2010 IEEE International Conference on Robotics and Automation
May 9-13, 2011 Shanghai China

Head Pose Estimation by Imperceptible Structured Light Sensing

Jingwen Dai and Ronald Chung

Dept. of Mech. & Automation Engineering , The Chinese University of Hong Kong

1. Introduction

This article describes a method of determining all the 6 DOFs of head pose in space by the use of an imperceptible structured light system. We employ projection that appears as white light (and thus is less annoying) to humans, but underneath actually embeds coding patterns that can facilitate 3D reconstruction. The method tracks 3D salient feature points on human face accurately without the need of going through prior training. Firstly, through an elaborate pattern projection strategy and camera-projector synchronization, a pattern-illuminated image and the corresponding scene-texture image are captured. Then in the point cloud generated by structured light sensing, the facial feature points in the scene-texture image — all localized by AAM — will have their 3D positions interpolated. Correspondences between such facial features in 3D with those associated with the previous or reference image frame can then be constructed. Finally, from such point pairs in 3D, the head orientation and translation are determined. Extensive experiments show that the proposed method is effective, accurate, and fast in determining the 6 DOFs ofthe head pose, making it suitable for use in real-time applications.

2. Method

❑ Pattern Projection Strategy for Imperceptible Structured Light Sensing

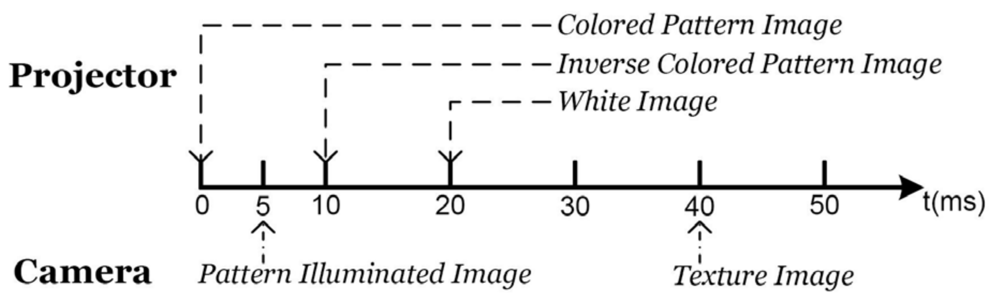
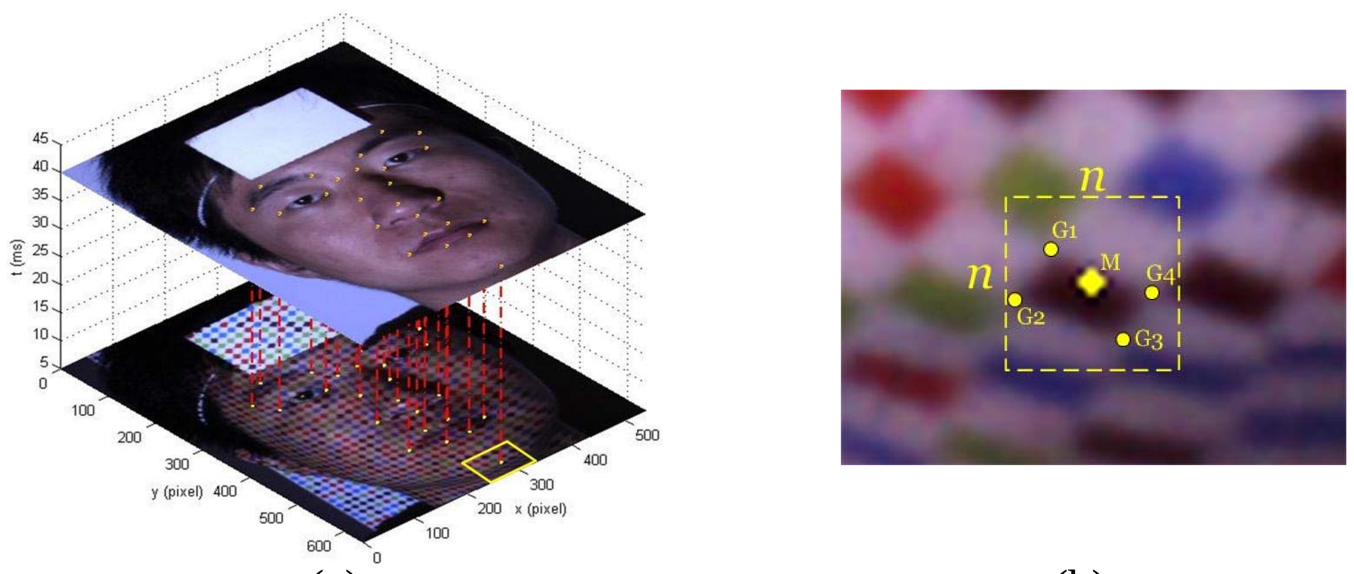


Fig. 1. Capture-Projection Synchronization Strategy

Here one capture-projection cycle, illustrated in Fig. 1, is used to describe the strategy of pattern projection. To achieve imperceptible structured light projection, the frequency of projection must exceed the flicker fusion threshold, which is 75*Hz* for most of the people.



(a) (b)

Fig. 2. 3D facial feature landmarking by interpolation: (a) Feature points in the scene-texture image and the corresponding mirror points in the pattern-illuminated image. (b) One mirror point and its neighboring grid points in an $n \times n$ window.

❑ Facial Feature Localization

- Locating 2D Positions of Key Facial Feature Points in Scene-texture Image

Firstly, Adaboost face detection method is employed to extract the position of the face in the image. Then AAM method is applied to locate the facial features in the segmented face image, as shown in Fig. 2 (a).

- Determining 3D Positions of Grid Points in Pattern-illuminated Image

Color coding scheme is used to determine the 3D position of the grid points in the pattern-illuminated image.

- Inferring 3D Positions of Key Facial Features

3D positions of the key facial feature points are derived by a combined use of the 2D facial feature points in the scene-texture image that are localized by AAM, and the point cloud generated by structured light sensing. The interpolation process is illustrated in Fig. 2.

❑ 6-DOF Head Pose Estimation

Head orientation and translation are estimated by SVD of a correlation matrix that is generated from 3D point pairs between consecutive frames.

$$H = \sum_{i=1}^{N} p'_{c_i} {p_{c_i}}^T = U \Lambda V^T,$$

$$\hat{R} = UV^T,$$

$$\hat{T} = \bar{p}' - \hat{R}\bar{p}.$$

where $p_{c_i} = p_i - \bar{p}$, $p'_{c_i} = p'_i - \bar{p}'$, $\bar{p} = \frac{1}{N}\sum_{i=1}^{N} p_i$, $\bar{p}' = \frac{1}{N}\sum_{i=1}^{N} p'_i$.

3. Experimental Results

To assess the feasibility of the proposed head pose estimation method using imperceptible structured light sensing, we conducted an accuracy evaluation experiment.

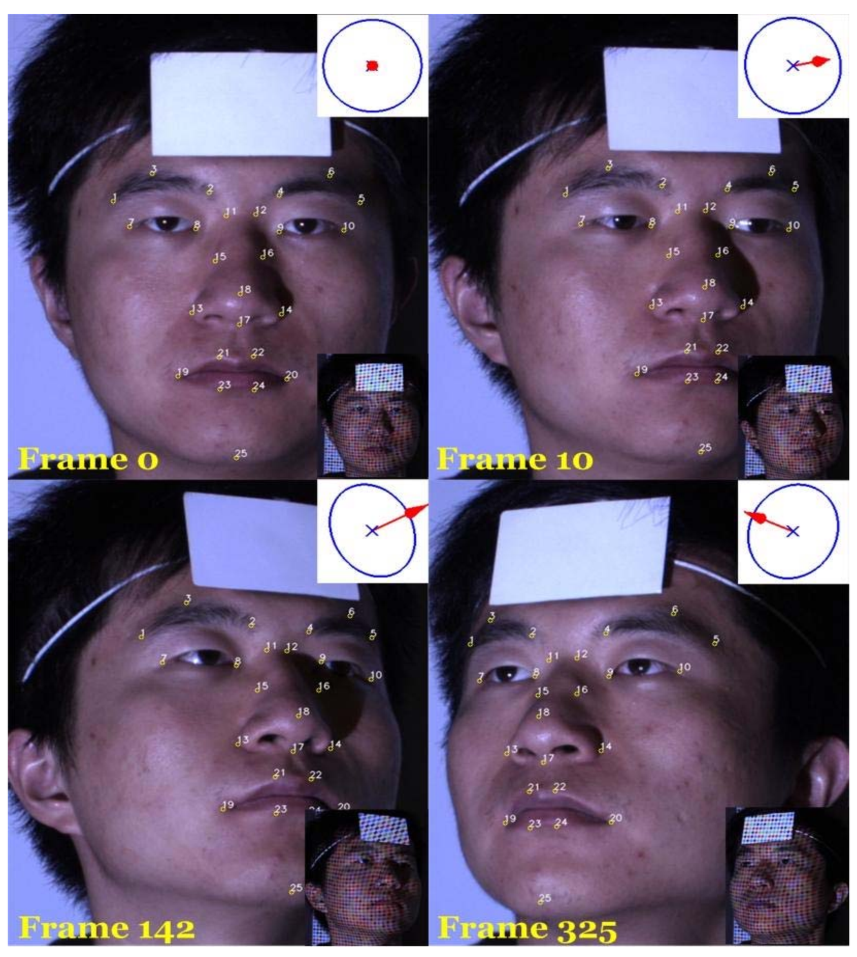


Fig. 3. Experimental results

Experimental results at some frames of a subject are presented in Fig. 3. In each sub-figure the AAM-located feature points are indicated by yellow circles in the corresponding scene-texture image. The inset image at the bottom-right corner of each sub-figure shows the corresponding pattern-illuminated image, while the inset image at the top-right represents a qualitative description of the estimated head pose, in which the ground-truth and the estimated head pose are implied by a blue circle (or ellipse) and a red arrow respectively.

The mean absolute estimation error of the proposed method, along with those of three other systems, are shown in Table I. The comparison should be considered as a reference only, since the evaluation data-sets and the systems used to obtain the ground-truth are not exactly the same.

Table I. Comparison of Pose Estimation Errors

Method	Sensing	Mean Absolute Error (°)		
		Yaw	Pitch	Roll
Murphy-Chutorian [3]	Monocular	3.39	4.67	2.38
Morency [7]	Stereo	3.50	2.40	2.60
Jimene [8]	Stereo	1.85	1.61	1.20
Our method	ILS	2.02	1.18	0.76

Acknowledgment

This work is affiliated with the Microsoft-CUHK Joint Laboratory for Human-centric Computing & Interface Technologies. The work described was partially supported by the Chinese University of Hong Kong 2009-2010 Direct Grant (Project No. 2050468).

Paper ID: 0104

Contact email: jwdai@cuhk.edu.hk