V&L Net 2014

The 3rd Annual Meeting Of The EPSRC Network On Vision & Language
and
The 1st Technical Meeting of the European Network on Integrating Vision and Language

A Workshop of the 25th International Conference on Computational Linguistics (COLING 2014)

Proceedings

August 23, 2014
Dublin, Ireland

ISBN 978-1-873769-28-1

This workshop is partly supported by ICT COST Action IC1307, the European Network on Integrating Vision and Language (iV&L Net): Combining Computer Vision and Language Processing For Advanced Search, Retrieval, Annotation and Description of Visual Data, and partly by the EPSRC Network on Vision and Language (V&L Net).

ESF provides the COST Office through an EC contract

COST is supported by the EU RTD Framework Programme

Preface

The Workshop on Vision and Language 2014 (VL'14) took place in Dublin on 23rd July 2014, as part of COLING'14. It was the joint 3rd meeting of the EPSRC Network On Vision and Language and 1st technical meeting of the new European Network on Integrating Vision and Language which is funded as a European COST Action. The VL workshops have the general aims:

1. to provide a forum for reporting and discussing planned, ongoing and completed research that involves both language and vision; and
2. to enable NLP and computer vision researchers to meet, exchange ideas, expertise and technology, and form new research partnerships.

As funding for the V&L EPSRC Network (EP/H018557) ends and funding for the iV&L Net European COST Action (IC1307) starts, the focus of the VL workshops will shift onto integration and joint modelling of language and vision. iV&L Net will take over the organisation of annual VL workshops for the next four years as the flagship workshop of this new COST Action.

The call for papers for VL'14 was issued in May 2014 and elicited a good number of high-quality submissions, each of which was peer-reviewed by three members of the programme committee. The interest in the workshop from leading NLP and computer vision researchers and the quality of submissions was high, so we aimed to be as inclusive as possible within the practical constraints of the workshop. In the end we accepted 14 submissions as long papers, and eight as short papers.

The resulting workshop programme packed a lot of exciting content into one day. We were delighted to be able to include in the programme a keynote presentation by Alex Jaimes of Yahoo! Inc., an internationally leading vision researcher. Our technical programme combined seven oral papers, seven long poster papers and seven short poster papers. Some thematic clusters emerged: combined text and image processing (Nguyen et al., Sakaki et al., Jones et al., Zhang et al., HaCohen-Kerner et al.), image description, annotation and labelling (Elliott, Liparas et al., Wang et al., Jokinen and Wilcock), data set creation (Weiland et al., Le et al., McGuinness et al.), situated dialogue (Summers-Stay et al., Schütte et al.), video analysis (Bhat and Olszewska, Shrestha et al.), aids for visually impaired people (Safi et al., Belz and Bharath), and visual analysis supported by text/speech features (Anbarjafari and Aabloo). The programme also included a discussion session on future directions for the VL community and workshops, including plans for shared task competitions.

We would like to thank all the people who have contributed to the organisation and delivery of this workshop: the authors who submitted such high quality papers; the programme committee for their prompt and effective reviewing; our keynote speaker, Alex Jaimes; the COLING 2014 organising committee, especially the workshops chairs, Jennifer Foster, Dan Gildea, and Tim Baldwin; the participants in the workshop; and future readers of these proceedings for your shared interest in this exciting new area of research.

August 2014 *Anja Belz, Marie-Francine Moens and Alan F. Smeaton*

Organising Committee

Anja Belz, University of Brighton
Darren Cosker, University of Bath
Frank Keller, University of Edinburgh
Marie-Francine Moens, University of Leuven
Alan F. Smeaton, Dublin City University
William Smith, University of York

Program Committee:

Yannis Aloimonos, University of Maryland, US
Tamara Berg, Stony Brook, US
Desmond Elliot, University of Edinburgh, UK
Erkut Erdem, Hacettepe University, Turkey
Sergio Escalera, Autonomous University of Barcelona, Spain
Claire Gardent, CNRS/LORIA, France
Jordi Gonzales, Universita Autonoma de Barcelona, Spain
Lewis Griffin, UCL, UK
Julia Hockenmaier, University of Illinois, US
John Kelleher, Dublin Institute of Technology, Ireland
Brian Mac Namee, Dublin Institute of Technology, Ireland
Dimitrios Makris, Kingston University, UK
Margaret Mitchell, University of Aberdeen, UK
Ray Mooney, University of Texas at Austin, US
Lucia Specia, University of Sheffield, UK
Chris Town, University of Cambridge, UK
Isabel Trancoso, INESC-ID, Portugal
David Windridge, University of Surrey, UK

Invited Keynote Speaker:

Alex Jaimes, Yahoo! Inc.

Table of Contents

Conference Program

Saturday, 23 August, 2014

(09.00 - 09.15) Introduction and Welcome to Workshop

(09.15 - 10.30) Interaction

The Effect of Sensor Errors in Situated Human-Computer Dialogue
Niels Schütte, John Kelleher and Brian Mac Namee

Joint Navigation in Commander/Robot Teams: Dialog & Task Performance When Vision is Bandwidth-Limited
Douglas Summers-Stay, Taylor Cassidy and Clare Voss

TUHOI: Trento Universal Human Object Interaction Dataset
Dieu-Thu Le, Jasper Uijlings and Raffaella Bernardi

(10.30 - 11.00) Morning Coffee

(11.00 - 11.40) Invited Keynote Talk - Alex Jaimes, Yahoo ! Inc.

(11.40 - 12.30) Language Descriptors

Concept-oriented labelling of patent images based on Random Forests and proximity-driven generation of synthetic data
Dimitris Liparas, Anastasia Moumtzidou, Stefanos Vrochidis and Ioannis Kompatsiaris

Exploration of functional semantics of prepositions from corpora of descriptions of visual scenes
Simon Dobnik and John Kelleher

Saturday, 23 August, 2014 (continued)

(12.30 - 13.30) Lunch

(13.30 - 14.20) Visual Indexing

A Poodle or a Dog? Evaluating Automatic Image Annotation Using Human Descriptions at Different Levels of Granularity
Josiah Wang, Fei Yan, Ahmet Aker and Robert Gaizauskas

Key Event Detection in Video using ASR and Visual Data
Niraj Shrestha, Aparna N. Venkitasubramanian and Marie-Francine Moens

(14.20 - 15.00) Poster Boasters

(15.30 - 17.00) Long Poster Papers (Parallel session)

Twitter User Gender Inference Using Combined Analysis of Text and Image Processing
Shigeyuki Sakaki, Yasuhide Miura, Xiaojun Ma, Keigo Hattori and Tomoko Ohkuma

Semantic and geometric enrichment of 3D geo-spatial models with captioned photos and labelled illustrations
Chris Jones, Paul Rosin and Jonathan Slade

Weakly supervised construction of a repository of iconic images
Lydia Weiland, Wolfgang Effelsberg and Simone Paolo Ponzetto

Cross-media Cross-genre Information Ranking based on Multi-media Information Networks
Tongtao Zhang, Haibo Li, Hongzhao Huang, Heng Ji, Min-Hsuan Tsai, Shen-Fu Tsai and Thomas Huang

Speech-accompanying gestures in Russian: functions and verbal context
Yulia Nikolaeva

DALES: Automated Tool for Detection, Annotation, Labelling, and Segmentation of Multiple Objects in Multi-Camera Video Streams
Mohammad Bhat and Joanna Isabelle Olszewska

A Hybrid Segmentation of Web Pages for Vibro-Tactile Access on Touch-Screen Devices
Waseem SAFI, Fabrice Maurel, Jean-Marc Routoure, Pierre Beust and Gaël Dias

Saturday, 23 August, 2014 (continued)

(15.30 - 17.00) Short Poster Papers (Parallel session)

Expression Recognition by Using Facial and Vocal Expressions
Gholamreza Anbarjafari and Alvo Aabloo

Formulating Queries for Collecting Training Examples in Visual Concept Classification
Kevin McGuinness, Feiyan Hu, Rami Albatal and Alan Smeaton

Towards Succinct and Relevant Image Descriptions
Desmond Elliott

Coloring Objects: Adjective-Noun Visual Semantic Compositionality
Dat Tien Nguyen, Angeliki Lazaridou and Raffaella Bernardi

Multi-layered Image Representation for Image Interpretation
Marina Ivasic-Kos, Miran Pobar and Ivo Ipsic

The Last 10 Metres: Using Visual Analysis and Verbal Communication in Guiding Visually Impaired Smartphone Users to Entrances
Anja Belz and Anil Bharath

Keyphrase Extraction using Textual and Visual Features
Yaakov HaCohen-Kerner, Stefanos Vrochidis, Dimitris Liparas, Anastasia Moumtzidou and Ioannis Kompatsiaris

Towards automatic annotation of communicative gesturing
Kristiina Jokinen and Graham Wilcock

Saturday, 23 August, 2014 (continued)

(17.00 - 17.30) Discussion and Closing