

# Guest Editorial: Analysis and Retrieval of Events/Actions and Workflows in Video Streams

Anastasios Doulamis <sup>1</sup> & Nikolaos Doulamis <sup>1</sup> &  
Marco Bertini <sup>2</sup> & Jordi Gonzalez <sup>3</sup> & Thomas Moeslund <sup>4</sup>

---

\* Anastasios Doulamis  
adoulam@cs.ntua.gr

Nikolaos Doulamis  
ndoulam@cs.ntua.gr

Marco Bertini  
marco.bertini@gmail.com

Jordi Gonzalez  
poal@cvc.uab.es

Thomas Moeslund  
tbm@create.aau.dk

<sup>1</sup> National Technical University of Athens, Athens, Greece

<sup>2</sup> Dipartimento di Ingegneria dell'Informazione, Florence University, Florence, Italy

<sup>3</sup> Departamento de Ciencias de la Computación, Autonomous University of Barcelona, Barcelona, Spain

<sup>4</sup> Department of Architecture, Design and Media Technology, Aalborg University, Aalborg, Denmark

Cognitive video supervision and event analysis in video sequences is a critical task in many multimedia applications. Methods, tools, and algorithms that aim to detect and recognize high-level concepts and their respective spatiotemporal and causal relations in order to identify semantic video activities, actions, and procedures have been in the focus of the research community over the last years.

This research area has strong impact on many real-life applications such as service quality assurance, compliance to the designed procedures in industrial plants, surveillance of people-dense areas (e.g., thematic parks, critical public infrastructures), crisis management in public service areas (e.g., train stations, airports), security (detection of abnormal behaviors in surveillance videos), semantic characterization, and annotation of video streams in various domains (e.g., broadcast or user-generated videos). For instance, the dynamic capture of situational awareness concerning crowds in specific mass gathering venues and its intelligent enablement into emergency management information systems, using smart communication devices and spaces, is critical for achieving rapid, timely guidance and safe evacuation of people out of dangerous areas. This is proven under the research outcomes of eVacuate Project [2]. Additionally, event analysis is critical in tunnel inspection applications in which image analysis tools and computer vision algorithms can be applied to identify cracks and other defects within such critical structural engineering infrastructures [1].

The traditional approaches for event detection in videos assume well-structured environments, and they fail to operate in largely unsupervised way under adverse and uncertain conditions from those in which they have been trained. Another drawback of current methods is the fact that they focus on narrow domains using specific concept detectors such as “human faces, cars, and buildings”.

This issue of *Multimedia Tools and Applications* (MTAP) journal covers research aspects from the aforementioned application areas. The issue is mainly focused on visual event analysis and understanding with applications on visual retrieval topics. A call for papers was published on the MTAP journal webpage. In parallel, the Guest Editors disseminated the event in several groups and societies researching on this topic.

The call attracted many submissions all over the world, from Europe, America (mainly from the USA and Canada), and Asia. All articles underwent a peer review process for at least three reviewers, while most of them received four or even five reviews. All non-rejected articles underwent revision to meet the reviewers’ and editors’ comments, while some of them were sent for a second or even a third revision at least all comments were carefully addressed. Failure to meet these tough conditions led to rejection. After completion of this hard process, only six papers were finally accepted for publication in MTAP.

In particular, the work of G. Marin et al. “Hand gesture recognition with jointly calibrated Leap Motion and depth sensor” (10.1007/s11042-015-2451-6) proposes a novel hand gesture recognition system through incorporation of novel low-cost depth camera devices such as Kinect™ and leap motion. Initially, a set of novel features is introduced while various distance metrics are examined. The classification scheme is based on Random Forests and multiclass Support Vector Machine.

The second article is entitled “3D measures exploitation for a monocular semi-supervised fall detection system” (10.1007/s11042-015-2513-9) and deals with the development of a novel system able to categorize falls and non-falls within complex human activities. The system is supported by semi-supervised learning schemes and is capable of discriminating humans’ actions in real time. Again, the work of Makantasis et al. “Semi-supervised vision-based maritime surveillance system using fused visual attention maps” (10.1007/s11042-015-2512-x) combines semi-supervised learning, visual attention maps, and foreground/background separation tools to

discriminate ships, boats, and other objects from the sea background. The developed algorithms have been tested on long-time maritime surveillance video sequences recorded in Limassol, Cyprus, and Chania, Greece, ports.

The work of Xiao-jun Chen et al. “Complex video event detection via pairwise fusion of trajectory and multi-label hypergraphs” (10.1007/s11042-015-2514-8) introduces a multi-label hypergraph method for complex event detection via pairwise fusion. Based on the hypergraph theory, this paper proposes to construct trajectory and multi-label hypergraphs considering the features of moving targets. The two hypergraphs are fused to detect complex events. Anomaly event detection in crowded scenes is described in the work of Chen et al. “An efficient subsequence search for video anomaly detection and localization” (10.1007/s11042-015-2453-4). For anomaly detection, one-class support vector machine with Bayesian derivation, is applied to detect unusual events. The work supports the maximum subsequence search that integrates local anomaly scores into a global consistent detection. The experiments have been conducted on public datasets and it reaches 80 % localization rate which approximately doubles the accuracy of recent works.

Finally, the last two papers refer to multimedia content search and summarization, “An automatic event-complementing human life summarization scheme based on a social computing method over social media content” (10.1007/s11042-015-2454-3). This paper proposes a human life summarization scheme based on multimedia content published on social media. In this context, the term “life” includes the events, occasions, and activities users post on their walls. The summarization module produces a meaningful video clip that includes the top moments of one’s life without completely disregarding the less important.

The Guest Editors express their thanks to the Editor-in-Chief of *Multimedia Tools and Applications* (MTAP) journal, Prof. Borko Furht, for supporting this special issue. The Guest Editors would also like to give special thanks to all the anonymous reviewers for their excellent work in reading and judging on the quality of the papers.

## References

1. Makantasis K, Protopapadakis E, Doulamis A, Doulamis N, Loupos C (2015) Deep convolutional neural networks for efficient vision based tunnel inspection. Proceedings—2015 I.E. 11th International Conference on Intelligent Computer Communication and Processing, ICCP 2015, art. no. 7312681, pp 335–342
2. Sabeur Z, Doulamis N, Middleton L, Arbab-Zavar B, Correndo G, Amditis A (2015) Multi-modal computer vision for the detection of multi-scale crowd physical motions and behavior in confined spaces. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9474, pp 162–173



Anastasios D. Doulamis - Technical University of Crete, Greece (adoulam@cs.ntua.gr) Prof. Anastasios Doulamis received the Diploma degree in Electrical and Computer Engineering from the National Technical University of Athens (NTUA) in 1995 with the highest honor (first ranked among all classmates) and the PhD degree in Electrical and Computer Engineering from NTUA in 2000. He is currently Assistant Professor at the National Technical University of Athens (NTUA). Prof. Anastasios Doulamis has received several awards and prizes during his studies, including the Best Greek Student in all fields of engineering in national level in 1995, the Best Graduate Thesis Award in the area of Electrical Engineering in 1996 and several prizes from the National Technical University of Athens, the National Scholarship Foundation and the Technical Chamber of Greece. In 1997, he was given the NTUA Medal as Best Young Engineer. In 2000, he received the best Phd thesis award by the Thomaïdion Foundation. He is author of more 200 papers in the area of multimedia processing and artificial intelligence among them more than 20 in IEEE/ACM journals papers and more than 60 journal papers. He has also more than 2500 citations in the respective field. He has served as Area Chair of the IEEE International Conference on Pattern Recognition (ICPR), organizer of ECCV, ICCV, and ACM Multimedia workshops and in the technical committee of more than 20 international workshops/conferences. Prof. Anastasios Doulamis is the Coordinator of the EU funded projects TERSPICHORE and 4D CH World. He is involved in many other European research projects, such as Robospect, Rangers, WaterSpy, EXPERIMEDIA, +Spaces, SOCIOS, in the area of multimedia processing, computer vision and artificial intelligence.



Nikolaos D. Doulamis – National Technical University of Athens, Greece (ndoulam@cs.ntua.gr) Prof. Doulamis received the Diploma degree in Electrical and Computer Engineering from the National Technical University of Athens (NTUA) with the highest honour (first ranked among all classmates) and the PhD degree in electrical and computer engineering from NTUA. His PhD thesis was supported by the Bodosakis Foundation Scholarship. He is currently Assistant professor at the National Technical University of Athens. Dr. Doulamis was awarded as the Best Greek Student in the field of engineering at the national level by the Technical chamber of Greece. He was awarded the Best Graduate Thesis Award in the area of electrical engineering. During his studies he has also received several prizes and awards from the National Technical University of Athens, the National Scholarship Foundation and the Technical Chamber of Greece. He was given the NTUA Medal as Best Young Engineer. He has also served as program committee in several international conferences and workshops. He was given the Thomaïdion Foundation best journal paper award. Prof. Nikolaos D. Doulamis is an Author of more than 60 journals papers, more than 25 book chapters and more than 200 conference papers. Among the 60 journals papers 20 papers have been published in IEEE journals, while most of the others in journals of high impact factor. In total, the cumulative sum of journals impact factors is 108.345. He has received 2635 citations in his work with h index =26 and g-Index =62 (see Scholar Google). One of his work has cited as “Doulamis Model” in the literature. . One of his work was cited as Doulamis Model in the literature and was taught at the Carnegie Mellon University. Prof. Nikolaos Doulamis is in the coordination group of FP7 VIMSEN, FP7 INTN-DCH and H2020 ViMM projects. He is currently involved in many European Projects like, H2020 TERPSICHORE, H2020 Rangers, H2020 WaterSpy, H2020 ViMM, FP7 eVacuate, FP7 Robospect, FP7 ZoneSec, and the national projects of e-Democracy, Viopolis, ABC



Marco Bertini - University of Florence, Italy, ([bertini@dsi.unifi.it](mailto:bertini@dsi.unifi.it)) Dr. Marco Bertini received the Laurea Degree in Electronic Engineering from the University of Florence in 1999, and Ph.D. in 2004. He is working at the Media Integration and Communication Center of the University of Florence. His interests are focused on digital libraries, action recognition and anomalous action detection. On these subjects he has addressed semantic analysis, action and event detection and recognition, semantic retrieval and techniques for video surveillance transcoding. He is author of 17 journal papers and more than 75 peer-reviewed conference papers, with h-index: 16 (according to Google Scholar). He has been involved in 5 EU research projects as WP coordinator and researcher. He has co-organized the 2010 Int. Workshop on Multimedia and Semantic Technologies (MUST 2010) and the 2nd and 3rd IEEE Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Streams (ARTEMIS) in conjunction with ICCV 2011 and ECCV 2012. He was chair of the ACM MM 2010 Open Source Software Competition, and was co-chair of the Workshop on Web-scale Vision and Social Media in conjunction with ECCV 2012.



Jordi González – Universitat Autònoma de Barcelona, Spain ([poal@cvc.uab.es](mailto:poal@cvc.uab.es)) Dr. Jordi González completed his PhD in Computer Engineering in 2004 at Universitat Autònoma de Barcelona (UAB). At present, he is Associate Professor in Computer Science and responsible of doctoral studies at the Computer Science Department, UAB. He is also a research fellow at the Computer Vision Center. The topic of his research is the cognitive evaluation of human behaviours in image sequences, or video-hermeneutics. The aim is the generation of both linguistic and visual descriptions, which best explain those behaviours observed in imagery streams. Dr. Jordi González co-organized the first and second international THEMIS workshops, and the 1st and the 3rd ARTEMIS workshop. He was member of the Workshop Chair and Local Arrangement Chair at ICCV 2011; Publicity Chair at AVSS 2011; and Tutorial Chair at ibPRIA 2011. He is a member of the Program Committee of IET IPR 2012, IHUS-12, ETRA 2012, VISAPP 2012, AMDO 2012, CIARP 2012 and IWCI 2012. He is member of the Editorial Board of the Computer Vision and Image Understanding (CVIU) journal and the IET journal in Computer Vision (IET-CVI).



Thomas Moeslund University of Aalborg, Denmark (tbm@create.aau.dk) Prof. Thomas B. Moeslund received the M.Sc. and Ph.D. degrees in 1996 and 2003, respectively, from Aalborg University in Denmark. He is head of the section for Media Technology, Aalborg and head of the Visual Analysis of People Laboratory both at Aalborg University. His research is focused on all aspects of automatic analysis of people using cameras and other sensors. He has been involved in 11 national and international research projects, both as coordinator, WP leader and researcher. He serves as associate editor and editorial board member for three international journals (Machine Vision and Application, The OpenCybernetics and Systemics Journal, Int. Journal of Advanced Robotic Systems). He has co-chaired six international workshops/tutorials and acted as PC member/reviewer for a number of conferences and workshops. He has published more than 100 peer reviewed journal and conference papers. Awards include a most cited paper award in 2009, a best IEEE paper award in 2010, a teacher of the year award in 2010 and a Most Suitable for Commercialization award in 2012. For further details on his activities please refer to <http://vbn.aau.dk/>