture has a strong component at a specified center frequency, is small. Hence, we claim that the Rayleigh pdf model for filter output distributions is valid with a higher probability than the Gaussian pdf model that inspires the descriptor in (3). This claim is consistent with the experimental results in Sec. 4.

Fig. 1b-d shows the result of filtering a texture image with Gabor filters of two different center frequencies. Fig. 1c is closer to a Rayleigh pdf, and Fig. 1d is closer to a Gaussian pdf. In the latter case, the input has strong frequencies in the vicinity of the center frequency of the filter. In the former case, it does not.

The Rayleigh pdf, $p(z) = \frac{z}{\gamma^2} \exp\left(-\frac{z^2}{2\gamma^2}\right)$ has only one parameter γ. Therefore, instead of (3), we propose the following texture descriptor with dimensionality $sk + 2$,

$$f_\gamma = [\gamma_{00}, \gamma_{01}, ..., \gamma_{s-1,k-1}, \mu_I, \sigma_I], \qquad (6)$$

where γ_{mn} is the Rayleigh parameter of the output distribution when the Gabor filter g_{mn} is applied. The maximum likelihood estimates of the parameters are given by

$$\gamma_{mn}^2 = \frac{1}{2N} \sum_{i=1}^{N} |z_i|^2, \qquad (7)$$

where z_i are corresponding filter outputs, and N is the number of output coefficients.

If we have precomputed $f_{\mu\sigma}$ (MPEG-7) features, it is easy to compute f_γ without having to perform the filtering step. Since, in (3),

$$\mu_{mn} = \frac{1}{N} \sum_{i=1}^{N} |z_i|;\ \sigma_{mn}^2 = \frac{1}{N} \sum_{i=1}^{N} |z_i|^2 - \mu_{mn}^2, \qquad (8)$$

it can be shown that $\gamma_{mn}^2 = \frac{1}{2}\left(\mu_{mn}^2 + \sigma_{mn}^2\right)$. Thus we can compute the new features from the old ones without having to repeat the computationally expensive filtering step. This observation is significant because many databases already use MPEG-7 texture features for content-based access. Also, we can see from (7) and (8) that, for large values of N, f_γ needs 50% fewer additions in its computation than $f_{\mu\sigma}$.

3.3. Normalization

The distance metric used for measuring visual dissimilarity is more sensitive in dimensions with larger dynamic ranges. Since this is undesirable, the descriptors are normalized so that each dimension has the same dynamic range. We experiment with two normalization methods (along each dimension d separately):

1. *Standard normalization (S.N.)*: Forcing the distribution to have zero mean and unit variance, i.e. $X_d^{(n)} = (X_d - \mu_d)/\sigma_d$,
2. *Rayleigh equalization (R.E.)*: Forcing the distribution to have a uniform distribution $U(0,1)$, i.e. $X_d^{(n)} = 1 - \exp\left(-\frac{X_d^2}{2\gamma_d^2}\right)$, where γ_d is estimated in a manner similar to (7). This does not apply to the last two (μ_I and σ_I) dimensions.

The latter is based on the observation that, over a large dataset, the values along each dimension (other than μ_I and σ_I) of both $f_{\mu\sigma}$ and f_γ follow a skewed distribution that can be well modeled by a Rayleigh pdf (see Fig. 2). *R.E.* enables the use of uniform space partitioning instead of data partitioning in a high dimensional space, thus improving indexing efficiency.

4. EXPERIMENTAL RESULTS

We compare the similarity retrieval performance of $f_{\mu\sigma}$ and f_γ on the Brodatz texture dataset, which is widely

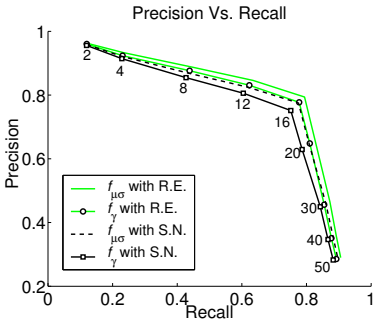


Fig. 3. Precision vs. Recall curves over the Brodatz dataset.

used for this purpose. Since we consider 5 scales and 6 orientations, the dimensionality of $f_{\mu\sigma}$ is 62 and that of f_γ is 32. The dataset consists of 1856 images (16 from each of 116 texture classes). Each texture q is used in turn as the query. Let $A(q)$ denote the set of T retrievals (based on the smallest L_1 distances from q in the descriptor space) and $R(q)$, the set of images in the dataset relevant to q. The *precision* is defined by $P(q) = \frac{|A(q) \cap R(q)|}{|A(q)|}$, and the *recall* by $C(q) = \frac{|A(q) \cap R(q)|}{|R(q)|}$, where $|\cdot|$ denotes cardinality. Fig. 3 shows the precision vs. recall curves for each descriptor and normalization method. The curves are plotted by averaging precision and recall over all q, for different values of T (shown below the curves). In the case of *R.E.*, the μ_I and σ_I dimensions are normalized using *S.N.* While the dimensionality of f_γ is smaller by almost 50% (for large sk), the drop in precision (equivalently, the increase in error rate) is below 3% for a wide range of T. This supports our claim that the Rayleigh pdf assumption for filter outputs is valid with high probability when we consider a wide range of textures. Also, note that *R.E.* gives better performance than *S.N.*

To compare the indexing performance of *S.N.* and *R.E.*, we use an aerial image dataset with 90,744 subimages of 128×128 pixels. f_γ is computed for each subimage (using previously computed MPEG-7 features) and a standard VA file index [11] (with S bits per dimension) is constructed for the dataset. The indexing efficiency is quantified by the number of candidates (the smaller the better) obtained after VA filtering for K nearest neighbors. Table 1 shows these numbers (averaged over 20 queries) for *S.N.* and *R.E.*, for $K = 20$, and different values of S.

5. DISCUSSION

When texture images are passed through Gabor filters, the filter outputs have a strong tendency to follow a Rayleigh distribution. Based on this, we have modified the MPEG-7 texture descriptor to have lower dimensionality and computational complexity. This benefits content-based retrieval systems by significantly reducing storage, computational expense, indexing overhead, and retrieval time. We support this approach by demonstrating that the new descriptor performs comparably with the MPEG-7 descriptor.

Table 1. Number of candidates obtained after a 20-NN VA filtering (using VA-file index of f_γ, with S bits per dimension).

S	**8**	**7**	**6**	**5**	**4**
S.N.	74	189	863	3466	9176
R.E.	66	147	440	1460	3481

Another observed phenomenon is that the values along each dimension of both descriptors follow a skewed distribution that can be well modeled by a Rayleigh pdf. Exploiting this behavior, we have proposed a new normalization scheme for the descriptors. We demonstrate that the new scheme improves similarity retrieval performance and indexing efficiency. We are working on a more accurate characterization of the aforementioned distribution.

References

[1] W. Y. Ma and B. S. Manjunath, "NeTra: a toolbox for navigating large image databases," *Multimedia Systems*, 7(3):184–198, May 1999.

[2] F. Liu and R. W. Picard, "Periodicity, directionality, and randomness: Wold features for image modeling and retrieval," *IEEE Trans. PAMI*, 18(7):722–733, 1996.

[3] J. Mao and A. Jain, "Texture classification and segmentation using multiresolution simultaneous autoregressive models," *Patt. Recog.*, 25(2):173–188, 1992.

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

[5] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. PAMI*, 18(8):837–842, Aug 1996.

[6] B. S. Manjunath, P. Salembier, and T. Sikora, Eds., *Introduction to MPEG7: Multimedia Content Description Interface*, John Wiley & Sons Ltd., 2002.

[7] J. G. Daugman, "Complete discrete 2D gabor transforms by neural networks for image analysis and compression," *IEEE Trans. ASSP*, 36:1169–1179, Jul 1988.

[8] S. Marcelja, "Mathematical description of the responses of simple cortical cells," *J. Opt. Soc. Am*, 70(11):1297–1300, Nov 1980.

[9] D. Dunn and W. E. Higgins, "Optimal gabor filters for texture segmentation," *IEEE Trans. Image Proc.*, 4(7):947–964, Jul 1995.

[10] P. Z. Peebles, *Probability, Random Variables, and Random Signal Principles*, McGraw-Hill, 1987.

[11] R. Weber, H.-J. Schek, and S. Blott, "A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces," in *Proc. VLDB*, Aug 1998, pp. 194–205.