

Editorial

# Statistical Machine Learning for Human Behaviour Analysis

Thomas B. Moeslund <sup>1</sup>, Sergio Escalera <sup>2,3</sup> , Gholamreza Anbarjafari <sup>4,5,6</sup> ,  
Kamal Nasrollahi <sup>1,7,\*</sup>  and Jun Wan <sup>8</sup> 

<sup>1</sup> Visual Analysis of People Laboratory, Aalborg University, 9000 Aalborg, Denmark; tbm@create.aau.dk

<sup>2</sup> Computer Vision Centre, Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola),  
08193 Barcelona, Spain; sergio@maia.ub.es

<sup>3</sup> Department of Mathematics and Informatics, Universitat de Barcelona, 08007 Barcelona, Spain

<sup>4</sup> iCV Lab, Institute of Technology, University of Tartu, 50411 Tartu, Estonia; shb@ut.ee

<sup>5</sup> Department of Electrical and Electronic Engineering, Hasan Kalyoncu University, 27900 Gaziantep, Turkey

<sup>6</sup> PwC Finland, Itämerentori 2, 00100 Helsinki, Finland

<sup>7</sup> Research Department of Milestone Systems A/S, 2605 Copenhagen, Denmark

<sup>8</sup> National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences,  
Beijing 100190, China; jun.wan@ia.ac.cn

\* Correspondence: kna@milestone.dk

Received: 22 April 2020; Accepted: 6 May 2020; Published: 7 May 2020



**Keywords:** action recognition; emotion recognition; privacy-aware

Human behaviour analysis has introduced several challenges in various fields, such as applied information theory, affective computing, robotics, biometrics and pattern recognition. This Special Issue focused on novel vision-based approaches, mainly related to computer vision and machine learning, for the automatic analysis of human behaviour. We solicited submissions on the following topics: information theory-based pattern classification, biometric recognition, multimodal human analysis, low resolution human activity analysis, face analysis, abnormal behaviour analysis, unsupervised human analysis scenarios, 3D/4D human pose and shape estimation, human analysis in virtual/augmented reality, affective computing, social signal processing, personality computing, activity recognition, human tracking in the wild, and application of information-theoretic concepts for human behaviour analysis. In the end, 15 papers were accepted for this special issue [1–15]. These papers, that are reviewed in this editorial, analyse human behaviour from the aforementioned perspectives, defining in most of the cases the state of the art in their corresponding field.

Most of the included papers are application-based systems, while [15] focuses on the understanding and interpretation of a classification model, which is an important factor for the classifier's credibility. Given a set of categorical data, [15] utilizes multi-objective optimization algorithms, like ENORA and NSGA-II, to produce rule-based classification models that are easy to interpret. Performance of the classifier and its number of rules are optimized during the learning, where the first one is obviously expected to be maximized while the second one is expected to be minimized. Testing on public databases, using 10-fold cross-validation, shows the superiority of the proposed method against classifiers that are generated using other previously published methods like PART, JRip, OneR and ZeroR.

Two published papers ([1,9]) have privacy as their main concern, while they develop their respective systems for biometrics recognition and action recognition. Reference [1] has considered a privacy-aware biometrics system. The idea is that the identity of the users should not be readily revealed from their biometrics, like facial images. Therefore, they have collected a database of foot and hand traits of users while opening a door to grant or deny access, while [9] develops a privacy-aware method for action recognition using recurrent neural networks. The system accumulates reflections of

light pulses omitted by a laser, using a single-pixel hybrid photodetector. This includes information about the distance of the objects to the capturing device and their shapes.

Multimodality (RGB-depth) is covered in [14] for sign language recognition; while in [11], multiple domains (spatial and frequency) are used for saliency detection. Reference [14] has applied restricted Boltzmann machine (RBM)s to develop a system for sign language recognition from a given single image, in two modalities of RGB and depth. Two RBMs are designed to process the images coming from the two deployed modalities, while a third RBM fuses the results of the first two RBMs. The inputs to the first two RBMs are hand images that are detected by a convolutional neural network (CNN). The experimental results reported in [14] on two public databases show the state-of-the-art performance of the proposed system. Reference [11] proposes a multi-domain (spatial and frequency)-based system for salient object detection in foggy images. The frequency domain saliency map is extracted using the amplitude spectrum, while the spatial domain saliency map is calculated using the contrast of the local and global super-pixels. These different domain maps are fused using a discrete stationary wavelet transform (DSWT) and are then refined using an encoder-decoder model to pronounce the salient objects. Experimental results on public databases and comparison with state-of-the-art similar methods show the better performance of this system.

Four papers in this special issue have covered action recognition [6,9,12,13]. Reference [12] has proposed a system for toe-off detection using a regular camera. The system extracts the differences between consecutive frames to build silhouettes difference maps, that are then fed into a CNN for feature extraction and classification. Different types of maps are developed and tested in this paper. The experimental results reported in [12] on public databases show state-of-the-art performance. Reference [6] proposes a system for individuals and then crowd condition monitoring and prediction. Individuals participating in this study are grouped into crowds based on their physical locations extracted using GPS on their smartphones. Then, an enhanced context-aware framework using an algorithm for feature selection is used to extract statistical-based time-frequency domain features. Reference [13] focuses on utilizing recurring concepts using adaptive random forests to develop a system that can cope with drastically changing behaviours in dynamic environments, like financial markets. The proposed system is an ensemble-based classifier comprised of trees that are either active or inactive. The inactive ones keep a history of market operators' reactions in previously recorded similar situations, while either an inactive tree or a background tree that has recently been trained replaces the active ones, as a reaction to drift.

In terms of face analysis, in [10] a system is proposed for detecting fuzziness tendencies and utilizing these to design human-machine interfaces. This is motivated by the fact that humans tend to pay more attention to sections of information with fuzziness, which are sections with greater mental entropy. The work of [4] proposes a conditional random field-based system for segmentation of facial images into six facial parts. These are then converted into probability maps, which are used as feature maps for a random decision forest that estimates head-pose, age, and gender.

The method introduced in [3] uses singular value decomposition for removing background of fingerprint images. Then, it finds fingerprints' boundaries and applies an adaptive algorithm based on wavelets extrema and Henry system to detect singular points, which are widely used in applications related to fingerprint, like registration, orientation detection, fingerprint classification, and identification systems.

Three papers have covered emotion recognition, one from body movements [5], and two from speech signals [2,7]. In [2] a committee of classifiers has been applied to a pool of descriptors extracting features from speech signals. Then, it is used as a voting scheme on the classifiers' outputs to get to a conclusion about the emotional status from the used speech signals. The paper in [2] shows that the committee of classifiers outperforms the single individual classifiers in the committee. The system proposed in [7] builds 3D tensors of spectrogram frames that are obtained by extracting 88-dimensional feature vectors from speech signals. These tensors are then used for building a 3D convolutional neural network that is employed for emotion recognition. The system has produced state-of-the-art

results on three public databases. The emotional recognition system of [5] does not use facial images or speech signals, but body movements, which are captured by Microsoft Kinect v2 under eight different emotional states. The affective movements are represented by extracting and tracking location and orientation of body joints over time. Experimental results, using different deep learning-based methods, show the state-of-the-art performance of this system.

Finally, two databases have been introduced in this special issue, one for biometric recognition [1] and one for detecting sleeping issues and fatigue [8], the later containing a database of patients suffering from Fibromyalgia, which is a situation resulting in muscle pain and tenderness, accompanied by few other signs including sleep, memory, and mood disorders. It uses similarity functions with configurable convexity or concavity to build a classifier on this collected database in order to predict extreme cases of sleeping issues and fatigue.

**Acknowledgments:** We express our thanks to the authors of the above contributions and to the journal Entropy and MDPI for their support during this work. Kamal Nasrollahi's contribution to this work is partially supported by the EU H2020-funded SafeCare project, grant agreement no. 787002. This work is partially supported by ICREA under the ICREA Academia programme.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Jahromi, S.M.N.; Buch-Cardona, P.; Avots, E.; Nasrollahi, K.; Escalera, S.; Moeslund, T.B.; Anbarjafari, G. Privacy-Constrained Biometric System for Non-Cooperative Users. *Entropy* **2019**, *21*, 1033. [[CrossRef](#)]
2. Kamińska, D. Emotional Speech Recognition Based on the Committee of Classifiers. *Entropy* **2019**, *21*, 920. [[CrossRef](#)]
3. Le, N.T.; Le, D.H.; Wang, J.-W.; Wang, C.-C. Entropy-Based Clustering Algorithm for Fingerprint Singular Point Detection. *Entropy* **2019**, *21*, 786. [[CrossRef](#)]
4. Khan, K.; Attique, M.; Syed, I.; Sarwar, G.; Irfan, M.A.; Khan, R.U. A Unified Framework for Head Pose, Age and Gender Classification through End-to-End Face Segmentation. *Entropy* **2019**, *21*, 647. [[CrossRef](#)]
5. Sapiński, T.; Kamińska, D.; Pelikant, A.; Anbarjafari, G. Emotion Recognition from Skeletal Movements. *Entropy* **2019**, *21*, 646. [[CrossRef](#)]
6. Sadiq, F.I.; Selamat, A.; Ibrahim, R.; Krejcar, O. Enhanced Approach Using Reduced SBTFD Features and Modified Individual Behavior Estimation for Crowd Condition Prediction. *Entropy* **2019**, *21*, 487. [[CrossRef](#)]
7. Hajarolasvadi, N.; Demirel, H. 3D CNN-Based Speech Emotion Recognition Using K-Means Clustering and Spectrograms. *Entropy* **2019**, *21*, 479. [[CrossRef](#)]
8. Sabeti, E.; Gryak, J.; Derksen, H.; Biwer, C.; Ansari, S.; Isenstein, H.; Kratz, A.; Najarian, K. Learning Using Concave and Convex Kernels: Applications in Predicting Quality of Sleep and Level of Fatigue in Fibromyalgia. *Entropy* **2019**, *21*, 442. [[CrossRef](#)]
9. Ofodile, I.; Helmi, A.; Clapés, A.; Avots, E.; Peensoo, K.M.; Valdma, S.-M.; Valdmann, A.; Valtna-Lukner, H.; Omelkov, S.; Escalera, S.; et al. Action Recognition Using Single-Pixel Time-of-Flight Detection. *Entropy* **2019**, *21*, 414. [[CrossRef](#)]
10. Bao, H.; Fang, W.; Guo, B.; Wang, P. Supervisors' Visual Attention Allocation Modeling Using Hybrid Entropy. *Entropy* **2019**, *21*, 393. [[CrossRef](#)]
11. Zhu, X.; Xu, X.; Mu, N. Saliency Detection Based on the Combination of High-Level Knowledge and Low-Level Cues in Foggy Images. *Entropy* **2019**, *21*, 374. [[CrossRef](#)]
12. Tang, Y.; Li, Z.; Tian, H.; Ding, J.; Lin, B. Detecting Toe-Off Events Utilizing a Vision-Based Method. *Entropy* **2019**, *21*, 329. [[CrossRef](#)]
13. Suárez-Cetrulo, A.L.; Cervantes, A.; Quintana, D. Incremental Market Behavior Classification in Presence of Recurring Concepts. *Entropy* **2019**, *21*, 25. [[CrossRef](#)]

14. Rastgoo, R.; Kiani, K.; Escalera, S. Multi-Modal Deep Hand Sign Language Recognition in Still Images Using Restricted Boltzmann Machine. *Entropy* **2018**, *20*, 809. [[CrossRef](#)]
15. Jiménez, F.; Martínez, C.; Miralles-Pechuán, L.; Sánchez, G.; Sciavicco, G. Multi-Objective Evolutionary Rule-Based Classification with Categorical Data. *Entropy* **2018**, *20*, 684. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).