

Band-Sifting Decomposition for Image-Based Material Editing

IVAYLO BOYADZHIEV and KAVITA BALA

Cornell University

SYLVAIN PARIS

Adobe

and

EDWARD ADELSON

MIT

Photographers often “prep” their subjects to achieve various effects; for example, toning down overly shiny skin, covering blotches, etc. Making such adjustments digitally after a shoot is possible, but difficult without good tools and good skills. Making such adjustments to video footage is harder still. We describe and study a set of 2D image operations, based on multiscale image analysis, that are easy and straightforward and that can consistently modify perceived material properties. These operators first build a subband decomposition of the image and then selectively modify the coefficients within the subbands. We call this selection process *band sifting*.

We show that different siftings of the coefficients can be used to modify the appearance of properties such as gloss, smoothness, pigmentation, or weathering. The band-sifting operators have particularly striking effects when applied to faces; they can provide “knobs” to make a face look wetter or drier, younger or older, and with heavy or light variation in pigmentation. Through user studies, we identify a set of operators that yield consistent subjective effects for a variety of materials and scenes. We demonstrate that these operators are also useful for processing video sequences.

Categories and Subject Descriptors: I.3.7 [**Computer Graphics**]: Imaging and Video - Image Processing—*Material editing*

General Terms: Algorithms, Design, Human Factors

Additional Key Words and Phrases: Image-based material editing, multiscale decomposition

ACM Reference Format:

Ivaylo Boyadzhiev, Kavita Bala, Sylvain Paris, and Edward Adelson. 2015. Band-sifting decomposition for image-based material editing. ACM Trans. Graph. 34, 5, Article 163 (October 2015), 16 pages.

DOI: <http://dx.doi.org/10.1145/2809796>

1. INTRODUCTION

Photographers care a great deal about the surface appearance of objects they photograph; indeed, much of the craft of traditional

Authors' addresses: I. Boyadzhiev (corresponding author), K. Bala, Cornell University, Ithaca, NY 14850; email: iboy@cs.cornell.edu; S. Paris, Adobe; E. Adelson, MIT, 77 Massachusetts Ave, Cambridge, MA 02139.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

2015 Copyright is held by the author/owner(s).

0730-0301/2015/10-ART163

DOI: <http://dx.doi.org/10.1145/2809796>

photography involves controlling material appearance using physical techniques. Portrait photographers control the appearance of skin wrinkles by adjusting the lighting, and apply makeup to hide variation in skin color (e.g., blemishes or mottling) and powder to make skin appear less shiny. In product photography, dulling spray is used to reduce specular highlights, while in food photography, where specularity may be desirable, a glycerine spray may be used to make the food look fresher or juicier.

Such adjustments can be performed digitally after the photo was taken, rather than physically during the photo session, which greatly simplifies the process and enables more control on the result. However, altering the appearance of material properties such as wetness, gloss, wrinkles, or mottled coloration remains a tedious task that requires advanced skills that are beyond the reach of casual users. Further, in the case of video, laborious manual retouching is simply impractical; not only are there multiple frames, but it is difficult to get the effects to align and adjust across the sequence without introducing temporal artifacts.

An alternative route to manipulating material appearance is to build a fully renderable 3D description of the scene and to change the physical parameters as needed. For example, multi-image capture with multiple cameras and light sources can create a highly detailed model of an object. However, this acquisition pipeline is not helpful to a casual or professional photographer working with a single image from an ordinary camera.

Our goal is to work with an ordinary photograph, and to allow the photographer to alter the appearance of a 3D surface without using a 3D representation. By using 2D image operations, we gain both speed and simplicity. We want our image-based technique to accept many kinds of source images as input, and to avoid the errors that can arise when attempting a full 3D scene analysis. At the same time, we must recognize that 2D operations can be limited and work best when there is a straightforward mapping between 3D surface properties and 2D image properties.

This article explores a space of image operators that can be used to modify a variety of visual surface properties. The operators decompose an image by first applying a series of splitting operations based on frequency, amplitude, and sign, and then *sifting* through these decompositions and recombining them to compose a new image. This sifting operation can create a range of visual effects that affect perceived material properties, for example, by changing perceived shininess/gloss, aging/weathering, and glow. Figure 1 shows an example of one such sifting procedure for two example images.

By selectively modifying coefficients in different subsets, one can achieve a variety of distinct image operators that we call *sifting operators*. While some of these are known and well studied (for instance, increasing high-frequency coefficients enhances detail in an image), the combination of the several criteria has not been

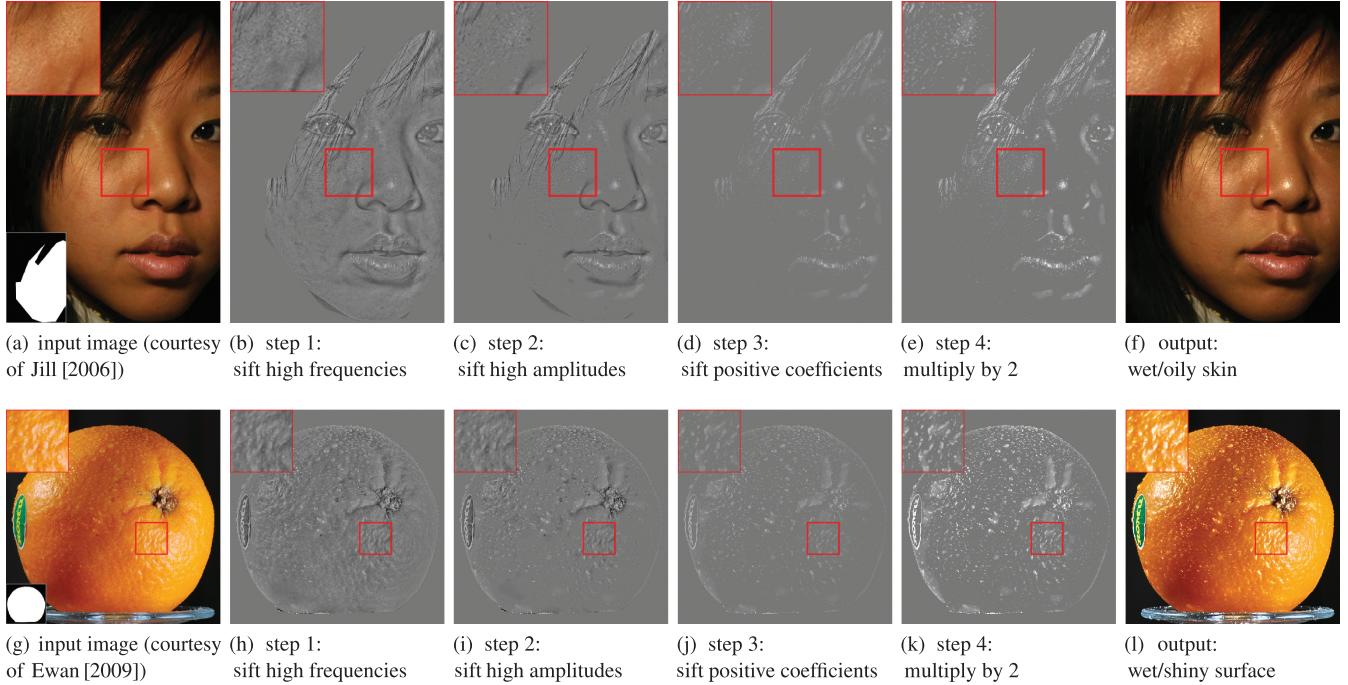


Fig. 1. Starting from a single input image and a mask (a), we selectively manipulate the subband coefficients of the luminosity channel by sifting them through a cascade of decisions based on the scale, amplitude, and sign of the coefficients. Here we show one of these decision paths. First, we sift the high-spatial frequencies from the low-spatial frequencies (b). Then we sift the high amplitudes from the low amplitudes (c), and finally we sift the positive from the negative coefficients (d). Multiplying the sifted coefficients (e), adding them back, and reconstructing the image gives the skin a more oily or wet look (f). In the second row, we show that a similar perceptual effect is achieved on a nonface object, where the orange is given a more shiny or wet look. We found that sifting subband coefficients allows to produce a variety of physically plausible effects that lead to perceptually consistent modifications across a variety of scenes.

explored so far. This article seeks to fill in this gap and focuses on aspects related to material perception in particular. That is, our goal is to characterize which band-sifting operators generate physically plausible change that lead to perceptually consistent effects. For example, in Figure 1 we show that the same band-sifting operator makes both the human skin (row 1) and the orange surface (row 2) look more shiny and wet.

We explored the space of band-sifting operators and found that, depending on the selected coefficients, operators modify properties such as the material shininess or its degree of weathering. On faces, the effects were particularly interesting with, for instance, variations in the appearance of oiliness, glow, wrinkles, and pigmentation of the skin. Figure 2 shows examples of these various effects. As expected, we also observed that applying a modification too strongly yields unnatural-looking results. This motivated two user studies. First, for each band-sifting operator, we characterized how strongly it can be applied before producing an unnatural look. And second, we studied how human observers describe the effect of each operator. This allowed us to isolate a subset of band-sifting operators that produce consistent effects across images. Finally, we also demonstrate the use of the band-sifting operators on videos. They are stable enough to achieve temporally coherent results without additional processing, they are fast enough to run at interactive rate, and they naturally “follow the scene content” without the need to estimate the optical flow explicitly. As an example, in a video of someone talking, we can add some glow onto them or make them look sweaty by simply applying our operators frame by frame, which is both simple and efficient.

Contributions. In this article we introduce a new approach to image-based material editing based on multiscale image decomposition and sifting.

—We propose and study a space of band-sifting operators that act along several criteria at the same time, based on scale, amplitude, and sign.

—We study the perceptual effects of our band-sifting operators; we validate their perceptual consistency through user studies; and we demonstrate their usefulness for both image and video post-process material editing.

1.1 Related Work

Image decomposition. Splitting an image into components is a standard strategy to manipulate some properties independently of others. For instance, one can convert RGB colors into YIQ or CIE-Lab to edit luminance and chrominance independently. The coring operation used for denoising drives the low-amplitude coefficients of a multiscale decomposition towards zero without changing the high-amplitude coefficients [Donoho 1995; Simoncelli and Adelson 1996]. The classic Retinex algorithm by Land and McCann [1971] use a similar amplitude threshold in the gradient domain to separate the illumination from the reflectance of a scene. Malic et al. [2006] describe a technique for separating specular and diffuse reflection components in images and videos. Durand and Dorsey [2002] separate large-scale variations from the small-scale ones for the purpose of HDR tone mapping, Bae et al. [2006] rely on a similar split for

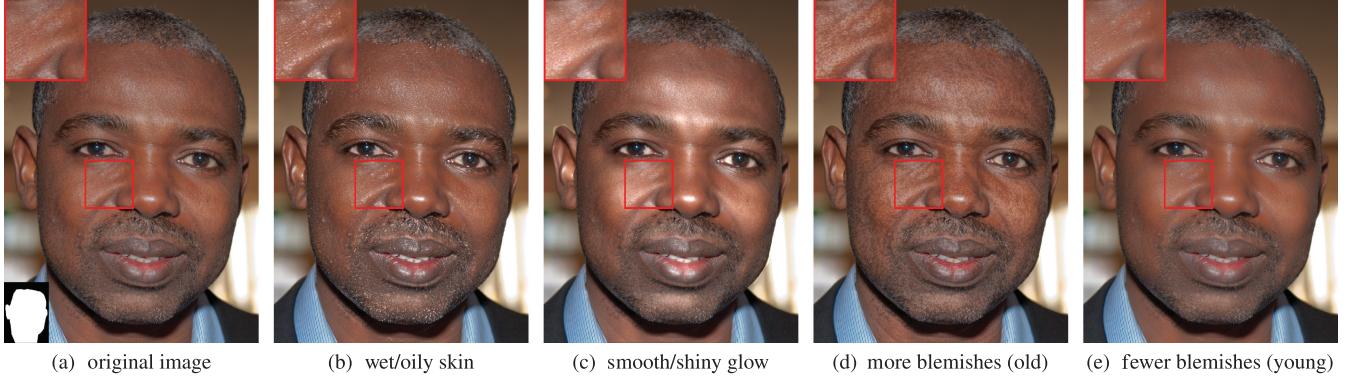


Fig. 2. Our band-sifting operators are particularly useful for manipulating material properties in human faces. (a) Original image courtesy of Bychkovsky et al. [2011] with detail inset at upper left, and mask inset at lower left. (b) We sift and then boost the high-amplitude, positive coefficients in the high-spatial frequencies, which gives the skin a more shiny or wet look. (c) We manipulate the positive low-spatial frequencies coefficients, which gives the skin a soft glow. (d) We produce an aging effect by emphasizing blemishes and pores that are not noticeable in the input image. We achieve this by sifting and then boosting the low-amplitude coefficients in the high-spatial frequencies. (d) We reverse the effect, that is, reduce blemishes and pores, by decreasing the sifted coefficients from (d).

style transfer, and Farbman et al. [2008] for a variety of photo edits such as detail enhancement and local sharpening. Heeger and Bergen [1995] and Portilla and Simoncelli [2000] also use a multiscale decomposition for their texture synthesis techniques. Motoyoshi et al. [2007] showed that manipulating the skewness of the coefficient distribution of the high-frequency bands affects the perceived gloss of materials.

Our work is related to this body of work since it splits images into components that are later modified separately. However, our purpose here is to look, more systematically, at the range of material-related manipulations that can be attained by doing modifications within the subband domain. Furthermore, material editing is different from other editing tasks because the visual information is distributed across space and subbands, and cannot be easily untangled.

Photo editing. Many image operators exist to manipulate the level of texture in photographs, for example, Tomasi and Manduchi [1998], Farbman et al. [2008], Fattal [2009], He et al. [2010], Paris et al. [2011], Gastal and Oliveira [2011, 2012], Xu et al. [2011, 2012], and Karacan et al. [2013]. These works focus on the signal processing challenges, for instance, they improve the output or accelerate the computation. While these methods change the perceived properties of the materials, this aspect is not discussed in those papers. Fattal et al. [2007] implicitly use this effect to reveal details that would be hard to see otherwise. However, the perceptual aspects themselves are not studied. Our work is complementary to these articles and focuses on how image operators alter the observer’s material perception.

In parallel, the perceptual effects of some editing tools have been quantified. For instance, Mantiuk et al. [2006] studied contrast changes and whether image changes are visible [2011]. Trentacoste et al. [2011] studied the interaction between blur size and image resolution, and showed that boosting the high frequencies of an image can be perceived as sharpening, halos, or countershading depending on the selected cutoff [2012]. In comparison, our work focuses on material perception.

Hybrid 2D/3D material editing. Khan et al. [2006] and Vergne et al. [2012] edit images to alter materials and their properties. The main difference with our work is their use of 3D data

provided by users, as in Vergne et al. [2012] or inferred 3D orientation as in Khan et al. [2006], from the images themselves—this allows Khan et al. [2006] to render new materials using standard 3D rendering, and Vergne et al. [2012] to warp images to convey shape and material properties. In comparison, our approach relies solely on the content of the input images and our operators are purely two dimensional, thereby avoiding any sort of 3D reconstruction that can be brittle on scenes with complex materials like those in which we are interested.

Weathering 3D models. Several techniques exist to modify 3D models and simulate aging and weathering, such as Golovinskiy et al. [2006], Mertens et al. [2006], and Glondu et al. [2012]. One of the effects we demonstrate makes people look older and objects more worn out, with the major difference being that we work purely in 2D. We also study several other effects besides aging, such as shininess and wetness.

Material perception. A few techniques recognize materials depicted in photos, for instance, to differentiate plastic from wood, as in Liu et al. [2010], Sharan et al. [2013], and Bell et al. [2015]. Fleming et al. [2013] conduct user studies to explore the interactions between material classification and judgments of material qualities, such as glossiness, roughness, and hardness in the visual and semantic domains. In comparison, we are interested in altering the properties of a given material like its shininess or roughness. Researchers have also studied the interplay between physical sources, such as 3D geometry, surface reflectance, and the light field in the perception of surface properties, such as gloss [Kim et al. 2011; Marlow et al. 2012; Marlow and Anderson 2013]. In our work we are interested in changing perceived surface properties, such as gloss, based on entirely image-based operations. More related to our approach, a few studies have shown a correlation between image statistics and the perception of properties such as translucency [Fleming and Bültlhoff 2005] and lightness [Motoyoshi et al. 2007; Sharan et al. 2008], and have proposed image filters that manipulate these statistics to alter the specific property that they study. Our work is inspired by these techniques and we build upon some of their findings. However, whereas those papers focus on a single effect, we explore a larger spectrum of effects and systematically characterize how they affect material perception.

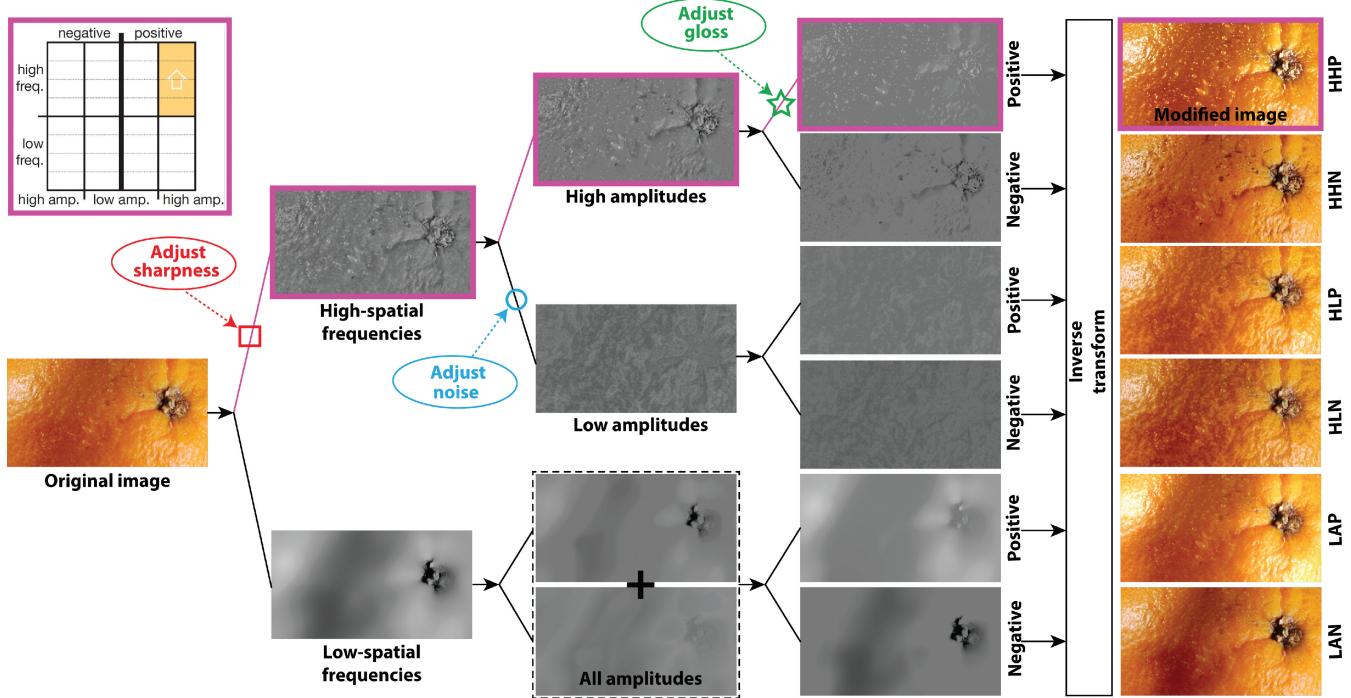


Fig. 3. Conceptual diagram of our band-sifting space. Input image courtesy of Ewan [2009]. Given an image, we split it into high- and low-frequency subbands. These are then split into high- and low-amplitude parts. These are further split into positive and negative parts. For visualization purposes we show only two frequency splits, but in practice we create $\log_2(\min(\text{width}, \text{height}))$ -frequency subbands and work on each one of them. Further, in order to make the size of the space more tractable, we “compress” the set of possible choices by looking at two categories of frequencies. We consider the high-to-mid frequencies as one category, which we refer to as “high-spatial frequencies”, and we look at the mid-to-low frequencies as another category, which we call “low-spatial frequencies”. Further, as shown in the diagram, we do not split the low-spatial frequencies category based on the amplitude of the coefficients since, numerically, sifting based on this criterion does not give much differentiation. However, the sign of the coefficients is still a useful sifting criterion along the low-spatial frequencies paths, for instance, it differentiates between broad-gloss and broad-shadow effects. With colored text and arrows we show how various operators can be mapped into paths in our space. With purple borders we show the path of sifted coefficients that was used to generate the orange result in Figure 1. In the upper-left border we show an alternative, more compact diagram, of the same path. On the far right, next to each path, we show acronyms which we use to describe paths in our space. For example, we use HHP for paths that manipulate (H)igh-spatial frequency, (H)igh-amplitude, (P)ositive coefficients.

2. BAND-SIFTING OPERATORS

In this section, we describe the space of band-sifting operators that is at the core of our work. We strike a balance between two objectives. We define a space that is both expressive enough to include an interesting variety of effects and concise enough to allow for an exhaustive study.

2.1 Motivation for Band-Sifting Operators

As we discussed in Section 1.1, many existing techniques can be interpreted as splitting an image according to a specific criterion like scale and amplitude, manipulating one of the generated components, usually with a simple operation like a multiplication, and recombining the result to form the final image. Our work extends this approach by decomposing images using several criteria at the same time. Figure 3 shows how we build our operators. We first split the original image into high- and low-frequency subbands. We then separate the subbands into their high- and low-amplitude parts. And we finally split the coefficients according to their sign, positive or negative. By placing multiplicative “control knobs” at specific

points in this flow diagram, one can modify sharpness (shown with a square), noise (shown with a circle), or gloss (shown with a star), similarly to previous work (Section 1.1). This illustration also suggests that there are many other ways to use our image decomposition, which raises several questions. What can one do by putting control knobs in other places? Are there more useful operators waiting to be found? What about yoked control knobs working on more than one component at a time? And what subband transforms are best?

Of course, there are any number of ways to increase the efficacy and complexity by adding in other techniques from image processing, computer vision, and machine learning. However, our purpose here is to understand what is possible while staying within this scheme. Even with this restriction, there is plenty of territory to explore, and useful operators can serve as a starting point for later improvements.

Our goal here is to ask what can be done by simple manipulations of multiscale transforms. By staying close to the original image data, we maintain locality and avoid any propagation of artifacts. We also avoid the fragility that can occur, for example, when imposing specific physical models or elaborate priors. We accept arbitrary

images as input and, in our experience, the image modifications look “natural” as long as they are not pushed too far.

2.2 Three Sifting Stages

We now describe the stages that we use to decompose images. We start by constructing a multiscale image decomposition and sift the subband coefficients based on three criteria: scale, amplitude, and sign.

Scale. Our design space follows a common trend and acts upon a multiscale image decomposition [Burt and Adelson 1983; Chen et al. 2007; Fattal et al. 2007; Farbman et al. 2008; Fattal 2009; Hanika et al. 2011; Paris et al. 2011]. Such decomposition provides a set of subbands that can be thought of as an overcomplete wavelet representation in the sense that each coefficient represents details at a given location and scale. This latter aspect is our first sifting stage: we allow our “sieve” to act selectively upon the large-scale or small-scale coefficients, or on all of them (i.e., both large and small scale). Intuitively, this separation differentiates between small elements such as skin pores on a face, and bigger ones like large-scale shading and shadow variations.

Amplitude. Our second sifting criterion is the amplitude of the coefficients. We separately manipulate coefficients with a high or low amplitude, or both (high and low). This sieve separates low-contrast from high-contrast features. It is related to wavelet coring [Donoho 1995; Simoncelli and Adelson 1996], with the major difference that we keep and process the low-amplitude coefficients instead of discarding them.

Sign. The third sifting criterion differentiates coefficients based on their sign: positive or negative. Recent studies [Sharan et al. 2008] have shown that the skewness of the subband coefficient distributions, that is, the asymmetry of the coefficient histograms, is correlated with the perception of lightness and gloss. From a numerical perspective, there is not a single well-defined way to alter skewness, that is, the same skewness value can be achieved with many different transforms. In our work we modify coefficients based on their sign, which gives direct control on the distribution symmetry. This approach has an intuitive interpretation that the positive subband coefficients describe bright features like highlights, and the negative ones capture features like crevices, holes, and shadows in wrinkles.

2.3 Refining the Scale-Sifting Criterion

The scale criterion raises two nontrivial issues: how many subbands to use and how to compute them. The rest of this section discusses these two issues.

Constructing the scale subbands. We started our study using a standard Laplacian pyramid that has the advantage of great simplicity. However, early in our investigations it became apparent that the Laplacian pyramid introduced artifacts at edges, which is a common problem when using linear filters on natural images. We therefore investigated pyramids based on edge-aware filters. We tried three such filters: the bilateral filter [Tomasi and Manduchi 1998], the weighted least-squares filter [Farbman et al. 2008], and the guided filter [He et al. 2010]. All gave a significant reduction in edge artifacts (see the supplemental material for comparison). We chose to use the guided filter (used for all results in this article), but other filters would presumably give similar results.

Number of subbands. For our study, we used images at the resolution of typical monitors (e.g., the longer side set to 512 pixels). Using a factor of 2 in resolution between each subband, this yields 8

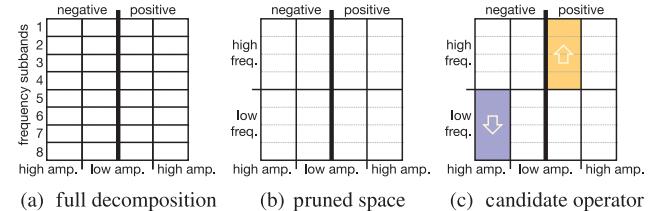


Fig. 4. In (a), we show a full decomposition into 8 subbands with each subband partitioned into 4 parts based on amplitude and sign. This gives 32 cells. Each cell can have one of three knob settings: boost, reduce, or leave the same. This gives 3^{32} potential configurations, which is hopelessly large. So we cluster the bands into two categories (high-spatial frequencies and low-spatial frequencies), giving 8 cells as shown in (b). For each band-sifting operator, we show the knob setting of each cell with an arrow and a color, as shown in (c), where the low-amplitude positive high-frequency coefficients have been boosted, and the high-amplitude low-frequency negative coefficients have been reduced. The set of 3^8 configurations is still large; see text for further methods to reduce dimensionality.

subbands. Then, the sign and amplitude sifting generates 4 components for each subband, and each of these components can either be boosted, reduced, or left unchanged. This would leave 3^{32} operators to explore, which is impractical. We take a few steps to make this number more tractable. The first one is to group the subbands into two sets: the high- and low-frequency subbands, which leaves 3^8 possible combinations. Figure 4 illustrates the decomposition we use. However, this number is still too large for the purposes of our exhaustive perceptual studies. In the next section, we further discuss how to reduce the space to a more manageable size while ensuring that we still have a variety of distinct nontrivial effects to study.

2.4 Early Pruning

Even with the subband grouping described in the previous section, the space of possible band-sifting operators remains challenging to explore. In this section, we explain how we structured the space to make its exploration tractable.

Independent criteria. First, we apply the sifting criteria independently of each other. For instance, for the sign, we choose between positive, negative, or both, and apply this choice to all the subbands. This gives us 3 sifting criteria (scale, amplitude, and sign), with 3 options for each of them: {high (H), low (L), all (A)}, {high (H), low (L), all (A)}, and {positive (P), negative (N), all (A)}, respectively. Once we have selected which coefficients to modify, we can either *boost* (B) them or *reduce* (R) them. This defines $3 \times 3 \times 3 \times 2 = 54$ combinations in total. Figure 5 illustrates these 3 criteria.

Removing redundancy. We explored and evaluated the space of operators in a set of pilot studies. We found that it could be reduced to a more useful set due to some redundancy in the effects achieved. Therefore we applied the following pruning based on our observations.

—High frequencies tend to mask low frequencies and there is no visually significant difference between paths that sift only the high frequencies and those that sift all of them. Therefore we do not include the latter in our study (pruned space: $2 \times 3 \times 3 \times 2 = 36$).

—All-amplitudes and high-amplitudes paths also produce visually similar results because they differ only due to the low-amplitudes coefficients that are small by construction. We do not include the

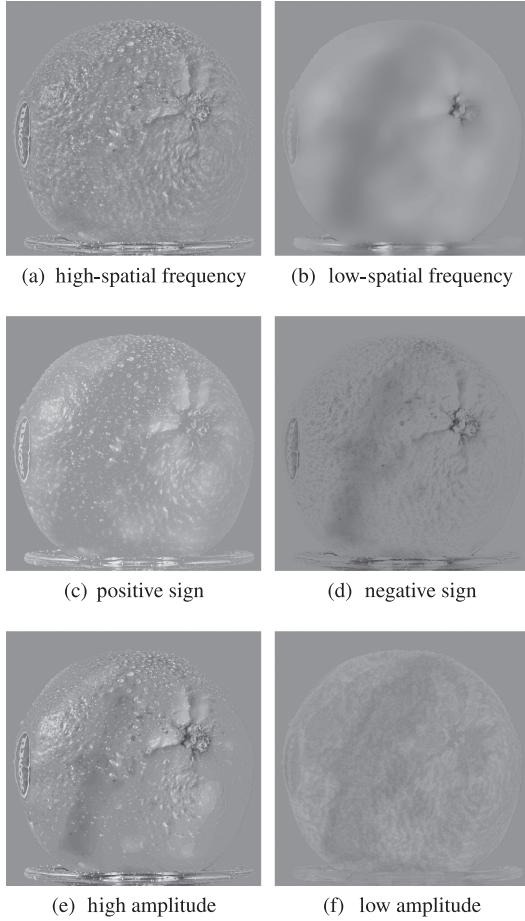


Fig. 5. Visualization of the subband coefficients sifted by each independent criterion. We visualize each possible selection by setting to 0 all the non-selected coefficients. We show the individual subbands for this input image in the supplemental material. High-spatial frequencies (a) capture features like small-scale bumps and wrinkles, and low-spatial frequencies (b) mostly represent large-scale shading variations. Positive (c) and negative (d) coefficients show highlights and shadows, respectively. Finally, high-amplitude coefficients (e) represent specular highlights and deep shadows, while low-amplitude values (f) capture more subtle reflectance variations.

all-amplitudes paths in our study (pruned space: $2 \times 3 \times 2 \times 2 = 24$).

—Low-spatial frequency coefficients come from repetitively smoothing the input image, and most of them are very small. Sifting these coefficients based on their amplitudes leaves only very few significant values, and the corresponding modifications have almost no effect as can be seen in Figure 3. We avoid the high-versus low-amplitudes paths and only include the all-amplitudes ones for the low-spatial frequency paths in the study. So, for high-spatial frequencies there are $(P|N|A) \times (H|L) \times (B|R) = 1 \times 3 \times 2 \times 2 = 12$ operators. And for low-spatial frequencies there are $(P|N|A) \times (A) \times (B|R) = 1 \times 3 \times 1 \times 2 = 6$ operators. This gives a total of 18 possible operators.

—Also, reducing the low-amplitude coefficients does not have a noticeable effect since their value is already low. We do not include these paths in our study. This eliminates 3 possible operators, giving a final total of 15 operators.

Another way of thinking of the space is: 9 operators where boost is applied, and 6 operators where reduce is applied.

2.5 Physical Observations

Appearance properties, such as luminance variations on surfaces, come from many physical sources, and a key reason why our approach works well is that in many cases these sources correspond to different bins of our subband decomposition. We discuss a few such examples next.

- Specularities typical of wet and glossy surfaces are bright and small, and mostly fall in the high-spatial frequencies, with high amplitude and positive sign.
- Pits and grooves, including the wrinkles and pores of the skin, tend to be dark due to self-shadowing, and their magnitude is often medium or large. Because of this, they appear in the high-spatial frequencies, high-amplitude, negative coefficients.
- Variations in albedo, caused by dirt, stains, age, wear, or other degradations, tend to be low in amplitude compared to dark pits and bright highlights, and often show up in the low-amplitude negative coefficients.

Such characteristics of physical objects are common and provide the basis for our approach. This also points at a limitation of our operators. If an object does not exhibit such properties, our operators are not effective. For instance, we cannot make a perfectly smooth object look rough. To do so, one would need to hallucinate surface details that do not exist in the original image. We believe this is an interesting direction for future work.

3. IMPLEMENTATION

In this section, we describe the actual implementation of the band-sifting operators that we described in the previous section. We provide detailed pseudocode in Algorithm 1.

Our multiscale decomposition is akin to that of Farbman et al. [2008]. We repetitively process the input image with an edge-aware filter, doubling its spatial extent each time. This produces a series of images of increasing smoothness. Taking the difference between two such successive images gives frequency bands, otherwise known as subbands, that contain details of a given size. Since we preserve edges in this construction, we do not downsample the subbands to prevent aliasing, that is, each subband has the same resolution as the input. We use the guided filter with its default regularization parameter ($\sigma_r = 0.1^2$) [He et al. 2010] for edge-aware filtering.

In our prototype, the multiscale decomposition and the subband-sifting procedure on the GPU were implemented using C++ and OpenCL. To accelerate the guided filter on the GPU, we implemented an efficient summed-area table algorithm [Harris et al. 2007] that we use as a building block for all the box filters, mean, and standard deviation computations required by the guided filter approach. This allows us to achieve interactive frame rates (5–6fps) for 1-megapixel videos, which is sufficient for preview purposes before running the full resolution computation offline.

For our study, we fix the long edge of the input image to 512 pixels, and compute 8 subbands. We split them into 4 low-frequency subbands and 4 high-frequency ones. For the amplitudes, we use the standard deviation of each subband as the threshold between the high and low categories. To avoid the artifacts that a hard threshold would introduce, we use a soft transition spanning $\pm 20\%$ around the standard deviation. For instance, if an operator multiplies by 2 the high-amplitude coefficients in a subband where the threshold is

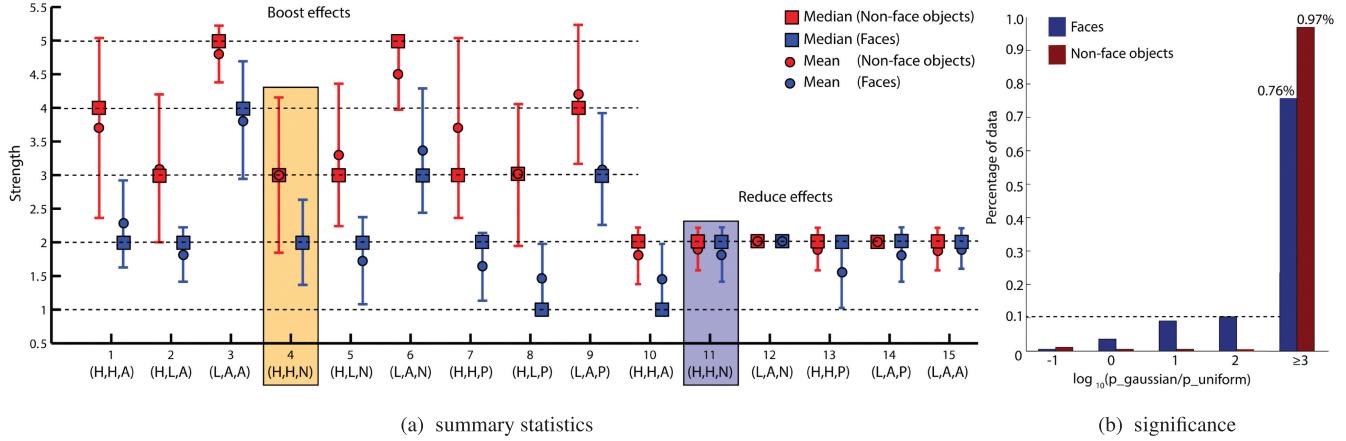


Fig. 6. The plot on the left reports the mean and standard deviations of the votes on the natural-vs.-unnatural study (a). We use the notation introduced in Section 2.4 on the horizontal axis. We also indicate the median vote that we use later in the second user study. These votes confirm that the threshold between natural and unnatural settings is statistically significant since the uniform distribution hypothesis is much less probable than the Gaussian distribution hypothesis in most cases (b). With yellow and purple rectangles we indicate the boost and reduce statistics for one of the operators, for which we show a concrete qualitative example in Figure 7. See the text for details.

σ_t , the multiplication factor is 1 below $0.8\sigma_t$ and 2 above $1.2\sigma_t$, and smoothly varies in between. Finally, the increasing or decreasing of the selected coefficients is performed with a simple multiplication factor greater or lower than 1.

4. USER STUDIES AND RESULTS

In this section, we describe the user studies that we performed to characterize which operators produce effects that are natural and how they affect material appearance. We then present more results on still images and videos.

To understand the visual impact of our band-sifting operators, we conducted two user studies to validate their perceptual effects. The goal of the first study was to find which operators are natural, that is, for a given image, what is the range of multiplication factors within which an operator produces a discernible and natural-looking change? This study tells how much we can boost or reduce an effect before it starts to look unnatural on a certain image.

The second study asks users to describe the visual change that operators produce. This task shares some similarities with a recent line of work in computer vision where the goal is to describe images through high-level attributes [Patterson et al. 2014]. In our work, we are interested in assigning attributes related to the perceptual changes produced by the band-sifting operators. We use 16 categories of words describing various material-specific properties. We designed the set of words in a pilot study between the authors by looking at the perceptual effects produced by the operators on tens of examples. Figure 8 lists these words. Some of those categories of words describe low-level features, such as “wrinkled, pitted, bumpy” and other categories describe high-level properties, such as “young, new, fresh”. Then, in our study with casual users, participants were shown the original image and the modified image, and were asked to pick all categories of words that apply.

4.1 Study 1: Natural vs. Unnatural

The goal of this study is to find whether there is a reasonable range of multiplication factors where the operators produce natural-looking results. We test a few multiplication factors and run a study to find the threshold between natural and unnatural.

Given a pair of an input image I and a band-sifting operator F , we seek to sample a few versions of the operator, $F(I, m_1), F(I, m_2), \dots, F(I, m_s)$, acting on the original image with different multiplication factors m_1, m_2, \dots, m_s . Our early experiments showed that using the same m factors across operators performs poorly; the same value can produce a strong effect with one operator and a weak one with another. Instead, we define the factor as $m_0 = 1$ and $m_{i+1} = \arg \min_m \|F(I, m_i) - F(I, m)\| > 1$ using the CIE-Lab L_2 norm. We use binary search to efficiently find m_{i+1} . This procedure generates samples regularly spaced in the CIE Lab color space akin to Ngan et al. [2006], which approximates a perceptually uniform sampling. We observed that, for the effects that boost the coefficients, 5 iterations of the preceding procedure were usually enough to produce too-strong results. For the effects that reduce the coefficients, 2 iterations were usually enough to make them close to 0.

With the aforesaid sampling procedure, we produced 5 images of different strength for each of the 9 operators that increase the coefficients, plus 2 images for the 6 operators that decrease the coefficients, for a total of $5 \times 9 + 2 \times 6 = 57$ variations per image. We used 21 images, 11 faces of various genders and races, and 10 nonface objects with uniform materials, such as metal, leather, ceramic, and fruits. Users were shown a single modified image at a time and were asked whether it looked natural to them. We provide a snapshot of this task as well as the full set of images in the supplemental material.

Every user was shown 15 sets of 21 images. Each set was made of each of the 21 test images modified using a randomly picked setting. Users saw the same scene 15 times. Occurrences of any scene were separated by 20 other images to limit the effect of users getting trained by the previous viewing of that scene. Users were asked to base their decision only on the current image, and had no reference original image. This study had a total of 47 participants and, on average, we got 7.5 votes per setting, since we assigned them uniformly across participants.

Figure 6 summarizes the results of the study and confirms our initial observation that our band-sifting operators can produce nontrivial natural-looking variations, even for familiar objects such as human faces. We also show separate statistics for face and nonface objects, which reveals some interesting trends. For example, in essentially all cases where we boost the signal, images

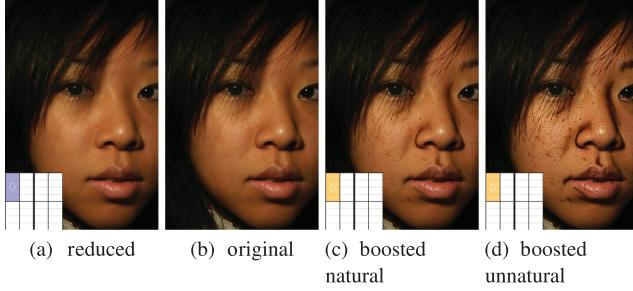


Fig. 7. Our first user study characterized how much we can reduce (a) or boost (c) a set of coefficients while maintaining a natural look. Increasing the coefficients past this point eventually generates unnatural images (d). The insets explain which coefficients are affected; see Table I.

of nonface objects withstand larger modifications than images of face objects. Another interesting observation is about reducing the high-frequency high-amplitude coefficients on faces. If we manipulate the positive or negative coefficients separately, we produce larger modifications compared to reducing them both at the same time. This happens because modifying the positive and negative coefficients at the same time leads to dampening of all high-frequency features, such as dark skin pores and bright skin gloss, which quickly produces unrealisitc smoothening of the skin compared to modifying skin pores independent of skin gloss. In the supplemental material, we show an example image that demonstrates this effect.

Figure 7 shows examples of natural and unnatural adjustments using our band-sifting operator that manipulates the high-frequency high-amplitude negative coefficients. We also mark the data points in Figure 6(a) that correspond to this operator. For example, as we pass beyond the realistic threshold of the boost operator, for instance, 3 steps, it starts to produce unrealisitc-looking images as we show in Figure 7(d).

Statistical significance. To confirm the statistical significance of our results, we assumed each user has a naturalness threshold. We compared two hypotheses: purely random thresholds, that is, uniformly distributed over the tested range, and Gaussian distributed thresholds centered on the value that we reported in Figure 6(a) (we used a unit variance for simplicity). We compared the probabilities of obtaining the users' answers under these two assumptions. As shown in Figure 6(b), for 76% of the face images and 97% of the nonface images, the results of our study are more than 300× more probable under the Gaussian hypothesis than under the uniform one, which confirms the hypothesis of a consistent threshold across images and users.

4.2 Study 2: Name the Effects

The goal of the second study is to determine the perceptual effects associated with the band-sifting operators and to evaluate their consistency across different users and scenes. Users were shown pairs of images where image A was the original input image I , and image B was a modified version $F(I, m)$. We seek a parameter value m that produces a visible and natural effect, which we achieved with the quasi-median of the votes in the first study, that is, the multiplication factor with an equal number of natural and unnatural votes above and below. We showed users the 16 groups of keywords and, for each group, asked them to choose between 3 options for the direction of the perceptual change: "Less", "More", or "N/A". Twenty users participated in this study and half of them had not taken part in Study 1. On average, we got 60 responses per operator (30 for faces and 30 for non-face objects) for a total of 1100 responses. Figure 8

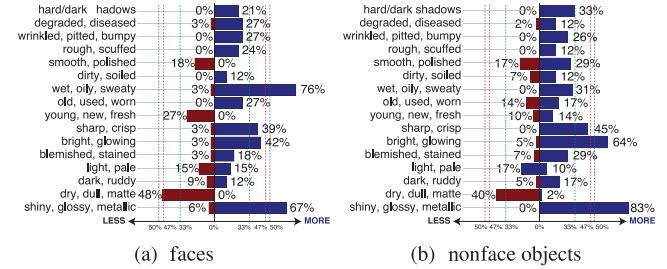


Fig. 8. Our second user study determined the perceptual effects associated with our band-sifting operators. In these plots we show results for one of our more consistent effects: boost high amplitude, positive-valued, high-spatial frequencies. The red bars show the percentage of votes for the "Less" option, the blue bars show the percentage of votes for the "More" option, and the difference to 100%, which we do not show on the plot, is the percentage of votes for the "N/A" option. The green lines indicate the probability of chance, that is, where the results have been generated by picking between the three options {"Less", "More", or "N/A"} uniformly at random. The blue lines indicate the 95% confidence interval, that is, results above this threshold, ≈50%, are statistically significant with high probability. The red lines, at ≈47%, indicate the 90% confidence interval. The majority of participants agreed that this band-sifting operator tends to make human faces more wet, oily, or sweaty, whereas, for objects, the band-sifting operator tends to make them look more shiny, glossy, or metallic. See the text for details about the test of significance.

summarizes the responses for one of our operators for both face and non-face objects. We show similar plots for all the 15 operators in the supplemental material.

Statistical significance. We tested our results against the null hypothesis that the choice between the 3 options is uniformly random. This hypothesis corresponds to a standard multinomial distribution with a 33% mean. For 30 votes, the standard deviation is 8.6%, and using a 95% confidence interval we can rule out the null hypothesis for results below 16% and above 50%. For a 90% confidence, the interval is [19%, 47%]. We give the detailed derivation of these numbers in the supplemental material. We show the 50% and 47% thresholds in Figure 8 and use them to report the results in Table I.

Consistent effects. We found 7 operators that produce consistent and perceptually discriminative effects: boost/reduce shininess, boost/reduce roughness, boost weathering patterns, and boost/reduce glow. Table I summarizes this finding. The number of word sets above the significance threshold ranges from 1 to 4. In general, operators have a more consistent effect on faces. We hypothesize that the diversity of scenes and materials present in the non-face images makes it "harder" for an operator to be consistent. In comparison, the only material in face images is skin and, although human observers recognize subtle differences, these are not as dramatic as those between bronze and potatoes, for instance.

From a photo-editing perspective, these band-sifting operators cover several common tasks on objects such as reducing or increasing weathering, smoothness, and shininess. For faces, they provide a simple and effective means for attenuating blemishes and wrinkles, controlling the dryness of the skin, and adding a photographic glow typically observed in studio portraits.

4.3 Image Results

We now demonstrate the effects of band sifting on a range of static scenes and qualitatively discuss the results.

Table I. Recap of The Band-Sifting Operators with a Consistent Effect

	coefficient selection	associated properties	input	output												
O1.	<table border="1"> <tr> <td></td><td>negative</td><td>positive</td></tr> <tr> <td>high freq.</td><td></td><td style="background-color: orange;">↑</td></tr> <tr> <td>low freq.</td><td></td><td></td></tr> <tr> <td>high amp.</td><td>low amp.</td><td>high amp.</td></tr> </table>		negative	positive	high freq.		↑	low freq.			high amp.	low amp.	high amp.	Objects (strength: 4) more shiny/glossy/metallic (83%) more bright/glowing (64%) Faces (strength: 2) more wet/oily/sweaty (76%) more shiny/glossy/metallic (67%) less dry/dull/matte (48%)		
	negative	positive														
high freq.		↑														
low freq.																
high amp.	low amp.	high amp.														
O2.	<table border="1"> <tr> <td></td><td>negative</td><td>positive</td></tr> <tr> <td>high freq.</td><td></td><td style="background-color: blue;">↓</td></tr> <tr> <td>low freq.</td><td></td><td></td></tr> <tr> <td>high amp.</td><td>low amp.</td><td>high amp.</td></tr> </table>		negative	positive	high freq.		↓	low freq.			high amp.	low amp.	high amp.	Objects (strength: 2) more dry/dull/matte (62%) less shiny/glossy/metallic (54%) Faces (strength: 2) more dry/dull/matte (62%) less shiny/glossy/metallic (55%)		
	negative	positive														
high freq.		↓														
low freq.																
high amp.	low amp.	high amp.														
O3.	<table border="1"> <tr> <td></td><td>negative</td><td>positive</td></tr> <tr> <td>high freq.</td><td style="background-color: orange;">↑</td><td></td></tr> <tr> <td>low freq.</td><td></td><td></td></tr> <tr> <td>high amp.</td><td>low amp.</td><td>high amp.</td></tr> </table>		negative	positive	high freq.	↑		low freq.			high amp.	low amp.	high amp.	Objects (strength: 3.5) more sharp/crisp (48%) Faces (strength: 2) more hard/dark shadows (60%)		
	negative	positive														
high freq.	↑															
low freq.																
high amp.	low amp.	high amp.														
O4.	<table border="1"> <tr> <td></td><td>negative</td><td>positive</td></tr> <tr> <td>high freq.</td><td style="background-color: blue;">↓</td><td></td></tr> <tr> <td>low freq.</td><td></td><td></td></tr> <tr> <td>high amp.</td><td>low amp.</td><td>high amp.</td></tr> </table>		negative	positive	high freq.	↓		low freq.			high amp.	low amp.	high amp.	Objects (strength: 2) more smooth/polished (56%) less wrinkled/pitted/bumpy (53%) less hard/dark shadows (53%) more young/new/fresh (47%) Faces (strength: 2) more blemished/stained (62%) more old/used/worn (55%) more wrinkled/pitted/bumpy (55%)		
	negative	positive														
high freq.	↓															
low freq.																
high amp.	low amp.	high amp.														
O5.	<table border="1"> <tr> <td></td><td>negative</td><td>positive</td></tr> <tr> <td>high freq.</td><td style="background-color: orange;">↑</td><td style="background-color: orange;">↑</td></tr> <tr> <td>low freq.</td><td></td><td></td></tr> <tr> <td>high amp.</td><td>low amp.</td><td>high amp.</td></tr> </table>		negative	positive	high freq.	↑	↑	low freq.			high amp.	low amp.	high amp.	Objects (strength: 3.5) more sharp/crisp (57%) more old/used/worn (49%) more blemished/stained (47%) Faces (strength: 2) more blemished/stained (62%) more old/used/worn (55%) more wrinkled/pitted/bumpy (55%)		
	negative	positive														
high freq.	↑	↑														
low freq.																
high amp.	low amp.	high amp.														
O6.	<table border="1"> <tr> <td></td><td>negative</td><td>positive</td></tr> <tr> <td>high freq.</td><td></td><td></td></tr> <tr> <td>low freq.</td><td style="background-color: orange;">↑</td><td style="background-color: orange;">↑</td></tr> <tr> <td>high amp.</td><td>low amp.</td><td>high amp.</td></tr> </table>		negative	positive	high freq.			low freq.	↑	↑	high amp.	low amp.	high amp.	Objects (strength: 5) more shiny/glossy/metallic (79%) more bright/glowing (56%) Faces (strength: 4) more bright/glowing (70%) more shiny/glossy/metallic (68%)		
	negative	positive														
high freq.																
low freq.	↑	↑														
high amp.	low amp.	high amp.														
O7.	<table border="1"> <tr> <td></td><td>negative</td><td>positive</td></tr> <tr> <td>high freq.</td><td></td><td></td></tr> <tr> <td>low freq.</td><td style="background-color: blue;">↓</td><td style="background-color: blue;">↓</td></tr> <tr> <td>high amp.</td><td>low amp.</td><td>high amp.</td></tr> </table>		negative	positive	high freq.			low freq.	↓	↓	high amp.	low amp.	high amp.	Objects (strength: 2) more dry/dull/matte (58%) less shiny/glossy/metallic (52%) Faces (strength: 2) more dry/dull/matte (64%) less shiny/glossy/metallic (57%) less bright/glowing (57%) less sharp/crisp (54%)		
	negative	positive														
high freq.																
low freq.	↓	↓														
high amp.	low amp.	high amp.														

Band-sifting operator no. 1 makes objects look shinier and more metallic, and faces wetter and oilier. No. 2 has the opposite effect. No. 3 reveals small defects in objects and makes face pores look deeper. No. 4's effect on objects is not consistent but it renders people younger. No. 5 reveals object details and makes them look stained. On faces, it emphasizes blemishes and makes people look older. No. 6 renders objects more metallic and gives a shiny glow to people. No. 7 has the opposite effect. The strength numbers in the middle column are the result of our first user study. For our second user study, we report the statistically significant perceptual votes. Votes larger than 50% are statistically significant at the 95% confidence interval. In gray we show votes that are statistically significant at the 90% confidence interval. The input image in row 3 is courtesy of Bychkovsky et al. [2011]. The input image in row 6 is courtesy of Bell et al. [2013].

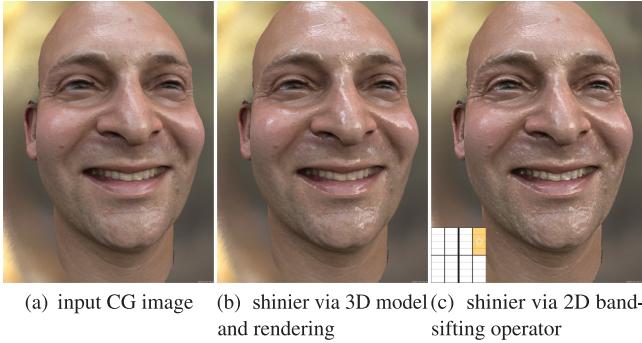


Fig. 9. Case study on a photorealistic scanned CG model of a face, courtesy of von der Pahlen et al. [2014]. In (a) and (b) we rendered the face under natural lighting conditions [Debevec and Malik 1997], using the isotropic Ward BRDF model with two different values of the parameter α that controls the spread of the specular lobe. In (c) we show that our band-sifting “wet/oily/shiny” operator, when applied to image (a), can produce a perceptually similar change in shininess.

In Figure 9, we show that our purely image-based band-sifting operators can produce results visually similar to what can be achieved with a 3D model rendered with a physically inspired BRDF model. We used a photorealistic scanned 3D model of a face, where we control perceptual parameters related to shininess by changing physical parameters of the underlying rendering model [Ward 1992]. We rendered the face with two different values of the α parameter that controls the spread of the specular lobe. Smaller values of α increase the sharpness of the reflected image and make the object look shinier. The question is whether we can get a similar effect using just 2D image manipulations. We show that our band-sifting operator that boosts the positive high-amplitude high-spatial frequencies produces a perceptually similar change in shininess.

In Figure 10(b) we show a failure case of one of our band-sifting operators. When the visual cues are not present in the original input image or they are not well isolated by our sifting criteria, our band-sifting operators fail to convey consistent perceptual effects (b). Even though one operator may not work well on an image, we found that usually others might still be useful. For example, we can reduce skin defects (c) and blemishes (d) or add a smooth skin glow (e). Furthermore, the independent band-sifting operators that we studied can be combined to achieve even more advanced material editing effects. In (f) we combine the previous three operators to achieve the combined perceptual effect, a younger-looking face with a nice skin glow which is often seen in professional magazines.

Figure 11 illustrates the diversity of effects that can be achieved with band-sifting operators. For brevity’s sake, we use the notation previously introduced where the amplitude is selected in $\{H, L, A\}$, the frequency in $\{H, L, A\}$, and the sign in $\{P, N, A\}$.

Gargoyle. Boosting the HHP coefficients enhances the gloss and also brings out whitish “distress” marks, which gives an overall shinier look. Boosting the AHN coefficients produces a patina with dark mottling.

Grapes. We show a combination of two of operators: boosting the ALP coefficients while reducing the AHN coefficients gives the grapes a luminous glow. We also achieve a weathering effect by boosting the LHN coefficients to bring out the patterning on the grape skins.

Onion. Reducing the HHA coefficients removes texture details while retaining the smooth shiny appearance of the onion, whereas

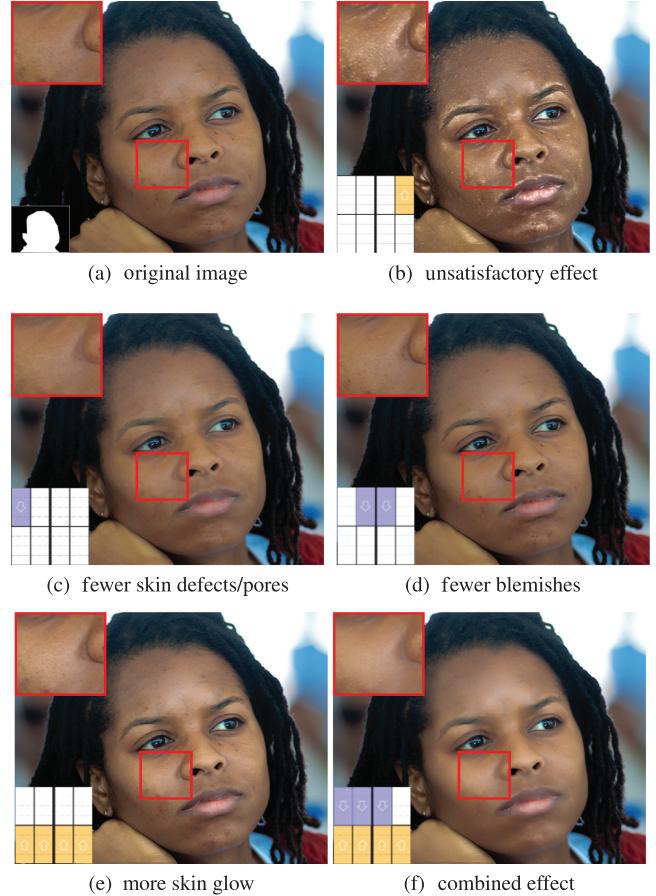


Fig. 10. Starting from an input image of a very dry and matte face (a), our “wet/oily/shiny” band-sifting operator fails to produce a plausible-looking effect (b). Even though one operator may fail to produce a satisfactory effect on some image, others might still work. In (c) we reduce skin defects and pores. In (d) we reduce skin blemishes and pigmentation. In (e) we add a smooth skin glow. Furthermore, simple combinations of band-sifting operators can be used to achieve advanced material-editing tasks. In (d) we show the combined effect which achieves a younger look with a nice skin glow, an advanced effect often seen in professional magazines. Original image is courtesy of Bychkovsky et al [2011].

boosting the same coefficients reveals the mottled coloration of the onion skin.

Sweet potatoes. Boosting the LHN coefficients reveals dark blotchy patches, while boosting the HHN ones reveals sharp dark spots.

Orange. Boosting the HHP coefficients emphasizes the highlights and makes the orange look shinier. Alternatively, we can emphasize pores and dark spots by boosting the HHN coefficients.

4.4 Video Results

We now demonstrate the effects of the band-sifting operators on video sequences. Editing videos consistently is particularly challenging and typically requires many hours of painstaking manual editing. In Figures 12 and 13, we show example input frames from the video sequences and the results on two different frames. We show results on three video categories where post-process material



Fig. 11. Showing a variety of effects produced with our band-sifting operators. In row (1) we make the gargoyle look more glossy (1,b) or more weathered, by emphasizing the patina (1,c). The grapes can be given a more glowing (2,b) or more dirty look (2,c) by emphasizing the patterning on the skin. The skin of the onion can be given a more clean (3,b) or a more mottled look (3,c). In row (4) we show that, by treating the low- and high-amplitude coefficients separately, we can get very different perceptual effects, weathering patterns (4,b) vs. surface roughness (4,c). Finally, we can make the orange look more shiny and wet (5,b) by manipulating the positive coefficients independently from the negative. In (5,c) we show that pores and dark spots are well captured in the negative coefficients which can be used to manipulate those material properties independently from highlights and gloss.



Fig. 12. Results on videos of faces, downloaded from the Internet. Our band-sifting operators can be used to efficiently post-process material appearance in videos without introducing temporal artifacts; see the supplemental video. For example, we can make the actor's skin look more oily (column 2) or add more skin glow (column 3). In column 4 we control wrinkles and blemishes by manipulating coefficients in the corresponding combination of band-sifting paths. We boost the coefficients in rows 1 and 2, which gives the face a more aged look, whereas in rows 3 and 4 we reduce them to render a more clean- and young-looking face. The simplicity of our model allows all this to be done interactively by manipulating a few sliders, without having to model the effects pixel by pixel on every frame. We show the abbreviations of the used operators in brackets. Video courtesy of OSU [2012].

editing would be a desired tool: (1) closeup views of people giving interviews; (2) 360 spins for product photography; and (3) static objects under dynamic lighting. In each case, we applied our operators frame by frame. As can be seen in the supplemental video, the results are artifact free and temporally consistent, which demonstrates the robustness and stability of our band-sifting operators.

Interview A. First, we boost the HHP coefficients to give the skin a more wet/oily look. Then we demonstrate our skin glow effect, which is a common appearance professional photographers aim to achieve through a combination of lighting and facial cosmetics. We achieve that effect through entirely image-based manipulations by boosting the ALP coefficients. Finally, we emphasize blemishes and

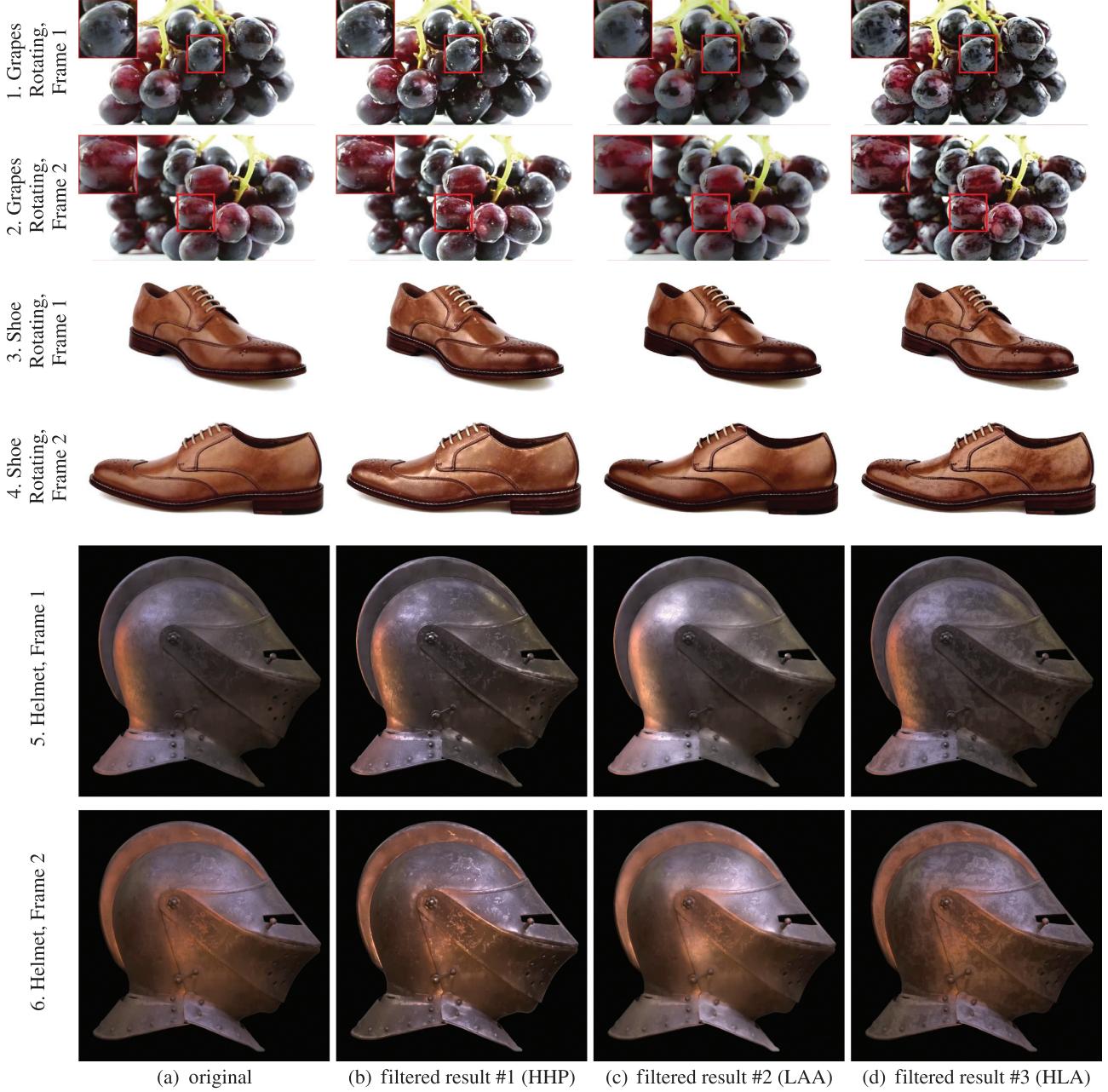


Fig. 13. We evaluate our band-sifting operators on videos of 360 spins for product photography (rows 1 through 4) and videos of static objects under dynamic lighting (rows 5 and 6). The helmet scene is courtesy of [Debevec et al. 2000]. In column (a) we show example input frames from the video sequences, and the next columns show some of the effects our band-sifting operators can achieve. In column (b) we demonstrate that our “shiny/glossy/metallic” effect, which boosts the high-amplitude positive-valued, high-spatial frequencies, produces a perceptually consistent effect across different scenes. In column (c), rows 1 and 2, we reduce the corresponding coefficients, which gives the grapes a more diffuse look. In the rest of column (c), rows 3 through 6, we show our “bright/glowing” effect which boosts the positive-valued, low-spatial frequency coefficients. This gives the leather shoes a more smooth and polished appearance (rows 3 and 4) and makes the metal helmet look more smooth and less rough (rows 5 and 6). In column (d) we show our “old/used/worn” effect which boosts the low-amplitude, high-spatial frequencies. In agreement with our user Study #2, this effect produces a persistent perceptual effect of aging or damage by bringing out weathering patterns such as spots and dust on fruits (rows 1 and 2) leather stains (rows 3 and 4) and patina on metals (rows 5 and 6). The grapes scene is courtesy of Videvo [2014]. The shoes scene is courtesy of Arqbll [2014] reproduced with permission from Arqbll LLC. The helmet scene is courtesy of Debevec et al. [2000].

spots by boosting the LHA coefficients. We did not use a detailed mask around the face in this case to show that the operators could be directly applied and used in settings like this one.

Interview B. We demonstrate the perceptual consistency of our wet/oily and skin glow effects used in the previous example by applying them onto a different subject. Then, we reduce blemishes and spots to produce a cleaner-looking face. To localize the effects on the face only, we created and tracked a detailed facial mask using the Roto Brush tool in Premiere Pro.

Leather shoes. Boosting the HHA coefficients emphasizes the highlights and some scratches, which gives an overall shinier-looking leather. Then, we give the leather a smoother and more polished look by boosting the ALP coefficients. Finally, we achieve a weathering effect by boosting the LHA to bring out the patterning on the leather.

Grapes. Boosting the HHP coefficients emphasizes the highlights and makes the grapes look shinier and wetter. Then, we reduce the same coefficients to produce a more diffuse look. Finally, we bring out the weathering patterns which also make the grapes look a bit more dirty.

Helmet. We demonstrate that the operators produce a consistent look under dynamic lighting conditions. Boosting the HHP coefficients emphasizes the highlights and some scratches, which makes the helmet appear shinier. Boosting the ALP brings out the broad gloss, which gives a smoother, less rough-looking metal. Finally, boosting the LHA coefficients brings out the weathering patterns, which makes the metal look like it has more patina.

5. DISCUSSION AND LIMITATIONS

Our purpose in this article is to devise 2D image operators which can make visually plausible modifications of surface properties. Such operators are simple to implement and can be applied to arbitrary images. However, they are only useful for properties that are manifested simply within the distributions of subband coefficients. Prior research indicates that such descriptions are useful for certain tasks involving natural images (e.g., in denoising [Donoho 1995; Simoncelli and Adelson 1996], and in texture analysis [Heeger and Bergen 1995; Portilla and Simoncelli 2000]). Here we have tested the utility of similar representations in modifying material appearance.

We have identified several kinds of material appearance that are commonly associated with certain subband properties and that can often be manipulated when the signal is already present in the image. In our experience, the observed perceptual changes are stable for materials where the albedo does not vary too rapidly, such as skin texture, leather texture, metals, and the texture of some fruits. For example, specularities from fine-scale features show up in positive-valued, high-amplitude, high-spatial frequencies. Boosting them leads to an enhanced “glistening” appearance which may be interpreted as oiliness or wetness on skin. Specularities from smooth, large-scale features show up in positive-valued, medium-spatial frequency coefficients. Boosting them leads to a smooth shine or sense of skin glow. Small spots, pits, and wrinkles typically manifest themselves as small dark features that show up in the high-amplitude negative coefficients of high-spatial frequencies. Boosting them often emphasizes the visibility of fine-scale texture (both fine-scale geometry and fine-scale albedo). These features are often associated with the aging of human skin, or the weathering of natural surfaces. In addition, the low-amplitude coefficients of the high-spatial frequencies (both negative and positive) are often associated with splotchy or mottled pigmentation. These can also enhance the

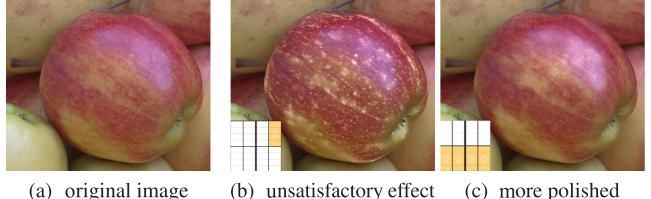


Fig. 14. Working on an object that has high-frequency albedo (a), our “shiny/glossy/metallic” band-sifting operator is less effective in conveying the more shiny look since it picks mainly on the albedo (b). Although one band-sifting operator may fail, others might still be useful. In (c) we manipulate material properties related to smoothness, which makes the apple look more polished. Original image courtesy of Bychkovsky et al. [2011].

sense of age, weathering, and discoloration. In the past, those coefficients were typically artifacts of the imaging chain, that is, sensor noise or image coding artifacts. However, modern digital cameras offer clean images, and the low-amplitude signals tell us about the scene, not about the camera. The perceptual effects of our operators are less effective in the presence of noise or lossy compression artifacts. We show an example of this in the supplemental material.

We find it remarkable that these effects tend to look natural and realistic rather than being the result of some artificial manipulation. The realism presumably results from the fact that the band-sifting operators are not inventing any information that is not already there; they are just emphasizing or de-emphasizing visual patterns that are already part of the image. As we have discussed earlier, when the visual cues are not present in the original input image or not well isolated by our sifting criteria, our band-sifting operators may fail to convey consistent perceptual effects. High-frequency albedo can also lead to unsatisfactory effects for band-sifting operators that manipulate those frequencies. In Figure 14, starting from an input image that has high-frequency texture (a), our band-sifting operator fails to produce a satisfactory “shiny” effect (b). Although one operator may fail on a certain image, other band-sifting operators might still be useful. In (c) we manipulate material properties related to smoothness by boosting the low-spatial frequency coefficients, which gives the apple a more polished look.

6. CONCLUSION AND FUTURE WORK

We present band-sifting operators and demonstrate their use in manipulating surface appearance. The band-sifting operators selectively alter coefficients within a subband decomposition, where the selection is based on spatial scale, sign, and amplitude. We explored a reasonable subspace of such operators and demonstrated their ability to modify a variety of surface properties in natural scenes. We use only 2D operations, but they can give the visual impression of acting on the materials in 3D scenes. We found some operators that were useful in controlling smoothness or gloss, which could alter the appearance of wetness, shininess, or degree of polish. Other operators altered the apparent pigmentation, roughness, or weathering of surfaces. We performed user studies and determined that there are certain operators that lead to consistent perceptual effects across various images and across multiple observers. Image class does matter: with images of faces, subjects reported that the filters would change face-specific properties such as oiliness, blemishes, wrinkles, skin age, and skin health. Given the importance of perceptual surface qualities, we expect that these band-sifting operators can offer an important tool for photography. Our band-sifting operators can also be used with video sequences. The visual effects tend

to be consistent across a sequence, making it possible, for example, to change the apparent shininess of an actor's skin. In the future, further exploration of band sifting, for example, by a finer subdivision of our proposed space or by introducing new sifting criteria, could open the door for many more operators for image-based material editing. Identifying the conditions when an operator would succeed or fail to produce a desired effect, based on the content of an arbitrary given image, is an interesting avenue for future work.

ELECTRONIC APPENDIX

The electronic appendix to this article can be accessed in the ACM Digital Library.

APPENDIX

In this appendix we provide a detailed pseudo-code implementation of our band-sifting operators. In practice all major steps can be implemented efficiently on a GPU using a fast summed-area table algorithm [Harris et al. 2007].

ALGORITHM 1: BANDSIFTINGOPERATOR

input: image I , multiplication factor λ , $sign \in \{\text{pos}, \text{neg}, \text{all}\}$, $freq \in \{\text{high}, \text{low}, \text{all}\}$, $amp \in \{\text{high}, \text{low}, \text{all}\}$
output: image O

- 1: $L \leftarrow \text{luminance}(I)$ // process only the L channel of Lab
- 2: $\{S_\ell\} \leftarrow \text{DECOMPOSE}(\log(L + \epsilon))$ // multiscale decomposition of log luminance (Alg. 2), use small ϵ to avoid 0 values
- 3: $n \leftarrow \#\{S_\ell\}$ // define n , the number of subbands
- 4: **for** all levels $\ell \in [1; n]$ **do**
- 5: $R \leftarrow S_\ell$ // keep a copy of the subband as reference
- 6: **for** all coefficients $c \in S_\ell$ **do**
- 7: // check if c selected by its sign and frequency (Alg. 3)
- 8: **if** SIGNANDFREQUENCYSELECTED($c, sign, freq$) **then**
- 9: // check if c is selected by its amplitude
- 10: **if** $amp = \text{all}$ **then**
- 11: $c \leftarrow \lambda c$ // if all amplitudes, directly apply multiplier
- 12: **else**
- 13: // else smooth transition between high and low amplitudes
- 14: $\sigma = \text{STDDEV}(S_\ell)$
- 15: $\alpha \leftarrow \text{SMOOTHSTEP}(0.8\sigma, 1.2\sigma, |c|)$ // returns 0 if $|c| < 0.8\sigma$, 1 if $> 1.2\sigma$; transitions in between
- 16: // orient transition depending on amplitude selection
- 17: **if** $amp = \text{high}$ **then**
- 18: $c \leftarrow c \times (1 + \alpha(\lambda - 1))$
- 19: **else if** $amp = \text{low}$ **then**
- 20: $c \leftarrow c \times (1 + (1 - \alpha)(\lambda - 1))$
- 21: **end if**
- 22: **end if**
- 23: **end if**
- 24: **end for**
- 25: // smooth applied gain map
- 26: $Gain \leftarrow S_\ell / R$
- 27: $Gain \leftarrow Gain \otimes G(2^\ell)$
- 28: $S_\ell \leftarrow R \times Gain$
- 29: **end for**
- 30: $O = \exp(\sum_1^n S_\ell) - \epsilon$ // sum subbands to get output

ALGORITHM 2: DECOMPOSE

input: single-channel image C
output: multiscale stack $\{S_\ell\}$

- 1: $n \leftarrow \log_2(\min(C.\text{width}, C.\text{height}))$ // number of layers
- 2: $Tmp_1 \leftarrow \text{GUIDEDFILTER}(C, 0.1^2, 2)$ // first operand is input image, second is range sensitivity, third is spatial extent
- 3: **for** $\ell = 1 \dots n$ **do**
- 4: $Tmp_2 \leftarrow Tmp_1$
- 5: $Tmp_1 \leftarrow \text{GUIDEDFILTER}(Tmp_2, 0.1^2, 2^\ell)$ // double the spatial extent each time
- 6: $S_\ell = Tmp_2 - Tmp_1$ // a subband is the difference of two successively filtered versions of the image
- 7: **end for**

ALGORITHM 3: SIGNANDFREQUENCYSELECTED

input: $c \in \mathbb{R}$, $sign \in \{\text{pos}, \text{neg}, \text{all}\}$, $freq \in \{\text{high}, \text{low}, \text{all}\}$
output: $isSelected \in \{\text{true}, \text{false}\}$

- 1: $signSelected \leftarrow sign = \text{all}$
or ($c < 0$ **and** $sign = \text{neg}$)
or ($c > 0$ **and** $sign = \text{pos}$)
- 2: $freqSelected \leftarrow freq = \text{all}$
or ($\ell \leq n/2$ **and** $freq = \text{low}$)
or ($\ell > n/2$ **and** $freq = \text{high}$)
- 3: $isSelected \leftarrow signSelected \text{ and } freqSelected$

REFERENCES

- Arqbball. 2014. Shoes. <https://arqspin.com/gallery/shoes/>.
- S. Bae, S. Paris, and F. Durand. 2006. Two-scale tone management for photographic look. *ACM Trans. Graph.* 25, 3, 637–645.
- S. Bell, P. Upchurch, N. Snavely, and K. Bala. 2013. OpenSurfaces: A richly annotated catalog of surface appearance. *ACM Trans. Graph.* 32, 4.
- S. Bell, P. Upchurch, N. Snavely, and K. Bala. 2015. Material recognition in the wild with the materials in context database. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'15)*.
- P. J. Burt and E. H. Adelson. 1983. The Laplacian pyramid as a compact image code. *IEEE Trans. Comm.* 31, 4, 532–540.
- V. Bychkovsky, S. Paris, E. Chan, and F. Durand. 2011. Learning photographic global tonal adjustment with a database of input/output image pairs. In *Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR'11)*. 97–104.
- J. Chen, S. Paris, and F. Durand. 2007. Real-time edge-aware image processing with the bilateral grid. *ACM Trans. Graph.* 26, 3.
- P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar. 2000. Acquiring the reflectance field of a human face. In *Proceedings of the 27th Annual ACM Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'00)*. ACM Press/Addison-Wesley, New York, 145–156.
- P. E. Debevec and J. Malik. 1997. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th Annual ACM Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'97)*. ACM Press/Addison-Wesley, New York, 369–378.
- D. Donoho. 1995. De-noising by soft-thresholding. *IEEE Trans. Inf. Theory*. 41, 3, 613–627.
- F. Durand and J. Dorsey. 2002. Fast bilateral filtering for the display of high-dynamic range images. *ACM Trans. Graph.* 21, 3.
- T. Ewan. 2009. Orange. https://www.flickr.com/photos/the_ewan/3958779607/.

- Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski. 2008. Edge-preserving decomposition for multi-scale tone and detail manipulation. *ACM Trans. Graph.* 27, 3.
- R. Fattal. 2009. Edge-avoiding wavelets and their applications. *ACM Trans. Graph.* 28, 3.
- R. Fattal, M. Agrawala, and S. Rusinkiewicz. 2007. Multiscale shape and detail enhancement from multi-light image collections. *ACM Trans. Graph.* 26, 3.
- R. W. Fleming and H. H. Bültlhoff. 2005. Low-level image cues in the perception of translucent materials. *ACM Trans. Appl. Percept.* 2, 3.
- R. W. Fleming, C. Wiebel, and K. Gegenfurtner. 2013. Perceptual qualities and material classes. *J. Vis.* 13, 8, 9.
- E. S. L. Gastal and M. M. Oliveira. 2011. Domain transform for edge-aware image and video processing. *ACM Trans. Graph.* 30, 3.
- E. S. L. Gastal and M. M. Oliveira. 2012. Adaptive manifolds for real-time high-dimensional filtering. *ACM Trans. Graph.* 31, 4.
- L. Glondou, L. Muguerza, M. Marchal, C. Bosch, H. Rushmeier, G. Dumont, and G. Drettakis. 2012. Example-based fractured appearance. *Comput. Graph. Forum* 31, 4.
- A. Golovinskiy, W. Matusik, H. Pfister, S. Rusinkiewicz, and T. Funkhouser. 2006. A statistical model for synthesis of detailed facial geometry. *ACM Trans. Graph.* 25, 3.
- J. Hanika, H. Dammertz, and H. P. A. Lensch. 2011. Edge-optimized a-trous wavelets for local contrast enhancement with robust denoising. *Comput. Graph. Forum* 30, 7.
- M. Harris, S. Sengupta, and J. D. Owens. 2007. Parallel prefix sum (scan) with CUDA. In *GPU Gems 3*, H. Nguyen, Ed., Addison-Wesley, 851–876.
- W. Hartshorn. 2008. Sweet potatoes. <https://www.flickr.com/photos/wallyhartshorn/2850269642/>.
- K. He, J. Sun, and X. Tang. 2010. Guided image filtering. In *Proceedings of the European Conference on Computer Vision (ECCV'10)*.
- D. J. Heeger and J. R. Bergen. 1995. Pyramid-based texture analysis/synthesis. In *Proceedings of the Annual ACM Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'95)*. ACM Press, New York.
- R. Jill. 2006. Self portrait. <https://www.flickr.com/photos/98972490N04/958207891>.
- E. Judy. 2009. Good grape crop. <https://www.flickr.com/photos/659247400N00/5140439795/>.
- L. Karacan, E. Erdem, and A. Erdem. 2013. Structure preserving image smoothing via region covariances. *ACM Trans. Graph.* 32, 6.
- E. A. Khan, E. Reinhard, R. W. Fleming, and H. H. Bültlhoff. 2006. Image-based material editing. *ACM Trans. Graph.* 25, 3.
- J. Kim, P. Marlow, and B. L. Anderson. 2011. The perception of gloss depends on highlight congruence with surface shading. *J. Vis.* 11, 9, 4.
- E. H. Land and J. Mccann. 1971. Lightness and retinex theory. *J. Optic. Soc. Amer.* 1–11.
- C. Liu, L. Sharan, E. H. Adelson, and R. Rosenholtz. 2010. Exploring features in a Bayesian framework for material recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'10)*. 239–246.
- S. P. Mallick, T. Zickler, P. N. Belhumeur, and D. J. Kriegman. 2006. Specularity removal in images and videos: A PDE approach. In *Proceedings of the European Conference on Computer Vision (ECCV'06)*. 550–563.
- R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich. 2011. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph.* 30, 4.
- R. Mantiuk, K. Myszkowski, and H.-P. Seidel. 2006. A perceptual framework for contrast processing of high dynamic range images. *ACM Trans. Appl. Percept.* 3, 3, 286–308.
- P. J. Marlow and B. L. Anderson. 2013. Generative constraints on image cues for perceived gloss. *J. Vis.* 13, 14, 2.
- P. J. Marlow, J. Kim, and B. L. Anderson. 2012. The perception and misperception of specular surface reflectance. *Curr. Biol.* 22, 20, 1909–1913.
- T. Mertens, J. Kautz, J. Chen, P. Bekaert, and F. Durand. 2006. Texture transfer using geometry correlation. In *Proceedings of the Eurographics Conference on Rendering Techniques (EGSR'06)*. 273–284.
- I. Motoyoshi, S. Nishida, L. Sharan, and E. H. Adelson. 2007. Image statistics and the perception of surface qualities. *Nature* 447, 206–209.
- Ngan A, F. Durand, and W. Matusik. 2006. Image-driven navigation of analytical BRDF models. In *Proceedings of the Eurographics Conference on Rendering Techniques (EGSR'06)*. 399–407.
- F. Osu. 2012. How to answer “behavior based interview questions” - Interview tip. <https://www.youtube.com/watch?v=qKBubK0-798>.
- S. Paris, S. W. Hasinoff, and J. Kautz. 2011. Local Laplacian filters: Edge-aware image processing with a Laplacian pyramid. *ACM Trans. Graph.* 30, 4.
- G. Patterson, C. Xu, H. Su, and J. Hays. 2014. The sun attribute database: Beyond categories for deeper scene understanding. *Int. J. Comput. Vis.* 108, 1–2, 59–81.
- J. Portilla and E. P. Simoncelli. 2000. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. J. Comput. Vis.* 40, 1, 49–70.
- L. Sharan, Y. Li, I. Motoyoshi, S. Nishida, and E. H. Adelson. 2008. Image statistics for surface reflectance perception. *J. Optic. Soc. Amer.* A25, 4.
- L. Sharan, C. Liu, R. Rosenholtz, and E. H. Adelson. 2013. Recognizing materials using perceptually inspired features. *Int. J. Comput. Vis.* 103, 3, 348–371.
- E. P. Simoncelli and E. H. Adelson. 1996. Noise removal via Bayesian wavelet coring. In *Proceedings of the IEEE International Conference on Image Processing (ICIP'96)*. 379–382.
- E. Stochter. 2013. Gargoyle. <https://pixabay.com/en/gargoyle-fountain-lion-lion-head-205999/>.
- C. Tomasi and R. Manduchi. 1998. Bilateral filtering for gray and color images. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV'98)*.
- M. Trentacoste, R. Mantiuk, and W. Heidrich. 2011. Blur-aware image downsizing. *Comput. Graph. Forum* 30, 2, 573–582.
- M. Trentacoste, R. Mantiuk, W. Heidrich, and F. Dufrot. 2012. Unsharp masking, countershading and halos: Enhancements or artifacts? *Comput. Graph. Forum* 30, 2, 573–582.
- R. Vergne, P. Barla, R. Fleming, and X. Granier. 2012. Surface flows for image-based shading design. *ACM Trans. Graph.* 31, 3.
- Videvo. 2014. Fresh grapes close up. <http://www.videvo.net/video/fresh-grapes-close-up/2350>.
- J. Von Der Pahlen, J. Jimenez, E. Danvoye, P. Devevec, G. Fyffe, and O. Alexander. 2014. Digital IRA and beyond: Creating real-time photoreal digital actors. In *ACM SIGGRAPH Courses*. ACM Press, New York, 1:1–1:384.
- G. J. Ward. 1992. Measuring and modeling anisotropic reflection. *ACM Trans. Comput. Graph.* 26, 2, 265–272.
- L. West. 2006. Onion. <https://www.flickr.com/photos/calliope/104314753/>
- L. Xu, C. Lu, Y. Xu, and J. Jia. 2011. Image smoothing via L0 gradient minimization. *ACM Trans. Graph.* 30, 6.
- L. Xu, Q. Yan, Y. Xia, and J. Jia. 2012. Structure extraction from texture via relative total variation. *ACM Trans. Graph.* 31, 6.

Received December 2014; revised June 2015; accepted July 2015