Liu, M., Ndjiki-Nya, P., Le Quintrec, J-C., Nikolaidis, N., & Pitas, I. (2015). Efficient automatic detection of 3D video artifacts. In 2014 IEEE 16th International Workshop on Multimedia Signal Processing (MMSP 2014): Proceedings of a meeting held 22-24 September 2014, Jakarta, Indonesia. (pp. 1-6). Institute of Electrical and Electronics Engineers (IEEE). DOI: 10.1109/MMSP.2014.6958787

Efficient Automatic Detection of 3D Video Artifacts

Mohan Liu[#1], Ioannis Mademlis[†2], Patrick Ndjiki-Nya[#1], Jean-Charles Le Quintrec[*3], Nikos Nikolaidis[†2], Ioannis Pitas[†2]

[#] *Interactive Media - Human Factors Department, Fraunhofer Institute for Telecommunication - HHI*
Einsteinufer 37, 10629 Berlin, Germany
[1] {mohan.liu, patrick.ndjiki-nya}@hhi.fraunhofer.de

[†] *Department of Informatics, Aristotle University of Thessaloniki - AUTH*
Box 451, 54124 Thessaloniki, Greece
[2] {imademlis, nikolaid, pitas}@aiia.csd.auth.gr

[*] *ARTE G.E.I.E., Association Relative à la Télévision Européenne*
4 Quai Du Chanoine Winterer CS 20035, 67080 Strasbourg Cedex, France
[3] jean-charles.lequintrec@arte.tv

***Abstract*—This paper summarizes some common artifacts in stereo video content. These artifacts lead to poor even uncomfortable 3D viewing experience. Efficient approaches for detecting three typical artifacts, sharpness mismatch, synchronization mismatch and stereoscopic window violation, are presented in detail. Sharpness mismatch is estimated by measuring the width deviations of edge pairs in depth planes. Synchronization mismatch is detected based on the motion inconsistencies of feature points between the stereoscopic channels in a short time frame. Stereoscopic window violation is detected, using connected component analysis, when objects hit the vertical frame boundaries while being in front of the virtual screen. For experiments, test sequences were created in a professional studio environment and state-of-the-art metrics were used for evaluating the proposed approaches. The experimental results show that our algorithms have considerable robustness in detecting 3D defects.**

I. Introduction

Three-dimensional (3D) videos are not only a big success in cinemas, but also entered into ordinary households. 3D experience is highly related to the quality of the 3D content. Thus, quality assessment of 3D videos has become a rising topic. In comparison to two-dimensional (2D) image quality assessment, 3D quality is also highly related to depth perception and the visual comfort. Although techniques for automatic quality assessment of 2D images have been extensively developed for years, prior research has shown that 2D quality metrics cannot be directly used to estimate 3D quality features [1].

There are two major ways to create 3D sequences: filming with stereo cameras and converting from 2D videos using depth maps. Quality control is an important task in such workflows. Although 3D quality assessment has been widely studied for several years, there is still no formal definition of 3D defects. In the MSU project [2], eight measures for 3D artifacts are proposed. Similarly, fifteen 3D quality issues have been identified in the Certifi3D project [3].

This paper proposes offline automatic analysis methods for the detection and assessment of 3D artifacts. Three important artifacts, i.e. Sharpness Mismatch (SM), SYNchronization Mismatch (SYNM) and Stereoscopic Window Violation (SWV), are further detailed. The proposed SM detector framework is based on measuring the width deviations between edge pairs in valid depth planes. The SYNM is estimated based on statistics on motion inconsistencies of object feature points across stereoscopic channels. SWVs happen when objects appearing in front of the virtual screen hit the vertical frame boundaries. Connected component analysis is used for detecting SWV. A robust disparity estimation algorithm [4], which computes Disparity Maps (DM) in the horizontal and vertical directions without rectifications [4], is integrated in the proposed algorithms. The remaining of this paper are organized as follows: Section 2 describes the proposed approaches in detail; Section 3 discusses the experimental results; Section 4 concludes the paper and describes future work.

II. Proposed Approaches

For observers, stereo 3D artifacts are not only undesirable but sometimes also painful. Common defects in real stereo 3D videos include (similar definitions of some defects are also introduced in [2] and [3]):

- ***Vertical misalignment***: For depth illusion, vertical disparity is unwanted. Imperfect horizontal alignment of stereo cameras can cause this defect.
- ***Sharpness mismatch***: Sharpness mismatch can be caused, among others, by focus/aperture setup errors, inconsistent light environment, compression, denoising.
- ***Colorimetric mismatch***: Common causes of colorimetric mismatches are different point-of-views, changing light conditions, malfunctioning or non-calibrated acquisition system or even the use of bad color grading in post-processing.
- ***Synchronization mismatch***: Non-genlock cameras or bad post-processing can cause asynchronism artifacts between stereo channels.
- ***Hyper divergence/convergence***: Excessive positive or negative parallaxes on inappropriate viewing devices can lead to these artifacts.

- ***Cross-talk level***: Artifact caused by imperfect view separation, such that one view can be partially seen in the other view.
- ***Stereoscopic window violation***: Objects appearing in front of the virtual screen in theatre space and hit the left or right frame boundary cause retinal rivalry, that is erroneously interpreted by the viewer as occlusion.
- ***Bent window effect***: Sometimes, objects appearing in front of the virtual screen in theatre space extend vertically across the entire frame and hit both the top and bottom frame boundaries. This is interpreted by the brain as an occlusion cue, causing the perception of the stereoscopic window as being bent towards the viewer.
- ***Depth jump cut***: During editing, video cuts between two shots with very different average depth cause a temporary loss of the viewer's 3D perception.

There are also other less common defects, e.g. view reversal and reflection, and 2D to 3D conversion defects, e.g. depth mismatch and visual mismatch. In this paper, techniques developed for sharpness mismatch, synchronization mismatch and stereoscopic window violation detection are presented. There is no significant connection among these three artifacts.

A. Disparity map correction

The automatically estimated DMs are usually noisy, which must be corrected for further use. A valid disparity mapping in the horizontal direction from the Left view to the Right view ($L2R$) views can be defined as

$$|DM_{L2R}(i,j) - DM_{R2L}(i, j + DM_{L2R}(i,j))| \leq \delta, \quad (1)$$

where δ denotes the disparity estimation error tolerance and (i,j) denotes the pixel coordinates in the disparity map. The validation of disparity maps from the Right view to the Left view ($R2L$) is similar to eq. 1.

B. Sharpness mismatch

The proposed approach estimates the SM by analyzing width deviations of corresponding edge pairs. According to the Epipolar geometry, the edge widths of edge pairs in a depth plane are consistent between LR views, when the focuses of stereo cameras are well calibrated. SM usually leads to cross-talk/ghost effect and unexpected blurriness, which can impair the 3D experience for observers.

We use the Sobel filters to extract edge pixels E. They are further segmented in depth planes based on the estimated disparities as

$$\tilde{E}^d(i,j) = E(i,j) \ \& \ (DM(i,j) = d), \quad (2)$$

where d denotes an estimated disparity value.

Only the vertical edge widths are measured in this work since the major disparities occur in the horizontal direction. For each edge pixel, a pixel sequence centered at the target edge pixel is then selected. The width of a pixel sequence is set to 64 pixels given human visual activity. Human visual activity is related to the size of the fovea region, which covers about 2% of the visual angle [5]. The central visual field covers approximately 30^o [5]. The width of an activity region on a high-definition (HD) image can be approximately calculated as about 64 pixels at a comfortable viewing distance. The method proposed in [6], is used to measure the edge width by locating the pixel positions of the local minimum and the local maximum of luminance intensities centered on the target edge pixel within the allocated pixel sequence. The edge width is then computed as the Euclidean distance between the positions of the local minimum and the local maximum pixels.

The perceived blurriness is also related to the local contrast [7]. The edge width of just noticeable blur w_{JNB} [7] is estimated in a 64×64 pixels block. If the deviations of edge widths are larger than w_{JNB}, the SM artifact cannot be noticed. The cumulative probability of noticeable SM can be calculated as

$$P_{sm} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{I}_{dw > w_{JNB}}, \quad (3)$$

where N denotes the number of edge pixels and dw denotes the width deviation of an edge pair between stereoscopic views. $\mathbf{I}_{dw>w_{JNB}}$ is an indicator function if the condition $dw > w_{JNB}$ is met. However, the cumulative probability must be corrected considering the lack of edge pixels caused by the disparity estimation errors. Then, the Probability of Sharpness Mismatches PSM is estimated and smoothed with the correction coefficients k, which are calculated considering the number of valid disparity mappings, as

$$PSM = 1 - exp\left(-\frac{P_{sm}^2 + k^2}{2 \cdot \tilde{\sigma}}\right), \quad (4)$$

where $\tilde{\sigma}$ denotes the standard deviation between P_{sm} and k.

C. Synchronization mismatch

The annoyance of the synchronization distortion depends on the 3D scene. SYNM can be very annoying in a shot with strong object motions. If the scene is still, SYNM is almost imperceptible. Thus, the first step of our approach is to analyze the perceptibility of SYNM. Furthermore, shot detection is required for the framework.

Single-valued spatial and temporal perceptual information, defined by ITU [8], is computed to estimate the perceptibility of SYNM. The Spatial Information (SI) describes the level of spatial details in textured images and is calculated as the maximum standard deviation of each Sobel-filtered frame within a time duration. The Temporal Information (TI) describes the strength of the motion in a sequence and is computed as the maximum standard deviation of the pixel luminance differences at the same location between two neighboring frames. The SI and TI values for a shot are respectively calculated for the LR channels. The reversed perceptibility score (RPS) is defined as

$$RPS = max\left\{\frac{SI_c}{TI_c}\right\}, \ c = \{L, R\}. \quad (5)$$

If RPS is smaller than a threshold, it is necessary to detect the synchronization mismatch.

SYNM is detected at frame and shot levels. If a frame is synchronous at time point t, the object motions in a depth plane between the LR views are consistent. Conversely, motion inconsistency can be observed between two asynchronous views. The computation of motion consistencies is based on relative displacements of feature point pairs, which are extracted and matched using SIFT [9] as well as RANSAC [10]. The matched feature points are segmented according to validated depth planes. The relative displacement $\tilde{d}$ of a matched feature point $fp(i)$ in a depth plane D_j, $j \in [0, 255]$, is determined as

$$\tilde{d}(fp(i)) = P_L - DM_{L2R}(P_L) - P_R, fp(i) \in D_j, \quad (6)$$

where P_L and P_R denote the coordinates of $fp(i)$ in the L and R views respectively. The variances of the relative displacements of all feature points are calculated to describe the motion consistencies. Please note that if there is only slight motion in depth but no noticeable motion in the horizontal (H) and vertical (V) directions between neighboring frames, the corresponding SI is significant larger than TI.

To measure the synchronization of frame $f(t)$ at time point t, two neighboring frames $f(t-1)$ and $f(t+1)$ are required. The synchronism probability is measured from the left view ($f_L(t)$) to the right views ($f_R(t-1, t, t+1)$) and from the right view ($f_R(t)$) to the left views ($f_L(t-1, t, t+1)$) respectively. The motion related displacements of the feature points are decomposed in HV directions. Then, HV synchronization probabilities are respectively estimated based on the variances of the motion displacements of the matched feature points. Hence, six histograms of the displacement variances in all valid depth planes are constructed to rank the synchronization probabilities between the frame $f(t)$ of one view and the frames $f(t-1)$, $f(t)$, $f(t+1)$ of the other view. Z_α, which is computed considering a confidence level α, is used as the threshold for the estimation of the synchronization probability. The synchronization probability in one direction p_r is calculated as,

$$p_r = \frac{\sum_{i=1}^{n} h_i}{n}, h_i \leq Z_\alpha, r = \{H, V\} \quad (7)$$

where n is the total number of the variance histograms h. In consideration of the statistical accuracy, the outliers of the variance histogram are first detected and removed. The overall synchronization probability p is computed as the geometric mean of p_r, since the geometric mean is more robust than the arithmetic mean, if outliers can be observed in test samples [11]. A frame is judged as synchronous when the maximum p between views occur at time point t in both estimation direction:

$$\begin{cases} max\{p(f(t)_L, f(t')_R)\} = p(f(t)_L, f(t)_R), \\ max\{p(f(t)_R, f(t')_L)\} = p(f(t)_R, f(t)_L), \end{cases} \quad (8)$$

where $t' = \{t-1, t, t+1\}$. If most of the frames within a shot are estimated as asynchronous, the shot is judged as asynchronous.

D. Stereoscopic window violation

In 3D cinematography, we observe the 3D world through the so-called Stereoscopic Window (SW) [12], namely the TV or cinema screen. In other words, the viewer watches objects floating in a space defined by the screen edges. If the left disparity of a 3D point is positive/zero/negative, the eyes converge to a point either behind the screen, on screen or within the theatre space (in front of the screen), respectively. Retinal rivalry occurs on the left or right frame edges, when object regions positioned close to the left image's left or right border do not have correspondence (are not displayed) in the right frame and vice versa. For objects with zero disparity, no retinal rivalry is observed. When an object region is cut off by the edge of the display, it results in the so-called Stereoscopic Window Violation (SWV) and is interpreted as occlusion by the viewer.

A SWV does not create any problems when it occurs behind the screen (i.e., for objects with positive left disparity), because both disparity and occlusion cues dictate that the object lies behind the screen. However, when SWV involves objects perceived as appearing in front of the screen (i.e. they have negative left disparity) the occlusion cue conflicts the disparity one. Generally, as occlusion supersedes the disparity cue, the object is finally perceived as lying behind the screen plane [12]. The above are true for a mild SWV, where only a small region of the object that interferes with the left or right frame edge is missing from the other image. In a severe SWV, the missing object region is so extended, that the human brain cannot fuse the images and eventually see 3D.

SWV in negative disparities is not only undesirable, but may also prove painful. The rule regarding SWV states that a cinematographer has to avoid breaking the stereoscopic window, while an object is being filmed with negative left disparity. There is one notable exception, related to object speed [13]. Objects entering or exiting the frame in no more than half a second cause no problem, since, by the time the brain localizes the object in front of the screen, the entire object is either fully visible in the frame or has disappeared, respectively.

A simple, yet effective, algorithm that detects the Stereoscopic Window Violation using disparity maps has been developed in this work. We assume the existence of left and right dense disparity maps for each stereoscopic video frame, i.e., $DM_{L2R}(u, v)$ and $DM_{R2L}(u, v)$, $u = 0, ..W-1$, $v = 0, ..H-1$, where W, H are the width and height of the video frame (in pixels). At the first step of the algorithm, pixels u, v are selected, having left disparity $DM_{L2R}(u, v) < -T_1$ and right disparity $DM_{R2L}(u, v) > T_1$. In order to exclude objects that do not appear in front of the screen, we set the threshold T_1 to a suitable value and perform connected component analysis with an 8-point neighbourhood to extract objects (connected components) that are displayed significantly in front of the screen. A value of $T_1 = 0.0025W$ worked well in our experiments. To reduce noise, objects with small width (less than T_w) or height (less than T_h) are rejected. Threshold values

of $T_w = 0.02W$ and $T_h = 0.04H$ have been found to work well. The detected objects are then enclosed into rectangular bounding boxes (Regions of Interest, ROIs). Thus, two sets of ROIs $\mathbf{R}^r = \{R_1^r, R_2^r, ..., R_n^r\}$ and $\mathbf{R}^l = \{R_1^l, R_2^l, ..., R_k^l\}$ are created for the left and right channel, respectively. These ROIs are represented by their upper left and lower right coordinates $[X_{i,min}^j, Y_{i,min}^j]^T$ and $[X_{i,max}^j, Y_{i,max}^j]^T$, where $j = \{r, l\}$ and i is the ROI index.

Two types of disturbing SWVs can be defined. In the first type, namely left SWV, the violation occurs on the left frame border, since there is a region in the left image which is missing from the right one. Its detection is performed as follows. If one or more object ROIs R_i^r, with disparity characteristics such as those previously described, lie on the left border of the right image, that is, if $X_{i,min}^r = 0$, a SWV is present. This is because $X_{i,min}^l = X_{i,min}^r + DM_{R2L}(i,j) > 0$ and, thus, the region $[0, DM_{R2L}(i,j)]$ in the left image is not present in the right one. In order to reduce false alarms, arising from inaccuracies in the disparity maps, another condition is introduced. The number of pixels that belong to the object in the two leftmost ROI columns must be greater than a threshold T_2, expressed as a percentage of the ROI height, to decide that this object signals a SWV. In our experiments T_2 is set to $0.3h_{ROI}$, where h_{ROI} is the ROI height.

A similar procedure is followed for the detection of a right SWV. In this case, a region appearing in the rightmost border of the right image is absent from the left image. Thus, if one or more object ROIs detected in the left disparity map R_i^l lie on the right border of the left image, i.e., if $X_{i,max}^l = W - 1$, a SWV is present. This is because $X_{i,max}^r = X_{i,max}^l + DM_{L2R}(i,j) < W - 1$. Therefore, the region $[W + DM_{L2R}(i,j), W - 1]$ in the right image is not present in the left one. The false alarm reduction approach regarding small regions (noise) is applied to right SWV detection, as well.

When a left or right SWV of duration d_{SWV} frames is detected, the condition $d_{SWV} > \frac{fps}{2}$ is checked, where fps is the video frame rate, to determine whether the violation is perceived as annoying or not. The satisfaction of this condition implies that the duration of the violation is more than half a second.

III. Experimental Results

In order to evaluate the performance of the proposed approaches, several HD 3D sequences (cf. TABLE I) were used for the experiments. The disparity maps of all test sequences were automatically estimated by the algorithm proposed in [4]. δ in Eq. 1 was set to 1 in the experiments considering the clipping problem between the integer and floating values.

S1 - S3 were realized in a professional studio using two Sony HDC 1400 cameras and broadcasted by ARTE (Association Relative à la Télévision Européenne). The cameras were set in convergence mode. Focuses were manually calibrated with a sharpness chart. Luminance and color parameters were corrected with Remote Control Panel (RCP) and the use of a waveform monitor and a video monitor. Adobe Première was used to remove the useless parts at the beginning and the end of each sequence. The focus of the left camera for S2 and S3 were manually modified to generate global sharpness mismatches. There is no modification of the right camera setups among S1, S2 and S3. The focus of the left camera was manually set to +2m for S2 and to +3m for S3 in comparison to the right camera. There is no camera motion in S1, S2 and S3. S4 - S6 contain densely meshed textures and local motion types. S7 is a sequence containing SWVs.

TABLE I
ORIGINAL TEST SEQUENCES USED IN THE EXPERIMENTS

ID.	Seq. name	ID.	Seq. name
S1	ARTE studio setup	S2	ARTE left camera +2m
S3	ARTE left camera +3m	S4	Badminton
S5	BeergardenNoFlag	S6	BeergardenFlag
S7	The Magician		

A. Results of sharpness mismatch estimation

The proposed SM framework was evaluated with two kinds of experiments. One experiment used S1 - S3 to evaluate the performance on global SMs. Fig. 1 (a) - (c) show example views with focus mismatches. S3 contains some local object motions. The aim of the second experiment is to measure the performance on depth-of-field (DOF) mismatches. Experimentally, DOF mismatches were generated in the right channels of S4 - S6 by Gaussian low-pass filters from a specific depth plane on (cf. Fig. 1 (d) - (f)) since defocus-based effects of lens aberrations can be modeled as a Gaussian blur [14]. The standard deviations σ of the Gaussian low-pass filter was varied from 0.0 to 6.0 with a step width of 0.4. The images are totally distorted if $\sigma > 6.0$. The radius of Gaussian filters was set to 3σ. Deviations ($CPBD_d$) of sharpness scores, which were estimated by the sharpness metric CPBD [15], were used as a reference in the experiment. CPBD is a well performing 2D objective no-reference sharpness metric for Gaussian blur distortions.

Mean SM scores of S1 - S3 are shown in TABLE II. It can be observed that both the proposed approach (PSM) and $CPBD_d$ can detect slight SMs between stereo cameras, as the SM scores of both metrics increase with the enlargement of focus deviations. Although S3 contains object motion, the variances of SM scores of both metrics are very small. Thus, local motion does not affect SM predictions estimated by the PSM and $CPBD_d$, when small variances are considered. However, PSM is more sensitive to SMs than $CPBD_d$ as can be seen in TABLE II. The range of the both metrics is [0, 1].

The results of the second experiment are shown in Fig. 2. The SM scores of PSM, monotonically increase with the increase of the standard deviations of Gaussian low-pass filters. Significant changes can be observed when $\sigma = [1.2, 4.4]$. However, PSM is not very sensitive for slight blur distortions (e.g $\sigma < 1.2$). Significant modifications of SM scores of S6 can be observed even from $\sigma = 1.6$ since S6 contains homogenous textures and strong local motion blur. CPBD

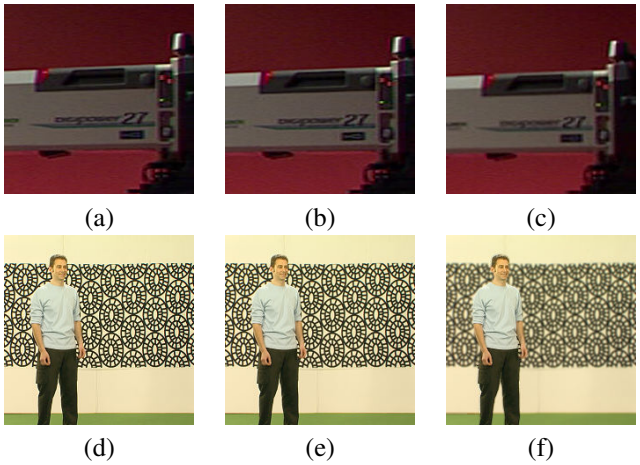


Fig. 1. Focus mismatch of stereo cameras. (a) a frame of an example left view of S1; (b) a frame of an example left view of S2 with camera focus +2m; (c) a frame of an example left view of S3 with camera focus +3m; (d) a frame of an original right view of S4; (e) a frame of local blurred right view with $\sigma = 2.0$ from positive parallax +5 pixels; (f) a frame of local blurred right view with $\sigma = 4.0$ from positive parallax +5 pixels. (The sharpness mismatch might be difficult to distinguish in the printed version)

produces inconsistent results for local SMs. The SM scores of S4 decrease between $\sigma = [0.4, 0.8]$. The scores of S6 decrease from $\sigma = 2.0$. The result curve of S5 is even worse and shows wavy behavior.

TABLE II
MEAN SCORES OF SHARPNESS MISMATCHES OF GLOBAL MISMATCHES

Metric	PSM		$CPBD_d$	
	score	variance	score	variance
S1	0.201	4.003e-5	0.019	1.752e-5
S2	0.240	5.052e-5	0.020	2.479e-5
S3	0.274	5.343e-5	0.023	2.664e-5

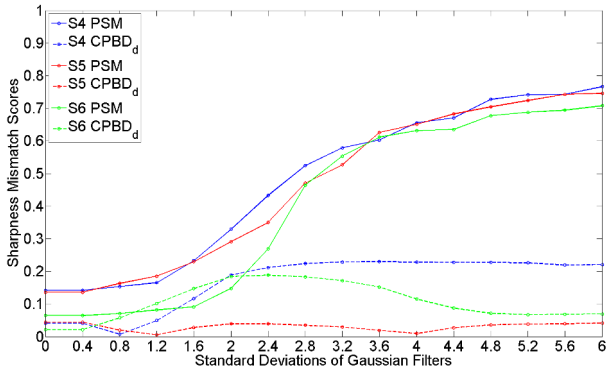
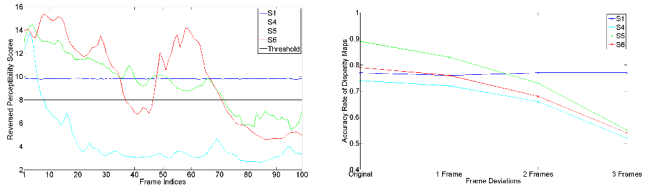


Fig. 2. Sharpness mismatch scores of local distortions

B. Results of synchronization mismatch estimation

The proposed SYNM approach was implemented and evaluated with 16 sequences. S1, S4, S5, S6 were selected for the evaluation (S2 and S3 are similar to S1 but with noticeable sharpness mismatches). 12 distorted sequences were generated from the original test sequences. S1 is a shot of a still scene. Other selected test sequences contain different object motions. Some sequences also contain several still scenes. The synchronization distortion was manually generated by temporally shifting the view pairs of original sequences 1, 2 and 3 frames. The frame rate of all test sequences is 30FPS. Thus, a 1 frame shift corresponds 0.033s temporal deviation.

According to the study of the connection between SI and TI in [16], RPS of the synchronization analysis was set to 8. If $RPS > 8$, the analysis for the corresponding frame is unnecessary. The trends of the RPSs of the test sequences are illustrated in Fig. 3 (a). The RPSs of S1 are almost constant since S1 is a shot of still scene. Several still scenes can be observed in S4 - S6. The detection results of these still frames were used as references in further evaluations. In order



(a) (b)

Fig. 3. Reversed perceptibility scores and accuracy rate of disparity maps. (a) RPS of the first 100 frames of the test sequences; (b) Accuracy rate of disparity maps with different frame shifts.

to estimate the synchronization probabilities, $\alpha = 25\%$ was selected as the confidence level in the experiments considering five-number summary [17]. If more than 75% of the frames ($RPS < 8$) within a shot is detected as asynchronous, the corresponding shot is judged as asynchronous. Precision-recall scores are used to evaluated the performance. If a shot is asynchronous, all frames within it are also considered as asynchronous. Thus, the computation of the precision is at the frame level, and the recall is calculated at the shot level. Fig. 4 shows the performance of the proposed approach with and without the perceptibility check. About 20%-30% gains of the precision scores with the perceptibility check can be observed considering the perceptibility correction of still scene frames. The detection accuracy reduces from 1 frame shift to 3 frames shift due to the increase of estimation errors in the disparity maps with increasing temporal deviations (cf. Fig. 3 (b)). Massive errors in the disparity maps reduce the valid samples of matched feature points. The accuracy rate of disparity maps of the S1 sequences are independent of frame shifts. If the temporal deviation is so large that the estimated disparity maps are too noisy, the proposed approach fails. The precision-recall scores of the original sequences are slightly lower than that of the 1 frame shifted sequences according to the constant setting of confidence level α. However, this estimation error does not affect the result at the shot level. The proposed approach did not detect the mismatch of S1 sequences. Thus, the recall of frame shifted shots without perceptibility checks are lower than that with the perceptibility checks.

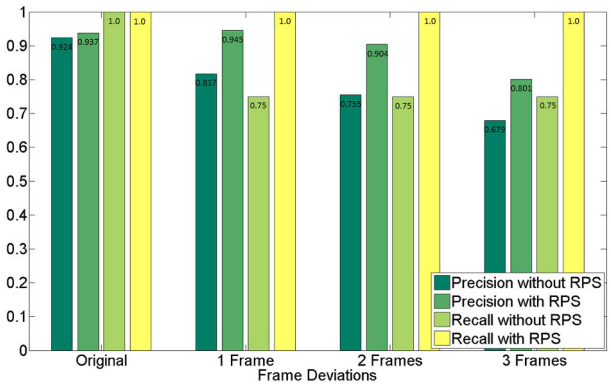


Fig. 4. Precision-recall scores of the synchronism analysis

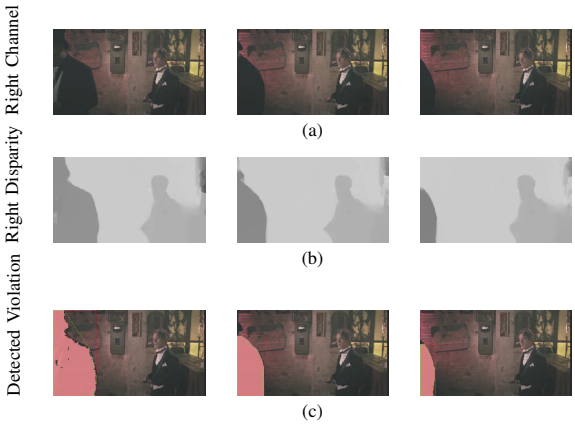


Fig. 5. A left stereoscopic window violation example. Right-channel images, the corresponding right disparity maps and right-channel images with the object (connected component) causing the detected violation marked in red, are shown.

C. Results of stereoscopic window violation

An example of left stereoscopic window violation is shown in Figure 5. The beginning of the violation is signalled when the man with the hat hits the left frame border while being in front of the screen, in the sixth frame ($n = 6$). The detected SWV ends when the man disappears at frame $n = 36$. The connected component involved in the SWV detection is presented in red in the third row. The duration of the violation is 30 frames or $30/25 = 1.2$ seconds (given that the video fps rate is 25). Thus, the duration threshold $T_d = FPS/2 = 25/2 = 13$ frames is exceeded and the violation is considered annoying.

IV. Conclusions and Future Work

This paper presented automatic detection frameworks for three stereo 3D artifacts: sharpness mismatch, synchronization mismatch and stereoscopic window violation. Sharpness mismatch is estimated by measuring the width deviations of edge pairs in depth planes. Synchronization mismatch is detected based on the motion inconsistencies of feature point pairs between the LR views within a short time frame. Stereoscopic window violation is detected, using connected component analysis, when objects hit the vertical frame boundaries while being in front of the virtual screen. Test sequences for experimental evaluation were created in professional studio environments. The experimental results show that our algorithms have a considerable robustness of analyzing appropriate 3D defects.

Future work will focus on improving the performance of the proposed approaches and developing new approaches for further 3D defects. Furthermore, the generation of additional 3D test sequences will be also targeted.

Acknowledgment

The authors would like to thank Camelot Broadcast and Max Hemmo for providing the movie "The Magician", used in the experimental evaluation of the stereoscopic window violation detection algorithm. The research leading to these results has received funding from the European Union Seventh Framework Program (FP7/2007-2013) under grant agreement number 287674 (3DTVS). This publication reflects only the author's views. The European Union is not liable for any use that may be made of the information contained herein.

References

[1] M. Barkowsky, K. Brunnström, T. Ebrahimi, L. Karam, P. Lebreton, P. L. Callet, A. Perkis, A. Raake, M. Subedar, K. Wang, L. Xing, and J. You, "Subjective and objective visual quality assessment in the context of stereoscopic 3d-tv," in *3D-TV System with Depth-Image-Based Rendering*, 2013, pp. 413–437.

[2] "http://www.compression.ru/video/vqmt3d/."

[3] "http://www.technicolor.com/en/solutions-services/technology/technology-licensing/3d-excellence/certifi3d."

[4] N. Atzpadin, P. Kauff, and O. Schreer, "Stereo analysis by hybrid recursive matching for real-time immersive video conferencing," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 14, no. 3, pp. 321–334, 2004.

[5] M. Yaqub, "Visual fields interpretation in glaucoma: a focus on static automated perimetry," *Community Eye Health Journal*, vol. 25, no. 79, pp. 1–8, 2013.

[6] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Applications to jpeg2000," *Signal Process.: Image Communication*, vol. 19, no. 2, pp. 163–173, 2004.

[7] R. Ferzli and L. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (jnb)," *IEEE Trans. on Image Processing*, vol. 18, no. 4, pp. 717–728, 2009.

[8] P.910, *Subjective video quality assessment methods for multimedia applications*. ITU, 2000.

[9] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. ICCV 1999*, Kerkyra, September 1999, pp. 1150–1157.

[10] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[11] W. Feller, *Introduction to Probability Theory and its Applications*. Wiley, 1950.

[12] B. Mendiburu, Y. Pupulin, and S. Schklair, *3D TV and 3D cinema. Tools and processes for Creative Stereoscopy*. Focal Press, 2012.

[13] B. Mendiburu, *3D movie making. Stereoscopic digital cinema from script to screen*. Focal Press, 2009.

[14] T. Sakamoto, "Model for spherical aberration in a single radial gradient-rod lens," *Applied Optics*, vol. 23, no. 11, pp. 1707–1710, 1984.

[15] N. Narvekar and L. Karam, "A no-reference image blur metric based on the cumulative probability of blur detection (cpbd)," *IEEE Trans. on Image Processing*, vol. 20, no. 9, pp. 2678–2683, 2011.

[16] T. Jussi, N. Mikko, and O. Pirkko, "Spatial and temporal information as camera parameters for super-resolution video," in *Proc. ISM 2012*, Irvine, December 2012, pp. 302–305.

[17] D. C. Hoaglin, F. Mosteller, and J. W. Tukey, *Understanding robust and exploratory data analysis*. Wiley, 1983.