Segmentation using Appearance of Mesostructure Roughness

Yaser Yacoob and Larry Davis

Computer Vision Laboratory

University of Maryland

College Park, MD 20742

Abstract

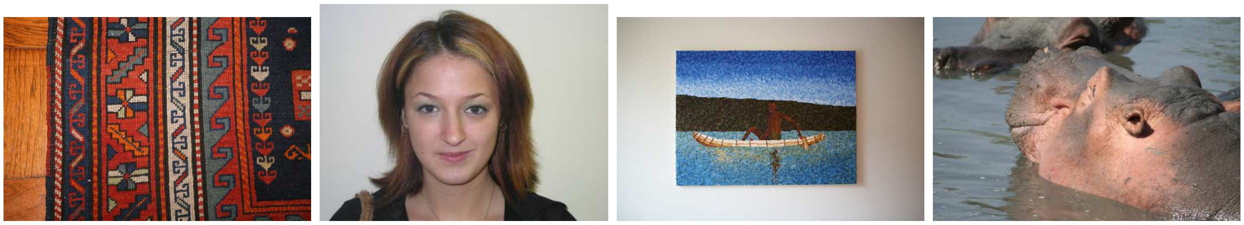
This paper introduces mesostructure roughness as an effective cue in image segmentation. Mesostructure roughness corresponds to small-scale bumps on the macrostructure (i.e., geometry) of objects. Specifically, the focus is on the texture that is created by the projection of the mesostructure roughness on the camera plane. Three intrinsic images are derived: reflectance, smooth-surface shading and mesostructure roughness shading (meta-texture images). A constructive approach is proposed for computing a meta-texture image by preserving, equalizing and enhancing the underlying surface-roughness across color, brightness and illumination variations. We evaluate the performance on sample images and illustrate quantitatively that different patches of the same material, in an image, are normalized in their statistics despite variations in color, brightness and illumination. We develop an algorithm for segmentation of an image into patches that share salient mesostructure roughness. Finally, segmentation by line-based boundary-detection is proposed and results are provided and compared to known algorithms.

Key words: Texture analysis, image segmentation, intrinsic images.

1 Background

Image understanding often involves scenes in which man-made objects are presented in their diversity of appearance. For example, while clothing is, perhaps, the most diverse of such objects, its basic components can be simply reduced to: the material of the thread (i.e., the fiber such as cotton, wool, etc.), the thickness of the thread and the particular weaving pattern. Similarly, human hair may be colored to achieve a diverse appearance, while the underlying material structure of hair remains typically unchanged. Figure 1 shows examples of images we are interested to analyze, a: (a) rug on a hardwood floor, (b) woman with multiple hair-colors, (c) textured painting hanging on a wall and (d) hippopotamus in water. In all these images, our objective is to delineate regions that have a basic common underlying structure despite significant variations in color, brightness and illumination attributes. Specifically, we seek the separation of the rug, hair, painting and hippopotamus from the rest of the image based on the fine-level structure of the scenes despite the intertwining of texture, color and shading. These images are taken at medium and high resolution (1600x1200 for the face and 4368x2912 otherwise) to capture the detailed surface-appearance of materials. It is important to note that image segmentation techniques that focus only on edges, color and monochromatic texture attributes are not well suited to handle these types of images since these cues are not the channels that convey the critical visual information.

The visual cue that we consider is the projection of *mesostructure roughness* on the camera plane. Mesostructure roughness is recognized as an important cue in diverse areas of science from space, geology, material, computer graphics, food to nano sciences. In computer graphics significant research has been invested in synthesis of real-looking scenes. It has been observed (e.g., [5, 14]) that real-world object appearance is a function of geometry, reflectance, and illumination. Geometry in turn is divided into macrostructure, mesostructure and microstructure ([11]). Macrostructure represents the coarse surface geometry and typically involves planar or deformable surfaces that can be studied and represented by classic or differential geometry. Mesostructure captures the bumps in the geometry and conveys a measure and pattern of *roughness*. In all but few materials roughness is an integral attribute that conveys lack of perfect surface uniformity of the shape at an observed scale. Microstructure represents shape at inter-molecular level. Koenderink and van Doorn [11] observed that these geometric structures are *scale-defined* and are typically 1-2 orders of magnitude apart in the case of macro and meso structures. In other words, the ability to simultaneously observe these structures requires capturing details at several orders or magnitude at once. The increase in camera resolution has made this possible although its benefit has gone unnoticed in image



(a) (b) (c) (d)

Figure 1: (a) a rug on a wood floor, (b) hair colored with different shades (c) a richly texture painting on a wall and (d) hippopotamus.

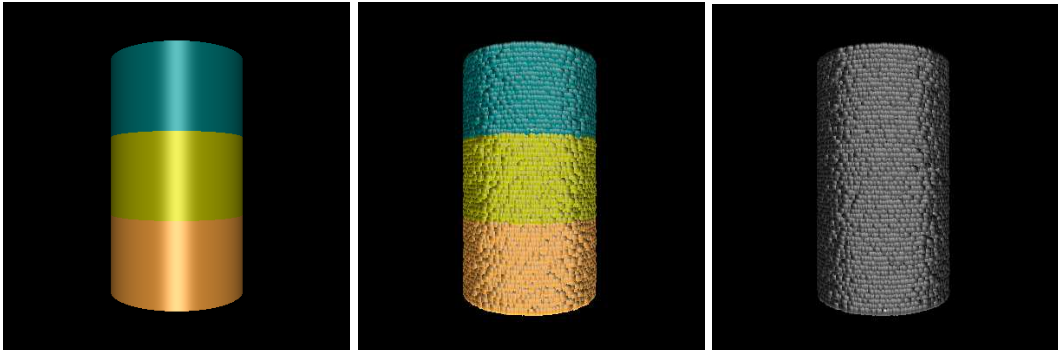


Figure 2: (Left) A smooth-surface cylinder with 3 distinct reflectance regions, and (Middle) a mesostructure roughness replaces the smooth surface while reflectance and shading are unchanged, (Right) Removing reflectance leaves a combined smooth and rough surface shading image of the cylinder.

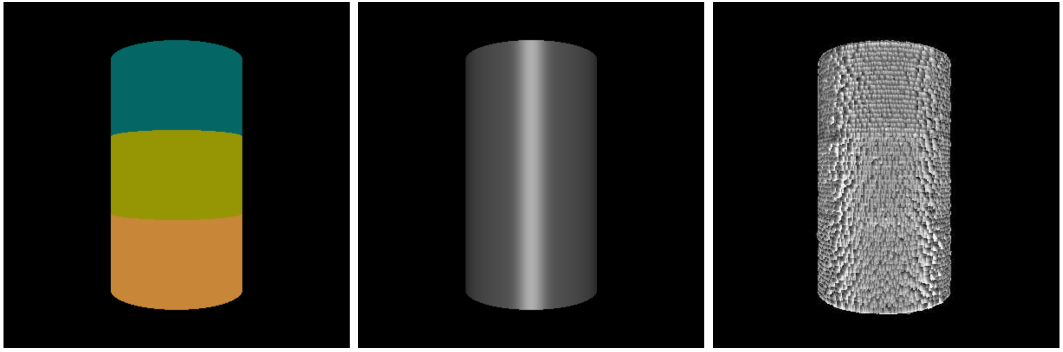


Figure 3: (From Left) The reflectance, smooth surface shading and mesostructure roughness images of the cylinder as proposed by the proposed intrinsic image formulation.

understanding.

Koenderink and van Doorn [11] studied the variation in illumination as a function of macro and meso structure attributes. They observed that macrostructure gives rise to smooth shading whereas mesostructure leads to illuminance texture. Their paper focused on these structures in a synthesis setting such as in computer graphics. A comparable analysis-oriented study remains open for research but in this paper we take an empirical approach to explore this complex subject.

Dana et al. [8] observed that the appearance of real-world surfaces at macroscale geometry is captured by a BRDF (bidirectional reflectance distribution function) while mesostructure is captured by a BTF (bidirectional texture function). For a fixed illumination and viewpoint these function become simply reflectance and texture. The co-occurrence of macro and meso structures is part of the real world but is not captured by this division.

In the rest of the paper we use the term smooth-surface to correspond to macro-structure and mesostructure roughness to indicate surface bumps.

Figure 2 illustrates the different appearance of macro and meso structures on a graphically rendered cylinder. The left image shows the smooth surface shading of a cylinder when mesostructure is invisible at the imaged scale. The cylinder is striped with three different reflectances. The middle image illustrates the appearance of the same cylinder with surface mesostructure at a scale that can be observed as visible roughness. The rendering parameters are kept constant in these two cases. The right image shows the combined smooth and rough surface shading without the reflectance component. The combined shading can be contrasted with the shading of smooth-surface cylinder in Figure 3 (middle).

Figure 2 illustrates the interplay of geometry, reflectance and illumination as the primary components in computer graphics. Both macro and meso structures are present, but the mesostructure is at a scale that can be observed in the middle image. In terms of computer vision, however, the confounding of these components creates an enduring challenge. Figure 3 shows the decomposition of the rough-surface cylinder into three components, reflectance, smooth-surface shading and roughness shading, respectively. Reflectance variations do not provide information about the shape of the cylinder. Smooth-surface shading convey a characteristic illumination of a smooth cylinder and is normally posed as a shape-from-shading problem. The image of the mesostructure roughness provides an opportunity to recover the cylinder shape by formulating a shape-from-texture problem, however, it is more challenging than the shape-from-shading problem.

Mesostructure roughness has attracted limited interest for image understanding of real-world scenes (e.g., [8, 19]) since low-resolution images rarely capture sufficient mesostructure roughness details. Note that given the 1-2 order of magnitude hypothesis [11] it is not surprising that only when sufficient image resolution became available did mesostructure become viable for visual analysis. The emergence of high-resolution digital cameras create an opportunity to observe and analyze this geometric cue. It is now common to capture images at 10 MPixels and it is likely that 100 MPixels cameras will be common in a decade. We argue that high resolution images reveal a new visual cue, mesostructure roughness, and can improve visual analysis.

The visibility of projected mesostructure roughness in images inspires our research to develop a transformation that converts a color image into an image stripped of all information other than what intuitively can be described as *meta-texture* information (or fine-structure of the surface) where image patches of similar underlying mesostructure roughness appear uniformly textured despite original differences in color, brightness or illumination. Note that mesostructure roughness is a 3D attribute but we focus on the projection of mesostructure roughness as an apparent image texture. We employ the term meta-texture to differentiate our image analysis from common texture analysis which is dependent on color, brightness variations and, to some extent, illumination (excepting the 3D texton-based approaches).

It is important to recognize that since macro and meso structures are confounded in real-world scenes, the respective smooth and rough surface shading are confounded as well in images. We observe that the surface of an object in the real-world is more likely to exhibit meso-structure uniformity rather than macrostructure uniformity (even simple shapes like sphere or cylinder satisfy this). As a consequence it follows that it is better to segment a scene based on the mesostructure and not the macrostructure. The challenge therefore is to isolate and estimate the texture that corresponds to the mesostructures in a scene.

1.1 Color and texture-based segmentation

There is a large body of research on texture-based segmentation of color images (e.g., [2, 4, 16, 18]). Segmentation is commonly posed as boundary detection of visual cue changes between neighboring pixels. Cues such as color, brightness and texture are commonly used. In the case of brightness, discontinuities in the brightness are detected as edges and indicate occurrence of boundaries. However, if mesostructure roughness is present then brightness discontinuities may be detected at local changes in brightness due to changes in facet geometry and not global changes in the surfaces. Detecting boundaries using color requires quantifying a notion of color constancy that enables recognizing when two neighboring pixels have different colors. Mesostructure roughness creates significant changes in color between neighboring pixels since the local shading can drastically change due to the change in surface normals of each facet. Texture boundary detection requires computing attributes over a set of pixels (typically a square) and as result can be affected by global shading changes that happen to occur within the pixels that the texture descriptor is computed over. In conclusion treating each visual cue separately has proven to be challenging for detection boundaries in real-world images.

Boundary detection using multiple cues benefits from the assumption that true boundaries typically change both color, brightness and texture. Martin et al. [18] formulated a gradient-based approach that uses brightness, color and texture to compute a set of local features that identify boundaries.

When applied to the images in Figure 1, state-of-art segmentation algorithms result in the identification of multiple regions due to color or apparent texture variations in spite of the similarity of the underlying surface texture (see Tables 1 and 2). This lead us to revise the notion of boundaries in some images. For example, the image of the carpet has many boundaries due to color changes, but content-wise these are not boundaries we are interested in.

1.2 Material perception and 3D Textons

Adelson et al. [1] studied the properties of materials under different illumination to determine classification by humans and study statistical texture measures. Leung and Malik [13] (also [7, 24]) proposed an approach for classifying materials based on 3D texture attributes, 3D textons, computed over small patches to capture local geometric and photometric properties of monochromatic images. Recognition of different materials under different lighting and viewing conditions was shown. The experimental results typically involve a classification process, where images are compared to a gallery of textures.

Note that a 3D texton-based approach can be adapted for image segmentation if the color textured areas have clear boundaries and large sizes which in the monochromatic image may be amenable to detection, normalization and analysis. However, it has to account for the change in texture due to reflectance and smooth surface shading variations. Reflectance variations appear as variable texture in monochromatic images. As a result filter responses will vary due to reflectance change in addition to texture variation. Smooth surface shading variations also change the texture appearance but generally across larger areas of the image. It should be noted that the meta-texture image proposed in this paper is a potential input for a 3D-texton based recognition process and not an equivalent approach for processing image texture.

Physics-based models for rough-surface image formation have been proposed for highly simplified scenes (overviews can be found in [15, 20]). Oren and Nayar [19] proposed a model for predicting reflectance from rough diffuse surfaces by employing a Torrance and Sparrow [23] V-cavities model. The key assumption was that multiple facets (i.e., rough surface components) are projected onto a single pixel in the image. In contrast, we assume that a facet is projected onto several pixels in the image. Measurement of the appearance of rough surfaces have been proposed using the Bidirectional Texture Functions (BTF) [8] and analysis and recognition [7].

In conclusion, analysis of rough surfaces in real-world images without the benefits of BTF, simplified physics models or learned models remains open for research and is an objective of our research.

1.3 Intrinsic Images

Intrinsic images (e.g., [3, 9, 12, 21, 22]) aim to reveal the underlying physical properties of a scene by estimating independent attributes such as shading (i.e., a function of lighting, viewer location and surface normals) and reflectance (i.e., surface color) images. Intrinsic image analysis has so far focused on coarse-detail images (i.e., when mesostructure roughness is invisible) and ignored fine-detail images that convey mesostructure roughness. Intrinsic images analysis involves scenes which arise from albedo or color variations on smooth surfaces while Figure 1 involves in addition rough surfaces with complex dependencies on color, viewing and illumination directions that destabilize simultaneously the shading and albedo due to variations at fine scales.

Estimating intrinsic images from a single image has typically involved computing the image derivatives and associating their values with surface attributes. Land and McCann [12] associated large image derivative magnitudes with albedo (i.e., reflectance) changes and small magnitudes with illumination (i.e., shading) changes. Tappen et al. [21, 22] proposed learning-based approaches for the classification of image derivatives, so that a less-stringent criterion is applied to image derivative magnitudes. In [21] a classifier is trained using Adaboost to disambiguate locally ambiguous areas of the image by propagating information along edges in the image using MRF. The challenge is to train the classifier on real images where the division into two classes, shading and reflectance components, is unnatural or unknown. The work in [22] builds on [21] by removing the binary discretization and adding a weighting of estimates in a mixture of experts framework.

1.4 Paper's Contribution

This paper investigates the potential of the mesostructure roughness cue for image segmentation. Image segmentation has been dominated by edge, color and general texture analysis. The emergence of digital imaging introduced a new and rapidly evolving parameter, the continuous increase in image resolution. Given that Moore's Law (i.e., the number of transistors on an integrated circuit doubles in less than two years) applies to digital cameras, it is important to develop algorithms that use all the information embedded in high resolution images. This paper introduces a new cue, mesostructure roughness, that becomes visible as resolution increases and as long as image blur is not present. With a few exceptions (e.g., air, water), all surfaces have mesostructure roughness if imaged at an appropriate scale.

The paper is focused on the diverse appearance of real-world rough-surfaces on top of smooth-surfaces and their surface texture functions (e.g., BTF, 3D Textons) are unknown. We define a meta-texture-image (MTI) to be a grey-scale image derived from a color image so that scene mesostructure roughness similarity translates into similarity of the image texture regardless of *reflectance* (e.g., color, absorbence, etc.), *lighting* (e.g., intensity, location) and *smooth-surface shading* (e.g., slant of the smooth surface with respect to the viewer and light source as long as roughness detail remains visible) variations *within* the image. Note the focus on image-centricity as opposed to inter-image variations in [7].

We propose that the intrinsic images framework [3] be expanded to convey mesostructure roughness as an independent intrinsic image. Whereas the traditional shading image aimed to support shape-from-shading, the shading image of mesostructure roughness reveals some material properties and could also serve as the basis for shape-from-texture. An important tenet of the intrinsic images framework is the *independence* of the shading and reflectance images for Lambertian scenes. Mesostructure roughness is an attribute related to shading since it reflects local-shading variations (i.e., local change in the surface normal) and can generally be viewed as independent of reflectance. Mesostructure roughness requires a wider interpretation of local derivatives in contrast to the bi-modality of the derivatives in the Retinex framework [12] (i.e., large derivatives at changes of reflectance and small derivatives at smooth changes of the surface normal). In reality, when mesostructure roughness is present, large derivatives can occur at rough-surface shading changes as well as at reflectance changes while small derivatives are likely to occur at smooth and rough surface shading changes. Therefore, we propose a patch-based approach and a formulation to derive the information from intrinsic images.

It is worth noting that it may be possible to embed mesostructure roughness treatment in a late-stage processing or post-processing of images. For example, in image segmentation, the mesostructure roughness attributes can be computed for patches after color/edge segmentation have been accomplished. We choose, however, to tackle mesostructure roughness as a pre-processing step of image analysis, and avoid entangling mesostructure roughness with segmentation. Therefore, we anticipate additional uses for the meta-texture image such as detection, synthesis, shape-from-texture, etc. which this paper does not explore.

The paper's contributions are characterizing the problem, proposing an approach to derive three intrinsic images from a color image (i.e., smooth-surface shading, reflectance, and meta-texture), proposing and developing the concept of salient meta-texture image (MTI) via transforming an image into a grey-level image in which the projected mesostructure roughness is *preserved, equalized* and *enhanced* while other properties such as color, brightness variations, smooth-surface shading, etc. are normalized. This MTI is evaluated quantitatively and image segmentation by texture-boundary detection is developed.

In Section 2 the approach for deriving the intrinsic images, smooth and rough surface shading images and reflectance image, is provided. Section 3 details the segmentation approach we developed and is followed by experimental results in Section 4.

2 Approach

Assume an ideally diffusing surface under a distant point source and that light is distributed uniformly independent of the viewing direction (i.e., Lambertian). Also assume that the camera response is linear with respect to the incoming energy and that different colors are equally reflected. Let N_L represent a unit vector of the light direction, F represent the flux density of the light source, and $N_S(x, y)$ be the unit surface normal to the smooth surface at a 3D point, P that is the source point of (x,y) in image I. In Figure 4 we represent the smooth surface as a cylinder without any loss to generality. For a smooth surface, the intensity image, $I(x, y)$, captured by a camera is

$$I(x, y) = F \times (N_S(x, y) \cdot N_L) \times R(x, y) \tag{1}$$

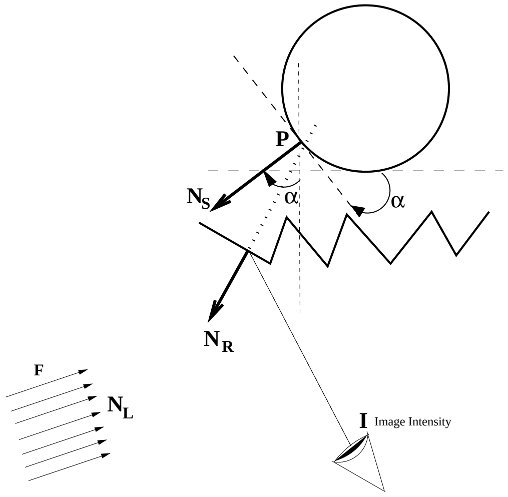


Figure 4: The geometry of rough surface appearance when placed on a smooth surface. The normal N_R on the undeformed rough surface is displaced by the smooth surface normal N_S.

where $\cdot$ is the dot product and $R(x, y)$ is the reflectance of the surface (i.e., albedo). This equation can be rewritten as

$$I(x, y) = L(x, y) \times R(x, y) \tag{2}$$

where $L(x, y)$ is the incident illumination also known as shading.

Consider the mesostructure roughness depicted in Figure 4 as it exists prior to its deformation to lie on the smooth surface of the cylinder. Let N_R represent the surface normal of a point prior to deformation. This point is mapped to a point P at the surface of the cylinder after deformation of the rough surface on the cylinder. Let N_S represent the normal to the cylinder at P. It is straightforward to see that the tangent to P forces a rotation of α on the base horizontal direction of the rough surface (the rough surface is aligned with the horizontal direction). The angle α is also represented by the direction N_S since we consider P the center of the coordinate system of these vectors. Therefore, the rough surface normal N_R changes direction by angle α which is equal to vector summation of N_S and N_R. Equation 1 can now be rewritten as (using also Equations 1 and 2)

$$I(x, y) = F \times ((N_S(x, y) + N_R(x, y)) \cdot N_L) \times R(x, y) \tag{3}$$

Note that $R(x, y)$ now represents the reflectance of the combined smooth and rough surfaces and not the smooth surface as in Equation 1. Since vector addition is distributive, it follows that

$$I(x, y) = (F \times (N_S(x, y) \cdot N_L) + F \times (N_R(x, y) \cdot N_L)) \times R(x, y). \tag{4}$$

Let $G(x, y) = (F \times (N_R(x, y) \cdot N_L)$ represent the shading of the mesostructure roughness, then Equation 4 can be rewritten as

$$I(x, y) = (L(x, y) + G(x, y)) \times R(x, y). \tag{5}$$

This suggests that the perceived shading is a summation of two shading components, one stemming from macrostructure and the other form mesostructure. Note that Equation 5 does not take into account the deformation of the shading of the mesostructure due to the (1) perspective camera projection, and (2) compression in the mesostructure at the sides of the cylinders due to self occlusions.

Untangling the three component functions (L, R and G) is a grand challenge, especially given the well-known difficulties of recovering smooth surface shading and reflectance even after four decades of research interest. Moreover, Equation 5 is under-constrained and therefore infinite solutions are mathematically plausible but few express or reveal visual meaning. Note that evaluating correctness of the derived three intrinsic images for real-world images cannot be done (a problem we share with existing work on intrinsic images). The meta-texture image we develop in this paper represents a realization of the shading of the mesostructure but it is possible to derive other estimates of G.

2.1 Expansion of the Intrinsic Images Framework

Deriving intrinsic images by extending known formulations is not possible since mesostructure roughness violates the basic assumptions of derivative-based approaches (e.g., [12, 21]). Figure 5 (left) shows an image of a T-shirt that is used to illustrate the basis of our analysis. The T-shirt is deformed to create a smooth 3D surface. The right side of the image shows the T-shirt surface bending away from the camera. At the same time, the mesostructure roughness of the T-shirt can be seen as variations of shading within

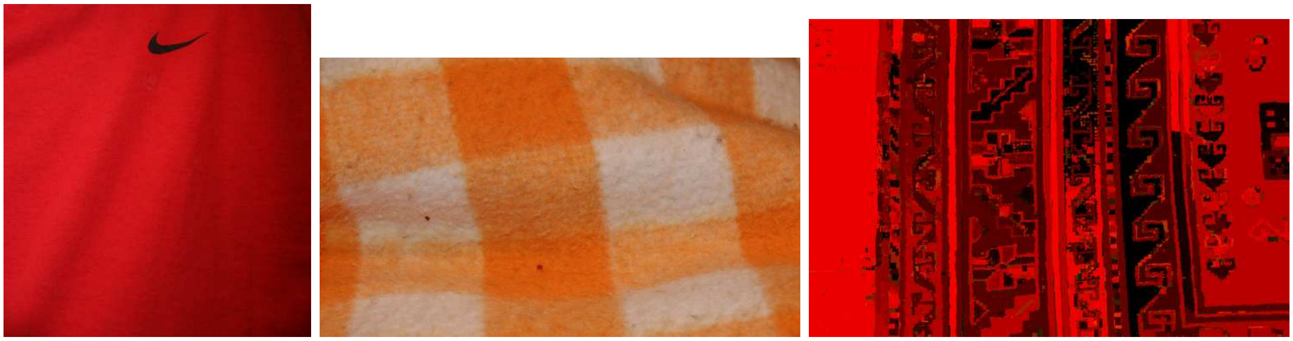


Figure 5: (Left) An image of a T-shirt exemplifying the smooth-surface and mesostructure roughness of a scene. (Middle) An image of a blanket exemplifying the smooth-surface and mesostructure roughness of a scene with varying reflectance. (Right) The patches computed for the rug on hardwood floor image (see Figure 1). The patch number is color-coded for display, from the Red channel through the Green to the Blue.

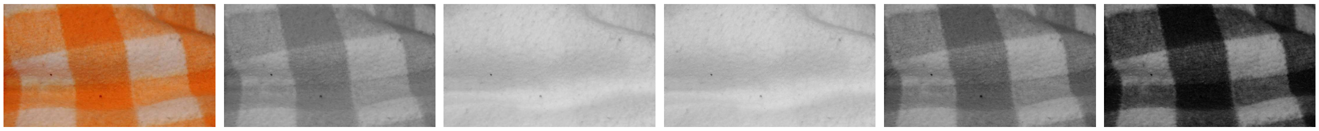


Figure 6: From left to right, respectively, input, intensity, V, Red, Green, Blue images.

local neighborhoods. Figure 5 (middle) shows an image of a blanket that also illustrates smooth-surface deformations, mesostructure roughness and reflectance variations. We observe that:

Observation 1. *Global (i.e., whole image), local (i.e., fixed-size window-based) and pixel-based operations are unlikely to enable estimating the three intrinsic images since reflectance and shading variations are deeply intertwined within these spatial structures.*

It is worth noting that Tappen et al [21] also argued and showed that simple whole-image low and high pass filtering are unable to untangle shading and reflectance. Therefore, they developed a discriminative-based approach focused on local image derivatives and a global Markov Random Field to propagate the classification to the whole image. Similarly, fixed-size window-based operators are not suitable since they arbitrarily impose a lattice on the image. Pixel-based approaches (e.g., [21]) are challenged since shading and reflectance are completely intertwined at single pixels and as a result methods for propagating information between pixels are devised (e.g., using MRF). Consequently, we explore an approach that divides an image into *homogeneous patches* and then derives intrinsic images based on analysis of these patches.

Observation 2. *While smooth and rough surface shading are reflected independently in the RGB channels* *the channel with the highest amplitudes provides the best basis for shading reconstruction.*

In the following we use Equation 2 to illustrate the observation but the treatment is equally applicable to Equation 5. Equation 2 can be rewritten based on the intrinsic images of the 3 RGB wavelengths,

$$\begin{aligned} RED(x,y) &= L(x,y) \quad F_{red}(x,y) \\ GREEN(x,y) &= L(x,y) \quad F_{green}(x,y) \\ BLUE(x,y) &= L(x,y) \quad F_{blue}(x,y) \end{aligned} \tag{6}$$

where $RED(x,y), GREEN(x,y)$ and $BLUE(x,y)$ are the R,G, and B wavelength images and $F_{red}(x,y), F_{green}(x,y)$ and $F_{blue}(x,y)$ are the reflectance images in each channel. Note that the shading image is the same in the three equations since shading is independent of reflected wavelengths. Assuming scene material disperses all wavelengths in the same direction, and given homogeneity of the patch, all wavelengths reflect *independently* the surface shading. As a result we have three equations with four unknown images (the shading and three reflectance images). The equations indicate that shading image can be computed from any single wavelength. Figure 7 shows an image part of the T-shirt treated as a single patch in which the shading images are computed independently from each of the RGB channels using Equation 6 where the reflectance in each channel is estimated to be a constant equal to the mean R, G and B values in each channel (since the T-shirt is presumed to have a single reflectance color). Clearly the shading images are equal up to a scalar multiplier. The wavelength that has the highest overall amplitude provides the strongest signal. In Figure 6, the red channel has the highest amplitude and therefore it provides the best signal-to-noise data. In general, different channels may be most suitable in different parts of the image. Therefore, a patch-based analysis is needed where a single channel is likely to be dominant within the patch and therefore it can be used for the intrinsic image estimation. This contrasts with grayscale intensity image use in [9, 12, 21].

To illustrate Observation 2, we show a simple scene of an image of a blanket (see Figure 6) where mesostructure roughness is uniform, smooth- surface shading changes in a few places while reflectance is, in principal, of three colors (white, light orange and deep orange). Figure 6 shows the input, grayscale intensity, V , Red, Green and Blue images respectively. Note that the intensity image shows intertwining of reflectance

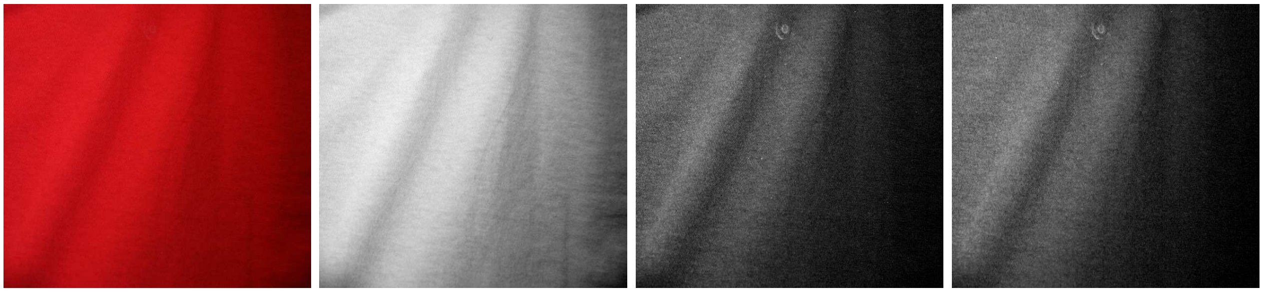


Figure 7: (Left to right) An image of a T-shirt treated as a single patch, and the three shading images computed using each of the RGB channels, respectively (Equation 6.

and shading while the V and Red channels (Red happens to be the channel where the highest amplitude luminance occurs, so that V and the Red channel are, in fact, equal) show exclusively the shading variations without reflectance. The smooth and rough surface shading are reflected independently and similarly in the Red, Green and Blue channels since the interaction between the light wavelength and the normal to the surface is equal in the three channels. However, the grayscale intensity image (due to the weighting of RGB) entangles the reflectance (present in the RGB data) with the shading which tend to make the reconstruction of the shading more challenging. In contrast, the Red channel (as well as the V) appears to readily reflect shading without being affected by reflectance.

In the following subsections we answer the following questions:

- How should homogeneous patches be determined in an image?
- How should intrinsic images be derived?
- How should patches be transformed so that the mesostructure roughness be optimally revealed?

2.2 Reflectance-Homogeneous Patch Selection

The discussion above concluded that deriving intrinsic images may benefit from a patch-based analysis. Patch-selection is challenging since it is an instance of image segmentation which is the objective of our research. Based on the earlier observations we define a simple criterion for detecting homogeneous patches:

Homogeneous Patch: *A patch is homogeneous if variations of its pixels are likely to be associated with smooth or rough-surface shading variations but not reflectance variations.*

Note that the definition aims to classify pixels into patches as long as their reflectance appears unchanged without estimating the actual reflectance and shading directly. It does not involve explicitly using derivatives as the criterion for reflectance change although it relies on the differences between pixels.

Using this definition, we can analyze reflectance of different surfaces in the image by an *inter-patch* process and analyze smooth and rough surface shading variations as an *intra-patch* process to finally arrive at a novel mesostructure roughness image.

We first convert the RGB color image into (Hue, Saturation, Value) (HSV) space. In this space hue mostly corresponds to reflectance and it describes the dominant wavelength at a pixel while the smooth and rough surface shading images are typically visible in the saturation and value images. It should be noted that it is possible to use the RGB representation using linear shadow/highlight color models such as the ones employed by [10, 21]. However, we consider HSV space more suitable since it reflects more intuitively and readily the reflectance and shading attributes.

We employ a conservative approach for patch delineation since some over-segmentation typically has only limited negative impact because the statistics of multiple patches sharing an underlying mesostructure roughness structure are similar. Each pixel, $P_{i,j}$, in HSV space will belong to a single patch P. Given a seed patch P with a single pixel $P_{i,j}$, a connected component expansion of P under constraints on the values $P^H_{i,j}$, $P^S_{i,j}$, $P^V_{i,j}$ is performed. A pixel, $P_{k,l}$, 8-connected neighbor of a pixel, $P_{i,j}$, is added to P if

$$\begin{array}{rcl} |P^H_{i,j} - P^H_{k,l}| & < & H_{threshold} \\ |P^S_{i,j} - P^S_{k,l}| & < & S_{threshold} \\ |P^V_{i,j} - P^V_{k,l}| & < & V_{threshold} \end{array} \tag{7}$$

where $H_{threshold} = 10$, $S_{threshold} = 40$, and $V_{threshold} = 50$. These values were determined empirically and are applied to all images in this paper, taking into consideration the nonlinearity of the HSV space.

A two-pass algorithm for segmenting the image into patches is implemented. In the first pass each pixel is allowed to be part of as many spatially-connected patches as it conforms to using Equation 7, while in the second pass the optimal patch of a pixel is chosen based on the patches' sizes and similarity of the pixel to all patches' statistics. Typically, hundreds or thousands of patches are found in a complex image.

Figure 5 (right) shows the patch determination for the image of a rug on a hardwood floor (see Figure 1). The patch number is color-coded so that the first patch has a color $(0, 0, 0)$ in RGB space, the 256 patch has $(255, 0, 0)$, after which the green channel is used, etc. It should be noted that the boundary between two neighboring patches occasionally forms a separate patch. This occurs since the boundary area often shows a mixture of the color properties of its neighboring patches and thus it appears different if these patches are sufficiently different.

2.3 Deriving Intrinsic Images

The visual information of a patch P is divided into smooth-surface shading, reflectance and meta-texture. We will use the 'shading' to express smooth-surface shading in the rest of the paper whereas shading encompassed both smooth and rough surface shading in earlier work [21, 22].

Reflectance. Given the homogeneity of P, reflectance should convey the substrate color of the patch independent of mesostructure roughness and smoothness of P. If we assume that the patch has a unimodal distribution of color along which mesostructure roughness and shading values vary, reflectance can be computed as the mean hue, saturation and brightness over P (i.e., $(H_{avg}$,S_{avg},$V_{avg}))$.

Shading: Since shading reflects illumination variations of the smooth-surface of P it can be computed by discarding the high-frequencies (e.g., by low and middle pass filtering) of the brightness V since these stem from mesostructure roughness. It is critical to note that we associate shading with the *maximum* amplitude of the light waveform of the patch (i.e., V) as opposed to the *combined* RGB amplitudes of light waveforms (i.e., intensity); Note that homogeneity presupposes a dominant color which implies that one of the RGB channels is likely to dominate V which was discussed in Observation 2.

Meta-texture. Meta-texture reflects the removal of the smooth-surface variations from the intensity image of the patch (e.g., by a high-pass filtering). Note that V is used but if one of the RGB channels is saturated the intensity image preserves the local texture while V eliminates it.

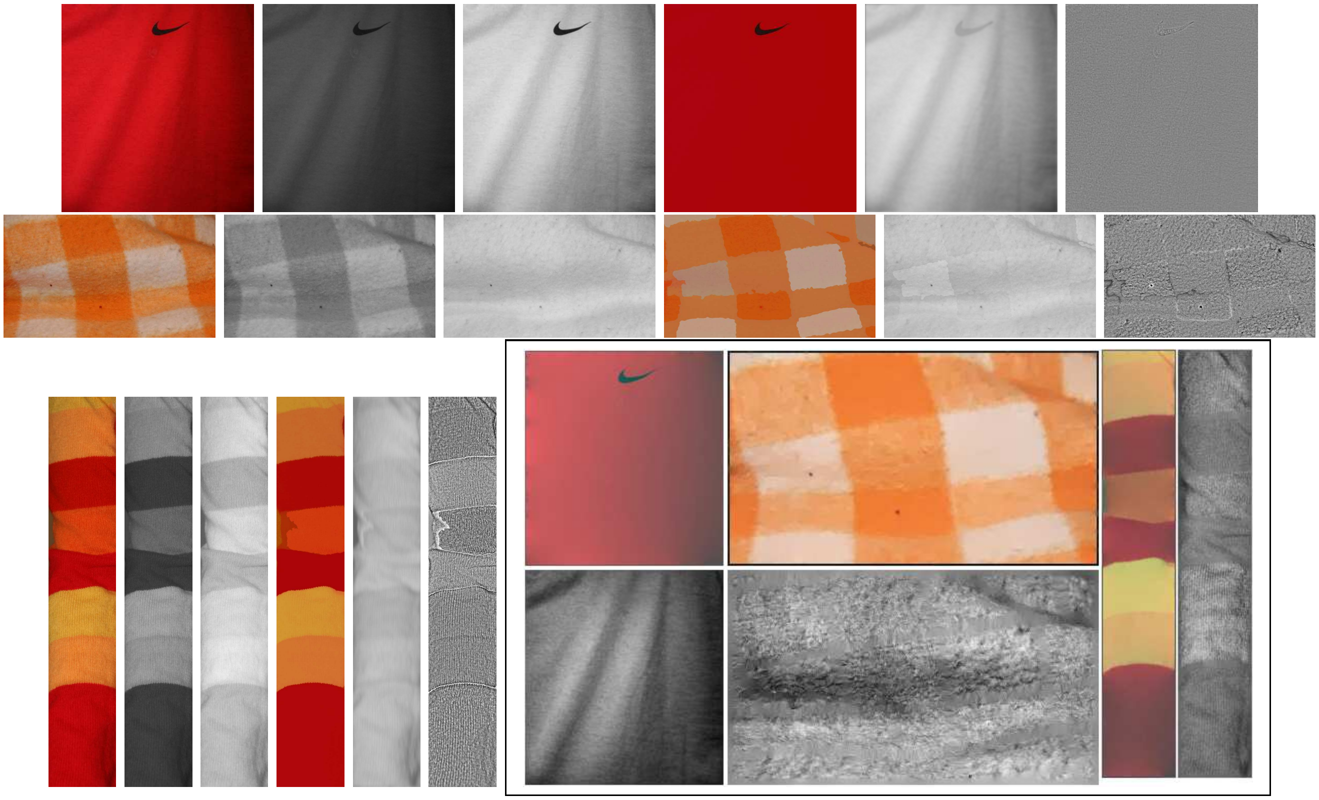


Figure 8: From top to bottom (raster-wise) input, intensity, V, reflectance, shading and MTI for each image followed by the respective reflectance and shading images computed by [21].

Normalizing values across patches is necessary for the shading and MTI images (for now we use simple mean normalization).

Figure 8 shows the intrinsic images derived for three sample images: T-shirt, blanket and a wool sweater over an arm. The images convey color variations, smooth and rough surface shading. For each image, the input, intensity, V, reflectance, shading and MTI are shown, respectively. The reflectance images show a completely flat image with surface smoothness and roughness removed, the shading images show the global smooth surface deformations and the MTIs show the mesostructure roughness excluding the shading information. Notice that patch boundary artifacts appear in the meta-texture and shading images.

The bottom right side images (surrounded by a rectangle) show the reflectance and shading images as computed by [21]. There are significant differences in the results. The reflectance images of [21] show traces of shading (especially in the case of the Red T-shirt), while our approach meets the expectation that the T-shirt has a Red reflectance with the black logo. Similarly in the case of the blanket, mesostructure roughness appears in the reflectance image despite the expectation of uniform color distributions. The shading images computed using [21] confound smooth and rough-surface shading, while our approach separates them.

It is important to note that the proposed derivation of the three intrinsic images is consistent with the Lambertian formulation expressed in Equation 5. Equation 5 is ill-constrained and therefore infinite solutions are mathematically plausible but few express or reveal visual meaning. As defined above, the summation of the MTI and shading, (i.e., $L(x, y)$ and $G(x, y)$) for each patch and for the aggregation of patches in the image is equal to summing the high and the remaining frequencies filtered components, respectively. Thereby preserving the information content of the grayscale image (in our case it is the V image). Moreover, the multiplication by the reflectance image leads to reconstitution of the input image. We verified this for the three images used in Figure 8.

Since Equation 5 is underconstrained there are other solutions that may provide greater visual meaning but are not explored in this paper since our focus is on mesostructure roughness. Specifically:

- Given a reflectance-homogeneous patch as defined above, what is the most accurate value that reflectance should be assigned? Using the average RGB (or HSV) can be easily improved upon if the smooth-surface shading is analyzed to determine which areas are darkly or brightly shaded and thus converge to the reflectance value for a patch with a surface normal that is facing the camera (or any other criterion). Possible approaches include statistical analysis, physics-based analysis and Retinex-based analysis to determine how the surface within the patch is affecting reflectance values at each pixel. Note that mesostructure roughness locally impacts the true reflectance value but is presumed to be averaged-out across the patch.

- The smooth-surface shading of a patch was computed by discarding its high-frequency components and keeping the low and medium frequencies. The result is that at coarse level the smooth-surface shading appears realistic. However, at a fine level it appears that the mesostructure roughness can also contribute to low-frequencies and therefore a subtle trace of the mesostructure roughness may remain visible. A more accurate smooth-surface shading using frequency analysis and derivatives analysis remains open for research.

- The smooth and rough surface shading were normalized across patches by a simple mean normalization. This process has a clear shortcoming as can be seen in the smooth surface shading images in Figure 8 (e.g., the Nike logo patch stands out from the T-shirt). Although the means of the patches are the same, the mean of the red patch is biased by smooth-surface shading. A similar problem exists for MTI patch boundaries. These artifacts create shading differences where there should be none and the elimination of these artifacts remains open for research.

In the rest of the paper we focus on the MTI and its potential for image segmentation.

2.4 Deriving a Meta-texture Image for Arbitrary Shaped Patches

The frequency-based estimation of mesostructure roughness shading cannot be applied to arbitrarily shaped patches. Therefore, a derivation of MTI by a construction-based approach is developed. The objective is to transform a patch into a grey-scale patch in which mesostructure roughness is *preserved*, *equalized* and *enhanced*. Preservation ensures that the meta-texture is not weakened. Equalization means that patches with the same underlying mesostructure roughness but perhaps different color, brightness or illumination are transformed into a similar meta-texture. Finally, where mesostructure roughness appears weak due to color, brightness or illumination it is enhanced (i.e., amplified) to reflect its prototypical appearance. Figure 9 shows our algorithm applied to a colorful cotton sweater. Preservation reveals the meta-texture of the red patch. Equalization shows that the meta-texture is equalized across the five shades of green and blue. Enhancement is shown for the black region where the Mesostructure roughness is revealed, enhanced and equalized with respect to other meta-textures. The MTI of these patches are similar and thus could easily be classified into the same material.

We first consider the distribution of brightness values within homogeneous patches. Figure 10 (top two rows) shows sample patches and histograms of the brightness values of each patch. The patches were taken

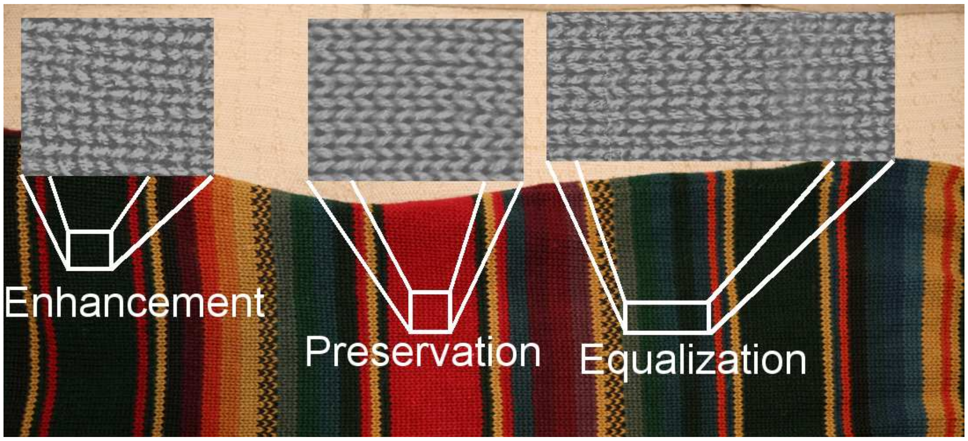


Figure 9: Achieving the requirements of MTI, preservation, equalization and enhancement on a cotton sweater.

from the brightness images, V, of the rug, hair and wool and cotton sweater images, respectively (shown in Figure 1 and 9). The histograms of the brightness are unimodal and coarsely fit a Gaussian distribution which is not surprising since the patches are assumed to be homogeneous. Each patch, however, has a distribution determined by its underlying texture. It is critical to note that ths Gaussian assumption is only used to employ the terms of mean and standard deviation and can be replaced by different approximations.

The MTIs of homogeneous patches should satisfy:

Preservation. The histogram of the brightness values of the MTI should *qualitatively* preserve the shape of the distribution of the brightness values.

Equalization. The mean of each meta-texture patch should be close to the means of other patches. Let MTI_{mean} denote the overall desired mean of the MTI.

Enhancement. The standard deviation of any patch should be close (not necessarily equal) to other patches. Let MTI_{sd} be the overall desired standard deviation of the MTI.

Let R_{mean}, R_{sd} denote the average and standard deviation of the histogram of the brightness values of pixels of P. Let P^v denote the brightness values of P. The MTI will be constructed as a grey-scale image that takes on values between $(V_{min}, V_{max}) = (0, 255)$. Let $MTI_{mean} = (V_{max} + V_{min})/2$ which places the mean brightness at the center of the range of brightness values. The brightness value, $P^v_{i,j}$, of a pixel (i, j),

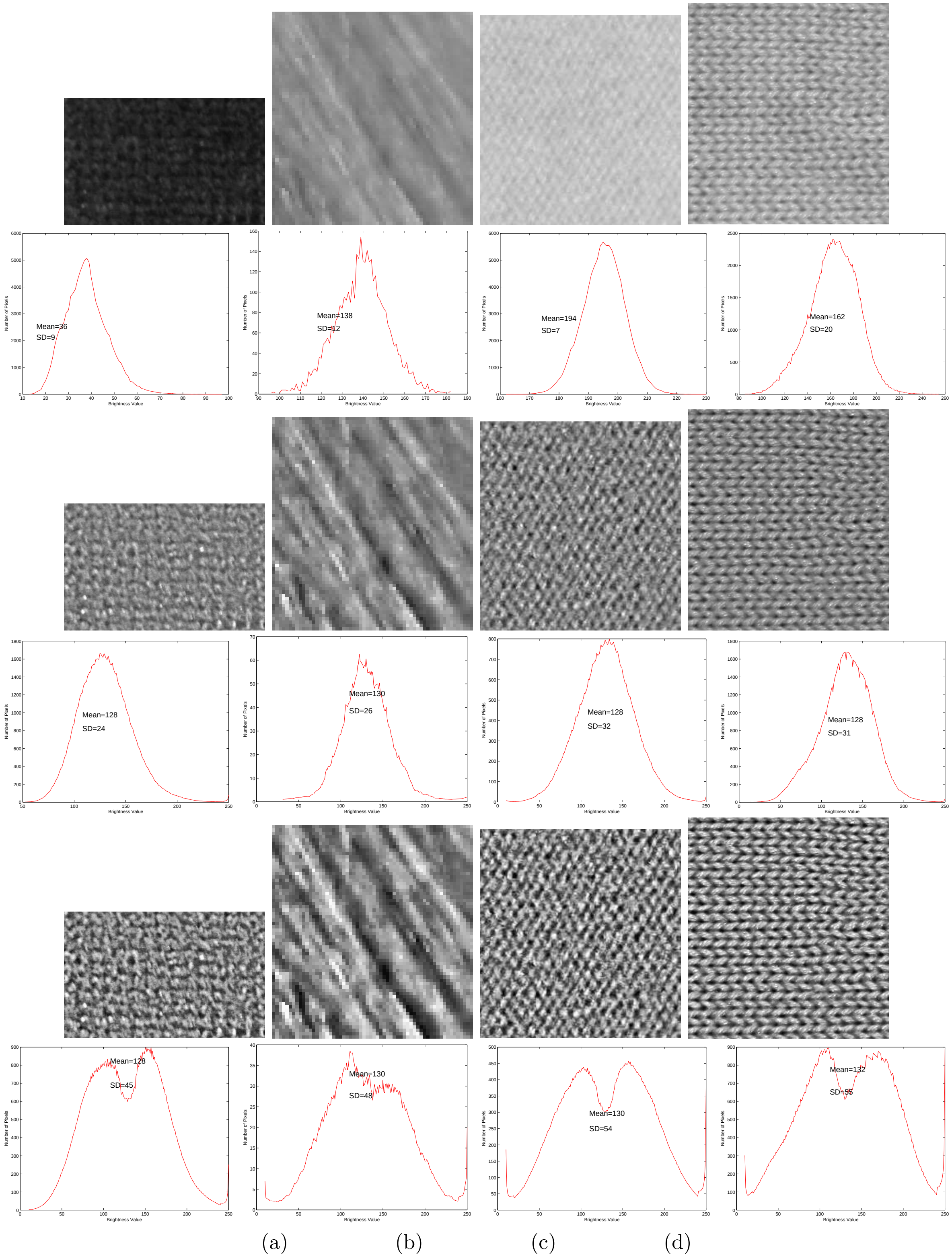


Figure 10: From top to bottom rows, (1) original brightness, (2) histogram of the brightness values, (3) MTI using $\gamma = 1$ (4) histogram of the brightness values for MTI $\gamma = 1$, (5) MTI using $\gamma = 0.8$ (6) histogram of the brightness values for MTI $\gamma = 0.8$. (a) rug (b) hair colored with different shades (c) a wool sweater and (d) cotton sweater.

is transformed into a value, $P_{i,j}^{MTI}$, in the MTI,

$$P_{i,j}^{MTI} = ((P_{i,j}^{v} - P_{i,j}^{D} - R_{mean})/(3R_{sd})) \cdot ((V_{max} - V_{min})/2) + MTI_{mean} \tag{8}$$

where $P_{i,j}^{D}$ is a floating point value of pixel (i,j) in the image P^D. P^D is the Gaussian smoothed deviation of the brightness values of pixels in P from the average brightness in the patch P,

$$P^D = \mathcal{S}(P^{v0}) \quad and \quad P_{i,j}^{v0} = P_{i,j}^{v} - R_{mean} \tag{9}$$

where $\mathcal{S}$ is a Gaussian smoothing function. Since we approximate the brightness histogram as a normal distribution, the area within $\pm 3R_{sd}$ approximately covers 99.73% of the points. It should be noted that this normal-distribution assumption is an approximation and points beyond $\pm 3R_{sd}$ are treated as being at $\pm 3R_{sd}$.

In Figure 10, the third and fourth rows show the transformation of the four different patches into a MTI using Equation 8 and the histograms of the values of the MTIs. The appearance of the underlying textures is preserved, enhanced and equalized. Quantitatively, the means and standard deviations of the MTIs are close and thus reflect roughness differences between these patches.

Equation 8 can be modified to enhance the contrast of the underlying brightness distribution by non-linear amplification of the brightness instead of the linear scaling. A gamma-correction model accentuates brightness differences within the patch,

$$P_{i,j}^{MTI} = sign(P_{i,j}^{G}) \cdot |P_{i,j}^{G}|^{\gamma} \cdot ((V_{max} - V_{min})/2) + MTI_{mean} \tag{10}$$

where $P_{i,j}^{G} = (P_{i,j}^{v} - P_{i,j}^{D} - R_{mean})/(3R_{sd})$, the function *sign* is 1 if $P_{i,j}^{G}$ is positive and -1 otherwise, and $|P_{i,j}^{G}|$ is the absolute value of $P_{i,j}^{G}$. Note that if $\gamma = 1$ then Equation 10 is equal to Equation 8. If $\gamma < 1$ then $P_{i,j}^{G}$ values closest to R_{mean} are amplified more than values far from R_{mean} and contrast is enhanced within the patch. In Figure 10, the last two rows show the images and histograms of brightness distributions for $\gamma = 0.8$. The enhancement in texture contrast is clearly visible. The graphs in Figure 10 (bottom row) show

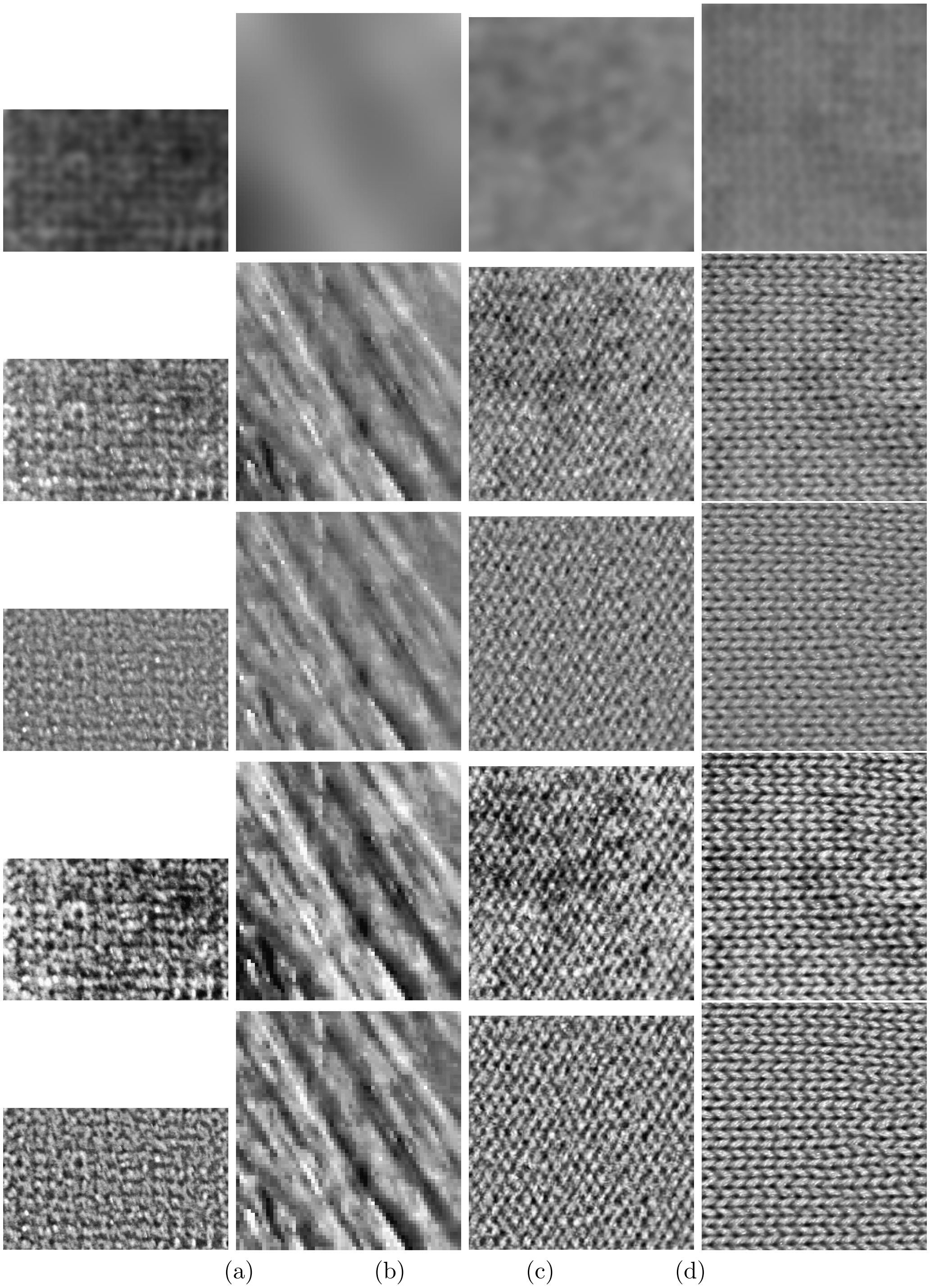


Figure 11: From top row to bottom, (1) the P^D image which is a smoothed version of the deviation from the average brightness of the patch (2) the MTI with $P^D = 0$ and $\gamma = 1$, (3) the MTI with P^D and $\gamma = 1$, (4) the MTI with $P^D = 0$ and $\gamma = 0.8$, (5) the MTI with P^D and $\gamma = 0.8$. (a) rug (b) hair colored with different shades (c) a wool sweater and (d) cotton sweater

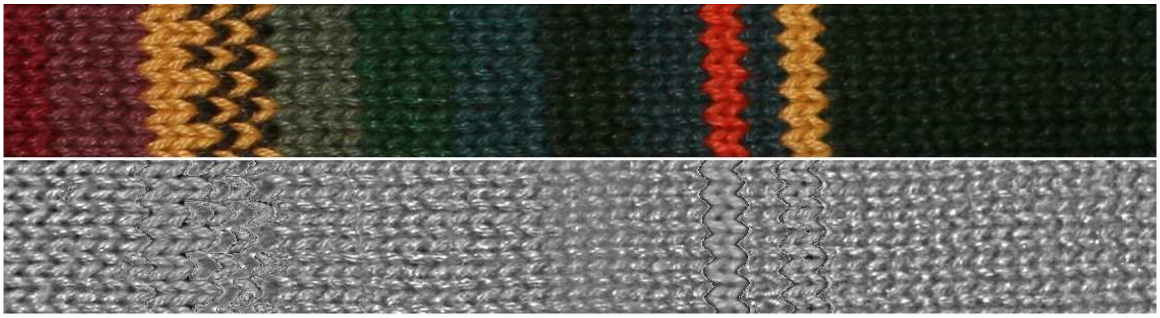


Figure 12: A colorful region in the sweater and its MTI.

that at the extreme low and high end of the brightness the number of pixels in the histogram have increased (the rising tails of both ends). This occurs since more points are shifted to the edges of the brightness range. Similarly, the mean area, which is the center of the transformation, lost points that have moved farther to increase the brightness contrast. It is worth noting that $\gamma = 0.8$ does not preserve the normal distribution (or even the unimodal distribution) but it provides greater discrimination power in texture analysis later on. If a smaller γ is used these phenomena increase and overall it is not a desirable outcome since it can excessively distort the histogram. If $\gamma > 1$ a suppression of the texture contrast occurs (not relevant to our research objective). In the rest of the paper we use $\gamma = 0.8$ in all experiments.

As defined in Equation 9, P^D is a Gaussian-smoothed image of the mean-subtracted brightness values. In our experiments we use a Gaussian smoothing filter of 41 units. P^D removes low-frequency brightness differences (e.g., smooth-surface shading) within a patch whether these were a result of material brightness or illumination differences. In Figure 11, top row, the value of P^D is shown as a grey-scale image although the actual values are positive and negative numbers. The second and fourth rows show the MTI images for $P^D = 0$ ($\gamma = 1$ and $\gamma = 0.8$, respectively), while the third and fifth rows show the MTI images using the actual P^D ($\gamma = 1$ and $\gamma = 0.8$, respectively).

Figure 12 shows MTI of an area taken from Figure 9. The MTI reflects appropriately the underlying salient weave of the sweater. Very small patches such as the black/yellow interweaved thread conform to the general texture pattern. The meshing of patch boundaries remains an open challenge (e.g., observe the boundaries of the red and yellow patches where small patches were detected between them and their neighbors). Boundary pixels are occasionally a blend of two colors, and depending on whether they are associated with one patch or another (or form a third patch) their meta-texture value will be different.

2.5 Evaluation of the MTI Transform

While ground truth MTIs of real scenes are not computable, we provide several evaluations of the MTI transform with respect to the stated objective of recovering mesostructure roughness. Specifically, we (1) compare the MTI transform to relevant image transformations, (2) demonstrate robustness to linear transformations, (3) evaluate quantitatively how different color regions are normalized, and (4) evaluate the texture variation within the MTIs with respect to normalizing across smooth-surface shading variations.

Comparative Performance. We compare MTI to histogram-equalization and low-frequency elimination using Fourier Transform with normalization. The latter approach eliminates the lower-frequencies and normalizes values. Figure 13 shows the original V channel patches in (a) columns, MTIs computed using Equation 10 in (b) columns, histogram equalized patches (c) columns, and low-frequency eliminated and normalized images in (d) columns. The MTIs achieve far more uniformity of texture appearance . Brightness variations due to illumination differences, material creases, or other anomalies are removed or reduced significantly so that the underlying texture appearance is uniform. Note that the low-frequency removal algorithm weakens the sharpness of the texture in (d) columns.

Robustness to Linear Transformations. Figure 14 illustrates the texture robustness to linear transformations. In the columns (top to bottom): the patch, its transformed MTI, histogram equalization transformation and low-frequency removed and normalized filtering shown for: (a) the given patch, (b) a constant is added to all brightness values, (c) values determined by a plane are added from the top of the image to the bottom (for a pixel (i, j), the added value is determined by $a * i + b * j + c$), (d) values determined by a plane are added only to the bottom half of the image, (e) values determined by a plane are added only to the bottom right quadrant. The boundary area in (d) and (e) have minor artifacts due to the discontinuities that are created in the data. The MTIs show constancy to these local and global linear transformations and the oversaturation in some patches. In contrast, histogram equalization does not remove the brightness

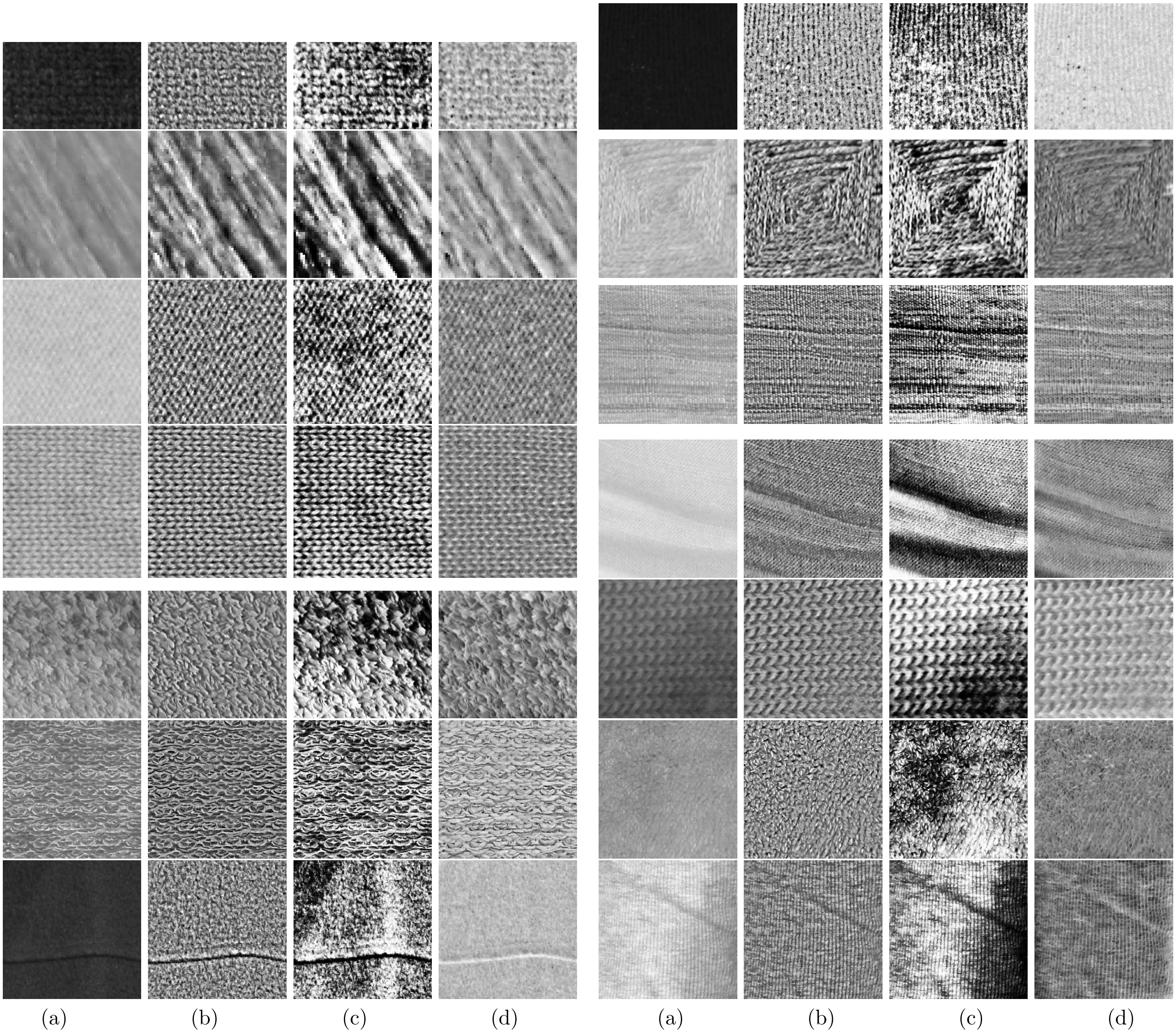


Figure 13: Columns show (a) the original V image, (b) the MTI of the patches using Equation 10, (c) the transformed patches using histogram equalization, (d) the low-frequency removal and normalized filtering.

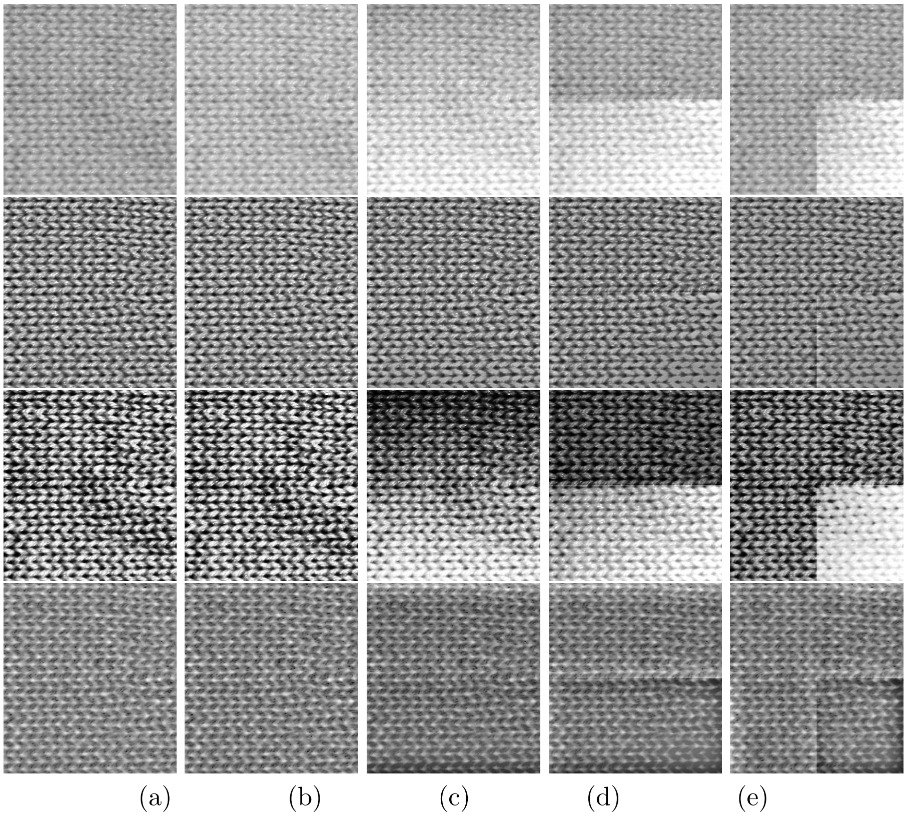


Figure 14: For each column the original, meta-texture, histogram equalized and low-frequency filtering are shown. Columns (a) original V image (b) scalar added to the image (c) linear plane in the vertical direction (d) linear plane added to the bottom half (e) a plane added to the bottom-right quadrant

variations and low-frequency removal filtering only partially copes with these variations and weakens the texture. Robustness of MTI to linear transformations can be shown mathematically but is omitted due to space constraints.

It is important to note that the linearly-transformed images of Figure 14 (c,d and e) violate the homogeneity assumption. Therefore, these are not typical images, but at the same time such images may be encountered due to gradual illumination changes that allow patches to increase their brightness diversity.

Inter-patch normalization. We next compare the similarity of MTIs of different patches in which mesostructure roughness is presumed to be equal. Two sets of five patches of different colors were selected from both the wool and cotton sweaters (Fig. 1 (c) and Fig. 9). Fig. 15 (bottom row) shows the histograms of the original brightness and the MTI values (left and right, respectively for each pair) for the wool ((a) and (b)) and cotton ((c) and (d)) sweater patches. An ideal transformation should bring each set of 5 patches to perfect alignment (assuming the original patches were indeed of identical mesostructure roughness). The MTI comes close to this ideal given likely intrinsic differences between the patches.

Intra-patch Normalization We next evaluate how well the MTI-transformation normalizes texture across smooth-surface shading variations within a homogeneous patch. An ideal MTI should remove all smooth-surface shading and preserve mesostructure roughness. We employ Principal Component Analysis (PCA) to measure the uniformity of the texture over the whole image. We associate uniformity with linear dependence within the data.

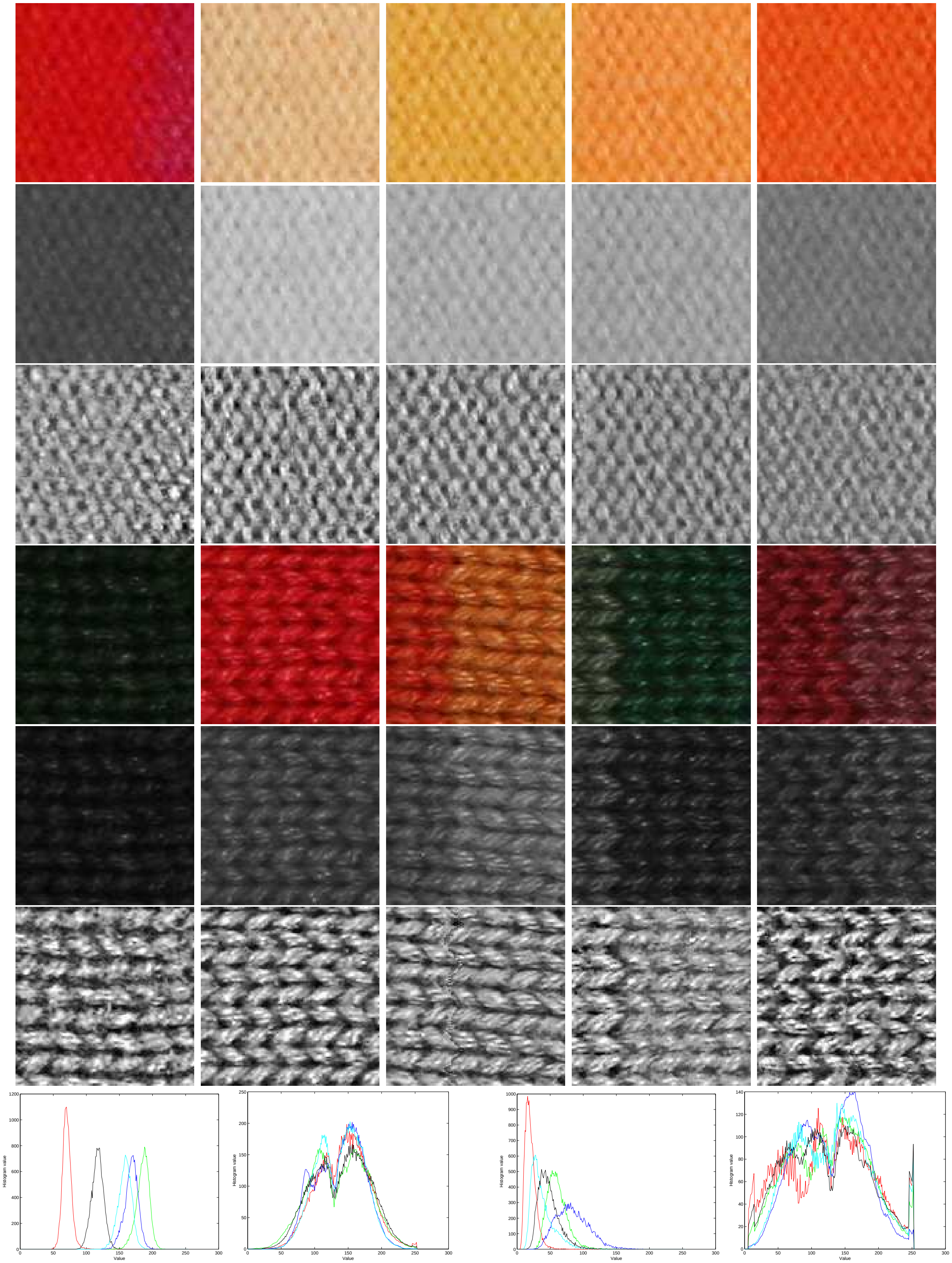
For an image I we compute its D Gabor-wavelet filtered images (9 scales with varied wavelengths and 6 orientations, resulting in $D = 54$). Performing PCA on all the pixels D-dimensional features provides a measure of information variability in I as a function of the values of the eigenvalues. It is worth noting that

- Since PCA is a linear representation, linearly dependent features do not change the subspace representation. As a result linearly-dependent (in the filter space) pixels do not penalize the representation.

- Mesostructure roughness appearance on a smooth-surface undergoes complex deformations (at a minimum a projective transformation but typically a higher-order deformation). As a result the Gabor-wavelet filtering representation will be affected and therefore the uniformity of texture is naturally penalized. The MTI does not compensate for these complex deformations.

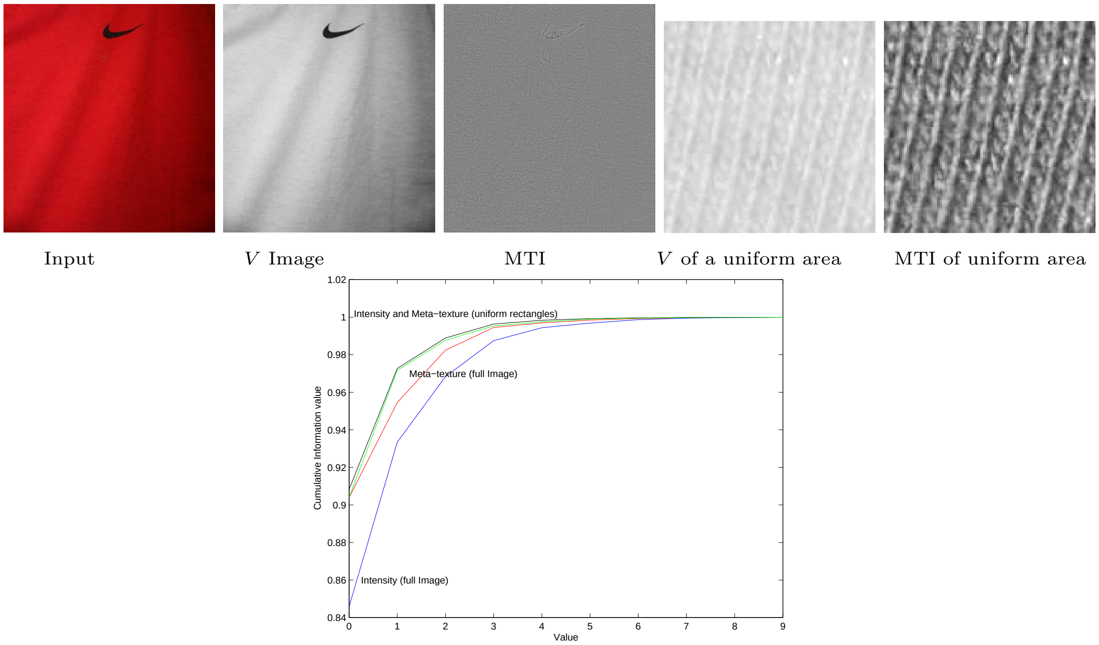
The amount of information captured by a set of eigenvectors is easily computed from the eigenvalues. The less the number of eigenvectors needed to capture 99% or more (for example) of the information the *simpler* and more uniform the representation and the more linearly dependent the data.

Figure 16 shows an input image of a T-shirt that conveys both smooth-surface deformations and mesostructure roughness. The second image (from the left) is the V image followed by the meta-texture image, a small uniform (texture-wise) region from V and its respective meta-texture region, and at the bottom is a graph for the cumulative information captured as a function of the number of eigenvectors. The Black and Green graphs show the cumulative information captured for the small uniform MTI and V regions, respectively. As each region is uniform, the graphs are almost identical, and in both cases the largest three eigenvalues



(a) (b) (c) (d)

Figure 15: The top three rows show the color, V and MTI for a different patches of wool material followed by similar patches for the cotton material. The bottom row show (a) histogram for the five wool patches, (b) histogram of the patches in the MTI (c) histogram for the five coton patches, and (d) histogram of the patches in the MTI.



Input V Image MTI V of a uniform area MTI of uniform area

cumulative information content

Figure 16: The top row shows the histogram of brightness values for 5 wool (left) and cotton (right) patches, the bottom graphs show the histograms for the MTI values.

capture about 99% of the variation in the data. The Red and Blue graphs show the respective information for the full MTI and V images, respectively. The Red graph is close to the Black graph, implying that the linear-dependence (i.e., information variability) in the large MTI image is nearly similar to that in the small MTI uniform region. This clearly suggests that the smooth-surface variations in the original image have been normalized quite effectively. In comparison, the Blue graph representing the V image shows more variability and therefore more eigenvectors are needed to capture variations in the data.

Another informative measure is the relationship between the eigenvectors of the image features. The angle between the first eigenvectors of the large and small MTI images is 4.2 degrees, suggesting a close alignment in the two subspaces. In contrast, the angle between the first eigenvectors of the large and small V images is 8.45 degrees suggesting a divergence between the two subspaces. The MTI and V angles if we consider the first three eigenvectors (summed up) are 60.75 and 74.51 degrees, respectively.

Note that an ideal MTI should lead to the Red graph being aligned with the Black graph and the angles being zero for the first and cumulative three angles. But such an MTI will have to flatten the mesostructure

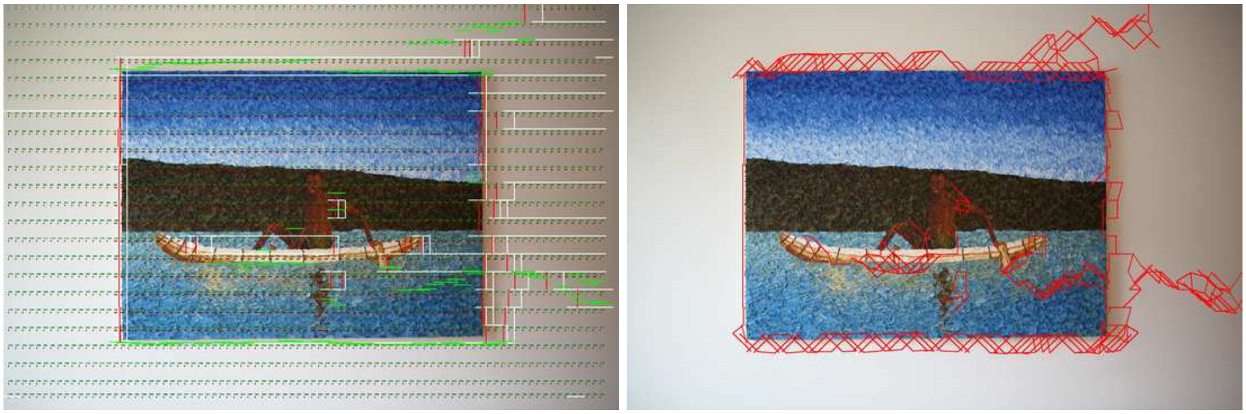


Figure 17: (Left Image) Illustration of the texture-boundary driven segmentation of an image. The dark green numbers denote the numeric ordering of 128x128 rectangles, the dark red denote their region label after the coarse texture analysis. The white lines show the coarse-level texture boundary. The red lines and green lines denote the optimal texture-boundary as estimated using the fine boundary detection process for vertical and horizontal boundaries, respectively. (right image) Illustration of the generation of regions after connecting proximate texture-boundary edges of the estimated texture boundary.

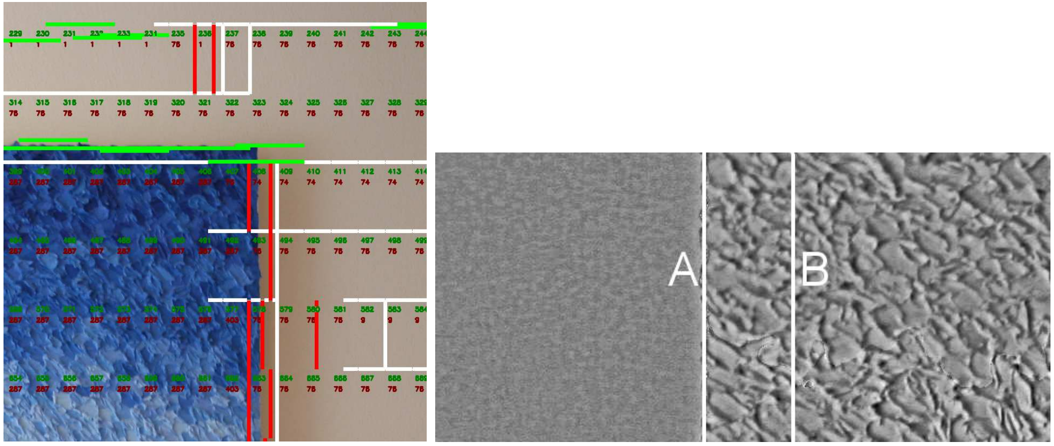


Figure 18: Left, zoom-in of top right corner of the painting of Figure 17. Right, the distance between the textures on the sides of A is larger than the distances on the sides of B.

roughness across the smooth-surface deformation as well as other natural variation in the material.

3 Segmentation of the Meta-texture Image

Given the high-resolution of our images, it is memory and computationally challenging to classify MTI pixels individually into separate patches. Therefore, we develop a window-based segmentation approach that reduces the memory and computation requirements to somewhat manageable level. Specifically, the texture features in the rest of the paper are all computed for NxN windows. Moreover, the search for texture boundaries is done using lines with length N at different locations and orientations in the image.

There are a variety of approaches for computing texture features [1, 4, 7, 13, 17, 24]. We adopt the Gabor wavelets method applied at square regions (details can be found in [17]). We re-pose the image segmentation problem as a *line-based* texture-boundary detection. We make a critical assumption that the true boundary between two different textures *coincides* with the maximum distance between texture windows on the sides of the boundary. In Figure 18 (right) it means that the distance between the texture windows on the sides of line A is larger than the distance between the textures on the sides of any other line such as B. The basis of this assumption is that the difference between the textures is maximized when the textures are as different from each other as possible.

This assumption can be observed in the boundary detection results expressed by the red and green lines in Figures 17 and 18(left) (left edge of the painting and the top right side edge of the painting). Figure 18(left) shows a zoom-in the region at the top right side of the painting shown in Figure 17. Note that some of the boundary lines are slightly displaced from the real boundary since boundary lines are computed only at pixels that are factors of 5 to reduce computation cost. However, there are situations in which the assumption is violated since textures are not normally as distinct as illustrated. In these cases the boundary may be slightly displaced with respect to the correct texture boundary. Note that since line-based boundaries are computed, inaccuracies and jagged boundaries appear at curved boundaries. The texture-boundary detection consists of the following:

Coarse texture representation. Given the MTI, we compute all NxN texture descriptors for a sliding square window (in raster scan). The vertical regions do not overlap, while the horizontal regions overlap by N/2. The Gabor wavelet features are computed for each NxN square rotated between ±75 degrees by 15 degrees steps to cope with the Gabor wavelet variance with respect to rotations.

Coarse texture-boundary detection. The evidence of a texture boundary between any two neighboring squares (both horizontal and vertical neighbors) is evaluated. If the minimum distance between their texture features (across orientations) exceeds T then a boundary is assumed to exist.

Fine texture-boundary detection. Since the boundary between textures is unlikely to fall at the edge

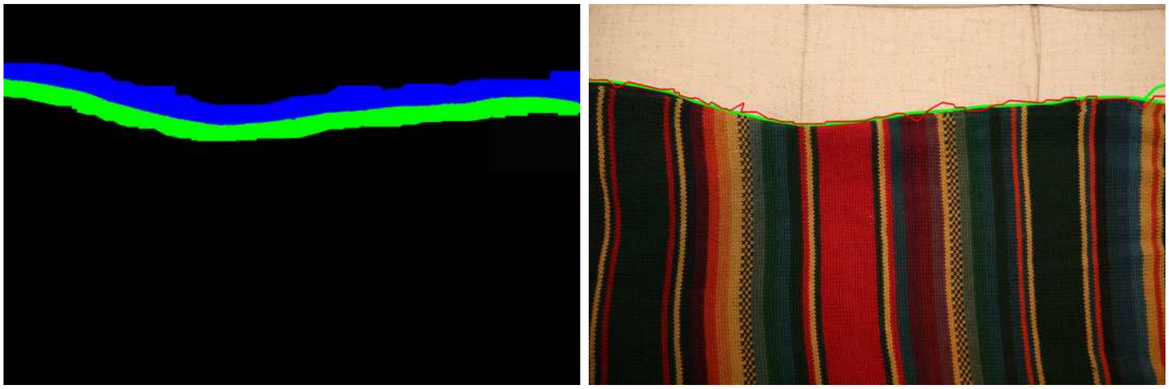


Figure 19: (Left Image), the classification of boundary pixels into two classes using EM, (Right Image) The red contour shows the initial jagged boundary and the Green contour shows the refined boundary using the EM algorithm.

between adjacent squares it is necessary to estimate the optimal line-boundary *location* and *orientation*. The optimal boundary is the one that maximizes the distance between the textures on its sides even when all image locations and orientations are considered. A scan of all possible boundary locations and orientations (±45) is done, where for each candidate boundary the minimum distance between the textures on the two sides of the candidate boundary is computed, and the maximum of these minima is selected as the optimal boundary. To reduce computations, this process is done raster-wise skipping N-pixel rows for vertical line boundaries and N columns for horizontal line boundaries.

Figure 17 shows the texture-boundary segmentation process on a sample image. The dark green numbers denote the numeric order of 128x128 pixel squares (in raster scan). The dark red denote the texture label of each of these squares as estimated by the coarse texture analysis. These numbers already identify and cluster many squares into uniform-texture regions that form the basis for the fine-level texture-boundary detection. The white lines show the texture boundaries (i.e., the square boundaries where the neighboring squares have different labels) as computed in the coarse-level process. Note that some of the white lines were overwritten by green and red lines since the fine and coarse processes of texture-boundary selection led to the same locations in the image (e.g., lower left side of the painting happened to coincide with the boundary of coarse squares). At the same time, some transition squares (e.g., the top edge the painting) led to detection of two coarse-level texture boundaries reflected by the two white horizontal lines apparent along the top edge of the painting. The red and green lines denote the optimal texture-boundary as estimated by the fine texture-boundary detection process for vertical and horizontal boundaries, respectively. These lines typically trace the correct texture-boundary of the painting. However, in some cases these lines are slightly off the optimal location since the texture-feature distance turned out to be higher at these locations than at the optimal location.

Patch Mergers. The texture-boundary detection is followed by delineation of patches of uniform texture and merger of patches that have similar textures. Delineation of patches is achieved by connecting proximate line texture boundaries. As a result some small patches may be created as shown in Figure 17, these patches have their texture feature recomputed. Two patches are merged if the distance between their textures is less or equal than $T + 3$.

Refined texture-boundary detection. We next employ a machine-learning classifier to refine the lines-based boundary between regions based on each pixel features and reduce the jaggedness of boundaries. For an image I we compute its D Gabor-wavelet filtered images (9 mixed scales/wavelengths and 6 orientations, resulting in $D = 54$). For each pixel, the D Gabor-wavelet responses plus the R, G, B are joined into a $D+3$ dimensional feature-vector. Given a boundary, B_{12} between two patches R_1, R_2, we create the set of all points within a distance N from the boundary B_{12}. This set of points is re-labeled so that each pixel belongs to either R_1 or R_2. We employ an Expectation-Maximization approach to estimate the parameters of the multivariate probability density function of the Gaussian mixture distribution with a specified constraint that the data represents two distinct sets.

Figure 19 demonstrates the EM-based classification of pixels on a sample image. The original jagged boundary is shown by the Red contour (image on the right). The Green contour shows the boundary as computed by the refined boundary process. The image on the left shows the classification of the pixels within and distance N from the initial boundary into two labels.

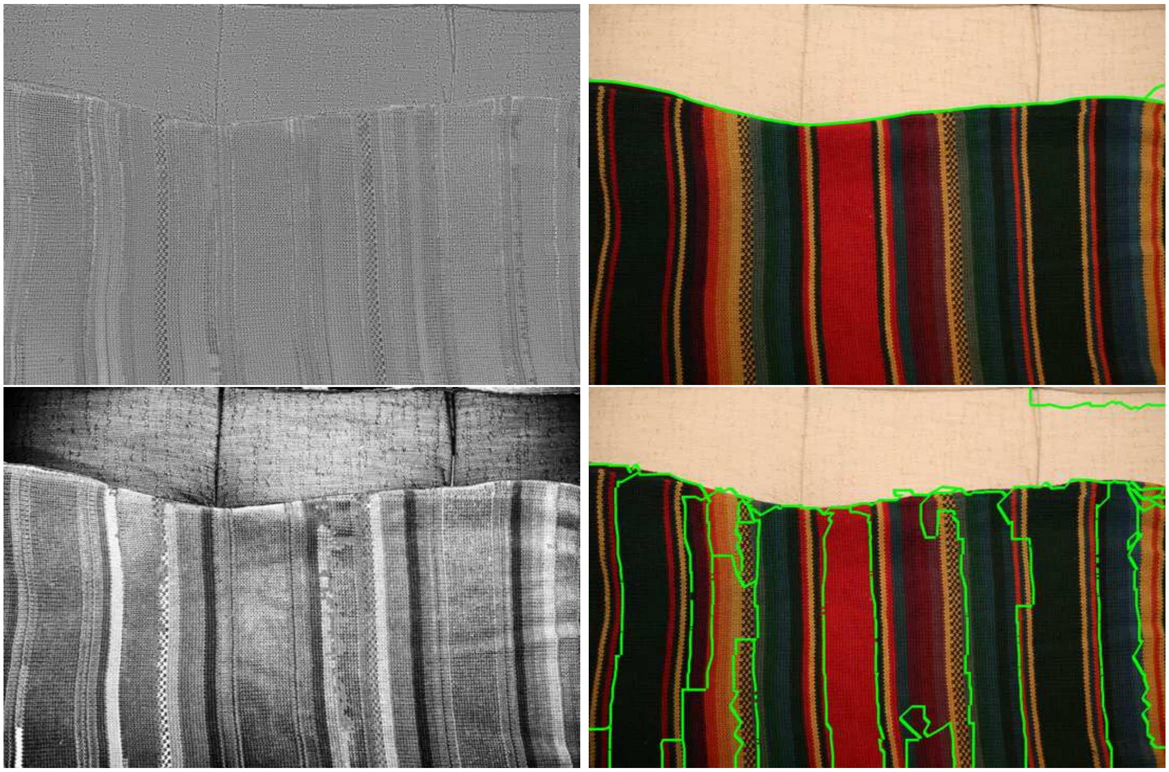


Figure 20: Left Column, the MTI normalization and the histogram equalization (applied at the same patches), respectively. Right column, segmented images using MTI and histogram equalization, respectively.

4 Results

Two parameters need to be set in the image segmentation approach: the window size for texture feature extraction is $N = 128$ and the distance threshold between texture features determines whether two texture regions should have a line-boundary detected, $T = 7$ ($N = 64$ and $T = 6$ for the lower resolution images).

Figure 20 shows a comparison between MTI-based and histogram equalization based segmentations (top row and bottom row, respectively). All parameters were kept the same and the only difference is using the MTI normalization described in Equation 10 versus histogram equalization applied to the same patches. The left column shows the texture images and the right column shows the segmentation results.

Tables 1,2 show MTIs (left column), segmentation using MTI (second column from left), and the output of three segmentation algorithms, Cour et al [6], Martin et al [18] and Arbelaez [2] (top to bottom, respectively, two right-side columns). The MTI images show in Red contours a manual texture segmentation of the image according to the boundaries between mesostructure roughness patches. These Red contours are considered the groundtruth in segmentation in the rest of section (despite the obvious subjectivity). In Table 1, the top image shows hair with highlights and shaded regions. The MTI clearly reveals the hair as having a uniform texture and accentuates the texture of the fleece. The cotton sweater, rug and painting show near optimal segmentation (including texture differences within the wood). The multi-colored pillow is in front of the textured painting and to its left the cream color fabric has visible texture while the other colorful pillow region (on the left edge) has no meta-texture visible at this resolution. The segmentation result is reasonable as the pillow, painting, wall and cream-color fabric are properly segmented. Finally, three outdoor images show appropriate segmentation of a lion, hippopotamus, and vegetation where surface detail is visible. The reflection of the hippopotamus in the still water appears similar to its mesostructure roughness and therefore it was merged with the hippopotamus. The vegetation is segmented by the size of leaves not their color.

In Table 2, the roughness of the sweaters is sufficient for segmentation. Parts of the background scenes lack surface detail so the segmentation typically follows color edges (except the carpet in the third image where mesostructure roughness is visible). Parts of the sweater are not merged properly since the 3D creases of the texture are not accounted for by our texture analysis algorithm. The second image from the top is notable since the background and foreground share the same colors, and the top of the pants at the waist is separated despite having the same white color.

The experiments suggest that MTI-segmentation provides better results than the three approaches [2, 6, 18] since mesostructure roughness saliency transcends edges and color attributes which lead to over-segmentation in these approaches. Our approach suffers from inaccurate boundary detection brought about by using a single cue, line-boundary constraint and sparse boundary search. These are not core issues and can be remedied in future research. The absence of roughness detail leads to reflecting differences in color and as a result the segmentation suffers.

Figure 21 shows segmentation results of images taken from the world-wide-web. The resolution of these images is in the range of 2-5MPixels, and a range of image quality can be observed. The results are short of human segmentation but are reasonable considering only mesostructure roughness is used and no examples

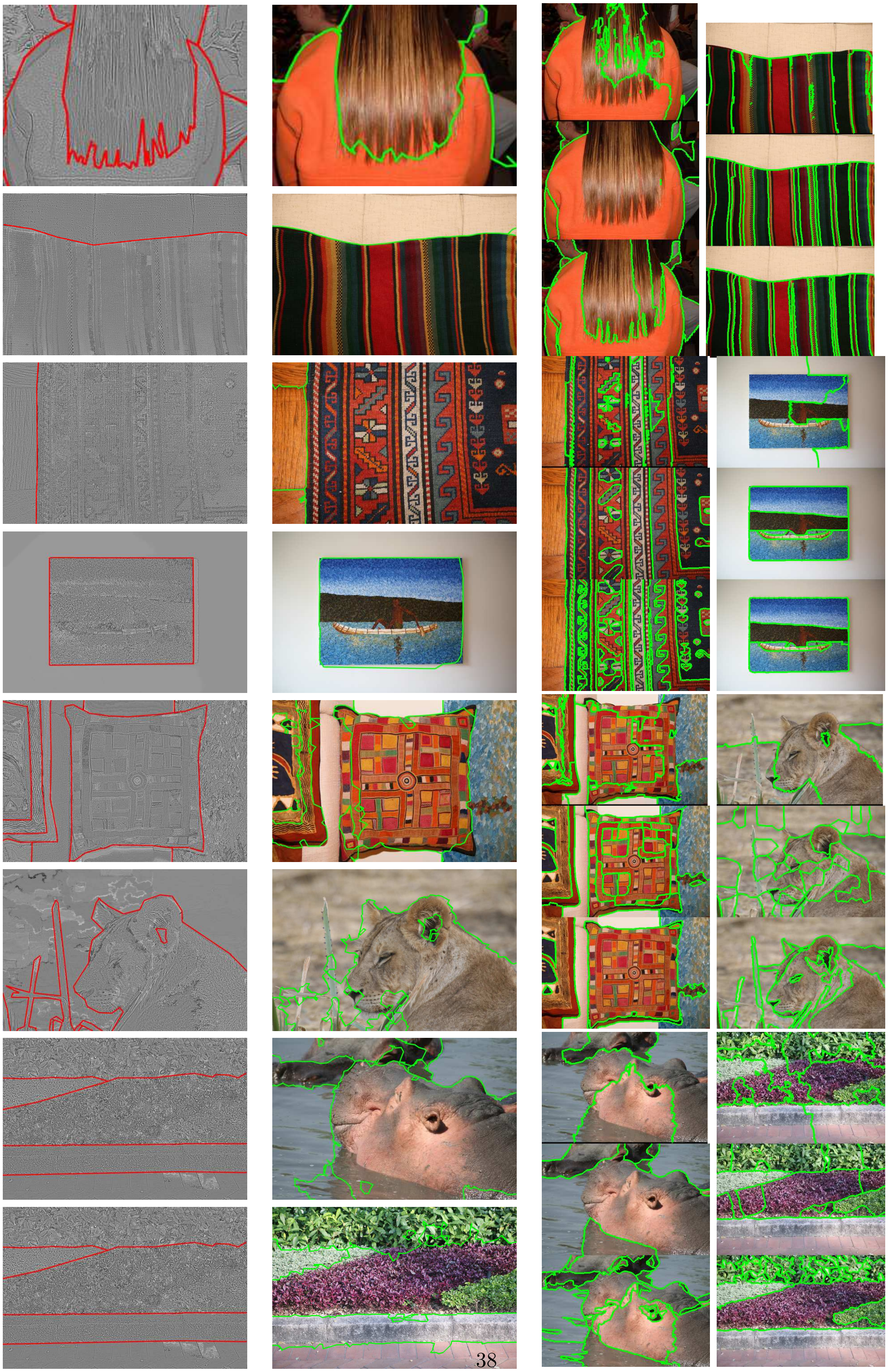


Table 1: MTI (Left) with manual texture segmentation shown by Red contours, segmentation of MTI (middle) and segmentation using [6] [18] [2], top to bottom respectively.

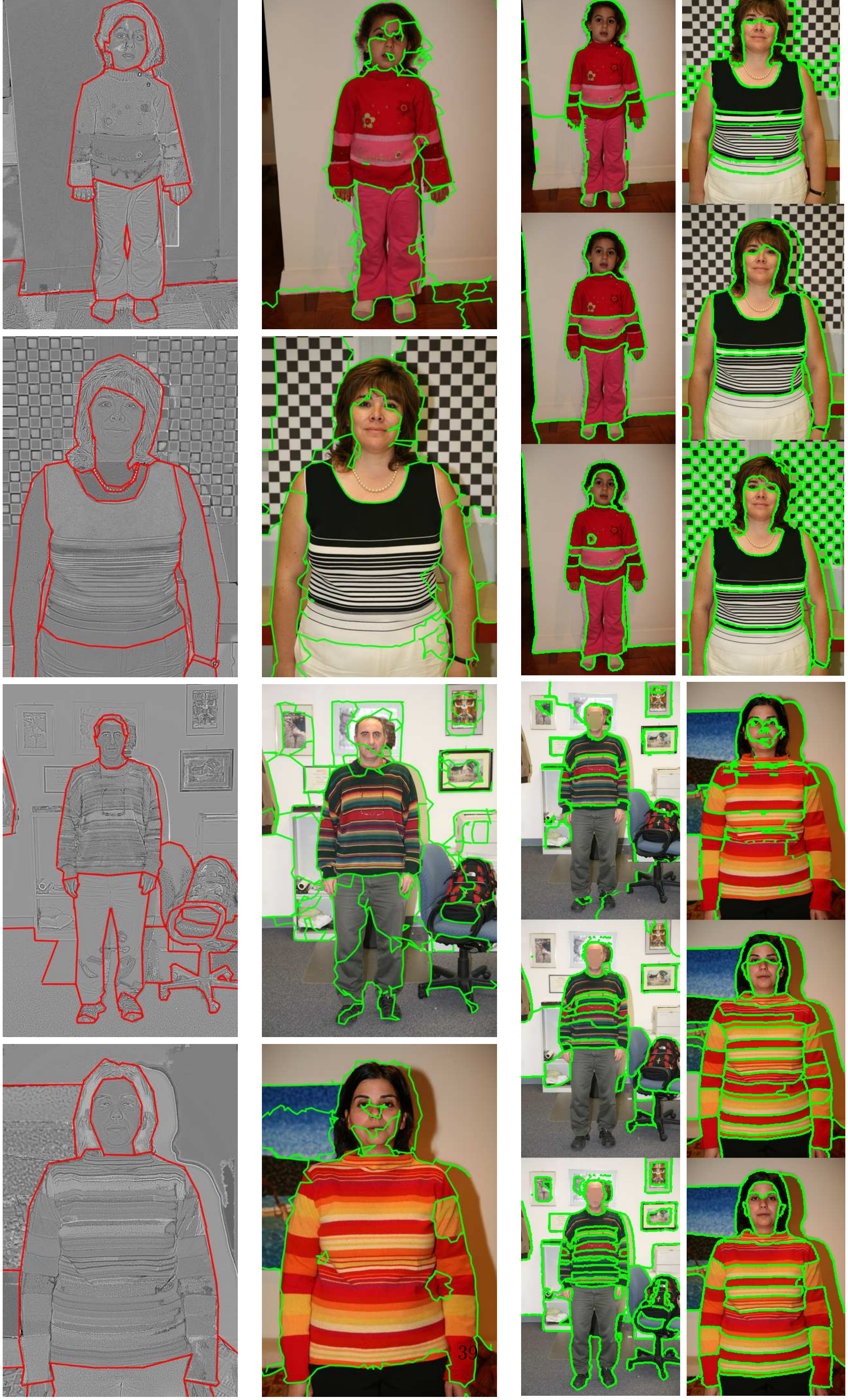


Table 2: MTI (Left) with manual texture segmentation shown by Red contours, segmentation of MTI (middle) and segmentation using [6] [18] [2], top to bottom respectively.

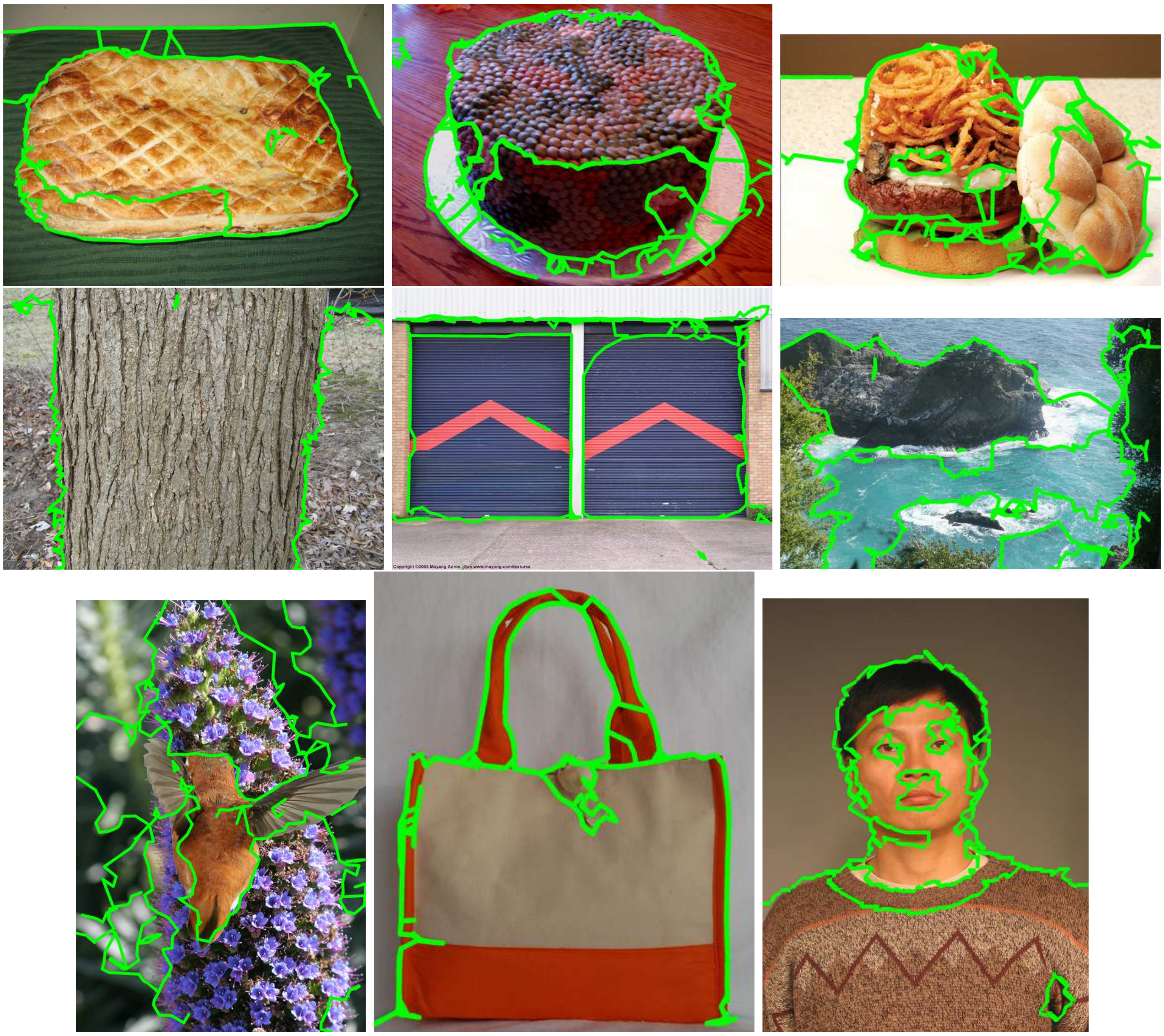


Figure 21: MTI-based segmentation examples of images taken from the WWW.

have been involved. The key-point to observe is that surfaces with different colors or illumination are segmented into a single patch if their underlying mesostructure roughness is consistent. The three food items have varying reflectance but mostly the same mesostructure roughness. The garage doors are segmented despite the red stripes and are separated from the top metal surface, bricks and concrete surfaces. The hand bag is segmented as one object, the two outdoor scenes and the person image are reasonably segmented as well.

The algorithm we developed has few parameters that were fixed for all experiments. Given the diversity and complexity of these images it is possible to improve the results of each image with tweaking of the these parameters as well as other implementation choices. The texture-boundary detection assumptions and process are relatively simplistic and the search for line-boundaries equal in length to the window size leads to jagged boundaries that at times are not correctable by the machine-learning 2-class labeling algorithm.

5 Discussion

This paper set out to study the utility of mesostructure roughness as a cue in image analysis. Mesostructure roughness has become easier to capture with the increase of image resolution. We argue that mesostructure roughness holds a great potential for image understanding, perhaps equivalent to the roles of edges and color analysis.

In contrast to earlier work on mesostructure roughness (e.g., BTF, 3D-textons, physics-based, etc.) we aimed to handle real-world scenes and employ as little priors as possible. As a result we demonstrated the performance on a wide range of images and quite favorably compared our performance to three state-of-the-art algorithms. Although different cues were used in these approaches, the experiments illustrate the potential of the mesostructure roughness cue.

We chose the intrinsic image analysis as the basis for our formulation. Perhaps for the first time we showed that the intrinsic image analysis provides a bridge between image formation and a high-level objective such as image segmentation. Considerable effort needs to be performed to improve the computation of the intrinsic images as well as integrating them into fundamental image understanding problems such as segmentation, detection, recognition, image synthesis, etc.

Given the well-known underconstrained formulation for intrinsic image estimation we pursued a new direction by focusing on patches. We argued that dividing the image into reflectance-homogeneous patches simplifies the process of deriving three meaningful images, smooth-surface shading, mesostructure roughness shading and reflectance. Although we proposed a method for computing these images, we are aware of the opportunity and need to improve this estimation (as we discussed in section 2). We also suggested that since infinite solutions exist for the intrinsic image analysis, those solutions that can be shown to derive *useful* information with respect to image understanding objectives are perhaps the best.

Since patches have arbitrary shape, the use of spatial frequency analysis to derive the mesostructure roughness cue as an MTI image is not possible. Instead, a transformation of an image into MTI was proposed by posing the constraints: preservation, equalization and enhancement of meta-texture properties of patches across color, brightness or illumination differences.

The MTI was used as input to a texture boundary segmentation process. Due to computational complexity window-based texture features were computed and a line-based boundary search was conducted. Given that our images are about two orders of magnitude larger than those typically used in other segmentation algorithms (one order in each spatial dimension), the computational tools (e.g., graph-cuts, machine-learning, etc.) are computationally challenged and we have not been able to adapt them to our data.

Experiments on realistic scenes were reported with respect to three state-of-the-art algorithms. While our results are better than these algorithms, they suffer from jagged and inaccurate line-boundaries since we used only mesostructure roughness (as opposed to edges and color and texture) and we approximated the boundary by lines and performed sparse computation to reduce complexity.

There are weak points in our mesostructure roughness computation approach that require further research (the texture-based segmentation approach is not considered a critical part of the approach).

- The pre-segmentation of an image into reflectance-homogeneous patches is a critical step and can affect the appearance of the MTI. Although we used fixed parameters throughout our work, these parameters occasionally lead to over or under segmentation of patches. While we have a classic segmentation problem (i.e., illusive optimal solution), the main objective is to avoid breaking patches due to significant smooth-surface shading changes.

- Images with a smoothly varying reflectance (e.g., tie-dye patterns) are challenging since reflectance gradual variations make it difficult to pre-segment an image into reflectance-homogeneous patches.
- The proposed approach does not account for specularities due to smooth-surface or mesostructure roughness. As result the patch-segmentation leads to oversegmentation, as well as normalization errors.
- The normalization that leads to MTI is not currently sensitive to intensity range. As a result, near-constant intensity patches are transformed into noise-amplified patches. Such patches could be identified and normalized differently to avoid creating roughness that is purely due to the imaging sensor.
- Patch boundaries are subject to different normalizations and as a result the true roughness is corrupted at these boundaries. A process of equalization across patch boundaries can preserve roughness across patches.
- Blur in part of the image is common due to limited depth-of-field imaging. It eliminates the appearance of mesostructure roughness. It is necessary to redesign our approach to detect and cope with such blur.

6 Acknowledgements

This work has been supported in part by the US Office of Navy Research (ONR), grant N000140610102.

References

[1] E.H. Adelson, Y. Li and L. Sharan. Image statistics for material perception. *Journal of Vision,* Volume 4, Number 8, Abstract 123, 2004.

[2] P. Arbelaez. Boundary Extraction in Natural Images Using Ultrametric Contour Maps. Workshop on Perceptual Organization in Comp. Vision, 2006.

[3] H.G. Barrow and J.M. Tenenbaum. Recovering Intrinsic Scene Characteristics from Images. *Compuer Vision Systems*, A. Hanson and E. Riseman, eds., Academic Press, 1978, 3-26.

[4] J. Chen, T. N. Pappas, A. Mojsilovic, B. E. Rogowitz. Adaptive perceptual color-texture image segmentation, *Tr. Im. Processing*, 14, 2005, 1524-1536.

[5] T. Chen, M. Goesele, H.P. Seidel, Mesostructure from Specularity. *CVPR* 2006, 1825- 1832.

[6] T. Cour, F. Benezit, J. Shi. Spectral Segmentation with Multiscale Graph Decomposition. CVPR, 2005, 1124-1131.

[7] O. Cula and K.J. Dana, 3D Texture recognition using bidirectional feature histograms, *IJCV*, (59)1, 2004, 33-60.

[8] K.J. Dana, B. van Ginneken, S.K. Nayar, and J.J. Koenderink. Reflectance and texture of real world surfaces. *ACM Transactions on Graphics*, 18(1), 1999, 1-34.

[9] G.D. Finlayson, M.S. Drew, and C. Lu. Intrinsic Images by Entropy Minimization. *ECCV* 2004, 582-595.

[10] T. Horprasert, D. Harwood, and L.S. Davis, A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection, ICCV FRAME-RATE Workshop, 1999.

[11] J.J. Koenderink, A.J. van Doorn, Illuminance Texture Due to Surface Mesostructure, *JOSA-A* (13), No. 3, March 1996, 452-463.

[12] E.H. Land and J.J. McCann, Lightness and Retinex Theory. *J. of Optical Soc. Am.*, 61, 1971, 1-11.

[13] T. Leung J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *IJCV*, 43(1),2001.

[14] X. Liu, Y. Yu, and H. Shum, Synthesizing bidirectional texture functions for real-world surfaces, *SIGGRAPH*, 2001, 97-106.

[15] A.J. Lundberg, L.B. Wolff, D.A. Socolinsky. New perspectives on geometric reflection theory from rough surfaces *ICCV 2001.* Vol. 1, 225-232.

[16] T. Maenpaa and M. Pietikainen. Classification with color and texture: jointly or separately? *Pattern Recognition*, 37(8):1629-1640, 2004.

[17] B.S. Manjunath and W.Y. Ma. Texture features for browsing and retrieval of image data. *PAMI*, (18)8, 1996, 837-842.

[18] D. Martin, C. Fowlkes, J. Malik. Learning to Detect Natural Image Boundaries Using Local Brightness, Color and Texture Cues, TPAMI 26 (5) 530-549.

[19] M. Oren and S.K. Nayar, Generalization of the Lambertian Model and Implications for Machine Vision. *Int'l. J. Comp. Vis.*, vol. 14, no. 3, 1995, 227-251.

[20] H. Ragheb and E.R. Hancock. Estimating Surface Characteristics using Physical Reflectance Models *CVPR*, 2003, 177-184.

[21] M.F. Tappen, W.T. Freeman, and E.H. Adelson. Recovering Intrinsic Images from a Single Image. *TPAMI*, (27)9, 2005, 1459-1472.

[22] M.F. Tappen, W.T. Freeman, and E.H. Adelson. Estimating Intrinsic Component Images using Non-Linear Regression. *CVPR*, 2006, 1992-1999.

[23] K. Torrance and E. Sparrow. Theory for Off-Specular Refection from Roughened Surfaces. *J. Opt. Soc. Am.*, vol. 57, 1967, 1105-1114.

[24] M. Varma and A. Zisserman. Classifying images of materials: Achieving viewpoint and illumination independence. *ECCV*, 2002, 3, 255–271.