Introduction to the Special Issue on Representation, Analysis, and Recognition of 3D Humans

Modeling, processing, recognizing, searching, and retrieving 3D human data (shapes, gestures, interactions) is a well-established research area in multimedia. In the last decade, there has been a significant increase in opportunities for using 3D human data in medicine, security, and human computer interaction, largely driven by the development of effective devices and algorithms for recovering 3D data (e.g., Microsoft Kinect, Intel RealSense, Google Project Tango, and Apple Prime-Sense). Such rich information opens the way to new modes of experiential computing and interactive environments, as well as new multimedia content. Based on this, this special issue (SI) aims to be of interest to an interdisciplinary target audience, as well interdisciplinary teams of contributors, spanning applied math, multimedia experiential computing, computational science and engineering, and application domain experts. This SI intends to bring together researchers interested in defining new and innovative solutions that advance the way 3D human data are used in multimedia computing, communications and applications such as human behavior understanding from 3D sensors, animation and entertainment, sports analytics, natural interaction, and virtual and augmented reality.

The SI received 16 submissions all of good quality. Among these, we accepted eight articles plus one survey article after rigorous peer review. The topics, main goals, and contributions of these articles are summarized below.

The SI is opened by the survey article “Representation, Analysis, and Recognition of 3D Humans: A Survey” by Berretti et al., which provides an updated overview of the most significant and recent works on the topic of this SI. In addition to proposing and discussing existing works according to a taxonomical view, current issues and future research perspectives are prospected.

The initial three articles of this SI investigate new solutions for **3D face and facial expression recognition**. In the first article, entitled “Deformation-Based 3D Facial Expression Representation,” Demisse et al. propose a deformation-based representation for analyzing expressions from 3D faces. In their approach, a point cloud of a 3D face is decomposed into an ordered deformable set of curves that start from a fixed point. Subsequently, a mapping function is defined to identify the set of curves with an element of a high-dimensional matrix Lie group, specifically the direct product of $SE(3)$. In the subsequent article, “Texture and Geometry Scattering Representation-Based Facial Expression Recognition in 2D + 3D Videos,” Huang et al. propose a novel and effective approach to facial expression recognition using multimodal 2D and 3D videos, which encode both static and dynamic clues by scattering convolution network. The articles about 3D faces are closed by the work of Azzakhnini et al., “Combining Facial Parts for Learning Gender, Ethnicity, and Emotional State Based on RGB-D Information,” where the authors address the problem of face classification in the context of RGB images and depth data. Inspired by the psychological results for human face perception, this article focuses on (1) finding out which facial parts are most effective at making the difference for some social aspects of face perception (gender, ethnicity, and emotion state); (2) determining the optimal decision by combining the decision rendered by the

2018 Copyright is held by the owner/author(s).

1551-6857/2018/03-ART15

<https://doi.org/10.1145/3181709>

individual parts; and (3) extracting the features from RGB-D faces to exploit all the potential that this data provide.

The subsequent three articles focus on different aspects of **3D body analysis and recognition**. In “Early Recognition of 3D Human Actions,” Li et al. focus on early recognition of ongoing 3D human actions. The goal is to infer the class label information of 3D human actions with partial observation of temporally incomplete action executions. Duan et al., in the work entitled “A Unified Framework for Multimodal Isolated Gesture Recognition,” investigate isolated gesture recognition and explore different modalities by involving the RGB stream, depth stream, and saliency stream for inspection. The proposed unified framework exploits the advantages of multimodal fusion. Specifically, a spatial-temporal network architecture based on consensus voting is proposed to explicitly model the long-term structure of the video sequence and to reduce estimation variance when confronted with comprehensive interclass variations. In the work entitled “Gait Recognition from Motion Capture Data,” Balazia and Sojka contribute to the state of the art on gait recognition from motion capture data, with a statistical approach for directly extracting robust gait features from raw data by a modification of the Linear Discriminant Analysis with Maximum Margin Criterion.

The SI is closed by two works on **3D body reconstruction**. Su et al., in “Sequential Articulated Motion Reconstruction from a Monocular Image Sequence,” present a sequential approach to reconstructing 3D body articulated poses from monocular sequences. This task is challenging due to the inherent depth ambiguities, high dimensionality, and complexity of human movement. The proposed approach models the human movement on kinematics manifold with tangent bundle, which allows for a natural geometrical representation of the articulated motion. Combined with a second-order stochastic dynamic model, which is based on the Markov hypothesis, the Extended Rauch Tung Striebel Smoother is generalized to a Riemannian manifold to simulate the process of human motion. Finally, Afzal et al., in “Full 3D Reconstruction of Nonrigidly Deforming Objects,” discuss enhanced full 360° 3D reconstruction of dynamic scenes containing nonrigidly deforming objects using data acquired from commodity depth or 3D cameras. In this work, the authors target online and template-free enhancement of the quality of noisy and low-resolution full 3D reconstructions of dynamic nonrigid objects.

Finally, we would like to thank all reviewers for their professional work. We appreciate the support of Professor Alberto Del Bimbo, the Editor-in-Chief of ACM TOMM, during the editing of this special issue. We hope this special issue can motivate additional efforts for overcoming major challenges in the representation, analysis, and recognition of 3D humans.

Stefano Berretti
University of Florence

Mohamed Daoudi
IMT Lille Douai, University of Lille

Pavan Turaga
Arizona State University

Anup Basu
University of Alberta