Computer Vision and Image Understanding 117 (2013) 250–268

Contents lists available at SciVerse ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Generative face alignment through 2.5D active appearance models ☆

Pedro Martins *, Rui Caseiro, Jorge Batista

Institute of Systems and Robotics, University of Coimbra, DEEC – Polo II, 3030-290 Coimbra, Portugal

ARTICLE INFO

Article history:
Received 15 October 2011
Accepted 29 November 2012
Available online 10 December 2012

Keywords:
2.5D Models
Active appearance models
AAM
2.5D AAM
Image alignment
Lucas–Kanade
Simultaneous forwards additive
Normalization forwards additive
Robust fitting

ABSTRACT

This work addresses the matching of a 3D deformable face model to 2D images through a 2.5D Active Appearance Models (AAM). We propose a 2.5D AAM that combines a 3D *metric* Point Distribution Model (PDM) and a 2D appearance model whose control points are defined by a *full perspective* projection of the PDM. The advantage is that, assuming a calibrated camera, 3D metric shapes can be retrieved from single view images. Two model fitting algorithms and their computational efficient approximations are proposed: the Simultaneous Forwards Additive (SFA) and the Normalization Forwards Additive (NFA), both based on the Lucas–Kanade framework. The SFA algorithm searches for shape and appearance parameters simultaneously whereas the NFA projects out the appearance from the error image and searches only for the shape parameters. SFA is therefore more accurate. Robust solutions for the SFA and NFA are also proposed in order to take into account the self-occlusion or partial occlusion of the face. Several performance evaluations for the SFA, NFA and theirs efficient approximations were performed. The experiments include evaluating the frequency of converge, the fitting performance in unseen data and the tracking performance in the FGNET Talking Face sequence. All results show that the 2.5D AAM can outperform both the 2D + 3D combined models and the 2D standard methods. The robust extensions to occlusion were tested on a synthetic sequence showing that the model can deal efficiently with large head rotation.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Facial image alignment is a key aspect in many computer vision applications, such as advanced human computer interaction, face recognition, head pose estimation, facial expression analysis, surveillance or realistic graphical animation. Detecting and tracking faces in video is a challenging task due to the non-rigidity structure of faces and also due to the large variability in shape, texture, pose and lighting conditions of their images.

The Active Appearance Model (AAM), introduced by [1], is one of the most effective face alignment technique with respect to fitting accuracy and efficiency. The standard AAMs are intrinsically 2D models, combining a 2D Point Distribution Model (PDM) [2,3] and a 2D appearance model into a single formulation using a fitting process that rely on a precomputed regression matrix.

The AAM has been reformulated with true analytical derived gradients by Matthews et al. [4], achieving a better fitting accuracy and real-time performances using the Inverse Compositional (IC) [5] approach. Their solution is probably the fastest introduced so far, where its key to efficiency is that both the Jacobian and the Hessian matrices are constant and can be precomputed. A dual inverse compositional algorithm was also proposed in [6], dealing with both the geometric and photometric transformations in image registration under varying lighting conditions.

Although the excellent performance of the 2D AAM, its convergence ability is severely affected under large 3D head pose variations. To deal with this issue, several solutions have been proposed [7–10]. View-Based AAM [7] uses multiple 2D AAMs taken from each view, while issues related to self-occlusion are solved by using multiple view-specific templates. Similarly, the solution proposed by [8] uses multiple view appearance models although combined with a sparse 3D PDM. In [11] a IC algorithm for simultaneously fitting a 2D and a 3D PDM to multiple images is proposed. Their fitting methodology, instead of relying on multiple independent optimizations, is formulated in a single-objective optimization by enforcing the same 3D model across all the views. In [9,12], a 3D PDM derived from the Candide model [13] is used, being combined with a weak perspective model. In that work, head occlusions are handled by exploiting facial texture symmetry and the model fitting is based on a numerically estimated gradient.

☆ This paper has been recommended for acceptance by J.-O. Eklundh.

* Corresponding author. Fax: +351 239 406672.
E-mail address: pedromartins@isr.uc.pt (P. Martins).
URL: http://www.isr.uc.pt/~pedromartins (P. Martins).

1077-3142/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cviu.2012.11.010

Natural extensions to 3D have also been proposed [12,14–17], with the 3D Morphable Model (3DMM) [18] one of the most popular. There are several differences between AAMs and 3DMMs. The 3DMMs are built from 3D range scans, therefore are usually constructed to be denser, including several thousands of vertices whereas the AAMs use only a few tens. The appearance model consists of 3D cylindrical folded textures that are densely aligned between all samples in the training set. This huge alignment step involves a modified optical flow, designed to operate on cylindrical coordinates, and smooth interpolation methods to fill in the registration holes. A reflectance model (the Phong model) is also used, i.e. the appearance model also uses surface normals. The large amount of data, due to the density of the 3DMMs, makes the algorithm quite slow, requiring several minutes to fit per frame (50 min using a SGI R10000 processor). Efficient 3DMMs, working under a scaled orthographic projection model and based on the IC algorithm, have also been proposed [19]. Still, its Jacobian and Hessian are only locally valid and take an average of 30 s per frame to fit, making it impracticable for real-time applications.

This paper addresses the fitting of a 3D shape deformable face model from a single view through 2.5D AAM. The 2.5D model can be viewed as a 3D sparse PDM whose projections define 2D control points for the 2D appearance. This means that 2.5D data has components of both 2D image data and 3D volumetric shape data. Consequently, the 2.5D model combines the advantages of both 3DMMs and 2D AAMs, in particular the robustness to pose changes and the fitting speed. Face alignment on this 2.5D dimensional space will carry an extra level of complexity since the IC approach is invalid in this case [20]. To deal with this problem, Matthews et al. [21] proposed a 2D + 3D AAM work around by exploiting the 2D and 3D shape models simultaneously. The shape instance generated by the 2D AAM is constrained to be consistent with the projection of a 3D *affine* shape (a 3D PDM is used, build from non-rigid structure from motion [22]). This constraint is formulated as part of the cost function, where a balancing weight is added and the value of this weighting constant is determined manually. In [22] is also showed that any 2D projection of a 3D shape model can be represented by a 2D shape model but at the expense of using up to six times more parameters than using a 3D model. However, a weak perspective projection model was used in this demonstration and this property does not hold for the perspective projection model. The solution described in this paper explores the advantages of using a single 3D model to constrain the possible 2D shape projection under the assumption of a full perspective model.

1.1. Paper contributions

The proposed solution extend the active appearance model approach to deal with matching a 3D face shape model to a single 2D image using a perspective projection model, whereas previous approaches have generally only dealt with scaled orthographic projections. This approach uses a single 3D metric PDM combined with a full perspective model. The use of a full perspective model carries an important advantage over the state of the art solutions. Assuming a calibrated camera, an estimation of the 3D Euclidean shapes can be obtained from a single image and face tracking can be performed by using cameras with short focal length and strong radial distortion (e.g. a low cost webcam). Compared to [21], no balancing weight is required since the approach is based on a single, low dimensional, 3D PDM.

Two algorithms to fit a 3D deformable shape model to a 2D image are proposed. Both algorithms seek to minimize the difference between the projected model and the target image using slightly different strategies: The Simultaneous Forwards Additive (SFA) and the Normalization Forwards Additive (NFA), both based on the Lucas-Kanade forwards additive [23] update step. The SFA algorithm is computationally expensive but more accurate. It searches for shape and appearance parameters simultaneously whereas the NFA projects out the appearance from the error image and searches only for the shape parameters. Although both solutions require evaluating several components per iteration, efficient approximations are proposed leading to an efficient update step. By comparison, our fitting solution is based on analytically derived gradients ("true gradients") rather than gradients approximated by numerical differences as in [9], genetic algorithms in [16] or generic optimization methods like the simplex in [15]. Finally, real-time performance can be achieving when using the efficient approximations, unlike the 3DMMs [18,19]. Moreover the methods used to acquired 3D dense shapes and textures normally demand very time consuming 3D reconstruction approaches or the use of expensive and cumbersome laser scan hardware.

Expanded solutions for the SFA and NFA are also proposed to handle self and partial occlusion, namely the Robust Simultaneous Forwards Additive (RSFA) and the Robust Normalization Forwards Additive (RNFA). These fitting methods use robust weighting functions that combine outlier estimation with pixel visibility extracted from the 3D pose.

In short, the main contributions in this paper are as follows:

- The use of a 2.5D AAM that combines a 3D metric Point Distribution Model (PDM) and a 2D appearance model whose control points are defined by full perspective projection of the PDM.
- A unique shape model is used where all the six degrees of freedom (6DOF) are modeled using a simple linear parametric model.
- Two model fitting algorithms and their computationally efficient approximations are proposed: the Simultaneous Forwards Additive (SFA) and the Normalization Forwards Additive (NFA).
- Robust solutions for the SFA and NFA are also proposed in order to take into account head partial and self occlusions.

Other 2D AAM related extensions such as using Light-Invariant theory to deal with external shading [24], multi-band appearance models [25–28] or modifying the cost function in order to include the previously aligned frame as an additional constraint (SICOV) [29] can be easily incorporated into the proposed algorithms with expected improvements on the overall performance.

1.2. Paper outline

This paper is organized as follows: Section 2 explains the 2.5D parametric model building process. The 3D PDM and 2D appearance models are both described in detail, as well as the full perspective camera model involved. Section 3 presents two model fitting algorithms, their respective efficient approximations and also the robust approaches to self and partial occlusion. In Section 4 is described how to efficiently evaluate the Jacobian of the warp for both shape and pose parameters, and the 2.5D AAM initial estimate problem is discussed in Section 5. Experimental results comparing both robust and non-robust fitting performances are presented in Section 6 and the results are discussed. Finally, Section 7 summarizes the paper.

As final note, we highlight that this 2.5D AAM image alignment method was first described in [30]. This journal paper describes the technique in more detail and includes the full derivation of the fitting algorithms and both Jacobians of the warp. New experiments and performance evaluation in new data sets are also presented.

Download English Version:

https://daneshyari.com/en/article/525656

Download Persian Version:

https://daneshyari.com/article/525656

Daneshyari.com