Image Based Information Retrieval Using Geo-Tagging

Prof. Prachi Abhyankar[1], Munaf M. Kapdi[2], Vishwa V. Mehta[3], Vihar V. Mehta[4]
[1, 2, 3, 4] Dept of Information Technology
[1, 2, 3, 4] Mumbai University, Ratnagiri, India

Abstract- *The project is aimed at developing an application/website for providing the rental services and details. The application verifies the user through OTP. These verified users can upload the details regarding rents, location of area, contact details, etc. The user will be able to provide reviews and feedback. The application provides filter so that users can find the area based on their preferences. The application and website will provide geo-tagged locations of the rental areas. The application will be able to recognize the well-known infrastructure by image processing and provide the details accordingly.*

Keywords- geo tagging, image retrieval, content based search, augmented reality.

I. INTRODUCTION

The popularity of digital cameras and online societies has led to an outburst of personal and web images. It has become a serious task to manage such an overwhelming amount of image data. currently, commercial search engines and web albums rely on text explanations associated with images for indexing and retrieval tasks. Comfortable and more correct semantic annotation would benefit many applications including image search, sharing, organization, and management. recognizing the need for semantic annotation, the latest version of the google™ picasa™ now enables users to label images in terms of faces, places, and user-specified tags. While face recognition has been widely used in existing systems, inference about the location of images has of late received increasing attention in research as well as commercial photo management systems.

Human beings, over the years, have built rich vocabularies to describe sets, objects, people, and places taken in pictures. Most such words rapidly strike geographical associations in our minds. These geographical connotations may vary from being rather specific (e.g., for Paris) to being fairly general (e.g., for beach). For human beings, building such families is natural and results from conditioning and education. Additionally, humans possess the unique capability to analyze the visual content of pictures and make fairly educated guesses as to their where abouts. In fact, Google has recently introduced an online game “Where in the World” (Fig. 2) to tap this human potential. Making and preserving these geographical associations with pictures is an age-old process. During the “film camera” days, people would write the place where the picture was taken on the back of the print. Today a user can map his pictures exactly using community image management systems such as Google Picasa, Google Earth, and Yahoo® Flickr.[1]

The capture of geographic coordinates or the accessibility of geographically pertinent tags with pictures opens up new data mining possibilities for better recognition, classification, and retrieval of images in personal collections and the Web. In a recent work, Luo et al. got satellite images consistent to picture location data, and proposed an approach to event recognition by fusing information from the ground image and the co-located satellite image. The powerful union of the balancing views results in significant performance improvement over the ground view baseline. With integrated GPS-capable cameras coming to the consumer market, it is expected that this line of research should transform event recognition and media annotation in years to come. Instinctively, reverse geocoding using entries in a georeferenced namespace database can be supportive for classifying the picture-taking location and help classify an event. However, there is a need to advance more precise and more specific semantic knowledge of a location to help classify an image captured at that location, because of four main problems: (1) a place is represented as a point (e.g., the central office in a zoo) in the database without any definition of the actual spatial extent; (2) multiple environments can potentially co-locate in close proximity of each other; (3) many georeferenced namespace databases are rather crude (e.g., no database marks a tennis court, which can be part of a school or park, or mark an area as residential or commercial); and (4) geotagging can be rather inexact because of the blast in sensor or human tagging.

An important research question that encourages our current work is how this huge size of public geotagged image data on the Web can be leveraged to geotag or allocate geographic locations to images, especially legacy pictures that were taken before cameras could line directly with GPS receivers.

II. RELATED WORK

Content understanding in images has been studied for periods in the vision research community. Content understanding in images can translate to understanding scene semantics or occasion semantics. Invariably, image content understanding algorithms involve building classifiers for a finite number of semantic groups. A strong application of image understanding is image retrieval. However, learning-based retrieval is controlled by the cardinality of semantic groups. Another line of research has, for a long time, explored unsupervised similarity-based search and retrieval using low-level visual features alone. Recently, brute force searches using huge image databases have been shown to be valuable for image understanding tasks as well. Such methods, which rely on retrieval for semantic understanding, whole a full circle in linking the fields of image retrieval and image understanding. However, all of the above systems still focus on only the image satisfied. With rapid advances in skills related to digital imaging, digital cameras also bring with them a powerful source of data little abused previously for scene classification: camera metadata fixed in the digital image files. Camera metadata (or “data about data”) records information related to the image capture situations and includes values such as date/time stamps, subject distance, and GPS coordinates. They contain rich background information that is usually opposite to the image features for the purpose of semantic understanding. The research community progressively turns to metadata and picture-taking setting in the pursuit to solve the semantic understanding problematic.

Significant metadata can be collected also as a result of user contribution. Online photo-sharing websites such as Flickr have observed a surge of cooperative tagging from users, resulting in folksonomies. Recently, there have been research efforts to understand user image tagging performance [1] and to characterize this behaviour over time [8]. When users associate geographic content with media on the Web, it becomes an example of geotagging. With the growing popularity of geotagging, mining, organizing, and making sense of georeferenced data and relating geo-content to visual content has become essential. Initial attempts to identify geo-relevant content on Web pages in order to assign a geographic focus to pages were made in [2]. Retrieval of geographical landmarks from the Flickr dataset was studied using a combination of visual features and geotags in. An algorithm to create summaries of georeferenced collections was proposed in to improve browsing and visualization of images. Season and location context was found to be useful for region labelling in. The problem of finding relations between places and picture semantics was studied in. Of late, the stress has changed to landmark recognition in very large image datasets.

III. METHODOLOGY

The first stage was getting everyone in the group acclimated to the new development platform. We did this by finding and sharing resources (documentation, sample code, etc.), as well as running through the set-up process for the Android SDK as a group. Before beginning development, we had established documentation guidelines and testing frameworks in order to enhance maintainability and prevent bugs in the long run.

Then, we worked up specifications, established deadlines, defined roles, and allocated tasks to each group member. We planned on dividing up the development process into two central phases, core development and feature development. Core development involved implementing the essential aspects of our application, whereas feature development has encompass any additions which extended the app’s capabilities. We have implemented these additional features in order of priority.

IV. WORKING PRINCIPLE

When an image is manually tagged, the user associates remarks with the image, which are expressive and may transmit information related to the location of the image. In some cases, the association is direct: An image tagged “Chicago” is quite probably taken in the Illinois city. However, in other cases the association is more understated but still informative. For example, an image tagged “snow” is not likely to have been captured near the equator. Other tags, such as “smile,” contain little information regarding the location of the image capture. Readers can confirm the advantage of user tags, where image content alone could have run to incorrect reading about the location of the images shown in the figure; neither picture was taken at the most likely location for the subject. Even if one thinks that one knows the location of an image from the content, the tags collectively can provide valuable information.

Fig 1 significance of user tags in geo localization assignment. An algorithm that uses only visual cues is extremely likely to wrongly predict the location of the images to be (left) New York City, and (right) Africa [1].

Augmented reality (**AR**) is a direct or indirect live view of a physical, real-world environment whose elements are "augmented" by computer-generated perceptual information, ideally across multiple sensory modalities, including visual, auditory, haptic, somatosensory,and olfactory .[1] The overlaid sensory information can be constructive (i.e. additive to the natural environment) or destructive (i.e. masking of the natural environment) and is spatial registered with the physical world such that it is perceived as an immersive aspect of the real environment.[2] In this way, Augmented reality alters one's current perception of a real world environment, whereas virtual reality replaces the real world environment with a simulated one.[3][4] Augmented Reality is related to two largely synonymous terms: mixed reality and reality.

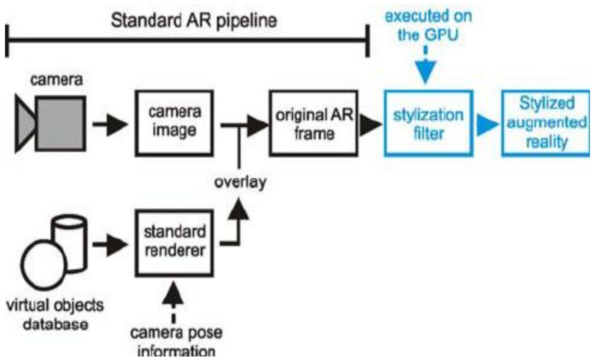

Fig 2 : Augmented Reality working principle

Types of AR technology are:

Marker-based augmented reality: Also called Image Recognition uses a camera and some type of visual marker, such as a QR/2D code, to produce a result only when the marker is sensed by a reader. Marker based applications use a camera on the device to distinguish a marker from any other real world object. Distinct, but simple patterns (such as a QR code) are used as the markers, because they can be easily recognized and do not require a lot of processing power to read. The position and orientation is also calculated, in which some type of content and/or information is then overlaid the marker.

Fig 3 : Marker based augmented reality

Markerless augmented reality: As one of the most widely implemented applications of augmented reality, markerless (also called location-based, position-based, or GPS) augmented reality, uses a GPS, digital compass, velocity meter, or accelerometer which is embedded in the device to provide data based on your location. A strong force behind markerless augmented reality technology is the wide availability of smartphones and location detection features they provide. It is most commonly used for mapping directions, finding nearby businesses, and other location-centric mobile applications.

Fig 4: Markerless augmented reality

Visual features and matching

In our work, we have adopted visual feature extraction and matching practices from [10]. While [10] used six groups of visual features. We began with the set of features used in [10] and assessed the performance of individual landscapes on a small random subset of our exercise corpus. Performance was determined in terms of the quality of graphic matching performed using the feature alone as judged by humans. Additionally, we tried to minimize idleness and include features that encode opposite information (color, structure, and texture). The four features that we have selected are widely used in computer image and are effective for matching a large spectrum of visual content.

A K-nearest-neighbor search is active for visual matching. Distances between images are calculated another way for different features. The GIST descriptors are related across images using Euclidean distance. The tiny images are compared using regularized cross correlation. We employ a $\chi 2$ distance measure to match color histograms and texton histograms as these features are integrally probability distributions. A combination of distances using multiple features is performed linearly by using feature-specific encumbrances learned from small random subsets of data.

Geographical location prediction

Our tag-baseline consists of the best informative tags inclusion plan on the basis of accuracy within 100 km range of prediction.

Once the K-nearest neighbors are retrieved for a query image, their geographical locations are represented as (longitude, latitude) pairs. These are then mapped onto the 900×1800 geographic regions.

Once the geo-maps based on visual nearest neighbors have been built, the integration can be handled by extending the inference way. The geo-map obtained from KNN is treated as another probability map and integrated.

V. FEATURES

1. This application will help user to gain information regarding any place by just using the device's camera
2. This will give you further a option to find the best possible route to that location..
3. It also has an option to view other details regarding place such as any photo, video, etc.
4. A full-fledged Augmented Reality experience the user.
5. The app also checks for internet connection to ensure appropriate functioning.
6. It has a better use interface.

VI. RESULTS AND DISCUSSION

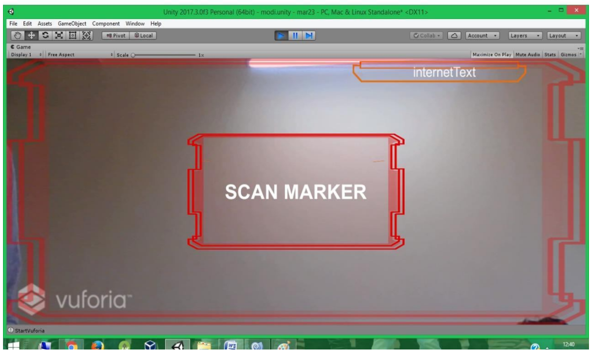

Fig a: Scan the Marker

Fig. (a) shows the marker where the place(in this case an object) will be scanned and than based on the object the corresponding results will be displayed.

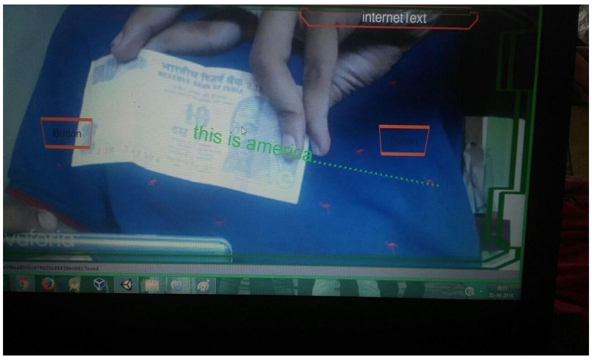

Fig b: Object displayed after scan

Fig. (b) by scanning the object(rupee note) a message giving information is displayed along with two buttons. This buttons will help user to get directions and more information of place such as images, videos, etc.The internet connection check is displayed on top right corner.

IX. CONCLUSION AND FUTURE SCOPE

There have been many discussions, and blogs posting regarding augmented reality (AR). As the concept of AR was recently introduced few years ago, some review papers are published as still only few effective applications have been implemented. This paper will help realize end users that how this application is simple and no special efforts are required to learn how to use this application. Users will also realize the need of this application as it makes traveller in a new city guide about the places around them.

X. ACKNOWLEDGEMENT

This research paper was made possible by the support of Dr. Vinayak Bharadi, HOD, IT Department, FAMT;Prof. Prachi Abhyankar, Project Guide. We would like to express our great gratitude to Prof. Prachi Abhyankar for his kind advice on the project and precious information.

REFERENCES

[1] " Inferring photographic location using geotagged web images" by H Dhiraj Joshi, Andrew Gallagher, Jie Yu, Jiebo Luo, Multimedia Tools and Applications,Springer. January 2012, Volume 56, Issue 1, pp 131–153.
[2] Ames M, Naaman M (2007) Why we tag: motivations for annotation in mobile and online media. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems
[3] Amitay E, Har'El N, Sivan R, Soffer A (2004) Web-a-where: geotagging web content. In Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval
[4] Chen Y, Chen XY, Rao FY, Yu XL, Li Y, Liu D (2004) LORE: an infrastructure to support location-aware services. IBM J Res Develop 48(5/6):601–616
[5] Cilibrasi RL, Vitanyi PMB (2007) The Google similarity distance. IEEE Trans Knowl Data Eng 19(3):370–383
[6] Comaniciu D, Meer P (2002) Mean shift: a robust approach toward feature space analysis. IEEE Trans Pattern Anal Mach Intell 24(5):603–619
[7] Crandall D, Backstrom L, Huttenlocher D, Kleinberg J (2009) Mapping the world's photos, WWW
[8] Datta R, Joshi D, Li J, Wang JZ (2008) Image retrieval: ideas, influences, and trends of the new age. ACM Comput Surv 40(65)
[9] Dubinko M, Kumar R, Magnani J, Novak J, Raghavan P, Tomkins A (2006) Visualizing tags over time. In Proceedings of the World Wide Web
[10] Gallagher A, Joshi D,Yu J, Luo J (2009) Geolocation inference from image content and user tags, CVPR Workshop in Internet Vision
[11] Hays J, Efros A (2008) IM2GPS: estimating geographic information from a single image. IEEE International Conference on Computer Vision and Pattern Recognition