Available online at www.sciencedirect.com

ScienceDirect

Pattern Recognition Letters 28 (2007) 895–903

Pattern Recognition Letters

www.elsevier.com/locate/patrec

Integrating monolithic and free-parts representations for improved face verification in the presence of pose mismatch

Simon Lucey *, Tsuhan Chen

Advanced Multimedia Processing Laboratory, Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Received 23 February 2005; received in revised form 10 October 2006
Available online 23 December 2006

Communicated by P.J. Flynn

Abstract

This paper concentrates specifically on the task of verifying faces when the gallery set stems from frontal face images with the probe set stemming from a number of alternate poses (i.e. pose mismatch). An argument is put forward for attempting to recognize faces through integrating holistic/monolithic and free-parts representations of the face. A contribution is made via the analysis of what traits, in a face, are most useful for each representation. As a result we are able to demonstrate that there is: (a) benefit in combining free-parts and monolithic representations, and (b) further benefit can be obtained by varying the weight placed on each representation as a function of viewpoint.

Keywords: Face recognition; Pose mismatch; Gaussian mixture models; Monolithic representation; Parts representation

1. Introduction

Face verification with a change in viewpoint, between 2D gallery and 2D probe images, is inherently a difficult task. Images taken of the face from one pose, for the same subject, are markedly different to images captured under another pose. One can tell from visual inspection that pixel variation due to pose change is far greater than the variation seen due to changes in identity. An example of this problem can be seen in Fig. 1. In this paper we will be dealing specifically with the problem of trying to verify clients from non-frontal viewpoint probe images given that only a single frontal view image of that client exists in the gallery.

In cognitive science, theories abound over whether humans recognize faces based on component parts or holistic representations. In fact there is a large amount of literature (Tanaka and Farah, 2003; Murray et al., 2003) indicating that both types of representations of the face are important in human face recognition in the presence of pose mismatch. We use the term *monolithic* in this paper to describe the holistic vectorized representation of the face based purely on pixel values within an image array, which can be associated with the holistic mechanism used in a human face recognition system. Similarly, we use the term *parts* to denote a representation of the face that can be considered as an ensemble of image patches of the image array. The employment of parts representations for object/face detection has recently gained much attention and success in machine vision literature (Schneiderman et al., 2000; Weber et al., 2000a,b). For the task of face recognition we additionally categorize parts representations into two subsets namely *rigid*- and *free*-parts. Rigid-parts representations assume the position/structure of the patches within the image is preserved. Free-parts representations assume that the position/structure of patches within the image can be relaxed so they can "freely" move to varying

* Corresponding author. Tel.: +1 412 268 2889; fax: +1 412 268 3890.
E-mail addresses: slucey@ieee.org (S. Lucey), tsuhan@cmu.edu (T. Chen).

0167-8655/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2006.12.006

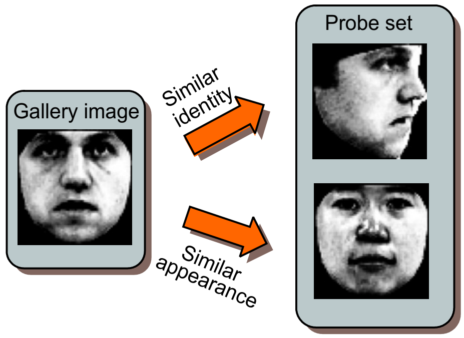

Fig. 1. Example of the difficulty in recognizing subjects from a different pose as images from the same pose, irrespective of identity, are more similar in terms of their pixel representation.

extents. Both rigid- and free-parts representations assume there is minimal dependence between the appearance of other patches within the image.

Considerable work has already been performed with monolithic face representations, for automatic face recognition, in the presence of pose mismatch. Most notably techniques like Tensorfaces (Vasilescu and Terzopoulos, 2002), Eigenlight fields (Gross et al., 2004) and Fisherfaces (Lee and Kim, 2004) have been employed with varying degrees of success. There has also been some preliminary work by Kanade and Yamada (2003) demonstrating the benefit of a rigid-parts representation. In this method weightings for each patch in the face are learnt off-line, from a world set, as a function of pose. Hitherto, the benefit of employing a free-parts representation has not been fully investigated for the task of automated face verification in the presence of pose mismatch. Free-parts representations have an inherent advantage over monolithic and rigid-parts representations in that they compare "distributions" which are naturally able to cope with appearance variation. In this paper we will be focussing on comparing free-parts and monolithic representations as they are representative of "point" and "distribution" style classification mechanisms for verification.

Recent work (Lucey et al., 2004; Sanderson and Paliwal, 2003; Eickeler et al., 2000) has been conducted demonstrating that good performance can be attained by employing a free-parts representation in the task of frontal view face verification. Some generative models that have been previously employed to model these free-parts face distributions are: pseudo 2-D hidden Markov models (HMMs) (Eickeler et al., 2000) and Gaussian mixture models (GMMs) (Lucey et al., 2004; Sanderson and Paliwal, 2003). GMMs can be thought of as a special subset of HMMs where no positional constraints are placed on the patch observations whatsoever. This is a highly desirable characteristic when trying to verify clients across pose as patch positions can vary wildly across viewpoints.

In this paper we will attempt to address the following two questions with respect to face verification via monolithic and free-parts representations:

- Q1: Are areas of the face which are often associated with being the most salient and discriminative (i.e. eyes, nose and mouth) equally important for all representations of the face? Or can other traits such as skin texture play a larger role depending on the representation employed?
- Q2: Is there any benefit in combining the match-scores resulting from a free-parts and monolithic representation? Can additional benefit be gained by combining these scores in an unequal manner?

As a result of answering the above questions we will also be presenting an algorithm which we refer to as the free-parts and holistic integration (FHI) strategy. The FHI strategy is able to give substantial performance improvement in comparison to current monolithic and free-parts approaches in the presence of pose mismatch.

2. Monolithic representations

It is outside the scope of this paper to perform a large scale evaluation of all possible monolithic approaches. Instead we will be taking a sample of techniques that are representative of current paradigms in pose robust face recognition. These paradigms differ largely by how they employ the world set in their off-line training. We define the world set as the set of observations used to obtain any data-dependent aspects of the verification algorithm (e.g. subspace, distribution, classifier, etc.), but does *not* provide any client specific information like those found in the on-line gallery and probe sets.

Specifically, we will be considering the Eigenface algorithm (Turk and Pentland, 1991) as a baseline due to its ubiquitous nature in face recognition literature. The Eigenface algorithm can be thought of as being representative of a paradigm that make matches based purely on pixel appearance. The Fisherface algorithm (Belhumeur et al., 1997) is also considered as a baseline due to its simplicity and high performance in recent evaluations (Navarrete and Ruiz-del-Solar, 2002; Ruiz-del-Solar and Navarrete, 2002; Sadeghi et al., 2003). This algorithm can be thought of as being representative of a paradigm that attempts to learn the within-class and between-class differences between poses in the world set. Finally, the Eigenlightfields technique will be used as a baseline due to its specificity to pose and its similar nature to other popular approaches such as Tensorfaces (Vasilescu and Terzopoulos, 2002) as well as the pose transformation technique of Lee and Kim (2004). These types of algorithms are representative of a paradigm that attempts to learn the relationships/transformations between each pose in the world set.

2.1. Eigen and Fisherfaces

Eigen and Fisherface approaches have been around for quite some time and have enjoyed much success in frontal pose face recognition. In this paper we will be evaluating a