

Image and 3D structure based ontology for object recognition

Paulo Dias Almeida and Jorge Gustavo Rocha

Abstract Image data is used constantly to express and convey information. Modern trends in technology and social media have made pictures a central part of human communication. However, the lack of semantic knowledge in image data severely limits its application. In this paper we present our research proposal for an improved object recognition method that uses semantic data over images that takes into consideration the structural environment of where the picture was taken.

1 Introduction

The concept of smart city is gaining increasing importance in the search for a sustainable society. Modern smart city implementations strive to function as an integrated system, capable of assessing its conditions and needs in real time, for the entire system [1, 3]. This requirement leads to a heavy reliance on sensors.

It is based on the data collected by sensors that informed, intelligent decisions can be made [4]. The sensors that can be used for this goal are of a varied nature, depending on what one desires to monitor. A particular sensor that is beginning to gain increasing importance in the smart city environment is the citizen.

The advances in technology and connectivity have made the general public knowledgeable, capable, and willing of being integrated in the smart city ecosystem. With the integration of citizens as a source of data, and because of the general availability of devices equipped with digital cameras, like smartphones and tablets,

Paulo Dias Almeida
Minho University, Braga, Portugal, e-mail: diasalmeida.paulo@gmail.com

Jorge Gustavo Rocha
Minho University, Braga, Portugal, e-mail: jgr@di.uminho.pt

Copyright (c) by the paper's authors. Copying permitted for private and academic purposes.
In: A. Comber, B. Bucher, S. Ivanovic (eds.): Proceedings of the 3rd AGILE Phd School, Champs sur Marne, France, 15-17-September-2015, published at <http://ceur-ws.org>

the city now has at its disposal thousands of moving camera sensors capable of providing images as a potential source of data for the smart city.

The use of cameras is in fact becoming so omnipresent in the modern lifestyle, that images are beginning to replace text as the mean of communication in several social media implementations. In fact, nowadays if we do a search in social networks for a particular current event we will find large quantities of images of that event provided by citizens. This data is being provided in real-time and is frequently complemented with text from which more information could be extracted.

This means that we have an abundance of image data available and capable of enhancing the knowledge of the city conditions essential for the smart city. However, the data present on an unprocessed image comes with almost no metadata or information on its contents and structure.

We propose that image data provided by the citizen has the potential to be used as an important sensor for the smart city environment. Specifically, we want to provide an algorithm for the classification of objects and structures present in an image. For this classification we want to take advantage of extracted knowledge of the city infrastructure.

In the remainder of this extended abstract we present our research proposal, starting with a short presentation of the most relevant related work in section 2, we will elaborate on our approach and research question in 3, and finally, in 4, we present the conclusions and future work.

2 Related Work

In order to combine semantic information over image data with structural knowledge of the surrounding geographical area, we first need to have available a reconstruction of the 3D structure of the city.

2.1 *Structure-from-Motion (SFM)*

Despite the desire for our method to work with any representation of 3D structure of a city, we found it beneficial to start by focusing in a specific method for image-based reconstruction of 3D structure.

SFM solutions have the goal of generating 3D reconstructions from image sequences. The algorithm is capable of solving the camera position and orientation simultaneous with the geometry of the scene, using a iterative bundle adjustment procedure based on a database of image features extracted from the sequence of overlapping images [8].

An approach of significant interest is presented in [7, 8]. These projects set out to provide an interactive browsing of large photo collections available online, integrated with the 3D model reconstructed from these photos. The nature of this dataset

of images implies a challenging set of conditions, including the use of different cameras, zoom levels, resolutions, etc. The approach proposed is capable of successfully solving this problem, however, it can take a couple of weeks for larger collections of photos.

With the success of SFM in the reconstruction of 3D structure using increasingly large and diverse sets of images, attention began to be shifted to its time complexity. In [9] several steps are taken in order to achieve close to linear time complexity. This implementation presents great gains in performance, being able to generate an accurate large-scale reconstruction (32000 images) in approximately 2 hours.

2.2 Ontology-based image processing

The data present on an unprocessed image has almost no metadata and there is no information on its structure or degree of importance. This fact contrasts with the user perception, where an image is capable of conveying a large amount of information. This problem is called the semantic gap [6]. In order to address this concern, several approaches utilize ontologies in image processing.

A relevant example of a tool looking to explore the potential gains of the use of semantic knowledge in image processing is presented in [5]. In this project the authors present the framework OntoPic, capable of automatically annotating images with relevant keywords, enabling a content based image retrieval that functions on a semantic level. The OntoPic framework takes a provided domain ontology and uses supervised learning techniques in order to train its classification.

Also of particular interest is the work presented in [2], seeing as it focus on object recognition in the urban environment. In order to achieve this goal the authors combine the efforts of experts with machine learning tools in order to build their ontology. This ontology is then used for object recognition in segmented images using a similarity measure with interesting results.

3 Approach

With our research, we aim to combine the results of 3D structure information, provided by reconstructions of the city structure, with semantic knowledge in the domain of the urban environment in order to provide an improved image-based object and structure classification system capable of enabling and enhancing a smart city implementation.

In our proposed approach, an image is first localized geographically and spatially in the city and only then we proceed to the detection and classification of objects and structures, enhancing this detection with extracted knowledge of the involving infrastructure. Figure 1 presents an overview of our proposed approach.

With our research we are mainly trying to provide answers for the question: "To what extent can we create a semantic representation of spaces/objects of a city, taking advantage of the images contributed by the citizen?"

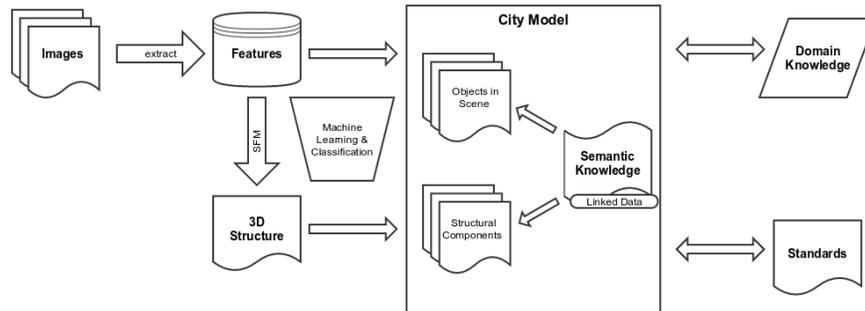


Fig. 1 Overview of the proposed system.

4 Conclusion

Image data is capable of conveying a large amount of information and can be supplied by the citizens across the city, making it a potential sensor for a smart city environment. The low amount of metadata and structure associated with this source of information originates the need for semantic knowledge to be added to image processing. With our proposed approach we will complement the semantic knowledge extracted from the image data with information relating to the 3D structure of its involving space to provide an improved method for object and structure recognition.

References

1. Caragliu, A., Del Bo, C., Nijkamp, P.: Smart cities in europe. *Journal of Urban Technology* **18**(2), 65–82 (2011). DOI 10.1080/10630732.2011.601117. URL <http://dx.doi.org/10.1080/10630732.2011.601117>
2. Durand, N., Derivaux, S., Forestier, G., Wemmer, C., Gañarski, P., Boussaid, O., Puissant, A.: Ontology-based object recognition for remote sensing image interpretation. In: *Tools with Artificial Intelligence, 2007. ICTAI 2007. 19th IEEE International Conference on*, vol. 1, pp. 472–479. IEEE (2007)
3. Giffinger, R., Fertner, C., Kramar, H., Kalasek, R., Pichler-Milanovic, N., Meijers, E.: *Smart cities-ranking of european medium-sized cities*. Tech. rep., Vienna University of Technology (2007)

4. Hancke, G.P., Silva, B.d.C.e., Hancke Jr., G.P.: The role of advanced sensing in smart cities. *Sensors* **13**(1), 393–425 (2012). URL <http://www.mdpi.com/1424-8220/13/1/393>
5. Schober, J.P., Hermes, T., Herzog, O.: Content-based image retrieval by ontology-based object recognition. In: Proc. Workshop on Applications of Description Logics, Ulm, Germany (2004)
6. Smeulders, A.W., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **22**(12), 1349–1380 (2000)
7. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: Exploring photo collections in 3d. *ACM Trans. Graph.* **25**(3), 835–846 (2006). DOI 10.1145/1141911.1141964. URL <http://doi.acm.org/10.1145/1141911.1141964>
8. Snavely, N., Seitz, S.M., Szeliski, R.: Modeling the world from internet photo collections. *Int. J. Comput. Vision* **80**(2), 189–210 (2008). DOI 10.1007/s11263-007-0107-3. URL <http://dx.doi.org/10.1007/s11263-007-0107-3>
9. Wu, C.: Towards linear-time incremental structure from motion. In: *3D Vision - 3DV 2013, 2013 International Conference on*, pp. 127–134 (2013). DOI 10.1109/3DV.2013.25