Image Decomposition into Structure and Texture Subcomponents with Multifrequency Modulation Constraints

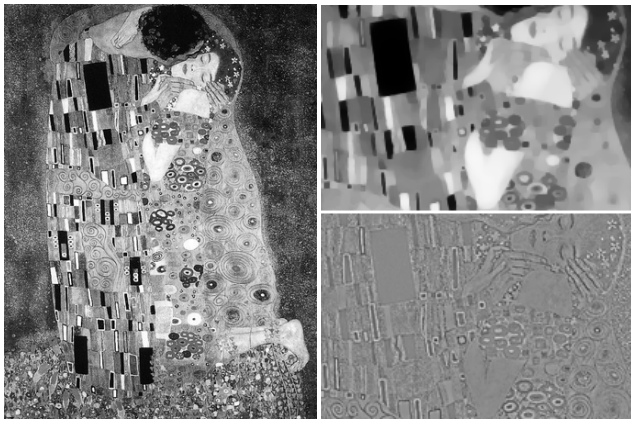
Georgios Evangelopoulos Petros Maragos *
School of Electrical and Computer Engineering
National Technical University of Athens, Greece
gevag@cs.ntua.gr,maragos@cs.ntua.gr

Abstract

Texture information in images is coupled with geometric macrostructures and piecewise-smooth intensity variations. Decomposing an image f into a geometric structure component u and a texture component v is an inverse estimation problem, essential for understanding and analyzing images depending on their content. In this paper, we present a novel combined approach for simultaneous texture from structure separation and multiband texture modeling. First, we formulate a new, variational decomposition scheme, involving an explicit texture reconstruction constraint (prior) formed by the responses of selected frequency-tuned linear filters. This forms a '$u + Kv$' image model of $K + 1$ components. Subsequent texture modeling is applied to the estimated v component and its consistency is compared to using the complete, initial image f. The decomposition step, functioning as an advanced texture-front end, improves clustering and classification performance, for various multiband features. The proposed method can be generalized to other texture models or applications.

1. Introduction

Visual texture is perceived as small-scale intensity variations and patterns that represent the finer characteristics of depicted scenes. Such information in natural images coexists with geometric macrostructures such as object contours, shapes and boundaries, or is embedded in coarser structures formed by lighting and shading conditions or smooth area/volume variations. Image decomposition is the process of separating an image in conceptually and theoretically different components, primarily under two perspectives i) addressing content-specific vision applications (*e.g.* separate structure and detail analysis) and ii) studying image pattern formation (*e.g.* multilevel representations, atomic decompositions). Inverse problems such as diffusions, scale-spaces and simplification techniques [4, 19], wavelet projections [14, 16, 17] and representations in bases [8, 20] or learned dictionaries (textons, primitives) [11], may be considered special cases of decompositions.



(a) Image (b) 'Cartoon' and texture (detail)

Figure 1: Painting (a) is a mixture of geometric structures ('cartoon') and textural patterns (*"The Kiss", Gustav Klimt, 1908*)

In the $u + v$ models for structure-texture decomposition [3, 20, 21], images are composed by a smoothly-varying, geometry-containing part and a small-scale, oscillatory part capturing textures, details and noise (Fig. 1). The two different in nature components require content-specific modeling and representations, and motivate an analysis in two parallel channels. The texture component can be used for texture segregation and classification, surface analysis, shape/orientation from texture, while the structure part for 1D feature detection (edges, ridges), segmentation, object recognition and shape analysis. In addition to emerging applications like digital inpainting [5], image abstraction and computational photography [4], decomposition has provided solutions to problems like image restoration [10], matching, compression, segmentation, classification [2, 21] and face recognition [7].

*This work was supported by grants ΠΕΝΕΔ-2001 and ΠΕΝΕΔ-2003-ΕΔ865 [co-financed by E.U.-European Social Fund (75%) and the Greek Ministry of Development-GSRT (25%)], and by the European Commission FP6-IST research projects MUSCLE, HIWIRE and ASPI.

In treatises of image texture, linear filter models augmented by appropriate nonlinear operators constitute primate texture detectors, in par with biologically motivated mechanisms [9, 14]. Multifrequency representations interpret texture as a linear combination of multiple frequency and orientation selective components [6, 17, 18]. Such narrowband components have been further modeled by spatial modulations of the image function [15].

Motivated by the conceptual analogy of texture defined as an oscillating function [16, 21] (Sec. 2) and modeled by spatial modulations [12, 13, 15] (Sec. 3), we propose a novel decomposition scheme, building on the $u+v$ image models, formulated as constrained total variation minimization (Sec. 4). This is shown to be a special case of decomposition in $K+1$ components $(u, v_1, ...v_K)$. An explicit model prior is included to penalize the dissimilarity between the texture part and reconstruction from a sum of narrowband image components. This is achieved from the responses of Gabor filters in multiple frequency bands with variants of the scheme depending on the component selection method (Sec. 4.2). Model fitting proves to be more consistent on the estimated texture component and results in more descriptive features for classification (Sec. 5).

2. Image decompositions

Variational decomposition and u+v models: In the '$u+v$' framework, an image $f : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ is modeled by the sum of two independent components: a piecewise smooth function u, the 'cartoon,' with quasi-flat intensity plateaus and jump discontinuities, that contains geometric information (edges, contours, large-scale features and illumination effects) and an oscillating function v that captures texture variations and small-scale features, *e.g*. Fig. 6 (c). A scale parameter controls the amount of detail in the two perceptually complementary components (Fig. 1).

The partial-differential equation (PDE-based) approach [1, 3, 10, 21], that generalizes the Mumford-Shah and Rudin-Osher-Fatemi (ROF) models [19], solves the inverse component estimation problem by minimization of a general convex functional

$$\inf_{(u,v)\in(U\times V)} \{E(u,v)=J(u)+\lambda F(u,v)+\mu L(v)\}, \quad (1)$$

given $f = u + v$, where $U, V \subset \Omega$ are functional spaces fit for each component and $\lambda, \mu \geq 0$ tuning constants. The first, regularizing term is normally the Total Variation (TV) norm $J(u) = \int_\Omega ||\nabla u|| dxdy$, minimized for functions in the space of Bounded Variation (BV). A second, fidelity term $F(u, v) = ||f - u - v||^2_X$ penalizes the approximation of f by $u + v$ and $L(v)$ is a metric of texture variations in a normed functional space. The decomposition coefficients λ, μ control respectively, the amount of detail in u, *i.e*. scales larger than $1/\lambda$, and the amount of variation in v.

The texture component was defined by Meyer [16] in a Banach space G of oscillating functions that satisfy $v = \mathrm{div}(\vec{g}) = \partial_1 g_1 + \partial_2 g_2$, where $g_1, g_2 \in L^\infty(\mathbb{R}^2)$. Space G is in some sense the dual of the BV space and equipped with a norm $\|\cdot\|_G$, computed as the infimum of the L^∞ norms of $|\vec{g}|$ of all possible decompositions. Norm $\|v\|_G$ was approximated computationally by Vese and Osher [21] using Sobolev norms $L^p,\ 1 \leq p \leq \infty$ in the functional

$$E(u,\vec{g}) = \int_\Omega (\|\nabla u\| + \lambda |f - u - \mathrm{div}(\vec{g})|^2) dxdy + \mu \|\vec{g}\|_p \quad (2)$$

with $(u, v) \in BV \times L^p$, leading to a three-part decomposition where the residual $w = f - u - v$ models image noise. Letting $\lambda, p \rightarrow \infty$ the solution approximates Meyer's model.

Aujol *et al*. [1] approached Meyer's model using the exact definition of $\|\cdot\|_G$ and the indicator function of the set $\|v\|_G \leq \mu$ as a constraint. Whereas, a selective decomposition w.r.t. texture frequency and orientation [3] was formulated by minimizing

$$E(u, v) = \int_\Omega ||\nabla u|| dxdy + \lambda ||v||^2_{\mathcal{H}}, \quad (3)$$

where $(u, v) \in BV \times \mathcal{H}$ and the Hilbert norm $\|f - u\|_{\mathcal{H}}$ is estimated by the projection of $v = f - u$ onto predefined Gabor wavelets g_k, *i.e*. $\|f - u\|_{\mathcal{H}} = \langle f - u, g_k * (f - u)\rangle$.

Multifrequency decomposition: Image decompositions in multiple spectral bands [6, 14] have been extensively studied in early vision and low level processing due to similarities with biological mechanisms [9, 14], information coding efficiency and descriptive capabilities of the representation [17, 18]. They are obtained via linear transformations, wavelet projections or filtering by a set of frequency and orientation tuned functions. Gabor filters are optimum w.r.t. uncertainty in space-frequency localization [6] and their complex responses come in quadrature pairs

$$g_k(x,y) = \frac{1}{2\pi\sigma_k^2} \exp\left(-\frac{x^2+y^2}{2\sigma_k^2}\right) \exp\left(jw_{k1}x + jw_{k2}y\right), \quad (4)$$

where σ_k determines the spatial support and bandwidth of the isotropic filter and $\vec{w}_k = (w_{1k}, w_{2k})$ the spatial frequency tuning. Arranged in polar tessellations, of geometrically progressing frequencies with octave bandwidths [9], the filters cover densely the spectral domain [12].

3. Texture modeling

Through the multiple filter model, texture is modeled as a composition of simpler, scale and orientation eclectic formations. This set of narrowband signals, have highly concentrated spatial frequency content [6, 14] and account for the periodicity, directionality, spatial extent and texture scale. A filterbank performs a rough decoupling of the components and its design specifies a-priori their number, loca-

tion and frequency extend. For small filter overlaps, components are assumed to be globally isolated and form a multidimensional representation in the response space [12].

Modulation components: Narrowband texture components are modeled in the AM-FM framework [12, 13] by nonstationary amplitude and frequency-modulated sines

$$t_k(x,y) = \alpha_k(x,y)\cos(\vec{\omega}_{k0}\cdot(x,y)+\phi_k(x,y)), \quad (5)$$

where the *amplitude modulating signal* $\alpha_k(x,y)$ accounts for the component spatial extend and local contrast and the *instantaneous frequency vector* $\vec{\omega}_k(x,y) = \nabla\phi(x,y)$ for the local scale and orientation. A component is assumed to dominate pointwise a filter response $(g_k * t)(x,y) \approx (g_k * t_k)(x,y),\ \forall(x,y)\in\Omega$, and under certain approximations, its parameters can be equivalently estimated from that response using image demodulation [12].

A robust demodulation approach was proposed in [13] based on the energy separation algorithm [15] and a regularized version of the image *energy operator* $\Psi(f) \triangleq \|\nabla f\|^2 - f\nabla^2 f$. The amplitude and instantaneous frequencies are estimated by energy ratios

$$\sqrt{\Psi(t_{k,x})/\Psi(t_k)} \approx |\omega_{k1}|,\ \sqrt{\Psi(t_{k,y})/\Psi(t_k)} \approx |\omega_{k2}| \quad (6)$$

$$\Psi(t_k)/\sqrt{\Psi(t_{k,x})+\Psi(t_{k,y})} \approx \alpha_k \quad (7)$$

where $t_{k,x} = \partial t_k/\partial x$ and $t_{k,y} = \partial t_k/\partial y$ are the partial derivatives of component t_k and $\Psi(\cdot)$ is the *regularized energy operator*

$$\Psi(t_k) = \Psi(t*g_k) = \|t_k*\nabla g_k\|^2 - (t_k*g_k)(t_k*\nabla^2 g_k), \quad (8)$$

that bypasses the finite difference approximation of the spatial derivatives through regularization $t_{k,x} = t_k * g_{k,x}$.

Dominant component: Analysis by dominant components (DCA) [12] assumes that a single component dominates locally the image spectrum. The texture dominant component, is a locally narrowband in space, smoothly varying function $d(x,y) = \alpha_{\rm d}(x,y)\exp\{j\phi_{\rm d}(x,y)\}$, derived by reconstructing the image at each $\mathbf{x} = (x,y)$ from a single component (5)

$$d(\mathbf{x}) = \{t_i(\mathbf{x}) : i(\mathbf{x}) = \arg\max_k\{\Gamma_k(\mathbf{x})\}, k\in[1,K]\}. \quad (9)$$

Function d maximizes an energy criterion $\Gamma_k(x,y)$ which is either the channel amplitude envelope $\Gamma_k = |\alpha_k(x,y)|$ [12] or the complex energy operator response $\Gamma_k = \Psi(t * g_k)(x,y)$ [13]. The latter results in a more coherent component, with improved localization in texture and object boundaries, and is better-suited for scale-selective texture separation (Sec. 4.2).

The modulation functions of $d(x,y)$ result from the K-to-1 backward mapping of the estimated channel parameters $(a_k, \nabla\phi_k)(x,y)$ to the dominant

$$\alpha_{\rm d}(x,y) = \alpha_{i(x,y)}(x,y),\ \ \vec{\omega}_{\rm d}(x,y) = \vec{\omega}_{i(x,y)}(x,y). \quad (10)$$

which is theoretically equivalent to directly demodulating the dominant component. DCA reduces the description from the K-dimensional Gabor response space (or the $3\times K$ modulation space) to the low-dimensional $(\alpha_{\rm d}, \nabla\phi_{\rm d})$ vector, while retaining essential texture information like extent, scale and orientation.

4. $u + Kv$: cartoon & texture subcomponents

Consider an arbitrary image function $f : \Omega \to \mathbb{R}$, with Ω an open and bounded subset of $\mathbb{R}^2$, and the model $f = u + v$. Vaguely stated, $u \in \Omega$ is a piecewise smooth, image approximation and $v \in \Omega$ a low amplitude oscillating function. To integrate multifrequency and content decomposition, we formulate the following minimization w.r.t. (u,v) functional

$$E(u,v) = \int_\Omega (\|\nabla u\| + \lambda|f-u-v|^2 + \mu|v-\sum_k^K t_k|^2) \quad (11)$$

where $u \in BV(\Omega)$ and $v \in \mathcal{H}(\Omega)$ where $\mathcal{H}$ some Hilbert space . The additional texture constrain is a proximity term of v to reconstruction by a set of narrowband components t_k of a function $t \in L^2(\Omega)$, given a set of frequency-tuned functions $\{g_k, k \in \{1\cdots K\}\} \in L^2(\Omega)$. We define function t to be a linear $t = \mathbf{T}f$ or non-linear $f \mapsto t$ image mapping. Thus, the extra term requires v to be close, in the L^2 sense, to a multifrequency decomposition of t.

The components (number, localization, tuning, bandwidth) are controlled by the parameters of the g_k functions. Choosing Gabor filters (4), arranged in a filterbank manner and setting $t = f$ ($\mathbf{T} = \mathbf{I}$), then t_k form the multiband representation, *i.e.* the set of narrowband components of f. Each component, the projection of t onto each of the K basis, is given by convolution of t with the filter response $t_k = (t * g_k)(x,y)$.

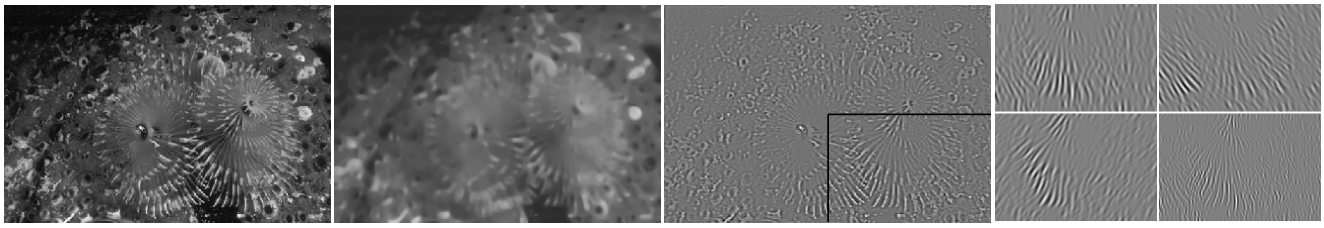
The functional $E(u,v) = \int_\Omega \Phi(u,v,u_x,u_y)dxdy$ is formally minimized by the associated pair of Euler-Lagrange equations. By imposing natural boundary conditions, the gradient-descent flow solution is

$$\frac{\partial u}{\partial t} + u = (f-v) + \frac{1}{\lambda}\mathrm{div}\left(\frac{\nabla u}{|\nabla u|}\right), \quad (12)$$

$$\frac{\partial v}{\partial t} + v = \frac{\lambda}{\mu+\lambda}(f-u) + \frac{\mu}{\mu+\lambda}\sum_{k=1}^K t_k, \quad (13)$$

$$(u,v)(x,y,0) = (f,0), \quad \partial u/\partial\vec{N} = 0,\ u \in \partial\Omega$$

where $\vec{N}$ the outward unit normal on the boundary $\partial\Omega$ and $\kappa(u) = \mathrm{div}\,(\nabla u/|\nabla u|)$ the level curvature of $u = u(x,y)$. Note that the solution of equation (12) is the formal minimizer of the ROF model [19]. The steady-state solutions



(a) Image (b) Cartoon (c) Texture (d) Subcomponents (detail)

Figure 2: Image (a) decomposition in cartoon (b) and texture (c) components, using the $K+1$ equations scheme. Four of the k texture subcomponents are shown in (d) for the lower right image detail

are given by eliminating v from (12):

$$u = f - \sum_{k=1}^{K} t_k + \frac{\mu+\lambda}{2\mu\lambda}\kappa(u), \tag{14}$$

$$v = \frac{1}{\mu+\lambda}\left(\lambda(f-u) + \mu\sum_{k=1}^{K} t_k\right), \tag{15}$$

from which we make the following observations: a) The solution is given by PDE (14) alone, with the second equation giving the texture component as an enhanced, weighted residual. b) As the residual $f - \sum_k t_k$, approximates the lowpass image component $g_0 * f$ where g_0 a Gaussian, the cartoon (14) is approximately $u \approx g_0 * f + \lambda'\kappa(u)$, which reads as the sum of the slow intensity variations and the curvature-based edge regularization. c) Texture equation (15) can be seen as the ROF residual, enhanced by the narrowband sum.

The constant parameters $\lambda, \mu > 0$ control the decomposition and modeling residuals. Large values of λ result in small residuals as $u + v \to f$ for $\lambda \to \infty$ with finer scales included in u. In any case, the residual $w = f - u - v$ is considered highpass noise, excluded from v. In analogy μ penalizes the texture modeling error and for large values $\mu \to \infty$ this error is almost identically zero as $v \to \sum_k t_k$.

Existence, uniqueness, convergence: For problem (11) exists a unique solution and the approximating sequence converges. Existence is proved by convexity and coerciveness of the functional [3]. Strict convexity implies that a solution $(\hat{u}, \hat{v})$ of (11) is unique, except in direction $(u, -u)$, where by inequalities it is verified that $\forall\tau \neq 0, (\hat{u}+\tau\hat{u}, \hat{v}-\tau\hat{u}) = (\hat{u}, \hat{v})$. A minimizing sequence $u^{(n)}$ that solves (12) or (14), converges to the unique solution since $E(u^{(n)}, v^{(n)})$ is non-increasing and convergent in $\mathbb{R}$.

4.1. Multiple texture subcomponents

Generalizing the decomposition problem in (11), we consider a $f = u + \sum_k v_k$ model that assigns to texture a set of K frequency-localized narrowband subcomponents v_k. We seek this $u + Kv$ model of $K+1$, cartoon and texture image components by minimizing the functional

$$E(u, \{v_k\}) = \int_\Omega (\|\nabla u\| + \lambda|f-u-v|^2) + \mu\sum_k^K \int_\Omega |v_k - t_k|^2 \tag{16}$$

where $k \in \{1, K\}$ and $(u, \{v_k\}) = (u, v_1, \cdots, v_K)$ are K+1 unknown variables. Solution by Euler-Lagrange, gives the set of $K+1$ steady state equations

$$u = (f - \sum_{k=1}^{K} v_k) + \frac{1}{2\lambda}\kappa(u) \tag{17}$$

$$v_k = \frac{\lambda}{\mu}(f - u - \sum_{k=1}^{K} v_k) + t_k,\ k \in \{1, K\}. \tag{18}$$

The texture component v is reconstructed by summation of the K individual subcomponent equations (18)

$$\sum_{k=1}^{K} v_k = \frac{1}{\mu - K\lambda}\left(K\lambda(f-u) + \mu\sum_{k=1}^{K} t_k\right). \tag{19}$$

According to this set of equations: (a) The PDE for u is given in closed analytic form by substituting (19) in (17) and depends only on (f, u, t_k). Each individual v_k can be directly estimated from the solution u using reconstruction (19) in (18). (b) Letting $\mu \gg \lambda K$, equation (19) gives $\sum_{k=1}^{K} v_k \approx \sum_{k=1}^{K} t_k$, *i.e.* the reconstruction of v from the respective reconstruction of t, which can also be seen inversely. Suppose in the $u + v$ scheme of (14, 15) the subcomponents are fixed to some of the filterbank responses, *i.e.* $v_k = g_k * v$. Then at iteration step n of a numerical solution, the term $\sum_k t_k$ can be approximated by projection of the most recent texture estimate $v^{(n)}$ onto the specified set of filters. (c) Problem (11) for the $f = u + v$ model is derived as a special case of the general decomposition in $K+1$ components $(u, v_1, ...v_K)$ and $v = \sum_k v_k$.

Theoretical comparison with TV-Gabor: In the model (3), the projection of the residual $v = f - u$ is minimized on specific Gabor wavelets g'_k, tuned to non-texture frequencies [3]. Noting that $\langle f-u, g'_k * (f-u)\rangle = \|(f-u) * (\delta(x,y) - g_k)^{1/2}\|_2^2$, we draw the analogies with the filtering

formulation here by writing

$$\|f-u\|_{\mathcal{H}}=\|f-u\|_2^2-\|(f-u)*g_k\|_2^2=\|v\|_2^2-\|v*g_k\|_2^2 \quad (20)$$

where the Gabor kernel g_k is a-priori chosen. Multiple orientations are accounted for by summing over a set of projections, *i.e.* $||f-u||_{\mathcal{H}} = K||v||_2^2 - \sum_k^K ||v*g_k||_2^2$. In the $u+Kv$ scheme, if we set $t=f$, meaning that $t_k = g_k * f$, and approximate the texture subcomponents by $v_k \approx g_k * v$, then the constraint in (16) becomes

$$\sum_k \|v_k-t_k\|_2^2 \approx \sum_k \|(f-v)*g_k\|_2^2 = \sum_k \|u*g_k\|_2^2. \quad (21)$$

Thus, while TV-Gabor minimizes the L_2-energy difference between v and its projection on the chosen 'texture channels,' the proposed $u+Kv$ minimizes the energy of the cartoon projection on them. One obvious advantage of the latter then is that in a three-part model, where $f-v=u+w$, the residual noise w energy will be minimized on the texture channels along with u, while in TV-Gabor it will in some amount be included in v, especially for white noise.

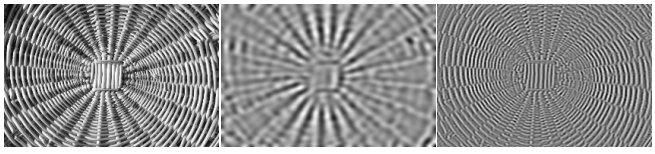
4.2. Texture reconstruction

The sum $\sum_k t_k$ in formulation (11) is, in its crudest form, a reconstruction of t from the set of responses to the filters g_k, with the number of components K defined by the filterbank design. Different model-based methods are proposed here to select a mapping $f \mapsto t$ and a subset of the K components, appropriate to reconstruct texture-specific information, in order to 'guide' the decomposition process and remove redundancy.

Finer scales: Intuitively and experimentally, texture is associated with the middle-higher frequency bands (finer scales), and important edges, corresponding to salient object contours, with the lower bands (coarse scales). This motivates a rough partition of scales in small, texture-important and large, structure-important ones. By rewriting the sum in terms of the R scales and P orientations of the filterbank

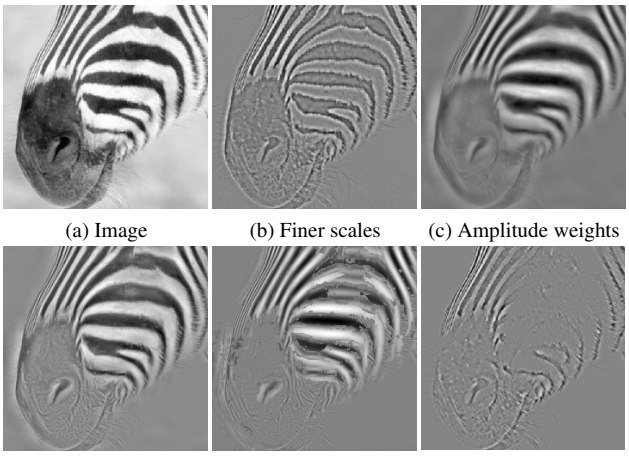
$$\sum_{k=1}^{K} t_k(x,y) = \sum_{\sigma=\sigma_1}^{\sigma_R} \sum_{\theta=\theta_1}^{\theta_P} t_{\sigma,\theta}(x,y),\ k = R(\theta-1)+\sigma, \quad (22)$$

where σ, θ are the scale and orientation indices, the reconstruction is rendered an 'iso-scale' sum of 'iso-orientation'



(a) Image (b) scale 2, ($\sigma = 9$) (c) scale 4, ($\sigma = 3$)

Figure 3: Sum of components over all orientations for a coarse (b) and a finer scale (c).



(a) Image (b) Finer scales (c) Amplitude weights

(d) Lower-order (e)Dom. component (f) v_{VO}

Figure 4: Component selection from a 40-filter bank (5 scales, 8 orientations). Image (a) and reconstruction from (b) finest 3 scales ($K = 24$), (c) amplitude weights ($K = 40$), (d) lower-order dominant components ($D = 9$), (e) dominant component, (f) texture component from Vese-Osher $u+v$ model.

narrowband components. Such reconstruction sums are given for a wideband texture image in Fig. 3.

For Gabor filter channels $g_{\sigma,\theta} = G_\sigma C_{\sigma,\theta}$, each component is $t_{\sigma,\theta} = f * g_{\sigma,\theta}$, with G_σ the Gaussian of standard deviation σ that defines component scale and $C_{\sigma,\theta}$ the complex carrier of angle θ that controls its orientation. In order to primarily reconstruct texture information the lower-frequency components, below a certain scale are omitted

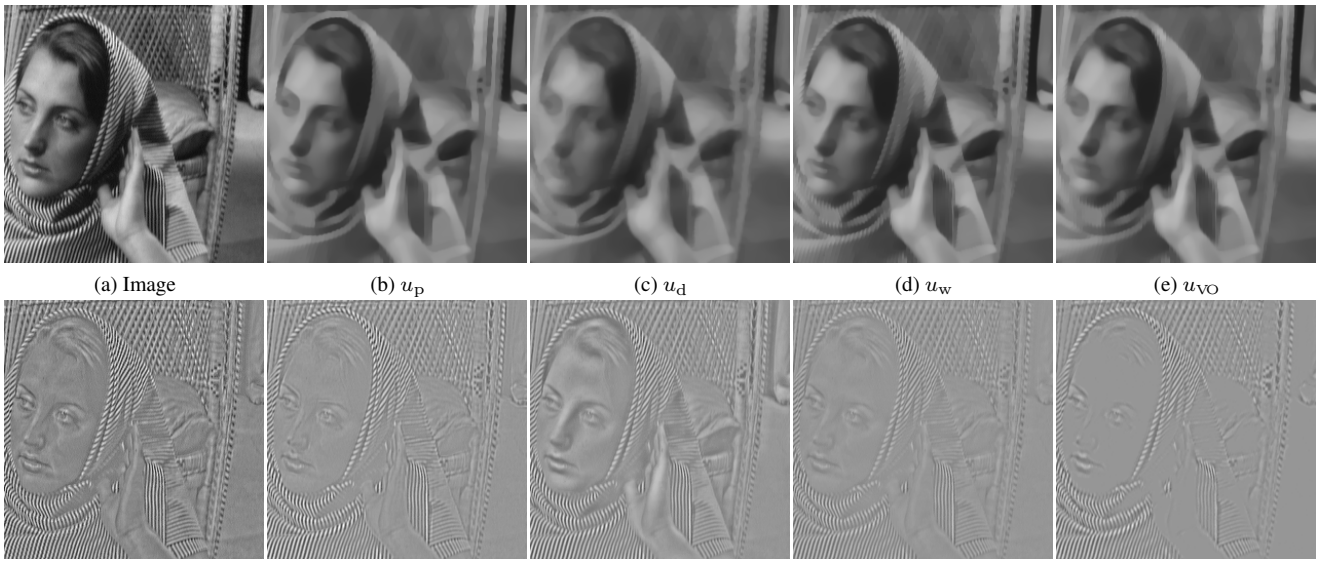
$$\sum_{k=1}^{K} t_k \triangleq \sum_{\sigma=\sigma_T}^{\sigma_R} (G_\sigma * f) \sum_{\theta=\theta_1}^{\theta_P} C_{\sigma,\theta}. \quad (23)$$

This scale range selection can be thought as a scale-based 'component shrinkage.' The choice of the coarsest texture scale, the threshold σ_T, can be done by spatially adaptive scale measures [10, 13]. We restrict the case to choosing globally a fixed range of scales, above a heuristic threshold, as in the reconstruction example of Fig. 4 (b).

Amplitude weights: Narrowband components $f_k = f * g_k$ contain contributions by both the texture and *non-texture* image part, *i.e.* $f_k = u*g_k + v*g_k$. In a weighted reconstruction scenario, each response is weighted by a spatially adaptive measure that enhances locally the contributions of the texture part. A plausible measure, is the component amplitude envelope (7), available through channel demodulation

$$\sum_{k=1}^{K} t_k \triangleq \sum_k \frac{\alpha_k (f * g_k)}{\sum_k \alpha_k}. \quad (24)$$

Thus the reconstruction sum results in a weighted average component. The amplitude $\alpha_k(x,y)$ attains large values in



(a) Image (b) $u_{\rm p}$ (c) $u_{\rm d}$ (d) $u_{\rm w}$ (e) $u_{\rm VO}$

(f) $f - u_{\rm ROF}$ (g) $v_{\rm p}$ (h) $v_{\rm d}$ (i) $v_{\rm w}$ (j) $v_{\rm VO}$

Figure 5: Top: Cartoon components of (a) by the proposed $u + Kv$ with alternative reconstruction sums in (b)-(d) and the Vese-Osher scheme in (e). Bottom: residual of ROF (f) and texture components. $(u_{\rm p}, v_{\rm p})$ denotes the higher-frequency partial sum (23), $(u_{\rm w}, v_{\rm w})$ the amplitude weighted (24) and $(u_{\rm d}, v_{\rm d})$ lower-order dominant reconstruction (25). Parameters $(\lambda, \mu) = (10, 5)$ and $(\lambda_{\rm VO}, \mu_{\rm VO}) = (5, 0.1)$.

high contrast regions that signify large oscillations tuned at the component frequency and orientation, and small in smooth areas. Amplitude weighting then penalizes the contribution of the piecewise-smooth u by $\alpha_k u$. Sharp edge transitions, 'seen' by the model as small-extend oscillations, may also give high amplitude values, however they are removed from v by the regularization process.

As opposed to keeping the finer scales (*spectral selection*), amplitude weighting enhances texture components at each image point, without thresholds. It performs an automatic *energy-based selection*, by suppressing locally the weakly-oscillating components. Further the normalized weights $\alpha_k^2 / \sum_k \alpha_k^2$ approximate the analysis by dominant component (9), for if $\forall (x,y)\ \exists k : \alpha_k(x,y) \gg \alpha_{k'}, \forall k' \neq k$, then t_k is the texture dominant component.

Lower-order dominant components: Using the DCA criterion as a measure of representational importance at each image location, a hierarchical ordering of components is constructed, in which increase in order equals a decrease in dominance. The mapping from the narrowband to the dominant space is $\mathbf{F}_{\rm D} = \mathbf{F} \cdot \mathbf{W}$, where $\mathbf{F} = [f_1, \cdots, f_K]$ and $\mathbf{F}_{\rm D} = [d_1, \cdots, d_K]$ the arrays of filtered and dominant image vectors and $\mathbf{W}$ a transformation matrix. It is defined by the criterion that is maximized across channel responses, with $d_1 = d$ being the dominant component (9). Since $\Gamma_{k+1}(d_{k+1}(\mathbf{x}) < \Gamma_k(d_k(\mathbf{x}), \forall(\mathbf{x})$, the ordering relation among the components arises naturally.

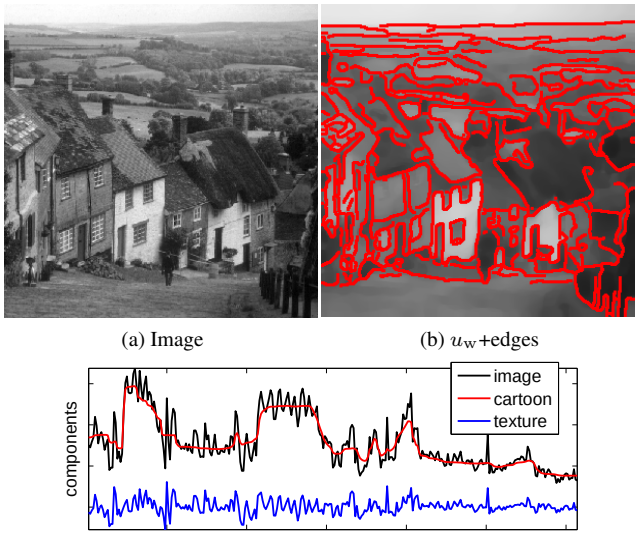
The lower-order components are the most important for representing the information conveyed by the choice of Γ_k. Rejecting the higher-order components of the ordering (*e.g.* by minimal contribution to the total variance)

$$\sum_{k=1}^{K} t_k(x,y) \triangleq \sum_{k=1}^{D} d_k(x,y),\ D : \frac{\mathrm{Var}\{d_k\}}{\mathrm{Var}\{f\}} > \epsilon, \forall k < D, \tag{25}$$

where ϵ a constant near unity (*e.g.* $\epsilon > 0.8$), results to a reduced-dimension representation compared to using the total number of components, simultaneously with a better reconstruction of f compared to DCA that employs only the dominant component d_1 (Fig. 4 (d), (e)). The use of the energy operator as opposed to the amplitude criterion, gives fewer maximally contributing variance components.

4.3. Image examples

Decomposition in Fig. 1 was obtained by the two equation scheme (12, 13) and the weighted reconstruction scenario (24), while in Fig. 2 the $K + 1$ equation scheme was employed. Comparisons of the proposed $u + Kv$, with the various texture reconstruction terms, are presented in Fig. 5 with an implementation of the Vese-Osher (VO) reference scheme (2). Image (a) is characterized by oriented, high frequency texture oscillations, captured in the texture components by all the alternative reconstruction schemes. Perceptually the smoother, yet sharp cartoon with highly-localized edges is obtained by the partial reconstruction scenario (b). The crisper oscillatory texture is given by the weighted scheme (i). Such reconstruction however retains some visible texture elements in (d), the same way DCA maintains some contrast structures in v (h). However, both were derived without heuristics and performed superior in images



(a) Image (b) $u_{\rm w}$+edges

Figure 6: Decomposition (307×307)by $u + Kv$ with amplitude weighted texture reconstruction. (b) Cartoon with edges extracted from u superimposed, (c) 1D component profiles (row 154).

of mixed frequency, strength and orientation textures.

5. Texture classification

Component v is free from albedo discontinuities, intensity transitions, shading and luminance variations and more consistent with subsequent texture modeling. Thus, for content-oriented image analysis, texture features and statistics can be pooled explicitly from v. This will be demonstrated for texture classification, by comparing texture modeling, *i.e.* extracted features, on v and on full image f.

Feature extraction: Five feature vectors, all based on multiband filtering are considered:

a) DCA(4×1)**:** dominant modulation components (10) from v, augmented by the intensity of the cartoon $\vec{V}_{\rm D} = [\alpha_{\rm d}(v),\ \vec{\omega}_{\rm d}(v),\ u]$.

b) Gabor$(K\times1)$**:** quadrature energies of the outputs $\vec{V}_{\rm G} = [\{({\rm Re}\{g_k\} * v)^2 + ({\rm Im}\{g_k\} * v)^2\}_{k=1:K}]$.

c) Teager$(K\times1)$ nonlinear complex operator [13] applied on the filter responses $\vec{V}_{\rm T} = [\{\Psi(g_k * v)\}_{k=1:K}]$.

d) Pooled$((K+4)\times1)$**:** augmented $\vec{V}_{\rm P} = [\vec{V}_{\rm D}, \vec{V}_{\rm G}]$.

e) Pooled/PCA: feature selection (PCA) on pooled vector.

Full-component vectors $\vec{V}_{\rm G}$ and $\vec{V}_{\rm T}$ are K-dimensional energy features by nonlinear operators, while DCA is a dense, low-dimensional description of the texture signal [13].

Experiments & discussion: Four Brodatz-pair images, *i.e.* photographic images of surfaces and materials, from [18] were considered for classification. In a *supervised* scenario, train and test regions are non-overlapping and five classifiers (parametric and non-parametric) are employed: Maximum Likelihood quadratic (Bayes NQ), nearest mean (NM), 1-nearest neighbor (1-NN), Fisher linear discriminant (FLD), mixture of (2) Gaussians (MoG) and support vector classifier (SVC). *Clustering* was also performed by k-means with the number of clusters a-priori defined.

Testing features were extracted both from the full f and the estimated with $u + Kv$ component v. The performance was quantified in terms of the overall classification error (supervised) and overall probability of error (clustering), after minimum-error cluster alignment. Table 1 lists the errors and relative change after decomposition. As the choice of (λ, μ) affect the classification performance, by controling the passing of details between components, the errors on v are the minimum per classifier *and* feature vector, achieved over a quantized range $(l, m) = 10\log_{10}(\lambda, \mu) \in \{-3, 2\}$. Note especially the minimum errors per analyzed component (bold), the average (over method) supervised error and the clustering error. Except for the DCA vector, feature extraction from v improved average performance, reaching a 22.7% error decrease over all methods. For clustering by k-means, the results are consistent across all vectors with the best performance of 4.24% error for the pooled vector and a maximum error decrease of 41% on Gabor features. Although DCA features on v did not improve average supervised performance, they gave *consistently* the best trained classification results in both f and v.

Figure 7, shows the k-means labeling in two of the test images, with the clusters on v on the last column. These are the minimum error (top) and the maximum relative error decrease (bottom) cases. In the difficult D5D92 image, the DCA vector decreases the error when extracted from v. The same experiments were run on the 5-texture patches from [18] (*e.g.* Fig. 8), where improvement by decomposition was similarly observed, reaching for the pooled vector, a 15.9% maximum relative clustering error decrease.

6. Conclusions

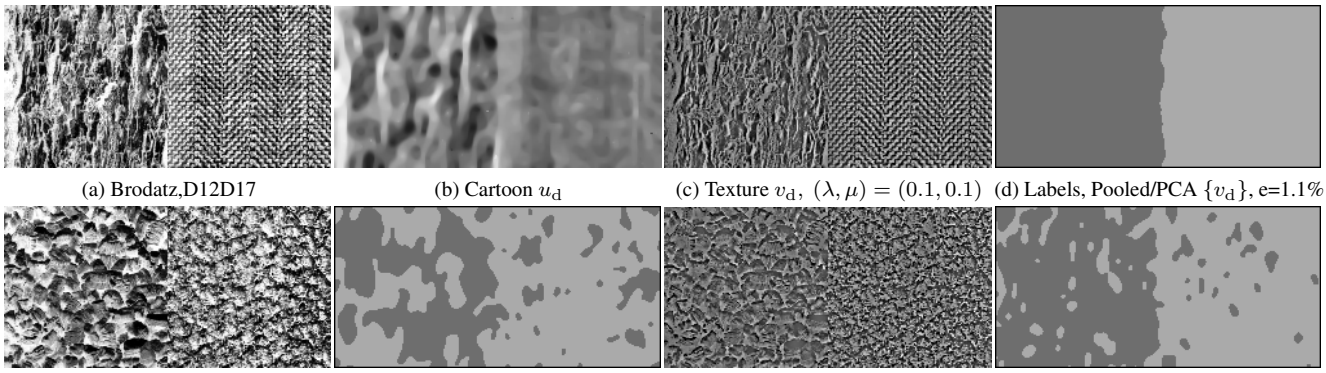
A novel $u + v$ model for image decomposition was proposed using explicit texture reconstruction constraints, in an approach to incorporate modeling priors in the inverse variational estimation problem. In addition, the scheme can further separate texture in narrowband subcomponents, a $u + Kv$ model, and texture scale can be automatically or a-

(a) Image (b) Pooled $\{v_{\rm d}\}$ (c) DCA $\{v_{\rm d}\}$

Figure 8: Clustering on 5-texture image using features from v.

Classifier (supervised/ clustering)	**Feature vectors, analyzed components & relative error change**														
	DCA			Gabor			Teager			Pooled			PCA(Pooled)		
	$\{f\}$	$\{v\}$	D	$\{f\}$	$\{v\}$	D	$\{f\}$	$\{v\}$	D	$\{f\}$	$\{v\}$	D	$\{f\}$	$\{v\}$	D
MoG, SVC, 1-NN	**13.4**	**16.6**	24.0	**29.0**	**25.4**	-12.7	**28.6**	**24.9**	-12.9	23.5	20.3	-13.5	49.5	39.4	-20.3
Bayes NQ	18.2	18.1	-0.37	36.5	27.4	-24.9	37.4	28.3	-24.5	26.9	21.5	-20.0	48.6	41.1	-15.5
NM	17.4	16.9	-3.09	31.9	26.8	-16.1	32.2	27.1	-15.8	**18.8**	**19.6**	4.42	**48.1**	**29.9**	-37.8
FLD	16.4	19.5	18.3	37.1	30.7	-17.4	40.4	31.9	-21.2	30.2	25.7	-15.0	45.5	34.4	-24.3
average	15.1	17.3	15.8	31.7	26.6	-15.6	32.1	26.7	-16.1	24.3	21.2	-12.1	48.6	37.6	-22.7
k-means	18.6	11.8	- 36.6	8.76	5.17	-41.0	9.06	6.18	-31.8	7.11	4.24	- 40.4	7.16	4.34	-39.4

Table 1: Average Classification Errors (%) on 2-texture images (min-error (λ, μ) for each case)



(a) Brodatz,D12D17 (b) Cartoon $u_{\rm d}$ (c) Texture $v_{\rm d}$, $(\lambda, \mu) = (0.1, 0.1)$ (d) Labels, Pooled/PCA $\{v_{\rm d}\}$, e=1.1%

(e) Brodatz, D5D92 (f) Labels, DCA $\{f\}$, e=36.4% (g) Texture $v_{\rm d}$, $(\lambda, \mu) = (100, 0.1)$ (h) Labels, DCA $\{v_{\rm d}\}$, e=19%

Figure 7: Classification (k-means clustering) results in 2-texture images ([18]). Top row corresponds to the minimum-error result, over all parameter values and feature vectors. Bottom row to the maximum relative error decrease results, when using v for feature extraction.

priori defined. Various content-oriented applications can be fertilized by component-wise processing like texture classification from v, or edge detection from u (shown in Fig. 6), while the combined approach can be generalized to other decomposition or texture modeling schemes.

References

[1] J.-F. Aujol and A. Chambolle. Dual norms and image decomposition models. *IJCV*, 63(1):85–104, June 2005.

[2] J.-F. Aujol and T. F. Chan. Combining geometrical and textured information to perform image classification. *JVCIR*, 17:1004–1023, 2006.

[3] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher. Structure-texture image decomposition–modeling, algorithms and parameter selection. *IJCV*, 67(1):111–136, Feb. 2006.

[4] S. Bae, S. Paris, and F. Durand. Two-scale tone management for photographic look. In *ACM SIGGRAPH 06*, 2006.

[5] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher. Simultaneous structure and texture image inpainting. *IEEE IP*, 12(8):882–889, Aug. 2003.

[6] A. C. Bovik, M. Clark, and W. S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE PAMI*, 12(1):55–73, Jan. 1990.

[7] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. S. Huang. Total variation models for variable lighting face recognition. *IEEE PAMI*, 28(9):1519–1524, 2006.

[8] I. Daubechies and G. Teschke. Variational image restoration by means of wavelets. *Applied and Computational Harmonic Analysis*, 19(1):1–16, 2005.

[9] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *JOSA A*, 4(12):2379–2394, Dec. 1987.

[10] G. Gilboa, N. Sochen, and Y. Y. Zeevi. Variational denoising of partly-textured images by spatially varying constraints. *IEEE IP*, 15(8):2281–2289, Aug. 2006.

[11] C. E. Guo, S. C. Zhu, and Y. N. Wu. Towards a mathematical theory of primal sketch and sketchability. In *ICCV*, volume 2, pages 1228–1235, 2003.

[12] J. P. Havlicek, D. S. Harding, and A. C. Bovik. Multidimensional quasi-eigenfunction approximations and multicomponent AM-FM models. *IEEE IP*, 9(2):227–242, Feb. 2000.

[13] I. Kokkinos, G. Evangelopoulos, and P. Maragos. Texture analysis and segmentation using modulation features, generative models and weighted curve evolution. *IEEE PAMI*, 2008. *in press*.

[14] S. G. Mallat. Multifrequency channel decompositions of images and wavelet models. *IEEE ASSP*, 12(37):2091–2110, Dec. 1989.

[15] P. Maragos and A. C. Bovik. Image demodulation using multidimensional energy separation. *JOSA A*, 12(9):1867–1876, 1995.

[16] Y. Meyer. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*. University Lecture Series. AMS, 2001.

[17] J. Portilla and E. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *IJCV*, 40(1):49–70, Oct. 2000.

[18] T. Randen and J. H. Husoy. Filtering for texture classification. *IEEE PAMI*, 21(4):291–310, Apr. 1999.

[19] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, Nov. 1992.

[20] J.-L. Starck, M. Elad, and D. L. Donoho. Image decomposition via the combination of sparse representations and a variational approach. *IEEE IP*, 14(10):1570–1582, Oct. 2005.

[21] L. A. Vese and S. J. Osher. Modeling textures with total variation minimization and oscillating patterns in image processing. *J. Scien. Comp.*, 19(1-3):553–572, Dec. 2003.