Neurocomputing 161 (2015) 1–2

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Editorial

Recognition and action for scene understanding

1. Motivation

Perceiving and understanding scenes and human behaviors is a key technology for smart environments, human–machine interaction and robotics and all are growing research fields with many future applications. Surveillance of populated environments, recognition of human activities and intentions, tracking of pedestrians in urban areas or detection of intruders are examples of tasks that rely on the ability to robustly detect abnormal. There is a great demand for even more robust systems, especially over a wider range of conditions for indoor and outdoor environments. There is also an increasing interest from industry for small-scale, low cost and robust surveillance systems.

Scene and human behaviour analysis and understanding have become popular topics in Computer Science, which combines issues such as attention in cognitive systems, object detection and recognition, global scene recognition and human sensing. Hence, it involves joining efforts and sharing knowledge from different research areas such as Computer Vision, Pattern Recognition, Machine Intelligence, Software Engineering and Cognitive Sciences. Besides, it is closely related to several emerging application areas (e.g., smart environments, video surveillance, visual-based mobile robot navigation, wearable sensors, and the detection and understanding of events and activities in video data). This Special Issue constitutes a collection of original works on these topics, mainly focused on the artificial understanding of both human actions and everyday scenarios, but also covering other domains such as video-text understanding or curve reconstruction.

2. Special issue overview

In an envisaged future in which we live in smart homes and intelligent robots populate our world we imagine an everyday phone conversation with our home ambient intelligent vision system (or robot), asking it "Is everything fine?" which the system happily confirms positively. However, this confirmation is provided with some degree of uncertainty because it is a fact partially observable by a machine and it requires cognitive interpretation. For an artificial intelligent agent reach a more reliable conclusion the agent needs to be able to observe its environment and interpret the information perceived in a cognitive manner. The paper by Zoe Falomir and Ana-Maria Olteteanu proposes an approach for scene understanding based on qualitative descriptors, domain knowledge and logics. In this paper, qualitative image descriptions are combined with domain knowledge and feature detectors for improving the categorization of objects (i.e. target objects or unknown objects with surfaces without textures). Moreover, semantics are provided to target objects for describing their affordances, mobility and other functional properties of objects. Logics have been also defined for reasoning about the provided semantics and interpreting scenes for detecting normal and abnormal/threatening situations. Tests were carried out at the *Interact@Cartesium* scenario and promising results were obtained. Focused on the artificial perception and understanding of the human as an interactive partner, cognition is also the key concept in the work by M. Glodek et al. Thus, this work presents the information fusion principles employed by cognitive approaches to human–computer interaction. Combination of information generally goes along with the level of abstractness, time granularity and robustness, such that large cognitive architectures must perform fusion gradually on different levels – starting from sensor-based recognitions to highly abstract logical inferences. In their application, the authors divide up information fusion approaches into three categories: perception-level fusion, knowledge-based fusion and application-level fusion. For each category, they introduce examples of characteristic algorithms and provide a detailed protocol on the implementation performed in order to study the interplay of the developed algorithms.

Human action recognition is typically a challenging process for scene understanding. Modeling a vocabulary of local image features in a bag of visual words (BoW) is a common approach to extract the components of an action video. In order to extract intrinsic shape bases and to consider temporal structure of an action, F. Moayedi et al. propose to take the advantages of group sparse coding methods. The main contribution of this work is to explore the geometry of action components via structured sparse coefficients of visual words in a real-time manner. In comparison with the conventional BoW models, this approach has other advantages including much less quantization error, higher level feature representation which leads to reduction in model parameters and memory complexity while considering temporal structure. The method is evaluated on standard human action datasets including KTH, Weismann, UCF-sports and UCF50 human action datasets. The experimental results are significantly improved in comparison with previously presented results methods. The paper by A. Iosifidis et al. focuses on the problem of distance-based, multi-class classification of human actions and specifically on the Nearest Class Centroid (NCC) classification scheme. Specifically, the authors introduce a new metric learning approach based on logistic discrimination for the determination of a low-dimensional feature space of increased discrimination power. Experimental results denote that the performance of the proposed distance-based classification schemes is even better to that of Support Vector Machine classifier, which is currently the standard choice for human action recognition. On the other hand, there are numerous instances in which, in addition to the direct observation of a human body in motion, the characteristics of related objects can also contribute to the identification of human actions. The aim of the paper by Nouf Al Harbi and Yoshihiko Gotoh is to address this issue and suggest a multi-feature method of determining human actions. This study demonstrates that region descriptors can be

http://dx.doi.org/10.1016/j.neucom.2015.02.054

attained for the action classification task. A cutting-edge human detection method is applied to generate a model incorporating generic object foreground segments. These segments have been extended to include non-human objects, which interact with a human in a video scene to capture the action semantically. Experiments using KTH, UCF sports and Hollywood2 dataset show that the approach achieves the state-of-the-art performance.

But scene understanding is not only related to objects or humans. For instance, scene text extraction in images and videos is of prime importance for automated scene understanding and video content analysis. The paper by Joanna I. Olszewska describes a new optical character recognition approach, which allows real-time, automatic extraction and recognition of digits in images and videos. The system shows excellent results when applied to the automated identification of team players' numbers in sport datasets, outperforming related state-of-the-art methods. The paper by Kavita Khanna and Navin Rajpal presents an approach for reconstruction of curves from an unorganized dense point cloud with noise. The method is based on the concepts of fuzzy logic, fuzzy clustering and the traveling salesman path generated by ant colony optimization. The experimental results show that the performance of the proposed algorithm is good, as the reconstructed curve resembles the original curve.

The availability of affordable RGB-D sensors, significant effort has been recently put into RGBD scene understanding. Such 3D representations are able to incorporate physical information, such as the 3D volume of the objects, supporting relations and stability.

The paper by R. Marfil, E. Antunez and A. Bandera presents a novel method for representing RGB-D images into a hierarchy of layers of abstraction. At the higher layers, the image is encoded with parametric models, a topic which is playing a central role in fields such as scene understanding or robotic grasping. The algorithm incorporates a model of the sensor into the processing, where the range image is described. Then, it combines a region growing and a model-based strategy inside the framework of the combinatorial pyramid. The approach is able of distinguishing between different types of surfaces while segmenting the range image. The implementation on parallel machines that can exploit one of the main features of the approach and the possibility to built the layers of the hierarchy using local kernels are considered topics to explore in the future.

The list of reviewers of this special issue

Jean Christophe Nebel – Kingston University, UK
David Filliat – ENSTA ParisTech, France
Andrea Torsello – Ca' Foscari University of Venice, Italy
Moulay A. Akhloufi – Center of Industrial Robotics and Vision, Canada
Manuel J. Marín Jiménez, University of Córdoba, Spain
Juan P. Bandera – University of Málaga, Spain
Raquel Viciana – University of Jaén, Spain
Pedro Núñez – University of Extremadura, Spain
Ricardo Vázquez Martín, CITIC, Spain
Luis Santos – Institute of Systems and Robotics, University of Coimbra, Portugal
Vittorio Murino – University of Verona, Italy
Cristiano Premebida – Institute of Systems and Robotics, University of Coimbra, Portugal
Anthony R. Dick – University of Adelaide, Australia
Pierluigi Casale – IMEC, Netherlands
Antonio Bandera – University of Málaga, Spain

Table of contents

Acknowledgments

We wish to thank to all the people that enabled the publication of this special issue. First of all, we wish to thank Doctor Tom Heskes the Editor-in-chief of this journal, for accepting the idea and for his support, patience and motivation. Our gratitude also goes to the Journal Manager and to all the staff from Elsevier, for the impeccable and timely logistical support. The papers in this issue were invited based on an open call for submission. They typically went through two rounds of reviews. We wish to equally thank the authors and the reviewers for all their hard work and contribution for the excellence of this special issue.

Rebeca Marfil
Grupo ISIS, Department of Electronic Technology, University of Málaga, Spain
E-mail address: rebeca@uma.es

Jorge Dias
University of Coimbra, Portugal & *Khalifa University, Abu Dhabi, United Arab Emirates*
E-mail addresses: jorge@deec.uc.pt, jorge.dias@kustar.ac.ae

Francisco Escolano
University of Alicante, Spain
E-mail address: sco@dccia.ua.es

Received 16 February 2015; accepted 16 February 2015

Available online 23 February 2015