Scenemash: Multimodal Route Summarization for City Exploration

van den Berg, J.; Rudinac, S.; Worring, M.

DOI
10.1007/978-3-319-30671-1_75

Publication date
2016

Document Version
Final published version

Published in
Advances in Information Retrieval

License
Article 25fa Dutch Copyright Act

Link to publication

Citation for published version (APA):
van den Berg, J., Rudinac, S., & Worring, M. (2016). Scenemash: Multimodal Route Summarization for City Exploration. In N. Ferro, F. Crestani, M-F. Moens, J. Mothe, F. Silvestri, G. M. Di Nunzio, C. Hauff, & G. Silvello (Eds.), *Advances in Information Retrieval: 38th European Conference on IR Research, ECIR 2016, Padua, Italy, March 20–23, 2016 : proceedings* (pp. 833-836). (Lecture Notes in Computer Science; Vol. 9626). Springer. https://doi.org/10.1007/978-3-319-30671-1_75

Scenemash: Multimodal Route Summarization for City Exploration

Jorrit van den Berg[1](✉), Stevan Rudinac[2](✉), and Marcel Worring[2](✉)

[1] TNO, Den Haag, The Netherlands
`jorrit.vandenberg@tno.nl`
[2] Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands
`{s.rudinac,m.worring}@uva.nl`

Abstract. The potential of mining tourist information from social multimedia data gives rise to new applications offering much richer impressions of the city. In this paper we propose Scenemash, a system that generates multimodal summaries of multiple alternative routes between locations in a city. To get insight into the geographic areas on the route, we collect a dataset of community-contributed images and their associated annotations from Foursquare and Flickr. We identify images and terms representative of a geographic area by jointly analysing distributions of a large number of semantic concepts detected in the visual content and latent topics extracted from associated text. Scenemash prototype is implemented as an Android app for smartphones and smartwatches.

1 Introduction

When visiting a city, tourists often have to rely on travel guides to get information about interesting places in their vicinity or between two locations. Existing crowdsourced tourist websites, such as TripAdvisor primarily focus on providing point of interest (POI) reviews. The available data on social media platforms allows for new use-cases, stemming from a much richer impression about places. Efforts to utilize richness of social media for tourism applications have been made by e.g., extracting user demographics from visual content of the images [3], modelling POIs and user mobility patterns by analysing Wikipedia pages and image metadata [2] or by representing users and venues by topic modelling in both text and visual domains [7].

We propose Scenemash[1], a system that supports way-finding for tourists by automatically generating multimodal summaries of several alternative routes between locations in a city and describing geographic area around a given location. To represent geographic areas along the route, we make use of user-contributed images and their associated annotations. For this purpose, we systematically collect information about venues and the images depicting them from location-based social networking platform Foursquare and we turn to content sharing website Flickr for a richer set of images and metadata capturing

J. van den Berg—This research was performed while the first author was a student at the UvA.

[1] Scenemash demo: https://www.youtube.com/watch?v=oAnj6A1oq2M.

N. Ferro et al. (Eds.): ECIR 2016, LNCS 9626, pp. 833–836, 2016.
DOI: 10.1007/978-3-319-30671-1_75

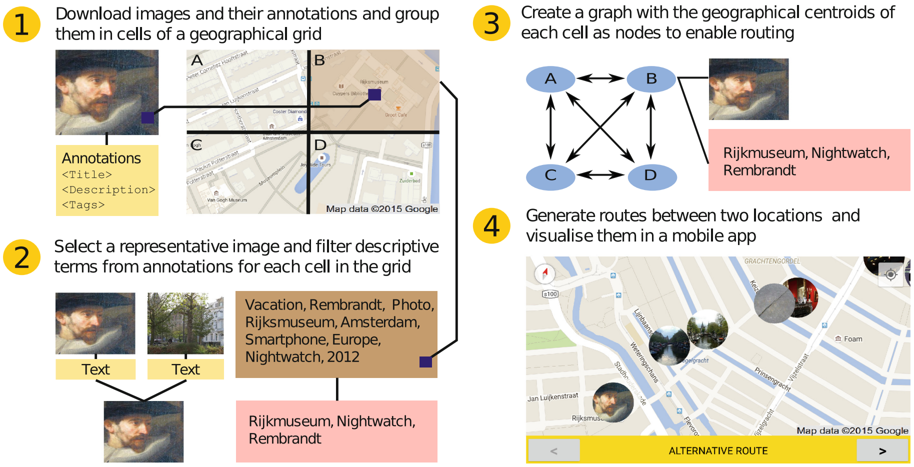

1
Download images and their annotations and group them in cells of a geographical grid
A
B
C
D
Annotations
<Title>
<Description>
<Tags>
Map data ©2015 Google
2
Select a representative image and filter descriptive terms from annotations for each cell in the grid
Text
Text
Vacation, Rembrandt, Photo, Rijksmuseum, Amsterdam, Smartphone, Europe, Nightwatch, 2012
Rijkmuseum, Nightwatch, Rembrandt
3
Create a graph with the geographical centroids of each cell as nodes to enable routing
A
B
C
D
Rijkmuseum, Nightwatch, Rembrandt
4
Generate routes between two locations and visualise them in a mobile app
Map data ©2015 Google
<
ALTERNATIVE ROUTE
>

Fig. 1. Overview of our approach to multimodal summarization of tourist routes.

a wide range of aspects users find interesting. We create summaries by jointly analysing distributions of semantic concepts detected in the visual channel of the images and the latent topics extracted from their associated annotations. To demonstrate the effectiveness of our approach, using Amsterdam as a showcase, we implement Scenemash prototype as an Android app for smartphones and smartwatches.

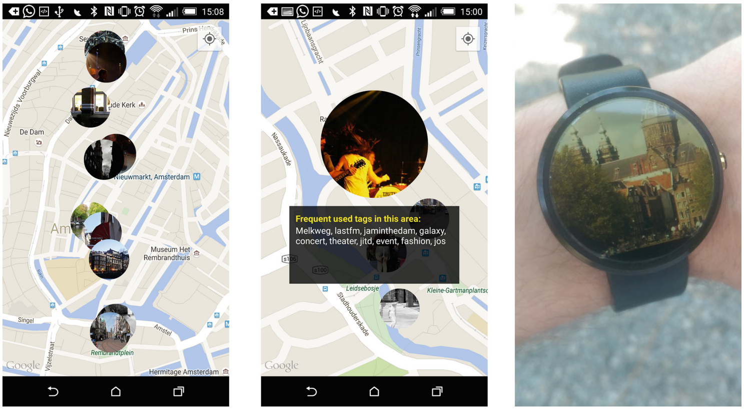
2 Approach Overview

The pipeline of our approach is illustrated in Fig. 1. In this section we describe data collection and analysis steps as well as the procedure for generating multimodal representations of the geographic areas (i.e., steps 1 and 2).

2.1 Data Collection and Analysis

We first queried the Facebook Graph and Foursquare APIs and compiled a list of all POIs within the radius of 9 km from the centre of Amsterdam. Then, we crawled georeferenced Flickr images along with their annotations (i.e., title, description and tags) within 500 m from each POI. To further enlarge the collection, we downloaded more images taken in Amsterdam by already known Flickr users. Finally, we crawled images of all verified Foursquare venues. The resulting dataset consisted of 157,000 images and their associated annotations.

We represent visual content of each image in the collection with a distribution of 15,293 semantic concept scores output by a customised implementation of Google "Inception" net [6]. We tokenize the text associated with the images and remove stopwords, unique words and HTML markup. After preprocessing step, we represent each image in the text domain using 100 LDA topics extracted using Gensim framework [8].



Frequent used tags in this area:
Melkweg, lastfm, jaminthedam, galaxy, concert, theater, jitd, event, fashion, jos

Fig. 2. Smartphone and smartwatch user interfaces.

2.2 Summarization of a Geographic Area

We use a rectangular geographical grid with 125 × 125 meter cells to define geographic areas and group the images. For each grid cell, we compute pairwise cosine similarity matrices for both distributions of visual concepts and LDA topics, extracted as described in previous section. We then combine such created unimodal similarity matrices using the weighting fusion scheme proposed by Ah-Pine et al. [1]. The resulting multimodal similarities serve as an input into the affinity propagation clustering, which aims at automatically selecting an optimal number of clusters [4]. Finally, we sort clusters in decreasing order of their size and select the first centroid available under a Creative Commons license as a representative image for a given geographic area.

As a starting point for generating description of the area we make use of preprocessed text associated with the images (cf. Sect. 2.1). To identify the terms representative of a particular geographic area, we apply tf-idf weighting, considering each grid cell as a single document. In general, tf-idf discriminates well between the terms that are used on a certain location and those used in the entire city. However, it also has the tendency to give a high weight to rare (often unwanted) terms. To mitigate this effect and improve alignment between visual and text representation of the area, we utilize tag ranking approach similar to the one introduced by Li et al. [5] and weight the terms by their frequency of occurrence in the k-nearest visual neighbours of the selected representative image. The ranked lists of terms produced by the two above-mentioned weighting schemes are combined and the top-10 ranked terms are selected.

3 Scenemash Prototype Design

We implement the Scenemash prototype as a native Android app for smartphones and smartwatches, which the attendees will be able to test at the conference. The app interface allows the user to query locations in the city or use the current location provided by GPS sensor. Scenemash features "explore" and "get route" functions. On the server side, a graph illustrated in step 3 of Fig. 1 is used to get the neighbouring nodes/grid cells of a node containing user coordinates when explore function is selected. In the get route mode, we apply the breadth-first search algorithm on the same graph for computing a route between two locations. Alternative routes are computed by selecting different neighbour nodes of the origin node. To give the users an opportunity to avoid crowded places, we create a weighted version of the same graph which uses the number of images captured in a geographic cell as a proxy for crowdedness. If the crowd avoidance feature is selected, we deploy Dijkstra's shortest path algorithm for computing the route between two locations.

The data collection and analysis steps described in Sect. 2 are precomputed offline, in order to reduce online computation load. Figure 2 illustrates the user interfaces. Each relevant geographic area (i.e., on the route or in user's vicinity) is represented by a circular thumbnail displayed in Google Maps. If the smartphone is paired with a smartwatch, the images are shown as a slideshow on the smartwatch. When a user interacts with the map by tapping on one of the images, the image is enlarged and an info-box with the most relevant terms for the area is shown. The effectiveness of the prototype gives us confidence that the Scenemash could be implemented in other cities as well.

References

1. Ah-Pine, J., Clinchant, S., Csurka, G., Liu, Y.: Xrce's participation in imageclef. In: Working Notes of CLEF 2009 Workshop Co-located with the 13th European Conference on Digital Libraries (ECDL 2009) (2009)
2. Brilhante, I., Macedo, J.A., Nardini, F.M., Perego, R., Renso, C.: TripBuilder: a tool for recommending sightseeing tours. In: de Rijke, M., Kenter, T., de Vries, A.P., Zhai, C.X., de Jong, F., Radinsky, K., Hofmann, K. (eds.) ECIR 2014. LNCS, vol. 8416, pp. 771–774. Springer, Heidelberg (2014)
3. Cheng, A.-J., Chen, Y.-Y., Huang, Y.-T., Hsu, W.H., Liao, H.-Y.M.: Personalized travel recommendation by mining people attributes from community-contributed photos. In: Proceedings of the 19th ACM International Conference on Multimedia, MM 2011, pp. 83–92. ACM, New York (2011)
4. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. Science **315**(5814), 972–976 (2007)
5. Li, X., Snoek, C., Worring, M.: Learning social tag relevance by neighbor voting. IEEE Trans. Mult. **11**(7), 1310–1322 (2009)
6. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9, June 2015
7. Zahálka, J., Rudinac, S., Worring, M.: New yorker melange: interactive brew of personalized venue recommendations. In: Proceedings of the ACM International Conference on Multimedia, MM 2014, pp. 205–208. ACM, New York (2014)
8. Řehůřek, R., Sojka, P.: Software framework for topic modelling with large corpora. In: Proceedings of LREC Workshop New Challenges for NLP Frameworks, pp. 46–50. University of Malta, Valletta (2010)