Detecting Good Quality Frames in Videos from Mobile Camera for Blind Navigation

Long Tian, Chucai Yi, and Yingli Tian, *Senior Member, IEEE*

***Abstract*—The development of smart mobile device with cameras makes it possible to build wearable and portable blind assistant navigation systems. However, it is difficult for blind user to capture high-quality images and videos of surrounding environments without motion blur or de-focus blur. To avoid this blur effect in camera-based videos, this paper proposes a method of high-quality frame detection based on image quality assessment. It distinguishes blurred frames from unblurred frames in the videos. In this method, an image frame is divided into 80×70 patches, where edge and luminance information is modeled for high-quality frame detection. 4 types of features based on gradient statistic are adopted to determine whether an image patch is blurred or not. The classification process is carried out in a support vector machine (SVM) based learning model. The unblurred frames are used to extract essential information for navigation and information collection such as scene text information. We collect a video dataset of natural scenes containing blurred and unblurred frames in both indoor and outdoor environments. Experimental results demonstrate that our proposed method is able to robustly handle video motions and extract surrounding text and signage in wearable and portable blind assistant navigation systems.**

***Index Terms*— Blind assistive navigation, Way-finding, Motion blur, Video quality, Frame selection, Scene text extraction.**

I. Introduction

WITH the development of mobile cameras in the form of smartphones, laptops, wearable cameras, it becomes very convenient to capture surrounding scenes and objects. Based on the camera-captured image or video, we are able to extract valuable information from surrounding environments. This information can be used in many practical applications, such as assistant navigation of blind or visually impaired people.

According to the World Health Organization investigation in 2010 [1], 4.24% of the world's total population is suffering visual impaired and 0.58% persons are blind. This number is very likely to increase as the baby boomer generation gets into old ages. In the daily life of blind people, one challenging task is to access unfamiliar environments. Most of them fell into trouble in finding the correct ways, and were even thrown into life threatening situations. If the blind or visually impaired people were able to access, understand, and explore unfamiliar environments, they would live a much better life, since this capability could enhance employment opportunities, foster independent living, and produce economic and social self-sufficiency.

To help visually impaired people, the wearable and portable cameras are used to develop many assistant systems over the last fifty years, including video magnification [10], reading machines [18], text-to-speech (TTS) and screen readers [2], sonic travel guide [14, 17]. There are many camera-based aids such as signage recognition [39], text extraction [43], bill recognition [12], and navigation [7, 16, 20, 23, 24, 36]. The research work in [37] considered which features were necessary to make 3D visual worlds usable for blind or visually impaired persons. A haptic indicator [3] was developed to help visually impaired persons find their ways, by delivering simple navigational information. The work in [40] attempted to design universal auditory graphs to help visually impaired and sighted listeners. A multimodal video game was developed in [29] to train navigation skills of blind children. These technologies were proposed to increase quality of life for millions of individuals with vision loss by allowing their independent access into unfamiliar environments. Recent technical developments in computer vision, digital cameras, and portable computers make it possible to develop better products to assist blind people. Our group has worked out a number of solutions, including robust scene text information extraction, for assistive technologies to help blind people [12, 36, 42]. However, the proposed methods of information retrieval from surrounding environments mainly focus on clear static images or videos without blur, so it limits practical applications of blind assistance.

Even though the wearable and portable cameras are available for people in normal vision, it is difficult for blind or visually impaired people to capture high-quality images that are well prepared for computer-vision-based object detection or recognition. In blind assistant systems, blind users are usually required to capture a video clip of surrounding environment, and a frame in good quality should be selected from the video clip for blind assistant information retrieval. In this paper, the quality is defined as the blurry extent of the image. The less the frame is blurred, the better its quality is.

Blur effect, including motion blur and de-focus blur, is a common problem which is usually encountered by the users of mobile cameras. They are mainly caused by the improper

This work was supported in part by NSF grant EFRI-1137172, IIP-1343402, and FHWA DTFH61-12-H-00002.

Long Tian is with Mississippi State University. This work was done when he was with the City College, City University of New York, New York, NY 10031 USA (e-mail: lt766@msstate.edu).

Chucai Yi is with Here North America LLC. This work was done when he was with the Graduate Center, University of New York, New York, NY 10031 USA (e-mail: gschucai@gmail.com).

YingLi Tian is with the City College and Graduate Center, City University of New York, New York, NY 10031 USA (phone: 212-650-7046; fax: 212-650-8249; e-mail: ytian@ccny.cuny.edu)

camera motion, camera shake or inaccurate camera shooting direction, all of which are unavoidable when a blind or visually impaired user wears the camera. People in normal vision can accurately aim at the target within the camera view and take the image while keep the camera static, but blind people cannot see the target and they are not able to keep the camera stable while taking images. This blur will significantly decrease the performance of surrounding information retrieval.

No existing blind assistant systems can handle motion blur and extract surrounding text and signage for Way-finding. To address this issue, we proposed a method to detect good quality frames from video clip [36] and then detect text information from surrounding signage [42]. This paper will combine the two techniques to generate a novel blind assistant system working in practical applications. Figure 1 shows our prototype system for blind-assistant navigation in unfamiliar environment, including a wearable camera mounted on a sun glasses, a mini computer for data processing, a microphone for speech command, and a Bluetooth ear piece for providing feedback to the blind user. Some example images captured by a wearable camera, as shown in Figure 2, are blurred due to improper camera motion and camera shake.

Figure 1. Testing our prototype system with a wearable camera for blind navigation.

In [36], we proposed a method to detect good quality frames from blurred frames in videos captured by wearable cameras while the blind users are moving. Our method was able to handle both indoor and outdoor environments. It combined gradient features and statistical features of frequency, entropy, etc. Then the SVM learning model was applied to distinguish the frames in good quality (unblurred) from those in low quality (blurred) frames. This process of image quality assessment is defined as high-quality frame detection in this paper. The unblurred frames would be further processed to extract essential information such as sign and text from surrounding environment. In this paper, our previous method is improved by dividing the whole frame into a group of patches and performing high-quality frame detection in each patch. This novel method can not only speed up the calculation time, but also increase the accuracy of high-quality frame detection. More details can be found in Section 4.

This rest of this paper is organized as follows. In Section 2, we describe the related work of image quality assessment. Section 3 provides an overview of the proposed method. Section 4 explains the features and classification to identify blurred and unblurred frames from video sequences. The experimental results and evaluation are presented in Section 5. Section 6 demonstrates some results of the text detection from blurred and unblurred images. Section 7 concludes the paper and discusses our future work.

Figure 2. Low-quality images captured by a wearable camera and they are blurred. It will be difficult to extract necessary information from these blurred images for navigation such as text and signage etc.

II. Related work

There are three types of methods for the image quality assessment including full-reference based [5, 30], reduced-reference based [19] and no-reference based methods [6, 8, 9, 11, 22, 26, 34, 41]. For the full-reference based algorithms, the original sharp image with good quality was used as the reference image to compare with the test images for the quality difference. In reduced-reference methods, instead of using the whole image, part of information in original image was used as reference information. Recently, researchers have developed no-reference algorithms without using any knowledge of the original high quality images. Our task is to detect blur frames in video, and there is no original high quality frames available. Therefore our method belongs to no-reference method.

The most popular features for non-reference image quality assessment are the edge features [9, 26]. Usually, the edge width of blurred images is larger than the edge width from clear images with good quality. In addition to edge width, gradient and slopes are also used for image quality assessment [8]. Furthermore, frequent space, transform-based and hybrid metrics are investigated by some researchers [13, 28, 38]. Considering the human visual perception system for edge changes, Ferzli and Karam proposed an algorithm based on cumulative probability of blur detection (CPBD) by introducing a measurement called "just noticeable blur (JNB)" to indicate the tolerance of human visual system for edge width changes [9]. Mittal *et al.* demonstrated that the statistics of all pixels illumination in an image can be applied for the image quality assessment [22]. Ruderman proposed mean subtracted contrast normalized (MSCN) coefficients for image representation that is the other form of normalized luminance coefficients [27]. The image quality distortion affects the MSCN coefficients distribution function, which will be quantized with the parameter change in a distribution function. This blind and reference-less image spatial quality evaluator (BRISQUE) is efficient enough to be applied to the real-time image quality assessment.

Although many methods have been developed, the accuracy of the existing methods is not sufficiently high for our high-quality frame detection task in blind navigation application. In this paper, we propose a new method for high-quality frame detection by combining the edge, luminance and statistical information.

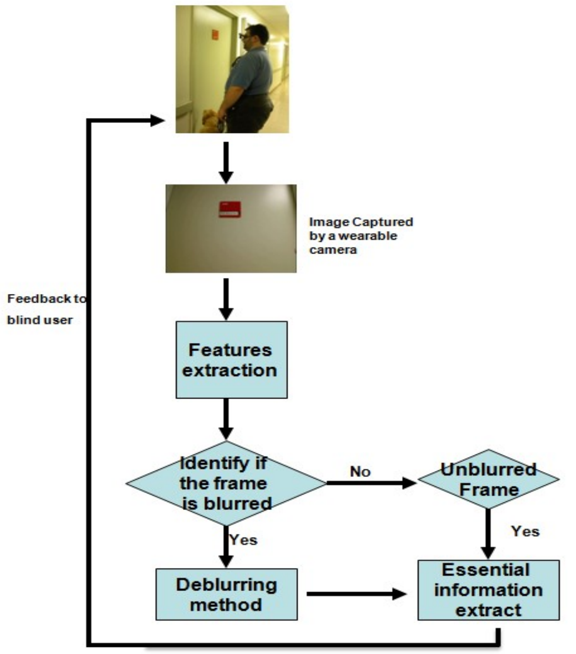

Figure 3. The flowchart of the proposed method for detecting good quality frames from blurred frames from videos captured by a wearable camera.

III. Overview of The Proposed High-quality Frame Detection Method

The flowchart of our proposed method is shown in Figure 3. It demonstrates the process from camera-based video frames to quality identification. As we observe, the edges in blurred frames are always wider than the edges in sharp frames. Therefore, edge information is employed to generate features to detect blurred frames and unblurred frames. For each frame, we first extract edge and statistic features. Then an SVM-based classifier is employed to determine if the frame is blurred or not. If the frame is unblurred, the essential information (text signage) in this frame is extracted and provided to blind users as feedback in speech. If all the frames in a short period time are blurred, a deblur process based on maximum a posterior and blind deconvolution framework will be applied to restore the frames and the essential information will then be extracted from the deblurred image. The deblur method is beyond the topic of this paper.

In the process of high-quality frame detection, specific patches are cropped from image for feature extraction, instead of the whole image. The specific patches will be less computationally expensive and the work in [15] pointed out that the blank background and no-edge patch could adversely affect the accuracy of high-quality frame detection. Taking Figure 4 (a) for example, the blank background contains no information on blur, and these parts would unexpectedly reduce the average edge width if we adopt the whole image for high-quality frame detection.

In our experiments, the image patch is defined to be 80×70 sizes, and it moves all over the image. The edge and luminance information of each patch will be calculated. Then, the patch containing the largest number of edges will be selected for the edge processing. Next, the patch that contains the highest luminance value will be chosen for the luminance processing later.

According to our daily life experience, the central part of an image catches our much more attention than the margin of the image, as shown in Figure 4 (b). Therefore, a weight is set in each position according to its distance from image center, and this weight is applied to the edge score for each patch in image. In our definition, the central patch weight will be 1 and the margin weight will be 0.01, while the weights of other parts are set along with position linearly.

IV. Blurred and Unblurred Image Classification

A. *Features for Blurred and Unblurred Image Classification*

Unblurred frames are distinguished from blurred frames in the video clips captured by wearable cameras. In this process of high-quality frame detection, we adopt 4 types of features on the basis of edge and luminance information: (1) average edge width, (2) average luminance, (3) cumulative edge blur probability, and (4) MSCN distribution [22]. This section will present the 4 types of features in detail.

Figure 4. The weak factor from blank patches from blurred frame (a), and (b) non-central patch in unblurred frame.

Average Edge Width: As shown in Figure 5, although the edges in blurred frames are much less than the unblurred frames, the width of edges in blurred frames are usually wider than that from frames with good quality. Therefore, edge width can be used to measure the image quality. To calculate the average edge width, we first apply a Sobel edge detector to the image. Then, in the perpendicular direction of the edge gradient angle, the edge width for each edge pixel is calculated by the distance between the start and end positions of the edge pixel, which are defined as the local extreme locations closest to the edge pixel [25]. At last, the average value of width of each edge pixel for the whole frame is calculated as the average edge width feature. As shown in the top-right images of Figure 5(a) and 5(b), the maps of edge width distribution demonstrate that the maximum edge width for the blurred image (Figure 5(a)) is about 30 pixels, while the edge width for the unblurred image (Figure 5(b)) for most pixels is smaller than 15 pixels. Therefore, the average edge width is a good feature to classify unblurred and blurred images.

Average Luminance: Edge-based features are necessary but not sufficient conditions to distinguish high-quality sharp images from blurred images. Given two images, one from indoor environment with sharp edges and the other from outdoor environment with blur, it is very possible that edge-based features always predict that the outdoor image has higher quality, because it should contain much more edge pixels than the indoor one. To eliminate the edge-based difference of indoor and outdoor environments, average luminance is modelled to differentiate them in the estimate of image quality. Generally, the average luminance of video frames for an outdoor environment in day time is larger than an indoor environment, because the distance between the camera and objects in outdoor environments is normally larger than that in indoor environments. Therefore, the feature of average luminance is applied in our method to model the environment luminance situation.

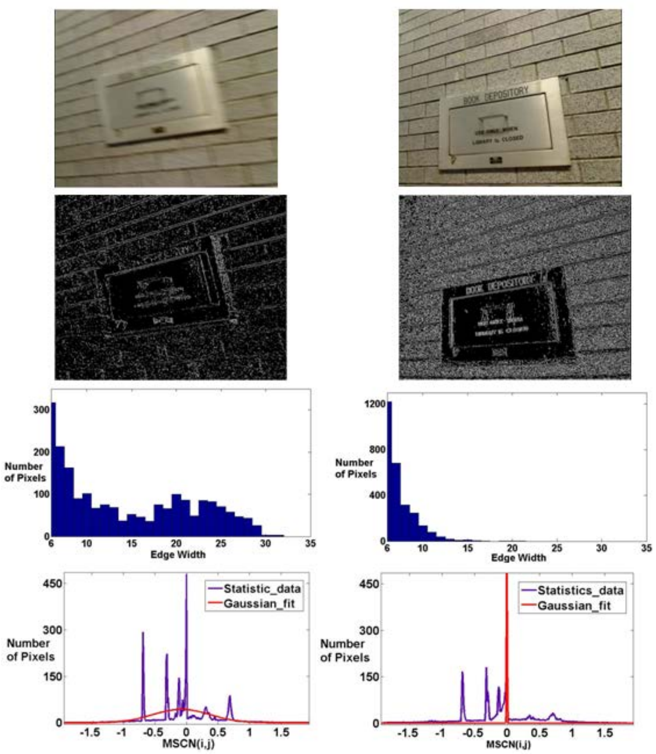

Figure 5. The edge width and MSCN distribution for the blurred (left column) and unblurred (right column) frames. In the two examples, from top to bottom, there are original images; edge width distributions, edge images, and MSCN distributions.

Cumulative Edge Blur Probability: In addition to the global edge and luminance information, the distribution of edge pixels in an image also plays an important role in the human perception of image quality. Thus we divide an image into patches and assign them different weights according to their edge densities in the estimate of image quality. At first, the image will be uniformly divided into patches (the patch size is selected as 64×64 in our implementation). If the patch contains enough number of edges pixel (0.2% of the total pixels in patch), we will label it as 'edge patch' and process it in later steps. Then, the probability that models the blur extent at each edge in 'edge patch' is calculated. At last the edge values of the whole image are pooled by the calculating the cumulative edges blur probability of image for high-quality frame detection. More details of the calculation for the cumulative edge blur probability can be found in [26].

The Mean Subtracted Contrast Normalized (MSCN) Luminance Distribution: Ruderman's work [27] demonstrates that the distribution of MSCN coefficients for an image is close to a generalized Gaussian distribution model (GGD). For the diagonal neighboring pixels in MSCN luminance image, the distribution follows an asymmetric generalized Gaussian distribution (AGGD) model [22]. In the bottom-right images of Figure 5(a) and 5(b), it is observed that the distributions of MSCN have very different distributions between blurred and unblurred frames. As shown in the following section, the parameters of GGD and AGGD are employed to measure the blur level of images.

B. Computation of MSCN Distributions

The mean subtracted contrast normalized (MSCN) luminance for pixel (i,j) is $\hat{I}(i,j)$ which was defined in [22] as Eq. (1)

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C}, \tag{1}$$

where $\mu(i,j)$ is the local mean of position (i,j), calculated from a 3×3 neighborhood around it, and $\sigma(i,j)$ denotes the standard deviation of $I(i,j)$within the block.

$$\mu(i,j) = \sum_{k=-1}^{1} \sum_{l=-1}^{1} w_{k,l}\, I_{k,l}(i,j)\,, \tag{2}$$

$$\sigma(i,j) = \sqrt{\sum_{k=-1}^{1} \sum_{l=-1}^{1} w_{k,l}\, \left(I_{k,l}(i,j) - \mu(i,j)\right)^2}\,, \tag{3}$$

where $w_{k,l}$ is a 2D circularly-symmetric Gaussian weighting function sampled out to 3 standard deviations and rescaled to unit volume. The GGD model with zero mean is

$$f(x;\alpha_1;\sigma^2) = \frac{\alpha_1}{2\beta\Gamma\left(\frac{1}{\alpha_1}\right)} exp\left(-\left(\frac{|x|}{\beta}\right)^{\alpha_1}\right), \tag{4}$$

where $\beta = \sigma\sqrt{\frac{\Gamma(1/\alpha_1)}{\Gamma(3/\alpha_1)}}$ and $\Gamma(\cdot)$ is the gamma function. For neighbouring MSCN coefficients, the AGGD model is:

$$f(x;\alpha_2;\sigma_l^2;\sigma_r^2) = \begin{cases} \dfrac{\alpha_2}{(\beta_l + \beta_r)\Gamma\left(\frac{1}{\alpha_2}\right)} exp\left(-\left(\dfrac{-x}{\beta_l}\right)^{\alpha_2}\right) & x < 0 \\ \dfrac{\alpha_2}{(\beta_l + \beta_r)\Gamma\left(\frac{1}{\alpha_2}\right)} exp\left(-\left(\dfrac{-x}{\beta_r}\right)^{\alpha_2}\right) & x \geq 0\,, \end{cases} \tag{5}$$

where $\beta_l = \sigma_l\sqrt{\Gamma\left(\frac{1}{\alpha_2}\right)/\Gamma\left(\frac{3}{\alpha_2}\right)}$ and $\beta_r = \sigma_r\sqrt{\Gamma\left(\frac{1}{\alpha_2}\right)/\Gamma\left(\frac{3}{\alpha_2}\right)}$.

The parameter σ indicates the GGD variance and α_1 controls the GGD shape. We employ (σ,α_1) and $(\sigma_l,\sigma_r,\alpha_2)$ as our features, where σ_l indicates the left side AGGD variance, σ_r respects the right side AGGD variance and α_2 controls the AGGD shape.

C. SVM-based Blurred and Unblurred Image Classification

There are 4 types of features involved in our method of high-quality frame detection, and each of them has a specific range. For example, the average edge width value is from 5 to 30, while the average luminance is around 50. To ensure each feature to make equivalent contribution to the learning process, we normalize them into the range from 0 to 1.

SVM-based learning model is adopted to classify blurred from unblurred image. The training dataset contains samples represented by the features and labels (x_i, y_i) where x_i denotes the designed features and $y_i \in \{-1,1\}$ denotes the label of categories. In this paper, the C-SVM classification is applied and it should complete the solution for the optimization problem.

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_i \xi_i \, , \tag{6}$$

subject to

$$y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \text{ and } \xi_i \geq 0 \, , \tag{7}$$

where ξ_i is non-negative slack variables, which measure the degree of misclassification of the data x_i, w is the weight vector, b is the bias and C is per-chosen parameter which will affect the accuracy performance of SVM.

The kernel function is $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$, where ϕ is the function to map training vectors x_i into a high dimensional space. The radial basis function (RBF) is applied as kernel function because it was found to be the most effective function.

$$K(x_i, x_j) = exp\left(-\gamma \|x_i - x_j\|^2\right) . \tag{8}$$

The SVM classifier needs RBF kernel function with two parameters (C, γ), which needs to be adjusted for the small number of support vectors that will reduce the calculating time.

V. IMPROVEMENT OF BLIND NAVIGATION AND WAY-FINDING BY BLUR DETECTION

Signage with text is the most reliable labels for destination recognition in the application of blind navigation. Text information in natural scene is generally surrounded by all kinds of background outliers. To help blind users find their destinations through surrounding text information in an unfamiliar environment, we have proposed several text detection algorithms to find the text from scene images [42]. In this text detention algorithms, color uniformity and linear alignment of text characters and strings are regarded as two significant layout conditions of scene text. They are used for building layout and structural models of scene text, and separating text characters and strings from background outliers that do not satisfy the two conditions.

Although our proposed algorithm of scene text extraction is able to handle text strings in different fonts, sizes, and colors, they cannot obtain good performance from low-resolution text signage in blurred images. It means that the quality of captured images is a bottleneck of scene text extraction in vision-based blind assistance.

Figure 6. Example results of text information extraction from blurred images (left column) and unblurred images (right column). It shows that the scene text extraction results are more reliable on high-quality frames without blur effect.

The method described in this paper can be used as a pre-processing component for assistive blind navigation system. To demonstrate the benefits of our proposed method in image-based information retrieval, we apply scene text detection to the low-quality blurred images and the good quality images selected by the proposed method. Some results are shown in Figure 6. The localized text regions are marked by blue rectangle boxes, which will be further processed for text recognition. We observe that more true positive text regions are successfully detected in the good quality images.

To build a wearable and portable blind assistant system, we have transplanted our method of scene text extraction into Android platform for a prototype demo system of blind assistance. As shown in Figure 7, the demo system runs on a Samsung Galaxy smart phone with Android platform and proves the feasibility of our wearable and portable blind-assistant navigation system. It also provides us some insights into algorithm design and performance improvement of blind-assistant navigation. First, blur detection plays an important role in scene text extraction in wearable/mobile devices which is very likely to generate low-quality image frame. Second, in scene text extraction, we focus only on text strings in approximately horizontal orientation, because most text strings in natural scene are horizontal and the wearable cameras can be adaptively rotated to fit the horizontal strings from different view angles. Third, in blind-assistant systems, we can simultaneously perform video capture and information retrieval in two independent modules to accelerate the processing speed.

Figure 7. The demo system of text extraction runs on a smart phone with Android platform.

VI. EXPERIMENTS

A. Database

To evaluate the proposed method as shown in Figure 3, we develop a blind assistant prototype system including a Logitech HD webcam mounted on a pair of sunglasses and a mini laptop for data analysis. Videos are captured from both outdoor and indoor environments while the user is moving around. Then it is decomposed into a sequence of frames. Among these frames, we can obtain both blurred frames and unblurred frames.

We totally capture 22 videos including 10 outdoor videos and 12 indoor videos, where the resolution is 1240×1024 pixels. The blurred and unblurred frames are manually labeled for the algorithm evaluation based on human visual perception. The frames we selected from these videos for training and testing mainly consist of the meaningful information, such as entrances, characters, text and signage information which could be valuable for blind person. Table I shows that the total number of frames selected from all the captured videos as blurred and unblurred group. The total number of frames for our method is 3138, including 1802 blurred frames and 1336 unblurred frames. For the training group, 1652 frames are selected, and the testing group contains 1486 frames.

TABLE I: NUMBER OF FRAMES OF BLURRED AND UNBLURRED FOR TRAINING AND TESTING

Data	Training Group		Testing Group	
	Blurred	*Unblurred*	*Blurred*	*Unblurred*
Indoor	469	332	491	496
Outdoor	421	430	421	258

TABLE II: PERFORMANCE COMPARISON OF OUR METHOD AND PREVIOUS METHODS OVER BLURRED AND UNBLURRED IMAGES

	Blurred	Unblurred	Total
CPBD	69.40%	78.30%	74.80%
BRISUE	61.50%	78.30%	69.40%
Method in [36]	82.30%	**90.50%**	86.20%
Our Method	**91.10%**	88.70%	**89.70%**

B. Experimental Results and Analysis

As mentioned in Section IV.A., 4 types of features are extracted from each frame, which are average edge width, average luminance, cumulative edge blur probability, and MSCN luminance distributions. The CPBD [9], BRISQUE [27] algorithms and our previous method in [36] are adopted in our experiments for performance comparisons. CPBD method has the best performance among the non-SVM high-quality frame detection in videos, BRISQUE could perform high-quality frame detection with SVM in real time and our previous method has the best performance with SVM. The method presented in [44] is able to extract unblurred regions from out-of-focus outdoor images, but does not work as well on indoor images with motion blur as out-of-focus images.

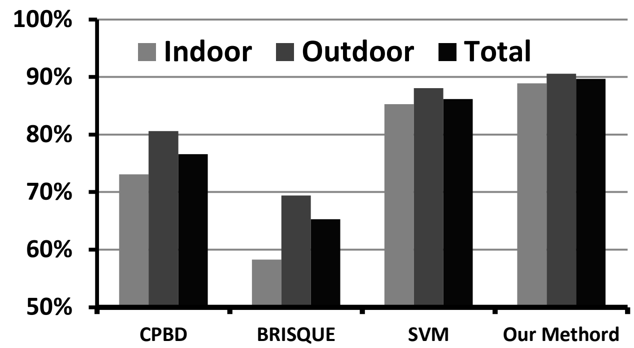

Figure 8. The blurred detection accuracy in outdoor and indoor environment.

TABLE III: PERFORMANCE COMPARISON OF OUR METHOD AND PREVIOUS METHODS OVER INDOOR AND OUTDOOR IMAGES.

	Blurred	Unblurred	Total
CPBD	73.10%	80.60%	76.60%
BRISUE	58.30%	69.40%	65.30%
Method in [36]	85.30%	88.10%	86.20%
Our Method	**88.90%**	**90.60%**	**89.70%**

The performance comparison of our method with the state-of-the-art methods is displayed in Table II. The C-SVM is applied as Eq. (6) and Eq. (7), and the parameters C and γ are set as 30 and $1/7$ respectively. Our method outperforms both CPBD and BRISQUE algorithms. In comparison with the CPBD method, the average accuracy of our method is about 15% higher accuracy, and the accuracy for the unblurred detection (good quality) frames is 22% higher while the accuracy in blurred frame detection is around 11% better, and there are around 4% higher result when it is compared with our previous method (SVM with whole frame information).

According to our results, the accuracy of detecting blurred frames is 2.4% higher than that of detecting unblurred frames. We infer that this is caused by the different sensitive of human visual system for blurred and unblurred frames. Usually, human visual system is more sensitive to the variance of blurred level, but it is hard to tell the different of two unblurred frames. Thus it is easier to detect blurred frames based on the pre-labeled dataset by human vision. In our method, the patch that contains the highest edge number is chosen for the SVM processing. In this case, the distraction effect of blank patches could be reduced. On the other hand, selecting good quality (unblurred) frames is the main focus on our application to directly extract text information directly from them.

We further evaluate our method for indoor and outdoor environments respectively. The evaluation results are shown in Figure 8 and Table III. Compared with previous methods, our experimental results achieve the best performance in both the indoor environment (around 14% higher than CPBD, 30% higher than BRISQUE and 3% higher than SVM), and in the outdoor environment. For the outdoor environment, our method achieves the best performance that is 10% higher than CPBD, 22% higher than BRISQUE and 3% higher than SVM. For the indoor situation, the background luminance is relatively low, which could affect the contrast just notable blur judgment in CPBD algorithm. Our method combines the luminance features in our system that is able to overcome the limitations of CPBD method. On the other hand, we pick up the highest luminance patch for the SVM, which also improves our result.

VII. CONCLUSION

In this paper, we have proposed a robust method to handle the motion blur in blind navigation and way-finding systems by selecting good quality frames in videos. To improve the accuracy and robustness of high-quality frame detection, we extract features by combing edge information, average luminance, and luminance distributions in patch that contain useful information. The proposed method has been evaluated by the self-collected database and a primary navigation prototype system. The evaluation results demonstrate that our method outperforms the state-of-the-art methods. Our future work will focus on improving the accuracy, robustness, and efficiency of the proposed method, as well as the human interface study for blind users.

REFERENCES

[1] Factsheet of World Health Organization. Visual impairment and blindness. from http://www.who.int/mediacentre/factsheets/fs282/en/

[2] A. Arditi and A. Gillman. 1986. Computing for the blind user. *Byte*, vol. 11, no. 3, pp. 199–211.

[3] T. Amemiya, and H. Sugiyama. 2010. Orienting kinesthetically: A haptic handheld wayfinder for people with visual impairments. *ACM Trans. Access. Comput*. 3, 2, Article 6 (November 2010), 23 pages. http://doi.acm.org/10.1145/1857920.1857923.

[4] X. Chen J. Yang, J. Zhang and A. Waibel. 2004. Automatic detection and recognition of signs from natural scenes. *IEEE Transactions on Image Processing*, 13:87-99.

[5] Q. Chen, Y. Xu, C. Li, N. Liu, and X. Yang. 2013. An image quality assessment metric based on quaternion wavelet transform. *IEEE International Conference on Multimedia &Expo*.

[6] A. Ciancio, A. L. N. T. da Costa, E. A. B. da Silva, A. Said, R. Samadani and P. Obrador. 2009. Objective No-Reference Image Blur Metric Based on Local Phase Coherence. *Electronics Letters* 45(23): 1162-1163.

[7] Y. Danilov and M. Tyler. 2005. Brainport: an alternative input to the brain. *Journal of integrative neuroscience*, vol. 4, no. 4, pp. 537–550.

[8] C. Feichtenhofer, H. Fassold and P. Schallauer. 2013. A perceptual image sharpness metric based on local edge gradient analysis. *IEEE Signal Processing Letters*, 20: 379-382.

[9] R. Ferzli and L. J. Karam. 2009. A No-Reference Objective Image Sharpness Metric Based on the Notion of Just Noticeable Blur (JNB). *IEEE Transactions on Image Processing*, 18(4): 717-728.

[10] S. M. Genensky, P. Baran, H. Moshin, and H. Steingold. 1968. A closed circuit TV system for the visually handicapped.

[11] K. Gu, G. Zhai, X. Yang, W. Zhang and L. Liang. 2013. No-Reference image quality assessment metric by combining free energy theory and structural degradation model. *IEEE International Conference on Multimedia &Expo*, 2013.

[12] F. Hasanuzzaman, X. Yang, and Y. Tian. 2012. Robust and Effective Component-based Banknote Recognition for the Blind. *IEEE Transactions on Systems, Man, and Cybernetics--Part C: Applications and Reviews*, 42(6), 1021-1030.

[13] R. Hassen, Z. Wang and M. Salama. 2010. No-reference image sharpness assessment based on local pahse conherence measurement. *IEEE Int. Conf. Acoust., Speech and Signal Processing*.

[14] A. D. Heyes. 1984. The Sonic Pathfinder: A New Electronic Travel Aid. *Journal of Visual Impairment and Blindness*, vol. 78, no. 5, pp. 200–02.

[15] Z. Hu, M. Yang. 2012. Good regions to deblur. *Proceedings of European Conference on Computer Vision* (ECCV 2012) 5: 59-72.

[16] V. Ivanchenko, J. Coughlan, and H. Shen. 2008. Crosswatch: A camera phone system for orienting visually impaired pedestrians at traffic intersections. *Computers Helping People with Special Needs*, pp. 1122–1128.

[17] L. Kay. 1974. A sonar aid to enhance spatial perception of the blind: Engineering design and evaluation. *Radio and Electronic Engineer*, vol. 44, no. 11, pp. 605–627.

[18] R. Kurzweil. 1976. The Kurzweil reading machine: A technical overview. *Science, Technology and the Handicapped*, pp. 3–11.

[19] Q. Li and Z. Wang. 2009. Reduced-Reference Image Quality Assessment Using Divisive Normalization-Based Image Representation. IEEE Journal of Selected Topics in Signal Processing, 3(2): 202-211.

[20] R. Manduchi, J. Coughlan, and V. Ivanchenko. 2008. Search strategies of visually impaired persons using a camera phone wayfinding system. *Computers Helping People with Special Needs*, pp. 1135–1140.

[21] R. Manduchi and J. Coughlan. 2012. (Computer) vision without sight. *Communications of the ACM*, vol. 55, no. 1, pp. 96–104.

[22] A. Mittal, A. K. Moorthy and A. C. Bovik. 2012. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Transactions on Image Processing 21(12) 4695-4708*.

[23] P. Meijer. Seeing with sound -- the vOICe. [Online]. from http://www.seeingwithsound.com.

[24] "VizWiz - Take a Picture, Speak a Question, and Get an Answer." 2013. [Online]. from http://vizwiz.org/. [Accessed: 10-Mar-2013].

[25] P. Marziliano, F. Dufaux, S. Winkler and T. Ebrahimi. 2002. A no-reference perceptual blur metric. *IEEE International Conference on Image Processing*, 3 57-60.

[26] N. D. Narvekar and L. K. Karam. 2011. A No-Reference Image Blur Metric Based on the Cumulative Probability of Blur Detection (CPBD). *IEEE Transaction on Image Processing* 20(9): 2678-2683.

[27] D. L. Ruderman. 1994. The statistics of natural images. Network: Computation in Neural Systems 5:517-548.

[28] M. Saad, A.C. Bovik and C. Charrier. 2011. Blind image quality assessment: A natural scene statistics approach in the DCT domain. *IEEE Trans. Image Process*. 21:3350-3364.

[29] J. Sanchez, M. Saenz, and J. Garrido. 2010. Usability of a multimodal video game to improve navigation skills for blind children. *ACM Trans. Access. Comput*. 3, 2, Article 7 (November 2010), 29 pages. http://doi.acm.org/10.1145/1857920.1857924

[30] H. R. Sheikh, M. F. Sabir and C. Bovik. 2006. A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms. *IEEE Transactions on Image Processing*, 15(11): 3411-3452.

[31] H. Shen and J. Coughlan. 2007. Grouping using factor graphs: an approach for finding text with a camera phone. *Graph-Based Representations in Pattern Recognition*, pp. 394–403.

[32] Y. Shen and L.Z. Ma. 2010. Detecting and Removing the Motion Blurring from Video Clips," *International Journal of Modern Education and Computer Science*, 1:17-23.

[33] Y. C. Shiu and C. Huang. 1991. Locating cylindrical objects from perspective projections. Aerospace and Electronics Conference, 1991,NAECON 1991, Proceedings of the IEEE 1991 National.

[34] S. Suresh, R. V. Babu and H. J. Kim. 2009. NO-reference image quality assessment using modified extreme learning machine classifier. *Applied Soft Computing* 9: 541-552.

[35] Y. Tian, X. Yang, C. Yi, and A. Arditi. 2013. Toward a Computer Vision-based Wayfinding Aid for Blind Persons to Access Unfamiliar Indoor Environments. *Machine Vision and Applications*, Issue 24, No. 3, April 2013, pp. 521-535.

[36] L. Tian, C. Yi and Y. Tian. 2013. Detecting good quality frames in videos captured by a wearable camera for blind navigation. *IEEE International Conference on bioinformatics and Biomedicine (BIBM'13*).

[37] S. Trewin, M. Laff, V. Hanson, and A. Cavender. Exploring visual and motor accessibility in navigating a virtual world. In ACM Transactions on Accessible Computing (TACCESS), Vol. 2, No. 2, 2009

[38] P. Vu and D. Chandler. 2012. A fast wavelet-based algorithm for global and local image sharpness estimation. *IEEE Signal Processing letter*, 19: 423-426.

[39] S. Wang, C. Yi, and Y. Tian. 2012. Signage Detection and Recognition for Blind Persons to Access Unfamiliar Environments. *Journal of Computer Vision and Image Processing*, vol. 2, no. 2.

[40] B. Walker, and L. Mauney. 2010. Universal design of auditory graphs: A comparison of sonification mappings for visually impaired and sighted listeners. *ACM Trans. Access. Computing*. 2, 3, Article 12 (March 2010), 16 pages. DOI = 10.1145/1714458.1714459.

[41] W. Xue, L. Zhang and X. Mou. 2013. Learning without human scores for blind image quality assessment. *IEEE Computer Vision and Pattern Recognition*.

[42] C. Yi and Y. Tian. 2011. Assistive text reading from complex background for blind persons. *ICDAR Workshop on Camera-based Document Analysis and Recognition (CBDAR), Springer LNCS-7139*, 15-28.

[43] C. Yi and Y. Tian. 2011. Text string detection from natural scenes by structure based partition and grouping. IEEE Transactions on Image Processing, Vol. 20, Issue 9.

[44] J. Shi, L. Xu and J. Jia. Discriminative Blur Detection Features. In IEEE Conference on Computer Vision and Pattern Recognition, pp. 2965 - 2972, 2014.

Long Tian received his M.S. degrees in Department of Electrical Engineering from The City College of New York, 2014. He is currently in Ph.D. program in Department of Electrical and Computer Engineering at Mississippi State University. His research interests include image detection and recognition, and Hyper-spectral image classification.

Chucai Yi received his B.S. and the M.S. degrees in Department of Electronic and Information Engineering from Huazhong University of Science and Technology, China, in 2007 and 2009, and his Ph.D. degree from The Graduate Center City University of New York, USA. He is currently working for HERE North America as Senior Research Engineer. His research interests include object detection and recognition, surveillance video event analytics, and high definition road feature extraction from LIDAR data.

YingLi Tian (M'99–SM'01) received her BS and MS from TianJin University, China in 1987 and 1990 and her PhD from the Chinese University of Hong Kong, Hong Kong, in 1996. After holding a faculty position at National Laboratory of Pattern Recognition, Chinese Academy of Sciences, Beijing, she joined Carnegie Mellon University in 1998, where she was a postdoctoral fellow of the Robotics Institute. Then she worked as a research staff member in IBM T. J. Watson Research Center from 2001 to 2008. She is currently a professor in Department of Electrical Engineering at the City College of New York and Department of Computer Science at the Graduate Center of the City University of New York. Her current research focuses on a wide range of computer vision problems from human activity analysis, motion detection and analysis, to facial expression analysis, video surveillance, and assistive technology. She is a senior member of IEEE.