Generating Stereoscopic HDR Images Using HDR-LDR Image Pairs

ELMEDIN SELMANOVIĆ, KURT DEBATTISTA, THOMAS BASHFORD-ROGERS,
and ALAN CHALMERS, University of Warwick

A number of novel imaging technologies have been gaining popularity over the past few years. Foremost among these are stereoscopy and high dynamic range (HDR) Imaging. While a large body of research has looked into each of these imaging technologies independently, very little work has attempted to combine them. This is mostly due to the current limitations in capture and display. In this article, we mitigate problems of capturing Stereoscopic HDR (SHDR) that would potentially require two HDR cameras, by capturing an HDR and LDR pair and using it to generate 3D stereoscopic HDR content. We ran a detailed user study to compare four different methods of generating SHDR content. The methods investigated were the following: two based on expanding the luminance of the LDR image, and two utilizing stereo correspondence methods, which were adapted for our purposes. Results demonstrate that one of the stereo correspondence methods may be considered perceptually indistinguishable from the ground truth (image pair captured using two HDR cameras), while the other methods are all significantly distinct from the ground truth.

Categories and Subject Descriptors: I.4.0 [**Image Processing and Computer Vision**]: General—*Image processing software; Image displays*; I.3.3 [**Computer Graphics**]: Picture/Image Generation—*Display algorithms*

General Terms: Design, Experimentation, Performance

Additional Key Words and Phrases: Stereoscopy, high dynamic range imaging, stereo correspondence, expansion operators

ACM Reference Format:

Selmanović, E., Debattista, K., Bashford-Rogers, T., and Chalmers, A. 2013. Generating stereoscopic HDR images using HDR-LDR image pairs. ACM Trans. Appl. Percept. 10, 1, Article 3 (February 2013), 18 pages.
DOI = 10.1145/2422105.2422108 http://doi.acm.org/10.1145/2422105.2422108

1. INTRODUCTION

Current developments in visual computing have led to a number of emerging new technologies that have the potential of revolutionizing the way digital media is created and consumed. This article presents an investigation into enabling the combination of two such technologies: HDR and stereoscopy.

This work was supported in part by EPSRC grant EP/1038780/1. E. Selmanović is funded by goHDR Ltd. within TSR project 14345-87267. This work is part of EU COST Action IC1005.
Authors' address: International Digital Lab, WMG, The University of Warwick, Coventry, CV4 7AL UK; email: elmedins@gmail.com.

DOI 10.1145/2422105.2422108 http://doi.acm.org/10.1145/2422105.2422108

HDR imaging is a relatively novel imaging method [Banterle et al. 2011; Reinhard et al. 2010] that enables the capture, storage, processing, and delivery of real-world lighting. The traditional method of imaging, commonly termed low dynamic range (LDR), is not capable of capturing and displaying all the real-world luminance simultaneously, leaving many parts of a scene under- or over-exposed. Until recently, most image capture devices would have been LDR devices. HDR video capture has only become a reality recently, and even so it is limited to a few research-focused camera developments [Chalmers et al. 2009; Tocci et al. 2011]. HDR displays, such as the one developed by Seetzen et al. [2004], remain unavailable to the mass market. However many methods, called tone mappers, map the HDR content to the lower dynamic range of traditional displays to provide an enhanced viewing experience [Delvin et al. 2002; Banterle et al. 2011].

Stereoscopy is an imaging technique which enables or improves the illusion of depth by presenting two offset images to each of the viewer's eyes [Crone 1992]. The traditional representation of depth, present in paintings, photographs, and television relies on monoscopic cues (object size, linear perspective, shading, texture gradients) only. Stereoscopic cues such as binocular disparity [Shirley and Marschner 2009] and occlusion [Nakayama and Shimojo 1990] have been shown to improve distance judgments [Servos et al. 1992] and task performance [Ware and Mitchell 2008]. Lately, the entertainment industry has started bringing stereo technology to mass market once more (there was an attempt in 1950s [Mendiburu 2009]) through stereoscopic cinema showings, dedicated stereoscopic TV channels, and a wide range of stereoscopic display devices.

Stereoscopic high dynamic range (SHDR) imaging has the ability of bringing these diverse technologies together, exploiting the advantages of both. This novel imaging method with an unprecedented level of realism has the potential to deliver both improved depth perception and a realistic representation of the scene lighting. This article investigates the creation aspect of SHDR based on capturing an HDR and LDR image pair, and compares their performances in a detailed user study.

Since HDR cameras are rare and expensive at the moment, we propose to generate SHDR content using an HDR and LDR pair; see Figures 1(a) and 1(b). The LDR image is artificially boosted to an HDR image using one of the four methods proposed below. Two of the methods use expansion operators (EOs), which expand a generic LDR image to HDR. The other two are adapted from stereo-matching algorithms. They generate a new HDR image from the LDR image using the original HDR image and the LDR image as guidance.

In this article, we compare the four proposed methods in a user study. The SHDR images are composed of two images: the first being a true HDR image and the other, either an HDR image created using one of the four techniques, or a true HDR, used as reference image. The SHDR content was displayed using two HDR displays and a stereo rig. The paired comparisons method was used, as it allows the possibility of comparing all the techniques with each other and the reference, providing a detailed ranking. Results show that most of the methods are significantly different. However (importantly), the reference was not considered significantly different from one of the methods for the scenes used, indicating that such a method produces a result which was perceptually indistinguishable from an HDR-HDR pair. This means that the same quality would be available at decreased cost and the adoption of SHDR could be accelerated.

The major contribution of this work is demonstrating that creating SHDR from an HDR-LDR pair is feasible. Other contributions include applications of stereo-matching algorithms to creating an acceptable HDR image from an LDR image. We also present a detailed user study clearly demonstrating the potential of this approach.

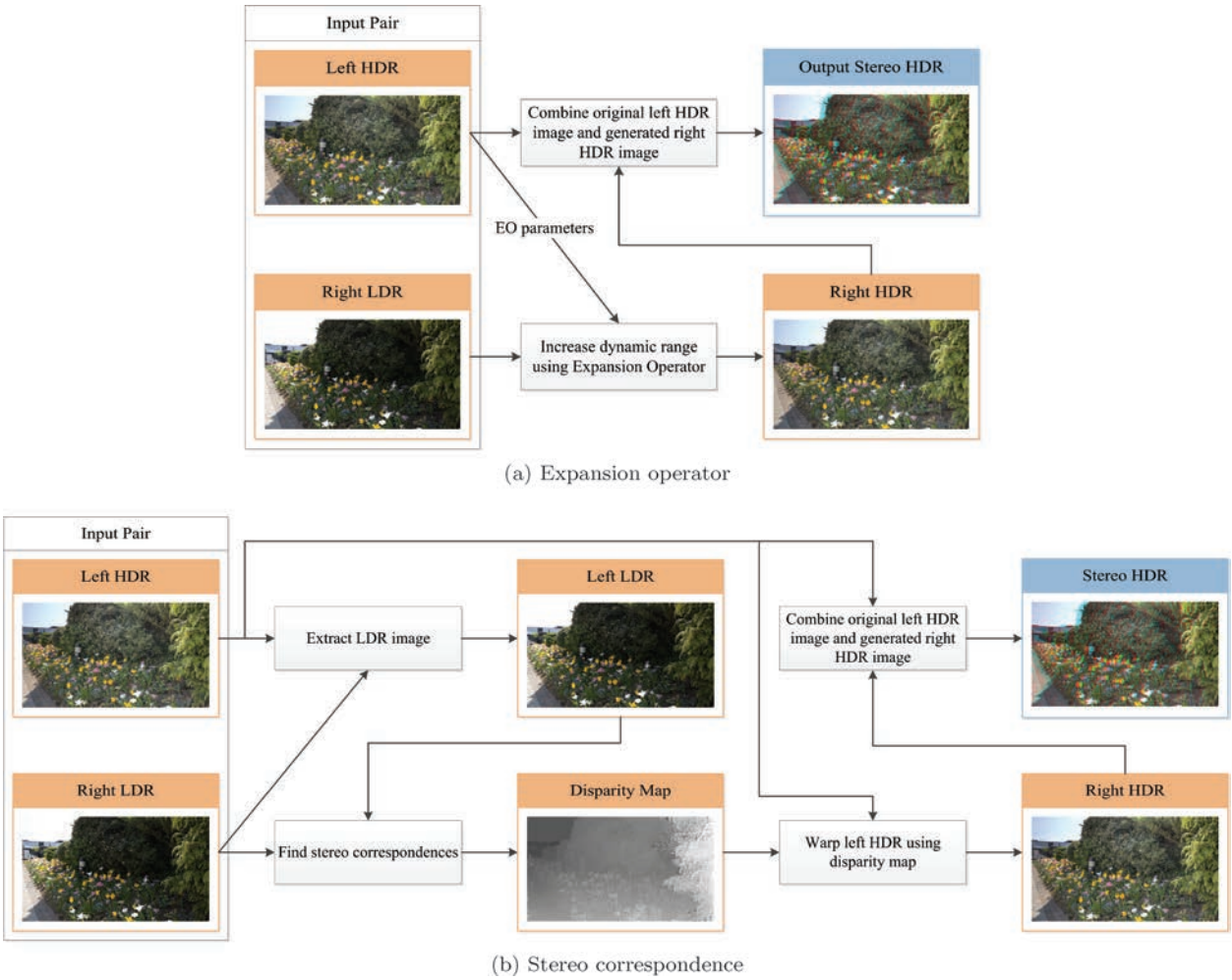


Fig. 1. SHDR capture methods.

2. RELATED WORK

The concept of generating or capturing a stereo pair where each image has different attributes has been employed before. Sawhney et al. [2001] used images which differed in spatial resolution. Their motivation was to speed up rendering and reduce capture cost. Stereo correspondence was utilized to guide the warp and the transfer of high quality data to the low quality one, thus enhancing it. In addition, as the technique was used to process videos it was possible to apply temporal matching and increase the reliability of the correspondences. Errors caused by unreliable matching were solved by up-sampling the low quality regions.

A similar technique which used images with different spatial resolutions was proposed by Lo et al. [2009]. However, instead of seeking to improve the low quality one they tested how the human visual system (HVS) copes with the dissimilarity. They suggest that HVS is capable of fusing two views into a perceptually equivalent one, as long as the resolution difference of the pair is within a certain range.

However, this range was shown to be limited, and the technique of Sawhney et al. [2001] is able to cope with bigger differences.

Bhat et al. [2007] proposed a generic approach for improving low quality video using high quality images. This method used image correspondences to guide transfer of data from images to video. These correspondence maps are obtained using structure from motion and multiview stereo algorithms. The maps are refined and data is transferred using spatial and temporal gradient fields. The method has a wide range of applications, and the ones suggested are the improvement of spatial resolution, dynamic range and lighting, removal of objects and camera shake from the video, and more efficient video editing. However, this method is limited to static scenes only. Also, the authors do not provide any quantitative results and only a single example for each application is given. Finally, very slow computation speeds (five minutes per single low resolution image) are reported, although the authors suggest that there may be space for improvement.

A multiple sensor approach has also been used by Wilburn et al. [2005]. Large arrays of low quality video cameras were set-up to produce high quality video. The authors argue that with multiple low cost cameras it is feasible to generate similar or better video than using high-end consumer products. Authors explored different applications for such arrays including increasing the resolution, frame-rate and dynamic range of the video, and simulating camera motion and large camera aperture. Even though, with modifications, this system could be used for capturing SHDR video, this had not been examined. Individual components that make the system are inexpensive, but the proposed solution contained one hundred camera sensors, lenses and processing boards. These were connected and controlled by four PCs. Engineering an entire system may require considerable assembly. The reported method could store both compressed and uncompressed video before processing, and the authors reported that two and a half minutes takes up 2GB using MPEG compression.

Lin and Chang [2009] suggested a form of creating HDR images using stereo. In their work, each image of the pair was taken at a different exposure level and combined to generate HDR images. They also used stereo correspondence to merge data between images. While they use this to generate HDR images, it is possible to use it for generating SHDR. Capturing SHDR using this technique would require only two LDR cameras, making the approach rather appealing. However, in order to generate a satisfactory disparity map used for calculating HDR values, over- and under-exposed regions need to be avoided in both images. This forces the image pair to be captured using very close exposure values, significantly limiting the potential dynamic range of generated images. This can be seen form the examples provided by the authors. In contrast, we are able to capture the whole dynamic range using one camera.

Techniques for compressing SHDR images have been proposed by Selmanović et al. [2012]. In their work they examined five different methods which were backwards-compatible with traditional and LDR stereo image viewers. Initially, each image in the pair was compressed using backwards-compatible HDR compression. In their work, authors used JPEG-HDR [Ward 2006] coding, but similar approaches could have been used as well. After the initial step, two of the methods relied on LDR stereo techniques to store images in side-by-side and half side-by-side fashion. Other two methods exploited low frequency, low range, and single channel attributes of the stereo image disparity map for coding. Here, the disparity map was saved together with a single image. During the decoding stage, the disparity map was used to reconstruct the missing HDR image. The two techniques differed in a way they generated these maps. The first used a low frequency disparity map which compressed better, while the second used high frequency disparities for problematic areas and produced better visual results. The final technique, which produced the best quality per bit when compared to ground truth, used motion compensation.

The method of paired comparisons, used for the experimental design in this article, is explained comprehensively by David [1988]. The method allows comparison of entities in pairs based on a given property, and results in their ranking. Details of the method, relevant for this research, and its application, are discussed further in Section 4. Additional statistical tests and derivations of formulas used for data analysis can also be found in the work by David [1988].

This method has also been successfully applied in a manner similar to ours by Ledda et al. [2005], Banterle et al. [2009b], Rubinstein et al. [2010], and Navarro et al. [2011]. Ledda et al. examined which tone-mappers generated an LDR image that was perceptually closest to the original HDR image. Banterle et al. used the method to compare which of the five EOs produced an image closest to the native HDR one. Rubinstein et al. applied the method of paired comparisons in a similar user study comparing image retargeting operators. Navarro et al. explored how high-level properties of rendering and different rendering parameters influence perceptional quality of motion blur in CG images.

3. LDR TO HDR METHODS

We propose the technique of acquiring an SHDR image that transforms a captured stereo pair of HDR-LDR images to an HDR-HDR pair. We consider two general methods for generating the HDR image from the LDR one. The first relies on established methods for increasing dynamic range of an image by applying EOs [Banterle et al. 2011] (Figure 1(a)). Expansion is performed using parameters obtained from the available HDR image. The second uses stereo correspondence to transfer HDR data to an LDR image (Figure 1(b)). For clarity, the figures shown depict a case where the left image is captured in HDR while the right is LDR, but sides can be interchanged.

Restoration of an HDR image is an ill-posed problem, as there is a lack of data in any over- and under-exposed areas of the image. This means that in most circumstances the reconstructed HDR image will not perfectly correspond to the captured scene. Still, it has been shown that the HVS can cope with discrepancies in a stereo image [Lo et al. 2009], and thus we hypothesise that these methods could provide a viable and efficient solution for generating SHDR.

3.1 Expansion Operators

The general class of operators that take the LDR image as an input and produce an HDR image as output have been given various titles such as expansion operators, inverse tone mappers (iTMOS) or inverse tone reproduction operators. Broadly, they can be divided into two categories: linear and nonlinear. Linear EOs apply a single expansion function over all of the image pixels, while nonlinear ones try to classify an image into regions and expand them based on local image information. Linear EOs have been shown to be less successful, as they introduce more image artifacts such as contours and halos and produce less precise colors and luminance values [Banterle et al. 2011].

Two EOs were tested in our experiment; one from each group. The linear operator selected was linear scaling (LS), as suggested by Akyüz et al. [2007], while the nonlinear used expand maps (EM) proposed by Banterle et al. [2006].

3.1.1 *Linear Scaling (LS).* Linear scaling is a straightforward way of expanding dynamic range of the LDR image. The luminance values of pixels are normalized and then scaled so that the maximum luminance corresponds to the desired value. Formally, this can be written as in Eq. (1):

$$L_{\mathrm{o}}(\mathbf{x}) = k\left(\frac{L_{\mathrm{i}}(\mathbf{x}) - L_{\min}}{L_{\max} - L_{\min}}\right)^{\gamma_{\mathrm{A}}}, \tag{1}$$

where k is the maximum luminance to be achieved, $\mathbf{x}$ are the coordinates of the processed pixel, L_i is the luminance value of the input, L_min and L_max are the minimum and maximum luminances of the input image, respectively, γ_A is the nonlinear scaling factor, and L_o is the output luminance value.

The main advantage of this operator is its low computational cost. It can easily be implemented to work in real time [Banterle et al. 2009a]. However, by its definition, quantization artifacts are to be expected. It lacks a step which would try to alleviate these problems or correct for luminance inconsistencies that can be expected to occur, especially in the over-exposed and under-exposed areas.

3.1.2 *Expand Maps (EM).* The expand maps technique starts, similarly to the linear scaling, by expanding the input's image range. This technique used a more advanced operator suggested by Reinhard et al. [2002], which is shown in Eq. (2).

$$L_\mathrm{o}(\mathbf{x}) = \frac{1}{2} L_\mathrm{max} L_\mathrm{white} \left(L_\mathrm{i}(\mathbf{x}) - 1 + \sqrt{(1 - L_\mathrm{i}(\mathbf{x}))^2 + \frac{4}{L_\mathrm{white}^2} L_\mathrm{i}(\mathbf{x})} \right), \tag{2}$$

where L_max is the maximum luminance to be achieved in cd/m^2, L_white controls the shape of the expansion curve, and L_i and L_o are input and output luminances, respectively. A subsequent step, following the inverse tone mapping, estimates missing information in the over-exposed areas and tries to minimize problems of quantization. This is done using the expand map which is a low frequency version of the input image. Further details can be found in Banterle et al. [2006].

So far, the EM operator has been shown to be the closest at reproducing a reference HDR image [Banterle et al. 2009b]. In addition, it is fully automated, therefore not requiring any user input. On the other hand, it is slower than LS, as it requires a high-end GPU to achieve real-time speeds. Also, there is still no guarantee that values are going to correspond to the ones in the captured scene or the other HDR image, especially given that there is no handling of under-exposed regions.

3.2 Stereo Correspondence

The techniques that rely on stereo correspondence utilize the high degree of correlation in stereo images. The correlation occurs because both images of the stereo pair represent the same scene taken simultaneously from slightly different positions. Therefore, most of the HDR data is already captured in one of the images and its respective place in the other needs to be found.

The process starts by choosing an exposure from the HDR image that matches the exposure at which the LDR image was captured. This can be achieved by traversing through different exposures of the HDR and extracting LDR slices. The slice which minimizes the difference of number of under- or over-exposed pixels between the available LDR image and itself gets selected. If the color difference between images is large, color matching can be performed using one of the available techniques (e.g., work by Pouli and Reinhard [2011]). Stereo correspondences between the extracted exposure and the available LDR image are subsequently found, resulting in a disparity map of the pair. This map is then used to copy pixel values from the original HDR image to their corresponding positions on the target HDR image.

Stereo correspondence has been widely addressed in previous publications. The Middlebury stereo testbed [Scharstein et al. 2001], which evaluates performance of dense stereo correspondence algorithms, has tested more than one hundred of them. These algorithms can be divided into two groups: local and global. Local algorithms try to find a disparity for each pixel separately, while global ones attempt to minimize an energy function for all the pixels. Local algorithms are efficient and easy to implement, but may be less accurate. Results of global methods contain fewer errors, but take more time to process images.

We have chosen a representative of each of the methods. The first one, the minimization of the sum of absolute differences (SAD), was chosen for its efficiency and simplicity. The second, a more advanced one, was proposed by Kolmogorov and Zabih [2001], and selected due to the method's ability to explicitly identify occlusions.

3.2.1 *Sum of Absolute Differences (SAD) Correspondence.* SAD is a basic method to decide if two pixel values correspond [Cyganek and Siebert 2009]. In its most straightforward implementation, the closest match for each pixel is found by selecting the pixel in the second image that gives a minimal sum of absolute RGB value differences. It is assumed that stereo images are rectified, reducing the search to a single horizontal line of pixels. In addition, the search is usually limited to a certain range. In order to reduce the number of equally good matches, instead of a single pixel, a window of pixels is matched between the images. This is calculated by

$$SAD(x,y) = \sum_{k\in R,G,B} \sum_{(i,j)\in W(x,y)} |I_{k,1}(x+i, y+j) - I_{k,2}(x+d_x+i, y+j)|, \tag{3}$$

where $W(x,y)$ are point coordinates of a window located at (x,y), $I_{k,l}(x,y)$ are the intensity values of the k-th channel of the l-th image at (x,y), d_x is a horizontal image displacement, and $SAD(x,y)$ is the value representing the difference between the compared regions.

Using SAD for stereo matching provides results quickly and can be implemented in real time. However, this method provides inconsistent matches, with a high signal to noise ratio, occasionally resulting in wrong pixel values being transferred.

3.2.2 *Correspondence with Occlusion via Graph Cuts (COGC).* To overcome the problem of noise, more sophisticated algorithms such as COGC impose a smoothness constraint. They assume that disparities vary smoothly in the scene, with occasional sharp discontinuities. In addition, COGC is able to detect occluded pixels. COGC works by using graph cuts to minimize the energy function shown in Eq. (4).

$$E(f) = E_{\text{data}}(f) + E_{\text{occ}}(f) + E_{\text{smooth}}(f), \tag{4}$$

where f is the configuration of disparity values, the E_{data} term represents differences between corresponding pixel values, the E_{occ} term penalizes occluded pixels, and E_{smooth} imposes smoothness between neighboring pixels. For a detailed explanation of each term, please refer to Kolmogorov and Zabih [2001].

The strong point of this method is the generation of reliable disparity maps. However, the feature that proved useful to this work was occlusion recognition. Advanced correspondence algorithms are good at assigning correct disparities to occluded regions using the smoothness constraint, which imposes depth values of adjacent regions. However, these values cannot be used for transferring pixel data between images because occlusions mean that the data is hidden/unavailable. If these values are used, the edges of the occluding objects get mistakenly transferred and shifted, violating the stereo effect. The ability to explicitly recognize and identify these areas enables correction. Values for these regions are found using SAD as described above.

COGC is still prone to errors, and when they do happen they might affect bigger regions. Also, function minimization is time consuming, and tens of minutes are needed to compute a full high definition image on a modern standard PC.

4. THE EXPERIMENT

The purpose of this experiment is to identify a technique, from the four mentioned in the previous section, that is the most appropriate to enable SHDR from an HDR and LDR pair and to compare

Table I. Example Preference Table

Technique	T1	T2	T3	T4	T5	**Score**
T1	–	0	1	0	1	**2**
T2	1	–	1	0	1	**3**
T3	0	0	–	0	0	**0**
T4	1	1	1	–	1	**4**
T5	0	0	1	0	–	**1**

such methods with a fully captured SHDR pair. The use of the paired comparisons method provides a straightforward choice between only two images at a time. This allows for small differences between compared objects to show. Additional statistics that reveal data consistency both within-participant and between-participant are available. Furthermore, the method of paired comparisons has been successfully used in other similar studies [Banterle et al. 2009b; Ledda et al. 2005; Rubinstein et al. 2010; Navarro et al. 2011]. For these reasons paired comparisons was chosen as the experimental method.

4.1 Design

A forced choice paired comparisons method was used to present all of the HDR from LDR techniques in pairs, applied to a number of different scenes, to all the participants. The task was to choose one of the techniques that looks more similar to the ground truth (GT). The chosen technique was said to be preferred.

More specifically, the *balanced paired comparison* method, which required every participant to perform every possible paired comparison, was used. For t techniques, n participants, and s scenes the total number of comparisons is $ns(t(t-1)/2)$.

In the experiment there were five different scenes and five methods. Besides the methods described in Section 3, GT was also compared. This work aimed to investigate whether there is significant difference between GT and the proposed methods. With 26 participants doing 50 comparisons each, the total number of comparisons was 1300.

We recorded the choices for each participant and for each of the scenes using a two-way preference table (e.g., Table I). If *Technique 1* is preferred to *Technique 3* (written $T1 \rightarrow T3$), value one is recorded in the row T1 and column T3 of the preference table and zero in the row T3 and column T1. In the example table, T1 is preferred to T3 and T5, but not T2 and T4. The table's principal diagonal is left empty, as a method cannot be preferred to itself. Entries below the diagonal are redundant but are recorded nevertheless. The last column of the table gives the score of a technique (denoted a_i) which measures how many times a method has been preferred. The total score in the preference table then is

$$\sum_{i=1}^{t} a_i = \frac{t(t-1)}{2}, \tag{5}$$

A preference table allows for the calculation of two measurements, namely: coefficient of consistence and coefficient of agreement. If these two coefficients are significantly high, it is possible to proceed and perform a test of equality and a range test, as explained below.

4.1.1 *Coefficient of Consistence (ζ).* When comparing three techniques, there are eight possible outcomes. Six outcomes have one of the techniques scoring two wins, another scoring one, and the last having none. In two cases, however, each technique scores one win (e.g., $T1 \rightarrow T2$, $T2 \rightarrow T3$, and $T3 \rightarrow T1$). These are called *circular triads* and they express an inconsistency of the participant, possibly caused by a small difference between very similar methods which requires a guess. The number of circular triads for any number of techniques can be calculated using Eq. (6), as proposed by Kendall

and Smith [1940].

$$c = \frac{t}{24}(t^2 - 1) - \frac{1}{2}T, \tag{6}$$

where $T = \sum(a_i - \bar{a})^2$, and $\bar{a} = \sum \frac{a_i}{t} = \frac{1}{2}(t-1)$.

After computing the number of circular triads it is possible to obtain *coefficient of consistence* ζ, which for an odd number of methods is defined by Kendall and Smith [1940] as

$$\zeta = 1 - \frac{24c}{t(t^2-1)}. \tag{7}$$

The coefficient ζ ranges from zero to one. A value of one means that no circular triads are present and preferences can be expressed as rankings, and as this value approaches zero, the number of triads increases and so do the inconsistencies.

4.1.2 *Coefficient of Agreement (u).* It is possible to test if participants doing comparisons make the same choices between themselves, that is, to calculate the *coefficient of agreement* u. Again, a preference table is used, but this time entries signify how many participants preferred each of the methods. If all the participants completely agree, half of the entries have value n (number of the participants) while the other half are zero.

The sum of agreements between pairs of participants Σ defined in Eq. (8) is calculated first.

$$\Sigma = \sum_{i \neq j} \binom{\alpha_{ij}}{2}, \tag{8}$$

where α_{ij} is the number of times *Technique i* is preferred to *Technique j*. The summation extends over $t(t-1)$ terms. Having found Σ, we calculate u as

$$u = \frac{2\Sigma}{\binom{n}{2}\binom{t}{2}} - 1. \tag{9}$$

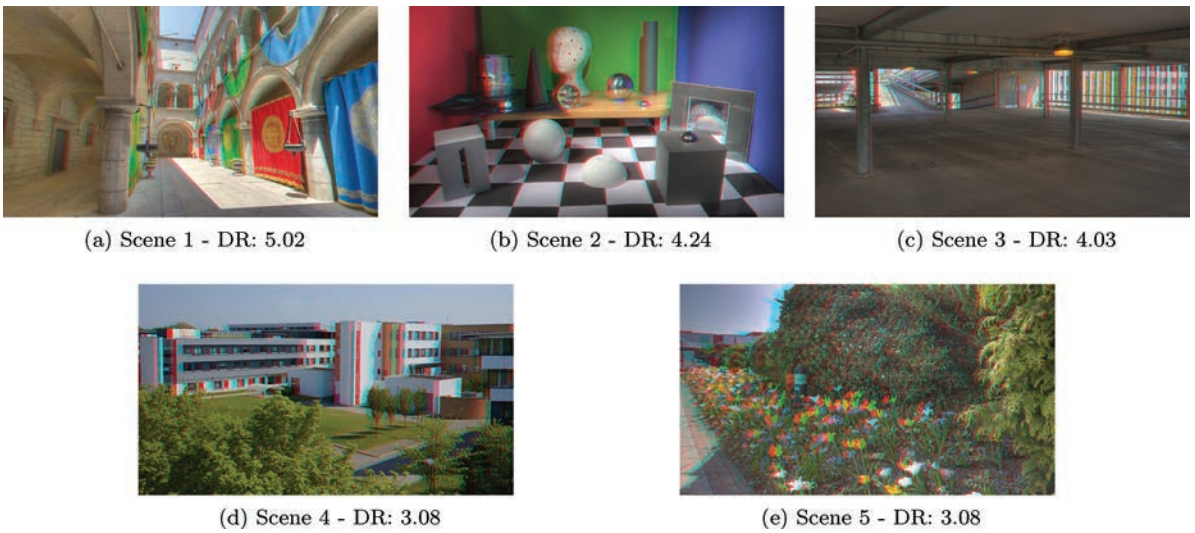
The maximum value u can take is one, which signifies all the participants completely agree. For complete disagreement, u takes its minimum value of $-1/(n-1)$ for an even number of participants or $-1/n$ for an odd number, and all preference table entries are $\frac{1}{2}n$ or $\frac{1}{2}(n \pm 1)$, respectively.

4.1.3 *Test of Equality and Range Test.* The overall test of equality examines whether the score differences are simply by chance or due to actual perceptual dissimilarities of methods. It verifies overall significance in data and assesses differences in preference scores (a_i) obtained by a specific method. This is similar in principle to ANOVA test, but more specific for use with ordinal data. Initially, a standardized sum of squares of the scores D_n using Eq. (10) is calculated.

$$D_n = 4\frac{\left[\sum_{i=1}^{t} a_i^2 - \frac{1}{4}tn^2(t-1)^2\right]}{nt}. \tag{10}$$

For a detailed derivation of Eq. (10), see the book by David [1988]. The null hypothesis $H_0 : \pi_i = \frac{1}{2}$, where π_i is the average preference probability for method i, can be rejected if D_n exceeds or equals the critical value which is obtained from χ^2 tables using the desired significance level and $t-1$ degrees of freedom.

The test of equality shows that if there are statistical differences between the methods compared but it cannot tell where these differences, if present, lie. To this end, the multiple comparison range test is used. It determines the significance of score differences between the methods compared. This is



(a) Scene 1 - DR: 5.02 (b) Scene 2 - DR: 4.24 (c) Scene 3 - DR: 4.03

(d) Scene 4 - DR: 3.08 (e) Scene 5 - DR: 3.08

Fig. 2. SHDR scenes (tone-mapped anaglyph) with dynamic range (Weber Contrast).

the analog of a post-hoc comparison test such as Tukey's test used in ANOVA. Any pairwise difference in scores that exceeds or equals R can be declared significant. R is calculated using Eq. (11).

$$R = \frac{1}{2} W_{t,\alpha} \sqrt{nt} + \frac{1}{4}, \tag{11}$$

where $W_{t,\alpha}$ is the upper α significance point of the W_t distribution of the variance-normalized range.

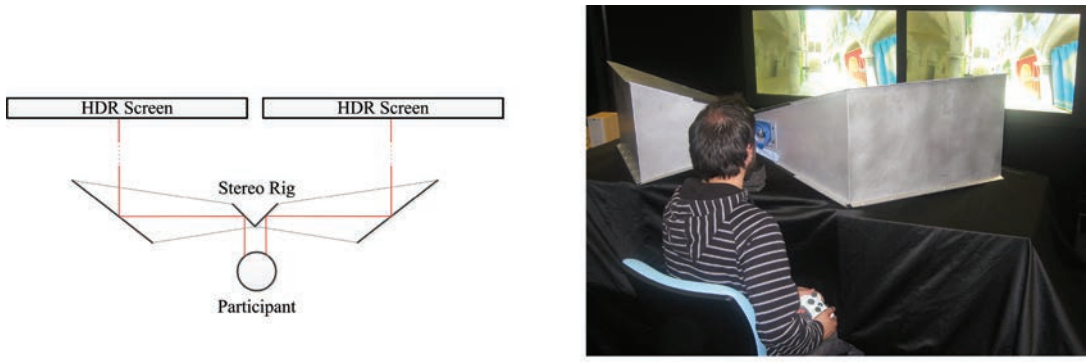
4.2 Participants

The number of tested participants was 26 (21 males and 5 females) with an age range between 20 and 52 (mean 31). All the participants were volunteers. They all had normal or corrected to normal vision and were able to perceive stereoscopy.

4.3 Materials

Five different HDR (see Figure 2) scenes in high definition resolution (1920 $\times$ 1080 pixels) were used for the experiment. The first was computer-generated using a physically-based renderer. Images were rendered using path tracing. The other four were captured using a *Canon 1Ds Mark II* camera and the multiple exposures technique [Deb]. For each scene and each eye position, seven exposures, separated by two f-stops, were taken and combined. To achieve stereoscopy, the camera was offset horizontally by 63 *mm* (mean interpupillary distance [Dodgson 2004]) and kept parallel. The middle exposure image was used as the the input for the LDR image.

SHDR images obtained as stated above were used as the GT. In addition, for each of the scenes, four other SHDR images were generated using the methods described in Section 3. To generate HDR from LDR using the LS method, γ_A was set to 1 as it produced the best results in the original paper [Akyuz et al. 2007]. The maximum luminance from the available HDR image was used to set parameter k. The same available HDR value was used when setting L_{max} and L_{white} of the EM method as suggested by the authors [Reinhard et al. 2002]. Stereo correspondence was performed between an LDR image and the matching single exposure of HDR (the middle image out of the seven combined for HDR). For the SAD technique, the window size was set to five. Horizontal search was constrained between



(a) Setup representation (not to scale) (b) Photo of setup

Fig. 3. Experimental setup.

± 40 pixels and vertical between ± 3 pixels for both the SAD and COGC techniques. For occlusion correction in COGC, the window was set to one. In addition, in the graph cut part of the algorithm, the order of labels was randomized every iteration, and the parameter λ, which controlled smoothness, was calculated automatically using heuristics for estimating noise in the images, as described by Kolmogorov [2004].

Images were presented using two custom-made 46" HD HDR displays based on the Dolby DR-37P HDR displays, with a luminance range of $0.15cd/m^2$ to $3000cd/m^2$. They were coplanar and separated by $4cm$. The displays were calibrated using procedures suggested by Ruppertsberg et. al. [2007]. A custom-built stereo rig with four highly reflective mirrors transferred images from the screens to the participants. The schematic and the real experiment setup are shown in Figure 3. The rig was positioned $1.6m$ from the screens and was centered both horizontally and vertically. Fine adjustments in image alignment were made by controlling the two front mirrors. Participants used a Microsoft XBox controller as the input device. Feedback information was provided using 2.1 stereo speakers and controller vibration.

4.4 Procedure

Each participant was presented with fifty randomized image sets, as five methods and five scenes were tested. All image sets contained three images: two random techniques to be compared (called A and B) and the GT. Participants were instructed to decide which of the two, A or B, appeared more similar to GT. Images were displayed for five seconds each, in the following sequence: A, GT, and B. During this period participants were only able to observe the displayed images, ensuring that each image was seen at least once. Subsequently, the controller vibrated, informing the participant that switching freely between any of the three images and selecting either image A or B was possible. We decided not to use side-by-side compassion, as perspective would change the perception of the stimuli. Revisiting any of the image pairs was permissable giving the participant the opportunity of providing a better informed decision. Fifteen seconds were allocated for this part and, after the time was up, a short sound was played, meaning that selection had to be made even if it required guessing. This would allow enough time for observing the images while still keeping the experiment length relatively short, avoiding participant fatigue. Every time an image was displayed, a sound notified which of the image pairs A, B, or GT was presented. Sound was also used to inform the participant that the next set had started. A uniformly grey color was displayed for 0.3 seconds prior to each image pair for eye desensitization. The screen showing the generated HDR image of the pair was chosen at random.

Table II. Experiment Results. Methods Within the Same Circle Cannot be Considered Perceptually Different

	mean ζ	u	χ^2	sig. u	D_n	sig. D_n	1st	2nd	3rd	4th	5th
Scene 1	0.769	0.279	79.692	$p < .05$	78.215	$p < .05$	GT	SAD	COGC	EM	LS
Scene 2	0.846	0.596	158.923	$p < .05$	141.107	$p < .05$	GT	COGC	SAD	EM	LS
Scene 3	0.938	0.602	160.615	$p < .05$	145.723	$p < .05$	GT	COGC	SAD	EM	LS
Scene 4	0.885	0.513	138.154	$p < .05$	131.815	$p < .05$	GT	SAD	COGC	EM	LS
Scene 5	0.931	0.562	150.462	$p < .05$	147.630	$p < .05$	SAD	COGC	GT	EM	LS
Average	0.874	0.510	137.569	$p < .05$	128.898	$p < .05$	GT	SAD	COGC	EM	LS

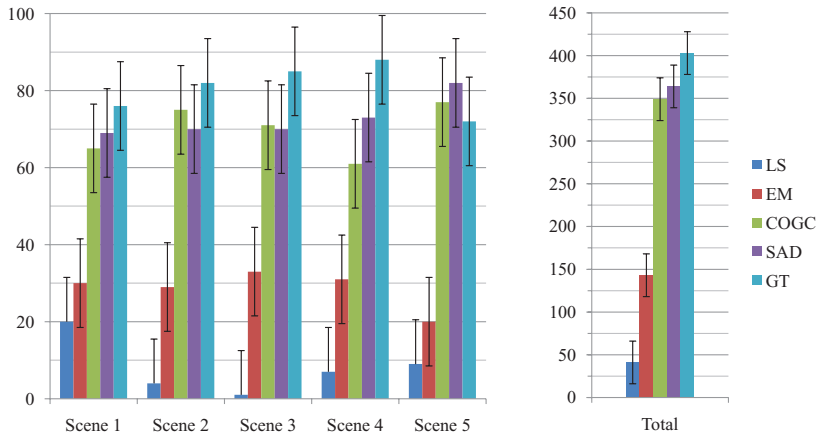


Fig. 4. Method preference. Black bars represent range R from Eq. (11). The value of R is 23 for single scenes and 50 for total.

The experiment was conducted in a dark room to minimize ambient light. Participants were allowed five minutes to adjust to the environment. Initially, a training set was shown to the participants such that they could familiarize themselves with the task. The set consisted of ten images, unrelated to the sets used in the actual experiment. After a one minute pause, the experiment started and lasted between twenty and thirty minutes, depending on how quickly participants made their choices.

5. RESULTS

The results of the experiment are shown in Table II. The mean coefficient of consistence (ζ) values were high and statistically significant for the given degrees of freedom. This indicated that participants understood the task, that the difference between some of the methods was big enough to be detected, and that results were reliable.

Preference tables for each scene and all the participants were generated. For a better visualisation, this data is represented as a graph in Figure 4. Each bar shows how many times a corresponding method was preferred. Aggregated data for all the scenes (labelled *Total*) is also presented.

The coefficient of agreement u was used to test the null hypothesis, which states that *selections were made at random*. The large sample approximation to the sampling distribution (χ^2) was used to determine significance of u (see Table II). Details of this test statistic are described by David [1988]. At

Table III. Root Mean Square Error of Log HDR Values as in Eq. 12

Operator	SAD	COGC	EM	LS
Scene 1	0.0021	0.0025	0.0324	0.0130
Scene 2	0.0057	0.0045	0.1457	0.3004
Scene 3	0.0183	0.0142	0.1401	0.1859
Scene 4	0.0076	0.0081	0.0713	0.1947
Scene 5	0.0024	0.0027	0.0705	0.0520

the standard $\alpha = 0.05$ level and for $\binom{t}{2} = 10$ degrees of freedom, the null hypothesis was rejected, and it was concluded that there is agreement between participants.

The sum of squares of the scores, D_n, was used for the overall test of equality. D_n was compared to a critical value of 9.45 (for $\alpha = 0.05$ significance level and $t = 5$ different methods). D_n exceeded the critical value for all the scenes causing rejection of the null hypothesis, $H_0 : \pi_i = \frac{1}{2}$. This implied that there is statistical difference between some of the methods.

The multiple comparison range test was used to determine which methods were statistically different from the others. The range, R, was calculated using Eq. (11). For $W_{t,\alpha} = 3.86$ (from Table 22 by Pearson and Hartley [1976]), $\alpha = 0.05$ and $t = 5$ value of range R was 23. To calculate R for the aggregated result, the total number of comparisons between two methods had to be accounted for. The number of participants n in Eq. (11) was multiplied by the number of scenes s, and gave $R = 50$. R is represented in Figure 4 with black range bars. The preference score of each method was compared with scores of all the other methods. If the difference was greater than R, it was declared significant. Visually, in Table II, methods which were not statistically different from each other were circled to represent a group. If a method is not grouped, it means it is significantly different by itself. The final five columns of Table II represent method rankings.

In addition to perceptual evaluation of methods, a *root mean square Error* measure on logarithmically scaled HDR values (RMSEL) was used to obtain an objective measurement of the differences between each method and GT. Values were scaled in order to avoid biasing the result towards high intensity:

$$\sqrt{\frac{\sum_{i=1}^{n} \lg(I_1(i)+1) - \lg(I_2(i)+1)}{n}}, \tag{12}$$

where n is the total number of HDR values (including all three channels) and $I_k(i)$ is the i-th value of the k-th image. The values are incremented by one to avoid negative results, which occur in the range between 0 and 1. RMSEL showed how much a generated HDR image differed from the GT. Here higher measurements represent a larger difference/error. The results are presented in Table III.

Figure 5 shows a comparison among all four methods for the reconstructed image in the stereo pair. This scene has many challenging regions including dark and bright regions, reflections, and refractions. It can be seen how the correspondence operators are superior to the chosen EOs for this scene. Figure 6 shows further details for Scene 2 for the two correspondence methods and GT. Individual and appropriate single exposures are selected for illustrative purposes for each inset, as some of them lie in very dark regions. Further analysis of this scene is provided in the next section.

6. DISCUSSION

Unsurprisingly, GT, was ranked in the first group for all the scenes and considered perceptually similar to itself. The SAD method was also continually found in the first group with GT, which was reflected in

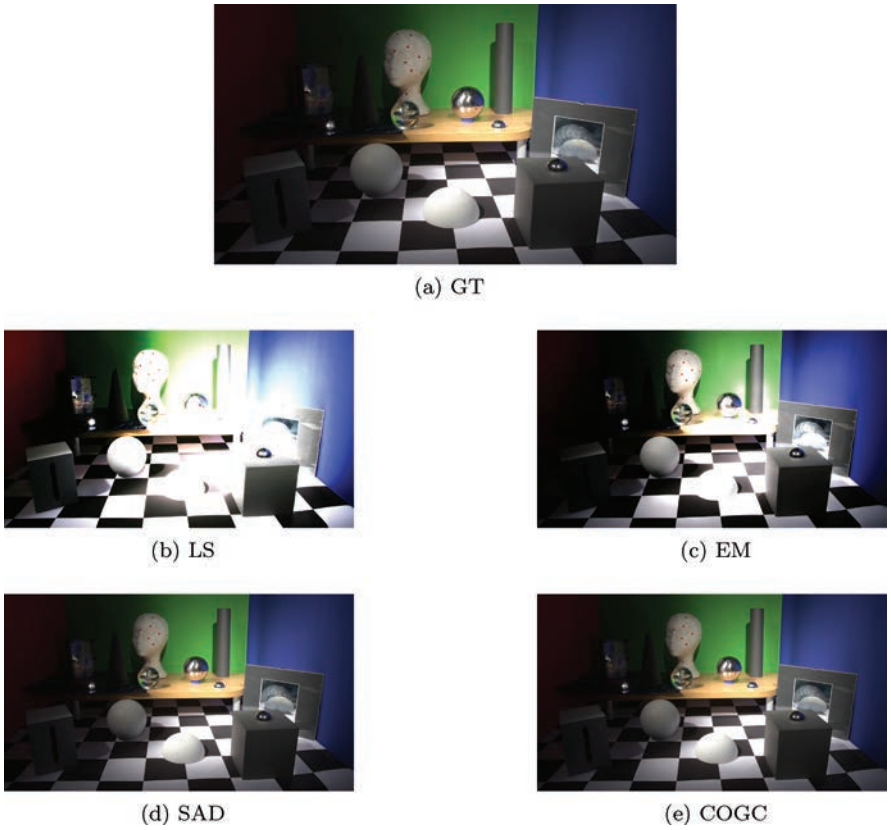


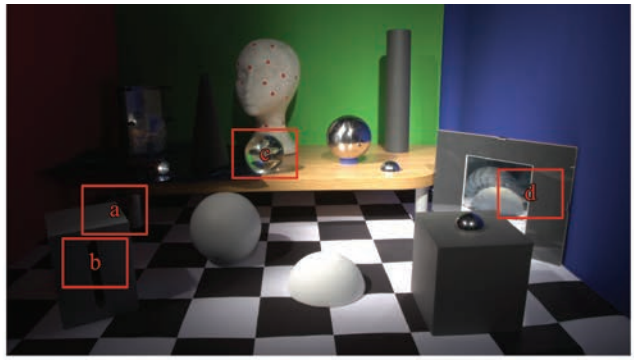
Fig. 5. Reconstructed image from the SHDR pair for all methods for Scene 2, shown at the same single exposure level.

the aggregated result. This implies that, for all the scenes, SAD is statistically indistinguishable from the ground truth and could, in theory, be used instead; hence reducing the cost of SHDR capture.

The COGC method was in the first group for four scenes, but in Scene 4 it was ranked third and was not in the same group as the GT. This result was reflected in RMSEL as well. The aggregated result was affected and COGC was regarded as perceptually different to GT, but considered similar to the SAD method. Although this might seem counterintuitive, it means that any perceptual differences between GT and SAD were too small to be detected, as were the differences between SAD and COGC. However, the difference between COGC and GT was substantial enough to be detected. Even if COGC did manage to get in the first group overall, usage of the SAD method would be preferred due to the high computational cost of COGC.

As expected, the correspondence maps generated by the SAD method contained more noise compared to COGC. While this resulted in more pixels being transferred from an incorrect position using the SAD method, it did not affect the results. It is important to note that, while COGC creates better disparity maps, SAD will find a pixel with similar RGB values. The values chosen by SAD, although less robust for traditional disparity calculations, may appear less distracting.

Both of the expansion operator-based methods performed less well with considerable differences when compared to the stereo correspondence methods. The LS method was ranked last throughout the experiment, making it the last choice for LDR enhancement. A possible reason for poor performance of



Scene 2 - Single Exposrue

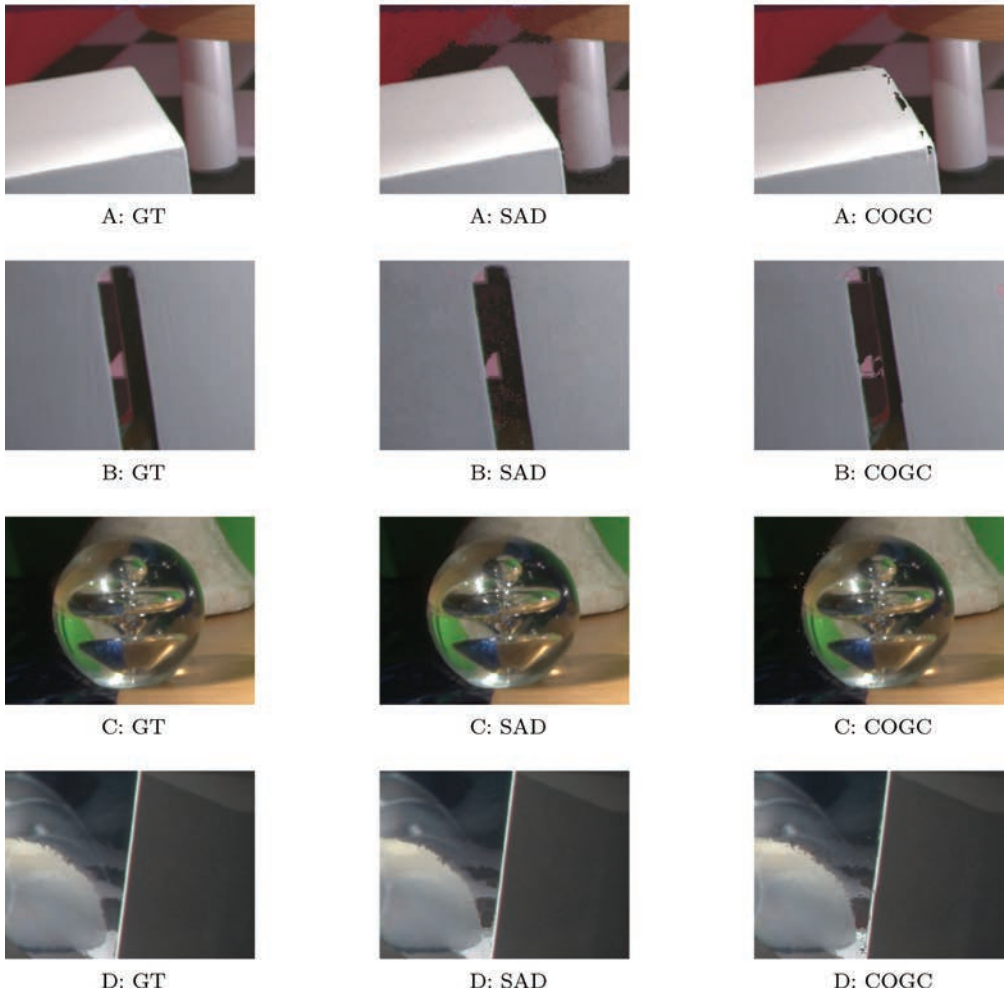


Fig. 6. Detailed insets for the reconstructed SHDR chosen from Scene 2, showing GT, SAD and COGC. All images are shown at a single exposure.

expansion operators might have been expanded brightness, which did not correspond well enough to the actual brightness of the other (HDR) image. In addition, there was no step that would reconstruct the detail that was lost in over: and under-exposed areas of the image.

The RMSEL measure was consistent with the results of the perceptual experiment. For stereo correspondence methods, RMSEL scores coincide with the ranking for all scenes (the method with the lower score was ranked better). In the case of EOs, LS had lower measurement than EM twice. However, in both cases (*Scene 1* and *Scene 5*) EOs were in the same ranking group. This suggests that RMSEL could be used to predict rankings for new scenes and new methods.

6.1 Limitations

The number of scenes tested in the study was limited by the need to maintain a reasonable amount of time for the participants' involvement with the study. The scenes where chosen as a representative sample of those which could be encountered in everyday situations. In addition, some of the scenes had specific properties which we suspected would be challenging for the proposed methods. For example Scene 2 (see Figure 6) included transparencies, object reflections, low-frequency and high frequency regions, specular highlights, under-exposed and over-exposed regions, and Scene 1 included a large over-exposed region. These areas tested the limit of the stereo correspondence approach, as here the matches were not guaranteed, and even if present they could not be considered reliable. However, in practice, algorithms performed well.

For example, in case of transparencies it is uncertain which depth value was going to be assigned (foreground or background object). The correctness of a depth map is not the primary concern for our application, as long as it provides a good correlation for the data to be transferred between the images (Figure 6, inset *C* and *D*). The reliable correlation is achieved using both algorithms. It is inherent for SAD, which looked for the closeness in intensity value, and, for COGC, it was imposed using the data energy minimization. These strong data correlations handle other, already mentioned, challenging areas, including depth of field, relatively well.

Another parameter, which was limited by the experiment's length, was the number of evaluated operators. In this case, a balanced solution was sought by selecting straightforward and advanced operators from each category. We expect that potential future operators may outperform the ones suggested here, and we do indeed encourage such investigations. A local operator worth exploring could be a more advanced one, suggested by Mei et al. [2011]. It expands on, the SAD technique by using additional cost measures, dynamic regions for aggregation, and error correction. An interesting global operator was proposed by Lang et al. [2010]; their hybrid approach finds initial, sparse and robust matches which are then used as a support for optical flow algorithms, ultimately yielding dense correspondences. We suspect that the current methods will not significantly improve over other global methods, as the fundamental nature is not too dissimilar. Future work will investigate such methods and any newer operators in the context presented in this article.

7. CONCLUSION AND FUTURE WORK

The aim of this work was to facilitate the capture and creation of SHDR images, avoiding the need for two rare and expensive HDR cameras. To this end, a technique for capturing an HDR-LDR pair that is enhanced to HDR-HDR (SHDR) was proposed. Four different, methods were proposed and compared using an experimental method of balanced paired comparisons. Two methods were based on expansion operators and the other two on stereo correspondence. For the scenes tested, results showed that most of the methods were statistically different, except the one using SAD. In our study, SAD was not significantly different from the reference image. This indicates that SAD, or similar methods, may be used for enabling SHDR content using an LDR and HDR camera setup. The proposed method makes it

significantly less expensive and easier to enable SHDR than using an HDR-HDR camera pair. To focus on the potential of the method, we have chosen to use only static images for this experiment. However, we also intend to apply stereo-matching methods to video sequences.

REFERENCES

AKYUZ, A., FLEMING, R., RIECKE, B., REINHARD, E., AND BULTHOFF, H. 2007. Do HDR displays support LDR content?: A psychophysical evaluation. In *ACM ACM SIGGRAPH '07*. ACM, New York, 38–44.

BANTERLE, F., ARTUSI, A., DEBATTISTA, K., AND CHALMERS, A. 2011. *Advanced High Dynamic Range Imaging: Theory and Practice*. AK Peters (CRC Press), Natick, MA.

BANTERLE, F., DEBATTISTA, K., ARTUSI, A., PATTANAIK, S., MYSZKOWSKI, K., LEDDA, P., CHALMERS, A., AND BLOJ, M. 2009a. High dynamic range imaging and low dynamic range expansion for generating HDR content. *Comput. Graph. Forum 28*, 8, 2343–2367.

BANTERLE, F., LEDDA, P., DEBATTISTA, K., BLOJ, M., ARTUSI, A., AND CHALMERS, A. 2009b. A psychophysical evaluation of inverse tone mapping techniques. *Comput. Graph. Forum 28*, 1, 13–25.

BANTERLE, F., LEDDA, P., DEBATTISTA, K., AND CHALMERS, A. 2006. Inverse tone mapping. In *Proceedings of the 4th International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia (GRAPHITE '06)*. 349.

BHAT, P., ZITNICK, C., AGARWALA, N., AGRAWALA, M., COHEN, M., CURLESS, B., AND KANG, S. 2007. Using photographs to enhance videos of a static scene. In *Proceedings of the Eurographics Symposium on Rendering*. 327–338.

CHALMERS, A., BONNET, G., BANTERLE, F., DUBLA, P., DEBATTISTA, K., ARTUSI, A., AND MOIR, C. 2009. High-dynamic-range video solution. In *ACM SIGGRAPH ASIA Art Gallery & Emerging Technologies: Adaptation*. ACM, New York, 71.

CRONE, R. 1992. The history of stereoscopy. *Documenta Ophthalmologica 81*, 1, 1–16.

CYGANEK, B. AND SIEBERT, J. P. 2009. *An Introduction to 3D Computer Vision Techniques and Algorithms*. Wiley, Chichester, UK.

DAVID, H. 1988. *The Method of Paired Comparisons* 2nd Ed. Charles Griffn .

DELVIN, K., CHALMERS, A., WILKIE, A., AND PURGATHOFER, W. 2002. Tone reproduction and physically based spectral rendering. In *Proceedings of Eurographics '02: State of the Art Reports*, 101–123.

DODGSON, N. A. 2004. Variation and extrema of human interpupillary distance. In *Proceedings of the SPIE.* vol. 5291, 36–46.

KENDALL, M. AND SMITH, B. 1940. On the method of paired comparisons. *Biometrika 31*, 3/4, 324–345.

KOLMOGOROV, V. 2004. Graph based algorithms for scene reconstruction from two or more views. Ph.D. dissertation, Cornell University, Ithaca, NY.

KOLMOGOROV, V. AND ZABIH, R. 2001. Computing visual correspondence with occlusions using graph cuts. In *Proceedings of the 8th IEEE International Conference on Computer Vision (ICCV '01)*. IEEE, Los Alamitos, CA, 508–515.

LANG, M., HORNUNG, A., WANG, O., POULAKOS, S., SMOLIC, A., AND GROSS, M. 2010. Nonlinear disparity mapping for stereoscopic 3D. In *ACM SIGGRAPH Papers.* ACM, New York.

LEDDA, P., CHALMERS, A., TROSCIANKO, T., AND SEETZEN, H. 2005. Evaluation of tone mapping operators using a high dynamic range display. In *ACM SIGGRAPH Papers*. ACM, New York, 640–648.

LIN, H. AND CHANG, W. 2009. High dynamic range imaging for stereoscopic scene representation. In *Proceedings of the16th IEEE International Conference on Image Processing (ICIP 09)*. IEEE, Los Alamitos, CA, 4305–4308.

LO, C.-H., CHU, C.-H., DEBATTISTA, K., AND CHALMERS, A. 2009. Selective rendering for efficient ray traced stereoscopic images. *Visual Comput. 26*, 2, 97–107.

MEI, X., CUI, C., SUN, X., ZHOU, M., WANG, Q., AND WANG, H. 2011. On building an accurate stereo matching system on graphics hardware. In *Proceedings of the Computer Vision Workshops (ICCV Workshops)*. 467–474.

MENDIBURU, B. 2009. 3D *Movie Making: Stereoscopic Digital Cinema from Script to Screen*. Focal Press.

NAKAYAMA, K. AND SHIMOJO, S. 1990. Da Vinci stereopsis: Depth and subjective occluding contours from unpaired image points. *Vision Res. 30*, 11, 1811–1825.

NAVARRO, F., CASTILLO, S., SERON, F. J., AND GUTIERREZ, D. 2011. Perceptual considerations for motion blur rendering. *ACM Trans. Appl. Percept. 8*, 3, 1–15.

PEAESON, E. AND HAETLET, H. 1976. *Biometrika Tables for Statisticians* 2nd Ed. Biometrika Trust.

POULI, T. AND REINHARD, E. 2011. Progressive color transfer for images of arbitrary dynamic range. *Comput. Graph. 35*, 1, 67–80.

REINHARD, E., HEIDRICH, W., PATTANAIK, S., DEBEVEC, P., WARD, G., AND MYSZKOWSKI, K. 2010. *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. Morgan Kaufmann.

REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. 2002. Photographic tone reproduction for digital images. *ACM Trans. Graph. 21*, 3, 267–276.

RUBINSTEIN, M., GUTIERREZ, D., SORKINE, O., AND SHAMIR, A. 2010. A comparative study of image retargeting. *ACM Trans. Graph. 29*, 6, 1.

RUPPERTSBERG, A., BLOJ, M., BANTERLE, F., AND CHALMERS, A. 2007. Displaying colourimetrically calibrated images on a high dynamic range display. *J. Visual Commun. Image Represent. 18*, 5, 429–438.

SAWHNEY, H., GUO, Y., HANNA, K., KUMAR, R., ADKINS, S., AND ZHOU, S. 2001. Hybrid stereo camera: an IBR approach for synthesis of very high resolution stereoscopic image sequences. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*. ACM, New York, 451–460.

SCHARSTEIN, D., SZELISKI, R., AND ZABIH, R. 2001. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In *Proceedings of the IEEE Workshop on Stereo and Multi-Baseline Vision* 1. IEEE, Los Alamitos, CA, 131–140.

SEETZEN, H., HEIDRICH, W., STUERZLINGER, W., WARD, G., WHITEHEAD, L., TRENTACOSTE, M., GHOSH, A., AND VOROZCOVS, A. 2004. High dynamic range display systems. *ACM Trans. Graph. 23*. 760–768.

SELMANOVIĆ, E., DEBATTISTA, K., BASHFORD-ROGERS, T., AND CHALMERS, A. 2012. Backwards compatible JPEG stereoscopic high dynamic range imaging. In *Theory and Practice of Computer Graphics*.

SERVOS, P., GOODALE, M., AND JAKOBSON, L. 1992. The role of binocular vision in prehension: A kinematic analysis. *Vision Res. 32*, 8, 1513–1521.

SHIRLEY, P. AND MARSCHNER, S. 2009. *Fundamentals of Computer Graphics.* AK Peters, Ltd.

TOCCI, M. D., KISER, C., TOCCI, N., AND SEN, P. 2011. A versatile HDR video production system. *ACM Trans. Graph.* 30, 4.

WARD, G. 2006. JPEG-HDR: A backwards-compatible, high dynamic range extension to JPEG. In *ACM SIGGRAPH Courses*, 8.

WARE, C. AND MITCHELL, P. 2008. Visualizing graphs in three dimensions. *ACM Trans. Applied Percept. 5,* 1, 1–15.

WILBURN, B., JOSHI, N., VAISH, V., EMILIO, E.-V. T., BARTH, A., ADAMS, A., HOROWITZ, M., AND LEVOY, M. 2005. High *performance imaging using large camera arrays*. *ACM Trans. Graph. 24*, 3, 765–776.

Received March 2012; revised August 2012; accepted September 2012