MIQM: A NOVEL MULTI-VIEW IMAGES QUALITY MEASURE

*Mashhour Solh and Ghassan AlRegib**

School of Electrical and Computer Engineering
Georgia Institute of Technology
{msolh,gregib}@ece.gatech.edu

ABSTRACT

Although several subjective and objective quality assessment methods have been proposed in the literature for images and videos from single cameras, no comparable effort has been devoted to the quality assessment for multi-camera images. With the increasing popularity of multi-view applications, quality assessment for multi-camera images and videos is becoming fundamental to the development of these applications. The quality of images, which are captured by a multi-view system, are affected by multiple factors such as camera configuration, number of cameras, and the calibration process. In the process of developing an objective metric specifically designed for multi-camera systems, we identified two types of visual distortions in multi-view images: photometric distortions and geometric distortions. In this paper, we show that in the presence of well defined reference these distortions can be translated into luminance, contrast, spatial motion and edge-based structure components. Then, we propose different index values that can quantify these components. We provide several examples to demonstrate the correlation between each of these components and the corresponding index metric. Then, we combine these indexes into one Multi-view Image Quality Measure (MIQM). The results and examples show that not only MIQM can capture the perceptual quality of multi-view images it also outperforms the Structural SIMilarity(SSIM) measure for single-view images quality assessment.

Index Terms— multi-view,quality assessment, geometric distortion, photometric distortion, immersive communication

1. INTRODUCTION

In the last decade, multi-view capture of events has gained increasing interest as a vital tool to satisfy the demand for advanced immersive multimedia products. Application fields include surveillance, sightseeing, advertisement, distance learning, medical training, and entertainment. Multi-view is a set of images or videos captured by a set of two or more cameras. The key advantage of multi-view applications is interactivity. The user of these applications has the freedom of choosing the viewpoint within the captured scene. The processing chain of multi-view applications includes image capturing, camera calibration, scene presentation, coding, transmission, multi-view rendering, and display. Where each one of these affects the perceived quality of the image or video at the output side. Over the last decade, subjective evaluation has been the dominant performance metric in multi-view videos and images processing. Ideally, image and video quality is best assessed through subjective evaluation; however the use of subjective testing is both inefficient and time consuming. Subjective methods are also not applicable in environments that require real-time processing. Therefore, the definition of an objective metric or set of metrics that can reliably predict the perceived quality of images and videos in multi-view applications is vital to the development of these applications.

A great effort has been devoted by academic and industrial communities to develop objective quality metrics for single-view images and videos. The amount of work dedicated for objective multi-view image quality assessment, on the other hand, is much less. Due to the nature and applications of multi-camera systems there are some multi-view distortions that are not common in single-camera images and videos. In [1], we characterized these distortion types and provided examples showing that single-view objective image quality measures are not adequate for multi-view perceptual quality assessment. In this paper, we show that geometric and photometric distortions can be analyzed in terms of luminance, contrast, spatial motion and edge-based structure components. Then, in the process of defining an objective metric for multi-view images we present three index measures that could objectively quantify these components. We combine these indexes into one multi-view image quality measure (MIQM). Finally, we provide experimental results comparing MIQM to both subjective ratings and Structural Similarity (SSIM) on a database of single-view images [2]. The results and examples show that not only MIQM can capture the perceptual quality of multi-view images it also outperforms SSIM for single-view images quality assessment.

The rest of this paper is organized as follows. In Section 2 we provide a background on distortions types in multi-camera

*This work is funded in part by HP laboratory.

systems and related work on multi-view quality assessment. The components of multi-view distortion are presented along with the proposed measures in Section 3. Section 4 describes the simulation results. Finally, conclusions are drawn in Section 5.

2. PRIOR ART

In [1], we classified the multi-camera image distortions into photometric and geometric distortions. Photometric distortions are the visible variations in brightness levels and color gamut across the entire displayed image. The source of this variation can be the non-uniformity between individual camera properties or the post processing applications such as compression. Geometric distortions are the visible misalignments, discontinuities and blur in the processed image. These distortions could result from noticeable calibration errors between adjacent cameras, affine/linear corrections, and error in scene geometry estimations. In manually built multi-camera arrays, these errors could also result from the mismatch in the vertical and horizontal directions among images and irregular camera rotations.

Leorin et al.[3] used subjective tests to show that current single video camera quality assessment techniques are not adequate for quality assessment of omnidirectional panorama video generated by multiple cameras. The authors then proposed an objective quality metric for omnidirectional video. The metric assessed the general quality of the video using no-reference blockiness and blur measure, and structural simimilarity (SSIM) [4] for each camera and then assigned higher weights to regions where motion is present. Their proposed work mainly addressed the color calibration problem across the seam and concentration of motion in limited areas of the panorama. However, geometric distortions and photometric variations such as blur and quantization were not considered.

Several objective quality metrics were proposed for multi-view video in stereoscopic 3D applications [5][6][7] and in 3D reconstructions [8]. The authors in [5] performed quality assessment of stereo image pairs using single-view quality metrics on each view. Several combination methods of the quality scores from each view were then evaluated to determine the ones that best correlate with the subjective scores. The same level of distortion was applied on both images of the stereo pair and the distortion types were limited to blur and compression artifacts. A similar approach was adapted by the authors in [7]. Along these efforts, Ozbek et al. in [6] assumed that the Peak Signal to Noise Ratio (PSNR) of the second view is less important for 3D visual experience and the new measure was composed of weighted combination of the two PSNR values for the two views and a jerkiness measure for temporal artifacts. An objective metric for free-viewpoint video production was proposed in [8]. The metric can be used as full-reference measure of fidelity for aligning structural details in presence of an approximate scene geometry of the 3D shapes. In [9] the authors proposed using conventional single camera quality measures (PSNR and SSIM) for 3DTV video as a quality measure for video plus depth content by measuring the quality of the virtual views that are rendered from the distorted color and depth sequences. The undistorted reference sequence is obtained by rendering virtual views from the original color and depth maps. The metric optimizes the visual quality of encoded 3D video, thus it assumed a geometric distortion-free video sources.

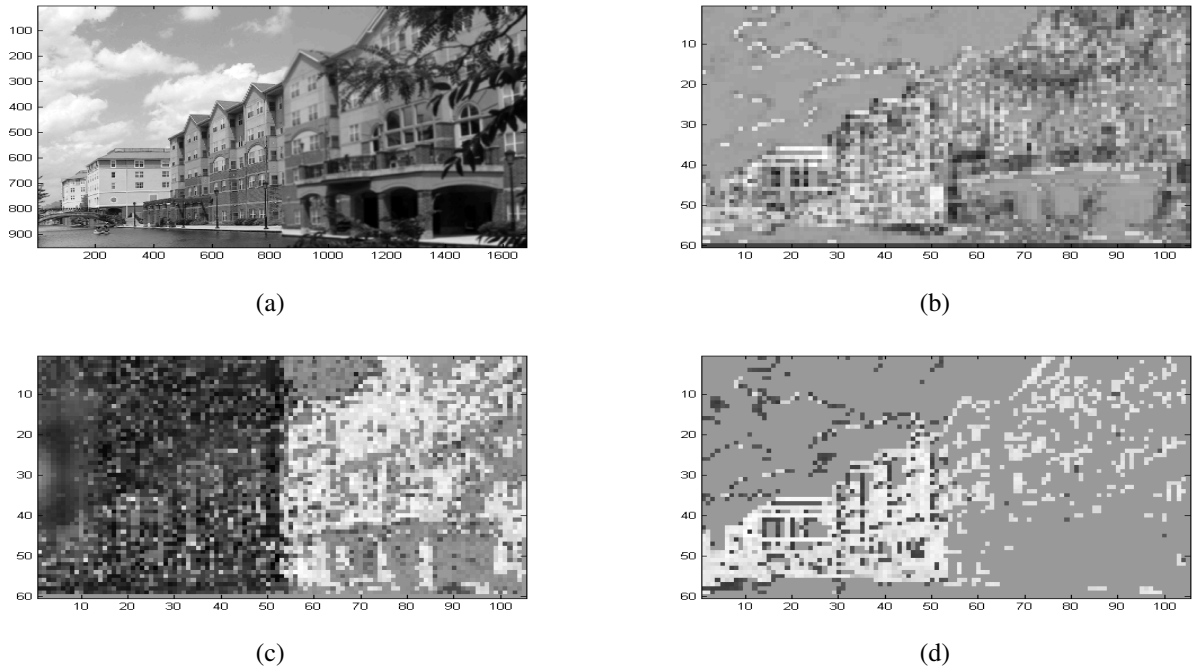
The limited work in the literature on multi-view image and video quality assessment has been dominated by attempts to define the multi-view quality metric as a combination of conventional single-view quality metrics. It has been shown that conventional single-view quality metrics do not correlate with the quality of multi-view images and videos [1][5][7]. Due to the nature and applications of multi-camera systems there are some multi-view distortions that are not common in single-camera images and videos. The limited work in the literature on multi-view image and video quality assessment did not address the variational photometric and geometric distortions in multi-camera systems. In the process of defining a quality metric that captures all types of distortion in multi-camera systems we arrived into three index measures that would reflect the visual properties of these distortions. None of these index measures alone can fully capture the perceptual distortions in multi-view images. However the combination of the three measures is necessary to capture the impact of these distortions on multi-view perception. We will call the combined measure the Multi-view Image Quality Measure (MIQM). In the next section we will present these index measures and the rational behind adopting each one of them. Then, we will show how to compute the MIQM.

3. QUALITY ASSESSMENT OF MULTI-VIEW IMAGES

Multi-camera applications are numerous [10] and each application has specific means for presentation and post processing. In these applications, a single camera is usually chosen as a reference for estimating the imaging plane or geometry [11]. The measures we are about to present are full-reference and aim at assessing the image quality for multi-camera systems. We define the reference as the set of images captured by perfectly identical set of cameras, and the planes of these cameras are perfectly aligned horizontally and vertically with the camera chosen to be the reference for the imaging plane or geometry. You can think of such perfect imaging to be performed by a high-definition camera with a single sensor.

3.1. Luminance and Contrast Index

Human perception is sensitive to abrupt local changes in luminance and contrast around structured regions [1]. Such changes are more common in multi-view images. Multi-view



(a) (b) (c) (d)

Fig. 1. Index maps: (a) Distorted multi-view image (b) Luminance and contrast index map (c) Motion index map (d) Edge-based structural index map

images captured by cameras looking at different parts of the scene are subject to non-uniform levels of distortion due to the difference between different cameras or different levels of view processing. To capture this phenomena we employ a measure that is a combination of luminance and contrast comparison function from [4] adjusted to give higher weights for structured regions. The combined luminance and contrast comparison function on a macroblock (x, y) is

$$k_{x,y}(I, J) = \frac{4(\sigma_I \sigma_J)(\mu_I \mu_J) + C}{(\sigma_I{}^2 + \sigma_J{}^2)(\mu_I{}^2 + \mu_J{}^2) + C} \qquad (1)$$

where I is the reference image, J is a distorted image, μ is the mean intensity and σ is the standard deviation. Both σ and μ are computed on macroblocks of dimension $s \times s$ and (x, y) is the coordinate of the upper left corner of the macroblock. C is a constant added to avoid instability when the denominator is close to zero. The overall quality index is calculated as a weighted average of individual macroblock quality index values. These weights are extracted from the reference image through employing a modified version of the texture model for bit rate-allocation in [12]. First the texture randomness index at macroblock (x, y) of the image I is computed by

$$t(x, y) = EI(x, y) \times MI(x, y) \qquad (2)$$

where EI is an edge intensity binary image with 1's where the function finds edges in I and 0's elsewhere, and MI is the mean intensity of I. Both EI and MI are computed on non-overlapping macroblocks of dimension $s \times s$. The texture randomness index is then mapped to the object index as in

$$T(x, y) = \begin{cases} K_1 + (0.5 \times K_1 \times \frac{log_2 t(x,y)}{log_2 \beta_1}) & \beta_1 \leq t(x, y) < \beta_2 \\ K_2 + (0.5 \times K_2 \times 2^{-(t(x,y)-\beta_2)}) & t(x, y) \geq \beta_2 \\ K_1 & \text{otherwise} \end{cases} \qquad (3)$$

where K_1 and K_2 are constant parameters chosen to control the weights assigned to the structured regions and randomly textured regions, respectively. By setting $K_1 \gg K_2$, higher weights are assigned to the structured regions. Parameters β_1 and β_2 are the edge detector thresholds. The human visual system is less sensitive to intensity variations in randomly textured region corresponding to values greater than β_2 in $t(x, y)$. $T(x, y)$ is designed to drop quickly around this region and to increase exponentially around structured regions corresponding to the values between β_1 and β_2. Low textured or smooth regions where $t(x, y)$ is less than β_1 are assigned a constant value. The luminance and contrast index for $M \times N$ total macroblocks is $K(I, J)$ where

$$K(I, J) = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} k_{x,y}(I, J) \times T(x, y)}{\sum_{x=1}^{M} \sum_{y=1}^{N} T(x, y)} \qquad (4)$$

$K(I, J)$ values range between 1 for minimum distortion and 0 for maximum distortion.

Fig. 2. Abrupt local changes in luminance and contrast can be very perceptually distrubing (face).

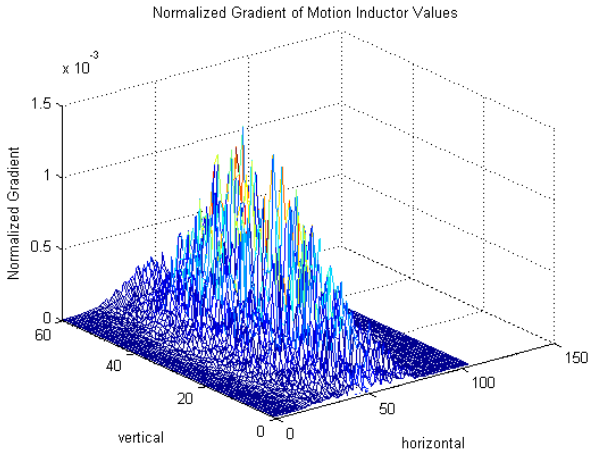


Fig. 3. Normalized gradient of the relative motion inductor values for image Fig.1.

The correlation between the three index measures and image properties is demonstrated in the example of Fig.1. The image shown in Fig.1(a) is a mosaic image of two-views. Prior to composting, perspective distortion was applied to the left view and the right view was blurred. The luminance and contrast index map is shown in Fig.1(b). The darker regions refer to areas of higher distortions. From Fig.1(b), the right half of the image corresponding to the the blurred view has darker regions than once observed in the left half corresponding with perspective distortion. Hence, the luminance and contrast index captures the perceptual distortion at the blurred side of the image with emphasis on structured objects. The latter is very important because abrupt local changes in luminance and contrast around structured regions can be very disturbing as shown in Fig.2.

3.2. Spatial Motion Index

A complex wavelet domain image similarity was proposed in [13] that is insensitive to spatial translations. The proposed model assumes that in single-view image perceptual distortions due to spatial scaling, rotation and translation are insignificant. However, this assumption is not true for multi-view images where significant distortions such as discontinuities, misalignments, blur and double image can result. Geometric distortions in multi-view images are the result of displacements or shifts at the pixel locations with respect to the reference image. In [12] a motion attention model was adopted for bit allocation in single-view video compression. In 2D these displacements are comparable to spatial motion of single-view videos. Hence, the motion attention model can be used to quantify geometric distortions. To compute the motion index, first the relative motion inductor at macroblock (x, y) is computed as

$$m(x,y) = \frac{\sqrt{mvx^2 + mvy^2}}{\sqrt{2p^2}} \tag{5}$$

where (mvx, mvy) is the motion vector of a macroblock (x, y) in the distorted image J relative to the reference image I. The motion search area is $p \times p$ macroblocks. The motion index at (x, y) is then computed as

$$M_{x,y}(I,J) = |1 - \frac{m(x,y) \times en(x,y)}{max(m(x,y) \times en(x,y))}| \tag{6}$$

where $en(x, y)$ is the entropy of the $m(x, y)$ values calculated within a spatial window of $w \times w$ macroblocks where $w >> p$. The entropy is low for regions with coherent displacements and high for regions with random displacements. The values of motion index are designed to be higher at regions of random displacements caused by photometric distortions such as blur. On the other hand, these values are designed to be lower at regions of coherent displacements corresponding to geometric distortions. The spatial motion index map of (6) is shown in Fig.1(c). The darker values over the left half indicates the spatial displacements due to the geometric errors. The perceptual distortion due to these geometric errors is presented in form of visible misalignments, discontinuities and blur around overlapping areas and across the seam of intersections. In order to accurately measure the perceptual distortion due to geometric errors, the spatial displacements at these locations must be assigned higher weights. This can be achieved by calculating the gradient of the relative motion inductor values. The weighting factors can be calculated in terms of normalized gradients:

$$w(x,y)(I,J) = \frac{\nabla m(x,y) \times n(x,y)}{\sum_{x=1}^{M} \sum_{y=1}^{N} \nabla m(x,y) \times n(x,y)} \tag{7}$$

where $n(x, y)$ is a gaussian low pass filter and

$$\nabla m = \sqrt{(\frac{\partial m}{\partial x})^2 + (\frac{\partial m}{\partial y})^2} \quad (8)$$

The above function assigns higher weights across the seam of intersections and the overlap areas as show in Fig. 3. The spatial motion index is then computed for an image of $M \times N$ total macroblocks as

$$M(I,J) = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}|1 - \frac{m(x,y)\times en(x,y)\times w(x,y)}{max(m(x,y)\times en(x,y)\times w(x,y))}|}{M \times N} \quad (9)$$

$M(I, J)$ values range between 1 for minimum distortion and 0 for maximum distortion.

3.3. Edge-based structural Index

Evaluating the structural similarity over edge maps instead of the actual images leads to better correlation with subjective quality for SSIM [14]. Spatial edges are defined as the locations of variations of intensity values and the relative intensity values at these locations. When an image is blurred or quantized the locations of the spatial edges are preserved however the intensity values of these edges change. In geometric distortions such as translations and rotations the spatial edge locations change where there relative intensity is preserved. Hence, by comparing the local edge information we can capture the loss of structural information due to both photometric and geometric distortions. The edge-based structural index for $M \times N$ total macroblocks is computed as

$$E(I,J) = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(1 - |\frac{(T_{x,y}(I) - T_{x,y}(J))}{T_{x,y}(I)}|)}{M \times N} \quad (10)$$

where $T_{x,y}(I)$ and $T_{x,y}(J)$ are defined as in (3) for images I and J respectively. $E(I, J)$ values range between 1 for minimum distortion and 0 for maximum distortion. It can be observed from Fig.1(d) that the structural losses represented by the edge-based structural index are mainly concentrated on the blurred view at the right, with some dark dots on the left view. Geometric distortion preserves global structures however the positioning and orientation of the structure are changed. Structural losses in geometric distortions may occasionally occur around a macroblock boundary in a low structured region (the clouds region in the left view).

3.4. Multi-view Image Quality Measure (MIQM)

The measures presented in the previous subsections can be combined over various dimensions, or all dimensions, to yield a single measure that summaries the visual distortions in multi-view images. In this paper, MIQM is computed as the multiplication of the fore mentioned three index measures:

$$MIQM(I,J) = K(I,J) \times M(I,J) \times E(I,J) \quad (11)$$

	Noise Type		
Image	Left View	Middle View	Right View
Test1	rotated	rotated	none
Test2	none	none	perspective
Test3	high blur	none	high blur
Test4	quant. Q=15	quant. Q=10	quant. Q=15
Test5	perspective + high blur	none	none
Test6	perspective	high blur	high blur

Table 1. Summary of the types of noise that were applied to the various views of the test images.

Image	$K(I,J)$	$M(I,J)$	$E(I,J)$	$MIQM(I,J)$
Test1	0.9615	0.9887	0.9882	0.9403
Test2	0.8957	0.9669	0.9731	0.8428
Test3	0.7938	0.9999	0.9657	0.7614
Test4	0.5853	0.9690	0.9360	0.5308
Test5	0.8491	0.9625	0.9610	0.7854
Test6	0.6637	0.9590	0.8895	0.5662

Table 2. Luminance/contrast index ($K(I, J)$), motion index ($M(I, J)$) and edge-based structural index ($E(I, J)$) values computed for a set of six images with different distortions

where values range between 1 for minimum distortion and 0 for maximum distortion.

4. SIMULATION RESULTS

The parameter settings for our simulations in this paper are stated as follows. The block size is $s = 16$, the motion search parameters are $p = 7$ and $w = 9$, the constants are $C = 2.5$, $K_1 = 128$, $K_2 = 64$, $\beta_1 = 10$, and $\beta_2 = 100$. Our simulations have shown that a different choice of parameters does not significantly impact the results. The Canny edge detection method was used for edge intensity calculations in (2) and (10). Canny's method is less likely than the others to be fooled by noise, and more likely to detect true weak edges[15].

We used six test images from our own database and we created three views for each image. The original unsplit image is the reference perfect image. For each of the views of the split images, we applied a certain type of noise as summarized in Table 1. Table 2 shows the three index values for a set of six multi-view images randomly chosen from our database. The $M(I, J)$ values are always less than one reflecting some displacement in intensity values. However, these index values are lower for images with geometric distortions (1, 2, 5 and 6). The $K(I, J)$ values indicate changes in local contrast and luminance in all images. These values are lower in photometrically blurred and highly quantized images (3,4, 5 and 6). $E(I, J)$ values indicate very close structural losses in all the images. The $MIQM(I, J)$ values reflect the combination of these attributes into a single measure. This measure does correlate with the overall perceived quality of the sample images. The validation of this result requires a comparison against comprehensive subjective testing.

	$PSNR$	$SSIM$	$MIQM$
RMSE	13.43	9.369	9.0568

Table 3. RMSE Comparison - PSNR, SSIM and MIQM

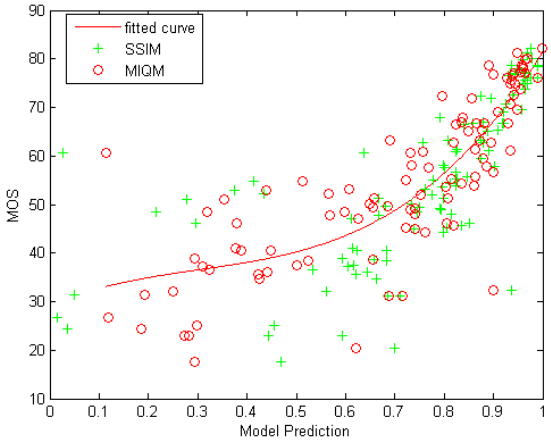


Fig. 4. Scatter plot of subjective mean score (MOS) versus the prediction model.

While our current work is directed toward building a database for subjective evaluation of multi-view images, we evaluated MIQM on the LIVE database used by [2]. Our purpose is to show the effectiveness of MIQM in evaluating the perceptual fidelity in single images and to compare the performance of MIQM versus the SSIM measure. The LIVE database contains 29 reference single-view images distorted by one of the five distortion types: JPEG2000 (227 images), JPEG (233 images), white noise - WN(174 images), Gaussian blur - GBlur (174 images), and fast fading- FF (174 images). Table 3 shows the improvement in RMSE performance of MIQM over both PSNR and SSIM. In Fig.4,we plot the subjective score values (MOS) and the normalized image quality values generated by SSIM, and MIQM when applied to a set of images from the database. The closer the measure values to the fitted curve the closer the SSIM and MIQM is to human judgment. The figure shows that MIQM predictions tends to be much closer to the human judgement for higher distortion values than SSIM.

5. DISCUSSION

In this paper, we introduced a multi-view image quality measure (MIQM) as a combination of three index measures that quantify the physical nature of multi-view image distortions. Then we presented simulation results of several multi-view images showing that none of these measures alone can fully capture the perceptual distortions in multi-view images, however the combination of the three measures captures the physical attributes of these distortions. Finally, we evaluated MIQM against database of single-view images. The results and examples show that MIQM outperforms SSIM for single-view images quality assessment. We consider single-view as a special case of multi-view images. However, in order to validate our measure for multi-view images MIQM must be evaluated a gainst MOS of a database of multi-view images. Our on going research is to build a database of MOS for sets of multi-view images with various geometric and photometric distortion. The database will be used to evaluate MIQM performance and determine which combination of the index measures is optimal for multi-view images quality prediction.

6. REFERENCES

[1] M. Solh and G. AlRegib, "Characterization of Image Distortions in Multi-Camera Systems," in *IMMERSCOM09*, May 2009, accepted.

[2] H. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE Image Quality Assessment Database Release 2," *http://live.ece.utexas.edu/research/quality*.

[3] S. Leorin, L. Lucchese, and R. Cutler, "Quality Assessment of Panorama Video for Videoconferencing Applications," in *Workshop on Multimedia Signal Proc.* IEEE, November 2005, pp. 1–4.

[4] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Trans. on Image Proc.*, vol. 13, pp. 600–612, 2004.

[5] P. Campisi, A. Benoit, P. Callet, and R. Cousseau, "Quality Assessment of Stereoscopic Images," in *(EUSIPCO)*. EURASIP, September 2007.

[6] N. Ozbek, A.Tekalp, and E. Tunali, "Rate Allocation Between Views in Scalable Stereo Video Coding using an Objective Stereo Video Quality Measure," in *ICASSP*, April 2007, pp. 1045–1048.

[7] C. Hewage, S. Worrall, S. Dogan, and A. Kondoz, "Prediction of Stereoscopic Video Quality Using Objective Quality Models of 2-D Video," *IEEE Elect. Letters*, vol. 44, pp. 963–965, July 2008.

[8] J. Starck, J. Kilner, and A. Hilton, "Objective Quality Assessment in Free-Viewpoint Video Production," in *3DTV08*, 2008, pp. 225–228.

[9] A. Tikanmaki, A. Gotchev, A. Smolic, and K. Miller, "Quality assessment of 3D video in rate allocation experiments," in *IEEE symposium on Consumer Electronics*, April 2008, pp. 1–4.

[10] "ISO/IEC JTC1/SC29/WG11 Applications and Requirements for 3DAV," Tech. Rep. N5877, Trondheim, Norway, July 2003.

[11] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge Uni. Press, 2000.

[12] C. Tang, C. Chen Y. Yu, and C. Tsai, "Visual sensitivity Guided Bit Allocation for Video Coding," *IEEE Trans. on Multimedia*, vol. 8, pp. 11–18, 2006.

[13] Z. Wang and E. Simoncelli, "Translation Insensitive Image Similarity in Complex Wavelet Domain," in *ICASSP-05*, March 2005, pp. 573–576,.

[14] L. Cui and A. R. Allen, "An Image Quality Metric Based on Corner, Edge and Symmetry Maps," in *British Machine Vision Conf.*, 2008.

[15] F. John Canny, "A Computational Approach to Edge Detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.