MAKING STEREO PHOTO CROPPING EASY

Fan Zhang, Yuzhen Niu, Feng Liu

Department of Computer Science, Portland State University
{zhangfan,yuzhen,fliu}@cs.pdx.edu

ABSTRACT

The increasing popularity of stereoscopic 3D brings the demand for tools for editing and authoring stereoscopic images and videos. This paper shows that even a simple task like cropping is difficult for amateur users with little stereoscopic photography knowledge. Unlike regular monocular (2D) images, cropping a stereoscopic image needs to be carefully executed to avoid stereoscopic violations, which otherwise cause an unpleasant stereoscopic viewing experience. In this paper, we present a system that assists in stereoscopic photo cropping by automatically measuring the stereoscopic photography violations and alerting users with the potential violations. Our study shows that compared to a popular stereoscopic photo editing system, our system makes stereoscopic photo cropping easier even for amateur users with little stereoscopic photography knowledge and provides a good user experience.

***Index Terms*—** Stereoscopic photography, photo cropping, photo editing

1. INTRODUCTION

Stereoscopic 3D enhances the feeling of depth in an image by feeding two different images to a viewer's left and right eye. It becomes increasingly popular these years. For example, consumer stereoscopic devices are now accessible to common users, such as 3D monitors, 3D HDTVs, and consumer-level 3D cameras. In contrast, high-quality stereoscopic content is still difficult to produce. Stereoscopic images and videos, if not properly produced, will lead to "3D fatigue", such as eyestrain and headaches [1]. The demand for stereoscopic content authoring and editing tools is emerging.

While a rich set of editing and authoring tools are available for monocular (2D) images and videos, only a few tools exist for stereoscopic ones. Most of them, however, are intended for professional users and require their users to have a good understanding of stereoscopic photography and cinematography [2, 3, 4]. Quintus and Halle developed a system for optimizing stereoscopic viewing experience for 3D graphics images [5]. Their system does not work for regular images and videos. Some advanced tools for stereoscopic content processing become available recently, such as video stabilization [6], photo display adaptation [7, 8], automatic quality optimization [9], warping [10], stereo keystone correction [11], and object copy and paste [12]. However, the support for amateur users in editing stereoscopic content is still insufficient. For users with little background in stereoscopic content production, even cropping a stereoscopic photo can possibly introduce stereoscopic violations and cause 3D fatigue.


(a) Ideal 3D perception (b) Actual 3D perception

Fig. 1. Stereo window violation. When an object with negative disparities is cut by the screen edge, it suffers from the stereo window violation. That is, the object is perceived in front of the screen, but is occluded by the screen edge.

This paper presents an Assisted Stereo Photo Cropping system that avoids stereoscopic violations. When a user crops an image region, our system automatically measures the potential stereoscopic violations and alerts a user to these violations. The user can then either re-crop the image or change the disparity to minimize these violations as guided by our system. Our experiments show that compared to a popular stereoscopic photo editing system, StereoPhoto Maker[1], our system makes stereoscopic photo cropping easier even for users with little stereoscopic photography knowledge.

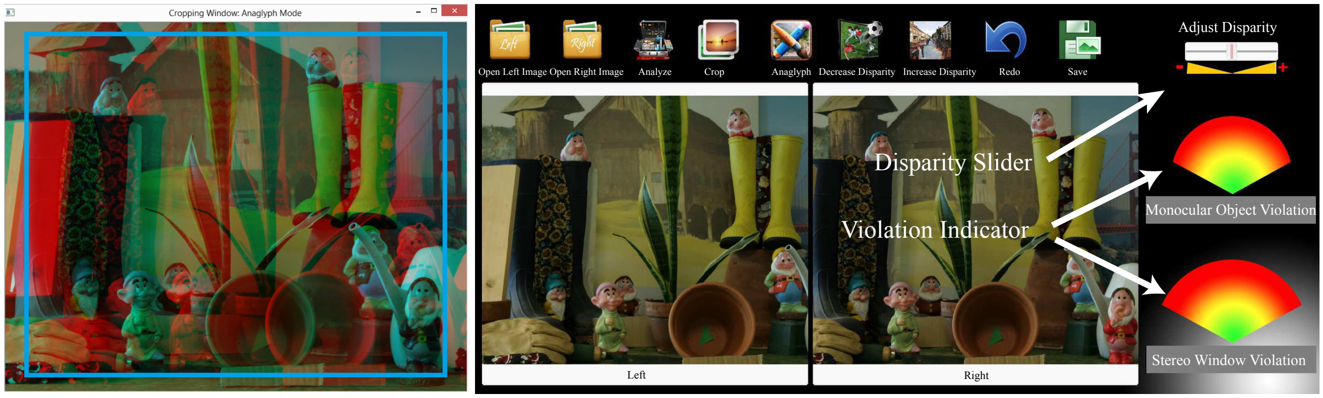
This paper shows that amateur users with little stereoscopic photography knowledge need to be assisted with proper tools in stereoscopic content production. Our work shows that detecting and alerting users to stereoscopic violations can effectively help them create stereoscopic content that delivers a pleasant stereoscopic 3D viewing experience.

2. ASSISTED STEREO PHOTO CROPPING

Our Assisted Stereo Photo Cropping (ASPC) system implements the basic stereoscopic image editing functions from StereoPhoto Maker (SPM), a popular freely available soft-

---

[1] http://stereo.jpn.org/eng/stphmkr/



(a) Anaglyph view (with the blue cropping window) (b) Side-by-side View

Fig. 2. Interface for our ASPC system. Our system provides both the red-cyan anaglyph in (a) and side-by-side stereoscopic viewing mode in (b). A user can crop a stereo photo in both views. In this figure, the user crops the photo using the anaglyph view as shown in (a). The stereoscopic violations are visualized by the indicators in (b). The user can use the disparity adjustment slider on the top right in (b) to eliminate or relieve the stereoscopic violations.

ware tool for stereoscopic image editing. Besides, ASPC provides assisted cropping technologies to help a user avoid stereoscopic violations during cropping. In this section, we first introduce the relevant stereoscopic photography concept, then describe the standard cropping interfaces and the unique assisted cropping interface, and finally elaborate the computer vision algorithms that support the assisted cropping interface.

2.1. Stereoscopic Photography Basics

Compared to monocular (2D) images, stereoscopic images have an extra dimension of disparity. Disparity directly affects the perceived distance of an object in a stereoscopic image. If the disparity of an object is zero, it is perceived on the screen. If its disparity is negative, the object pops out of the screen. On the other hand, if the disparity is positive, the object is perceived behind the screen. Disparity needs to be correctly taken care of to deliver a pleasant viewing experience. Professionals often carefully adjust the disparities to position content of interest in the stereoscopic comfort zone [13].

There are two stereoscopic photography rules that are particularly applicable to cropping: monocular object violation and stereo window violation. Monocular object violation occurs when salient visual content only shows in one of the two views of a stereoscopic photo. This sometimes brings in *retinal rivalry* to viewers. Stereo window violation happens when an object has a negative disparity value and thus is perceived in front of the screen, but is occluded by the screen edge, as shown in Fig. 1 (b). This often confuses our brain: an object cannot be occluded by something behind it.

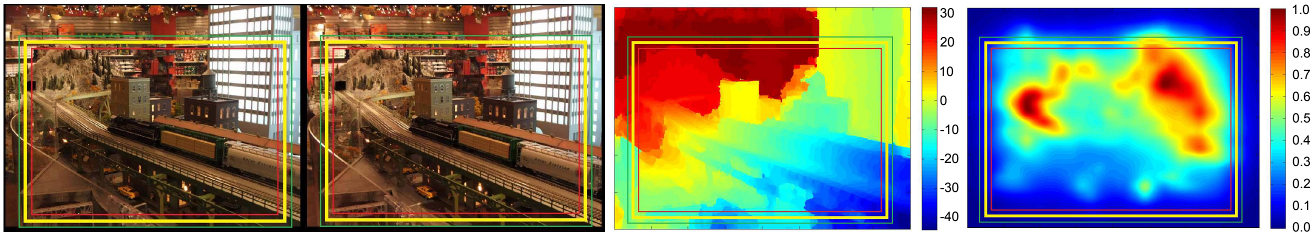
2.2. Standard Cropping Interface

We borrowed the following standard interfaces for stereoscopic photo cropping from SPM. First, ASPC provides two options for a user to display a stereoscopic image: the red-cyan anaglyph mode and side-by-side mode. These two stereoscopic viewing modes are shown in Fig. 2 (a) and (b), respectively. A user can select a region of interest and crop it in either display mode. For cropping, the anaglyph mode provides an intuitive disparity visualization, as the user can sense the disparity by looking at the shift between the red and cyan color channels, as shown in Fig. 2 (a). Its disadvantage is that sometimes the image details in some regions are difficult to tell. The side-by-side mode displays the left and right view of a stereoscopic image. When a user crops an image region in one view, the same cropping window will be propagated to the same location in the other view. With this mode, the user can fully appreciate the image details and colors, but it is difficult for the user to know the disparities.

Second, ASPC implements the standard horizontal image translation (HIT) [14]. HIT is widely used in stereoscopic photography to uniformly change the disparity distribution of a stereoscopic image by horizontally shifting the left and right image. ASPC allows a user to perform HIT through a horizontal slider interface.

2.3. Assisted Cropping Interface

Our ASPC system automatically detects and measures the stereo window violation and monocular object violation whenever a cropping window is specified, and visualizes these measurements to alert users. As shown in Fig. 2 (b), our system uses the size and color of fans to visualize the violation measurements: as the violation increases, the fan grows larger and the outer part of the fan takes a hotter color. A typical cropping scenario with our system is as follows. A user first selects and crops a region of interest. If our system predicts stereoscopic violations with hot colors, the user can re-crop the image or apply HIT in a trial-and-error way. With HIT, a user can either uniformly increase or decrease the disparity. Typically, a user can first try to change the disparity in



(a) Input: left and right image (b) Disparity map (c) Saliency map

Fig. 3. Stereo window violation detection. The yellow rectangle in each image is the cropping window, and the red and green rectangles are its inner and outer boundaries. In this example, the train at the bottom right has negative disparity values and is perceived popping out of the screen. But it is cut by the screen edge.

one direction and if it deteriorates the violations, the user can change the disparity in the other direction. This does not require much stereoscopic photography knowledge from users. Our experiments showed that simple trial-and-error in a few seconds will significantly relieve the stereoscopic violations.

It is worthwhile to note that a user may be able to detect these stereoscopic violations by watching the cropping result on a stereoscopic display. This eyeball test, however, has two major drawbacks. First, the user can often suffer from 3D fatigue from viewing the bad cropping result in the processing of tweaking the cropping result. This actually motivated us to develop our ASPC system to minimize the fatigue from inspecting the cropping result. Second, violation detection via the eyeball test is not very accurate. Most amateur users can only detect and correct the most obvious violations and miss some violations which can cause 3D fatigue. As 3D fatigue from viewing bad stereo images is accumulative, those missed violations can also be problematic for viewers.

2.4. Stereoscopic Violation Detection

Our method uses a data-driven approach to detect and measure the two stereoscopic violations. Below we describe how our method encodes the relevant stereoscopic photography rules to define features and uses them in machine learning techniques to detect the violations.

2.4.1. *Stereo Window Violation*

Since stereo window violation only happens when objects along image borders have negative disparities, the percentage of pixels that have negative disparities in the cropping window border region is a good indication of this violation. The cropping window border region is the region between the cropping window and its inner boundary as shown in Fig. 3 (a). The border width is 5 pixels in our system. We first estimate a dense disparity map [15] as shown in Fig. 3 (b) and then compute this percentage.

An object pops out more when it has a larger negative disparity value. If such an object is cut by the screen edge, the worse the stereo window violation is. Accordingly, we use the magnitude of the negative disparities to measure this violation. Specifically, we compute the average magnitude of negative disparities for each of the four cropping window border regions and select the maximal one as a feature.

A visually salient object catches the viewer attention more than those less salient ones. We use the graph-based visual saliency method to estimate the saliency from the input image [16], as shown in Fig. 3 (c). We then compute the average image saliency inside each cropping window border region and select the maximal one as a feature.

The stereo window violation occurs only when an object is cut by the screen edge. We measure whether an object is cut or not by computing the similarity between the inner and outer regions around each boundary, as shown in Fig. 3 (a). The similarity between these two regions is measured based on both the color histogram and spatiogram [17].

$$sim_1 = \max_{\mathcal{B}} f(H_c^i(\mathcal{B}), H_c^o(\mathcal{B})) + f(H_s^i(\mathcal{B}), H_s^o(\mathcal{B})) \quad (1)$$

where $H_c^i(\mathcal{B})$ and $H_c^o(\mathcal{B})$ are the color histogram of the inner and outer boundary region along boundary $\mathcal{B}$. $H_s^i(\mathcal{B})$ and $H_s^o(\mathcal{B})$ are the spatiograms. f computes the similarity between two histograms using the Bhattacharyya coefficient [18]. Similarly, we measure whether an object is cut or not by computing the disparity difference between pixels separated by each image boundary. If two pixels belong to the same object, they tend to have a similar depth value. We measure the depth difference using the disparity difference.

$$sim_2 = \min_{\mathcal{B}} \frac{\sum_{i,j} |d_i - d_j|}{l_b} \quad (2)$$

where i and j are two neighboring pixels separated by the boundary $\mathcal{B}$, and l_b is the boundary length.

We use these features in the Logistic Regression method [19] to detect whether an image suffers from the stereo window violation and use the probability that the image suffers from the violation as the violation measurement.

2.4.2. *Monocular Object Violation*

Monocular object violation occurs when an object appears only in one of the two views of a stereoscopic image. As

(a) Input left image and its edge map (b) Input right image and its edge map

(c) Input left image and its edge map (d) Input right image and its edge map

Fig. 4. Monocular object violation detection. The yellow box in each image indicates the cropping window. The small rectangle with a red edge is the monocular region. The edge rate for the image in the first row and second row are 0 and 0.143, respectively.

this violation increases with the monocular region size, we estimate the monocular region size as a feature. Specifically, we detect feature point correspondences [20] between the two views of a stereoscopic image in the image boundary area and compute the width of the local monocular region near each feature point as the disparity size. Then we compute the maximal and minimal widths as features. We assign the signs for these values from the disparities. In order to consider the original image size, we also compute the relative widths by dividing these monocular region widths with the image width.

If a monocular region has a large amount of visually salient image features, such as edges, it tends to cause serious retinal rivalry. In contrast, if a monocular region is textureless, such as the uniform background shown in Fig. 4 (a) and (b), the retinal rivalry is less significant. Therefore, our method estimates an edge map for each view using the Canny algorithm [21] as shown in Fig. 4 and computes the edge rate in the monocular region as a feature to measure the violation.

$$er = \max\left(\frac{\sum_{i\in\mathcal{M}_l} E_i}{\sum_{i\in\mathcal{I}_l} E_i}, \frac{\sum_{i\in\mathcal{M}_r} E_i}{\sum_{i\in\mathcal{I}_r} E_i}\right) \quad (3)$$

where $\mathcal{M}_l$ and $\mathcal{M}_r$ are the monocular region in the left view $\mathcal{I}_l$ and right view $\mathcal{I}_r$. E_i is the edge value of pixel i.

When the border regions in the two views on the same side differ significantly, retinal rivalry becomes serious. We therefore measure the similarity between the two border regions based on the color histogram and spatiogram [17].

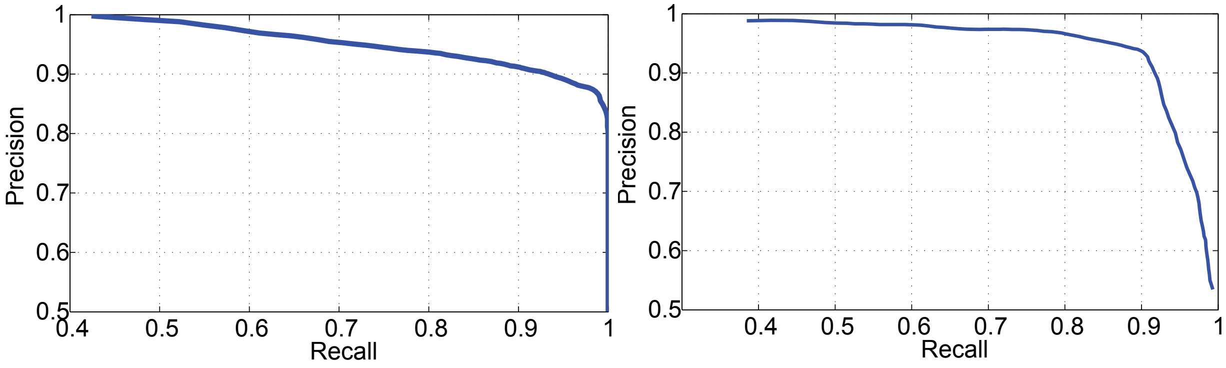
$$bsim = \min(f(\mathcal{HB}_l^l, \mathcal{HB}_r^l), f(\mathcal{HB}_l^r, \mathcal{HB}_r^r)) \quad (4)$$

where $\mathcal{HB}_l^l$ and $\mathcal{HB}_r^l$ are the histogram of the left border region of the left view and right view, respectively. $\mathcal{HB}_l^r$ and $\mathcal{HB}_r^r$ are the histogram of the right border region of the left view and right view, respectively. f computes the histogram similarity using the Bhattacharyya coefficient [18].

3. EXPERIMENTS

We first quantitatively evaluate the performance of our stereoscopic violation detection methods. To our best knowledge, there is no large-scale benchmark available for stereoscopic violation detection. We collected 3780 stereo images from *Flickr* and Muttyan's Stereo Galleries to test our algorithms. We randomly generated cropping results from these images and asked five viewers to label whether each cropping result suffers from the stereo window violation or not. The viewers examined these results using an ASUS VG236H 120 Hz 3D monitor with shuttered glasses and Nvidia GeForce 3D Vision Solution. Each cropping result was assigned a label that has maximal votes. We collected a set of 984 cropping results that have the stereo window violation and another set of 984 results that do not. We randomly allocated 75% of the dataset into a training set and the rest into a testing dataset, and then tested the detection performance of our method. We repeated the random partition and the corresponding test for 100 times and report the average result in the Fig. 5 (a). Our method achieves a precision value 91.2% at a recall value 90.0%.

We also identified 550 images that suffer from the monocular object violation and another 550 images that are free from this violation. We then randomly assigned 75% of this dataset into a training set and the rest into a testing set, and then tested the performance of our detection algorithm. Again, we repeated the random partition and test for 100 times and report



(a) SWV detection

(b) MOV detection

Fig. 5. Precision-recall curves for stereo window violation (SWV) and monocular object violation (MOV) detection.

the result in Fig. 5 (b). Our method achieves a precision value 93.6% at a recall value 90.0%.

3.1. User Study

We designed a user study to evaluate the user experience with our ASPC system and compare it to a popular stereoscopic photo editing tool, SPM. We collected 10 stereoscopic images where cropping, if not properly conducted, will violate the stereoscopic rules described above. We asked each participant to use both ASPC and SPM to crop a region of interest and after cropping, we asked them to compare the two results using the above-mentioned stereoscopic display and glasses.

There are 22 participants in our study, including 12 females and 10 males. They are all students covering a wide variety of background and coming from 14 departments, including Business, Education, Biology, English, Mathematics, Physics, Psychology, Computer Science, Economics, etc. Their ages range from around 20 to 55 years old. While 19 of them watched stereoscopic movies before, they all have little knowledge about stereoscopic photography.

Before the study, we warned the participants that if cropping is not well conducted, it can lead to an unpleasant viewing experience. Our study is within-participant. Each participant uses both systems to crop all the 10 images. We randomize the order of images for each participant and use the Latin square to counterbalance the order for the two systems. Before the study, each participant was given instructions on how to use each system. With ASPC, a user first identifies and crops a region of interest. Whenever a cropping window is generated, ASPC will instantaneously alert the user to the violation measurement. The user can then use the disparity adjustment slider to quickly find a cropping window that they feel good with. With SPM, the user can first crop the image in the same way as ASPC. Then, a user can press the left or right arrow key to adjust the disparity. SPM does not give any violation warning. Alternatively, a user can use the auto-adjustment function to optimize the cropping result by clicking the corresponding button. Besides the final results, we save the initial cropping results without any adjustment.

After cropping all the images, each participant was asked to rate her/his own cropping results. For each input image, the participant was first shown the initial cropping result and final cropping result from the same system side by side. The placement of the results is randomized. The participant was then asked to rate each result using a Likert scale ranging from 1 (strongly disagree/very dissatisfied) to 5 (strongly agree/very satisfied). The participant was then shown the initial cropping result and final cropping result from another system side by side and rated them in the same way. The order for the results from the two systems was counterbalanced using the Latin square. These tests were designed to compare the initial cropping results and final cropping results, and particularly to study whether naïve cropping without disparity adjustment is acceptable or not. After these tests, the participant was shown the final cropping results from ASPC and SPM side by side. The placement of the results was also randomized. The participant was asked to rank the preference after viewing the two results using the same Likert scale.

We conclude the study by asking the participant to rank their experiences with each system from 1 to 5 and if the alerts from ASPC were helpful or not.

3.1.1. Result

Table 1 shows the comparison between the initial cropping results without disparity adjustment and final cropping results with disparity adjustment. Naïve cropping without disparity adjustment tends to produce unsatisfactory results, which shows the need of cropping tools dedicated for stereoscopic photos. Disparity adjustment, especially from ASPC, can significantly improve the cropping results. The p-values of the paired two-sample t-tests between the initial and final cropping results for SPM and ASPC are both smaller than 0.0001.

The direct comparison between the final cropping results from SPM and ASPC shows the strong preference of ASPC over SPM (M = 4.66 (out of 5), STD = 0.69). The p-value of the t-test is smaller than 0.0001. This is consistent with the satisfaction tests on these two systems, where the average score for ASPC was 4.82 (STD = 0.39) and the average score

	Initial: mean	Initial: std	Final: mean	Final: std
SPM	2.07	0.33	2.80	0.58
ASPC	2.02	0.32	4.49	0.33

Table 1. The comparison between the initial cropping results and final cropping results after disparity adjustment.

for SPM was 1.82 (STD = 0.39). The *p*-values of the paired two-sample t-test between SPM and ASPC is smaller than 0.0001. Finally, the participants unanimously and strongly agree that the alerts from ASPC are helpful.

4. CONCLUSION

This paper shows that even cropping a stereo image is still beyond the capability of common users without stereoscopic photography knowledge. The presented Assisted Stereo Photo Cropping system uses computer vision technologies to detect stereoscopic violations and warns users against them during cropping. Our experiments show that this ASPC system can effectively help amateur users obtain good stereo photo cropping results that lead to a satisfactory 3D viewing experience. In the future, we plan to extend our system to support more stereo photo editing tasks for amateur users.

Acknowledgments. We would like to thank Flickr users *Chung Chu* and *ortygia* for letting us use their photos under a Creative Commons license. Fig. 1 uses materials from Blender Foundation / www.bigbuckbunny.org under a Creative Commons license. This work was supported by NSF CNS-1205746 and CNS-1218589.

5. REFERENCES

[1] B. Mendiburu, *3D Movie Making: Stereoscopic Digital Cinema from Script to Screen*, Focal Press, 2009.

[2] S.J. Koppal, C.L. Zitnick, M. Cohen, Sing Bing Kang, B. Ressler, and A. Colburn, “A viewer-centric editor for 3d movies,” *IEEE Comput. Graph. Appl.*, vol. 31, no. 1, pp. 20–35, 2011.

[3] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross, “Nonlinear disparity mapping for stereoscopic 3d,” *ACM Trans. on Graphics*, vol. 29, 2010.

[4] The Foundry, “Nuke,” http://www.thefoundry.co.uk/products/nuke/.

[5] K. Quintus and M. Halle, “A composition tool for creating comfortable stereoscopic images,” in *Stereo. Displays and Virtual Reality Sys.*, 2008, pp. 11:1–10.

[6] C. Liu, T. Huang, M. Chang, K. Lee, C. Liang, and Y. Chuang, “3d cinematography principles and their applications to stereoscopic media processing,” in *ACM Multimedia*, 2011, pp. 253–262.

[7] Y. Niu, F. Liu, W. Feng, and H. Jin, “Aesthetics-based stereoscopic photo cropping for heterogeneous displays,” *IEEE Transactions on Multimedia*, vol. 14, no. 3, pp. 783–796, 2012.

[8] C. Chang, C. Liang, and Y. Chuang, “Content-aware display adaptation and interactive editing for stereoscopic images,” *IEEE Transactions on Multimedia*, vol. 13, no. 4, pp. 589–601, 2011.

[9] H. Lin, S. Guan, C. Lee, and M. Ouhyoung, “Stereoscopic 3d experience optimization using cropping and warping,” in *Siggraph Asia Sketches*, 2011, pp. 40:1–2.

[10] Y. Niu, W. Feng, and F. Liu, “Enabling warping on stereoscopic images,” *ACM Transactions on Graphics*, vol. 31, no. 6, pp. 183:1–183:7, 2012.

[11] F. Liu, Y. Niu, and H. Jin, “Keystone correction for stereoscopic cinematography,” in *IEEE Workshop on 3D Cinematography*, 2012, pp. 1–7.

[12] W. Lo, J. van Baar, C. Knaus, M. Zwicker, and M. Gross, “Stereoscopic 3d copy & paste,” *ACM Trans. on Graphics*, vol. 29, no. 6, pp. 147:1–147:10, 2010.

[13] D.M. Hoffman, A.R. Girshick, K. Akeley, and M.S. Banks, “Vergence–accommodation conflicts hinder visual performance and cause visual fatigue,” *Journal of Vision*, vol. 8, no. 3, pp. 1–30, 2008.

[14] R. Allison, “The camera convergence problem revisited,” in *Stereoscopic Displays and Virtual Reality Systems XI*, 2004, pp. 167–178.

[15] D. Sun, S. Roth, and M.J. Black, “Secrets of optical flow estimation and their principles,” in *IEEE CVPR*, 2010, pp. 2432–2439.

[16] J. Harel, C. Koch, and P. Perona, “Graph-based visual saliency,” in *NIPS*, 2007, vol. 19, pp. 545–552.

[17] C.Ó. Conaire, N.E. O’Connor, and A.F. Smeaton, “An improved spatiogram similarity measure for robust object localisation,” in *IEEE ICASSP*, 2007, pp. 1069–1072.

[18] A. Basu, H. Shioya, and C. Park, *Statistical Inference: The Minimum Distance Approach*, CRC Press, 2011.

[19] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2007.

[20] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[21] J. Canny, “A computational approach to edge detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, 1986.