Cross-Modal Mining of Objects and Events from Community Photo Collections

Till Quack
BiWi Computer Vision Lab
ETH Zurich
tquack@vision.ee.ethz.ch

Abstract

We describe an approach for mining images of objects (such as landmark buildings) from community photo collections in an unsupervised fashion. Our approach relies on retrieving geotagged photos from those web-sites using a grid of geospatial tiles. The downloaded photos are clustered into potentially interesting entities through a processing pipeline of several modalities, including visual, textual and spatial proximity. The resulting clusters are analyzed and are automatically classified into objects and events. Using data mining techniques, we then find text labels for these clusters, which are used to again assign each cluster to a corresponding Wikipedia article in a fully unsupervised manner. A final verification step uses the contents (including images) from the selected Wikipedia article to verify the cluster-article assignment. We demonstrate this approach on several urban areas covering over 5 million photos.

Biography

Till Quack is a post doctoral researcher at BiWi Computer Vision Lab at ETH Zurich. He obtained a M.Sc. degree in Information Technology and Electrical Engineering from ETH Zurich in 2004. From 2004 to 2009, he was research assistant and Ph.D. student at the Computer Vision Group of the Swiss Federal Institute of Technology (ETH), Zurich, Switzerland with Prof. Luc Van Gool. He received his Ph.D. in Computer Science (Dr.sc.techn.) from ETH Zurich for his work on mining and retrieval of visual data in a multimodal context in 2009. He is also co-founder and CTO of the startup kooaba in the field of mobile visual search. His recent research interests include multi-modal mining of data from community photo collections.