Quality of Experience for 3D video streaming

Chaminda T. E. R. Hewage and Maria G. Martini
Wireless Multimedia and Networking Research Group
SEC Faculty, Kingston University London
Penrhyn Road, KT1 2EE, UK
{c.hewage, m.martini}@kingston.ac.uk

***Abstract*—New 3D video applications and services are emerging to fulfill increasing user demand. This effort is well supported by the increasing 3D video content including user generated content (e.g., through 3D capture/display enabled mobile phones), technological advancements (e.g., HD 3D video capture and processing methods), affordable 3D displays, and standardization activities. However, not much attention has been given to how these technologies, along the end-to-end chain, from content capture to display, affect user perception and whether the overall experience of 3D video users is satisfactory or not. 3D video streaming also introduces artifacts on the reconstructed 3D video at the receiver end, leading to inferior quality and user experience. In this article we present and discuss in detail how artifacts introduced during 3D video streaming affect the end user perception and how we could use real-time quality evaluation methodologies to overcome these effects. The observations presented can underpin the design of future QoE-aware 3D video streaming systems.**

I. Introduction

Interactive 3D video streaming will enable seamless, more involving and adaptable delivery of 3D content to end users. However, 3D video streaming over band-limited and unreliable communication channels can introduce artifacts on the transmitted 3D content. The effect could be much more significant compared to conventional 2D video streaming. For instance, the nature of 3D video source format (e.g. colour plus depth images vs. left and right views) and the way our Human Visual System (HVS) perceives channel introduced artifacts in 3D video is different from 2D video; as an example, colour plus depth map 3D video presentation may have to utilize impaired depth map information at the receiver-side to render novel views.

In the remainder of this paper, after an introduction on interactive video streaming and 3D artifacts, we discuss how we can quantify the overall user experience in 3D viewing and how our HVS reacts to different artifacts. Moreover, we elaborate on how we could exploit the measurement of the quality at the receiver side to overcome these effects through QoE-driven system adaptation.

A. Interactive 3D video streaming

Video streaming over the Internet has become one of the most popular applications and Internet 3D video streaming is expected to become more popular in the future, also thanks to the recently standardized wireless systems, including WIMAX, 3GPP LTE / LTE advanced, the latest 802.11 standards, and advanced short range wireless communication systems, enabling the transmission of high bandwidth multimedia data. For such applications the target of the system design should be the maximization of the final quality perceived by the user, or Quality of Experience (QoE), rather than only of the performance of the network in terms of "classical" quality of service (QoS) parameters such as throughput and delay. 3D video services, and in particular those delivered through wireless and mobile channels, face a number of challenges due to the need to handle a large amount of data and to the possible limitations due to the characteristics of the transmission channel and of the device. This can result in perceivable impairments originated in the different steps of the communication system, from content production to display techniques (see Fig. 1), and influence the user's perception of quality. For instance channel congestion and errors at the physical layer may result in packet losses and delay, whereas compression techniques introduce compression artifacts. Such impairments could be perceived by the end user and result to a different extent in the degradation of the quality of the rendered 3D video. Some of these artifacts are common to 2D video applications as well. In addition, artifacts which are specific to 3D can be introduced during end-to-end chain of 3D video delivery such as *cross talk*, *keystone distortion*, etc. [1].

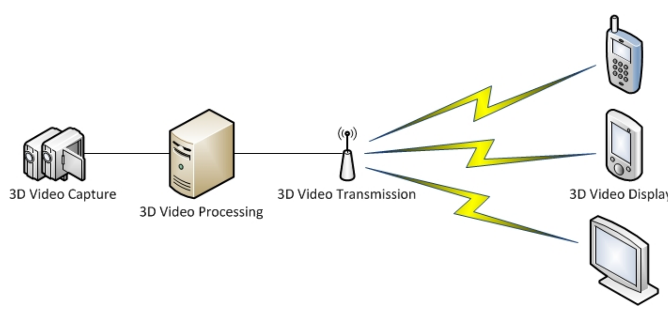

Fig. 1. End-to-end 3D video processing chain.

B. 3D video QoE

The overall enjoyment or annoyance of 3D video streaming applications or services is influenced by several factors such as human factors (e.g., demographic and socio economic background), system factors (e.g., content and network related influences) and contextual factors (e.g., duration, time of the day and frequency of use). The overall experience can be analyzed and measured by QoE related parameters which quantify the user's overall satisfaction about a service [2] [3]. Quality of Service (QoS) related measurements only measure performance aspects of a physical system, with main focus on telecommunications services. Measuring QoS parameters is straightforward since objective, explicit technological methods can be used, whereas measuring and understanding QoE requires a multi-disciplinary and multi-technological approach. The added dimension of depth in 3D viewing influences several perceptual attributes such as overall image quality, depth perception, naturalness, presence, visual comfort, etc. For instance, an increased binocular disparity enhances the depth perception of viewers, although in extreme cases this can lead to eye fatigue as well. Therefore, the overall enjoyment of the 3D application could be hindered by the eye strain experienced by the end user. The influence of these attributes on the overall experience of 3D video streaming users is yet to be investigated. Figure 2 outlines a comparison between QoE and QoS.

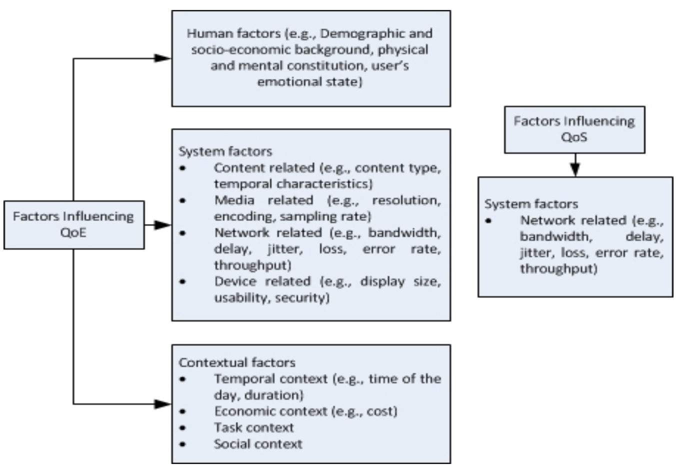

Fig. 2. Quality of Experience vs. Quality of Service.

C. 3D video artifacts

The different processing steps along the end-to-end 3D video chain introduce image artifacts which may affect 3D perception and the overall experience of viewers [1]. Even though much attention has been paid into analyzing and mitigating the effects of 3D image/video capture, processing, rendering and display techniques, the effects of artifacts introduced by the transmission system have not received much attention compared to the 2D image/video counterpart. Some of these artifacts influence the overall image quality, for instance blurriness, luminance and contrast levels, similar as in 2D image/video. The effect of transmission over band-limited and unreliable communication channels (such as wireless channels) can be much worse for 3D video than for 2D video, due to the presence in the first case of two channels (i.e., stereoscopic 3D video) that can be impaired in a different way; as a consequence the 3D reconstruction in the human visual system may be affected. Some networks introduce factors directly related to temporal domain de-synchronization issues. For instance delay in one view could lead to temporal de-synchronization and this can lead to reduced comfort in 3D viewing.

The methods employed to mitigate these artifacts (e.g., error concealment) need to be carefully designed to suit 3D video applications. The simple application of 2D image/video methods would not work effectively in this case, as discussed in [4] for different error concealment algorithms for 3D video transmission errors. In [4] it is observed that is some cases switching back to the 2D video mode is preferred to applying 2D error concealment methods separately for left and right views to recover missing image information during transmission. There could be added implications introduced by these artifacts into our HVS. Therefore artifacts caused as a result of 3D video streaming can be clearly appreciated only by understanding how our HVS perceives different 3D video artifacts.

1) Binocular vision: The HVS is capable of aligning and fusing two slightly different views fed into the left and right eyes and hence of perceiving the depth of the scene. Both binocular and monocular cues assist our HVS to perceive different depth planes of image objects [5]. Binocular disparity is the major cue used by the HVS to identify the relative depth of objects. Other monocular depth cues include perspective, occlusions, motion parallax, etc. ([5]). During 3D video streaming, one view or both views could be badly affected by channel impairments (e.g., bit errors and packet losses caused by adverse channel conditions, delay, jitter). For instance, frame freezing mechanisms employed to tackle missing frames caused by transmission errors or delay could lead to temporal de-synchronization where one eye sees delayed content compared to the other eye. There are two implications associated to the case where one view is affected by transmission impairments:

- Binocular suppression
- Binocular rivalry

Our HVS is still capable to align and fuse stereoscopic content if one view is affected by artifacts due to compression, transmission, and rendering. Binocular suppression theory suggests that in these situations the overall perception is usually driven by the quality of the best view (i.e., left or right view), at least if the quality of the worst view is above a threshold value. However this capability is limited and studies show that additional cognitive load is necessary to fuse these views [6]. Increased cognitive load leads to visual fatigue and eye strain and prevents users from watching 3D content for a long time. This directly affects user perception and QoE. If one of the views is extremely altered by the transmission system, the HVS will not be able to fuse the affected views, and this causes binocular rivalry. This has detrimental effects on the final QoE perceived by the end user. Recent studies on 3D video transmission [4] have found that binocular rivalry is causing the overall perception to be affected and this effect prevails over the effect of binocular suppression. To avoid the detrimental effect of binocular rivalry, the transmission system could be designed appropriately taking this issue into account. For instance, the transmission system parameters can be updated "on the fly" to obtain 3D views with minimum distortions, according to the feedback on the measure of 3D video quality at the receiver-side. In case of low quality due to different errors in the two views, if the received quality of one of the views is significantly low, the transmission system could be informed to allocate more resources to the worse view or to increase the error protection level for that 3D video channel to mitigate the quality loss in the future. This increases the opportunity to fuse the 3D video content more effectively and improve the final QoE of users. The next section of this article discusses in detail how we can measure 3D video quality.

II. Measuring 3D video quality

Immersive video quality evaluation is a hot topic among researchers and developers at present, due to its complex nature and to the unavailability of an accurate objective quality metric for 3D video. 3D perception can be associated with several perceptual attributes such as “overall image quality”, “depth perception”, “naturalness”, “presence”, “comfort”, etc. Currently, 3D quality evaluation studies focus only on one specific aspect such as overall quality, depth perception or visual comfort. A detailed analysis is necessary to study how these 3D percepts influence the overall perceived 3D image/video quality in general (i.e., 3D QoE). For instance, these attributes can be in conflict (e.g., increased disparity can cause eye fatigue), resulting in affected user experience. Mostly, appreciation oriented psychophysical experiments are conducted to measure and quantify 3D perceptual attributes. At present, electro-physiological devices are also being used to record user’s overall excitement about a 3D service.

A few standards define subjective quality evaluation procedures for 3D video (e.g., ITU-R BT1438, ITU-R BT500-12 and ITU-T P.910). However, these procedures are not competent enough to measure 3D QoE parameters and show several limitations; for instance these are not able to measure the combined effect of different perceptual attributes. Current standardization activities on 3D quality evaluation are discussed in Subsection IV.A. Subjective quality evaluation studies under different system parameter changes have been reported in a number of studies [7]. However, these studies are limited to certain types of image artifacts (e.g., compression artifacts) and have limited usage in practical applications. Furthermore, subjective quality evaluation requires time, effort, controlled test environments, money, human observers, etc. and cannot be deployed in a live environment where quality is measured “on the fly”.

Objective quality evaluation methods for 3D video are also emerging to provide accurate results in comparison to the quality ratings achieved with subjective tests [8]. The performance of these metrics is most of the time an approximation of the results of subjective quality assessments. 3D objective quality metrics are designed to account for both disparity and texture related artifacts based on extracted image features. The final score should therefore reflect the quality degradation in terms of both 2D image and depth perception. It is important to have a reliable ground truth 3D dataset to evaluate and verify the performance of emerging 3D objective metrics. These metrics need to be verified and validated using sequences with different characteristics and under different environmental settings. Studies have also found out that there is a high correlation between subjective ratings and individual objective quality ratings of 3D video components (e.g., average PSNR and SSIM of left and right video or colour and depth video) [9]. For instance, depth perception is highly correlated to the average PSNR of the rendered left and right image sequences [9]. This could be due to the loss of correspondence between left and right objects and reduction of monocular depth cues as a result of compression and transmission errors. This means that we could use individual objective quality measures of different 3D video components to predict the true user perception in place of subjective quality evaluation, through a suitable approximation derived based on correlation analysis. However, with some 3D source representations such as the colour and depth map 3D image format, it may be difficult to derive a direct relationship between objective measures and subjective quality ratings. For instance, the objective quality of the depth map may have a very weak correlation on its own with the overall subjective quality, because the depth map is used for projecting the corresponding colour image into 3D coordinates and it is not directly viewed by the end users. Individual quality ratings of left and right views may not always account for depth reproduction of the scene. Therefore, the next phase of 3D objective quality metrics includes a methodology to quantify the effect of binocular disparity of 3D scenes in addition to a conventional image/video quality assessment methodology. For instance in [8], in addition to image quality artifacts, disparity distortion measures were also incorporated to evaluate the overall 3D video quality. The article showed improved performance over the method which does not account for the correspondence information of stereoscopic views. The latest 3D image/video quality metrics evaluate depth reproduction in addition to usual image artifacts (such as blockiness) using specific image features (e.g., edge, disparity and structural information of stereoscopic images) which are important for the HVS in both 2D and 3D viewing. For instance the method proposed in [10] shows high correlation values with subjective quality results (Mean Opinion Score, MOS): the correlation coefficient with subjective quality ratings is as high as 0.95; this outperforms the method based on 2D image quality + disparity [8] and other conventional 2D quality metrics separately applied to left and right views (see Table I). The reported performance figures in Table I are obtained using the same 3D dataset. These observations confirm that accurate 3D image quality metrics should be designed to also consider binocular disparity distortions. All the methods described above are Full-Reference (FR) methods and need the original 3D image sequence to measure the quality by comparison, hence they are not suitable for the evaluation of the quality “on the fly” in real-time transmission applications such as interactive 3D video streaming. In this case the solution is to use Reduced-Reference (RR) or No-Reference (NR) metrics which do not require the original image for quality assessment, but either no information (NR) or just some side-information about it (RR) requiring few bits for its transmission. Most of the NR metrics are designed specifically for a known set of artifacts (e.g., JPEG compression) and cannot be deployed in a more general scenario. In case of RR metrics, side-information is generated from features extracted from the original 3D image sequence and sent to the receiver-side to measure 3D video quality. Since the reference side-information has to be transmitted over the channel, either in-band or on a dedicated connection, the overhead should be kept at a minimum level. The next section describes how we could measure 3D video quality "on the fly" using RR and NR methods and provides a brief description of the existing methods.

TABLE I
CORRELATION BETWEEN OBJECTIVE 3D IMAGE/VIDEO QUALITY MEASURE AND SUBJECTIVE QUALITY

Method	CC	SSE	RMSE
SSIM (Structural SIMilarity)	0.837	0.965	0.159
VQM (Video Quality Metric)	0.932	0.423	0.106
Proposed in [8]: 2D image quality + Disparity	0.901	0.608	0.126
Proposed in [10]	0.947	0.341	0.095

III. REAL-TIME 3D VIDEO QUALITY EVALUATION STRATEGIES

The measured image quality at the receiver-side can be used as feedback information to update system parameters "on the fly" in a "QoE-aware" system design approach [11] [2]. However, measuring 3D video quality in real time is a challenge mainly due to the complex nature of 3D video quality and also the fact that the amount of side-information to be sent to measure the quality with RR methods is larger compared to 2D image/video applications. The emerging RR and NR quality evaluation methods are based on image features associated to the characteristics of the HVS. Some of these features are related to image perception (e.g., luminance, contrast) and some are related to depth perception (e.g., disparity, structural correlations). An appropriate selection of these features is crucial to design an effective 3D image/video quality assessment method. The selected features should be able to quantify image and depth perception related artifacts with a minimum overhead. If the overhead is significant, the feasibility of deploying the designed RR method is reduced. Figure 3 shows how the extracted edge information is employed to measure 3D video quality in the RR method proposed in [12]. In this method, luminance and contrast details of the original and distorted images are utilized to count for conventional image artifacts, whereas edge information based structural correlation is employed to measure the structural/disparity degradation of the 3D scene, which is directly affecting rendering using colour plus depth map based 3D video. In order to reduce the overhead for side-information (i.e., extracted features of the reference image) lossless compression mechanisms can be deployed for its compression. An extra effort should be also made to send the side-information without corruption using a dedicated channel or highly protected forward channel. Visual attention models could also be utilized to find 3D image/video features which attract significant attention during 3D viewing. However, a direct relationship between visual attention and image perception for 3D images and video is yet to be found. NR methods are the most suitable for real-time 3D video applications since these do not consume any bandwidth for the transmission of side information. However, their performance and application domain is limited since they rely solely on the received 3D image/video sequence and other contextual information (e.g., Hybrid-NR methods: packet loss rate, bit-error rate). It may be impossible to count for all the artifacts imposed along the end to end 3D video chain without referring to the original image sequence. This is why most of the proposed NR metrics are limited to a specific set of artifacts [13].

Table II reports a few existing NR and RR quality metrics for 3D image/video. This table explains which image features are used to measure the overall perception and how much the different metrics are correlated with subjective quality scores (i.e., MOS) and with existing Full-Reference methods. It can be observed that most of these methods show a high degree of correlation with subjective MOS and Full-reference methods. However, these metrics are focused on one or two specific 3D perceptual attributes. The combined effect of these perceptual attributes which is directly related to user 3D QoE has not been addressed to date. The methods in [14] and [13] are evaluated using the same image database whereas others are evaluated using different data sets. Since some of these metrics, e.g., NR metrics ([13] and [15]) are designed for a particular types of image artifacts (e.g., JPEG compression), it is not always possible to compare the performance of a NR metric with another objective quality model in a common dataset. On the other hand, due to the overhead associated with RR metrics compared to zero overhead for NR metrics, the usage and advantages of these methods are significantly different. In addition, due to some practical reasons (intellectual property rights, different source 3D video formats, e.g., colour + depth vs. left and right images, unavailability of ground truth depth maps, etc.), it is not always feasible to compare the performance of two different 3D quality evaluation algorithms in a common dataset. The lack of reliable and comprehensive 3D image/video databases is another major challenge faced by researchers and developers, making difficult to effectively compare the performance of emerging objective and subjective quality evaluation methods with that of the existing methods.

IV. CHALLENGES AND FUTURE RESEARCH DIRECTIONS

The possibility to measure 3D image/video quality in real time, as requested by 3D video applications, is hindered by several issues. The major challenge is how we could measure the effect of all perceptual attributes (e.g., depth, presence, naturalness, etc.) associated with 3D viewing. The lack of availability of common 3D image/video databases is also detrimental for the advance in this discipline. The following paragraphs briefly discuss these challenges and possible solutions foreseen.

A. Measurement of different 3D perceptual attributes

Even though emerging 3D quality evaluation methods accurately predict a given quality attribute, the relationship among these perception attributes has not be thoroughly studied. The

TABLE II
NR AND RR METHODS FOR 3D IMAGE/VIDEO

Quality Metric	Method (NR or RR)	Artifacts	Features used to measure image artifacts (IA) and disparity (D)	CC	ROCC	OR	RMSE
Cyclop [14]	RR	JPEG symmetric and asymmetric coding artifacts	IA: Contrast sensitivity (spatial frequency and orientation); D: coherence of cyclopean images	0.981	0.950	0.050	-
Sazzad et al. [13]	NR	JPEG symmetric and asymmetric coding artifacts	IA: Blockiness and zero crossing of edge, flat and texture areas; D : average zero crossing of plane and non-plane areas	0.960	0.920	0.069	-
Solh et al. [15]	NR	Depth map and colored video compression, depth estimation (stereo matching), and depth from 2D to 3D conversion	IA D: Temporal outliers (TO), temporal inconsistencies (TI), and spatial outliers (SO) using ideal depth estimate for each pixel	0.916	0.1003	0.8	1.686
Hewage & Martini [12]	RR	H264 compression and random packet losses	IA: Luminance, structure and contrast D : edge based structural correlation	Colour: 0.9273 (vs. FR); Depth: 0.9795 (vs. FR)			Colour: 0.0110 (vs. FR); Depth: 0.0064 (vs. FR)

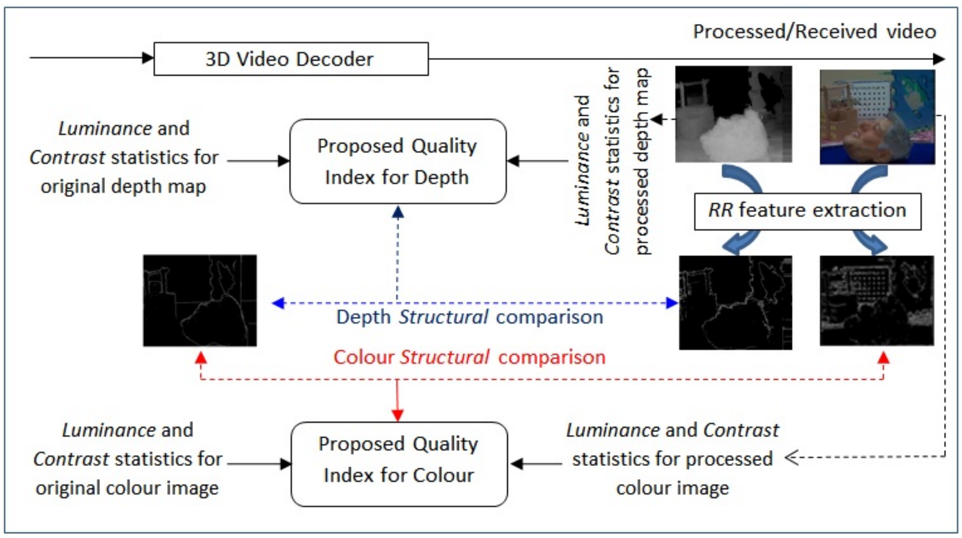

Fig. 3. Reduced reference edge based 3D video quality metric [12].

combined effect directly affects user experience and can be measured using emerging QoE indices. Therefore the current need is to understand how 3D audio/image processing and transmission artifacts affect the overall experience of the user, then identify audio, image and contextual features which can be used to quantify the overall effect on user experience. On the other hand, it is necessary to understand how the HVS perceives these 3D artifacts. For instance, there could be conflicts based on whether binocular suppression or binocular rivalry is taking place based on the artifacts in question. These aspects need extended attention in order to measure the overall experience of 3D viewing.

In order to enable a unified approach to 3D objective quality subjective quality evaluation studies, standardization of these procedures are necessary. Several standardization activities are being carried out by VQEG, ITU (Recommendations: ITU-T P- and J-series), European Broadcasting Union EBU (3D-TV Group) and other Standards Developing Organizations (SDOs) in relation to 3D video subjective and objective quality evaluations. Currently, the Video Quality Expert Group (VQEG) is working (3DTV project) on creating a ground truth 3D video dataset (GroTruQoE dataset) using the pair-comparison method. This ground truth database will then be used to evaluate other time-efficient 3D subjective quality evaluation methodologies and objective quality models. In addition, the project also addresses the objective quality assessment of 3D video, with the plan to evaluate 3D quality of experience in relation to the visual quality, depth quality and visual comfort dimensions. Most of these findings are reported to objective and subjective 3D video quality studies in ITU-T Study Groups (SG) 9 and 12. EBU is also working on 3D video production, formats and sequence properties for 3D-TV Broadcasting applications (e.g., EBU Recommendation R 135).

B. Lack of 3D image/video databases

There are several image/video quality databases for conventional 2D image/video artifacts, although only a few have been reported for 3D image/video artifacts. This prevents developers from using a common dataset to evaluate the performance of their metrics. Table III shows some of the reported 3D image/video databases in the literature. The amount of artifacts considered in these databases is limited. Most of them do not consider artifacts which could be introduced during transmission. Therefore it is a responsibility of the research community to produce comprehensive 3D video datasets covering a range of image and transmission artifacts and make available the developed 3D image/video dataset publicly.

C. Visual attention models to develop RR and NR quality metrics

The attention of users during 3D viewing can be influenced by several factors including spatial/temporal frequencies, depth cues, conflicting depth cues, etc. The studies on visual attention in 2D/3D images found out that the behaviour of viewers during 2D viewing and 3D viewing is not always identical (e.g., centre bias vs. depth bias). These observations are tightly linked with the way we perceive 3D video. Therefore, effective 3D video quality evaluation and 3D QoE enhancement schemes could be designed based on these observations. There are still unanswered questions such as whether quality assessment is analogous to attentional quality assessment and also how attention mechanisms could be properly integrated into design of QoE assessment methodologies. A thorough study has not been conducted to date in order to identify the relationship between 3D image/video attention models and 3D image/video quality evaluation.

Similar to the integrated model described above, attentive areas identified by visual attention studies can be utilized to extract image features which can be used to design No-Reference (NR) and Reduced-Reference (RR) quality metrics for real-time 3D video application (see Fig. **??**). Furthermore, since visual attention models can predict the highly attentive areas of an image or video, these can be integrated into source and channel coding at the sender side. Emerging 3D saliency models incorporate 2D image, depth and motion information which can be applied to 3D video sequences. Most of the reported 3D saliency models are extensions of 2D visual saliency models by incorporating depth information. Table IV summarises a few 3D saliency models reported in the literature. There are two main types of depth integrated saliency models, namely: Depth weighted 3D saliency model and Depth saliency model based methods. The depth weighted saliency models weight the 2D saliency map based on depth information. In depth saliency models, the predicted 3D saliency map is derived based on the chosen weights for 2D and depth saliency maps.

V. CONCLUSION

We presented the main concepts for the assessment of the Quality of Experience for 3D video streaming, highlighting the most recent achievements and current challenges. Studies on the human visual system together with the development of mathematical models will enable advances in this area, together with the availability of 3D video databases for the comparison of results among different research teams.

REFERENCES

[1] L. Meesters, W. IJsselsteijn, and P. Seuntiens, “A survey of perceptual evaluations and requirements of three-dimensional TV,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 3, pp. 381 – 391, March 2004.

[2] M. Martini, C. Chen, Z. Chen, T. Dagiuklas, L. Sun, and X. Zhu, “Guest editorial: QoE-aware wireless multimedia systems,” *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 7, pp. 1153–1156, 2012.

[3] P. Le-Callet, S. Moeller, and A. Perkis, “Qualinet white paper on definitions of quality of experience, version 1.1,” in *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)*, June 2012.

[4] K. Wang, M. Barkowsky, K. Brunnstrom, M. Sjostrom, R. Cousseau, and P. Le-Callet, “Perceived 3D TV transmission quality assessment: Multi-laboratory results using absolute category rating on quality of experience scale,” *IEEE Transactions on Broadcasting*, 2012, pub. ahead of print.

[5] J. Cutting and P. Vishton, “Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth,” *in Perception of Space and Motion, edited by W. Epstein and S. Rogers*, pp. 69 – 117, 1995.

[6] M. Lambooij, M. Fortuin, I. Heynderickx, and W. IJsselsteijn, “Visual discomfort and visual fatigue of stereoscopic displays: A review,” *Journal of Imaging Science and Technology*, vol. 53, no. 3, pp. 30 201–1 – 30 201–14, May 2009.

[7] P. Seuntiens, L. Meesters, and W. Ijsselsteijn, “Perceived quality of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation,” *ACM Transactions on Applied Perception (TAP)*, vol. 3, no. 2, pp. 95–109, 2006.

[8] A. Benoit, P. Le-Callet, P. Campisi, and R. Cousseau, “Quality assessment of stereoscopic images,” *EURASIP journal on image and video processing*, vol. 2008, 2008.

[9] C. Hewage, S. Worrall, S. Dogan, S. Villette, and A. Kondoz, “Quality evaluation of color plus depth map-based stereoscopic video,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 304–318, 2009.

[10] J. Seo, X. Liu, D. Kim, and K. Sohn, “An objective video quality metric for compressed stereoscopic video,” *Circuits, Systems, and Signal Processing*, pp. 1–19, 2012.

[11] M. Martini, M. Mazzotti, C. Lamy-Bergot, J. Huusko, and P. Amon, “Content adaptive network aware joint optimization of wireless video transmission,” *IEEE Communications Magazine*, vol. 45, no. 1, pp. 84–90, 2007.

[12] C. Hewage and M. Martini, “Edge-based reduced-reference quality metric for 3-D video compression and transmission,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 5, pp. 471–482, 2012.

[13] Z. Sazzad, S. Yamanaka, Y. Kawayokeita, and Y. Horita, “Stereoscopic image quality prediction,” in *Proc. International Workshop on Quality of Multimedia Experience (QoMEX)*, 2009, pp. 180–185.

[14] A. Maalouf and M. Larabi, “CYCLOP: a stereo color image quality assessment metric,” in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 1161–1164.

[15] M. Solh and G. AlRegib, “A no-reference quality measure for DIBR-based 3D videos,” in *Proc. IEEE International Conference on Multimedia and Expo (ICME)*, 2011, pp. 1–6.

PLACE PHOTO HERE

Chaminda T.E.R. Hewage Chaminda T.E.R. Hewage received the B.Sc. Engineering (Hons.) degree in Electrical and Information Engineering from the University of Ruhuna, Sri Lanka. During 2004-2005, he worked as a Telecommunication Engineer for Sri Lanka Telecom, Ltd.. He completed his Ph.D. (Thesis: Perceptual quality driven 3D video over networks) at Centre for Communication Systems Research, University of Surrey, UK. From September 2009 he is attached to the Wireless and Multimedia Networking Research Group at Kingston University, London, UK.

TABLE III
AVAILABLE 3D IMAGE/VIDEO DATABASES

3D image/video database	Creator	Artifacts
Mobile 3D video database	University of Tampere and Nokia	Crosstalk, blocking, colour mismatch and bleeding, packet losses for low-resolution video (only impaired sequences, no MOS values provided).
IRCCyN 3D image database	University of Nantes	JPEG, J2K, upsample/downsample, etc.
EPFL databases for images/videos	EPFL	Different camera distances
Kingston University video database	Kingston University-London	Packet losses
NAMA3DS1-COSPAD1	University of Nantes	H.264 and JPEG2000 compression artifacts
RMIT3DV	RMIT University	Uncompressed HD 3D video

TABLE IV
3D IMAGE/VIDEO VISUAL ATTENTION MODELS

3D visual attention model	Creator	Model type and considered image features
J. Wang, M.P. Da Silva, P. Le Callet, V. Ricordel. (2013)	University of Nantes	2D + Depth saliency (using both current and prior image information), motion information is not taken into account
Y. Zhang, G. Jiang, M. Yu, K. Chen (2010)	Ningbo University	2D + Depth + Motion
E. Potapova, M. Zillich, M. Vincze (2011)	Vienna University of Technology	2D + Depth saliency (based on surface height and relative surface orientation)
N. Ouerhani, H. Hugli (2010)	Neuchatel University	2D + scene depth

PLACE
PHOTO
HERE

Maria G. Martini Maria G. Martini is a Reader (Assoc. Prof.) in Kingston University (KU), London. She received the Laurea in electronic engineering (summa cum laude) from the University of Perugia (Italy) in 1998 and the Ph.D. in Electronics and Computer Science from the University of Bologna (Italy) in 2002. She has led the KU Wireless Multimedia Networking research group in national and international research projects. Her research interests include wireless multimedia networks, cross-layer design, 2D/3D video quality assessment.