Currency Recognition on Mobile Phones
Suriya Singh¹, Shushman Choudhury², Kumar Vishal¹ and C.V. Jawahar¹
¹IIIT Hyderabad ²IIT 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:
  \[ E(x, y) = \sum_{i} \log p(y_i | x_i) + \sum_{(i,j) \in \mathcal{E}} S(x_i, y_j | x) \]
  where \( x \) is the colour of the \( i^{th} \) pixel and \( y \) is +1 if the pixel belongs to the object, otherwise -1. \( S(x_i, y_j | x) \) favours neighbor pixels with similar color to have the same label.
- Segmentation results

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 sifting 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

<table>
<thead>
<tr>
<th>Component</th>
<th>Size</th>
<th>Time (in seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAM use (on average)</td>
<td>13.5MB</td>
<td>SVM: 0.25 s</td>
</tr>
<tr>
<td>Inverted index</td>
<td>20.5MB</td>
<td>ORB-FREAK: 0.15 s</td>
</tr>
<tr>
<td>Vocabulary (10K)</td>
<td>3.5MB</td>
<td>SIFT: 0.2 s</td>
</tr>
<tr>
<td>Keypoints location</td>
<td>11MB</td>
<td>SURF: 0.2 s</td>
</tr>
<tr>
<td>Annotations</td>
<td>6.9KB</td>
<td>Total Recognition (seconds)</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Vocabulary Size</th>
<th>2K</th>
<th>5K</th>
<th>10K</th>
<th>50K</th>
<th>100K</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIFT</td>
<td>81.2%</td>
<td>86.7%</td>
<td>87.6%</td>
<td>93.9%</td>
<td>96.1%</td>
</tr>
<tr>
<td>SURF</td>
<td>68.7%</td>
<td>71.4%</td>
<td>72.8%</td>
<td>79.6%</td>
<td>84%</td>
</tr>
<tr>
<td>ORB-FREAK</td>
<td>49.8%</td>
<td>55.8%</td>
<td>56.6%</td>
<td>65.2%</td>
<td>66.1%</td>
</tr>
</tbody>
</table>

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

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

- Android App, Code and Data are available on Project Web Page

http://researchweb.iiit.ac.in/~suriya.singh/Currency2014ICPR/

Centre for Visual Information Technology  |  http://cvit.iiit.ac.in
IIIT – Hyderabad, INDIA  |  http://www.iiit.ac.in