Currency recognition using a smartphone: Comparison between color SIFT and gray scale SIFT algorithms

Iyad Abu Doush *, Sahar AL-Btoush

Department of Computer Science, Yarmouk University, Irbid, Jordan

Received 9 April 2016; revised 21 May 2016; accepted 23 June 2016

Abstract  Banknote recognition means classifying the currency (coin and paper) to the correct class. In this paper, we developed a dataset for Jordanian currency. After that we applied automatic mobile recognition system using a smartphone on the dataset using scale-invariant feature transform (SIFT) algorithm. This is the first attempt, to the best of the authors knowledge, to recognize both coins and paper banknotes on a smartphone using SIFT algorithm. SIFT has been developed to be the most robust and efficient local invariant feature descriptor. Color provides significant information and important values in the object description process and matching tasks. Many objects cannot be classified correctly without their color features. We compared between two approaches colored local invariant feature descriptor (color SIFT approach) and gray image local invariant feature descriptor (gray SIFT approach). The evaluation results show that the color SIFT approach outperforms the gray SIFT approach in terms of processing time and accuracy.

1. Introduction

Physical banknotes (bills and coins) are used in most countries. Automatic currency recognition for both coins and bills is critical for many applications: for example, assist blind persons, banknote counting machines, or classify the money deposited in ATM machines (Toytman and Thambidurai, 2011).

Coins and bills have different properties concerning recognition. For coin detection various methods have been proposed: statistical approach, vector quantization, Eigen space decomposition, and image registration. All of these methods are sensitive to illumination conditions and many of them depend on taking images at fixed environment setting such as: image background and camera location which make the recognition task difficult for some applications as helping visually impaired people. On the other hand, bills exhibit much more tolerance to illumination due to the lack of specular reflection; also they have many details that need to be identified. However, bill recognition has other sort of problems such as...
as shape distortion due to wrinkling and/or folding (Toytman and Thambidurai, 2011).

The banknote recognition methods include: hidden Markov chains, artificial neural networks, and dynamic template matching that mimics the behavior of the human brain. Most of these methods are tested on personal computers. Scale-invariant feature transform (SIFT) algorithm can produce distinct key-points and feature descriptors for each object. It is considered one of the most robust feature extraction algorithms (Wang et al., 2013). Although many alternative algorithms to SIFT have been designed, such as SURF (Bay et al., 2006), it is still used by several researchers (Lin and Zhao, 2008).

SIFT algorithm is a feature selection technique which depends on the appearance of the object at specific interest points. These interest points are not changed by image scale or rotation. The technique is robust to illumination change, to noise on image, and to minor changes in the image viewpoint (Bastanlar et al., 2010; Geng and Jiang, 2009).

The computational power and camera availability of current smartphones make them a good candidate for currency recognition. However, a small number of literature have tackled such problem (Toytman and Thambidurai, 2011; Wang et al., 2013).

In this paper we introduce a mobile system for currency recognition that can recognize the Jordanian currency in different perspective views and scales. For example, images may not be ideally oriented and may have scale changes due to the variation of distance from the camera and illumination changes. Banknote recognition in a variable environment is a complex problem because we have many uncontrolled conditions that affect the image quality. We develop a smartphone application to identify currencies that are partially visible, folded, wrinkled or even worn by usage.

Scale-invariant feature transform (SIFT) is designed for gray images mainly. However, color provides significant information and important values in the object description and matching. Many objects cannot be classified correctly without their color features. We evaluate colored and gray local invariant feature descriptor in terms of processing time and accuracy ratio.

2. Literature review

A large number of researchers have made several contributions toward developing techniques for currency recognition. The different properties between coins and bills make researchers deal with the recognition task differently for each one of them. In this section, we review previous work in currency recognition techniques.

2.1. Coin currency recognition approaches

Mitsukura et al. (2000) proposed a method to design a neural network (NN) using a genetic algorithm (GA) and simulated annealing (SA). The similar characteristics of the coin images (i.e., size, weight, color and pattern) cause a trouble for currency recognition. The proposed scheme discovered several features and the recognition rate was approximately 98%. This technique needs long time and high computational power. So, it is not suitable for smartphone.

Modi and Bawa (2011) developed an Artificial Neural Network (ANN) Automated Indian Coin Recognition System with rotation invariance. Hough transformation and pattern averaging techniques are used to extract the image features. Then, the extracted features are passed as input to train neural network. The trained neural network is tested and validated using 5040 images of all the Indian coin values that are rotated at different angles. Experiments show that the system had 97.74% correct recognition rate.

Velu and Vivkanandand (2009) presented a simple Indian coin recognition method with rotation invariance. The HT algorithm combines the features of straight line detection HT algorithm, curve detection and circle detection HT algorithm. The results show that the method obtains 100% correct recognition.

Mahajan and Gaikwad (2014) presented an Indian coin recognition system. Segmentation technique is applied by traversing row wise and column wise through the image. The image is selected for comparison based on the radius matching. The paper mentioned that the proposed method achieved good accuracy of coin recognition with no more details about the accuracy rate.

These techniques assume a controlled environment for the coin image with no high illumination or portion of the image taken. Such techniques are not suitable to be applied on multiple variable conditions environment (e.g., using a smartphone) which could affect the quality of the obtained image.

Khashman et al. (2006) proposed a coin identification system based on coin surface patterns and a neural network for identification of rotated coins at intervals of 15°. This system uses image preprocessing as its first phase. A back propagation neural network is then used to understand the coin patterns. The results show 95.83% correct coin identification.

Meenakumari (Meenakumari, 2013) designed a coin recognition system by applying adaptive hinging hyperplanes (AHH) algorithm with rotation invariance to classify the Indian coins. The system takes the parameters of Indian coins such as size, shape, weight, surface and so on. The system provides good result of coin identification in different rotation degrees with no information about the achieved accuracy.

These techniques take one invariant of the taken image which is rotation. Other invariants when the image is taken were not considered. However, such multiple invariants are necessary to be considered to achieve accurate recognition in an uncontrolled environment.

Reisert et al. (2006) proposed coin recognition system using the direction of the gradient vectors and Fast Fourier Transform. The classification is done using the nearest neighbor search. The results show that the directional information is good enough to design a reliable classification system that achieves 97% accuracy.

2.2. Paper Currency Recognition approaches

Mirza and Nanda (Mirza and Nanda, 2012) use three extracted features from the banknote including identification mark, security thread and watermark. The features are extracted using edge based segmentation by Sobel operator.

An algorithm based on Local binary patterns (LBP) for recognition of Indian paper currency was proposed by Sharma et al. (2012). The results show that the system provides
a good performance for images with low noise with 99% accuracy.

Sargano et al. (2013) proposed a new intelligent system for Pakistani paper currency recognition. The proposed system required less time compared with other systems. Three layer feed-forward Back propagation Neural Network (BPN) is used for classification. The system is tested with 350 Pakistani banknotes. The results indicated that the system had 100% recognition accuracy. This technique is applied on banknotes without any distortion (e.g., wrinkled or folded).

Da Costa (da Costa, 2014) developed a banknote recognition system to recognize multiple banknotes in different perspective views and scales. Feature detection, description and matching are used to enhance the confidence in the recognition results. The banknote contour is computed using a homography. The system is evaluated with 82 test images, and all the Euro banknotes were successfully recognized. The system provides robust results to handle folded and wrinkled banknotes with several kinds of illumination. The algorithm needs several steps with high computation overhead which make it not suitable to be used by a smartphone.

Debnath et al. (2010) presented a paper currency recognition system using ensemble neural network (ENN). The individual neural networks (NNs) in an ENN are qualified via negative correlation learning (NCL). The used currencies are from three types new, old and noisy. The banknote image is converted into gray scale and then compressed. Then each pixel of the compressed image is passed as an input to the network. This method can recognize and match noisy currency images and it provides good results when compared to single network and ensemble network with independent training. The results show the recognition accuracy range from 100% to 54% depending on the noise level of the input image.

Vijay and Jain (2013) proposed an image processing technique to extract paper currency denomination. The extracted Region of Interest (ROI) is used with pattern recognition and neural networks matching technique. In this method they captured the images by the simple flat scanner with a fixed size and then some filters are applied to extract the denomination value of the banknote. The paper has no information about the accuracy of the proposed algorithm.

Reel et al. (2011) use heuristic analysis of characters and digits for Indian currency notes for recognition. This process is invariant to light conditions, use font type and deformations of characters caused by a skew of the image. Heuristic analysis of the characters is performed for this purpose to get the exact features of characters before feature extraction. One of the challenges raised in the character segmentation part is that two characters are sometimes joined together.

These techniques focus on extracting the number on the paper currency. However, such technique is not feasible in the case of wrinkled or folded banknote.

Paisios et al. (2012) developed a mobile currency recognition system using SIFT to recognize partial images. The system is evaluated using a limited sample set with different states as: folded, incomplete or had orientation and rotation. The results indicated that the nearest neighbor algorithm provides an accuracy of 75% and the nearest to second nearest neighbor ratio algorithm provides an accuracy of 93.83%.

Toytman and Thambidurai (Toytman and Thambidurai, 2011) proposed a banknote recognition system on Android with improved Speed up Robust Features SIFT. The algorithm solves the problems related to illumination conditions, scale and rotation. The approach was insensitive to clutter, occlusion tolerable and wrinkling of banknotes. The approach was tested and good result was obtained with respect to clutter and variations in illumination. On the other hand, the problem of inability to detect folded banknotes is not solved. The paper does not provide any information about the algorithm accuracy.

Rashid et al. (2013) propose models of classifiers (i.e., support vector machine (SVM), artificial neural network (ANN), and hidden markov model (HMM)) to be used on low cost embedded system. They used SIFT, bag-of-word, and SVM classifier on real time recognition process. Contact image sensors are used to get the two sides of the banknote image simultaneously. The experimental results show that SVM classifier outperforms ANN and HMM.

These techniques attempt to recognize paper banknotes only on a smartphone. None of these techniques compare the original SIFT with color SIFT.

2.3. Color SIFT descriptor and gray-scale SIFT descriptor

Van De Sande et al. (2010) presented a comparison between the local color descriptors with gray-value descriptors. They use the evaluation framework of Mikolajczyk and Schmid to the level of local gray value invariants. The results show that the method which combines color information and SIFT gives better results.

Abdel-Hakim and Farag (2006) introduced colored local invariant feature descriptor (CSIFT) as the result of combining both color and geometrical information in the object description. Evaluation results show high performance and good result of CSIFT when compared with the gray space SIFT descriptors.

Scale Invariant Feature Transform (SIFT) is used mainly for grayscale images. Many local features cannot be classified correctly without color information. Cui et al. (2010) used a new color space, called perception-based color space instead of traditional SIFT the proposed approach used the SIFT color descriptors and the result showed that the colored descriptor was more robust than the standard SIFT.

Rassem and Khoo (2011) designed different color SIFT descriptors in order to evaluate the object class recognition system performance and all possible combinations of these descriptors were implemented. The results show that some combinations of color SIFT descriptors obtained remarkable classification accuracy. Nonlinear $\chi^2$-kernel support vector machine is used as a learning classifier.

These techniques have not been evaluated or tested on a smartphone. The performance of such algorithms on a smartphone can raise the need to speed up the process by minimizing the number of steps or introducing enhancement on the algorithm feature detection process.

3. Methodology

The proposed banknote recognition system is based on the scale-invariant feature transform (SIFT) algorithm. Using SIFT algorithm requires comparing and classifying of a larger number of key-points. For this reason the operation on mobile phone is relatively slow (Paisios et al., 2012).
SIFT is designed to be applied on gray images. However, color provides significant information for object description. Many objects cannot be classified correctly without their color features (Abdel-Hakim and Farag, 2006). We compare between two approaches: a colored local invariant feature descriptor and gray image local invariant feature descriptor.

3.1. System design

The proposed system uses a set of sample images of Jordanian banknotes which are used as a training set for the classification algorithms. The training data are used to help the algorithm in recognizing Jordanian currency correctly. Our system is implemented in the mobile phone and it has the following phases shown in the Fig. 1:

1. Banknote images are captured by the mobile phone.
2. The captured image is pre-processed as follows: cropped to separate it from its background to get the regions of interest using discrete wavelet transform. The image is then compressed (size less than 20 KB) using the nearest neighborhood interpolation as SIFT computation is too expensive and such compression does not significantly affect accuracy. The images are converted to grayscale if the technique is color SIFT approach.
3. Key-point detection: identifies points in the image at which they are maximum or minimum pixels from its neighbor (Van De Sande et al., 2010).
4. Key-point description: each region around a detected key point location is converted into a more compact and stable descriptor that can be matched against other descriptor.

The dataset was collected by taking different images of Jordanian bills and coins (50 JD, 20 JD, 10 JD, 5 JD, 1 JD, 50 piaster, 25 piaster, 5 piaster, and 1 piaster). All the images were taken using Galaxy Grand 2 camera (8 MP 3264 x 2448 pixels) smartphone on different backgrounds. The pictures were captured by the phone’s camera held at different states in terms of angle, distance, scale, illumination, different image distortion (folded and wrinkled) and the images were taken from both sides. Fig. 2 demonstrates several sample images from each condition. In total, there were 100 training images used and 400 testing samples were captured by the same phone 20 image for each class of currency. The collected dataset is available in the link https://www.dropbox.com/sh/b7bmmkpumx4sutt/AAD6b-Yd3-YrmVOimDKBryna?dl=0.

3.2. A SIFT descriptor with color characteristics

3.2.1. Image preprocessing

To improve the identification of descriptive features and ensure that the system has effective recognition a preprocessing step is applied. Images from both the training and testing samples are compressed (size less than 20 KB) in order to reduce the processing time and increase the RAM efficiency. Then the image is cropped automatically from its background in order to detect the image boundary.

3.2.2. Image main color extraction

The extraction of banknote colors is not a trivial task since the images are sensitive to the surrounding environment lighting. The color that represents each banknote is identified using a set of 5 images with lighting variations. Then, to extract the three main colors for each image the image is converted to bitmap format. Bitmaps are defined as a regular rectangular cells called pixels, each pixel has a color value. They are represented by only two parameters the number of pixels and the color depth per pixel (Lu and Chang, 2007; Sawalha and Abu Doush, 2012).

3.2.3. Using SIFT descriptors

The key-points extracted by SIFT are not affected by rotation and scaling. This algorithm consists of two primary stages: key-point detection and creating descriptors.

The following are the steps of SIFT algorithm:

1. Constructing a scale space: the scale space can be created by taking the original image and produce blurred images. Next, the original image is resized to the half size...
and produces a number of blurred out images again that form an octave (vertical images of the same size) (Vidhi and Khushbu, 2014).

2) Laplacian of Gaussian Approximation: the Laplacian of Gaussian technique calculates the difference between two consecutive scales for the image. The difference of Gaussians (DOG) is a feature improvement algorithm that is computed from the subtraction of one blurred version of the original image from another version of the same image which is less blurred (Vidhi and Khushbu, 2014).

3) Finding Key-points: the major key-points for matching are recognized in the image. These key-points are usually chosen by analyzing the edges, corners, blobs or even ridges. The first step of finding the image key-points is to find the maximum and minimum pixels from all its neighbors (Vidhi and Khushbu, 2014).

4) Eliminate edges and low contrast regions: the edges need to be eliminated. For this reason, a concept of a Harris corner detector is used. A $2 \times 2$ Hessian matrix (H) is used to compute the principal curvature. If this ratio is greater than a threshold then that key-point is removed (Chaudhry and S., 2015).

5) Orientation Assignment: the orientation is assigned to each key-point to get invariance for image rotation. The maximum peak in the histogram is used to calculate the orientation and any peak higher than 80% of the histogram is considered (Vidhi and Khushbu, 2014).

**Figure 2** Several samples of recognition system dataset.

**Figure 3** Orientation assignment.
magnitude and orientation are calculated for all pixels around the key-point. An example of the orientation assignment on a 1 JD paper currency is shown in Fig. 3.

(6) Key-point Descriptor: after a key-point is detected a descriptor must be created by taking a $16 \times 16$ neighborhood around the taken key-point (Vidhi and Khushbu, 2014; Pan and Lyu, 2010).

### 3.3. Key-point matching

In order to represent the best reference image descriptors of each banknote class many techniques can be applied. The simplest technique is to find the total key-points of five training images for each class to represent the major class features that will be stored in the internal database as best reference image descriptors of each class. Then the Euclidean distance is computed between the test image key-points and reference image descriptors in the database to determine whether the two images belong to the same keypoint. Using this technique the results show that there were only 20 common features between the images.

Having more features extracted from the dataset will improve the classification accuracy. In order to improve the detection accuracy we increase the number of extracted features by improving the technique. The used technique takes the intersection of the first training image key-points with the second training image key-points for each class. Then take the intersection of the third training image key-points with fourth one and so on until five training images. Next, the summation of total intersection represents the major features that will be stored in the internal database as the best reference image descriptors of each banknote class for a matching. Then the Euclidean distance is computed between the test image key-points and reference image descriptors in the database. Using this technique the results show that there were 50 common features.

Each image key-point descriptor is associated with the 5 best reference image descriptor. This allows to decide if it is the correct matching or not by computing the Euclidean distance between reference key-points. In that case, minimum Euclidean distance for invariant descriptor vector will be taken (Pan and Lyu, 2010).

The authors find experimentally the best Euclidean distance threshold between these reference key-points which is 200. All values below 200 are discarded. Table 1 shows the Euclidian distance and the matching results from corresponding two images of the same banknotes with some difference in scale and illumination.

The steps of gray SIFT Descriptor technique are the same of SIFT with color, except in the image preprocessing steps in which there is no need to convert the image into a gray scale image.

### 3.4. Implementation

Java programming language is used to build the system with Android and OpenCV library. OpenCV is a computer vision and machine learning software library. It provides a set of important SIFT functionalities (Mitsukura et al., 2000). Object detection is done using the built-in object detector in the OpenCV. OpenCV also provides functions which draw the small circles on the locations of key-points and it will even show its orientation.

The recognition system classifies the currency by searching the internal database which contains the best reference image descriptors of each currency class. The recognition program uses C functions from the OpenCV library to recognize the different currency. In order to use C functions the Java Native Interface2 (JNI) is used in the application (Ledwich and Williams, 2004).

### 4. Experiments and evaluation

The performance of the proposed system has been evaluated using the developed dataset of 400 images. In the experiments, the authors tested a database of 400 Jordanian banknotes, which includes 10 kinds of Jordanian banknotes (50 JD, 20 JD, 10 JD, 5 JD, 1 JD, 50 piaster, 25 piaster, 10 piaster, 5 piaster, 1 piaster).

The processing time is computed for every recognition process. Fig. 4 shows a sample of the tested banknote and the detection of their key-points. Fig. 5 shows samples of testing banknote and the detection of their key-points using gray SIFT Approach.

#### 4.1. Comparisons

We compute the average correct recognition and the average processing time for coin and paper banknotes for color SIFT approach. Table 2 shows these results.

We compute the average correct recognition and the average processing time for coin and paper banknotes for gray SIFT approach. Table 3 shows these results.

The experimental results show that the accuracy recognition for coin currencies is less than recognition accuracy for paper currencies. This is because of the illumination that makes the key-points for the same coin class differs.

A comparison between the performances of Color SIFT with the performance of the gray SIFT is performed and the time required for the recognition process is calculated. The input of SIFT and color SIFT are the same set of images. It is clear from the results that the number of detected features in the images for color SIFT is larger than those in the gray SIFT. Color SIFT has a large number of repeated features, which leads to a more accurate estimation of the banknote values (Abdel-Hakim and Farag, 2006). This leads to the increase of detecting features and enhances the performance of the

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Euclidian distance and matching result of two different images.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold</td>
<td>Image1 features</td>
</tr>
<tr>
<td>----------</td>
<td>----------------</td>
</tr>
<tr>
<td>10</td>
<td>365</td>
</tr>
<tr>
<td>50</td>
<td>365</td>
</tr>
<tr>
<td>100</td>
<td>365</td>
</tr>
<tr>
<td>150</td>
<td>365</td>
</tr>
<tr>
<td>200</td>
<td>365</td>
</tr>
</tbody>
</table>
Figure 4  Sample of the tested banknote and the detection of their key-points (color).

Figure 5  Sample of the tested banknote and the detection of their key-points (gray scale).

Table 2  Correct recognition average and processing time average for color SIFT approach.

<table>
<thead>
<tr>
<th>Banknote type</th>
<th>Average correct recognition</th>
<th>Average processing time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper Currency</td>
<td>0.71</td>
<td>72.9</td>
</tr>
<tr>
<td>Coin Currency</td>
<td>0.25</td>
<td>78.2</td>
</tr>
</tbody>
</table>

Table 3  Correct recognition average and processing time average for gray SIFT approach.

<table>
<thead>
<tr>
<th>Banknote type</th>
<th>Average correct recognition</th>
<th>Average processing time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper Currency</td>
<td>0.53</td>
<td>72.4</td>
</tr>
<tr>
<td>Coin Currency</td>
<td>0.20</td>
<td>80</td>
</tr>
</tbody>
</table>

Please cite this article in press as: Abu Doush, I., AL-Btoush, S. Currency recognition using a smartphone: Comparison between color SIFT and gray scale SIFT algorithms. Journal of King Saud University – Computer and Information Sciences (2016), http://dx.doi.org/10.1016/j.jksuci.2016.06.003
recognition process. Also, the time required for the recognition process with gray SIFT was longer than the time required for color SIFT. Therefore, the color SIFT performs better with respect to the recognition process and processing time.

4.2. Discussion

The proposed approach provides good result in the case of paper banknotes except some cases as:

- Too wrinkled: it is difficult to match the exact descriptors (some key-points do not appear or changed).
- Folded several times: it is difficult to match the exact descriptors (some key-points do not appear or changed).
- Take image from close distance: crop large portion of the image which means it is difficult to match the exact descriptors (a large number of key-points do not appear).
- Take image from too far distance: crop large portion of the background which means it is difficult to match the exact descriptors (large numbers of redundant key-points appear).

The proposed approach provides good results in the case of coin banknotes except the following cases:

- Images with high illumination: it is difficult to match the exact descriptors (some key-points do not appear or changed).
- Take image from close distance: crop large portion of the image which means it is difficult to match the exact descriptors (a large number of key-points do not appear).
- Take image from too far distance: crop large portion of the background which means it is difficult to match the exact descriptors (large numbers of redundant key-points appear).

5. Conclusion

Currency recognition in an uncontrolled environment (i.e., using a smartphone) is not an easy task because of the multiple variable conditions that could affect the quality of the image. The experimental results have shown the effectiveness of SIFT algorithm in general for Jordanian banknote recognition, although our algorithm is tested in a more challenging dataset with the images taken in different conditions. The system depends on the appearance of the object at specific interest points. The results show that the recognition accuracy for coin currencies is less than recognition accuracy for paper currencies. This happened because of the illumination condition that affects the coins image.

We introduced color SIFT approach for the purpose of combining both color and local object features descriptor contrasting to the common existing methods (gray SIFT). It is clear from the results that the number of detected features in the color SIFT is larger than that in the gray SIFT. The evaluation results show high performance of color SIFT when compared with gray SIFT descriptors in terms of processing time and accuracy rate.

The authors’ future work will focus on the enhancement of the detection of folded or too wrinkled paper banknotes and coins with too illumination. The processing speed will be decreased to less than 1 min for each testing image to be more suitable for smartphone users.

References


