

A Study on Image Annotation Techniques

Reena Pagare

Assistant Professor

Department Of Computer Engineering
M. I. T. College Of Engineering

Anita Shinde

Department Of Computer Engineering
M. I. T. College Of Engineering

ABSTRACT

With the development of mobile device, large amount of digital images are generated every day. The requirements of effective indexing and searching image are growing rapidly. Image annotation is an effective way for content based image retrieval. This paper gives a review of different categories for image annotation. The paper also gives a brief overview of different methods for image annotation under Semi- Automatic and Automatic annotations. Manual annotation is a costly and time consuming work, especially for mobile device.

General Terms

Context, Image annotation, Retrieval.

Keywords

Ontology, Semi-automatic annotation, Automatic annotation.

1. INTRODUCTION

As the mobile communication technology and user interface technology are growing rapidly, people are concerned about various internet devices like notebook, smart phone, laptop, PDA. Ubiquitous computing consists of incorporation between the information space and the physical space. Because of this, information is accessible to people anytime & anywhere through the device, connected to internet. By this, it actually reduces the difficulty of working with the device & makes it simple & well organized. The frequently changing surrounding or situation in ubiquitous computing such as location is known as context [1].

With the improvement of mobile device, huge amount of digital images are produced every day. So there is need of successful indexing and searching image. Content based method is one of the best ways for image searching. But, the search via image content is still feasible field to enhance. The main approach to search image is a keyword based way. At the same time, creation of semantic metadata about photo content remains an elusive goal. The amount of annotation can greatly improve the usefulness of photo collections as they grow into the thousands. In addition to making the photo searchable by the contributing user, tags enable users to discover other users' photos. Similarly, the usual use for annotation, personal organization and retrieval, is improved by the skill for users to render their photos on-line to be viewed by other members of the society [2].

Image annotation is the professional way for content based image retrieval. The manual annotation is high-priced and it takes more time, particularly for mobile device. There are successful contexts for a mobile device, like photo captured context, private context and social network context. These contexts are combined to acquire functional semantic content of the photo. The produced results are taken as annotation

suggestions. Annotations can include event, time, person, location etc [3].

The paper is organized as follow. Section 2 related work in Image Annotation. Section 3 introduces some of the commonly used techniques in image annotations. Section 4 explains semi-automatic way to include annotation to image by using contextual information of the mobile device. Section 5 concludes the paper.

2. RELATED WORK

Image annotation & retrieval has been a very strong area of research nowadays. In image annotation & retrieval, text annotation acts as an important role. Efficient labeling of photos has been an active research at the present time. In [4], the authors spotlight on collections of personal photos and make use of the contextual information by the associated GPS and time metadata. The photos are annotated based on collections or groups rather than individual human being. In [5], a background photo annotation approach on mobile device is introduced by author. The spatial, temporal and public contexts of a photo are offered for automatic annotation. In [6], the inspiration of annotation is studied. It explains the different ways to be taken into account while tagging their photos. In [7], the author has used the incident model in which two types of annotation proposals are taken into account. It can be either for a particular user context or for a community network. In [8], model for photo sharing on mobile phone is explained. It provides annotation propositions based on web server computing and user interactivity.

3. IMAGE ANNOTATION TECHNIQUES

3.1 Making use of Textual Information

The huge numbers of images are available on the World Wide Web. In [9] order to categorize and competently retrieve them , background information of the images such as surrounding content and associations can be used for image annotation. Automatically we can obtain semantic knowledge for Web images. Similarly context can be assigned to web images by using page layout analysis method. At the same time, the accuracy of image retrieval will be low as it this can retrieve many relevant images. There are three reasons for low accuracy. Firstly, Web images can be used by anyone in the Web pages and there is no standard exists for the relationships between the texts and inserted images in the Web pages. Secondly, Web images are fairly wide-ranging in meaning, because they are created by different group for different reasons. Thirdly, the qualities of the Web images vary significantly. The users require passing through the whole list of retrieved images to search the preferred ones.

3.2 Manual Annotation

In manual annotation users have to enter some descriptive keywords when the images are loaded / registered/browsed. Manual annotation of image content is considered “best case” in terms of accuracy, since keywords are selected based on human determination of the semantic content of images. But at the same time, it is an effort intensive and monotonous process. Manual annotation can have a problem that at the retrieval, users can forget the annotations they have used after a long period of time.

3.3 Image Annotation Based On Ontology

Semantic Web technologies like ontologies can be used to annotate images with semantic descriptions. Ontology is a design of an abstract/summary. It actually defines a collection of figurative terms called concepts. Focus of ontology based semantic image annotation is on relating the contents of an image and tries to describe image contents as completely as possible. Three layer architecture [10] for image annotation is suggested. Low level features of images are selected by the bottom layer. These features are then mapped to semantically significant keywords in the middle layer. These keywords are then connected to schemas and ontologies on the top layer.

The keyword-based approach is user friendly and it can be easily applied with satisfactory retrieval accuracy, while semantically rich ontology concentrate on the need for complete descriptions of image retrieval and advances the accuracy of retrieval. Ontology performs well with the combination of low level image features with high level textual information due to effectiveness of visual information to sort out the most of imprecise results.

3.4 Semi Automatic Annotation

There is high semantic gap between the low level visual features of the image & high level human semantic due to which performance of traditional Content Based Information Retrieval (CBIR) systems degrades. In semi automatic image annotation, it requires some sort of user participation in the image annotation process.

It [10] explains machine learning algorithms for user supported image annotation. Three layer architecture is used for image annotation. Visual information taken from the raw image contents forms the bottom layer. These contents are then mapped with the semantically rich keywords at the middle layer. Top layer consists of by mapping of keywords to schemas (structure described in a formal language) & ontologies (a formal explicit description of concepts). Machine learning together with user feedback helps to make use of previously annotated images to increase the rate of annotation for images from the same domain. It is consistent, cost-effective, fast, intelligent annotation of visual data. Intelligent Image Indexing Web Service (I3WS) is used in this approach which will take a raw image repository (along with some voluntary restrictions and parameters such as schemas, keywords, ontologies, etc.) as an input and return its annotated version as an outcome.

In [11] this to acquire the information about the semantic meaning of an image called as keywords, the image is divided based on contents including objects consisting of its category, personality & its action. The resulting semantic classification of the image as semantic class is treated as the root of the hierarchical description structure. Sequence of keywords is used to annotate the image & selection of keyword is dependent on occurrence of concept in the image. Set of

training examples are used as input where each training example is described by its low level features & corresponding annotation to the image. It results in template for annotating image with set of relevant keywords for image. To bridge the gap between low-level features and high-level semantics in retrieval systems, user participation is required in semi automatic annotation. The user is supposed improve results by using negative as well as positive examples and revise the knowledge about image classes in the semantic space.

This semi automatic [12] annotation method is combination of efficiency of automatic annotation & accuracy of manual annotation. The user has to provide the feedback while examining retrieval results. This method has three main parts: The query firing interface (A keyword query), the image browser and the relevance feedback interface. When user submits a query, search results are returned as a ranked of images as relevance with query. According to ranked list order, images are displayed on the image browser where user can view them. After browsing images the user can give feedback through relevance feedback interface. The system returns the refined retrieval results based on the user’s feedback and presents the results in the browser. This method is particularly suitable in a dynamic database system, in which new images are constantly being introduced continuously.

3.5 Automatic Image Annotation

In this [13] automatic image annotation method, image segmentation algorithms are used to divide the images into a number of unevenly shaped ‘blob’ regions and to work on these blobs. It uses the ‘global’ features for automated image annotation. This modeling framework is based on nonparametric density estimation, using the technique of ‘kernel smoothing’. User annotator has to select the word for annotation of image with the some probability. This probability can be interpreted into probability density of image x & density of x conditional upon the assignment of annotation w .

In this approach [14] a training set of images are used for automatically annotating images. Vocabulary of blobs is used for describing the regions in the image. In this method by using the training set of images with annotated keywords, it is required to predict the probability of deriving the label for the blobs in the image. Image can be seen as collection of blobs. For each, there is probability distribution called as relevance model of image. This relevance model can be treated as container holding all possible blobs that exist in the image. It will contain the keywords that exist in the image. With the help of training set of images with annotated labels, the possibility of producing a tag specified the blobs in an image can be guessed.

This method [15] uses word to word correlation because sometimes image features are inadequate in establishing the corresponding word annotation. To integrate the word-to-word correlation, it needs to approximate the probability of annotating image with a set of words. This approach uses language model to produce annotation words for image. This model contains set of word probabilities. Probability means how probably the particular word will be used for annotation. Advantage to this approach is that it automatically determines the annotation length for a given image which in turn enhances precision of image retrieval.

This [16] approach improves the existing annotations of images i.e. it will refine the conditional probability so that more accurate annotations will have higher probabilities. As effect, the annotations with highest probabilities will be kept as the

final annotations. For query image, an existing image annotation method is used to obtain a set of applicant annotations. Then, the applicant annotations are re-ranked and only the top ones with the high probabilities are considered as the ultimate annotations. While re-ranking it actually calculates fixed probability of word annotating particular image.

Table 1. Comparative analysis of image annotation Techniques

Image annotation Technique	Advantages	Disadvantages
Manual	Reliable, Accuracy in Extracting semantic information at several levels.	Tedious, Requires a lot of time & efforts, Costly
Semi-automatic	Efficient than manual annotation More accurate , useful for dynamic database	Requires User interfaces refinements to improve the feedback process
Automatic	Speed (Saves time)	Less Reliable than Manual , More error-prone, Produces more general (less detailed) annotation as compared to manual method Less Accurate than automatic annotation

4. SEMI-AUTOMATIC ANNOTATIONS ON MOBILE PHONE USING CONTEXTUAL INFORMATION

In ubiquitous computing applications, three important aspects of context are as location, identity and time. One more parameter can be considered as event to deal with photos. The four important key features of photo are: time, location, person and event. Contextual information is used to help annotation on mobile phone. Three kinds of context are integrated: photo captured context, personal context and social network context.

Temporal Context Date, Time , Artist	Spatial Context Location Name GPS Co-ordinates
Social Context SNS relationships User name	Personal Context Calendar, Email Contacts

Fig 1: Photo Context [3]

4.1 Photo Captured Context

While taking a photo, the time gets recorded in camera. The time is documented in the photo metadata (EXIF: Exchangeable image file format). Similarly if the mobile device is prepared with a GPS receiver or other positioning system, it becomes location-aware. By using this GPS module, the GPS coordinates of the location where the photo is taken can also be recorded in EXIF of photo. In this way, temporal and spatial context from photo metadata can be achieved. Along with this data, there is also a field in EXIF called “artist” which will store data about photographer. So the photographer can be recorded on the camera software.

4.2 Personal Context

Being personal device, additional user context information from mobile device can be used for annotation purpose. A common service in mobile device is calendar that user utilize in making their timetable. So, when a picture is taken at some time which is within a planned event, the picture is possibly associated to the event. Similarly, more contexts from email and contacts can be taken. One of the best semantic contents of a photo is event. If the event can be acquired automatically, the annotation can become easier. So it is logical to use the scheduled event as an implication of event annotation. In Image retrieval, annotation of person in a picture is very helpful. When it is impossible to automatically identify person in the photo, it needs to be manually annotated. The candidates for annotations can be the contacts stored in the mobile device.

4.3 Social Network Context

As one of the important SNS (Social networking service) activities is photo sharing, Social context can be used for annotation. In social context based photo sharing system, the friendship between users is an implied annotation of image. This annotation needs to be added on the server side. Social relationship is very essential for a photo sharing system. The ways in which friends can be added in system:

- 1) Adding friend by Name: The request can be sent to someone known to be friend by giving his user name.
- 2) Adding friend by Photo: While browsing photos uploaded by other people, a friend request can be sent to make friendship with owner.
- 3) Importing friend from Email Contacts: Contacts of email can be retrieved with help of email service providers. If user wants to import friends, by inputting email and password, the system acquires contacts from email. Then the system evaluates these contacts with users in system, and recommends users who are the user’s contacts. The user will select them to send friend requests.

4.4 Annotation Process

There is limitation of input on mobile device so the functions of annotation process are compact. The annotation algorithm needs to be as simple as possible due to the restriction computing ability of mobile phone. To give annotations to an image, following are the steps:

- 1) Metadata Analysis: The metadata can be acquired from the image (EXIF), including time, GPS and artist fields.
- 2) Getting Personal Context: The time and GPS information is used to analyze valuable information in personal utilities such as calendar, contacts and email, and the produced results are presented as annotation implications;
- 3) Tagging: Images are annotated based on these suggestions and add other tags manually if required (such as emotion).
- 4) Uploading: The photo and the annotations are uploaded to the server. The server accumulates the photo and the annotations and generates multidimensional indices for the photos.

When a photo is selected to upload, the system reads date/time, GPS and photographer (artist) from the EXIF segment of the photo file. The event which is programmed at that time in the calendar through some API (application programming interface) is searched. The search results are listed as event annotation suggestions.

After uploading photo, the server obtains the GPS and other metadata and annotates the upload time and the owner of the photo. Server accumulates the photo and creates thumbnails for the photo. Along with this, it also generates multidimensional indices for the photo. These indices are used to retrieve the image.

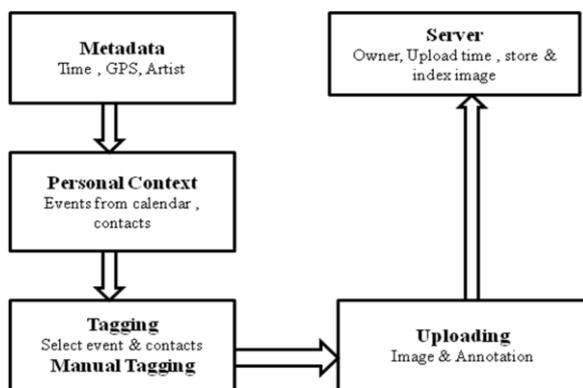


Fig 2: Annotation Process

- 1) Time Index
- 2) User Index
- 3) Emotion Index
- 4) Location Index
- 5) Relevance Index
- 6) Event Index

5. CONCLUSION

Manual annotation is a costly and time consuming work, especially for mobile device. We have discussed methods of automatic & semi-automatic image annotation. Semi-automatic performs better than other annotation techniques in terms of accuracy as user is participating in annotation process. The semi-automatic way to add annotation to image by using contextual information of the mobile device is discussed here. Machine learning mechanism if integrated into annotation

system can make the annotation method more intellectual and precise.

6. REFERENCES

- [1] Wei Liu , Xue Li, Daoli Huang “A Survey on Context Awareness” Computer Science and Service System (CSSS), International Conference on,29 June 2011 IEEE
- [2] B. Shevade, H. Sundaram, L. Xie. “Modeling Personal and Social Network Context for Event Annotation in Images”. In JCDL 2007, ACM Press (2007).
- [3] Shuangrong Xia, Xiangyang Gong, Wendong Wang, Ye Tia “Context-Aware Image Annotation and Retrieval on Mobile Device” 2010 IEEE.
- [4] L. Cao, J. Luo, H. Kautz, T. S. Huang. “Image Annotation within the Context of Personal Photo Collections Using Hierarchical Event and Scene Models”. In IEEE Multimedia 2009 11(2), 208- 219.
- [5] W. Viana, J. B. Filho, J. Gensel, M. Villanova-Oliver, H. Martin. PhotoMap: From location and time to context-aware photo Annotations. In Journal of Location Based Services 2008 2(3), 211-235
- [6] M. Ames, M. Naaman. “Why We Tag: Motivations for Annotation”. In proc. CHI 2007, ACM Press (2007), 971-980.
- [7] U. WESTERMANN and R. JAIN. Toward a Common Event Model for Multimedia Applications. In IEEE Multimedia 2007 14(1), 19-29
- [8] M. Davis, N. V. House, J. Towle, S. King, S. Ahern, C. Burgener, Perkel, M. Finn, V.Viswanathan, M. Rothenberg. “MMM2: Mobile Media Metadata for Media Sharing”, Ext. Abstracts CHI 2005, ACM Press (2005), 1335-1338
- [9] asullah Khalid Alham, Maozhen Li1, Suhel Hammoud and Hao Qi “Evaluating Machine Learning Techniques for Automatic Image Annotations” 2009 IEEE
- [10] O. Marques, N. Barman, Semi-Automatic Semantic Annotation of Images Using Machine Learning Techniques, Proc. of ISWC, pp. 550-565, 2003
- [11] J. Vompras, S. Conrad, A Semi-Automated Framework for Supporting Semantic Image Annotation, Proc. of ISWC, pp.105-109, 2005.
- [12] L. Wenyin, S. Dumais, Y. Sun, H. Zhang, M. Czerwinski and B. Field, Semi-Automatic Image Annotation, Proc. Of INTERACT, pp.326-333, 2001.
- [13] A. Yavlinsky, E. Schofield, S. M. Ruger, Automated Image Annotation using Global Features and Robust Nonparametric Density Estimation, Proc. of CIVR, pp. 507-517, 2005
- [14] Jeon, V. Lavrenko, and R. Manmatha, Automatic Image Annotation and Retrieval using Cross-Med Relevance Models, Proc. ACM SIGIR, pp. 119-126, 2003.
- [15] R. Jin, J. Chai, L. Si, Effective Automatic Image Annotation via a Coherent Language Model and Active Learning, Proc. of ACM Conference on Multimedia, pp. 892-899, 2004
- [16] C. Wang, F. Jing, L. Zhang, H. Zhang, Content-Base Image Annotation Refinement, Proc. of CVPR, 2007.