Discovering Geographic Regions in the City Using Social Multimedia and Open Data

Rudinac, S.; Zahálka, J.; Worring, M.

DOI
10.1007/978-3-319-51814-5_13

Publication date
2017

Document Version
Final published version

Published in
MultiMedia Modeling

License
Article 25fa Dutch Copyright Act

Link to publication

Citation for published version (APA):
Rudinac, S., Zahálka, J., & Worring, M. (2017). Discovering Geographic Regions in the City Using Social Multimedia and Open Data. In L. Amsaleg, G. Þ. Guðmundsson, C. Gurrin, B. Þ. Jónsson, & S. Satoh (Eds.), *MultiMedia Modeling: 23rd International Conference, MMM 2017, Reykjavik, Iceland, January 4-6, 2017 : proceedings* (Vol. 2, pp. 148-159). (Lecture Notes in Computer Science; Vol. 10133). Springer. https://doi.org/10.1007/978-3-319-51814-5_13

Discovering Geographic Regions in the City Using Social Multimedia and Open Data

Stevan Rudinac(✉), Jan Zahálka, and Marcel Worring

Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands
{s.rudinac,j.zahalka,m.worring}@uva.nl

Abstract. In this paper we investigate the potential of social multimedia and open data for automatically identifying regions within the city. We conjecture that the regions may be characterized by specific patterns related to their visual appearance, the manner in which the social media users describe them, and the human mobility patterns. Therefore, we collect a dataset of Foursquare venues, their associated images and users, which we further enrich with a collection of city-specific Flickr images, annotations and users. Additionally, we collect a large number of neighbourhood statistics related to e.g., demographics, housing and services. We then represent visual content of the images using a large set of semantic concepts output by a convolutional neural network and extract latent Dirichlet topics from their annotations. User, text and visual information as well as the neighbourhood statistics are further aggregated at the level of postal code regions, which we use as the basis for detecting larger regions in the city. To identify those regions, we perform clustering based on individual modalities as well as their ensemble. The experimental analysis shows that the automatically detected regions are meaningful and have a potential for better understanding dynamics and complexity of a city.

Keywords: Urban computing · Social multimedia · Open data · Human mobility patterns · Semantic concept detection · Topic modelling

1 Introduction

A modern city is a complex organism shaped by the dynamics of various processes, related to e.g., economy, infrastructure and demographics. Administrative divisions, therefore, often do not match the *actual* regions in the city determined by their common functionality, architectural resemblance and ever-changing human flows. Identifying those regions is of critical importance for better understanding and modelling the processes in a city. Recently, open data has been shown invaluable in solving various problems a modern metropolis is faced with. However, open data is often associated with a lack of content connecting various sources and the absence of "factual" information about human flows and their perception of the habitat, so it does not provide the full picture of city dynamics.

L. Amsaleg et al. (Eds.): MMM 2017, Part II, LNCS 10133, pp. 148–159, 2017.
DOI: 10.1007/978-3-319-51814-5_13

While much of social multimedia is spontaneously captured, we conjecture that useful information can still be encoded in it. User-contributed images and their associated metadata in particular may offer valuable insights about the properties of a geographic area. Indeed, users with similar background and interests are intuitively more likely to reside, work or seek entertainment in the same geographic area, which will be reflected in their mobility patterns. Regions of a city associated with a similar functionality, such as business, education or entertainment, are further more likely to share certain visual attributes and be described by the users in similar way. Similarly, regions of the city constructed during a particular epoch may bear resemblance in terms of architectural style, function or demographics. Social multimedia has recently been succesfully deployed in analysing various urban and geographic phenomena [5,6,11,15,17,21,29]. Its full utilization remains challenging due to e.g., a wide variety of user interests, which makes their contributed content extremely topically heterogeneous. Additionally, as compared to professional or curated content, valuable information is only implicitly embedded in social multimedia. Appropriate multimedia analysis techniques are needed to bring out the best from social multimedia.

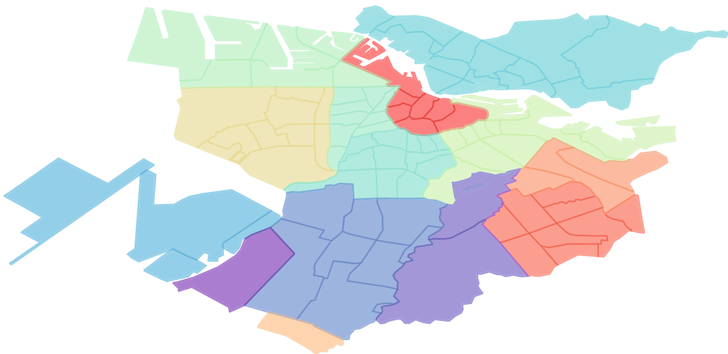


Fig. 1. A map of districts and postal regions in the Amsterdam Metropolitan area used in our study.

In this paper we investigate the potential of social multimedia and open data for identifying and characterizing functional regions of the city. For this purpose we systematically crawl the location-based social networking platform Foursquare and the content sharing platform Flickr. Additionally, we make use of open data related to various neighbourhood statistics, such as demographics, housing, transportation and services. In our analysis we focus on investigating both the descriptive power of each individual information channel, as well as its usefulness in interpreting results obtained by the other channels. As the test bed for our analysis we chose the Amsterdam Metropolitan Area (cf. Fig. 1), due to its moderate size, multicultural nature and a unique blend of historic and modern architecture. Next to the potential for aiding the analysis of various urban issues, our approach is applicable in a number of multimedia problems, such as location recommendation [26,28], recognition and exploration [5], region summarization [17] and content placing [12,13,24].

The remainder of this paper is organised as follows. Section 2 provides a brief overview of related work. In Sect. 3 we introduce our approach and in Sect. 4 we report on experimental results. Section 5 concludes the paper.

2 Related Work

Recently, a number of approaches to detecting characterizing regions within a city have been proposed. The majority of these methods focus on utilizing information about points of interest (POIs) and human mobility patterns. For example, to discover functional regions of a city Yuan et al. analyse the distribution of POI categories together with information about road network topology and the GPS trajectories generated by taxi vehicles [27], while Toole et al. utilize spatio-temporal information about mobile phone calls [23]. More recently, Andrienko et al. proposed a visual analytics approach, which jointly utilizes mobile phone usage data together with information about land use and points of interest for discovering place semantics [1].

While the above-mentioned approaches have been proven effective in various use cases, they often rely on scarce and difficult to obtain data, which limits their wider adoption. Following a different logic, Thomee and Rae propose a multimodal approach to detecting locally characterizing regions in a collection of geo-referenced Flickr images [22]. The regions are detected at several scales, corresponding to e.g., world, continent, and country level. With regard to data use and research goal, the analysis conducted by Cranshaw et al. [3] is probably most relevant to our study. The authors of the paper utilize information about location and users of a large number of Foursquare venues to identify dynamic areas comprising the city, which they name *livehoods*. While offering several original ideas and an interesting qualitative analysis, the study does not utilize open data and the content (i.e., text and visual) richness of social multimedia.

3 Region Detection

In this section we describe our data collection procedure and then elaborate on our approach for identifying regions within a city based on different modalities. The approach pipeline is illustrated in Fig. 2.

3.1 Data Collection

To ensure reasonable granularity and to allow easier visual interpretation of the results, we use postal code regions as the unit for data collection and analysis. As parameter for querying the Foursquare venue database we use geo-coordinates of the centres of the regions corresponding to street-level postal codes, and set the query radius to 100 m. For each venue we further download images as well as the information about the users who recommended it.

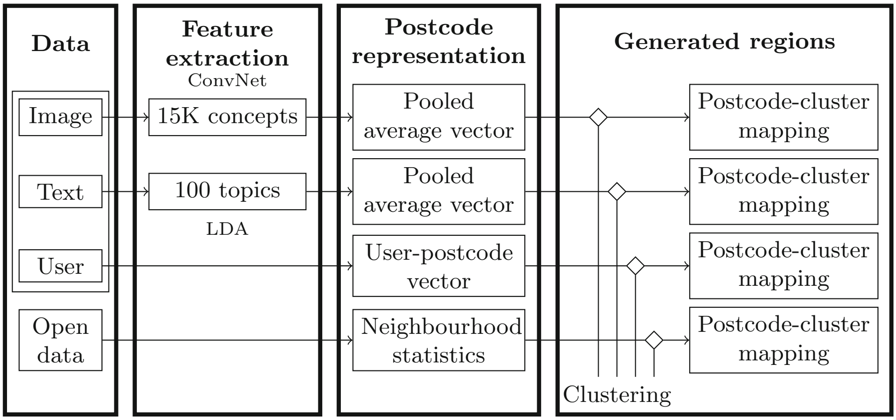


Fig. 2. The pipeline of our region detection approach.

As the Flickr API imposes fewer restrictions on the number of results per request, we query it using geo-coordinates of the centres of neighbourhood-level postal regions (cf. Fig. 1), modularly adjusting the radius to the size of the region. Images are downloaded together with the information about their uploaders as well as the title, tags and description. We restricted our search to the Creative Commons images assigned with the highest available location accuracy.

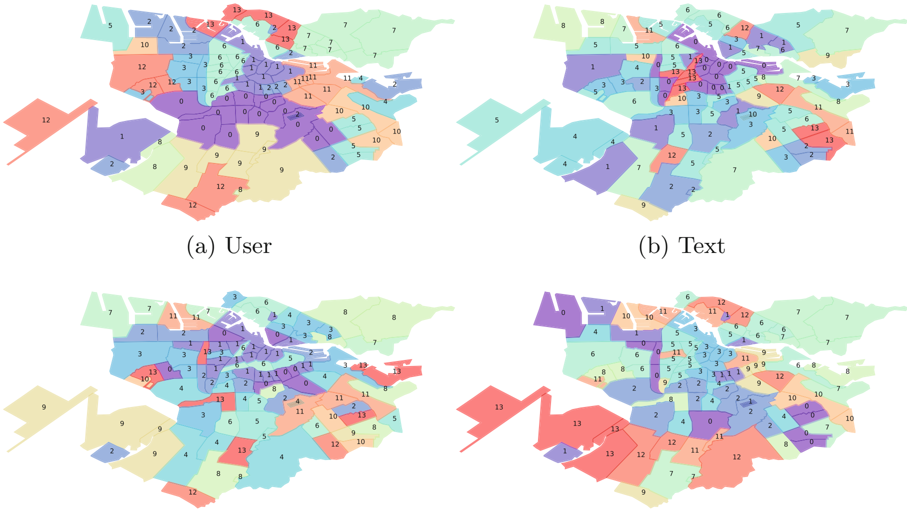
Due to an irregular shape of postal code regions, the above-mentioned data collection procedure can result in image or venue assignment to multiple postal code regions. To make precise assignments, we perform reverse geo-coding using postal code shapefiles downloaded from Google Maps [8].

Finally, we make use of official neighbourhood statistics, regularly compiled by the local government. The collection includes a large number of variables related to neighbourhood *demographics*, *companies*, *housing*, *energy consumption*, *motor vehicles*, *surface area*, and *facilities*. In Sect. 4.3 we will discuss the most discriminative variables from the above-mentioned categories.

3.2 Feature Extraction

Visual Features: Given the wide variety of potentially interesting visual attributes that should be captured and to make interpretation of results easier, we opt for image representation at an intermediate semantic level. For each image we extract 15,293 ImageNet concepts [4] output by a customised Caffe [10] implementation of "Inception" network [20]. A postcode region is then represented with a concept vector generated by applying average pooling [2] on concept vectors of individual images captured within it.

Text Features: We index the *title*, *description* and *tags* text of each Flickr image. Before computing a bag of words representation, we pre-process the text by removing HTML elements and stopwords. We further apply Online Latent



(a) User (b) Text

(c) Visual (d) Ensemble

Fig. 3. A map of Amsterdam regions produced by spectral clustering based on (a) user (b) text and (c) visual modalities as well as (d) their ensemble.

Dirichlet Allocation (LDA) [9] and extract 100 latent topics using the gensim framework [16]. Finally, we use average pooling over LDA vectors of individual images to produce a topical representation for a postal code region.

User Features: We represent each postal code region with a binary vector to indicate if a particular Flickr or Foursquare user visited it. We associate Foursquare users with a postal code region if they recommended a venue within it. Flickr users uploading an image captured within a postal code region are associated with it.

Open Data: For each postal code we create representations containing neighbourhood statistics of a particular category (e.g., *demographics* or *housing*), which we further combine in an early fusion fashion. The vectors are scaled to the [0,1] range.

3.3 Clustering

After the representations of postcode regions are computed for each social media modality (i.e., visual, text and user), we proceed by identifying the borders of larger geographic regions within the city. For that purpose, in the visual and text domain we independently compute cosine similarity between the semantic concept and LDA postcode representations. Similarly, for the user modality, we compute Jaccard similarity between binary user vectors representing postal code regions. The similarities are further used as an input into spectral clustering [14]. After the clustering in individual modalities is performed, we deploy an ensemble (consensus) clustering approach [19] to produce a single, reinforced clustering. In case of open data, we resort to k-means clustering, which is more appropriate for relatively short feature vectors.

We set the number of clusters to be equal to the number of administrative units in the city (e.g., 14 city quarters and municipalities visualised in Fig. 1). While performing clustering, unlike most related work, we intentionally chose not to impose a geographic constraint. This choice is motivated by our specific goal to investigate whether meaningful boundaries of geographic regions could be identified using individually information on human mobility patterns, visual properties and users' perception of the city. Since most clustering algorithms attempt to identify larger regions, imposing a geographic constraint would make conclusions of such analysis unreliable. For example, two neighbouring postal code regions would more likely belong to the same cluster, independently of their topical similarity, which could easily lead to misinterpretation of results.

4 Experimental Results

In this section we seek to answer the following questions:

1. What do the social media channels tell us about the regions in a city?
2. To what degree do the computed region boundaries agree with the official administrative division?
3. Which regions can be identified based on neighbourhood statistics?

Applying the procedure described in Sect. 3.1 to Amsterdam, we arrive at a collection consisting of 1,136 verified venues, 4,605 unique users and 34,419 images from Foursquare and 59,417 Flickr images uploaded by 2,342 users. The neighbourhood statistics are provided as open data by Statistics Netherlands (CBS) [18].

4.1 Regions Shaped by Social Multimedia

The maps of Amsterdam regions output by spectral clustering based on user, text and visual modality as well as the ensemble clustering are respectively shown in Figs. 3a–d. Our first observation is that relatively large and coherent regions can be identified in all four visualisations, most notably the one shown in Fig. 3a. The regions revealed by user modality are furthermore particularly easy to interpret. For example, the number "1" region, comprising a postal code in south-west corner and another ten in the central part of the figure corresponds to Amsterdam Airport Schiphol and the centre of the city (i.e., the Amsterdam Canal District). Such association is not surprising, since Amsterdam is a popular tourist destination[1] with more than 15,000,000 visits per year, many of which begin with arrival to Schiphol airport, followed by transfer to Amsterdam Central station. Additional three postcodes at the south of the same region correspond

[1] www.amsterdam.info/basics/figures/.

to the area around museum quarter, which is again, very popular among tourists. Another general observation related to Fig. 3a is that the user mobility patterns are normally limited to a particular quarter of the city.

The results of visual clustering reveal interesting patterns as well. For example, a region detected in south-west corner of Fig. 3c comprises visually characteristic large industrial areas. In the central quarter of the city, for example, a distinction is correctly made between the oldest city parts featuring typical canal houses and two postcode regions constructed more recently. Similar observations can be made about large regions in the north, east and west of the city, which were built in coherent architectural styles. Finally, a large cluster encompassing the postcodes in those three districts of the city corresponds to the decaying residential neighbourhoods.

Table 1. Evaluation of the agreement between partitioning into regions output by four different clustering algorithms and the official administrative division; the agreement is reported in terms of Adjusted Mutual Information (AMI).

	GRT	VIS	TXT	USR	ENS	RAN
GRT	•	0.134	0.115	0.452	0.265	0.014
VIS	0.134	•	0.106	0.114	0.367	0.005
TXT	0.115	0.106	•	0.106	0.295	0.035
USR	0.452	0.114	0.106	•	0.524	0.012
ENS	0.265	0.367	0.295	0.524	•	0.052
RAN	0.014	0.005	0.035	0.012	0.052	•

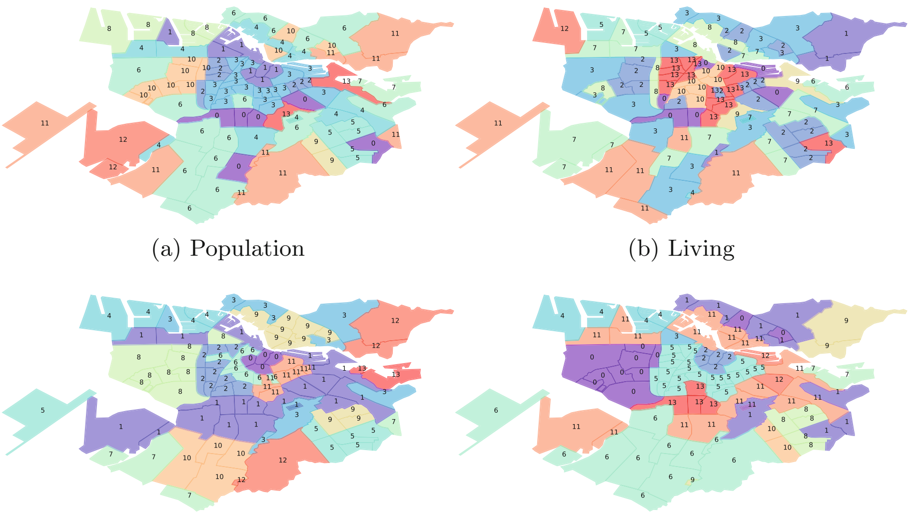
4.2 Agreement with the Official Division

Table 1 summarizes the agreement between the following area partitions: the ground truth official administrative division (GRT) as visualized in Fig. 1; the output of the four different clusterings: visual (VIS), text (TXT), user information (USR), and ensemble (ENS); the results of random clustering (RAN) serving as a reference point. As the measure of agreement we use Adjusted Mutual Information (AMI), which enumerates how the individual partitionings agree with each other [25]. An AMI of 1 means perfect agreement, while 0 corresponds to the average agreement with random partitioning.

Relatively low agreement between text, visual and user partitionings indeed indicates that all modalities provide complementary information. At the same time, all automatically generated partitionings score well above random, showing that the disagreement is topical, not manifesting by chance. The clustering by user modality is the closest to the administrative division, suggesting that the official administrative districts do roughly follow *vox populi*. A lower agreement with the administrative division observed in the case of visual and text modalities may reflect the fact that the large official administrative regions, for example, do not have uniform visual appearances or functionality. Ensemble clustering agrees with the constituent unimodal approaches reasonably, most dominantly with the user-based clustering, which is the least fragmented one (cf. Fig. 3). Ensemble thus does a good job in smoothing out the fragmentation brought to the table by visual and text clusterings, while including their complementary information. Making use of multimodal data and analysis is thus shown to bring rich information about the city, reflecting diverse aspects and showing promise for urban computing.

4.3 Regions by Neighbourhood Statistics

Figures 4a–d show the regions generated based on statistics about population, living, services and the combined neighbourhood statistics. Again, we observe large coherent and meaningful regions. For example, region "5" in Fig. 4d corresponds to some of the most exclusive neighbourhoods of the city, including the parts of Amsterdam Canal District and Old South. Further, large regions in the west, south-east and north of the city, corresponding to the known disadvantaged neighbourhoods are also correctly identified. Finally, region "6" encompasses several municipalities neighbouring Amsterdam. Similar observations can be made in Fig. 4c, where the regions with comparable socio-economic structure are grouped together.



(a) Population (b) Living

(c) Services (d) Combined neighbourhood statistics

Fig. 4. A map of Amsterdam regions produced by k-means clustering based on different categories of neighbourhood statistics.

Table 2. Evaluation of the agreement between partitioning into regions based on different categories of official neighbourhood statistics and social multimedia; the agreement is reported in terms of Adjusted Mutual Information (AMI).

	GRT	VIS	TXT	USR	ENS	RAN
ENE	0.145	0.124	0.051	0.090	0.054	0.007
COM	0.260	0.029	0.008	0.093	0.040	−0.004
LIV	0.239	0.120	0.119	0.161	0.148	0.010
SUR	0.149	0.175	0.059	0.125	0.120	0.004
VEH	0.233	0.150	0.043	0.125	0.110	−0.006
SER	0.444	0.132	0.086	0.242	0.143	0.013
POP	0.385	0.127	0.112	0.239	0.182	0.008
ALL	0.462	0.147	0.112	0.260	0.188	0.043

Table 3. A list of top-15 variables comprising neighbourhood statistics, sorted by their importance in discriminating between the regions shown in Fig. 4d.

Rank	Variable	Category
1	Number of department stores within 5 km	Services
2	Number of hotels within 5 km	Services
3	Number of secondary schools within 5 km	Services
4	Number of restaurants within 5 km	Services
5	Number of bars within 1 km	Services
6	Number of attractions within 50 km	Services
7	Percentage of surinamese immigrants	Population
8	Number of grocery stores	Services
9	Percentage of married inhabitants	Population
10	Number of cafeterias within 5 km	Services
11	Number of bars within 5 km	Services
12	Percentage of unmarried inhabitants	Population
13	Percentage of turkish immigrants	Population
14	Percentage of western immigrants	Population
15	Percentage of moroccan immigrants	Population

The results shown in Table 2 suggest a high agreement between the regions yielded by the analysis of user mobility patterns on one side and the statistics about demographics, services and the combined neighbourhood statistics on the other. This agreement is also apparent in Figs. 3a, 4a, c and d, which further confirms the importance of information about human flows for better understanding the city. Another interesting observation in Table 2 is that the visual modality in general does a better job than text in "emulating" official neighbourhood statistics. In particular, visual modality appears to be effective in capturing the characteristics of a neighbourhood related to surface use, motor vehicles, services and population.

Finally, we train an extra-tree classifier [7] to investigate which variables are the most useful for discriminating between the regions generated based on combined neighbourhood statistics. Based on the results from Table 3 we conclude that the demographics and services are important forces shaping up the city.

5 Conclusions

In this paper we have investigated the feasibility of utilizing social multimedia and open data for identifying actual regions of the city. Our analysis involved information about human mobility patterns, visual appearance and user perception of the city as well as the official neighbourhood statistics. The experiments show that the areas proposed by our approach reflect diverse semantic aspects of the city, with each information channel contributing unique information. Region detection based on social multimedia and open data thus shows promise in reflecting the vibrant and ever-changing nature of the city, making it a solid basis for further urban computing research.

References

1. Andrienko, N., Andrienko, G., Fuchs, G., Jankowski, P.: Scalable and privacy-respectful interactive discovery of place semantics from human mobility traces. Inf. Vis. **15**(2), 117–153 (2016)
2. Boureau, Y.-L., Ponce, J., LeCun, Y.: A theoretical analysis of feature pooling in visual recognition. In: Proceedings of the 27th International Conference on Machine Learning (ICML 2010), pp. 111–118 (2010)
3. Cranshaw, J., Schwartz, R., Hong, J., Sadeh, N.: The livehoods project: utilizing social media to understand the dynamics of a city. In: International AAAI Conference on Weblogs and Social Media (2012)
4. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 248–255, June 2009
5. Fang, Q., Sang, J., Xu, C.: Giant: geo-informative attributes for location recognition and exploration. In: Proceedings of the 21st ACM International Conference on Multimedia, MM 2013, pp. 13–22. ACM, New York (2013)
6. Boonzajer Flaes, J., Rudinac, S., Worring, M.: What multimedia sentiment analysis says about city liveability. In: Ferro, N., Crestani, F., Moens, M.-F., Mothe, J., Silvestri, F., Nunzio, G.M., Hauff, C., Silvello, G. (eds.) ECIR 2016. LNCS, vol. 9626, pp. 824–829. Springer, Heidelberg (2016). doi:10.1007/978-3-319-30671-1_74
7. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. Mach. Learn. **63**(1), 3–42 (2006)
8. Google Maps. Postcodes Amsterdam. http://goo.gl/hHoZWi. Accessed Nov 2015
9. Hoffman, M., Bach, F.R., Blei, D.M.: Online learning for latent Dirichlet allocation. In: Advances in Neural Information Processing Systems, NIPS 2010, pp. 856–864 (2010)

10. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia, MM 2014, pp. 675–678. ACM New York (2014)
11. Kennedy, L., Naaman, M., Ahern, S., Nair, R., Rattenbury, T.: How flickr helps us make sense of the world: context and content in community-contributed media collections. In: Proceedings of the 15th ACM International Conference on Multimedia, MM 2007, pp. 631–640. ACM, New York (2007)
12. Larson, M., Soleymani, M., Serdyukov, P., Rudinac, S., Wartena, C., Murdock, V., Friedland, G., Ordelman, R., Jones, G.J.F.: Automatic tagging, geotagging in video collections, communities. In: Proceedings of the 1st ACM International Conference on Multimedia Retrieval, ICMR 2011, pp. 51:1–51:8. ACM, New York (2011)
13. Luo, J., Joshi, D., Yu, J., Gallagher, A.: Geotagging in multimedia and computer vision–a survey. Multimedia Tools Appl. **51**(1), 187–211 (2011)
14. Ng, A.Y., Jordan, M.I., Weiss, Y.: On spectral clustering: analysis and an algorithm. In: Dietterich, T., Becker, S., Ghahramani, Z. (eds.) Advances in Neural Information Processing Systems 14, pp. 849–856. MIT Press, Cambridge (2002)
15. Porzi, L., Rota Bulò, S., Lepri, B., Ricci, E.: Predicting and understanding urban perception with convolutional neural networks. In: Proceedings of the 23rd ACM International Conference on Multimedia, MM 2015, pp. 139–148. ACM, New York (2015)
16. Řehůřek, R., Sojka, P.: Software framework for topic modelling with large corpora. In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, pp. 45–50. ELRA, Valletta, May 2010
17. Rudinac, S., Hanjalic, A., Larson, M.: Generating visual summaries of geographic areas using community-contributed images. IEEE Trans. Multimedia **15**(4), 921–932 (2013)
18. Statistics Netherlands. Neighbourhood statistics. https://www.cbs.nl/nl-nl/maatwerk/2015/48/kerncijfers-wijken-en-buurten-2014. Accessed Nov 2015
19. Strehl, A., Ghosh, J.: Cluster ensembles – a knowledge reuse framework for combining multiple partitions. J. Mach. Learn. Res. **3**, 583–617 (2003)
20. 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
21. Thomee, B., Arapakis, I., Shamma, D.A.: Finding social points of interest from georeferenced and oriented online photographs. ACM Trans. Multimedia Comput. Commun. Appl. **12**(2), 36:1–36:23 (2016)
22. Thomee, B., Rae, A.: Uncovering locally characterizing regions within geotagged data. In: Proceedings of the 22nd International Conference on World Wide Web, WWW 2013, pp. 1285–1296 (2013)
23. Toole, J.L., Ulm, M., González, M.C., Bauer, D.: Inferring land use from mobile phone activity. In: Proceedings of the ACM SIGKDD International Workshop on Urban Computing, UrbComp 2012, pp. 1–8. ACM, New York (2012)
24. Trevisiol, M., Jégou, H., Delhumeau, J., Gravier, G.: Retrieving geo-location of videos with a divide & conquer hierarchical multimodal approach. In: Proceedings of the 3rd ACM International Conference on Multimedia Retrieval, ICMR 2013, pp. 1–8. ACM, New York (2013)

25. Vinh, N.X., Epps, J., Bailey, J.: Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance. J. Mach. Learn. Res. **11**, 2837–2854 (2010)
26. Yin, H., Cui, B., Huang, Z., Wang, W., Wu, X., Zhou, X.: Joint modeling of users' interests and mobility patterns for point-of-interest recommendation. In: Proceedings of the 23rd ACM International Conference on Multimedia, MM 2015, pp. 819–822. ACM, New York (2015)
27. Yuan, J., Zheng, Y., Xie, X.: Discovering regions of different functions in a city using human mobility and POIs. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2012, pp. 186–194. ACM, New York (2012)
28. Zahálka, J., Rudinac, S., Worring, M.: Interactive multimodal learning for venue recommendation. IEEE Trans. Multimedia **17**(12), 2235–2244 (2015)
29. Zheng, Y., Capra, L., Wolfson, O., Yang, H.: Urban computing: concepts, methodologies, and applications. ACM Trans. Intell. Syst. Technol. **5**(3), 38:1–38:55 (2014)