Finding Objects for Assisting Blind People

Chucai Yi[2], Roberto W. Flores[1], Ricardo Chincha[1], and YingLi Tian[1,2]

[1] *Department of Electrical Engineering*
The City College of New York
New York, NY, 10031, USA
{rflores01, rchinch00, ytian}@ccny.cuny.edu

[2] *Computer Science, The Graduate Center*
The City University of New York
New York, NY, 10016, USA
cyi@gc.cuny.edu

Abstract – Computer vision technology has been widely used for blind assistance, such as navigation and wayfinding. However, few camera-based systems are developed for helping blind or visually-impaired people to find daily necessities. In this paper, we propose a prototype system of blind-assistant object finding by camera-based network and matching-based recognition. We collect a dataset of daily necessities and apply Speeded-Up Robust Features (SURF) and Scale Invariant Feature Transform (SIFT) feature descriptors to perform object recognition. Experimental results demonstrate the effectiveness of our prototype system.

1.   Introduction

The World Health Organization (WHO) estimated that in 2002, 2.6% of the world's total population was visually impaired. Also, the American Foundation for the Blind (AFB) approximates that there are more than 25 million people in the United States living with vision loss. Visually impaired people face much inconvenience when interacting with their surrounding environments, and the most common challenge is to find dropped or misplaced personal items (i.e., keys, wallets, etc.). While many literatures and systems have been focusing on navigation [2], wayfinding [9], text reading [4], bar code reading, banknote recognition [8], etc, there are rare camera-based systems available in the market to find daily necessities for the blind. According to the American Foundation for the Blind [1], visually impaired individuals can live comfortably in a house or apartment by following these principles: lightning increment, hazard elimination, color contrasts creation, item organization and labeling, and glare reduction. It means that blind or visually impaired people need effective assistance, including natural daylight simulation, fluorescent tape, colored objects against dark backgrounds, and the tagging of daily necessities for recognition. However, it is inconvenient and expensive to make these complex arrangements in most cases. On contrast, camera setup is easier and more economical. It costs less than $500 to buy a 1080p full High-Definition camera, which supports video in 1920×1080 resolution. Thus computer vision and pattern recognition technology could be an effective alternative to assist blind or visually impaired people.

Based on our knowledge, few blind-assistive systems are able to help blind or visually impaired people find their daily necessities. Previous blind assistant system focuses more on navigation, path planning and blind person tracking. In [12], a method was designed to track foot player under a network of multiple cameras. In [13], a method is designed to infer the relative locations of cameras and construct indoor layout by analyzing their respective observations. In [14], cooperated with first-view camera, a sensor network was built to assist robot to perceive obstacles and plan effective paths. In [15], a camera network was built for object retrieval, where SIFT descriptor was employed for object detection and recognition. Locomotion analysis [26] could be used to design specific room layout to help blind people live a more convenient life. These systems assume that blind user is located at unfamiliar environment, and provide general instruction to assist blind people reach their destination. However, the probability of blind or visually-impaired people in unfamiliar environment is low (even for normal-vision people), so blind-assistant system should care more about the daily life of blind people. They need better perception and control of their personal objects, so we employ the computer vision technology to help them manage their daily necessities.

In our proposed system, a multiple-camera network is built by placing a camera at the important locations of the indoor environment of blind user's daily life. The important locations are usually located around tables, cabinets, and wash sink. The cameras provide scene monitoring around these fixed locations, and inform blind users of the locations of his/her demanded objects. In this process, matching-based object recognition is performed to find out the objects. Fast and efficient object recognition is a very popular area in the computer vision community. It is to localize and certify given objects in images or video sequences. Humans are able to recognize a wide variety of object within images or videos with little effort, no matter the difference in viewpoint, scenarios, image scale, translation, rotation, range of distortion, and illumination [22]. It is proved that

human visual system can discriminate among tens of thousands of different object categories [23] in an efficient way, about100 to 200 ms [24, 25, 29]. Also human vision system can effectively handle background clutter and object occlusions. However, it is extremely difficult to build robust and selective computer vision algorithms of object recognition to handle objects with large similarity or variations. Fortunately, present and future technological innovations in object detection and recognition systems will contribute extensively to helping blind individuals. LookTel was presented in [8] to be an extensive platform that integrates state of the art computer vision techniques with mobile communication devices which return real-time banknote recognition results announced by a text-to-speech engine. OCR engines [4] were developed to transform the text information on camera-based scene image into readable text codes. Signage and text detection [30] were designed to extract abstract symbol information directly from natural scene images. Besides, some other recognition systems used FM sonar sensors [2] and camera-integrated walking canes [9] for blind navigation. Sensor modules could be used for searching tasks in the surrounding environments [16]. Some other prototypes for blind-assistant localization and recognition were introduced in [17, 18, 19]. These devices could offer an equivalent of raw visual input to blind people, via complex sounds capes (head-mounted camera and stereo headphones), thus leaving the recognition tasks to the human brain. Furthermore, many other techniques have been developed for efficient object detection and recognition. Jauregi *et al.* proposed a two-step algorithm based on region detection and feature extraction [10]. This approach aims to improve the extracted features by reducing unnecessary keypoints, and increasing efficiency through accuracy and computational time. Ta *et al.* presented an efficient algorithm for continuous image recognition and feature descriptor tracking in video [11].

2.   System Design

Our system consists of a wearable camera with multiple fixed cameras. Figure 1 illustrates the layout design of the system. Blind user is equipped with a wearable camera which is connected (wire or wireless) to a computer (PDA, or laptop), as shown in Figure 2. The user can send requests of finding an item by speech command and then wear the camera system to look for the item. When the system finds the requested item, an audio signal will be produced. A dataset of the personal items for the user is created as reference samples. In the dataset, multiple images for each item are captured for different camera views, scales, lighting changes, and occlusions.

Multiple cameras are fixed at the important locations where blind user probably leaves his/her items, and compose a blind-assistant network. When a blind user sends a request of object finding into the system, all the fixed cameras will start object recognition by comparing their captured objects with the reference objects in the dataset. Then the system will report the most similar object though matching distance and instruct blind user to get close to its location. Next, the camera attached to blind user will perform further recognition to certify the existence of the demanded object.
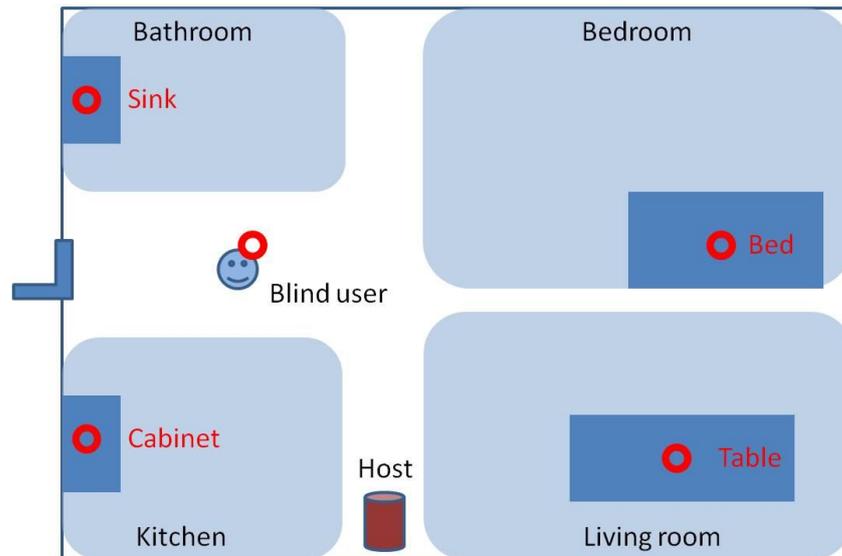
Figure 1: An illustration of the layout design of our prototype system. Cameras are fixed at the important locations where blind users usually leave their daily necessities. All the cameras have access to a local network, which is managed by a host.
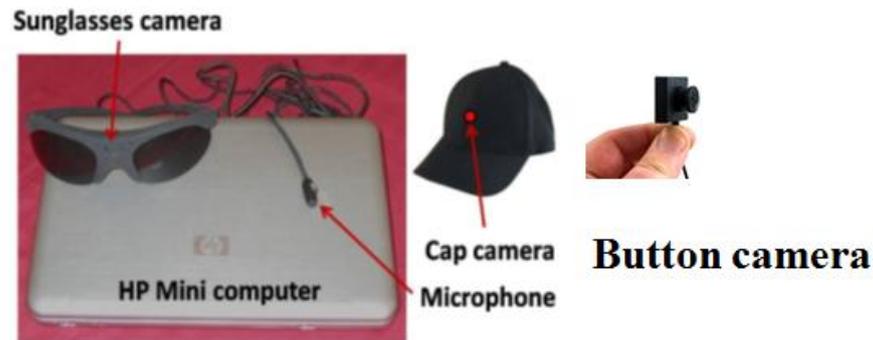

Figure 2: An illustration of the wearable devices for blind user, including sunglasses camera, mini-computer for data processing, and audio system to output recognition results.

The flowchart of our proposed object recognition algorithm is shown in Figure 3. First, based on the blind user's request, features from camera captured images are extracted by SURF or SIFT. Then, these features are compared to pre-calculated features from reference images of the request object in the dataset. If matches are found, the algorithm will output if the object has been found or not according to the pre-established thresholds for each object.
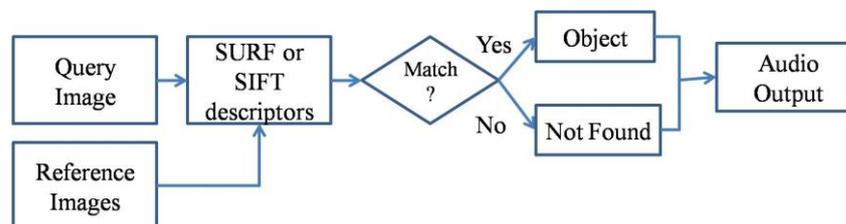

Figure 3: Flowchart of our proposed object recognition algorithm.

3.   Object Feature Extraction

A large variety of features can be extracted from camera-based scene image, so the combinations and selection of the features play an important role in data analysis [27, 28]. Interest point detectors and descriptors [5, 6] are able to extract representative and discriminative features

from reference images. Based on the state-of-the-art local feature descriptors SIFT [6] and SURF [5, 7], precise matching is allowed between images containing identical objects. Both SIFT and SURF are able to extract representative and discriminative keypoints from an image, which contain significant information of the object appearance and structure in the image.

SURF is a robust image detector and descriptor that can be used in computer vision tasks like object recognition or 3D reconstruction. The standard version of SURF is several times faster than SIFT. The SURF detector is based on the Hessian matrix which provides excellent performance in computational time and accuracy. Given point x(x,y) in an image I, the Hessian matrix H(x,σ) in x at scale σ is defined as Eq. (1).

$$\mathcal{H}(x,\sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix} \tag{1}$$

where $L_{xx}(x,\sigma)$, $L_{xy}(x,\sigma)$ and $L_{yy}(x,\sigma)$ are the convolution of the Gaussian second order derivative with image I in point x(x,y). Gausians are best for scale-space analysis, but in practice, they must be discretized and cropped. The detection of interest points in the image is determined by non-maximum suppression in a $3 \times 3 \times 3$ neighborhood.

On the other hand, the SURF descriptor is based on sums of approximated Haar wavelets responses within the circular interest point region, with radius $6S$, with S being the scale at which the interest point is detected. Since the wavelets are large at high scales, integral images are efficiently used for fast filtering. Once the wavelet responses are calculated, dominant orientation can be estimated within a sliding orientation window covering an angle of $\pi/3$. The horizontal and vertical responses originated from this window yield a vector, which lends it orientation to the interest point according to the longest one obtained. Square descriptors are then constructed around the interest points along the orientation determined by the vector. Each square then subdivides regularly into a smaller 4×4 sub-region keeping important spatial information in for the descriptor construction.

SIFT extracts local features by mapping gradient orientations within predefined blocks and cells into a histogram. It calculates the Difference of Gaussian (DOG) of image maps in scale space as Eqs. (2) and (3).

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \tag{2}$$

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma) \tag{3}$$

where x, y, and σ denote spatial coordinate and scale respectively, $G(x,y,\sigma)$ is a Gaussian filter with variable scale σ, and $D(x,y,\sigma)$ represent the DOG map.

The DOG function approximates the scale-normalized Laplacian of Gaussian, that is, there is $D(x,y,\sigma) \approx \sigma^2 \nabla^2 G$. Previous work shows that local maxima and minima of $\sigma^2 \nabla^2 G$ are the most stable image features in comparison with gradient, Hessian, Harris and so on. Thus, the local maxima and minima of $\sigma^2 \nabla^2 G$ are extracted as SIFT feature points.

The SIFT keypoints are in the form of oriented disks attached to representative structure of the objects in the image. The detected keypoints could keep invariant to transitions, rotations, scale changes, and other deformatins. They are then used as representative local features of the objects from image. Around each SIFT keypoint, a 4-by-4 block is defined and a histogram of gradient orientations is generated. Since gradient orientation is quantized into 8 values, the histogram has 8 corresponding bins. Then all histograms of the blocks are cascaded into a $16 \times 8 = 128$ dimensional vector, as shown in Figure 4. This feature vector will be used as SIFT feature descriptor at the keypoint.

SIFT descriptor plays an important role in many applications involved in content-based information retrieval, object matching, and object recognition *etc*. It is able to find out distinctive keypoints in images that are invariant to location, scale, rotation, affine transformations and illumination changes. Given two images, ones is a complete object with clean background and

5

standard viewpoint, and the other contains the same object under a different viewpoint, complex background or partial occlusion. We apply SIFT detector and descriptor to find out matches between the two images. Each match will be assigned a score based on the Euclidean distance between the SIFT descriptors of the two matched points. The number and score of keypoint matching will be used to design our algorithm of object recognition.
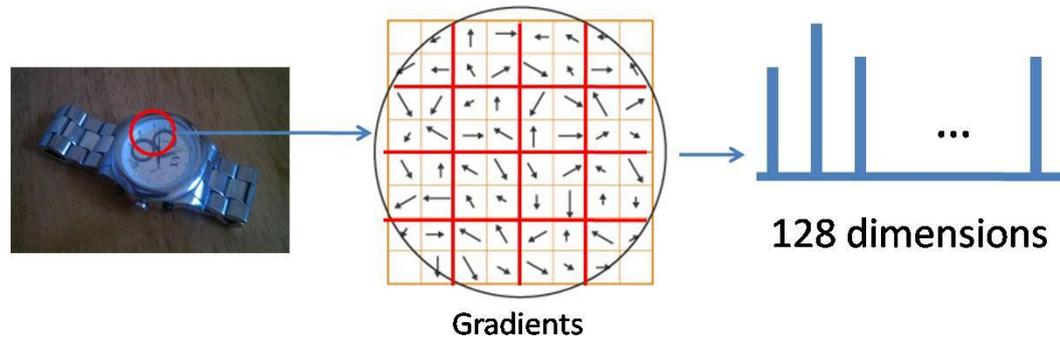
Figure 4. An illustration of SIFT descriptor. The left presents original image and a keypoint in the red circle. The middle presents the 16 blocks around the keypoint and their respectively gradient orientations. The right denotes a 128-dimenstional feature vector generated by the votes of quantized gradient orientations.

4. System Implementation

In order to effectively recognize objects, we collect a dataset of daily necessities as reference objects. This dataset contain personal items that are essential to visually-impaired individuals, such as keys and sunglasses. In addition, this dataset covers a variety of conditions such as image scale, translation, rotation, change in viewpoint, and partial occlusion. All these images are captured in the presence of cluttered backgrounds. Some examples are displayed in Figure 5.

4.1. Camera Network

Our proposed system is based on a network of cameras. A blind user sends his/her request of finding a specific object, and the system will start object recognition from the cameras to search their respective regions. Each camera will output a recognition score by comparing its captured objects with image samples of the reference object in the dataset. The score is calculated from the average matching distances of the SURF/SIFT key points. Thus several possible locations of the request object could be obtained from small matching distances.

In our system, all the cameras are connected to a local network for information share. Each camera reports the results of matching reference objects. A host takes charge of information collection and analysis. It will notify the blind user of the most probable locations of his/her expected objects.
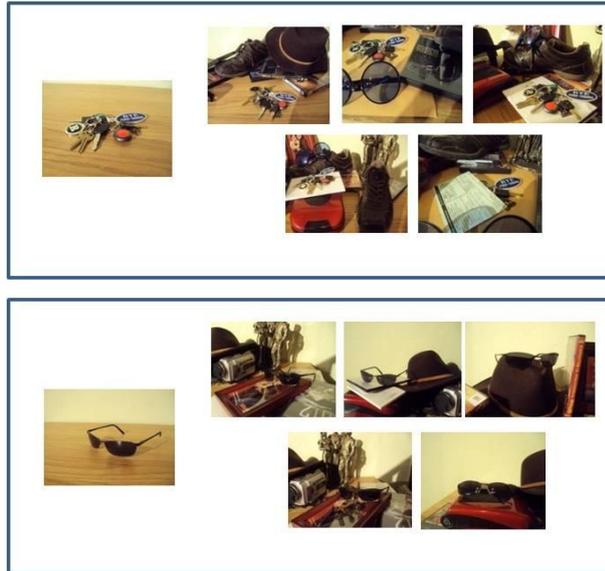
Figure 5: Samples of object images taken under different conditions. In each group, the first one is a standard reference image, and the later five respectively represent scaling change, transition, rotation, viewpoint change and partial occlusion.

4.2. Matching-based Recognition

To identify an object in every query image and without counting any false matches, thresholds are set up for every reference image based on experiments. In our system, a threshold level of 10 key matching points is employed for the cell phone, sunglasses and keys, and a threshold level of 25 key matching points is used for the Watch Wrist, and Juice Cup.

Figure 6 illustrates the detected interest points from different reference objects, marked in red circles. The same method is also applied to the query object. After the construction of descriptors around the interest points of reference and query images, SURF features from the reference images are extracted and matched against the features from the query images in our dataset. If the SURF features in the query image match the ones in the selected object reference image, then we employ the predefined thresholds to determine if the object is the expected item which the blind user looks for. Figure 8 illustrates that our algorithm can successfully predict the category of query image in the presence of background clutter, partial occlusion, and viewpoint changes.

Figure 6: Interest points of main reference images. For each group, the left one denotes original image and the right one denotes detected SURF features.

Next, SIFT descriptor is applied to the dataset for performance evaluation of object recognition through the keypoint matching. At first, we collect an object dataset of reference images, which consists of 10 categories of common-use objects, such as key, glasses, coffee cup, etc. To predict object category of the queried image I_Q, we compare it with each reference image I_R in the dataset. Then SIFT detector is applied to obtain two set of keypoints K_Q and K_R respectively from the two images. The corresponding 128-dimensional SIFT descriptor is calculated from each keypoint. Based on Euclidean distance between SIFT descriptor, each key point will be assigned a match from the other image. Firstly, a key point P ($P \in K_Q$) is measured to obtain its distance to each keypoint in K_R. If the keypoint P' ($P' \in K_R$) has the minimum distance to P, then they are regarded as a match of each other. Secondly, we only preserve the keypoints where the ratio between nearest neighboring distance and the second nearest neighboring distance is greater than 0.6. It is able to reduce false positives and increase robustness. The other keypoints and their corresponding matches will be removed. Thirdly, we calculate the mean distance $d(I_Q, I_R)$ from the rest keypoint matches. As above mention, I_R is one of the reference images in the dataset. Then we calculate all the mean distances from the query image I_Q to the sibling reference images of I_R, which belong to the same category, as Eq. (4),

$$L_C(I_Q) = \frac{1}{\|C\|} \sum_{I_R \in C} d(I_Q, I_R) \tag{4}$$

where $L_C(I_Q)$ represents the mean distances from query image to a category C of reference images, and $\|C\|$ represents the number of reference images in the category C. The minimum distance models similarity between the object in query image and those in the dataset, and it can be used to predict which category the object in query image belongs to. In the same way, we calculate the mean distance to each category in the dataset. Then the query image will be assigned to the category with minimum mean distance as Eq. (5),

$$C^* = \underset{C \in dataset}{\operatorname{argmin}} \ L_C(I_Q) \tag{5}$$

Figure 7 depicts the keypoint matching between the objects from identical category, but in different viewpoint, distance and background. This figure demonstrates that SIFT feature detector and descriptor can handle arbitrary transitions, rotations, and scale changes. It ensures that our system of object finding will not be influenced by relative positions of the blind user and the object, or the surrounding environments of the object.
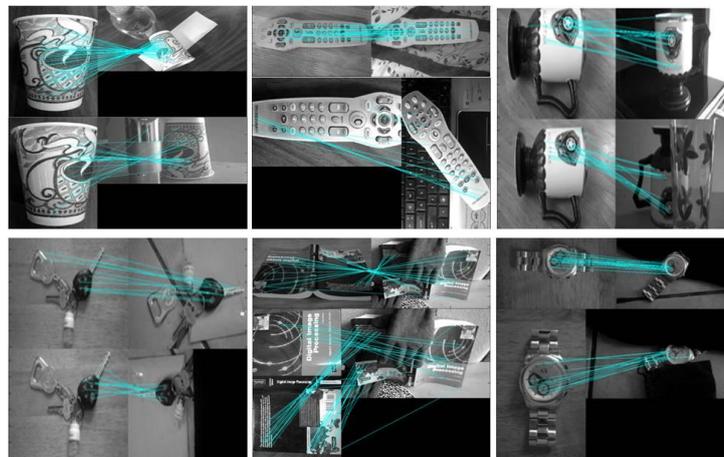

Figure 7. SIFT feature based interest points matching of main reference images in the dataset while blue lines represent matched keypoints.

5.   Experimental Results and Discussion

To evaluate the performance of SURF and SIFT-based object recognition, we collect a testing dataset for the 10 classes of daily necessities. Each class contains 10 image samples. It covers a variety of conditions such as image scale, translation, rotation, change in viewpoint, and partial occlusion.

TABLE 1: SURF-based recognition accuracy for each reference object

Objects	No. of Images	Correct detected	False Positive	Accuracy
Book	10	7	3	70%
TV-Control	10	5	5	50%
Antiseptic Pad	10	5	5	50%
Coffee Cup	10	8	2	80%
DVD Control	10	8	2	80%
Juice Cup	10	8	2	80%
Water Bottle	10	6	4	60%
Watch Wrist	10	10	0	100%
Keys	10	6	4	60%
Sunglasses	10	6	4	60%
Total	100	69	31	69%

The proposed algorithm can effectively distinguish different classes of objects from test images of the same objects in the presence of a cluttered background in different scenarios under a variety of conditions. When further testing is performed with multiple query images, the algorithm successfully identified objects as shown in Figure 8. The white lines demonstrate the traces from the reference image to the test image which represent the position where the matching features have been detected. If the number of the matching points is greater than the thresholds, the object is detected. Furthermore, there are no falsely identified items for all the classes of objects even when they are partially occluded by other objects.

We observe that under some situations, the algorithm fail to find enough matching points and identify an expected object when it is captured by the camera. Further, when the algorithm initially identifies an object of interest, based on the thresholds of each class of object, its matching features are not enough to correctly identify it. This happens due to the challenging conditions of some test images which are taken on a variety of conditions.

Figure 8. Matching between reference and test images in the presence of cluttered background under various conditions, keys (image scale) and sunglasses (translation). It shows that the SURF detector and descriptor can recognize the expected object on various environmental conditions.

SURF features are only robust to some degree of rotation and view point changes. These errors are caused by images with very large variations in illumination, background, scaling and/or rotation. Changes in illumination and cluttered background generate few matching points that would affect object recognition. The large scale change has been approved as the strong relief effect in [20, 21] and would also affect recognition. Rotation plays a big role as well, since the SURF descriptors are constructed around interest points whose largest orientation vector is within a sliding orientation window of $\pi/3$. Any orientation vector outside this range will result in mismatching. Therefore, if we were not dealing with images that contain extreme scaling changes and rotations, we would have a very robust and efficient object recognition algorithm.

Because we deal with images that contain a variety of conditions, the test accuracy for each class of objects is between 50% and 95%. Our algorithm achieves an average accuracy of 69% as shown in Table 1.

TABLE 2: SIFT-based recognition accuracy for each reference object

Objects	No. of Images	Correct detected	False Positive	Accuracy
Book	10	9	1	90%
TV-Control	10	8	2	80%
Antiseptic Pad	10	8	2	80%
Coffee Cup	10	9	1	90%
DVD Control	10	8	2	80%
Juice Cup	10	9	1	80%
Water Bottle	10	7	3	70%
Watch Wrist	10	6	4	60%
Keys	10	5	5	50%
Sunglasses	10	5	5	50%
Total	100	74	26	74%

Table 2 presents the results of object recognition. We can see the objects with discriminative surface texture, such as Book and Juice Cup, obtain higher accuracy of recognition. It is because SIFT descriptor is designed for texture matching. However, the recognition results of keys and sunglasses are lower, because their appearance depends much on the viewpoints of capture. The above two tables demonstrate that SIFT-based descriptor obtain better performance than SURF-based descriptor.

5.3   Comparisons between SURF-based recognition and SIFT-based recognition

The experimental results demonstrate that SURF-based recognition has higher efficiency, while SIFT-based recognition obtains higher accuracy over all testing object categories. SIFT detector extracts DOG maxima and minima as keypoints, which are more stable than SURF keypoints based on Hessian matrix. Besides, SIFT descriptor in our experiments has 128 dimensions, while SURF descriptor has only 64 dimensions. SIFT descriptor preserves more information of local features. From another aspect, SURF improves the computational efficiency of feature extraction, because 1) it lowers feature dimension compared to SIFT; 2) it simply sums first-order Haar wavelet responses to extract feature descriptor, instead of the statistics of gradients.

It is challenging tasks to combine SIFT and SURF, because the two detectors extract different groups of keypoints from an identical object, which correspond to different aspects of object appearance and structure. The simple cascade of the two descriptors cannot improve the recognition accuracy, but further lower the efficiency. In future work, we will integrate both SIFT and SURF into a histogram of visual words under the Bag-of-Words framework, and it will generate most robust and efficient object recognizer.

6.   Conclusion and Future Work

This paper presents a prototype blind-assistant system. It helps blind user to find their personal items in daily life through camera-based network and matching-based object recognition algorithm. We employ two types of local feature descriptors for detecting keypoint matches. The SURF and SIFT interest point detector and descriptor are scale-invariant and rotation-invariant, and provided our object recognition algorithm with ways to handle image scaling, translation, rotation, change in viewpoint, and partial occlusion between objects in the presence of cluttered backgrounds.

In order to evaluate the performance of our algorithm, a reference dataset is built by collecting daily necessities that are essential to visually impaired or blind people. Experimental results demonstrate that the proposed algorithm can effectively identify objects under conditions of cluttered background and occlusions without falsely identifying any reference object. However, based on pre-learned thresholds, the algorithm also failed to find enough matches to identify an object of interest when present in the image such as "cell phone" due to a lack of distinguishable features.

Our future work will focus on enhancing the object recognition system so that it can better detect and identify objects under extreme and challenging conditions. We will enhance cooperation of different cameras within the system, and address human interface issues for image capture and auditory display of the object recognition on computers and cell phones.

Acknowledgement

References

[1] American Foundation for the Blind, http://www.afb.org/

[2] G. Kao, P. Probert, and D. Lee, "Object Recognition with FM Sonar; An Assistive Device for Blind and Visually-Impaired People", *AAAI Fall Symposium* on ``Developing Assistive Technology for People with Disabilities," MIT, Cambridge, MA, 1996.

[3] G. Nikolakis, D. Tzovaras, and M. G. Strintzis, "Object recognition for the Blind", Proceedings of 13th European Signal Processing Conference (EUSIPCO2005), Antalya, Turkey, 2005.

[4] Mobile OCR, Face and Object recognition for the Blind - http://www.seeingwithsound.com/ocr.htm

[5] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded Up Robust Features", European Conference on Computer Vision, 2006.

[6] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints". International Journal of Computer Vision, 2004.

[7] SURF source code, http://www.vision.ee.ethz.ch/~surf/

[8] J. Sudol, O. Dialameh, C. Blanchard, T. Dorcey, "LookTel – A Comprehensive Platform for Computer-Aided Visual ssistance", IEEE Conference on Computer Vision and Pattern Recognition, 2010.

[9] R. Guide, M. Østerby, S. Soltveit, "Blind Navigation and Object Recognition", Laboratory for Computational Stochastics, University of Aarhus, Denmark, http://www.daimi.au.dk/~mdz/BlindNavigation_and_ObjectRecognition.pdf

[10] E. Jauregi, E. Lazkano and B. Sierra, "Object Recognition using Region Detection and Feature Extraction", Proceedings of Towards Autonomous Robotic Systems (TAROS), ISSN: 2041-6407, 2009

[11] D.N. Ta, W.C. Chen, N.Gelfand, and K. Pulli, "SURFTrac: Efficient Tracking and Continuos Object Recognition using Local Feature Descriptors", IEEE Conference on Computer Vision and Pattern Recognition, June 2009.

[12] J. Orwell, L. Lowey, and D. Thirde. Architecture and algorithms for tacking footable players with multiple cameras. In IEEE Proceedings of Vision, Image and Signal Processing, Vol. 152, No. 2, PP. 232-241, 2005.

[13] D. Marinakis and G. Dudek. Topology inference for a vision-based sensor network. In the Proceedings of Canadian conference on Computer and Robot Vision, pp. 121-128, 2005.

[14] A. Hoover and B. Olsen. Sensor network perception for mobile robotics. In IEEE International conference on Robotics and Automation, Vol. 1, pp. 342-347, 2000.

[15] D. Xie, T. Yan, D. Ganesan, and A. Hanson. Design and Implementation of a dual-camera wireless sensor network for object retrieval. In Proceedings of the 7th International Conference on Information processing in sensor networks, pp. 469-480

[16] A. Hub, J. Diepstraten, T. Ertl, "Design and Development of an Indoor Navigation and Object Identification System for the Blind", *In Proceedings of ASSETS*, pp147-152, 2004.

[17] A. Hub, T. Hartter, T. Ertl, "Interactive tracking of movable objects for the blind on the basis of environmental models and perception oriented object recognition methods", In Proceedings of ASSETS, pp111-118, 2006.

[18] S. Gehring, "Adaptive Indoor Navigation for the Blind", in Proc. GI Jahrestagung (1), pp.293-294, 2008.

[19] B.Bobo, R. Chellapa, C.Tang, "Developing a Real-Time Identify-and-Locate System for the Blind", Workshop on Computer Vision Applications for the Visually Impaired (2008)

[20] Guoshen Yu and Jean-Michel Morel. ASIFT: An Algorithm for Fully Affine Invariant Comparison. *Image Processing On Line*, 2011.

[21] ASIFT: An Algorithm for Fully Affine Invariant Comparison. http://www.ipol.im/pub/algo/my_affine_sift/

[22] G. Kreiman, "Biological object recognition," Scholarpedia, 3(6):2667, 2008. http://www.scholarpedia.org/article/Biological_object_recognition.

[23] Biederman, I. 1987, "Recognition-by-Components: A Theory of Human Image Understanding," Psychological Rev., vol. 94, pp. 115-147, 1987.

[24] M. Potter and E. Levy, Recognition memory for a rapid sequence of pictures. Journal of Experimental Psychology, 1969. 81: 10-15.

[25] S. Thorpe, D. Fize and C. Marlot, Speed of processing in the human visual system. Nature, 1996. 381: 520-522.

[26] W. Tang, and D. Su. Locomotion analysis and its applications in neurological disorders detection: state-of-art review. Network modeling analysis in health informatics and bioinformatics, 2012.

[27] Y. Xiang, D. Fuhry, K. Kaya, R. Jin, U. Catalyurek, and K. Huang. Merging network patterns: a general framework to summarize biomedical network data. Vol 1, No 3, pp. 103-116, 2012.

[28] J. Husle, M. Khoshgoftaar, A. Napolitano, and R. Wald. Threshold-based feature selection techniques for high-dimensional bioinformatics data. Network modeling analysis in health informatics and bioinformatics, Vol 1, No. 1-2, pp. 47-61, 2012.

[29] C. Hung, G. Kreiman, T. Poggio and J. DiCarlo, Fast Read-out of Object Identity from Macaque Inferior Temporal Cortex. Science, 2005. 310: 863-866.

[30] S. Wang, C. Yi, and Y. Tian. Signage detection and recognition for blind persons to access unfamiliar environment. Journal of Computer Vision and Image Processing, Vol. 2, No. 2, 2012.