ADVANCES IN TEXTURE ANALYSIS: ENERGY DOMINANT COMPONENT & MULTIPLE HYPOTHESIS TESTING

Iasonas Kokkinos, Georgios Evangelopoulos and Petros Maragos

National Technical University of Athens,
Computer Vision, Speech Communication and Signal Processing Group
School of Electrical and Computer Engineering, Athens, 15773, Greece.
{jkokkin,gevag,maragos}@cs.ntua.gr, http://cvsp.cs.ntua.gr/

ABSTRACT

Modelling textured images as AM-FM functions has been applied during the last years to texture analysis and segmentation tasks. In this paper we present some advances in two directions, namely the improvement of modulation feature extraction and texture vs. non texture discrimination. First we present a modified dominant AM-FM component analysis scheme based on non-linear energy operators. Subsequently we propose a novel approach to the discrimination between textured and non-textured image areas using multiband filter responses which is formulated as a statistical multiple hypothesis testing problem. The theoretical insights of this approach are supported by experimental validation using natural images.

1. INTRODUCTION

The detection and analysis of textured regions in images is a central research topic in image processing and computer vision, since textures are present in almost all natural images. A recently developed model for decomposing an image into its essential components is the AM-FM modulation model [4, 7] which describes intensity variations in magnitude and orientation in terms of instantaneous frequencies of spatially varying 2D modulated sinusoids. Extraction of modulation information has been widely used so far for image analysis, classification and segmentation [13, 4, 9, 15].

Multiband filtering approaches [3, 1, 14] for texture analysis are commonly used to derive a representation of an image which lends itself more easily to analysis. The Dominant Component Analysis (*DCA*) scheme [9, 10], which aims at capturing local image oscillations as laid out by the AM-FM image model, chooses at each pixel the most powerful of these channels and estimates the AM-FM model parameters using the outputs of that channel.

In this work we present some advances in feature extraction from modulation components and present a statistical hypothesis testing method for the task of discriminating between textured and non-textured image regions. In the first part we propose an alternative method for DCA based on 2D energy operators that is characterized by strong localization and boundary preservation. Subsequently we propose a multiple hypothesis testing formulation for the texture vs. non texture selection and for the channel selection criterion.

This work has been supported partially by the GME research program Hrakleitos, the GSRT research program PENED-2001, the NTUA research program Protagoras and the EC NoE MUSCLE.

2. IMAGE MODULATIONS ANALYSIS

2.1. Modelling Images as AM-FM Modulations

The use of multiple sinusoidal modulating signals for modelling images has been proposed and explored by Bovik [4] and Havlicek in [7, 9], according to which any 2D narrowband signal can be expressed in terms of spatially varying amplitude and frequency modulations, namely as a superposition of locally smooth, nonstationary sinusoids:

$$I(x,y) = \sum_{k=1}^{K} a_k(x,y)\cos[\phi_k(x,y)]. \qquad (1)$$

$a_k(x,y)$ and $\vec{\Omega}_k(x,y) = \nabla\phi_k(x,y)$ are the Amplitude and Frequency Modulation signals of the K image components respectively, modelling image contrast and locally emergent frequency variations respectively. The FM signal may be decomposed in the instantaneously varying horizontal $\Omega_{1,k} = \partial\phi_k/\partial x$ and vertical $\Omega_{2,k} = \partial\phi_k/\partial y$ frequency signals.

A set of bandpass filters is used to decompose the image into narrowband components and a demodulation algorithm is applied to each of them. A common choice for the filterbank are Gabor filters which achieve the optimal joint time-frequency uncertainty minimization in both 1-D [6] and 2-D [5, 3]. Demodulation techniques of single AM-FM component images include the 2D Energy Separation Algorithm (ESA) [13] based on the differential Teager-Kaiser energy operator and the analytic image extension [10].

2.2. Dominant Component Features

As an alternative to channelized analysis and demodulation of the narrowband single components of (1), DCA offers a rigorous and more compact way for capturing locally emerging modulation signals. After decomposing the image in narrowband components, the amplitude and frequency estimates of each pixel are estimated on a pointwise basis using the channel that maximizes an activity criterion:

0-7803-8554-3/04/$20.00 ©2004 IEEE.

$$\Gamma_i(x,y) = \frac{|(J * h_i)(x,y)|}{max_{\vec{\Omega}}|H_i(\vec{\Omega})|}, \;\; J(x,y) = I(x,y) + j\hat{I}(x,y). \tag{2}$$

In (2) * denotes convolution, h_i, H_i are the i_{th} filter's impulse and frequency response and J is the analytic counterpart of I with $\hat{I}$ its 2D directional Hilbert transform. According to the *analytic image model* [8], any real-valued image can be extended to a complex-valued one, using an adjusted directional Hilbert transform, uniquely assigning instantaneous amplitude and phase signals to an image AM-FM component of the form $a(x,y)e^{j\phi(x,y)}$. A demodulation algorithm has been developed to approximately extract the amplitude and signed frequency signals from the analytic image J. Usually post-processing and smoothing techniques are required in order to suppress error effects. A detailed descriptions of the above can be found in [10].

2.3. Modulation Energy and Separation

An efficient and computationally simple approach to extracting modulation features from AM-FM modelled images (1) was presented in [13], based on the 2D Energy Operator, $\Psi(I) \triangleq ||\nabla I||^2 - I\nabla^2 I$, an extension of the 1D Teager-Kaiser Energy Operator used for speech demodulation. Applying Ψ to a 2D AM–FM signal $I_k(x,y) = a_k(x,y)\cos[\phi_k(x,y)]$ yields

$$\Psi[a_k \cos(\phi_k)] \approx a_k^2 ||\vec{\Omega}_k||^2 \tag{3}$$

which equals the product of the instantaneous amplitude and frequency magnitude squared which may be interpreted as the *image modulation energy*. The above approximation error is negligible assuming that the instantaneous amplitude and frequency do not vary too fast in space or too greatly in value compared to the carriers. Furthermore, if we apply the energy operator on the image derivatives $\partial I/\partial z_1$ and $\partial I/\partial z_2$, where $\vec{z} = (x,y)$ are the image vector coordinates, it is possible to separate the energy into its amplitude and frequency components via the 2D ESA [13]:

$$\sqrt{\frac{\Psi\left(\frac{\partial I}{\partial z_n}\right)}{\Psi(I)}} \approx |\Omega_n(\vec{z})|, \quad \frac{\Psi(I)}{\sqrt{\sum_{n=1}^{2}\Psi\left(\frac{\partial I}{\partial z_n}\right)}} \approx |a(\vec{z})| \tag{4}$$

where $n = 1, 2$. This algorithm can estimate at each location $\vec{z}$ the amplitude envelope and the magnitude of instantaneous frequencies of the spatially-varying AM–FM signal.

3. IMPROVEMENTS AND A NOVEL VIEWPOINT FOR THE CHANNEL SELECTION PROCEDURE

3.1. Energy-based Dominant Component Selection

In this subsection we examine the usage of the image modulation energy measurement as a decision criterion for a modified energy-based DCA scheme. Motivated by the capability of the 2D Energy Operator to incorporate both amplitude and frequency magnitude information, we aim to combine various filters outputs in a multiband analysis process. By intuition, we believe that image modulation energy will succeed in choosing filter outputs with lower amplitude levels, where conventional DCA fails, due to a possible local high spatial frequency content. This is important in cases where edges and sharp texture variations are present, since the extracted features should retain boundary information.

We propose a DCA scheme in which the amplitude and frequency values for each pixel are chosen from the ESA - demodulated filter output that maximizes the Energy Operator response. We mention that to the best of our knowledge using the ESA algorithm resulted in improved results when compared with the analytic image approach. The dominant filter is indexed by i where:

$$i = \arg\max_j \Psi[(I * h_j)(x,y)] \tag{5}$$

We compare the performance of our approach with that of the decision criterion (2) in the extraction of dominant modulations features from synthetic, textured and natural images. We use a bank of 40 Gabor filters in 4 different scales and a radial frequency domain arrangement. Implementation of (2) is done without explicit use of the analytic image model, using the theoretically equivalent quantity:

$$\Gamma_i^*(x,y) = \frac{\sqrt{(I * h_{ei})^2(x,y) + (I * h_{oi})^2(x,y)}}{max_{\vec{\Omega}}|H_i(\vec{\Omega})|} \tag{6}$$

where h_{ei} and h_{oi} are the the real and imaginary parts of the complex Gabor filter $h_i = h_{ei} + jh_{oi}$. This way we bypass the explicit estimation of the analytic image representation as well as the singularities that may be caused by the division in (4).

3.2. A Detection Theoretic Perspective for DCA

It would be interesting to somehow relate a likelihood term with the channel selection procedure. For example this would make it possible to incorporate alternative hypotheses, like the *absence* of an AM-FM signal, into the DCA analysis; thereby spurious amplitude or frequency measurements can be avoided, by setting these to zero.

We cast the problem of channel selection as a multiple hypothesis testing problem [11], yet in a somehow different form than the one used e.g. in communications theory. For simplicity we use the 1-D setting, even though the arguments carry over immediately to two dimensions. In the proposed setting, selecting a channel amounts to detecting a sinusoidal with the same frequency as that of the channel and unknown phase and amplitude. The hypothesis that best explains the signal data corresponds to the winning channel.

An extra factor that comes into play is the locality of the decision process: assuming that we want to decide at a point whether a sinusoid of a specific frequency exists in its vicinity we test how well points near it match this hypothesis; far away points are modelled with a less informative model, namely a white noise process.

Pinning things down, we form the hypothesis $\mathcal{H}_\Omega$ that around the point $x = 0$, the signal is a sinusoidal of frequency Ω, unknown phase ϕ, amplitude a, and DC value b. We express the likelihood of the signal I at point x as:

$$P(I(x);a,b,\phi|\mathcal{H}_\Omega) = G(x)\frac{e^{-\frac{(I(x)-(a\cos(\Omega x+\phi)+b))^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} + (1-G(x))\frac{1}{\sqrt{2\pi}\sigma_n}e^{-(I(x)-C)^2/2\sigma_n^2}.$$

$G(x) = e^{-x^2/\sigma_G^2}$ is a weighting factor accounting for the locality of the decision process, σ is the variance of a white gaussian noise (WGN) process that is assumed to contaminate our signal, and σ_G models the spatial decay of the 'ability' of the hypothesis to explain the data. The second term is a background model which is assumed to be a white gaussian noise process of mean C and variance σ_n. As we shall see, setting σ_G equal to the spread of the Gabor filter that is used in the corresponding channel results in a link between the amplitude based channel selection and the likelihood based one.

Using the WGN assumption, the likelihood of a signal patch around point 0 under the hypothesis $\mathcal{H}_\Omega$ is given by:

$$P(I;a,b,\phi|\mathcal{H}_\Omega) = \prod_{x=\ldots-1,0,1,\ldots} P(I(x);a,b,\phi|\mathcal{H}_\Omega) \quad (7)$$

Taking the logarithm and using the fact that log is a concave function we derive the following lower bound on the likelihood $\ln(P(I)) = \sum_x \ln(P(I(x)))$ of the data:

$$\sum_x G(x)\left\{-\frac{(I(x)-(a\cos(\Omega x+\phi)+b))^2}{2\sigma^2} - \frac{\ln(2\pi\sigma)^2}{2}\right\} + \sum_x (1-G(x))\left\{\frac{-(I(x)-C)^2}{2\sigma_n^2} - \frac{\ln(2\pi\sigma_n^2)}{2}\right\}$$

Even though the last equation is a lower bound on the likelihood of the data, it can be used to choose among two conflicting hypotheses; henceforth we shall be referring to it as the (log)likelihood of the data, for simplicity. We use a test similar to the Generalized Likelihood Ratio Test (GLRT), which means we calculate the Maximum Likelihood (ML) estimates of the a, b, ϕ parameters, and insert them into the likelihood term. By maximizing (7) w.r.t. a, b, ϕ we derive:

$$\hat{a} = \sqrt{\frac{(h_e * I)^2 + (h_o * I)^2}{(\sum G(x))^2}}, \hat{\phi} = \tan^{-1}(-\frac{h_o * I}{h_e * I}) \quad (8)$$

and $\hat{b} = G * I$ where h_e, h_o are an even/odd filter pair $h_e(x) = G(x)\cos(\Omega x)$, $h_o(x) = G(x)\sin(\Omega x)$. Using a G function of the same spread with that of the one used in the corresponding Gabor filter, we observe that $\hat{a}$ is the estimated amplitude of the channel response, while $\hat{b}$ is what we would interpret as the mean of the data at that point. The first term in the bound can then be shown to equal $\hat{a}^2/(2\sigma^2/\sum_x G(x))$, which is up to a constant scaling factor the criterion (6) to choose a certain channel. Since the rest of the terms are constant for all channels at a specific scale, we can ignore them, when working at a single scale. Using exactly the above rationale, we can calculate the likelihood of the hypothesis that the image is piecewise constant around the point 0, by excluding the sinusoidal component; however, to compare this likelihood with that of the hypothesis that our data are sinusoidal-like we have to take into account that fewer parameters are used to build the image model: otherwise a sinusoidal with zero amplitude could equally well explain the data. This is accounted for by using the Minimum Description Length (MDL) criterion [11] which, modified for our case, becomes

$$MDL(\mathcal{H}_i) = -\ln(P(I|\mathcal{H}_i)) + n_i/2\ln(\sum_x G(x)) \quad (9)$$

with n_i the number of parameters used for each hypothesis.

4. EXPERIMENTAL RESULTS

In this section we present experimental results with the proposed methods. In Fig. 1(b,c) and (e,f) we compare the decisions made based on the amplitude measurement, (6) with the ones made using our 2D Teager Energy based approach, (5); henceforth we shall call the two approaches ADCA and EDCA for simplicity. EDCA generally outperforms ADCA in terms of localization accuracy. One can note for example in Fig. 1(b) how spread far away from the leopard boundaries the amplitude is nonzero, influenced by the existence of the leopard nearby, contrary to the case where EDCA is used for channel selection. In Figs.1 (e-f) we note that using ADCA the legs of the zebra are interpreted as a spread-out low frequency region, while using the EDCA the results are better localized. It should be noted, however, that in some cases EDCA tends to favor channels responding to sharp changes in the image, even when smooth changes visually prevail.

The second set of experiments demonstrates the application of the hypothesis testing algorithm mentioned previously. In Fig.2 (b)-(d) we present the nature of this texture vs. non-texture decision where the texture indexes are shown as an intensity image. White parts of the image correspond to areas where the non-texture hypothesis prevails, while gray level codes the orientation of the winning Gabor filter. It is intuitively appealing that at small scales fine structure is interpreted as texture (Fig. 2 (f,g)) while at larger scales (Fig. 2 (h)), larger image areas are interpreted as textured (e.g. the folds on the madrill's nose) and at the same time smaller scale areas are interpreted as a constant signal contaminated by noise, instead of texture.

As a first application, this can help refine the initial results by computing a texture vs. non-texture mask at each scale: for each pixel, after making a decision using a DCA algorithm we estimate the likelihood ratio of the winning hypothesis and the hypothesis that a non-textured model 'explains' the signal at the same scale near that point. Pixels where the non-texture hypothesis prevails are assigned with an amplitude and frequency value equal to zero. The refined features, extracted via modulations modeling and the EDCA methodology can be seen in Fig.3. Clearly the non-textured hypothesis prevailed in the case of the flat ground surface which was removed from the initial estimations for the dominant instant Amplitude and Frequency Magnitude (b)-(c). This is also evident by the black labelled pixels on the Channel Index image in (d) which, for this image, correspond to non-textured regions.

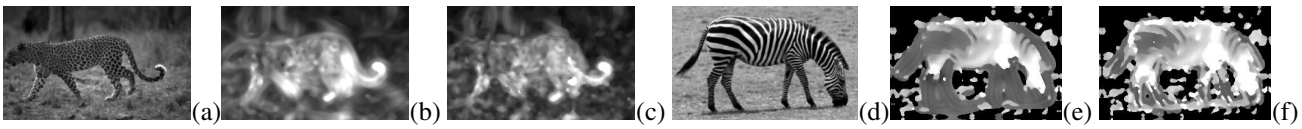

Fig. 1. Energy vs. Amplitude based Dominant Component Analysis: **(a)** *Leopard* Image Amplitude Estimate with **(b)** ADCA and **(c)** EDCA respectively. **(d)** *Zebra* Image and Frequency Magnitude Estimate with **(e)** ADCA and **(f)** EDCA.

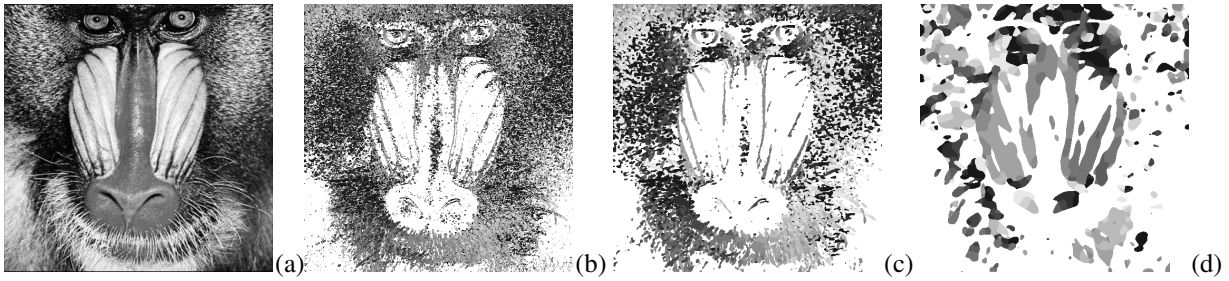

Fig. 2. **(a)** Madrill Image, **(b)-(d)**: Texture vs. Non-texture decision at increasing scales. White pixels stand for non-texture.

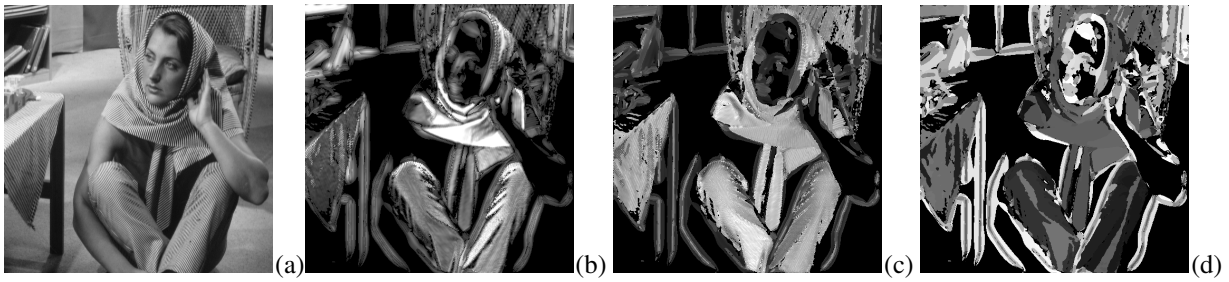

Fig. 3. Modulation Features by Energy-based DCA: **(a)** *Barb* Image, **(b)** Amplitude Estimate with EDCA and texture mask, **(c)** Estimated Frequency Magnitude, **(d)** Active Filter Index.

5. CONCLUSIONS

In this paper we have presented some advances in the AM-FM modelling of textured images, by examining different channel selection criteria to perform DCA, as well as a statistical hypothesis testing viewpoint of channel selection. We believe the proposed approach is of general interest, since it offers a different viewpoint for Gabor filtering, thresholding the filtering results, etc. Finally, we note that the extracted features have been used in a companion paper [12] for curve evolution-based textured image segmentation.

6. REFERENCES

[1] A.C. Bovik, "Analysis of Multichannel Narrow-Band Filters for Image Texture Segmentation", *IEEE Trans. on Sig. Proc.*, Vol. 39(9), 1991.

[2] A.C. Bovik, M. Clark and W.S. Geisler,"Texture Segmentation using a Class of Narrowband Filters",*Proc. ICASSP*, 1987.

[3] A.C. Bovik, M. Clark and W.S. Geisler, "Multichannel Texture Analysis using Localized Spatial Filters",*IEEE Trans. PAMI*, Vol. 12(1), pp. 55-73, 1990.

[4] A. C. Bovik, N. Gopal, T. Emmoth, and A. Restrepo, "Localized Measurement of Emergent Image Frequencies by Gabor Wavelets", *IEEE Trans. Info. Theory*, Vol. 38, pp. 691–712, 1992.

[5] J. E. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters", *J. Opt. Soc. Amer. A*, Vol. 2(7), pp.1160-1169, 1985.

[6] D. Gabor, "Theory of Communication", *Journal of the IEE*, Vol. 93(III), pp. 429-457, 1946.

[7] J.P. Havlicek, D.S. Harding, and A.C. Bovik, "The mutli-component AM-FM image representation", *IEEE Trans. Image Proc.*, Vol. 5(6), pp. 1094-1100, 1996.

[8] J.P. Havlicek, J.W. Havlicek, and A.C. Bovik, "The analytic image," in *Proc. ICIP*, 1997.

[9] J.P. Havlicek, D.S. Harding, and A.C. Bovik, "Multidimensional quasi-eigenfunction approximations and multicomponent AM-FM models", *IEEE Trans. Image Proc.*, Vol. 9(2), pp. 227-242, 2000.

[10] J.P. Havlicek and A.C. Bovik, "Image Modulation Models", in *Handbook of Image and Video Proc.*, A.C. Bovik, ed., pp. 305-316, 2000.

[11] S.M. Kay, "Fundamentals of Statistical Signal Processing: Detection Theory", Prentice Hall, N J, 1993.

[12] I. Kokkinos, G. Evangelopoulos and P. Maragos, "Modulation-Feature Based Textured Image Segmentation using Curve Evolution", in *Proc. ICIP* 2004.

[13] P. Maragos and A.C. Bovik, "Image demodulation using multidimensional energy separation", *J. Opt. Soc. Amer. A*, Vol.12(9), pp. 1867-1876, 1995.

[14] J. Malik and P. Perona, "Preattentive texture discrimination with early vision mechanisms", *J. Opt. Soc. Amer. A*, Vol.7(5), pp. 923-932, 1990.

[15] N. Ray, J. Havlicek, S. Acton, and M. Pattichis, "Active contour segmentation guided by AM-FM DCA", in *Proc. ICIP* 2001.