The Social Picture

Sebastiano Battiato*, Giovanni M. Farinella*, Filippo L.M. Milotta*, Alessandro Ortis*,
Luca Addesso†, Antonino Casella†, Valeria D'Amico†, Giovanni Torrisi†
*Universitá degli Studi di Catania, Viale A. Doria 6, 95125 - Catania, Italy
Email: {battiato, gfarinella, milotta, ortis}@dmi.unict.it
†Telecom Italia, Viale A. Doria 6, 95125 - Catania, Italy
Email: {luca.addesso, antonino.casella, valeria1.damico, giovanni.torrisi}@telecomitalia.it

***Abstract*—We present *The Social Picture*, a framework to collect and explore huge amount of crowdsourced social images about public events, cultural heritage sites and other customized private events. The collections can be explored through a number of advanced Computer Vision and Machine Learning algorithms, able to capture the visual content of images in order to organize them in a semantic way. The interfaces of *The Social Picture* allow the users to create customized collections by exploiting semantic filters based on visual features, social network tags, geolocation, and other information related to the images.**

Keywords–Social Media; Crowdsourcing; Multimedia; Image Collections; Image Understanding.

I. INTRODUCTION

Nowadays, the diffusion of social networks plays a crucial role in collecting information about people opinion and trends. In social events (e.g., concerts), the audience typically produces and share a lot of multimedia data with mobile devices (e.g., images, videos, geolocation, tags, etc.) related to what has captured their interest. The redundancy in these data can be exploited to infer social information about the attitude of the attending people by means of Machine Learning (ML) and Computer Vision (CV) algorithms. In [1] we introduced a framework called *The Social Picture* (TSP) to collect, analyze and organize huge flows of visual data, and to allow users the navigation of image collections generated by the community.

In this demo, we present some additional features that extend the work done in [1]: design of a new t-SNE (t-distributed Stochastic Neighbor Embedding) exploration tool suitable for very large collections, 3D reconstruction of heritage sites by means of the attending people's photos, more advanced statistics provided to event organizers, creation of private events collections and temporal extension of collection analysis. The rest of this paper is organized as follows. Section II describes the aims and the features of the framework presented in [1]. Section III describes some issues related to the first prototype of the framework, and presents the implemented improvements, as well as the new developed features. The acknowledgement closes the article.

II. OVERVIEW

TSP is a social framework populated by images uploaded by users or collected from other social media (Figure 1). Anyone registered to TSP can become an event manager and start a social collection accordingly to the "prosumer" paradigm, where the users are both producers and consumers of a service. Indeed, each collection has two kind of users: the event organizer and the event participant. Imagine an art-gallery manager who leases a famous Picasso's painting with

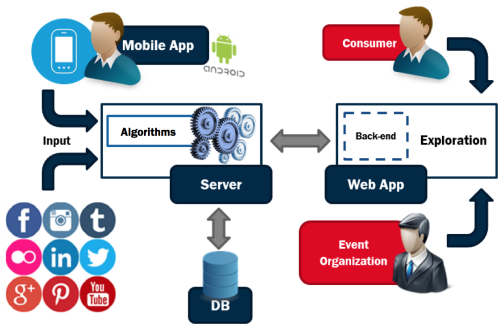

Figure 1. The Social Picture's architecture.

the aim to include it in a event exhibition, together with other famous and expensive artworks. How does he know he did a good investment? Which was the more attractive artwork? The collection of the uploaded images of an event, gives the sources analysed in TSP to answer the aforementioned questions. The obtained information can be then exploited by the event organizers for the event evaluation and further planning. These information could be exploited, for example, to perform aimed investments. The system can suggest what is the better subject to use for the advertising campaign of the event, or which of the attractions it worth to mainly reproduce in the souvenir shop products, to support merchandising strategies. Feedback about what is the most interesting part (i.e., the most photo captured) of a landmark building can help on taking decisions about renovating some parts rather than others as first investment. Users can add an image to a collection by using either a mobile application and a website interface. Furthermore, an event collection can be populated by selecting images from the most common social networks for images (e.g., Flickr, Panoramio, Instagram). Once an image is uploaded, it is analysed by a set of CV and ML algorithms. The web interface exhibits a range of filtering tools to better explore the huge amount of data. When an event manager creates a new collection, he is allowed to specify several options to customize the image gathering, the social analysis to be performed, and the visualization tools to be shown for the users of that collection. The event manager is also allowed to set a range of statistics, which will be available after the analysis of the collected images. The several exploration tools are based on both visual and textual information, such as EXIF (Exchangeable Image File Format) data and a number of ad-hoc extracted visual features. The visual analysis module of the system feds all the images into two different CNNs (Convolutional Neural Networks), *AlexNet* [2] and *Places205-*

 ISBN: 978-1-61208-534-0

AlexNet [3], in order to extract the classification labels and image representations. Furthermore, the system is able to distinguish pictures depicting food, and pictures captured in indoor/outdoor environments by exploiting visual features. The system provides different exploration tools, detailed in the followings. A demonstrative video is available at the following link: http://iplab.dmi.unict.it/TSP

A. Advanced Tools in The Social Picture

Among the tools included in TSP, there is the one useful to generate automatic subsets of images from a specific photo collection. This tool allows the user to set the number of images to obtain as output for a collection in TSP, and automatically generates the subset of images taking into account visual features as well as EXIF information related to the images composing the photo collection. In this way, the user is provided by a number of representative image prototypes related to the collection, which can be used for different purposes (e.g., printing the most significative pictures of paintings of a museum for a specific social group). For this tool, CNN representation used in [3] is employed.

In [1], we exploited the *fc7* feature extracted with the *AlexNet* architecture [2] for each image and exploited the t-SNE embedding algorithm [4] to compute a 2D embedding that respects the pairwise distances between visual features.

The landmark heatmap is a visualization tool used to depict the intensity of images at spatial points. The heatmap consists of a colored overlay applied to the original image of a specific landmark building or area of interest. Areas of higher intensity will be colored red, and areas of lower intensity will appear blue. The intensity of the heatmap is related to the number of collected pictures that contain that visual area. By clicking on a point of the heatmap, the user can retrieve and visualize the images that contribuited to generate the map intensity at that point.

Finally, the automatic image captioning, as described in [5], is another feature included in TSP. With the aim to help the user to include a description to an uploaded image, *The Social Picture* automatically generates and suggests a description to the user that can then refine it. The descriptions of images can be used for text based query performed by the user.

III. Platform Improvements

A. Hierarchical t-SNE

The first implementation of the t-SNE exploration tool in TSP was unable to scale with the number of the collections' images. The new tool presented in this demo implements an hierarchical version of the t-SNE embedding which allows to explore picture collections without limits on the amount of processed pictures. This helps the user to better explore the image distribution in a custom level of detail. Furthermore, the user can choose a subset of images and compute the t-SNE embedding of them directly on the browser. As the number of pictures of a collection is unpredictable, the computation of the t-SNE coordinates could be very expensive. Besides the t-SNE computation, which needs to be executed only one time per dataset, a huge number of pictures can affect the browser efficiency for the visualization of the 2D embedding. We organize the entire collection of pictures in a hierarchical structure. After the collection is analysed (i.e., the *fc7* features have been computed for all the images) the system performs a hierarchical k-means clustering of the image features. The algorithm divides the dataset recursively into k clusters, for each computation the k centroids are used as elements of a *k-tree* and removed from the set. When this new version of the t-SNE tool (hierarchical t-SNE) is executed, it shows to the user the t-SNE embedding computed only for the elements in the root of the *k-tree* (i.e., the picture centroids of the first *k-means* computation). When the user selects one of these pictures, the system computes the t-SNE of the pictures included in the child node corresponding to the selected picture element. This hierarchical exploration can be continued by selecting one of the shown pictures and computing the t-SNE embedding for its sub-elements in the hierarchy.

B. 3D Reconstruction

Starting from VSFM (Visual Structure From Motion) [6], we are able to compute a 3D sparse reconstruction of large photos collections. The models are augmented with colors for vertices, related to the frequency of been acquired in a photo, colors for cameras, related to the number of visual features acquired by each photo, and with a plane which show the spatial density of contributing users. We embedded in TSP the models through a 3D web viewer based on Threejs, allowing the users to browse the 3D sparse reconstructed models gaining a cue about what are the points of view and the subjects preferred by users when take photos. Moreover, the models in the 3D web viewer can also be browsed through Leap Motion system, an intuitive and fast interactive system.

C. Private Events

Private collections can be created in TSP: the private collections con be accessed only by the owner and the invited users. Owner invites users to contribute adding new photos to the collection, while users receive the invitation through an e-mail.

D. Temporal Update

We developed a temporal uddate for collection: owners of collections can launch collection update request to server. It is possible to check if new photos have been added since at most 1 year from the time of creation or last update of the collection. Once the window for update is set, server quiries social networks for new photos and add them to the collection.

Acknowledgments

The authors would like to thank Marco Cavallo and Michele Bellocchi for helping in the implementation of the web framework.

References

[1] S. Battiato et al., "The social picture," in Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval. ACM, 2016, pp. 397–400.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097–1105.

[3] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in Advances in Neural Information Processing Systems, 2014, pp. 487–495.

[4] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne," Journal of Machine Learning Research, vol. 9, no. 2579-2605, 2008, p. 85.

[5] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignment for generating image descriptions," in Computer Vision and Pattern Recognition, 2015.

[6] C. Wu, "Towards linear-time incremental structure from motion," in 3D Vision-3DV 2013, 2013 International Conference on. IEEE, 2013, pp. 127–134.

 ISBN: 978-1-61208-534-0