



Fish4Knowledge Deliverable D6.4

International scientific workshop 3

Principal Author: UCATANIA
Contributors: All partners
Dissemination: PU

Abstract: This document reports the delivery of the third international workshop for the Fish4Knowledge project, namely, the International Workshop on Video and Image Ground Truth in computer vision Applications (VIGTA'13).

Deliverable due: Month 36

1 The workshop concept

The Fish4Knowledge team organised a one day workshop on “Video and Image Ground Truth in computer vision Applications”, held on July 15, 2013 in conjunction with the 9th International Conference on Computer Vision Systems (ICVS 2013), July 16-18, St. Petersburg, Russia. The proposal was accepted by the ICVS conference organising committee. The call for papers included the following motivation:

“In the development of computer vision applications, a fundamental role is played by the availability of large datasets of annotated images and videos (ground truth) providing a wide coverage of different scenarios and environments. These are used both to train machine-learning approaches, which have been largely and successfully adopted for computer vision, but still strongly suffer the lack of comprehensive, large-scale training data, and to evaluate algorithms’ performance, which has to provide enough evidence, to the developers and especially to peer scientists reviewing the work, that a method works well in the targeted environment and conditions.

The main limitation to collect large scale ground truth is the daunting amount of time and human effort needed to generate high quality ground truth; in fact, it has been estimated that labeling an image may take from two to thirty minutes, depending on the task, and this is, obviously, even worse in the case of videos. Currently, most available datasets with the related ground truth are produced as the result of efforts of single research groups who have manually annotated such datasets, which, however, are too task-oriented and cannot be generalized.

Moreover, the large-scale ground truth gathering approaches, which have been experimented so far, suffer from many limitations, from incomplete or low-quality annotations (due to the lack of quality control) to interoperability issues, since no common representation schema has been adopted yet.

In addition, it is not always trivial to identify metrics for performance evaluation. A notable case is object tracking, for which some research groups have developed self-evaluation-based approaches. Therefore, the availability of massive ground truth would allow the development of such methods and make them in the long run independent of ground truth; this would be inline with the current wave of scientific development, which is “data-driven” in contrast to theory or simulation driven.

The aim of this workshop is to present and report on the most recent methods to support automatic or semi-automatic ground truth annotation and labeling as well as algorithms’ performance evaluation and comparison in many applications such as object detection, object recognition, scene segmentation and face recognition both in still images and in videos.

More specifically, the workshop will bring together researchers in computer vision, machine learning and semantic web to share and collect ideas with the aim to allow researchers to model and keep track of the whole research process, from dataset construction to performance evaluation.”

The research topics of the workshop included:

- Computer vision and machine learning methods for supporting human in the task of generating new ground truth more efficiently
- Computer vision and machine learning methods to combine annotations in the form of both textual labels and graphical items

- Framework for sharing of dataset, ground truth, features, algorithms and tools
- Web Semantic approaches for ground truth representation, sharing and harvesting
- Ontologies and vocabularies describing annotations, video and image data, algorithms' capabilities and performance
- Semantic Interactive and Collaborative Video Annotation tools
- Crowdsourcing and quality control mechanisms to generate high quality ground truth
- Comparative analysis of existing tools
- Methods for performance evaluation without ground truth data
- Tools and applications

The workshop organisers were:

Concetto Spampinato	University of Catania	Italy
Baas Boom.	University of Edinburgh	UK
Benoit Huet	Eurecom	France

Additional program committee members were:

Daniela Giordano	University of Catania (Italy)
Gabriella Sanniti di Baja	CNR (Italy)
Giovanni Farinella	University of Catania (Italy)
Guillaume Gravier	IRISA (France)
Isaak Kavasidis	University of Catania (Italy)
Jenny Benois-Pineau	University Bordeaux (France)
Lucia Ballerini	University of Edinburgh (UK)
Margrit Betke	Boston University (USA)
Monique Thonnat	INRIA (France)
Sebastiano Battiato	University of Catania (Italy)
Simona Ullo	IIT Genova (Italy)
Simone Palazzo	University of Catania (Italy)
Sotirios Tsaftaris	IMT Lucca (Italy)
Stefanos Vrochidis	Centre for Research and Technology Hellas (Greece)
Subramanian Ramamoorthy	University of Edinburgh (UK)
Vasileios Mezaris	Centre for Research and Technology Hellas (Greece)
Xuan Huang	University of Edinburgh (UK)
Xueliang Liu	EURECOM (France)
Yu Wang	Rensselaer Polytechnic Institute (New York)
Zheng-Jun Zha	NUS (Singapore)
Zuoguan Wang	3M (USA)

2 The workshop delivery

- The papers to be presented at the workshop were selected based on 8 page full papers formatted according to ACM template¹.
- The workshop proceedings are published by the ACM International Conference Proceedings Series, under the ISBN 978-1-4503-2169-3 and are available at <http://dl.acm.org/citation.cfm?id=2501105>.
- The workshop was advertised to about 2000 computer vision/image analysis/multimedia researchers by direct email plus several newsgroups and the main conference web site.
- 15 papers were received and each paper was reviewed by 3 members of the organisers and programme committee: 8 papers were accepted for oral presentation at the workshop.
- The workshop featured also two keynote talks: “Overview of Quality Assessment for Visual Signals and Newly Emerged Trends” given by Prof. Ngan, King Ngi from Chinese University of Hong Kong; and “Ground Truth Design Principles – An Overview” given by Dr. Kondermann from Heidelberg Collaboratory for Image Processing (HCI) of University of Heidelberg.
- 35 people attended the workshop out of a total of 50 people present at the conference (there were few tutorials running in parallel) that day.
- The authors of the papers accepted and presented at the workshop have been invited to submit a significant extension of their paper to the special issue “Large Scale Data-Driven Evaluation in Computer Vision” of the Computer Vision and Image Understanding journal.

¹<http://www.acm.org/sigs/publications/proceedings-templates>

3 The workshop programme

The programme was:

- 8:45 – 9:00 **Opening**
- 9:00 – 10:00 **Keynote Talk**
- Overview of Quality Assessment for Visual Signals and Newly Emerged Trends
Ngan, King Ngi
- 10:00 – 10:30 **Session 1**
- Semi-Supervised Evaluation of Face Recognition in Videos
V. Biaud, V. Despiegel, C. Herold, O. Beiler, S. Gentric
- 10:30–11:00 **Break**
- 11:00 – 12:30 **Session 2**
- Human Nonverbal Behavior Multi-sourced Ontological Annotation
B. Knyazev
 - Tell me more: How Semantic technologies Can Help Refining Internet Image Search
F. Setti, D. Porello, R. Ferrario, S. Abdulhak, M. Cristani
 - Approximate Nearest Neighbor Search to support Manual Image Annotation of large domain-specific datasets
B. Boom, P. X. Huang, R. Fisher
- 12:30–14:00 **Lunch**
- 14:00 – 15:00 **Keynote Talk**
- Ground Truth Design Principles – An Overview
D. Kondermann
- 15:00 – 15:30 **Session 3**
- Automatic Regions of Interest Selection by Image Matting Fusing Depth and Color Data Streams
L. Cinque, P. Dondi, L. Lombardi
- 15:30–16:00 **Break**
- 16:00 – 17:30 **Session 4**
- Comparative Study of Segmentation Methods for Tree Leaves Extraction
M. Grand-brochier, A. Vacavant, G. Cerutti, K. Bianchi, L. Tougne
 - A Crowdsourcing Approach to Support Video Annotation
R. Di Salvo, D. Giordano, I. Kavasidis
 - Evaluation of Manually Created Ground Truth for Multi-view People Localization
A. Kiss, T. Sziranyi
- 17:30 – 17:45 **Discussion and Conclusion**