J. Vis. Commun. Image R. 35 (2016) 198–208

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Sequentially adaptive active appearance model with regression-based online reference appearance template ☆

Ying Chen [a,*], Chunjian Hua [b], Ruilin Bai [a]

[a] Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi, Jiangsu 214122, PR China
[b] School of Mechanical Engineering, Jiangnan University, Wuxi, Jiangsu 214122, PR China

ARTICLE INFO

Article history:
Received 15 October 2014
Accepted 20 December 2015
Available online 28 December 2015

Keywords:
Active appearance model
Model fitting
Incremental learning
Kernel regression
Facial features tracking
Individual generalization
Fitting context sensitivity
Tracking drifts

ABSTRACT

Statistically motivated approaches, such as the active appearance model (AAM), have been widely used for non-rigid objects registration and tracking. As an extension of AAM, sequential AAM (SAAM) was proposed, in which both an incremental updated component and a reference component were employed simultaneously in the fitting scheme. To make SAAM more adaptive to facial context variations during tracking, a regression-based online reference appearance model (ORAM) is presented to update the subject-specific appearance of the SAAM. The spatial map between scattered local feature correspondences and structured landmark correspondences is learned via Kernel Ridge Regression (KRR). Additionally, a shape deformation and appearance model evaluation strategies help to improve the accuracy and efficiency of the algorithm. The approach is experimentally validated by tracking face videos with improved fitting accuracy.

1. Introduction

Model-based image registration/alignment is a fundamental topic in computer vision. Active appearance models (AAMs) [1] and its extensions [2–4] form a powerful approach to important computer vision problems such as deformable shape segmentation and appearance modeling of deformable objects [5,6]. Recently, with the abundance of surveillance cameras and greater need for face recognition from video, methods to effectively fit an AAM to facial images in videos are of increasing importance [7].

Although widely studied and applied, AAMs suffer from many difficulties in practical applications. These difficulties are mainly embodied in the following aspects:

(1) Individual generalization: It has be shown in [8] that even with training images from a large set of subjects, generic AAM fitting to faces of an unseen subject can be hard because the appearance model of generic AAM only learns the appearance variation retained in the training data, and cannot represent the appearance information of the unseen test subject sufficiently well. It is commonly known as the individual generalization problem of the conventional AAM.
(2) Fitting context sensitivity: AAMs are extremely sensitive to facial contexts variation such as poses, expressions, and illumination changes. The AAM fitting performance is drastically degraded when the facial context characteristic of an input image is deviated from that of the training images. One simple way to solve this problem is to train the AAM with a large amount of training images with all possible context conditions. However, it is unpractical to collect the training images with all possible conditions and to represent all possible variations.
(3) Tracking drifts: The underlying assumption behind the model-based tracking is that the appearance of the object remains the same throughout the entire video sequence. However, this assumption is often being violated. One simple solution is to update the model every frame (or every n frames) with a new model extracted from the current image at the current location of the template. The problem with this naive strategy is that small errors are introduced in the location of the template, and accumulate with each update, which pushes the model steadily drifts away from the object [9,10].

To tackle the problems, incremental learning approaches were introduced to AAMs. Ref. [11] presented an adaptive AAM that updates the appearance basis vectors with the input image by the incremental principal component analysis (PCA) in person

☆ This paper has been recommended for acceptance by Yehoshua Zeevi.
* Corresponding author. Fax: +86 0510 85910633.
E-mail address: chenying@jiangnan.edu.cn (Y. Chen).

http://dx.doi.org/10.1016/j.jvcir.2015.12.012
1047-3203/

specific AAM. However, the method needed images of same person for training, which is generally unpractical. Liu proposed an adaptive hybrid AAM (AAAM) including a generic component and a subject-specific one which was learned and updated in an online fashion by making use of the test video sequence [7]. However, the performance of the algorithm is also strongly influenced by training set selection because of the generic component of the hybrid AAM. A new hybrid active appearance model (AAM) called sequential AAM (SAAM) based on online instance learning is presented in [12], in which the subspace of the subject-specific AAM component was initially learned from sequential registration results of first frame and was periodically updated through incremental PCA, and the drift correction component of the AAM was updated by selecting previous good fitting frame as a reference image. Although it performs well in most cases, the performance may decrease when large variation between neighbor frames occurs while tracking.

In this paper, we introduce a novel online reference appearance model (ORAM) to the SAAM tracking, and call it adaptive SAAM in which training samples for incremental person-specific PCA and the reference model for tracking drifts correction are generated via a non-linear map from Local feature correspondences to Landmark matching (L–L). The L–L map is learned offline from training set via Kernel Ridge Regression (KRR) [13]. The diagram of the adaptive SAAM tracking is illustrated in Fig. 1. Given the first frame landmark to start tracking, the AAM is initialized by setting the mean shape $\boldsymbol{s}_0$ to be the shape vector of the first frame, and the mean appearance $\boldsymbol{A}_0^0$ to be the warped appearance of the first frame. After deformation evaluation to confirm the necessity for ORAM update, local feature correspondence is performed and the result is taken as the input of the learned L–L map, together with the annotation of the first frame. The output of the map is the landmark estimation of the current frame, which would be taken for ORAM and SAAM update after the model evaluation and selection process. The updated ORAM and SAAM is then used for the image fitting of the current frame. The fitting result may also update ORAM and SAAM after the fitting evaluation, and then goes to the next frame.

The remainder of this paper is organized as follows: Section 2 introduces the model learning and fitting methods of the SAAM. Section 3 describes the proposed online appearance template in detail. Then we present how to introduce the online appearance template to SAAM in Section 4. Sections 5 and 6 present the experimental results and the conclusions respectively.

2. Sequential active appearance model

SAAM [12] is a incremental person-specific AAM which was proposed for AAM-based tracking in video sequences. The shape and appearance model of SAAM are updated with an online fashion, using sequential tracking results. With the model, facial

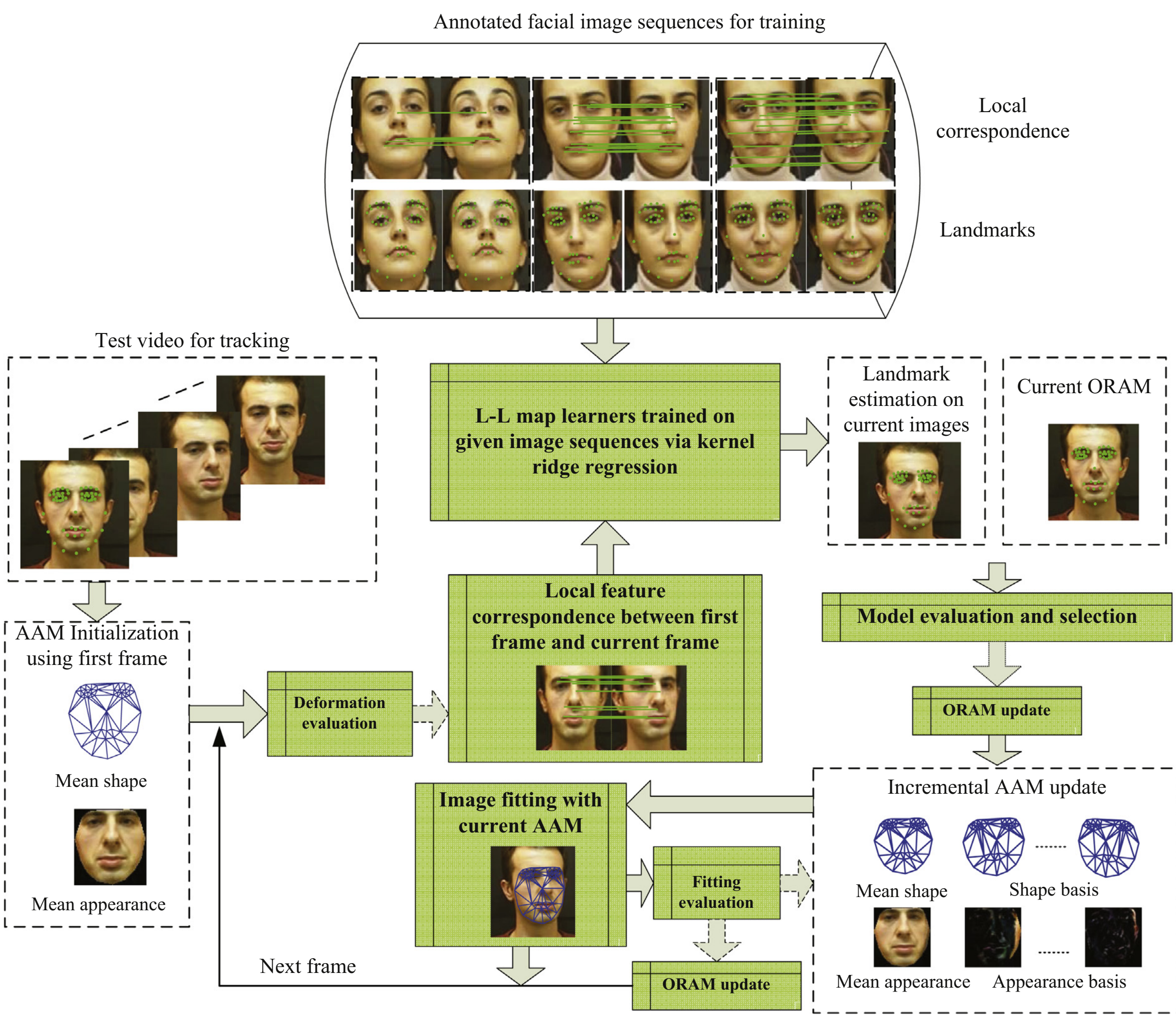

Fig. 1. Framework of the adaptive SAAM.