A ghosting artifact detector for interpolated image quality assessment

Kai Berger*, Christian Lipski*, Christian Linz*, Anita Sellent* and Marcus Magnor*
*Computer Graphics Lab
TU Braunschweig, Germany
Email: {berger, lipski, linz, sellent, magnor} @ cg.cs.tu-bs.de

***Abstract*—We present a no-reference image quality metric for image interpolation. The approach is capable of detecting blurry regions as well as ghosting artifacts, e.g., in image based rendering scenarios. Based on the assumption that ghosting artifacts can be detected locally, perceived visual quality can be predicted from the amount of regions that are affected by ghosting. Because the approach does not require any reference image, it is very suitable, e.g., for assessing quality of image-based rendering techniques in general settings.**

I. Introduction

Free-viewpoint navigation around real-world, dynamic scenes has recently received considerable attention in computer graphics as well as computer vision, either making use of reconstructed scene geometry [1], [2], [3], [4], [5] or being based on the input images alone [6], [7], [8]. In both approaches, the quality of the rendered virtual image can safely assumed to be lower than the input photos. Quality degradation in image interpolation typically manifests itself in image blurring and image ghosting, artifacts observers perceive as highly distracting.

In geometry-assisted systems, artifacts come from inexact reconstruction and/or camera calibration errors. Some of the visual degradation can be corrected by advanced reprojection techniques [9]. In purely image-based systems, the major sources of error are occlusions/disocclusions of different depth layers and inaccurate correspondences. The latter, in general, leads to prominent ghosting artifacts in the interpolated image. While those artifacts can be compensated for by correcting the correspondence fields [10], their correction in the novel views remains a tedious and often subjective manual task.

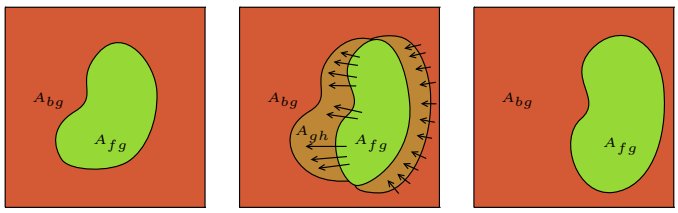


Fig. 1. Ghosting occurs in image interpolation when the corresponding pixels from two or more original images I_1, I_2 are warped to different pixel positions in the interpolated image I_{int}: At the left border of O the region A_{fg} is replicated and displaced by $\vec{d}_1$. The color values of A_{gh} appear alpha-blended from A_{fg} and A_{bg}. At the right border of O a similar displacement $\vec{d}_2$ occurs.

The identification of various artifacts including blurriness, blockiness and noisiness has been extensively studied in the field of image/video compression and transmission and can be applied to the analysis of interpolated images as well [11], [12], [13], [14], [15]. However, only little research has been devoted to image quality metrics tailored to the particularities of free-viewpoint systems, i.e., detection of blurring and ghosting where no ground-truth data is available. As a first step in this direction, Stark and Kilner [16], [17] proposed a quality assessment system for free-viewpoint video production. Yet, the metrics used essentially measure the quality of 3D reconstruction in image space and are thus only applicable to geometry-based free-viewpoint systems.

As main contribution, this paper presents a no-reference objective quality metric for purely image-based free-viewpoint systems, focusing on the detection of ghosting artifacts where no ground-truth reference images are available for comparison. Our approach is based on the observation that ghosting occurs when corresponding pixels from two or more original images I_1, I_2 are warped to different pixel positions in the interpolated image, Fig. 1. We validate the results of our ghosting detector against the results of a user study. The user study evaluation also confirms that ghosting and blurring are perceived as the most distracting artifacts in image interpolation.

The rest of this paper is organized as follows. After summarizing the related work in Sect. II, we give a detailed description of our ghosting detector in Sect. III. In Sect. IV, we present the subjective quality evaluation study of our image set. We compare the performance of our ghosting detector to the subjective evaluation in Sect. V, before the paper concludes in Sect. VI.

II. Related Work

Image quality assessment methods have been thoroughly studied over the past two decades. The main focus has been on the evaluation and improvement of image and video compression methods, such as JPEG2000, or MPEG-2.

Quality metrics have been developed to measure the quality of an image or video that has been altered due to compression or transmission over a communication channel. A good overview over existing metrics is provided by Engelke and Zepernik in [18]. Basically, metrics can be separated into three

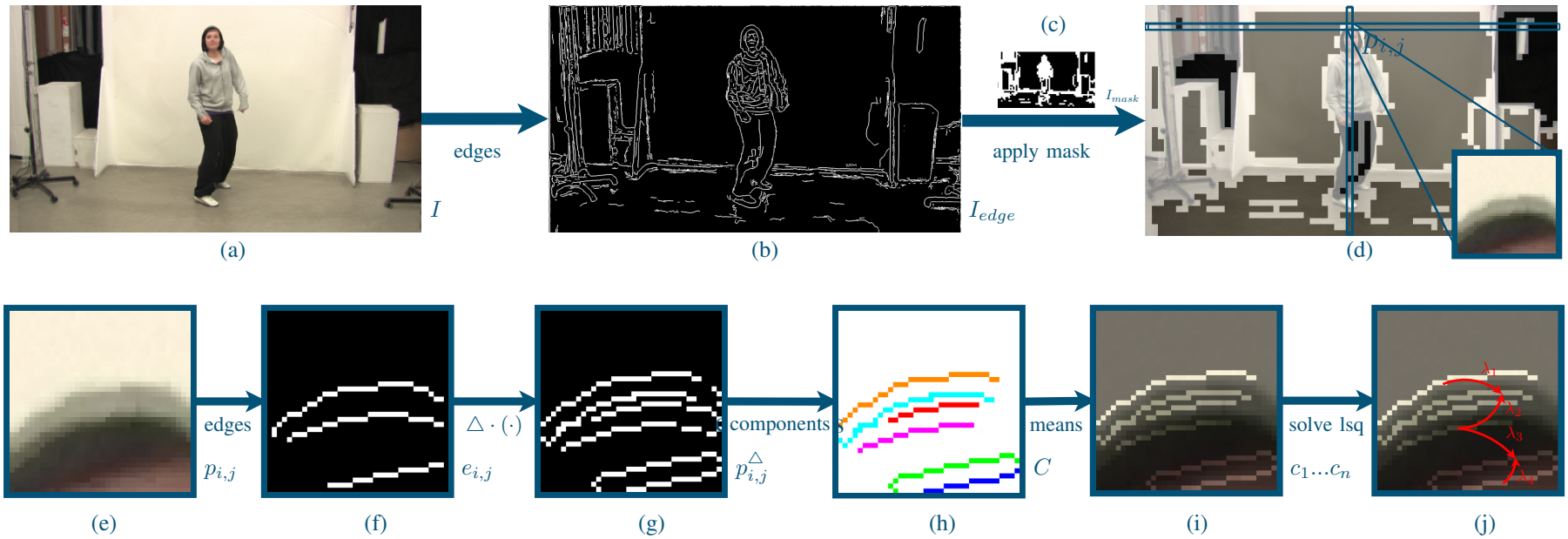


Fig. 2. The processing steps of the algorithm. **Top row**: the input image I (a) is smoothed and an edge detection is applied to get I_{edge} (b). A mask (c) is calculated to obtain the relevant patches of I (d, bright regions). **Bottom row**: (from left to right): For each patch $p_{i,j}$ (e) the algorithm performs an edge detection to get $e_{i,j}$ (f) and the Laplacian operator $p^{\triangle}_{i,j}$ (g) is applied to it. Connected components in $p^{\triangle}_{i,j}$ are grouped into sets C (h). For each group $c \in C$ the mean color value of the corresponding image pixels in I is calculated (i). Finally a least-squares problem is solved for each three components in C (j). The red variables λ_1,λ_2 sum to 1, the 2-norm of the residual is small. The same holds for λ_3,λ_4.



Fig. 3. The three test scenes used for the subjective user study. Form each scene original and interpolated images were evaluated.

classes: Full-reference (FR), Reduced-Reference (RR) and No-reference (RR) metrics.

Full-reference metrics compare a processed image and compare it to the original image [12], [19]. Distorted videos are evaluated in comparison to the original videos [20], [21], [22].

Reduced-reference metrics extract key features of the original image and compare them to key features extracted from the altered image. The key features are usually provided via an ancillary channel [23] or are embedded in the image [24].

No-reference metrics evaluate only the altered image and apply filters or heuristics to it. Liu et al. [15] detect the blockiness in compressed images and rely on the periodicity of artifacts due to the DCT transform. Sheikh et al. [14] use natural scene statistics in the wavelet domain to assess the image quality to tackle the DWT coefficient quantization introduced by JPEG2000. A perceptual blur metric is introduced by Marziliano et al. [12]. They measure the blurriness in terms of edge width, i.e. the pixel distance between local extrema around a detected edge pixel in gray-level images. Another blurriness metric searches for phase distortion in the Fourier spectrum[11]. Farias et al. [13] propose an artifact assessment in videos by combining a blockiness,blurriness and noisiness metric.

Only recently, quality measurements of free-viewpoint video results have been addressed. Kilner et al.[17], [16] investigate video errors caused by image rendering techniques and propose a reduced reference metric. They measured the pixel error of an image from a new viewpoint to the images of adjacent input cameras based on the Hausdorff-distance. However, their metric is only applicable in free-viewpoint systems based on geometric reconstruction. Our approach, in contrast, is purely image-based and does not rely of any geometric proxy and examines only the interpolated image itself.

III. The ghosting metric

In order to assess image quality based on ghosting artifacts, we make two assumptions. We assume that ghosting can be detected locally, and that ghosting artifacts are only visible in areas that contain strong object edges. In a first step, the input image I is subdivided into small patches $p_{i,j}$ of surface area d^2. In our image set, we chose $d = 15$ pixels. Since the most noticeable quality loss appears along object edges, the algorithm detects only patches near edges. According to our second assumption, object edges are usually predominant edges. In order to find them, an edge detection is applied to a low-pass filtered version of the input image I, retaining only the most prominent edges in I_{edge}, Fig. 2(b). In a second step, a binary mask I_{mask} is calculated from I_{edge} to determine the relevant patches, Fig. 2(c). The mask image is d^2 times smaller than the input image I. A pixel $I_{i,j}$ in I_{mask} is set to 1, if its corresponding patch $p_{i,j}$ contains at least d edge pixels, Fig. 2(d).

After this preprocessing, the algorithm iterates over all selected patches P and assigns a label $l_{i,j} \in L$ with $L = \{ghosting, crisp\}$ to each patch $p_{i,j} \in P$. In this classification step the algorithm performs edge detection on the patch $p_{i,j}$ of the input image I to obtain an edge patch $e_{i,j}$, Fig. 2(f). To this edge patch the Laplacian operator is applied

$$p^{\triangle}_{i,j} = \triangle \cdot (e_{i,j}).$$

The Laplacian-transformed patch $p^{\triangle}_{i,j}$ contains nonzero pixels only in the 8-neighborhood of an edge pixel in $e_{i,j}$, Fig. 2(g). Hence, we assume that for each edge in $e_{i,j}$ there exist two sets of connected pixels in $p^{\triangle}_{i,j}$, one set for each side of the edge. The algorithm then groups each set $c = \{I_{x_1,y_1}, \ldots, I_{x_m,y_m}\}$ of connected pixels into a set $C = \{c_1, \ldots, c_n\}$. Note that the color values of the input image are stored for each set of connected pixels, Fig. 2(h). Afterwards for each of the three sets $c_a, c_b, c_c, (a \neq b \neq c)$ of connected pixels in C, the mean color values $\vec{m}_a, \vec{m}_b, \vec{m}_c \in \mathbb{R}^3$ are computed,Fig. 2(i), and a least-squares problem is solved:

$$(\vec{m}_a \ \vec{m}_b) \cdot \vec{\lambda} = \vec{m}_c$$

with

$$\vec{\lambda} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}, 0 \leq \lambda_1, \lambda_2 \leq 1$$

If there is a $\vec{\lambda}$ with $\lambda_1 + \lambda_2 = 1$ for three sets c_a, c_b, c_c and with a small 2-norm of the residual, then the colors of c_a and c_b can be blended into the colors of c_c, Fig. 2(j). Hence c_a, c_b, c_c are considered to belong to a ghosting artifact and the label $l_{i,j}$ of the patch is set as $ghosting$. This is due to the fact that ghosting occurs when a surface area is repeated, slightly shifted within the image, or blended with the background.

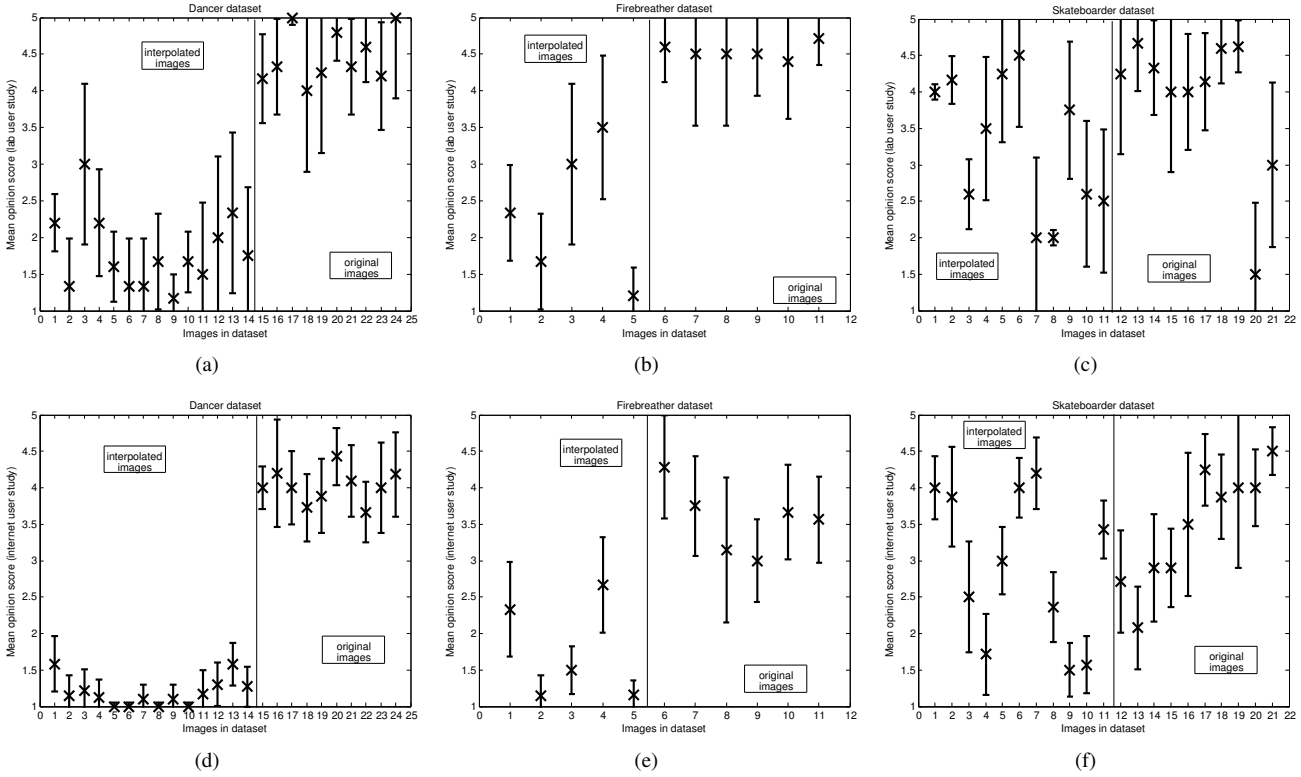


Fig. 4. The results of the user studies are plotted for each image dataset. The first row shows the lab user study (a - c), the second row shows the internet user study (d - f). The plot shows the mean opinion score for a given image in the dataset. The vertical bars for each data point mark the 95% confidency interval. The vertical line distinguishes the interpolated images from the original images.

The overall numerical quality $g(I)$ of the image is finally computed as the percentage of patches labeled as $ghosting$ to the overall number of detected patches in the image:

$$g(I) = \frac{\sum_{p_{i,j} \in P} g(p_{i,j})}{\|P\|},\ 0 \leq g(I) \leq 1$$

where

$$g(p_{i,j}) = \begin{cases} 1, & l_{i,j} = ghosting \\ 0, & l_{i,j} = crisp \end{cases} .$$

IV. Subjective Evaluation

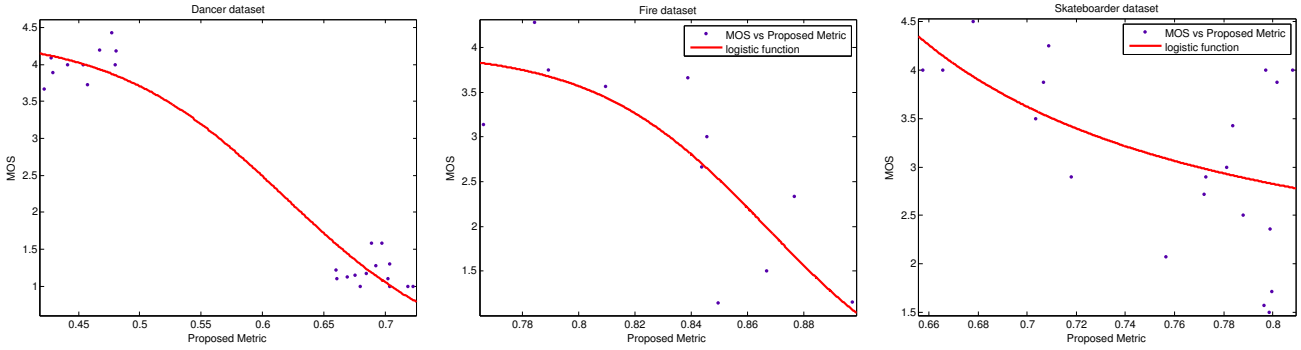
The presented metric has been evaluated against a subjective user study. For this study a set of 20 images per scene has been randomly assembled from 3 different scenes, Fig. 3, consisting of 50 % images from a video camera and 50 % rendered images from a novel viewpoint. The interpolated images have been generated with the algorithm proposed by Stich et al. [25]. The rendered images from a novel viewpoint show quality alterations of different severity. In accordance with the international recommendations for subjective video quality assessment [26], the image set was presented to 15 human observers in a laboratory environment. The observers were given the task to grade each of the presented images with one of the following values according to the the ITU-R quality scale: ”excellent” (highest score), ”good”, ”fair”, ”poor”, ”bad” (lowest score).

In a second step, an online evaluation system was set up and advertised university-wide. Again the observers could grade the images with the same scores. The online evaluation was performed within a week, 61 students participated.

In both studies the results have been analyzed as proposed by [26]. Fig. 4 shows for each dataset the mean opinion score and the 95% confidence interval per image for the lab environment user study (a-c) and the internet user study (d-f). Comparing for each dataset the plot for the lab environment user study to the internet user study, it can be observed, that the confidence interval of the mean opinion score for most images in the lab user study is broader than in the internet evaluation. This can be explained by the small number of participants. Furthermore it can be stated for both user studies, that the mean opinion scores for interpolated images are lower than for original images. This statement holds for the skateboarder, firebreather, and dancer image set. The difference is smaller for the skateboarder since there are only few artifacts in the shadow region which seem to have less impact on perceived quality.

V. Results

We have applied the algorithm described in Sect. III to the presented image dataset which consists of 3 different scenes with roughly 50% interpolated images and 50% original images each. The image resolution of each image in the set



(a) (b) (c)

Fig. 5. For each dataset the results of the proposed metric are compared to the results of the internet user study. For each image in the dataset the percentage of patches containing ghosting ($g(I)$) is plotted against the mean opinion score for the image. The red line indicates the approximation by a logistic function.

was 960×540 pixels, the evaluation time varied between 10 and 30 seconds on a PC. In order to predict a subjective evaluation with the proposed metric, [26] suggests that their result values have to be approximated by a logistic function $y = \frac{a \cdot b}{(a-b) \cdot \exp(-cx)+b}$, where x is the range of the metric, y is the mean opinion score and a, b, c are the parameters of the logistic function. In Fig. 5 such an approximation is depicted for each dataset. In each graph the results of the metric are approximated to the MOS values for each image in the dataset. A data point represents an image of the data set. The plots for the dancer, Fig. 5(a), the skateboarder, Fig. 5(b), and the firebreather, Fig. 5(c), dataset show a reasonable approximation by a logistic function. Hence the metric is capable of predicting a subjective value for an image in these datasets. From these results we deduce, that the proposed metric performs well on images which contain ghosting artifacts on the outline of opaque objects.

VI. Conclusion

We have presented a versatile image quality metric which succeeds in detecting ghosting artifacts. The new metric is purely image-based and can be classified as no-reference metric. A laboratory study based on the international recommendations for subjective video quality assessment [26] and a university-wide internet user study have been done to evaluate the confidence of the metric. Both studies are based on the same image dataset, which consist of 3 different scenes, roughly composed with 50% original and 50% interpolated images. Best accuracy is achieved for scenes with opaque objects and occlusion edges.

In the future, we want to extend our metric from still images to interpolated image sequences with a short time span to investigate the temporal evolution of ghosting artifacts. Secondly, we want to improve our metric towards non-opaque objects, like foam and water to get more accurate statements about their change of appearance during interpolation.

References

[1] J. Carranza, C. Theobalt, M. A. Magnor, and H.-P. Seidel, "Free-viewpoint video of human actors," *ACM Trans. on Graphics*, vol. 22, no. 3, pp. 569–577, July 2003. [Online]. Available: http://dx.doi.org/10.1145/882262.882309

[2] C. Zitnick, S. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-Quality Video View Interpolation Using a Layered Representation," *ACM Trans. on Graphics*, vol. 23, no. 3, pp. 600–608, 2004.

[3] S. Vedula, S. Baker, and T. Kanade, "Image Based Spatio-Temporal Modeling and View Interpolation of Dynamic Events," *ACM Trans. on Graphics*, vol. 24, no. 2, pp. 240–261, 2005.

[4] J. Starck and A. Hilton, "Surface Capture for Performance Based Animation," *IEEE Computer Graphics and Applications*, vol. 27(3), pp. 21–31, 2007.

[5] E. de Aguiar, C. Stoll, C. Theobalt, N. Ahmed, H.-P. Seidel, and S. Thrun, "Performance Capture from Sparse Multi-View Video," *ACM Trans. on Graphics*, vol. 27, no. 3, pp. 1–10, 2008.

[6] M. Levoy and P. Hanrahan, "Light Field Rendering," in *Proc. of ACM SIGGRAPH'96*. New York: ACM Press/ACM SIGGRAPH, 1996, pp. 31–42.

[7] W. Matusik and H. Pfister, "3D TV: A Scalable System for Real-Time Acquisition, Transmission, and Autostereoscopic Display of Dynamic Scenes," *ACM Trans. on Graphics*, vol. 23, no. 3, pp. 814–824, 2004.

[8] H. Wang, M. Sun, and R. Yang, "Space-Time Light Field Rendering," *IEEE Trans. Visualization and Computer Graphics*, pp. 697–710, 2007.

[9] Eisemann, M., D. Decker, B., Magnor, M., Bekaert, P., D. Aguiar, E., Ahmed, N., Theobalt, C., and A. Sellent, "Floating textures," *Computer Graphics Forum*, vol. 27, no. 2, pp. 409–418, April 2008. [Online]. Available: http://dx.doi.org/10.1111/j.1467-8659.2008.01138.x

[10] T. Stich, C. Linz, C. Wallraven, D. Cunningham, and M. Magnor, "Time and View Interpolation in Image Space," *Computer Graphics Forum (Proc. Pacific Graphics'08)*, vol. 27, no. 7, pp. 1781–1787, 10 2008.

[11] Z. Wang and E. P. Simoncelli, "Local phase coherence," in *in Adv. Neural Information Processing Systems (NIPS03*, vol. 16, 2003, pp. 786–792. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.9.3565

[12] P. Marziliano, "Perceptual blur and ringing metrics: application to jpeg2000," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 163–172, 2004. [Online]. Available: http://dx.doi.org/10.1016/j.image.2003.08.003

[13] M. C. Q. Farias and S. K. Mitra, "No-reference video quality metric based on artifact measurements," in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 3, 2005, pp. III–141–4. [Online]. Available: http://dx.doi.org/10.1109/ICIP.2005.1530348

[14] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No-reference quality assessment using natural scene statistics: Jpeg2000," *Image Processing, IEEE Transactions on*, vol. 14, no. 11, pp. 1918–1927, 2005. [Online]. Available: http://dx.doi.org/10.1109/TIP.2005.854492

[15] H. Liu and I. Heynderickx, "A no-reference perceptual blockiness metric," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 2008, pp. 865–868. [Online]. Available: http://dx.doi.org/10.1109/ICASSP.2008.4517747

[16] J. Starck, J. Kilner, and A. Hilton, "Objective quality assessment in free-viewpoint video production," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008*, 2008, pp. 225–228. [Online]. Available: http://dx.doi.org/10.1109/3DTV.2008.4547849

[17] J. Kilner, J. Starck, J. Y. Guillemaut, and A. Hilton, "Objective quality assessment in free-viewpoint video production," *Signal Processing: Image Communication*, vol. 24, no. 1-2, pp. 3–16, 2009.

[18] U. Engelke and H. J. Zepernick, "Perceptual-based quality metrics for image and video services: A survey," in *Next Generation Internet Networks, 3rd EuroNGI Conference on*, 2007, pp. 190–197. [Online]. Available: http://dx.doi.org/10.1109/NGI.2007.371215

[19] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004. [Online]. Available: http://dx.doi.org/10.1109/TIP.2003.819861

[20] Z. Wang and A. C. Bovik, "A human visual system-based objective video," in *Intl Conf. On Multimedia Processing and Systems*, 2001. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.10.8803

[21] K. Seshadrinathan and A. C. Bovik, "An information theoretic video quality metric based on motion models," in *Proc. Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2007, pp. 25–26. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.76.9341

[22] H. R. Sheikh and A. C. Bovik, "A visual information fidelity approach to video quality assessment," in *The First International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2005, pp. 23–25. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.124.8865

[23] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in *in Proc. of SPIE Human Vision and Electronic Imaging*, vol. 5666, 2005, pp. 149–159. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.75.3

[24] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E.-H. Yang, and A. C. Bovik, "Quality-aware images," *Image Processing, IEEE Transactions on*, vol. 15, no. 6, pp. 1680–1689, 2006. [Online]. Available: http://dx.doi.org/10.1109/TIP.2005.864165

[25] T. Stich, C. Linz, C. Wallraven, D. Cunningham, and M. Magnor, "Perception-motivated interpolation of image sequences," in *APGV '08: Proceedings of the 5th symposium on Applied perception in graphics and visualization*. New York, NY, USA: ACM, 2008, pp. 97–106. [Online]. Available: http://dx.doi.org/10.1145/1394281.1394299

[26] R. ITU, "Recommendation BT. 500-11," *methodology for the subjective assesment of the quality of televission pictures, tech. rep., International telecommunication union, Geneva, Switzerland*, 2002.