Affordable 3D Face Tracking Using Projective Vision

D.O. Gorodnichy [1]*, S. Malik [2], G. Roth [1]
[1] Computational Video Group, National Research Council, Ottawa, Canada K1A 0R6
[2] School of Computer Science, Carleton University, Ottawa, Canada, K1S 5B6
http://www.cv.iit.nrc.ca/research/Stereotracker

Abstract

For humans, to view a scene with two eyes is clearly more advantageous than to do that with one eye. In computer vision however, most of high-level vision tasks, an example of which is face tracking, are still done with one camera only. This is due to the fact that, unlike in human brains, the relationship between the images observed by two arbitrary video cameras, in many cases, is not known. Recent advances in projective vision theory however have produced the methodology which allows one to compute this relationship. This relationship is naturally obtained while observing the same scene with both cameras and knowing this relationship not only makes it possible to track features in 3D, but also makes tracking much more robust and precise. In this paper, we establish a framework based on projective vision for tracking faces in 3D using two arbitrary cameras, and describe a stereo tracking system, which uses the proposed framework to track faces in 3D with the aid of two USB cameras. While being very affordable, our stereotracker exhibits pixel size precision and is robust to head's rotation in all three axis of rotation.

1 Introduction

We consider the problem of tracking faces using a video camera and focus our attention on the design of the vision-based perceptual user interface systems [28]. The main applications of these systems are seen in HCI, teleconferencing, entertainment, security and industry for disabled [27, 30].

Being a high-level vision problem, face tracking problem poses four major challenges: robustness, precision, speed and affordability. While the last two have become much less critical over the last few years due to the significant increase of computer power and decrease of camera cost, the first two remain unresolved.

The approaches to face tracking can be divided into two classes: global (image-based) and local (feature-based) approaches [10, 31]. Global approaches use global cues like skin colour, head geometry and motion, are more robust, but cannot be used for pixel-size precision tracking. On the other hand, local approaches, which are based on tracking facial features, can theoretically track very precisely. In practice however they do not, because they are very unrobust, due to the variety of motions and expressions a face may exhibit.

While for humans it is definitely easier to track objects with two eyes than with one eye, in computer vision face tracking is usually done with one camera only. A few authors do use stereo for face tracking [14, 15, 20, 29]. They however use the second camera mainly for the purpose of acquiring the third dimension rather than making tracking more robust, precise or affordable. In fact, conventional stereo setups are usually precalibrated and quite expensive.

The problem is that human brain knows and makes use of the relationship between the images while processing them [2, 3], while computers do not. Recent advances in computer vision though provided the methodology based on projective vision that allows one to compute this relationship for any two cameras [9, 23]. This relationship is naturally obtained while observing the same scene with both cameras and is represented by the fundamental matrix which relates the two images to one another.

This paper describes how to use the projective vision techniques for tracking faces with two arbitrary cameras. The most significant result is that computing the fundamental matrix not only allows one to recover the 3D position of the object with low-cost cameras, but also makes tracking much more robust. Combining the proposed projective-vision-based tracking approach with the robust convex-shape nose tracking technique described in [5] allowed us to build a stereo tracking system, which is able to track faces in 3D using two generic USB cameras. The robustness of the system, the binary code of which can be downloaded from our website, is such that the rotations of a head of up to 40 degrees in all three axis of rotation can be tracked.

The paper is organized as follows. After presenting the outline of our framework for affordable stereotracking (Section 2), we recap the projective vision properties which make the framework possible (Section 3). This is followed by the description of the stereo selfcalibration procedure

*The author the correspondence should be sent to.