Engineering Applications of Artificial Intelligence 51 (2016) 71–91

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

A survey on Flickr multimedia research challenges

Evaggelos Spyrou [a,*], Phivos Mylonas [b]

[a] Computational Intelligence Laboratory (CIL), Institute of Informatics and Telecommunications, National Center for Scientific Research Demokritos, Athens, Greece

[b] Department of Informatics Ionian University, Corfu, Greece

ARTICLE INFO

Available online 15 February 2016

Keywords:
Flickr
Retrieval
Recommendations
Localization
Travel applications
Knowledge extraction

ABSTRACT

Multimedia content sharing within social networks has become one of the most interesting and trending research fields over the last few years. This undoubted emerge of related research works is rather twofold, namely it includes both the analysis and management techniques of the content itself, as well as new ways for its accompanied meaningful interpretation and exploitation. In this paper, we review the recent advances in the above fields in the humanistic framework of the popular Flickr social network. In addition, the major research challenges in the area are demonstrated and discussed, which include current state-of-the-art approaches with respect to interesting humanistic data collection and interpretation research fields, such as multimedia information retrieval, (semi-) automatic tag manipulation, travel applications, semantic knowledge extraction, human activity tracking, as well as related benchmarking efforts. At the end of this survey, we also discuss the main challenges and propose a number of future research directions for interested fellow researchers to continue investigation in the field.

1. Introduction

We all agree that our digital era is characterized and rather dominated by a single, yet very important observation, namely *extremely large amounts of digital multimedia content are being produced everyday and almost instantly shared online by people interacting with others*, within the framework of the social network of their preference. This online social media networking explosion has gone to unprecedented lengths, with people redefining their lives with the aid of social media characteristics. In principle, a social network is considered to be a digital, on-line place, where people create profiles and build a personal network that connects them to other people. During recent years, such networks have emerged into a phenomenon that engages tens of millions of Internet users every day. Moreover, the number of people that use the Internet to share their own generated multimedia content has been continuously increasing. More specifically, 73% of online users having a social networking profile, compared to 8% in 2005, 16% in 2006, and 37% in 2008, respectively (Lenhart et al., 2010).

Flickr[1] is an image and video hosting website created by a Vancouver-based company named "Ludicorp" back in 2004 and currently owned by "Yahoo!", an American multinational Internet corporation headquartered in Sunnyvale, California, USA.[2] What makes it special among other multimedia sharing websites is its aspect as an online community, within which users are able to interact by sharing comments about photography and create groups of particular interests. The technology news and media network "The Verge"[3] reported in March 2013 that Flickr had a total of 87 million registered members and 3.5–10 million new photos uploaded daily (The Verge, 2013; Jordan, 2015). Each photo may contain metadata added by its photographer, such as tags that describe either its visual content, its geolocation, or a free text description somehow related to the photo contents (Fig. 1). It also contains metadata added by the camera that has been used, such as the actual date the photo was taken, specific camera settings, and the camera model. Few GPS-enhanced cameras automatically geo-tag the photos they take, but in principal this is still done manually by photographers themselves, after the actual photo has been captured. In principle, textual metadata associated to a photo often serve as a reminder of the context of the image for the photographer and his social circle (Nov et al., 2010; Van House, 2007).

* Corresponding author. Office address: NCSR Demokritos Institute of Informatics and Telecommunications Patriarchou Grigoriou and Neapoleos St. GR-15310, Aghia Paraskevi, Attiki Greece. Tel.: +302106503175.
E-mail addresses: espyrou@iit.demokritos.gr (E. Spyrou), fmylonas@ionio.gr (P. Mylonas).

http://dx.doi.org/10.1016/j.engappai.2016.01.006
0952-1976/

[1] http://www.flickr.com
[2] http://www.yahoo.com
[3] http://www.theverge.com

Fig. 1. Flickr images sample.

The biggest majority of such photos uploaded and shared to the aforementioned popular social network is taken by common users or amateur photographers. The fact that such personal multimedia content has become easy to grasp and capture by rather cheap hardware aided to its mass expansion. In addition, as Van House (2007) empirically identified, this content is being used according to four basic humanistic axes, namely (a) memory, narrative and identity; (b) relationships; (c) self-representation; and (d) self-expression. Eventually, other types of problems arose as well, i.e., it is nowadays acknowledged that user-generated multimedia content is difficult to efficiently get access to, or to be processed and effectively manipulated by humans within a meaningful amount of time and effort spent. As a result tasks have emerged that become day-by-day very difficult and challenging to tackle.

Up to now it is common research knowledge that the overwhelming distribution of such dynamically generated humanistic content over its end users, online communities and consumption devices requires ways for efficient representation and, more importantly, *organization*, in order to be exploited in applications and services. In another empirical study, Zeng and Wei (2013) investigated potential relationships between social ties and similarities of the type of digital content that people create and share online. Among their discoveries they found that around the time that a social tie between two individuals is formatted, they begin to create more similar content compared to what they have created before. Interestingly, this similarity tends to evolve in different ways when observing different subgroups of user pairs. In the framework of the latter observation, traditional analysis approaches focusing on analyzing data in terms of objects, and/or concepts and other isolated entities are often quite insufficient, since they do not take into account important properties and relations of online shared multimedia content, the so-called "content metadata". Flickr enables the latter by providing a full set of metadata managing options for its uploaded photos, allowing its users to edit and refine their photo content metadata.

Although in general the application of qualitative and quantitative multimedia content analysis techniques to assess generic metadata records goes back in time and does not advance significantly current research state-of-the-art, it is the nature of today's "instant content capture and sharing" conditions that point out a special interesting role in the process that makes the difference, i.e., the role of a special kind of photo metadata: *geo-tags*. In the dynamic environment of a social network, human behavior and activities are better described and exploited in terms of enriched content metadata. Consequently, geo-tags are considered important for online multimedia analysis and annotation. It is worth noting at this point that the roots of this annotation process lie within the analog photo era, where users wrote some "metadata" information, such as place and date, behind paper photos!

Now, in principle every part of a photo may be tied to a geographic location, but in most typical applications only the position of the photographer is associated with the entire digital photo. As the reader may imagine, this small detail implicates and significantly burdens most humanistic data search and retrieval tasks. In the most typical example, photos of a landmark may have been taken from very different positions apart and in order to identify all photos of this particular landmark within an image database, all photos taken within a reasonable circular distance from it must be considered. Now, when such geo-tagged photos are uploaded to online multimedia content sharing communities like Flickr, which enables the construction of infinite connections among its users (Van Dijck, 2011), a photo can be placed onto a map to view the location where the photo was taken. In this way, social network users can browse photos directly from the map, search for photos of a given area, and find related photos of the same place from other users. The aforementioned tasks are considered elementary in order to build additional, ad hoc value-added digital services on top, like automated route/trip planning or like, to our most recent knowledge, the popular "NOW" app. The latter uses geo-tagged photos to find nearby events happening now.[4]

As expected, the act of automatically providing or calculating meaningful photo geo-tags (the so-called "geo-tagging" process) opens a huge research topic for the researchers' community, mainly to the directions of being able to analyze them, identify and determine social patterns amongst them. However, issues of credibility on the volunteered user-generated geo-tagging should become of broader research interest in various areas (Flanagin and Metzger, 2008; Spielman, 2014), motivating us to further investigate this topic in the following. At this point, it should be noted that our work clearly differentiates from existing surveys in the field, e.g., the one of Luo et al. (2011) and the one of Wu et al. (2013). The former focused on geo-tagged content in general (i.e., originating not only from Flickr but even from other sources) and research within the fields of multimedia and computer vision. Moreover, the latter focused on the social aspect of social media research in general. On the other hand, our survey paper deals with a broader aspect of research using Flickr derived datasets and taking into account the special characteristics of Flickr within the process.

The structure of the rest of our paper is as follows. In Section 2 we explain in more detail the motivation behind selecting Flickr as the social network under investigation, as well as provide some basic aspects of its functionalities, while in Sections 3–7 we provide the main details on the research opportunities that exist in the research fields identified. More specifically we start by presenting several approaches in the multimedia content retrieval field (Section 3), as well as approaches that focus on the interpretation of multimedia content (Section 4), both in the sense of automatic tag/geo-tag generation (Section 4.1) and of knowledge extraction from Flickr metadata (Section 4.2). Section 5 focuses on the humanistic aspect of related applications in a twofold approach, namely by tackling touristic-oriented travel applications (Section 5.1), and applications dealing with human activity tracking (Section 5.2). Additional

[4] http://techcrunch.com/2013/01/11/now-app/