A Stereo Camera Distortion Detecting Method for 3DTV Video Quality Assessment

Quanwu Dong [*] , Tong Zhou [*†] , Zongming Guo [*] and Jianguo Xiao [*]
[*] Institute of Computer Science and Technology, Peking University, Beijing, P. R. China.
E-mail: dqw@pku.edu.cn Tel: +86-10-82529726
[†] Zhongguancun Haidian Science Park Postdoctoral Workstation
[†] Peking University Founder Group Co., Ltd.

Abstract— **For 3DTV systems, camera distortions, such as vertical misalignment, camera rotation, unsynchronized zooming, and color misalignment, are introduced during the process of video capturing using stereo cameras. They are important factors affecting the perceptual quality of 3D videos. This paper proposes a stereo camera distortion detecting method based on the statistical model. Experiment is set up on a database which consists of video clips from an on-air 3DTV channel. Subjective assessment is also performed to find out its relation with the experimental results of the proposed method.**

I. INTRODUCTION

During these years stereoscopic displays have been more and more widely used at home and 3DTV services have been presented to a large audience in many countries. For example, the 3DTV experimental channel of China[1] began broadcasting in Jan., 2012. While providing a dramatic enhancement in the viewer's experience, 3DTV services suffer from low-quality contents. The perceptual evaluation of 3D video content plays a key role for the success of 3DTV systems.

The 3DTV system consists of the chain from content generation, 3D video compression, transmission to 3D video display. The 3D content is exposed to various distortions during these chains [1] [2]. The majority of 3D broadcast content has been captured using stereo cameras. Depending on the precision of the alignment, several camera distortions may be introduced such as vertical misalignment, camera rotation, unsynchronized zooming and color misalignment et al. [3].

The subjective evaluation of 3DTV video is multidimensional. In addition to the visual quality of separated views as traditional 2D image quality evaluation, depth perception , 3D image impairments, and visual comfort should also be considered [1] [3]. There are also some objective 3D video quality assessment methods. Yasakethu et al. applied some 2D objective image quality assessment methods to evaluate the two stereoscopic images separately and then combined them into an overall score as the quality metric of 3D video [4]. An objective assessment algorithm which used the image spatial frequency was proposed in [5]. Anish Mittal et al. proposed a no-reference algorithm to assess the comfort associated with viewing stereo images and videos by extracting statistical features from disparity and disparity gradient maps as well as indicators of spatial activity from images [6]. However, most objective quality assessment methods do not consider the effect of stereo camera distortions.

To detect and calibrate camera distortions, H.-M. Wang et al. proposed a procedure which contains six steps, including feature point extraction, bidirectional feature point matching, relative distance checking, image transformation, hole-filling and reshaping [7]. Other methods were proposed to correct zoom mismatch [8], or to examine the horizontal and vertical disparity characteristics of 3D images [9]. However, the above methods are mainly designed for stereo video calibration, not for video quality assessment.

In this paper, we focus on the impact of stereo camera distortion on 3D video quality during video capturing. A method is presented to extract and match corresponding feature points of the left and right view, and then the statistical characters of the left and right view are calculated to find out the precisely amount of vertical misalignment, camera rotation, unsynchronized zooming, and color misalignment. At last, the method is applied on a database which consists of video clips from the 3DTV experimental channel of China. Subjective assessment shows that the method is effective to find out the distorted shots of the 3D videos.

The rest of the paper is organized as follows. Section II gives a brief discussion on the camera distortions of stereo images. Section III describes the algorithm of the proposed method in detail. The experimental results of the method and the subjective assessment are shown in section IV. We conclude the paper in section V.

II. THE CAMERA DISTORTIONS OF STEREO IMAGES

Most 3D programs broadcasted in 3DTV channels are captured with a dual-camera system, where two cameras are fixed with some sort of stereoscopic 3D rig. If the two cameras are precisely aligned and synchronized, the fusion of the left view and the right view will bring perfect stereo visualization to human beings, as see in Fig 1(a). However, the calibration of the stereo camera is a complex process, and unskilled calibration leads to various camera distortions. Four common stereo camera distortions are discussed here.

This work was supported by National Natural Science Foundation of China under contract No. 61071082 , Beijing Nova program under contract No. 2010B001 and Postdoctoral Foundation of Beijing under contract No.108385

[1] http://3d.tv.cn/

Vertical misalignment appears when the left camera and the right camera are not positioned at the same height. Therefore the corresponding objects in the captured stereo images have vertical displacement. See Fig 1(b). **Camera Rotation** occurs when the stereo cameras rotate around the camera direction and the rotation angles are different. See Fig 1(c). **Unsynchronized zooming** is a kind of stereo distortion when the camera zoom parameters of the two cameras are not changed in unison. It will lead to different scaling between the captured left view and right view. See Fig 1(d).

The above distortions are all geometrical misalignment of the stereo images. **Color misalignment** is caused by different electrical parameters between the stereo cameras. It will lead to different luminance and/or chrominance.

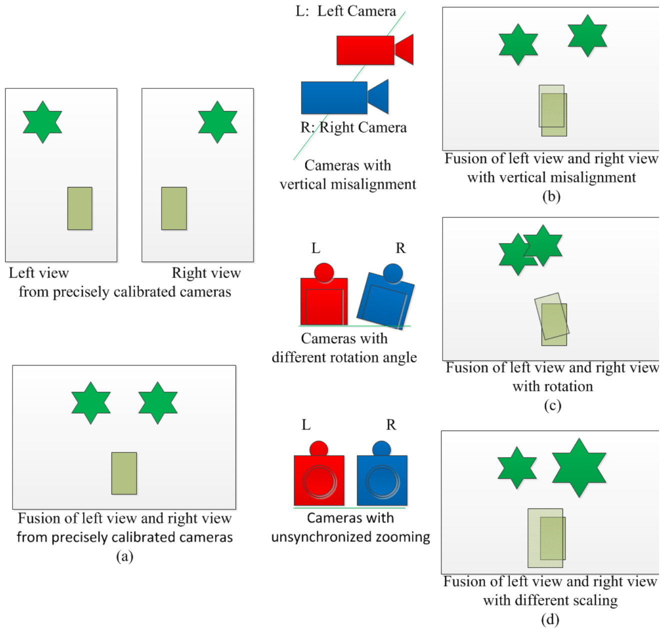

Fig. 1 Geometrical camera distortions [10].

A number of researches have been conducted to find out the relations between the above distortions and the amount of perceptual visual fatigue. For 3DTV systems, it is proposed in [10] that the image vertical displacement should be less than 10 pixels, the rotation angle difference be less than 1°, the scaling factor be less than 1%, and the luminance difference be less than 20%.

III. The Proposed Method

In this section, a method is proposed to detect and measure the above stereo camera distortions. Firstly the stereo image is split in half to get the left image and right image. Then several specific image processing algorithms are applied on the left and right images to extract various features. Lastly the statistical characters of the features are calculated and defined as the precise metrics of the distortions above.

A. Vertical Misalignment Detecting

To estimate the vertical misalignment of the stereo image, we need to find out the corresponding points of the left image and right image at first. There are two kind of stereoscopic correspondence methods: dense matching and sparse matching. The sparse matching method is chosen due to its subpixel precision and reliability. As a widely used sparse matching algorithm, the Scale-Invariant Feature Transform (SIFT) [11] is adopted to get the corresponding feature points, as shown in Fig. 2.

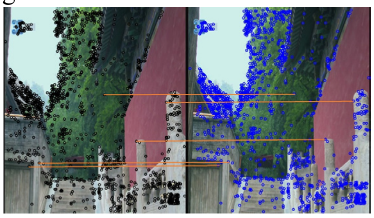

Fig. 2 Extracted feature points and the matched point pairs between the left view and right view of a stereo image.

For every matched point pairs $P_{li}(X_{li}, Y_{li})$, $P_{ri}(X_{ri}, Y_{ri})$, the vertical disparity in pixels is:

$$d_{vi} = Y_{ri} - Y_{li} \tag{1}$$

and the horizontal disparity in pixels is :

$$d_{hi} = X_{ri} - X_{li} \tag{2}$$

where $i = 1, 2, \ldots, N$, and N is the number of all the matched feature point pairs.

We then count the number of the matched point pairs with the same vertical disparity value. Those outlier matches (d_{vi} >20 or d_{vi} <20) are omitted since a normal stereo camera will not generate such vertical disparity. Experiment shows that the outlier matches occupy very small fraction of the overall matched point pairs. A histogram is given as shown in Fig. 3, where N_{dv} denotes the number of the point pairs for the vertical disparity d_v.

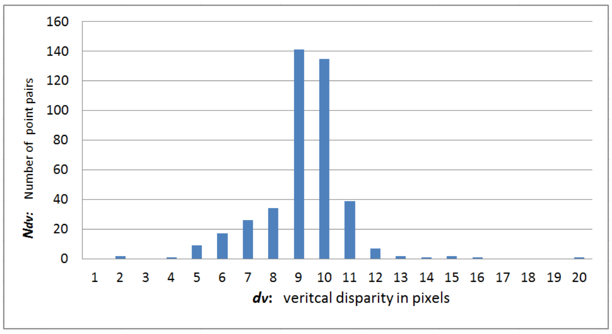

Fig. 3 Vertical disparity histogram.

If there is no vertical misalignment between the stereo images, the maximum of N_{dv} will appear near the position where d_v = 0. Otherwise, the horizontal position which maximizes N_{dv} indicates the vertical displacement of the stereo images. Therefore, the vertical misalignment can be assessed by the metric VD:

$$VD = dv^*, \text{ where } N_{dv^*} = max\{ N_{dv} \} \tag{3}$$

dv^* is the vertical disparity value which most matched point pairs possess. For example, VD = 9 in Fig. 3.

We can also count the number of the matched point pairs with the same horizontal disparity value, and give the histogram as shown in Fig. 4. It can be used to assess the distribution of horizontal disparity and excessive parallax.

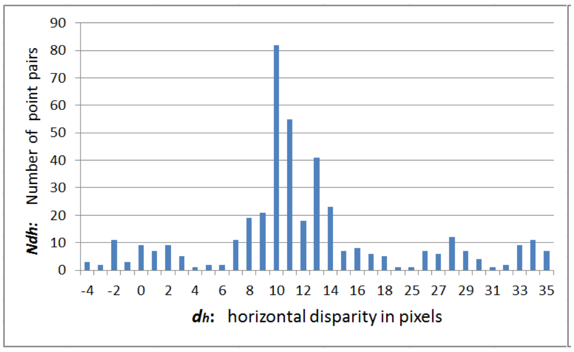

Fig. 4 Horizontal disparity histogram.

B. *Unsynchronized Zooming Detecting*

The unsynchronized zooming detecting algorithm is also based on the result of the SIFT matching method above. In addition to the coordination of the matched points, the SIFT method can also extract the scaling of the matched points, denoted as S_{li}, S_{ri} for point P_{li} , P_{ri}. The Scale Factor (SF) of the point pair (P_{li} , P_{ri}) is defined as :

$$SF_i = S_{li} / S_{ri} \qquad (4)$$

if $S_{li} = S_{ri}$, then $SF_i = 1$. This means the left view and right view have the same scaling, and the zooming of the stereo cameras is perfectly synchronized. As [10] indicated, SF_i should be kept between 0.99 and 1.01.

Because there is not an apparent maximum value in all SF values of the matched point pairs, we define the unsynchronized zooming metric SFC as:

$$SFC = N_g / N \qquad (5)$$

where N_g is the number of the "good" point pairs with $0.99 < SF < 1.01$, and N is the number of all the matched feature point pairs.

C. *Camera Rotation Detecting*

To get the rotation angle between the left and right images, we rotate the left image by θ_i, where:

$$\theta_i = -10^\circ + i * 0.1^\circ, \ i = 0, 1, 2, 3, \ldots 20 \qquad (6)$$

Then the difference of the rotated left image and the original right image is calculated by :

$$\Delta I_{\theta_i} = \sum_{x,y} \left(I_{l_{\theta_i}}(x,y) - I_r(x,y) \right) \Big/ N_g \qquad (7)$$

where $I_{l\theta_i}(x,y)$ represents the (R,G,B) value of the rotated left image by θ_i at point (x,y). $I_r(x,y)$ represents the (R,G,B) value of the original right image at point (x,y). The edge of the rotated image contains some blank points, which cannot find the corresponding points in the right image. Such points are omitted in this algorithm, and N_g is the number of those "good" points. This process can be illustrated in Fig.5.

The Rotation Angle (RA) between the left and right images can be estimated by the angel θ_i^* which minimize ΔI_{θ_i}.

$$RA = \theta_i^*, \text{ where } \Delta I_{\theta_i^*} = min\{ \Delta I_{\theta_i} \} \qquad (8)$$

Since there are horizontal disparity between the left view and the right view, they are not exactly the same and normally $\Delta I_{\theta_i^*}$ will not be zero. The angel θ_i^* is an approximation of the real rotation angle.

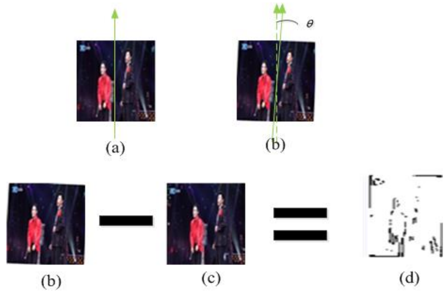

Fig. 5 Rotation detecting method. (a) original left image. (b) left image rotated by θ. (c) original right image. (d) the difference between the rotated left image and the original right image

D. *Color Misalignment Detecting*

We estimate the color misalignment by summing up the (R,G,B) value of the stereo images separately, then calculate the difference between the corresponding sum value of the left image and the right image. The color misalignment can be denoted by the following metrics:

$$\Delta R = \frac{\left| \sum_{x,y} R_l(x,y) - \sum_{x,y} R_r(x,y) \right|}{\sum_{x,y} R_l(x,y)} \times 100\%$$

$$\Delta G = \frac{\left| \sum_{x,y} G_l(x,y) - \sum_{x,y} G_r(x,y) \right|}{\sum_{x,y} G_l(x,y)} \times 100\%$$

$$\Delta B = \frac{\left| \sum_{x,y} B_l(x,y) - \sum_{x,y} B_r(x,y) \right|}{\sum_{x,y} B_l(x,y)} \times 100\% \qquad (9)$$

IV. EXPERIMENTAL RESULTS

A. *3DTV Video Database*

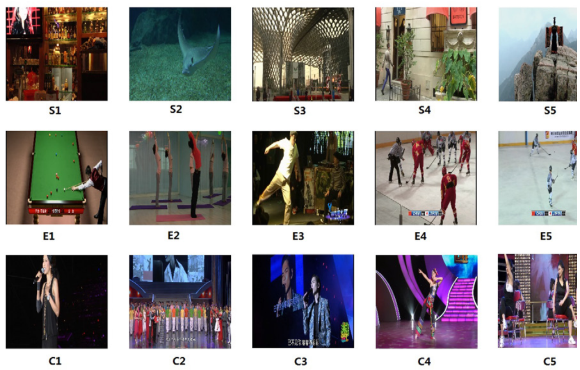

Fig. 6 Experimental clips from the 3DTV video database.

Our 3DTV video database consists of clips which are captured from an on-air TS video stream by a satellite receiver. The videos are all in full HD 1080i /25fps format, with a bit rate of 16M bps. The contents are from the 3DTV experimental channel of China, which are produced by six TV stations across China. About 300 hours of stereoscopic videos have been captured since 2012.

Fifteen video clips are chosen for the experiment, as shown in Fig. 6. They are divided into 3 groups: group S for landscape, group E for sport and group C for entertainment. The duration of each clip is 30 seconds.

B. Subjective Quality Assessment Experiment

A 55 inch full HD active shutter 3D LED TV Display (SONY KDL-55EX720) is used for the subjective assessment. The video clips are played back by a HD player (HiMedia HD910A 3D Player) .

Ten non-expert assessors participate in the subjective experiment. The viewing distance is 3 times the height of the screen. They watch any of the 3 clip groups and score up their quality on a scale of 1 to 5, which are labeled with "bad", "poor", "fair", "good", and "excellent".

C. Experimental Results

The metrics of different stereo camera distortions are calculated for each clip, see *⊿R, ⊿G, ⊿B, RA, SFC* in table I. As the vertical displacement of the stereo images in all the chosen clips is below 5 pixels, we do not list the metric *VD* in the table. The experimental results of the subject assessment are shown as *MOS* in table I.

TABLE I
METRICS OF THE POPOSED METHOD AND RESULTS OF THE SUBJECTIVE ASSESSMENT

Clip#	*⊿R*	*⊿G*	*⊿B*	*RA*	*SFC*	*MOS*
S1	2.68	3.91	3.73	0.2	0.28	3.88
S2	0.13	0.12	0.80	0	0.20	2.88
S3	1.41	1.24	2.24	0	0.24	2.25
S4	0.36	5.94	1.59	0.9	0.31	4.00
S5	7.25	1.48	3.00	2.4	0.19	2.00
E1	0.28	0.39	0.36	1	0.26	3.44
E2	1.49	1.36	1.68	0	0.23	2.44
E3	2.30	4.22	3.57	2.9	0.21	3.22
E4	3.22	1.16	6.34	0.5	0.30	3.56
E5	5.58	5.71	5.68	1.1	0.28	2.44
C1	0.42	0.75	0.13	0.5	0.25	2.80
C2	1.71	4.68	0.37	0	0.23	3.20
C3	6.16	0.76	0.90	0.9	0.21	3.20
C4	3.11	9.21	0.19	0	0.22	2.60
C5	4.53	1.15	11.32	0.1	0.34	2.80

From table I we can find that color misalignment and rotation distortion are common in the 3DTV videos. Some clips have more than one kind of distortion, such as S5, and their *MOS* values are low. Slightly distorted images have little effect on the perceptual quality of the video, such as S4.

V. CONCLUSIONS

This paper presents a method to detect four types of stereo camera distortions during 3D video capturing: vertical misalignment, unsynchronized zooming, camera rotation, and color misalignment. The method is based on extracting statistical characters of the stereo images. It is effective to find out the distorted frames in the stereo videos. With the rapid development of 3DTV, the method can be used in the quality control process of broadcast organizations to help producing better 3D contents.

REFERENCES

[1] Lydia M. J. Meesters, Wijnand A. IJsselsteijn, and Pieter J. H. Seuntiëns, "A Survey of Perceptual Evaluations and Requirements of Three-Dimensional TV", *IEEE Trans. Circuits Syst. Video Technol.,* vol. 14, no. 3, March 2004.

[2] Junming Zhou, Gangyi Jiang, Xiangying Mao, Mei Yu, Feng Shao, Zongju Peng, et al., "Subjective Quality Analyses of Stereoscopic Images in 3DTV System", *IEEE Visual Communications and Image Processing (VCIP)*, 2011.

[3] Quan Huynh-Thu , Patrick Le Callet, Marcus Barkowsky, "Video Quality Assessment: From 2D TO 3D –Challenges and Future Trends", *IEEE International Conference on Image Processing (ICIP)*, 2010.

[4] S. L. P. Yasakethu, C. T. E. R. Hewage, W. A. C. Fernando, and A. M. Kondoz, "Quality Analysis for 3D Video Using 2D Video Quality Models", *IEEE Transactions on Consumer Electronics*, Volume: 54 , Issue: 4 , 2008.

[5] Feng Lu, Haoqian Wang, Xiangyang Ji, Guihua Er, "Quality assessment of 3D asymmetric view coding using spatial frequency dominance model", *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, 2009.

[6] A. Mittal, A.K. Moorthy, J. Ghosh, A.C. Bovik, "Algorithmic Assessment of 3D Quality of Experience for Images and Videos", *IEEE Digital Signal Processing Workshop*, *Sedona, Arizona,* January 04-07, 2011.

[7] H.-M. Wang, C.-W. Chang, J.-F. Yang, "An effective calibration procedure for correction of parallax unmatched image pairs", *IET Image Processing.*, Vol. 3, Iss. 2, pp. 63–74, 2009.

[8] Colin Doutre, Mahsa T. Pourazad, Alexis Tourapis, Panos Nasiopoulos and Rabab K. Ward, "Correcting unsynchronized zoom in 3D video" , *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2010.

[9] Donghyun Kim, Kwanghoon Sohn, "Depth adjustment for stereoscopic image using visual fatigue prediction and depth-based view synthesis" , *IEEE International Conference on Multimedia and Expo (ICME)*, 2010.

[10] The state administration of radio film and television of the People's Republic of China, *3DTV Technology Guidance Part One: Program Production and Broadcast*, 2012.

[11] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, Nov. 2004.