Information Sciences 328 (2016) 270–287

Contents lists available at ScienceDirect

Information Sciences

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

Multimodal latent topic analysis for image collection summarization

Jorge E. Camargo [a,*], Fabio A. González [b]

[a] LACSER Research Group, Facultad de Ingeniería de Sistemas, Universidad Antonio Nariño, Calle 22 Sur No. 12D-81 Bloque 2 Piso 6, Bogotá, Colombia
[b] MindLab Research Group, Universidad Nacional de Colombia, Bogotá, Colombia

ARTICLE INFO

Article history:
Received 28 January 2015
Revised 18 August 2015
Accepted 21 August 2015
Available online 4 September 2015

Keywords:
Multimodal image collection summarization
NMF
Latent topic analysis
Multimodal analysis

ABSTRACT

This paper presents a multimodal latent topic analysis method for the construction of image collection summaries. The method automatically selects a set of prototypical images from a large set of retrieved images for a given query. We define an image collection summary as a subset of images from a collection, which is visually and semantically representative. To build such a summary we propose MICS (Multimodal Image Collection Summarization), a method that combines textual and visual modalities in a common latent space, which allows to find a subset of images from which the whole collection can be reconstructed. Experiments were conducted on two collections of tagged images demonstrating the ability of the approach to build summaries with representative visual and semantic contents. The method was evaluated using objective measures, reconstruction error and diversity of the summary, showing competitive results when compared to other summarization approaches.

1. Introduction

The large amount of images produced every day requires systems able to efficiently and effectively manage them. Photo-sharing systems, such as Flickr[1], pose important challenges to organize, browse and query large image collections. The typical scenario to search images within Flickr consists of providing a query by means of keywords, which the system processes to return a set of images with tags matching the query. Although this paradigm has been satisfactorily used in search engines for searching textual content, it is not necessarily the most suitable way to interact with large image collections [24]. One of the problems with this approach is that, in general, textual queries are not enough to express the visual richness of images and therefore the most relevant images are not necessarily at the top of the search results. Typically a user explores only the first page of results, so he/she will not see other images that could be of interest in subsequent pages. Fig. 1 shows the top 24 images returned by Flickr for the query *apple*. Note that the returned images have some relation with the *apple* term since the associated tags contain such term. However this term can be used to describe different semantic concepts such as fruit, computers, food, cake, etc. The returned images in this example are not representative (iconic) of the complete set of results, so the user only has access to a small portion of them in a first view, and some relevant images may be missed by the user. This traditional navigation method is the common behavior of image search systems such as Flickr, Google Similar Images, Yahoo Images, and Bing. The image collection summarization algorithm proposed in this paper can be applied to any of these systems.

* Corresponding author. Tel.: +5712781430.
E-mail address: jorgecamargo@uan.edu.co, jecamargom@unal.edu.co, camargoj@gmail.com (J.E. Camargo).

[1] http://www.flickr.com.

http://dx.doi.org/10.1016/j.ins.2015.08.044
0020-0255/

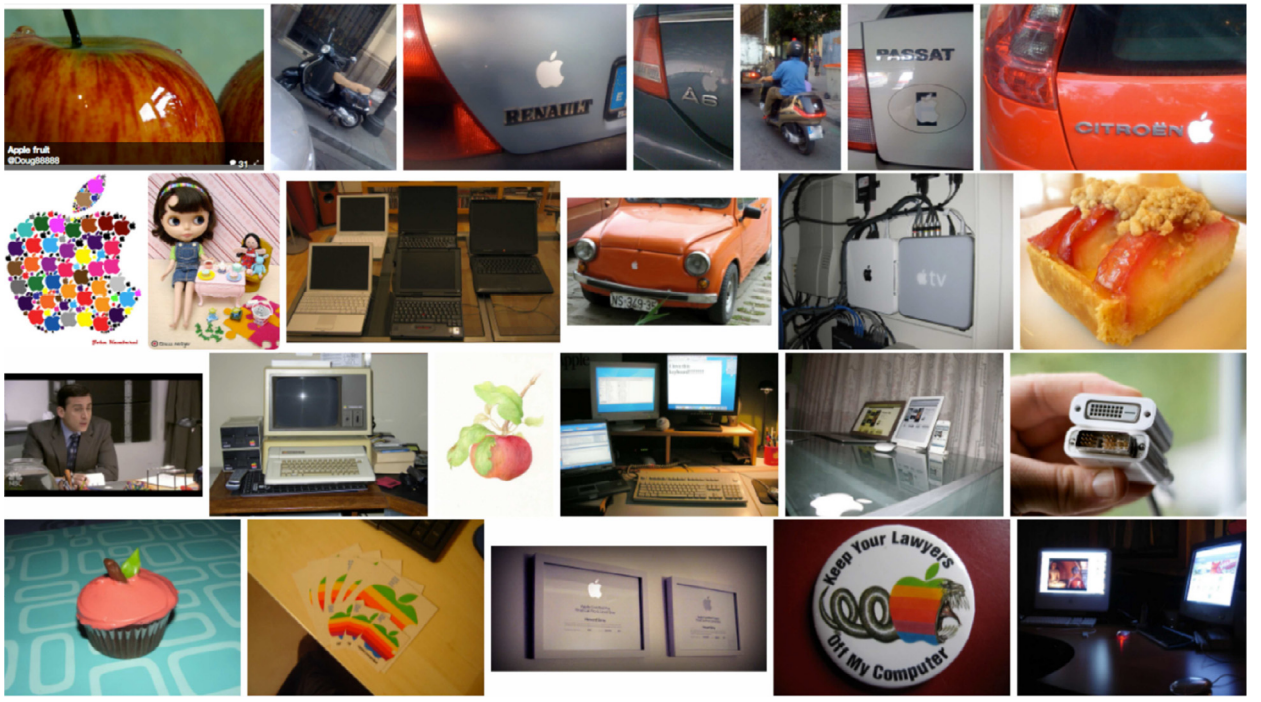

Fig. 1. The first 24 images retrieved by Flickr for the query *apple* (retrieved on September 1, 2014). (For interpretation of the references to color in this figure, the reader is referred to the web version of this article).

Automatic image collection summarization is the process of selecting images which are good representatives of a larger set. This process is fundamental to enable interactive navigation and exploration of large-scale image collections [22,37].

This paper proposes a new method to automatically build *multimodal image collection summaries* in which both textual and visual contents are combined in the same latent semantic space. This method is based on Convex Non-Negative Matrix Factorization (Convex-NMF), which is an algorithm that allows to generate a latent space in which visual and textual contents can be simultaneously analyzed to generate a multimodal image collection summary.

Most of the proposed summarization methods found in the literature extract visual features such as color, texture and edges to represent the image content, and then apply a clustering algorithm to perform the summarization process [18,32,35]. However, images are commonly accompanied by other information such as text, audio, links, etc. This information complements the visual content of the image providing additional hints to characterize the image semantics. We define a *modality* as one of these additional information sources; therefore, the concept *multimodal* refers to the use of multiple information sources to represent the information content of an image. The proposed method also provides a mechanism to project images that do not have associated text, which addresses the problem of images that are not accessible because of the lack of textual information associated with them. Notwithstanding that user satisfaction studies are widely used to evaluate summarization algorithms, we favored more objective and quantitative evaluation metrics for assessing the performance of the proposed summarization method. Consequently, the paper also presents a method to quantitatively measure the quality of an image collection summary by estimating its ability to reconstruct the multimodal collection content as well as its ability to represent the semantic diversity of the collection.

This paper is organized as follows: Section 2 presents introductory concepts of image collection summarization; Section 3 describes related work; Section 4 presents the proposed method; Section 5 presents experimental evaluation of the proposed strategy; Section 6 presents a discussion of the proposed method and the obtained results; and finally, Section 7 concludes the paper.

2. Image collection summarization

Image collection summarization is the process that aims at selecting a small set of the most representative images (summary) to faithfully represent the information content of a larger set [8]. This process has become very important to enable interactive navigation and exploration of large-scale image collections [22]. Many applications can benefit from results of image collection summarization: (1) image search engines such as Google Similar Images, Flickr and Yahoo, which use the conventional page-based navigation paradigm; (2) on-line shopping web sites [47], which visualize representative images for each product category; and (3) personal photo collection systems, which automatically organize image categories based on metadata information. Such