Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Stereoscopic medical data video quality issues

Foteini Patrona
Ioannis Mademlis
Fotios Kalaganis
Ioannis Pitas
Kleoniki Lyroudia

Foteini Patrona, Ioannis Mademlis, Fotios Kalaganis, Ioannis Pitas, Kleoniki Lyroudia, "Stereoscopic medical data video quality issues," *J. Med. Imag.* **3**(2), 025501 (2016),
doi: 10.1117/1.JMI.3.2.025501.

Stereoscopic medical data video quality issues

Foteini Patrona,^{a,*} Ioannis Mademlis,^a Fotios Kalaganis,^a Ioannis Pitas,^{a,b} and Kleoniki Lyroutdia^c

^aAristotle University of Thessaloniki, Faculty of Sciences, Department of Informatics, University Campus, Thessaloniki 54124, Greece

^bUniversity of Bristol, Faculty of Engineering, Department of Electrical and Electronic Engineering, Senate House, Tyndall Avenue, Bristol BS8 1TH, United Kingdom

^cAristotle University of Thessaloniki, Faculty of Health Sciences, School of Dentistry, University Campus, Thessaloniki 54124, Greece

Abstract. Stereoscopic medical videos are recorded, e.g., in stereo endoscopy or during video recording medical/dental operations. This paper examines quality issues in the recorded stereoscopic medical videos, as insufficient quality may induce visual fatigue to doctors. No attention has been paid to stereo quality and ensuing fatigue issues in the scientific literature so far. Two of the most commonly encountered quality issues in stereoscopic data, namely stereoscopic window violations and bent windows, were searched for in stereo endoscopic medical videos. Furthermore, an additional stereo quality issue encountered in dental operation videos, namely excessive disparity, was detected and fixed. The conducted experiments prove the existence of such quality issues in stereoscopic medical data and highlight the need for their detection and correction. © 2016 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/J.MI.3.2.025501]

Keywords: medical videos; stereoscopic video quality assessment; stereoscopic window violation; bent window effect; excessive disparity.

Paper 15203LR received Dec. 14, 2015; accepted for publication May 3, 2016; published online May 27, 2016.

1 Introduction

Recent technological advances in stereo cameras, microcameras, fiber optic devices, and computer vision algorithms have gradually led to a new era in the video recording of preoperative as well as intraoperative medical operations. Such recordings allow the estimation of tissue surface geometry and tissue deformation recovery, and help in accurate decision-making in minimally invasive surgery (MIS).¹

Maximally stable extremal regions (MSER), detected using the lowest image intensity as the starting point along with image gradient features, are employed as an aid to salient landmark selection on stereo laparoscopic data.² Feature matching is subsequently performed and three-dimensional (3-D) point reconstruction as well as temporal tracking are accomplished based on epipolar constraints. MSER and image gradient-related features are also leveraged in a GPU implemented system,³ thus enabling the accurate combination of preoperative and intraoperative endoscopic images for better tissue surface geometry estimation. Edge-based snake segmentation of the produced phantom/available surgical data is performed, followed by manual segmentation, surface mesh simplification techniques, and point tracking through linear node interpolation. Stereoscopic preoperative and multiview intraoperative endoscopic data can be combined for a multiorgan segmentation system.⁴ The stereoscopic medical video data, along with the incorporation of the camera motion prior to this system, enable more accurate MIS by stabilizing the 3-D organ segmentation, 3-D tissue pose tracking, and tissue-deformation estimation.

Dynamic expansion of the original field-of-view, using optical flow information, has been applied to stereo endoscopic video data, captured from the peritoneal or thoracic cavity during natural orifice transluminal endoscopic surgery,⁵ in a way similar to the one used in image mosaicing.⁶ In this case,

projection on nonplanar surfaces was also considered. Parallax correction, combined with holography fitting, enables optical flow estimation beyond the image boundaries. Real-time laparoscopic 3-D tissue surface reconstruction has also been proposed for robotically assisted MIS and improved surgeon decision-making.⁷ The proposed 3-D surface reconstruction is based on propagating a sparse set of stereo point correspondences, used as seeds, to correspondences of a semidense set of neighboring points of similar color, subject to their correlation scores. This method has been implemented in parallel.⁷

As a consequence of their widely expanding use, quite a low effort has been paid to the quality assessment of entertainment-oriented stereoscopic videos as well as to the devising of methods capable of fixing the most commonly encountered quality issues. Considering the way they approach stereoscopic video quality assessment, the devised methods vary to those originating from two-dimensional (2-D)-video quality assessment and are modified in such ways that they can be applied to stereoscopic videos, to those specifically devised for stereoscopic videos and finally to methods based on the human visual system and perception.⁸ Among the main causes of stereoscopic video quality issues like stereoscopic window violation (SWV), focus and color mismatch, geometry distortion, time desynchronization, crosstalk and excessive or insufficient disparity, lie camera setup and synchronization,⁹ compression and rendering.¹⁰ Finally, it is worth mentioning that defects like the bent window (BW) effect, SWVs, and excessive disparity¹¹ can cause visual fatigue and discomfort,^{10,12–14} thus hindering the viewing process.

Video quality assessment methods for target recognition tasks have also been thoroughly investigated^{15,16} in recent years, and even standards concerning the procedure recommended for performing subjective evaluation experiments to this end have been proposed by the ITU-T.¹⁷ Two of the most representative examples of task-based video usage, not

*Address all correspondence to: Foteini Patrona, E-mail: fotinip@aiia.csd.auth.gr

related to entertainment, are surveillance and medical videos for accurate target (e.g., event, object, and person) detection and credible diagnosis prior/during surgery (e.g., bronchoscopy), respectively. However, both the conducted research and the proposed standards are oriented to 2-D videos, mainly focusing on the perceived image quality, and on issues mostly related to it, like compression and transmission,^{16,18,19} not investigating the degree of recognition effectiveness in these videos, as it is the case with quality assessment for recognition tasks.²⁰

The aforementioned issues motivated us to get involved with stereoscopic medical data and investigate the existence of quality defects in them more thoroughly. Some of the defects already mentioned are subsequently presented in detail and examples of their automatic detection¹¹ are provided, using publicly available stereoscopic medical videos, as well as dental videos recorded at the School of Dentistry, Aristotle University of Thessaloniki. Additionally, in the case of excessive disparity, the detected quality issues are fixed.

2 Stereoscopic Video Quality Issues

In natural human vision, disparity refers to differences in the two object views, as perceived by the two eyes. In a stereoscopic image pair, composed of a left and a right video channel, a dense disparity map that assigns a depth-related disparity value to each image pixel can be estimated from detected pixel correspondences between the two video channels,²¹ as shown in Fig. 1. Two different disparity maps can be extracted from a single stereo image pair, associated with the left/right image channel, respectively. Figure 2 displays the simplest parallel stereo rig geometry. The left and right camera centers of projection are marked as O_l and O_r , respectively, their projection planes as I_l , I_r , while P and P' denote two 3-D world points. When captured by the stereo rig, these points project on the image point pairs p_l , p_r and p'_l , p'_r , respectively. For each left/right-channel image point $p = [x, y]^T$, in pixel coordinates, the corresponding horizontal disparity values are $d_{x,y}^l \leq 0$ and $d_{x,y}^r = -d_{x,y}^l$, while vertical disparities are zero. The closer an imaged object lies to the cameras during image acquisition, the larger its disparity is in absolute value. In contrast, objects considered to be lying at infinity, i.e., positioned very far from the cameras, are projected on pixels with near-zero disparity.

Fig. 2 Stereo image disparities.

Fig. 3 Stereo image perception.

During video display, such objects appear in front of the display screen or, in the case of objects at infinity, on the display screen itself, as shown in Fig. 3.

However, the stereo-pairs are typically processed to allow a perceived placement of imaged objects during video display, both in front of and behind the screen plane. Therefore, the disparity maps estimated from postprocessed 3-D visual content typically contain both positive and negative disparity values. Pixels associated with negative left disparity are to be displayed in front of the screen, pixels with positive left disparity are to be

Fig. 1 (a) Left and (b) right SWVs.

displayed behind the screen, and pixels with zero disparity will be displayed on the screen plane itself.

The 3-D world observation screen, also called stereoscopic window (SW), sets limitations to the 3-D perception. Several problems may occur, stemming from the existence of screen edges, in combination with the projected object disparity values and the point to which each image is projected in space (i.e., the point where human eyes have to converge). Unpleasant effects may thus emerge, affecting viewing comfort and causing visual fatigue. Three representative examples, which can also be encountered in medical 3-D data, are presented in short in the following sections.¹⁴

2.1 Stereoscopic Window Violation

As previously mentioned, the 3-D viewing space is defined by the screen edges and the relative positions of the viewer eyes. Due to the disparities between the positions of displayed corresponding left and right image pixels, though, pixels located near the screen side borders in one of the two images may not correspond to pixels in the other one, as shown in Fig. 1. Indicative image pixels having no correspondence in the other image channel are depicted by $\times$. This effect causes retinal rivalry and occurs only for points with disparity not equal to zero. Objects appearing near the image edges are, therefore, cut off by the screen borders and interpreted by the brain as lying behind them, thus being occluded by them (i.e., by the screen). However, this depth cue is at odds with the one stemming from disparity, which suggests that they are positioned in front of the screen plane. This stereo perception discrepancy is widely known as an SWV and, apart from being annoying, may also lead to eye strain, visual discomfort and/or loss of depth perception, since the image pair cannot be fused to a single image.¹¹ This effect can only be ignored when fast moving objects enter or leave the scene, as their position in space has not yet been decided by the brain at the time the violation occurs.¹⁴

SWVs are detected using the disparity map of each video frame pair.¹¹ The method begins by identifying the disparity map regions displayed significantly in front of the screen, by detecting image segments and comparing their mean disparity values with a predefined threshold. A rectangular region of interest (ROI) that is represented by its upper left and lower right corner coordinates is then employed in order to enclose the detected objects. The violation presence is decided based on the ROI position with respect to the image borders. Thus, if such an ROI lies on the left border of the right image, a left SWV occurs. Similarly, right SWVs occur when an ROI lies on the right border of the left image. Two representative examples of left and right SWVs are presented in Figs. 1(a) and 1(b), respectively.

2.2 Bent Window Effect

The BW effect is encountered when objects projected in front of the screen extend vertically across the entire video frame and are cut off by the top and bottom screen borders, as shown in Fig. 4. However, this would normally happen only if they lay behind the screen. The cues perceived by the brain are, thus, rather contradictory, causing visual discomfort and a feeling that the middle section of the screen is bent toward the viewer,¹¹ so that both disparity and upper/lower screen border occlusion can be reconciled by the brain.¹⁴ Even though each border violation causes different visual discomfort, they all do affect depth

Fig. 4 BW effect.

perception to a great extent. While the violations occurring at the left and right edges are the most distracting, as in the case of SWV, those occurring at the lower and mostly at the upper edge tend to affect 3-D perception as well.

BW effects are detected in a way similar to the one used for SWV detection.¹¹ Objects having significant disparity are detected, their disparity values are compared to an appropriate threshold and their respective ROIs are defined. ROIs touching both the top and the bottom image edges are thus considered to enclose objects causing the BW effect.

2.3 Excessive Disparity

One last quality defect observed in stereoscopic videos is the so-called excessive disparity. As can be easily deduced by its name, it is closely related to large horizontal disparity values. Our eyes are positioned in parallel to one another at a distance of ~ 2.5 cm, as shown in Fig. 3. If the disparity of the displayed objects is excessive, as in the case of point **P** in Fig. 3, the object is perceived to be outside the so-called comfort zone,¹⁴ thus causing eye fatigue due to excessive eye convergence. Therefore, moderate disparity values are in general desirable. Otherwise, 3-D perception deteriorates, viewer discomfort emerges and, in extreme cases of excessive disparity, double-vision arises, and 3-D perception is entirely lost. The most common reasons for excessive disparity are inadequate camera calibration and extreme zooming in.

Excessive disparity can be easily detected, simply by displaying the 3-D content onto the screen it is aimed for, and can be automatically or semiautomatically fixed with the aid even of freely available existing tools, like stereo photo maker, which allows both manual image alignment and automatic alignment, based on the scale-invariant feature transform algorithm in order to extract corresponding points of interest from the images to be aligned. A sample dental stereo image pair with excessive disparity before and after its automatic fixing with stereo photo maker is presented in Figs. 5(a) and 5(b), respectively.¹⁴

3 Experiments

3.1 Objective Stereoscopic Video Quality Assessment

The aforementioned quality issue detectors were applied to the publicly available Hamlyn Centre Laparoscopic/Endoscopic video data sets, consisting of *in vivo* patient data sets and validation data sets.¹ Sample stereo video frame pairs along with their respective disparity maps are presented in Figs. 6 and 7.

Fig. 5 (a) Excessive disparity and (b) fixed excessive disparity.

Fig. 6 Tissue-tool interaction in abdomen.

Twenty five videos from the validation data set, with resolutions ranging from 320×240 to 720×288 , were evaluated and their evaluation results are presented in Table 1. In brief, both left and right SWVs were detected in the vast majority of the videos, 92.0% and 96.0%, respectively, while just 4.0% of the videos were proven not to exhibit quality issues. The BW effect was not detected so often, as only 52.0% of the examined videos were found having this effect. All three issues were encountered in 52.0% of the videos and just 40.0% of the videos were found to have only left and right stereoscopic violations.

3.2 Subjective Stereoscopic Video Quality Evaluation

After ascertaining that stereoscopic medical data do exhibit quality issues, subjective psychophysical experiments were conducted, aiming to the identification of the effects these issues have both on the video viewers and on the recognition tasks themselves.

The videos used for the subjective evaluation were two clips from the dental operation we recorded, as well as three clips from the Hamlyn Centre Laparoscopic/Endoscopic data set.¹ As far as the dental video clips is concerned, their original and disparity corrected 3-D versions were presented, while one laparoscopic video with no defects, one with SWVs, and a third one in which both SWVs and BW were detected and chosen. Stereoscopic as well as nonstereoscopic versions of all the videos were used for the subjective evaluation.

Thirty undergraduate, postgraduate, and PhD candidate students participated in the evaluation process, half of them studying at the Department of Informatics and half at the Department of Dentistry of Aristotle University of Thessaloniki. The sample was purposely chosen in this way, so that feedback from equal numbers of subjects of different expertise levels on 3-D video quality and medical imaging for recognition tasks could be acquired. Both male and female subjects participated in the subjective evaluation, and they all had normal or corrected to normal vision.

Fig. 7 Laparoscopic acquisition of liver deformation due to respiration.

During the experiment, the participants were shown pairs of the 2-D version of each of the five videos we chose followed by the corresponding 3-D version, while in the cases of the dental videos, a triplet consisting of the nonstereoscopic video, the original stereoscopic recording, and finally the stereoscopic video after the excessive disparity adjustment was presented. After watching all the versions of each video, the participants were asked to fill in a questionnaire, with questions mainly focusing on the adverse symptoms usually induced by 3-D viewing as well as on the subjective attributes that could be used to express their impressions concerning the videos.²² Finally, the subjects were asked whether they preferred the stereoscopic or the nonstereoscopic version of each video.

Some questions not related to the presented videos were also included in the questionnaire filled in by the participants in the subjective evaluation. These additional questions aimed at the induction of their emotional and cognitive state, i.e., their attitude toward the content of the videos they were going to

watch, toward new technologies, especially stereoscopy, as well as their familiarity with it. The reason for the inclusion of these questions is that such factors have been shown to affect the subjects' judgment.²²

Table 2 summarizes the physiological sensations²² included in our questionnaire along with the percentage of the users who reported experiencing each one of them while watching our dental operation video clips. In brief, adverse symptoms seem to be encountered during watching stereoscopic videos with excessive disparity to a greater extent. They subside when disparity values are appropriately adjusted, not getting eliminated, although, while the fact that symptoms are also mentioned for nonstereoscopic video watching could be regarded as incidental and maybe content-related.

As far as the attributes used in the comparison of the three conditions²² is concerned, it can be easily noticed from Table 3 that nonstereoscopy tends to be regarded as ordinary and boring by the majority of the participants, while stereoscopic data exhibiting quality defects induce unpleasant reactions/effects. On the contrary, when the quality issues are fixed, stereoscopy becomes interesting and allows a more detailed scene depiction,

Table 1 Quantitative results on stereoscopic medical video quality issue detection.

Quality issue	Appearance percentage (%)
Left SWV	92.0
Right SWV	96.0
BW	52.0
No quality issue	4.0
Only left SWV + right SWV	40.0
Left SWV + right SWV + BW	52.0

Table 2 Adverse physiological sensations.

Symptoms	2-D (%)	3-D (%)	Fixed 3-D (%)
General discomfort	6.7	56.7	23.3
Eye stain	3.3	80.0	30.0
Dizziness	0.0	53.3	16.7
Double images	0.0	100.0	6.7
Focusing difficulties	0.0	86.7	33.3

Table 3 Subjective attributes.

Attributes	2-D (%)	3-D (%)	Fixed 3-D (%)
Tiring	0.0	90.0	26.7
Ordinary	73.3	6.7	3.3
More details noticeable	0.0	0.0	63.3
Weird feeling	0.0	96.7	10.0
Depth impression	0.0	0.0	56.7
Presence	0.0	0.0	33.3
Interesting	0.0	3.3	73.3
Boring	86.7	23.3	16.7

making the viewers perceive depth very well and feel as if they are part of the scene.

4 Conclusions

In this paper, we investigated the existence of some of the most commonly encountered stereoscopic data quality issues in the more and more widely used stereoscopic medical data, recorded before or during medical operations. To this end, both objective and subjective quality evaluation methods were employed. The conducted experiments proved that such defects do appear in stereoscopic medical data and highlighted the need for their detection and fixing, as they impinge not only on the appropriateness of the data for the recognition tasks they are aimed at but also on the doctors' viewing experience, confusing their visual system and inducing adverse physiological sensations.

Acknowledgments

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 287674 (3DTV). This publication reflects only the authors' views. The European Union is not liable for any use that may be made of the information contained therein. Research Policies: During this research on human subjects, informed consent was obtained from all subjects, according to the institutional policies prescribed by the Research Committee, Aristotle University of Thessaloniki.

References

1. P. Mountney, D. Stoyanov, and G.-Z. Yang, "Three-dimensional tissue deformation recovery and tracking," *IEEE Signal Process Mag.* **27**(4), 14–24 (2010).
2. D. Stoyanov et al., "Soft-tissue motion tracking and structure estimation for robotic assisted MIS procedures," in *Proc. of the 8th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '05)*, pp. 139–146 (2005).
3. P. Pratt et al., "Dynamic guidance for robotic surgery using image-constrained biomechanical models," in *Proc. of the 13th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '10)—Part I*, Vol. **6361**, pp. 77–85 (2010).
4. M. Nosrati et al., "Efficient multi-organ segmentation in multi-view endoscopic videos using pre-operative priors," in *Proc. of the 17th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '14)—Part II*, Vol. **8674**, pp. 324–331 (2014).

5. M. Lerotic et al., "Dynamic view expansion for enhanced navigation in natural orifice transluminal endoscopic surgery," in *Proc. of the 11th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '08)—Part II*, Vol. **5242**, pp. 467–475 (2008).
6. I. Pitas, *Digital Video and Television*, 1st ed., CreateSpace/Amazon (2013).
7. D. Stoyanov et al., "Real-time stereo reconstruction in robotically assisted minimally invasive surgery," in *Proc. of the 13th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '10)*, Vol. **6361**, pp. 275–282 (2010).
8. A. Voronov et al., "Towards automatic stereo-video quality assessment and detection of color and sharpness mismatch," in *Int. Conf. on 3D Imaging*, pp. 1–6 (2012).
9. A. Voronov et al., "Methodology for stereoscopic motion-picture quality assessment," *Proc. SPIE* **8648**, 864810 (2013).
10. M. Lambouij and W. IJsselstein, "Visual discomfort and visual fatigue of stereoscopic displays: a review," *J. Imaging Sci. Technol.* **53**(3), 030201 (2009).
11. S. Delis et al., "Automatic detection of 3D quality defects in stereoscopic videos using binocular disparity," *IEEE Trans. Circuits Syst. Video Technol.* **PP**(99), 1–1 (2016).
12. M. Emoto, Y. Nojiri, and F. Okano, "Changes in fusional vergence limit and his hysteresis after viewing stereoscopic TV," *Displays* **25**, 67–76 (2004).
13. S. Sharples et al., "Virtual reality induced symptoms and effects (VRISE): comparison of head mounted display (HMD), desktop and projection systems," *Displays* **29**, 58–69 (2008).
14. B. Mendiburu, *3D Movie Making: Stereoscopic Digital Camera from Script to Screen*, Focal Press (2009).
15. D. Strohmeier, S. Jumisko-Pyykkö, and K. Kunze, "Open profiling of quality: a mixed method approach to understanding multimodal quality perception," *Adv. Multimedia* **35**, 36 (2010).
16. M. Leszczuk, "Assessing task-based video quality—a journey from subjective psycho-physical experiments to objective quality models," *Multimedia Commun. Serv. Secur.* **149**, 91–99 (2011).
17. ITU-T, "ITU-T recommendation P.910: subjective video quality assessment methods for multimedia applications" (1999).
18. ITU-T, "ITU-T recommendation P.912: subjective video quality assessment methods for recognition tasks" (2008).
19. M. Duplaga et al., "Evaluation of quality retaining diagnostic credibility for surgery video recordings," *Visual Inf. Syst.: Web-Based Visual Inf. Search Manage.* **5188**, 227–230 (2008).
20. M. Leszczuk and J. Dumke, "Quality assessment for recognition tasks (QART)," in *4th Int. Conf. on Emerging Network Intelligence* (2012).
21. D. Scharstein and R. Szeleiski, "A taxonomy and evaluation of dense two frame stereo correspondence algorithm," *IEEE Int. J. Comput. Vision* **47**(1–3), 7–42 (2002).
22. J. Häkkinen et al., "Measuring stereoscopic image quality experience with interpretation based quality methodology," in *Proc. SPIE* **6808**, 68081B (2008).

Foteini Patrona received her BSc degree in applied informatics from the University of Macedonia, Thessaloniki, Greece, in 2012 and her MSc degree in digital media from Aristotle University of Thessaloniki, Greece, in 2014. She was a research assistant at the Artificial Intelligence and Information Analysis Laboratory of the Department of Informatics at Aristotle University of Thessaloniki from 2013 to 2015 and has participated in two research projects financed by national and European funds. Currently, she is a research associate at the Centre of Research & Technology–Hellas/Information Technologies Institute.

Ioannis Mademlis received her BSc degree in 2007 and her MSc degree in computer science from the University of Ioannina, Greece, in 2010. He also received an MSc degree in the area of intelligent systems from the School of Electrical and Computer Engineering at the Aristotle University of Thessaloniki in 2014. Currently, he is a PhD candidate and a teaching assistant at the Department of Informatics of the same institution and is also employed as a research assistant at the Artificial Intelligence and Information Analysis Laboratory of the aforementioned department. He has coauthored two journal papers and three papers in international conferences.

Fotios Kalaganis received his BSc degree in informatics from Aristotle University of Thessaloniki in 2013. Since 2014, he has been an MSc student in the aforementioned university, specializing in digital media. He was awarded for excellence in academic studies during the second semester of the master's program. His research interests include digital signal processing, computational intelligence, neuroinformatics, and brain-computer interaction.

Ioannis Pitas: Biography is not available.

Kleoniki Lyroudia received her diploma of dentistry in 1980 and her PhD in dentistry in 1982, both from the Aristotle University of Thessaloniki, Greece. Since 2009, she has been a professor in the Department of Endodontology, Dental School, of the same university. She has published 2 monographs, over 70 scientific/research articles in national and international journals, most of them related to endodontics, and has over 100 presentations in national, European, and international meetings and congresses.