Object Detection using a Handheld Android Device for Processing Image Data in Real Time

Ishan Agarwal
Jaypee Institute of Information Technology
Noida, India
ishanag93(at)gmail.com

Bansidhar Joshi
Jaypee Institute of Information Technology
Noida, India
bansidhar.joshi(at)jiit.ac.in

Shalabh Kumar
Jaypee Institute of Information Technology
Noida, India
shalabhkumar100(at)gmail.com

ABSTRACT
All through decades and decades of time since the tourism industry even came into existence, we have all known about assistants both digital and personal guides that walk us through the famed attractions of tourist spot. At the end of the day, one is left with little information, lesser money and absolutely no excitement that tourist usually come to enjoy at these places. It would be simpler if the visitor could further personalize the experience of visiting and experiencing the destination. Old and obsolete traditions of obnoxious, damaged headphones and travel guides could be given some relief making the tourist's exploit not only more enjoyable but also wholesome. Our landmark recognition application will cater to the same resulting in a never before experienced encounter for the end user

General Terms
Algorithms, Documentation, Performance, Reliability, Experimentation

Keywords
ORB (Oriented FAST and Rotated BRIEF), RANSAC, Landmark Detection, Object Detection, Classification of image data

1. INTRODUCTION
When people go to tourist attractions and/or landmarks, first thing they look for is a tour guide. Which is a person or an audio device or a book that can give the tourists the complete details about the place. But often it becomes frustrating for the tourist to stand in long queues and wait to purchase the same, despite shelling out several bucks to procure it. On an average, a human tourist guide costs around Rs. 300 (whose credibility is not guaranteed), an AV tourist guide costs around Rs. 200 to Rs.300. Even after the money has been spent on the part of the tourist who travel around the visiting city in excitement to come across something new, all they end up getting is the bookish knowledge that is already available on the internet which can be accessed sitting at home or worse still, they are given a mugged up summary of what is already written on the explanatory boards that record tiny accounts of the place as managed by the authorities of the monuments.

Moreover, availed guide services at the visited tourist attractions are pretty much rushed up since the resource is already in for bids by the next customer that may generate added monetary benefits for the provided information. Tourists end up with a half – hearted experience not just being denied of value for money services but also ruining the leisure time they have planned for their anticipated outing.

We realized that it could be an unparalleled alternative for the tourist if he could use his cellular or mobile device as a replacement for all these services he is exploiting at the venue. He would end up saving money, time and would also be able to run on the schedule he has planned. Not only would it eliminate a major hurdle this industry is facing in terms of client interaction but would also bring about a revolution in the machinery of this niche. Not only would the user of such an application exploit the benefits he was being bestowed upon already but he would be subject to vast range of new options that will come to him once the destination he is visiting and descriptive information he desires about it is linked via the application to the abundant knowledge stores of the internet.

The main objective of this project is to convert all android devices to personal guides by completely eliminating the above mentioned traditional guides with an application by which users can instantly retrieve the required information about the tourist attractions by simply hovering their device's camera towards the landmark. This application would be more like a window to globalize the hospitality and tourism industry of today. The much hurdling gap between publicizing social media and the thrilling experiences tourism industry caters to clients will be fulfilled.

The application, or WhatsThat as we call it, is capable of recognizing a landmark matching / comparing a fixed set of descriptors that have been pre-fed into a created dataset. A query image with an ideal resolution of 320x240 pixels is generated. The query image will then be matched with all candidate landmark images from the dataset. Now, a homography will be estimated between the query image and the top retrieved image from the matching phase. The RANSAC algorithm will then be used to estimate the homography. If the number of inliers in the homography model exceeds a threshold, a match will be declared, otherwise the next query frame will be processed to look for a potential match.

ICTCS '16, March 04-05, 2016, Udaipur, India
 ISBN 978-1-4503-3962-9/16/03..$15.00
DOI: http://dx.doi.org/10.1145/2905055.2905224

The application has been designed keeping in mind concepts of ORB (Oriented FAST and Rotated BRIEF). ORB is two orders of magnitude faster than SIFT and hence a better option when it comes to working on a mobile application development.

The introduction section deals with a basic overview of the application and concise idea of what developers are aiming at. The technical overview in section 2 builds upon this idea and further explains the resources and the research works that may be utilized in achieving the same. The test results of the application have been incorporated under the Experiments and Results in section 3. Section 4 concludes the paper.

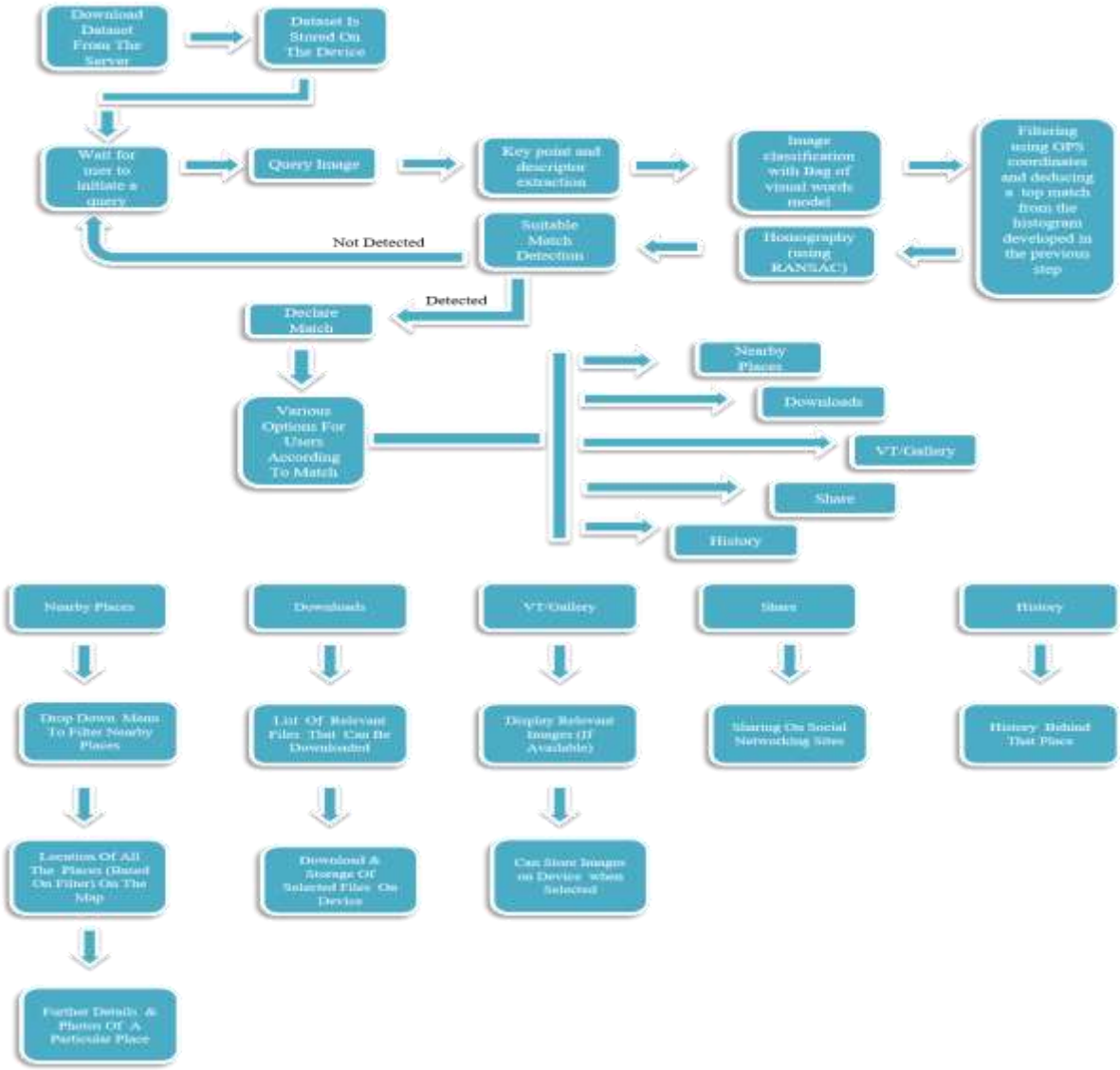

Figure 1: Overview Diagram

2. TECHNICAL OVERVIEW

WhatsThat! A Mobile Based Landmark Recognition Application as the name suggests is designed for tourists to give them a new experience to their travelogue.. An overview of the process goes as follows.

The application opens up with a camera frame, so that the user wastes no time when he arrives at the desired location. One touch toggle button, does all the work that is intended for the application. Once the toggle button is activated it captures the hovered image as a query. The query image is generated and is sent to the backend for a match and the location is recognized. Since now the user would desire some feature to guide him along, make his experience last longer and ecstatic, a menu appears that lets him use leisure functionality he might like. The menu opens up with the following features- Virtual Tours/Gallery, History, nearby places, Downloads and Share. These have been included keeping in mind a tourist's requirements/objectives while visiting a tourist destination. For example, a tourist would like a guide to walk him through the place with appropriate information. Such functionality is being provided by Virtual Tour. Nearby places would let the user discover places near the current location that he might like to visit or that may be related to the spot he is visiting. The Share option lets you share your experience online. The History tab highlights the historical precedence as well as all important news down the timeline about the landmark.

The dataset, trained using 500 images of Jaypee Institute of Information Technology, Noida, is stored in the device memory.

This dataset is referenced into the application and the application is trained on the aforesaid. Thresholds to decide matches obtained against these reference images are also declared and stored in the form of an array.

An array already stores the GPS coordinates of these selected destinations in the dataset. The image is downsized for efficient processing. Through GPS coordinates of the device and the pre-stored coordinates of the stored sets, possible matched sets are filtered out for processing. FAST [8] and BRIEF [10] implemented within the application provide with the keypoints and the descriptors of the images which are then matched with those of the query image to get a shortlisted set of matches. RANSAC [5, 7] further shortens this set of good matches to get the minimum best possible match for this query image. This then returns a string that holds the name of the identified landmark and displays on the application interface along with the return string that holds a short description of the same. Beyond this, the user interfaces makes use of the standard Google APIs to detected and display nearby locations on a map.

Figure 2: Application interface

2.1 ORB (ORIENTED FAST AND ROTATED BRIEF) [6]

Implementations of ORB in the backend provide us with pixel data in a format that can be processed to evaluate matches in the form of keypoints and descriptors.

FAST (Features from Accelerated Segment Test), runs a series a binary tests between the center pixel in a patch and those in a circular ring around the center, with threshold variants to extract good keypoints to cumulatively detect and identify a distant image.

BRIEF (Binary Robust Integrated Elementary Features) gives us the string representations of these keypoints obtained after "n" such binary tests on different patches. Color gradient and color information is taken into account. The relative orientation of a particular patch with reference to the initial frame can also be stored in terms of a tan function value that can later be compared to judge the relevance of a differently oriented similar patch in the query image. Thus descriptors and keypoints provide all relevant information about the image.

2.2 RANSAC (RANDOM SAMPLE CONSENSUS) [5,7]

RANSAC, or the Random Sample Consensus, given by Fischler and Bolles copes up with a large number of outliers in the input data. It uses the minimum number of points from within the underlying model to give the model parameters. This set is then enlarged to cover the entire model.

RANSAC follows a simple process. It selects a minimum number of points and checks if they surpass a minimum tolerance e. The fraction of these points or inliers that result from this test are required to meet with another parameter t, both parameters defined for the model at hand. This process is repeated for different sets of points. Thus a concise set that represents the model is obtained.

3. EXPERIMENTS AND RESULTS

We performed a series of tests on the application in order to gain head to toe performance measure. We ran the application through different devices taking multiple query images of the same object. The observations could be summed up as under.

The application was able to generate a result, match found or not found in less than one second on Android Platform Version 4.0 and above.

Multiple test cases for a single object using HTC One X device provided with correct results 88 out of 100 times thus putting the efficiency quotient at 88%

Different devices may yield slightly different results depending upon the hardware and quality of query image. When tested with Sony Xperia P device, the efficiency was lowered down to 85%.

4. CONCLUSIONS AND FUTURE WORK

The aim behind this project was to bring an influx of modernization and technology into the lagging tourism sector industry that is involved with the client and is yet waiting to be globalized in tune with the progressing world. Efficient image processing algorithm like ORB have significantly helped in bridging similar gaps in other application areas. An application of such technology will once again benefit the science worlds and the financial markets as well as the users that require such support.

Also introducing several other features into the application that make it more connected with the user's other areas of interest is a possibility. One radical amendment would be replacing the static dataset provided to the user from developer's end to a dynamic self-maintaining dataset. Since now dataset would acquire query images and build its own dataset with repetitive use, the intelligent dataset would make this application further user oriented and the user would be able to add new places and hence result in successfully enriching the application with new places.

4. ACKNOWLEDGMENTS

We would like to extend a gesture of thanks to the developers of ORB. Their labor and efforts provided the large scientific world with a medium to explore further depths in image processing applications.

5. REFERENCES

[1] Ethan Rublee, Vincent Rabaud, Kurt Konolige, Gary Bradski "ORB: an efficient alternative to SIFT or SURF", Willow Garage, Menlo Park, California, International Conference on Computer Vision. Pages 2564-2571,2011

[2] M. A. Fischler, R. C. Bolles. "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", Comm. of the ACM, Vol 24, pages 381-395,1981

[3] C. Grana, D. Borghesani, M. Manfredi, R. Cucchiara, "A Fast Approach for Integrating ORB Descriptors in the Bag of Words Model" in Proceedings of IS&T/SPIE Electronic Imaging: Multimedia Content Access: Algorithms and Systems, vol. 8667, San Francisco, California, US, pp. 091-8, Feb. 2-7, 2013

[4] D. Langdon, A. Soto, and D. Mery, Automatic Selection and Detection of Visual Landmarks Using Multiple Segmentations. Lecture Notes in Computer Science, LNCS 4319:601-610.

[5] Overview of the Ransac Algorithm by Konstantinos G. Derpanis. Available : http://www.cse.yorku.ca/~kosta/CompVis_Notes/ransac.pdf

[6] ORB (Oriented FAST and Rotated BRIEF), Available : http://docs.opencv.org/trunk/doc/py_tutorials/py_feature2d/py_orb/py_orb.html

[7] RANSAC, Available : http://docs.opencv.org/trunk/doc/py_tutorials/py_feature2d/py_feature_homography/py_feature_homography.html

[8] FAST algorithm for Corner Detection, Available : http://docs.opencv.org/trunk/doc/py_tutorials/py_feature2d/py_fast/py_fast.html

[9] For Developing the Android Application Available : http://developer.android.com/

[10] BRIEF, Available : http://docs.opencv.org/trunk/doc/py_tutorials/py_feature2d/py_brief/py_brief.html

[11] Arc Menu, Available : https://github.com/daCapricorn/ArcMenu