



First International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications

Principal Author: C. Spampinato, B. Boom and J. He
Contributors: UCAT, UEDI, CWI
Dissemination: PU

Abstract: The First International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications was held as part of ACM AVI 2012, May 21-25, 2012 Capri, Italy. It specifically reports on the state of the art methods to support automatic or semi-automatic ground truth annotation and labeling in many applications such as object detection, object recognition, scene segmentation and face recognition both in still images and in videos.

1 Motivation and Description

Ground truth data plays a central role both for classifiers' training and for quantitative performance evaluation, and it has received significant attention by the the whole machine vision community. Indeed, one of the most significant efforts during the training and evaluation processes is represented by the collection of accurate truth on which algorithms will be trained and tested. However, the cost of providing labelled data, which implies asking a human to examine images and provide labels, becomes impractical and it is error prone as training sets grow.

In order to address this need, the computer vision community has directed its attention to the development of methods for collecting large scale datasets by exploiting also collaborative efforts of annotators. Although there exist functional, task-oriented requirements for tools supporting ground truth labeling, the research still lacks in user-oriented tools. In fact, annotators, must be at the centre of such tools since it has been estimated that labeling an image may take from two to thirty minutes, depending on the task, and it is, obviously, even worse in the case of videos.

The aim of the First International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications¹ was to report on tools, interfaces and methods able to speed-up the ground truth creation process by supporting users, through the integration of computer vision methods, to build up a reliable truth in a reasonable amount of time.

The call for papers attracted 17 papers (10 different countries) from which the program committee selected 6 for oral presentations, 6 for poster presentations and 1 for demo presentation. In addition, it featured a keynote talk delivered by Dr. Benoit Huet from the Multimedia Communications Department of EURECOM on “Multimedia Data Collection using Social Media Analysis”.

The workshop proceedings appear in the ACM International Conference Proceeding Series published by ACM. Selected and revised papers presented at the workshop are under review for publication on the special issue “Methods and Tools for Ground Truth Collection in Multimedia Applications” of the Multimedia Tools and Applications (MTAP) Journal, Springer with C. Spampinato, B. Boom and J. He as guest editors.

Regarding the paper contents, we received contributions on:

- Automatic methods (e.g. clustering or combinatorial approaches) for the generation of large scale ground truth starting from small datasets [1, 2];
- User-oriented tools supporting annotators mainly in the task of object detection, recognition, face detection and image segmentation in still images and in video streams [3, 4, 5, 6, 7, 8];
- Tools for supporting medical image annotation [9, 10];
- Integration of computer vision methods and ad-hoc hardware for video annotation [11, 12].

A detailed description of these papers can be found in the workshop proceedings [13]

¹<http://vigta2012.dieei.unict.it/>

References

- [1] Julia Moehrmann and Gunther Heidemann. Efficient annotation of image data sets for computer vision applications. In *Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications*, VIGTA '12, pages 2:1–2:6, New York, NY, USA, 2012. ACM.
- [2] Akhil R. Shah and Siddhartha R. Dalal. Combinatorial enlargement of ground-truth datasets and efficient evaluation of segmentation algorithms. In *Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications*, VIGTA '12, pages 12:1–12:4, New York, NY, USA, 2012. ACM.
- [3] Cesar Isaza, Joaquin Salas, and Bogdan Raducanu. Synthetic ground truth dataset to detect shadows cast by static objects in outdoors. In *Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications*, VIGTA '12, pages 11:1–11:6, New York, NY, USA, 2012. ACM.
- [4] Benjamin Barbour and Karl Ricanek, Jr. An interactive tool for extremely dense landmarking of faces. In *Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications*, VIGTA '12, pages 13:1–13:5, New York, NY, USA, 2012. ACM.
- [5] Ákos Utasi and Csaba Benedek. A multi-view annotation tool for people detection evaluation. In *Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications*, VIGTA '12, pages 3:1–3:6, New York, NY, USA, 2012. ACM.
- [6] Ozan Şener, Kemal Uğur, and A. Aydin Alatan. Robust interactive segmentation via coloring. In *Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications*, VIGTA '12, pages 9:1–9:4, New York, NY, USA, 2012. ACM.
- [7] Boguslaw Cyganek and Katarzyna Socha. A multi-tool for ground-truth stereo correspondence, object outlining and points-of-interest selection. In *Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications*, VIGTA '12, pages 4:1–4:4, New York, NY, USA, 2012. ACM.
- [8] I. Kavasidis, S. Palazzo, R. Di Salvo, D. Giordano, and C. Spampinato. A semi-automatic tool for detection and tracking ground truth generation in videos. In *VIGTA '12: Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications*, pages 1–5, New York, NY, USA, 2012. ACM.
- [9] P. M. Ferreira, T. Mendonça, J. Rozeira, and P. Rocha. An annotation tool for dermoscopic image segmentation. In *Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications*, VIGTA '12, pages 5:1–5:6, New York, NY, USA, 2012. ACM.
- [10] Valentina Pedoia, Alessandro De Benedictis, Giuseppe Renis, Emanuele Monti, Sergio Balbi, and Elisabetta Binaghi. Manual labeling strategy for ground truth estimation in mri

- glial tumor segmentation. In *Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications*, VIGTA '12, pages 8:1–8:4, New York, NY, USA, 2012. ACM.
- [11] Tobias Zimmermann, Markus Weber, Marcus Liwicki, and Didier Stricker. Covida: pen-based collaborative video annotation. In *Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications*, VIGTA '12, pages 10:1–10:6, New York, NY, USA, 2012. ACM.
- [12] Jose M. Mossi, Antonio Albiol, and Alberto Albiol. Efficient annotation of traffic video data. In *Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications*, VIGTA '12, pages 7:1–7:2, New York, NY, USA, 2012. ACM.
- [13] Concetto Spampinato, Bas Boom, and Jiyin He, editors. *VIGTA '12: Proceedings of the 1st International Workshop on Visual Interfaces for Ground Truth Collection in Computer Vision Applications*, New York, NY, USA, 2012. ACM.