MIQM: A Multi-camera Image Quality Measure

Mashhour Solh, Student Member, IEEE, and Ghassan AlRegib, Senior Member, IEEE

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 of multi-camera images and videos is becoming fundamental to the development of these applications. Image quality is affected by several factors such as camera configuration, number of cameras, and the calibration process. In order to develop an objective metric specifically designed for multi-camera systems we identified and quantified two types of visual distortions in multi-camera images: photometric distortions and geometric distortions. The relative distortion between individual camera scenes is a major factor in determining the overall perceived quality. In this paper, we show that such distortions can be translated into luminance, contrast, spatial motion and edge-based structure components. We propose three different indices that can quantify these components. We provide examples to demonstrate the correlation among these components and the corresponding indices. Then, we combine these indices into one Multi-camera Image Quality Measure (MIQM). The results and comparisons with other measures such as PSNR, MSSIM and VIF show that MIQM outperforms other measures in capturing the perceptual fidelity of multi-camera images. Finally, we verify the results against subjective evaluation.

Index Terms

multi-view imaging, multi-camera arrays, perceptual quality, image quality assessment, fidelity measure

I. INTRODUCTION

With the rapid improvement in electronics and computing technologies and dropping costs of cameras, multi-capture of events has gained increasing interest as a vital tool to satisfy the demand for advanced

M. Solh and G. AlRegib are with the Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 30332 USA e-mail: {msolh, alregib}@gatech.edu
immersive multimedia products. Applications for such products include video conferencing, surveillance, sightseeing, advertisement, distance learning, medical training, and entertainment. Multi-view is a set of images or videos captured by 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. Multi-view applications include but are not limited to panoramic videos, free-viewpoint video, 3DTV, virtual view synthesis, object tracking, and stereoscopic video [1] [2]. The processing chain of these applications consists of image capturing, camera calibration, scene presentation, coding, transmission, multi-view rendering, and display [2]. Each step in the processing chain affects the perceived quality of the image or video at the output. 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 [3]. 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 of multi-view applications is vital to the development of these applications.

Multi-camera applications are numerous and each application has its specific means of acquisition, representation and display. The definition of the quality of the perceived multi-view video or image is dependent on the means of presentation. There are several means of presentation of multi-view videos and images: panoramic, interactive stereo, free viewpoint, and 3DTV [1]. In panoramic video applications multiple cameras are used to capture a particular scene. The outputs of cameras are then combined to emulate the performance of a much costlier multi-megapixel wide-angle video camera. In interactive stereoscopic video two cameras are used to capture two views of an object from slightly different positions. Then a 3D impression of the scene is created by projecting the 2D slightly different scenes on the retina of each eye. The human brain creates the impression of depth through physiological fusing of the stereoscopic pair. In the case of free viewpoint video a scene is captured by multiple cameras. Through a combination of video sources and some information about camera calibration and scene geometry (e.g. disparity data) free viewpoint video allows the user to navigate through the image by choosing his/her own viewpoint. Finally, in 3DTV a scene is captured as in multiple view video, and one or more 3D video objects are created. The cameras are arranged with relatively short baseline to synthesize virtual views directly from camera images.

Although each of the previously mentioned applications suffers from synthetic visual artifacts that are exclusive to its mean of presentation, they all share similar acquisition apparatus and pre-compositing...
processing block. The acquisition apparatus involves multiple cameras placed under specific arrangement to capture multiple views of a real world scene. The captured views are then photometrically and geometrically calibrated before being composited to be displayed. Different views captured by different cameras may vary in terms of color, brightness, noise level and orientation. The calibration process derives the necessary information to map each of the views dimensions into the real world or the reference view dimensions. The perceived scene for each of the multi-camera applications is an output of the compositing algorithm which is generally a function of the captured scenes, camera calibration and scene geometry. Hence, defining a single quality measure that would capture the perceived quality of all multi-camera applications is impossible considering the difference in the means of presentation and the view compositing algorithms for each application. In this paper we define a multi-camera image quality measure (MIQM). MIQM was tested and refined for ultra-high resolution panoramic image applications. The resulting measure captures the visual effects of artifact introduced at the acquisition and pre-compositing processes to predict the composited image quality. The measure was developed based on acquisition and pre-compositing artifacts to serve as a basis from which we could extend the results to develop quality measures for stereoscopic, free viewpoint, and 3DTV applications after taking into consideration stereoscopic impairments and synthetic artifacts.

Distortions in multi-camera systems can be classified into geometric and photometric [4]. MIQM is composed of three index measures: the luminance and contrast index, the spatial motion index and the edge-based structural index. Photometric distortions are mostly measured by the luminance and contrast index and the edge-based structural index. Geometric distortions on the other hand are mostly measured by the spatial motion index and the edge-based structural index.

In this paper we first discuss the state of the art work done in literature on multi-view quality assessment. We then review the multi-camera distortions as presented in a previous work [4] with examples in section 3. In section 4 we present our three indexed multi-camera image quality measure (MIQM) with reasoning leading to each index. In section 5 we analyze the performance of MIQM against subjective data collected on a set of distorted images. Finally we conclude our findings in section 6.

II. STATE OF THE ART

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 (see Table I for a summary of existing measures). Leorin et al. [5] used subjective tests to show that current single video camera quality assessment tech-
TABLE I: A summary of existing multi-view quality measures. The table shows the application for each quality measure and the applicability of each measure to the geometric and the photometric distortions.

<table>
<thead>
<tr>
<th>Quality Measure</th>
<th>Application</th>
<th>Applicable for Geometric Distortions</th>
<th>Applicable for Photometric Distortions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leorin et al. [5]</td>
<td>panoramic video</td>
<td>No</td>
<td>partially (only across the seam)</td>
</tr>
<tr>
<td>Campisi et al. [8]</td>
<td>stereoscopic images</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Ozbek et al. [9]</td>
<td>stereoscopic images</td>
<td>No</td>
<td>partially (limited to PSNR of better quality view)</td>
</tr>
<tr>
<td>Hewage et al. [10]</td>
<td>stereoscopic images</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Starck et al. [11]</td>
<td>free-viewpoint video</td>
<td>partially</td>
<td>No</td>
</tr>
<tr>
<td>Tikkanmaki et al. [12]</td>
<td>3DTV</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Techniques are not adequate for quality assessment of omnidirectional panorama video generated by multiple cameras. Panoramic video image plane could be spherical, cylindrical, or even hyperbolic. The number of cameras and configuration of the cameras for multi-view panoramic video application are two parameters that depend on the scene geometry and the desired resolution. Increasing the number of cameras, for instance, would cover larger scene areas if cameras are widely spread and a denser arrangement of the cameras would cover a smaller area with better resolution. In multi-camera panoramic video applications different camera setting could be possible. Fig. 1 shows three possible camera configurations, i.e. parallel view, convergent and divergent view. In addition to common artifacts present in digital video streams such as blur, blocking artifacts, noise and ringing, quality assessment of panoramic videos has to emphasize problems in multi-view panorama such as noticeable calibration and intrinsic differences between adjacent cameras, concentration of motion in limited regions of the scene, combined emphasis problems, error in the image mosaicking and double image effects [5] [6]. The authors in [5] 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 similarity (SSIM) [7] for each camera and then assigned higher weights to regions where motion is present. The 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 compression artifacts were not considered.

Several objective quality metrics were proposed for multi-view video in stereoscopic 3D applications [8] [9] [10] and 3D reconstructions [11]. The authors in [8] 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 [10]. Ozbek et al. [9] assumed that PSNR of the second view is less important for 3D visual experience and the new measure was composed of weighted combination of two PSNR values and a jerkiness measure for temporal artifacts. An objective metric for free-viewpoint video production was proposed in [11]. The metric can be used as full-reference measure of fidelity of aligning structural details in presence of approximate scene geometry of the 3D shapes. In [12] 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 proposed optimizing 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 [8] [10] [4]. Due to the nature and applications of multi-camera systems there are multi-view distortions that are not common in single-camera images and videos. In a previous work [4] we classified multi-camera distortions and defined the distortions that have been overlooked by multi-view quality assessment literature. In the next section we will summarize our findings presented in [4] and provide an analysis of the multi-camera distortion properties.
III. DISTORTION TYPES IN MULTI-CAMERA IMAGES

A. Simulating Multi-Camera Images

In order to simulate distortions in multi-camera images a single digital camera was used to capture high-resolution images. Each image was then split into multiple sub-images with overlap areas. The overlap areas where varied with each image, however they were all in the range of 5% – 10% of the original image. Distortion was then applied separately on each individual sub-image. The multi-camera image was then simulated by compositing the sub-images into a single image mosaic using a multi-resolution spline [13]. The reference image is created by combining all sub-images without any distortion.

B. Photometric Distortion

Photometric distortion in a single camera is defined as the degradations in perceptual features that are known to attract visual attention such as noise, blur, and blocking artifacts. Photometric distortion can be intrinsic due to the acquisition device or extrinsic due to applications such as lossy compression, transmission over error prone channels, or image enhancements. Quantifying the perceptual quality of these distortion types is essential to the improvements or developments of new video or image applications and hence has motivated the development of contemporary image and video quality metrics.

In multi-camera systems, 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. This type of distortion will be referred to as the variational photometric distortion.

In order to simulate photometric distortions in multi-camera images targeted distortion was applied on each sub-image independently prior to reconstruction. Fig. 2 shows four examples of images with variational photometric distortion. The images in Fig. 2a and Fig. 2b are composed each of two sub-images. The right view of the image in Fig. 2a was distorted by applying JPEG compression with $Q = 5$, while the left view was left undistorted. Both the left and right views of Fig. 2b were distorted by applying Gaussian blur, however higher level of blur was applied to the right view. Images in Fig. 2c and Fig. 2d are composed each of three sub-images. The left view of Fig. 2c was distorted with a Gaussian blur and middle view was distorted by applying JPEG compression with $Q = 10$. The left and right views of image in Fig. 2d were both distorted by applying JPEG distortion with $Q = 10$ and $Q = 5$ respectively. The right view of Fig. 2c and the middle view of Fig. 2d are both left undistorted.

\footnote{each sub-image is a single view in a multi-view setting}
C. Geometric Distortion

The second type of image distortions in multi-camera systems is *geometric distortions*. In multi-camera systems a scene captured by $N$ cameras can vary with each individual camera’s position and orientation. Geometric distortions are the visible misalignments, discontinuities and blur in the processed image.
Fig. 3: Example of geometric distortion in a single view image: (a) Original (no distortion), (b) Planar (rotation), and (c) Perspective.

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. There are two types of geometric distortions: planar and perspective distortions. Planar distortions can occur during the mapping, which may include rotation and translation. Perspective distortions can occur in the mapping from the 3D world to the 2D plane of the image. Fig. 3 shows examples that illustrate the types of geometric errors as well as the original image. The image in Fig. 3c is subject to perspective distortion. The columns look closer than the original image Fig. 3a. The image in Fig. 3b is rotated clockwise by 3 degrees. In multi-camera systems such errors can also occur when mapping a certain camera plane to another reference camera plane in the system.

To simulate the geometric distortions in multi-camera system, geometric distortions were applied to the generated views independently and then reconstructed into a single image mosaic. Fig. 4 shows two examples of geometric distortions in multi-view images. The image Fig. 4a is composited of two sub-images with a 5% overlap. The left view of image Fig. 4a was perspective-distorted whereas the right view was left undistorted prior to reconstruction. The result is severe perceptual distortion that
Fig. 4: Example of geometric distortion in a multi-view images: (a) Perspective and (b) Planar (rotation).

is very obvious on the face. The image Fig. 4b is composed of three sub-images with a 20% overlap between each two adjacent views. Two levels of perspective and planar distortions were applied to the left and right views respectively. The middle view was not distorted. The resulting multi-view image has noticeable misalignments and discontinuities. Hence, the geometric distortions in single camera translate to misalignment and discontinuities in the reconstructed multi-view image. Unlike photometric distortions where distortions translate as abrupt changes that occur across the whole image, geometric distortions attracts perceptual attention especially around connecting edges and overlapping areas. Geometric distortions in single view images have been considered in [14]. The authors proposed a complex wavelet domain image similarity 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 imaging can result. Therefore, geometric distortions must be accounted for in multi-camera image visual quality assessment.

D. Properties of Multi-Camera Distortions

The properties of multi-camera images that influence the design of the proposed quality measure can be summarized as follows:

- Unlike single view images, the perceived quality of a multi-view image may vary across the entire
display area. Human perception is sensitive to such abrupt changes and these changes become significant around structured regions as compared to smooth and highly textured regions.

- The geometric misalignments, blurs, and discontinuities are visible around overlapping areas and seams of intersection.
- Geometric distortions are more noticeable around structured regions and less noticeable around smooth and highly textured regions.

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).

IV. QUALITY ASSESSMENT OF MULTI-CAMERA IMAGES

Multi-camera applications are numerous [1] 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 [15]. 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.

A. Luminance and Contrast Index

This index measures abrupt local changes in luminance and contrast around structured regions. Such changes are common in multi-camera images. Multi-camera 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 observation we employ a measure that is a combination of luminance \( l_{I,J} \) and contrast \( c_{I,J} \) comparison functions from [7] adjusted to give higher weights for structured regions. \( l_{I,J} \) is the luminance comparison function of \( I \) and \( J \) computed on each macroblock. The matrix \( l_{I,J} \) of all macroblocks is calculated as follows:

\[
l_{I,J} = \frac{2\mu_I\mu_J + C_1}{\mu_I^2 + \mu_J^2 + C_1},
\]

where \( c_{I,J} \) is the contrast comparison function of \( I \) and \( J \) computed on each macroblock. The matrix \( c_{I,J} \) of all macroblocks is calculated as
\[ c_{I,J} = \frac{2\sigma_I \sigma_J + C_2}{\sigma_I^2 + \sigma_J^2 + C_2}, \]  

(2)

where \( I \) is the original image and \( J \) is the distorted image. \( \mu_I \) is the mean intensity of image \( I \), and \( \sigma_I \) is the standard deviation of the intensity values of \( I \). The mean and standard deviation are all calculated on macroblock levels. \( C_1 \) and \( C_2 \) are constants included to avoid instability when the denominator is close to zero.

We derive the combined luminance and contrast function for each macroblock \([m,n]\) as

\[ k_{I,J}[m,n] = l_{I,J}[m,n]c_{I,J}[m,n] \]

\[ = \frac{2\mu_I \mu_J + C_1}{\sigma_I^2 + \sigma_J^2 + C_2} \]

\[ = \frac{4(\sigma_I \sigma_J)(\mu_I \mu_J) + 2C_2(\mu_I \mu_J) + C_1}{(\sigma_I^2 + \sigma_J^2)(\mu_I^2 + \mu_J^2) + C_1 C_2}. \]

(3)

By choosing \( C_1 \) and \( C_2 \) to be small enough the combined index at \([m,n]\) can be approximated by

\[ k_{I,J}[m,n] \approx \frac{4(\sigma_I \sigma_J)(\mu_I \mu_J) + C}{(\sigma_I^2 + \sigma_J^2)(\mu_I^2 + \mu_J^2) + C}. \]

(4)

where \( I \) is the reference image, \( J \) is the distorted image, \( \mu \) is the mean intensity and \( \sigma \) is the standard deviation. Both \( \sigma \) and \( \mu \) are computed on macroblocks of dimension \( s \times s \) and \([m,n]\) in \( k_{I,J} \) is the mapping of the upper left corner of the macroblock in \( I \) whose coordinates are \([1 + ms, 1 + ns]\). \( C \) is a constant to avoid instability when the denominator is close to zero.

To calculate the overall index across the image, we will adopt a texture structure model to detect the structured, smooth and randomly textured regions of an image. Regions with structured texture and a region with random texture can be distinguished based on the distribution of edge pixels in the region. A randomly-textured region is composed of small edges with random orientations on the other hand a region with structured texture is composed of long edges with consistent orientations. An edge based texture model was proposed in [16] for visual distortion sensitivity in video bit allocation algorithm. Based on edge based texture model in bit allocation we derive a visual sensitivity model for multi-camera images.

First the texture randomness index at macroblock \([m,n]\) of the image \( I \) is computed by

\[ R_I[m,n] = \mu_I[m,n] \mu_B[m,n], \]

(5)

where \( B \) is an edge intensity binary image with 1’s where the function finds edges in \( I \) and 0’s elsewhere, \( \mu_B[m,n] \) is the mean edge intensity for macroblock of dimension \( s \times s \) at location \([1 + ms, 1 + ns]\) of \( B \) and \( \mu_I[m,n] \) is the mean intensity value of macroblock of dimension \( s \times s \) at location \([1 + ms, 1 + ns]\)
Fig. 5: Texture Randomness Index: (a) Face (Before Mapping) (b) Water Front (Before Mapping) (c) Face (After Mapping) (d) Water Front (After Mapping).

of image $I^1$. If we look at two examples of the texture randomness index values in Fig. 5a and Fig. 5b, the index value is large in randomly-textured regions but small in structured regions.

$^1$Location refers to the upper left corner of the macroblock
The texture randomness index is then mapped using the following mapping function:

\[ T_I[m,n] = \begin{cases} 
\alpha_1 + (0.5 \times \alpha_1 \times \frac{\log_2 R_I[m,n]}{\log_2 \beta_1}) & \beta_1 \leq R_I[m,n] < \beta_2 \\
\alpha_2 + (0.5 \times \alpha_2 \times 2^{-(R_I[m,n] - \beta_2)}) & R_I[m,n] \geq \beta_2 \\
\alpha_1 & \text{otherwise}
\end{cases} \] (6)

where \( \alpha_1 \) and \( \alpha_2 \) are constant parameters chosen to control the weights assigned to the structured regions and randomly textured regions, respectively. By setting \( \alpha_1 > \alpha_2 \), higher weights are assigned to the structured regions. Parameters \( \beta_1 \) and \( \beta_2 \) are the edge detector thresholds. The human visual system is less sensitive to intensity variations in randomly textured region corresponding to values greater than \( \beta_2 \) in \( R_I[m,n] \). \( T_I[m,n] \) is designed to drop quickly around this region and to increase exponentially around structured regions corresponding to the values between \( \beta_1 \) and \( \beta_2 \). Low textured or smooth regions where \( R_I[m,n] \) is less than \( \beta_1 \) are assigned a constant value. A plot of the mapping in (6) is shown in Fig. 6. The index maps for the two examples of Fig. 5 after mapping are shown in Fig. 5c and Fig. 5d.

The combined luminance and contrast index \( LC_{I,J} \) for \( M \times N \) total macroblocks is the average weighted average of \( k_{I,J} \) values where the weights are mapped texture randomness index values of (6) as follows:

\[ LC_{I,J} = \frac{\sum_{m=0}^{M} \sum_{n=0}^{N} k_{I,J}[m,n]T_I[m,n]}{\sum_{m=0}^{M} \sum_{n=0}^{N} T_I[m,n]} \] (7)
Fig. 7: Index maps: (a) Distorted multi-view image, (b) Luminance and contrast index map, (c) Motion index map, and (d) Edge-based structural index map.

$L_{C_{1,3}}$ values range between 1 for minimum distortion and 0 for maximum distortion.

The correlation between the three index measures and image properties is demonstrated in the example of Fig. 7. The image shown in Fig. 7a is a mosaic image of two-views. Prior to compositing, perspective distortion was applied to the left view while the right view was blurred. The luminance and contrast index map is shown in Fig. 7b. The darker regions refer to areas of higher distortions. From Fig. 7b, the right half of the image corresponding to the blurred view has darker regions than ones 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.

B. Spatial Motion Index

Geometric distortions in multi-view images are the result of displacements or shifts at the pixel locations with respect to the reference image. In 2D these displacements are comparable to spatial motion of single-view videos. Hence, a motion model can be used to quantify geometric distortions. We will use motion vectors to compute the pixel displacements relative to the reference image. First the motion vector \( \mathbf{v} = [v_m, v_n] \) at a macroblock location \([1 + ms, 1 + ns]\) of the distorted image \( J \) relative to the reference image \( I \) is computed over a search area of \( p \times p \). The values of displacement are then normalized leading to the relative motion inductor at \([m, n]\) is computed as

\[
\eta[m, n] = \frac{\sqrt{v_m^2 + v_n^2}}{\sqrt{2}p^2}.
\] (8)

Photometric distortions can cause changes in intensity values which can also lead to non-zero motion inductor values. The motion inductor values resulting from a photometric distortion are random and spatially inconsistent. Geometric distortions, on the other hand, have spatial consistency in the directions of the motion vectors. The latter is due to the fact that the pixel displacements due to rotations, translations, and scaling occur in one consistent direction or orientation. The motion index shall be designed to be higher at regions of random displacements caused by photometric distortions and to be lower at regions of coherent displacements corresponding to geometric distortions. To measure magnitude of randomness in these displacements, we use the entropy of the motion inductor values. The entropy is calculated over the probability distribution function, \( p(\eta_i) \), generated from the motion induction values within a spatial window of \( w \times w \) macroblocks. The entropy values are low for regions with coherent displacements and high for regions with random displacements, and hence can be used to suppress the effect of motion inductor values resulting from photometric distortions. The entropy \( \varepsilon[m, n] \) of \( \eta[m, n] \) values at \([m, n]\) is calculated within a spatial window of \( w \times w \) macroblocks for \( w >> p \) as

\[
\varepsilon(m, n) = -\sum_{i=0}^{L} p(\eta_i) \log_2(p(\eta_i)),
\] (9)

where \( L \) is the number of distinct inductor values. We then multiply the relative motion inductor at each macroblock, \( \eta[m, n] \), with the entropy calculated at the very same macroblock. We will call the new product \( \varsigma[m, n] \) the motion consistency index calculated as
Fig. 8: Gradient of image in Fig. 7. The horizontal and vertical axis refer to the horizontal and vertical dimensions of the image.

\[ \varsigma[m, n] = \varepsilon[m, n]\eta[m, n]. \]  

The spatial motion index map of (8) is shown in Fig. 7c. The darker values over the left half indicate the spatial displacements due to the geometric errors. The perceptual distortion that are due to these geometric errors is presented in form of visible misalignments, discontinuities and blur around overlapping areas and across the seam of intersections. To account for this observation we calculate the gradient of motion inductor values smoothed using a low pass Gaussian filter. This can be achieved by calculating the gradient of the relative motion inductor values. The filter coefficients can be calculated as follows:

\[ \lambda[m, n] = \frac{\nabla \eta[m, n]g[m, n]}{\theta}, \]  

where \( g(x, y) \) is a Gaussian low pass filter and

\[ \nabla \eta = \sqrt{\left(\frac{\partial \eta}{\partial m}\right)^2 + \left(\frac{\partial \eta}{\partial n}\right)^2}, \]

\[ \theta = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \nabla \eta[m, n]g[m, n]. \]

The above function assigns higher coefficients across the seam of intersections and the overlap areas as show in Fig. 8. The spatial motion index is then computed for an image of \( M \times N \) total macroblocks.
as follows:

\[ S_{I,J} = \frac{1}{MN} \sum_{m=0}^{M} \sum_{n=0}^{N} |1 - \frac{\varsigma[m,n] \lambda[m,n]}{\kappa}| \]

\[ \kappa = \arg \max_{\varsigma[m,n] \lambda[m,n]} \{ m, n | 0 \leq m \leq M, 0 \leq n \leq N \} \] (13)

\( S_{I,J} \) values range between 1 for minimum distortion and 0 for maximum distortion.

C. Edge-based structural Index

The two indices presented so far capture the distortions in terms of changes in contrast and luminance and pixel displacements in an image. Photometric and geometric distortions might cause loss in structural information. Such information includes degradation in texture quality or lost image components on intersection or overlapping areas. Evaluating the structural similarity over edge maps instead of the actual images leads to better correlation with subjective quality for SSIM [17]. 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. To calculate the edge-based structural index we reuse the mapped texture randomness index. For \( M \times N \) total macroblocks the index is computed as follows:

\[ E_{I,J} = \frac{1}{MN} \sum_{m=0}^{M} \sum_{n=0}^{N} \left( 1 - \frac{(T_I[m,n] - T_J[m,n])}{T_I[m,n]} \right), \] (14)

where \( T_I[m,n] \) and \( T_J[m,n] \) are defined as in (6) 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. 7d that the structural losses represented by the edge-based structural index are mainly concentrated on the blurred view at the right; notice that the majority of the pixels are gray indicating structural losses. The figure also shows some scattered dark pixels on the left side. These pixels are caused by the geometric distortions. 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).
D. 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 summarizes the visual distortions in multi-view images. In this paper, $MIQM$ is computed as the multiplication of the aforementioned three index measures:

$$MIQM_{I,J} = LCI_{I,J}S_{I,J}E_{I,J},$$

(15)

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

V. SIMULATION RESULTS

Peak Signal-to Noise Ratio (PSNR) is the most widely used objective metric due to its low complexity and clear physical meaning. It quantifies the image quality by measuring the error in intensity between two different images. $SSIM$ proposed by Wang et al. [7] is based on the assumption that the human visual system is highly adapted to extract structural information from the viewing field. $SSIM$ is defined as follows:

$$SSIM_{I,J} = l_{I,J}^\alpha c_{I,J}^\beta s_{I,J}^\gamma,$$

(16)

where $l_{I,J}$ and $c_{I,J}$ are defined in (1) and (2) respectively. $s_{I,J}$ is the structure comparison function of $I$ and $J$ computed on each macroblock. The matrix $s_{I,J}$ of all macroblocks is calculated as follows:

$$s_{I,J} = \frac{\sigma_{IJ} + C_3}{\sigma_I \sigma_J + C_3},$$

(17)

where $\sigma_{IJ}$ is the correlation coefficient between $I$ and $J$. The correlation is calculated on macroblock level. $C_3$ is a constant included to avoid instability when the denominator is close to zero. $\alpha$, $\beta$ and $\gamma$ are three positive parameters used to adjust the relative importance of the three components. The overall image quality is calculated as the mean of all $SSIM$ values and it is referred to as the mean $SSIM$ ($MSSIM$)$^1$.

Looking into the results shown in Fig. 9, image Fig. 9a and image Fig. 9b are the original undistorted images. Both images Fig. 9c and Fig. 9d suffer from geometric distortions, however distortion in image Fig. 9d is hardly noticeable compared to the distortion in image Fig. 9c. When looking into the $MSSIM$ values we notice that $MSSIM$ for image Fig. 9c is much higher than image Fig. 9d which contradicts with the actual perceived quality. Similarly when comparing image Fig. 9d to image Fig. 9f we notice

$^1$For the rest of the paper we will use either $MSSIM$ or $SSIM$ but both refer to $SSIM$. The term $MSSIM$ is usually used in result evaluations to stress the fact that it is the mean $SSIM$. 

August 31, 2011 DRAFT
that Fig. 9f has higher PSNR and MSSIM values implying image Fig. 9f has better quality when in fact image Fig. 9f subjectively looks more distorted than Fig. 9d. The same applies for PSNR values when comparing image Fig. 9e and image Fig. 9c. These examples show that objective values by quality measures such as MSSIM and PSNR designed to capture the quality of single-view images contradict with the actual perceived quality of multi-camera images.

In order to test the performance of MIQM, we conducted an extensive subjective quality assessment study. First we produced a database of multi-camera images generated using the techniques described earlier in the paper where various combinations of geometric and photometric distortions were applied. For our tests we then prepared a similar setup for subjective testing as in [18]. In these experiments, 22 human subjects were asked to assign each image with a score indicating their assessment of the quality of that image. The subjects were not screened for color blindness or vision problems, and their verbal expression of the soundness of their (corrected) vision was considered sufficient. Most of the participants were young male students of engineering background but they had no previous experience of multi-camera images. Their opinions on multi-camera image quality may differ from those of people accustomed to this technology. We defined quality as the extent to which the distortions were visible and annoying. In this experiment, a total of 64 images, out of which 7 were the reference images, were evaluated by student volunteers, and the raw scores for each subject were processed to give Mean Opinion scores (MOS) and a Difference Mean Opinion Score (DMOS) for each distorted image. The test images had varying types and levels of distortions.

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 $w = 9$, and the constants are $C = 2.5$, $\alpha_1 = 128$, $\alpha_2 = 64$, $\beta_1 = 10$, 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 (5) and (14). Canny’s method is less sensitive to noise, and more likely to detect true weak edges.

In the plots of Fig. 10 and Fig. 11 and in the results of Table II the DMOS scores obtained from the subjective experiments are compared against the multi-view image quality measure (MIQM), the peak signal-to-noise ratio (PSNR), the mean structural similarity (MSSIM), and the visual information fidelity (VIF). MIQM, PSNR, and MSSIM are as previously defined in this paper. However, VIF is a full-reference

\[1\] We did not account for the exact the female/male distribution of our participants, as gender difference plays no role in quality of vision.
Fig. 9: \textit{PSNR} and \textit{MSSIM} values for images with various distortion types.
Fig. 10: Scatter plots for the four objective quality criteria: PSNR, MSSIM, VIF, and MIQM. The Image Quality Ratings were all scaled to the MOS range $[0, 5]$ for comparison. The lines in red indicate the outliers’ boundary and line in blue (middle) indicate the ideal image quality rating. A point is considered an outlier if the distance from the ideal is greater than twice the DMOS standard deviation [20]. In our results the standard deviation of the DMOS values was: $\delta_{DMOS} = 0.8424$. 
image quality introduced in [21] that quantifies the mutual information that is present in the reference image and how much of this reference information can be extracted from the distorted image. \textit{VIF} has shown to perform better than \textit{MSSIM} and \textit{PSNR} for single view images [21].

The scatter plots of \textit{DMOS} versus the image quality ratings for four objective quality measures (\textit{PSNR}, \textit{MSSIM}, \textit{VIF} and \textit{MIQM}) are shown in Fig. 10. The lines in red indicate the outliers’ boundary and line in blue (middle) indicate the ideal image quality rating. A point is considered an outlier if the distance from the ideal is greater than twice the \textit{DMOS} standard deviation [20]. The plots show that the number of outlier points for \textit{MIQM} is much less than those of \textit{PSNR}, \textit{MSSIM}, and \textit{VIF}. The plots in Fig. 10 also show that the points outside the outlier points of \textit{MIQM} are very close to the boundaries and they all fall within a half \textit{DMOS} standard deviation. The percentage of outlier points in a quality measure is an indicator for consistency. The results are a proof that \textit{MIQM} ratings have less outlier points and hence are significantly more consistent than the other quality measures.

Table II shows the validations scores for the objective quality measures. Following the \textit{VQEG} recommendations in [20], the validation scores that are used in this paper are the root mean squared error (RMSE), the Pearson linear correlation coefficient (CC), the Spearman rank order correlation coefficient (ROCC), the mean absolute error (MAE), and the Outlier Ratio (OR). These validation scores express the relationships between each quality measure and the subjective ratings. A higher CC and ROCC values mean an increased coherency for the objective quality measure predictions. ROCC is also a metric used evaluate the monotonicity of the objective quality measure predictions. The RMSE and MAE on the other hand are measures of accuracy of the predictions, where lower RMSE and MAE values mean a

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE</th>
<th>CC</th>
<th>ROCC</th>
<th>MAE</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>1.1249</td>
<td>0.2746</td>
<td>0.2147</td>
<td>0.8680</td>
<td>0.1475</td>
</tr>
<tr>
<td>MSSIM</td>
<td>0.9438</td>
<td>0.9487</td>
<td>0.5612</td>
<td>0.8749</td>
<td>0.2459</td>
</tr>
<tr>
<td>VIF</td>
<td>1.6718</td>
<td>0.5298</td>
<td>0.4034</td>
<td>1.3130</td>
<td>0.2951</td>
</tr>
<tr>
<td>MIQM</td>
<td>0.7014</td>
<td>0.9506</td>
<td>0.6671</td>
<td>0.6643</td>
<td>0.0819</td>
</tr>
</tbody>
</table>

TABLE II: Validation scores for different quality assessment methods. The methods tested were \textit{PSNR}, \textit{MSSIM}, \textit{VIF}, and \textit{MIQM}. The methods were tested against \textit{DMOS} from the subjective study after a fitting into non-linear regression. The validation criteria are: root mean squared error (RMSE), Pearson linear correlation coefficient (CC), Spearman rank order correlation coefficient (ROCC), mean absolute error (MAE), and Outlier Ratio (OR).
The image quality measure results after fitting the results into a non-linear regression function are shown in Fig. 11. The resulting curves show that MIQM is the closest fit to the ideal quality rating. The Image Quality Ratings were all scaled to the MOS range [0, 5] for comparison.

More accurate predictions. OR is a measure of consistency and values closer to zero indicate better consistency in the quality measure predictions. The validations scores were calculated after fitting the results into nonlinear regression function from [20]:

$$DMOS_p = B_1/(1 + \exp(-B_2 \times (IQR - B_3))) .$$

Where IQR is the image quality rating obtained using the objective quality measures and DMOS_p is the resulting predicted DMOS values. The reason this is done is to remove any nonlinearity due to the subjective rating process and to allow comparison of the quality measure in a common analysis space. The resulting curves after applying the non-linear regression fit are shown in Fig. 11. Looking into the curve we notice that MIQM is the closest fit to the ideal image quality rating represented by the middle 45% line.

The results in the Table II shows that MIQM values has the least RMSE and MAE values among all quality measures. In addition, the RMSE for MIQM is less than the one standard deviation of the DMOS values ($\delta_{DMOS} = 0.8424$) which actually is an indication that MIQM is relatively an accurate prediction of image quality for multi-camera systems. The Pearson linear correlation (CC) and Spearman rank order correlation coefficient (ROCC) values for MIQM also outperforms the three other quality measures. CC
values mean that \textit{MIQM} is more coherent than \textit{VIF}, \textit{PSNR} and \textit{MSSIM}. ROCC values also indicate a significant gain in monotonicity of quality predictions using \textit{MIQM} over the closest quality measure \textit{MSSIM}. The results also show \textit{MIQM} has a significantly lower outlier ratio (OR) and therefore is the most consistent quality measure.

Overall \textit{MIQM} is the most accurate, coherent and consistent among the objective measures represented in this paper for multi-camera images. The results also show that \textit{PSNR} has a lower OR value than \textit{MSSIM} and \textit{VIF}, which may indicate that the \textit{PSNR} is more consistent. \textit{MSSIM} is second to \textit{MIQM} in accuracy (RMSE and MAE) and coherency (CC and ROCC) however it comes at very big disadvantage in terms of consistency (the outlier ratio). \textit{VIF} on the other hand is more coherent than \textit{PSNR}; however it has the least accuracy and consistency. We attribute this randomness in performance to the fact these measures, unlike \textit{MIQM}, were actually designed for single camera images where photometric distortion is spatially coherent and geometric distortions are not significant.

\section*{VI. Conclusion}

In this paper, we have introduced the various multi-camera applications and the different type of distortions affecting each one of them. Then we studied two particular types of distortions that are unique to multi-camera images. We provided examples on how each can influence multi-camera image perceived quality. All examples were taken for a panoramic image application. Then we introduced a multi-camera image quality measure (\textit{MIQM}) as a combination of three index measures. We presented the derivation and reasoning for each index measure. Finally, we compared \textit{MIQM} against a database of multi-camera images. We ran a set of subjective tests to evaluate the quality of the images in the database and the MOS score was calculated for each image. The results and examples show that \textit{MIQM} outperforms \textit{SSIM}, \textit{VIF} and \textit{PSNR} for multi-camera images quality assessment.

\textit{MIQM} was tested and refined for panoramic image applications. However, the measure is designed to capture the visual effects of artifact introduced at the acquisition and pre-compositing processes to predict the composited image quality. Hence, we can build on the findings of this work to develop quality measures for stereoscopic, free viewpoint, and 3DTV applications after taking into consideration stereoscopic impairments and synthetic artifacts. Therefore, we consider MIQM a particular implementation based upon which we will expand these concepts to include other forms of multi-camera presentation.

\textbf{References}


Mashhour Solh is a Ph.D. candidate in electrical and computer engineering (ECE) at Georgia Institute of Technology. He received a degree M.S. degree in ECE from Florida Institute of Technology in 2006, and a B.E. in computer and communications engineering from American University of Beirut in 2005. His research scope includes image and video coding, multimedia quality assessment, multi-camera imaging, multimedia signal processing (image, video and audio), 3DTV techniques, signal processing for infrasound signals and TI's Digital Signal Processors. Mr. Solh is currently the web chair of third international conference on Immersive Telecommunication (IMMERSCOM2011). He has also served as web chair of IMMERSCOM 2007 and IMMERSCOM 2009. He is a member of Video Quality Experts Group (VQEG) standardization efforts. Mr. Solh was also vice president of IEEE student branch of Georgia Institute of Technology for 2009, chair of ECE seminar student committee at Georgia Tech 2007-2009, and a member of Georgia Techs student advisory committee 2009-2010. He is a member of IEEE signal processing society and IEEE communications society. Mr. Solh has earned several awards including travel grant award for the IEEE International Conference on Multimedia and Expo(ICME 2011), and the Franking Antonio Scholarship Award for his outstanding performance during a internship in Qualcomm in summer of 2010. He also earned the Deans Honor List Award several times and the pepsi cola international scholarship while being a student at the American University of Beirut. He has also been awarded the outstanding achievement award for his service as a Florida Tech diplomat.

Ghassan AlRegib received the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology in 2003. He joined the Georgia Tech faculty in 2003 and is currently an associate professor in the School of Electrical and Computer Engineering (ECE). His research group is working on projects related to multimedia processing and communications, immersive communications, collaborative systems, quality of experience, and social media processing.

Dr. AlRegib received the ECE Outstanding Graduate Teaching Award in spring 2001 and both the Center for Signal and Image Processing (CSIP) Research Award and the CSIP Service Award in spring 2003. In 2008, he received the ECE Outstanding Junior Faculty Member Award at Georgia Tech.

Dr. AlRegib was the general co-chair and co-founder of the First International Conference on Immersive Telecommunication (IMMERSCOM) that was held in November 2007. He is currently the steering committee co-chair for IMMERSCOM and was the chair of the Special Sessions Program at the IEEE International Conference on Image Processing (ICIP) in 2006. Dr. AlRegib has also served as an associate editor of the IEEE Signal Processing Magazine, and in 2008, he became the area editor for the IEEE Signal Processing Magazine and the editor-in-chief of the ICST Transactions on Immersive Communications. Dr. AlRegib has served as a session chair and technical program committee member for several international conferences and workshops. He has authored over 60 journal and conference technical papers and has been issued four U.S. patents. He has conducted several consulting jobs for several companies and organizations. Dr. AlRegib is a senior member of IEEE.