

Editorial

Computer Vision in Human Analysis: From Face and Body to Clothes

Mohamed Daoudi ^{1,2}, Roberto Vezzani ³, Guido Borghi ^{4,*}, Claudio Ferrari ⁵, Marcella Cornia ⁶, Federico Becattini ^{7,8} and Andrea Pilzer ⁹

¹ CNRS, Centrale Lille, Institut Mines-Télécom, UMR 9189 CRISTAL, University of Lille, F-59000 Lille, France; mohamed.daoudi@imt-lille-douai.fr

² IMT Nord Europe, Institut Mines-Télécom, Centre for Digital Systems, F-59000 Lille, France

³ Department of Engineering “Enzo Ferrari”, University of Modena and Reggio Emilia, 41100 Modena, Italy; roberto.vezzani@unimore.it

⁴ Department of Computer Science and Engineering, University of Bologna, 47521 Cesena, Italy

⁵ Department of Engineering and Architecture, University of Parma, 43121 Parma, Italy; claudio.ferrari2@unipr.it

⁶ Department of Education and Humanities, University of Modena and Reggio Emilia, 42121 Reggio Emilia, Italy; marcella.cornia@unimore.it

⁷ Media Integration and Communication Center, University of Florence, 50100 Florence, Italy; federico.becattini@unifi.it

⁸ Dipartimento di Ingegneria dell’Informazione e Scienze Matematiche, University of Siena, 53100 Siena, Italy

⁹ Department of Computer Science, Aalto University, 02130 Espoo, Finland; andrea.pilzer@aalto.fi

* Correspondence: guido.borghi@unibo.it

1. Introduction

For decades, researchers of different areas, ranging from artificial intelligence to computer vision, have intensively investigated human-centered data, i.e., data in which the human plays a significant role, acquired through a non-invasive approach, such as cameras. This interest has been largely supported by the highly informative nature of this kind of data, which provides a variety of information with which it is possible to understand many aspects including, for instance, the human body or the outward appearance. Some of the main tasks related to human analysis are focused on the body (e.g., human pose estimation and anthropocentric measurement estimation), the hands (e.g., gesture detection and recognition), the head (e.g., head pose estimation), or the face (e.g., emotion and expression recognition). Additional tasks are based on non-corporal elements, such as motion (e.g., action recognition and human behavior understanding) and clothes (e.g., garment-based virtual try-on and attribute recognition). Unfortunately, privacy issues severely limit the usage and the diffusion of this kind of data, making the exploitation of learning approaches challenging. In particular, privacy issues behind the acquisition and the use of human-centered data must be addressed by public and private institutions and companies.

Thirteen high-quality papers have been published in this Special Issue and are summarized in the following: four of them are focused on the human face (facial geometry, facial landmark detection, and emotion recognition), two on eye image analysis (eye status classification and 3D gaze estimation), five on the body (pose estimation, conversational gesture analysis, and action recognition), and two on the outward appearance (transferring clothing styles and fashion-oriented image captioning). These numbers confirm the high interest in human-centered data and, in particular, the variety of real-world applications that it is possible to develop.

2. Overview of Contribution

The human body represents one of the most investigated elements in the literature and in our Special Issue. In [1], the authors propose a system that can predict the future



Citation: Daoudi, M.; Vezzani, R.; Borghi, G.; Ferrari, C.; Cornia, M.; Becattini, F.; Pilzer, A. Computer Vision in Human Analysis: From Face and Body to Clothes. *Sensors* **2023**, *23*, 5378. <https://doi.org/10.3390/s23125378>

Received: 31 May 2023

Accepted: 31 May 2023

Published: 6 June 2023



Copyright: © 2023 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 (<https://creativecommons.org/licenses/by/4.0/>).

skeleton sequence through the integration of the surrounding situation directly into the presented model. In particular, the accuracy is improved for motions related to humans and objects. Amadi et al. [2] analyze the segmentation of human body parts through the usage of optimized 2D poses, validating the approach on the *Transportation Security Administration Passenger Screening Dataset (TSA-PSD)*. The task of 3D human pose estimation is addressed in [3], in which the authors propose the use of bidirectional gated recurrent units to predict the global motion sequence from the local pose sequence. Gestures are investigated in [4], where a method for capturing gestures automatically from videos and transforming them into stored 3D representations is proposed. In [5], the authors exploit body joints to predict action progress.

Another topic investigated in this Special Issue is human face analysis. In particular, the published papers address different topics, including the problem of machine interaction using voice commands and facial movements [6], 3D face and body geometry reconstruction [7,8], and dyadic interaction analysis based on facial expressions [9]. Focusing on eye images, Gibertoni et al. [10] propose a system to automatically classify the eye status in images acquired through an ophthalmic tool. The authors suggest that this solution can help to improve the quality of future datasets acquired in this field, also simplifying the operations of non-technical figures, such as doctors. The second work concerning the human eye is described in [11] and consists of a framework developed to identify the user's attention in a corneal imaging system. The proposed system is based on infrared and RGB images and, through an eyeball model, a final prediction of the 3D direction of the gaze is output.

Finally, two papers focus on the problem of outward appearance and fashion. In particular, Fontanini et al. [12] propose a method for transferring clothing styles across images of people, while Moratelli et al. [13] propose an image captioning approach for fashion retrieval applications.

3. Conclusions

The main goal of this Special Issue is to improve the communication between companies and researchers belonging to both private and public institutions regarding the opportunities (and limitations) of the use of human-centered data in the development of future artificial intelligence applications. The above-mentioned papers contribute to stimulating new ideas, motivations, and methodologies that can shape the future of this area, also outlining potential future industrial applications and prospective trends. Again, we remark on the importance of the proper use of data concerning humans, which must be compliant with privacy and ethical regulations.

Author Contributions: Conceptualization, G.B., C.F., M.C., F.B. and A.P.; writing—original draft preparation, G.B., C.F., M.C., F.B. and A.P.; writing—review and editing, G.B., C.F., M.C., F.B. and A.P.; supervision, G.B., C.F., M.C., F.B., A.P., M.D. and R.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

References

1. Fujita, T.; Kawanishi, Y. Future Pose Prediction from 3D Human Skeleton Sequence with Surrounding Situation. *Sensors* **2023**, *23*, 876. [[CrossRef](#)] [[PubMed](#)]
2. Amadi, L.; Agam, G. Weakly Supervised 2D Pose Adaptation and Body Part Segmentation for Concealed Object Detection. *Sensors* **2023**, *23*, 2005. [[CrossRef](#)] [[PubMed](#)]
3. Kim, S.H.; Jeong, S.; Park, S.; Chang, J.Y. Camera Motion Agnostic Method for Estimating 3D Human Poses. *Sensors* **2022**, *22*, 7975. [[CrossRef](#)] [[PubMed](#)]
4. Močnik, G.; Kačič, Z.; Šafarič, R.; Mlakar, I. Capturing Conversational Gestures for Embodied Conversational Agents Using an Optimized Kaneda–Lucas–Tomasi Tracker and Denavit–Hartenberg–Based Kinematic Model. *Sensors* **2022**, *22*, 8318. [[CrossRef](#)] [[PubMed](#)]
5. Pucci, D.; Becattini, F.; Del Bimbo, A. Joint-Based Action Progress Prediction. *Sensors* **2023**, *23*, 520. [[CrossRef](#)] [[PubMed](#)]

6. Ramos, P.; Zapata, M.; Valencia, K.; Vargas, V.; Ramos-Galarza, C. Low-Cost Human–Machine Interface for Computer Control with Facial Landmark Detection and Voice Commands. *Sensors* **2022**, *22*, 9279. [[CrossRef](#)] [[PubMed](#)]
7. Young, P.; Ebadi, N.; Das, A.; Bethany, M.; Desai, K.; Najafirad, P. Can Hierarchical Transformers Learn Facial Geometry? *Sensors* **2023**, *23*, 929. [[CrossRef](#)] [[PubMed](#)]
8. Gallucci, A.; Znamenskiy, D.; Long, Y.; Pezzotti, N.; Petkovic, M. Generating High-Resolution 3D Faces and Bodies Using VQ-VAE-2 with PixelSNAIL Networks on 2D Representations. *Sensors* **2023**, *23*, 1168. [[CrossRef](#)] [[PubMed](#)]
9. Sham, A.H.; Khan, A.; Lamas, D.; Tikka, P.; Anbarjafari, G. Towards Context-Aware Facial Emotion Reaction Database for Dyadic Interaction Settings. *Sensors* **2023**, *23*, 458. [[CrossRef](#)] [[PubMed](#)]
10. Gibertoni, G.; Borghi, G.; Rovati, L. Vision-Based Eye Image Classification for Ophthalmic Measurement Systems. *Sensors* **2022**, *23*, 386. [[CrossRef](#)] [[PubMed](#)]
11. Mokatren, M.; Kuflik, T.; Shimshoni, I. 3D Gaze Estimation Using RGB-IR Cameras. *Sensors* **2022**, *23*, 381. [[CrossRef](#)] [[PubMed](#)]
12. Fontanini, T.; Ferrari, C. Would Your Clothes Look Good on Me? Towards Transferring Clothing Styles with Adaptive Instance Normalization. *Sensors* **2022**, *22*, 5002. [[CrossRef](#)] [[PubMed](#)]
13. Moratelli, N.; Barraco, M.; Morelli, D.; Cornia, M.; Baraldi, L.; Cucchiara, R. Fashion-Oriented Image Captioning with External Knowledge Retrieval and Fully Attentive Gates. *Sensors* **2023**, *23*, 1286. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.