IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 6, Ver. V (Nov – Dec. 2014), PP 53-56
www.iosrjournals.org

ComparativeAnalysis of Automated Image-Tagging Techniques

Mayuresh Jakhotia[1],Dheeraj Agarwal[2],Mithun Sanghvi[1],Shubham Shah[1]
[1]*ComputerDepartment, VIIT, PuneUniversity*
[2]*ComputerDepartment, MESCOE, PuneUniversity*

Abstract: *Smart phones with cameras having powerful built-in sensors tend to meet at a point towards the problem of automatic image-tagging. So such an out-of-band approach is valuable, especially with increasing device density and greater sophistication in sensing and learning algorithms. This paper shows that automatic tagging of images with relevant tags is possible by using a combination of the capture of location, the date/time when the image was captured, and an image category. Comparison of various automatic image-tagging approaches has been done so as to develop a new system with a dynamic approach for using multiple images when possible, and fewer images when not many relevant images are found.*
Keywords: *Annotation, context, digital images, image tagging, sensing.*

I. Introduction

In the past, roll of film was given to a professional photographer so as to develop images into handheld photographs or to prepare a photo-album. In today's world, all the images are captured and stored digitally. It is still a tedious job for annotating and tagging a given picture automatically. It is vital to handle large number of pictures getting stored in online content repositories. It will be very much subjective and time consuming to tag and annotate images manually. Thinking practically, people simply do not bother or have time to tag their images. Furthermore, human beings think differently with different innovative opinions, meaning that similar images will be tagged differently by different people. This can be due to variations in language, mood, vocabulary, education, culture, taste etc. Some of the digital cameras already have built-in GPS like Panasonic TZ10 and the number is increasing in a voluptuous manner. There also exist solutions where a GPS receiver is attached to the flash connector of digital cameras. The images are then geo-coded when they are transferred from the digital camera to a computer with Internet connection. Furthermore, several of the mobile phones on the market today like iPhone 5S is equipped with both accurate GPS systems and cameras able to take images with high and better quality.

Thus, it is very likely that a lot of images in near future will have GPS coordinates available, generated either automatically or manually. The location where an image is taken can be a very valuable asset when tagging images. Location can be combined with other contextual information sources such as weather information, nearby buildings and facilities, date/time (in case of an event taking place), other images taken nearby and geo-referenced articles. This information can be helpful when automatically tagging images. The increase in digital images and research in image related topics together with the problems concerning manually tagging of images indicates that there is a need for automatic image tagging. Mainly for two reasons we believe smart phones could make a difference in solving this auto tagging problem. The first is that today's smart-phones have very powerful built-in sensors, we all know that, and the second is people always carry their phones with them. There are number of sensors including Accelerometer, Gyroscope, Compass, GPS, Camera, Microphone and Cloud. These different sensors have the capability capturing the "moments" across multiple sensing dimensions because the microphone might be able to detect laughter more naturally, dancing can exhibit an accelerometer signature, and light sensors may easily differentiate between indoor and outdoor environments.

II. Literature Survey

A. SpiritTagger

Spirit Tagger is a geo-aware tag suggestion tool which uses Flickr that depicts geographically relevant tags for images with GPS coordinates [1]. It is done by combining the geographical context with content-based image analysis. Geographic mining is done by collecting a set of images that are within a certain radius of the candidate image to be tagged. This set of images is narrowed down by using visual similarity techniques. The tags of the images in the set are then compared to their global frequency. Local frequency refers to the frequency of a tag in the result set, whereas global frequency refers to the frequency of a tag in all images on Flickr. Tags with higher local frequency than global frequency are assumed to be relevant for the query image. Experiments have found that SpiritTagger works well as compared to baseline methods that only use geographical context [9].

B. MonuAnno

MonuAnno automatically annotates landmark images [2]. They refer to landmarks as geographically situated objects or small areas such as Eiffel Tower and Big Ben. An important part of the system is a reference database of landmarks generated based on image locations and visual similarity from images on Flickr and Panoramio. The annotation of a query image consists of two steps. The first step is to decide which of the nearest landmarks the query image belongs to. The second step is to verify that the query image indeed belongs to that landmark. Visual similarity and location is used in both steps. Finally, whereas MonuAnno only tag with the name of the landmark, the focus of this work will be to find a set of relevant tags for a query image.

C. ZoneTag

ZoneTag is a mobile phone application allowing and encouraging users to easily upload images taken with their mobile phones directly to Flickr at the time of capture [3]. It also suggests tags based on context information such as previously used tags and names on nearby attractions. The client (mobile phone) communicates with a server that performs computational and time consuming tasks unsuitable for mobile platforms. It differs from the work in this thesis mainly in that ZoneTag does not use categories. Further, location information is only approximated if exact GPS location is not available. Also, ZoneTag is a mobile application that uploads images as they are taken. Ames and Naaman performed a user study using Flickr and ZoneTag and exposed that the main motivation for tagging images is functionality [6]. People want to tag and organize images to make it easier both for them and others to search, browse and retrieve images. Sigurbjornsson et al. the image was taken. This was found by performing an experiment classifying tags from a set of images on Flickr with the use of the classification system used by WordNet [7].

D. AnnoSearch

AnnoSearch is a system that annotates images based on search using a keyword and the image itself [4]. First, a text-based web search is performed to find a set of semantically similar images. AnnoSearch then use the query image to find a set of visually similar images. Next, the two set of images are clustered into sets of keywords (for example castle, cloud and tree). Finally, these keywords are ranked according to frequency and visual similarity to the query image. The top ranked keywords are assigned as tags to the query image. Experiments on 2.4 million images proved the effectiveness and efficiency of the system [3].

E. Other Surveys

Rattenbury et al. shows that it is possible to check whether an existing tag on Flickr represents an event or a place [5]. They demonstrate that if a certain tag represents an event or a place, then that tag must have a significant higher frequency in a certain time scale and/or in a certain area compared to its general frequency outside this time scale or area. Date and time in itself is often not enough to base automatic image tagging systems on. But it can be very useful when used in addition to other approaches. When fewer images are available to work with, it is obvious that it is more difficult to find relevant tags. This seems to agree with the general consensus in the field of image retrieval and image annotation; that handling events is a difficult task.

F. Context-Based Image Analysis Approach

For systems using content-based image analysis/visual similarity techniques (e.g. Spirit Tagger, MonuAnno and AnnoSearch), non-relevant images can be discarded based on their low similarity score compared to the query image [9]. On the same basis, images that get low similarity score com-pared to the majority of the returned images (or the normal of the returned image set) are also likely to be non-relevant. This is an advantage compared to systems not using visual similarity because it is harder to discard images that are not relevant when the visual content of the image is not analyzed. However, a severe problem with the content-based image analysis is that images are taken from different views and angles. This makes it harder to find visual similarity among images of the same attraction. The back side of a building is not necessarily very similar to the front side of the building, and the background can also be significantly different on images taken in opposite directions (or in a different season, time of day etc.). Similarly, images from an event does not necessarily have to be visually similar (consider a concert where images are taken both of artists and spectators). Another related problem is that people often take images of themselves in front of attractions, which will disturb the visual similarity techniques. It was possible to collect relevant context information related to an image using date/time of image capture, capture location and a user-defined image category. These works demonstrate that it is possible to use category, date/time and location to collect relevant information about the image.

G. TagSense, LoCaTagr, LoTagr and SimpleTagr

TagSense is a system wherein users time will be saved tremendously compared to Facebook and Flickr which ask users to provide the tags (at least in the initial stages) [8]. TagSense overcomes this by making auto generated tags on-the-fly for smart phones taking pictures rather than asking human to do it. It does it by using the multiple sensing domains of current smart phones coupled with its powerful built-in sensors. Using a WiFi Adhoc network all the sensors of nearby smart phones in a group are activated, and on shutter press all the phones reply with their sensing data which is now processed on the camera phone to generate the tags [8]. LoCaTagr use Flickr to find a set of relevant images and retrieve tags from these relevant images [9]. LoTagr and SimpleTagr use only location and these systems therefore only find information relevant for the location where the image was captured [9].

III. Comparative Analysis of Survey Papers

Recall (number of relevant tags found / number of relevant tags that could have been found) is not used in the evaluation as it near impossible to decide how many tags that could be used to tag an image. It would possibly be even more correct to use total number of relevant tags that are available on Flickr for a certain image. But again, it is near impossible to find all relevant tags that could and should be used for a certain image.

Precision is the number of relevant tags divided by total number of tags found .A high precision score will prove that most of the tags found are relevant. A perfect precision score (1.00) means that all tags found are relevant. Precision will regard the unsure tags as noisy.

SpiritTagger will always suggest exactly 20 tags whereas the other systems have a more dynamic approach which results in selecting tags only when they have a certain frequency, i.e. when they appear in at least 20 % of the images in the result set. This is a big disadvantage for SpiritTagger with regards to the precision scores [9]. The idea in SpiritTagger is that the system suggests a set of tags and that the user is supposed to pick the most relevant. Hence, Picasa is giving the best result with highest average precision [8].

TABLE I
COMPARATIVE ANALYSIS OF SURVEY PAPERS

Sr.No	Technique	Technology and Parameter	Average precision
1	SpiritTagger	Flickr, location, context-based image analysis.	0.23
2	SimpleTagr	Location, optimizations	0.27
3	LoTagr	Location, Flickr	0.44
4	TagSense	Sensors like Accelerometer, Gyroscope, GPS, Camera, Microphone , Compass , Cloud	0.79
5	LoCaTagr	Location, Image Category, date/time Flickr	0.86
6	iPhoto	Digital Camera, red-eye filter , Sensors	0.90
7	Picasa	Digital Camera, Exporter for iPhoto.	0.98

IV. Problem Definition

The main motive of the work in this paper is to develop a comparative analysis of survey papers with the help of average precision in order to design, implement and evaluate a system that automatically finds tags for images based on image category, date/time and location (GPS co-ordinates). The idea is giving image category and query image as an input to the system. The tags have been collected from an online image sharing database (Flickr or any other) with images that are already tagged. However, the information in these community based collections is often highly unreliable and noisy. Therefore, the paper will further focus on the significance of the collected tags, and how this is affected by using a combination of the context sources (location, date/time and image category) as input to the system.

Location is generally a widely used context-source and there exist several location based image tagging systems. However, as far as I know, no previous work has looked into the possibility of combining location with image categories and date/time. The approach helps to handle similar types of images (belonging to the same image category) in a specific way giving the desired results for that specific type of image.

The contribution of this work is to explore the possibility of making an automatic image tagging system based on combining category, location and date/time. The hypothesis is that using categories together with location and date/time will result in more relevant and less non-relevant tags than by using other approaches.

V. Conclusion and Future Work

Now-a-days, mobile phones are becoming inseparable from humans, replacing traditional cameras. After the comparative study it is found that, LoCaTagr finds very few noisy tags despite using a noisy image database [9]. The noise level is kept low because the category approach restricts the usage of images that are not relevant. Further, usage of tags from same users is restricted and the system utilizes a dynamic approach for using many images when possible, and fewer images when not many relevant images are found. It also performs very good compared to SpiritTagger, which use both location and content-based image analysis. However, content-based image analysis proved to be useful for detecting certain tags such as architecture, sky and city. This shows that visual similarity techniques can be used to detect tags which are not necessarily linked directly to the image category or location.

Smartphones are becoming context-aware with personal sensing and are replacing point and shoot cameras. The granularity of localization will approach a foot.

References

[1] Moxley, E., J. Kleban, and B.S. Manjunath,"SpiritTagger: A geo-aware tagsuggestion tool mined from flickr, in Proc.1st ACM international conference on Multimedia information retrieval. 2008, ACM: Vancouver, British Columbia, Canada. p. 24-30.

[2] Popescu, A., et al., "MonuAnno: Automatic annotation of geo-referenced landmarks images", in Proc.ACM International Conference onImage and Video Retrieval. 2009, ACM: Santorini, Fira, Greece. p. 1-8.

[3] Ahern, S., et al., "ZoneTag: Designing Context-Aware Mobile Media Capture to Increase Participation, Pervasive Image Capture and Sharing", in Proc.Eight International Conference on Ubiquitous Computing (UbiComp2006), 2006.

[4] Wang, X.-J., et al., "AnnoSearch: Image Auto-Annotation by Search", in Proc.2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2. 2006, IEEE Computer Society.p.1483-1490.

[5] Rattenbury, T., N. Good, and M. Naaman, "Towards automatic extraction of event and place semantics from flickr tags", in Proc. 30th annualinternational ACM SIGIR Conference on Research and Development in Information Retrieval. 2007, ACM: Amsterdam, The Netherlands. p. 103-110.

[6] Ames, M. and M. Naaman, "Why we tag: Motivations for annotation in mobile and online media", in Proc. SIGCHI conference on Human Factors in computing systems (CHI 2007), 2007.

[7] Miller, G.A.,"WordNet: A Lexical Database for English.Communications", in Proc ACM, 1995. Vol. 38(No. 11): p. 39-41.

[8] Qin et.al," TagSense: Leveraging smartphones for Automatic Image Tagging," in Proc.IEEE Transactions on Mobile Computing, VOL. 13, NO. 1, January 2014.

[9] Martin Haetta Evertsen, "Automatic Image Tagging based on Context Information," Master's Thesis in Computer Science, Faculty of Science and Technology, University of Tromsø, June 2010.