Exchanging Cash with no Fear: A Fast Mobile Money Reader for the Blind

Nektarios Paisios* Alexander Rubinsteyn Lakshminarayanan Subramanian

ABSTRACT
Despite the rapidly increasing use of credit cards and other electronic forms of payment, cash is still widely used for everyday transactions due to its convenience, perceived security and anonymity. However, the visually impaired might have a hard time telling each paper bill apart, since, for example, all dollar bills have the exact same size and, in general, currency bills around the world are not distinguishable by any tactile markings. We experiment with the use of a broadly available tool, the camera of a smart-phone, and several methods of classifying SIFT key-points to recognize partial and even distorted images of paper bills. Our results show that our system can be used in real-world scenarios to recognize unknown bills with a high accuracy.

Keywords
Camera Phone, Currency Identification, Visually Impaired.

Categories and Subject Descriptors
K.4.2 [**Computers and Society**]: Social Issues - Assistive Technologies for Persons with Disabilities; H.5.2 [**User Interfaces**]: User Centered Design, Prototyping.

General Terms
Algorithms, Design, Performance, Experimentation

1. INTRODUCTION
The ability to correctly and effortlessly exchange cash is fundamental for our everyday needs. Many simple transactions which are concluded each day, such as buying groceries or dining at a restaurant, require not only the handing over of paper bills but, most often than not, the return of some change. Unfortunately, in certain countries, visually impaired individuals cannot distinguish banknotes of different denominations from one another. In the United States, for example, dollar bills of different denominations all have the same size, shape and color. They feel exactly the same using the sense of touch. This has been a source of anguish and even dispute amongst the visually impaired community, sparking a lawsuit against the government of the United States, a lawsuit which the government recently lost [3]. Visually impaired people are dependent on the goodwill, the honesty and kindness of others to help them organize and recognize their banknotes. Given the central role that money and more specifically cash plays in our social and professional lives, this dependence on others places the visually impair at a stark disadvantage. To overcome this dependence, the visually impaired have come up with various forms of coping methods which can help them, to an extent, deal with the situation. These practical but often fragile methods include folding some bills in their wallets while keeping those of other denominations straight, or even storing different denominations in different pockets. These procedures are easy to mix up and forget to carry out, given that they must be performed with great detail and care every time a new banknote has been received. Relying on the assistance of others when using these currencies, has become therefore a fact of life for most visually impaired individuals, robbing their independence.

*Computer Science, New York University. Email:{nektar,alexr,lakshmi}@cs.nyu.edu

The popularity of smart phones and feature phones with built-in high-resolution cameras can provide a possible solution to the currency recognition problem. In this work, we propose a smart phone based solution that leverages the phone's camera in conjunction with image processing techniques to recognize currency bills. The system takes a continuous stream of images of a currency bill placed in front of the camera's lens and analyzes the provided images to recognize the denomination of the banknote. It subsequently uses synthetic speech to announce the results of its recognition. The software employs the Scale Invariant Feature Transform (SIFT) [6] for feature extraction but uses a faster approach for feature classification so that it can work effectively on the mobile phone. Our system yields high classification accuracy for real pictures of currency bills taken by a visually impaired user through the phone's camera.

2. PROBLEM SETTING
We face several practical and algorithmic challenges for recognizing currency bills on a mobile device. Some of the practical challenges include:

Size: In the United States, dollar bills of different denominations have all the same size and so cannot be easily separated by measuring or comparing one against another. In 120 other countries, including the countries which have adopted the Euro, paper currency comes in different sizes [9]. Despite different currency bill sizes, it might still not be easy to separate them quickly and effortlessly especially if the person has low finger dexterity, a difficulty faced by many seniors [2].

No tactile marks: To make things worse, there are no tactile markings or other forms of identification on each bill to enable a blind person to discover without any sighted assistance its value. The various engravings or embossed pictures currently found on such paper bills are too thin and lightweight to be felt by the sense of touch [3].

Problems with markings: Countries which have used tactile markings have discovered that such markings disappear after bills have been in circulation for a while as they tend to wear out with time and the extensive usage of the bill. Employing a pattern of distinctive holes or cut corners is undesirable, as this could lessen the lifetime of the bill by allowing it to be more easily torn.

Electronic devices: Though electronic devices which can identify dollar bills are available on the market, these devices can cost up to $300 and are not always accurate, especially with older or worn out bills. Meanwhile, visually impaired individuals do not want to carry around with them yet another specialized gadget just to help them distinguish their paper currency.

From the algorithmic standpoint, the image or video captured by a mobile phone camera can be of poor quality due to several factors: (a) Poor image quality due to focus and lighting issues; (b) Camera positioning issues which can result in partially captured images (not containing the entire currency bill); (c) Apart from camera alignment, the currency bill can be crumpled or folded. When dealing with visually impaired users, we have to make very limited assumptions regarding the quality of the images and make the system design robust.

3. RELATED WORK

In prior work [5, 4, 7, 8], the authors have attempted to use very specific and specialized techniques when tackling the problem of unhindered access to paper currency. However, any specialized technique can be fragile and hard to port from one type of currency to another. The AdaBoost learning algorithm was used in [5, 4] to train a set of weak classifiers in order to determine the value of dollar bills. These weak classifiers consisted of a set of specific pixel pairs on each bill which were picked by the researchers. However, even though this system works on both faces of the bill, it seems that it still requires the particular face to be exposed fully to the camera and so would not work on folded bills or bills with distortion. The blind user would still need to orient each bill and take care where the phone's camera was being pointed, a procedure hard to perform when standing in line at a store. In comparison to a robust and tried image recognition algorithm which can reliably extract distinguishing features from any image presented, the authors' procedure of selecting numerous distinct pixel pairs for each bill cannot easily scale, especially when having to port the system to more currencies of other countries. Similarly, in [8], specific regions which contain "characteristic saliencies that differ enough from one bill to the other" were selected. These characteristic image regions were also rotated in order to create more training samples which were then fed to an image recognition algorithm. However, although identifying bills under multiple orientations is very useful to a visually impaired individual, the distinguishing regions for each bill are still picked by hand and do not cover the whole of the bill's surface. In fact, users of the system had to first locate the plastic strip on each bill before taking a picture of that specific area in order that the recognition would work. The idea of finding specific unique characteristics which could easily separate one banknote from another was taken a step further in [7]. The authors employed an algorithm which would help the user move the bill until the denomination's value printed on it would be exposed to the camera. This procedure of asking the user to move the banknote around until a region of interest has been located, however, can be tedious and time consuming. Also, the algorithm for recognizing the printed denomination would need to be adapted extensively for it to work on other currencies. Using real humans as a part of the image recognition process was proposed in [10]. In this work, images are first labeled by using an automatic approach such as a database or an image search engine and the results are validated in a distributed manner by paid human workers through Amazon Mechanical Turk. This system can certainly be adapted and used by the visually impaired to identify their banknotes. However, the use of human workers means that the users would need to incur at least some financial cost, a fact which can be avoided by building a completely automatic and high-accuracy currency recognition system which could also run without any connection to an online service.

4. SYSTEM DESIGN

Our system works by using a set of sample images of dollar bills which are used to train a set of classification algorithms. The system is not trained by hand and does not rely on any hand-picked distinguishing characteristics usually found on such bills. Instead a more robust machine learning approach is followed whereby the training data is used to guide the algorithm in recognizing similar bills when they are later presented to it by the visually impaired user. Our current system design focuses on US currency bills but our techniques can be easily extended for other currencies.

4.1 Image Pre-Processing

Our system stores and uses an array of training images for each supported denomination, ($1, $5, $10 and $20) that are captured by a visually impaired user (same for the testing set). Each image from both the training and testing samples is first turned into a raw format by uncompressing the original pictures produced by the phone's camera. Each image is scaled so that its height is 300 pixels, whilst its width is scaled in an equal amount. A 200 pixel white border is added around the image. For experimentation and to account for image distortions, we optionally create additional artificial training images by taking the existing images and rotating them through 90, 180 and 270 degrees, in addition to scaling each one by 0.5 and by 1.5 its original size. Finally, for all the training images, an implementation of the

SIFT algorithm is used to extract a collection of key-points for each one of them.

Fixing lighting variations: Images captured under light conditions with different colors will look much different from one another. To handle this problem, we use the Gray World Assumption to remove the effects of the light source, [1]. This assumption uses the fact that in any realistic and thus sufficiently diverse scene, we expect to have a diverse amount of colors and color differences which should, on a very simplistic level, average out to gray. Given that each pixel has 3 color components in the RGB representation, we adapt the image to have on average an approximate gray value for each of its 3 color components. The product of the mean values of each red/green/blue (RGB) color component is use to calculate the value of the gray color for that image.

4.2 Using SIFT Descriptors

We employ an adaptation of the Scale Invariant Feature Transform (SIFT) algorithm [6] which is widely used in the vision community for detecting similarities across images. The SIFT algorithm identifies *key-points* or *descriptors* within any image and generates a multi-dimensional feature vector representation of an image taking into consideration a large number of image-specific and image-agnostic parameters. SIFT is an ideal choice since it can perform robust classification in the face of positional, orientation, rotation and scale variants.

The collection of SIFT key-points for each training sample, along with the sample's denomination, are transferred in bulk from memory into binary files which could then be loaded quickly by the phone's software. When instructed by a visually impaired user to recognize a banknote, our system takes snapshots continuously from the phone's camera which are pre-processed 4.1 and are fed into SIFT for key-point extraction. The extracted key-points from each captured image are then compared with the key-points of the training samples and are classified according to denomination. The more key-points that the system classifies as belonging to a specific denomination, the higher the confidence that the object currently being photographed is a banknote of that specific denomination. After a sufficiently high confidence value has been attained, the system stops taking snapshots and announces, using the phone's built-in synthetic speech, the denomination of the recognized banknote. On the other hand, if the object being captured is not a banknote, the system's classification confidence will not reach the required built-in threshold. The system will continue taking snapshots (up to a certain bounded time) and nothing will be announced until a banknote is put before the camera. The classification confidence measure should indicate how close the image to be recognized belongs to a certain currency denomination.

To compute the confidence measure, we first find a collection of measures each one indicating the distance of the captured image to each class. This can simply be the ratio of the image's key-points which have been classified as belonging to each particular denomination. We compute the confidence of a particular currency denomination as a function of the distance measures to the training images in the same denomination.

Comparing and classifying the SIFT key-points on a mobile phone is a time-consuming process as it potentially involves performing floating-point calculations on thousands of key-point pairs belonging to the training sample set and the images to be recognized. This might make the operation slow to the point of being infeasible, since the input data from the phone's camera will arrive at a faster speed than it can be processed. However, achieving high classification accuracy is paramount and it should not be sacrificed for faster performance. Next, we describe our attempts of achieving high classification accuracy and subsequently outline how to deal with the performance issue.

4.2.1 Nearest Neighbor

The simplest approach for determining the class of the SIFT key-points of the image to be recognized (testing image) is to employ an approach whereby each key-point is assigned the class of the closest key-point in the training set. Firstly, for all the key-points of the testing image, the Euclidean distance, the angle or the *Histogram Similarity* is computed to all the key-points in the training set. Subsequently, for each testing key-point, we find the closest training key-point and note its class. We also take the distance to the nearest neighbor and invert it in order to create a similarity measure. We keep a total of all the similarity measures for each class. In order to remove any possible skewing effects that may be created if training classes do not contain the same number of images, we divide each similarity total with the total number of training SIFT key-points for that class. For each testing key-point, we assign the class with the highest similarity total as its denomination label. A normalized sum of all the similarity measures from all the testing key-points which have been classified to a specific class is then used as the distance measure of the testing image to that class. The *Histogram Similarity* is a similarity measure between two feature vectors computed as follows: (a) The two vectors are compared element-by-element and from each pair the minimum is kept; (b) The sum of these minimums gives the Histogram Similarity.

4.2.2 Nearest to Second Nearest Neighbor Ratio

To improve classification accuracy of a testing key-point, we observe that we would be more confident to accept the class label of the nearest neighbor in the training key-points set if that nearest neighbor happened to be much closer than any other training key-point. Otherwise, if several of the training key-points are within a small radius from our testing key-point, then the class label of the nearest neighbor cannot be trusted, as any one of several training key-points might have happened to be the closest simply by chance. For this reason, a classification result for a specific testing key-point is only accepted if the ratio between the distance of the nearest and the second nearest neighbor is below a specific threshold. This threshold has been determined empirically to be 0.85.

4.3 Execution Environment

The image processing algorithm in practice can be either executed on the mobile phone or can be executed in a mobile cloud environment. To run image processing algorithms on a mobile phone environment may require us to make assumptions about the minimum processing power capabilities

Figure 1: Incomplete but clear images

Table 1: Accuracy and Speed of each Key-Point Classification Method

Classification Method	Accuracy
Nearest Neighbor	75%
Nearest to Second Nearest Neighbor Ratio	93.83%

of the phone. While the compute power of mobile devices are continuously increasing, we cannot also assume that all visually impaired users may have access to a phone with high compute power. An alternative environment is to use a mobile cloud service that allows a user to upload a mobile image and use the cloud service for classification. An image can be reduced less than a few KB and should take very limited time for data transfer and the compute time in the cloud should take much less than a second. Hence, the mobile cloud service can be easily performed in real time and potentially be faster than the execution time on a mobile phone.

5. EVALUATION

Training data was collected by taking clear and complete pictures of bills through the phone's camera under a uniform light source and on a white background. The banknotes used for the training set where placed on a white sheet of paper and exposed to an electric lamp which shone from above in an otherwise darkened room. The pictures were taken by the phone's camera held at the same angle and distance for all samples. The images were then post-processed to remove the surrounding white background so as to leave the clear and complete image of the bill on its own. In total, there were 91 training images used. Of them, 21 were of 1 dollar bills, 28 of 5 dollar bills, 22 of 10 dollar bills and 20 of 20 dollar bills. All 81 testing samples were captured by a totally blind user using the same phone. Naturally, these images are full of occlusions and may be blurred. The user was asked to take pictures, one bill at a time, in any way he deemed desirable. Thus, many of the testing samples contain partial and distorted images of banknotes, in addition to pictures of the users fingers and of the background furniture embedded in them (Figure 1). The user was also asked to randomly turn on and off the light and move about in the room while he was taking the pictures. More specifically, from the 82 testing images captured, 17 were of 1 dollar bills, 28 of 5 dollar bills, 17 of 10 dollar bills and 20 of 20 dollar bills.

Table 1 shows the accuracy of each classification method. While the nearest neighbor algorithm yields an accuracy of only 75%, the nearest to second nearest neighbor ratio algorithm yields an accuracy of 93.83%. These numbers represent the accuracy for a single image. In reality, when we capture a mobile video stream of several images, the overall accuracy significantly shoots up for both techniques as long as at least a few images have reasonable quality for analysis. If multiple images match to the same class, then the false negative ratio of such a system can also be significantly reduced. This shows that such a system can be made highly practical with very high accuracy of currency bill recognition.

6. CONCLUSIONS

We have presented the design, implementation and evaluation of a mobile currency recognition system that provides high detection accuracy for visually impaired users. In our limited sample set, our algorithm exhibited good classification accuracy with high confidence for images that were clear and properly taken even if the currency bill was folded, incomplete or had orientation and rotation effects.

7. REFERENCES

[1] G. Finlayson and E. Trezzi. Shades of gray and colour constancy. In *IS&T/SID Twelfth Color Imaging Conference*, pages 37–41, 2004.

[2] V. for the Blind and V. Impaired. Quick Facts | VISIONS/Services for the Blind and Visually Impaired. `http://www.visionsvcb.org/statistics.html`, December 2010.

[3] R. J. U. S. D. Judge. American Council Of The Blind, et al., Plaintiffs, v.Henry M. Paulson, Jr., Secretary of the Treasury Defendant. Civil Action No. 02-0864 (JR). `http://www.dcd.uscourts.gov/opinions/2006/2002-CV-0864~12:3:41~12-1-2006-a.pdf`, November 2006.

[4] X. Liu. A camera phone based currency reader for the visually impaired. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*, pages 305–306. ACM, 2008.

[5] X. Liu, D. Doermann, and H. Li. Mobile visual aid tools for users with visual impairments. *Mobile Multimedia Processing*, pages 21–36, 2010.

[6] D. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[7] S. Papastavrou, D. Hadjiachilleos, and G. Stylianou. Blind-folded recognition of bank notes on the mobile phone. In *ACM SIGGRAPH 2010 Posters*, page 68. ACM, 2010.

[8] R. Parlouar, F. Dramas, M. Macé, and C. Jouffrais. Assistive device for the blind based on object recognition: an application to identify currency bills. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*, pages 227–228. ACM, 2009.

[9] M. Williams and R. Anderson. Currency design in the united states and abroad: counterfeit deterrence and visual accessibility. *REVIEW-FEDERAL RESERVE BANK OF SAINT LOUIS*, 89(5):371, 2007.

[10] T. Yan, V. Kumar, and D. Ganesan. Crowdsearch: exploiting crowds for accurate real-time image search on mobile phones. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pages 77–90. ACM, 2010.