TEXTURE MODULATION-CONSTRAINED IMAGE DECOMPOSITION

Georgios Evangelopoulos and Petros Maragos

School of ECE, National Technical University of Athens, Zografou, 15773 Athens, Greece

[gevag,maragos]@cs.ntua.gr

ABSTRACT

Texture modeling and separation of structure in images are treated in synergy. A variational image decomposition scheme is formulated using explicit texture reconstruction constraints from the outputs of linear filters tuned to different spatial frequencies and orientations. Relevant to the texture image part information is reconstructed using modulation modeling and component selection. The general formulation leads to a $u + Kv$ model of $K + 1$ image components, with multiple texture subcomponents.

Index Terms— texture, image decomposition, variational methods, PDEs, modulations, reconstruction

1. INTRODUCTION

Texture in images coexists with geometric macrostructures such as contours, shapes and boundaries, or is embedded in coarser structures formed by lighting and shading conditions or smooth area/volume variations. Image decomposition refers to 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. Decomposition approaches include image diffusions and simplifications [1], wavelet projections [2, 3] and representations in bases [4] or non-linear filtering of the bilateral [5] or morphological, leveling type [6].

In the $f = u + v$ decomposition models, an image f, is treated as the sum of two independent components: a piecewise smooth function u with quasi-flat intensity plateaus and jump discontinuities, the 'cartoon', that contains geometric structure information (edges, contours, large-scale features and illumination effects) and a small-scale oscillatory function v that captures texture and possibly noise (Fig. 1). Inverse methods have been proposed for image decomposition in structure u plus texture v components [7, 4, 8, 6]. The texture component can be used for solving texture-depended problems (classification, surface analysis, shape/orientation from texture), while the structure part for feature detection, segmentation, shape analysis and object recognition. Image decompositions have motivated solutions to classic problems like image restoration, segmentation, matching, classification and compression by adapting algorithms for the two components, in addition to emerging applications like image inpainting [9] or computational photography [5].

Building on the $u + v$ image models, a decomposition scheme is formulated as a constrained total variation minimization problem. An explicit model prior is included on the constructed functional, through a term penalizing dissimilarity between the texture part and a reconstruction from a sum of narrowband image components. Such components reside in multiple frequencies and orientations, and are further modeled by spatial amplitude and frequency modulations of the image function [10, 11, 12]. The texture constraint is thus derived from the responses of Gabor filters and a component-amplitude weighted image reconstruction. The developed scheme is shown to be a special case of a decomposition in $K + 1$ components.

2. BACKGROUND

2.1. Image decomposition

Estimation from an image $f : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ of the components u, v is posed in the variational paradigm [1, 7, 9, 13, 8], as a minimization of general convex functionals of the form

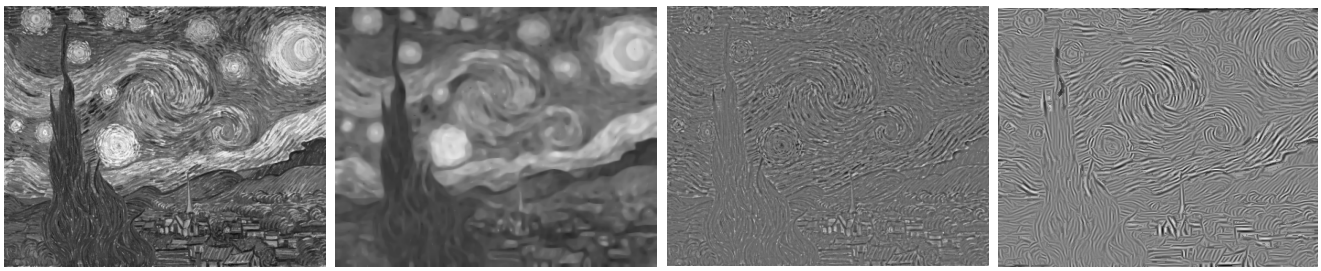
$$\inf_{(u,v)\in(U\times V)} \{E(u,v) = J(u) + \lambda F(u,v) + \mu L(v)\}, \quad (1)$$

where $U, V \subset \Omega$ and $\lambda, \mu \geq 0$ tuning constants. The first, regularizing term is normally the Total Variation (TV) norm $J(u) = \int_\Omega ||\nabla u|| dxdy$. A second, fidelity term $F(u,v) = ||f - u - v||_X^2$ 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, defined by Meyer in a space of oscillating functions $v = \mathrm{div}(\vec{g}) = \partial_1 g_1 + \partial_2 g_2$, where $g_1, g_2 \in L^\infty(\mathbb{R}^2)$ [3], was approximated computationally by Vese and Osher [7] using Sobolev norms L^p, $1 \leq p \leq \infty$. The formed functional

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

This work was supported by the Greek GSRT grants ΠΕΝΕΔ-2001 and ΠΕΝΕΔ-2003-ΕΔ865 and by the European Union FP6-IST research NoE MUSCLE and projects HIWIRE and ASPI.



(a) Image (b) 'Cartoon' (c) Texture (d) Image dominant component

Fig. 1: Painting (a) is a mixture of geometric structures ('cartoon') and textural patterns (*"The Starry Night", Vincent van Gogh, 1889).*

with $(u, v) \in BV \times L^p$, leads to a three-part model where the residual $w = f - u - v$ giving image noise. Aujol and Chambolle [13] used the exact definition of Meyer's texture norm $\|\cdot\|_G$ and the indicator function of the set $\|v\|_G \leq \mu$ as a constraint. A two-part decomposition, selective to texture frequency and orientation was proposed by minimizing the projection of $v = f - u$ onto predefined Gabor wavelets [8].

2.2. Texture modeling

Texture variations are described in the multiple filter model by a set of narrowband signals, of highly concentrated spatial frequency content, that account for texture periodicity, directionality, spatial extent and scale [2]. A wideband image may be decoupled to such components in the output of a Gabor filterbank, covering densely the frequency plane [11].

Narrowband texture components are modeled in the AM-FM framework [11, 12] by nonstationary amplitude and frequency-modulated sinusoids

$$t_k(x, y) = \alpha_k(x, y) \cos(\vec{\omega}_{k0} \cdot (x, y) + \phi_k(x, y)), \quad (3)$$

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. Demodulation of the components in the amplitude and instantaneous frequencies is estimated by energy ratios via the energy separation algorithm [10] and a regularized version of the image *energy operator* $\Psi(t_k) = \Psi(t * g_k) = \|t_k * \nabla g_k\|^2 - (t_k * g_k)(t_k * \nabla^2 g_k)$, where g_k the response of the k-th Gabor filter channel [12].

The texture *dominant component* [11], 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 (3)

$$d(\mathbf{x}) = \{t_i(\mathbf{x}) : i(\mathbf{x}) = \arg\max_k \{\Gamma_k(\mathbf{x})\}, k \in [1, K]\}. \quad (4)$$

Function d maximizes an energy criterion $\Gamma_k(x, y)$ which is either the channel amplitude envelopes $|\alpha_k(x, y)|$ [11] or the complex energy operator response $\Psi(t * g_k)(x, y)$ (Fig. 1(d)).

3. A DECOMPOSITION SCHEME USING TEXTURE MODEL RECONSTRUCTION

Consider a mapping the initial image $f \mapsto t \in L^2(\Omega)$ and its narrowband components t_k from a set of frequency-tuned functions $\{g_k, k \in \{1 \cdots K\}\} \in L^2(\Omega)$. Texture modeling and content decomposition are integrated through the constrained minimization functional

$$E(u, v) = \int_\Omega (\|\nabla u\| + \lambda |f - u - v|^2 + \mu |v - \sum_k^K t_k|^2) \quad (5)$$

where $u \in BV(\Omega)$ and v belongs to to a Hilbert space $v \in \mathcal{H}(\Omega)$. The texture constrain term requires v to be close, in the L^2 sense, to a reconstruction by a set of narrowband components t_k, whose parameters (localization, tuning, bandwidth, maximum number) are controlled by the g_k functions. The problem is then to select the mapping of f and a subset of the K components, appropriate to reconstruct texture-specific information. For example, if $t = f$ and g_k are Gabor functions, t_k are the narrowband components of f, given by $t_k = (f * g_k)(x, y)$.

One option for the reconstruction sum would be to omit the lower-frequency components from the responses to a filterbank g_k, below a certain scale, by considering a partition of scales, in small, texture-important and large, structure-important ones. Unless an efficient, spatially-adaptive scale-selection mechanism is used, this would require a heuristic threshold selection, possibly trained in a large set of images. As opposed to keeping the finer scales (*spectral selection*), we propose an automatic, *energy-based selection* using modulation modeling [14].

Bearing in mind that narrowband image components contain contributions by both the texture and *non-texture* image part, *i.e.* $f_k = u * g_k + v * g_k$, we wish to suppress u in the reconstruction sum. The filtered component amplitude envelope is used for that purpose as a spatially adaptive measurement to weight the responses. Then, the reconstruction sum results in a weighted average component

$$\sum_{k=1}^K t_k \triangleq \sum_k \frac{\alpha_k (f * g_k)}{\sum_k \alpha_k}. \quad (6)$$

(a) Image (b) Reconstruction (c) Dom. component

Fig. 2: Reconstruction by amplitude-weighted component sum ($K = 32$) (b) and energy dominant component (c), from 32 Gabor filters (4 scales, 8 orientations).

The amplitude $\alpha_k(x, y)$ attains large values in highly contrasted regions that signify large oscillations tuned at the component frequency and orientation, and small in smooth areas. An example of that effect can be seen in Fig. 2. TInterestingly, the normalized weights $\alpha_k^2 / \sum_k \alpha_k^2$ approximate the analysis by dominant component (4), 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 [14].

3.1. Solution and PDEs

The gradient-descent flow solution that minimizes the functional (5) $E(u, v) = \int_\Omega \Phi(u, v, u_x, u_y)dxdy$ is given by the pair of Euler-Lagrange equations

$$\frac{\partial u}{\partial t} + u = (f - v) + \frac{1}{2\lambda}\text{div}\left(\frac{\nabla u}{|\nabla u|}\right), \tag{7}$$

$$\frac{\partial v}{\partial t} + v = \frac{\lambda}{\mu + \lambda}(f - u) + \frac{\mu}{\mu + \lambda}\sum_{k=1}^{K} t_k, \tag{8}$$

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

where $\vec{N}$ is the outward unit normal on the boundary $\partial\Omega$ and $\kappa(u) = \text{div}\,(\nabla u/|\nabla u|)$ the level curvature of $u = u(x, y)$. From the steady-state solutions

$$u = f - \sum_{k=1}^{K} t_k + \frac{\mu + \lambda}{2\mu\lambda}\kappa(u), \tag{9}$$

$$v = \frac{1}{\mu + \lambda}\left(\lambda(f - u) + \mu\sum_{k=1}^{K} t_k\right), \tag{10}$$

we observe that a) the texture component is the residual of the ROF model [1] enhanced by the reconstruction sum and b) as the residual $f - \sum_k t_k$, approximates the lowpass image component $g_0 * f$ where g_0 a Gaussian, the cartoon (9) is approximately $u \approx g_0 * f + \lambda'\kappa(u)$, *i.e.* the large-scale intensity variations plus a curvature-based edge regularization.

A fixed point (Gauss-Seidel) iteration numerical scheme is used to solve (9, 10), with a discretization similar to [7] (symmetric finite differences for derivatives with the most recent u^n value used at each point and a discrete curvature $\kappa_n(u^n, u^{n+1})$). At step $n + 1$ the cartoon estimate is

$$u^{(n+1)} = f - \sum_k t_k^{(n+1)} + \nu\kappa^{(n)}(u^{(n)}, u^{(n+1)}) \tag{11}$$

where $\nu = (\mu+\lambda)/2\mu\lambda$, with an analogous $v^{(n+1)}$. Iterations stop when changes in the values of the estimated components are small.

3.2. Extension to a general $u + Kv$ model

Generalizing the decomposition problem (5), we consider a model $f = u + \sum_k v_k$ that associates texture with a set of frequency-localized narrowband subcomponents v_k. This $u + Kv$ model is the minimizer of functional

$$E(u, \{v_k\}) = \int_\Omega (\|\nabla u\| + \lambda|f - u - v|^2) + \mu\sum_k^K \int_\Omega |v_k - t_k|^2 \tag{12}$$

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

$$u = (f - \sum_{k=1}^{K} v_k) + \frac{1}{2\lambda}\kappa(u) \tag{13}$$

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

The texture component v is further reconstructed by summing up the K individual subcomponent equations (14)

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

The solution equations (13, 15) depend only on the reconstruction sum $\sum_k t_k$. Letting $\mu \gg \lambda K$, equation (15) gives $\sum_k v_k \approx \sum_k t_k$, which in a numerical solution (11) can serve as an update of $\sum_k t_k^{n+1}$, *i.e.* by projection at each step $n+1$ of the most recent texture estimate v^n onto the configuration's set of filters g_k.

4. EXAMPLES

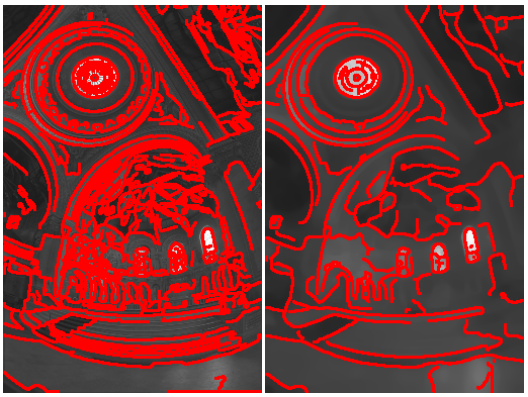
In Fig. 3 we present decomposition by the developed scheme along with an implementation of the reference Vese-Osher (VO) model (2), using empirical parameter tuning. The image in (a) is characterized by oriented, multi-frequency texture areas and large-scale scene contours. The color image was decomposed in a simple channel-by-channel manner. The two cartoons are similar, with clearly visible shapes and smooth intensity variations preserved. However, more small-scale features are 'seen' by $u + Kv$ as texture, compared to VO where these are 'wiped out' and included in the residual. Further, texture oscillations in (c) have a smaller dynamic range, as opposed to (e) where larger peaks are visible.

5. CONCLUSIONS

The developed scheme is an effort in the direction of inserting prior texture model information in $u + v$ models. The scheme

(a) f (b) u (c) v (d) u_{VO} (e) v_{VO}

Fig. 3: Cartoon and texture components of image (a) by the proposed $u + Kv$ with amplitude weighted texture reconstruction in (b),(c), and Vese-Osher decomposition in (d),(e). Parameters $(\lambda, \mu) = (10, 5)$ and $(\lambda_{\mathrm{VO}}, \mu_{\mathrm{VO}}) = (5, 0.1)$. Decomposition was performed in each color channel separately and results are displayed as color images. Texture images are displayed added the mean cartoon value (*Stanford Memorial Church* (high-dynamic-range image) [5]).



(a) edges$\{f\}$ (b) edges$\{u\}$

Fig. 4: Edges extracted from the initial image in Fig. 3(a) and its cartoon component Fig. 3(b).

has been applied to content-oriented component processing, *e.g*. improved edge detection from u (shown in Fig. 4), image restoration (shown for an ancient wallpainting in Fig. 5), texture feature extraction from v and texture classification [14]. Future work can be done towards including other model constraints or focusing on separating more efficiently texture from noise in the residuals. Also, extensions of the model are considered for the treatment of vector-valued (color or complex) images.

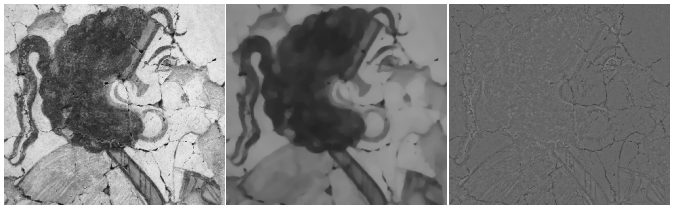


Fig. 5: Decomposition for image restoration (Left original is part of the "Potnia" wallpainting, from *ancient Thera, Acrotiri*).

6. REFERENCES

[1] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, Nov. 1992.

[2] S. G. Mallat, "Multifrequency channel decompositions of images and wavelet models," *IEEE Trans. ASSP*, vol. 12, no. 37, pp. 2091–2110, Dec. 1989.

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

[4] J.-L. Starck, M. Elad, and D. L. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Trans. IP*, vol. 14, no. 10, pp. 1570–1582, Oct. 2005.

[5] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," in *ACM SIGGRAPH*, 2002, pp. 257–266.

[6] P. Maragos and G. Evangelopoulos, "Leveling cartoons, texture energy markers, and image decomposition," in *Int'l Symp. Mathematical Morphology (ISMM)*, Rio de Janeiro, 2007, pp. 125–138.

[7] L. A. Vese and S. J. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing," *J. Scien. Comp.*, vol. 19, no. 1-3, pp. 553–572, Dec. 2003.

[8] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher, "Structure-texture image decomposition–modeling, algorithms and parameter selection," *IJCV*, vol. 67, no. 1, pp. 111–136, Feb. 2006.

[9] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, "Simultaneous structure and texture image inpainting," *IEEE Trans. IP*, vol. 12, no. 8, pp. 882–889, Aug. 2003.

[10] P. Maragos and A. C. Bovik, "Image demodulation using multidimensional energy separation," *JOSA A*, vol. 12, no. 9, pp. 1867–1876, 1995.

[11] J. P. Havlicek, D. S. Harding, and A. C. Bovik, "Multidimensional quasi-eigenfunction approximations and multicomponent AM-FM models," *IEEE Trans. IP*, vol. 9, no. 2, pp. 227–242, Feb. 2000.

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

[13] J.-F. Aujol and A. Chambolle, "Dual norms and image decomposition models," *IJCV*, vol. 63, no. 1, pp. 85–104, June 2005.

[14] G. Evangelopoulos and P. Maragos, "Image decomposition into structure and texture subcomponents with multifrequency modulation constraints," in *CVPR*, Anchorage, AK, 2008.