MULTIMODAL STATISTICS OF ADAPTIVE WAVELET PACKET COEFFICIENTS: EXPERIMENTAL EVIDENCE AND THEORY

Roberto Cossu, Ian Jermyn, Josiane Zerubia

Ariana (joint research group CNRS/INRIA/UNSA), INRIA Sophia Antipolis
2004 route des Lucioles, B.P. 93, 06902 Sophia Antipolis Cedex, France
firstname.lastname@inria.fr

ABSTRACT

In recent work, it was noted that although the subband histograms for standard wavelet coefficients take on a generalized Gaussian form, this is no longer true for wavelet packet bases adapted to a given texture. Instead, three types of subband statistics are observed: Gaussian, generalized Gaussian, and interestingly, in some subbands, bi- or multi-modal histograms. Motivated by this observation, we provide additional experimental confirmation of the existence of multimodal subbands, and provide a theoretical explanation for their occurrence. The results reveal the connection of such subbands with the characteristic structure in a texture, and thus confirm the importance of such subbands for image modelling and applications.

1. INTRODUCTION

Wavelet and wavelet packet bases have been widely used for image modelling, and in particular for texture description and analysis [1]. This is no accident: these bases are well-adapted to texture description because they can capture spatial dependencies between pixels while remaining localized (unlike the Fourier basis), and are thus adaptable to regions of different shapes. Probabilistic models of textures based on wavelet bases have tended to assume that the wavelets themselves capture all relevant dependencies, and thus that the wavelet coefficients themselves are independent. This has the added advantage of rendering the models simple and efficient to use in practice. However, many textures possess long-range dependencies in rather narrow frequency bands, and independent models of standard wavelet bases are unable to capture such behaviour. One notable exception to this rule of independence are Markov and hidden Markov tree models. These capture dependencies between scales, and are thus well adapted to the description of edge structures in images, since these produce large wavelet coefficients at all scales. The presence of a large wavelet coefficient at one scale thus indicates the likelihood of large coefficients at the same spatial location but different scales. This very suitability to the description of edges means, however, that these models are not well adapted to capturing the dependencies that exist in many textures.

This work was partially supported by EU FP5 project MOUMIR, HP-99-108 (www.moumir.org)

In order to get around the limitations of these two types of model, while retaining the computational tractability associated with them, Brady *et al.* [2] developed models based on wavelet packet bases adapted to each texture being described. The coefficients in these bases were assumed independent, with the difference that the basis itself could adapt to capture the dependencies present in the texture. Thus long-range dependencies in a certain frequency band can be captured not by introducing dependencies between the wavelet coefficients at different spatial locations, but by using a basis containing a wavelet packet with large spatial extension supported in that frequency band. Since the distances between such wavelet packet basis elements are now much larger, the assumption of independence makes more sense: dependencies decrease with distance in most textures.

In [2], the marginal distributions of the wavelet packet coefficients were assume to be Gaussian. This requires some comment, since it is well known that in standard wavelet bases, the subband histograms of natural images often assume a leptokurtic form, frequently modelled by a generalized Gaussian or two-component Gaussian mixture with one large and one small variance [3, 4]. Although there is no reason to expect this behaviour to survive conditioning on a particular texture class, in practice the generalized Gaussian model remains accurate even when one examines the histograms derived from coherent textures. However, the form of such distributions is not preserved under a change of basis, and thus the distribution of the adapted bases is not known *a priori*. Gaussian distributions then provide the maximum entropy choice in the absence of further prior information.

While the models used were Gaussian, it was noted briefly in [5] that the histograms of the resulting adapted subbands did not necessarily conform to the models. As was to be expected, the subbands in the adapted bases that were also standard wavelet subbands (these tended to be at high frequencies/small scales) showed typical leptokurtic behaviour. Many of the other subbands were mesokurtic, and in fact very close to Gaussian. Most interestingly, however, some of the subbands showed a remarkably different behaviour. They were bi-, and sometimes multi-modal, with maxima at finite positive and negative values.

The importance of this fact for texture modelling is clear. In the absence of such phenomena, *i.e.* if all subband his-

tograms (excluding the scaling coefficient subband) have their maxima at zero, the most probably image of that texture class is constant and thus 'untextured'. This conflicts strongly with our prior knowledge: the most probably images of raffia [6], or of forest in a remote sensing image for example, are certainly not constant. Strikingly, most if not all probabilistic models of texture in the literature possess this unintuitive property: the most probable image under these models is constant.

These new statistics are very difficult to observe without the use of adaptive models, because they are swamped by 'noise' that is unrelated to the structures that produce the statistics. The adaptive models act as a microscope, drawing out the important structures in the texture and enabling the study of their properties. Motivated by the observation of the new statistics, and by the efficacy of the adaptive models in the rendering them observable, the approach used in [2, 5] was recently extended in [7] to deal with this new behaviour. Instead of modelling each subband by a Gaussian, three possible models were considered for each adaptive subband: Gaussian, generalized Gaussian, and a constrained mixture of three Gaussians with one component with mean zero, and two components with means of equal magnitude but opposite sign, equal variances (not necessarily the same as that of the Gaussian centred at zero), and equal mixing probability. As in [2, 5], a Bayesian methodology was used to compute MAP estimates for the adaptive basis, for the subband models, and for the subband model parameters.

The aim of this paper is to present further empirical evidence, garnered through the use of these new models, for the existence of bi- and multi-modal subbands in both Brodatz and remote sensing images, and to provide a theoretical explanation, based on simple texture models, of the origins of this behaviour. Section 2 discusses the empirical evidence, while section 3 the theory. The theory conforms that the multimodal subbands are closely related to the principal periodicities present in the texture. Conclusions are drawn in section 4.

2. EXPERIMENTAL EVIDENCE FOR MULTIMODAL SUBBAND HISTOGRAMS

The models used in [7] were trained on a variety of textures. The resulting models are described by several parameters. First, the adapted basis, which is given by a dyadic partition of one quarter of the Fourier domain, which represents the approximate frequency support of the wavelet packets in the different subbands. Second, a choice of one of the three models for each subbands. Subbands for which the constrained mixture of Gaussians was estimated to be the correct model are defined as 'multimodal'. Finally, for each subband, there are the parameters corresponding to the model chosen for that subbands: variance for the Gaussian subbands; variance and shape factor for the generalized Gaussian subbands; and for the mixture of Gaussians, two variances, one mean, and a mixing probability.

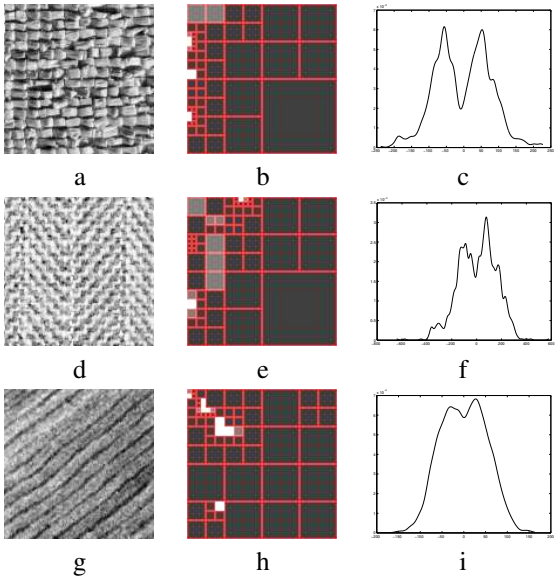


Fig. 1. Textured image, decomposition, and example of multimodal subband (first, second, and third column, respectively). First, second and third row correspond to Raffia, Herring and remote-sensing textures, respectively.

In figure 2, we show the results obtained by training the model on the Raffia and Herring textures from the Brodatz album [6], and on a textured patch extracted from a remote sensing image. Figure 2(b), figure 2(e), and figure 2(h) show the adaptive decomposition obtained by [7] for the three considered textures. The white colour denotes multimodal subbands, *i.e.* those for which the mixture of Gaussians was selected. Examples of histograms from these subbands are shown in the third column of the figure.

The multimodal subbands are typically those that are 'furthest' from a standard wavelet basis, *i.e.* they are the most adapted to the texture at hand. They are also typically narrow in frequency content. The presence in the subband histograms of maxima at non-zero coefficient values thus seems to indicate the likely presence of approximate periodicities running throughout images belonging to that texture class, with frequencies in the support of these subbands. This intuition will be confirmed by the theoretical models to be discussed in the next section.

3. THEORETICAL EXPLANATION FOR MULTIMODAL SUBBAND HISTOGRAMS

The purpose of this section is to show that a number of simple texture models lead directly to the type of bimodal behaviour observed in the experiments above. Based on probabilistic models of images/textures, we will predict the observed histograms of wavelet packet coefficients, and show that they are bimodal.

In order to predict histograms of wavelet coefficients, we need an image model, which is to say a probability distrib-

ution $\Pr(\phi)$ on the space of images Φ. (We assume that all images are functions on a d-dimensional, discrete or continuous torus, of side T.) Given such a model, we can predict the observed histogram using an estimator, for example the mean histogram, under this distribution.

Assuming that the images used to compile the histogram are independent, and using the necessary condition that $\Pr(\phi)$ be translation invariant, we find that the mean histogram under a distribution $\Pr(\phi)$ is given by

$$H(\omega) = Z^{-1} \int_{\Phi} \Pr(\phi)\, \delta(\omega, W(\phi)) \quad ,$$

where ω is the value of a wavelet coefficient in the subband of interest (any one will do, since all the histograms in one subband are the same by translation invariance), W is the linear operator that calculates this coefficient from the image ϕ, and Z is a normalization constant, to be dropped hereafter. Clearly the mean histogram only depends on the marginal probability of an image in the spatial and frequency support of the operator W.

We will construct our models in the Fourier domain, in which case

$$W(\phi) = \sum_{k \in K} w_k^{\dagger} \phi_k = \sum_{k \in K^+} \tilde{\phi}_k = \sum_{k \in K^+} A_k \cos(\theta_k) \quad .$$

Here w_k (resp. ϕ_k) is the Fourier coefficient of W (resp. ϕ) at frequency k, while K (resp. K^+) is the frequency support of the wavelet coefficient (resp. the intersection of this support with an arbitrarily chosen 'positive' half of the Fourier domain). A_k and θ_k are the amplitude and phase of ϕ_k respectively, while $\tilde{\phi}_k$ is the real part. In the second and third equations, we have set all the $w_k = 1/2$, giving an ideal bandpass filter. Although wavelet and wavelet packet filters do not have this behaviour, we do not expect this simplification to change the statistics greatly.

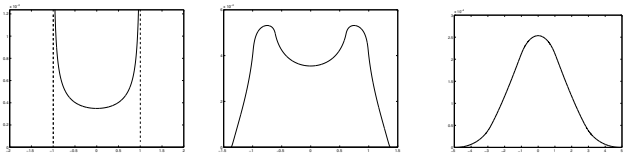
For the mean histogram, we thus have that

$$H(\omega) = \int \Pr(\tilde{\phi})\, \delta(\omega, \sum_{k \in K^+} \tilde{\phi}_k) \tag{1}$$

$$= \int \Pr(A, \theta)\, \delta(\omega, \sum_{k \in K^+} A_k \cos(\theta_k)) \quad . \tag{2}$$

We note that the form of equation 1 means that if the $\tilde{\phi}_k$ are all independent, then the full mean histogram is given by the convolution of their individual mean histograms.

In the following, we derive the mean histogram for a number of simple models $\Pr(A, \theta)$, showing that this histogram is multimodal. Two extreme cases can be distinguished, based on the dependencies between the phases θ_k. The phases may be independent, in which case there is no spatial coherence between the different Fourier components contributing to the wavelet coefficient, or they may be deterministically related, in which case the Fourier components have fixed translations with respect to one another, but may have different amplitudes. If the amplitudes are fixed too, then the distribution represents all possible translations of a single fixed signal.



a b c

Fig. 2. Histograms corresponding to different image models: a) one non-zero Fourier coefficient with fixed amplitude and uniformly distributed phase; b) three non-zero coefficients with fixed amplitudes of 1, 0.01, and 0.01, and independently and uniformly distributed phases; c) five non-zero coefficients of equal amplitude and independently and uniformly distributed phases.

Subsection 3.1 deals the special case in which all the amplitudes A_k are zeros but one. The two extreme cases of independent and deterministically related phases are analysed in subsections 3.2 and 3.4, respectively.

3.1. Single sinusoid

Let us suppose that all the A_k are zero, except for $A_0 = A_{k_0}$, which is equal to a fixed value A, while the θ_k are all uniformly distributed on the unit circle:

$$\Pr(A, \theta) = DA D\theta\, \delta(A_0, A) \left[\prod_{k \in K^+ \setminus k_0} \delta(A_k) \right] \quad ,$$

where $DA = \prod_{k \in K^+} dA_k$ and $D\theta = \prod_{k \in K^+} (d\theta_k / 2\pi)$. The set of images thus consists of a sinusoid of fixed frequency and amplitude, but with unknown translation. In this case, equation (2) is easy to calculate. The result is

$$H_A(\omega) = \int_{-\pi}^{\pi} \frac{d\theta_0}{2\pi}\, \delta(\omega, A\cos(\theta_0))$$

$$= \begin{cases} \frac{1}{\pi}(A^2 - \omega^2)^{-1/2} & \text{if } -A \le \omega \le A \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Figure 2(a) shows a plot of this function. The intuitive reason for the result is clear: a sinusoid 'spends more of its time' near its maximum value than it does near zero.

If we add 'noise' by relaxing the constraint on A_0, for example by replacing the delta function in equation (3) by a narrow Gaussian (to avoid problems with the positivity of the A_0), we still find that the mean histogram is bimodal, although the peaks become progressively smoother as the variance of this Gaussian increases, until eventually bimodality is lost. We can also use a Gamma distribution for A_0^2, which again produces a bimodal mean histogram for all values of the index greater than $3/2$. When the index is $3/2$, the distribution for A_0 becomes Rayleigh, at which point, as is well known, the distributions for ϕ_k and $\tilde{\phi}_k$, and hence the mean histogram of ω, become zero-mean Gaussians, and bimodality is lost. Similar gradual loss of bimodality is observed by

adding complex noise of increasing variance to ϕ_k. Thus, as might be expected, even with a single sinusoidal signal, too much uncertainty about the value of its amplitude destroys the bimodality of the histogram.

The case of a single sinusoid of frequency k_0 in white noise is similar. The variance of the noise at frequencies other than k_0 but in the support of the wavelet operator W adds to the variance of the noise at k_0 itself, so that even if the latter is not sufficient in itself to destroy bimodality, the cumulative effect of the other frequencies can easily become large enough to do so.

3.2. Independent phases

We consider now the case in which more than one of the A_k is non-zero. First, let them all have fixed values B_k, while the phases θ_k are all independently and uniformly distributed, as before:

$$\Pr(A, \theta) = DA\, D\theta\, \delta(A, B) \quad . \tag{4}$$

The A_k integrals in equation (2) are again trivially performed, giving

$$H_{\{B_k\}}(\omega) = \int \left[\prod_{k \in K^+} \frac{d\theta_k}{2\pi} \right] \delta(\omega, \sum_{k \in K^+} B_k \cos(\theta_k)) \quad .$$

The resulting mean histogram is therefore the convolution of a number of mean histograms of the form in equation (3) for different values of A.

If among the non-zero B_k values, one of them has an energy significantly higher than the others, then the effect of this convolution is to round the double peak function in equation (3). An example is shown in figure 2(b). This corresponds to the case where the signal possesses a strong periodicity in the frequency support of the subband considered, but some residual noise is present also.

On the other hand, if the B_k all have approximately the same value, then the individual mean histograms are all the same, and thus, as the number of non-zero B_k values increases, their convolution tends towards a Gaussian, as guaranteed by the central limit theorem. The double peak disappears using only a few values. An example is shown in figure 2(c). This result clearly holds whenever the $\tilde{\phi}_k$ are iid.

To validate the above, we carried out the following numerical experiment. The values B_k were computed from

$$B_{k_n} = \exp\{-(k_n - k_0)^2 / 2\sigma^2\} \quad ,$$

where $\{k_n = k_0 + n\delta k : n \in [-N, \ldots, N]\}$ were a number of equally spaced frequencies centred on k_0, and σ, when it is small enough, effectively controls the number of non-zero B_k. We then sampled, for each k_n, many random phases $\theta_{n,i}$, and computed the histograms of the resulting values of $B_{k_n} \cos(\theta_{n,i})$. In figure 3, these histograms are plotted against σ. For small σ, that is with only a few non-zero B_{k_n}, and with B_{k_0} much larger than $B_{k_{\pm 1}}$, the histograms are strongly bimodal. As σ is increased, the histogram is smoothed, until for large enough sigma, the bimodality is destroyed and the histogram becomes unimodal.

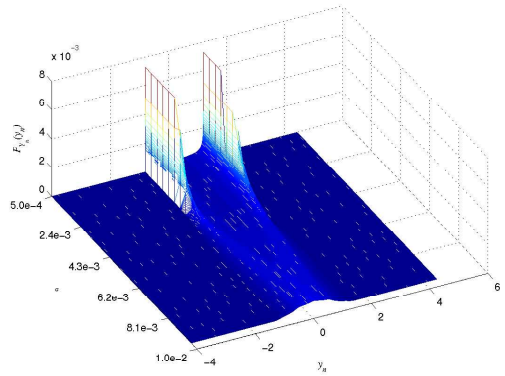


Fig. 3. Histograms of wavelet coefficients for a signal whose Fourier coefficient amplitudes at neighbouring frequencies take on a Gaussian shape, plotted against σ.

3.3. Comment

The behaviour described at the end of section 3.1 and in the above is directly relevant to adaptive wavelet packets. In the above case, even though the values of the amplitudes A_k were fixed, a number of nearby frequencies with similar amplitudes but independent and uniformly distributed phases could destroy the bimodality. In the case of section 3.1, it was the inclusion of nearby noisy frequencies that was responsible. These factors are of course controlled by the image probability distribution, but importantly they are also cut off by the support of the wavelet operator W. The point of adaptive wavelet packets is precisely to narrow the support of W in an appropriate way in order to exclude such confounding factors, thus making them more likely to reveal bimodality than unadapted wavelets or wavelet packets. Indeed the ease with which bimodality was destroyed in the above makes it unlikely that this behaviour could be discovered without the use of adaptive bases. In the next subsection, we illustrate another factor that supports bimodality: phase coherence.

3.4. Single translated signal

In all the image models considered so far, the $\tilde{\phi}_k$ were independent. We now consider a case in which the phases are dependent, while the amplitudes remain fixed. In particular, we consider a distribution whose support is all translations of a given signal:

$$\Pr(A, \theta) = DAD\theta\, \delta(A, B) \int_T \frac{dt}{T} \prod_{k \in K^+} \delta(\theta_k, \theta_{k,0} + kt) \quad ,$$

As a consequence of the dependence between the phases, the histogram can no longer be computed as a convolution. It is not easy to obtain general analytical results, but numerically it is simple. We proceeded as follows. In order to construct the A_k and $\theta_{k,0}$, we chose a finite-length discrete sinusoid for which the signal length was not a multiple of the period. The discrete Fourier transform of this signal presents a number of

non-zero coefficients with fixed phase relationships. If these phase relationships are not preserved, then the structure of the 'texture' is lost. We then created a large number of shifted versions of the signal by uniformly sampling values of t, each shifted version then being filtered and downsampled (using an ideal bandpass filter). We then computed the histogram of the resulting signals.

Figure 4(a) shows the DFT of a discrete sinusoid of length 64 and frequency 0.18. We considered 6400 equally spaced translations of the signal, obtaining 6400 different values of the phase for each DFT coefficient. At this point we filtered the signal, extracting the coefficients with wavenumbers from 9 to 16. The resulting histogram is plotted in figure 4(b). As one can see, despite considering several non-zero coefficients, the distribution takes on a bimodal form. The preservation of the bimodality is due to the relations that exist between the phases, as can be seen by looking at figure 4(c), which shows the histogram of a signal whose DFT amplitudes are the same as before but whose phases were sampled independently. The histogram is unimodal, in agreement with the results shown in figure 3.

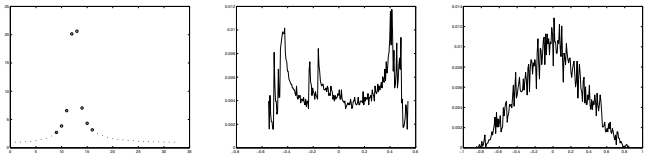


Fig. 4. Absolute values of the DFT of a finite-length discrete sinusoidal signal (a). Histogram obtained when: b) retaining phase information; c) discarding phase information

4. CONCLUSIONS

In this paper we have provided experimental evidence and a theoretical explanation for the multimodal statistics of adaptive wavelet packet coefficients. In addition to confirming this behaviour, the theoretical explanation reveals the importance of such multimodal subbands for texture description. Multimodal distributions, usually occurring in subbands narrow in frequency content, indicate the likely presence of approximate periodicities running throughout images belonging to that texture class, with frequencies in the support of these subbands. Multimodal subbands thus represent the characteristic structure, *i.e.* the main periodicities, present in images of a texture class.

The microscope effect of the new models goes beyond the ability to identify bi- and multimodality. In [8], it was observed that the joint statistics of adapted wavelet packet bases also show remarkable new behaviour, and our current research efforts are focused on modelling these joint statistics.

Such a search for more accurate and more specific image models, *i.e.* the untangling of the mixture distributions that one sees when one looks at images as a whole into separate components that model well individual entities in an image, is important because the resulting models are applicable not just to texture description and segmentation, but to any problem of image processing. It seems highly likely, for example, that the multimodal models used in this paper would work well in denoising problems, since they separate important structure from noise.

5. REFERENCES

[1] S. Livens, P. Scheunders, G. Van de Wouwer, and D. Van Dyck, "Wavelets for texture analysis, an overview," in *Proc. IEE Int'l Conf. Im. Proc. and its Applications*, Dublin, Ireland, July 1997, pp. 581–585.

[2] K. Brady, I. H. Jermyn, and J. Zerubia, "Texture analysis: An adaptive probabilistic approach," in *Proc. IEEE Int'l Conf. Image Procesing*, Barcelona, Spain, September 2003.

[3] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 2^{nd} edition, 1999.

[4] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Liebler distance," *IEEE Trans. Image Processing*, vol. 11, pp. 146–158, 2002, 2.

[5] K. Brady, I. H. Jermyn, and J. Zerubia, "Adaptive probabilistic models of wavelet packets for the analysis and segmentation of textured remote sensing images," in *Proc. Brit. Mach. Vis. Conf.*, Norwich, U. K., September 2003.

[6] P. Brodatz, *Textures: A Photographic Album for Artists and Designers*, Dover, New York, USA, 1966.

[7] R. Cossu, I. H. Jermyn, and J. Zerubia, "Texture analysis using probabilistic models of the unimodal and multimodal statistics of adaptive wavelet packet coefficients," in *Proc. IEEE Int'l Conf. Acoustics, Speech and Signal Processing*, Montreal, Canada, May 2004.

[8] R. Cossu, I. H. Jermyn, and J. Zerubia, "Texture discrimination using multimodal wavelet packet subbands," in *Proc. IEEE Int'l Conf. Image Procesing*, Singapore, October 2004.