Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computer Vision and Image Understanding 99 (2005) 303–331
Computer Vision and Image Understanding
www.elsevier.com/locate/cviu

Visual tracking and recognition using probabilistic appearance manifolds

Kuang-Chih Lee [a,*], Jeffrey Ho [b], Ming-Hsuan Yang [c], David Kriegman [b]

[a] *Beckman Institute and Computer Science Department, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*
[b] *Computer Science and Engineering Department, University of California at San Diego, La Jolla, CA 92093-0114, USA*
[c] *Honda Research Institute, 800 California Street, Mountain View, CA 94041, USA*

Received 13 August 2004; accepted 28 February 2005
Available online 12 April 2005

Abstract

This paper presents an algorithm for modeling, tracking, and recognizing human faces in video sequences within one integrated framework. Conventional video-based face recognition systems have usually been embodied with two independent components: the tracking and recognition modules. In contrast, our algorithm emphasizes an algorithmic architecture that tightly couples these two components within a single framework. This is accomplished through a novel appearance model which is utilized simultaneously by both modules, even with their disparate requirements and functions. The complex nonlinear appearance manifold of each registered person is partitioned into a collection of submanifolds where each models the face appearances of the person in nearby poses. The submanifold is approximated by a low-dimensional linear subspace computed by principal component analysis using images sampled from training video sequences. The connectivity between the submanifolds is modeled as transition probabilities between pairs of submanifolds, and these are learned directly from training video sequences. The integrated task of tracking and recognition is formulated as a maximum a posteriori estimation problem. Within our framework, the tracking and recognition modules are complementary to each other, and the

* Corresponding author.
E-mail address: klee10@uiuc.edu (K.-C. Lee).

1077-3142/$ - see front matter
doi:10.1016/j.cviu.2005.02.002

capability and performance of one are enhanced by the other. Our approach contrasts sharply with more rigid conventional approaches in which these two modules work independently and in sequence. We report on a number of experiments and results that demonstrate the robustness, effectiveness, and stability of our algorithm.

Keywords: Face tracking; Face recognition; PCA; Appearance manifold

1. Introduction

In the past few decades, there has been intensive research and great strides in designing and developing algorithms for face recognition with still images. Only until recently has the problem of face recognition with video sequences started to attract the attention of the research community [1–4]. This can be partly attributed to the recent advance in computer hardware. In particular, with low cost cameras and sufficiently powerful personal computers, it is now possible to inexpensively implement a real-time face tracking system (e.g. [5,6]) with good performance. This capability is the prerequisite for developing real-time video face recognition applications.

Compared with conventional still image face recognition, video face recognition offers several challenges and opportunity. First, there is the "alignment" problem between the tracking and the recognition modules. A video-based face recognition system invariably has two components, i.e., tracking and recognition modules. Since tracking and recognition problems have been studied intensively but separately in the past, these two modules are usually implemented independently and work in sequence. Without any alignment between the two modules, the images returned by the tracker generally are not be in good agreement with the appearance model used by the recognition module,[1] i.e., misaligned images. Unfortunately, virtually all appearance-based recognition techniques are sensitive to misalignments. Therefore, some mechanism should be in place to ensure that the images returned by the tracking module can be correctly processed by the recognition module.

Second, there is the problem of modeling appearance variation of faces for both the tracking and recognition modules. At the heart of any tracking or recognition algorithm is an internal representation which defines the allowable variation in appearances of the object to be tracked or recognized. Factors such as changes of viewpoint, shape (deformations, articulations), and illumination, individually or combined, can cause significant image variations in a dynamic environment. (See Fig. 1). For appearance-based methods, some (if not all) of these image variations should be modeled in order to produce robust results. However, due to their different missions, tracking and recognition modules generally place different emphasis and requirement on their internal model or representation. For

[1] In this paper, our main focus is on appearance (or image)-based recognition methods. For face recognition, it has been argued [7,8] that feature-based techniques are generally less stable and accurate.