

Characterization of Image Distortions in Multi-Camera Systems

Mashhour Solh
School of Electrical and Computer Engineering
Georgia Institute of Technology
msolh@ece.gatech.edu

Ghassan AlRegib
School of Electrical and Computer Engineering
Georgia Institute of Technology
gregib@ece.gatech.edu

ABSTRACT

Multi-camera systems have been evolving as next generation video cameras with applications including 3D reconstruction, image-based rendering, free viewpoint, and 3D TV. Quantifying visual quality of multi-camera systems is fundamental in developing these applications. In this paper, image distortion types in multi-camera systems are investigated where distortion is classified as either *geometric* or *photometric*. Examples and measurements are presented showing that single-view objective image quality measures are not adequate for perceptual assessment of multi-camera images. A new algorithm that characterizes the type of distortion in a given image captured by a multi-camera system is proposed and evaluated. The new algorithm is based on the edge intensity summation (*EIS*). A new *EIS*-based structural similarity (*EISSM*) quality measure is proposed. *EISSM* is shown to capture the perceptual fidelity that is not fully grasped by PSNR and SSIM.

Keywords

Image quality, Multi-Camera system, Geometric distortion, Photometric distortion, Structural similarity

1. INTRODUCTION

With the rapid improvement in electronic and computing technologies and dropping costs of cameras, multi-camera systems have been evolving as next generation video cameras. The goal of these systems is to expand users experience beyond what can be offered by a single camera. The multi-view video includes video sequences captured by several cameras simultaneously but at different locations or from different angles [1]. Applications of multi-camera systems include but are not limited to panoramic videos, free-viewpoint video, 3D TV, virtual view synthesis, object tracking, and stereoscopic video [1] [2]. In panoramic videos the image plane is expanded to cover larger areas and other kinds of planes such as cylindrical and spherical. Free-viewpoint video allows the user to navigate through the

image by choosing his/her own viewpoint. For example, in 3DTV a scene is captured using a multiple camera. In this application, 3D scene reconstruction gives the user the impression of a 3D view. Similarly, stereoscopic video produces a 3D impression by providing two views, one for each eye. These applications and others can be useful in many fields including surveillance, sightseeing, advertisement, distance learning, medical training, and entertainment [1] [3].

Challenges in multi-camera applications are numerous where the processing chain may include image capturing, camera calibration, scene presentation, coding, transmission, rendering, and display. Although each of these processes is a challenge by itself, all of them require some type of image quality measurement to provide a feedback to the designer or the user [2]. For example, if video compression algorithms are designed for such video sequences, a quality measure is needed to compare these algorithms. Similarly, an image quality measure may prove valuable for accurate calibration of the multi-camera systems for both camera position and color properties.

Unlike images captured by single cameras where the distortion can be homogeneous across the image plane, images generated by multiple cameras may not only differ in content but also in quality of the captured images. Moreover, the quality of the rendered views is influenced by the camera configuration, the number of cameras, and the calibration process [2]. A special feature of images captured by multi-camera systems is that visual distortion may result from geometric corrections or calibrations. In this paper, distortion types in multi-camera systems are characterized and compared. Examples and measurement are presented showing that single-view objective image quality measures are not adequate for perceptual assessment of images produced by multi-camera systems. Moreover, a new algorithm that characterizes the type of distortion in a given image captured by a multi-camera system is proposed and evaluated. To achieve this goal, an edge intensity summation (*EIS*) is defined and a new *EIS*-based structural similarity (*EISSM*) quality measure is proposed. *EISSM* is shown to capture the perceptual fidelity that is not fully grasped by PSNR and SSIM.

The rest of the paper is organized as follows. Section 2 provides an overview of prior work on image quality assessment for single and multi-camera systems. Distortion types in multi-camera systems are classified in Section 3. The per-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or to publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Immerscom 09, May 27-29, 2009, Berkeley, California, USA.
Copyright C 2009 ICST ISBN 978-963-9799-39-4

formance of some contemporary objective quality measures (PSNR and SSIM) is evaluated in section 4. Then we propose a technique to detect distortions of multi-camera using the edge intensity summation (*EIS*) in section 5. In section 6 we propose a *EIS*-based structural similarity to improve SSIM performance over multi-camera distortion types. Finally, we conclude our work in Section 7.

2. PRIOR ART

Degradation of visual quality of images may occur during acquisition, processing, compression, and transmission. Video and image processing algorithms are evaluated using objective metrics or through subjective testing in a controlled environment. The best method of quantifying perceptual image quality is subjective evaluation. Subjective quality assessment is expensive, tedious and not applicable in environments, which require real-time processing. Objective image quality on the other hand automatically predict the perceived image quality and are more desirable [4].

Quantifying the visual fidelity of a distorted image has primarily been based either on the properties of the images or on specific information the human visual system (HVS) attempts to extract when viewing an image. A well-known example of the former is the mean-squared error (MSE), which is known as the peak signal-to-noise ratio (PSNR). PSNR is solely based on the differences in intensity. It is evident that PSNR is not the most accurate metric for image quality [4]. On the other hand, HVS-based metrics employ a frequency-based decomposition, which take into account the detectable visual thresholds of distortions[5]. Other metrics quantify visual fidelity based on the structural content such as object boundaries and regions of high entropy. Some of the HVS-based metrics include Sarnoff just noticeable differences (JND) [6], Watson's digital video quality (DVQ)[5] and Wang et al's Structural Similarity (SSIM)[7]. SSIM index computes the mean, variance and covariance of small blocks inside an image. SSIM assumes that the human visual system is highly adapted to extract structural information from the viewing field [7]. In [8] an edge based-structural similarity was proposed to improve performance of SSIM over highly blurred images.

User experience and image quality are dependent on human perception in different environmental conditions and multimedia device capabilities. One such environment is multi-camera system. In stereoscopic 3D video systems humans perceive good quality 3D video if one of the eyes is experiencing a high quality video [9]. The authors in [10] assumed that PSNR of the second view is less important for 3D visual experience and proposed a new quantitative measure for stereo video quality as weighted combination of two PSNR values and a jerkiness measure. In [11] the authors addressed the need to decide which camera or subset of available cameras to use in a given instant of time in surveillance applications. They introduced new measure named Quality-Of-View (QOV) that takes into account view angle and distance from subjects to evaluate each camera view performance. The authors in [12] proposed an objective quality measure for 3D free-view point video quality. The measure attempts to quantify visual artefacts arising from inaccurate sampling of appearance in capturing system or through an inexact 3D scene representation. The

quality measure adopts Hausdorff distance to quantify the mis-registration between two images.

The fore mentioned quality metrics of multi-camera systems are application specific and tend to measure only one related visual aspect of multi-camera image visual quality. Therefore, a quality measure is required that can quantify all multi-camera image distortion types. This measure has to account for perceptual fidelity in such images beside taking into account how they were acquired and displayed.

Photometric and geometric distortions are common in multi-camera systems. Photometric distortions are the variations in brightness levels and color gamut across the entire displayed image. The source of this variation can be the camera system or the post processing applications. Geometric distortions are the misalignments and discontinuities due to geometric errors. The source of these errors can be the camera system internal properties, positioning or post processing applications. We will discuss these distortions in detail next.

3. DISTORTION TYPES IN MULTI-CAMERA SYSTEMS

In this paper, the distortion types in multi-camera systems are classified into *photometric* or *geometric* distortion. Each of these types have a different impact on the overall image quality. The overall image in the paper is defined as the full mosaiced image or a set of images projected independently. As a first step toward deriving an effective image quality measure for multi-camera systems, one has to characterize each of these distortion types and better understand their nature and impact.

3.1 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 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 improvements or developments of new video or image applications and hence have motivated the development of contemporary image and video quality metrics.

In *multi-camera* systems individual cameras looking at different parts of the scene may experience different types and levels of distortion. These variations may be intrinsic due to the differences among individual cameras or extrinsic due to different level of coding or post-processing that each camera-video stream may experience. Human perception is sensitive to abrupt local changes in images. This type of distortion will be referred to as the *variational photometric distortion*. This distortion is especially obvious around overlapping and content rich areas of the captured images.

In order to explain the effect of photometric distortion on an image captured by a multi-camera system, simulation is used. A single digital camera was used to capture high-resolution images. The captured image was split into multiple sub-images. Targeted distortion was then applied on

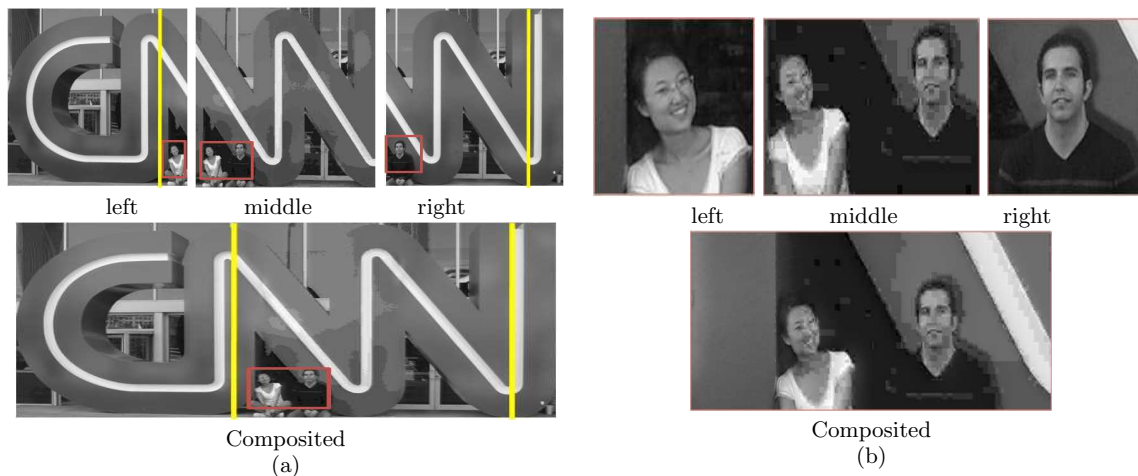


Figure 1: (a) Top: original images before compositing. The middle image is JPEG compressed with $Q = 5$. Bottom: the resulting image after compositing. The yellow markers show the borders of the overlapping areas. (b) Close up views of the regions in red boxes in the images in (a).

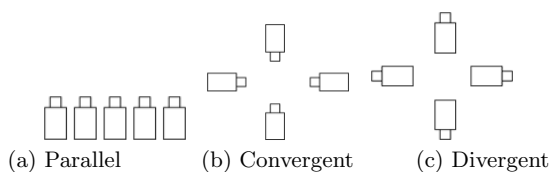


Figure 2: Examples of spatial camera configuration for multi-camera systems.

each sub-image independently. The effect of the different types of distortion on a multi-camera application are demonstrated by an image mosaic algorithm that applies a multi-resolution spline [13]. An example of photometric distortion in multi-camera applications is shown in figure 1. Figure 1(a) demonstrates a possible variational photometric distortion in a three-camera with different levels of JPEG distortions. The original image was split into three sub-images: left, middle, and right. The overlap regions are as shown in the figure between yellow markers. The left and right image are kept unchanged. The middle is distorted by applying JPEG compression with $Q = 5$. The three images (left, JPEG-middle, and right) were mosaiced together using the multi-resolution spline algorithm. The resulting composited image is shown at the bottom Figure 1(a). The images shown in Figure 1(b) are close-ups that show that the variation in the image quality is reflected into perceptually noticeable abrupt change in the composited image. Abrupt changes in image quality are perceptually annoying and can affect negatively our judgement of the overall quality of the image especially at areas with important details such as the human face in Figure 1(b) at the bottom.

3.2 Geometric Distortion

In multi-camera systems a scene is captured by N cameras where each individual camera's position and orientation can vary depending on the application. In Figure 2, some of the possible camera configurations are shown. In multi-camera

applications and reconstructions, the knowledge of the camera calibration and scene geometry is important. However, changes in a camera position or orientation as well as errors in estimating the camera parameters or the scene geometry can lead to geometric distortions. Geometric distortion in multi-camera system is defined as structural disparity such as discontinuity and misalignment in the observed image due to geometric errors. Geometric errors can occur during the mapping, which may include rotation and translation. This type of distortion is going to be referred to as *planar distortion*. Geometric distortion can also occur in the mapping from the 3D world to the 2D plane of the image. This type of distortion is going to be referred to as *perspective distortion*. Figure 3 shows examples that illustrate the types of geometric errors as well as the original image. The image in Figure 3(b) is subject to perspective distortion which can be seen at the white buildings in the back. The buildings look smaller and more to the right side than the original image in 3(a). The image in Figure 3(c) is rotated clockwise by 10 degrees. In multi-camera systems such errors can also occur when mapping a single camera plane to another reference camera in the system.

To simulate the geometric distortions in multi-camera system geometric errors were applied to the generated views and then composited into a single image mosaic. Figure 4 is a close-up view that demonstrates perspective and planar distortions in a three-camera system. The perspective and planar distortions in Figure 4 are visible on the right eye in the top image and the thumb in the bottom image. We notice the disparity due to perspective distortions is represented by distortion in the face (top) and a shrinkage on the thumb (bottom). On the other hand planar distortion results in an misalignment seen all over the face and the hand area. Hence, geometric distortions in single camera translates to misalignment and discontinuities in the multi-camera image. Unlike photometric distortions where distortions translate as abrupt changes occur across the whole image, geometric distortions are percep-



Figure 3: Geometric distortions in single cameras: (a) No distortion (b) Perspective (c) Planar(rotation)

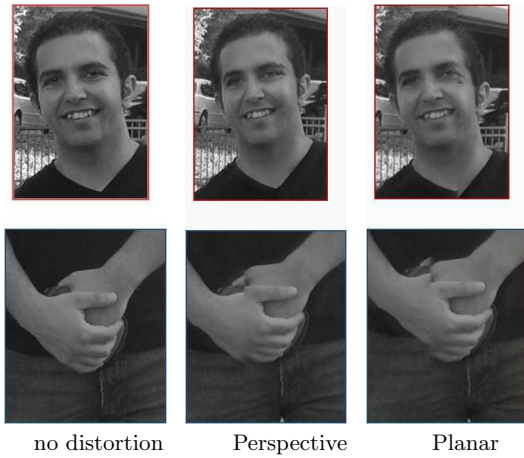


Figure 4: Geometric distortions in a multi-camera system.

tually attracting around connecting edges and overlapping areas. Next we will study the performance of two objective single view quality metrics over multi-camera images.

4. PSNR AND SSIM PERFORMANCE WITH MULTI-CAMERA IMAGES

In the previous section we investigated the types and properties of distortions in multi-camera systems. In this section we will investigate if the two objective *PSNR* and *SSIM* can quantify the perceptual quality of these distortions. Peak signal-to noise ratio (*PSNR*) is the most widely used objective metric due to its low complexity and clear physical meaning. It quantify the image quality by measuring the error in intensity between two different images. *SSIM* proposed by Wang et al [7] and has attracted a lot of attention in literature. *SSIM* is based on the assumption that HVS is highly adapted to extract structural information from the view. *SSIM* is defined as:

$$SSIM(i, j) = f(l(i, j), c(i, j), s(i, j)) \quad (1)$$

where $l(i, j)$ is the luminance comparison, $c(i, j)$ is the contrast comparison, $s(i, j)$ is the structure comparison, i is the reference image and j is a distorted image.

Looking into the results shown in Figure 5, image (b) has lower *SSIM* and *PSNR* than images (c) and (d). However, the visual quality of image (b) is very similar to original image (a), unlike images (c) and (d) where the perceptual quality is obviously different than (a). Also we notice the in both images (c) and (d) the distortion level is not uniform where the left and right side of the image experience two different levels of distortion. Similarity, we applied these measures to several images in our database. In all these experiments we concluded that both *SSIM* and *PSNR* do not adequately capture the perceptual fidelity of the multi-camera system's image. In next section we will present a new technique that allows us to detect the distortion type in a multi-camera image.

5. DETECTION OF DISTORTION TYPES

5.1 EIS - Edge Intensity Summation

In photometric distortion a captured image is subject to intensity errors at various pixel locations. These errors can result in erosion of image details. A fair estimate of these erosions are the spatial edges. In [14] the edge information is described by the locations of variations of intensity values and the related intensity values at these locations. When an image is blurred the locations of the intensity variations are preserved however the intensity values of these variations are reduced. Also photometric distortions such as noise and quantization introduce new intensity variations without changing the location of the original edge information. In contrary to photometric distortions, in geometric distortions the intensity values do not change only the pixel locations do. As a result, the location of the spatial edges are changed but the relative intensity variations in the image are preserved. Considering this observation, we propose using the edge intensity summation to predict the type distortion in a camera system. Edge information of an image can be derived by applying edge detection filters on the luminance component of the image. The edge intensity summation (EIS) of an $M \times N$ image is defined as:

$$E = edge(I) \quad (2)$$

$$EIS = \sum_{i=1}^M \sum_{j=1}^N E(i, j) \quad (3)$$

where, E is an edge intensity binary image of the same size as I with 1's where the function finds edges in I and 0's

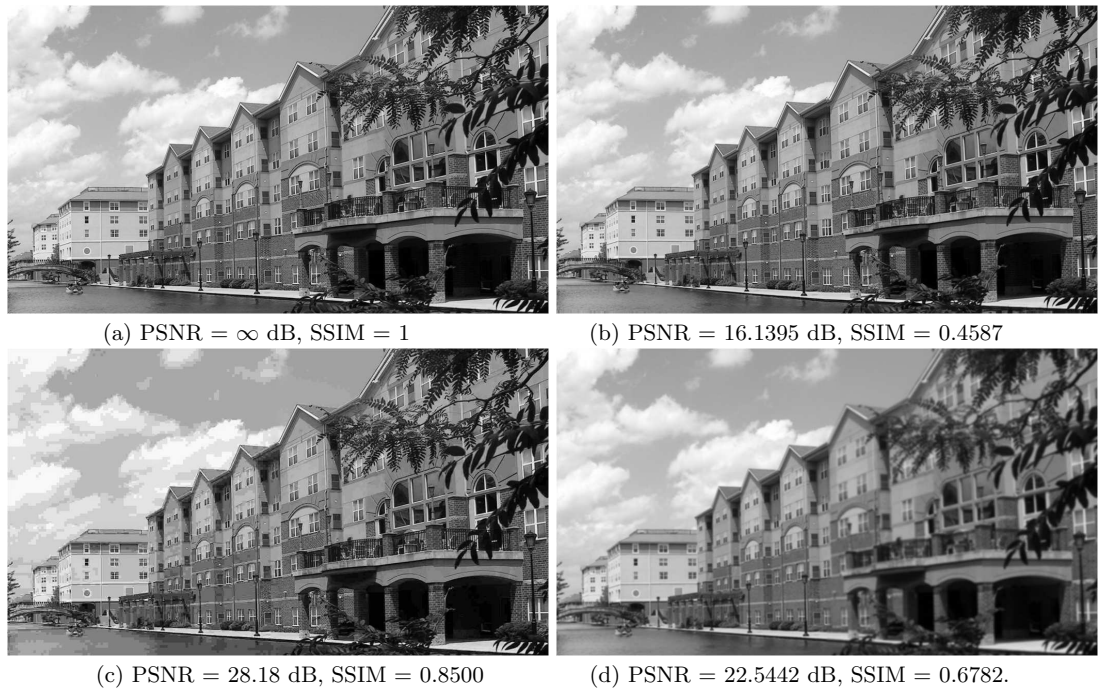


Figure 5: Reconstructed Images:(a) No distortion (b) Perspective distortion (c) JPEG distortions (d) Blurred.

elsewhere. The edge intensities are computed on 32×32 non-overlapping blocks. In this paper the Canny edge detection method was used [15]. Canny's method is less likely than the others to be fooled by noise, and more likely to detect true weak edges. Hence, it is a proper method to compute the edge intensity. The results of edge intensity summation for seven different images are shown in Table 1. For each image eight different instances of distortion were applied. The EIS for the original image, and the distorted images were calculated. The variations in the distorted image EIS relative to the original image were also computed.

5.2 Using EIS to Identify Source of Distortion

At first look into results, we notice that for geometric distortions there is a very slight loss in the $|EIS|$ value (3% to 7%) and these results reflects the amount of information that might be lost around the borders of the image in planar and perspective distortions. On the other hand we notice that the $|EIS|$ value decreases in the order of (24% to 50%) in a blurred image and increases in the order of (14% to 209%) for JPEG distortions.

By setting a threshold of 10% to the changes in EIS values we can predict the source of distortion with high level of accuracy. In this we assumed that there is only one type of distortion at each instant. However, when an image is experiencing more than one type of distortion the changes in EIS reflect the most dominant type distortion. For instance when an image experiences a geometric distortion and blur at the same time the EIS values can be the blur. The choice of the block size was motivated by the processing speed. Table 2 shows the variations in EIS value for an 8×8 block size. The results are similar to 32×32 block size case but

Table 3: EISSM values for examples of Figure 5. Subjective evaluation of these images show that (b) has the best quality and (d) the worst.

Image	PSNR	SSIM	Δ_{EIS}	EISSM
(b)	16.14	0.4587	0.0199	0.8118
(c)	28.18	0.8500	0.2983	0.6721
(d)	22.54	0.6782	0.3853	0.4957

with different range values. Next, we will use EIS to derive an EIS -based structural similarity for multi-camera system.

6. EIS-BASED STRUCTURAL SIMILARITY

In this paper we will use the EIS information to improve $SSIM$ performance for geometric type distortions. $SSIM$ is defined as:

$$SSIM(i, j) = [l(i, j)]^\alpha \cdot [c(i, j)]^\beta \cdot [s(i, j)]^\gamma \quad (4)$$

where $\alpha > 0$, $\beta > 0$ and $\gamma > 0$ are parameters that depends on the relative importance of the three components. To obtain the EIS -based structural similarity ($EISSM$) the structural comparison component is replaced by EIS comparison component. We define ($EISSM$) as:

$$EISSM(x, y) = (l(x, y))^\alpha \cdot (c(x, y))^\beta \cdot (1 - \Delta_{EIS})^\gamma \quad (5)$$

$$\Delta_{EIS} = \left| \frac{EIS(I) - EIS(J)}{EIS(I)} \right| \quad (6)$$

where, I is the undistorted image and J is the distorted version of the image. The $EISSM$ values of the example of Figure 5 are shown in Table 3. The $EISSM$ values in the table better reflect the order perceptual quality of the distorted images with more accurate estimate of the perceptual distortions due to geometric distortions.

Table 1: The percentage change in EIS values for various distortion sources with 32×32 blocks.

Image	Undistorted EIS	Geometric Distortion				Photometric Distortion			
		Planar		Perspective		Blur		JPEG	
		2 degrees	5 degrees	Outward	Inward	$r=6, \sigma=3$	$r=4, \sigma=2$	$Q=5$	$Q=10$
Emily	369724	-4	-6	-5	-3	-24	-19	109	47
Cabinet	188774	-6	-6	-5	-5	-46	-36	140	43
Group	177156	-4	-4	-4	-3	-40	-30	63	14
F18	89479	-7	-7	-5	-5	-51	-41	205	123
Waterfront	102468	-4	-5	-3	-4	-44	-32	96	40
CNN	204093	-5	-5	-4	-7	-41	-34	193	130
CNN2	208509	-6	-6	-4	-5	-42	-34	209	135

Table 2: The percentage change in EIS values for various distortion sources with 8×8 blocks.

Image	Undistorted EIS	Geometric Distortion				Photometric Distortion			
		Planar		Perspective		Blur		JPEG	
		2 degrees	5 degrees	Outward	Inward	$r=6, \sigma=3$	$r=4, \sigma=2$	$Q=5$	$Q=10$
Emily	399674	-4	-5	-4	-5	-8	-9	221	174
Cabinet	190059	-2	-1	-1	-1	-5	-12	289	196
Group	176128	-3	-2	-3	-3	-12	-11	267	191
F18	78457	0	0	-1	-1	-14	-13	316	297
Waterfront	108237	-6	-6	-6	-6	-13	-14	249	204
CNN	180636	-1	-2	-1	-1	-14	-16	310	285
CNN2	183261	-2	-2	-2	-2	-16	-18	311	297

7. CONCLUSION

In this paper we characterized the distortion types in multi-camera systems and classified them into geometric or photometric. Then we show that each distortion type affects the edge information. We proposed using the edge intensity summation (EIS) to detect the type of distortion. Then, we proposed EIS-based structural similarity (EISSM) to improve the performance of *SSIM* to quantify the geometric and photometric distortions. The ongoing investigation aims at using the conclusion from this paper to develop a robust quality measure for multi-camera images.

8. ACKNOWLEDGMENT

This work is funded in part by HP laboratory.

9. REFERENCES

- [1] "ISO/IEC JTC1/SC29/WG11 Applications and Requirements for 3DAV," Tech. Rep. N5877, Trondheim, Norway, July 2003.
- [2] A. Kubota, A. Smolic, M. Magnor, M. Tanimoto, T. Chen, and C. Zhang, "Multiview Imaging and 3DTV," *IEEE Signal Proc. Magazine*, vol. 24, no. 6, pp. 10–21, Nov 2007.
- [3] Jian-Guang Lou, Hua Cai, and Jiang Li, "A real-time interactive multi-view video system," in *MULTIMEDIA '05*, NY, USA, 2005, pp. 161–170.
- [4] U. Engelke and H.-J. Zepernick, "Perceptual-based Quality Metrics for Image and Video Services: A Survey," in *EuroNGI07*, May 2007, pp. 190–197.
- [5] Andrew B. Watson, James Hu, and John F Iii, "DVQ: A digital video quality metric based on human vision," *Jour. of Elect. Imaging*, vol. 10, pp. 20–29, 2001.
- [6] Jeffrey Lubin, "Sarnoff jnd vision model," Tech. Rep. T1A1.5, T1 Standrads Committee, 1997.
- [7] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Trans. on Image Proc.*, vol. 13, pp. 600–612, 2004.
- [8] G.H. Chen, C.L. Yang, and S.L. Xie, "Edge-Based Structural Similarity for Image Quality Assessment," in *ICIP06*, 2006, pp. 2929–2932.
- [9] S. F. Chang and A. Vetro, "Video Adaptation: Concepts, Technologies, and Open Issues," in *Proc. of the IEEE*, January 2005, vol. 93, pp. 148–158.
- [10] A. Murat Tekalp Nukhet Ozbek and E. Turhan Tunali, "Rate Allocation Between Views in Scalable Stereo Video Coding using an Objective Stereo Video Quality Measure," in *ICASSP*, April 2007, pp. 1045–1048.
- [11] C.S. Shen, C. Zhang, and S. Fels, "A Multi-Camera Surveillance System that Estimates Quality-of-View Measurement," in *ICIP07*, 2007, pp. III: 193–196.
- [12] J. Starck, J. Kilner, and A. Hilton, "Objective Quality Assessment in Free-Viewpoint Video Production," in *3DTV08*, 2008, pp. 225–228.
- [13] Peter J. Burt and Edward H. Adelson, "A Multiresolution Spline with Application to Image Mosaics," *ACM Trans. on Graphics*, vol. 2, pp. 217–236, 1983.
- [14] Xiaonong Ran and Nariman Farvardin, "A Perceptually Motivated Three-component Image Model-Part I: Description of the Model," *IEEE Trans. on Image Proc.*, vol. 4, no. 4, pp. 401–415, 1995.
- [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.