Tone Reproduction: A Perspective from Luminance-Driven Perceptual Grouping

HWANN-TZONG CHEN

Institute of Information Science, Academia Sinica, Nankang, Taipei 115, Taiwan;
Department of CSIE, National Taiwan University, Taipei 106, Taiwan

pras@iis.sinica.edu.tw

TYNG-LUH LIU

Institute of Information Science, Academia Sinica, Nankang, Taipei 115, Taiwan

liutyng@iis.sinica.edu.tw

CHIOU-SHANN FUH

Department of CSIE, National Taiwan University, Taipei 106, Taiwan

fuh@csie.ntu.edu.tw

Abstract.   We address the tone reproduction problem by integrating the local adaptation effect with the consistency in global contrast impression. Many previous works on tone reproduction have focused on investigating the local adaptation mechanism of human eyes to compress high-dynamic-range (HDR) luminance into a displayable range. Nevertheless, while the realization of local adaptation properties is not theoretically defined, exaggerating such effects often leads to unnatural visual impression of global contrast. We propose to perceptually decompose the luminance into a small number of regions that sequentially encode the overall impression of an HDR image. A piecewise tone mapping can then be constructed to region-wise perform HDR compressions, using local mappings constrained by the estimated global perception. Indeed, in our approach, the region information is used not only to practically approximate the local properties of luminance, but more importantly to retain the global impression. Besides, it is worth mentioning that the proposed algorithm is efficient, and mostly does not require excessive parameter fine-tuning. Our experimental results and comparisons indicate that the described framework gives a good balance in both preserving local details and maintaining global perceptual impression of HDR scenes.

1. Introduction

The attempt to reproduce the visual perception of the real world is at the heart of painting and photography. Artists have long been endeavoring to develop skills in simulating actual reflected light within the limitation of the medium, since our world generally delivers a much wider range of luminance than pigments can reflect. Apart from artistic concern, recreating real-scene impressions on limited media is also inevitable in many vision and graphics applications (see Tumblin and Rushmeier, 1993 for a more detailed discussion). For example, the contrast of a today's fine LCD monitor could be around 1,000 to 1, but we may need to use it to display an indoor scene comprising very *bright* windows that results in a contrast of 250,000 to 1.

By a tone reproduction problem, we are interested in establishing a method to *satisfactorily* reconstruct the

high dynamic range (HDR) radiance on a low dynamic range (LDR) image. The *dynamic range* of a digital image typically refers to the contrast ratio between its brightest and darkest parts, though a more comprehensive definition should include issues such as noise sensitivity and possible nonlinear distortion in encoding the pixel values. In Ward (1991), Ward proposes a floating-point picture format to record HDR radiance in 32 bits per pixel, and designs a graphics rendering system to output images in the format. Debevec and Malik (1997) have shown that HDR radiance of *real* scenes may also be captured using regular digital cameras. They describe a technique to combine a series of pictures taken with different exposure settings into a single HDR image, called the *radiance map*. In this context, the focus of our work can be stated as solving the tone reproduction problem of radiance maps, that is, generating a displayable standard RGB image that preserves visually significant properties of the original HDR radiance map.

1.1. Related Work

Several previous works have been devoted to producing HDR images of real scenes. The method by Debevec and Malik (1997) requires multiple photos of a scene taken under different exposures. To simultaneously capture luminance under multiple exposures, Nayar and Mitsunaga (2000) develop a camera, including a mask with cells of different optical transparencies. The grid mask provides different exposures in neighboring pixels, which are averaged to output an HDR image. Naturally, panoramic imaging also involves high dynamic range. The process of collecting a broader range of luminance at each location can be coupled with the process of constructing a wider field of view by adding spatially varying optical filters to a camera (Aggarwal and Ahuja, 2001; Schechner and Nayar, 2003). Yet, on a different focus, Kang et al. (2003) propose to generate HDR videos by varying the exposure of each frame and then by stitching consecutive frames.

Concerning displaying HDR images with good perceptual fidelity, tone reproduction addresses *visibility* and *impression* by finding an appropriate mapping to compress the contrast into a displayable range. For the convenience of our discussion, we categorize tone-mapping functions of related work into global mappings and local mappings, according to whether spatial information is incorporated in their respective formulations.

1.1.1. Global Tone Mapping. Global tone mappings are spatially uniform functions. They uniformly compress the luminance values of pixels into a displayable range, regardless of pixels' spatial or local properties. The great advantage of using a global mapping is its efficiency Tumblin and Rushmeier, 1993. Since global mappings are often conveniently designed as monotonically increasing, these functions therefore preserve the order of pixel luminance values. Such a property gives another advantage of using global mappings— for avoiding halo artifacts, which are undesirable in displaying HDR images. Halos are due to reversed contrasts in a neighborhood, i.e., an area originally darker than its neighbors becomes brighter after tone mapping, or vice versa. However, strictly and globally preserving the order of luminance is not always necessary (though sufficient) for avoiding halos, and may not be preferable in some circumstances. Every so often we may want to emphasize local contrasts to display more details, even at the risk of re-arranging the global order of luminance, as long as no perceivable artifacts are produced. Ward et al. (1997) describe a more sophisticated approach to globally adjust contrast with respect to luminance histograms. Nevertheless, their method still smooths out image details in areas of flat histograms.

1.1.2. Local Tone Mapping. To better preserve important visual features, such as image details, of high contrast scenes, several tone reproduction methods have exploited local (or spatially nonuniform) mappings, as human visual system operates more likely this way. Particularly, human visual cells are organized in a center-surround manner so that we can see a broad range of luminance values by discriminating locally (Palmer, 1999). Simulating the local adaptation effect is a common tactic to keep image details for tone mapping. Its basic idea is to attenuate high contrast by computing locally-averaged luminance values for effectively adjusting the tone-mapping parameters, e.g., the scale factor (Chiu et al., 1993; Tumblin et al., 1999), or the luminance gain control (Pattanaik et al., 1998). Since the term *adaptation luminance* has already been used in global tone-mapping methods (Ferwerda et al., 1996; Tumblin and Rushmeier, 1993; Ward, 1994), the locally-averaged luminance in Ashikhmin (2002) and Yee and Pattanaik (2003) is referred to as *local adaptation luminance* to emphasize its spatially-varying property. A related approach to non-uniformly reducing pixels' luminance values relies on extracting the

illumination from an HDR image (Kimmel et al., 2003; Tumblin et al., 1999). Since the estimated illumination usually has a wider dynamic range, a displayable image can be obtained by reducing the dynamic range via compressing the illumination, and then by recombining the compressed one with the rest of the original HDR image.

When displaying synthesized HDR images, the problem of extracting the layers of illumination and surface properties is inherently solved because images in graphics are usually rendered from illumination and reflectance (e.g., Tumblin et al., 1999). In dealing with images of natural scenes, Land's Retinex theory (Land and McCann, 1971) can be used to compute the illumination, and has been implemented with multi-scale Gaussian filters (Jobson, 1997). There are other tone-mapping methods also using multi-scale Gaussian filters to derive the locally-averaged luminance for simulating the local adaptation effect (Ashikhmin, 2002; Pattanaik et al., 1998), or the photographic process (Reinhard et al., 2002). The concern here is that since the global order of luminance is no longer guaranteed by local tone mappings, strategies to avoid halo artifacts must be explicitly formulated. To gain insight into this issue, we look at a scenario of applying a large-scale Gaussian filter to smooth two neighboring and monotone areas that occupy significantly different luminance levels. Then, tone mapping with the resulting over-smoothed adaptation luminance would exaggerate the contrast near the boundary of the two areas, and consequently yield halos. The example suggests that choosing an appropriate scale of the smoothing operator is critical for computing the locally-averaged luminance and avoiding halos. Ashikhmin (2002) and Reinhard et al. (2002) adopt similar schemes to determine the scales of Gaussian filters. At each pixel location, they gradually increase the scale of a Gaussian filter until that large contrasts are encountered.

Besides multi-scale Gaussians, *anisotropic diffusion* is another popular choice as the smoothing operator for computing the locally-averaged luminance. Tumblin and Turk (1999) establish the *low curvature image simplifier* (LCIS) to blur an HDR image without smearing the edges. Anisotropic diffusion can also be implemented in the form of *bilateral filtering* (Tomasi and Manduchi, 1998). Specifically, a bilateral filter includes two weights (such as Gaussians): one for the spatial domain, and the other for the range domain to explain intensity differences between each pixel and its neighbors. Nearby areas with dissimilar

luminance values thus yield small weights, which prevent from blurring the edges. DiCarlo and Wandell (2000) have applied this concept to HDR tone mapping. More recently, Durand and Dorsey (2002) develop a fast bilateral filtering to efficiently compute the locally-averaged luminance. Though seemingly different from the methods mentioned above, the gradient-domain approach of Fattal et al. (2002) is also a local tone-mapping. Their formulation emphasizes attenuating large gradients, and recovers the luminance by solving a Poisson equation.

Since estimating local adaptation luminance relates to the issues of how to separate dissimilar areas and how to determine an adequate region for averaging, it is worthwhile to address these problems from a segmentation viewpoint (Krawczyk et al., 2004; Schlick, 1994; Yee and Pattanaik, 2003). Schlick (1994) has proposed to divide the picture into zones of similar intensity values and then compute the average of each zone. He notes that although image segmentation is itself a challenging one, the segmentation problem in tone reproduction is a simplified case—on a gray-level image with few regions. Schlick further suggests to use gradient or histogram thresholding techniques to partition the image, but concludes without elaborating further details in Schlick (1994) that the improvement on anti-aliasing around high contrast is imperceptible. Yee and Pattanaik (2003) develop a multi-layer partitioning and grouping algorithm to compute the local adaptation luminance. Their method computes *layers* under different intensity resolutions, i.e., to quantize intensities by different bin widths, and then for each intensity resolution, to collect pixels of the same bin into a same group. Each pixel's value in the local adaptation luminance can be obtained by averaging the values of all pixels from the groups in different layers with which the underlying pixel is associated. Though not directly related to the problem of HDR tone reproduction, the anchoring model proposed by Gilchrist et al. (1999) also considers the connection between the decomposition of an image and the perception of lightness. They propose to decompose a complex image into multiple local *frameworks* (in terms of the *Gestalt grouping principles*, e.g., see Palmer, 1999, p. 257). The perceived reflectance is estimated by combining the local and global anchoring properties according to the decomposed frameworks. For practically implementing the model of Gilchrist et al., the main difficulties to be overcome include finding the right grouping factors to reasonably explain lightness

perception, and establishing a computational approach to suitably decompose an image for the anchoring.

1.2. Our Approach

We aim to develop an efficient tone-mapping operator to compress the HDR luminance, accounting for the revealing of image details and the preserving of visual impressions. Often these two effects are hard to be dealt with in a compatible way, but rather as two competitive factors in formulating a tone-mapping function. More precisely, since only a limited dynamic range is allowed to display an image, achieving a good trade-off between local and global contrast is not always straightforward. This difficulty does reflect in our survey that most recent approaches have focused mainly on local adaptation properties of tone reproduction (Ashikhmin, 2002; Durand and Dorsey, 2002; Fattal et al., 2002; Reinhard et al., 2002; Yee and Pattanaik, 2003), but somewhat overlooked an equally important issue about maintaining the consistency in global contrasts. Motivated by this aspect of consideration, we propose to perceptually decompose the luminance into a small number of regions that encode the overall impression based on the sequential order of luminance levels, and thus term the process as *perceptual grouping* for deriving a *sparse decomposition* of an HDR image. Accordingly we gain the advantage of concentrating on designing a local tone mapping for each region, without the burden of worrying about breaking the overall impressions carried by the original luminance. Furthermore, it becomes feasible to estimate the local adaptation luminance with the perceptual-based region information. Having established the piecewise tone mapping, we can conveniently perform region-wise HDR compressions using local mappings constrained by the desirable global property, and produce a displayable LDR image of good quality.

In the following sections, we shall describe first how we derive the sparse decomposition of an HDR image based on a luminance-driven grouping process. We also discuss how the region information is embedded into the computation of the local adaptation luminance. We then explain how we region-wise construct a local tone-mapping function, and modify the mapping with the global property encoded in the decomposition. Finally, we conclude with several experimental results and comparisons to illustrate the characteristics of HDR compressions yielded by our method.

2. A Sparse Decomposition for HDR Images

The important link between visual grouping and lightness perception in human visual system has been empirically studied in, e.g., Adelson (2000) and Gilchrist et al. (1999). Still an attempt to devise overall corresponding computational models that explain the theories of lightness perception and the responses of human visual system would be extremely challenging, though worthwhile. For the problem of HDR tone reproduction, our formulation is not to build a complicated algorithm emulating the machinery/process of human visual system, but mainly to establish an efficient method yielding satisfactory results based on useful properties of the proposed perceptual grouping.

To reduce the dynamic range by tone mapping, we decompose an HDR image into a small number of regions through a grouping process on luminance, using a perceptual similarity measure for image retrieval developed by Rubner and Tomasi (2001) and Rubner et al. (2000). We regard such a result of perceptual grouping on luminance as a sparse decomposition for an HDR image. Besides giving preference to fewer regions, the eventual decomposition would comprise regions that are brighter inside and darker near the boundary. As we will elaborate later, these two properties are essential to our approach that locally and globally considers the reduction of dynamic range.

A number of works have explored the idea to compute the local adaptation luminance by referencing image segmentation (e.g., Krawczyk et al., 2004; Schlick, 1994; Yee and Pattanaik, 2003). In particular, while Schlick (1994) and Krawczyk et al. (2004) take advantage of existing segmentation techniques, Yee and Pattanaik (2003) specifically develop a fast segmentation algorithm for computing local adaptation luminance. They describe a multi-layer grouping process (quantized with multiple bin widths in the intensity domain) to obtain the locally-averaged luminance value of each pixel. Nevertheless, all these aforementioned approaches simply focus on how to estimate local adaptation luminance. None of them seeks to use the region or segmentation information for modulating tone-mapping functions.

In our formulation the region information is used not only to compute local adaptation luminance values, but also to establish more comprehensive tone-mapping functions for maintaining global contrasts. We shall introduce a new perceptual grouping algorithm that can fast construct the sparse decomposition for an HDR
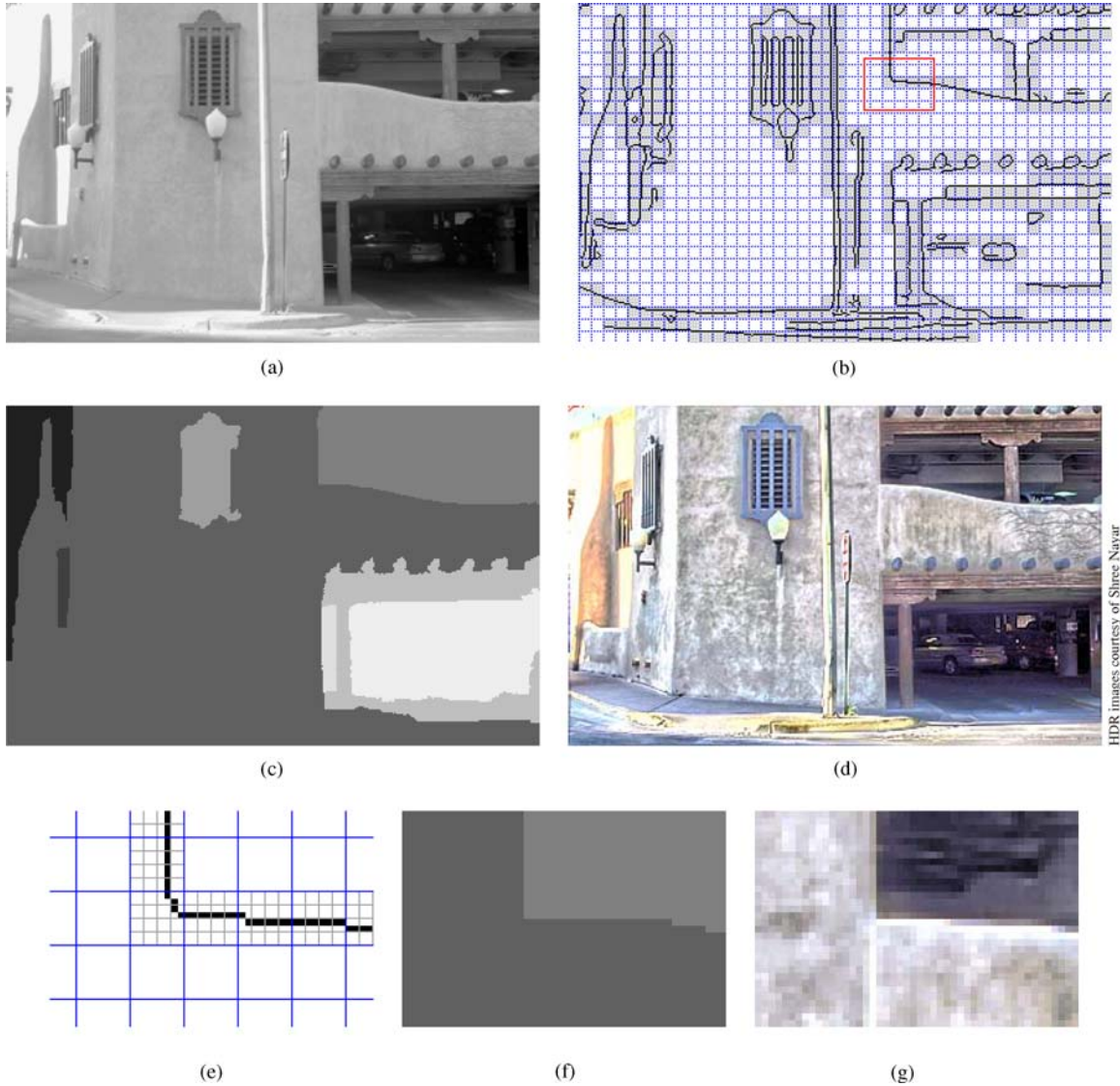
Figure 1. Garage. (a) The corresponding log-luminance $\tilde{L}$. Note that values of $\tilde{L}$ are normalized to [0, 1] for displaying. (b) The adaptive block partition of $\tilde{L}$, where blocks of the smaller size are located in those shaded areas on top of the Canny edges. (c) A sparse decomposition of 8 regions for $\tilde{L}$. (d) The displayable garage image derived by our method. (e) A close-up of the block partition inside the red rectangle. (f) The corresponding regions in close-up. (g) The resulting displayable image patch in close-up.

image. The advantages of using the proposed decomposition to encode the overall perceptual impression will be thoroughly discussed in Section 3. Suffice it to say now that working with an adequate number of regions, we can perform the HDR compression without incurring excessive overheads in assembling the results across different regions. Furthermore, with the region information, we can devise tone-mapping functions accounting for local and global contrast adjustments. On deriving such a sparse/compact decomposition, it takes

two steps: *adaptive block partitioning* and *perceptual grouping*, detailed in what follows.

2.1. Adaptive Block Partitioning

The luminance values of an HDR image can be computed from its R, G, and B channels by

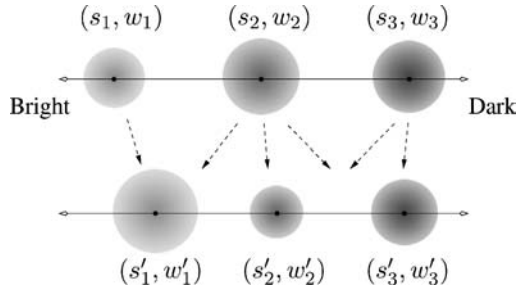$$L(x, y) = 0.2126R(x, y) + 0.7152G(x, y) + 0.0722B(x, y), \qquad (1)$$

Figure 2. An illustration of a possible transformation from one set of weighted features to another. The EMD between these two feature sets is the minimum amount of "work" (with respect to $\|s_i - s'_j\|$ and the weights) needed to transform $\{(s_1, w_1), (s_2, w_2), (s_3, w_3)\}$ to $\{(s'_1, w'_1), (s'_2, w'_2), (s'_3, w'_3)\}$, e.g., moving the whole of w_1 from s_1 to s'_1, moving parts of w_2 from s_2 to s'_1, s'_2, and s'_3, and moving w_3 from s_3 to s'_2 and s'_3. Note that in our algorithm we have $\sum_i w_i = \sum_i w'_i = 1$, and the transformation is symmetric.

where the sRGB color space and a gamma of 2.2 are assumed (ITU, 1990; Stokes et al., 1996). Previous experience on studying perception has suggested that human visual system senses the contrast of light based on *intensity ratio* rather than *intensity difference* (e.g., the Weber's law discussed in Palmer 1999, p. 672). Following this observation, we consider the decomposition of an HDR image by examining its luminance property in the logarithmic domain. More precisely, the luminance L is transformed into log-luminance by

$$\tilde{L}(x, y) = \log L(x, y). \tag{2}$$

A modification to the scene's key value, as is used in Reinhard et al. (2002), has been performed on the (original) log-luminance. Its effect is equivalent to shifting the log-luminance values for providing a preferred middle-gray tone, a technique commonly used by photographers. Notice that the adjustment will not affect the result of the perceptual grouping process we are about to discuss, because the difference in log-luminance values between any two pixels remains the same. Hence, we will assume $\tilde{L}$ denotes the log-luminance after the key value adjustment hereafter.

While partitioning on a pixel level is both time-consuming and sensitive to noise, we also find working with blocks of uniform size often leads to unsatisfactory segmentation results. Though the situation could be improved by using small-size blocks, such a tactic again has the drawback of inefficiency. We thus design an adaptive scheme to partition the image with blocks of two different sizes. The smaller blocks are placed in the areas of strong log-luminance gradients; the larger ones are in the areas of less log-luminance variation. Precisely we use Canny edge detector (Canny, 1986) to obtain the edge information, and then divide $\tilde{L}$ into blocks of larger size $b_\ell \times b_\ell$. For those image blocks containing Canny edges, they are further split into blocks of smaller size $b_s \times b_s$. A complete example to illustrate these steps is given in Fig. 1 with block sizes $b_\ell = 8$ and $b_s = 2$. Note that the two specified sizes are used in all our experiments reported throughout this work too.

2.2. Luminance-Driven Perceptual Grouping

Among the many possible ways to group the blocks, we are interested in finding a sparseness one, driven

(a) Local adaptation log-luminance $\tilde{V}$ (b) Local adaptation luminance V

Figure 3. Intermediate stages of displaying **garage**: $\tilde{V}$ and V. Both images are normalized to fit into the range between 0 and 1 for displaying. Generally, a local adaptation luminance takes a significant portion of the dynamic range but leaves out the details.
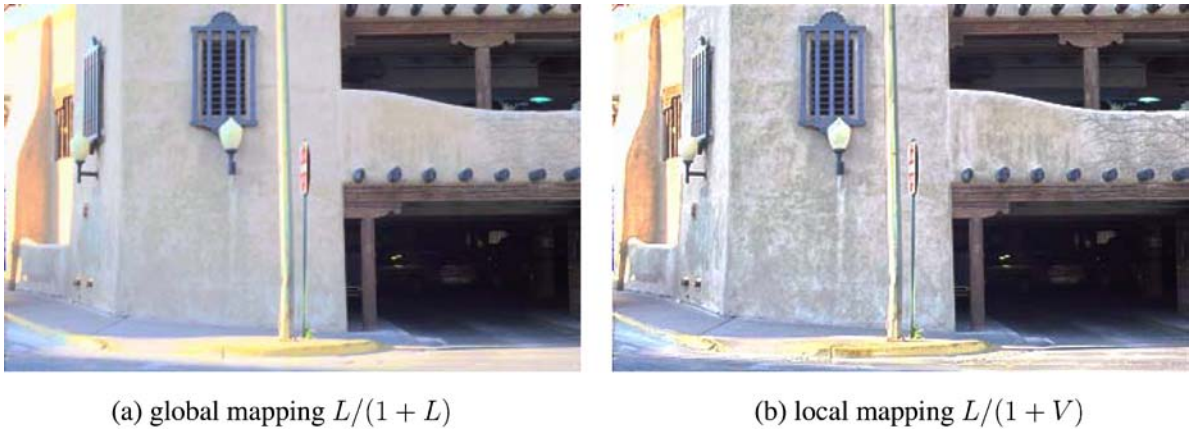
(a) global mapping $L/(1 + L)$

(b) local mapping $L/(1 + V)$

Figure 4. Comparison of the global and local mappings. The result of local mapping provides more details of the garage image. However, in both (a) and (b), the top-right and bottom-right areas (inside the garage) are still too dark.
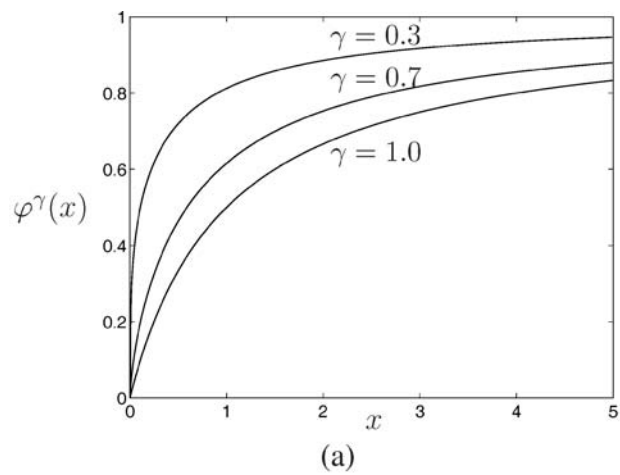

(a)

(b) $(L/(1 + L))^{0.3}$

(c) $(L/(1 + V))^{0.3}$

Figure 5. (a) The mapping $\varphi^{\gamma}(x) = (x/(1 + x))^{\gamma}$ with different γ values. (b) and (c) The results of garage by $\varphi^{0.3}(L)$ and $\psi(L, V; \rho = 0.3, \gamma = 0.3)$, cf. Figs. 4(a) and 4(b). Although the details in the bottom-right area are visible now, the dynamic range of the rest area is also *squeezed*.
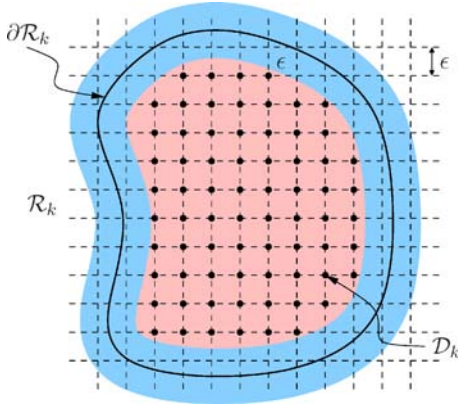
Figure 6. The setting for estimating ρ in $\mathcal{R}_k$. Let $\mathcal{D}_k$ denote the set of $\epsilon \times \epsilon$ grid points that are inside the boundary $\partial\mathcal{R}_k$, and for all points $\mathbf{x} \in \mathcal{D}_k$ we have $\min_{\mathbf{y}\in\partial\mathcal{R}_k} \|\mathbf{x} - \mathbf{y}\|_2 \geq \epsilon$. Later on, we will assign pixels $n \in \mathcal{D}_k$ some preliminary ρ_n values from which the ρ values for all pixels in $\mathcal{R}_k$ will be derived.

by examining the log-luminance values. For that, we look at two important matters: (1) how to define an appropriate distance function to measure the degree of similarity between two regions, and (2) how to proceed with a reasonable grouping process to derive the decomposition. Our choice to seek a sparse decomposition for segmenting an HDR image is mostly due to computational concern: We would like to region-wise perform tone mappings, and then integrate them into the final solution. It is thus useful to partition the image into a few (perceptually) significant regions, provided that the decomposition reasonably encodes the overall impression.

2.2.1. Perceptual Distance. We apply the *earth mover's distance* (EMD) to evaluate the perceptual similarity between two image regions (Rubner and Tomasi, 2001; Rubner et al., 2000). Used mostly in image retrieval, EMD has been shown to be an effective measure to perceptually correlate two images. The physiological experiments in Allred et al. (2003, 2004) further support that the image similarity measured by EMD pertains to the neural responses elicited by the same set of images. The EMD between two sets of weighted features is computed as the minimum amount of "work" (pertaining to the weights and the Euclidean distances between features) that is required to transform one feature set to the other (see Fig. 2). In our formulation, we generate features by equally dividing the dynamic range of a region, which could be an image block or a set of neighboring blocks,

into three bins. The mean value s_i and the number h_i of the pixels in each bin are then calculated. It follows that the weighted-feature set $\mathbf{p}$ of each region contains three pairs $\{(s_1, w_1), (s_2, w_2), (s_3, w_3)\}$ where $w_i = h_i / \sum_{j=1}^{3} h_j$, and they accordingly represent the *bright*, the *middle-gray*, and the *dark* parts of that region. Clearly, the weighted-feature description is more general than the histogram representation in that the feature s_i is not necessary the bin center. Details about the computation and the analysis of EMD can be found in Rubner et al. (2000). For our formulation, the perceptual distance between two regions $\mathcal{R}_1$ and $\mathcal{R}_2$ is defined by

$$D(\mathcal{R}_1, \mathcal{R}_2) = \text{EMD}\,(\mathbf{p}_1, \mathbf{p}_2), \qquad (3)$$

where $\mathbf{p}_i$ is the weighted-feature set of region $\mathcal{R}_i$.

2.2.2. Luminance-Driven Grouping. While analytic arguments on how to optimally decompose an image are difficult to establish, we prefer a compact/sparse representation Barlow (1961) to decompose an image into *a few* regions (see Fig. 1(c)). Specifically, we adopt a greedy approach to grow a new region as large as possible, starting each time from the location of the *brightest* log-luminance value in the unvisited areas. That is, the algorithm follows a *brightest-block-first* rule to determine the seeds and to merge image blocks. Since the feature set of each block includes three clusters, namely, the bright, the middle-gray, and the dark parts, the brightest block can be simply identified as the one with the largest s_1 value. Unlike ours, other grouping algorithms for HDR luminance such as Yee and Pattanaik (2003) and Krawczyk et al. (2004) do not specify the order of accomplishing the process. Owing to the brightest-block-first rule, each region obtained by our algorithm tends to be brighter in the inner area and darker in the peripheral. This property will be further exploited in regulating tone-mapping functions and smoothly piecing together adjacent regions.

Our grouping process proceeds as follows. All image blocks are initially marked as *unvisited*. Later on as the grouping process iterates, the number of unvisited blocks decreases. At iteration k, we pick the brightest and unvisited block, say, block $\mathcal{B}$, and start to grow the region $\mathcal{R}_k$ from it. These steps of region grouping are summarized in Algorithm 1. Upon termination, the process will yield a decomposition that each derived region consists of connected blocks of similar luminance distributions. Indeed, our algorithm works

by balancing the local and global similarity in a region. Similar blocks are pulled into the same region if the EMD between two neighboring blocks is smaller than δ. On the other hand, a region will stop growing when all blocks right beside the region boundary are not close enough to the region within θ. (See Algorithm 1 for details.) The typical values of EMD threshold θ are from 1.5 to 2.0, and those of δ are between 0.5 and 1.0.

2.3. Local Adaptation Luminance

Human visual system attains the HDR perception by locally adapting to different levels of luminance to ensure a proper dynamic range can be recreated by the responses of visual cells. For tone reproduction, the local adaptation effect is often simulated by computing the local adaptation luminance, and then using the information to extract the details and to reduce the dynamic range. Thus it is important to have a reliable way for pixel-wise estimating the local adaptation luminance

of an HDR image. And that in turn can be done by investigating the average log-luminance of a *suitable neighborhood* about each pixel. As we will see, a good choice of a suitable neighborhood to this purpose is the sparse decomposition derived in Section 2.2.

Let $\tilde{V}(x, y)$ be the local adaptation log-luminance at pixel (x, y). To compute $\tilde{V}$, we consider a generalized version of *bilateral filtering* (Durand and Dorsey, 2002; Tomasi and Manduchi, 1998) by constructing a region-dependent scheme such that the computation of $\tilde{V}(x, y)$ takes account of bilateral effects from different regions. Specifically, for each pixel (x, y) in region $\mathcal{R}_k$, we have

$$
\tilde{V}(x, y) = \frac{1}{\tilde{Z}_{x,y}} \left\{ \sum_{(i,j) \in \mathcal{R}_k} \tilde{L}(i, j)\, G_{x,y}(i, j)\, K_{x,y}(i, j) \right.
$$
$$
\left. + \sum_{(i,j) \notin \mathcal{R}_k} \tilde{L}(i, j)\, G_{x,y}(i, j)\, K'_{x,y}(i, j) \right\},
$$
(4)

Algorithm 1: Luminance-Driven Grouping with EMD.

Input : The adaptive block partition $\mathcal{B}$ of $\tilde{L}$.

Output : A sparse decomposition $\mathcal{R}$ of $\tilde{L}$.

Initialization: Create an empty priority queue $\mathcal{Q}$ on D as in (3); Compute the
weighted-feature sets of all blocks; Label all blocks *unvisited*; Let $k \leftarrow 1$;
Choose $\theta, \delta > 0$.

while $\exists$ *unvisited block* **do**
    Create a new region $\mathcal{R}_k \leftarrow \emptyset$, and let $\mathcal{Q} \leftarrow \emptyset$;
    Select the brightest, *unvisited* block $\mathcal{B}_{i^*}$ and add it into $\mathcal{Q}$;
    Let $D(\mathcal{B}_{i^*}, \mathcal{R}_k) \leftarrow 0$;
    while $\mathcal{Q} \neq \emptyset$ **do**
        Retrieve $\mathcal{B}_i$ from $\mathcal{Q}$ with the smallest $D(\mathcal{B}_i, \mathcal{R}_k)$;
        if $D(\mathcal{B}_i, \mathcal{R}_k) < \theta$ **then**
            Label $\mathcal{B}_i$ *visited*, and add $\mathcal{B}_i$ into region $\mathcal{R}_k$;
            Update the feature set of region $\mathcal{R}_k$;
            Update $\mathcal{Q}$ by recomputing the Ds according to the new feature set of $\mathcal{R}_k$;
            Find $\mathcal{B}_i$'s *unvisited* neighbors $\{\mathcal{B}_j \mid \mathcal{B}_j \notin \mathcal{Q}\}$;
            foreach $\mathcal{B}_j$ *satisfies* $D(\mathcal{B}_j, \mathcal{B}_i) < \delta$ **do**
                Insert $\mathcal{B}_j$ into $\mathcal{Q}$;
        else
            Goto 1.
1     $k \leftarrow k + 1$.
Output all $\mathcal{R}_k$s.

where

$$G_{x,y}(i, j) = \exp\left\{ -((i - x)^2 + (j - y)^2)/2\sigma_s^2 \right\}, \quad (5)$$

$$K_{x,y}(i, j) = \exp\left\{ -(\tilde{L}(i, j) - \tilde{L}(x, y))^2/2\sigma_r^2 \right\}, \quad (6)$$

$$K'_{x,y}(i, j) = \exp\left\{ -(\tilde{L}(i, j) - \tilde{L}(x, y))^2/2\sigma_{r'}^2 \right\}, \quad (7)$$

and the normalization term is

$$\tilde{Z}_{x,y} = \sum_{(i,j)\in\mathcal{R}_k} G_{x,y}(i, j)\, K_{x,y}(i, j)$$
$$+ \sum_{(i,j)\notin\mathcal{R}_k} G_{x,y}(i, j)\, K'_{x,y}(i, j). \quad (8)$$

Equation (4) comprises two aspects of bilateral filtering. The first part calculates the averaging in the same region, and the second evaluates the contributions from other regions. We have conveniently used the same spatial-domain filter $G_{x,y}$ to regulate the effects to pixel (x, y) from all regions. Such an option makes it possible to apply the fast bilateral filtering described in Durand and Dorsey (2002), which downsamples the image and linearly interpolates the filter responses in range domain. Meanwhile, we have $\sigma_r \geq \sigma_{r'}$ to ensure a flatter and more expanded range-domain filter $K_{x,y}$ for pixels in the same region of (x, y), and to lessen the influences from pixels of different regions with $K'_{x,y}$. If $\sigma_r = \sigma_{r'}$, our proposed scheme in (4) is reduced to the one used in Durand and Dorsey (2002). A suitable value for σ_s can be set to 4% of the image size, and we also have $\sigma_r = 2 \times \sigma_{r'} = 0.4$. Our discussion so far has been focused on the log-luminance domain. The local adaptation luminance, denoted as V, can now be recovered by $V = \exp(\tilde{V})$, e.g., see the examples in Fig. 3.

3. Piecewise Tone Mapping

Our approach to dynamic range reduction relies on region-wise constructing suitable tone-mapping functions that depend on the estimations of local adaptation luminance. It is also critical that the resulting *piecewise tone mappings* could be smoothly pieced together to produce a good-quality displayable image without violating the overall impression of the original HDR radiance map. In this section, we will show that all these issues can be adequately addressed by considering the region information encoded in a luminance-driven sparse decomposition.

3.1. Global and Local Mappings

A handy choice of simple functions for compressing the high luminance values into the displayable range [0, 1] is the nonlinear mapping used in Reinhard et al. (2002)

$$\varphi(x) = \frac{x}{1 + x}. \quad (9)$$

If φ is considered for the whole luminance map, i.e., $L' = \varphi(L) = L/(1 + L)$, we will actually obtain a displayable but smoother image. This type of compression scheme is called *global mapping* or *spatially uniform mapping*. Another tone-mapping method is to extract from L the details H by $H = L/V$. Then, only the local adaptation luminance is compressed by $V' = \varphi(V) = V/(1 + V)$. Recombining the details H with the compressed V', we have

$$L' = H \times V' = \left(\frac{L}{V}\right) \times \left(\frac{V}{1 + V}\right) = \frac{L}{1 + V}, \quad (10)$$

which is a *local mapping* used by Reinhard et al. (2002) for preserving details. Equation (10) can also be interpreted as adjusting the luminance through a position-dependent scale function. Let $m(V) = 1/(1 + V)$ be the position-dependent scale function. Then the local mapping (10) becomes $L' = m(V)L$, conformed to that in Tumblin et al. (1999). Figure 4 illustrates the results of compressing garage by global and local mappings. The result of local mapping provides more details of the original HDR image.

Even though a global mapping like φ in (9) often has the drawback of losing the details in brighter areas, its monotone property ($d\varphi/dx > 0$) is desirable for preventing halos and other artifacts. It would be favorable if the monotonicity can be incorporated into a local tone-mapping method. Nonetheless, one still needs to figure out a reasonable way to *monotonize* a local mapping and determine an appropriate neighborhood for each such a monotonization.[1]

Another issue is concerning the brightness of those large dark areas in Fig. 4. In both results, the top-right and bottom-right parts look quite dark, and occupy a substantial proportion of each image. This impression is not very realistic because not enough details can be *seen* inside the large dark areas. (For comparison, refer to our result of garage shown in Fig. (1).) The *alternative illumination correction* proposed by Kimmel et al. (2003) can solve the brightness problem satisfactorily. However, as to be described later, our method includes
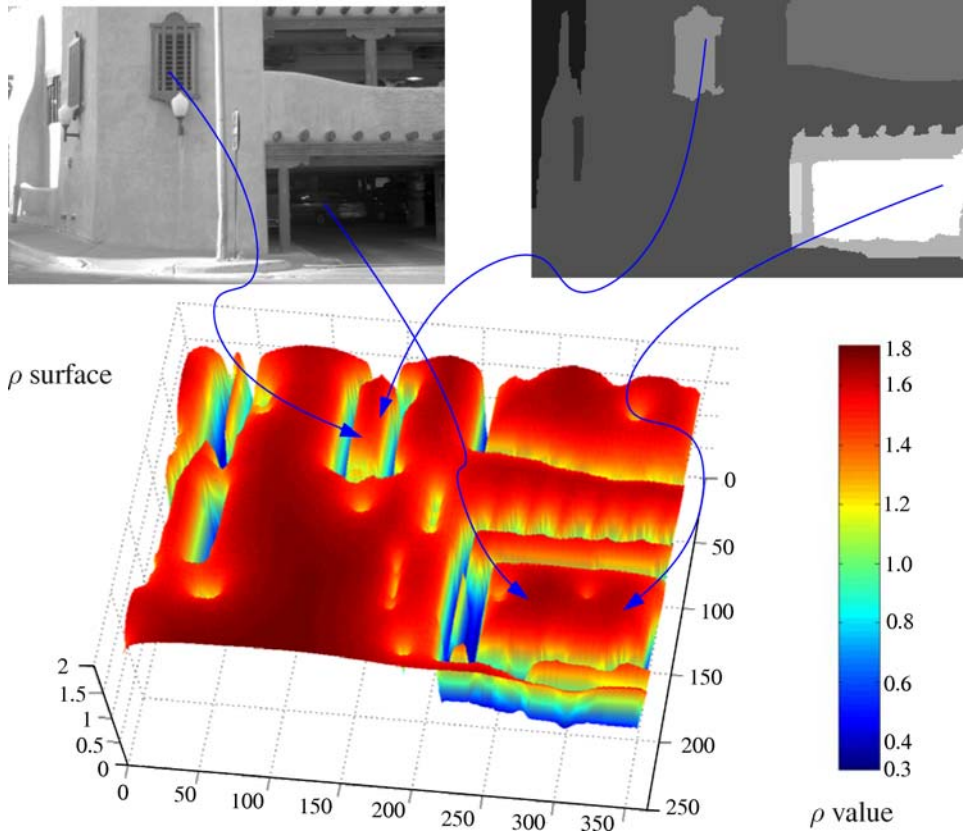
Figure 7. The ρ values after kernel smoothing (e.g., Bishop, 1995). Examples of correspondences to the log-luminance and to the region decomposition are illustrated. Note that larger ρ values will amplify more the subtleties in luminance.

the brightness adjustment into a unified tone-mapping approach that a more general luminance-based *brightness correction* will be implicitly carried out subject to the overall impression constraint.

3.2. Modulating the Tone-Mapping Functions

Perceptually, the proposed region decomposition correlates with an overall visual impression about the scene, including, for example, the contrasts and the order of brightness levels of different areas. And maintaining this impression after compressing the dynamic range should be a good criterion. We thus consider a *piecewise tone mapping* scheme that region-wise performs the monotonization, and globally retains the relative brightness among different regions, i.e., those $\mathcal{R}_k$s in Algorithm 1. The complete idea of deriving such tone mappings can be realized by the following steps.

3.2.1. Design a Local Mapping ψ. As remarked earlier, the local adaptation luminance V is the key

element in compressing the luminance L. We define a local mapping

$$\psi(L, V; \rho, \gamma) = \left(\frac{L}{V}\right)^{\rho} \varphi^{\gamma}(V) = \left(\frac{L}{V}\right)^{\rho} \left(\frac{V}{1+V}\right)^{\gamma},$$
(11)

where $0 < \rho < 2$ and $0 < \gamma \leq 1$ are spatially-dependent parameters to adjust the image quality resulting from the HDR reduction. Specifically, when $\gamma < 1$, the darker areas in an HDR radiance map will be compressed into a larger dynamic range, compared with the case $\gamma = 1$. On the other hand, since L/V contains the details, the ρ values would have direct impacts on preserving the image details after the dynamic-range reduction. When $\rho = \gamma$, we have

$$\psi(L, V; \gamma, \gamma) = \left(\frac{L}{V}\right)^{\gamma} \left(\frac{V}{1+V}\right)^{\gamma} = \left(\frac{L}{1+V}\right)^{\gamma}.$$
(12)

To illustrate the effect of γ, we plot three global mappings with $\gamma = 0.3, 0.7$ and 1.0 in Fig. 5(a), and show the results of directly applying $(L/(1+L))^{0.3}$ and $(L/(1+V))^{0.3}$ to the garage image in Figs. 5(b) and 5(c), respectively. While the visibility of dark areas is boosted (cf. Figs. 4(a) and 4(b)), the overall outcomes are still unsatisfactory. (Notice that in this work we have used a spatially uniform value of $\gamma = 0.3$ for all the experiments.) We shall improve the results by investigating the ρ values and other related factors through the succeeding steps.

3.2.2. Reshape the Global Mapping φ^γ.

Besides the local mapping ψ in (11), we also need a global mapping like φ^γ to embed the overall impression into the resulting region-wise tone mappings. It is therefore useful to make sure that for an arbitrary HDR image, the mapping values of φ^γ take up the complete displayable range $[0, 1]$. We enforce this requirement through a reshapping step by

$$\varphi^\gamma \mapsto \hat{\varphi}^\gamma \quad \text{and} \quad \hat{\varphi}^\gamma(L) = \alpha\varphi^\gamma(L) + \beta, \quad (13)$$

where α and β are the scaling and the shifting variable, respectively, and their values can be analytically derived. More precisely, let $L_{\max}$ and $L_{\min}$ be the maximum and minimum luminance values of a given radiance map. We then solve the following simple linear system to obtain the proper values of α and β in (13):

$$\begin{bmatrix} \left(\dfrac{L_{\max}}{1+L_{\max}}\right)^\gamma & 1 \\ \left(\dfrac{L_{\min}}{1+L_{\min}}\right)^\gamma & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \quad (14)$$

3.2.3. Estimate ρ by Kernel Smoothing.

Appropriate values of ρ have to be assigned for the local mapping $\psi(L, V; \rho, \gamma)$ in (11). We plan to give larger ρ values to most pixels for revealing details, as long as no halos will occur. For each region $\mathcal{R}_k$, we first construct a grid $\mathcal{D}_k$ within $\mathcal{R}_k$. The $\mathcal{D}_k$ is assumed to be the largest grid of resolution $\epsilon \times \epsilon$ with all its grid points at least ε-pixel away from the region boundary $\partial\mathcal{R}_k$. (The idea can be better understood by referencing the illustration in Fig. 6.) We then assign some preliminary ρ_n value

to each pixel $n \in \mathcal{D}_k$ by

$$\rho_n = \begin{cases} \gamma, & \text{if } \log(L_n/V_n) \leq -1, \\ 0.5 \times (\gamma + \rho_{\max}), & \text{if } \log(L_n/V_n) \geq 1, \\ \rho_{\max}, & \text{otherwise.} \end{cases} \quad (15)$$

The first two rules of (15) simply reflect that if L_n is already *very different* from its local adaptation luminance V_n, using a larger ρ_n could only cause an inconsistent overemphasis by further amplifying the log-ratio. In addition, for all pixels on $\partial\mathcal{R}_k$, their ρ_n values are also set to γ, and those of the pixels between $\mathcal{D}_k$ and $\partial\mathcal{R}_k$ can be computed by interpolations. It will later become clear that putting such constraints near $\partial\mathcal{R}_k$ is useful in constructing an overall smooth ρ surface across regions. With all these ρ_n values predefined, we now apply *kernel smoothing* to adjust and to derive the ρ values for all pixels in $\mathcal{R}_k$. See Fig. 7 for an example of ρ surface of the garage image. Note that we have $\rho_{\max} = 1.8$ and $\epsilon = 4$ for all the experiments.

3.2.4. Monotonize Local Tone Mappings.

For each region $\mathcal{R}_k$, a monotonization based on the global mapping $\hat{\varphi}^\gamma$ in (13) is to estimate the variables α_k and β_k so that the local mapping ψ in (11) can be elevated to

$$\hat{\psi}(L, V; \rho, \gamma) = \alpha_k\psi(L, V; \rho, \gamma) + \beta_k, \quad (16)$$

and to account for both local and global factors in region-wise compressing the luminance. We begin by sampling N pixels from $\partial\mathcal{R}_k$ according to the sorted $|\log(L_n/V_n)|$ values in ascending order. In our experiments, using the first 5% of boundary pixels will be sufficient to give good results. For each sample pixel n, we have its ψ_n from (11) and $\hat{\varphi}_n^\gamma$ from (13). The values of α_k and β_k can be obtained by calculating the least squares solution of

$$\begin{bmatrix} \psi_1 & \psi_2 & \cdots & \psi_N \\ 1 & 1 & \cdots & 1 \end{bmatrix}^T \begin{bmatrix} \alpha_k \\ \beta_k \end{bmatrix} = \begin{bmatrix} \hat{\varphi}_1^\gamma & \hat{\varphi}_2^\gamma & \cdots & \hat{\varphi}_N^\gamma \end{bmatrix}^T. \quad (17)$$

With (16) and (17), $\hat{\psi}$ becomes a proper tone mapping for reducing the dynamic range in $\mathcal{R}_k$, and computes the displayable luminance value of each pixel $(x, y) \in \mathcal{R}_k$ by

$$L'(x, y) = \hat{\psi}(L(x, y), V(x, y); \rho(x, y), \gamma)$$

$$= \alpha_k \left(\frac{L(x, y)}{V(x, y)} \right)^{\rho(x,y)} \left(\frac{V(x, y)}{1 + V(x, y)} \right)^{\gamma} + \beta_k. \quad (18)$$

Once the foregoing procedures have been performed for each $\mathcal{R}_k$, the resulting LDR image is ready for displaying. This is accomplished by multiplying the compression ratio L'/L to each of the HDR *RGB* channels with suitable gamma correction for a specific CRT or LCD.

We now give some reasoning for the proposed piecewise tone mapping, especially for the justifications of monotonizations. Notice that only pixels on the region boundaries are considered for the monotonization in (17), and their ρ values after kernel smoothing are still close to γ. It implies that $\psi(L, V; \rho, \gamma) \approx (L/(1 + V))^{\gamma}$ at and near each region boundary. These properties justify our compression scheme in two ways. First, the least squares fitting (17) to reshape ψ into $\hat{\psi}$ can be reasonably solved owing to that the underlying process is indeed to fit a general function of the form $(L/(1 + V))^{\gamma}$, valid to all the sampled points, to the global mapping $\hat{\varphi}^{\gamma}(L) = \alpha(L/(1 + L))^{\gamma} + \beta$. Had we arbitrarily sampled pixels within a region, those ρ values of the selected points could vary significantly ($\rho \not\approx \gamma$). The least squares fitting would become both difficult to interpret its meaning and sensitive to the sampling perturbations. (In this case, the monotonization is no longer implicitly concerning a function of a general form but of the comprehensive form $(L/V)^{\rho}(V/(1 + V))^{\gamma}$.) Second, since $\hat{\varphi}^{\gamma}$ is monotone, the reshaping of a local mapping ψ into $\hat{\psi}$ should *approximately* retain the global brightness consistency among the regions of a given sparse decomposition. Of course, the validity of our claim establishes upon the assumption that the perimeter of an image region would be brighter than that of a less bright region. This property is generally true because the region decomposition for an HDR image is obtained through a grouping process based on the brightest-block-first rule. Therefore, it is fair to conclude that compressing the dynamic-range with (18) has the advantage of emphasizing the local details of each region without breaking the global visual consistency.

4. Experiments and Discussions

Having detailed each step of our approach, depicted in Fig. 8, we are now in a position to go through some of our experimental results and comparisons with other related works. In all our experiments, the HDR im-
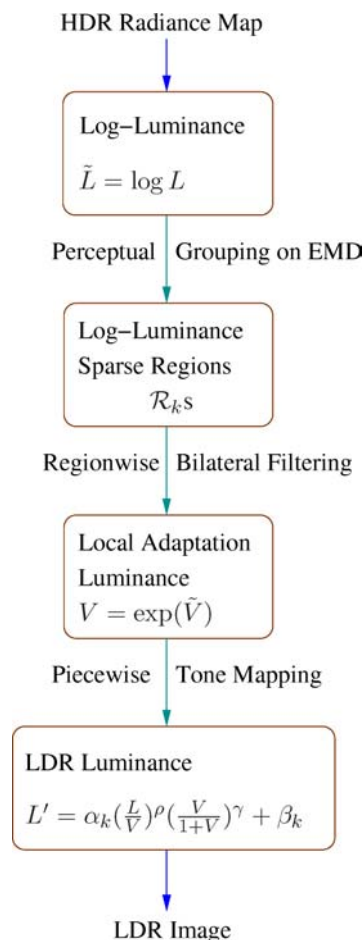

HDR Radiance Map

Log–Luminance
$\tilde{L} = \log L$

Perceptual | Grouping on EMD

Log–Luminance
Sparse Regions
$\mathcal{R}_k$s

Regionwise | Bilateral Filtering

Local Adaptation
Luminance
$V = \exp(\tilde{V})$

Piecewise | Tone Mapping

LDR Luminance
$L' = \alpha_k(\frac{L}{V})^{\rho}(\frac{V}{1+V})^{\gamma} + \beta_k$

LDR Image

Figure 8. The steps of our tone reproduction method.

ages are downloaded from the Web, and stored in the radiance map format. To illustrate, a typical radiance map with multiple exposure values is shown in Fig. 9, where, in this example, the dynamic range of Stanford memorial is about $250,000 : 1$.

4.1. Related Issues and Experimental Results

The efficiency bottlenecks of our method lie in the steps of EMD perceptual grouping and region-wise bilateral filtering. For the memorial image of size 512 by 768 in Fig. 9, the elapsed time of grouping and of filtering on a 3.4 GHz PC are 2.70s and 4.17s, respectively. Clearly, the time complexity of region-wise bilateral filtering should depend on the number of regions. We handle this issue by generalizing the fast bilateral filtering in Durand and Dorsey (2002) to incorporate region

Table 1. Timing results for grouping and filtering.

HDR Image	Size (pixels)	No. of regions	Grouping (seconds)	Filtering (seconds)
garage	352×224	8	0.58	0.69
memorial	512×768	10	2.70	4.17
door	224×336	3	0.23	0.83
park	744×405	6	0.95	2.25
chairs	344×224	14	0.29	1.22
groveC	712×480	17	1.89	5.27
groveD	712×480	18	4.63	4.94
office	1000×656	68	2.71	22.36
Tahoe	1024×768	5	2.74	3.24
clock	768×1024	34	5.70	15.38
designCenter	656×1000	13	2.06	7.76
Tintern	768×1024	14	5.08	4.97
Belgium	1024×768	32	3.20	15.58
synagogue	1024×768	9	3.05	3.48

The corresponding results are shown in Figs. 11, 13, 14, 15 and 17.

support. As a result, a region-wise bilateral filter can be implemented in a way that the slowdown is not proportional to the number of regions. For instance, ten regions are constructed by perceptual grouping for Stanford memorial, but region-wise bilateral filtering simply doubles the time needed for a typical bilateral filtering ($\sim$1.86s). More timing results are listed in Table 1.

We like to emphasize most of the parameters for producing the results in this paper are fixed. In particular, for region-wise bilateral filtering, we have set the spatial scale σ_s to 4% of the image size, and the range-domain scales $\sigma_r = 2 \times \sigma_{r'} = 0.4$; for tone-mapping,

we have $\gamma = 0.3$ and $\rho_{\max} = 1.8$. Therefore, in effect, only two parameters need to be tuned: the EMD grouping thresholds θ and δ. Moreover, to select the values of θ and δ is not complicated at all. In our experiments, typical values of θ are chosen between [1.5, 2.0], and those of δ are between [0.5, 1.0]. Within these ranges the choice of values for θ and δ is not crucial to achieving satisfactory results. Figure 10 shows the outcomes of applying different values of θ and δ.

It can happen that, with bilateral filtering, pixels around large gradients may still undergo small halos due to that too few neighboring pixels are included by the scale of range-domain filter in evaluating proper range-domain responses. This problem is remarked by Durand and Dorsey (2002) as *uncertainty*, a consequence of computing a statistical estimator on insufficient data, and thus causing a fairly high variance value. We have experimented with a similar scheme as in Durand and Dorsey (2002) to resolve the uncertainty by considering a low-pass version $\tilde{V}_l$ of $\tilde{V}$. More precisely, in our experiments a 3×3 small-scale Gaussian kernel with $\sigma < 1$ is used to slightly blur $\tilde{V}$ into $\tilde{V}_l$ as the improved local adaptation log-luminance. This tactic not only fixes the uncertainty problem but also gives a bit more details without generating perceptible halos.

In Fig. 11, we demonstrate several LDR images derived by our tone-mapping method. (The respective sparse decompositions are shown in Fig. 12.) The compression results are characterized by two types of visually pleasing effects: maintaining overall impressions of luminance and preserving details and local contrasts. To examine the first effect we can compare the relative brightness levels of different parts in each pair of an HDR and its compressed LDR images. The quality of

Figure 9. By setting a fixed range and cutting off out-of-range luminance values, the Stanford memorial radiance map can only be displayed like these images.
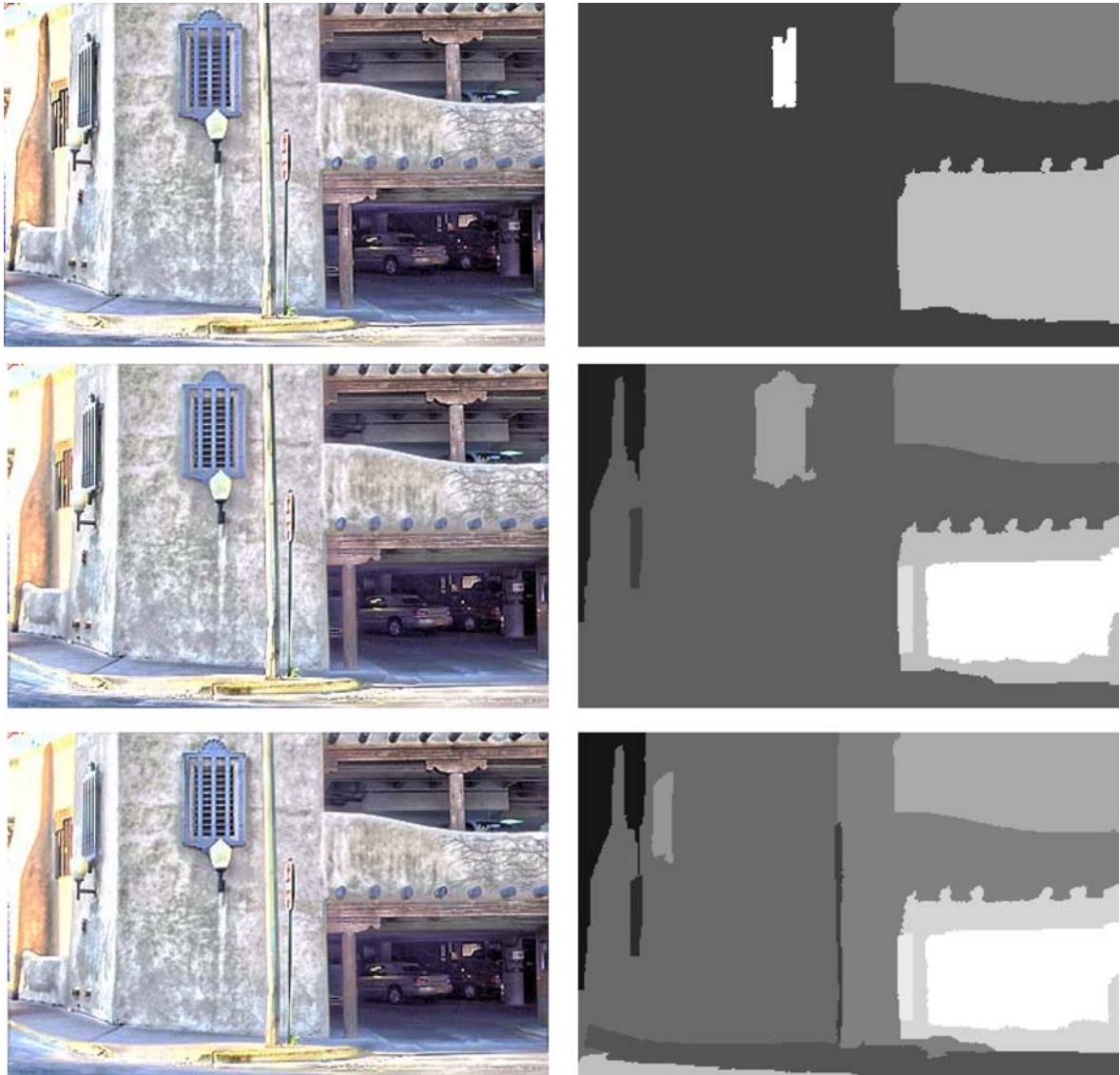
Figure 10. The only two parameters we need to select are the EMD grouping thresholds θ and δ. Typical values of θ are from 1.5 to 2.0, and those of δ are from 0.5 to 1.0. Left column, from top to bottom: the results of choosing $\{\theta = 2.0, \delta = 1.0\}$, $\{\theta = 1.55, \delta = 0.85\}$, and $\{\theta = 1.5, \delta = 0.5\}$. Right column: the corresponding sparse decompositions of the left column. The respective number of regions is 4, 8, and 12 from top to bottom.

the second effect can be inferred, say, from those images in Fig. 11—*local contrasts are not degraded and fine details in dark areas are revealed.*

4.2. More on the Sparse Decomposition

It should be clear now the sparse decomposition plays an important role in our formulation for solving tone reproduction problems. By encoding the overall impression this way, we gain the advantages of focusing on performing local tone mappings over (a few) regions, and meanwhile keeping the global visual consistency through the proposed monotonizations. This effect of retaining relative brightness among different regions is depicted in Figs. 13 and 14.

Suppose we disregard the region information, and use only the local adaptation luminance and the local
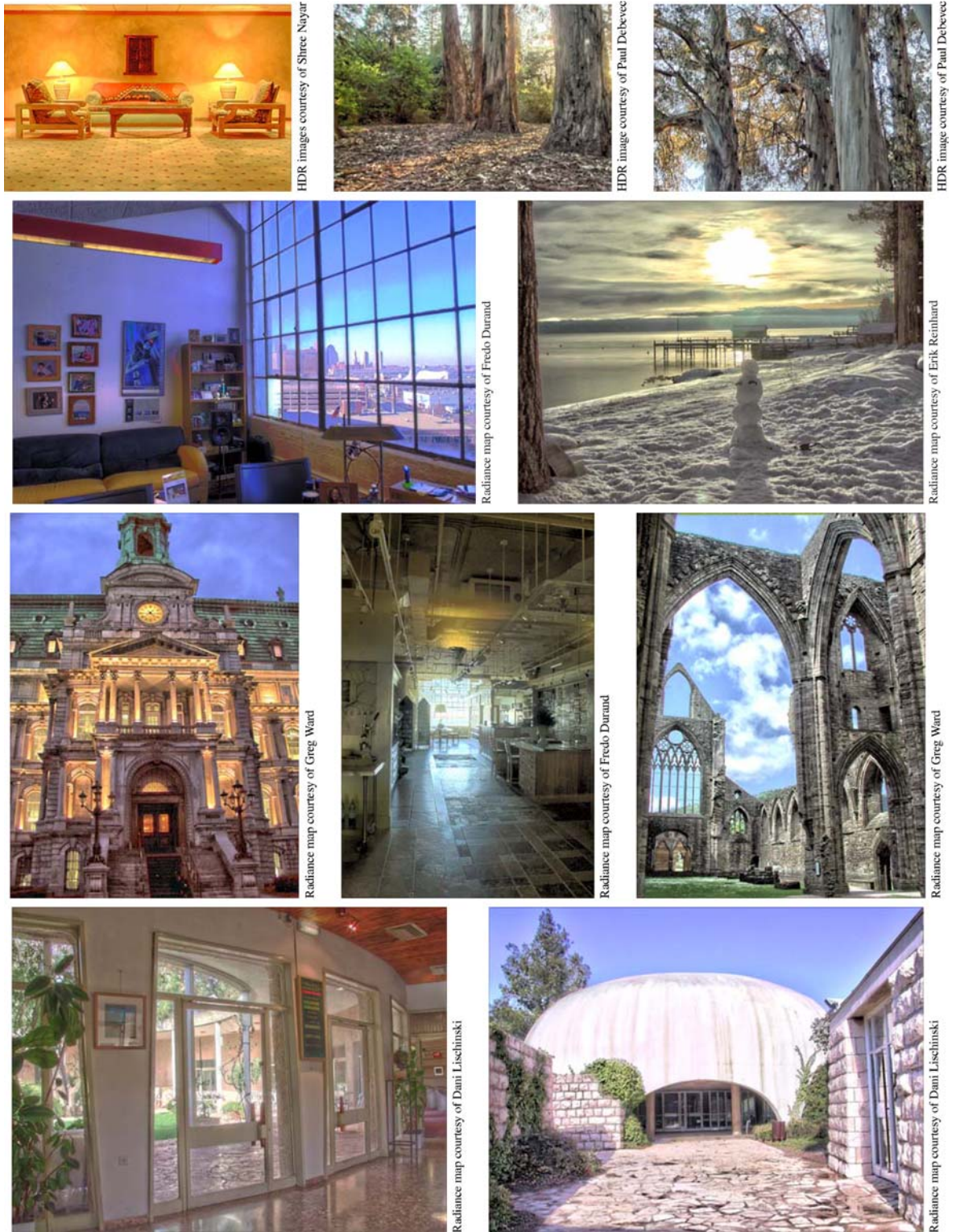
Figure 11. HDR tone-reproduction results. From left to right and top to bottom: chairs, groveC, groveD, office, Tahoe, clock, designCenter, Tintern, Belgium, and synagogue.
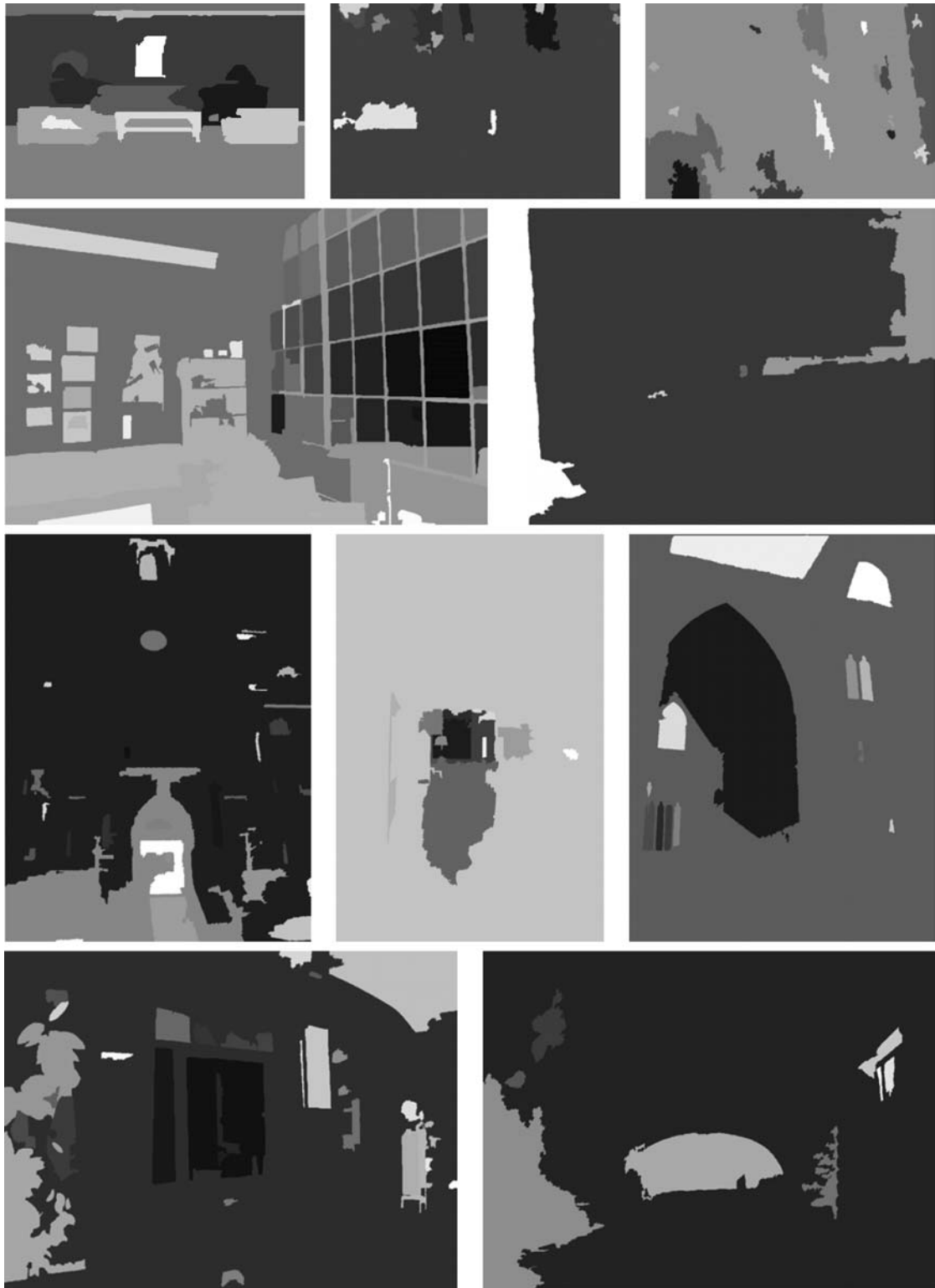
Figure 12. Sparse decompositions. From left to right and top to bottom (with the number of derived regions): chairs (14), groveC (17), groveD (18), office (68), Tahoe (5), clock (34), designCenter (13), Tintern (14), Belgium (32), and synagogue (9).
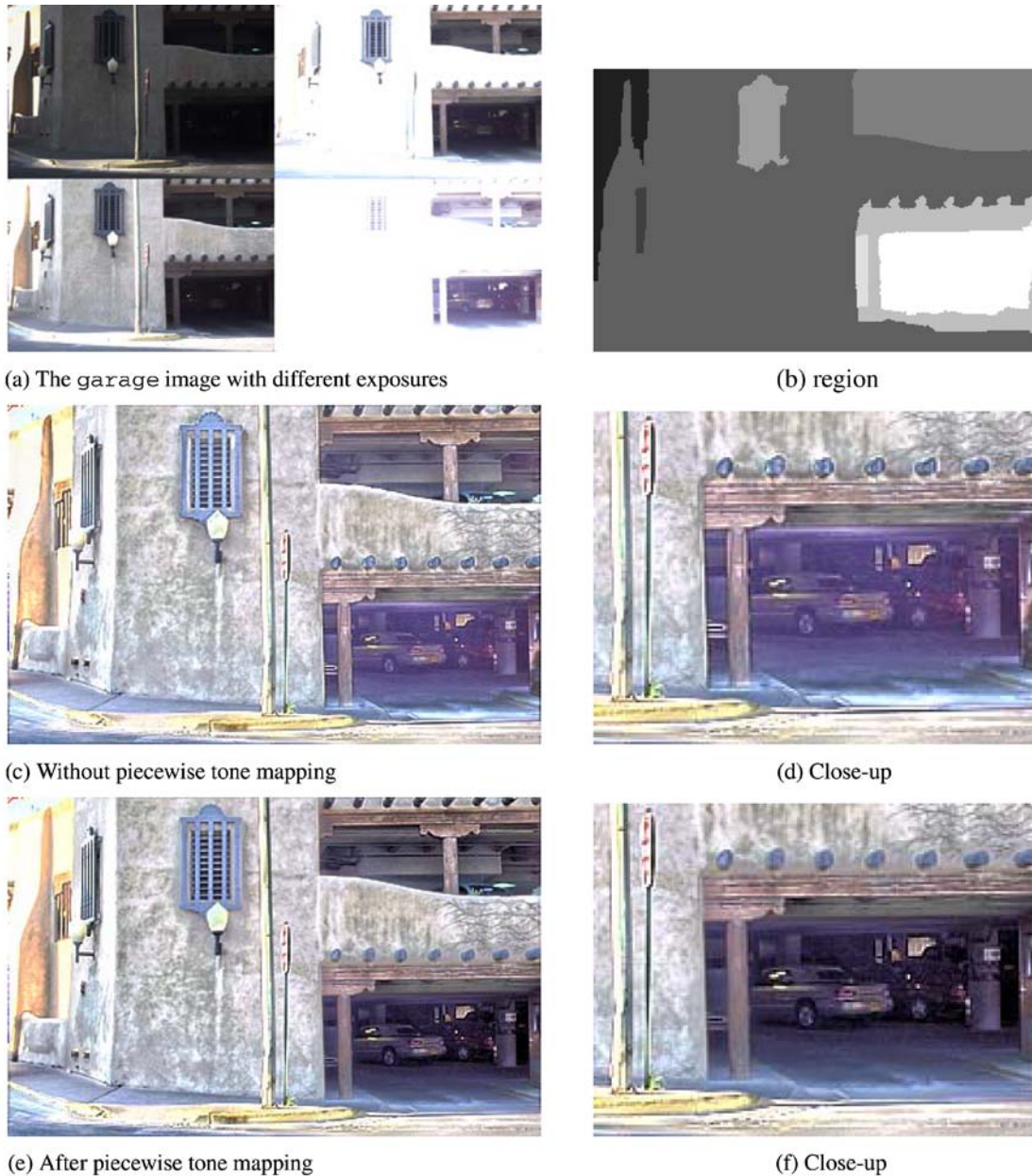
(a) The `garage` image with different exposures

(b) region

(c) Without piecewise tone mapping

(d) Close-up

(e) After piecewise tone mapping

(f) Close-up

Figure 13. The effect of sparse decomposition and piecewise tone mapping. (a) The original HDR image is displayed with different exposures. (b) An eight-region sparse decomposition is obtained after the perceptual grouping stage. (c) Without piecewise tone mapping, the brightness around some areas is somewhat abnormal, e.g. the wall on the left and the inside of the garage. (Here we apply the local mapping $(\frac{L}{V})^{\rho_{\max}}(\frac{V}{1+V})^{\gamma}$ to the whole image with $\rho_{\max} = 1.8$ and $\gamma = 0.3$.) (d) A close-up of (c). (e) Based on the region information of (b), the relative brightness is *monotonized* after piecewise tone mapping. Nonetheless, the details of the dark areas are still noticeable. (f) A close-up of (e).

mapping defined in (10), the tone reproduction results are like the example in Fig. 13(c). That is, we apply the local mapping $(\frac{L}{V})^{\rho_{\max}}(\frac{V}{1+V})^{\gamma}$ to the whole image rather than individual regions. (Recall that $\rho_{\max} = 1.8$ and $\gamma = 0.3$.) As can be seen in Figs. 13(c) and (d), the

yellow exterior on the left is darkened, and the inside area is too bright and lack of contrast. On the contrary, the results in Figs. 13(e) and (f) yielded by applying our piecewise tone mapping to the sparsely decomposed image reveal a good compromise between local and
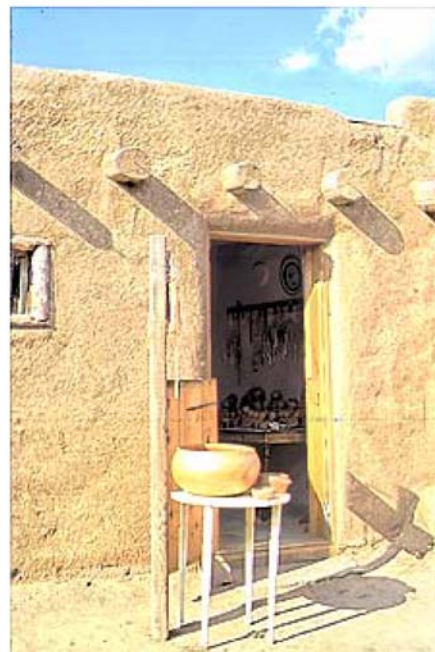
(a) The door image

(b) Its sparse decomposition

(c) Without piecewise tone mapping

(d) After piecewise tone mapping

Figure 14. (a) The original HDR image is displayed with an acceptable exposure setting. (b) A three-region sparse decomposition. (c) Without piecewise tone mapping, the brightness around some areas is somewhat abnormal, e.g. the cloud and the indoor scene. (Directly mapped by $(\frac{L}{V})^{\rho_{\max}}(\frac{V}{1+V})^{\gamma}$ with $\rho_{\max} = 1.8$ and $\gamma = 0.3$.) (d) After piecewise tone mapping.
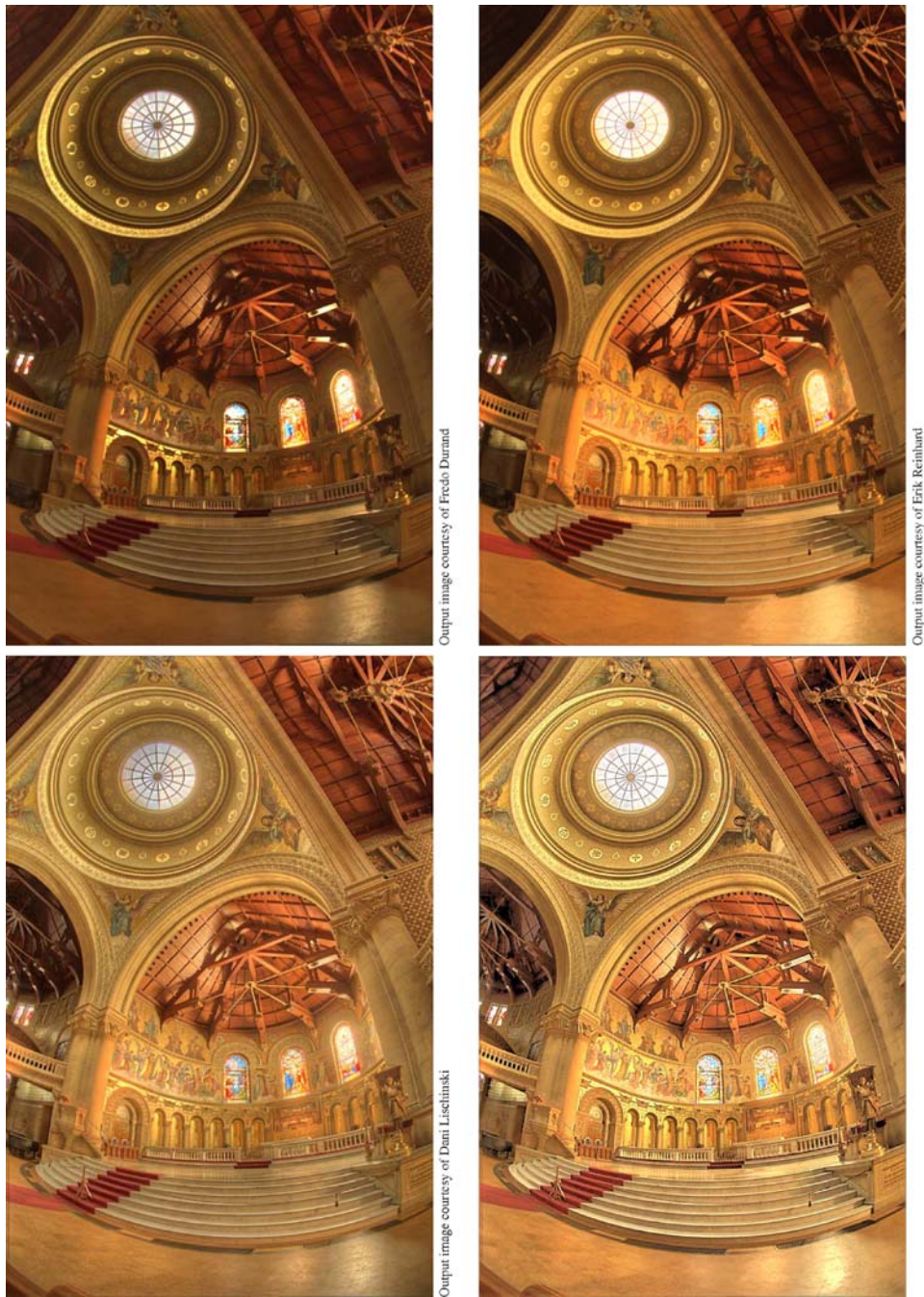
Figure 15. **Stanford memorial**. From left to right, top to bottom: the respective compression results derived by *bilateral filtering* (Durand and Dorsey, 2002), *photographic tone reproduction* (Reinhard et al., 2002), *gradient domain* (Fattal et al., 2002), and *our method*. Our method can reveal the fine details of the scene and maintain a global impression similar to that of the photographic (Reinhard et al., 2002).

global contrasts. Taking the door image as another example, a similar effect produced by piecewise tone mapping can again be observed in the indoor area, cf. Figs. 14(c) and (d).

4.3. Comparisons

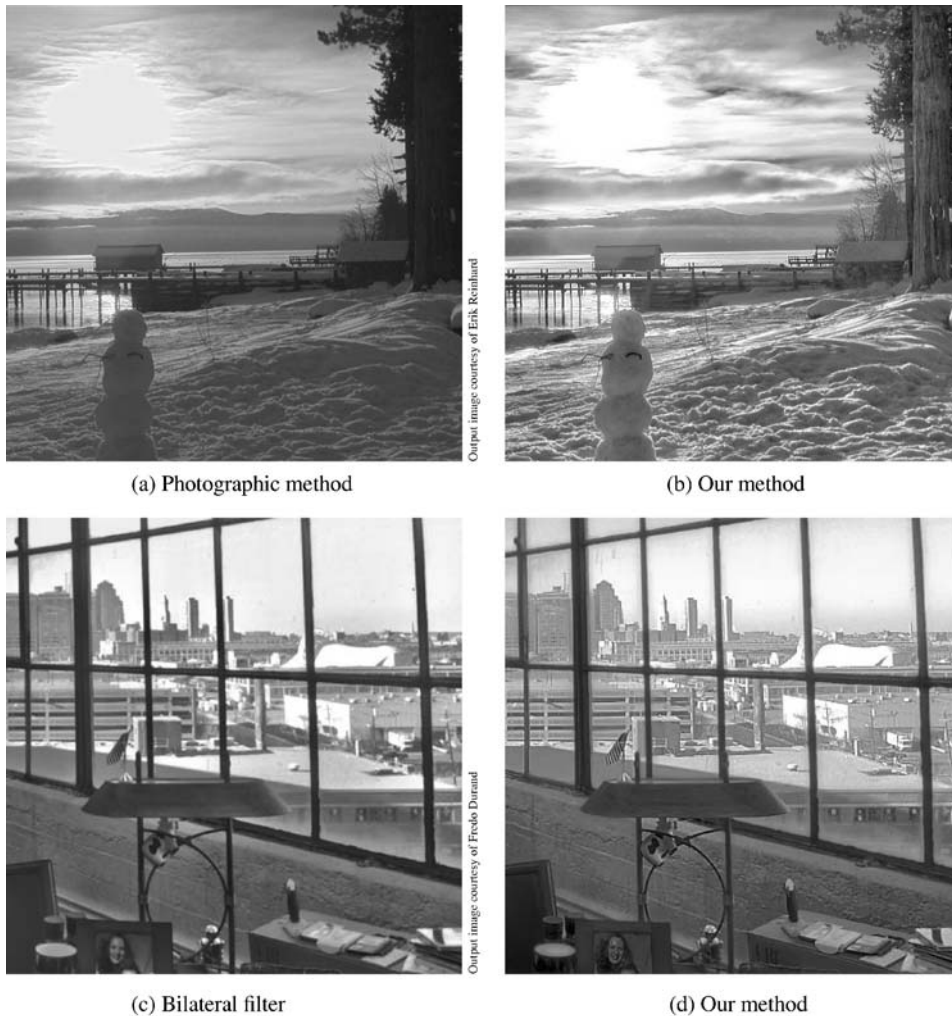Some recent methods (Durand and Dorsey, 2002; Fattal et al. 2002; Reinhard et al., 2002) on displaying

Figure 16. (a) The sun is too dark in the result of Reinhard et al. (2002). (b) Corrected by piecewise tone mapping. (c) Bilateral filtering (Durand and Dorsey, 2002) produces some halos near the skyline and the shadows. (d) Our method not only preserves more details but prevents most halos.

HDR images also can eliminate halos and preserve details. Despite different quality of visual impressions, their results generally look good. It is indeed hard and biased to rank the performances of different methods. So far, the fairest but impractical way to evaluate a tone-reproduction method may only depend on the *feeling* of someone who has been present when the photos were taken. Thus, in what follows, we shall focus on the typical differences and characteristics between the outcomes of our method and others, illustrated with the examples shown in Figs. 15–17.

The gradient-domain compression by Fattal et al. (2002) performs well in preserving local contrasts and details. However, their method does not explicitly deal with the balance between local contrasts and global consistency. Thereby a noticeable difference between their compression outcomes and ours is that in their results global contrasts are more often distorted. For instance, brighter areas may not be bright enough as they should be. A good example is the circular window of Stanford memorial in Fig. 15. Note that, in the original radiance map, the area of the circular window is at least 200 times brighter than the top-right area of the image.

While our tone-mapping approach relates to photographic tone reproduction (Reinhard et al., 2002) and fast bilateral filtering (Durand and Dorsey, 2002) in computing the local adaptation luminance, the
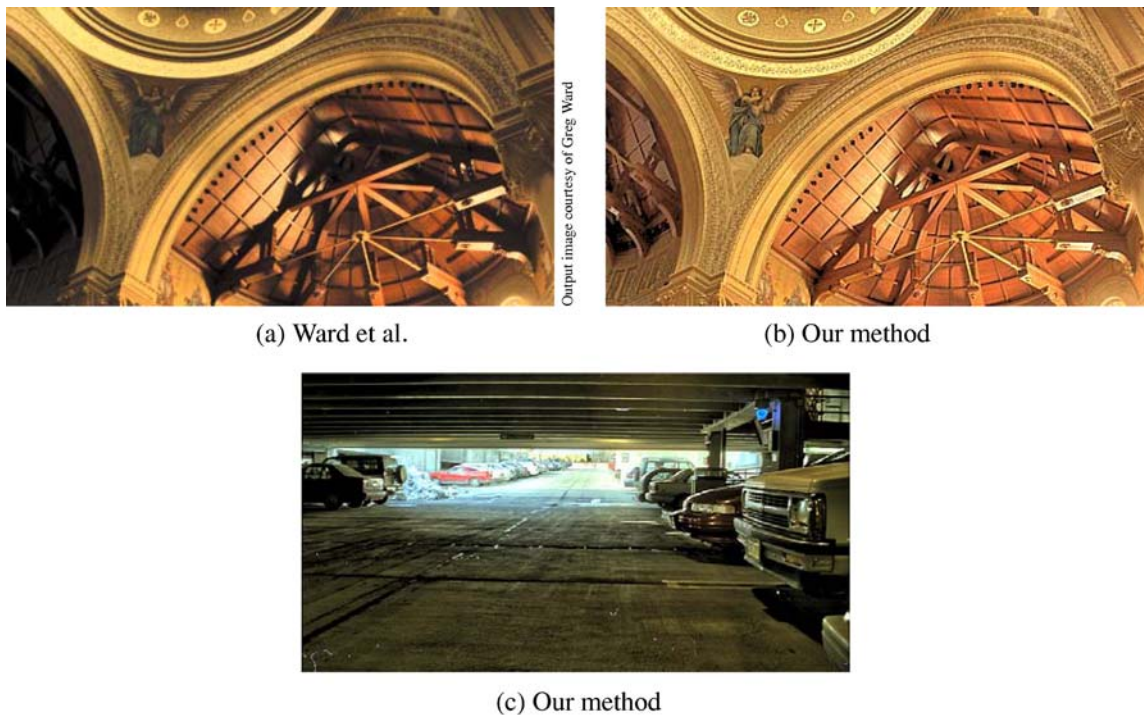
(a) Ward et al.

(b) Our method

Output image courtesy of Greg Ward

(c) Our method

Figure 17. (a) The result of Ward et al. (1997). (b) The details of the medallion and the carvings are preserved by our method. (c) The result of our method (cf. Fig. 3 of Yee and Pattanaik, 2003).

proposed method usually yields better contrasts within the displayable range and preserves more image details. The examples in Fig. 15 should well demonstrate these common phenomena. In addition to the foregoing differences in quality, we further discuss another two noticeable distinctions between our results and those produced by Durand and Dorsey (2002) and Reinhard et al. (2002), using the examples in Fig. 16.

The first distinction is that although the photographic method of Reinhard et al. (2002) basically maintains overall impressions, sometimes the approximate local adaptation luminance is too *local*—the scales of Gaussian kernels are not large enough. Consequently, only those pixels with similar luminance values are included for computing the local adaptation luminance, and thus the contrast of a nearby area cannot be fully expressed. In particular, this effect often occurs in very bright areas, and causes the brightness of these areas to decrease. For instance, the sun in Fig. 16(a) appears to be too gray. Alternatively, owing to piecewise tone mapping, our results do not have this anomaly, e.g., Fig. 16(b). The second distinction is that the bilateral filtering method of Durand and Dorsey (2002) still generates some halos near the boundaries between bright

and dark regions. The differences can be observed near the skyline and near the shadows in Figs. 16(c) and (d). Via a smooth ρ-surface adjustment, our method produces a bit lower contrasts along boundaries, and can avoid the problem of halos.

Our approach differs from global operators, e.g., Ward et al. (1997), by the effectiveness on preserving details. Ward et al. (1997) adopt histogram adjustment to design the tone-mapping function. They use a linear ceiling that constrains visible contrast distortions and thus prevents halos. The resulting function is global and monotonically increasing. Although an operator of this kind is fast and halo-free, it tends to blur details. The proposed piecewise tone mapping instead maintains the global contrasts via regulating local mapping functions. Local contrasts are preserved by (18), which includes the local adaptation luminance $V(x, y)$ and the ρ surface. These differences between global mappings and our method are illustrated in Figs. 17(a) and (b).

Yee and Pattanaik (2003) develop an algorithm consisting of segmentation, grouping, and assimilation for estimating the local adaptation luminance: Pixels are categorized by their log-luminance values based on a given bin-width. The grouping process is akin to

computing connected components for pixels that belong to the same category. Small singleton groups are then merged into their larger neighbors, and the resulting groups form a *layer*. These processes are repeated several times with different bin-widths to yield multiple layers of an HDR image. Hence every pixel in each layer belongs to some group, and has its averaged group luminance value. The local adaptation luminance at each pixel location is obtained by averaging its corresponding group luminance values from all layers. Like bilateral filtering, the formulation in Yee and Pattanaik (2003) is aimed to compute the local adaptation luminance for preserving details and preventing halos. Indeed Yee and Pattanaik do not propose a new tone-mapping operator, but employ existing ones (e.g. Tumblin and Rushmeier, 1993), to display HDR images with the local adaptation luminance derived by their scheme. However, we use the grouping process to obtain region information (in order of luminance) for adjusting piecewise tone-mapping functions, not just to estimate the local adaptation luminance. One of our experimental results for comparing the two methods is given in Figs. 17(c) (cf. Fig. 3 of Yee and Pattanaik (2003).

5. Conclusions

We have thoroughly investigated the tone-reproduction problem in two aspects: (1) deriving local adaptation luminance for preserving details, and (2) region-wise reducing the high dynamic range into a displayable range without breaking overall impressions. For a visual model, the *correctness* of local adaptation mainly depends on the definition of *locality*. Only by an overall consideration of both luminance and perception factors can one reproduce local high contrasts in the form of visually pleasing details rather than halos.

Our proposed sparse decomposition for the luminance has been playing a pivotal role in our formulation. Through the image decomposition, we are able to analyze both the local adaptation luminance and the tone-mapping functions. Specifically, we formulate a region-wise bilateral weighting scheme to enhance the local adaptation effect inside a region. Pertaining to the tone-mapping functions, the decomposition implicitly encodes the globally relative contrasts into a small number of perceptually significant regions. Since the grouping process is driven by a brightest-block-first rule, the derived regions are arranged roughly in order of the luminance value. These properties are useful

in modulating an arbitrary local tone mapping of a region into the one strengthened with the desirable global property by our so-called monotonization step.

The piecewise tone-mapping scheme indeed cultivates various tone-reproduction qualities. Besides maintaining global contrasts among different regions by fitting to the global mapping $\hat{\varphi}^{\gamma}(L)$, it includes the two parameters, γ and ρ, in the mapping functions (18) to regulate brightness and local contrasts of tone-reproduction results. We have well illustrated this point by using an appropriate γ value to effectively elevate the brightness levels in dark areas. And the ρ surface tempers high contrasts in too dark and too bright areas, as well as alongside region boundaries (which usually imply strong edges). The near-boundary ρ values are uniformly close to γ, so that we can smoothly stitch regions and prevent possible halos.

Finally, as expected and verified by our other experiments, directly applying the proposed tone-reproduction algorithm to LDR images would also considerably enhance image qualities. Therefore, the piecewise tone mapping is effective in not just compressing high dynamic range but also expanding low dynamic range. However, in both cases the dynamic range is considered to be continuous. It is possible that a radiance map has a bipolar distribution of extremely large and small luminance values, i.e., there is a large gap in the dynamic range. For future work, it will be interesting to devise specific algorithms for this kind of situation. Another extension we are interested in as future work is to apply graphical models to the tone reproduction problem. Graphical models are shown to be useful tools for low-level vision modeling. Their application to high dynamic range reduction appears to be a new and promising direction.

Note

1. By a monotonization within an appropriate neighborhood, we mean to monotonize a local mapping so that it *almost* has the monotone property of φ in that region.

References

Adelson, E.H. 2000. Lightness perception and lightness illusions. *The New Cognitive Neurosciences*, 2nd edition, pp. 339–351.

Aggarwal, M. and Ahuja, N. 2001. High dynamic range panoramic imaging. In *Proc. Eighth Int'l Conf. Computer Vision*, vol. 1, pp. 2–9.

Allred, S.R., Thompson, J.Y.S., and Jagadeesh, B. 2003. Color-based estimates of stimulus similarity predict perceptual similarity of image pairs to monkeys [abstract]. *Journal of Vision*, 3(9):511a.

Allred, S.R., Liu, Y., and Jagadeesh, B. 2004. Algorithms for image database navigation and tuning of object selective nneurons in the non-human primate. In *APGV '04: Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization*, pp. 161–161.

Ashikhmin, M. 2002. A tone mapping algorithm for high contrast images. In *13th Eurographics Workshop on Rendering*, pp. 145–156.

Barlow, H.B. 1961. *The Coding of Sensory Messages: Current Problems in Animal Behaviour*, W.H. Thorpe and O.L. Zangwill (Eds.), Cambridge University Press.

Bishop, C.M. 1995. *Neural Networks for Pattern Recognition*. Oxford University Press.

Canny, J.F. 1986. A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8(6):679–698.

Chiu, K., Herf, M., Shirley, P., Swamy, S., Wang, C., and Zimmerman, K. 1993. Spatially nonuniform scaling functions for high contrast images. In *Graphics Interface*, pp. 245–253.

Debevec, P.E. and Malik, J. 1997. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH*, pp. 369–378.

DiCarlo, J.M. and Wandell, B.A. 2000. Rendering high dynamic range images. In SPIE: Image Sensors, vol. 3965, pp. 392–401.

Durand, F. and Dorsey, J. 2002. Fast bilateral filtering for the display of high-dynamic-range images. In *SIGGRAPH*, pp. 257–266.

Fattal, R. Lischinski, D., and Werman, M. 2002. Gradient domain high dynamic range compression. In *SIGGRAPH*, pp. 249–256.

Ferwerda, J.A., Pattanaik, S., Shirley, P., and Greenberg, D.P. 1996. A model of visual adaptation for realistic image synthesis. In *SIGGRAPH*, pp. 249–258.

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., Spehar, B., Annan, V., and Economou, E. 1999. An anchoring theory of lightness perception. *Psychological Review*, 106(4):795–834.

International Telecommunication Union (ITU). 1990. Basic parameter values for the HDTV standard for the studio and for international programme exchange. In *ITU-R Recommendation BT.709*, Geneva, Switzerland.

Jobson, D.J., Rahman, Z.U., and Woodell, G.A. 1997. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans. IP: Special Issue on Color Processing*, 6(7):965–976.

Kang, S.B., Uyttendaele, M., Winder, S., and Szeliski, R. 2003. High dynamic range video. In *SIGGRAPH*.

Kimmel, R., Elad, M., Shaked, D., Keshet, R., and Sobel, I. 2003. A variational framework for retinex. *Int'l J. Computer Vision*, 52(1):7–23.

Krawczyk, G., Mantiuk, R., Myszkowski, K., and Seidel, H.-P. 2004. Lightness perception inspired tone mapping. In *Proc. Symp. Applied Perception in Graphics and Visualization*, p. 172.

Land, E.H. and McCann, J.J. 1971. Lightness and retinex theory. *J. Optical Society of America*, 61:1–11.

Nayar, S.K. and Mitsunaga, T. 2000. High dynamic range imaging: Spatially varying pixel exposures. In *Proc. Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 472–479.

Palmer, S.E. 1999. *Vision Science: Photons to Phenomenology*. MIT Press.

Pattanaik, S.N., Ferwerda, J.A. Fairchild, M.D., and Greenberg, D.P. 1998. A multiscale model of adaptation and spatial vision for realistic image display. In *SIGGRAPH*, pp. 287–298.

Reinhard, E. Stark, M., Shirley, P., and Ferwerda, J. 2002. Photographic tone reproduction for digital images. In *SIGGRAPH*, pp. 267–276.

Rubner, Y. and Tomasi, C. 2001. *Perceptual Metrics for Image Database Navigation*. Springer.

Rubner, Y., Tomasi, C., and Guibas, L.J. 2000. The earth mover's distance as a metric for image retrieval. *Int'l J. Computer Vision*, 40(2):99–121.

Schechner, Y.Y. and Nayar, S.K. 2003. Generalized mosaicing: High dynamic range in a wide field of view. *Int'l J. Computer Vision*, 53(3):245–267.

Schlick, C. 1994. Quantization techniques for visualization of high dynamic range pictures. In *5th Euro-graphics Workshop on Rendering*, pp. 7–18.

Stokes, M., Anderson, M., Chandrasekar, S., and Motta, R. 1996. A standard default color space for the internet—sRGB. In http://www.w3.org/Graphics/Color/sRGB

Tomasi, C. and Manduchi, R. 1998. Bilateral filtering for gray and color images. In *Proc. Sixth Int'l Conf. Computer Vision*, pp. 839–846.

Tumblin, J., Hodgins, J.K., and Guenter, B.K. 1999. Two methods for display of high contrast images. *ACM Trans. Graphics*, 18(1):56–94.

Tumblin, J. and Rushmeier, H. 1993. Tone reproduction for computer generated images. *IEEE Computer Graphics and Applications*, 13(6):42–48.

Tumblin, J. and Turk, G. 1999. LCIS: A boundary hierarchy for detail-preserving contrast reduction. In *SIGGRAPH*, pp. 83–90.

Ward, G. 1991. Real pixels. *Graphics Gems II*, pp. 80–83.

Ward, G. 1994. A contrast-based scale factor for luminance display. *Graphics Gems IV*, pp. 412–421.

Ward, G., Rushmeier, H.E., and Piatko, C.D. 1997. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Trans. Visualization and Computer Graphics*, 3(4):291–306.

Yee, Y.H. and Pattanaik, S.N. 2003. Segmentation and adaptive assimilation for detail-preserving display of high-dynamic range images. *The Visual Computer*, 19(7/8):457–466.