Currency Classification System for visually impaired

Sujit J. Singh
University at Buffalo
University at Buffalo, Buffalo, N.Y, U.S.A
sujitjit@buffalo.edu

Gautam A. Shende
University at Buffalo
University at Buffalo, Buffalo, N.Y, U.S.A
gautamav@buffalo.edu

Abstract

The use of currency is an absolute matter for exchange in every – day life of an individual. This exchange become more difficult to people with visual impairments. The proposed system makes use of image classification technique to solve difficulties of the partially sighted. The proposed system is developed using Bag of words (BOW) approach to classify currencies of two countries, USA and India.

1. Introduction

Paper currency is a vital mode of exchange for performing every-day commercial and non-commercial activities. The main difficulty faced by specially challenged people needs to be addressed using the modern computer science approach. Computer vision and image processing in its extensive applicability addresses some of these problems. One such problem related to Paper currency recognition is addressed in this proposed system. The proposed system uses Classification technique combined with image processing to build a system that can recognize authenticity, type, and nature of currency (In which country it lies). For this, we are using Bag of Words approach as explained below.

Bag of words, also known as vector – space model is a way of representing textual data in the form of dictionary. This dictionary can be obtained from the methods like clustering. The idea can even be extended from text to image and hence it is a significant development in the field of computer vision. It is generally used in the domains of information retrieval, document classification and natural language processing. Along with Bag of words, different smoothening filters are used that helps in feature detection and extraction. These filters are formed at different scale forming a filter bank. The concept of bag of words is thoroughly discussed in the upcoming sections.

2. Literature Review

Computer vision and image processing plays a vital role in solving some of the problems of specially challenged people, especially the one with the visual impairment. This has led to many studies in this field. Some of the examples include using image processing techniques to further apply it in various systems like scanners and camera to increase its applicability as represented in papers discussed in the literature.

In a research paper[1] (by Naura Semary, 2015), the simple image processing techniques like histogram equalization, noise removal, image segmentation, etc. techniques is used. The region of interest (ROI) extraction gives the features that are matched to recognize currency. Our system use a set of 20 filters, namely, Gaussian, Laplacian of Gaussian (LOG), dx-scale and dy-scale filter etc. to extract features (visual words) as the filter responses that are used for training. The guessing of test images are done against the labels formed from these visual words.

In another research paper[2] (by Kuldeep Verma, Bhupesh Kumar Singh, and Anupam Agarwal, 2011), the recognition of Indian currency is done based on text analysis. The main drawback of this system is that the feature extraction just depends upon the intrinsic feature. This has been addressed in the proposed system as applying different filters to various scales not only extract texture but also gradient and other properties of an image. For e.g. : LOG is also used for extracting gradient properties of an image.

3. Methodology

The classification algorithm used (BOW) is applied on a self - developed dataset. The training dataset consisted of a set of 397 images and the test set consisted of 35 images. It goes through the following stages as shown in the diagram below:

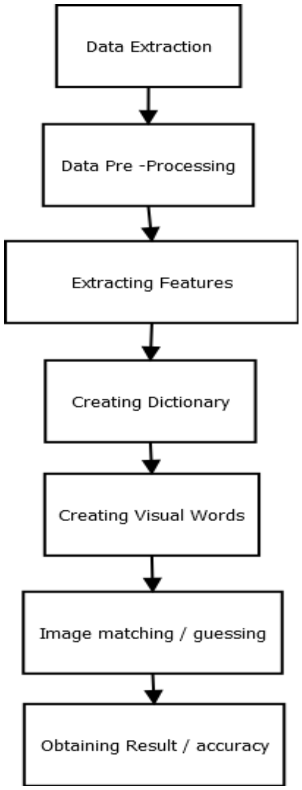

Fig 3.1 Phases of the proposed system

4.1 Dataset Extraction

The data was extracted manually for the 12 different currency notes. For training set, we downloaded the standard front and back images from online sources and then performed various transforms on these images to create a training set of 12 images.
For testing, to make it more realistic, we took real time pictures from our mobile phones and used them as input for testing.

4.2 Dataset Pre-processing

Python preprocessing -> using local file path as input and 1 to 12 as output labels mapped with IN50 and so on. The training dataset was synthetically generated using image augmentation techniques of translation, rotation, scaling, cropping, resizing and affine transformation. The implementation of the proposed system is done in python – OpenCV and the processing of the algorithm and result formulation is done in Matlab.

4.3 Extracting features from images

For feature extraction, we have used a combination of 4 different filters.

4.3.1. Gaussian Filter

The default Gaussian filter using the fspecial function has the following 3 x 3 representation in Matlab:

0.0113	0.0838	0.0113
0.0838	0.6193	0.0838
0.0113	0.0838	0.0113

The filter is high at the centre and it falls off when we move towards the edges and further drops down as we move to the corners. This creates a smoothening effect and smoothens any sharp, thereby also trying to remove or reduce noise.

4.3.2. Laplace of Gaussian filter:

The default Laplace of Gaussian filter using the fspecial function has the following 5 x 5 representation in Matlab:

0.0448	0.0468	0.0564	0.0468	0.0448
0.0468	0.3167	0.7146	0.3167	0.0468
0.0564	0.7146	-4.9048	0.7146	0.0564
0.0468	0.3167	0.7146	0.3167	0.0468
0.0448	0.0468	0.0564	0.0468	0.0448

The filter is highly negative only at the center and then has decreasing positive values towards the edges that further decrease down the corners. The filter thus has kind of a reverse effect when compared to the Gaussian filter. On convolution, the borders have negative values and the pixels next to the borders have higher values. This helps in edge and line detection in the image.

4.3.3. dxscale filter

The default dxscale filter is obtained by just filtering the original Gaussian filter with the matrix [-1 0 1]. It has the following 3 x 3 representation in Matlab:

0.0838	0	-0.0838
0.6193	0	-0.6193
0.0838	0	-0.0838

The dxscale filter tries to highlight the edges and lines along the x axis i.e. along the vertical direction in image processing.

4.3.4. dyscale filter

The default dyscale filter is obtained by just filtering the original Gaussian filter with the transpose matrix [-1 0 1]. It has the following 3 x 3 representation in Matlab:

0.0838 0.6193 0.0838
0 0 0
-0.0838 -0.6193 -0.0838

The dyscale filter tries to highlight the edges and lines along the y axis i.e along the horizontal direction in image processing.

4.4. Creating Dictionary

The dictionary is created from the various responses that we get after applying different filters to the images in the training set.

4.5. Creating Visual words

In bag of words, when we have to classify an image, we apply filters on it and then represent it using a wordmap which is simply representing the image by using only the words that are present in the created dictionary. This is known as a visual word.

4.6. Image matching / Image guessing

For image matching, we are using histogram matching along with knn. Essentially, while classifying the incoming images, instead of finding just the closest instance in the training data set, we found out k closest instances to the input test image and then used the k nearest neighbors' algorithm to predict the class. We have used unweighted voting where in each of the neighbors votes is given equal weightage and thus the label with maximum number of votes is selected as the output label. This can be simply achieved by using the mode function.

4.7. Obtaining result / accuracy – The result of the Image matching stage is realized into the confusion matrix and accuracy is calculated as follows:

percent = trace(conf) / sum (conf(:));

5. Experiments / Results:

5.1. Accuracy at different kNN values:

We tried applying the k nearest neighbor (kNN) for histogram comparison and image guessing. Essentially, while classifying the incoming images, instead of finding just the closest instance in the training data set, it was found that k closest instances to the input test image and then used the k nearest neighbors' algorithm to predict the class. We have used unweighted voting where in each of the neighbors votes is given equal weightage and thus the label with maximum number of votes is selected as the output label. This can be simply achieved by using the mode function

The algorithm was tested with different values of kNN (K nearest neighbors) and the experimental results are displayed as follows:

S.N.	kNN value	Accuracy (in %)
1.	2	35.6
2.	5	42.8571
3.	6	45.7143
4	7	45.7143
5.	8	40
6.	10	37.1429
7.	15	31.4286
8.	20	28.5714

Table 4.1 Accuracy at different kNN values

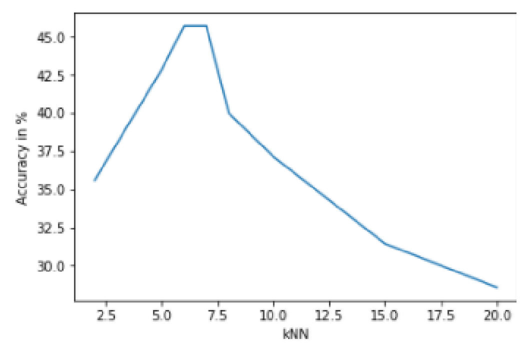

Fig 4.1 Graph showing accuracies

5.2. Analyzing accuracy:

The confusion matrix of the model having high accuracy is given below:

	1	2	3	4	5	6	7	8	9	10	11	12
1	3	0	0	0	0	0	0	0	0	0	0	0
2	0	2	0	0	0	0	0	1	0	0	0	0
3	0	0	3	0	0	0	0	0	0	0	0	0
4	0	2	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	2	0	0	1	0	0	0	0
6	0	0	0	0	0	1	1	0	0	0	0	0
7	1	0	0	0	0	0	3	0	0	0	0	0
8	0	0	0	0	1	0	0	1	0	0	1	0
9	0	0	0	0	1	0	0	0	1	0	1	0
10	0	0	0	0	0	1	0	0	0	2	0	1
11	0	1	0	0	0	0	0	0	0	0	2	0
12	0	0	0	0	0	1	1	0	0	0	0	0

Table 4.2 Confusion matrix at kNN = 7

From the above confusion matrix we come to know that the label 12 i.e. USD 100 is incorrectly classified as label 6 and label 7 USD1 and USD10. This may be due to the fact that USD1, USD 10, and USD 100 has very minute feature difference. Thus the confusion matrix appears little sparse due to the lack of training and testing data. More data can give better results if the proposed system is practiced in future.

6. Application

The proposed system provides the base structure for wide range of real – life applications. One, it can be used in commercial areas like malls, banks, etc. to give assistance to visually impaired. In addition to that, it can be used by the first time visitors to identify the authenticity (fake or real) of the foreign currency.

7. Conclusion / Future Work

Although the Bag of Words (BOW) is one of the most popular and flexible approach in Computer Vision, the approach is data – dependent and incurs low accuracy when compared to modern deep learning algorithms. Thus, the future scope of the proposed system can be extended by classifying currency through more comprehensive algorithms like Convolutional Neural Network (CNN). Also, the proposed system is scale and rotation invariant, hence, more efficient algorithms like SIFT can be used for feature extraction. This project can also be extended to currency belonging to different countries.

References

[1] N. A. Semary, S. M. Fadl, M. S. Essa and A. F. Gad, "Currency recognition system for visually impaired: Egyptian banknote as a study case," 2015 5th International Conference on Information & Communication Technology and Accessibility (ICTA), Marrakech, 2015, pp. 1-6. doi: 10.1109/ICTA.2015.7426896

[2] K. Verma, B. K. Singh and A. Agarwal, "Indian currency recognition based on texture analysis," 2011 Nirma University International Conference on Engineering, Ahmedabad, Gujarat, 2011, pp. 1-5. doi: 10.1109/NUiConE.2011.6153231

[3] URL:https://docs.opencv.org/3.0-beta/doc/py_tutorials/py_gui/py_image_display/py_image_display.html

[4] URL:http://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_imgproc/py_geometric_transformations/py_geometric_transformations.html