Robust Stereoscopic Crosstalk Prediction

Jianbing Shen, *Senior Member, IEEE*, Yan Zhang, Zhiyuan Liang, Chang Liu, Hanqiu Sun, *Member, IEEE*, Xiaopeng Hao, Jianhong Liu, Jian Yang, and Ling Shao, *Senior Member, IEEE*

***Abstract*— We propose a new metric to predict perceived crosstalk using the original images rather than both the original and ghosted images. The proposed metrics are based on color information. First, we extract a disparity map, a color difference map, and a color contrast map from original image pairs. Then, we use those maps to construct two new metrics ($V_{\mathbf{dispc}}$ and $V_{\mathbf{dlogc}}$). Metric $V_{\mathbf{dispc}}$ considers the effect of the disparity map and the color difference map, while $V_{\mathbf{dlogc}}$ addresses the influence of the color contrast map. The prediction performance is evaluated using various types of stereoscopic crosstalk images. By incorporating $V_{\mathbf{dispc}}$ and $V_{\mathbf{dlogc}}$, the new metric $V_{\mathbf{pdlc}}$ is proposed to achieve a higher correlation with the perceived subject crosstalk scores. Experimental results show that the new metrics achieve better performance than previous methods, which indicate that color information is one key factor for crosstalk visible prediction. Furthermore, we construct a new data set to evaluate our new metrics.**

***Index Terms*— Color contrast information, crosstalk perception, disparity map, objective metric.**

I. Introduction

With dramatic advances in the modern display devices, the 3D display technology has been widely used. Since our eyes are located in two different positions on the head, we perceive slightly different information from left and right views. This difference between the left image and the right image allows the human visual system (HVS) to perceive the relative depth of objects. Based on the processes of the HVS, the stereoscopic 3D techniques deliver two offset images for respective eyes to make the end user perceive a more realistic scene. Although the 3D display technology has been rapidly developed in recent years [3], [7], [19], [22], [24], [25], [31], [32], many issues remain open challenges, such as the visual fatigue and the visual uncomfortableness. These phenomena arise from the conflict between the accommodation and convergence of human eyes. Crosstalk is one of the most serious problems related to the imperfect separation in a stereoscopic 3D display, where the image for one eye remains dimly visible to the other eye. Crosstalk can be perceived as ghosts, shadows, or double contours, resulting in the degradation of the image quality.

The mechanisms by which crosstalk occurs vary between different stereoscopic display technologies [1], [2], [4]–[6], [40], [41]. In time-sequential display systems, the screen displays the left-eye and right-eye images alternatively at high frame rates. The viewer can wear a pair of glasses, which block each eye in an alternating fashion, synchronizing to the content being displayed. The main factors to crosstalk in these systems are slow shuttering, shutter leakage, and persistence of the image [1], [5], [7]. In multiplexed spectrum systems, the polarization state of one eye image is orthogonal to that of the other eye image. The eye-wearing appropriate polarizers have been used to separate the stereoimages by blocking the image, which is not intended for that eye. The main factors to crosstalk are the imperfect spectral performance of the filters and the mismatch with the spectral emission of the displays [1], [6], [7]. In color-multiplexed stereoscopic displays, the most common anaglyph method uses different channels for each eye (e.g., the red channel for the left eye and the cyan channel for the right eye). The views for the left and right eyes can be separated by wearing a pair of colored glasses. The spectral response of the display and the anaglyph glasses has been cited as the main source of crosstalk [1], [7]. Besides, crosstalk can also occur during the stereoscopic image acquisition stage and the manipulation stage.

In order to reduce the amount of perceived crosstalk with a particular stereoscopic display, it is necessary to perform a detailed analysis of these mechanisms. The analysis can help to characterize and measure the effect of the components to crosstalk in many domains (temporal, spatial, and spectral). Therefore, crosstalk can be reduced by adjusting one or more of these components [7]. Since crosstalk of displays cannot be eliminated completely with current display devices, researchers attempt to use image processing methods to conceal crosstalk before display, which is also known as crosstalk cancelation [8]–[10]. The methods of crosstalk cancelation hide the visibility of crosstalk by subtracting the amount of leakage from the intended view. The image we perceive is the results of the modified intended view plus the leakage from the unintended image. It is equivalent to

This work was sup- ported in part by the National Basic Research Program of China (973 Program) under Grant 2013CB328805; in part by the National Natural Science Foun- dation of China under Grant 61272359, Grant 61672099, Grant 61527827, Grant 61379087, and Grant 61602183; in part by the RGC Research under Grant 416212; in part by the UGC Direct Grant for Research under Grant 4055060; in part by the Fok Ying-Tong Education Foundation for Young Teachers; and in part by the Specialized Fund for Joint Building Program of the Beijing Municipal Education Commission.

J. Shen, Y. Zhang, Z. Liang, C. Liu, X. Hao, and J. Liu are with the Beijing Key Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing 100081, China (e-mail: shenjianbing@bit.edu.cn).

H. Sun is with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong (e-mail: hanqiu@cse.cuhk.edu.hk).

J. Yang is with the Beijing Engineering Research Center of Mixed Reality and Advanced Display, School of Optics and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: jyang@bit.edu.cn).

L. Shao is with the School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, U.K. (e-mail: ling.shao@ieee.org).

the original intended image. However, we have to note that crosstalk cancelation does not always work effectively in all situations [7].

There is already a substantial amount of literatures on the perceptual consequences of crosstalk. In [4], experiments address the effect of crosstalk on the perceived magnitude of depth from two aspects, disparity, and monocular occlusion. The results show that even at fairly low levels, the perceived depth is significantly reduced from both cues. Besides, ghosting from crosstalk is implicated as a major factor of influencing visual comfort [11], [12]. Moreover, WIlcox and Stewart [12] reported that the crosstalk has an important influence on determining the image quality for 75% of their observers, and they also found that crosstalk over 5% cause visual comfort reduction.

Therefore, it is beneficial to study the quantitative measurement of crosstalk. There is far less existing work on this topic. Among them, the crosstalk is reported more annoying in regions with high contrast, large disparity, and sharp edges in still images [11], [13]. As a consequence, crosstalk will be more visible when the contrast and disparity of the image increase. In [13], an acceptability threshold of crosstalk is provided. This threshold is examined by computer-generated static images with several levels of image contrast and binocular disparity. Luminance comparison between the ghost image and the original image is often used in most of the existing crosstalk metrics [6], [7], [11]. However, the calculated crosstalk values were not always consistent with the perception of the observer. In [14], some 3D crosstalk metrics are defined using either the CIE uniform color coordinate or the gray scale level instead of the luminance, because these quantities are known to agree better with human perception. However, only some pure color patches are used to measure the effect of color on crosstalk in their experiments. Besides, Seuntëens *et al.* [11] use two similar natural scenes with varying crosstalk levels (0%, 5%, 10%, and 15%) and camera baselines (0, 4, and 12 cm) to investigate the effect of crosstalk on perceived image distortion, perceived depth, and visual discomfort. Xing and You [15] give a detailed analysis on the effect of 2D and 3D perceptual attributes on crosstalk. Then, they integrate the structural SIMilarity map (SSIM) and the filtered depth map to build objective metric for crosstalk perception. Although the metric in [15] could be used to predict the subjective judgment by humans with a high correlation, the authors use the comparison between the crosstalk images and the original images, which is not useful in practice. For instance, they have to synthesize crosstalk images from original pairs, which would introduce errors inevitably.

This paper proposes a new objective metric, which can better represent the subjective judgment of humans. To the best of our knowledge, the existing methods use the difference (e.g., the SSIM map) between the ghost images and the original images. In contrast, we use disparity information and color contrast information between the original left and right images to reflect the perceived crosstalk. Besides, the crosstalk level (p) of a special stereoscopic 3D display can be measured by using the four cross combinations of full white and full black in the left-eye and right-eye channels. There are also some other methods to measure crosstalk in different stereoscopic displays [6]. Once we get the parameter (p), we can use it to measure the visibility of crosstalk for all images displayed in the special device. We test the proposed metrics on both the data set in [16] and our new data set. The data set provided in [16] contains seven image scenes with four different crosstalk levels and three different camera baselines, while our new data set consists of 23 nature image scenes with four different crosstalk levels (0%, 5%, 10%, and 15%), which have various color information and depth structures. The experimental results demonstrate that our metric has a higher correlation with the objective judgment. On the other hand, we can see that the metrics using color information are more effective than the existing metrics using structural information. It indicates that color contrast information is more important for crosstalk perception. Our source code and supplement materials will be available at http://github.com/shenjianbing/crosstalk

To summarize, this paper has the following contributions.

1) We propose new metrics to predict the perceived crosstalk using the color information of original images. Our new metrics can predict the perceived crosstalk using original images rather both the original and the ghosted images.
2) A new data set is constructed to contain more natural image scenes, rich color information, depth structures, and complicated background. The prediction performance of those metrics is evaluated using various types of stereoscopic crosstalk images in our new data set.
3) We have trained a support vector regression (SVR) to measure the stereoscopic crosstalk prediction. We also invite subjects and train them to collect subjective crosstalk visibility scores, and provide a more challenging and complete data set for the future research.

II. Related Work

In this section, we first introduce how we measure system-introduced crosstalk and synthesize crosstalk images with different crosstalk levels. Then, we review the main works with crosstalk in recent years, which provide us clues to choose features for crosstalk visibility prediction.

There are two main methods, which measure the crosstalk in stereoscopic displays. One uses optical sensors and the other uses visual measurement charts [7]. Maximum crosstalk often occurs when the left-eye and right-eye images have the maximum difference in brightness. So, the traditional measure of crosstalk is displaying full black and full white in left-eye and right-eye channels, and using an optical sensor to measure the amount of leakage between channels. In this metric, four cross combinations of full white and full black in each eye channel have been used, and the system-introduced crosstalk can be modeled as follows [7], [15], [17] :

$$\begin{cases} CL = \dfrac{(BW - BB)}{(WB - BB)} \\ CR = \dfrac{(WB - BB)}{(BW - BB)} \end{cases} \tag{1}$$

where WB means the brightness when displaying full white in the left-eye channel and full black in the right-eye channel,

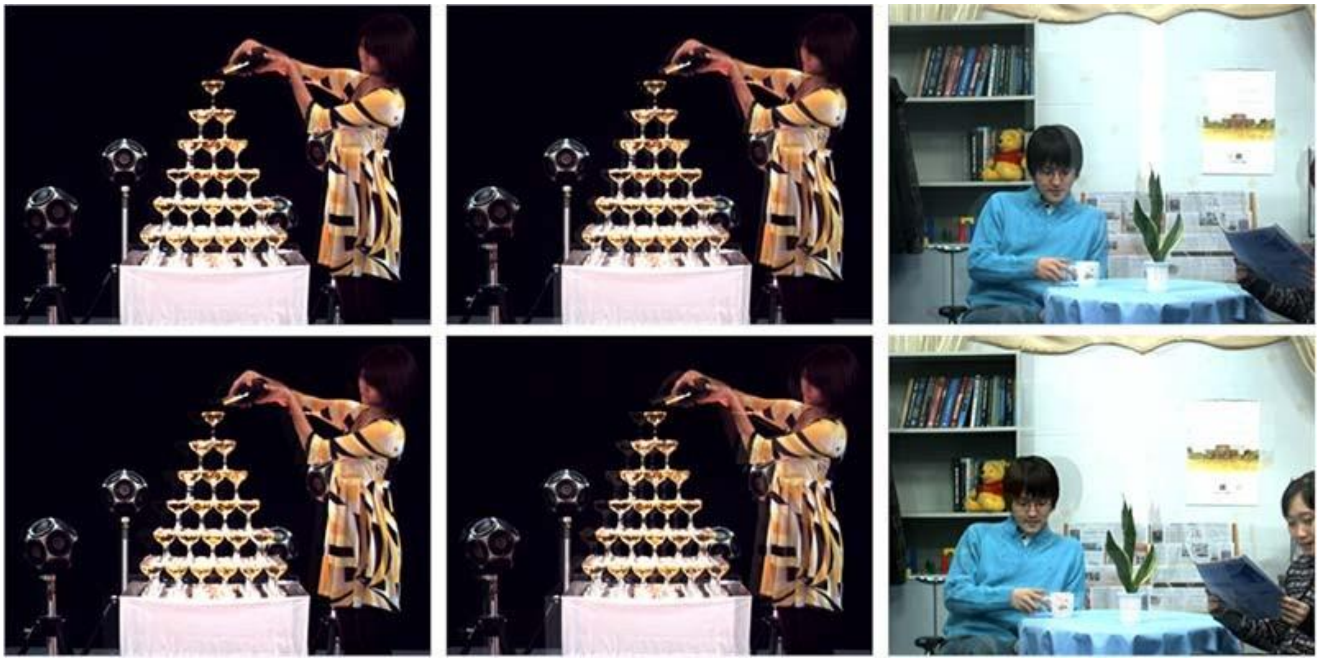


Fig. 1. Left images for Champagne and Newspaper with different crosstalk visibility levels, which is impacted by the crosstalk level and the camera baseline. The crosstalk level is 8% and 18% in the first two columns. The camera baseline is 50 mm in the top row and 100 mm in the bottom row.

BW denotes the brightness when using black as the left image and white as the right image, BB is the brightness when using full black as both eye images, and WW represents the brightness when displaying full white in both channels. CL and CR are the ghosting images for the left and right eyes.

For time-sequential 3D LCDs, the metric recently proposed to measure crosstalk is gray-to-gray crosstalk [18]. In this method, crosstalk mainly occurs from the slow response time of liquid crystal, which is determined by the different gray level changes. Though using visual measurement to evaluate crosstalk is very quick and effective, there are still some limitations. Once the parameter *p* (crosstalk level) of a specific device has been measured by appropriate method, we can use it as a main factor to predict crosstalk visibility with this device.

Conventionally, various crosstalk metrics have been proposed to quantify the crosstalk effect using the luminance channel alone [7], [13]. In [13], the visibility and acceptability threshold of crosstalk is formalized using contrast and binocular disparity, where the display contrast is signed as the ratio between the luminance difference and the background luminance. In their experiments, the crosstalk images are generated by a computer based on the hypothesis that the crosstalk is the luminance leakage from one eye channel to the other eye channel. Their results show that the crosstalk visibility decreases with the increment of the contrast and disparity. Moreover, there exists various crosstalk cancelation methods using this mathematical model [8], [10], [20], [21], [37]. However, the color information is also important to predict the crosstalk visibility. In [14] and [24], the CIE XYZ tristimulus values are used to characterize crosstalk models.

In stereoscopic 3D televisions with shutter glasses, the left and right images are independent during the process of 3D image generation. Based on that assumption, the computer-generated color patches used in [23] satisfy the left and right views' additivity. According to the results of the experiments, three different crosstalk characterization models are proposed. They are liner combinations of the original color and the leaked light from the opposite view. Moreover, various crosstalk cancelation technologies have been proposed in [9] and [10] using RGB channels. As mentioned before, we adopt the definition of system-introduced crosstalk as the unexpected light leakage of the image from one eye to the other, which is often perceived as ghosts, shadows, and double contours. A high level of the system crosstalk can significantly degrade the quality of stereoscopic images and cause visual discomfort because of the mismatches of color, luminance, and structure [8], [26]. The data set provided by [15] and [16] contains 72 test images. They are generated by using various crosstalk levels (3%, 8%, 13%, and 18%) to six natural scene image pairs acquired either indoor or outdoor using three different camera baselines. In this data set, the crosstalk model in a stereoscopic 3D display has been chosen as follows [15], [16]:

$$L_c = L_o + p \cdot R_o, \qquad R_c = R_o + p \cdot L_o \tag{2}$$

where p is the overall crosstalk level by combining the system-introduced crosstalk level and simulated crosstalk level. L_o and R_o denote the luminance for the left-eye and right-eye images without crosstalk, respectively, while L_c and R_c represent the light reaching the viewer's eyes.

The most reliable approach to evaluate the perceived crosstalk is the subjective testing, but it is not applicable in practice. We need to design a computational method, which is more suitable in practice and has a high correlation with the subjective judgment of humans. In this paper, we propose a new crosstalk metric using crosstalk level (p), disparity information, and color information. We test our metric on our new data set and the data set provided in [15] and [16], and the results show that our metrics have a higher correlation with the objective perceived values than the existing methods.

III. Proposed Method

Crosstalk is one of the main factors that degrade the quality of the stereoscopic images. In the previous literature, it has been well known that the perceptual crosstalk is influenced by crosstalk level [11], camera baseline [11], [27], and scene content [15], [27], [28]. We notice that color information plays an important role in crosstalk perception detection. Based on this observation, we develop two new metrics to define the objective perceptual metric, using either the color difference information or the color contrast information of the original stereopairs. Furthermore, we construct a new comprehensive metric by combining these two metrics together, which has a better performance.

A. Crosstalk Level and Color Difference Map

As shown in Fig. 1, crosstalk is more annoying, when we increase the crosstalk level p from 8% in the first column to 18% in the second column, which implies that the crosstalk level is one of the main factors affecting the crosstalk perception. When we increase p, there is more light leakage from the unintended image to the intended image, which increases the shadow degree in turn. From Fig. 1, we can also see that crosstalk is more annoying in the areas surrounding the edges. For one reason, the colors of the objects in the foreground and background regions are different. In the regions surrounding the edges, the colors of the pixels in the same position of the left and right images are different, so we can easily distinguish the crosstalk. Given the same amount of light leakage, crosstalk could be more serious in the regions, where different colors occur in the same position of the left and right images. On the other hand, we could hardly see the crosstalk in the regions, where the colors are the same. It has been generally agreed that the crosstalk visibility also increases when the image contrast increases with a certain disparity [8], [14], [15], [33]. In Fig. 2(c), crosstalk in the red regions is more serious than that in the green regions. Intuitively, when given the crosstalk level (p), we have to tolerate more crosstalk in the regions with a bigger color difference (e.g., the red regions). The color difference map we used is the maximum difference of the color channels R, G, and B as

$$\text{deta\_map} = \text{MAX}\left(|R_r - R_l|, |G_r - G_l|, |B_r - B_l|\right) \tag{3}$$

where the MAX operation is used to choose the biggest difference from three channels for every pixel. R_l, G_l, and B_l denote the pixel values in R, G, and B channels of the left view respectively, and R_r , G_r , B_r are the color values in the right image. deta_map is the color difference map.

The regions around the edges have a larger color difference in Fig. 2(b). Furthermore, the crosstalk is more serious in the regions with a larger color difference from Fig. 2(c).

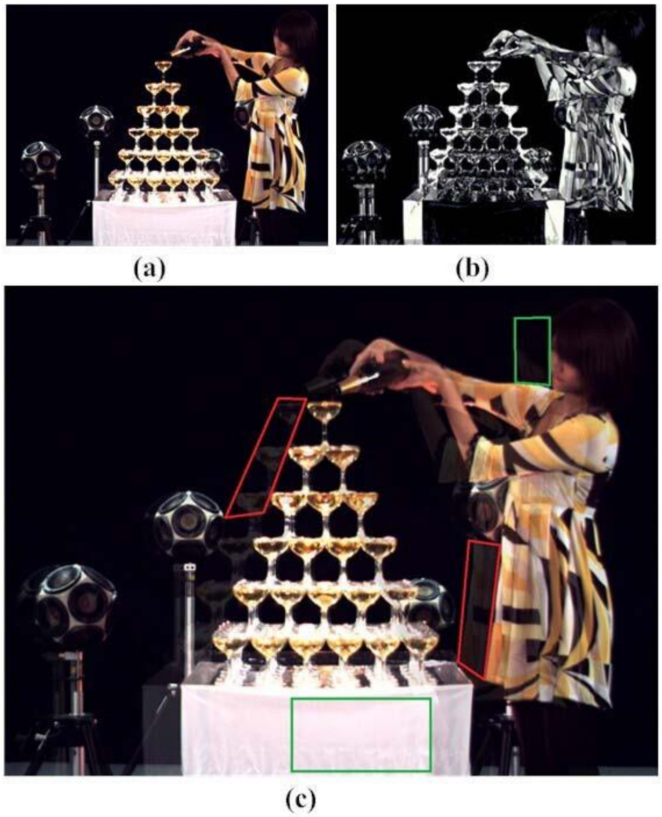


Fig. 2. Influence of image color information to the crosstalk visibility. (a) Original left image. (b) Color difference image. (c) Crosstalk image, where crosstalk is more visible around high contrast regions (red) than low contrast regions (green).

Though the pixels in the bottom green region [Fig. 2(c)] have more light leakage, the crosstalk is almost invisible. This is because the colors of these pixels are similar. On the other hand, the color difference of the pixels in the top green region is large, but the perceived crosstalk is slight. This is caused by less light leakage in that region.

B. Disparity Map

When viewing each column from top to bottom in Fig. 1, we can see that the crosstalk becomes more annoying with an increased separation distance for the same proportion of leakage. The camera baseline becomes larger from top to bottom in Fig. 1. It indicates that the camera baseline has an impact on the separation distance of crosstalk. Furthermore, it is obvious that the separation distance of crosstalk varies from the foreground regions to the background regions for different depth structures. When comparing the images of Champagne with Newspaper, the separation distance is also different. Actually, the separation distance of crosstalk is determined by both camera baseline and relative depth structure together, which is also known as disparity. Therefore, crosstalk level, color information, and disparity map can be used to construct the metric for crosstalk perception.

We adopt the algorithm in [30] to compute the disparity map in this paper. Fig. 3 shows the disparity maps of Love Bird, where the pixel value 0 (black) denotes the smallest disparity while 255 (white) means the biggest disparity.

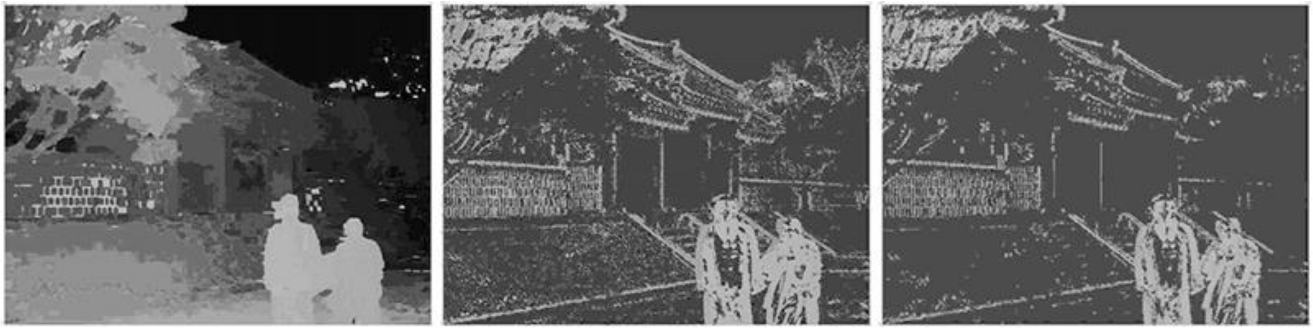


Fig. 3. Disparity maps and filtered disparity maps of Love Bird (CB = 100 mm and p = 8%). From left to right: disparity maps, filtered maps by (4), and filtered maps by color difference maps by (5). We enhance these images using histogram equalization for better visualization.

The nearer objects have larger disparity values, but the objects of background have smaller disparity values. We can also observe that there are some wrong disparity values in the background caused by the mismatching between the left and right images, and most of these wrong regions have similar structures.

As mentioned before, disparity is one of the main factors impacting the crosstalk visibility. Because our eyes pay more attention to the foreground objects, the regions with larger disparity values have a higher impact on the crosstalk perception, especially for the regions with high contrast in the foreground. However, the disparity maps we obtain by using the existing methods have some regions with wrong values. For instance, some regions in Fig. 3 have bigger disparity values in the background while they are supposed to have smaller values. The disparity maps should be filtered before they are used to compute the perceived crosstalk. We know that the crosstalk is almost invisible in the regions with a little leakage, so we can first set the values to zero in these regions. Besides, the background regions have similar colors between the left and right images, for the shift values in these regions are small. Moreover, crosstalk can be hidden in the regions with a small color difference between the stereopairs, whereas it will be more annoying in the regions with a large color difference and disparity. Thus, the disparity map can be filtered by using the color difference map, which means that smaller weights are assigned to the regions with smaller color differences. The filtered maps are shown in the middle and right images of Fig. 3, where the pixel value 0 (black) means no perceived crosstalk while 1 (white) denotes serious crosstalk. The filtered disparity map is formulated as

$$R_{\text{disf}}(i, j) = \begin{cases} 0, & \text{if } p * \text{deta\_map}(i, j) < \theta \\ R_{\text{dis}}(i, j), & \text{otherwise} \end{cases} \tag{4}$$

where i and j are the pixel indices, and R_{dis} denotes the disparity map, $R_{dis\ f}$ is the filtered disparity map by setting zero to the pixels with little leakage. We use (4) to obtain the filtered disparity map, and we empirically set the threshold $\theta = 6$ in (4) according to our experiment. The filtered disparity maps are shown in the middle column of Fig. 3, where the pixel values of the regions with similar colors are set to zero (black) so as to reduce the influence of crosstalk evaluation.

The difference map is first normalized into the interval [0,1] by dividing the maximum color difference 255, and then, we use it to filter the result map by (4). We then use the filtered map and crosstalk level to build a new metric as

$$\text{disc\_map} = \frac{\text{deta\_map}}{255} \cdot * R_{\text{disf}}$$
$$V_{\text{dispc}} = \text{AVG}(\text{disc\_map}) \times (\sqrt{p}) \tag{5}$$

where deta_map is the color difference map built by (3) and disc_map denotes the filtered disparity map using the color difference map, V_{dispc} is one of the new metrics we propose, p means crosstalk level, and AVG denotes the average operation. disk_map and crosstalk p are two main factors in our experiment, and they have influence on the performance of metric V_{dispc}. In order to balance the impact of these two factors, we use the square root operator for variable p.

C. Contrast Map

Crosstalk is more annoying with a dark intended region than a bright intended region when given the amount of light leakage. The reason is that the relative intensity difference from an original dark region is larger than the value from a bright one [8]. For example, given the light leakage 50, the perceived intensity is 60 in the regions with intended intensity 10, while the perceived intensity is 200 in the regions with original intensity 150. But, crosstalk in the darker region is more annoying than in the brighter region. Weber's law states that subjective sensation is proportional to the logarithm of the stimulus intensity [34]. Similar to the phenomenon in luminance channel, crosstalk is more visible in the regions with a larger color changing amplitude. The contract map is constructed by using the values in R, G, and B channels as

$$\text{dlog\_map} = \frac{\text{deta\_map}}{\log_{10}(R_l + G_l + B_l + a)}$$
$$V_{\text{dlogc}} = \text{AVG}(\text{dlog\_map}) \times (\sqrt{p}) \tag{6}$$

where dlog_map is the contrast map, and a small value means slight influence on crosstalk perception, and vice versa. We empirically set $a = 30$ in our experiments to ensure the value of $\log_{10}(R_l + G_l + B_l + a)$ is positive. In order to ensure the range of values in *dlog_map* to be the same as those of disc_map in (5), we use the log-sigmoid function in (6).

p denotes the crosstalk level and it is a parameter to simulate the real-world stereoscopic images by (9). We use four different values of p (0.1%, 5.1%, 10.1%, and 15.1%) to simulate the real-world stereoscopic artifacts and denote the crosstalk level of these stereoscopic pairs. V_{dlogc} is one of our new metrics to measure the crosstalk visibility. As shown in Fig. 3, the regions with larger intended values have smaller influence on crosstalk visibility with the given intensity difference. The regions with smaller intensity differences have a smaller influence as well.

As shown in Tables II and III, we can see that the metrics V_{dispc} and V_{dlogc} perform better than the existing metrics, which indicate that both disc_map and dlog_map have important influences on crosstalk perception. But, the metric V_{dispc} cannot reflect the influence of the contrast map to crosstalk perception. In addition, the metric V_{dlogc} does not use the information of the disparity map. We can further combine these two metrics to construct a new metric, which performs better. The two metrics are combined as

$$V_{\mathrm{pdlc}} = \beta \times V_{\mathrm{dispc}} + (1 - \beta) \times V_{\mathrm{dlogc}} \tag{7}$$

where V_{dispc} is the metric we proposed by using the disparity map, color difference information, and crosstalk level in (5), and V_{dlogc} is the metric constructed by the color contrast map and crosstalk level in (6). β is the weighting factor to balance the contributions of disc_map in (5) and dlog_map in (6). We set $\beta = 0.1$ in our experiment.

D. Crosstalk Prediction

We use the SVR ($z - SVR$) to predict the perceived crosstalk. The $z - SVR$ is expressed as

$$y = w \cdot \varphi(x) + b + z \tag{8}$$

where x is the feature (such as V_{dispc}, V_{dlogc}, or V_{pdlc}). $\varphi()$ is the kernel function, and we chose radial basis function in this paper. z is a margin of tolerance. The z-SVR model is trained using training samples (x_i and y_i). After we get the parameters w and b from the training session, we can predict the value y using the z-SVR model and input feature x .

In our experiments, the prediction performance of our proposed metrics is evaluated by a 100 times tenfold cross validation. At training stage, the prediction function, which defines the relationship between the MOS$_p$ of our new metrics and the mean opinion scores (MOSs) of the crosstalk perception, is constructed using training samples. Then, the prediction performance is measured on the test data by Pearson correlation (Pcor), Spearman correlation (Scor), and root-mean-squared error (RMSE) between the predicted value and MOS$_s$. The data are divided into ten parts evenly and randomly. During the tenfold cross validation, each of the subsets is used once as test data while the others are used as training samples. The tenfold cross validation was repeated 100 times. We describe the whole pseudocode of this algorithm in Algorithm 1.

We use z-SVR to predict the crosstalk perception for our new data set, where the results are shown in Table I. From Table III, we can see that predicted values of the metrics V_{dispc} and V_{dlogc} have a higher correlation with the subject adjust values of crosstalk than the existing methods on our new data set. These metrics by the color contrast information significantly improve the prediction performance of the metrics using the structure information (V_{dis}, V_{pdis}, and V_{ssim}).

Algorithm 1 Pseudocode of the Proposed Metrics

```
Require: Crosstalk level $p$ and a group of stereo pairs $I = \{L_1, R_1, L_2, R_2, \cdots, L_n, R_n\}$;
Ensure: The vector of perception scores $y$;
for i = 1:n do
    if $(i - 1)\%4 == 0$ then
        Initialize the crosstalk level $p$ ;
        Obtain the disparity maps $disp_i$;
        Extract the color difference maps $deta\_map_i$ ;
        Extract the color contrast maps $log\_map_i$ ;
    else
        $p = p + 0.05$;
    end if
    Filter the disparity map by (4);
    Extract the filtered disparity map and compute the
        metric $Vdispc_i$ using (5);
    Compute the metric $Vdlogc_i$ by (6);
    Obtain the metric $Vpdlc_i$ by (7);
end for
Train the $z - SVR$ model using training samples ($Vpdlc_j$,
    MOS $_j$ ), obtain the parameters $w$ and $b$ ;
Compute the crosstalk perception values y for testing
    samples by (8).
```

TABLE I

PREDICTION PERFORMANCE OF THE PROPOSED METRICS

Metrics	V_{dis}	V_{pdis}	V_{psnr}	V_{ssim}	V_{dispc}	V_{dlogc}
$Pcor$	0.5647	0.5849	0.5260	0.5454	0.8868	0.8810
$Scor$	0.4776	0.4584	0.3861	0.4251	0.8410	0.8105
$RMSE$	0.7449	0.7200	0.7518	0.7599	0.4102	0.4169
STD	0.5045	0.5452	0.5776	0.4930	0.7780	0.7976

IV. EVALUATION DATA SET DESCRIPTION

In order to evaluate the prediction performance of our metrics, it is necessary to obtain subjective crosstalk perception scores for each stereoscopic 3D scene. The similarity between subjective scores and predicted scores obtained by our metrics is used for that evaluation.

A. Composition of Our Stereoscopic Data Set

We construct a stereoscopic 3D image data set, which contains a wide variety of content characteristics. Xing *et al.* [16] introduced a useful data set to assess the crosstalk perception. But, it has only seven scenes that lead to the lack of diversity of scene contents. To improve this, we choose 30 stereoimages from the data set in [37]. We pick high comfortable score images to reduce the influence of visual discomfort and fatigue. These images consist of eight indoor and 15 outdoor scenes (Fig. 4). These scene contents cover a wide range of depth structures, contrasts, edges, and textures, which are

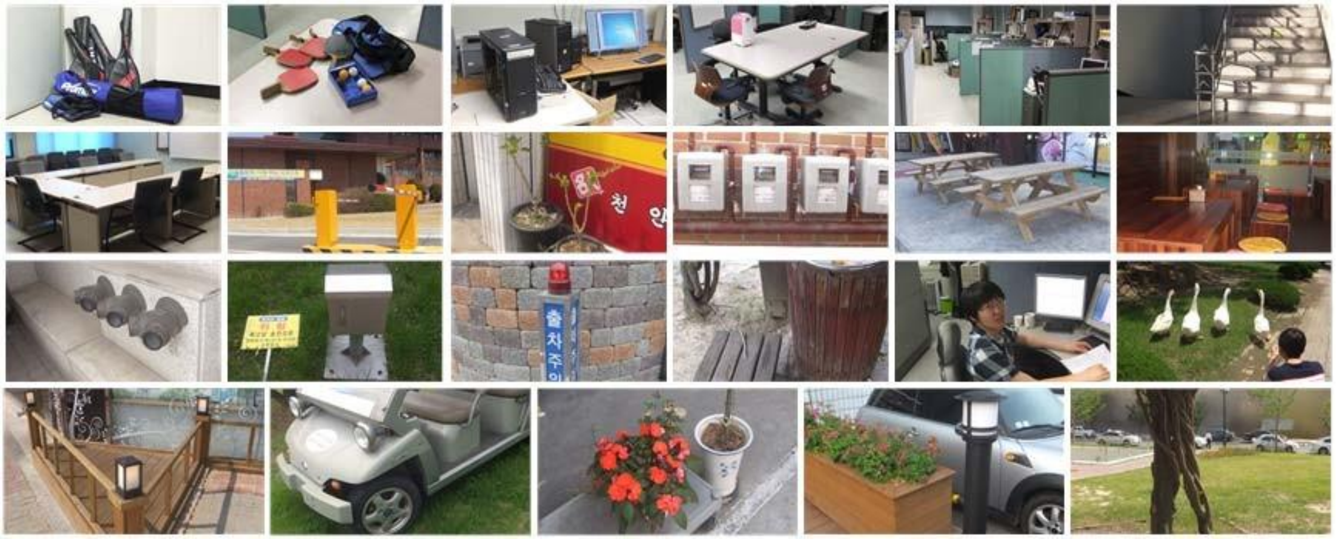


Fig. 4. Our data set of stereoimages for evaluating the crosstalk performance. We use the last row as training set to train subjects and use the rest of images as testing set to obtain crosstalk visibility scores.

TABLE II

EVALUATION RESULTS OF PROPOSED METRICS ON THE SUBJECTIVE DATA SET AND THE RESULTS PROVIDED BY [15]

Metrics	RMSE	Pearson	Spearman
V_{psnr}	0.465	0.821	0.763
V_{ssim}	0.461	0.825	0.784
V_{pdep}	0.382	0.884	0.859
V_{dep}	0.448	0.836	0.844
V_{pdis}	0.416	0.860	0.808
V_{dis}	0.574	0.709	0.688
V_{dispc}	0.421	0.858	0.921
V_{dlogc}	0.295	0.932	0.917
V_{pdlc}	0.290	0.934	0.919

considered as potential factors that impact on perception of crosstalk. Compared with the *crosstalk stereoscopic* data set provided by [15] and [16], our new data set has a more complicated background. For example, it is difficult to distinguish the foreground and the background for the first pair of stereoimages in the bottom row of Fig. 4. The illumination intensity of stereoimages in *crosstalk stereoscopic* data set tends to be constant, and our data set is quite different. Our data set contains not only constant illumination images but also more challenging stereoimages with unstable illumination. For example, the illumination conditions of the indoor stairs are very different between the left half and right half of the image for the top-right scene in Fig. 4. Our data set contains 23 scenes (five for training and 18 for testing), which have more scene content characteristics than *crosstalk stereoscopic* data set, which makes the evaluation of the metric more reasonable and convincing.

Then, we use the algorithm developed in [38] to simulate different levels of system-introduced crosstalk for different displays. Boev *et al.* [38] present a framework for simulating real-world stereoscopic artifacts using the original stereoimages. This algorithm can be summarized by

$$\begin{aligned} Channel\,L^d &= Channel\,L + p \times Channel\,R \\ Channel\,R^d &= Channel\,R + p \times Channel\,L \end{aligned} \qquad (9)$$

TABLE III

EVALUATION RESULTS OF METRICS ON OUR NEW DATA SET

Metrics	RMSE	Pearson	Spearman
V_{dis}	0.752	0.571	0.538
V_{pdis}	0.728	0.607	0.527
V_{psnr}	0.685	0.663	0.563
V_{ssim}	0.784	0.518	0.490
V_{dispc}	0.421	0.888	0.900
V_{dlogc}	0.406	0.896	0.881
V_{pdlc}	0.400	0.900	0.884

where *Channel L* and *Channel R* denote the original left and right views, *Channel* L^d and *Channel* R^d are the distorted views by simulating system-introduced crosstalk distortions.

The parameter p denotes the crosstalk level and it is a real value in the interval [0,1]. In our new data set, four different crosstalk levels p (0%, 5%*,* 10%*, and*15%) are introduced to each 3D image pair, so that there are 72 test stimuli in total. In addition, the crosstalk level of our 27-in patterned retarder 3D display is about 0*.*1%. This crosstalk level is lower than the visibility threshold reported in the literature (about 1% to 2%). So, we combine both system-introduced crosstalk and simulated crosstalk as the total crosstalk levels, which are 0*.*1%, 5*.*1%*,* 10*.*1%, and 15*.*1%. The resolution of these images is 1920 × 1080 pixels.

B. Subjective Crosstalk Visibility Scores

In order to collect subjective crosstalk visibility scores for 32 subjects, totally 18 males and 14 females are invited to participate in our experiments. Among them, one subject who failed the Ishihara test is excluded from our subjective assessment. The rest of the 31 subjects have normal or corrected binocular vision tested by the Snellen chart, and could perceive color (tested by the Ishihara) and the binocular depth [42]. According to the guidance [43], at least fifteen subjects are needed to obtain reliable subjective assessment results. The viewing distance between a subject and the 3D display is fixed to three times of the picture height (about 1.5 m). In order to avoid light pollution, our experiments are carried out in dim environment.

The experiments contain the training sessions and the testing sessions. Among the 23 scenes of our data set, five scenes are used as the training images to train subjects (the bottom row of Fig. 4). A special interface is developed using the *psychtoolbox* [39] to conveniently display the stereoscopic images in a random order. Subjects could conveniently and freely decide when they moved to the next image pairs by pressing keyboard. During the training sessions, a subject could conveniently move to the next image pairs by pressing the "spacebar." In the testing sessions, after the subject gives a score for the current image pair, the assistant presses the corresponding "numerical" key—by doing this, the subject could concentrate on the 3D perception, rather than entering the scores using the keyboard. The score displays in the top-left for 1 s. Then, the program moves to the next pairs. In this developed interface, we also display the images in full screen, and disable unnecessary keys.

During the training sessions, a modified version of the single stimulus [43] is used with a five-point grading scale (5: imperceptible, 4: perceptible but not annoying, 3: slightly annoying, 2: annoying, and 1: very annoying). Five image pairs from five scenes (the last row of Fig. 4) are selected by expert viewers in such a way that each quality level is represented by an example image. We display each example and explain the corresponding quality level to the testers, until they could distinguish the five different quality levels. After that, three dummy 3D images from the image scenes we used in the training sessions are presented to testers to stabilize their judgment. Then, a total number of 72 stereoscopic images (18 scenes with four different crosstalk levels) are randomly presented to the subjects in the testing sessions.

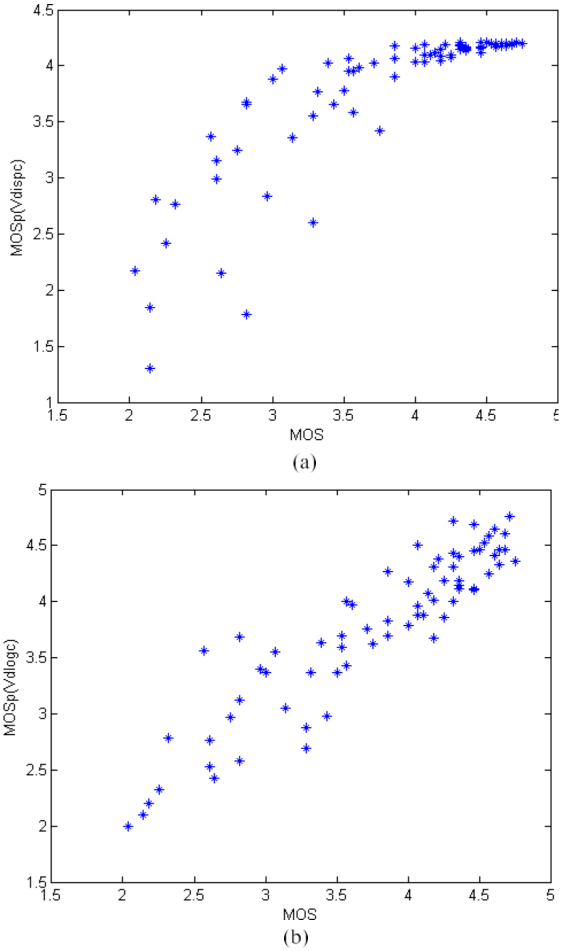
After subjective assessment, one outlier is eliminated according to the screening methodology recommended by ITU-R BT.500-11 [43]. We use the β_2 test [43] to determine whether the subjective scores are normal. The results show that 53 stimuli are normally distributed ($2 \leq \beta_2 \leq 4$), and 17 stimuli are close to normally ($1 \leq \beta_2 < 2$ or $4 < \beta_2 \leq 5$), while the remaining two stimuli are not ($\beta_2 < 1$ or $\beta_2 > 5$). We can safely assume that the scores are subject to the normal distribution. Finally, the crosstalk perception score for each test image is represented by MOSs from subjects.

V. Performance Evaluation

In this section, we evaluate the performance of the proposed methods using the real-world images of our new data set and the *crosstalk stereoscopic* data set, provided by [15] and [16].

A. Evaluation on Crosstalk Stereoscopic Data Set

In the *crosstalk stereoscopic* data set, 72 test stimuli are acquired by applying three camera baselines and four crosstalk levels to six scene contents. Based on the definition that crosstalk is the leakage of luminance from one eye channel to the other eye channel, the crosstalk images are simulated by a computer on the luminance channel. The corresponding MOSs of crosstalk visibility are obtained by subjective assessments. To grade the degree of crosstalk visibility, an example of five categorical adjectival levels is used, and a total of 28 subjects participated in the tests.



(a)

(b)

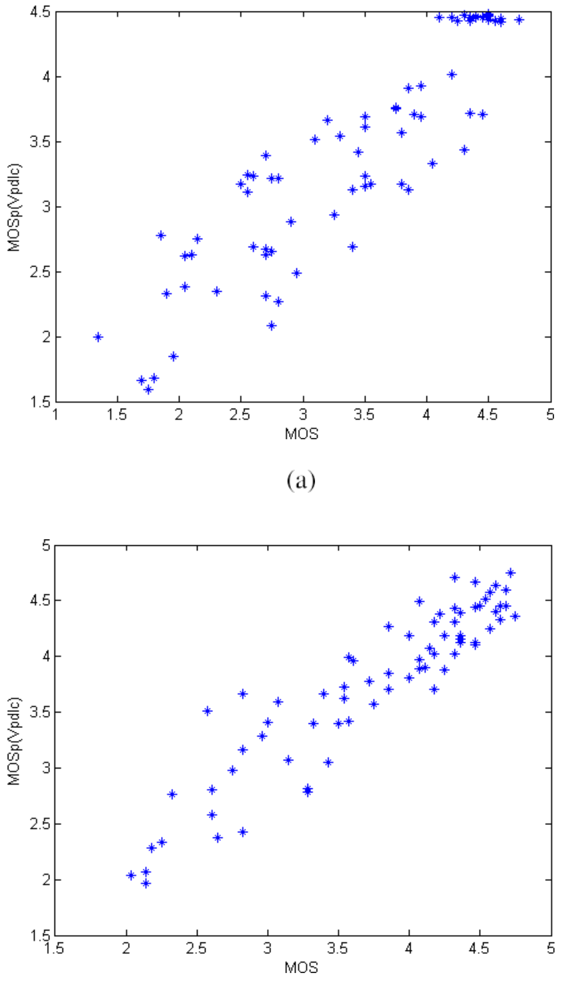
Fig. 5. Scatter plot of *MOS* of crosstalk perception versus predicted values MOS $_p$ of V_{dispc} and V_{dlogc} on the *crosstalk stereoscopic* data set provided by [15]. (a) Pcor=0.8580, Scor=0.9206, RMSE=0.4214. (b) Pcor=0.9320, Scor=0.9172, RMSE=0.2955.

First, we use a nonlinear regression in (10) suggested by Video Quality Expert Group (VQEG) to transform the results of each metric (V) to the predicted MOS values (MOS$_p$), then calculate the RMSE, Pearson correlation (Pcor) coefficient, and Spearman correlation (Scor) coefficient between the objective values MOS$_p$ and the subjective values MOS$_s$. Equation (10) normalizes the value of each metric to the range of MOS. The nonlinear regression suggested by the VQEG is defined as

$$\text{MOS}_p = \frac{b_1}{1 + \exp(-b_2 \times (r(V) - b_3))} \qquad (10)$$

where b_1, b_2, and b_3 are the regression coefficients, which can be initialized by 0, and $r(V)$ is the raw value calculated from the proposed metrics.

Fig. 5 shows the scatter plot of *MOS* versus MOS$_p$ of the proposed metrics on the *crosstalk stereoscopic* data set in [15], where the performance of metric V_{dlogc} has a better performance than metric V_{dispc}. It can be seen from bottom row of Fig. 5 that the metric V_{dlogc} has better performance



(b)

Fig. 6. Scatter plot of MOS of crosstalk perception versus predicted values MOS $_p$ of V_{pdlc} on (a) our data set and (b) data set in [15].

in predicting crosstalk perception of stereoscopic images with low and high impairments. The metric V_{dispc} also has good performance in predicting crosstalk perception of stereoscopic images with high and medium impairments. Six scene contents of crosstalk stereoscopic data set have simple background and remarkable foreground, especially for scenes "Champagne," "Dog," and "Pantomime," the background of these scenes is near pure color. Therefore, contrast maps of the stereoscopic images have significantly impact on this data set.

We compare our metrics with six other popular metrics. The proposed metric is compared with traditional 2D metrics V_{psnr} and V_{ssim} as well as other three metrics V_{dep}, V_{pdis}, and V_{dis}. The SSIM [29] is a perceptual quality metric, which takes the characteristics of stereoscopic images into account for predicting quality levels of crosstalk perception in stereoscopic images. It is based on an understanding of three main factors: crosstalk level, camera baseline, and scene content. V_{psnr} and V_{ssim} are calculated between the original-crosstalked and the left-crosstalked images. V_{dep} combines the SSIM map [29] and the depth map calculated by Depth Estimation Reference Software. V_{pdis} and V_{dis} are proposed by [15]. The metric V_{dis} is a combination of the SSIM map and the disparity map estimated by the sum of squared difference plus min filter [35]. V_{pdis} and V_{pdep} are built by using the disparity map and the depth map filtered by the SSIM map, respectively. The similarity between subjective and predicted scores obtained by the above metrics and our metrics using the crosstalk stereoscopic data set are presented in Table II. Compared with other metrics, our metrics V_{dispc} and V_{dlogc} have low values of RMSE and high values of Pearson correlation and Spearman correlation coefficients. In particular, the proposed metric V_{dlogc} has better performance in RMSE and Pearson than above metrics, because of high contrast of the crosstalk stereoscopic data set. Furthermore, we combine our two metrics to construct a new metric V_{pdlc}, which performs better than other well-known metrics.

B. Evaluation on Our New Data Set

Fig. 6 shows the scatter plot of MOS of crosstalk perception versus predicted values MOS$_p$ of V_{pdlc} on our new data set and crosstalk stereoscopic data set. Our metrics has a great performance on the crosstalk stereoscopic data set, and we can see that MOS$_p$ in Fig. 6(b) is more accurate than the one in Fig. 6(a). This is because our data set has more challenging scenes with vivid content characteristics, such as complicated background, unstable illumination, rich color information, and depth structures. The more discrete points concentrate on the diagonal, the more correct predicted scores are. The similarity between subjective scores and predicted scores obtained by the exciting metrics and our metrics using the new data set can be seen in Table III. Our new metrics V_{dispc} and V_{dlogc} perform better than the existing methods in our data set. The metric V_{dlogc} has the largest Person correlation (0.896) and Spearman correlation (0.881), which indicates that the color contrast between the original left and right images and crosstalk level have an important influence on the crosstalk perception.

VI. Conclusion and Future Work

In this paper, we have proposed a novel metric to predict the crosstalk visibility. To build this metric, we used crosstalk level, disparity map, and color contrast map, which are calculated between the original left and right images. The existing methods usually use the difference between the ghost and the original images. But, the ghost images generated by a computer are not accurate, because the mechanism behind crosstalk is still not clear. Our metric does not require the ghost images. The color information has an important influence on crosstalk perception, and it builds our new metrics. Experimental results show that our metrics using color information yield higher correlation against the subjective MOS values than previous metrics. Our results indicate that color information is the most critical factor to achieve better crosstalk visible prediction.

References

[1] H. Urey, K. V. Chellappan, E. Erden, and P. Surman, "State of the art in stereoscopic and autostereoscopic displays," *Proc. IEEE*, vol. 99, no. 4, pp. 540–555, Apr. 2011.

[2] S. Ryu and K. Sohn, "No-reference quality assessment for stereoscopic images based on binocular quality perception," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 4, pp. 591–602, Apr. 2014.

[3] Z. Zhang, C. Zhou, Y. Wang, and W. Gao, "Interactive stereoscopic video conversion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 10, pp. 1795–1808, Oct. 2013.

[4] I. Tsirlin, L. M. Wilcox, and R. S. Allison, "The effect of crosstalk on the perceived depth from disparity and monocular occlusions," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 445–453, Jun. 2011.

[5] A. J. Woods and S. S. L. Tan, "Characterizing sources of ghosting in time-sequential stereoscopic video displays," *Proc. SPIE*, vol. 4660, pp. 66–77, May 2002.

[6] A. J. Woods and C. R. Harris, "Comparing levels of crosstalk with red/cyan, blue/yellow, and green/magenta anaglyph 3D glasses," *Proc. SPIE*, vol. 7524, p. 75240Q, Feb. 2010.

[7] A. J. Woods, "Understanding crosstalk in stereoscopic displays," presented at the Three-Dimensional Syst. Appl. Conf., Tokyo, Japan, 2010, pp. 19–21.

[8] H. Sohn, Y. J. Jung, and Y. M. Ro, "Crosstalk reductiong in stereoscopic 3D displays: Disparity adjustment using crosstalk visibility index for crosstalk cancellation," *J. Opt. Exp.*, vol. 22, no. 3, pp. 3375–3392, 2014.

[9] C. Doutre and P. Nasiopoulos, "Crosstalk cancellation in 3D video with local contrast reduction," in *Proc. EUSIPCO*, 2011, pp. 1884–1888.

[10] B. Lacotte, J. Konrad, and E. Dubois, "Cancellation of image crosstalk in time-sequential displays of stereoscopic video," *IEEE Trans. Image Process.*, vol. 9, no. 5, pp. 897–908, May 2000.

[11] P. J. H. Seuntiëns, L. M. J. Meesters, and W. A. IJsselsteijn, "Perceptual attributes of crosstalk in 3D images," *Displays*, vol. 26, nos. 4–5, pp. 177–183, 2005.

[12] L. M. Wilcox and J. A. D. Stewart, "Determinants of perceived image quality: Ghosting vs. brightness," *Proc. SPIE*, vol. 5006, pp. 263–268, May 2003.

[13] L. Wang *et al.*, "Crosstalk evaluation in stereoscopic displays," *IEEE J. Display Technol.*, vol. 7, no. 4, pp. 208–214, Apr. 2011.

[14] D.-H. Kang, E.-J. Lee, J.-H. Lee, and J.-K. Song, "Perceptual strength of 3-D crosstalk in both achromatic and color images in stereoscopic 3-D displays," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3253–3261, Jul. 2012.

[15] L. Xing, J. You, T. Ebrahimi, and A. Perkis, "Assessment of stereoscopic crosstalk perception," *IEEE Trans. Multimedia*, vol. 14, no. 2, pp. 326–337, Apr. 2012.

[16] L. Xing, J. You, T. Ebrahimi, and A. Perkis, "Stereoscopic quality datasets under various test conditions," in *Proc. 5th Int. Workshop Quality Multimedia Exper. (QoMEX)*, 2013, pp. 136–141.

[17] J.-C. Liou, K. Lee, F.-G. Tseng, J.-F. Huang, W.-T. Yen, and W.-L. Hsu, "Shutter glasses stereo LCD with a dynamic backlight," *Proc. SPIE*, vol. 7237, p. 72370X, Jan. 2009.

[18] S. Shestak, D. Kim, and S. Hwang, "Measuring of gray-to-gray crosstalk in a LCD based time-sequential stereoscopic display," in *SID Symp. Dig. Tech. Papers*, May 2010, vol. 41. no. 1, pp. 132–135.

[19] J. Shen, X. Yang, Y. Jia, and X. Li, "Intrinsic images using optimization," in *Proc. IEEE CVPR*, Jun. 2011, pp. 3481–3487.

[20] T. Kim, J. M. Ra, J. H. Lee, S. H. Moon, and K.-Y. Choi, "3D crosstalk compensation to enhance 3D image quality of plasma display panel," *IEEE Trans. Consum. Electron.*, vol. 57, no. 4, pp. 1471–1477, Nov. 2011.

[21] M. Zeng and T. Nguyen, "Crosstalk modeling, analysis, simulation and cancellation in passive-type stereoscopic LCD displays," in *Proc. IEEE ICASSP*, May 2013, pp. 1840–1844.

[22] J. Shen, Y. Du, W. Wang, and X. Li, "Lazy random walks for superpixel segmentation," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1451–1462, Apr. 2014.

[23] Y. Kwak, S. Lee, and S. Yang, "Crosstalk characterization method for stereoscopic three-dimensional television," *IEEE Trans. Consum. Electron.*, vol. 58, no. 4, pp. 1411–1415, Nov. 2012.

[24] J. Shen, D. Wang, and X. Li, "Depth-aware image seam carving," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1453–1461, Oct. 2013.

[25] J. Shen, Y. Du, and X. Li, "Interactive segmentation using constrained Laplacian optimization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 7, pp. 1088–1100, Jul. 2014.

[26] Y. J. Jung, H. Sohn, S.-I. Lee, Y. M. Ro, and H. W. Park, "Quantitative measurement of binocular color fusion limit for non-spectral colors," *Opt. Exp.*, vol. 19, no. 8, pp. 7325–7338, 2011.

[27] S. Pastoor, "Human factors of 3D displays in advanced image communications," *Displays*, vol. 14, no. 3, pp. 150–157, 1993.

[28] L. Lipton, "Factors affecting 'ghosting' in time-multiplexed pianostereoscopic CRT display systems," *Proc. SPIE*, vol. 0761, pp. 75–78, Jun. 1987.

[29] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[30] J. Peng, J. Shen, and X. Li, "High-order energies for stereo segmentation," *IEEE Trans. Cybern.*, vol. 46, no. 7, pp. 1616–1627, Jul. 2016.

[31] W. Wang, J. Shen, X. Li, and F. Porikli, "Robust video object cosegmentation," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3137–3148, Oct. 2015.

[32] X. Dong, J. Shen, L. Shao, and L. Van Gool, "Sub-Markov random walk for image segmentation," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 516–527, Feb. 2016.

[33] A. J. Woods, "Crosstalk in stereoscopic displays: A review," *J. Electron. Imag.*, vol. 21, no. 4, p. 040902, 2012.

[34] M. W. Levine and J. M. Shefner, "Fundamentals of sensation and perception," *Color Res. Appl.*, vol. 26, pp. 324–325, 1991.

[35] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, nos. 1–3, pp. 7–42, Apr. 2002.

[36] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011.

[37] H. Sohn, Y. J. Jung, S.-I. Lee, and Y. M. Ro, "Predicting visual discomfort using object size and disparity information in stereoscopic images," *IEEE Trans. Broadcast.*, vol. 59, no. 1, pp. 28–37, Mar. 2013.

[38] A. Boev, D. Hollosi, and A. Gotchev, "Software for simulation of artefacts and database of impaired videos," Mobile3DTV, Project Rep. 216503, 2008.

[39] D. H. Brainard, "The psychophysics toolbox," *Spatial Vis.*, vol. 10, no. 4, pp. 433–436, 1997.

[40] W. Wang, J. Shen, and F. Porikli, "Saliency-aware geodesic video object segmentation," in *Proc. IEEE CVPR*, Jun. 2015, pp. 3395–3402.

[41] W. Wang, J. Shen, Y. Yu, and K.-L. Ma, "Stereoscopic thumbnail creation via efficient stereo saliency detection," *IEEE Trans. Vis. Comput. Graphics*, to be published, doi: 10.1109/TVCG.2016.2600594.

[42] *Test Your Stereo (3D) Vision*. [Online]. Available: http://3d.mcgill.ca/

[43] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document ITU-R BT-500.11, 2002.