Regression-based Active Appearance Model initialization for facial feature tracking with missing frames ☆

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

[a] Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi, Jiangsu 214122, PR China
[b] Key Laboratory of System Control and Information Processing (Ministry of Education), Shanghai Jiaotong University, Shanghai 200240, PR China
[c] School of Mechanical Engineering, Jiangnan University, Wuxi, Jiangsu 214122, PR China

ARTICLE INFO

ABSTRACT

The Active Appearance Model (AAM) is receiving considerable attention in the field of facial analysis as a powerful method for modeling and segmenting deformable visual objects. Several extensions and improvements have been proposed on the original AAM, but AAMs maintain their dependence on the good initialization of model parameters to achieve accurate fitting results. AAMs are usually used directly in video tracking by searching on each subsequent frame that employs the fitting result of the previous frame for initialization. However, this model sometimes fails when large movements exist between two frames. This mechanism occurs when frames are dropped from the video due to the use of a lossy multimedia network. A regression-based approach for automatic AAM initialization is presented in this paper. After undergoing a scattered feature correspondence based on a dual-threshold matching strategy, the AAM shape points are initialized by the spatial map between local-landmark (L2L) correspondences. The map is learned based on Kernel Ridge Regression (KRR). The proposed method can successfully track the frames that are not identified with the general AAM trackers by establishing spatial relationship between local and landmark points. The initialization is robust to disturbances, which enables it to outperform key-feature-tracking or detection-based methods. We demonstrate the efficacy of the approach on two challenging facial videos with different training data and report a detailed quantitative evaluation of its performance.

1. Introduction

The Active Appearance Model (AAM) (Cootes et al., 2001) receives a significant amount of attention from the computer vision community in terms of deformable visual objects registration. A variety of applications are possible, including dynamic head pose and gaze estimation for real-time user interfaces, expression recognition, and lip reading.

AAMs generally treat registration as an optimization problem solved by local minimization methods. However, the gradient-descent-based fitting scheme is inherently dependent on good initializations. Poor initializations of model parameters cause the optimization to stick easily into the local minimum and diverge away from the target. This problem is prominent in the registration of images with considerable shape variations, which often occurs in lossy multimedia mobile network where some frames are unavailable due to unstable wireless connection and narrow bandwidth. Therefore, robust initializations for AAM tracking are extremely desirable to deal with frame loss in such a lossy environment.

The simplest method to initialize the model is through brute-force searching, which iteratively tests every possible configuration. However this method is extremely time-consuming because of the huge number of possible initializations. Stegmann (2000) suggested performing an AAM search in parallel with different initialization parameters, i.e., perturbed pose and model parameters. This method is less time-consuming compared to brute-force searching. However, it is still far from efficient, especially for real-time applications.

In most AAM fitting cases whose targets are facial images, the parameters of the shape model are roughly estimated by the detected face and facial features such as eyes, mouth centers, and nose tips (Rara et al., 2009; Rabie et al., 2008; Wong and Chung, 2010). After detecting these facial features, the AAM base mesh is warped to these points for AAM initialization. The initialization accuracy increases with more detailed facial features. However, the performance of these methods highly depends on the accuracy of feature detection, which will decrease at the presence of varying facial contexts and complex backgrounds. Wimmer (2008) used a

---

☆ This paper has been recommended for acceptance by C. Luengo.

* Corresponding author at: Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi, Jiangsu 214122, PR China. Tel.: +86 0510 85910631; fax: +86 0510 85910633.

E-mail address: chenying@jiangnan.edu.cn (Y. Chen).

learned Active Shape Model (ASM) fitting for AAM initialization because it provides stable results for the entire set of experiments, even in cases of poor initial parameter estimates as determined by a face detector.

In fitting an AAM to video sequences, conventional methods directly fit the AAM to each frame using the fitting results, i.e., shape and appearance parameters, of the previous frame as the initialization of the current frame (Ionita et al., 2011; Liu, 2010; Cristinacce and Cootes, 2008; Saragih et al., 2011; Sung and Kim, 2009). However, this method is only suitable for small movements between frames. Cui and Jin (2012) used the Lucas–Kanade optical flow to track several salient feature points, which were then used to constrain AAM shape initialization. The method considered interframe correspondences. However, its performance relied on salient feature tracking results, and only the similarity preservation of the two frames was considered. This approach limited the initialization performance because internal and external changes may disable salient feature tracking, and the transformation between two frames is usually more complicated.

One possible solution is to locate a sparse set of local points on each image and use them to conduct the initialization. In Feng et al. (2011), local feature matching between neighboring frames was adopted to predict the initial three-dimensional (3D) AAM parameters, wherein 3D pose estimation was conducted using a 3D-shaped model constraint. This method is also time-consuming and is designed for 3D-based face tracking.

A regression-based approach is proposed in the current paper for efficient two-dimensional AAMs initialization. Instead of looking for salient features like eyes and mouth, we use local sparse/scattered feature correspondence for AAMs initializations. The relationship between the local features and the global shape is obtained from the training data. By establishing an inter-frame relationship that combines the local and global facial features, compared to a previous work, the approach becomes more robust to external facial context changes (illuminations, viewpoints, etc.) and internal changes (expressions, glass-bearing, etc.). This improvement makes this approach more suitable for lossy network where the variation between neighbor frames may be large.

Fig. 1 shows a diagram of the proposed method. The landmarks in the first frame are manually annotated to initialize tracking. The AAM for the rest frames during tracking is initialized by a local-landmark (L2L) mapping based on Kernel Ridge Regression (KRR) (Trevor Hastie and Friedman, 2009). This method exploits the spatial relationship between scattered local invariant features and structured facial annotation points. To improve initialization accuracy, an improved local feature correspondence strategy called dual-threshold scale invariant feature transform (SIFT) matching is also presented in this paper as one of the supporting strategies.

The proposed AAM initialization framework has two main contributions: (1) a data-driven approach is proposed to identify the shape correspondence between sequential images of faces from their scattered local feature matching; and (2) an accurate match strategy for local feature correspondences of consecutive frames, which improves the accuracy of tracking results. The proposed initialization method helps AAMs to converge and accurately localize the facial features during tracking. This method outperforms other AAM initializations in terms of convergence rate and tracking accuracy, especially in lossy network where some frames are missing.

The remainder of this paper is organized as follows: Section 2 describes the proposed regression-based AAM initialization approach in detail. Section 3 introduces some supporting strategies for performance improvement. Section 4 and Section 5 present the experimental results and the conclusions respectively.

2. Regression based AAM initialization

We briefly introduce KRR in this section before investigating the initialization process through a map obtained from the scattered/sparse local feature correspondence space to the structured landmark space.

2.1. KRR

Let $X = [x_1, x_2, \ldots, x_N]$ be the data matrix of dimension(D_x, N), where N is the number of training samples, $x_i \in \mathbb{R}^{D_x \times 1}$ being a D_x-dimensional input. Let $Y = [y_1, y_2, \ldots, y_N]$ be the target matrix
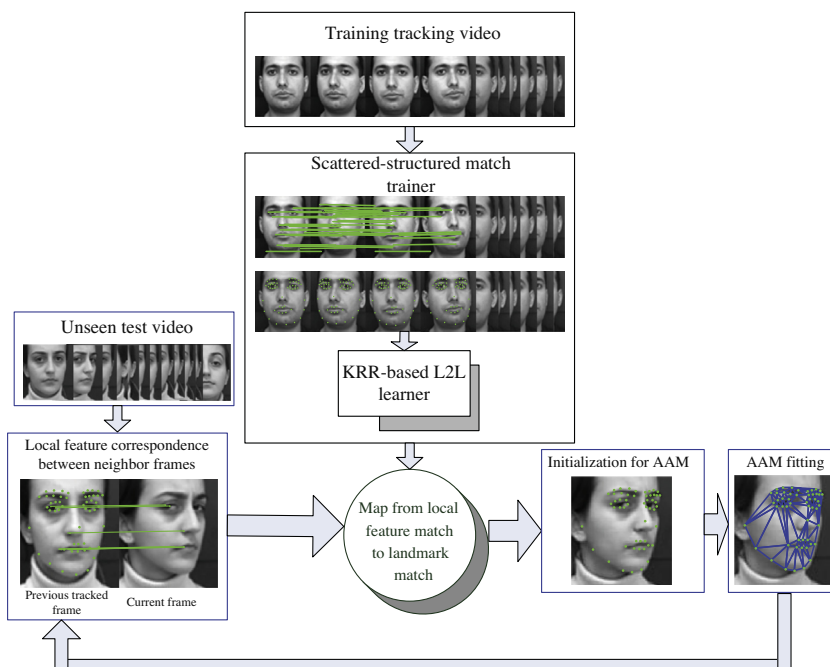
Fig. 1. Framework of the proposed method.