Wearable Vision Assistance System Based on Binocular Sensors for Visually Impaired Users

Bin Jiang, Jiachen Yang, *Member, IEEE*, Zhihan Lv, *Member, IEEE*, and Houbing Song, *Senior Member, IEEE*

***Abstract*—Blind or visually impaired people face special difficulties in daily life. With the advances in vision sensors and computer vision, the design of wearable vision assistance system is promising. In order to improve the life quality of the visually impaired group, a wearable system is proposed in this paper. Typically the performance of visual sensors is affected by a variety of complex factors in practice, resulting in a large number of noise and distortion. In this paper, we will creatively leverage image quality evaluation to select the captured images through vision sensors, which can ensure the input quality of scenes for the final identification system. First, we use binocular vision sensors to capture images in a fixed frequency and choose the informative ones based on stereo image quality assessment. Then the captured images will be sent to cloud for further computing. Specially, the detection and automatic result will be done for all the received images. Convolutional neural network based on big data will be used in this step. According to image analysis, the cloud computing can return the requested information for users, which can help them make a more reasonable decision in further action. Simulations and experiments show that the proposed method can solve the problem effectively. In addition, statistical results also demonstrate that wearable vision system can make visually impaired group more satisfied in visual needed situations.**

***Index Terms*—Binocular vision sensors, convolutional neural network (CNN), stereo image quality assessment (SIQA), wearable assistant system.**

I. Introduction

WITH the advances in computer vision, it is possible for people with visual impairment to be better integrated into normal daily life [1]–[3]. In the past, a variety of medical tools have been developed to help solve this problem [4]–[6]. However, the real-time visual recognition system based on wearable devices will enhance life experience to this group, thus bridging the gap between visual impairments and ordinary people's sense of experience [7]. In addition, the senior people will encounter similar problems in their life, and how to help them live better is also a subject worthy of investigation [8]–[10].

Many research efforts have been made on design of visual impairment auxiliary algorithms. Ross [11] first proposed to use wearable computers as a virtual environment interface for people with visual impairment. These preliminary explorations laid a solid foundation for this paper. Neto *et al.* [12] introduced a face detection system for people with visual impairments, which can work in real time. In this system, Microsoft Kinect was chosen as the main sensor for wearable device. Chua *et al.* [13] investigated on binocular luster effect. Based on the theory, they proposed an augmenting visual information system which aimed at helping colorblind people. Carroll *et al.* [14] put forward a shopping assistant system which can also help low-vision people in shopping. Cardin *et al.* [15] proposed a wearable system which can help visually impaired people in mobility improvement. The system detected obstacles that surround the user by using multisonar system and sending appropriate vibro-tactile feedback. Elmannai and Elleithy [16] conducted a comprehensive review of the state-of-the-art and the state-of-the-practice of sensor-based assistance systems. Terven *et al.* [17] conducted a comprehensive review of the state-of-the-art and the state-of-the-practice of sensor-based assistance systems. Pieralisi *et al.* [18], [19] proposed a system based on an electromagnetic sensor, which can be used for autonomous running in the athletes with low-vision. Based on RGB-D sensors, Pieralisi *et al.* [18], [19] studied navigation assistance for the visually impaired people. Other than using several sensors, they chose a consumer RGB-D camera and took advantage of both range and visual information. In a word, visual health assistance system has become an active research area. And the above research also serves as a reference for the progress of this area.

In addition to the study of visual obstacle groups, other researches on wearable vision sensors have been also worthy of reference. Terven *et al.* [17] mainly studied the topic of gesture recognition based on wearable vision sensors, which can improve the museum experiences for visitors. Zhu *et al.* [20] designed a behavior recognition system based on wearable sensor, which aimed to be a smart assisted living system. Wan *et al.* [21] proposed a wearable sensor localization scheme which can be a health monitoring system. Mayol-Cuevas *et al.* [22] compared on wearable

Manuscript received January 9, 2018; revised April 28, 2018; accepted May 27, 2018. Date of publication May 31, 2018; date of current version May 8, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61471260 and Grant 61572231, and in part by the Natural Science Foundation of Tianjin under Grant 16JCYBJC16000. *(Corresponding author: Jiachen Yang.)*

B. Jiang and J. Yang are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: jiangbin@tju.edu.cn; yangjiachen@tju.edu.cn).

Z. Lv is with the School of Data Science and Software Engineering, Qingdao University, Qingdao 266071, China (e-mail: zhihanlv@gmail.com).

H. Song is with the Department of Electrical, Computer, Software, and Systems Engineering, Embry-Riddle Aeronautical University, Daytona Beach, FL 32114 USA (e-mail: h.song@ieee.org).

Digital Object Identifier 10.1109/JIOT.2018.2842229

vision sensors for industry and research. Wu and Jafari [23] proposed a method for seamless vision-assisted calibration which can improve the performance of wearable inertial sensors. Brutti and Cavallaro [24] designed a cross-modal adaptation scheme based on wearable sensors. In their research, the online system can work online for audio-visual person identification. Bolanos *et al.* [25] conducted a comprehensive overview of visual lifelogging considering storytelling, which acquires images that capture the daily experiences of the user by wearing a camera over a long period of time.

Other researchers have also investigated in other aspects of the auxiliary sensors. Lin *et al.* [26] designed a noninvasive and continuous blood pressure monitoring method using wearable body sensor networks. Nemati *et al.* [27] designed a wireless wearable ECG sensor, which aimed at long-term applications. Fensli *et al.* [28] put forward a sensor acceptance model which can be used for measuring the status of patients. In these studies, they focus on the performance and application environment of the sensor.

Based on above discussions, three issues that restrict the development of wearable vision assistance system require for further research. First of all, hardware devices such as poor quality visual sensors cannot meet the development requirements. Therefore, the selection of sensors is very important. Proper use will help improve the performance of the whole system. Second, the noise interference in the acquisition process poses a challenge to the image quality selection [29]. If all the images are processed, bad information will introduce some interference [30]. Finally, the complexity of the recognition algorithm results in implementation difficulty in the wearable terminal in real time. At present, the traditional feature matching algorithm is difficult to meet the requirements of recognition. All of these problems are the motivations of the project.

In this paper, we propose a wearable system for blind or visually impaired users based on big data and image quality assessment (IQA). In the proposed system, the following contributions of this paper can be summarized to improve the performance of wearable vision assistance system for health.

1) The selection of binocular visual sensors will help the whole system. Most of the previous recognition systems collect data through monocular camera, which is not consistent with the basic features of the human eye vision system. Therefore, the use of a binocular vision sensor will help to solve such a problem and establish a more stable and practical way to collect visual information. In other areas, such a sensor is also very worthy of consideration. The application of the binocular vision will greatly improve the performance of the related applications.
2) We derive a stereo-IQA (SIQA) model for data choice. Experiments show that the proposed SIQA method can provide complementary information for image quality perception and choose the required images for the following detection steps. Captured images will be selected for further cloud computing, which will choose the required data for convolutional neural network (CNN) in recognition.

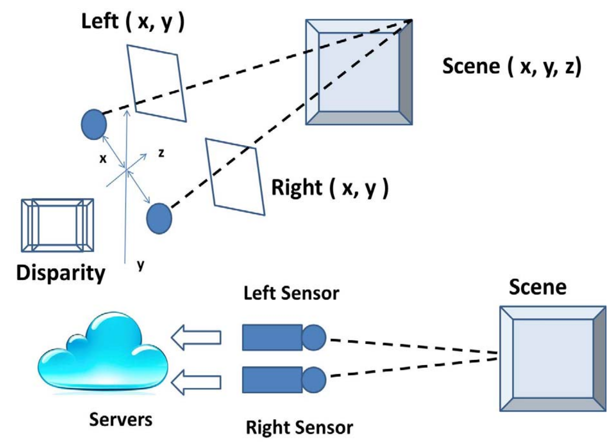


Fig. 1. Human binocular vision system.

3) Image recognition will be the core of the system, and the fast response of cloud computing will be the key to the overall performance. Specially, the detection and automatic update will be conducted detection and automatic innovation for all the received images. Based on the image analysis, the cloud computing can return the requested information for users, which can help them make a more reasonable decision. And CNN will be used in this step.

The rest of this paper is organized as follows. Section II presents the background and motivation. In Section III, the proposed algorithm framework will be presented. In Section IV, the experimental design and experimental results are shown. At last, the conclusions and future research direction will be given in Section V.

II. Background and Motivation

The objective of this paper is to propose a wearable vision assistance system for blind or visually impaired users based on big data and binocular sensors. Considering the previous discussions in related work, three aspects are relevant to the background and motivation. In Section II-A, binocular sensor design will be compared with monocular sensor. Specially, binocular vision theory in human vision system will be presented. In Section II-B, the research in stereoimages quality assessment will be reviewed, which can help the choice of captured images from wearable sensor. Section II-C focuses on objects detection.

A. From Monocular to Binocular

In the field of traditional computer vision, image data is usually obtained by only one camera. On the one hand, it can not well simulate the human vision system. On the other hand, binocular stereo vision is an important model for human visual system, which is based on the principle of parallax and uses the imaging device to acquire two images of objects from different locations [31]. In this way, it can obtain the stereo geometric information for objects, which is shown in Fig. 1.

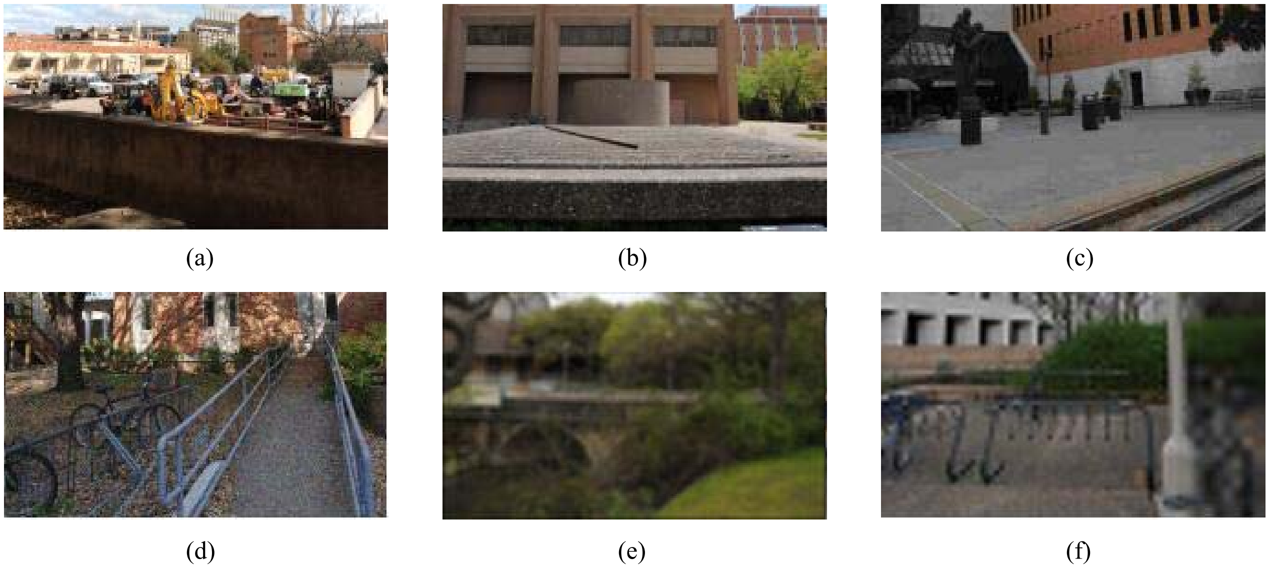


Fig. 2. Stereopairs in different image qualities and the corresponding differential mean opinion scores (DMOS) in LIVE 3-D phase 1 database. Because of symmetrically distorted pairs, left images of stereopairs are only displayed. (a) Pristine image. (b) DMOS = 5.3646. (c) DMOS = 17.7409. (d) DMOS = 26.6732. (e) DMOS = 36.8685. (f) DMOS = 46.4388. The higher DMOS, the lower quality.

Compared with monocular images, binocular images can better deal with human vision problems. As an emerging research field, the binocular stereo vision is almost built through 3-D structure extraction from a digital image. Based on binocular vision, the shape and spatial relationships of objects can be described easily. Then the simple 2-D image in the past analysis is extended to the complex stereo scene, which marks the birth of the stereo vision technology. In (1), it give one example of the previous binocular models. And the disparity computing is shown in (2) as follows:

$$C = \frac{E_L}{1+E_L+E_R} \cdot I_L + \frac{E_R}{1+E_L+E_R} \cdot I_R \quad (1)$$

$$\text{Disparity} = f\frac{x_c}{z_c} - f\frac{(x_c - 2r)}{z_c}. \quad (2)$$

Further research has extended to edge features, corner features, and geometric elements. Lines, planes, and surfaces are analyzed with image's shading, texture, motion, and geometry. Based on the above features, all kinds of data structure and inference rules are established. After that, researchers generalized the research achievements of image processing, psychophysics, neurophysiology, and clinical psychiatry from the field of information processing [32]. In addition, some publications established the theoretical framework of binocular visual computing, which played a great role in promoting the development of the basic theory in stereo vision technology. In other words, a complete ecosystem from image acquisition to stereo visual surface reconstruction has been formed. In this way, the stereo vision in computer vision has become a very important area [33].

Retinal images will be obtained by the brain, which can get the final vision. In fact, the human visual system has two important visual neural pathways that control the behavior of people: 1) ventral pathway and 2) dorsal pathway. On the one hand, ventral pathway contains the primary visual cortex $V1$, secondary visual cortical areas $V2$ and $V3$. In addition, it contains ventral extrastriate cortex $V4$ for consciousness and perception [34]. On the other hand, dorsal pathway is mainly built by $V1$, $V2$, and $V3$. Based on the temporal lobe MT and MST composition, visual stimuli in a specific location and movement status is determined [35].

B. *Stereo-Image Quality Assessment for Data Choice*

In the process of acquisition and transmission of stereoscopic images, distortion is inevitable [36]. However, the introduction of image distortion and noise greatly reduces the quality of the image, thus affecting further processing of the acquisied image. If a large number of poor quality images are mixed, the processing efficiency will be greatly reduced, thus affecting the overall performance of the corresponding system, in both accuracy and time consuming. Fig. 2 gives the different stereoimage examples with different quality. Therefore, how to select data from a large number of images to meet the requirements is an important research area. Based on the above discussion, the IQA is becoming more important. In particular, such evaluation methods can be divided into two categories: 1) IQA based on traditional feature extraction and 2) IQA based on deep learning [37]–[41].

In the early stage of the study, peak signal to noise ratio was widely used. Structural similarity is another important index for the evaluation methods. With the development of the study, the SIQA algorithm has been improved, and researchers extract different features to represent the quality of images. Further more, the use of shallow learning is used to make score prediction, such as SVM.

Recently, researchers have begun to use deep learning in image quality evaluation. For different information (image, voice, and text), it is necessary to use different network models to achieve excellent results. Deep learning can be used to evaluate the quality of the images more accurately. In addition,

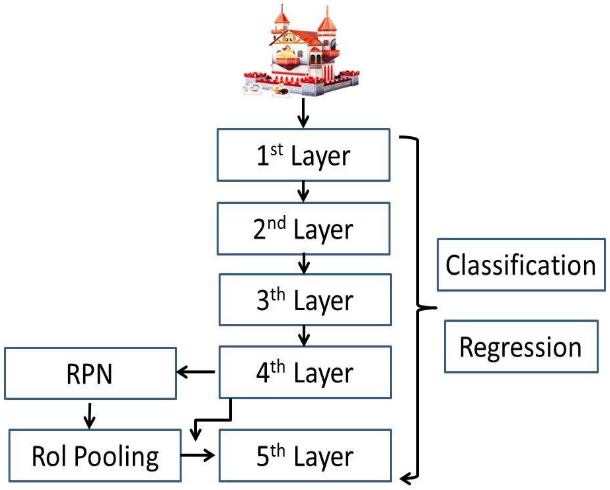


Fig. 3. Typical deep learning model: CNN.

it also has a very significant effect on the quality evaluation of stereoscopic images.

C. Object Detection Based on Convolutional Neural Network

Object detection is an important area in image processing, which is useful for the wearable vision system in this paper. In the traditional object detection algorithm, feature extraction is always a very effective method. For a potential object that needs to be identified, the distance from the eigenvector of a control system is also a reasonable method. However, the limitations of this method are also very obvious.

The concept of deep learning derives from the study of artificial neural networks. Multi perceptrons with multiple hidden layers are a kind of deep learning structure. Deep learning forms more abstract highlevels by combining low layer features to represent attribute categories or features, in order to identify distributed characteristics of data [42]–[44].

Deep learning is a method based on the representation of data in machine learning. Observations can be expressed in many ways, such as vectors of each pixel's intensity value, or more abstractly to represent a series of edges and specific shape areas. It is easier to learn tasks from an instance such as face recognition or facial expression recognition [45], [46]. The advantage of deep learning is to use unsupervised or semi supervised feature learning and layered features to extract efficient algorithms to replace handmade features.

CNN is a kind of artificial neural network, which is important in the field of speech analysis and image recognition. In Fig. 3, we show a typical example on pattern recognition based on CNN. Its weight sharing network structure makes it similar to the biological neural network, which reduces the complexity of the network model, and reduces the number of weights. This advantage is more obvious when the input of the network is multidimensional image [46], [47]. The image can be directly used as the input of the network, which avoids the complex feature extraction and data reconstruction process in the traditional recognition algorithm. Convolutional network is a multi layer perceptron for recognizing 2-D shape and special design. This kind of network structure has a high degree of invariance to translation, scaling, tilting, or deformation [48].

III. System Design

In this paper, we propose a wearable vision assistance system for blind or visually impaired users based on big data and binocular sensors. The framework is shown in Fig. 4. Specially, three sections will be integrated in this system.

A. Binocular Image Acquisition for Wearable System

Sensors choice is very important for the performance of the whole system. In this paper, binocular sensor will be used for the wearable vision system. The way of binocular detection is to directly measure the distance of the front scene by calculating the parallax of the two images.

In this system, sensors based on binocular vision contain two sets of CCD cameras with the same performance. Based on the principle of parallax, it can complete the 3-D measurement of all feature points in the field of view, especially the visual measurement tasks that other types of sensors cannot complete, such as the center of the round hole and triangular vertex position measurement. Therefore, the binocular vision sensor is chosen in the detection system. The key point to directly measure the 3-D measurement in large object implementation is binocular vision sensor, which is necessary to know the internal parameters of the sensor, structural parameters relationship, and sensor coordinate system.

The binocular vision probe in this paper is made up of two CCD cameras and one semiconductor laser. As a light source, semiconductor laser sends a light source to columnar lens which can become a straight light line. The laser beam is projected onto the surface of the workpiece as a measuring line. The laser wavelength is 650 nm, and the laser line width is about 1 mm. Two ordinary CCD cameras are placed in a certain angle to finish the depth measurement. The length of the CCD lens will affect the angle of the lens axis, the distance between the probes, the object to be measured, and the depth of the field.

Before the actual measurement, the camera parameters should be calibrated first. In this paper, the sensor is provided to the system before offline calibration, which can complete the internal parameters and the structure parameters. In addition, it can make use of standard 2-D and 1-D precision target guide. Through the mobile guide, it can determine a coordinated system, the relationship between these parameters can be obtained.

In this paper, the visual measurement belongs to a noncontact measurement based on the principle of laser triangulation. The purpose of camera calibration is to establish effective imaging models and determine intrinsic parameters of cameras, so as to correctly establish corresponding relations between pixels in spatial coordinates and pixels on the image plane.

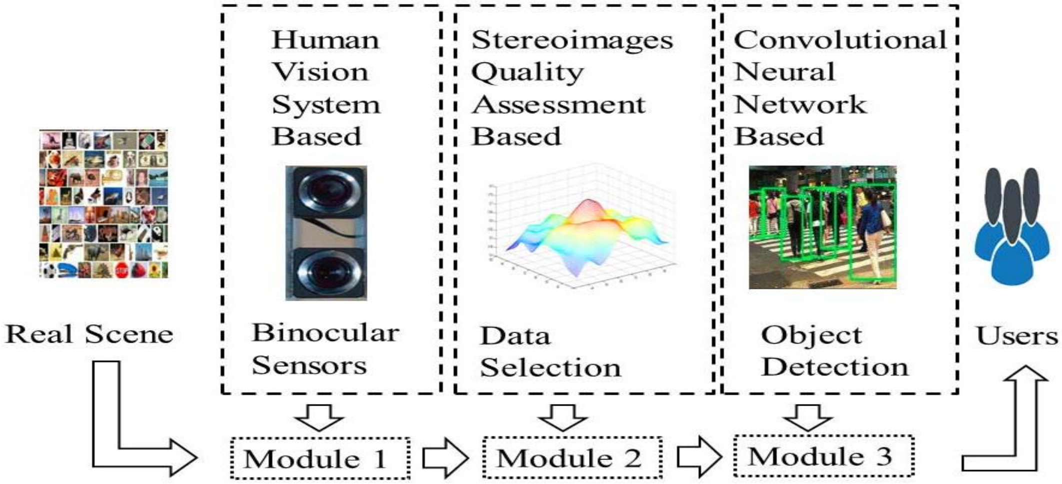


Fig. 4. Flow of the proposed wearable assistant visual system.

For binocular vision, the log-Gabor filter $G_{s,o}(\omega, \theta)$ can be used for the analysis, as shown in

$$G_{s,o}(\omega, \theta) = \exp\left[-\frac{(\log(\omega/\omega_s))^2}{2\sigma_s^2}\right] \cdot \exp\left[-\frac{\theta - \theta_0}{2\sigma_o^2}\right] \tag{3}$$

where s and o are, respectively, spatial scale and orientation information, θ represents direction angle information, ω_s and ω_o are used to determine the filter energy, ω and ω_s represent the normalized radial filter frequency and the center frequency. In this way, we can calculate the signal X position in the size of s, o direction of the local energy information

$$E_o(X) = \sqrt{F_o(X)^2 + H_o(X)^2} \tag{4}$$

where

$$F_o(X) = \sum_s \eta_{s,o}(X), H_o(X) = \sum_s \zeta_{s,o}(X). \tag{5}$$

Based on this method, the real-time image will be collected by the binocular sensor to form stereo image sources. In addition, the frequency of image acquisition will be controlled at least 10 frames/s.

B. Required Images Choice Based on SIQA

As the second part of the system, it is not reasonable for all the images to be sent for the cloud computing. At the same time, a large number of noise and other interference will greatly reduce the quality of some stereo images. So, selecting images will be the second task we have to do. In this way, we introduce the IQA to solve this issue. For stereoscopic images, depth information will be easier to calculate, which is useful for object detection. So the quality evaluation of stereoscopic image will be divided into two directions. On the one hand, there is a need to evaluate the basic quality of the image. On the other hand, there is a need to evaluate the stereoscopic realism of the image. In the quality evaluation of 2-D image, several models have been proposed. In this system, some existing image quality evaluation models will be used. In addition, disparity map is often used to evaluate the degree of matching of stereoscopic images. Therefore, we also use disparity map to build stereoscopic and non stereoscopic parts of stereoscopic images.

According to the above analysis, the binocular part will determine the depth and authenticity of the stereoscopic image, then the rest will determine the basic quality of the image [49]. In order to fully decompose the image, we use the Hough transform to detect the edge information so as to select the most effective stereo information. In traditional image processing, Hough transform is an effective feature extraction method, which can be used for object shaping detection based on a voting algorithm. In Hough transform, line detection in images will be extended to the recognition of objects of arbitrary shape, circle, and ellipse, which can be used in this paper

$$Q = \omega_1 Q_{\text{stereo}} + \omega_2 Q_{\text{nonstereo}}. \tag{6}$$

For stereo parts, we choose features extracted from disparity map, which can be defined as Q_{stereo}. For non stereo parts, we only deal with it as a single 2-D image, which can be defined as $Q_{\text{nonstereo}}$. Then the evaluation results can be obtained through different parts. Based on the experimental test, it can be defined that $\omega_1 = 0.623$ and $\omega_2 = 0.377$ in (6).

C. Image Recognition Based on Cloud Computing

When the image selection is completed, the recognition process will be started under the support of the cloud computing. In the server, the appropriate recognition algorithm needs to be used. In this paper, we choose ResNet as the main network as the detection method [50]. In our task, the stereo images are obtained by data choice, so the resolution is high. In order to apply ResNet better, we first analyze the difficulties in application process. In the last layer of ResNet, the output of the standard ResNet is regarded as a coarse tool for obtaining the effective feature maps, and the weight change is shown in (7). As known, the final pooling value must represent a large receptive field. For example, a potential detective object with 48×48 in resolution can be transformed as only 3×3 in resolution through convolutional computing in ResNet network. So

TABLE I
PRECISION COMPARISON OF DIFFERENT METHODS IN ACTUAL DATABASE

	Persons	Trees	Monuments	Motors	Food	Buildings	Animals	Flowers	Traffic	Mountains	Average
PMWBS	0.453	0.398	0.374	0.741	0.369	0.915	0.304	0.852	0.568	0.293	0.527
PMWIQC	0.424	0.446	0.411	0.852	0.427	0.587	0.426	0.898	0.589	0.268	0.533
PMBTPR	0.683	0.540	0.562	0.764	0.733	0.812	0.658	0.891	0.803	0.522	0.727
Proposed	**0.730**	**0.612**	**0.597**	**0.890**	**0.781**	**0.990**	**0.713**	**0.916**	**0.862**	**0.573**	**0.766**

TABLE II
RECALL COMPARISON OF DIFFERENT METHODS IN ACTUAL DATABASE

	Persons	Trees	Monuments	Motors	Food	Buildings	Animals	Flowers	Traffic	Mountains	Average
PMWBS	0.115	0.121	0.127	0.092	0.129	0.072	0.132	0.087	0.102	0.135	0.111
PMWIQC	0.126	0.113	0.132	0.099	0.122	0.104	0.119	0.093	0.103	0.152	0.116
PMBTPR	0.141	0.192	0.174	0.121	0.132	0.101	0.149	0.112	0.134	0.213	0.146
Proposed	**0.173**	**0.221**	**0.197**	**0.132**	**0.135**	**0.113**	**0.159**	**0.134**	**0.146**	**0.229**	**0.164**

we find the inaccurate localization when facing object detection. Based on this consideration, we have improved ResNet to better adapt to the requirements under the cloud computing framework

$$y = F(x, W_i) + W_s x. \tag{7}$$

Through the cloud computing framework, we can calculate the detection results in the image on the server. Then, it is also a very important step to feed the results to the user. Of course, the needs of each user are different. For example, some users only focus on the subject objects in the current field of vision, so the feedback will pick out the results of the object with the highest score. In another scenario, the customer needs a comprehensive analysis, so the feedback will pick up more detection results.

IV. TEST PROTOCOL AND EXPERIMENTAL RESULTS

In this section, we propose the experiment benchmark and performance protocol, which can identify the accuracy for the designed wearable system. Due to the related lack in research on vision assistance system, it is very hard to choose a convinced method to evaluate the designed system. Based on this situation, we have selected different actual scenarios to test and then verify it based on standard image datasets. Based on the two standard stereo image datasets, the results in both performance and user experience are given to verify the efficiency of the designed system. In addition, some detailed performance will also be discussed in different sections.

A. Real Scenarios and Stereo Datasets

The first proposed verification is based on real scene dataset, in which 60 different scenarios are selected as the test basis. In this process, scenarios with different objects are collected. In every scenario, multiobjects construct the complex changes for the test process.

In addition, we use stereoscopic image collection as a test database with a larger amount of data. In this part, some databases with stereoscopic images are imported. Specially, more than 200 stereoscopic images are used. During the test process, stereoscopic images will be played by stereo displaying devices. And the user can sit or stand in front of the screen. In this way, it can simulate the real scenarios.

B. Performance Protocol

At present, there is no strict evaluation system in this field to estimate this kind of system performances. So, we use two important indexes in the image retrieval process as the basis for this research topic. Specially, precision and recall to carry out test analysis [51]. These two performance metrics can fully demonstrate the accuracy of the system in judging the object process and make an objective evaluation. The larger value of the above two indicators means the better performance of the system.

In addition to demonstrate the advantages of sensor selection and algorithm design based on the experimental results of this system, we also compared with other methods. Here, we choose some other methods by giving up some critical steps as references. Specially, PMWBS denotes proposed method without binocular sensor, PMWIQC denotes proposed method without image quality choosing, and PMBTPR denotes proposed method based traditional pattern recognition algorithm.

C. Performance Results

In this section, we will present the data results for the performance, which can evaluate the system in objective view. For each evaluation, we conduct 100 experiments to obtain the average value to offset the impact of accidental factors. As discussed in Section IV-B, the precision and recall will be calculated, respectively.

Based on 100 experiments for average, the precision and recall in real database are shown in Tables I and II. As seen in the data statistic, we choose ten typical kinds of objects to be tested, and give the precision in Table I and recall in Table II. For every object category, the proposed method can get the highest precision and recall. The last column for both the two indexes gives the average for the ten objects recognition. The average precision is 0.766, compared with 0.527, 0.533, and 0.727. On the other hand, the average precision

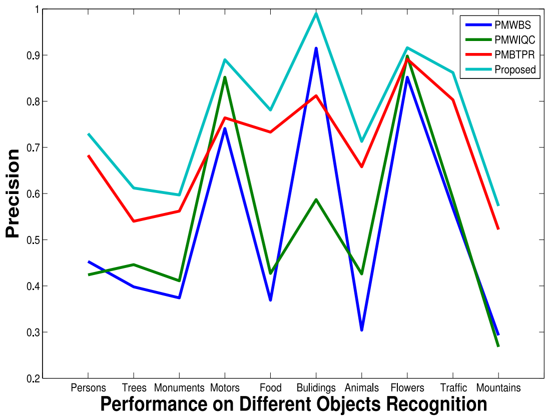


Fig. 5. Precision comparison of different methods in actual database.

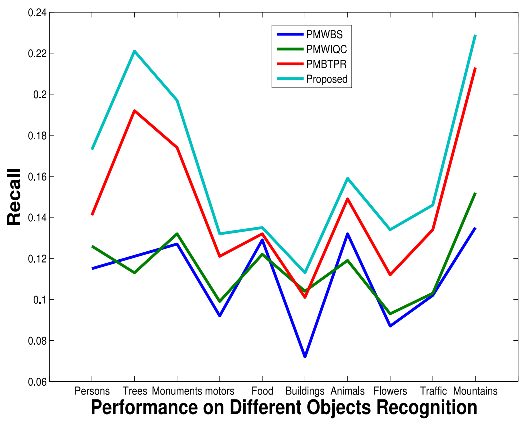


Fig. 6. Recall comparison of different methods in actual database.

TABLE III
COMPARISON OF DIFFERENT METHODS IN COLLECTED STEREOIMAGES

	PMWBS	PMWIQC	PMBTPR	Proposed
Precision	0.573	0.421	0.531	**0.675**
Recall	0.132	0.176	0.199	**0.287**

is 0.164, compared with 0.111, 0.116, and 0.146. Obviously, the proposed system can accomplish the task better in helping people with visual impairments in real scenarios. In addition, we show the results in Figs. 5 and 6, which can facilitate the comparison.

In addition, the performance based on the collected stereoimages are shown in Table III and the results can embody the simulation results. As shown, the performance in simulation can also get the best precision and recall by the designed system. The average precision is 0.675, compared with 0.573, 0.421, and 0.531. On the other hand, the average precision is 0.287, compared with 0.132, 0.176, and 0.199. In order to show it in another way, Fig. 7 can help compare the four different systems in simulation environment.

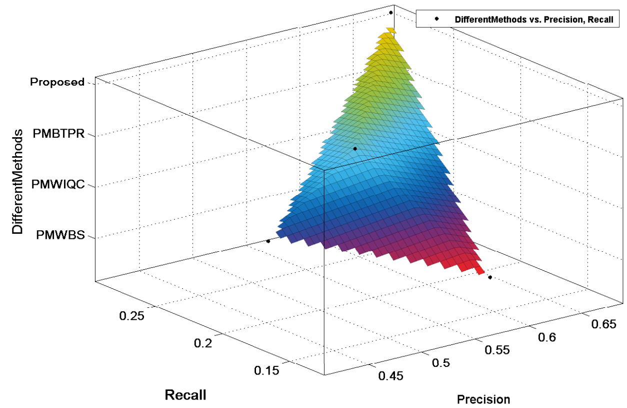


Fig. 7. Comparison of the four different systems in simulation environment.

TABLE IV
COMPARISON OF DIFFERENT METHODS IN COLLECTED STEREOIMAGES

	Group One	Group Two	Group Three	Average
PMWBS	2.32	2.12	2.74	2.39
PMWIQC	3.24	3.44	3.11	3.26
PMBTPR	2.11	2.40	2.22	2.24
Proposed	**3.67**	**3.84**	**3.97**	**3.83**

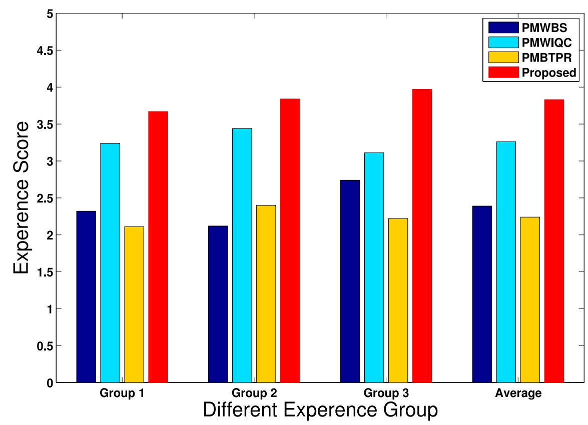


Fig. 8. Comparison of different methods in collected stereoimage databases.

D. User Experience Results

In addition to the statistical data, we also give the assessment on experience quality based on participants for the performance, which can evaluate the system in subjective view. In our research, we apply the designed system in both the group of visually impaired people and the group without visual trouble. After that, we can collect the feedback from them. Specially, we set a standard for the feedback data, which ranges from 0 to 5. In addition, we divide the test people into three groups. Group One is the people with visually impaired trouble; Group Two is the people without it; and Group Three is the researchers who are familiar with the system design. At last, the average results can be obtained. Using same data statistic methods, we collect the feedback with different designed systems which are same as last section. And the data results are shown in Table IV and Fig. 8.

In conclusion, the designed system proposed in this paper can get the best quality of experience.

V. Conclusion

With the advances in vision sensors and computer vision, the design of wearable vision assistance systems is promising. First, we use binocular vision sensors to capture images in a fixed frequency and choose the informative ones based on IQA. Second, captured images are sent to cloud for further computing. Specially, it can complete detection and automatic update for all the received images. CNN in big data is used in this step. Based on the image analysis, the cloud computing can return the requested information for users, which can help them make a more reasonable decision. Statistical results also demonstrate that wearable vision system can make visually impaired group more satisfied in visual needed situation. For future research, many potential works should be made to refine this system. Among them, how to apply the other characters of binocular vision to the system will be a key problem. The recognition of stereoscopic images will also improve the performance of the system.

References

[1] S. Chen, H. Xu, D. Liu, B. Hu, and H. Wang, "A vision of IoT: Applications, challenges, and opportunities with China perspective," *IEEE Internet Things J.*, vol. 1, no. 4, pp. 349–359, Aug. 2014.

[2] D. Pascolini and S. P. Mariotti, "Global estimates of visual impairment: 2010," *British J. Ophthalmol.*, vol. 96, no. 5, pp. 614–618, 2012.

[3] D. Peters, A. Heijl, L. Brenner, and B. Bengtsson, "Visual impairment and vision-related quality of life in the early manifest glaucoma trial after 20 years of follow-up," *Acta Ophthalmologica*, vol. 93, no. 8, pp. 745–752, 2016.

[4] Y. Zhang, L. Sun, H. Song, and X. Cao, "Ubiquitous WSN for healthcare: Recent advances and future prospects," *IEEE Internet Things J.*, vol. 1, no. 4, pp. 311–318, Aug. 2014.

[5] K. L. Hanna, L. R. Hepworth, and F. Rowe, "Screening methods for post-stroke visual impairment: A systematic review," *Disability Rehabil.*, vol. 39, no. 25, pp. 2531–2543, 2016.

[6] N. Luo *et al.*, "Direct medical costs associated with clinical and health characteristics in visually impaired individuals in Singapore," *Value Health*, vol. 18, no. 3, p. A254, 2015.

[7] A. Ometov *et al.*, "Facilitating the delegation of use for private devices in the era of the Internet of Wearable Things," *IEEE Internet Things J.*, vol. 4, no. 4, pp. 843–854, Aug. 2017.

[8] J. R. Evans *et al.*, "Prevalence of visual impairment in people aged 75 years and older in Britain: Results from the MRC trial of assessment and management of older people in the community," *British J. Ophthalmol.*, vol. 86, no. 7, pp. 795–800, 2002.

[9] Y. Jiang *et al.*, "Data-centered runtime verification of wireless medical cyber-physical system," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 1900–1909, Aug. 2017.

[10] M. Hosseini, Y. Jiang, R. R. Berlin, L. Sha, and H. Song, "Towards physiology-aware DASH: Bandwidth-compliant prioritized clinical multimedia communication in ambulances," *IEEE Trans. Multimedia*, vol. 19, no. 10, pp. 2307–2321, Oct. 2017.

[11] D. A. Ross, "Wearable computers as a virtual environment interface for people with visual impairment," *Virtual Reality*, vol. 3, no. 3, pp. 212–221, 1998.

[12] L. B. Neto *et al.*, "A Kinect-based wearable face recognition system to aid visually impaired users," *IEEE Trans. Human–Mach. Syst.*, vol. 47, no. 1, pp. 52–64, Feb. 2017.

[13] S. H. Chua *et al.*, "ColorBless: Augmenting visual information for colorblind people with binocular luster effect," *ACM Trans. Comput. Human Interact.*, vol. 21, no. 6, pp. 1–20, 2015.

[14] J. M. Carroll, M. Mcmanus, S. Lee, P. A. Zientara, and V. Narayanan, "The third eye: A shopping assistant for the visually impaired," in *Proc. CHI Conf. Extended*, 2017, p. 465.

[15] S. Cardin, D. Thalmann, and F. Vexo, "A wearable system for mobility improvement of visually impaired people," *Vis. Comput.*, vol. 23, no. 2, pp. 109–118, 2007.

[16] W. Elmannai and K. Elleithy, "Sensor-based assistive devices for visually-impaired people: Current status, challenges, and future directions," *Sensors*, vol. 17, no. 3, pp. 565–606, 2017.

[17] J. R. Terven, J. Salas, and B. Raducanu, "New opportunities for computer vision-based assistive technology systems for the visually impaired," *Computer*, vol. 47, no. 4, pp. 52–58, Apr. 2014.

[18] M. Pieralisi *et al.*, "An electromagnetic sensor for the autonomous running of visually impaired and blind athletes (Part I: The fixed infrastructure)," *Sensors*, vol. 17, no. 2, pp. 364–373, 2017.

[19] M. Pieralisi *et al.*, "An electromagnetic sensor for the autonomous running of visually impaired and blind athletes (Part II: The wearable device)," *Sensors*, vol. 17, no. 2, pp. 381–391, 2017.

[20] C. Zhu, W. Sheng, and M. Liu, "Wearable sensor-based behavioral anomaly detection in smart assisted living systems," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 4, pp. 1225–1234, Oct. 2015.

[21] L. Wan *et al.*, "Wearable sensor localization considering mixed distributed sources in health monitoring systems," *Sensors*, vol. 16, no. 3, pp. 368–392, 2016.

[22] W. W. Mayol-Cuevas, B. J. Tordoff, and D. W. Murray, "On the choice and placement of wearable vision sensors," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 39, no. 2, pp. 414–425, Mar. 2009.

[23] J. Wu and R. Jafari, "Seamless vision-assisted placement calibration for wearable inertial sensors," *ACM Trans. Embedded Comput. Syst.*, vol. 16, no. 3, pp. 1–22, 2017.

[24] A. Brutti and A. Cavallaro, "Online cross-modal adaptation for audio-visual person identification with wearable cameras," *IEEE Trans. Human–Mach. Syst.*, vol. 47, no. 1, pp. 40–51, Feb. 2017.

[25] M. Bolanos, M. Dimiccoli, and P. Radeva, "Toward storytelling from visual lifelogging: An overview," *IEEE Trans. Human–Mach. Syst.*, vol. 47, no. 1, pp. 77–90, Feb. 2017.

[26] H. Lin *et al.*, "Noninvasive and continuous blood pressure monitoring using wearable body sensor networks," *IEEE Intell. Syst.*, vol. 30, no. 6, pp. 38–48, Nov./Dec. 2015.

[27] E. Nemati, M. J. Deen, and T. Mondal, "A wireless wearable ecg sensor for long-term applications," *IEEE Commun. Mag.*, vol. 50, no. 1, pp. 36–43, Jan. 2012.

[28] R. Fensli, P. E. Pedersen, T. Gundersen, and O. Hejlesen, "Sensor acceptance model—Measuring patient acceptance of wearable sensors," *Methods Inf. Med.*, vol. 47, no. 1, pp. 89–95, 2008.

[29] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell, "Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition," in *Proc. Comput. Vis. Pattern Recognit. Workshops*, 2011, pp. 74–81.

[30] D. Gur *et al.*, "Forced choice and ordinal discrete rating assessment of image quality: A comparison," *J. Digit. Imag.*, vol. 10, no. 3, pp. 103–107, 1997.

[31] S. Wei, S. Wang, C. Zhou, K. Liu, and X. Fan, "Binocular vision measurement using Dammann grating," *Appl. Opt.*, vol. 54, no. 11, pp. 3246–3251, 2015.

[32] W. J. Zhou, L. Yu, and M. W. Wu, "Simulating binocular vision for no-reference 3D visual quality measurement," *Opt. Express*, vol. 23, no. 18, pp. 23710–23725, 2015.

[33] B. Montrucchio, C. Celozzi, and P. Cerutti, "Thresholds of vision of the human visual system: Visual adaptation for monocular and binocular vision," *IEEE Trans. Human–Mach. Syst.*, vol. 45, no. 6, pp. 739–749, Dec. 2015.

[34] M. A. Georgeson, S. A. Wallis, T. S. Meese, and D. H. Baker, "Contrast and Lustre: A model that accounts for eleven different forms of contrast discrimination in binocular vision," *Vis. Res.*, vol. 129, pp. 98–118, Dec. 2016.

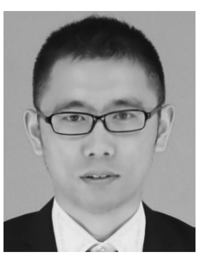
[35] D. Saint-Amour, V. Walsh, J. P. Guillemot, M. Lassonde, and F. Lepore, "Role of primary visual cortex in the binocular integration of plaid motion perception," *Eur. J. Neurosci.*, vol. 21, no. 4, pp. 1107–1115, 2015.

[36] W. Hou, X. Gao, D. Tao, and X. Li, "Blind image quality assessment via deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 6, pp. 1275–1286, Jun. 2015.

[37] Y. Liu *et al.*, "Stereoscopic image quality assessment method based on binocular combination saliency model," *Signal Process.*, vol. 125, pp. 237–248, Aug. 2016.

[38] Y.-H. Lin and J.-L. Wu, "Quality assessment of stereoscopic 3D image compression by binocular integration behaviors," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1527–1542, Apr. 2014.

[39] F. Shao, K. Li, G. Jiang, M. Yu, and C. Yu, "Monocular–binocular feature fidelity induced index for stereoscopic image quality assessment," *Appl. Opt.*, vol. 54, no. 33, pp. 9671–9680, 2015.
[40] F. Shao, W. Lin, S. Wang, G. Jiang, and M. Yu, "Blind image quality assessment for stereoscopic images using binocular guided quality lookup and visual codebook," *IEEE Trans. Broadcast.*, vol. 61, no. 2, pp. 154–165, Jun. 2015.
[41] W. Zhang, C. Qu, L. Ma, J. Guan, and R. Huang, "Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network," *Pattern Recognit.*, vol. 59, pp. 176–187, Nov. 2016.
[42] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[43] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.
[44] J. Yang, B. Jiang, B. Li, K. Tian, and Z. Lv, "A fast image retrieval method designed for network big data," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2350–2359, Oct. 2017.
[45] Z.-H. Ling *et al.*, "Deep learning for acoustic modeling in parametric speech generation: A systematic review of existing techniques and future trends," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 35–52, May 2015.
[46] D. Holden, J. Saito, and T. Komura, "A deep learning framework for character motion synthesis and editing," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–11, 2016.
[47] Z. Dong, Y. Wu, M. Pei, and Y. Jia, "Vehicle type classification using a semisupervised convolutional neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2247–2256, Aug. 2015.
[48] T. He, W. Huang, Y. Qiao, and J. Yao, "Text-attentional convolutional neural network for scene text detection," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2529–2541, Jun. 2016.
[49] B. Jiang, J. Yang, N. Jiang, Z. Lv, and Q. Meng, "Quality assessment for virtual reality technology based on real scene," *Neural Comput. Appl.*, vol. 29, no. 5, pp. 1199–1208, 2016.
[50] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, 2016, pp. 770–778.
[51] H. Müller, W. Müller, D. M. Squire, S. Marchand-Maillet, and T. Pun, "Performance evaluation in content-based image retrieval: Overview and proposals," *Pattern Recognit. Lett.*, vol. 22, no. 5, pp. 593–601, 2001.



Bin Jiang received the B.S. and M.S. degrees in communication and information engineering from Tianjin University, Tianjin, China, in 2013 and 2016, respectively, where he is currently pursuing the Ph.D. degree at the School of Electrical and Information Engineering.

He is also a Visiting Scholar with Ember-Riddle Aeronautical University, Daytona Beach, FL, USA. His current research interests include computer vision, including stereo image, image quality, pattern recognition, and cyber-physical image processing.

Jiachen Yang (M'15) received the M.S. and Ph.D. degrees in communication and information engineering from Tianjin University, Tianjin, China, in 2005 and 2009, respectively.

He is currently a Professor with Tianjin University. He was also a Visiting Scholar with the Department of Computer Science, School of Science, Loughborough University, Loughborough, U.K. His current research interests include image quality evaluation, stereo vision research, pattern recognition, and virtual reality.

Zhihan Lv (M'16) received the Ph.D. degree from the Ocean University of China, Qingdao, China, in 2012.

He is an Assistant Professor with Qingdao University, Qingdao. He was a Research Engineer with CNRS, Paris, France, a Post-Doctoral Fellow with Umeå University, Umeå, Sweden, and an Experience Researcher with Fundacion FIVAN, Valencia, Spain. He was a Marie Curie Fellow in European Union's Seventh Framework Programme LANPERCEPT. His current research interests include multimedia, computer vision, 3-D visualization, graphics, serious game, networks, big data, and software engineering.

Houbing Song (M'12–SM'14) received the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA, USA, in 2012.

In 2017, he joined the Department of Electrical, Computer, Software, and Systems Engineering, Embry-Riddle Aeronautical University, Daytona Beach, FL, USA, where he is currently an Assistant Professor and the Director of the Security and Optimization for Networked Globe Laboratory (www.SONGLab.us). He was a faculty member of West Virginia University, Morgantown, WV, USA, from 2012 to 2017. He has edited 4 books and has authored over 100 papers. His current research interests include cyber-physical systems, Internet of Things, cloud computing, big data analytics, connected vehicle, wireless communications and networking, and optical communications and networking.

Dr. Song has been serving as an Associate Technical Editor for *IEEE Communications Magazine* since 2017. He is a Senior Member of the ACM.