Journal of The Korea Society of Computer and Information
Vol. 24 No. 6, pp. 67-72, June 2019

www.ksci.re.kr
https://doi.org/10.9708/jksci.2019.24.06.067

A Computer Vision-Based Banknote Recognition System for the Blind with an Accuracy of 98% on Smartphone Videos

Gustavo Adrian Ruiz Sanchez*

Abstract

This paper proposes a computer vision-based banknote recognition system intended to assist the blind. This system is robust and fast in recognizing banknotes on videos recorded with a smartphone on real-life scenarios. To reduce the computation time and enable a robust recognition in cluttered environments, this study segments the banknote candidate area from the background utilizing a technique called Pixel-Based Adaptive Segmenter (PBAS). The Speeded-Up Robust Features (SURF) interest point detector is used, and SURF feature vectors are computed only when sufficient interest points are found. The proposed algorithm achieves a recognition accuracy of 98%, a 100% true recognition rate and a 0% false recognition rate. Although Korean banknotes are used as a working example, the proposed system can be applied to recognize other countries' banknotes.

▸Keyword: Banknote recognition, Real-time, Computer Vision, Video, SURF

I. Introduction

There are 36 million blind people worldwide [1] and 252,632 blind people just in Korea [2]. We all have to deal in our daily lives with banknotes to do our financial transactions. Unfortunately for the visually impaired, this simple activity is impossible to perform without relying on the goodwill of the counterpart in the transaction, giving that differences of size and texture of the bills are not enough for enabling the blind to discriminate the denomination of the banknotes.

Fortunately, computer vision applications on mobile devices can provide tools for the blind to make their lives easier and to allow them to become more independent. This paper proposes a novel method for the recognition of banknotes on mobile devices for the blind. Fig. 1 shows the system's schematic. First, the proposed system segments the foreground from a video sequence hoping to find the banknote, the resulting image is pre-processed using morphology operations and its Speeded-Up Robust Features (SURF) [3] keypoints and descriptors are computed to be matched with the descriptors stored in the database extracted from banknote template images. Finally, the banknote denomination corresponding to the template with the higher number of good matches over a threshold is given as the recognition output.

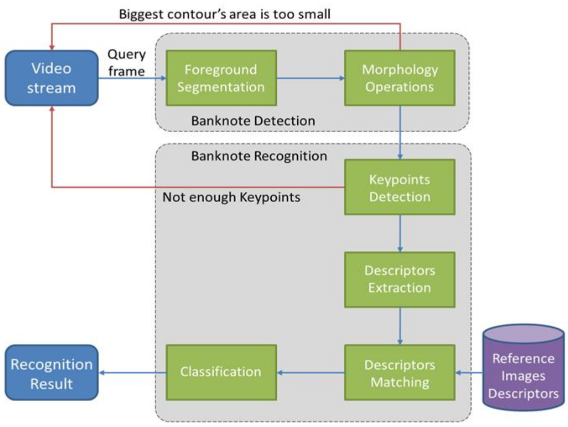


Fig. 1. A conceptual schematic of the banknote recognition system

• First Author: Gustavo Adrian Ruiz Sanchez, Corresponding Author: Gustavo Adrian Ruiz Sanchez
*Gustavo Adrian Ruiz Sanchez (gustavo.ruiz@kmu.ac.kr), Dept. of Computer Engineering, Keimyung University
• Received: 2019. 04. 23, Revised: 2019. 05. 28, Accepted: 2019. 05. 29.

II. Preliminaries

1. Related works

Although banknote recognition research is less studied than other computer vision areas, the problem has been addressed in multiple researches with relative success as shown in [4]. Next, several works are introduced as a review of the state of the art on banknote recognition.

An embedded system was designed in [5] to assist blind people in recognizing Australian banknotes. The device called the MoneyTalker uses the reflection and transmission properties of the light to take advantage of the largely different colors and patterns on each of the Australian bank notes. Different color lights are transmitted through the inserted banknote and the amount of light that goes through the banknote is measured by light sensors. These values are later compared with the values stored in memory from training examples in order to classify the banknote.

Using computer vision techniques on mobile phones, [6] presented an application for recognizing Indian National Rupee banknotes. The system utilizes an iterative graph cuts algorithm for segmenting the banknote from the cluttered background, and a visual Bag of Words (BoW) based method for recognition.

To assist visually impaired people, a computer vision-based system for the automatic recognition of American dollar banknotes was proposed in [7]. The system utilizes a component-based framework based on SURF for recognition of the banknotes.

With the aim of allowing blind people to detect and recognize Euro banknotes, [8] developed a portable system which utilizes a modified Viola and Jones algorithms for detection of the banknotes, while the banknote recognition algorithm relies on the SURF technique.

To recognize the Mexican banknotes, [9] uses the banknotes' color and texture features described via the RGB space and the Local Binary Patterns (LBP) respectively, this work is done under the assumption that there are no illumination variations between the banknotes' images on the training and testing set.

For the Series 7 New Zealand banknotes [10], with the assumption of no illumination variations, banknotes are recognized utilizing color and texture features described with a composite feature vector containing the HSV color histogram and the uniform LBP histogram. Principal Component Analysis (PCA) is utilized to reduce the dimensionality of the composite feature vector.

In contrast to traditional approaches using handcrafted features, deep learning based banknote recognition highly reduced the dependence on appearance-physics based models and other pre-processing techniques by enabling "end-to-end" learning to occur in the pipeline directly from the input images [11].

As a recent algorithm using the deep learning technique, [12] proposed a banknote classification method that simultaneously classifies banknotes from multiple national currencies in both types and fitness levels using the combination of infrared transmission (IRT) and visible-light reflection (VR) images of the input banknote and Convolutional Neural Networks (CNN).

The above-mentioned researches focused on analyzing single pictures taken by the user. However, in real-life scenarios it can be challenging for a blind user to take pictures on focus or even to take pictures of a banknote inside the camera frame, forcing the user to take multiple pictures before a satisfactory recognition is obtained. In addition, although very deep or wide networks-based banknote recognition approaches usually perform reasonably well, they still require amount of processing time and memory consumption in the training and inference processes.

Therefore, this paper operates on a dataset of videos obtained through a smartphone and proposes a banknote recognition algorithm based on a SURF keypoints matching algorithm capable of operating in low specification smartphone devices while achieving a comparable recognition performance with fast speed.

III. The Proposed Scheme

The proposed approach to the problem is different from the above introduced researches as instead of analyzing a single picture; this study works on a video stream. Foreground segmentation is performed on the video stream and SURF features are extracted and matched only if a banknote appears on camera and has a sufficient number of keypoints to be successfully recognized.

1. Database of reference images

The reference database consists of 8 images in which discriminative regions are carefully selected. Front and back images of the four different Korean banknote denominations (₩50000, ₩10000, ₩5000, and ₩1000). From each of these images, 64-dimensional SURF descriptors [3] are computed only over a discriminative region of the banknote as indicated by the red boxes in Fig. 2. This region contains the most discriminative information of the banknote while having a smaller area; obtaining a speed up in the matching process and more robustness against geometric transformations and photometric changes.

Fig. 2. Reference regions on Korean banknotes indicated by red boxes

2. Video test dataset

A dataset was created consisting of 100 short videos of a duration ranging from three to five seconds and a resolution of 720 x 480 pixels. The videos were recorded under conditions the system will be subject to in real-life situations, meaning that the banknote may appear blurred, partially occluded, over cluttered background, rotated, scaled, and under different illuminations, also negative images like banknotes from other currencies and different objects are included. At the beginning of the videos the banknote is out of the camera frame, later it is introduced to disappear again from the camera's view. Table 1 shows the dataset structure and Fig. 3 shows some sample frames from the video dataset.

Table 1. Video test dataset used in the experiments

Banknote	No. of videos
₩50000	20
₩10000	20
₩5000	20
₩1000	20
OTHER	20

Fig. 3. Sample frames from the video test dataset

3. Foreground segmentation

The goal of a foreground segmentation algorithm is to come to a binary decision indicating which pixels belong to the foreground and which to the background. As stated by [6] the segmentation of a banknote from the image is not just important for reducing the data to process and therefore reducing the processing time, but also for reducing irrelevant features that could affect the recognition results introducing false matches. To cope with camera jitter and/or slight illumination changes, a technique called Pixel-Based Adaptive Segmenter (PBAS) [13] is utilized in this study. A non-parametric approach in which the background is modeled by an array of recently observed color pixel values. The foreground decision depends on a decision threshold and the background update is based on a learning parameter. To speed up the segmentation, the size of the video frame is reduced by four. Then, some morphology operations are applied to the segmented images for removing small contours, and if the biggest contour area is bigger than 1000 pixels, the next stage is processed; otherwise, the frame is discarded for further processing. Fig. 4 shows a video frame and its corresponding foreground binary matrix result of the segmentation algorithm.

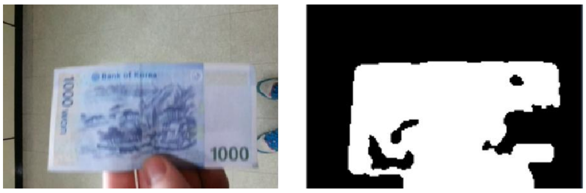


Fig. 4. Input video frame (left) and it's corresponding foreground binary matrix (right)

4. Best-fitted frame selection and feature extraction

To select the best-fitted frames of the input videos, the keypoints on the resulting image from the banknote detection stage are detected, and if the number of keypoints is bigger than the experimentally found threshold, descriptor computation proceeds; otherwise the image is discarded, preventing further extra processing. This study uses SURF as a blob detector and descriptor because it is largely invariant to changes in scale, rotation, illumination, and affine distortions. SURF implements various performance optimizations like the use of "Integral images" [14] obtaining similar results in performance with more robust descriptors but in a much shorter time, characteristic that is desirable for real-time or mobile applications.

5. Descriptor matching

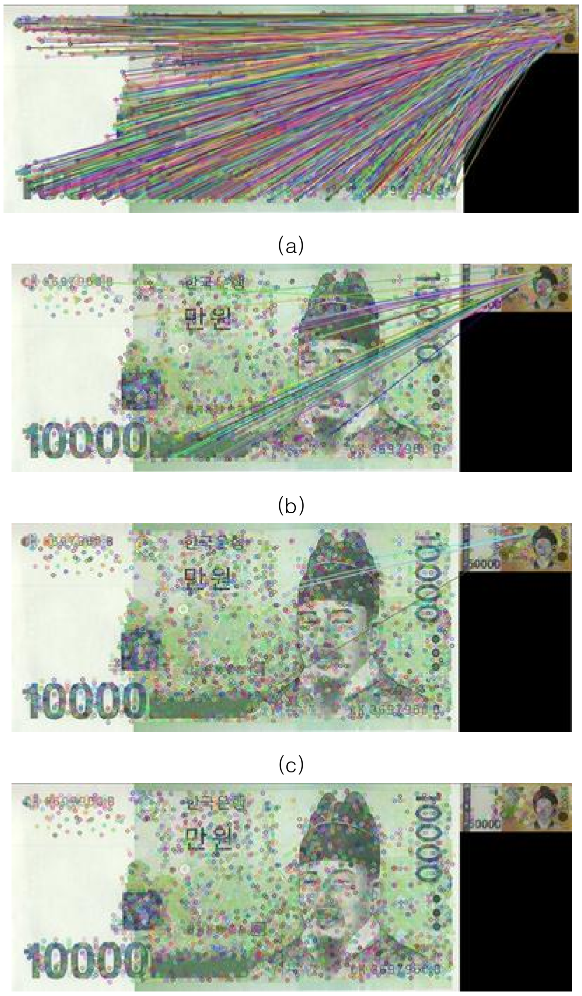
This study finds the two Nearest Neighbors from the database for each 64-dimensional SURF descriptor of the query image. Later a ratio test is performed [15] to calculate the ratio of distance between the two closest nearest neighbors. This study uses it as a threshold.

Rejecting all the matches with a distance ratio greater than 0.8 eliminates 90% of the false matches and only 5% of correct matches [15]. Moving this ratio of distances below the 0.8 threshold, the number of correct matches-discarded rapidly increases but the number of incorrect matches-discarded decreases at a much more smaller rate. This study chooses a threshold of 0.6 which is the minimum reasonable threshold.

Despite the above thresholds, there still remain a considerable amount of false matches that can possibly lead to misclassifications. Even if this study had decided to use a 0.5 threshold there would still exist misclassifications, with a threshold as low as 0.4 there would not be misclassifications but too many banknotes would be ignored.

Therefore in case that one feature matches with more than one feature, it is discarded. Moreover, this study verifies the quality of the matches by discarding all matches with Euclidean distance bigger than 0.2.

It was demonstrated by [16] that it is possible to verify the matches between images by thresholding on the Euclidean distance between features; this is due to the fact that the Euclidean distance between true matches is generally smaller than the distance between false matches. This study found that for purposes a threshold of 0.2 gave the best results, as despite reducing the number of correct matches, it removed completely any false detection while choosing a bigger threshold like 0.3 could lead to occasional false detections. Fig. 5 shows how the false matches are discarded as the thresholds described above are introduced.



(a)

(b)

(c)

(d)

Fig. 5. Rejection of SURF descriptor false matches.
(a) SURF descriptor matches
(b) After applying a 2-KNN ratio test (0.6)
(c) After discarding matches with Euclidean distance > 0.2
(d) After discarding features with more than 1 match

This study uses the Fast Library for Approximate Nearest Neighbors (FLANN) [17], this library allows to speed-up the matching of these high dimensional vectors by automatically choosing the best algorithm (K-D Tree, Randomized K-D Tree or Hierarchical K-Means Tree) and optimum parameters for the dataset.

6. Classification

The recognition result of this study is given by the banknote denomination associated with the reference image which has the largest number of matches over a threshold obtained by analyzing the whole video dataset. From Fig. 6 which shows the descriptor matches on a video of the dataset showing the front of a Chon-won (₩1000) banknote, it is seen how introducing a threshold over 2 matches can completely eliminate misclassifications.

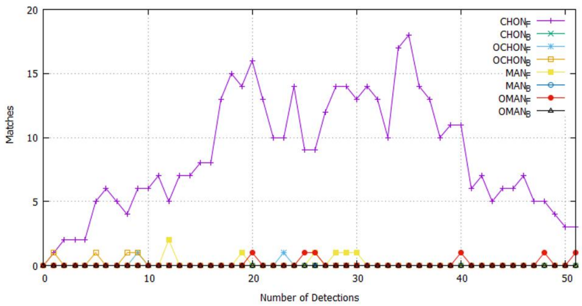


Fig. 6. Descriptor matches during a video sequence showing the front of a Chon-won (₩1000) banknote

IV. Experiments and results

We tested the performance of our algorithm on our test database of 100 real-life short videos of 480x720 pixels and an average duration of four seconds, using a computer with an Intel® Core™ i5 CPU 760 @ 2.80GHz CPU.

This study achieved a recognition accuracy of 98%, a 100% true positive rate and a 0% false positive rate. In two videos out of the 100 videos there was no recognition when there actually was a bill shown in the video, this explains the 10% false negative rate for ₩5000 banknotes in Table 2.

Table 2. Banknote recognition results

Banknote	No. of videos	False Positive Rate (%)	True Positive Rate (%)	False Negative Rate (%)
₩50000	20	0	100	0
₩10000	20	0	100	0
₩5000	20	0	100	10
₩1000	20	0	100	0
OTHER	20	0	100	0

In average, the system processed a video frame in 0.094 seconds (11 FPS) and processed a video frame in a median of 0.019 seconds (53 FPS). The reason being that frames with no banknotes present were quickly discarded. Taking into account only the video frames which were fully processed (SURF features computed and matched); we achieved an average of 0.205 seconds (5 FPS).

In addition, as a result of the foreground segmentation stage, the size of the query image from which to calculate the SURF descriptors was reduced in average to an image of 480x358 pixels. A reduction of more than half of the original video frame size, directly contributing to speeding up the system.

Deep learning-based methods are not suitable for low-specification systems such as smartphones running on CPUs instead of high-end GPUs to run in real time. For example, the AlexNets [18] method achieves a similar performance but requires 61 million parameters and over 1 million operations, which is not suitable for execution on a smartphone.

V. Conclusions

This paper succeeded to implement a system for banknote recognition to assist the blind. The system is highly accurate and fast in recognizing banknotes present on videos recorded with a smartphone on real-life scenarios. This level of performance was achieved by video foreground segmentation and computing the SURF descriptors just on those images with a sufficient area and a minimum number of keypoints. By changing the database of reference images, this system is expected to be able to recognize other countries' banknotes with a similar level of accuracy and speed. Future work will focus on creating an embedded device for multinational banknote recognition utilizing the method presented here.

REFERENCES

[1] "Blindness and vision impairment," http://www.who.int/mediacentre/factsheets/fs282/ (accessed Apr. 10, 2019)

[2] "Number of Registered Disabled Persons," http://kosis.kr/statHtml/statHtml.do?orgId=117&tblId=DT_11761_N004&conn_path=I2 (accessed Apr. 10, 2019)

[3] Bay, H., et al., "Speeded-up robust features (SURF)," Computer vision and image understanding, Vol. 110, No. 3, pp. 346-359, Nov. 2008.

[4] Pawade, D., P. Chaudhari, and H. Sonkambale, "Comparative study of different paper currency and coin currency recognition method," International Journal of Computer Applications, Vol. 66, No. 23, Jan. 2013.

[5] Hinwood, A., P. Preston, G. Suaning, and N. Lovell, "Bank note recognition for the vision impaired," Australasian Physics & Engineering Sciences in Medicine, Vol. 29, No. 2, pp. 229-233, June 2006.

[6] Singh, S., S. Choudhury, K. Vishal, and C.V. Jawahar, "Currency recognition on mobile phones," 2014 22nd International Conference on Pattern Recognition, IEEE, pp. 2661-2666, Aug. 2014.

[7] Hasanuzzaman, F.M., X. Yang, and Y. Tian, "Robust and effective component-based banknote recognition for the blind," IEEE Transactions on Systems, Man, and Cybernetics, Vol. 42, No. 6, pp. 1021-1030, Apr. 2012.

[8] Dunai Dunai, L., Chillarón Pérez, M., Peris-Fajarnés, G., and Lengua Lengua, I., "Euro Banknote Recognition System for Blind People," Sensors, Vol. 17, No. 1, pp. 1-14, Jan. 2017.

[9] García-Lamont, F., J. Cervantes, and A. López, "Recognition of Mexican banknotes via their color and texture features," Expert Systems with Applications, Vol. 39, No. 10, pp. 9651-9660, Aug. 2012.

[10] Ren, Y., M. Nguyen, and W.Q. Yan, "Real-Time Recognition of Series Seven New Zealand Banknotes," International Journal of Digital Crime and Forensics (IJDCF), Vol. 10, No. 3, pp. 50-65, July 2018.

[11] Ko, B.C., "A Brief Review of Facial Emotion Recognition Based on Visual Information," Sensors, Vol. 18, No. 2, pp. 1-20, Jan. 2018.

[12] Pham, T.D., D.T. Nguyen, C. Park, and K.R. Park, "Deep Learning-Based Multinational Banknote Type and Fitness Classification with the Combined Images by Visible-Light Reflection and Infrared-Light Transmission Image Sensors," Sensors, Vol. 19, No. 4, pp. 1-28, Feb. 2019.

[13] Hofmann, M., P. Tiefenbacher, and G. Rigoll. "Background segmentation with feedback: The Pixel-Based Adaptive Segmenter," Computer Society Conference on Computer Vision and Pattern Recognition Workshops, IEEE, June 2012.

[14] Viola, P. and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, Dec. 2001.

[15] Lowe, D.G., "Distinctive image features from scale-invariant keypoints," International journal of computer vision, Vol. 60, No. 2, pp. 91-110, Nov. 2004.

[16] Zhou, W., H. Li, and Q. Tian, "Multimedia Content-Based Visual Retrieval," Academic Press Library in Signal Processing, Vol. 5, pp. 383-416, Jan. 2014.

[17] Muja, M. and D. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," International Conference on Computer Vision Theory and Applications VISAPP, Vol. 2, pp. 331-340, Feb. 2009.

[18] Krizhevsky, A., I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, pp. 1097-1105, Dec. 2012.

Author

Gustavo Adrian Ruiz Sanchez received the B.S. in Mechatronics Engineering from Chiapas Polytechnic University, Mexico, in 2011, and M.S. in Computer Science from Yonsei University, Korea, in 2015. Mr. Ruiz joined the faculty of the Department of Computer Engineering at Keimyung University, Daegu, Korea, in 2018. He is currently an Associate Professor in the Department of Computer Engineering, Keimyung University. He is interested in computer vision, sensor data analytics and machine learning.