

MTAP special issue on methods and tools for ground truth collection in multimedia applications

Concetto Spampinato · Bastiaan J. Boom · Jiyin He

Published online: 29 January 2014
© Springer Science+Business Media New York 2014

The importance of having high quality ground truth annotations for a variety of multimedia applications is widely recognised. Indeed, one of the most time-consuming steps in methods' development is represented by the generation of accurate truth and comparing this truth to the output of applications to provide evidence that the devised methods are performing well in the targeted domain. However, the cost of creating labeled data, which implies asking a human to examine multimedia data thoroughly and provide labels, becomes impractical as datasets to be labeled grow. This can lead to the creation of disparate datasets which are often too small for both learning and evaluating the underlining data distribution. To build up large scale datasets, recently, methods exploiting the collaborative effort of a large population of users annotators (e.g. Labelme, CalTech, Pascal VOC, Trecvid) have been devised. Nevertheless, the creation of a common and large scale ground truth data to train, test and evaluate algorithms for multimedia processing is still a major concern. In particular, the research in ground truth labelling still lacks both in developing user-oriented tools and in automatic methods for supporting annotators in accomplishing their labelling tasks. In fact, tools for ground truth annotation must be user-oriented, providing visual interfaces and methods that are able to guide and speed-up the process of ground truth creation. Under this scenario, multimedia processing methods and collaborative methods play a crucial role. Further, setting up requirements and standards for the creation of multimedia dataset allows other researchers in the field to continue efforts and to contribute to the creation and annotation of multimedia data. This allows researchers to share and extend each others' work, which is beneficial for the research community.

C. Spampinato (✉)
University of Catania, Viale Andrea Doria, 6 - 95125 Catania, Italy
e-mail: cspampin@diit.unict.it

B. J. Boom
University of Edinburgh, 10 Crichton Street, EH8 9AB Edinburgh, UK
e-mail: bboom@inf.ed.ac.uk

J. He
Centre of Mathematics and Informatics, Science Park 123, 1098 XG Amsterdam, The Netherlands
e-mail: j.he@cw.nl

The special issue specifically addresses the development of multimedia processing methods for supporting automatic ground truth generation, methods and tools for combining and comparing ground truth labeled by multiple users in any field of multimedia where ground truth is required, interfaces for collecting ground truth, obtaining groundtruth by simulation and domain requirements/standardization of groundtruth data.

The paper “An Innovative Web-Based Collaborative Platform for Video Annotation” by Kavasidis et al. presents a web based tool for supporting collaborative generation of ground truth for object detection, tracking and image segmentation. The authors, moreover, introduce a new approach to combine annotations of multiple users and show how the quality of the annotations increases incrementally as more users create annotations (independently from user skills and the quality of their individual annotations).

The paper “A web-based platform for biosignal visualisation and annotation” by Lourenco et al. develops a new web-based platform for visualisation, retrieval and annotations of biosignals for non-technical users allowing them to provide ground truth labels for biomedical applications. This allows automated machine learning algorithms to use the ground truth information as input for both learning and evaluation. To evaluate the usability of the system, non-technical users were asked to perform certain task to assess the functionalities of this annotation system.

The “Ground Truth annotation of Traffic Video Data” by Mossi et al. describes an application to annotate traffic surveillance videos (although the method can be extended to other application domains). The main novelty of the method is the use of a jog shuttle wheel to navigate through the video frames which results in a substantial gain of efficiency in the video annotation task.

The paper “From Global Image Annotation to Interactive Object Segmentation” by Giro et al. deals with annotations of still images at two different scales: 1) at a *global scale*, the proposed method allows users to tag semantically images with labels taken from an ontology and 2) at a *local scale*, for interactive segmentation of objects starting from automatically segmented regions using a hierarchical partition obtained by using a Binary Partition Tree.

The paper “Robust Semi-automatic Head Pose Labeling for Real-World Face Video Sequences” by Meltem Demirkus et al. present a methodology to annotate the temporal head pose of face in real-world video sequences. In order to annotate the head pose efficiently, a semi automatic framework is used, where faces are automatically detected in a subset of the frames, afterwards the head pose of the face is labelled for those frames and using an interpolation technique the head pose of the faces in the remaining frames is accurately obtained. An head pose dataset is created, which is of interest to evaluate head pose methods that are used as preprocessing step for face detection and classification.

The paper “On the automatic online collection of training data for visual event modeling”, by Liu et al. proposes a framework to collect ground truth data for modeling visual appearance of social events by analyzing the temporal and spatial context of online media data and events. In particular, the authors proposed a ranking based approach to address the problem of finding representative negative examples, which is a more difficult task compared to the task of finding positive examples in the context of social media.

The paper entitled “A new benchmark image test suite for evaluating colour texture classification scheme ” by Porebski et al. studies two existing benchmark image test suites for

colour texture classification evaluation. They conclude that the partitioning used to build these two test sets consists of training and validation sub-images extracted from a same original image, which leads to biased classification results when combined with a classifier such as nearest neighbour. The authors then proposed a new image test suite where the training and validating sub-images come from different original images and are therefore weakly correlated.

The paper “Rendering Ground Truth Data sets to Detect Shadows Cast by Static Objects in Outdoors” by Isaza et al. focuses on creating an evaluation dataset for shadow detection methods in particular geographical locations. In order to create an evaluation dataset, the authors proposed to render virtual objects in a virtual environment using the precise longitude, latitude and elevation given an object’s GPS location, as well as the suns position for a given time and day.

The paper “Requirements for Multimedia Metadata Schemes in Surveillance Applications for Security” by van Rest et al. studies the requirements for metadata schemes in security surveillance applications. The authors proposed a terminology and used it to present these requirements. The author also show that no existing metadata schemes fulfils all requirements.

We would like to thank, first, the authors for their contribution to this special issue, then, all the reviewers for the effort and time spent to provide thorough reviews and valuable suggestions to the submitted manuscripts. Finally, we also would like to extend thanks to the Editor in Chief, Professor Borko Furth and the entire editorial staff of Multimedia Tools and Applications for recognising the importance of the subject of this special issue, namely that “Methods and Tools for Ground Truth Collection in Multimedia Applications” are necessary for researchers to train and evaluate the methodologies. We hope that the selected papers will serve as valuable reference for future investigation on this research field.



Concetto Spampinato received the Laurea (grade 110/110 cum laude) degree in computer engineering and the Ph.D. degree from the University of Catania, Catania, Italy, in 2004 and 2008, respectively, where he is currently research assistant. His research interests include mainly image and video analysis, medical image analysis and medical informatics. He has particular interest in ecological data analysis, being involved in the EU project Fish4Knowledge. He has coauthored more than 90 publications in international refereed journals and conference proceedings. As further research activities, he have organised and chaired dedicated workshops (VIGTA 2012 and 2013, MAED 2012 and 2013), several special sessions at mainstream conferences and guest-edited six special issues of international journals with impact factor.



Bastiaan J. Boom is currently Lecturer at the University of Edinburgh in the School of Informatics. In 2005 he received a Master degree from the Free University of Amsterdam in Computer Science on a thesis entitled "Fast Object Detection". This thesis was the result of a successful internship at the company PrimeVision, where he developed methods for fast detection (localisation) of license plates, faces and addresses in images. He received his PhD at the University of Twente, in the field of face recognition with special interests in face registration and illumination correction. His current research interests are domain-specific image retrieval, collection of image-based ground-truth annotations, discovering the illumination in a scene, object detection and recognition. He has been organising several scientific workshops (VAIB 2012, VIGTA 2012 and 2013) and is a guest editor for the related special issues. He has published several journals and conference articles on biometrics and computer vision.



Jiying He is a postdoctoral researcher at Centrum Wiskunde & Informatica (CWI) in Amsterdam, the Netherlands. She received her PhD degree from the University of Amsterdam in 2011 and M.Sc. degree in Artificial Intelligence at K.U. Leuven, Belgium (2006). Her research includes information retrieval, topic mining, semantic link generation, query log analysis, and user aspects of information access. She has been actively publishing and serving as reviewers and program committees at international conferences and journals in the field of information retrieval and data management since 2007.