A Novel System for the Semi Automatic Annotation of Event Images

Philip J. McParlane[*]
School of Computing Science
University of Glasgow (UK)
p.mcparlane.1@research.gla.ac.uk

Joemon M. Jose
School of Computing Science
University of Glasgow (UK)
Joemon.Jose@glasgow.ac.uk

ABSTRACT

With the rise in popularity of smart phones, taking and sharing photographs has never been more openly accessible. Further, photo sharing websites, such as Flickr, have made the distribution of photographs easy, resulting in an increase of visual content uploaded online. Due to the laborious nature of annotating images, however, a large percentage of these images are unannotated making their organisation and retrieval difficult. Therefore, there has been a recent research focus on the automatic and semi-automatic process of annotating these images. Despite the progress made in this field, however, annotating images automatically based on their visual appearance often results in unsatisfactory suggestions and as a result these models have not been adopted in photo sharing websites. Many methods have therefore looked to exploit new sources of evidence for annotation purposes, such as image context for example. In this demonstration, we instead explore the scenario of annotating images taken at a large scale events where evidences can be extracted from a wealth of online textual resources. Specifically, we present a novel tag recommendation system for images taken at a popular music festival which allows the user to select relevant tags from related Tweets and Wikipedia content, thus reducing the workload involved in the annotation process.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing

Keywords

Twitter; Wikipedia; Photo tag recommendation

1. INTRODUCTION

Taking and sharing images is cheaper, easier and more accessible than ever; with the advancement of smart phone camera technology, photographers no longer need expensive equipment. This change has increased the amount of visual content uploaded to image sharing websites such as Flickr[1]. Of this content, an ever increasing number of users are uploading photographs taken at large social (e.g. London 2012 Olympics) and world (e.g. Philippines Typhoon) events, where the user acts the role of the amateur photo journalist. Organising these images is difficult, however, as a result of the semantic gap [3] and lack of annotations provided by users [4]; an entire field of work has focused on the automatic annotation of images [3, 2]. Despite the progress made in the last two decades, due to the presence of the semantic gap [3], fully automatic methods still perform lower than what is required for industry. Therefore, real life applications have instead adopted semi-automatic tag recommendation approaches, allowing users to annotate their images from a list of suggested tags. Aside from Flickr's recommendation approach, there have been many photo tag recommendation methods proposed in recent years [4, 1].

These recommendation approaches suggest tags based on *historical* Flickr data, however, which is often sparse, out-of-date (due to the *time lag problem*, where users upload images long after they are actually taken) and lacking in coverage. Aside from recommending based on historical Flickr data, there now exists extensive, up-to-date textual content related to a wide spectrum of events (e.g. Twitter, Wikipedia) which can also be exploited for tag recommendation purposes; despite this, no existing photo annotation approaches have considered these streams for this purpose. In this demo, we propose a photo tag recommendation system, designed for the amateur photographer, which offers automatic tag suggestions alongside novel annotation strategies based on related Tweets and Wikipedia data for images taken at the Austin City Limit 2012 music festival, which aim to reduce the effort and time required in comparison to existing image annotation approaches.

2. DEMONSTRATION FEATURES

The web interface (build using HTML5, Javascript and PHP) presents the user with an image, the tags assigned by the user (e.g. `acl`, `aclfest`, `austincitylimits`) and various
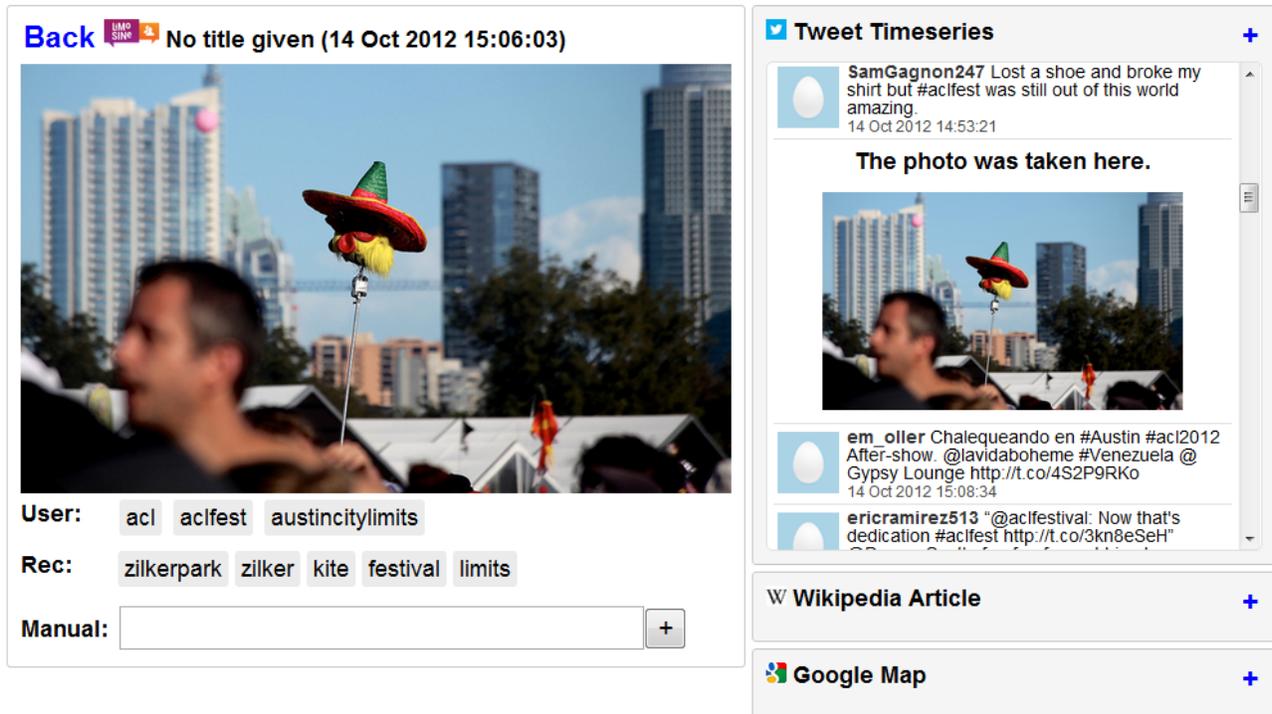
[1]http://www.flickr.com

Figure 1: Semi automatic image annotation interface exploiting social media and Wikipedia data

annotation strategies. In total, the user is able to use four different tagging approaches, as described in the following sections:

1. **Manual Tagging:** Firstly, the user can add tags manually using the text box displayed underneath the image.
2. **Tag Recommendations:** Secondly, the user is offered tag recommendations based on the tf-idf model described in [1]. This model computes suggestions based on those tags already added by the user based on a tag co-occurrence matrix built on 1M Flickr images. Users are able to click-and-drag any relevant tags in order to annotate the given image.
3. **Related Tweets:** The user is also presented with a Twitter feed containing tweets tagged with the hash related to the event (i.e. #aclfest). Further, the tweets are displayed with temporal relevance to the given image i.e. the image is placed within the stream at the correct chronological position with respect to the time it was taken. Therefore, the user can browse potentially relevant tweets posted at the event around the same time. From this feed, the user is able to *double-click* any term within any tweet in order to quickly and easily annotate the image without any keyboard input.
4. **Related Wikipedia:** The user is also presented with the Wikipedia article relevant to the event. As before, the user is also able to *double-click* any term within the document in order to use it as an annotation for the image.

By immersing the user within a context of related Tweets and Wikipedia data, they are able to consider this evidence in the image annotation process. Further, by offering simple

double-click and click-and-drag interactions as a means of annotating an image, the laborious workload associated with traditional tagging approaches is reduced.

3. CONCLUSION AND FUTURE WORK

In this demonstration we presented a novel system, offering four tagging strategies, for the annotation of images taken at large scale event, which aim to streamline the traditionally laborious process. In comparison to existing manual annotation and tag recommendation methods, our interface offers a wider evidence scope which draws temporally significant suggestions from users in real time (in the form of Tweets) as well as structured encyclopaedia evidences (in the form of Wikipedia) which complement traditional methods. As future work we propose to exploit the geographical location of images by suggesting tags from images with a similar location.

4. REFERENCES

[1] N. Garg and I. Weber. Personalized, interactive tag recommendation for flickr. In *ACM RecSys 2008*.
[2] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS 2012*.
[3] A. Makadia, V. Pavlovic, and S. Kumar. Baselines for image annotation. *IJCV 2010*.
[4] B. Sigurbjörnsson and R. van Zwol. Flickr tag recommendation based on collective knowledge. In *WWW 2008*.