Signal Processing 110 (2015) 5–14

Contents lists available at ScienceDirect

Signal Processing

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

Accurate human face pose recovery from single image through generic shape regularization

Jian Zhang [a], Rui Sun [b], Xiaosi Zhan [a], Chaoqun Hong [c,*]

[a] School of Science and Technology, Zhejiang International Studies University, Hangzhou 310012, China
[b] Department of Electronic Technology, China Maritime Police Academy, Ningbo 315801, China
[c] Faculty of Computer Science, Xiamen University of Technology, Xiamen 361024, China

ARTICLE INFO

Article history:
Received 27 February 2014
Received in revised form
27 July 2014
Accepted 3 August 2014
Available online 11 August 2014

Keywords:
Face pose recovery
Rotation matrix
Generic shape
Newton method
Facial animation

ABSTRACT

In this paper, we propose a novel approach to recovering face pose from a single face image. The projection from 3D space to image plane is modeled by a scaled orthogonal projection, which contains the unknown pose parameters and face shape. To remove the ambiguity, the projection is regularized by prior information about generic face shape. Given a set of image features and a set of 3D features selected from a generic face, we can solve the unknown pose parameters conveniently through a Newton method. Owing to the shape regularization, the efficiency and accuracy of the proposed method precede the existed approach.

1. Introduction

Face pose recovery means discovering high-level concept of face direction from a given digital face image. The representation of face pose has a wide range of variation. From the view point of computer vision, 3D face pose recovery refers to estimating certain parameters on multiple degrees of freedom describing the position and orientation of a face.

Face pose recovery has important applications in various domains like digital entertainment [1–3] and video surveillance. For example, high quality facial animations in current 3D CG films and interactive video games are often created by performance driven technique. In this process, the actions of a human face are copied and transplanted to various virtual roles in films or games. Therefore, recovering 3D face pose parameters from given views is crucial to generating vivid virtual animations. In some applications of human computer interaction, facial expression recognition [4] plays an important role. The pose estimation and correction are expected to improve the recognition rate, because unknown pose of input facial expression has great impact on the recognition.

3D face pose recovery has received considerable attention in recent years. The approaches in this domain can be roughly classified into six categories which include image template matching methods, geometric feature methods, sample-based pattern classification methods, Shape-from-Motion (SfM) methods, regression methods and parameterized model fitting methods. Image template matching methods [5] compare the given face image with a set of sample images, each labeled with a pose, to find the closest image with the desired pose. Geometric feature methods [6] estimate the face pose using the geometric relationships between facial features. Pattern classification methods [7] build a binary classifier using face images labeled with a pose, and feed the input image to the classifier. If the classification result is positive, the pose of

* Corresponding author.
E-mail address: cqhong@xmut.edu.cn (C. Hong).

http://dx.doi.org/10.1016/j.sigpro.2014.08.001

input face is the same as that of the positive sample images. When multiple images of a human face are available, SfM methods [8] can be used to decompose the matrix of facial features taken from the images into matrix of face shape and matrix of face pose parameters. Regression methods [9] build a mapping from image or image features to pose parameters using a set of labeled face image samples, then they can recover the pose of any given face image through the mapping. Parameterized model fitting methods [10,11] need to construct a statistic model based on 2D or 3D face samples, then fit the model to given face image by adjusting the model and pose parameters. The desired pose parameters are obtained when the model fitting is accomplished.

The current face pose recovery methods have some obvious disadvantages. First, the accuracy of pose recovery cannot be ensured, since some methods are based on simple image comparison [5], and some other methods are reliable only when recognize several discrete poses [7]. Second, many methods adopt machine learning to achieve pose recovery, and the recovery result heavily depends on the amount and diversity of the training samples [9]. Last, some methods have high computational complexity which is caused by the training process or the model fitting process [11].

In this paper, we propose a novel face pose recovery method which is free from the above mentioned disadvantages. Unlike the existing methods, we aim to build equations comprising the pose parameters according to certain camera projection model, and solve for the parameters directly. To remove the ill-posedness of the problem, a generic face shape is used to regularize the model. The proposed method can recover face pose accurately from single input face image. It is worthwhile to highlight several advantages of our method: (1) It has high accuracy. With the generic shape regularization, the solution is very close to the globally optimized solution. (2) The model can be solved directly, so it does not need any training sample. Accordingly, it is not restricted by the amount and diversity of samples. (3) The computational efficiency is very high. We will demonstrate that the model can be solved by a simple Newton method.

The rest part of this paper is organized as follows: Section 2 shows the related works; in Section 3, we describe the proposed method in detail; Section 4 demonstrates the experimental results, and the last section concludes this paper.

2. Related works

The simplest pose recovery methods are the image template matching methods which compare the given image to a set of sample images labeled with face pose information, and the pose of given image is that of the closest sample image [12]. The comparison is based on certain distance metric [13]. These methods are quite simple, and the sample image set can be expanded to any scale to include more pose variation. However, they can just estimate discrete poses, and with the enlargement of the sample set, the computational complexity of comparison also rises. To speed up the comparison, some researchers tried to train Support Vector Machines (SVM) and used the support vectors instead of sample images for pose recovery [14,15]. Another problem is that these methods regard the similarity between images as the similarity between corresponding poses. This is not always true because the image variation may be caused by different illumination or expressions.

Geometric feature methods extract a few salient facial feature points from given image, and estimate the pose parameters according to the relative position of these feature points. Sheng et al. [16] recovered the roll, yaw and pitch angles of 3D face using the center points of two eyes, the middle points of two eyes and the mouth center of the planar face. Another estimate of pose can be obtained using a different set of five points (the inner and outer corners of each eye, and the tip of the nose) [17]. The methods are fast and simple, but the result is often affected by the precision of the feature detection. More over, the result of geometric-based methods is usually not very accurate.

Pattern classification methods [18] achieve pose recovery by training a binary classifier for each distinct pose using positive and negative sample images. Huang et al. [7] trained three SVMs to recognize three discrete yaw angles from given image. Zhang et al. [19] used five FloatBoost classifiers for five-view face pose recovery. To improve the computational efficiency of pose recognition, Jones and Viola [20] adopted a cascaded AdaBoost classifiers. To improve the performance of SVM classifier, Tao et al. [21] proposed an asymmetric bagging-based SVM to address the problem caused by small-sized training set. However, like template matching methods, the methods based on classifiers can just perform discrete pose recognition. To recognize more poses, multiple classifiers must be constructed. But the number of classifiers can not increase unrestrictedly. The reason is that when two classifiers are designed to recognize very similar poses, they may simultaneously classify an input image as a positive image. To avoid building a distinct classifier for each pose, Linear discriminant analysis (LDA) was adopted in [22] for head pose estimation. These methods usually have poor generalization since the recognition result is restricted in the scope of pose covered by the sample images. In addition, the difference between faces in different classes is usually caused not only by the pose discrepancy, but also by the illumination and expression discrepancy. This is a big obstacle to traditional classification algorithms. Recent improvements to LDA include the general tensor discriminant analysis (GTDA) [23] and geometric mean for subspace selection [24]. GTDA provides high recognition rate because it preserves the discriminative information of the training images. The geometric mean-based subspace learning projects the images from different classes apart in the subspace even when the dimension of the projected subspace is lower than the number of classes. To measure the image similarity, [13] proposed a semantic preserving distance metric that could encode the semantic contents. A high-order distance-based multimodal stochastic learning method was proposed in [25] to achieve image classification with image features from multiple sources. Also, an improved graph-based method was successfully applied to image classification in [26]. The aforementioned methods may find their usage in pose classification.