Currency Recognition on Mobile Phones

Suriya Singh[1], Shushman Choudhury[2], Kumar Vishal[1] and C.V. Jawahar[1]
[1] IIIT Hyderabad [2] IIT Kharagpur

Goal

Recognizing currency bills in cluttered scenes and challenging situations on a low-end mobile phone. This mobile application is intended for robust, practical and easy use by the visually impaired.

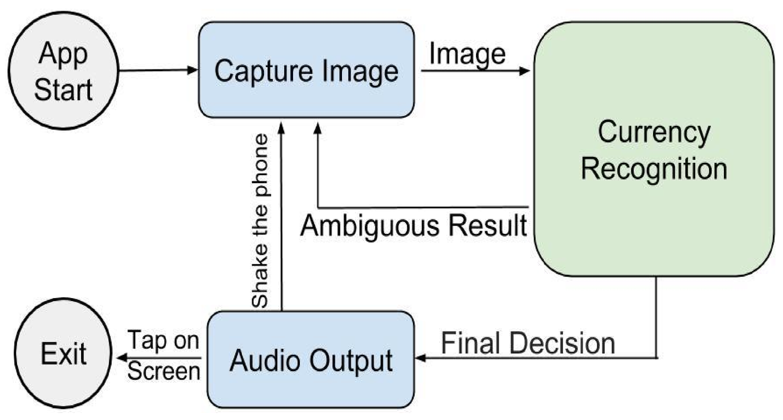
Motivation and Challenges

- We formulate the recognition problem as a task of fine-grained instance retrieval that can run on mobile devices.
- The real-world usage by the visually impaired introduces challenging queries in terms of the image quality, the portion of the bill visible, illumination and clutter.
- Strong restrictions in the memory, application size, and processing time.
- A thin index structure is used to make the application efficient and compact.

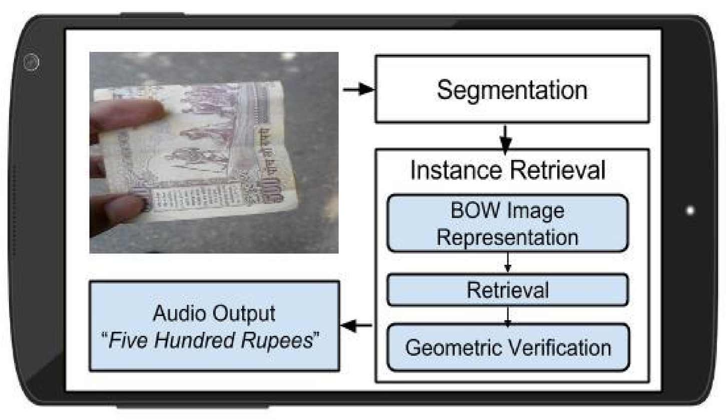
Method Overview

- **High-level control flow diagram**

1. The app once started does not need any input from the user.
2. It takes a picture when the phone is held stable for some time.
3. The app processes the image and gives audio feedback.



- **A conceptual schematic of the back-end**



A. Segmentation

- Image Segmentation reduces processing time and improves accuracy.
- It not only cuts down the data to process but also the likelihood of irrelevant features by eliminating much of the background.
- We use GrabCut, which involves energy minimization based on iterative graph cuts.
- The cost function for this is:

$$E(x,y) = -\sum_i \log p(y_i|x_i) + \sum_{(i,j)\in\mathcal{E}} S(y_i, y_j|x)$$

where x_i is the colour of the i^{th} pixel and y_i is +1 if the pixel belongs to the object, otherwise -1. $S(y_i, y_j|x)$ favours neighbor pixels with similar color to have the same label.

- **Segmentation results**

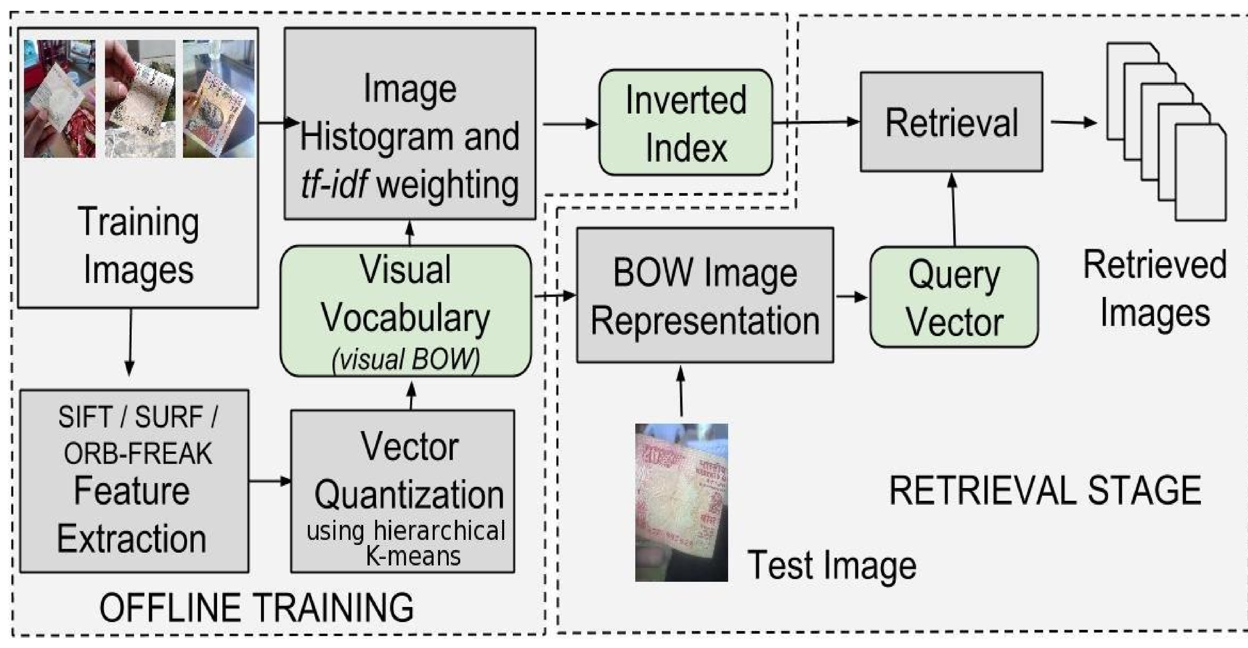
Method Overview

B. Instance Retrieval

Classification of bills in the image uses an instance retrieval pipeline:

1. Building a visual vocabulary - The set of clusters of features obtained from standard descriptors, forms the visual vocabulary of images.

2. Image Indexing using Text Retrieval Methods - Each image is represented by a histogram of visual words followed by *tf-idf* weighting.



3. Retrieval Stage - Each test image histogram is compared to the training set via cosine similarity. The ten most similar images are retained.

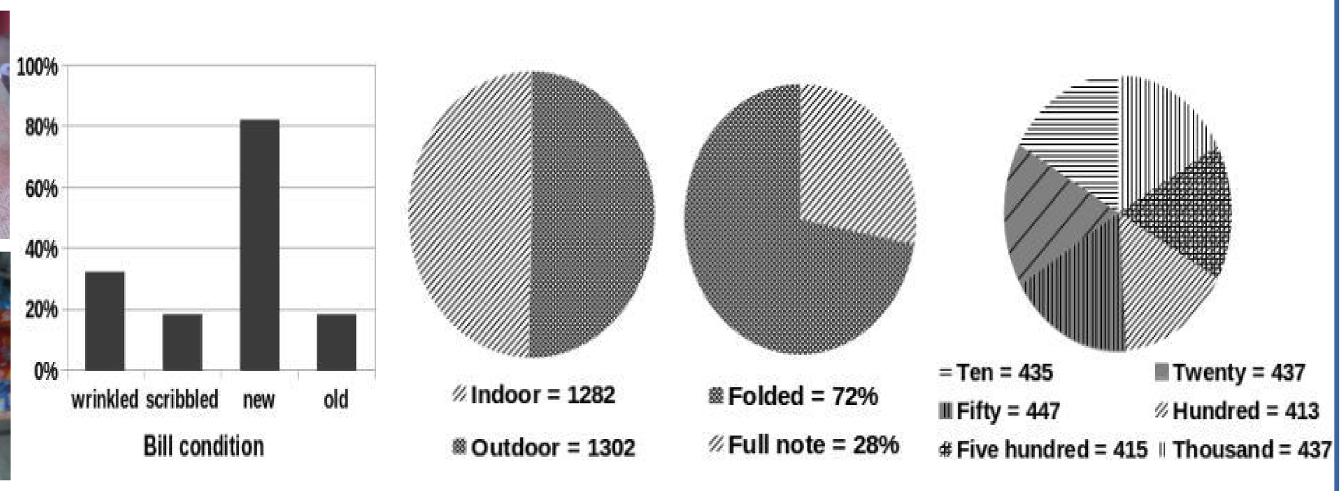
4. Spatial re-ranking - To ensure spatial consistency of keypoints, we use geometric verification (GV) by fitting the fundamental matrix.

5. Classification - Each retrieved image votes for its image class by the number of spatially consistent keypoints. The class with the highest vote is returned as the result.

Experiments and Results

***Dataset* # –** 2584 images captures the possible use cases of a visually impaired user.

Images from the dataset with bills in varying illumination and background.



Various statistics that reflects the dataset's comprehensiveness.

Results

1. Results of mobile adaptation (a) Storage and memory requirements (b) Time analysis

(a)

Component	Size
RAM use (on average)	23.5MB
Inverted index	20.5MB
Vocabulary (10K)	5.3MB
Keypoints location	11MB
Annotations	6.9KB

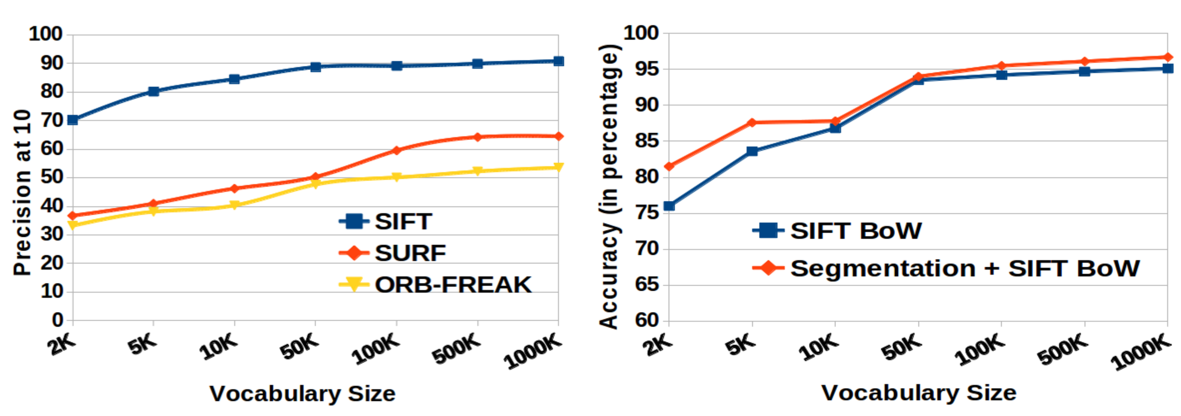
(b)

Module	SIFT BoW+GV Time in seconds without segmentation	SIFT BoW+GV Time in seconds with segmentation
Segmentation	-	0.27 s
SIFT keypoints detection	0.25 s	0.25 s
SIFT descriptor extraction	0.27 s	0.13 s
Assigning to vocabulary	0.01 s	0.01 s
Inverted index search	0.12 s	0.12 s
Spatial re-ranking	0.61 s	0.31 s
Total Recognition Pipeline	**1.26 s**	**1.09 s**

2. Classification Accuracy using SIFT, SURF and ORB-FREAK each as the feature, with segmentation, for various sizes of the vocabulary.

Feature	Vocabulary Size 2K	5K	10K	50K	100K	500K	1000K
SIFT	81.2%	87.6%	87.8%	93.9%	96.1%	96.3%	96.7%
SURF	68.7%	71.4%	72.8%	79.6%	84%	92%	92.4%
ORB-FREAK	49.8%	55.8%	56.6%	65.2%	66.1%	69.3%	71.1%

3. (a) Precision at 10 with segmentation. (b) Comparison between accuracy of SIFT BoW + GV and segmentation + SIFT BoW + GV.



(a) (b)

Conclusions

- Succeeded in developing a system that recognizes bills reliably, and ported the system to a mobile environment.
- With limited processing power and memory, the system still achieves high accuracy and low reporting time.
- Segmentation is particularly helpful for retrieval.
- Easily adaptable to other currencies, while maintaining performance.

References

1. Jayaguru Panda, Michael S. Brown, and C. V. Jawahar. Offline mobile instance retrieval with a small memory footprint. *ICCV*, 2013.
2. C. Rother, V. Kolmogorov, and A. Blake. GrabCut: interactive foreground extraction using iterated graph cuts. *SIGGRAPH*, 2004.
3. J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. *ICCV*, 2003.
4. Xu Liu. A camera phone based currency reader for the visually impaired. ASSETS, 2008.