Visualization of High Dynamic Range Images

Alvaro Pardo* Guillermo Sapiro†

Abstract

A novel paradigm for the visualization of high dynamic range images is presented in this paper. These images, real or synthetic, have luminance with typical ranges many orders of magnitude higher than that of standard output devices, thereby requiring some processing for visualization. In contrast with existent approaches, that compute a single image with reduced range, close in a given sense to the original data, we propose to look for a representative set of images. The goal is then to produce a minimal set of images capturing the information all over the high dynamic range data, while at the same time preserving a natural appearance for each one of the images in the set. A specific algorithm that achieves this goal is presented and tested on natural and synthetic data.

1 Introduction

High dynamic range images contain a wide range in luminance, many times in the order of tens of thousands different values. [1] These images could be natural, for instance obtained from multi-exposure photographs [1] or with a multiple exposure sensor [5, 11], or synthetic, in the case of computer graphic applications. These images have ranges that greatly exceed that of the output device. The question is then how can we *reproduce* and *visualize* such an image in a standard output device. This is the question we want to address in this paper.

Before proceeding, let us introduce some basic terminology. The *scene* is the real or synthetic picture we perceive without involving any output device between it and our eyes.[2] An *image*, on the other hand, is what we can see using the output device, or its internal computer representation as an array of digital values. The key problem is how to translate from scenes to images, while preserving the relevant scene information, producing a natural looking image, and avoiding common artifacts such as *halos* (which are due to local gradient reversals [2]).

Numerous applications exist for high dynamic range images. One is computer graphics and the production of synthetic images with realistic or hyper-realistic appearance. Another application covers high dynamic range photographs which are able to capture much more detailed scene information than standard photographs. Recently, methods to acquire such photographs have been developed [1, 11, 5]; in particular, in [11, 5] the described system is already included in a camera prototype. This allows the user to capture a high dynamic (detailed) range representation of the scene and later process the data in order to select the image/s that better fulfills the given requirements. Furthermore, these obtained images could improve computer vision and image analysis algorithms which usually rely on just 8 to 24 bits images with limited range. This is particularly

*IMERL & IIE, Facultad de Ingeniería, Universidad de la República, CC-30, Montevideo, Uruguay. apardo@iie.edu.uy

†Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455. guille@ece.umn.edu

[1]These images are also referred to in the literature as *high contrasted.*

[2]In the case of synthetic scenes it might be sometimes very difficult to directly visualize the data.

relevant in scenarios where we do not have complete control over the illumination, like medical applications for instance. There is then a need to develop algorithms that perform this translation from scene to images, algorithms as the ones discussed and presented in this paper.

When addressing this translation problem we have to consider the digital nature of output devices. Indeed, this is a very important issue when trying to display details. In some way, the problem of high dynamic range could be seen as a resolution problem, with a small number of output levels we have to display a highly detailed image. As we know, things become worst when the number of output levels is less than the minimum needed to obtain a good *quantization*.

We can divide the existing approaches to the translation from scene to image in two main groups (examples of these will be detailed in the next section). The first group consists of algorithms that map the original range to the output range while attempting to preserve the subjective perception of the scene, e.g., [2]. Although this idea of "tone mapping" works quite well, it has some caveats. First, it is not able to reproduce all the details present in the scene. The method breaks down when the input range is too wide compared with the available output range.

The second group favors the visualization of details instead of the subjective perception of the scene. In this group we have the works of Tumblin and Turk [10] and DiCarlo and Wandell [2]. Both apply a multiscale decomposition to discriminate between illumination and details. The main problem with this idea is that, although correct in theory, it usually introduces halos in the output image. Moreover, these approaches tend to have a large number of parameters, generally hard to control in an automatic fashion.

Our proposed paradigm attempts to have the best of both groups mentioned above. We propose a method that captures the details while preserving the natural appearance of the scene. As we will explain below, there is an intrinsic limitation in representing a high dynamic range image with only *one standard* output image. Furthermore, sometimes it is practically impossible to find an output image containing all the relevant information in the high dynamic range image that represents the scene. For this reason, we argue for a method to obtain a *set of images* containing all the relevant information of the original scene and displayed in a suitable way.

1.1 Related work

In this section we review some of the recent literature on the reproduction of high dynamic range images. First, we address methods which try to preserve the subjective perception of the scene and then the ones which favor the visualization of details.

In [9], Tumblin and Rushmeier developed a tone mapping operator using models of human perception. The main drawback of their algorithm is that they use a global brightness adaptation, dark and bright regions are somehow clipped. Schlick [7] concentrated on a simple method for computing a local tone mapping.

Ferweda et. al. [3] noted the connection between light levels, color, and acuity. Using this work, Larson et. al. [4] proposed a global tone mapping operator which adjusts the histogram of the scene based on psychophysical models for color, glare, and acuity perception. The results of this method are very good, specially under the hypothesis of preserving the subjective perception of the scene and considering its simplicity (just a global map). The problem is, as with most of the methods in this category, when the input range is too large. Note that since this method tries to preserve the original perception of the scene, details hard to see in the original scene will be difficult to see in the output image as well.

In [8] the authors introduce two methods to display high dynamic range images. The first one is intended to display synthetically generated images. The image is decomposed into layers of lighting and surface properties. The light layer, which contains most of the high contrast, is compressed and added back to the surface layers containing details and texture. In this way, high contrast is reduced while preserving the details and texture from the original image. The mayor problem is that this approach works only for synthetic images, for which we have easily available the different layers. We will come back to this point later when reviewing the work of Tumblin and Turk [10]. For natural scenes they proposed a locally adaptive method, denoted as the foveal display, which is inspired by eye movements. The user selects a point of attention and the algorithm computes an output image with preserved contrast in the foveal region. Is important to remark that this approach is dynamic in the sense that a set of images is generated with the aid of user interaction. The problems with this approach is that the user needs to select a point of attention (not automatic procedure). Moreover, the user should be able to "see" everywhere in the image to choose the points of attention. Problems could then arise if the image presented to the user is not sufficient for inspection. On the other hand, we should note that this approach is connected to ours in the sense that we also suggest to compute a set of images instead of a single one.

The work of Pattanaik et. al. [6] proposes a multiscale model for the representation of pattern, luminance, and color in the human visual system. The main problem with this work is that, although interesting for its detailed modeling of the human visual system, it can not avoid halos.

Continuing with the idea of segregating the image into layers of lighting and details [8], Tumblin and Turk [10] proposed a multiscale approach to extract a hierarchy of details and boundaries. Their idea is to mimic the way artists work from coarse to fine to recreate highly contrasted scenes in low contrasted mediums. Basically, artists start with a sketch of strong features and progressively add small details. They decompose the image into strong and weak features using a multiscale operator. Then, only strong features are compressed. Although the idea is very attractive, their algorithm can not avoid halos completely and, as is pointed out by the authors, needs the difficult tuning of some crucial parameters.

In our view of the problem, the two groups described are not equivalent, they try to solve different problems. Furthermore, although similar, their solutions can not be easily compared since one representation tries to capture the subjective appearance of the scene under the limitation of the output/display device, while the second group attempts to preserve the scene details. Basically, if we enhance details we may be adding information not perceptually present in the original scene. On the other hand, while preserving visual appearance, details might be omitted.

1.2 Our contribution

The common approaches described above produce a single image per-scene or per focused scene region. It is a quite optimistic approach that we can accurately represent the information of an image with tens of thousands of values with just a few hundred. This is what motivates the paradigm here proposed, meaning the use a set of images to represent such highly detailed information. We could say that while the algorithms described above deal with the *reproduction* of the scene, the technique here proposed deals with its *visualization*. Moreover, we argue that not only the set of images has to accurately visualize the relevant information present in the scene, but they have to do it in a visually pleasant form. We present a particular algorithm to exemplify this new paradigm.

2 The method

Our idea is to propose a simple but effective method to visualize *all* the information in the scene in a *pleasant* way. *All* remarks the fact that we would like to capture as many details as possible, and *pleasant* means a procedure which appears natural to the observer. And last but not least, we will try to avoid the introduction of halos and other artifacts in the solution.

Lets suppose we have a scene with dark and bright areas, and details all over it. If we want to visualize all the details we would need first to have a minimum resolution available (number of output levels), and second to be able to "see" in every region. In order to understand what is meant by "see," we are going to present a couple of simple examples. Consider the image of a dark room, in order to "see" the details we would usually turn on the lights. Now, suppose we are in the beach and everything is too bright, in this case to "see" we would probably wear sunglasses to reduce the amount of light. In both cases, the information is out there but it can not be seen. In photographic terms, this is due to under-exposure or over-exposure.

These problems are not at all new, specially in photography they are of great importance. If we take a picture of an object and there is light coming from behind it, we won't capture all its details. To solve this, photographers put extra light to the front of the object. We are going to borrow this idea and modify the luminance of the scene to capture all the details over it. All the operations could be seen as just contrast changes, and in this way, we won't be introducing artifacts.

The first observation is that if we illuminate a given region, we will be stretching its output range, thereby using more output levels. Therefore, on one hand we display this region with good light and resolution, while on the other hand we might be compressing and missing details in other regions. Hence, there is clearly a competition between the output fidelity of different regions, and unfortunately, it is difficult to find a solution with a single output. To overcome this, we propose to generate a sequence of images with different resolutions in each scene region (space varying resolution). This sequence could be either observed as a movie or a set of still images could be extracted from it. The key point here is that for many applications more than one output image is a reasonable solution.

The basic idea is then to distribute the existing resources among different output images. In the case of only two regions, we first display dark regions via adding some extra light on them and then we slowly swap resources to bright areas. This simple idea resembles the control of illumination during acquisition and is perceived as natural by human observers.

2.1 Outline of the algorithm

Before presenting our proposed algorithm, let's give some basic notation:

(r, g, b)	Input color primaries
L	Input luminance
$[Lwmin, Lwmax]$	Input luminance range
L'	Modified luminance
$[Ldmin, Ldmax]$	Output luminance range
(R, G, B)	Digital output values

We now describe the different steps of the algorithm.

1. **Compute image luminance:** From the (r, g, b) primaries compute the luminance L (in cd/m^2) and the color information $(r/L, g/L, b/L)$. We will process the luminance while preserving the color information.

2. **Segment the image:** Divide the image into two or more regions of interest. This is achieved splitting the histogram into sub-intervals $[Lwmin, L1, L2, ..., Ln, Lwmax]$. If we segment the image arbitrary then we could loose the monotonic restriction of the tone mapping. We choose a simple procedure which segments the image into bright and dark areas to illustrate the idea. For the rest of this section we assume two regions, $[Lwmin, Lw^*]$ and $[Lw^*, Lwmax]$.

3. **Modify the luminance:** Apply Larson's et al. histogram adjustment algorithm, [4], to each interval. Map $[Lwmin, Lw^*]$ to $[Ldmin, Ld^*]$ and $[Lw^*, Lwmax]$ to $[Ld^*, Ldmax]$. The important point here is that when using the human contrast sensitivity function, the mapping does not produce a contrast greater than the one present in the original scene. Since we are working in regions, this means controlling the contrast over that given region. This step is not a traditional histogram adjustment, it modifies the output range and the distribution within it.

   It is clear that if we select Ld^* close to $Ldmax$ we will be displaying the dark areas, $[Ldmin, Ld^*]$, with a wider range than the bright areas. Thus, by changing Ld^* we modify the resources assigned to each interval of the original luminance. A first solution to the problem of visualization is to construct a movie by increasing Ld^* from $Lwmin$ to $Lwmax$. This gives us a continuous movie which, starting from the image with all resources allocated to the dark areas, slowly moves to an image with all resources allocated to the bright areas. A second possibility is to choose just a certain fix number of images.

   In the case of three intervals or more, the idea is the same. We start with all resources allocated to the first interval and we swap them to the next interval on the right.

4. **Quantization:** Quantize and gamma correct the recomputed primaries $(\frac{r}{L} * L', \frac{g}{L} * L', \frac{b}{L} * L')$ to obtain the digital output values (R, G, B).

2.2 Information assessment

As we mentioned before, the evaluation of the output image is mostly subjective. However, it would be of interest to have an automatic procedure to asses the information content of each image in the set. With it, we will be able to extract the best or set of best images in the sequence.

If we consider each image as a message, its information can be measured with Shannon's entropy function. We can therefore use it as a plausible measure of the image information. From the histogram of the output image we find the probabilities of each output level, p_i, and with them we compute the Shannon's entropy as:

$$H = -\sum_i p_i \log p_i$$

Note the relation between histogram equalization and entropy. The entropy is maximum when all symbols are equally probable, which implies a flat histogram. Furthermore, entropy maximization captures our subjective preference towards well contrasted images.

3 Results

The first example is shown in figures 1-4. This image was captured with Devebec's method [1]. Before going into the results obtained for this image, we give specifics on the parameters of the algorithm. The first step is the segmentation of the original image (see Figure 1). We segmented it in four regions, detailed in Table (1). Secondly, we assign the original resources for each region. In this case we assumed $Ldmax = 100cd/m^2$ and $Ldmin = 2cd/m^2$, and selected initial ranges: $[74, 8, 8, 8]$ (cd/m^2). For the next image in the set, the resources are re-allocated moving them from left to right in steps of 6 cd/m^2. (The second image was computed with $[68, 14, 8, 8]$ cd/m^2 per region.)

From the data in Table 1 we can draw several conclusions. First of all, as we can see from the figures in the table, the first regions have very small luminance values. That means that is

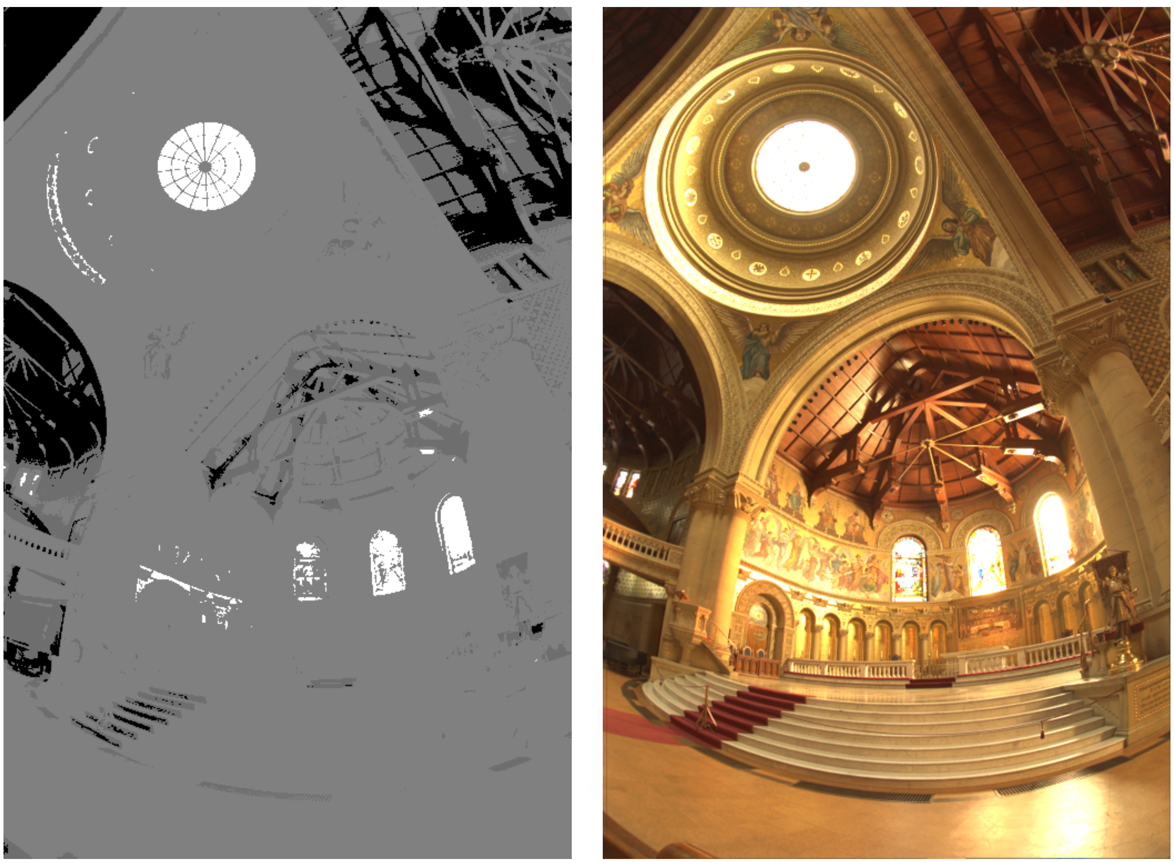

Figure 1: Left: Segmented image into four regions. Right: The image with a linear tone mapping.

Region	Range(cd/m^2)	%
0	(0.12,0.70)	1.45e-3
1	(0.70,1.80)	2.72e-3
2	(1.80,148.41)	0.36
3	(148.41,40427.39)	99.63

Table 1: Information for Memorial church image.

very difficult, even being in front of the real scene, to perceive something there. Furthermore, the dynamic range for this region is extremely small, representing only the 1.45e-3 % of the total input range. In other words, the region 0 expands an insignificant range from the total. Here, is clear that is not only a problem of output range, is in fact a problem of resolution within regions in the scene. Now, if we assume that we do want to "see" all over the image, we apply the algorithm discussed before. In every region we modify the luminance distribution.

In Figure 2 we show the images with maximum resources per region. With these images we can see more details than in the image processed with Larson's algorithm (Figure 4). The resulting images in Figure 2 are fairly natural, specially for the last two images. The four images selected visualize more information, specially in very dark and very bright regions. Hence not only we manage to display the original scene in one image, we make visible some information that was obscured in the original scene. Note for example the details in the dark corners and in the bright windows. In the first image observe the upper left corner, the details in the ceiling are clearly visible now. In the second image observe the wall at the left. The third one represents all the walls. And finally, the fourth image represents the windows. Note how we make visible the details in the windows. (This areas are completely lost with a linear tone mapping (simple linear map of the input range onto the output range), Figure 1) The simple algorithm presented here produces a movie, from which we extracted the images in Figure 2, that contains images that "add light" and capture details not present in Figure 4.

If we compute the entropy per region and select the image which maximizes it we obtain the images in Figure 3. The images which maximize the entropy are very close to the ones with maximum resources allocated to them.

If we take the image which maximizes the global entropy and we compare it with the image computed with Larson's algorithm, see Figure 4, we can see that it contains enough details everywhere. This is due to the modification of the luminance distribution, we expanded regions that were originally very compressed. This point it is important, shows how the method presented here can be used to obtain only one image if needed.

The same methodology was applied to the bathroom image (Figure 5). Here we selected two regions with limits $[3, 299, 177222]$. Figure 6 shows the segmented image together with the image obtained with a linear tone mapping. In this case is not so clear the need for two images. Only some details in the wooden frame are more visible in the image on the left upper corner. Obviously, the image with all resources allocated to the bright region captures in a better way the details in the lamps. Finally, the image with maximum global entropy balances both regions and is comparable with the image obtained by Larson's algorithm.

The last example is presented in Figure 7. The segmentation is with limits $[0.22, 12.18, 406.50]$. Figure 9 shows the segmented image and the image obtained with a linear mapping. The details in the floor and in the wooden desk are more clear in the top image on Figure 7, this details are

Figure 2: These four images correspond to maximum resources allocated to each region, that is: [74,8,8,8],[8,74,8,8],[8,8,74,8] and [8,8,8,74] cd/m^2.

Figure 3: For each region we show the image with maximum entropy in the region.

Figure 4: Left: image processed with Larson's algorithm. Right: the image with maximum global entropy among all the images computed. Note how this image shows more details.

difficult to visualize in the bottom image. Finally, once again, the image with maximum global entropy offers a balance between both images and outperforms in this respect the image obtained with Larson's algorithm (Figure 8).

4 Conclusions

In this paper we have presented a new paradigm for the reproduction and visualization of high dynamic range images. We argued for the use of a set of images instead of a single one as in traditional approaches. More than being this the last word about the problem of visualizing high dynamic range data, with this work we attempted to illustrate the intrinsic limitation of working with only one image. We showed how going for more than one image we can obtain a simple and nice solution to the problem of complete information visualization of high dynamic range images.

A number of questions remain open after this work. First of all, the specific algorithm here described for the computation of the set of images is just a particular example, and others should be developed. One of the crucial additional points is how to find the minimal number of images required to visualize all the relevant information. These images have also to be pleasant and hopefully with smooth transitions among them. We hope that the work here presented will open the door to works on these and other relevant questions.

Acknowledgments

The Memorial church image was taken from Paul Debevec's web page, www.cs.berkeley.edu/debevec. The office and bathroom images were taken from Larson's web page, positron.cs.berkeley.edu/gwlarson. The authors thank Jack Tumblin for explaining his work [10], and Marcelo Bertalmio for fruitful discussions. This work was partially supported by a grant from the Office of Naval Research ONR-N00014-97-1-0509, the Office of Naval Research Young Investigator Award, the Presidential Early Career Awards for Scientists and Engineers (PECASE), a National Science Foundation CAREER Award, by the National Science Foundation Learning and Intelligent Systems Program (LIS), and by CSIC - Universidad de la República, Uruguay.

References

[1] P. Debevec and J. Malik. Recovering High Dynamic Range Radiance Maps from Photographs. In *SIGGRAPH 97 Conference Proceedings*, pages 369–378, 1997.

[2] J. DiCarlo and B. Wandell. Rendering High Dynamic Range Images. In *Proceedings of the SPIE, Image Sensors.*, volume 3965, January 2000.

[3] J. Ferwerda, S. Pattanaik, P. Shirley, and D. Greenberg. A Model Of Visual Adaptation for Realistic Image Synthesis. In *SIGGRAPH 96 Conference Proceedings*, pages 249–258, 1996.

[4] G. W. Larson, H. Rushmeier, and C. Piatko. A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes. *IEEE Transactions on Visualization and Computer Graphics*, 3(4):291–306, October-December 1997.

[5] S. Nayar and T. Mitsunaga. High Dynamic Range Imaging: Spatially Varying Pixel Exposures. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 472–479, June 2000.

Figure 5: In the first row we have the images with maximum resources per region. In the second row we have, on the left the image with maximum global entropy and on the right the one computed with Larson's algorithm.

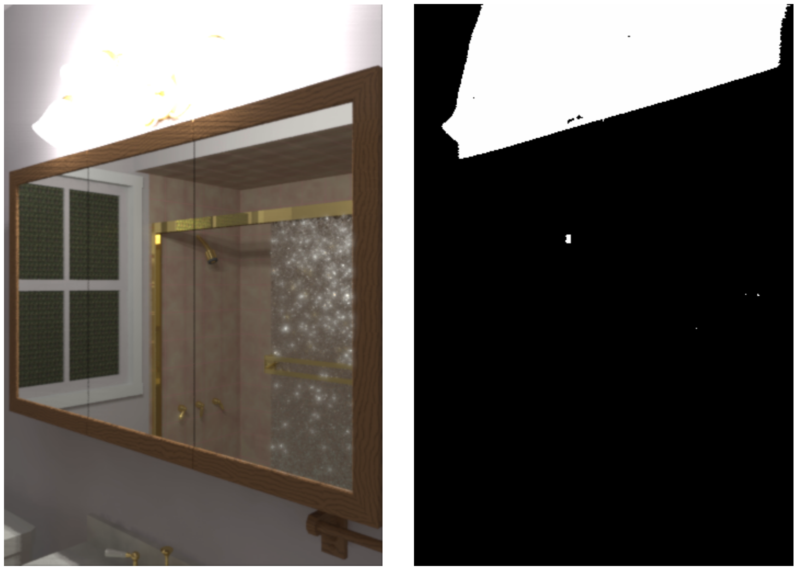

Figure 6: Left: Image with a linear tone mapping. Right: Segmented image.

[6] S. Pattanaik, J. Ferwerda, M. Fairchild, and D. Greenberg. A Multiscale Model of Adaptation and Spatial Vision for Realistic Image Display. In *SIGGRAPH 98 Conference Proceedings*, pages 287–298, 1998.

[7] C. Schlick. Quantization Techniques for Visualization of High Dynamic Range Pictures. In *Photorealistic Rendering Techniques, Proceedings of the 5th Eurographic Rendering Workshop*, pages 7–20. Springer-Verlag, June 1994.

[8] J. Tumblin, J. Hodgins, and B. Guenter. Two Methods for Display of High Contrasted Images. *ACM Transactions on Graphics*, 18(1):56–94, January 1999.

[9] J. Tumblin and H. Rushmeier. Tone Reproduction for Realistic Images. *IEEE Computer Graphics and Applications*, 13(6):42–48, November 1993.

[10] J. Tumblin and G. Turk. LCIS: A Boundary Hierarchy for Detail-Preserving Contrast Reduction. In *SIGGRAPH 99 Conference Proceedings*, 1999.

[11] B. Wandell, P. Catrysse, J. DiCarlo, D. Yang, and A. E. Gamal. Multiple Capture Single Image Architecture with a CMOS Sensor. In *Proceedings of the International Symposium on Multispectral Imaging and Color Reproduction for Digital Archives. Japan.*, pages 11–17, 1999.

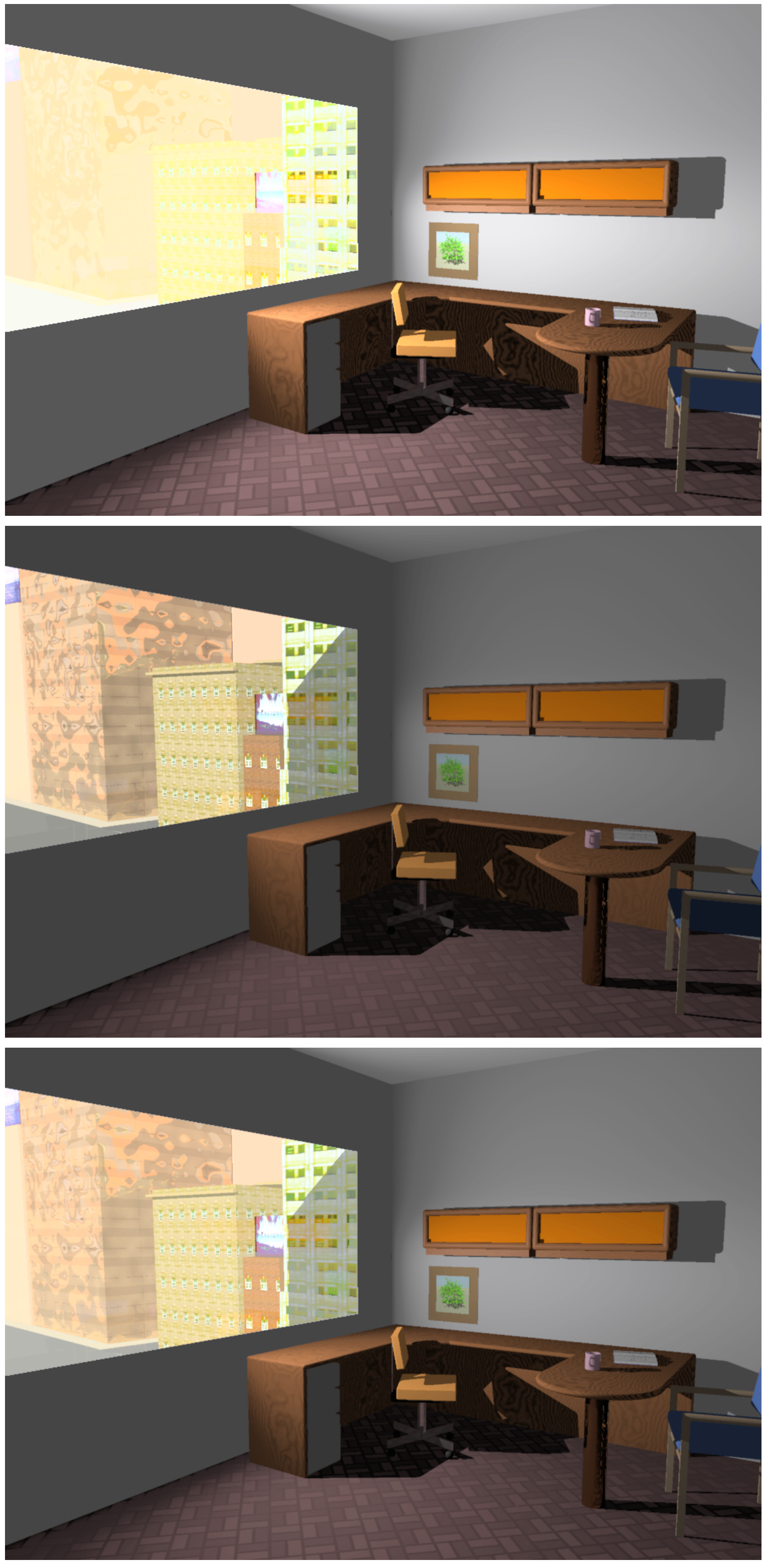

Figure 7: The top and middle images correspond to maximum resources per region. Bottom: image with maximum global entropy

Figure 8: Top: Image computed with Larson's algorithm. Bottom: Image with maximum global entropy.

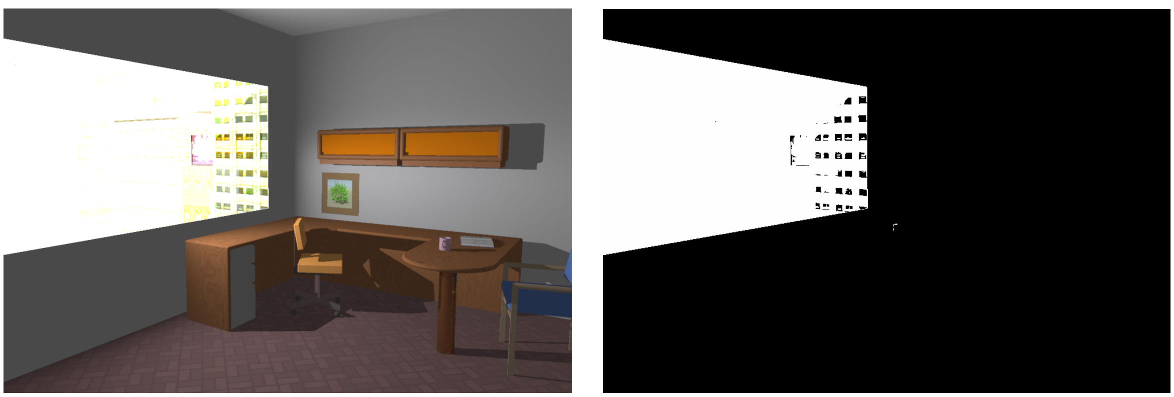

Figure 9: Left: Image with a linear tone mapping. Right: Segmented image