Chapter 10
Single View 3D Face Reconstruction

Claudio Ferrari

https://orcid.org/0000-0001-9465-6753

University of Florence, Italy

Stefano Berretti

https://orcid.org/0000-0003-1219-4386

University of Florence, Italy

Alberto Del Bimbo

University of Florence, Italy

ABSTRACT

3D face reconstruction from a single 2D image is a fundamental computer vision problem of extraordinary difficulty that dates back to the 1980s. Briefly, it is the task of recovering the three-dimensional geometry of a human face from a single RGB image. While the problem of automatically estimating the 3D structure of a generic scene from RGB images can be regarded as a general task, the particular morphology and non-rigid nature of human faces make it a challenging problem for which dedicated approaches are still currently studied. This chapter aims at providing an overview of the problem, its evolutions, the current state of the art, and future trends.

INTRODUCTION

Estimating the 3D information of a scene from 2D images using computer vision techniques is a research topic with a quite long tradition that dates back to '80. Now, remaining the 3D acquisition limited to a certain constrained domain, the deployment of powerful deep learning tools has pushed forward this research field, with innovative solutions that appeared recently.

Estimating the 3D geometry from single or multiple images under the most general conditions, where no *a priori* knowledge is available about the imaged scene and the capturing conditions is a very challenging task. Hence, to make the problem solvable to some extent, priors are usually assumed. In the case a 3D model of the face, the prior knowledge can be in the form of camera parameters and reflectance properties of the face considering either a single image, as in the shape from shading (SfS) solution (Horn and Brooks, 1989) or multiple images with different illuminations in the photometric stereo approach (Woodham, 1980). Though quite accurate reconstructions can be obtained with these solutions, the given assumptions are rarely verified in real contexts. However, recent technical advancements in scanning technologies are making it possible to acquire 3D data of sufficient quality at an affordable cost, so that reconstruction algorithms can leverage additional data. Not only 3D data but also the increased availability of 2D imagery can be exploited to this aim; in (Ioannides et al, 2013), authors employ online image repositories to collect multi-view imagery of cultural heritage sites to reconstruct the 3D structure. Multi-view reconstruction is usually achieved using multiple photographs of the same object taken at subsequent time steps from different view-points, which are not easily available. However, multiple images collected from the internet can still be used with some additional effort; clearly this reasonably implies less accurate reconstructions. In a similar way, researchers are putting effort in collecting 4D databases containing temporal sequences of 3D scans as in (Cheng et al, 2018). The temporal consistency is indeed an importan source of information that can be exploited in place of other constraints.

Nevertheless, to overcome such limitations, in 1999 the 3D Morphable face Model (3DMM) was first proposed (Blanz and Vetter, 1999) to be used as a statistical human face shape prior. Despite being originally developed as a generative model, it has been subsequently used in many applications as a means to limit the shape deformations to statistically plausible motions. The main idea behind the 3DMM is that of exploiting the statistics of 3D face shapes to generate new faces by controlling a set of parameters learned from a training set of real 3D face scans. This statistical model limits the shape of the reconstructed face to the combination, according to a set of parameters, of an average face model and some deformation components. Thanks to its intuitiveness and simplicity, it is still exploited as a means for relieving constraints on the general problem. Different solutions have been proposed in the literature for solving for these parameters. In the original 3DMM, as firstly proposed in (Blanz and Vetter, 1999), this was formulated as the problem of iteratively minimizing the difference between the 2D target image and the face image rendered from the 3D reconstruction. This line of research was extended by many others, as it allows a pretty accurate shape reconstruction, though computationally onerous. Later works proposed many alternatives to learn the parameters, for example via linear regression from the position of a set of corresponding 2D and 3D landmarks or exploiting geometrical constraints. These latter solutions, though efficient, often result in coarse reconstructions that can be sensitive to inaccurate landmarks detection in the 2D images.

Despite these drawbacks, the 3DMM has been the founding idea of several recent solutions that use deep neural networks to learn complex non-linear regression functions or as a tool for generating synthetic training data. Nonetheless, the main limitation of the 3DMM is that the results of such reconstructions appear still approximated, lacking fine-grained details of the face. The current trend is moving towards solutions that start from an initial smooth estimation of the face shape, then add local, fine-grained details.

Since the vast majority of works addressing the problem of single-view 3D face reconstruction are based on the morphable model, in the following the latter is detailed and relevant works are presented.

Related Content

Classification of Ovarian Cancer Subtype Using Histopathology Images

R. Parvathiand Xiaohui Yuan (2025). *Signal and Image Processing Techniques for Defense, Security, and Healthcare (pp. 59-86).*

www.irma-international.org/chapter/classification-of-ovarian-cancer-subtype-using-histopathology-images/372698

Early Perception-Action Cycles in Binocular Vision: Visuomotor Paradigms and Cortical-Like Architectures

Silvio P. Sabatini, Fabio Solari, Andrea Canessa, Manuela Chessaand Agostino Gibaldi (2013). *Developing and Applying Biologically-Inspired Vision Systems: Interdisciplinary Concepts (pp. 154-182).*

www.irma-international.org/chapter/early-perception-action-cycles-binocular/72028

Enhancement of Hybrid Multimodal Medical Image Fusion Techniques for Clinical Disease Analysis

Rajalingam B.and Priya R. (2018). *International Journal of Computer Vision and Image Processing (pp. 16-40).*

www.irma-international.org/article/enhancement-of-hybrid-multimodal-medical-image-fusion-techniques-for-clinical-disease-analysis/212374

Feature Extraction Methods for Intrusion Detection Systems

Hai Thanh Nguyen, Katrin Frankeand Slobodan Petrovic (2013). *Image Processing: Concepts, Methodologies, Tools, and Applications (pp. 1064-1092).*

www.irma-international.org/chapter/feature-extraction-methods-intrusion-detection/77588

Growing Neural Forest-Based Color Quantization Applied to RGB Images

Jesús Benito-Picazo, Ezequiel López-Rubioand Enrique Domínguez (2017). *International Journal of Computer Vision and Image Processing (pp. 13-25).*

www.irma-international.org/article/growing-neural-forest-based-color-quantization-applied-to-rgb-images/188758