

Structure-Preserving Image Smoothing via Region Covariances

Levent Karacan*

Erkut Erdem†

Aykut Erdem‡

Department of Computer Engineering
Hacettepe University



Input image

Structure component

Texture component

Figure 1: Our approach makes use of region covariances in decomposing an image into coarse and fine components. The coarse component correspond to prominent structures beneath the image, whereas the fine component only includes texture. Our smoothing method successfully captures the grain of the figures and the rocky texture while preserving the edges in the extracted structure (source image © reibai).

Abstract

Recent years have witnessed the emergence of new image smoothing techniques which have provided new insights and raised new questions about the nature of this well-studied problem. Specifically, these models separate a given image into its structure and texture layers by utilizing non-gradient based definitions for edges or special measures that distinguish edges from oscillations. In this study, we propose an alternative yet simple image smoothing approach which depends on covariance matrices of simple image features, aka the region covariances. The use of second order statistics as a patch descriptor allows us to implicitly capture local structure and texture information and makes our approach particularly effective for structure extraction from texture. Our experimental results have shown that the proposed approach leads to better image decompositions as compared to the state-of-the-art methods and preserves prominent edges and shading well. Moreover, we also demonstrate the applicability of our approach on some image editing and manipulation tasks such as image abstraction, texture and detail enhancement, image composition, inverse halftoning and seam carving.

Links: [DL](#) [PDF](#) [WEB](#) [VIDEO](#) [CODE](#)

1 Introduction

Natural images provide rich visual information about the world we live in, and typically contain various objects organized in a meaningful configuration. For instance, consider the image given in Figure 1, which shows a historical site consisting of highly textured figures on a rocky surface. While our visual system is very successful at extracting the prominent structures beneath the image without getting distracted by the texture, enabling a machine to perform the same task, i.e. decomposing the image into its structure and texture components, poses great challenges.

From a computational point of view, image decomposition can be expressed as an estimation problem in which a given image is separated into two components that respectively correspond to coarse and fine scale image details. Historically, Gaussian filter is the oldest and the most commonly used smoothing operator [Witkin 1984; Burt and Adelson 1983]. It provides a linear scale-space representation of an image where the input image is smoothed at a con-

CR Categories: I.4.3 [Image Processing and Computer Vision]: Enhancement—Smoothing

Keywords: image smoothing, structure extraction, texture elimination, region covariances

*e-mail: karacan@cs.hacettepe.edu.tr

†e-mail: erkut@cs.hacettepe.edu.tr

‡e-mail: aykut@cs.hacettepe.edu.tr

stant rate in all directions. Nonlinear scale-space operators extend linear operators by creating a scale space representation of images that consists of gradually simplified images where some image features such as edges are preserved [Perona and Malik 1990; Rudin et al. 1992; Tomasi and Manduchi 1998; Durand and Dorsey 2002; Buades et al. 2005; Farbman et al. 2008; Xu et al. 2011]. Each of these operators integrates a priori edge information in the smoothing process in a different way, with the aim of extracting or removing certain image details.

Edge-preserving smoothing approaches, namely Anisotropic diffusion filter [Perona and Malik 1990], Total Variation model [Rudin et al. 1992], Bilateral filter [Tomasi and Manduchi 1998; Durand and Dorsey 2002], NL-means filter [Buades et al. 2005], WLS filter [Farbman et al. 2008], L_0 smoothing [Xu et al. 2011] commonly employ differences in the brightness values or gradient magnitudes as the main cues for edge indicator at an image pixel, and make use of this information to guide the smoothing process. These local contrast-based definition of edges, however, might fail to capture high-frequency components that are related to fine image details or textures. Therefore these approaches can not fully separate textured regions from the main structures as they consider them as part of the structure to be retained during computations.

In this paper, we present a novel structure-preserving image smoothing approach which jointly eliminates texture. In the literature, only a few studies tackle this challenging problem of structure extraction from texture [Meyer 2001; Subr et al. 2009; Farbman et al. 2010; Buades et al. 2010; Xu et al. 2012]. In contrast to these previous models, our approach performs a patch-based analysis which depends on second order feature statistics. Specifically, we consider the region covariance matrices [Tuzel et al. 2006] of simple image features such as intensity, color and orientation to estimate the similarity between two image patches within a simple adaptive filtering framework. As demonstrated in Figure 1, the proposed model can effectively eliminate texture without distorting structure. Extracting structure from texture greatly improves the results of many image editing tools. Throughout the paper, we will also demonstrate several image editing and manipulation applications including image abstraction, texture and detail enhancement, image composition, inverse halftoning and seam carving.

2 Background

2.1 Previous work

In the literature, the traditional strategy for structure-preserving image decomposition is to perform joint image smoothing and edge detection [Perona and Malik 1990; Rudin et al. 1992; Tomasi and Manduchi 1998; Farbman et al. 2008; Xu et al. 2011]. These unified formulations simply decompose a given image into structure and detail by smoothing the image, simultaneously preserving or even enhancing image edges, and they differ from each other in how they define edges and how this prior information guides smoothing.

Anisotropic diffusion model [Perona and Malik 1990] employs a PDE-based formulation in which pixel-wise spatially-varying diffusivities are estimated from image gradients. These diffusivities prevent smoothing at image edges, and preserve important image structures while eliminating noise and fine details.

Bilateral filtering [Tomasi and Manduchi 1998; Durand and Dorsey 2002] is another widely used model for removing noise from images while simultaneously performing detail flattening and edge preservation. Due to its simplicity and effectiveness, bilateral filtering has been successfully applied to several computational photography applications [Fattal et al. 2007; Winnemöller et al. 2006].

However, as pointed out by Farbman et al. [2008], multi-scale image decompositions by bilateral filtering have some halo artifacts due to ongoing coarsening process. Weighted Least Square (WLS) filtering [Farbman et al. 2008] overcomes some of these problems by controlling the level of smoothing by forcing the filtered image to be smooth everywhere except at regions having large gradient values.

In a more recent work, Xu et al. [2011] introduced a robust filtering method which uses a sparse gradient measure. The optimization framework considers the number of image pixels having non-zero gradient magnitudes as a regularization constraint which can be linked to L_0 -norm. The proposed filter consequently removes image details with small gradient magnitudes while preserving and even enhancing the most salient edges in images.

The aforementioned studies all depend on gradient magnitudes or brightness differences. Therefore, the ongoing smoothing processes generally can not distinguish fine image details or textures from main image structures, and thus result in unsatisfactory image decompositions. Some studies attempted to improve this separation by using better similarity metrics, which are based on geodesic [Crimini et al. 2010] or diffusion [Farbman et al. 2010] distances instead of traditional Euclidean distances used in color or intensity comparisons.

Another edge-preserving regularization framework is the total variation (TV) model [Rudin et al. 1992], which uses L_1 -norm based regularization constraints to penalize large gradient magnitudes. In its original formulation, this model provides fairly good separations for structure from texture. Some studies extended the standard TV formulation with different norms for both regularization and data fidelity terms, and demonstrated that more robust norms could improve image decompositions [Aujol et al. 2006; Meyer 2001]. In [Buades et al. 2010], the authors proposed a relatively simple technique to decompose an image into structure and oscillatory texture components by using a nonlinear low pass-high pass filter pair. It is used to compute a local total variation of the image around a pixel and subsequently perform the decomposition. More recently, Xu et al. [2012] proposed another robust method with relative total variation measures and showed that better separations can be achieved with spatially-varying total variation measures.

An alternative model for multi-scale image decompositions has been proposed by Subr et al. [2009], which particularly aims at eliminating the oscillatory components of images that generally correspond to texture and extracting the layer containing salient structures. The suggested framework carries out this separation via an extrema analysis in which first minimal and maximal extremal envelopes are estimated from extrapolation of local minima and local maxima and then smoothing result is computed by the average of the extremal envelopes. Since this framework considers the oscillatory behavior of textures in its computations, it gives better results compared to most of the edge-preserving image decomposition models. However, in practice, this averaging idea might fail when the analysis is carried on image regions containing both texture and meaningful structures.

In summary, most of the existing image decomposition models aim at extracting structure from noise with edge-preserving capabilities. However, only a few of them has a specific goal of extracting structure from texture. The formulation that we present in this paper differs from these related works in that it is built upon a patch-based framework and employs similarity measures based on the region covariance descriptor. In Section 4, we provide result comparison and discussion for these methods.

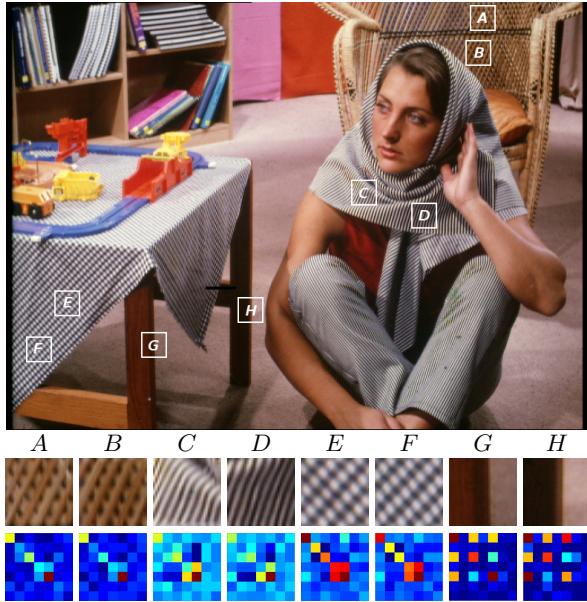


Figure 2: Region covariance descriptors for different regions of the publicly available Barbara image. Regions having similar visual characteristics are represented by similar covariance descriptors. In this example, the covariance representations are based on very simple image features, namely intensity, orientation, and pixel coordinates (Equation 6).

2.2 Region covariances

Expressing an image region by the covariance of features extracted from the pixels within it, known as the *region covariance* descriptor, was first proposed in [Tuzel et al. 2006]. In mathematical terms, let F denote the feature image extracted from an image I :

$$F(x, y) = \phi(I, x, y) \quad (1)$$

where ϕ defines a mapping function that extracts an d -dimensional feature vector (such as constructed from intensity, color, orientation, pixel coordinates, etc.) from each pixel $i \in I$. Then, a region R inside F can be represented with a $d \times d$ covariance matrix \mathbf{C}_R of the feature points:

$$\mathbf{C}_R = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{z}_i - \mu)(\mathbf{z}_i - \mu)^T \quad (2)$$

with $\mathbf{z}_{i=1\dots n}$ denoting the d -dimensional feature vectors inside R and μ being the mean of these feature vectors.

A covariance matrix provides a compact and natural way of fusing different visual features with its diagonal elements representing the feature variances and its non-diagonal elements representing the correlations among the features. Moreover, it provides surprising strong discriminative power in distinguishing local image structures. As illustrated in Figure 2, regions with similar texture and local structures are described by similar covariance matrices. Motivated by these properties, in this study we employ the region covariance descriptor to measure the similarity between two pixels with respect to local image geometry.

Despite these advantages, comparing two image regions with respect to their covariance descriptors tends to be computationally time consuming as covariance matrices do not live on an Euclidean space, but rather on a Riemannian manifold, and requires

non-trivial similarity measures [Tuzel et al. 2006; Cherian et al. 2011]. An interesting take on this issue was offered by Hong et al. [2009] where the authors make use of the property that every covariance matrix (symmetric positive semi-definite matrix) has a unique Cholesky decomposition and use it to transform covariance matrices into an Euclidean vector space.

More formally, let \mathbf{C} be a $d \times d$ covariance matrix, a unique set of points $\mathcal{S} = \{\mathbf{s}_i\}$, referred to as Sigma Points, can be computed as:

$$\mathbf{s}_i = \begin{cases} \alpha \sqrt{d} \mathbf{L}_i & \text{if } 1 \leq i \leq d \\ -\alpha \sqrt{d} \mathbf{L}_i & \text{if } d+1 \leq i \leq 2d \end{cases} \quad (3)$$

where \mathbf{L}_i is the i th column of the lower triangular matrix \mathbf{L} obtained with the Cholesky decomposition $\mathbf{C} = \mathbf{LL}^T$ and α is a scalar¹. Here, it is important to note that the set of columns of \mathbf{L} has the same second order statistics as the original covariance matrix \mathbf{C} [Hong et al. 2009].

3 Approach

Many natural textures lie in between the two extremes of regular and stochastic textures as they contain regular periodic structures as well as additional irregular stochastic components. Here we adopt a general definition of texture as being any visual pattern which has a distinct appearance and local statistics [Efros and Leung 1999]. In this regard, the region covariance descriptor [Tuzel et al. 2006] is a perfect candidate to represent texture information as covariance matrices effectively encode local geometry via second-order statistical relations among features. However, it should be noted that this descriptor has a minor drawback that it falls short on explaining differences in means. Therefore, in this paper, we investigated two different models which incorporate both first and second order statistics to come up with a simple yet effective filtering framework for extracting structure from different types of texture.

Our aim is to decompose a given image I into its structural (S) and textural (T) parts, that is:

$$I = S + T \quad (4)$$

In this study, we follow a patch-based approach, much like the NL-Means [Buades et al. 2005] method, and compute the structure component of a pixel \mathbf{p} as:

$$S(\mathbf{p}) = \frac{1}{Z_p} \sum_{\mathbf{q} \in N(\mathbf{p}, r)} w_{pq} I(\mathbf{q}) \quad (5)$$

where $N(\mathbf{p}, r)$ denotes a squared neighborhood centered at \mathbf{p} and of size $(2r+1) \times (2r+1)$ pixels, and the weight w_{pq} measures the similarity between two pixels \mathbf{p} and \mathbf{q} based on the similarity between $k \times k$ patches centered on these pixels, and $Z_p = \sum_q w_{pq}$ is a normalization factor.

The key to our adaptive filtering framework relies on how we define w_{pq} . In contrast to the NL-Means method, which computes w_{pq} based on Gaussian-weighted Euclidean distance between the patches, here we propose two alternative schemes based on the *region covariance* [Tuzel et al. 2006] descriptor, which make use of first and second-order statistics to encode local structure as well as texture information.

The proposed framework is quite general and does not depend on specific features. In our implementation, we use simple visual features, namely intensity, orientation and pixel coordinates so that an

¹In the experiments, we take $\alpha = \sqrt{2}$.

image pixel is represented with a 7-dimensional feature vector:

$$F(x, y) = \left[I(x, y) \quad \left| \frac{\partial I}{\partial x} \right| \quad \left| \frac{\partial I}{\partial y} \right| \quad \left| \frac{\partial^2 I}{\partial x^2} \right| \quad \left| \frac{\partial^2 I}{\partial y^2} \right| \quad x \quad y \right]^T \quad (6)$$

where I denotes the intensity of the pixel, $\left| \frac{\partial I}{\partial x} \right|, \left| \frac{\partial I}{\partial y} \right|, \left| \frac{\partial^2 I}{\partial x^2} \right|, \left| \frac{\partial^2 I}{\partial y^2} \right|$ are first and second derivatives of the intensity in both x and y directions, estimated via the filters $[-1 \ 0 \ 1]$ and $[-1 \ 2 \ -1]$, and (x, y) denotes the pixel location. Hence, the covariance descriptor of an image patch is computed as a 7×7 matrix. Including (x, y) into the feature set is important since it allows us to encode the correlation of other features with the spatial coordinates. The feature set can be extended to include other features, like for example rotationally invariant forms of the derivatives, if desired.

In the experiments, we handle color images by computing the patch similarity weights w_{pq} using the intensity information and taking the weighted average over the corresponding RGB vectors rather than the intensity values in Equation 5. We empirically found that including RGB components to the feature set does not change the results much but increases the running times.

Model 1

Using the set S defined by Equation (3), a vectorial representation of a covariance matrix can be obtained by simply concatenating the elements of S . Moreover, first-order statistics can be easily incorporated to this representation scheme by including the mean vector of the features μ . This enriched feature vector denote by $\Psi(\mathbf{C})$ is defined as:

$$\Psi(\mathbf{C}) = (\mu, \mathbf{s}_1, \dots, \mathbf{s}_d, \mathbf{s}_{d+1}, \dots, \mathbf{s}_{2d})^T \quad (7)$$

Then, we simply define the weight w_{pq} in Equation (5) as:

$$w_{pq} \propto \exp \left(-\frac{\|\Psi(\mathbf{C}_p) - \Psi(\mathbf{C}_q)\|^2}{2\sigma^2} \right) \quad (8)$$

with \mathbf{C}_p and \mathbf{C}_q denoting the covariance descriptors extracted from the patches centered at pixels p and q , respectively.

Model 2

As an alternative way to measure the similarity between two image pixels with respect to first and second-order feature statistics, we came up with a distance measure, which can be seen as an approximation of the Mahalanobis distance between two Normal distributions. More specifically, for two image pixels p and q , the corresponding distance measure is defined as:

$$d(p, q) = \sqrt{(\mu_p - \mu_q)^T \mathbf{C}^{-1} (\mu_p - \mu_q)} \quad (9)$$

with $\mathbf{C} = \mathbf{C}_p + \mathbf{C}_q$, μ_p and \mathbf{C}_p , μ_q and \mathbf{C}_q denoting the means and covariances of features extracted from the image patches centered at pixels p and q .

Based on this measure, the adaptive weight for the computations in Equation can be alternatively defined as follows:

$$w_{pq} \propto \exp \left(-\frac{d(p, q)^2}{2\sigma^2} \right) \quad (10)$$

A naive implementation of our structure preserving image smoothing algorithm is summarized in Algorithm 1. Our code is publicly available in the project website.

Algorithm 1 Structure preserving image smoothing

Input: image I , scale parameter k , smoothing parameter σ

- 1: extract visual features F via Eq. 6
- 2: **for each:** image pixel p **do**
- 3: compute first and second order region statistics, μ_p and \mathbf{C}_p
- 4: **end for**
- 5: **for each:** image pixel p **do**
- 6: **for each:** neighboring image pixel q **do**
- 7: compute weight w_{pq} using either Eq. 8 (Model 1) or Eq. 10 (Model 2)
- 8: **end for**
- 9: estimate structure component $S(p)$ using Eq. 5
- 10: **end for**

Output: structure image S

Figure 3 shows sample structure-texture decompositions obtained with our smoothing models (Model 1 and Model 2). The input image contains various textured regions with different characteristics, such as the cloth spread over the table, the pants and the scarf of the girl. It may be seen that both of the proposed models successfully separated texture from structure, with Model 2 slightly better than Model 1. Interestingly, the similarity measures defined in Equations 8 and 10 are so effective that they can differentiate local structures from the texture, without employing an explicit edge or texture definition. Moreover, as we will analyze in Section 4, one key difference of our approach is that both of our smoothing models also preserve shading information.

Effects of parameters

Both of our models have two main parameters, k and σ . The spatial parameter σ controls the level of smoothing as it implicitly determines the size of the neighbourhood window. For small values of σ , we have limited smoothing whereas increasing the value of σ causes blurriness. On the other hand, the parameter k controls the size of the patches from which the feature statistics are calculated and accordingly the local structure information to be captured. Hence, its value should be set by taking into account the scale of the texture elements. In that respect, it is more important for structure-texture separation than the spatial parameter σ . As demonstrated in Figure 4, with a proper value of k , the structure component of a mosaic image can be accurately separated from texture. Increasing the patch size too much might cause inaccurate information to be extracted from patches as it may blend texture and meaningful structures, leading to structures to be perceived as fine details. In all the experiments, we empirically set the neighborhood size to 21×21 pixels (See the supplementary material for an analysis on the effect of varying the size of the neighborhood window).

Multi-scale decomposition

While smoothing a given image I , our approach separates it into a structure component S and a texture component T . We use this process iteratively to obtain a multi-scale decomposition of an input image with each layer capturing different fine details of I . Specifically, we smooth the input image by increasing the patch size k (by increasing the scale of analysis) at each iteration and by using the extracted structure component at an iteration as an input for the smoothing process at the subsequent iteration: After n iterations, this yields the decomposition:

$$I(\mathbf{p}) = \sum_{i=0}^n T_i(\mathbf{p}) + S_n(\mathbf{p}) \quad (11)$$

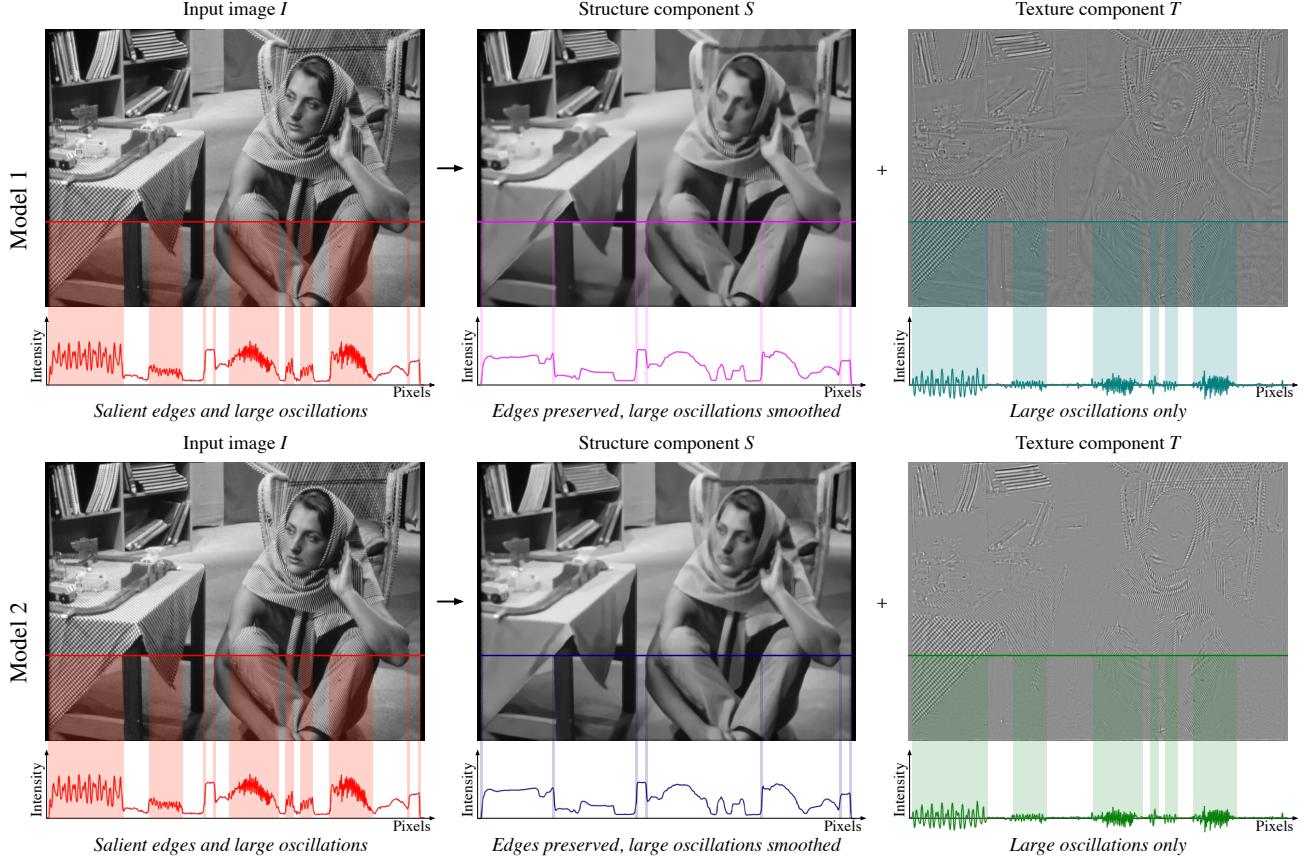


Figure 3: Top row: Model 1, Bottom row: Model 2. Textures of tablecloth, the pants and the scarf are extracted in an accurate way. During the smoothing process, Model 2 preserves image edges slightly better than Model 1 while eliminating texture.

Here, T_1, T_2, \dots, T_n represent the extracted texture components at increasing scales of coarseness and S_n is the coarsest structure component. Figure 5 illustrates the multi-scale representation of two Van Gogh paintings obtained by the proposed Model 1 (See the supplementary material for the results of Model 2). By progressively smoothing the image, some fine details are smoothed out from the original paintings at each layer. The coarsest layer representations do not contain any brush strokes and we obtain smoother versions of the original paintings.

Effects of noise

In Figure 6, we present the smoothing results of a sample image and its noisy counterpart. As can be observed, both of our models successfully extracted the structure even under noise. The decomposed structures from the clean and the noisy images are visually quite similar. This phenomenon can be credited to our patch-based similarity measures in Equation 8 and 10 being robust to noise.

Computational costs

Computationally, the most time-consuming part of our approach is the estimation of the adaptive weights which involves extracting region covariance descriptors and computing the distances between them. We note that feature covariance matrices of arbitrary rectangular regions can be computed efficiently in $\mathcal{O}(d^2)$ time by making use of summed area tables, also known as integral images [Tuzel et al. 2006].

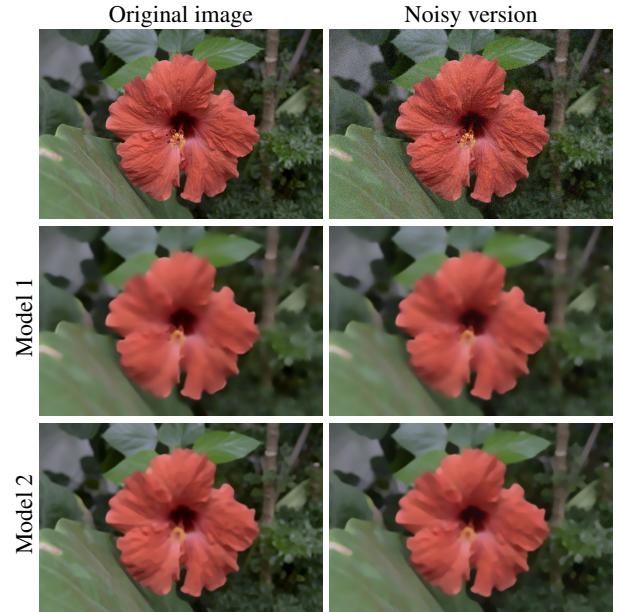


Figure 6: Smoothing results produced by the proposed models on a clean image and its noisy counterpart. Both of our models effectively recovers the structure information in the presence of noise.

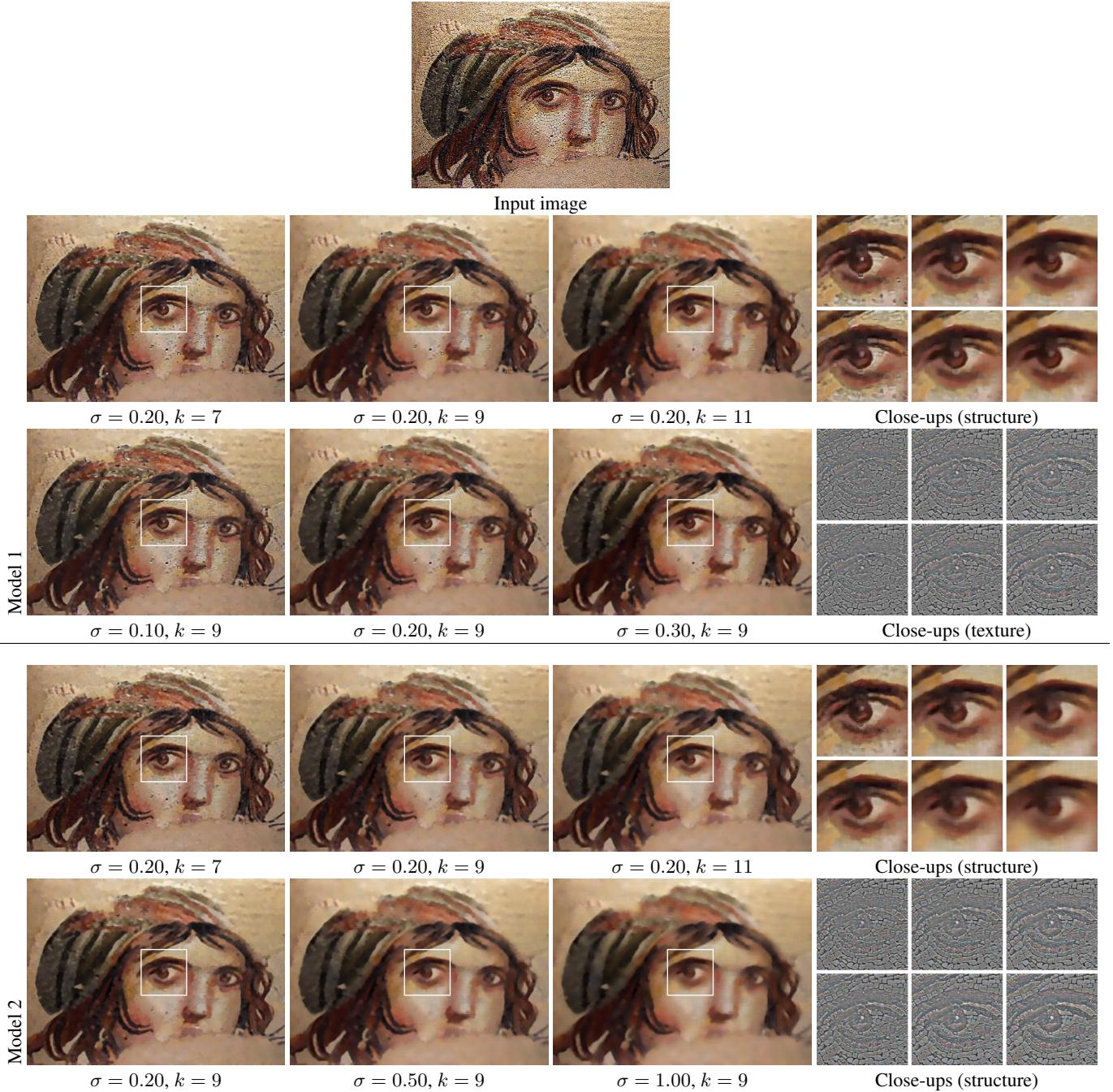


Figure 4: Effects of model parameters. The first rows show the effect of varying k and the second rows illustrate the effect of altering σ . The value of k determines which texture elements are smoothed out and thus it is more vital in structure decomposition from texture as compared to the spatial parameter σ . Setting k to a small value cannot remove texture components well and increasing the value of σ blurs edges.

Our current Matlab implementation is not heavily optimized and was executed on a 3.40 GHz Intel Core i7-2600 processor with 8 GB RAM. A single scale decomposition of a 313x296 color image takes 72 and 101 secs with our Model 1 and Model 2, respectively. The computational cost of Model 2 is somewhat higher than that of Model 1 simply because of the matrix inversion operation in Equation 9. We believe that the running time performances of our models can be greatly improved by a parallel GPU implementation or by using some clever sampling and/or hashing strategies [Baek and Jacobs 2010; Dowson and Salvado 2011].

4 Comparison

In our experiments, we compare our approach with some state-of-the-art edge preserving smoothing methods [Rudin et al. 1992; Tomasi and Manduchi 1998; Farbman et al. 2008; Subr et al. 2009; Buades et al. 2010; Xu et al. 2011; Xu et al. 2012]. Evaluation of the models is carried out qualitatively on the basis that a good method should only smooth fine details and textures and preserve structure, and the extracted texture or so-called detail component should be devoid of any information regarding the structure.

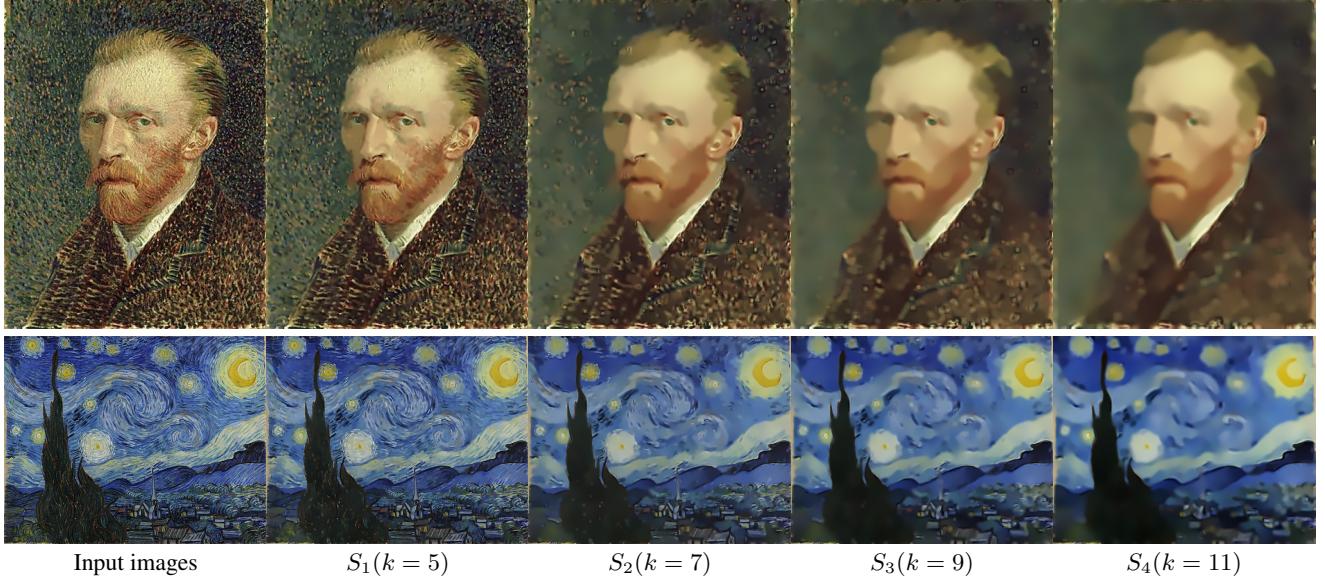


Figure 5: By increasing the value of k and progressively smoothing the images, we extract coarser representations of the original paintings. Each smoothing operation makes different fine details to be removed from the images and final-level smoothing fully eliminates the brush strokes exist in the paintings (source images © Wikimedia Commons).

In Figure 7, we provide the smoothing results of the Barbara image introduced in Figure 2, together with the corresponding detail parts for an image region containing different textures. For all the tested methods, we fine tuned their parameters. As expected, the models which are especially suited for texture smoothing, namely [Subr et al. 2009; Buades et al. 2010; Xu et al. 2012], have given satisfactory results. However, as compared to our approaches, the methods of [Subr et al. 2009; Xu et al. 2012] seem to degrade the main structures during smoothing as some structures are clearly noticeable in the detail components. The model of [Buades et al. 2010] captures the fine details and texture components relatively well but its structure component still includes some textures (especially on the table cloth and the straw chair). As for our models, Model 2 yields better structure-texture decomposition than Model 1. As can be seen from the close-up views given in Figure 8, both of our models preserve shading, capture the texture boundary extremely well, and do not suffer from any staircase-like edges on the table cloth. Moreover, most of the other models produce some spurious colors in their texture/detail layer. The color bleeding in this layer indicates that the corresponding smoothing approach is unable to preserve structure well and constantly deteriorates the color information. Another comparison is provided in Figure 9 for a mosaic image.

5 Applications

Many image editing and manipulation tasks could benefit from a well operated structure extraction from texture. In the following, we will demonstrate such applications which will highlight the effectiveness of our approach.

5.1 Image abstraction

Although textures enrich our visual world and make objects more rich and realistic, they are commonly considered as non-sketchable part of images [Marr 1982]. Hence, eliminating texture without degrading structure could be helpful to simplify an image and give a non-realistic look. In Figure 10, we present some image abstrac-



Figure 9: Smoothing results on the Gypsy girl mosaic image.

tion results which were obtained with the image abstraction scheme proposed by [Winnemöller et al. 2006] in which we replaced the bilateral filtering with our structure-preserving smoothing approach. In these examples, employing Model 1 resulted in a more cartoon-like results as compared with Model 2.



Figure 7: Structure-texture decomposition results on the Barbara image.

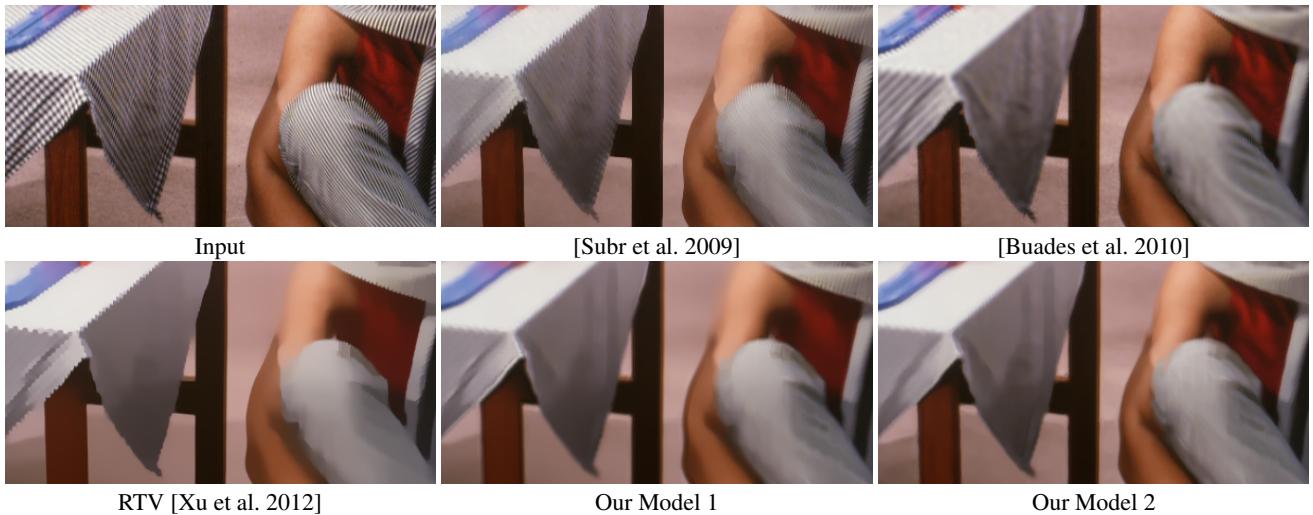


Figure 8: Some close-up views of the extracted structures from the Barbara image. Both of our models extracts the structure from the texture stunningly well while preserving shading information and without introducing any unintuitive edges.



Figure 11: Texture and detail enhancement results (Left: Model 1, Right: Model 2, input image by Stephen Wolf from the Kodak PhotoCD, Photo Sampler).

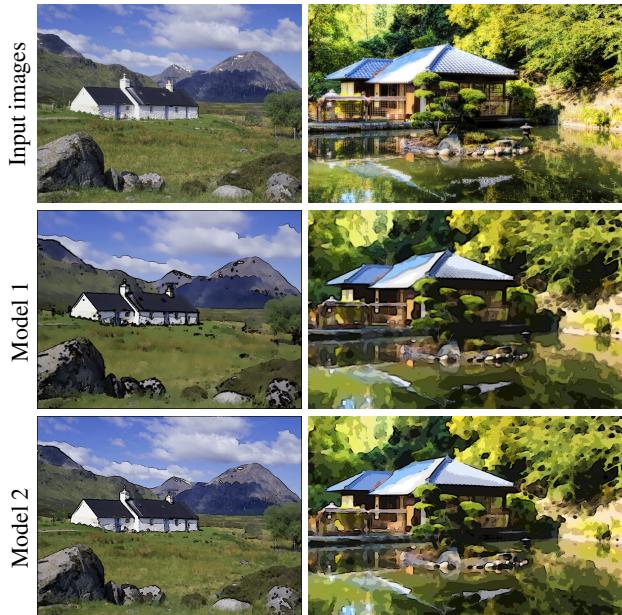


Figure 10: Image abstraction results. Our models effectively removes texture from structure and the extracted structure can be used to obtain a non-photorealistic rendering of the scene.

5.2 Texture and detail enhancement

We can decompose an image into several layers by considering structure and texture at different scales. This decomposition enables us to enhance and boost its texture and fine details or coarse details, as shown in Figure 11. The input image contains a living statue and we enhanced or decreased the details on the facial skin and the wrinkled clothing. For all our models, boosting the fine features effectively enhanced the details in the image without blurring the edges as they successfully capture the low-frequency structure component from the image. As a result, new images have high contrast and detailed textures, so texture enhancement makes input image more appealing with different scales textures obtained with our smoothing method. Similarly, when we boost the coarse features, the textures and fine details are eliminated accordingly and only the prominent structures remain visible in the resulting images.



Figure 12: Image composition results. Top row: An input image and its extracted structure component. Bottom row: Composed images obtained with the original and the structure images (source image © Sew Technicolor, destination image © designshard).

5.3 Image composition

Images of paintings, mosaics, textiles contain different types of texture. Composing such an image directly into another one may give visually unconvincing results since the texture information is transferred along with structure information. Our structure-texture decomposition approach can help obtain visually more plausible results by performing image composition on the structure layer of one of the input images. In Figure 12, we present such an example where one of the inputs is a paper image and the other one consists of a knitted pattern. As can be seen from the figure, when the composition is carried out by considering the structure layer of the pattern image (Model 1 result), the composed image has a more natural look since the details in the texture layer are left out.



Figure 13: Inverse halftoning results. A crop from Iron Man (input image © Marvel Comics).

5.4 Inverse halftoning

We performed our image smoothing models on scanned color comics to reproduce continuous tone shades from the halftones. Figure 13 presents some results in which we compared our results to a recently proposed inverse halftoning method for comics [Kopf and Lischinski 2012]. Our reconstructions have very few artefacts and the details are mostly well preserved. Moreover, we also show the results obtained with post-processing the smoothing results with a shock filter [Gilboa et al. 2002] to enhance the edges. Even though our approach does not explicitly detect dot patterns and extract black inks as done in [Kopf and Lischinski 2012], it provides visually plausible results.

5.5 Seam Carving

The Seam Carving method introduced in [Avidan and Shamir 2007] resizes a given image by taking into account its content. The related resizing process uses a gradient-based energy function which is calculated for every pixel. Natural scenes, however, have high gradient values not only at image edges but also on textured image regions. As a consequence, the seam carving method may provide unsatisfactory results since it preserves textured regions containing details such as waves, rocks, stones, grass, etc. even if they are the part of the background and less prominent than the foreground objects.

Figure 14 shows a sample image from the RetargetMe dataset [Ru-

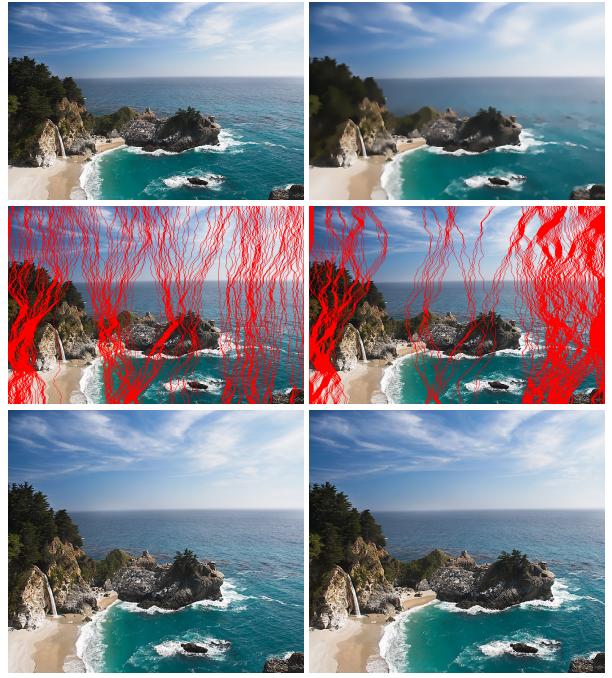


Figure 14: Seam Carving results. Left column: Original model. Right column: Model modified by our structure-texture decomposition approach. Rows top to bottom: Input image by Eric Chan from the RetargetMe dataset, eliminated seams and the resized output images.

binstein et al. 2010]. The ocean waves and rocks exist in this beach photograph have high gradient values and thus make the Seam Carving keep these regions almost entirely in the resized output image. On the other hand, if we estimate the energy function on not the original image but its structure component extracted by our Model 1, the resulting resized image becomes visually more pleasing. In this case, the method no more preserves the textured parts, selects the seams from them more, and accordingly let the visually important structures remain in the final result.

6 Discussion

We have presented the idea of using region covariances for structure preserving image smoothing. Our method employs first and second order feature statistics to obtain an implicit embedding of local image patches, which allows us to describe and distinguish local structure and texture within a single representation. We have demonstrated that using such statistical measures offers a better structure/textture separation than the previous work and improves results of many image editing and manipulation applications.

A shortcoming of our method is that it may sometimes misinterpret some image structures as texture when there are statistically similar structures nearby in terms of appearance and scale. Figure 15 illustrates an example where the text in the sign is smoothed out due to the repetitive nature of the characters. We believe that employing more complex features might help to alleviate this problem.

An important issue that needs further investigation is extending the approach for multi-scale analysis. Our current extension operates in an iterative manner by smoothing the input image by incrementally increasing the patch size, and gives good results in practice. An interesting direction for future work is to exploit the full poten-



Figure 15: Failure example. Our method mistakenly interprets the text in the sign as texture since the characters have similar characteristics in terms of appearance and repetitive behavior.

tial of region covariance descriptor that its size solely depends on the number of features not the size of the region. This opens up a possibility of a unified multi-scale formulation in which smoothing is simultaneously performed across different scales, as in [Zontak et al. 2013]. This can lead to further improvements.

Acknowledgements

We would like to thank all anonymous reviewers for their constructive comments. We also thank the Flickr users reibai, Sew Technicolor, designshard, the Wikimedia Commons for the photos used in the paper. This work was supported by a grant from The Scientific and Technological Research Council of Turkey (TUBITAK) – Career Development Award 112E146.

References

- AUJOL, J.-F., GILBOA, G., CHAN, T., AND OSHER, S. 2006. Structure-texture image decomposition–modeling, algorithms, and parameter selection. *Int. J. Comput. Vision* 67, 1, 111–136.
- AVIDAN, S., AND SHAMIR, A. 2007. Seam carving for content-aware image resizing. *ACM Trans. Graph.* 26, 3.
- BAEK, J., AND JACOBS, D. E. 2010. Accelerating spatially varying gaussian filters. *ACM Trans. Graph.* 29, 6.
- BUADES, A., COLL, B., AND MOREL, J.-M. 2005. A non-local algorithm for image denoising. In *CVPR*, vol. 2, 60–65.
- BUADES, A., LE, T. M., MOREL, J.-M., AND VESE, L. A. 2010. Fast cartoon + texture image filters. *IEEE Trans. Image Process.* 19, 8, 1978–1986.
- BURT, P. J., AND ADELSON, E. H. 1983. The laplacian pyramid as a compact image code. *IEEE Trans. Commun.* 31, 4, 532–540.
- CHERIAN, A., SRA, S., BANERJEE, A., AND PAPANIKOLOPOULOS, N. 2011. Efficient similarity search for covariance matrices via the Jensen-Bregman LogDet divergence. In *ICCV*, 2399–2406.
- CRIMINISI, A., SHARP, T., ROTHER, C., AND PÉREZ, P. 2010. Geodesic image and video editing. *ACM Trans. Graph.* 29, 5.
- DOWSON, N., AND SALVADO, O. 2011. Hashed nonlocal means for rapid image filtering. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 33, 3, 485–499.
- DURAND, F., AND DORSEY, J. 2002. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Trans. Graph.* 21, 3, 257–266.
- EFRON, A., AND LEUNG, T. 1999. Texture synthesis by non-parametric sampling. In *ICCV*, 1033–1038.
- FARBMAN, Z., FATTAL, R., LISCHINSKI, D., AND SZELISKI, R. 2008. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Trans. Graph.* 27, 3.
- FARBMAN, Z., FATTAL, R., AND LISCHINSKI, D. 2010. Diffusion maps for edge-aware image editing. *ACM Trans. Graph.* 29, 6.
- FATTAL, R., AGRAWALA, M., AND RUSINKIEWICZ, S. 2007. Multiscale shape and detail enhancement from multi-light image collections. *ACM Trans. Graph.* 26, 3.
- GILBOA, G., SOCHEN, N., AND ZEEVI, Y. 2002. Regularized shock filters and complex diffusion. In *ECCV*, 399–313.
- HONG, X., CHANG, H., SHAN, S., CHEN, X., AND GAO, W. 2009. Sigma set: A small second order statistical region descriptor. In *CVPR*, 1802–1809.
- KOPF, J., AND LISCHINSKI, D. 2012. Digital reconstruction of halftoned color comics. *ACM Trans. Graph.* 31, 6.
- MARR, D. 1982. *Vision*. W. H. Freeman and Company.
- MEYER, Y. 2001. *Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures*, vol. 22. Amer Mathematical Society.
- PERONA, P., AND MALIK, J. 1990. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 12, 629–639.
- RUBINSTEIN, M., GUTIERREZ, D., SORKINE, O., AND SHAMIR, A. 2010. A comparative study of image retargeting. *ACM Trans. Graph.* 29, 6.
- RUDIN, L., OSHER, S., AND FATEMI, E. 1992. Nonlinear total variation based noise removal algorithms. *Phys. D* 60, 259–268.
- SUBR, K., SOLER, C., AND DURAND, F. 2009. Edge-preserving multiscale image decomposition based on local extrema. *ACM Trans. Graph.* 28, 5.
- TOMASI, C., AND MANDUCHI, R. 1998. Bilateral filtering for gray and color images. In *ICCV*, 839–846.
- TUZEL, O., PORIKLI, F., AND MEER, P. 2006. Region covariance: A fast descriptor for detection and classification. *ECCV*, 589–600.
- WINNEMÖLLER, H., OLSEN, S. C., AND GOOCH, B. 2006. Real-time video abstraction. *ACM Trans. Graph.* 25, 3, 1221–1226.
- WITKIN, A. P. 1984. Scale-space filtering: A new approach to multi-scale description. In *ICASSP*, vol. 9, 150–153.
- XU, L., LU, C., XU, Y., AND JIA, J. 2011. Image smoothing via L_0 gradient minimization. *ACM Trans. Graph.* 30, 6.
- XU, L., YAN, Q., XIA, Y., AND JIA, J. 2012. Structure extraction from texture via relative total variation. *ACM Trans. Graph.* 31, 6.
- ZONTAK, M., MOSSERI, I., AND IRANI, M. 2013. Separating signal from noise using patch recurrence across scales. In *CVPR*.