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

Head Pose Estimation by Imperceptible Structured Light Sensing

Jingwen Dai and Ronald Chung

Abstract— **We describe a method of estimating head pose in space by imperceptible structured light sensing. Firstly, through elaborate pattern projection strategy and camera-projector synchronization, pattern-illuminated images of the subject and the corresponding scene-texture image are captured under imperceptible patterned illumination. 3D positions of the key facial feature points are then derived by a combined use of (1) the 2D facial feature points in the scene-texture image that are localized by AAM, and (2) the point cloud generated by structured light sensing. Eventually, the head orientation and translation are estimated by SVD of a correlation matrix that is generated from the 3D corresponding feature point pairs over the various image frames. Extensive experiments show that the proposed method is effective, accurate, and fast in 6-DOF head pose estimation, making it suitable for use in real-time applications.**

I. INTRODUCTION

Head pose estimation has continuously been an active research subject for its usefulness in a variety of applications. In human-computer interaction, head pose is an important cue for computer or robot to infer the intention of human [1]. For some face-related applications like face alignment, face recognition, and facial expression recognition, estimating the pose of the face is considered as a precondition or preprocessing step [2]. For driver-assistance systems, head pose estimation is essential for inferring the driver's focus of attention [3].

In the context of computer vision, head pose estimation is most commonly interpreted as the ability to infer the orientation and translation of a person's head from image data with respect to a camera. If the human head is regarded as a disembodied rigid object, the human head motion is limited to six degrees of freedom (DOFs), three for orientation that is characterized by pitch, roll, and yaw, and three for translation along say three orthogonal directions in space.

The adoption of structured light illumination has been proven to be an effective and accurate visual means for 3D reconstruction [4]. Structured light system (SLS) consists of a projector that projects controlled patterns to the target object, and a camera capturing images of the illuminated object. Once correspondences between positions on the projector's pattern panel and positions on the camera's image plane are established through the use of some coding strategies on the illuminated patterns, simple triangulation over the light rays from the projector and the corresponding light rays to the camera would recover 3D information about the target object. Recently, the availability of pico projectors with physical size of only $4 \times 2 \times 1$ inches has further extended the application domain of SLSs. Nikon Corp. has even integrated an ultra-small built-in projector into its latest digital camera COOLPIX S1000pj, making it possible to implement SLS in even hand-held consumer electronic products.

However, light projection device's intrinsic characteristics could lead to disadvantages in specific circumstances: it could lead to loss or corruption of colorimetrical and textual information of the lighted surfaces, to inconsistency of optical flow, and even to offensive, possibly hazardous aspects of illumination (e.g., direct LASER illumination to human face for face measurement etc. could arouse eye discomfort and even injury).

To benefit from the merits of SLS while avoiding the drawbacks, researchers have designed SLSs in the non-visible spectrum [5] or in an imperceptible way. Three major approaches are InfraRed Structured Light (IRSL), Filtered Structured Light (FSL), and Imperceptible Structured Light (ISL). In particular, ISL is easy to implement, since it requires similar equipments as those of regular projection: a digital projector, and cameras. The light source projects a light pattern (which is related to structure coding of the illumination) followed by its complement (the inverse pattern) onto the scene, at high frequency so as to hide the coding pattern from humans and make the illumination appear uncoded and uniform. The first camera is synchronized with the projection of the first illuminated pattern (the coding pattern) to achieve 3D reconstruction, just like in the traditional structured light methods; the second one has long integration time and observes the scene under uniform-like illumination to capture a gray-level or colored image representing scene texture.

This article describes a method of determining the 6-DOF head pose by the use of an imperceptible structured light system. The method is able to track accurate 3D positions of salient facial landmarks without the need of going through any training process. Firstly, through elaborate pattern projection strategy and camera-projector synchronization, a pattern-illuminated image and the corresponding scene-texture image are captured under illumination that appears as white light yet embeds coding patterns. Subsequently, in the point cloud generated by structured light sensing, the facial feature points in the scene-texture image 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, the head orientation and translation are estimated by SVD of a correlation matrix that is generated from such point pairs in 3D.

All authors are with the Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong. E-mail: {jwdai,rchung}@mae.cuhk.edu.hk

978-1-61284-385-8/11/$26.00 ©2011 IEEE

The remainder of this article is structured as follows. In Section II, related works on head pose estimation and imperceptible structured light sensing are briefly reviewed. The essential processes of the proposed method including pattern projection strategy, facial landmark localization, and 6-DOF head pose estimation are described in Section III. In Section IV, the system setup and experimental results are shown. Conclusion and possible future work are offered in Section V.

II. RELATED WORK

A. Head Pose Estimation

Due to the immense potential applications of head pose estimation, a variety of approaches to the problem have been proposed in the past decade. A comprehensive literature review has been recently carried out by Murphy-Chutorian and Trivedi [6]. Below we outline a few key works related to our work.

Compared to the use of 2D texture information (such as points, edges etc.) in the image data, the use of 3D information and 3D face models is inherently more direct and accurate for pose or motion estimation in 3D. Morency *et al.* used depth and intensity view-based eigenspaces to build a prior model from the first image frame that is then robustly tracked [7]. Jimenez *et al.* built a 3D face model using points chosen by SMAT in the first image frame [8]. Through stereo correspondence of the two cameras, the 3D coordinates of these points can be extracted, and the points are tracked in the subsequent frames and 3D pose are calculated at each frame by RANSAC and POSIT.

Some methods have been presented that work on range-scan data. Based on a novel shape signature to identify noses in range images, Breitenstein *et al.* generated candidates for the nose positions, and inferred and evaluated many pose hypotheses [9]. The pose is estimated using an error function that is employed to compare the input range image with the pre-computed pose image of an average face model.

In the above methods, though 3D acquisition systems are there to provide accurate and dense data, the vast amount of data needed also demands the use of powerful parallel processors (GPU), or else there could be difficulty in processing the data in real time.

Besides the above, hybrid methods that combine one or more methods have been studied and they showed good performance in pose estimation. Murphy-Chutorian *et al.* had a system based on localized gradient orientation histograms, that are integrated with support vector machines for regression [3]. However, some training processes on some previously prepared training sets are needed in the learning-based method, which could be tedious and time consuming.

In this work, we derive the 3D positions of key facial feature points from a sparse point cloud generated from an ISL system. The system requires no training process. The low computational complexity of the system also makes real-time performance possible.

B. Imperceptible Structured Light

A first proof-of-concept system that embeds invisible structured light patterns into DLP (Digital Light Processing) projections was introduced in the "Office of the Future" project [10]. In the work, binary codes are embedded by projecting coded images and their complements in temporally alternating manner. Provided that the frequency of projection reaches the *flicker fusion threshold* ($\geq 75Hz$), the coding pattern and the inverse pattern are visually integrated over time in human perception, and the illumination has the appearance of a flat field ("white" light) to humans. However, embedding structured light into DLP projections was made possible only with extensive modifications of the projection hardware and firmware, including removal of the color wheel and reprogramming of the controller. The resulting images were also in greyscale only. The implementation of such a setting was impossible without having full access to the projection hardware.

Cotting *et al.* introduced a coding scheme [11] that synchronizes a camera to a specific time slot of a DLP micro-mirror flipping sequence in which imperceptible binary patterns are embedded. However, not all mirror states are available for all possible intensities, and the additional hardware – DVI repeater with tapped vertical sync signal – is not an off-the-shelf instrument.

With the development of digital projection technology, some so-called 3D compatible DLP projectors with fresh rate of $120Hz$ or higher emerge recently. They make it possible to implement imperceptible structured light without any hardware modification or extra assisting hardware.

III. METHOD

A. Pattern Projection Strategy for Imperceptible Structured Light Sensing

The vital aspect of imperceptible structured light sensing is the synchronization between camera's image capture and projector's illumination projection. Here we take one capture-projection cycle as an example to describe the strategy of pattern projection, which is illustrated in Fig. 1. To achieve imperceptible structured light projection, the frequency of projection must exceed the flicker fusion threshold, which is $75Hz$ for most of the people. First of all, we ensure that the projector projects an image every $10ms$, i.e., at $100Hz$. As shown in Fig. 1, along the time axis, the colored pattern illumination, the inverse colored pattern illumination, and entirely white illumination are projected at the time instants $0ms$, $10ms$, $20ms$ respectively. The former two images are projected for ISL sensing, while the latter one is projected for capturing the scene-texture image at the closest time instant. On the camera side, the camera captures the pattern-illuminated image at $5ms$. With a refresh rate of the camera at about 30 frames per second (which is similar to that of most of the CCD cameras), the camera captures the scene-texture image at $40ms$, shortly after the projector projects the entirely white illumination on the object. At $70ms$ a new capture-projection cycle resumes. With the aforementioned

capture-projection strategy, the system could capture about 14 image pairs (pattern-illuminated image and scene-texture image) per second.

The colored pattern illumination in our system is designed after the principle of pseudorandom array [12]. The grid points at the intersection corners of neighboring rhombic pattern elements are chosen as the feature points. We employed an encoding mechanism described in [12] to assure the code uniqueness of each grid point. The 2D pseudorandom color pattern of 65×63 elements that have red, green, blue, or black colors for the pattern elements (the foreground), and white color for the background, together with the pattern's inverse, are shown in Fig. 2. To human's sensing the pattern and the inverse pattern are visually integrated over time. Thus the illumination appears like fluorescent light to humans.

Next, the 3D positions of the key facial landmarks are located by a combined use of the the pattern-illuminated image and the scene-texture image.

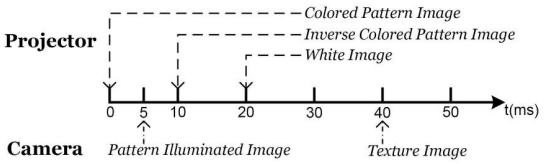
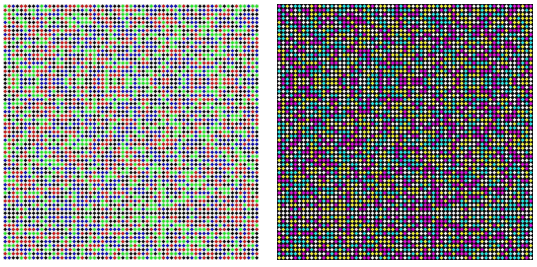


Fig. 1. Capture-Projection Synchronization Strategy.



(a) (b)

Fig. 2. Pattern-illuminated images: (a) image under the original illumination; (b) image under the inverse illumination.

B. Facial Feature Localization

An image pair composed of a pattern-illuminated image and the corresponding scene-texture image will be available in each projection-capture cycle. From the pattern-illuminated image, the 3D positions of the grid points can be determined from the inter-geometry of the projector and camera and the intrinsic parameters of the two instruments, through triangulation. From the scene-texture image, some salient facial landmarks can be located with ease. How to locate the 3D positions of the facial features from the two modalities is described below.

1) Localizing 2D Positions of Key Facial Feature Points in Scene-texture Image: Automatic face detection and facial feature localization in 2D image has been an actively researched subject for years, and many effective methods have been proposed in the literature.

For the sake of accuracy and efficiency of 2D facial feature localization in the scene-texture image, firstly, we employ the Adaboost [13] face detection method to extract the position of the face in the image. We then apply the AAM [14] method to localize the facial features in the segmented face image.

For instance, in Fig. 3, 25 feature points are shown that were defined from AAM localization. They lie on or around the salient features in the face, such as the inner corner and outer corner of the eyes, the corner of the eyebrows, the tip of the nose, and the corner of the mouth etc., which are relatively less affected by expression variation. In addition, all the feature points are distributed symmetrically in the frontal face, allowing at least half of them to be located accurately even if the face orientation in the operation stage is an extreme one.

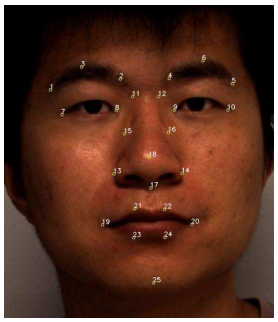


Fig. 3. 2D facial features located by AAM

2) Determining 3D Positions of Grid Points in Pattern-illuminated Image: How unique code can be attributed to each position of the illuminated pattern is a key question in SLS. On this, there are the temporal and spatial coding schemes. The spatial coding scheme has the advantage that 3D determination can be achieved with a single illumination and a single image capture. It is therefore particularly suitable for use in dynamic applications like head pose estimation. In this work, we employ the color coding scheme described in [12] to determine the 3D position of grid points in the pattern-illuminated image. In the illuminated pattern, each grid point is encoded by the color profile of the 2×3 rhombic elements surrounding it, and the code is generally preserved in the image data. Each of such grid points, once its position in the pattern-illuminated image is located, can thus have its corresponding position in the illuminated pattern (on the projector side) identified from the unique code. With knowledge of the inter-geometry of the projector and camera and the intrinsic parameters of the two instruments (that are acquired from an off-line calibration process), the 3D position of the grid point could be calculated by a simple triangulation step.

3) Inferring 3D Positions of Key Facial Features: Since the interval between the captures of the pattern-illuminated image and the scene-texture image is rather small (relative to the motion of the head), in this work we make the simplifying assumption that the head position is constant in the two images. With that, the grid point positions and the salient

facial features in 3D can be related through the rigidity of the human face. More precisely, we infer the facial features from a combined use of the facial features' positions in the scene-texture image, the grid points' positions in the pattern-illuminated image, and the grid points' 3D positions estimated from the structured light sensing step. For each feature point in the scene-texture image, a mirror point could be found in the pattern-illuminated image, as illustrated in Fig. 4(a). It would be most desirable that the mirror point coincides with one of the grid points, as that way the 3D position of the feature point can be read as the depth of the grid point determined from structured light sensing. However, in practice the coincidence would hardly occur, and the 3D positions of the facial feature points would need to be interpolated from the 3D positions of the nearby grid points.

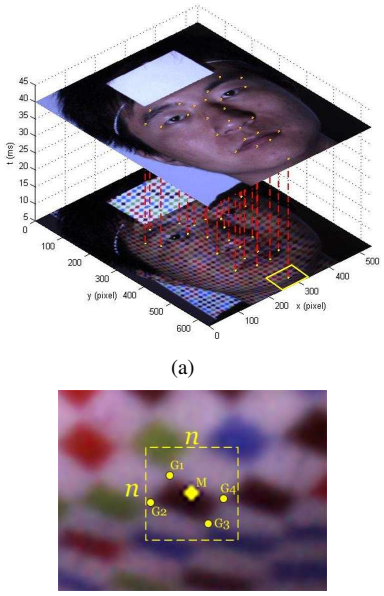
Consider a facial feature point and the image patch around it, which is illustrated by the yellow rectangle in Fig. 4(a). The window is magnified and shown in Fig. 4(b). Set an $n \times n$ window centered in the mirror point M. Assume that in this window, there are N grid points, denoted as $G_i, i = 1, \ldots, N$. Suppose the 3D position of G_i is X_i. Then the 3D position $\overline{X}$ of the feature point could be interpolated as the weighted average of the 3D positions of the nearby grid points in the selected window, which could be formulated by Eq. 1, where α_i is the weight, and d_i in Eq. 2 is the 2D Euclidean distance between the i-th grid point G_i and the mirror point M. For computational efficiency, here we need only the 2D positions of the feature point and the nearby grid points, and in the structured light sensing step we determine the 3D positions of not all grid points but only those that are in the immediate neighborhood of some key facial feature points. Despite that there should be certain discrepancy between the interpolated depth and the real depth of each facial feature point, the pose estimation algorithm described in the following subsection could embrace such discrepancies and determine the head pose with the minimum influence.

$$\overline{X} = \sum_{i=1}^{N} \alpha_i X_i, \tag{1}$$

$$\alpha_i = \frac{d_i}{\sum_{j=1}^{N} d_j}. \tag{2}$$

C. 6 DOF Head Pose Estimation

By the aforementioned method, the 3D positions of the predefined feature points could be determined in each frame. As a result, the correspondence between two sets of 3D points, each set from a consecutive image frame, can be established. Like other computer vision tasks, notably those that require the estimation of the motion of a rigid object from 3D point correspondences, here we encounter the following mathematical problem. We have two 3D point sets $\{p_i\}$ and $\{p'_i\}, i = 1, 2, \ldots, N$ (here, p_i and p'_i are considered as 3×1 column matrices), from which we need to determine the 3D rigid displacement (3×3 rotation matrix R, and 3×1 translation vector T) between them:



(a)

(b)

Fig. 4. 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.

$$p'_i = R p_i + T + N_i, \tag{3}$$

where N_i is a noise vector. We want to estimate R and T to minimize

$$\Sigma^2 = \sum_{i=1}^{N} \| p'_i - (R p_i + T) \|^2. \tag{4}$$

This problem is known as the *absolute orientation problem*, and there are a number of methods in the literature available to tackle it. The solution methods can be categorized into two classes: iterative form, and closed form [15]. Closed form solutions are generally more superior in terms of efficiency and robustness, because the iterative methods suffer from the problems of not guaranteeing convergence, becoming trapped in local minima of the error function, and requiring good starting estimate. For these reasons, we chose a closed form solution to solve this problem. With comprehensive consideration of accuracy, robustness, stability, and efficiency of a number of methods, we employed the method proposed by Umeyama [16], which is based on computing the singular value decomposition (SVD) of a correlation matrix defined by:

$$H = \sum_{i=1}^{N} p'_{c_i} {p_{c_i}}^T = U \Lambda V^T, \tag{5}$$

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

Then the optimal rotation matrix and translation vector could be calculated as

$$\hat{R} = U V^T, \tag{6}$$

$$\hat{T} = \bar{p}' - \hat{R}\bar{p}. \tag{7}$$

As long as more than three non-collinear corresponding point pairs are available, the method can determine the transformation parameters uniquely.

IV. EXPERIMENTS

A. Experimentation Setup

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

The projector-camera system used in the experiment consisted of a DLP projector (Mitsubishi EX240U projector) with a native resolution of 1024×768 and a refresh rate of $120Hz$, and a camera (Point Grey FL2G-13S2C-C CCD camera with Myutron FV1520 f15mm lens) of 1288×964 resolution at $30fps$, both being off-the-shelf equipments. The focal length of the camera was fixed in $15mm$, while that of the projector was in the range of $25-31mm$. The ILS was configured for a working distance (the distance from the camera to the mean position of the human face) of about $800mm$.

We first fixed the camera and projector rigidly, and the projector and camera were connected to a desktop computer through VGA and IEEE-1394b interfaces respectively. Then the projector-camera system was calibrated using an LCD monitor as the calibration object; the calibration method, detailed in [17], can derive the intrinsic and extrinsic parameters of the two instruments. Once the experimental system was set up, we could collect data for further experiments.

B. Test Dataset Collection

Because of the differences in the various sensing methods used (such as monocular vision, stereo vision, infrared vision etc.), there is no standard benchmark for evaluating the performance of head pose estimation, and researchers generally tested their algorithms on the databases collected by themselves. Through reviewing the literature, we found that the subjects in their databases range from one to less than 10, and for every subject, the video length is about several minutes. Because of the speciality of the proposed sensing method, we ought to collect our own experimental data. Our database are about 15 persons, of which nine are male and six are female, and six wearing glasses. The length of each video sequence is 1 minute, i.e., 1200 frames. The sequences start with the objects facing head-on to the cameras. Several sequences were recorded for each participant. The sequences were collected in the laboratory environment with some global illumination changes.

Performance assessment requires ground-truth of the orientation of the head in each image frame, yet such ground-truth about a real human subject is generally difficult to measure in practice. To make the ground truth accessible, we asked the human subjects to wear a headband to which a credit card sized white planar board has been attached, as shown in Fig. 6. The white board was adjusted to be parallel with the face, implying that the orientation of the face can be read from that of the white board. With color coded illumination, the 3D position of any three non-collinear grid points (named by P_1, P_2 and P_3) on the white board could be derived by the aforementioned approach, as depicted in Fig. 5. Let $\mathbf{X}_i$ be the 3D positions of $P_i, i = 1,2,3$, the surface normal of the white board could be formulated as $\mathbf{n} = \frac{(\mathbf{X_1}-\mathbf{X_2})\times(\mathbf{X_3}-\mathbf{X_2})}{\|(\mathbf{X_1}-\mathbf{X_2})\times(\mathbf{X_3}-\mathbf{X_2})\|}$. This way, the ground-truth of face orientation was made directly accessible. As for the position of the head, it was interpreted from the centroid position of the white board in space.

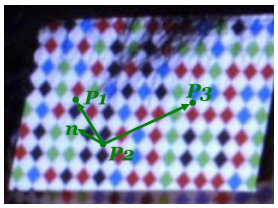


Fig. 5. Ground truth on the surface orientation of human face: it was made the same as that of a white board attached to the face, and the latter could be computed directly for each image.

C. Results

Experimental results at some frames of a subject are presented in Fig. 6. 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.

In the first image frame, the subject was required to have his face in the head-on orientation with respect to the camera, so that the orientation vector of the face was parallel with the optical axis of the camera. This is shown in the top-left sub-figure of Fig. 6. The 3D positions of all 25 facial feature points were derived for the first image frame and the subsequent frames, allowing the head poses relative to the camera to be estimated on the basis of the corresponding 3D point pairs.

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.

It should be noticed that the mean absolute error of yaw in the proposed method was generally larger than those of pitch and roll. We believe the reason for it lies in the asymmetric inaccuracy in localizing the 2D feature points by the AAM method, which was incurred by the illumination shadows around the eyes and nose caused by extreme pose variations.

For real-time applications, efficiency is of great importance, hence we implemented the proposed method in C++ using the Intel OpenCV Library to better evaluate its processing time. Through multi-thread programming, the projection-capture process and calculation process were executed in two different threads respectively, each of which was able

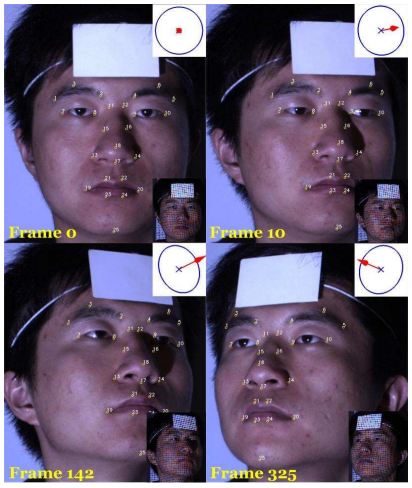


Fig. 6. Experimental results

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

to run in real time in a desktop with Intel Pentium D 3.0GHz CPU. Table II shows the average processing times for AAM facial features localization, 3D depth calculation, and head pose estimation in the given system. Facial feature localization with AAM is the most time-consuming process. Processing times varied slightly according to the number of iterations in the AAM algorithm. However, they all satisfied the requirement of real-time application.

V. CONCLUSION AND FUTURE WORK

We have described a method of estimating head pose using imperceptible structured light sensing. Through elaborate pattern projection strategy and camera-projector synchronization, pattern-illuminated images and the corresponding scene-texture images can be captured under imperceptible patterned illumination. The 3D positions of the facial feature points are then determined by putting together the 2D locations of the facial feature points in the scene-texture image (that are localized by AAM), and the point cloud generated by structured light sensing. Finally, the 6-DOF head motion is estimated from the 3D corresponding feature point pairs over the image sequence through SVD of a correlation matrix.

TABLE II

AVERAGE PROCESSING TIME

Subroutine	AAM	3D Depth Calc.	Pose Est.	Total
Time (ms)	17.43	1.82	2.56	21.81

The proposed method has been tested on video sequences captured by a prototype of the described system. Experimental results show that the proposed method is effective, accurate, and fast for 6-DOF head pose estimation. The processing time is short enough for real-time application.

Our future work will be about introduction of motion compensation between the pattern-illuminated image and the subsequent scene-texture image, and of the use of 3D deformable model that embraces facial expression variation, so that better estimation accuracy can be achieved.

VI. 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).

REFERENCES

[1] Y. Matsumoto et al, "3D Model-based 6-DOF Head Tracking by a Single Camera for Human-Robot Interacton", *IEEE International Conference on Robotics and Automation*, pp. 3194-3199, 2009.

[2] K.W. Bowyer, K. Chang and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition", *Computer Vision and Image Understanding*, 101(1):1-15, 2006.

[3] E. Murphy-Chutorian and M.M. Trivedi, "Head Pose Estimation and Augmented Reality Tracking: An Integrated System and Evaluation for Monitoring Driver Awareness", *IEEE Transactions on Intelligent Transportation Systems*, 11(2):300-311, 2010.

[4] S. Joaquim, P. Jordi and B. Joan, "Pattern codification strategies in structured light systems", *Pattern Recognition*, 37(4):827-849, 2004.

[5] D. Fofi, T. Sliwa and Y. Voisin, "A comparative survey on invisible structured light", In *Proceedings of Machine Vision Applications in Industrial Inspection XII*, pp. 90-98, 2004.

[6] E. Murphy-Chutorian and M.M. Trivedi, "Head Pose Estimation in Computer Vision: A Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(4):607-626, 2009.

[7] L. Morency, P. Sundberg and T. Darrell, "Pose estimation using 3D view-based eigenspaces", *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 45-52, 2003.

[8] P. Jimenez, and J. Nuevo, et al, "Face tracking and pose estimation with automatic three-dimensional model construction", *IET Computer Vision*, 3(2):93-102, 2009.

[9] M. Breitenstein, D. Kuettel *et al.*, "Real-time face pose estimation from single range images", *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-8, 2008.

[10] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs, "The office of the future: A unified approach to image-based modeling and spatially immersive displays", *Proceedings of SIGGRAPH 98*, pp. 179-188, July 1998.

[11] D. Cotting, M. Naef, M. Cross and H. Fuchs, "Embedding imperceptible patterns into projected images for simultaneous acquisition and display", *The IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp.100-109, 2004.

[12] Z. Song and R. Chung, "Determining Both Surface Position and Orientation in Structured-Light-Based Sensing", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(10):1770-1780,2010.

[13] P. Viola and M. J. Jones, "Robust Real-Time Face Detection", *International Journal of Computer Vision*, 57(2):137-154, 2004.

[14] I. Matthews and S. Baker, "Active Appearance Models Revisited", *International Journal of Computer Vision*, 60(2):135-164, 2004.

[15] D.W. Eggert, A. Lorusso and R.B. Fisher, "Estimating 3-D rigid body transformations: a comparison of four major algorithms", *Machine Vision and Applications*, 9:272-290, 1997.

[16] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):376-380, 1991.

[17] Z. Song and R. Chung, "Use of LCD Panel for Calibrating Structured-Light-Based Range Sensing System", *IEEE Transactions on Instrumentation and Measurement*, 57(11):2623-2630, 2008.