Global Alignment for Dynamic 3D Morphable Model Construction

Rui Tang, Darren Cosker and Wenbin Li
University of Bath, Bath, BA2 7AY, UK

I. Abstract

We propose a novel 3D Dynamic Morphable Model which utilizes the Dynamic 3D FACS Dataset (D3DFACS) [1]. Compared to the technique proposed in [1], the model provides a more accurate global representation and improved methods for aligning independent sequences of facial expressions. Firstly, we find the rigid correspondence between raw meshes by using Coherent Point Drift (CPD) [2]. In order to obtain more precise correspondences on UV texture maps, we then use a novel hybrid registration approach combining Thin-plate Splines (TPS) and optical flow [3]. We compare this approach of alignment of independent dynamic sequences against standard optical flow methods proposed in [1]. After both rigid and non-rigid alignment, a 3D dynamic morphable model is constructed using Principle Component Analysis (PCA). We aim to use the model to assess the ability of 3D dynamic morphable models on speech replications for talking head and other animation applications.

II. Introduction

In recent years, facial analysis using 3D models has been a central topic in computer vision and graphics. A main reason of this is that such models are more robust to pose and light invariance in recognition and allow for the estimation of 3D facial shape from 2D images [4]. To account for these advantages, the 3D morphable model (3DMM) was originally introduced by Blanz and Vetter [5] in 1999. This statistical model has been widely used to perform various tasks such as face recognition, expression transfer between individuals and reconstruction of 3D face from 2D images.

Facial expression recognition is also an active research area, with many works based on recognising facial movement descriptions based on the Facial Action Coding System (FACS) [6]. FACS was primarily introduced by psychologists to describe different configurations of facial actions or Action Units (AUs). It provides 44 individual AUs which form the basis of 6 prototypical facial expressions: happiness, sadness, fear, surprise, anger and disgust. Although FACS recognition is widespread in 2D facial analysis, there are only a limited number of 3D facial datasets available, and only one dynamic 3D facial expression database based on FACS [1].

In this paper, we propose a new method for dynamic 3D morphable model construction based on FACS data. One difficult aspect of using such data is that it will typically be based on multiple recordings of different facial expressions (AUs), as opposed to multiple e.g. neutral expressions of different people [5]. Each AU will have a different neutral, and each expression will contain a large degree of rigid head movement. Our method uses a global reference expression combined with both rigid and non-rigid alignment (in 2D and 3D) to construct a globally aligned non-rigid 3D space.

III. 3D Dynamic Morphable Model

We extract 63 AU sequences from the D3DFACS data set for one participant, and use our approach to build a 3D dynamic morphable model for this person.

A. 3D Rigid Registration

The first issue to overcome is rigid alignment for the data set. Iterative closest point (ICP) is commonly used for rigid 3D alignment. To align two meshes, ICP iteratively revises the transformation to minimize the distance between two point sets. However, it requires the initial position of the raw point sets to be adequately close. We therefore use CPD which is more robust for data with noise and a high degree of rigid variation. CPD converts the alignment of two point sets to a probability density estimation problem. As opposed to applying CPD on each individual 3D mesh, we estimate the transforms between neutral images for each AU and a global neutral. We then apply these to the remaining frames in the AU sequence, aligning it with the global neutral. The process is as follows:

a) Select an AU face mesh with neutral expression as our global reference mesh G.
b) For the ith AU sequence, we mark its first frame (neutral) as Fi.
c) Apply CPD between Fi and G to obtain a rigid transformation matrix RTi for each AU sequence.
d) Obtain global rigid registration by applying RTi for each mesh in the ith AU sequence.

B. 3D Global Non-Rigid Alignment

1) 3D to 2D projection: After rigid alignment, all 3D data is normalised into a standard space. Next, we apply a cylindrical projection on this data to obtain cylindrical UV maps for each aligned 3D scan. As a result, for the ith sequence with meshes \( \chi_i = [x_i, y_i, z_i] \), where \( X = [x_1, x_2, \ldots, x_n] \), \( x_i = [x_{i1}, x_{i2}, x_{i3}]^T \in \mathbb{R} \), we generate a set of UV texture maps \( \mathbb{U} = [U_1, U_2, \ldots, U_n] \), where \( U = [u_1^T, u_2^T, \ldots, u_n^T] \), and \( u_i = [u_i, v_i] \). Using \( \mathbb{U} \) and \( \chi \) we...
also generate an additional set of 3D images $I_{3D}(u) = x_i$. These map any point in the UV space $u_i = [u_i, v_i]$ to a 3D mesh coordinate $x_i = [x_i^1, x_i^2, x_i^3]^T$, allowing us to handle 3D deformation in image space. More details are described in [1].

2) Local Optical Flow based Non-Rigid Registration: After rigid alignment of the data set, the next aim is to create vertex correspondences through the set of meshes. Typically, each mesh has a different number vertices, so the problem is to non-rigidly deform a reference mesh to each other mesh using pixel tracking. We use a two step non-rigid alignment process to solve this problem, which first acts locally (intra-sequence) and then globally (inter-sequence).

Firstly, a local alignment is applied to each AU sequence individually. Given the UV texture maps from sequence $I_i$, we estimate the optical flow fields $f_1, f_2, f_3, \ldots, f_{n-1}$ between the pairs $(I_1, I_2), (I_2, I_3), (I_3, I_4), \ldots (I_{n-1}, I_n)$. We apply these fields to warp both the UV texture maps and 3D images such that they are all aligned back to the neutral expression of the AU sequence (typically the first frame). We denote the corresponding UV texture maps and 3D images as $I_i^{UV}$ and $I_i^{3D}$ respectively.

3) Global Hybrid Non-Rigid Registration: After local non-rigid alignment for an individual sequence, a global non-rigid alignment is then employed to align all UV textures and 3D images, for each AU sequence, to the global reference (global neutral expression image). In this procedure, we first compute the optical flow fields between the global reference image and the texture reference images of each sequence respectively. As the difference between the neutral expressions for AU sequences can be large, this can result in a noisy flow field, and hence lead to warping artifacts. We therefore combine optical flow with Thin Plate Splines (TPS) [7] in this stage to achieve a more consistent warp field. We manually landmark the global neutral reference with a set of landmarks, and use the flow field positions of these landmarks to obtain corresponding landmarks in the neutral references of each AU sequence. We then use TPS to warp $I_i^{UV}$ and $I_i^{3D}$ to the global neutral reference face, obtaining $\bar{I}_i^{UV}$ and $\bar{I}_i^{3D}$. Figure 1B shows visual comparison of using optical flow alone for global registration versus sparse optical flow and TPS.

C. Statistical Model

Once the UV maps and 3D images are fully corresponded, aligned 3D meshes (in non-rigid correspondence) are generated from the 3D images by regularly sampling them. A 3D morphable model can then be constructed in a similar way to [5]. We have:

$$T = \bar{T} + \alpha \ast P_T, \quad S = \bar{S} + \beta \ast P_S$$

where $\bar{T}$ is the mean UV Texture Map, $\bar{S}$ is the mean mesh, $P_T$ and $P_S$ are the eigenvectors of $\bar{I}_i$ and $\bar{U}_i$, and $\alpha$ and $\beta$ are vectors of weights. Given this representation, any new face meshes can be generated by varying $\alpha$ and $\beta$ which control shape and texture. Additionally, by rewriting equation(1), all 3D mesh and UV data can be represented into a lower dimensional space:

$$\alpha = P_T(T' - \bar{T}), \quad \beta = P_S(S' - \bar{S})$$

where $T'$ and $S'$ denote UV Texture Map and UV Geometric Map of any mesh. This lower dimensional representation provides a more efficient way for further 3D facial analysis.

IV. Further Work

We have described an improved approach for 3D morphable model construction given dynamic 3D data. The process considers the problem of building such models given multiple dynamic facial sequences. In future work, we look to apply this model for 3D tracking of facial movements for speech analysis, and the animation of 3D faces during speech.

References