CrossMark

Visual discomfort prediction on stereoscopic 3D images without explicit disparities

Jianyu Chen[a,b,*], Jun Zhou[b,a], Jun Sun[b], Alan Conrad Bovik[a]

[a] Laboratory for Image and Video Engineering (LIVE), The University of Texas at Austin, Austin, TX 78701, USA
[b] Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

A B S T R A C T

Almost all existing 3D visual discomfort prediction models are based, at least in part, on features that are extracted from computed disparity maps. These include such estimated quantities such as the maximum disparity, disparity range, disparity energy and other measures of the disparity distribution. A common first step when implementing a 3D visual discomfort model is some form of disparity calculation, whence the accuracy of prediction largely depends on the accuracy of the disparity result. Unfortunately, most algorithms that compute disparity maps are expensive, and are not guaranteed to deliver sufficiently accurate or perceptually relevant disparity data. This raises the question of whether it is possible to build a 3D discomfort prediction model without explicit disparity calculation. Towards this possibility, we have developed a new feature map, called the percentage of un-linked pixels (PUP), that is descriptive of the presence of disparity, and which can be used to accurately predict experienced 3D visual discomfort without the need for actually calculating disparity values. Instead, PUP features are extracted by predicting the percentage of un-linked pixels in corresponding retinal patches of image pairs. The un-linked pixels are determined by feature classification on orientation and luminance distributions. Calculation of PUP maps is much faster than traditional disparity computation, and the experimental results demonstrate that the predictive power attained using the PUP map is highly competitive with prior models that rely on computed disparity maps.

1. Introduction

Stereoscopic movies have gained popularity in recent years, as reflected by the large amount of 3D content being offered to consumers at the cinema and in home theatres. Although 3D displays effectively enhance viewers' visual experience by providing a more immersive, stereoscopic visualization, low-quality 3D images can induce feelings of discomfort such as eye strain, headache, fatigue, asthenopia, and other phenomena leading to a less pleasant viewing experience [1]. Several possible factors that may affect visual discomfort have been extensively studied, such as the vergence–accommodation conflict [2], duration of viewing, the viewing distance [3–5], the amount of defocus-blur [6–8], and the distribution of disparities along the vertical dimension and its relation to the shape of the empirical horopter [9,10]. Towards being able to predict and potentially reduce feelings of visual discomfort experienced when viewing 3D images, a large number of studies have focused on finding features that can be reliably extracted from 3D images (stereopairs) towards developing automatic 3D discomfort prediction algorithms [11–13].

In the absence of geometrical distortions and window violations, factors related to vergence and accommodation conflict induced by horizontal disparity are thought to be the dominant factors that cause visual discomfort [14]. Commonly, the features used in discomfort prediction models have included the disparity range, maximum angular disparity, disparity distribution, disparity location and disparity gradient [15–20], among various other quantities calculated from disparity maps. Hence, the predictive powers of existing discomfort assessment models generally depend strongly on the accuracy of disparity calculation.

Disparity calculation methods can be classified into algorithms that conduct sparse and dense stereo matching. Sparse stereo matching methods do not calculate disparity at every pixel, hence often offer the advantage of low complexity [17,18]. Dense stereo matching methods calculate disparity at every pixel. Most recent discomfort assessment models are built on dense stereo matching algorithms, such as the matching algorithm [21] used in [15], dynamic programming [22,19], and the Depth Estimation Reference Software [23] used in [16]. Kim and Sohn used both sparse and dense disparity calculations [20]. High-

quality dense disparity calculation is both expensive and difficult, yet the predictive capability of discomfort models could suffer if only sparse disparity features delivered by a low complexity algorithm were available. To improve the accuracy of disparity calculation results while also increasing calculation speed, additional information regarding the range of available depths or of the camera parameters may be exploited [22,23], but this kind of information may not be easy to obtain. These limitations raise the question of whether it is possible to build a 3D visual discomfort model that does not require disparity calculation. Indeed, there is no convincing perceptual evidence that disparity maps are the best source of information when defining features to drive visual comfort models. Disparity information provided by depth maps is neither a continuous nor a "linear" source of information [24,25].

Therefore, we ask: are there other discomfort-predictive features that can be quickly computed from stereopairs while enabling accurate discomfort prediction? Is it enough to infer the presence of disparity without computing it? Towards realizing these possibilities, we formulated a new discomfort predictive framework that relies on a simple, perceptually relevant implicit disparity tool, that we call the percentage of un-linked pixels map (PUP map).

2. Background

In a 3D scene viewed on a stereoscopic display, referred to here as Stereoscopic 3D (S3D), accommodation is fixed by the distance of the dichoptic images from the two eyes. However, vergence is free to adapt to the disparity-defined depth planes that occur when a fused image is achieved. This produces a perceptual conflict. Prolonged exposure to conflicts between vergence and accommodation when viewing S3D content produces visual fatigue and discomfort [26]. The binocular disparity signal is the primary cue for vergence. It is much stronger than blur or other factors in evoking vergence [27]. To simulate the vergence–accommodation conflict without the use of disparity maps, we instead seek to deploy other information descriptive of the binocular disparity signals.

2.1. Corresponding retinal images and linked images

When viewing natural scenes, the two eyes receive corresponding images that originate from the same locations in space (Fig. 1). Images that the visual system treats as corresponding images will be referred to as linked images [28].

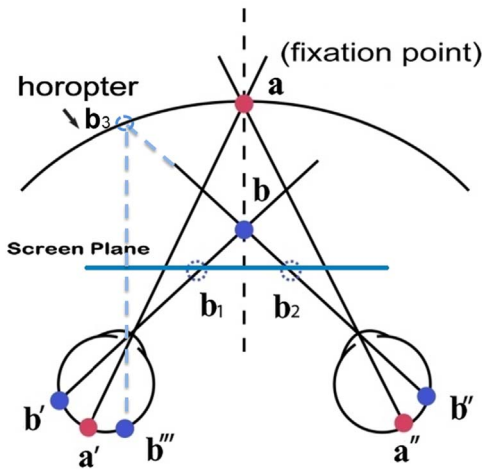Corresponding retinal images are images that have the same

Fig. 1. The relationship of corresponding images, linked images and corresponding retinal images. Point **a** is the fixation point, while **a′** and **a″** are the linked images of point **a** on the left and right retinae. Points **b′** and **b″** are the linked images of **b**. Point **b‴** is the corresponding retinal image of **b″**, also projected from point **b₃** on the horopter. **b₁** and **b₂** are corresponding images of **b**, both located on the screen plane.

projected location on the two retinae and that also project to the same location in visual cortex [29]. For example, the blue point **b** in Fig. 1 is a point in space. The points **b₁** and **b₂** are corresponding points displayed on a screen to the left and right eyes. The points **b′** and **b″** are the linked images of these points received by the stereoscopic vision system. The point **b‴** is the corresponding retinal image of **b″**.

The horopter is the surface containing all those points in space whose images fall on corresponding points of the retinae of the two eyes.

Given a set of distinct objects on the horopter, the visual system has no difficulty linking corresponding images, since each point in the image is linked to a corresponding retinal position. In this situation, the linked images are also corresponding retinal images. Hence the disparity between the linked images is zero and no disparity signal is triggered.

During stereoscopic vision, the fixation point is the location in 3D space that is fixated by the two eyes.

For example, in Fig. 1, the fixation point **a** which is on the horopter is projected to **a′** on the left eye and **a″** on the right eye. Since they are projected from the same point in space and located on the horopter, they are linked in the stereoscopic vision system, and no disparity signal is triggered.

However, for images which do not lie on the horopter, there exists a mismatch between the linked image and corresponding retinal image due to disparity. Images are not linked to their corresponding retinal images, but are instead linked to shifted images. For example, point **b** in Fig. 1, which is in front of the horopter, is projected to **b′** and **b″** on the two retinas. The point **b′**, which is linked to **b″**, is shifted leftwards from its corresponding retinal image **b‴**. These unlinked images will trigger disparity signals [28,30]. A useful and robust predictor of the presence of disparity signals may be formed by judging whether the corresponding retinal image is linked or unlinked.

2.2. Percentage of un-linked pixels

In Fig. 2, the red blocks in the right views are the retinal image (CRI) blocks that correspond to the blue blocks in the left view. The blue blocks in the right view are the linked image (LI) blocks. It is supposed that the pixels in the left view blocks have the same disparities.

Several possible conditions exist. In condition ① indicated in Fig. 2, since the disparity is 0, the LI block is the same as the CRI block. However, for conditions ② through ⑦, since the LI blocks are linked to shifted retinal image blocks because of horizontal disparities, mismatches arise between the LI blocks and CRI blocks. Further, the shifted distances between the LI blocks and the CRI blocks are the same as the disparities between the blocks. For LI blocks having no zero disparities, only part of the image is linked to the CRI block due to shift. The shaded area in the left view is that part of a blue LI block that can be linked to a sub-image in its corresponding retinal image. The shaded region in the right view is its linked area. Therefore in condition ①, the shaded region in the right view is as large as the CRI block. However when the disparity increases, the shaded region is reduced (conditions ② and ③). When the disparity is at least as large as the width of the CRI block (conditions ④ and ⑤), there is no shaded region, i.e., no part of the blue LI block is linked to its corresponding retinal image.

On the other hand, the size of the shaded region is also affected by the size of the image. For a larger block (conditions ⑥ and ⑦) of width of $2W_a$, the shaded region is larger than the smaller block having the same disparity. In this situation, most of the blue block is linked to its CRI block.

Hence given the disparity and the size of the block, the size of the shaded area can be described by them as follows:

$$\begin{cases} S_{Linked} = S_a\left(1 - \dfrac{d_a}{W_a}\right) & 0 \le d_a < W_a \\ S_{Linked} = 0 & d_a \ge W_a \end{cases} \tag{1}$$
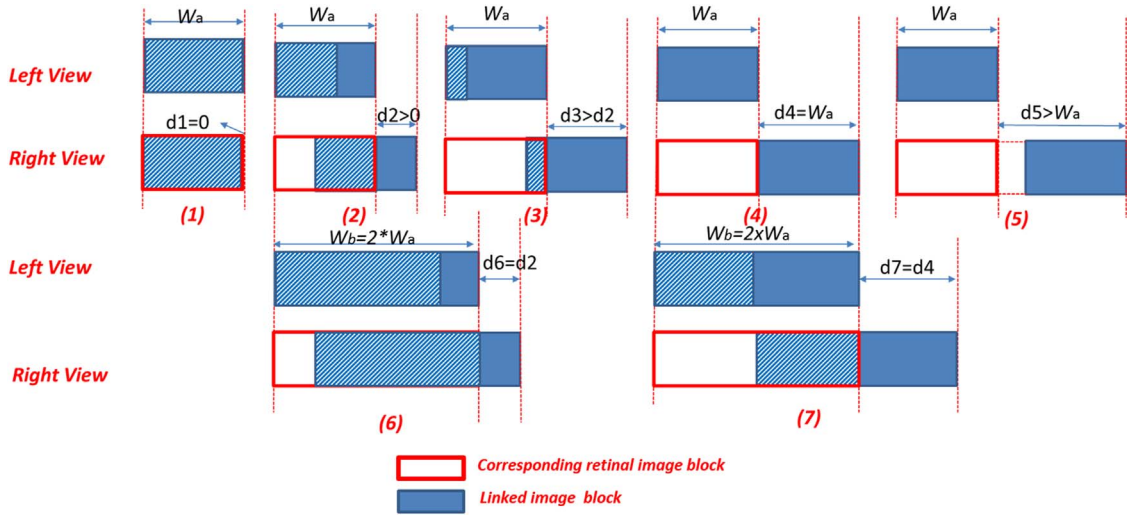
Fig. 2. Positional relationships between linked images and corresponding retinal images.

where S_a is the area of CRI block a, W_a is the width of CRI block a, d_a is the disparity of the block and S_{Linked} is the shaded area which is the part of the block that can be linked to its corresponding retinal image. Conversely, if the size of the shaded area can be found, then the disparity of a block having consistent disparity can be approximately expressed as follows:

$$\begin{cases} \overline{d_a} = \left(1 - \dfrac{S_{Linked}}{S_a}\right)W_a & d_a \leq W_a \\ \overline{d_a} = W_a & d_a > W_a \end{cases} \tag{2}$$

where $\overline{d_a}$ is the estimated disparity. From (2), it is apparent that for a given block width, the disparity has a proportional relationship with $\frac{S_a - S_{Linked}}{S_a}$. This parameter is the percentage of coordinates in the left image which do not have links to the corresponding retinal image. We call this quantity *the percentage of un-linked pixels* (PUP). If PUP = 0 over a patch, then the local disparity is zero and all of the patch pixels are linked. Conversely, if PUP = 1, the disparities will be equal to or larger than the width of the block, and all of the pixels in the block are un-linked.

We use this quantity to describe the likely presence of disparity. As we elaborated in last section, others have shown that visual discomfort induced by watching S3D images can be predicted by features extracted from disparity signals. Here we instead use quantities that are predictive of disparity, but that do not require expensive calculation of disparity values, to predict the degree of experienced visual discomfort induced by viewing S3D images.

3. Visual discomfort modeling using PUP

The stereoscopic vision system attempts to link images having similar luminance, size, shape, orientation, and color. Conversely, images having different orientations, luminances or other aspects are not linked [28,30]. Hence, a corresponding retinal image pair having more pixels with disparate luminances, orientations or other image aspects will generally have more un-linked pixels and a high PUP.

For instance, Fig. 3(a) shows a S3D image from the IEEE-SA database [31]. The red block has a large disparity while the blue one has a small disparity. Figs. 3(d) and (e) are CRI blocks in the left and right views corresponding to the red block, while Figs. 3(b) and (c) are CRI blocks corresponding to the blue block. Figs. 3(f)–(i) are the corresponding luminance histograms. From Figs. 3(b) and (c), since most of the pixels in blue block are linked to the CRI block, and linked pixels have similar luminances, the luminance histograms of the left view (Fig. 3(f)) and the right view (Fig. 3(g)) are similar. Otherwise,

since only a small percent of pixels in the red block are linked, the luminance histograms (Figs. 3(h) and (i)) differ significantly between views. The right view contains more low-luminance pixels (Figs. 3(e) and (i)) which cannot be linked with the pixels of higher luminance in the left view (Figs. 3(d) and (h)).

A similar result is obtained with orientation histograms. We captured orientation information using the responses of Gabor filters of different orientations [32,33]. The responses were normalized to the range [0, 1] before computing distribution histograms. Horizontal and vertical oriented Gabor filters were deployed with center frequencies 0.592 cycles/degree. The design of the Gabor filter tessellation and parameters was based on the perceptual design conducted by Su et al. [34] which uses a cortical model based on neuronal data. Figs. 3(j)–(m) show the horizontal orientation histograms. The horizontal orientation histograms of the blue block are very similar, but the orientation histograms of the red block are different. A very similar result is also obtained for the vertical orientation histograms. More pixels in the right view of the red block have large Gabor responses which cannot be linked to pixels with small Gabor responses in the left view.

To obtain PUP values, it is first necessary to classify pixels into groups such that pixels within each group have similar features. These groups are defined as feature groups. It is supposed that pixels within the same feature group are linked to each other, and that pixels from different groups cannot be linked. For example, the pixels in the left and right images are classified into two feature groups, FG1 and FG2. If the left image has 10 more pixels than its corresponding retinal image in group FG1, these 10 pixels cannot be linked to pixels in group FG2 in the corresponding retinal image, hence are un-linked.

Although in some cases (e.g., background and foreground images having very similar luminance, texture or other image features) un-linked pixels may be classified as linked pixels by feature grouping, in most cases, the number of un-linked pixels is well represented by counting the numbers of pixels in the different feature groups:

$$S_{unlinked} = \frac{\sum_{i=1}^{N_{hist}} |H^{l,N_{hist}}(i) - H^{r,N_{hist}}(i)|}{2} \tag{3}$$

where N_{hist} is the number of feature groups and $H^{l,N_{hist}}(i)$ and $H^{r,N_{hist}}(i)$ are the numbers of pixels in the ith group in the left and right views. Hence PUP is defined as:

$$PUP^{N_{hist}} = \frac{\sum_{i=1}^{N_{hist}} |H^{l,N_{hist}}(i) - H^{r,N_{hist}}(i)|}{2N_{total}} \tag{4}$$

It is important to observe that while (4) measures a type of similarity between blocks or patches, the similarity measure is made
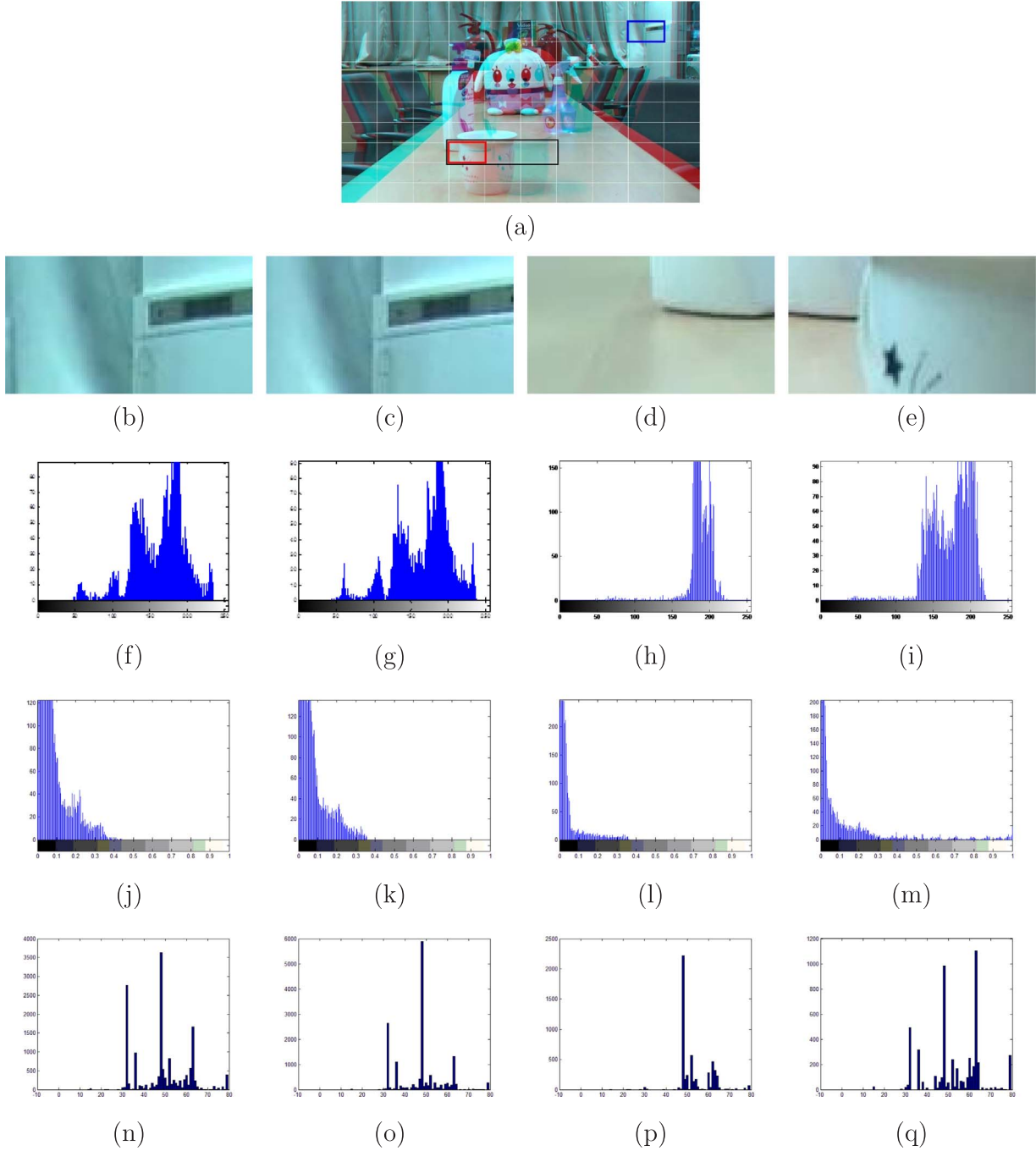
Fig. 3. Example image from IEEE-SA, from left column to right column: (b) and (c) are the corresponding blue blocks in the left and right images, (d) and (e) are the corresponding red blocks in the left and right images, respectively, (f)–(i) are the corresponding luminance histograms, (j)–(m) are the corresponding horizontal orientation histograms, and (n)–(q) are the feature distribution histograms.

only on corresponding blocks, i.e., there is no searching or "matching" process or attempt to estimate actual disparity values therefrom. The same is true of later expressions, e.g., (6) and (7).

3.1. Pixels grouping by feature classification

Following the analysis from earlier, the pixels can be classified by their orientations and luminance distribution. Pixels having both similar orientation and luminance distributions are classified into the same group. Orientation features are extracted along four different Gabor orientations (both cardinals and both diagonals) at a constant spatial frequency of 0.592 cycles/degree. The design of the Gabor filter

set was based on the perceptual model used by Su et al. [34]. The responses are normalized into the range [0, 1] before the classification process. If the normalized response is sufficiently large along an orientation, then the orientation feature along this direction is marked as high. Otherwise, the orientation is marked as low. The judgment is made according to comparison with a threshold as:

$$
\begin{cases}
O^l(x, y, \theta) = \text{high} & \text{if } G^l(x, y, \theta) \geq T_G(\theta) \\
O^l(x, y, \theta) = \text{low} & \text{if } G^l(x, y, \theta) < T_G(\theta)
\end{cases} \tag{5}
$$

where $O^l(x, y, \theta)$ is the orientation feature associated with the left view pixel at orientation θ at coordinate (x, y) and $G^l(x, y, \theta)$ is the normalized Gabor response at (x, y). $T_G(\theta)$ is the threshold applied on the Gabor

response at orientation θ. Hence if $G^l(x, y, \theta)$ is at least as large as its corresponding threshold $T_G(\theta)$, $O^l(x, y, \theta)$ is marked as high. Otherwise, $O^l(x, y, \theta)$ is marked as low. Here we simply take $T_G(\theta) = 0.5$. Hence pixels are classified into 16 different groups according to their orientation features.

After classification by orientation, the pixels are sub-grouped by luminance levels within each orientation feature group. The number of luminance levels was empirically fixed at 5. Finally, the pixels are classified into 80 ($N_{hist} = 80$) different feature groups. Figs. 3(n)–(q) plot the resulting feature distribution histograms for the red and blue blocks. For each red block, the number of pixels in each feature group differs between the two views. By contrast, the number of pixels in each feature group is very similar between the two views for blue blocks. Hence the same result is obtained: most pixels in the red block cannot be linked to pixels in its CRI block. However, most pixels in the blue block can be linked to the pixels in its CRI block.

3.2. The sign of PUP

In addition, as described in [17,35–37,14,9], excessive uncrossed disparities generally produce more discomfort than excessive crossed disparities of the same magnitudes. Hence the polarity of disparity seriously impacts the degree of experienced visual discomfort. If the local disparity is uncrossed, the locally linked image will be shifted to the left of the corresponding retinal image in the right view. The percentage of un-linked pixels between the left and shifted right view images are a useful measure of the direction of disparity. The sign of PUP can be used to represent the polarity of disparity, and can be estimated as follows:

$$\begin{cases} PUP_{L,(x,y)} = \dfrac{\sum_{i=1}^{N_{hist}} |H^l_{(x,y)}(i) - H^r_{(x-t,y)}(i)|}{2N_{total}} \\[2mm] PUP_{R,(x,y)} = \dfrac{\sum_{i=1}^{N_{hist}} |H^l_{(x,y)}(i) - H^r_{(x+t,y)}(i)|}{2N_{total}} \end{cases} \quad (6)$$

where t is a variable that controls the degree of shift in the corresponding right view image. For example, Fig. 4 depicts the spatial relationship between the right shifted right view image and the left view image. The value of t in Eq. (6) is estimated by the disparity value predicted by PUP:

$$t = PUP_{(x,y)} \times W \quad (7)$$

$$\begin{cases} PUP_{(x,y)} = -PUP^U_{(x,y)}, & (PUP_{L,(x,y)} \leq PUP_{R,(x,y)}) \\ PUP_{(x,y)} = PUP^U_{(x,y)}, & (PUP_{L,(x,y)} > PUP_{R,(x,y)}) \end{cases} \quad (8)$$

where $PUP_{(x,y)}$ is the signed PUP value and $PUP^U_{(x,y)}$ is the unsigned PUP value. At local coordinate (x, y), $PUP_{L,(x,y)}$ and $PUP_{R,(x,y)}$ are the values of PUP computed between the left and shifted right view images, respectively. When $PUP_{L,(x,y)} < PUP_{R,(x,y)}$, then the left image has more pixels linked to the left shifted image in the right view and the disparity signal is uncrossed. When $PUP_{L,(x,y)} = PUP_{R,(x,y)}$, the sign of $PUP_{(x,y)}$ is negative.

3.3. PUP map

For every sub-image in an S3D image pair being processed, a signed PUP value is obtained. To simplify calculation, the images are divided into sub-images of constant size rather than attempting image segmentation.

3.3.1. Height of patch

Firstly, the height of each patch is determined following Kooi and Toet [8], who provided a limit on vertical disparity. They recommended that viewing comfort is not reduced by keystone distortions of
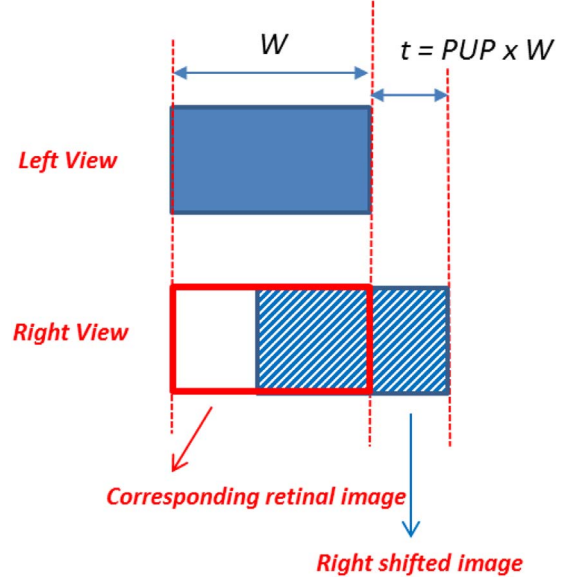

Fig. 4. Spatial relationship between the right shifted right view image and the left view image.

up to 0.57° of visual angle. The viewing distance between each subject and the 3D display was fixed at about 3 times the screen height as suggested in ITU-R BT.1438 [38] and as used in previous studies [39,16,14]. Hence the approximate patch height for an image of resolution 1920×1080 was 32 pixels.

3.3.2. Width of patch

From (2), the maximum disparity which PUP can describe is controlled by the width of the sub-image. Therefore, the width of the image patch should be set as reasonably large as possible to capture large disparities. However, as the width increases, an image patch may contain diverse objects and disparity levels which will affect the estimation of disparity signals. Since different images may have very different aspects, images with highly variable disparity levels and many details may be better analyzed using a smaller patch width, while images with few disparity levels and details should be analyzed using a larger width. Generally, multiple patch widths are needed for accurate modeling.

The region within which objects are fused binocularly is called Panum's fusional area. Objects located outside Panum's area induce double vision and binocular suppression may occur [40]. While Panum's fusional area marks the limit of the accommodative output under natural viewing conditions, comfortable viewing is not guaranteed.

A sub-region inside Panum's fusional area known as Percival's zone of comfort defines maximum retinal disparities under normal viewing conditions that lead to comfortable viewing [41]. Hence if the disparity of the image patch is smaller than Percival's zone of comfort, it will have little effect on the visual experience. If the disparity surpasses Panum's fusional area, the degree of visual discomfort may greatly increase.

Therefore, the width of the image patch is defined as a function of the size of the comfort zone and the fusional area. Three different patch widths are deployed in the model. The width of the largest patch corresponds to the disparity limit of the fusional area. The width of the smallest patch corresponds to the disparity limit of the comfort zone. The width of the average patch corresponds to the average value of the fusional limit and the comfort limit disparities.

Disparities larger than 60–70 min of arc are more likely to induce visual discomfort [42,43]. Hence 60 min of arc was determined as the disparity limit of the comfort zone. Disparities larger than about 4.93° generally cannot be fused without diplopia [40]. Hence 4.93° was

determined as the disparity limit of the fusional area. The block widths were determined using the pixel disparity limit of the fusional area and the comfort zone. The pixel disparity is transformed from angular disparity to pixel disparity as follows:

$$d = \frac{b/2 - V*\tan\left(\arctan\left(\frac{b}{2V}\right) - d_\theta\right)}{PP},$$

(9)

where d is the pixel disparity, b is the distance between the two eyes (65 mm), V is the viewing distance (three times the height of the monitor), PP is the size of each pixel on the screen, and d_θ is the angular disparity. For an image with a resolution of 1920×1080 pixels, the block widths of PUP-L, PUP-M, PUP-S were computed to be 280, 168 and 56 pixels, respectively, and PUP-M is found by averaging the block widths of the PUP-L and PUP-S maps. Given an image pair, three PUP maps with large, average and small patch widths (PUP-L, PUP-A and PUP-S) were extracted.

3.3.3. Overlapping patches

Typically, image patches partition the image along rows. For example, in Fig. 5, a piece of the image is subdivided into two patches p_1 and p_2 of widths W_b. However, objects may overlap multiple patches, like the green circle in Fig. 5, which may lie in different disparity planes than the background. This may adversely affect the accuracy of the PUP signals. Instead, overlapping patches may be used. In the example in Fig. 5 (lower part), adjacent patches have overlap of width $(1 - b_s) \times W_b$.

Hence for a patch of width W_b and height H_b, the PUP map is:

$$\begin{Bmatrix} PUP_{(1,1)} & PUP_{(1+b_s \times W_b, 1)} & \cdots & PUP_{(1+b_s \times (n-1) \times W_b, 1)} \\ PUP_{(1,1+H_b)} & \cdots & \cdots & \cdots \\ \cdots & \cdots & PUP_{(x,y)} & \cdots \\ PUP_{(1,1+(m-1) \times H_b)} & \cdots & \cdots & PUP_{(1+b_s \times (n-1) \times W_b, 1+(m-1) \times H_b)} \end{Bmatrix}$$

(10)

where $b_s < 1$. When W_b takes larger values, b_s is made smaller. In our implementation b_s was fixed at 1/5, 1/3, and 1/2 for PUP-L, PUP-A and PUP-S, thereby making the distances between adjacent blocks in PUP-L and PUP-A the same as the width of the blocks in PUP-S (56 pixels). Hence the total number of patches in the PUP map is:

$$N_{patch} = \frac{H_I}{H_b} \times \left(\frac{W_I - W_b}{b_s \times W_b} + 1\right)$$

(11)

where N_{patch} is the number of patches and H_I and W_I are the height and
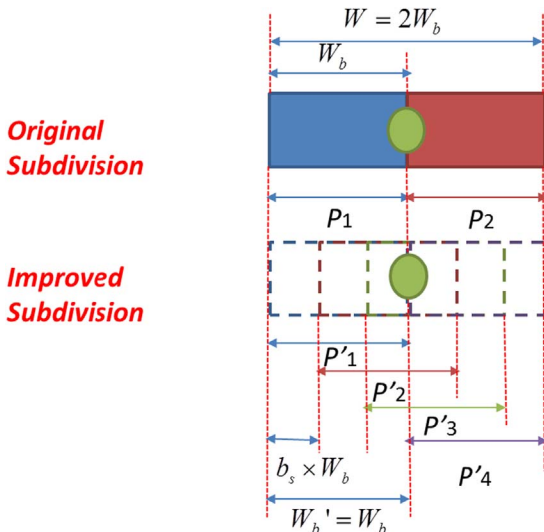

Fig. 5. Example of improved patch subdivision.

width of the image. Fig. 6 is the corresponding set of PUP maps obtained on the image in Fig. 3.

3.4. Features used in PUP map

In order to predict the degree of experienced visual discomfort on an S3D image, each of four kinds of features were extracted from the PUP-L, PUP-A and PUP-S maps following the feature extraction method in [15]. Thus, each kind of map has four kinds of features and the total number of features used in PUP map set is 12. The features include the mean values of positive and negative PUP, and the upper and lower 5% of the PUP values:

$$f_1 = \frac{1}{N_{Pos}} \sum_{PUP(n)>0} PUP(n)$$

(12)

$$f_2 = \frac{1}{N_{Neg}} \sum_{PUP(n)<=0} PUP(n)$$

(13)

$$f_3 = \frac{1}{N_{5\%}} \sum_{n \leq N_{total} \times 0.05} PUP(n)$$

(14)

$$f_4 = \frac{1}{N_{95\%}} \sum_{n \geq N_{total} \times 0.95} PUP(n)$$

(15)

where $PUP(n)$ is the nth smallest value in a PUP map, and N_{Pos} and N_{Neg} are the number of positive and negative values in a PUP map, respectively. If $N_{Pos} = N_{Neg} = 0$, then $f_1 = f_2 = 0$. $N_{5\%}$ and $N_{95\%}$ are the number of values lower and higher than 5% and 95% of the PUP values.

4. Performance evaluation

To evaluate the efficiency of the PUP map features in predicting visual discomfort, the compute time, correlation coefficient and outlier ratio between predicted results and MOS and robustness were tested in this section.

The features extracted from the set of PUP maps were tested on the IEEE-SA stereo image database [31], which consists of 800 stereo high-definition images, along with associated subjective visual discomfort scores. The IEEE-SA database was specifically built to test 3D visual discomfort predictor models [15,14] (more details can be found in [14]). We rigorously tested the PUP features on the IEEE-SA stereo image database and compared the results with prior work [15,17–20,44]. Among these models, Nojiri et al. [17] and Yano et al. [18] also considered temporal aspects of visual discomfort, hence are less comparable. The temporal parts of their models were removed by us for better comparison in this experiment. Choi et al. [19] targeted visual fatigue which is one of the main aspects of visual discomfort and may impact the comparison results. Table 1 overviews the depth estimation algorithm and the features extracted in these prior works.

To compare the discomfort prediction efficiency of the PUP maps with that of features extracted from disparity maps, the same types of features were also extracted from the results of four popular disparity extraction models using Eqs. (11)–(14): the mean values of crossed (MP) and uncrossed (MN) disparities (MN), and the top 5% of the largest (TL) and smallest (TS) disparities. The details of the feature extraction can be found in [15]. The disparity estimation models tested were: the optical flow based algorithm [21], dynamic programming [22], a low complexity sum-of-absolute difference stereoscopic lumi-nance matching model (Disparity SAD or DSAD), and a simple SSIM based stereo algorithm DSSIM [50]. DSSIM is a window-based stereo matching algorithm based on the SSIM index (DSSIM) [50]. The disparity map of a stereo pair is generated by using SSIM as the matching objective, resolving ties by a minimum disparity criterion. Both the computation times and prediction results were tabulated for comparison. The speed testing was done on the IEEE-SA stereo image database. The computing environment was an Apple MacPro4.1 with
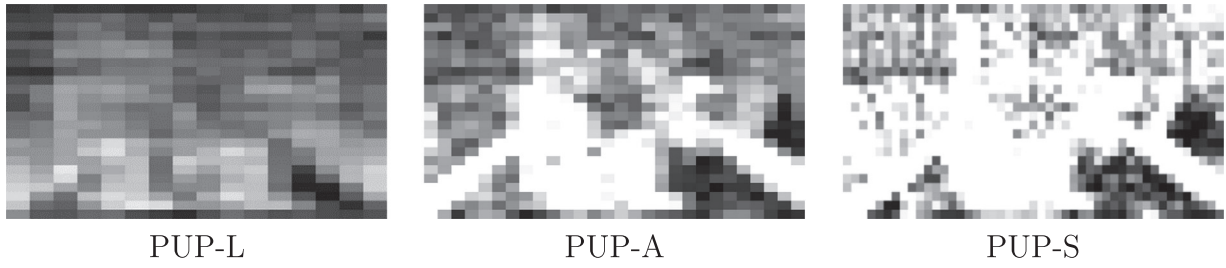
PUP-L　　　　　PUP-A　　　　　PUP-S

Fig. 6. Set of PUP maps for the image in Fig. 3.

Table 1
Overview of comparison models.

Models	Depth estimation	Extracted features
Nojiri et al. [17]	Block level phase correlation computation [17]	The minimum and maximum values, range, dispersion, absolute average, and average disparities
Yano et al. [18]	Block level correlation computation [18]	The ratio of sums of horizontal disparities near the screen and those far from the screen.
Choi et al. [19]	Dynamic programming [22]	Spatial depth complexity and depth position (the variance and absolute mean of the disparity map)
Kim and Sohn [20]	SIFT matching [45,46] and region-dividing technique with energy-based regularization [47]	The experienced horizontal disparity range and maximum angular disparity
Richardt et al. [44]	Normalized cross-correlation [48] and left–right consistency check [49]	The consistency of the two stereo half-images
Park et al. [15]	Optical flow software [21]	Anomaly of AV/A ratio, anomaly of VA/V ratio, absence of de-focus blur, and absence of differential blur

Table 2
Compute times of disparity calculation algorithms and PUP map extraction.

Algorithm	**PUP**	Flow	Dynamic	DSSIM	DSAD	Park et al. [15]
Time (h)	**2.41**	45.71	40.17	22.04	3.51	57.21

Intel Xeon CPU e5520 2.27 GHz (16CPUs) and 6 GB of RAM. All models were implemented in MATLAB. Table 2 shows the computation comparison between PUP map set extraction and prior work: Park et al. [15] and the four disparity algorithms. The computation times

were recorded in units of hours. From Table 2, it is apparent that PUP map set extraction is faster than all of the disparity calculation algorithms and much faster than Park et al. [15].

The PUP model and comparison prior models were all trained and tested on the IEEE-SA database using a support vector regressor (SVR). The IEEE-SA database with corresponding MOS was divided into test and training subsets. SVRs have been shown to perform well on high-dimensional regression problems. To implement the SVR, we used the LibSVM package [51] with the radial basis function kernel, whose parameters were estimated by cross-validation during the training session. 1000 iterations of the train–test process were used and the image database was randomly divided into 80% training and 20% test sets across 1000 iterations at each iteration. The training and testing sets did not overlap in content.

The performance was measured using Spearman's rank ordered correlation coefficient (SROCC), (Pearson's) linear correlation coefficient (LCC), and root mean square error (RMSE) between the predicted scores and the MOS following ITU-T P.1401 [52]. Higher SROCC and LCC and lower RMSE values indicate good correlation (monotonicity and accuracy) against human quality judgments. Before calculating the performance measures (other than SROCC), the algorithm scores were transformed using a four parameter non-linear (logistic) regression as recommended in [53].

To observe the relationship between the specific band of spatial frequencies that the Gabor filter is tuned to, the mean values of SROCC and LCC were computed as a function of spatial frequency over the 1000 iterations, and plotted in Fig. 7. Clearly, the predictive power of the PUP model is not seriously affected by the tuning of the Gabor filter.

In addition, to demonstrate that PUP is not highly dependent on the size of the training set, we also obtained the mean values of the LCC and SROCC as a function of the percentage of the overall database used, over 1000 iterations, as shown in Fig. 8. The LCC and SROCC slightly decreased with a reduction of the proportion of training data, but this was insignificant above 10% of the database.

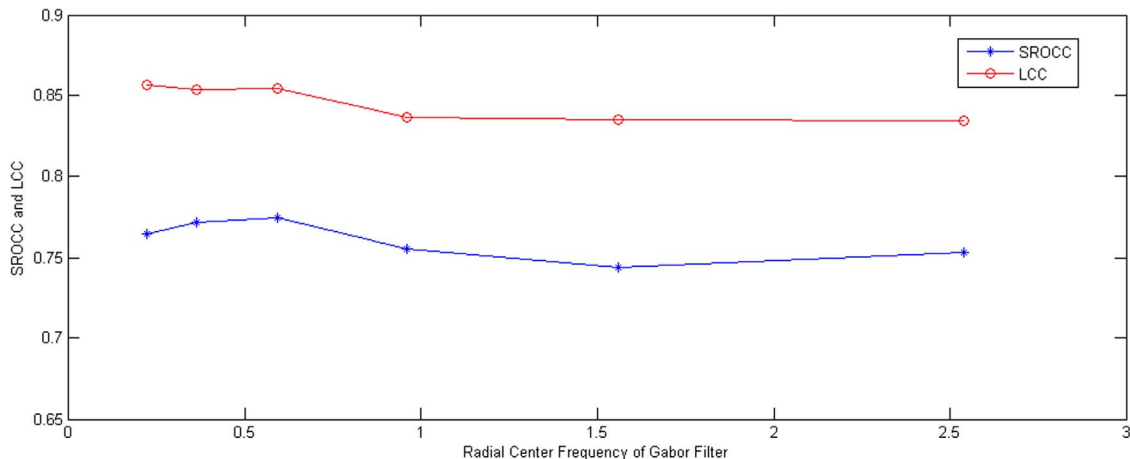We obtained the mean, median, and standard deviations of LCC,

Fig. 7. Mean SROCC and LCC against MOS as a function of the frequency tuning of Gabor filter.
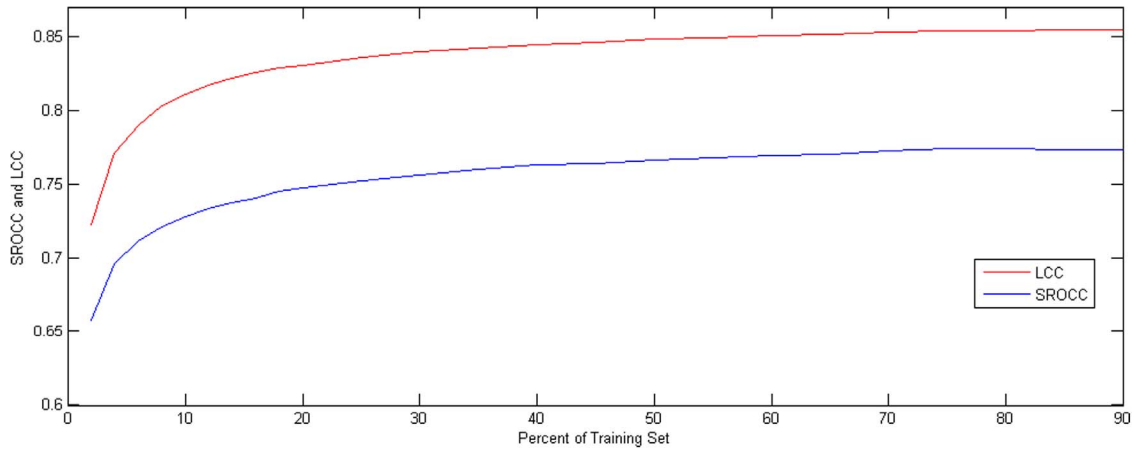
Fig. 8. Mean SROCC and LCC of PUP features against MOS as a function of the percentage of the training set used.

SROCC, and RMSE of the PUP features against MOS over all 1000 train-test trials, as tabulated in Tables 3, 4 and 5, respectively. Values of LCC and SROCC close to 1 mean superior linear and rank correlation with MOS, respectively. Obviously, the higher the mean and median, the better the LCC and SROCC performance. Conversely, a higher standard deviation implies more unstable performance. In order to examine the contribution of features from each PUP map, we also obtained correlation results for the constituent predictive features from PUP-L, PUP-A and PUP-S. For comparison, we also tested the models contributed by Park et al. [15], Nojiri et al. [17], Yano et al. [18], Choi et al. [19] and Kim and Sohn [20]; the same types of features extracted from the Flow algorithm [21], dynamic programming [22] and the DSAD and DSSIM based stereo algorithms [50], with all results tabulated in Tables 3–5.

The SROCC, LCC and RMSE results in the tables indicate that PUP has a high correlation with experienced visual discomfort. PUP ranks 2nd in the comparison and performed better than all of the disparity based algorithms. The predictive power drops dramatically when using simple, low complexity disparity calculations (DSAD and DSSIM).

To test the statistical efficacy of the model against other models, a *F*-tests was conducted to assess the statistical significance of the errors between the MOS scores and the model predictions on the IEEE-SA database. Table 6 shows the results of *F*-tests. The residual error between the predicted score of a model and the corresponding MOS value in the IEEE-SA database can be used to test the statistical efficacy of the model against other models. The residual errors between the model predictions and the MOS values are:

$$R = \{Q_i - MOS_i, i = 1, 2, \ldots, N_T\} \tag{16}$$

where Q_i is the *i*th objective visual discomfort score and MOS_i is the

corresponding *i*th MOS score. The *F*-test was used to compare one objective model against another objective model at the 95% significance level. (The *F*-value is 1.6378 and the degrees of freedom are 159 for both numerator and denominator.) A symbol value of 1 indicates that the statistical performance of the model in the row is better to that of the model in the column, while 0 indicates the performance in the row is worse to that in the column, and – indicates equivalent performance. The results indicate that the PUP map achieves a performance quite comparable to that of Park, and better than that of the other compared models.

Table 4
Mean LCC over 1000 trials of randomly chosen train and test sets on the IEEE-SA database.

LCC	Mean	Median	Standard deviation
Nojiri et al. [17]	0.69	0.69	0.079
Yano et al. [18]	0.40	0.40	0.075
Choi et al. [19]	0.65	0.66	0.070
Kim and Sohn [20]	0.70	0.71	0.077
Richardt et al. [44]	0.70	0.71	0.079
Park et al. [15]	**0.86**	**0.87**	**0.048**
Flow [21]	0.83	0.84	0.051
Dynamic [22]	0.79	0.79	0.062
DSAD	0.61	0.61	0.049
DSSIM [50]	0.70	0.70	0.042
PUP-L	0.78	0.78	0.033
PUP-A	0.80	0.80	0.029
PUP-S	0.83	0.83	0.028
PUP	**0.86**	**0.86**	**0.022**

Table 3
Mean SROCC over 1000 trials of randomly chosen train and test sets on the IEEE-SA database.

SROCC	Mean	Median	Standard deviation
Nojiri et al. [17]	0.61	0.62	0.073
Yano et al. [18]	0.34	0.34	0.073
Choi et al. [19]	0.59	0.59	0.080
Kim and Sohn [20]	0.62	0.62	0.070
Richardt et al. [44]	0.62	0.63	0.073
Park et al. [15]	**0.78**	**0.79**	**0.045**
Flow [21]	0.76	0.76	0.046
Dynamic [22]	0.71	0.72	0.069
DSAD	0.59	0.59	0.049
DSSIM [50]	0.66	0.66	0.042
PUP-L	0.70	0.70	0.039
PUP-A	0.72	0.73	0.037
PUP-S	0.75	0.75	0.037
PUP	**0.78**	**0.78**	**0.033**

Table 5
Mean RMSE over 1000 trials of randomly chosen train and test sets on the IEEE-SA database.

RMSE	Mean	Median	Standard deviation
Nojiri et al. [17]	0.61	0.61	0.085
Yano et al. [18]	0.76	0.76	0.071
Choi et al. [19]	0.71	0.71	0.085
Kim and Sohn [20]	0.58	0.58	0.061
Richardt et al. [44]	0.58	0.58	0.067
Park et al. [15]	**0.39**	**0.40**	**0.025**
Flow [21]	0.44	0.45	0.034
Dynamic [22]	0.52	0.52	0.036
DSAD	0.62	0.62	0.039
DSSIM [50]	0.58	0.57	0.039
PUP-L	0.50	0.51	0.031
PUP-A	0.47	0.47	0.029
PUP-S	0.45	0.45	0.030
PUP	**0.42**	**0.42**	**0.028**

Table 6
Results of the *F*-test preformed on the residuals between objective visual discomfort predictions and MOS values at a significance level of 95%.

	Nojiri	*Yano*	*Choi*	*Kim*	*Richardt*	*Park*	*PUP*
Nojiri	–	1	1	–	–	0	0
Yano	0	–	0	0	0	0	0
Choi	0	1	–	0	0	0	0
Kim	–	1	1	–	–	0	0
Richardt	–	1	1	–	–	0	0
Park	1	1	1	1	1	–	–
PUP	**1**	**1**	**1**	**1**	**1**	–	–

Table 7
Mean outlier ratio over 1000 trials of randomly chosen train and test sets on the IEEE-SA database.

Outlier ratio	Mean	Median	Standard deviation
Nojiri et al. [17]	0.32	0.32	0.033
Yano et al. [18]	0.42	0.43	0.047
Choi et al. [19]	0.33	0.33	0.033
Kim and Sohn [20]	0.32	0.32	0.037
Richardt et al. [44]	0.33	0.33	0.037
Park et al. [15]	**0.15**	**0.15**	**0.021**
Flow [21]	0.18	0.18	0.025
Dynamic [22]	0.23	0.23	0.029
DSAD	0.34	0.34	0.033
DSSIM [50]	0.29	0.30	0.031
PUP-L	0.25	0.25	0.032
PUP-A	0.22	0.23	0.031
PUP-S	0.19	0.19	0.030
PUP	**0.17**	**0.18**	**0.027**

Further, we compared the outlier ratio yielded by our model and the other models. The outlier ratio (OR) is defined as the number of points that fall outside of the 95% confidence intervals:

$$OR = \frac{Total \ no \ outliers}{N} \qquad (17)$$

$$outlier: |MOS_i - MOS_{pi}| > z \times \sigma(MOS_i)/\sqrt{M} \qquad (18)$$

where N is the total number of points, MOS_{pi} is predicted MOS_i and $\sigma(MOS_i)$ is the standard deviation of MOS_i. M is the number of subjects and z is the z-value corresponding to a two-tailed normal distribution with a significance level 0.05, $z = 1.96$. We obtained the mean, median, and standard deviations of outlier ratio of the PUP features against MOS over all 1000 train-test trials, as tabulated in Table 7. The outlier ratio was similar to that of Park et al. [15] and lower than the other methods. Fig. 9 shows the scatter plot of the predicted results of the PUP model of Park et al. [15] and of Yano et al. [18]. From the result we can see that the outlier ratio plots of PUP and Park models are small and that the plots are very similar. The outlier ratio plots of Yano is big.
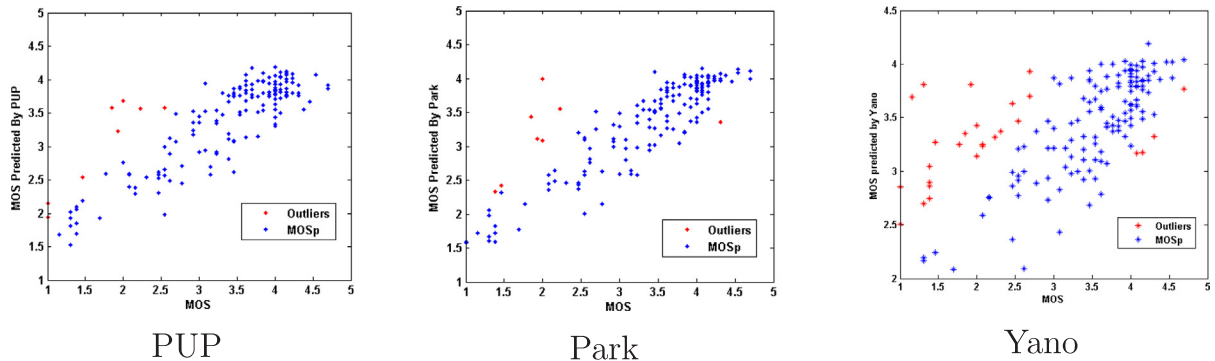
Table 8
Pearson's linear correlation coefficient between features from the four models (PUP-A, Dynamic, DSSIM, and DSAD) and the ground-truth model flow [21].

Features	TL	TS	MP	MN
PUP-A	0.80	0.79	0.79	0.62
Dynamic	0.75	0.76	0.66	0.64
DSSIM	0.41	0.10	0.12	0.46
DSAD	0.35	0.12	0.10	0.30

Features extracted using all three block widths (PUP-L, PUP-A, and PUP-S) correlated well with MOS (PUP-S delivered the best results) and delivered better results than features extracted by using DSAD or DSSIM [50], but worse than features extracted by Flow [21]. That is because Flow [21] delivers highly competitive predictions of disparity as compared to the state of the art on the Middlebury Stereo Evaluation [54], viz., disparities extracted by Flow [21] may be assumed to be close to accurate disparities.

As IEEE-SA does not provide ground truth depth maps, it is difficult to compute how well a PUP map is correlated to a depth map. Instead, we calculated Pearson's linear correlation coefficient between the four features (TL, TS, MP and MN) extracted from PUP-A and Flow [21], which we used as a proxy for ground truth. The correlation coefficients between DSAD, DSSIM [50] and Flow [21] results were also computed and tabulated for comparison. As may be observed from Table 8, the PUP-A features were significantly more correlated with features from Flow [21] than with DSSIM [50] and DSAD features.

The combination of PUP maps with three different block sizes performed much better than prediction using only one PUP map, and also better than prediction using Flow [21]. The low correlations obtained using DSSIM and DSAD, especially for TS and MP, occurred since the errors in their disparity calculation were distributed throughout the entire disparity range. The relationship between the block widths when creating a PUP map and predictive power is illustrated in Fig. 10, where the mean values of SROCC and LCC are plotted as a function of block width over 1000 iterations. The features were extracted from a single PUP map using a fixed patch width.

We also performed additional experiments on the EPFL stereo image database [39] which was designed to test quality of experience models, which is influenced by visual discomfort among other influencing parameters. This database has been used in the development of a few visual discomfort prediction models, in particular [15,14,55,56]. In this database the test subjects were asked to rate quality and not discomfort, so the results are only an indirect indication of model applicability to predict visual discomfort. We did this to address questions regarding possible database dependence. The entire IEEE-SA database was used to train the PUP features, then the model was tested on the EPFL database. Table 9 shows the performance results as well as comparison with the prior work and four disparity algorithms in Tables 4 and 5. The PUP map set delivered very good performance in regards to cross database
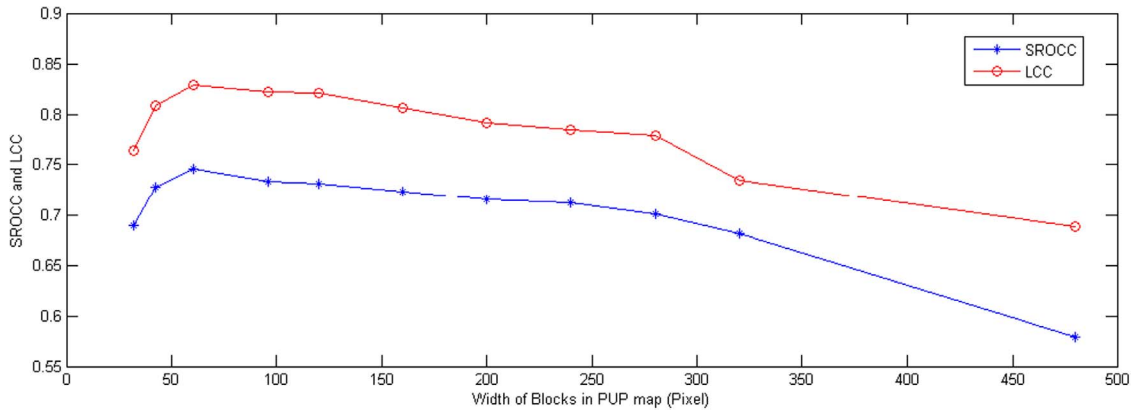

Fig. 9. Scatter plot of predicted results of PUP model, Park et al. [15] and Yano et al. [18].

Fig. 10. Mean SROCC and LCC as a function of block width.

Table 9
LCC, SROCC and RMSE on the EPFL database.

EPFL	SROCC	LCC	RMSE
Nojiri et al. [17]	0.87	0.85	0.48
Yano et al. [18]	0.86	0.84	0.69
Choi et al. [19]	0.82	0.77	0.74
Kim and Sohn [20]	0.88	0.87	0.47
Richard et al. [44]	0.86	0.85	0.69
Park et al. [15]	**0.91**	**0.90**	**0.43**
Flow [21]	0.88	0.87	0.48
Dynamic [22]	0.85	0.85	0.67
DSAD	0.43	0.43	0.78
DSSIM [50]	0.70	0.72	0.75
PUP	**0.92**	**0.91**	**0.44**

prediction and better results than the other models, even though the stereo image capture system, the disparity distributions, the image contents and the display apparatus of the EPFL database are different from those of the IEEE-SA database.

From all the performance evaluation test results, it is concluded that PUP map is efficient in predicting the degree of visual discomfort experienced when viewing S3D images and the computing speed of PUP map extraction is more competitive.

5. Conclusion and future work

We have described a first of a kind visual discomfort model that predicts the degree of visual discomfort experienced when viewing S3D images without computing disparity maps. The model is built based on a new concept that we call the percentage of un-linked pixels map (PUP map). Computation of the PUP map is much faster than most disparity calculation algorithms and the predictive power of PUP outperforms other visual discomfort models that are based on disparity features. The model was tested on both the IEEE-SA database and the EPFL database which used different average viewing distances of 170 cm and 200 cm. The prediction accuracy of the PUP map (or any other model, for that matter) at shorter or longer viewing distances remains an interesting open question. We plan to explore combinations of PUP maps with other image features, such as measurements of defocus-blur and of the detection of disparities along the vertical dimension. We also plan to investigate the temporal dimension of the stereoscopic signal in the future, where time-varying PUP maps would be used to predict the degree of visual discomfort experienced over time when watching S3D television, cinema, or other videos [3,18,26,9]. We also plan to explore using PUP maps to accomplish training-free discomfort prediction, following recent trends in image and video quality assessment [57–59]. As more generally QoE are still not known enough particular indicators to be used robustly for predicting the subjective experience (whether it is comfort, fatigue or any other QoE feature), further tests of PUP map

on predicting generally QoE will be conducted.

References

[1] M. Lambooij, M. Fortuin, I. Heynderickx, W. IJsselsteijn, Visual discomfort and visual fatigue of stereoscopic displays: a review, J. Imag. Sci. Technol. 53 (3) (2009) 1–14.
[2] M. Lambooij, M. Fortuin, W. IJsselsteijn, I. Heynderickx, Measuring visual discomfort associated with 3D displays, in: IS & T/SPIE Electronic Imaging, International Society for Optics and Photonics, Bellingham, WA, 2009, p. 72370K.
[3] S. Yano, M. Emoto, T. Mitsuhashi, Two factors in visual fatigue caused by stereoscopic HDTV images, Displays 25 (4) (2004) 141–150.
[4] R. Patterson, Human factors of 3-D displays, J. Soc. Inf. Disp. 15 (11) (2007) 861–871.
[5] K. Ukai, P.A. Howarth, Visual fatigue caused by viewing stereoscopic motion images: background, theories, and observations, Displays 29 (2) (2008) 106–116.
[6] Y. Okada, K. Ukai, J.S. Wolffsohn, B. Gilmartin, A. Iijima, T. Bando, Target spatial frequency determines the response to conflicting defocus-and convergence-driven accommodative stimuli, Vis. Res. 46 (4) (2006) 475–484.
[7] M. Torii, Y. Okada, K. Ukai, J.S. Wolffsohn, B. Gilmartin, Dynamic measurement of accommodative responses while viewing stereoscopic images, J. Mod. Opt. 55 (4–5) (2008) 557–567.
[8] F.L. Kooi, A. Toet, Visual comfort of binocular and 3D displays, Displays 25 (2) (2004) 99–108.
[9] Y. Nojiri, H. Yamanoue, S. Ide, S. Yano, F. Okana, Parallax distribution and visual comfort on stereoscopic HDTV, in: Proc. IBC, vol. 373, 2006.
[10] E.A. Cooper, J. Burge, M.S. Banks, The vertical horopter is not adaptable, but it may be adaptive, J. Vis. 11 (3) (2011) 20.
[11] S.-I. Lee, Y.J. Jung, H. Sohn, Y.M. Ro, Subjective assessment of visual discomfort induced by binocular disparity and stimulus width in stereoscopic image, in: IS & T/SPIE Electronic Imaging, International Society for Optics and Photonics, Bellingham, WA, 2013, p. 86481T.
[12] W.J. Tam, F. Speranza, S. Yano, K. Shimono, H. Ono, Stereoscopic 3D-TV: visual comfort, IEEE Trans. Broadcast. 57 (2) (2011) 335–346.
[13] R. Patterson, Review paper: Human factors of stereo displays: an update, J. Soc. Inf. Disp. 17 (12) (2009) 987–996.
[14] J. Park, H. Oh, S. Lee, A.C. Bovik, 3D visual discomfort predictor: analysis of disparity and neural activity statistics, IEEE Trans. Image Process. 24 (3) (2015) 1101–1114.
[15] J. Park, S. Lee, A.C. Bovik, 3D visual discomfort prediction: vergence, foveation, and the physiological optics of accommodation, IEEE J. Sel. Top. Signal Process. 8 (3) (2014) 415–427.
[16] H. Sohn, Y.J. Jung, S.-i. Lee, Y.M. Ro, Predicting visual discomfort using object size and disparity information in stereoscopic images, IEEE Trans. Broadcast. 59 (1) (2013) 28–37.
[17] Y. Nojiri, H. Yamanoue, A. Hanazato, F. Okano, Measurement of parallax distribution and its application to the analysis of visual comfort for stereoscopic HDTV, in: Electronic Imaging 2003, International Society for Optics and Photonics, Bellingham, WA, 2003, pp. 195–205.
[18] S. Yano, S. Ide, T. Mitsuhashi, H. Thwaites, A study of visual fatigue and visual comfort for 3D HDTV/HDTV images, Displays 23 (4) (2002) 191–201.
[19] J. Choi, D. Kim, S. Choi, K. Sohn, Visual fatigue modeling and analysis for stereoscopic video, Opt. Eng. 51 (1) (2012) 017206-1.

[20] D. Kim, K. Sohn, Visual fatigue prediction for stereoscopic image, IEEE Trans. Circuits Syst. Video Technol. 21 (2) (2011) 231–236.

[21] D. Sun, S. Roth, M.J. Black, Secrets of optical flow estimation and their principles, in: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, San Francisco, La, 2010, pp. 2432–2439.

[22] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, Int. J. Comput. Vis. 47 (1–3) (2002) 7–42.

[23] M. Tanimoto, T. Fujii, K. Suzuki, N. Fukushima, Y. Mori, Depth Estimation Reference Software (DERS) 5.0, ISO/IEC JTC1/SC29/WG11 M 16923 (2009), 2009.

[24] D. Stransky, L. Wilcox, Depth magnitude and binocular disparity: a closer look at patent vs. qualitative stereopsis, J. Vis. 10 (7) (2010) 330.

[25] L.M. Wilcox, R.S. Allison, Coarse-fine dichotomies in human stereopsis, Vis. Res. 49 (22) (2009) 2653–2665.

[26] M. Emoto, T. Niida, F. Okano, Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television, J. Disp. Technol. 1 (2) (2005) 328–340.

[27] A.M. Horwood, P.M. Riddell, The use of cues to convergence and accommodation in naïve, uninstructed participants, Vis. Res. 48 (15) (2008) 1613–1624.

[28] I. Howard, B. Rogers, Perceiving in Depth, Volume 2: Stereoscopic Vision, 2012.

[29] I.P. Howard, Seeing in Depth, Volume 1: Basic Mechanisms, University of Toronto Press, Toronto, 2002.

[30] G. Sperling, Binocular vision: a physical and a neural theory, Am. J. Psychol. (1970) 461–534.

[31] Standard for the Quality Assessment of Three Dimensional (3D) Displays, 3D Contents and 3D Devices Based on Human Factors, IEEE P3333.1 2012, doi:http://grouper.ieee.org/groups/3dhf.

[32] A.C. Bovik, M. Clark, W.S. Geisler, Multichannel texture analysis using localized spatial filters, IEEE Trans. Pattern Anal. Mach. Intell. 12 (1) (1990) 55–73.

[33] M. Clark, A.C. Bovik, Experiments in segmenting texton patterns using localized spatial filters, Pattern Recognit. 22 (6) (1989) 707–717.

[34] C.-C. Su, A.C. Bovik, L.K. Cormack, Natural scene statistics of color and range, in: 2011 18th IEEE International Conference on Image Processing (ICIP), IEEE, Brussels, Belgium, 2011, pp. 257–260.

[35] S. Ide, H. Yamanoue, M. Okui, F. Okano, M. Bitou, N. Terashima, Parallax distribution for ease of viewing in stereoscopic HDTV, in: Electronic Imaging 2002, International Society for Optics and Photonics, Bellingham, WA, 2002, pp. 38–45.

[36] J. Park, K. Seshadrinathan, S. Lee, A.C. Bovik, Video quality pooling adaptive to perceptual distortion severity, IEEE Trans. Image Process. 22 (2) (2013) 610–620.

[37] A.K. Moorthy, A.C. Bovik, Visual importance pooling for image quality assessment, IEEE J. Sel. Top. Signal Process. 3 (2) (2009) 193–201.

[38] ITU-R BT.1438, Subjective Assessment of Stereoscopic Television Pictures. 2000.

[39] L. Goldmann, F. De Simone, T. Ebrahimi, Impact of acquisition distortion on the quality of stereoscopic images, in: Proceedings of the International Workshop on Video Processing and Quality Metrics for Consumer Electronics, no. MMSPL-CONF-2009-022, 2010.

[40] Y.-Y. Yeh, L.D. Silverstein, Limits of fusion and depth judgment in stereoscopic color displays, Hum. Factors: J. Hum. Factors Ergon. Soc. 32 (1) (1990) 45–60.

[41] A.S. Percival, The relation of convergence to accommodation and its practical bearing, Ophthal. Rev. 11 (1892) 313–328.

[42] F. Speranza, W.J. Tam, R. Renaud, N. Hur, Effect of disparity and motion on visual comfort of stereoscopic images, in: Electronic Imaging 2006, International Society for Optics and Photonics, Bellingham, WA, 2006, p. 60550B.

[43] M. Wöpking, Viewing comfort with stereoscopic pictures: an experimental study on the subjective effects of disparity magnitude and depth of focus, J. Soc. Inf. Disp. 3 (3) (1995) 101–103.

[44] C. Richardt, L. Świrski, I.P. Davies, N.A. Dodgson, Predicting stereoscopic viewing comfort using a coherence-based computational model, in: Proceedings of the International Symposium on Computational Aesthetics in Graphics, Visualization, and Imaging, ACM, Vancouver, BC, Canada, 2011, pp. 97–104.

[45] D. Kim, D. Min, J. Oh, S. Jeon, K. Sohn, Depth map quality metric for three-dimensional video, in: IS & T/SPIE Electronic Imaging, International Society for Optics and Photonics, Bellingham, WA, 2009, p. 723719.

[46] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.

[47] H. Kim, Y. Choe, K. Sohn, Disparity estimation using a region-dividing technique and energy-based regularization, Opt. Eng. 43 (8) (2004) 1882–1890.

[48] H.R. Filippini, M.S. Banks, Limits of stereopsis explained by local cross-correlation, J. Vis. 9 (1) (2009) 8.

[49] G. Egnal, R.P. Wildes, Detecting binocular half-occlusions: empirical comparisons of five approaches, IEEE Trans. Pattern Anal. Mach. Intell. 24 (8) (2002) 1127–1133.

[50] M.-J. Chen, C.-C. Su, D.-K. Kwon, L.K. Cormack, A.C. Bovik, Full-reference quality assessment of stereopairs accounting for rivalry, Signal Process.: Image Commun. 28 (9) (2013) 1143–1155.

[51] C.-C. Chang, C.-J. Lin, LibSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. (TIST) 2 (3) (2011) 27.

[52] ITU-T P.1401, Statistical Analysis, Evaluation and Reporting Guidelines of Quality Measurements. 2012.

[53] V.Q.E. Group, et al., Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II (fr_tv2), ftp://ftp.its.bldrdoc.gov/dist/ituvidq/Boulder_VQEG_jan_04/VQEG_PhaseII_FRTV_Final_Report_SG9060E.doc, 2003.

[54] D. Scharstein, R. Szeliski, Middlebury Stereo Evaluation – Version 2, vision.middlebury.edu/stereo. 2006.

[55] A. Mittal, A.K. Moorthy, J. Ghosh, A.C. Bovik, Algorithmic assessment of 3D quality of experience for images and videos, in: 2011 IEEE Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE), IEEE, Sedona, AZ, 2011, pp. 338–343.

[56] J.-S. Lee, L. Goldmann, T. Ebrahimi, A new analysis method for paired comparison and its application to 3D quality assessment, in: Proceedings of the 19th ACM International Conference on Multimedia, ACM, Scottsdale, Arizona, 2011, pp. 1281–1284.

[57] A. Mittal, G.S. Muralidhar, J. Ghosh, A.C. Bovik, Blind image quality assessment without human training using latent quality factors, IEEE Signal Process. Lett. 19 (2) (2012) 75–78.

[58] A. Mittal, R. Soundararajan, A.C. Bovik, Making a completely blind image quality analyzer, IEEE Signal Process. Lett. 20 (3) (2013) 209–212.

[59] A. Mittal, M.A. Saad, A.C. Bovik, A completely blind video integrity oracle, IEEE Trans. Image Process. 25 (1) (2016) 289–300.