

Binocular Tone Mapping

Xuan Yang^{1,2}

Linling Zhang¹

Tien-Tsin Wong^{1,2}

Pheng-Ann Heng¹

¹ The Chinese University of Hong Kong* ² CUHK Shenzhen Research Institute

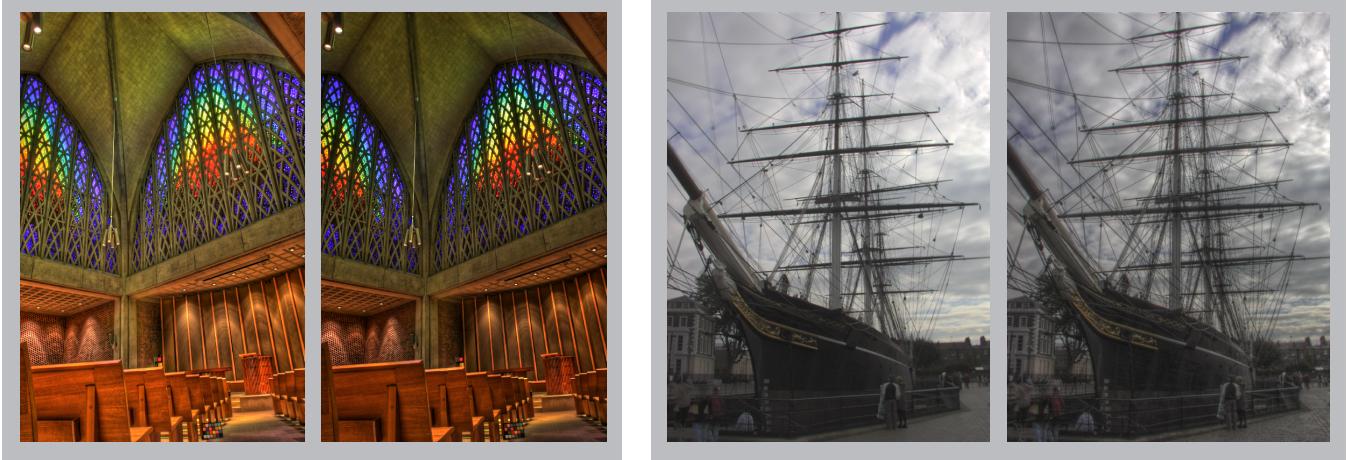


Figure 1: By utilizing the binocular single vision phenomenon, we can generate binocularly tone-mapped image pair that can collectively present more visual richness to human observers than just a single tone-mapped image. LDR image pairs can be better inspected on computer screen.

Abstract

By extending from monocular displays to binocular displays, one additional image domain is introduced. Existing binocular display systems only utilize this additional image domain for stereopsis. Our human vision is not only able to fuse two displaced images, but also two images with difference in detail, contrast and luminance, up to a certain limit. This phenomenon is known as *binocular single vision*. Humans can perceive more visual content via binocular fusion than just a linear blending of two views. In this paper, we make a first attempt in computer graphics to utilize this human vision phenomenon, and propose a binocular tone mapping framework. The proposed framework generates a binocular low-dynamic range (LDR) image pair that preserves more human-perceivable visual content than a single LDR image using the additional image domain. Given a tone-mapped LDR image (left, without loss of generality), our framework optimally synthesizes its counterpart (right) in the image pair from the same source HDR image. The two LDR images are different, so that they can aggregate present *more human-perceivable visual richness* than a single arbitrary LDR image, *without triggering visual discomfort*. To achieve this goal, a novel *binocular viewing comfort predictor* (BVCP) is also proposed to prevent such visual discomfort. The design of BVCP is based on the findings in vision science. Through our user studies, we demonstrate the increase of human-perceivable visual richness and the effectiveness of the proposed BVCP in conservatively predicting the visual discomfort threshold of human observers.

*{xyang, llzhang, ttwong, pheng}@cse.cuhk.edu.hk

Keywords: Binocular vision, singleness of vision, HDR, tone mapping, psychophysical, visual perception

1 Introduction

The popularity of 3D movies leads to the wide availability of low-cost binocular display devices. While the dual display domains (one for the left eye and the other for the right eye) double the space of visualization, existing binocular displays only serve for one type of binocular vision, stereopsis. Another commonly experienced binocular vision phenomenon in our daily life is *binocular single vision* (or *singleness of vision*), i.e. images from two eyes are fused and perceived as a single percept, even though these two images are different (Fig. 2) [Howard and Rogers 2002]. Such image fusion is *not* a simple blending, but a *complex non-linear neurophysiological* process [MacMillan et al. 2007]. The first two rows of Fig. 2 illustrate the difference between fusion (the third column) and the linear blending (the fourth column). In addition, it tends to retain higher contrast, sharply focused, and brighter content from either view during the single percept formation [Steinman et al. 2000]. In other words, via the dual display, it is feasible to present more human-perceivable visual content than any single image, as our vision can naturally combine two images without being aware of the difference between two images.

Unlike binocular display, high-dynamic range (HDR) display is less accessible to the general public. Even though tone mapping can be adopted to present the HDR content on a low-dynamic range (LDR) display, there is a tension between displaying large-scale contrast and fine-scale details. Striking a good balance is always challenging. In this paper, we explore the possibility to utilize existing LDR binocular display to simultaneously present the contrast and details in HDR images by proposing a *binocular tone mapping* framework. Given an HDR image and its tone-mapped LDR image (left view, without loss of generality) generated by one of the existing tone mapping techniques, our framework optimally syn-

thesizes its counterpart (right) LDR image in the binocular image pair. So that, through the phenomenon of binocular single vision, the two LDR images aggregate more human-perceivable visual content than any arbitrary single LDR image. For example, in Fig. 1, one image within each pair presents more global contrast while the other presents more local details. We demonstrate the effectiveness of increasing visual richness via a user study.

Obviously, we want two views to be as different as possible in order to retain more visual content from the source HDR. However, there is a limit on the difference between two views. When such limit is exceeded, binocular viewing discomfort appears, and even worse, binocular single vision may fail. Such viewing discomfort [Lambooij et al. 2009] is an important health issue receiving much attention, due to the wide availability of 3D displays. To guarantee a stable formation of binocular single vision and avoid viewing discomfort, we propose a novel metric called *binocular viewing comfort predictor* (BVCP) to guide the binocular tone mapping. Instead of studying the cause of visual discomfort, we derive our BVCP based on the findings in vision science and the experimental results in existing literature. As the first attempt to predict visual discomfort, we intentionally design the BVCP metric in a conservative manner, in order to avoid visual discomfort for most individuals. Our optimization-based framework generates the right tone-mapped image with the goal of maximizing the overall visual information content, under the guidance of BVCP. We demonstrate the validity of such metric by a user study. Our contributions can be summarized as follow:

- To our best knowledge, this is the first attempt in the graphics area to enrich the visual experience with binocular single vision.
- The design of a novel metric to measure the viewing discomfort due to binocular content difference.
- The utilization of an optimization-based binocular tone mapping framework to produce LDR image pairs that preserve more human-perceivable visual content than arbitrary single image, while simultaneously avoiding viewing discomfort.

2 Related Work

Binocular Single Vision Our visual system has the ability to combine different images from our two eyes into a single vision [von Helmholtz 1962]. Such phenomenon is a separate and independent process from stereopsis [O’Shea 1983]. While binocular single vision occurs only in a small volume of retinal area around where our eyes are fixating, stereopsis occurs at places even where our eyes are not fixating (when images of object appear double). It was discovered that such fusion process is a non-linear combination of luminance, contrast, and color. To prove this, MacMillan et al. [2007] measured the interocular brightness response using asymmetrical neutral density filters and Baker et al. [2007] measured the interocular contrast response using sine-wave gratings.

This non-linear fusion is a complicated neurophysiological procedure and is generally regarded as a combination of *binocular fusion* and *suppression* [Ono et al. 1977; Steinman et al. 2000]. Binocular fusion is a process of superimposing and combining similar content from the two views into one unified and stable percept, which happens when the two views are similar or identical (Fig. 2, upper row). Binocular suppression occurs when one view (submissive) is blank, significantly dimmer, much more blurry, or has significantly less contrast than the other (dominant). In this case, a single percept is formed in our vision system by smartly turning off all or part of the submissive view (Fig. 2, middle row). However, when the two views are too different (e.g. Fig. 2, bottom row), an undesir-

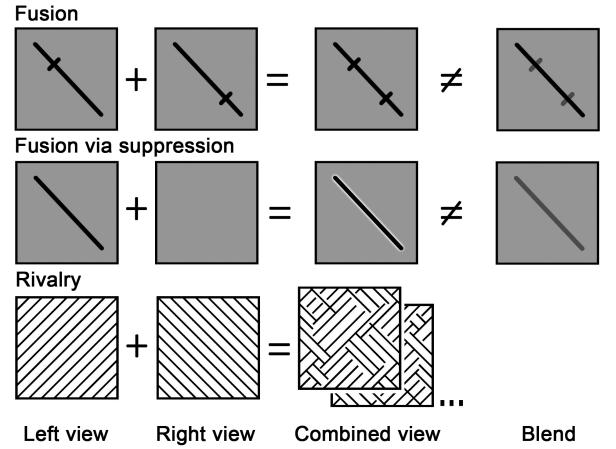


Figure 2: Fusion, suppression and rivalry. (Top row) Binocular fusion occurs when the two dichoptic images are similar; (middle row) binocular suppression happens when there is a submissive view; (bottom row) binocular rivalry, continuous alternation between the two views, occurs when the difference between two views is too large.

able phenomenon, *binocular rivalry*, occurs. In this case, the result is a non-converging percept composed of continuously alternating “patches” from the two views [Lei and Schor 1994], as both stimuli are too strong and none of them can suppress the other. Obviously, such continuous alternation can be noticed by viewers and cause viewing discomfort. Besides binocular rivalry, sometimes binocular suppression may also lead to visual discomfort when the stimulus is too strong. A halo or drifting can be observed as a result of inhibitory effect at the center-surround receptive fields excited by the contour [Lei and Schor 1994].

The above discomforts can greatly destroy the visual experience. Hence, we need an assessment for the binocular viewing comfort. This assessment sounds like an image similarity metric. There are several existing metrics, including mean squared error (MSE), structural similarity (SSIM) [Wang et al. 2004], perception-oriented metrics Visible Difference Predictor (VDP) [Daly 1993] and its extension High Dynamic Range Visible Difference Predictor (HDR-VDP, HDR-VDP-2) [Mantiuk et al. 2005; Mantiuk et al. 2011]. Note that existing metrics consider the visible difference between two images when the observers look at these images with *both of their eyes*. However, these existing metrics are not considering the binocular vision in which the left and right eyes of observers are presented with two different images. An obvious shortcoming of existing metrics can be illustrated by binocular suppression (Fig. 2, middle row) where two images are obviously different using any existing metric, even though a stable percept can be formed. Hence, none of the existing metrics can be applied. In this paper, we design a brand new metric, named Binocular Viewing Comfort Predictor (BVCP), based on theories and experimental results of vision science, to suit our purpose.

Tone Mapping Several sophisticated tone mapping techniques have been proposed to generate LDR images from HDR images. Reinhard [2006] provided a comprehensive survey on tone mapping techniques, ranging from sigmoidal compression to image appearance model, and to perception and engineering-based methods. Tone mapping methods can be roughly classified into global and local operators. Histogram adjustment methods and adaptive logarithmic mapping such as [Larson et al. 1997; Drago et al. 2003] are two main categories of global operators. On the other hand, there are also several prevalent local operators, such as bilateral fil-

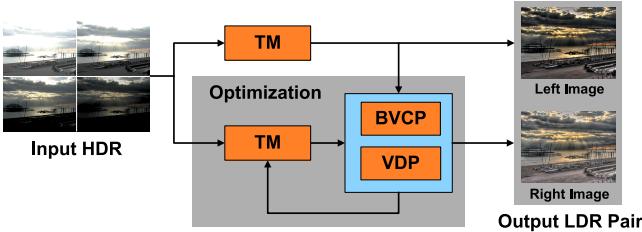


Figure 3: System overview.

tering approach [Durand and Dorsey 2002], gradient domain optimization [Fattal et al. 2002] and perceptual-based contrast processing [Mantiuk et al. 2006].

Instead of proposing a brand new tone mapping operator, in this paper we propose a binocular tone mapping framework that utilizes existing tone mapping operators as our building block to generate two LDR images that optimally increase the human-perceivable visual content without triggering discomfort. In particular, our framework has been evaluated with four state-of-the-art tone mapping methods, including bilateral filtering approach [Durand and Dorsey 2002], gradient domain HDR compression [Fattal et al. 2002], adaptive logarithmic mapping [Drago et al. 2003] and perceptual-based contrast processing [Mantiuk et al. 2006].

3 Overview

An overview of our binocular tone mapping framework is illustrated in Fig. 3. The input is an HDR image. A LDR image is first tone-mapped from this source HDR image using a selected tone mapping operator (the upper path in Fig. 3). The parameter(s) for generating the LDR image can be manually specified or automatically suggested by the operator (default parameters). Currently, our system supports four state-of-the-art tone mapping operators as described previously. Other tone mapping operators may also be adopted.

Without loss of generality, we refer to this LDR image as the left image. Our goal is to generate the optimal right LDR image using the same tone mapping operator as the left one, by maximizing the visual difference between two views and avoiding any visual discomfort. The optimization framework generates the optimal right view by iteratively adjusting the tone mapping parameters in a gradient ascent fashion. The iteration continues until the objective value converges (the lower path in Fig. 3).

Our objective function composes of two metrics, *visible difference predictor* (VDP) [Daly 1993] and a novel *binocular viewing comfort predictor* (BVCP) (Section 4), corresponding to the total amount of visual content and the viewing comfortability, respectively. The BVCP is put as a hard constraint such that any LDR image pair leading to discomfort is rejected. During each iteration, we adjust the tone mapping parameters to follow the gradient ascent direction, based on the VDP of the current LDR image pair. If the LDR image pair cannot pass the BVCP, we then reduce the step size of gradient ascent.

Since the gradient ascent approach is sensitive to initial values, it can be easily trapped by local optimum. To raise the chance of finding global optimum, we distribute n random seeds in the search space and search for n paths accordingly. The final output is the best among the n pairs. In our current experiments, $n=10$.

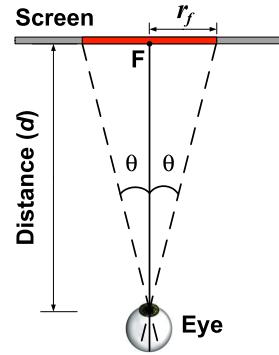


Figure 4: Fusional area.

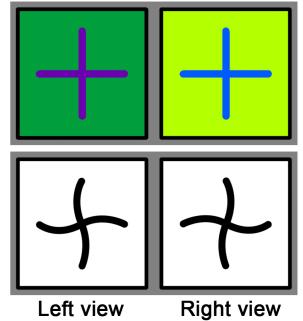


Figure 5: Contour dominance.

4 Binocular Viewing Comfort Predictor

While tone mapping limits the visual content for display, binocular tone mapping offers one additional domain for visualization. An obvious way to exploit the dual image domain is to ensure the two images span different dynamic ranges, in order to maximize the total amount of visual content. However, there is a limit on the difference between the two images. When such limit is exceeded, rivalry and high level suppression appear and lead to visual displeasures, such as flicker, nausea and fatigue.

To avoid such viewing discomfort, we introduce the novel BVCP metric, based on current psychophysical studies [Levelt 1965; Liu et al. 1992; Steinman et al. 2000; Kooi and Toet 2004]. In particular, we measure the difference between the left and right images, in terms of *contour density, contrast, luminance, and color*. The limit of contour density can be relaxed when the contrast is small. This phenomenon is called failure of rivalry, found by Liu et al. [1992]. Such relaxation allows more visual information to be represented. Hence, we shall utilize this phenomenon in our BVCP design.

Fusional Area Whether two corresponding points in the left (\tilde{L}) and right (\tilde{R}) images can be fused into a single percept or not requires a complex consideration. The decision is not solely based on the local colors of the two points, but based on the visual agreement of neighborhoods surrounding the two points. This neighborhood is called *Panum's fusional area* in literatures of vision science. Fusional area is an area on our retina. In graphics terminology, fusional area occupies a constant solid angle subtended at our eye. When it is stimulated together with a given single retinal point in the other eye, we form a single binocular percept [Steinman et al. 2000]. When both of viewers' eyes fixate at a pixel with position (i, j) in both left and right images, the whole fusional areas (both left and right) surrounding the position (i, j) have to be considered for fusion stability, in terms of contour, contrast, luminance and color differences.

By projecting the fusional area to the screen via our pupil (Fig. 4), we can compute the size of projected fusional area in terms of pixel units. For simplicity, when we say fusional area, we mean the projected fusional area, from now on. Suppose the viewer is sitting in front of the screen at distance d . We assume all pixels are projected onto the screen equally and have a square shape (aspect ratio = 1). The pixel density on the screen is γ pixels per inch (PPI). Fig. 4 explains the notations. The radius of the fusional area in pixel unit can be computed as,

$$r_f = \gamma \tan \theta \cdot d, \quad (1)$$

where θ is the maximal retinal disparity, which is around 60 to 70 arcmin [Wopking 1995] for most people. Hence, the fusional area ζ in the image is a circular neighborhood of radius r_f . To simplify

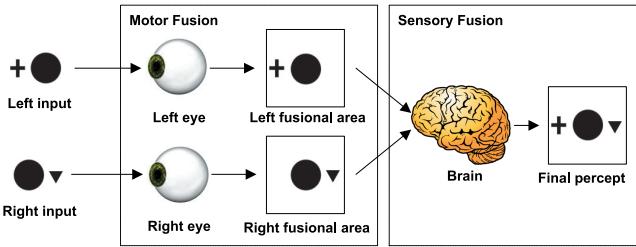


Figure 6: Motor fusion and sensory fusion.

our computation, we approximate the neighborhood by a rectangle of $2r_f \times 2r_f$ pixels instead. Note that r_f is a function of viewing distance d , which means the viewing distance affects the viewing comfort.

As viewers' eyes may fixate at an arbitrary pixel position in the image pair, we need to go over all pixel pairs from the LDR image pair and consider their corresponding fusional areas in order to measure the fusion stability. From now on, all following assessments consider the fusional area ζ at an arbitrary position (i, j) . The corresponding fusional areas in the left and right images are denoted as L and R .

Contour Fusion It has been found that the *contour* (edge) difference is more important than the contrast or color differences in determining binocular fusion [Treisman 1962]. Fig. 5 illustrates this idea. Although the images in the upper image pair are different in terms of color and contrast, they can still be fused because their contours are very similar. On the other hand, even the color and contrast are very similar, the lower image pair cannot be fused (i.e. rivalry) as their contours are noticeably different. Note that, when two areas have similar contours, only extremely large color difference or contrast inversion may result in rivalry (to be explained later).

Contour fusion does not require contours in the two corresponding fusional areas L and R to be exactly the same, because our eyes have two mechanisms, *motor fusion* and *sensory fusion*. Motor fusion superimposes corresponding points or similar contour by the movement of our eyes. After the alignment by motor fusion, a neurophysiological process, sensory fusion, combines the two views into one unified percept. As illustrated in Fig. 6, left and right views are first aligned to superimpose the ‘disk’ by motor fusion, then the ‘cross’ and ‘triangle’ are both fused into the final percept by sensory fusion. So the precondition of sensory fusion is that part of the contour can successfully trigger motor fusion. We need to evaluate the percentage of contour that can be aligned by motor fusion as a guidance of whether a stable fusion can be formed. Blake and Boothroyd [1985] demonstrated that areas containing 50% matched and 50% unmatched contour segments can still be successfully fused.

Contour (edge) has different definitions in different domains. In our paper, a meaningful contour segment is defined as obvious color difference expanding to, or beyond a specified visual degree. To figure out as many contour segments as we can, a scale space representation is applied to the fusional area. We construct a pyramid from the original fusional area. First, we Fourier transform this fusional area to the frequency domain. Then we apply a pyramid of low-pass filters in this frequency domain (Fig. 7, upper row). By inversely Fourier transforming each low-passed frequency images, we obtain a pyramid of low-passed fusional areas (Fig. 7, lower row). The low-pass filter we selected is called mesa filter [Watson 1987; Daly 1993]. It can be approximately regarded as a kind of truncation at half-amplitude frequency. Such frequency property

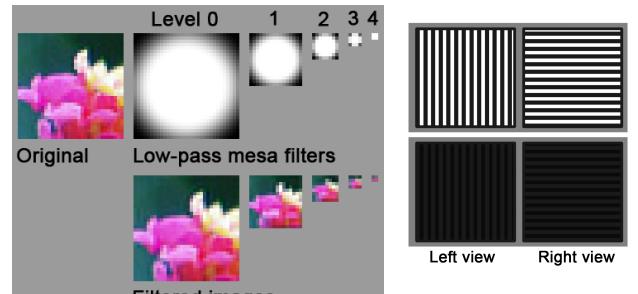


Figure 7: Mesa pyramid.

Figure 8: Failure of Rivalry.

confirms with our requirement in determining the *failure of rivalry*, which will be discussed later. This is the rationale why we selected the mesa filter. Suppose the pyramid has K levels in total, the radius of the level-0 kernel is r_f pixels while that of level- $(K - 1)$ kernel is 1 pixel.

We then define the contour based on the concept of visual acuity (VA). A well-designed diameter of letter ‘C’ or height of letter ‘E’ in Landolt C and Tumbling E Chart, is 5 times of the visual acuity. Hence, we define contour segment to be meaningful when its occupied visual degree reaches or exceeds 5-VA (a typical human eye has the visual acuity to separate 1 – 2 arcmin). Thus the lowest level of mesa pyramid involved in our computation should be level S with a width of less than or equal to $(\tan \theta / \tan (VA/24))$. With such mesa pyramid, obvious color difference between two adjacent pixels in each level is regarded as a segment of visible contour. For each pair of fusional areas L and R , two mesa pyramids are set up respectively. The k -th level of the pyramids are denoted as L^k and R^k , where $k \in [S, K - 1]$.

To identify a contour, we measure the color difference ΔE_c between the adjacent pixels. It is defined as a 2-norm distance of their colors in LAB color space. Consider Fig. 9, the color difference between a pair of pixels (red and blue pixels) with positions, p_1 and p_2 , in the left image is

$$\Delta L^k(p_1, p_2) = \Delta E_c(L^k(p_1), L^k(p_2)). \quad (2)$$

Similarly, we can obtain $\Delta R^k(p_1, p_2)$ for the right image. Now, we can predict their fusion state $S^k(p_1, p_2)$ (i.e. whether a contour is recognized by viewers) by looking up a decision table (Table 1), with $\Delta L^k(p_1, p_2)$ and $\Delta R^k(p_1, p_2)$ as query. In this table, *JND* stands for a constant called *just noticeable color difference* and *OCD* stands for another constant called *obvious color difference*. According to existing studies [Chen and Wang 2004; Lin and Jane 2009], we set *JND*=2.3 and *OCD*=6.0. If both $\Delta L^k(p_1, p_2)$ and $\Delta R^k(p_1, p_2)$ are less than *OCD*, no contour is recognized. So, the fusion state is set to be 0 (stands for ‘no contour’). If both of them reach *OCD*, two obvious contour segments are recognized and fused together. Hence the fusion state is set to be 1 (stands for ‘match’). Confusion appears only when one of the color differences

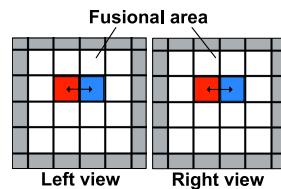


Table 1: Decision table for contour fusion where $J=JND$, $O=OCD$.

reaches OCD, while the other falls below JND, in that case the fusion state is set to be -1 (stands for ‘not match’). If one reaches OCD and the other falls between JND and OCD, it can still be regarded as support for existence of contour. Hence the state is 1 too.

The k -th level of two corresponding fusional areas (L^k and R^k) are regarded as fusible if the count of ‘1’ (match) is equal to or larger than the count of ‘-1’ (not match). On the other hand, if the count of ‘1’ is smaller than that of ‘-1’, L^k and R^k are not fusible. If both L^k and R^k contain no contour, contour fusion takes no effect, and the final fusion state is determined by other levels. Here, we record the contour fusion state of two fusional areas at the k -th level in the pyramid with a state variable B_{cf}^k as follows (0 stands for ‘no contour’, 1 stands for ‘fused’, and -1 stands for ‘not fused’),

$$B_{cf}^k = \begin{cases} 0, & \text{if } \sum_{(p_1, p_2) \in \zeta} |S^k(p_1, p_2)| = 0 \\ 1, & \text{else if } \sum_{(p_1, p_2) \in \zeta} (S^k(p_1, p_2)) \geq 0 \\ -1, & \text{otherwise} \end{cases}. \quad (3)$$

Note that the contour fusion state at higher levels override the lower ones, so

$$B_{cf}^{k-1} = B_{cf}^k \quad \text{if } B_{cf}^k \neq 0, \quad (4)$$

The final fusion state of two fusional areas is,

$$B_{cf} = B_{cf}^S. \quad (5)$$

Contour and Regional Contrasts The differences in luminance can be generally referred to as contrast. Two types of contrast can influence the binocular single vision, they are contour contrast and regional contrast. Contour contrast coexists with contour if it can be detected by human eye. Matched contour pair generally helps the fusion except when their contrasts are obviously inverted. So, we revise $S^k(p_1, p_2)$ before evaluating Eq.3 as follows. When a pair of matched contour has obviously inverted contrast, their fusion state changes to -1 (‘not match’). The revised function is:

$$\begin{aligned} S^k(p_1, p_2) = -1, & \quad \text{if } |C(L^k(p_1), L^k(p_2))| > OCD \\ & \quad \text{and } |C(R^k(p_1), R^k(p_2))| > OCD \\ \text{and } C(L^k(p_1), L^k(p_2)) \cdot C(R^k(p_1), R^k(p_2)) < 0, & \end{aligned} \quad (6)$$

where $C(c_1, c_2)$ computes the lightness difference between the pixel pair c_1 and c_2 .

Regional contrast refers to the contrast between two regions (in our case, the two corresponding fusional areas). It has a relatively smaller impact on the viewing comfort (compared to contour fusion), unless the two regions differ too much [Kooi and Toet 2004]. After studying existing literature, we adopt a restrictive constraint for regional contrast, such that the average color difference between two fusional areas L and R must be less than a *distinct color difference* (DCD). According to the existing study [Carter and Huertas 2010], we set $DCD = 34$. The viewing comfort due to the regional contrast (B_{rc}), between two corresponding fusional areas, is assessed by

$$B_{rc} = \begin{cases} 1, & \text{if } \sum_{p \in \zeta} \left(\Delta E_c(L(p), R(p)) \right) \cdot \frac{1}{4r_f^2} < DCD \\ -1, & \text{otherwise} \end{cases}, \quad (7)$$

where $L(p)$ and $R(p)$ are two corresponding pixels located at position p in L and R , respectively.

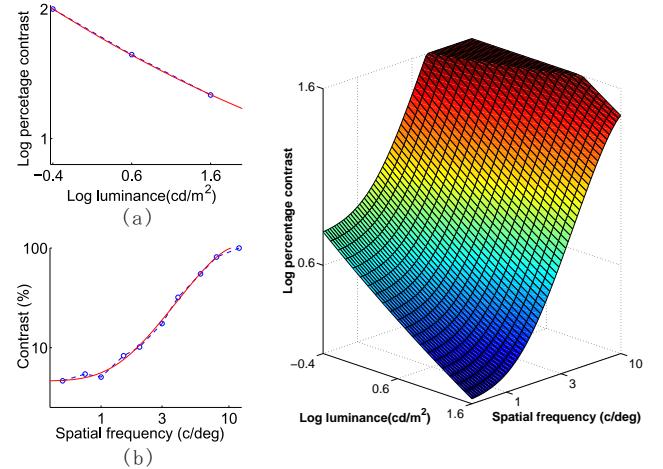


Figure 10: Luminance and freq. vs. contrast.

Figure 11: The constructed log percentage contrast threshold.

Failure of Rivalry Normally when contour fusion fails, rivalry or over-suppression occurs and leads to discomfort. However, it was discovered that the occurrence of rivalry depends also on contrast. When contrast is below a certain threshold, a stable single percept always forms regardless of the contour fusion criteria mentioned above. This phenomenon is called *failure of rivalry* and more likely to happen in low-contrast regions. Fig. 8 illustrates this phenomenon. The upper image pair cannot be fused because of the contour difference. Conversely, the lower image pair has the same contours but with much lower contrast and luminance. In this case, they can still be fused thanks to the failure of rivalry. The contrast threshold is a function of image luminance and spatial frequency [Liu et al. 1992]. It is negatively related to the luminance (Fig. 10(a), originated from Fig. 3 in [Liu et al. 1992]) while positively related to the spatial frequency (Fig. 10(b), originated from Fig. 4 in [Liu et al. 1992]). The contrast is measured in log space as *log percentage contrast* between a pair of pixels, c_1 and c_2 , expressed as,

$$P(c_1, c_2) = \log_{10} \left| \frac{\Gamma(Y_1) - \Gamma(Y_2)}{\Gamma(Y_1) + \Gamma(Y_2)} \right|, \quad (8)$$

where Y_1 and Y_2 are the normalized luma of c_1 and c_2 , respectively. $\Gamma(Y)$ maps the normalized luma Y in $[0, 1]$ to the physical measurement in the unit of cd/m^2 .

Whenever the contrast value is below the contrast threshold, a stable single percept always forms. However, existing literature either measure the contrast threshold against luminance or contrast threshold against spatial frequency. To our best knowledge, no existing method formulates such threshold as a function of both luminance and spatial frequency, as required in our application. So we construct the function based on the existing findings in the literature. We first fit the plots (blue dots) in Fig. 10 to obtain two continuous curves (red curves), expressed as,

$$T(l) = \min (2, 1.999 - 0.362 \log_{10}(l) + 0.026 \log_2^2(l)), \quad (9)$$

$$T(f) = \min (2, \log_{10}(3.557 - 1.334f + 1.881f^2 - 0.108f^3)), \quad (10)$$

where l is average luminance measured in cd/m^2 , f is the spatial frequency measured in cycles per degree, and T represents log percentage contrast threshold. When $l = 40cd/m^2$ and $f = 5cy/deg$, Eq. 9 and 10 fixate at the same percentage contrast threshold. By

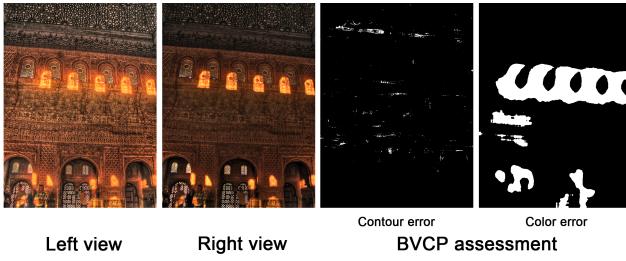


Figure 12: Visualization of BVCP assessment.

assuming the curvature is constant, we extrapolate a function of both luminance and spatial frequency as

$$T(l, f) = \min \left(2, \log_{10} \left(3.557 - 1.334f + 1.881f^2 - 0.108f^3 \right) + 0.514 - 0.362\log_{10}(l) + 0.026\log_{10}^2(l) + \delta \right), \quad (11)$$

where δ is user parameter and is set to a value in $[-0.15, 0.15]$. Its default value is 0. Fig. 11 visualizes Eq. 11 in 3D.

Obviously, including the failure of rivalry in our metric allows us to accept more pairs of LDR images and further maximize the information content of the optimal pair. To incorporate the failure of rivalry into our metric, we first compute the average luminance l_L in the left fusional area L as $\Gamma(\bar{Y}_L)$, where \bar{Y}_L is the average luma in L .

Recall that we Fourier transform the image into frequency domain and perform mesa filtering, the difference between two adjacent pixels in the k -th level of the filtered image pyramid can be roughly regarded as a frequency f_k of $r_f/2^{k+1}$ cy/deg. Thus, the contrast threshold T_L^k for the left fusional area L^k is obtained by feeding l_L and f_k into Eq. 11,

$$T_L^k = T(l_L, f_k). \quad (12)$$

Similarly, we can obtain T_R^k . So their common threshold T^k is $\min(T_L^k, T_R^k)$. When the contrast is below this threshold T^k , stable fusion always forms and hence $S^k(p_1, p_2)$ can never be negative ('not match'). Thus, we take a second revision to the fusion state variable $S^k(p_1, p_2)$, as follow

$$\begin{aligned} S^k(p_1, p_2) &= 0, & \text{if } S^k(p_1, p_2) < 0 \\ &\text{and } P(L^k(p_1), L^k(p_2)) \leq T^k \\ &\text{and } P(R^k(p_1), R^k(p_2)) \leq T^k. \end{aligned} \quad (13)$$

The Overall Fusion Predictor With the contour fusion predictor B_{cf} and the regional contrast predictor B_{rc} defined above, we can now predict the ultimate viewing comfort for an arbitrary pair of fusional areas by

$$B = \begin{cases} -1, & \text{if } B_{rc} = -1 \\ B_{cf}, & \text{otherwise} \end{cases}. \quad (14)$$

When $B = -1$ ('not fused'), viewing discomfort exists. Fig. 12 shows a pair of LDR images and their BVCP assessment result according to a viewing environment described in Section 6. Those white pixels indicate the areas triggering viewing discomfort when our eyes fixate at these locations.

So far, we have only discussed the BVCP test of two fusional areas. An image pair passes the BVCP test only when all pixels pass the BVCP test, i.e. no pixel in the image triggers visual discomfort ($B = -1$). Obviously, this is a conservative design, as human observers may be able to tolerate certain amount of pixels violating the BVCP.

5 Optimization

In general, the more different the left and right images are, the more visual content they can preserve aggregate. To measure the visual difference, we adopt the *visible difference predictor* (VDP) [Daly 1993], that has been utilized in various applications [Myszkowski 1998]. Given a pair of images, the output of VDP is a probability of detecting visible difference, $V(i, j)$, at each pixel location (i, j) . The overall binocular visual difference E is defined by

$$E = \frac{1}{\Omega} \sum_{i,j} \left(H[V(i, j) - \tau] \right), \quad (15)$$

where τ is a user-defined probability threshold and generally set as 75%. H is the Heaviside step function. Ω is the total number of pixels in the image.

With the VDP and the proposed BVCP, our optimization framework maximizes E without violating the BVCP (Fig. 3). Throughout the optimization, the left image remains unchanged. Only the right image is iteratively generated by adjusting the tone mapping parameter(s) in a gradient ascent fashion. The pair of the left image and the generated right image in the current iteration is evaluated with the BVCP test. If the pair fails the BVCP test, the current right image is rejected and the step size is reduced to generate a new one. Such process is repeated until the BVCP test is finally passed or the step size drops to zero. Fig. 13 plots E against the parameters of four tone mapping operators for an HDR example in Fig. 20. Since the upper two tone mapping operators have only a single parameter, their plots are 2D. On the other hand, the lower two tone mapping operators have two parameters, their plots are 3D. The red region in each plot indicates that the corresponding tone-mapped right image will trigger visual discomfort. They are predicted by the BVCP. The green dot corresponds to the tone mapping parameter(s) of the left view.

6 Results and Discussions

The proposed framework is independent of the adopted tone mapping operators. In our current implementation, we support four state-of-the-art tone mapping operators including bilateral filtering approach [Durand and Dorsey 2002], gradient domain HDR compression [Fattal et al. 2002], adaptive logarithmic mapping [Drago

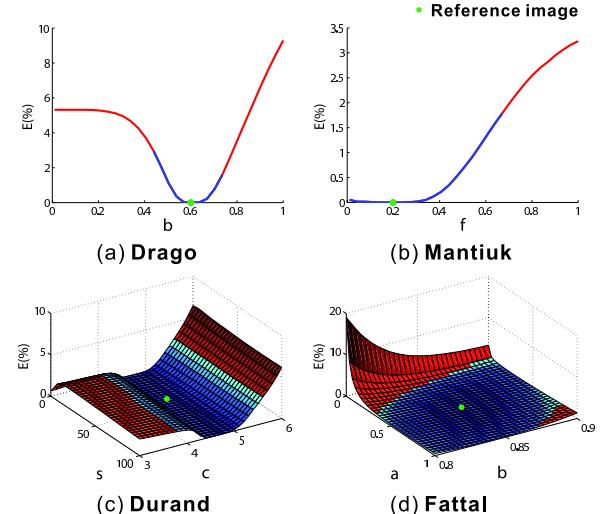


Figure 13: Visualization of binocular visual differences of four tone mapping operators for the HDR example in Fig. 20.

et al. 2003] and perceptual-based contrast processing [Mantiuk et al. 2006]. Both global and local tone mappings are included in these operators. The bilateral filtering approach [Durand and Dorsey 2002] is a two-scale decomposition of the image into a base layer, encoding large-scale variations, and a detail layer. Gradient domain HDR compression [Fattal et al. 2002] manipulates the gradient field of the luminance image by attenuating the magnitudes of large gradients. Adaptive logarithmic mapping [Drago et al. 2003] displays high-contrast scenes by logarithmic compressing the luminance values and imitating the human response to light. Perceptual-based contrast processing [Mantiuk et al. 2006] enhances or compresses the contrast in a visual response space, in which the contrast values directly correlate with their visibility in an image.

To demonstrate the effectiveness of the framework, we have experimented with all four operators on a rich variety of test images. Fig. 14 shows four LDR pairs mapped from the same source HDR, but with different tone mapping operators. In each pair, the left image is tone-mapped with user-specified tone mapping parameters. On the other hand, the right one is automatically optimized by our framework to maximize the overall visual content without triggering visual discomfort. Note that the left and right images can be interchanged without significantly affecting the visual experience. The perceived binocular single percept generally preserves more details (especially in Fig. 14(a)&(b)) and better contrast (especially in Fig. 14(c)&(d)) than a single tone-mapped image. More results can be found in the supplementary materials.

User Study To validate our method, we conducted two user studies. In our experiment set-up, we use a Zalman ZM-M215W 21.5" 3D LCD display with the highest luminance of around $300\text{cd}/\text{m}^2$ for displaying binocular images. The pixel density of the display is 102 PPI. The viewer is asked to sit at a distance of 0.5 meters from the display and to wear a pair of 3D polarized glasses (with transmittance of around 44%). All experiments are conducted indoor with an ambient illumination of around 200 lux. Detailed statistics of the user studies can be found in the supplementary materials.

Visual Richness The first user study evaluates the effectiveness of our binocular tone mapping in terms of visual richness. It compares bioptic image pairs (both views are identical) to dichoptic image pairs (the two views are different). To fairly compare, the image of bioptic pair is generated using the “best” parameter values, instead of the default parameter values which may not be satisfactory for certain HDR input. The “best” parameters are determined as follow. For each HDR input, we first randomly generate 10 LDR images with 10 sets of parameter values in the recommended parameter range of the particular tone mapping operator. Then, 8 participants are asked to choose the best (in terms of details and contrast) LDR



Figure 14: Optimal LDR image pairs generated by our framework using four tone mapping operators.

	Mean	Standard deviation	95% confidence interval	
			Lower Bound	Upper Bound
Drago	0.804	0.140	0.706	0.900
Durand	0.717	0.108	0.643	0.793
Fattal	0.753	0.101	0.683	0.823
Mantiuk	0.721	0.277	0.529	0.913

Table 2: User study of visual richness.

image among them. The top-rated one is then referred to as the image of bioptic pair in the user study. The same image is also referred to as one view of the dichoptic image pair, while the other view is optimally determined by our framework. Fig.19 to 22 show four such image pairs used in our user study.

All four tone mapping operators (Durand, Fattal, Drago and Mantiuk) and eight different HDR images (32 sets of image pairs in total) are experimented in the user study. Thirty-five participants are invited to evaluate these 32 sets of randomly displayed image pairs. In each round, a dichoptic image pair and the corresponding bioptic image pair are chosen for comparison. Note that the bioptic image pair is presented to the participants via the same 3D glasses. These two image pairs are shown in random order, i.e., the participants do not know which one is the dichoptic image pair. Each participant is then asked to select the one he/she prefers. We briefly explain to them that he/she may consider the visual richness and/or visual content clarity during selection. However, it is up to the participants’ decision in selecting the preferred image pair. To allow the participant to better inspect the image pairs, he/she can press a button to toggle between these two image pairs during the selection. Once the participant makes the decision, he/she can press a ‘select’ button.

Table 2 shows the statistics for four tone mapping operators. It is clear that most participants prefer our binocularly tone-mapped results (80.4% for Drago, 71.7% for Durand, 75.3% for Fattal and 72.1% for Mantiuk). It also shows that our binocular tone mapping can effectively preserve more visual richness than a single one, especially for the case with Drago’s operator. One possible explanation is that global operators, like Drago’s, are generally more constrained than the local ones, leading to either more severe loss in details or loss in contrast.

Binocular Symmetry Our second user study evaluates whether the effectiveness of our binocular tone mapping is symmetric to the left and right eyes. We conduct the previous visual richness experiment again on 20 participants with the same data set (four tone-mapping operators for eight different HDR images, 32 sets of image pairs in total). However, this time these 32 sets of image pairs are evaluated twice. In one of these two evaluations, the LDR image generated by our framework is shown to the participants’ left eyes. While in the other evaluation, our generated LDR image is shown to participants’ right eyes. Participants do not know which

		Mean	Standard deviation	95% confidence interval	
				Lower Bound	Upper Bound
Drago	R	0.788	0.169	0.671	0.904
	L	0.825	0.167	0.709	0.941
Durand	R	0.706	0.135	0.613	0.800
	L	0.738	0.095	0.671	0.804
Fattal	R	0.725	0.136	0.631	0.819
	L	0.763	0.109	0.687	0.838
Mantiuk	R	0.731	0.225	0.575	0.887
	L	0.738	0.285	0.540	0.935

Table 3: User study of binocular symmetry. L/R means the left/right eye sees the images generated by our framework.

of their eyes are shown with our generated LDR images.

Table 3 compares the statistics from these two sets of evaluations. It is clear that, no matter which optimal image is shown to which eye, our binocular tone mapping can always effectively preserve more visual richness than a single one. From the statistics, the left and right eyes are slightly asymmetric. This confirms to existing study on ocular dominance [Ehrenstein et al. 2005].

Predictability of BVCP Our third user study evaluates how well our proposed BVCP predicts the discomfort limit of binocular vision. We conduct the experiment based on the classical psychophysical methodology, *method of adjustment* [Norton et al. 2002]. Given an HDR image and a specific tone mapping operator, we produce a sequence of 10 LDR image pairs labeled from 0 to 9. The left and right LDR images of the 0th image pair are equivalent and are tone-mapped from the source HDR using a random parameter set. As the label number increases, the left LDR image remains unchanged throughout the whole sequence, while the right LDR images are generated by linearly increasing/decreasing the values of the most influential parameter of that particular tone mapping operator. The actual increasing/decreasing step size of the parameter value does not matter, provided that the sequence contains discomfort image pairs. Fig. 15 shows one such sequence. Five HDR images (1 to 5) and all four tone mapping operators are tested, resulting in 20 sequences.

Twenty-two participants are invited to take part in the experiment. They are asked to determine the discomfort limits of all 20 sequences. For each sequence, the participants are asked to search among the 10 image pairs to find the image pair of discomfort threshold, i.e. the image pair they begin to feel discomfort, and any image pairs with lower label numbers are still acceptable. The participants can move along the sequence by increasing or decreasing the label number. We then record the discomfort thresholds identified by participants. To avoid any bias, the first image pair within a sequence shown to the participants is randomly selected among the 10 pairs.

Fig. 16 plots the statistics of 20 test sequences. For each sequence, the mean of human-identified discomfort threshold is indicated as a dot, while the vertical interval corresponds to 95% confidence interval of human selections. In the mean time, we use the proposed BVCP to compute the predicted discomfort threshold. We plot each prediction as a vertical bar. Any image pair with label number above the bar will trigger visual discomfort. For all test sequences, our prediction (the blue bar) is generally below the lower bound of human-identified thresholds. In other words, our BVCP can always conservatively predict the discomfort threshold.

From the statistics, our BVCP metric does not perform equally well



Figure 15: One test sequence for evaluating BVCP predictability.

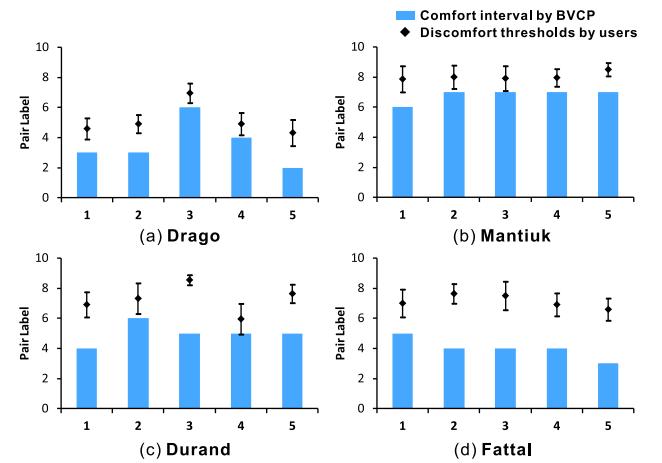


Figure 16: Statistics of the predictability of BVCP.

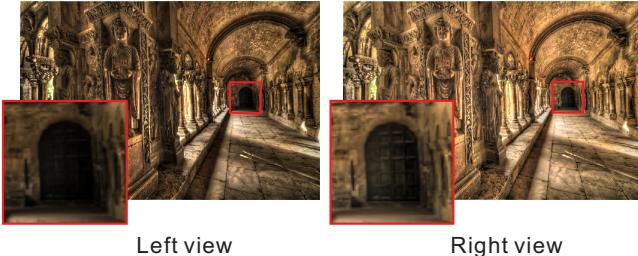
with all tested tone-mapping operators. Due to our current conservative design, its prediction is overly conservative for certain operator (Fig. 16(d)). Sometimes, it also may not be well correlated with user statistics (Fig. 16(c)). A possible explanation to this uncorrelation is the neglect of visual attention, as human observers may fail to identify those low-saliency sites while our predictor treats all sites equally.

Incorporating Stereopsis Note that the binocular fusion of color, luminance or contrast difference does not prevent the simultaneous perception of stereopsis [Steinman et al. 2000]. In other words, it is feasible to extend our binocular tone mapping to incorporate the stereopsis. Fig. 17 shows one example of binocularly tone-mapped stereo image pair from a stereo HDR image pair. It presents the depth cue and, simultaneously, raises the visual richness without triggering visual discomfort. To generate this example, we treat the left displaced view of the stereo HDR image pair as a standard HDR input just as before, and temporarily ignore the right displaced view. After applying our binocular tone mapping on the left displaced HDR image, we obtain a left LDR image and another LDR with the optimized tone mapping parameter(s). We then simply apply this optimized tone mapping parameter(s) to the right displaced HDR image. Since our framework is originally designed for single HDR input, the above approach implicitly assumes the luminances of the two HDR stereo images are similar and the disparity is small.

Limitations Our current BVCP design is very conservative. It rejects an image pair if any test of the contour fusion, contour con-



Figure 17: A stereo LDR image pair with left and right images tone-mapped differently.



Left view

Right view

Figure 18: A case with small improvement of overall visual richness (image courtesy of Jacques Joffre/Hdrsoft).

trast, or regional contrast fails. It rejects the image pair if there is obvious contour error at any level of the mesa pyramid. An image pair passes the BVCP test only if all pixels pass the BVCP test. All these add up to give a conservative metric. Although our prediction can effectively avoid visual discomfort for most individuals, it sometimes may not be very close to the user statistics (Fig. 16(d)). In other words, human-tolerable image pairs may sometimes be rejected. A future direction is to relax our current constraints to give predictions closer to the user statistics.

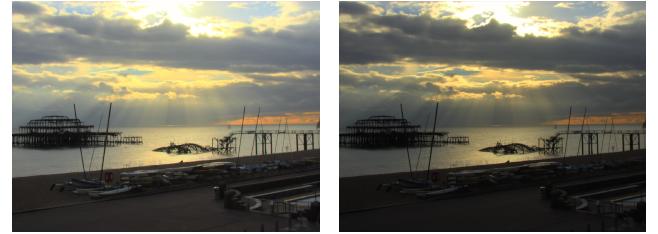
Our current framework treats all pixels in the image equally. But in reality, human vision is not equally sensitive to every pixel, due to visual attention. Fig. 18 shows one such image pair. While our method significantly increases the visual richness at the small door (boxed area), the attention of human observers is very likely to be attracted by another high-contrast region in the same image. This high-contrast region may, however, not be significantly improved, leading to an overall impression of insignificant improvement of visual experience. A possible extension to incorporate visual attention is to introduce an importance map based on the image content. Another limitation is that the contrast threshold function in modeling the failure of rivalry is approximated by extrapolation. Obviously, an independent psychophysical study is necessary in order to determine a more accurate contrast threshold function. Lastly, a more robust BVCP should account for occlusions due to the disparity, when applied to stereo input.

7 Conclusion

By applying the findings of vision science, we present a binocular tone mapping framework to generate a binocular LDR image pair that presents more visual richness than just a single tone-mapped image. We develop a novel BVCP metric that can conservatively predict the discomfort threshold. It guides our generation of binocularly tone-mapped image pairs, so that we can maximize the visual information content of the image pair without triggering visual discomfort. From the user studies, we demonstrate the effectiveness of our framework on a wide variety of images. Multiple tone mapping operators are experimented using our framework. Other than our current tone mapping application, the proposed BVCP may also be applied in any other applications requiring binocular display. Our work serves as the first attempt in graphics applications to maximize the utilization of stereo display system for binocular single vision. The next step is probably to extend our BVCP to support stereopsis when severe occlusion occurs.

Acknowledgments

This work is partially supported by the Research Grants Council of the Hong Kong Special Administrative Region, under RGC General Research Fund (Project No. CUHK 417411), CUHK Shun Hing Institute of Advanced Engineering, (Project No. SHIAE-MMT-P2-11), and CUHK Shenzhen Research Institute Start-up Fund. We



$b=0.625$

Figure 19: LDR pair of “BrightonPierOld” using Drago’s operator.



$s=13.000, c=3.500$

$s=91.300, c=3.600$

Figure 20: LDR pair of “GGmusicians” using Durand’s operator.



$a=0.734, b=0.826$

$a=0.857, b=0.880$

Figure 21: LDR pair of “WreathBuilding” using Fattal’s operator.



$f=0.600$

Figure 22: LDR pair of “Alhambra” using Mantiuk’s operator.

would like to thank Greg Ward (Fig. 1(right), 3, 12, 14, 15, 19–22), Jacques Joffre (Fig. 18), and Mark Fairchild’s HDR Photographic Survey (Fig. 1(left)) for allowing us to use their HDR images. We also thank all reviewers for their constructive comments and guidance in shaping this paper.

References

- BAKER, D., MEESE, T., MANSOURI, B., AND HESS, R. 2007. Binocular summation of contrast remains intact in strabismic amblyopia. *Investigative ophthalmology & visual science* 48, 11, 5332.
- BLAKE, R., AND BOOTHROYD, K. 1985. The precedence of binocular fusion over binocular rivalry. *Attention, Perception, & Psychophysics* 37, 2, 114–124.
- CARTER, R., AND HUERTAS, R. 2010. Ultra-large color difference and small subtense. *Color Research & Application* 35, 1, 4–17.
- CHEN, H., AND WANG, S. 2004. The use of visible color difference in the quantitative evaluation of color image segmentation. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, vol. 3, IEEE, iii–593.
- DALY, S. 1993. The visible differences predictor: an algorithm for the assessment of image fidelity. *Digital images and human vision* 11.
- DRAGO, F., MYSZKOWSKI, K., ANNEN, T., AND CHIBA, N. 2003. Adaptive logarithmic mapping for displaying high contrast scenes. In *Computer Graphics Forum*, vol. 22, Wiley Online Library, 419–426.
- DURAND, F., AND DORSEY, J. 2002. Fast bilateral filtering for the display of high-dynamic-range images. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, ACM, 257–266.
- EHRENSTEIN, W., ARNOLD-SCHULZ-GAHMEN, B., AND JASCHINSKI, W. 2005. Eye preference within the context of binocular functions. *Graefe's Archive for Clinical and Experimental Ophthalmology* 243, 9, 926–932.
- FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient domain high dynamic range compression. *ACM Transactions on Graphics* 21, 3, 249–256.
- HOWARD, I., AND ROGERS, B. 2002. *Seeing in Depth*, vol. 1. I Porteous.
- KOOI, F., AND TOET, A. 2004. Visual comfort of binocular and 3D displays. *Displays* 25, 2-3, 99–108.
- LAMBOOIJ, M., IJSELSTEIJN, W., FORTUIN, M., AND HEYNDERICKX, I. 2009. Visual discomfort and visual fatigue of stereoscopic displays: a review. *Journal of Imaging Science and Technology* 53, 030201.
- LARSON, G., RUSHMEIER, H., AND PIATKO, C. 1997. A visibility matching tone reproduction operator for high dynamic range scenes. *Visualization and Computer Graphics, IEEE Transactions on* 3, 4, 291–306.
- LEI, L., AND SCHOR, C. 1994. The spatial properties of binocular suppression zone. *Vision research* 34, 7, 937–947.
- LEVELT, W. 1965. Binocular brightness averaging and contour information. *British journal of psychology* 56.
- LIN, M., AND JANE, S. 2009. Analysis of color difference in digital proofing based on color management system. *Journal of Communication and Computer* 6, 10, 32.
- LIU, L., TYLER, C., AND SCHOR, C. 1992. Failure of rivalry at low contrast: Evidence of a suprathreshold binocular summation process. *Vision Research* 32, 8, 1471–1479.
- MACMILLAN, E., GRAY, L., AND HERON, G. 2007. Visual Adaptation to Interocular Brightness Differences Induced by Neutral-Density Filters. *Investigative ophthalmology & visual science* 48, 2, 935.
- MANTIUK, R., DALY, S., MYSZKOWSKI, K., AND SEIDEL, H. 2005. Predicting visible differences in high dynamic range images: model and its calibration. In *Proc. SPIE*, vol. 5666, SPIE, 204–214.
- MANTIUK, R., MYSZKOWSKI, K., AND SEIDEL, H. 2006. A perceptual framework for contrast processing of high dynamic range images. *ACM Transactions on Applied Perception (TAP)* 3, 3, 286–308.
- MANTIUK, R., KIM, K., REMPEL, A., AND HEIDRICH, W. 2011. Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. In *ACM Transactions on Graphics (TOG)*, vol. 30, ACM, 40.
- MYSZKOWSKI, K. 1998. The visible differences predictor: Applications to global illumination problems. In *Rendering techniques' 98: proceedings of the Eurographics Workshop in Vienna, Austria, June 29-July 1, 1998*, Springer Verlag Wien, 223.
- NORTON, T., CORLISS, D., AND BAILEY, J. 2002. *The psychophysical measurement of visual function*. Butterworth Heinemann.
- ONO, H., ANGUS, R., AND GREGOR, P. 1977. Binocular single vision achieved by fusion and suppression. *Attention, Perception, & Psychophysics* 21, 6, 513–521.
- O'SHEA, R. 1983. Does stereopsis have a fusional component? *Attention, Perception, & Psychophysics* 34, 6, 599–603.
- REINHARD, E. 2006. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann.
- STEINMAN, S., STEINMAN, B., AND GARZIA, R. 2000. *Foundations of binocular vision: A clinical perspective*. McGraw-Hill Medical.
- TREISMAN, A. 1962. Binocular rivalry and stereoscopic depth perception. *The Quarterly Journal of Experimental Psychology* 14, 1, 23–37.
- VON HELMHOLTZ, H. 1962. *Helmholtz's treatise on physiological optics*, vol. 3. The Optical Society of America.
- WANG, Z., BOVIK, A., SHEIKH, H., AND SIMONCELLI, E. 2004. Image quality assessment: From error visibility to structural similarity. *Image Processing, IEEE Transactions on* 13, 4, 600–612.
- WATSON, A. 1987. The cortex transform: rapid computation of simulated neural images. *Computer Vision, Graphics, and Image Processing* 39, 3, 311–327.
- WOPKING, M. 1995. Viewing comfort with stereoscopic pictures: An experimental study on the subjective effects of disparity magnitude and depth of focus. *Journal of the Society for Information Display* 3, 3, 101–103.