https://doi.org/10.2352/ISSN.2470-1173.2017.14.HVEI-125

Subjective evaluation of distortions in first-person videos

Chen Bai and Amy R. Reibman;
Purdue University, West Lafayette, IN, USA

Abstract

First-person videos (FPVs) recorded by wearable cameras have different characteristics compared to mobile videos. Video frames in FPVs are subject to blur, rotation, shear and fisheye distortions. We design a subjective test that uses actual captured images with real distortions, synthetic distortions or a combination of both. Results indicate shear is less sensitive to content than rotation. For fisheye, personal preference and content dependence affect the subjective results. The performance of 7 no-reference (NR) quality estimators (QEs) and our QE, local visual information (LVI) [1], are evaluated based on subjective results. We propose two mapping functions for rotation and shear that improve the ability of LVI and 4 NR QEs to accurately predict the subjective scores.

Introduction

As wearable cameras gain increasingly popularity, the so-called first-person videos (FPVs) grow to be an important video source over networks [2]. FPVs have different attributes compared to mobile videos. When recording mobile videos, people can see the ongoing scene in the screen, so they can be aware of how to fit the scene into the camera shot and intentionally control the video to be less shaky. However, FPVs are passive videos, and often lack real-time feedback about the field of view. As a result, the wearer rarely adjusts the camera to be stable or horizontal. Recently, the increased volume of uploading, transmission and storage of FPVs has given rise to a variety of applications. Human Activities in FPVs are classified in [3]. The detection of "snap points" was proposed to extract frames that are worthy to be captured and stored [4]. The ego-sampling method was developed to sample useful frames for faster video browsing [5].

Because of the capture process of FPVs, quality degradations are not limited to transmission or post-processing, and the resulting distortions in frames have not been subjectively evaluated. Motion blur and geometric distortions are two major distortions in FPVs [1]. Motion blur mainly arises from fast motion of the camera. When the camera keeps changing its positions in the scan time of one frame, the captured scene is blurred. Geometric distortions can be classified into 3 categories: rotation, shear and fisheye. Rotation results from head or body rotation. Wearers regularly move their bodies and shake their heads, and rarely are aware whether or not the camera is kept horizontal while recording videos. Shear is caused by camera panning. When the camera changes its positions in the scan time of one frame, the top rows of the frame are not vertically align with the bottom rows. For example, architecture in a sheared image is visually skewed. Fisheye images are captured by wearable cameras with ultra wide-angle lens (i.e. Gopro, Looxie Camera, Mobius). Instead of capturing a rectilinear image, the content appears to be convex. Fisheye is one of the lens distortions, called barrel transformation. The transformation warps the image to be bent; the magnification decreases from center to margins [6].

Most existing subjective tests [7, 8, 9] explored quality degradations starting from an undistorted image, which is considered to be the reference image. By synthetically adding distortions (i.e. Gaussian blur, JPEG, JPEG 2000, noise) to the reference image with different degradation levels, a series of distorted images of the same content but different severity are created. To design a subjective test for FPVs, however, applying each type of distortions separately fails to consider two issues:

1. Degradations in actual images captured by a camera may be subtly different than those created synthetically using a model [10]. In our test, images with real, synthetic or real plus synthetic distortions are evaluated. The real images are extracted from frames in FPVs.

2. Many blurry images are subject to geometric distortions (i.e. rotation, shear or fisheye) simultaneously in FPVs. The question is what is the visual impact on the overall quality when one image has multiple distortions. Multiply-distorted images have been evaluated in the subjective test, but only for blur, JPEG and noise [11]. In addition, these distortions, like blur, JPEG or noise, all have pixel-to-pixel correspondence, whereas geometric distortions do not. When constructing images for the subjective test for different amounts of geometric distortions, we keep the same content in the image center to maintain a consistent focus of attention.

Actual images captured by a wearable camera have no reference image, so full-reference (FR) quality estimators (QEs) [12, 13] and reduced-reference (RR) QEs [14, 15] cannot be used to measure their quality, but no-reference (NR) QEs [16, 17, 18] are commonly content dependent and show very different behaviors for different distortions applied to the same image. Using our subjective test, we evaluate the behaviors of different NR QEs for actual captured images and multiply-distorted images.

We design a subjective test to evaluate motion blur and geometric distortions in FPVs. Section 2 compares existing QEs with our QE, local visual information (LVI) [1]. The strategy of the quality estimation in LVI cannot be categorized into FR, RR or NR methods. In section 3, our subjective test evaluates actual captured images with real distortions, synthetic distortions or a combination of both using the paired comparison method. The types of distortions include motion blur, rotation, shear and fisheye. In section 4, we proposed two mapping functions for rotation and shear to compute the overall quality of images with blur and geometric distortions. Personal preferences and content dependence in fisheye are discussed.

Comparison of LVI with existing QEs

FR QEs and RR QEs often use a reference image to estimate the quality of a distorted image with pixel-to-pixel correspondence. They cannot be used for any image for which there is no corresponding reference. NR QEs uses only a single image without any reference to compute a quality score.

In [1], we introduced a QE, LVI, that does not fit in the typical categorization of QEs into FR, RR and NR [19]. LVI estimates the quality of a test image based on a pseudo-reference image. Unlike in FR QEs, the pseudo-reference image and the test image are not necessarily pixel-aligned but obey the constraint that they share overlapping content. For example, the two images could be captured by a camera in the same scene but different camera angles. The geometric relationship between the two images are usually modeled by a homography.

Since the quality of the pseudo-reference image is unknown, LVI is a relative quality estimation instead of an absolute quality score as computed by FR or NR QEs. LVI primarily measures the amount of relative blur of the test image based on the pseudo-reference. When the test image is more blurred than the pseudo-reference image, its LVI score would be in the range from 0 to 1. Otherwise, its LVI score would be larger than 1, which indicates the test image is sharper. If the two images have too few matching feature points, we then set the LVI score of the test image to be 0.

LVI is insensitive to rotation and shear [1]. To make LVI be able to measure geometric distortions, we need to include rotation and shear into its quality estimation. We design a subjective test to evaluate multiply-distorted (blur with geometric distortions) images. The subjective results are used to develop the rotation and the shear mapping functions for LVI. The mapping functions take the rotation or shear and the LVI score as inputs to compute the overall quality score of an image with simultaneous blur and geometric distortions.

Subjective Test

Our test employs the paired comparison method for still images containing both actual and synthetic distortions that are typical of images extracted from FPVs. Each pair of test images are simultaneously displayed on two monitors side by side. Viewing distance is kept to be around 3 times the height of test images. 9 videos have been recorded by a Pivothead camera (frame rate: 30fps, resolution: 1920×1080) including “billiards", “eating", “flight", “bell tower",“winter Hovde Hall", “parking lot", “autumn Hovde Hall", “apartment building" and “parking garage". 4 distortions including motion blur, rotation, shear and fisheye, are evaluated. Test images are either real frames from the 9 videos or created by adding synthetic distortions to selected frames. Because synthetic shear and fisheye change the image size compared to the original image, all test images are cropped to be 1600×900 to remove marginal regions with little content.

Test Method

In a paired comparison subjective test, the subject needs to indicate his or her preference among the two images according to their visual quality. 50 subjects including 43 males and 7 females participated in the test. All pairs of images are displayed in random order. To improve the efficiency of paired comparison, we use the “square design" in [20]. Given that we have n stimulus, a full comparison needs $0.5n(n-1)$ pairs. By using the “square design", the comparison number can be decreased to $n(\sqrt{n}-1)$.

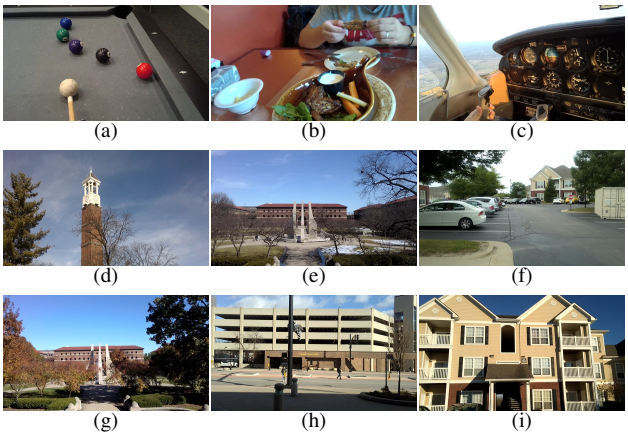

Figure 1: Reference images for each content: (a) billiards (b) eating (c) flight (d) bell tower (e) winter Hovde Hall (f) parking lot (g) autumn Hovde Hall (h) apartment building (i) parking garage

video content	real distortions	synthetic distortions	number of images	SI
(a)	blur (5 levels)	-	5	6.58
(b)	blur (5 levels)	-	5	8.57
(c)	blur (5 levels)	-	5	9.40
(d)	rotation (4 angles)	blur (4 levels)	16	9.69
(e)	rotation (4 angles)	blur (4 levels)	16	15.75
(f)	-	blur (4 levels) + shear (4 levels)	16	13.30
(g)	-	blur (4 levels) + shear (4 levels)	16	16.78
(h)	-	blur (4 levels) + fisheye (3 levels)	12	13.91
(i)	-	blur (4 levels) + fisheye (3 levels)	12	18.50

Table 1: Test images in subjective test

Test Setup

A pair of test images are presented on the two monitors (Dell U2415) side by side with time synchronization. The two monitors are calibrated (calibration tool: Spyder5ELITE) to have ignorable visual difference. The brightness after calibration is 120 cd/m^2. The monitor resolution is 1920×1200, and the test image size is 1600×900. The test image is displayed in the center of the monitor with a surrounding background that is uniformly gray 128.

The environment fixes the viewing conditions for each subject to minimize the influence from external stimuli other than the test images. For each pair of images, the subject needs to indicate which image has better viewing quality according to his or her visual evaluation by keyboard (“1" for choosing left image, “0" for choosing right image). The maximum time for the comparison of one pair of images is 10 seconds. Whenever

Figure 2: Test images captured in FPVs to have different amounts of motion blur

Figure 3: Test images intentionally captured to have different amounts of rotation

the subject fails to make a choice after 10 seconds, he or she must randomly choose one of the two test images as the better one. The time interval between each comparison is 1.5 seconds. The interface for this test is built on PsychoPy [21]. We also conducted informal post-test feedback discussions with some participants who were willing to share their opinions.

Test Sources

Figure 1 shows source images (actual captured frames) of each content and Table 1 lists all test images of different categories in our test. The index of each content is the same as in Figure 1. We take 3 distinct approaches to create test images for motion blur, rotation plus blur, shear plus blur and fisheye plus blur. First, for “billiards", “eating", “flight", 5 nearby frames of each content that have different amounts of blur are selected. Next, “Winter Hovde Hall" and “bell tower" are intentionally created by continuous head rotation in front of a scene. 4 sharp frames with different amounts of rotation are then selected from these two sequences and different amounts of synthetic motion blur are added. Finally, one frame is chosen to be the reference respectively from “Autumn Hovde Hall", “parking lot", “parking

Figure 4: Test images with different amounts of synthetic shear created from one reference image

Figure 5: Test images with different amounts of synthetic fisheye created from one reference image

garage" and “apartment building": distortions are applied to the reference with controllable amount. We also measure and report in Table 1 the spatial information (SI) [22] of each source image. SI is calculated as the mean of the gray-scale image filtered with both vertical and horizontal Sobel kernels. We experimentally find that the 3 geometric distortions, rotation, shear and fisheye, have very small influence on the SI of images in the same blur level.

Motion Blur: Our test of motion blur uses both nearby frames from FPVs and synthetic distorted images. These nearby frames are chosen to share at least half of their content and have minor difference in rotation and shear, but they differ in the amount of blur. Figure 2 shows chosen frames with the most and the least motion blur for each content. Synthetic motion blur are created by the motion model in [23]. The model can be used to create non-linear motion blur kernels by controlling motion trajectory and motion kernel size. In our test, the motion trajectory is clockwise 45° diagonally up to the right in a straight line. The motion kernels are created with size 2×2, 4×4, 8×8 to apply 3 levels of motion blur.

Rotation: Synthetic generation of rotation would require a significant area of the rotated image to be cropped to maintain a rectangular image. Therefore, we uses real images selected out of videos which were purposely created to contain rotated frames. The center of each image from the same video is selected to be almost the same location of the scene. This is intended to avoid

initial shear	$k_1 = 0.00$	$k_2 = 0.00$	$k_3 = 0.00$	$k_4 = 0.00$
incremental shear 1	$+0.105$	$+0.105$	$+0.105$	$+0.105$
incremental shear 2	-0.035	-0.035	$+0.035$	$+0.035$
incremental shear 3	-0.035	$+0.035$	-0.035	$+0.035$
incremental shear 4	-0.035	-0.035	$+0.035$	$+0.035$
final shear	$k_1 = 0.00$	$k_2 = 0.07$	$k_3 = 0.14$	$k_4 = 0.21$

Table 2: Symmetric transformation method to create shear images

a change in the location of focus of attention. Figure 3 shows sample test images of different amounts of rotation.

Shear: In geometric transformation, shear between two images can be modeled as

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \quad (1)$$

where $[x \quad y]^T$ are points in the reference image, and $[x' \quad y']^T$ are corresponding points in the sheared image. k is the shear parameter and $atan(k)$ indicates the angle of shear.

Shear transformation is a spatially varying filter. To create synthetic images with different levels of shear but almost the same amount of blur, we introduces a "symmetric transformation method" to add shear to the reference image as is shown in Table 2. Let i be image index, k_i is shear parameter for each image. The initial shear is 0. In each step of the process, we add the same amount of shear to each test image. The sign in front of shear amount is angle direction (i.e. "+" indicates shear to the right, "-" indicates shear to the left). At the conclusion of all steps, each test image has nearly similar blur but a distinct amount of shear. Figure 4 shows sample test images with different amounts of synthetic shear.

Fisheye: Fisheye distortion is a barrel transformation, which can be modeled as [24]

$$r' = a*r + b*r^2 + c*r^3 + d*r^4 + \mathcal{O}(r^5), \quad (2)$$

where r and r' are the distances of pixels to the image center in non-fisheye and fisheye images, respectively. $\mathcal{O}(r^5)$ are the higher order terms of r, which can be ignored. a, b, c, d are coefficients depending on the camera lens.

Barrel transformation introduces spatially varying blur to images. Therefore, we again use the "symmetric transformation method" shown in Table 2 to create image pairs with different amounts of fisheye. We set $a = 1$, $b = d = 0$, ignore $\mathcal{O}(r^5)$ and vary c only to get different amount of fisheye distortions. Fisheye images have decreasing scale factors from center to the image edges; our synthetically distorted images have the same scale in the image center compared to the original image. Figure 5 shows sample test images with different amounts of synthetic fisheye.

Results and Discussion

Our subjective test evaluates 4 distortions: motion blur, rotation, shear and fisheye. The test of motion blur uses nearby frames extracted from FPVs. Rotation, shear and fisheye are evaluated simultaneously with synthetic motion blur. By applying the Bradley-Terry model with maximum likelihood estimation [25], paired comparison results can be converted to relative subjective scores. Note that we use a logarithmic scale for the final subjective scores and set the score of the best image in each test to be 0. We also calculate the 95 percent confidence interval (CI) of each

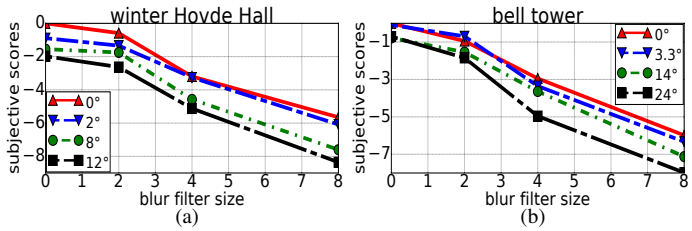

Figure 6: Subjective test - rotation and blur: (a) winter Hovde Hall (b) bell tower

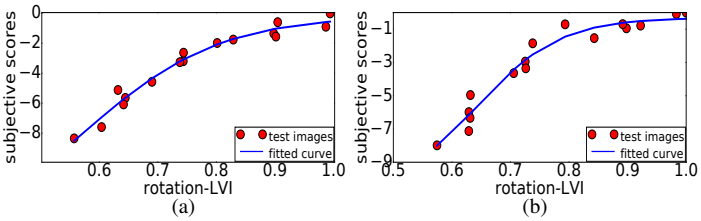

Figure 7: Curve fitted with logistic function between subjective scores and rotation-LVI: (a) winter Hovde Hall (p=4.53) (b) bell tower (p=1.13)

subjective score using the method presented in [25]. Each subjective score is represented as $q_s \pm q_r$, where q_s is the estimated subjective score, q_r is half the range of the 95 percent CI of q_s. If the CIs of two image scores overlap each other, the quality difference of the two images is not significant, or namely, their quality is similar.

The test results indicate (1) LVI is effective at measuring blur when two images are not perfectly aligned, in the absence of rotation and shear. (2) Rotation and shear degrade quality. With the amount of degradation, shear is less sensitive to different contents than rotation. (3) The preference of fisheye versus non-fisheye differs from person to person, and this preference is influenced by content.

Motion Blur

The test of motion blur uses all actual captured frames from 3 FPVs with different content. In each content, the 5 test frames are selected to be temporally nearby and have different amounts of blur as measured by LVI.

In particular, the results show that LVI is an effective metric to estimate blur among misaligned images with minor rotation or shear difference. Seven existing NR QEs (JNBM [26], BIQI [27] , CPBD [28], BRISQUE [16], CORNIA [29], IL-NIQE [17] and NIQE [18]) and LVI are evaluated by subjective scores, as is shown in Tables 3, 4 and 5. The Spearman rank-order correlation coefficient (SROCC), the Kendall rank-order correlation coefficient (KROCC) and the Pearson linear correlation coefficient (PLCC) are employed to measure the performance of these QEs. Note that the images with the most severe blur, in "billiards" and "eating" respectively, cannot be measured by LVI because they have too few matching feature points. The LVI scores of these two images are considered to be zero. Note that LVI is the only QE we evaluated that correctly rank-orders the subjective quality for these image sets.

Rotation

Test images with different amounts of rotation are captured in front of the same scene with minor differences of viewpoint. The images with different rotation are selected to have tiny differ-

QE name	SROCC	KROCC	PLCC
JNBM	0.9000	0.8000	0.8248
BIQI	0.6000	0.4000	0.2342
CPBD	0.3000	0.2000	0.1905
BRISQUE	0.1000	0.0000	0.1130
CORNIA	0.6000	0.4000	0.7446
IL-NIQE	0.9000	0.8000	0.8599
NIQE	0.9000	0.8000	0.8514
LVI	**1.0000**	**1.0000**	**0.8752**

Table 3: QE performance: motion blur - billiards

QE name	SROCC	KROCC	PLCC
JNBM	0.4000	0.4000	0.5701
BIQI	0.8000	0.6000	0.7137
CPBD	0.2000	0.2000	0.2732
BRISQUE	0.5000	0.2000	0.5467
CORNIA	0.3000	0.2000	0.1610
IL-NIQE	0.9000	0.8000	**0.9597**
NIQE	0.8000	0.6000	0.8167
LVI	**1.0000**	**1.0000**	0.8719

Table 4: QE performance: motion blur - eating

QE name	SROCC	KROCC	PLCC
JNBM	0.9000	0.8000	0.8914
BIQI	0.9000	0.8000	0.9637
CPBD	0.6000	0.4000	0.6068
BRISQUE	0.3000	0.2000	0.5452
CORNIA	0.7000	0.6000	0.6342
IL-NIQE	0.9000	0.8000	0.9028
NIQE	0.2000	0.2000	0.2052
LVI	**1.0000**	**1.0000**	**0.9958**

Table 5: QE performance: motion blur - flight

ences as measured by LVI. Motion blur degradations are synthetically added to images with different rotation. Figure 6 shows the subjective scores of content "winter Hovde Hall" and "bell tower", in which the average range of CIs is 0.22 for each content.

First, we explore the intra-relationship of both motion blur and rotation. For a fixed rotation angle, subjective scores monotonically decrease when the blur filter size increases. Only when the blur filter size increases from 0 to 2, the CIs for the respective subjective scores have overlap. For a fixed blur level, rotation introduces quality degradations, and the influence becomes larger as the blur level increases. The rotated images of the lowest and the second lowest blur levels have closer subjective scores than rotated images of higher blur levels, which is reflected as the overlap between the confidence regions in the first two blur levels compared to the non-overlapping of the other two higher blur levels.

To elaborate, let the score of a test image be (r,b), where r is the rotation level, b is the blur level, and $r,b = 1,2,3,4$. For content "winter Hovde Hall", (3,1) and (4,1) are -1.55 ± 0.29 and -1.98 ± 0.23, respectively. The overlap between their confidence regions indicates their quality is similar. As a comparison, (3,3) and (4,3) are -4.57 ± 0.20, and -5.12 ± 0.24, respectively. The non-overlapping indicates that they have significant quality difference.

Second, we explore the inter-relationship between motion blur and rotation. In both contents, the quality differences of images are not statistically significant when blur is small. One example is in content "bell tower", (2,2) and (3,1) are similar, with scores -0.69 ± 0.27 and -0.78 ± 0.30, respectively. As a comparison, (4,3) is -4.96 ± 0.22, while (1,4) is -5.99 ± 0.23 worse than the former. As an addition, the content "winter Hovde Hall" has higher spatial information (SI) than "bell tower", and its subjective quality is more sensitive to rotation.

To model the overall quality measure of an image with blur plus rotation, we propose a mapping function to combine LVI and rotation. The mapping function is given by

$$Q(\theta, q_{LVI}) = q_{LVI} \cdot (1 - p \cdot \exp(q_{LVI} - 1) \cdot \theta^2), \tag{3}$$

where the rotation angle θ (radian) is estimated relative to the reference image by affine estimation as described in [1]. q_{LVI} is the LVI score of the distorted image, and p is a constant parameter which needs to be optimized. equation (3) is called rotation-LVI.

From discussion with the participants in the subjective test, preference regarding rotation is content dependent. The same rotation angle for different content gives rise to different viewing quality. We optimize p for each content to maximize SROCC and KROCC. The optimized p is 4.53 for "winter Hovde Hall" and 1.13 for "bell tower". Figure 7 shows the nonlinear mapping curve between subjective scores and rotation-LVI with optimized p. The logistic function used for curve fitting is

$$f(x) = (t_0 - t_1)/(1 + \exp(-(x - t_2)/|t_3|)) + t_1 \tag{4}$$

where t_0, t_1, t_2 and t_3 are 4 unknown parameters for fitting.

For extension to other quality metrics, the term q_{LVI} can be replaced with any other quality measure q, given by

$$Q(\theta, q) = q \cdot (1 - p \cdot \exp(-\frac{|q - q_{best}|}{|q_{best} - q_{worst}|}) \cdot \theta^2). \tag{5}$$

The term $\exp(q_{LVI} - 1)$ is replaced with $\exp(-\frac{|q-q_{best}|}{|q_{best}-q_{worst}|})$, where q_{best} and q_{worst} indicate the quality scores for the best- and the worst-quality images based on the corresponding quality measure q, respectively.

Table 6 and 7 show the performances of 7 NR QEs and LVI. "rotation-" indicates the QE is mapped by equation (3) with corresponding optimized p. Note that we use the self-reported best and worst QE values when available, otherwise the observed best and worst QE values in [30] are used. One exception is that JNBM has maximum value at infinity, so we set its best score as the QE score of the best image for each content. The comparison shows that LVI, JNBM, CPBD, IL-NIQE and NIQE all improve their performance after including rotation mapping. To get a generalized model, we fix p to be 1.16; the resulting SROCC of rotation-LVI are 0.9588 and 0.9529 for "winter Hovde Hall" and "bell tower", respectively. However, since two scenes are not enough to allow generalization of the model, more subjective data needs to be collected.

Shear

Test images distorted with shear and motion blur are synthetically created from one reference image. Figure 8 shows the subjective scores of content from two sequences "autumn Hovde Hall" and "parking lot", with average range of CIs to be 0.28 and 0.75, respectively. The difference of their CI range results from content difference. The content "parking lot" has lower SI than information "autumn Hovde Hall", and many of its edges are highly

QE name	p	SROCC	KROCC	PLCC
JNBM	-	0.8441	0.6833	0.8920
BIQI	-	0.6941	0.5167	0.7140
CPBD	-	0.8205	0.5833	0.8412
BRISQUE	-	0.6441	0.4500	0.7361
CORNIA	-	0.5059	0.3833	0.4201
IL-NIQE	-	0.9176	0.7500	0.9429
NIQE	-	0.8941	0.7667	0.8561
LVI	-	0.8529	0.7000	0.9042
rotation-JNBM	6.06	0.9706	0.8833	0.9102
rotation-CPBD	8.48	0.9559	0.8333	0.8955
rotation-IL-NIQE	-1.60	0.9529	0.8500	**0.9574**
rotation-NIQE	-5.14	0.9382	0.8333	0.9050
rotation-LVI	4.53	**0.9853**	**0.9333**	0.9480

Table 6: QE performances: rotation - winter Hovde Hall

QE name	p	SROCC	KROCC	PLCC
JNBM	-	0.9117	0.7333	0.9109
BIQI	-	0.6471	0.4333	0.6413
CPBD	-	0.9441	0.8167	0.9009
BRISQUE	-	0.3088	0.1833	0.5503
CORNIA	-	0.3853	0.2833	0.2539
IL-NIQE	-	0.7411	0.5667	0.8557
NIQE	-	0.9558	0.8500	0.9269
LVI	-	0.8764	0.6500	0.9256
rotation-JNBM	0.47	0.9618	0.8500	0.9180
rotation-CPBD	1.1	0.9705	0.9000	0.9052
rotation-IL-NIQE	-1.64	0.8882	0.7667	0.9210
rotation-NIQE	-0.4	**0.9794**	**0.9333**	**0.9406**
rotation-LVI	1.13	0.9618	0.8833	0.9221

Table 7: QE performances: rotation - bell tower

curved or within texture. Since local orientation structure is visually more sensitive to straight edges than curved edges or textures [31], shear is visually less sensitive in “parking lot" than in “autumn Hovde Hall".

First, we explore the intra-relationship of both motion blur and shear. In both contents, subjective scores monotonically decreases as the blur level increases for any fixed shear level. Within each blur level, the image with the greater shear often has worse visual quality. Note that while in “parking lot", the CI of image scores from the same blur level has significant overlap, the score of the image with the least shear has no overlapping CI with that of the most shear.

Second, we explore the inter-relationship between motion blur and shear. Let the score of a test image be (s,b), where s is the shear level, b is the blur level, and $s,b = 1,2,3,4$. We find that (s,b) often has similar value with $(s-1,b+1)$. For instance, in content “autumn Hovde Hall", (2,1) and (1,2) has respective scores -0.89 ± 0.31 and -0.59 ± 0.42 with no significant difference. (4,2) is a little better than (3,3), with scores -3.36 ± 0.34 and -4.49 ± 0.34, respectively.

To model the overall quality measure of an image with blur plus shear, we propose a mapping function to combine LVI and shear. The overall quality is modeled as:

$$Q(k, q_{LVI}) = q_{LVI} \cdot (1 - g \cdot exp(q_{LVI} - 1) \cdot k^2) \quad (6)$$

where k is the shear in equation (1), q_{LVI} is the LVI score of the distorted image; g is a constant parameter. The mapping by equation (6) is called shear-LVI.

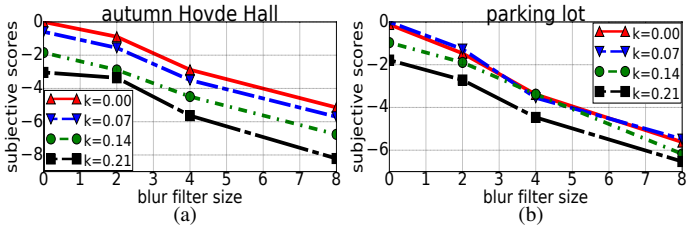

Figure 8: Subjective test - shear and blur: (a) autumn Hovde Hall (b) parking lot

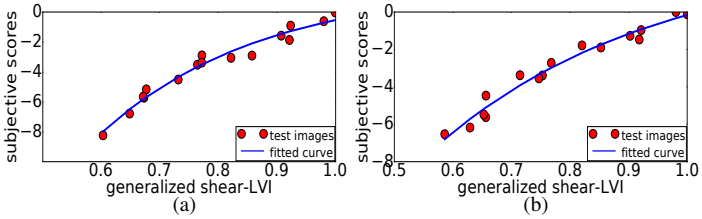

Figure 9: Curve fitted with logistic function between subjective scores and generalized shear-LVI (g=4.07): (a) autumn Hovde Hall (b) parking lot

To find the optimized value of g, we also maximize SROCC between shear-LVI scores and subjective scores. The optimized values of g are 5.34 and 2.59 for “autumn Hovde Hall" and “parking lot", respectively. To generalize equation (6) for all content without dramatic influence on SROCC and KROCC, g is experimentally chosen to be 4.07. Figure 9 shows the nonlinear mapping curve between the generalized shear-LVI and subjective scores. The fitted logistic function used is equation (4).

By using the same replacement as equation (5), we can extend equation (6) to to other quality metrics. The performances of 7 NR QEs and LVI are compared in Table 8 and 9. “shear-" indicates that the QE score is mapped by equation (6) with corresponding optimized g. JNBM, CPBD, IL-NIQE, NIQE and LVI improve their performances after mapping by equation (6). The generalized shear-LVI (g=4.07) shows competitive performance compared to other 7 QEs after mapping.

Fisheye

Test images distorted with fisheye and motion blur are synthetically created from one reference image. Figure 10 shows the subjective scores from two content “parking garage" and “apartment building", with average range of CIs to be 0.30 and 0.31.

Variations in quality due to different blur levels are stronger than those due to differences in the degree of fisheye. Specifically in Figure 10(a), the variances of subjective scores for 4 levels of blur with same fisheye are 5.56, 6.29 and 5.48, while the variances for 3 levels of fisheye with the same blur are 0.09, 0.11, 0.08 and 0.062. In Figure 10(b), the variances of subjective scores for 4 levels of blur with the same fisheye distortion are 7.40, 5.00 and 5.85, while the variances for 3 levels of fisheye with the same blur are 0.00, 0.06, 0.04 and 0.27. We can also draw the same conclusion from the CI of scores. For example, in content “parking garage", the scores of the 3 levels of fisheye in blur level 3 are -2.98 ± 0.43, -3.61 ± 0.16 and -3.56 ± 0.05, which have overlapping regions. In content “apartment building", the scores of the 3 levels of fisheye in blur level 3 are -3.12 ± 0.38, -3.08 ± 0.12 and -3.48 ± 0.19 with no significant difference. As a comparison, in both contents, the lowest score in blur level s is statistically

QE name	g	SROCC	KROCC	PLCC
JNBM	-	0.9235	0.8000	0.8853
BIQI	-	0.5706	0.3166	0.7395
CPBD	-	0.7529	0.5166	0.7787
BRISQUE	-	0.6441	0.5500	0.7011
CORNIA	-	0.2235	0.1333	0.2270
IL-NIQE	-	0.9294	0.8167	0.9283
NIQE	-	0.9382	0.8333	0.8734
LVI	-	0.7108	0.4602	0.8456
shear-JNBM	5.88	0.9735	0.9000	0.9470
shear-CPBD	23.38	0.9471	0.8333	0.9016
shear-IL-NIQE	-6.78	0.9735	0.9000	0.9601
shear-NIQE	-10.09	0.9647	0.8667	0.9528
shear-LVI	5.34	**0.9912**	**0.9500**	**0.9672**
shear-LVI	4.07	0.9853	0.9333	0.9694

Table 8: QE performances: shear - autumn Hovde Hall

QE name	g	SROCC	KROCC	PLCC
JNBM	-	0.8353	0.6333	0.8885
BIQI	-	0.6411	0.4000	0.6547
CPBD	-	0.8147	0.6167	0.8473
BRISQUE	-	0.8353	0.7000	0.8710
CORNIA	-	0.1529	0.1000	0.1963
IL-NIQE	-	0.9088	0.7667	0.9580
NIQE	-	0.9706	0.8667	0.9669
LVI	-	0.8992	0.7113	0.9481
shear-JNBM	2.46	0.9765	0.8833	0.9779
shear-CPBD	11.53	0.9735	0.9000	0.9189
shear-IL-NIQE	-1.7	0.9500	0.8333	0.9686
shear-NIQE	-0.37	0.9735	0.8833	0.9697
shear-LVI	2.59	**0.9853**	**0.9333**	**0.9858**
shear-LVI	4.07	0.9794	0.9000	0.9750

Table 9: QE performances: shear - parking lot

greater than the highest score in blur level $s+1$, when $s = 2,3,4$.

Not apparent from Figure 10, however, a personal preference exists for fisheye, and that preference is content dependent. The preference is extracted based on the percentage of times that the participant chose images with smaller fisheye levels in the same blur level. For content “parking garage", Figure 11 shows a comparison of subjective scores between two group of subjects: 35 people prefer non-fisheye while the remaining 15 prefer fisheye. Figure 12 shows that 33 people prefer non-fisheye while another 17 prefer fisheye images for content “apartment building". Among all subjects in the test, 23 prefer non-fisheye and 5 prefer fisheye for both contents, and 22 people show different preferences for the two scenes.

“Apartment building" has a relatively close view and higher spatial information (SI) compared to “parking garage". From post-test feedback, the bend of the scene around the edges in the former is not as obvious as in the latter. This feedback indicates the field of view of the content influences the viewing quality of a fisheye image. Some participants also indicate that the broad view of fisheye images could convey more information about the scene compared to non-fisheye images, especially for “apartment building".

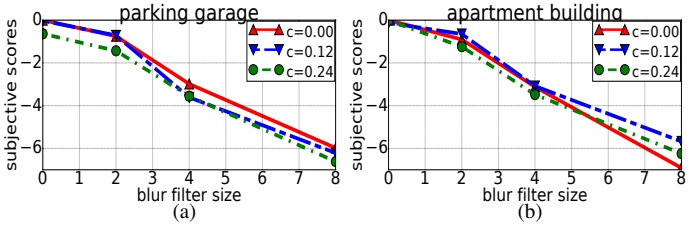

Figure 10: Subjective test - fisheye and blur: (a) parking garage (b) apartment building

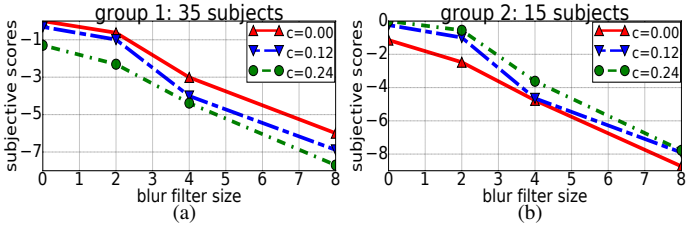

Figure 11: Parking garage: group 1 prefers non-fisheye, group 2 prefers fisheye

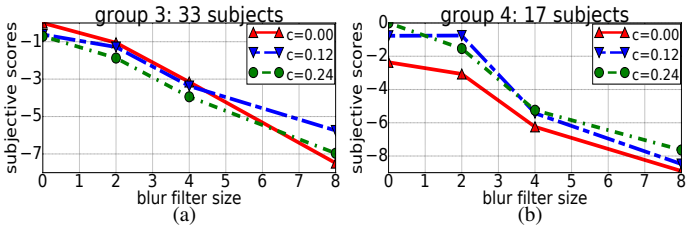

Figure 12: Apartment building: group 1 prefers non-fisheye, group 2 prefers fisheye

Conclusion

We design a subjective test using paired comparison to evaluate distortions in FPVs. We use real images and images with multiple distortions (motion blur and geometric distortions). We propose rotation-LVI and shear-LVI to estimate the quality of images with geometric distortions. The mapping functions for rotation and shear are also successfully tested on other NR QEs.

All subjective results of geometric distortions are influenced by content. Shear is less sensitive to different contents compared to rotation and fisheye. To get a generalized model for rotation, subjective data of more content is needed. Results also show the existence of a personal preference: one group of people prefers fisheye, another group prefers non-fisheye. Furthermore, the preference somewhat influenced by content. To explore the personal preference in fisheye, additional content should be considered.

References

[1] Chen Bai and Amy R. Reibman. Characterizing distortions in first-person videos. In IEEE International Conference on Image Processing (ICIP), 2016.

[2] Alejandro Betancourt, Pietro Morerio, Carlo S Regazzoni, and Matthias Rauterberg. The evolution of first person vision methods: A survey. IEEE Transactions on Circuits and Systems for Video Technology, 25(5), 2015.

[3] Michael S. Ryoo and Larry Matthies. First-person activity recognition: What are they doing to me? In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2013.

[4] Xiong Bo and Kristen Grauman. Detecting snap points in egocentric

video with a web photo prior. In European Conference on Computer Vision, pages 282–298, 2014.

[5] Yair Poleg, Tavi Halperin, Chetan Arora, and Shmuel Peleg. Egosampling: Fast-forward and stereo for egocentric videos. In IEEE Conference on Computer Vision and Pattern Recognition, pages 4768–4776, 2015.

[6] Zhongwei Tang, Rafael Grompone von Gioi, Pascal Monasse, and Jean-Michel Morel. High-precision camera distortion measurements with a calibration harp. J. of the Optical Society of America A, 29(10), 2012.

[7] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans. Image Proc., 15(11):3440–3451, Nov. 2006.

[8] E. C. Larson and D. M. Chandler. Most apparent distortion: Full-reference image quality assessment and the role of strategy. J. of Electronic Imaging, 9(1), March 2010.

[9] Ponomarenko Nikolay et al. Image database TID2013: Peculiarities, results and perspectives. Signal Processing: Image Communication, pages 57–77, 2015.

[10] D Michele A. Saad, M. Pinson et al. Image quality of experience: A subjective test targeting the consumers experience. In Human Vision and Electronic Imaging, February 2016.

[11] Dinesh Jayaraman, Anish Mittal, Anush K Moorthy, and Alan C Bovik. Objective quality assessment of multiply distorted images. In Forty-Sixth Annual Asilomar Conference on Signals, Systems, and Computers, pages 1693–1697, 2012.

[12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Proc., 13(4):600–612, April 2004.

[13] H. R. Sheikh and A. C. Bovik. Image information and visual quality. IEEE Trans. Image Proc., 15(2):430–444, Feb. 2006.

[14] Zhou Wang and Eero P Simoncelli. Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. In Human Vision and Electronic Imaging, Jan. 2005.

[15] Qiang Li and Zhou Wang. Reduced-reference image quality assessment using divisive normalization-based image representation. IEEE J. Sel. Top. Signal Process, 3(2), 2009.

[16] Anish Mittal, A. Krishna Moorthy, and Alan C. Bovik. No-reference image quality assessment in the spatial domain. IEEE Transactions on Image Processing, 21(12), 2012.

[17] Lei Zhang Zhang Lin and Alan C. Bovik. A feature-enriched completely blind image quality evaluator. IEEE Transactions on Image Processing, 24(8), 2015.

[18] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a completely blind image quality analyzer. IEEE Signal Processing Letters, 20(3), 2013.

[19] S. S. Hemami and A. R. Reibman. No-reference image and video quality estimation: Applications and human-motivated design. Signal Processing: Image Communication, Aug. 2010.

[20] Jing Li, Marcus Barkowsky, and Patrick Le Callet. Analysis and improvement of a paired comparison method in the application of 3DTV subjective experiment. In IEEE International Conference on Image Processing (ICIP), 2012.

[21] Jonathan W Peirce. Psychopypsychophysics software in python. Journal of neuroscience methods, 162(1):8–13, 2007.

[22] ANSI T1.801.03. Digital transport of one-way video signals parameters for objective performance assessment. In American National Standards Institute, 1996.

[23] Giacomo Boracchi and Alessandro Foi. Modeling the performance of image restoration from motion blur. IEEE Transactions on Image Processing, 21(8), 2012.

[24] W Neale, David Hessel, and Toby Terpstra. Photogrammetric measurement error associated with lens distortion. Accident Reconstruction, 2011.

[25] John C. Handley. Comparative analysis of Bradley-Terry and Thurstone-Mosteller paired comparison models for image quality assessment. In PICS, 2001.

[26] R. Ferzli and L. J. Karam. A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB). IEEE Trans. Image Proc., 18(4):717–728, April 2009.

[27] A. K. Moorthy and A. C. Bovik. A two-step framework for constructing blind image quality indices. IEEE Signal Processing Letters, 2010.

[28] N. Narvekar and L. J. Karam. An improved no-reference sharpness metric based on the probability of blur detection. In Wkshp. on Video Proc. and Quality Metrics, January 2010.

[29] Le Kang David Doermann Peng Ye, Jayant Kumar. Unsupervised feature learning framework for no-reference image quality assessment. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1098–1105. IEEE, 2012.

[30] He Liu and Amy R. Reibman. Software to stress test image quality estimators. In Quality of Multimedia Experience (QoMEX), 2016.

[31] Peter Bex. Sensitivity to spatial distortion in natural scenes. Journal of Vision, 8(6):688–688, 2008.

Author Biography

Chen Bai received his BS in engineering from the Huazhong University of Science and Technology (2014) and now is a PhD student in Electrical and Computer Engineering at Purdue University. His research focuses on image and video quality assessment.