氏名（本籍地）	石山　塁（いしやま　るい）
学位の種類	博　士（情報科学）
学位記番号	情　博　第　401　号
学位授与年月日	平成20年3月25日
学位授与の要件	学位規則第4条第1項該当
研究科、専攻	東北大学大学院情報科学研究科（博士課程）システム情報科学専攻
学位論文題目	Face Image Recognition by Three-Dimensional Modeling of Appearance （３次元アピアランスモデルの構築に基づく顔画像認識の研究）
論文審査委員	（主査）東北大学教授　出口光一郎 東北大学教授　橋本　浩一　　東北大学教授　田所　諭 東北大学教授　青木　孝文　　東北大学准教授　岡谷　貴之

論文内容の要旨

1. Introduction

There have been numerous studies in face recognition, and some are being put into practical use in applications such as user authentication systems. However, the applications have been limited to the scenes where the subject gives their cooperation and where it is possible to set environmental conditions thus limiting variations in the pose of the face and the illumination conditions. This is because the appearance of the same face is changed drastically when such variations become wide and it occasionally becomes quite difficult to identify the subject. Conventional methods are unable to cope with such variations.

Nowadays, there are growing demands for the applications of face recognition, including video surveillance and autonomous robot vision. In these applications, the target image is taken in various environments without his/her cooperation, and large variations in pose and illumination conditions are inevitable. It is mandatory to ensure the robustness of the recognition system against these variations.

2. Face Recognition under Variable Pose and Illumination Condition

One problem of the conventional methods (employing pattern recognition techniques) is that they require an enormous number of enrolled images to be able to cope with wide variations in pose and illumination conditions. Such a system is impractical, because most general users are unable to take such a large number of images under various environments. A few studies have proposed acquiring these images easily by using measurements of 3-D shapes and computer graphics. However 3-D measurements of human face were also impractical because the conventional 3-D measurements with high accuracies required an unacceptably long measurement time for humans to keep still. Another problem is that the conventional methods take impractically large costs in calculating and storing the templates to cope with wide variations of pose ranging from frontal to profiles. The view-based description, which is used in the conventional methods, requires the templates in many poses which densely sample the whole range of possible poses, and the number of templates becomes impractically large when the range of pose variations is large. Finally, there were not enough experimental evaluations in the conventional studies to verify the performance. The number of subjects enrolled in the database was small, and the range of pose and illumination variation was quite limited.

The aim of this thesis is to solve these problems by developing a new framework, 2D-3D face recognition system that employs three-dimensional modeling of appearance for recognition of 2-D images taken in variable environments (Figure 1). The goal is to achieve the following four targets: first, measure 3-D shape of human face with high density and accuracy within an acceptably short time; second, describe appearances in any pose and illumination conditions by a compact model; third, accurately and efficiently estimate the face pose; and four, verify the performances of the proposed methods by experiments using an extensive database with images of a large number of subjects taken in a variety of poses and illumination conditions.

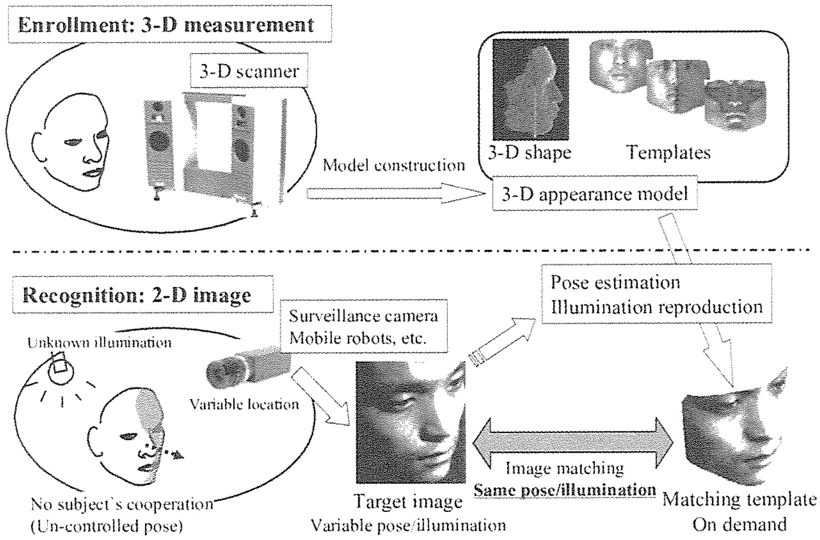

Figure 1: Outline of proposed approach: 2D-3D face recognition system.

3. 3-D Measurement of Human Face

The conventional methods could not achieve accurate and dense 3-D measurements within an acceptably short time for most people to keep still. Although the conventional phase shift method can achieve high-density 3-D measurements within a short time, its applications were limited to measurements of planar targets. This is because the conventional methods lack a solution to the absolute phase determination and the phase un-wrapping for measurements of a non-smooth surface such as a face.

New methods are proposed to solve these problems by using multiple cameras and projectors. In the first proposed method, geometrical constraints between multiple views are utilized to determine the absolute phase in wide-range measurements. The second proposed method uses the constraints to stabilize the phase un-wrapping in measurements of a non-smooth surface. These methods make it feasible to use the phase shift method in human face measurements (Figure 2).

The proposed methods achieved the 3-D measurement with a degree of high accuracy, 0.14 mm, for every pixel within 0.5 seconds. Using standard-speed 30fps cameras and projectors was enough to achieve a measurement in such a short period of time, thus using high-speed devices enables measurement within a shorter time. Consequently, employing the accurate and dense 3-D measurements in a face recognition system has become feasible.

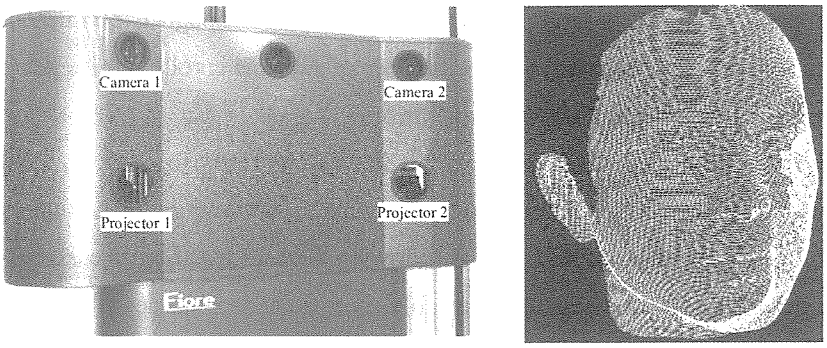

Figure 2: 3-D scanner employing proposed methods (left) and 3-D measurements of human face (right).

4. Accurate 3-D Measurement Using Un-Calibrated Pattern Projection

The assumption underlying conventional methods using pattern projection is that the projected pattern is exactly the same as that which is designed in the theory. However, the pattern actually projected is inevitably distorted from the designed waveform, because of nonlinear characteristics of the actual devices. This distortion induces systematic errors in the relationship between the measured phase and the corresponding position on the projector image, and these errors directly affect the accuracy of the 3-D measurements. To compensate for the errors, the conventional methods require an impractically large amount of calibration data.

New methods are proposed to ensure high levels of measurement accuracy in implementations of the 3-D measurements using pattern projection. The proposed methods make use of actual phase measurements of a plane to obtain 3-D measurements without being affected by the systematic errors. The invariance of the cross-ratio under perspective projections means that the cross-ratio of the depths between the calibrated/target points is equal to that of the distances between camera image points at which the same phases are measured when the plane is measured (Figure 3 (left)). To determine the points on the measurement of the plane the proposed method uses measured phase as a label, and then the distance is calculated on the basis of the invariance of the cross-ratio.

The proposed method achieved measurements that were 10% more accurate than conventional methods and removed the systematic noises, which cause inaccurate shading in computer graphics rendering (Figure 3 (center and right)). Since the measurement of the plane is already included in the conventional calibration process, no additional cost is required for the proposed methods.

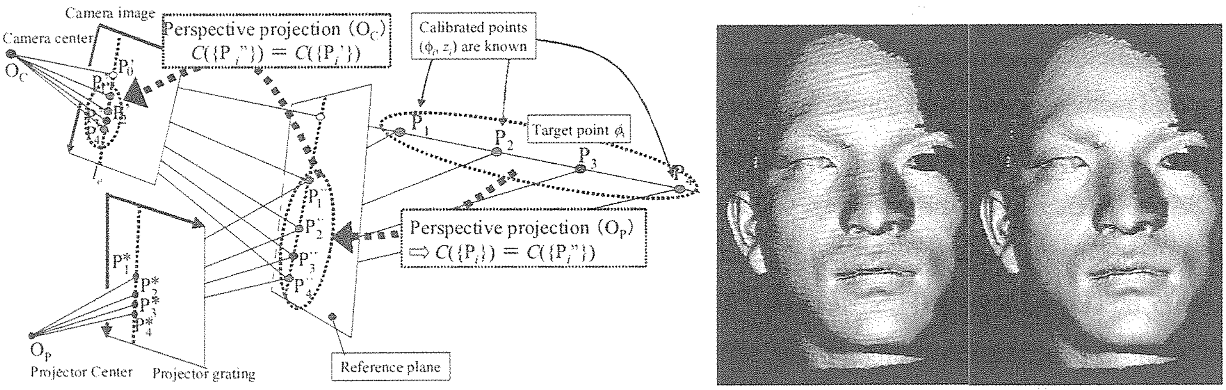

Figure 3: Geometries of depth measurement by proposed method (left), 3-D measurements by conventional method (center, resulting in stripe-like noises) and proposed method (right).

5. Construction of Extensive Database for Experimental Evaluation

To obtain practical evaluations of the performances of the algorithms, an extensive database was constructed. The database is comprised of images that were taken in a wide range of environments, as well as 3-D models, and annotated data. Experimental evaluations in conventional studies have been insufficient because the databases used did not have a large number of subjects and the environmental variations of the test images were slight. Our database is far more extensive than those used in the conventional studies. The number of enrolled subjects is 200, and it covers wider variations in pose from the frontal to profiles as well as wider variations in illumination conditions, including the presence of strong shadows and backlighting (Figure 4). The additional data was annotated to conduct a detailed analysis of the performances relating to pose errors.

Figure 4: Seventy conditions of test images of one person. Images of 200 persons (14,000 images in total) and 3-D face models were collected to test performances of algorithms.

6. Modeling Appearance in Variable Poses and Illumination Conditions

The conventional methods employ the view-based approach to cope with variations in the pose. This approach prepares the templates to describe illumination variations for each of a large number of pre-determined poses. The number of templates grows in proportional to the range of pose variations, thus the data storage and computational costs become impractically large. Moreover, as shown in our experiment, even a small difference between the target pose and the pre-determined one severely degrades the recognition performance.

In this chapter, the 3-D appearance model is proposed to efficiently describe appearance in any pose and illumination condition. The proposed model consists of 3-D shape data and geodesic illumination bases (GIB), which are the templates used to describe radiances of 3-D surface facets under various illuminations (Figure 5 (left)). In Lambertian reflectance model, the radiances are constant for any view. Therefore, illumination variations of an image in any pose can be described with GIBs. GIBs are projected onto the image in any pose on demand, and then the projected bases are equivalent to the illumination bases used by conventional view-based methods. The model can be used to reproduce the unknown illumination condition of a given target image.

The experimental results show that the 3-D appearance model reproduced the images under any illumination condition in various poses with a level of accuracy of more than 99% (Figure 5 (right)). It is achieved using only ten GIBs and 3-D data. The amount of the data is almost the same as that required by conventional methods for only one pose, and the proposed model is quite efficient to cope with wide range of poses.

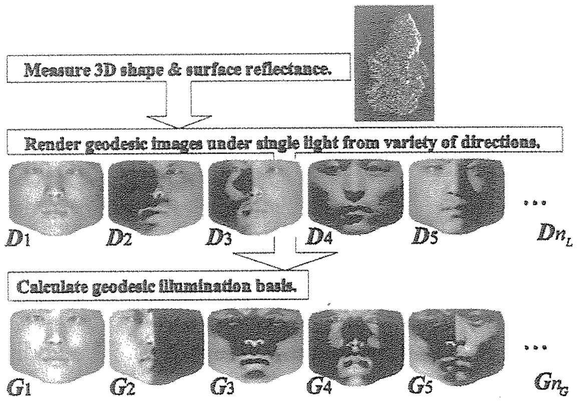

Figure 5: Flow of construction of 3-D appearance model from 3-D measurements (left), and accuracies of image reproduction by 3-D appearance model for images under any illumination condition in various poses (right).

7. Pose Estimation Using 3-D Appearance Model

A new method is proposed to estimate poses including a large rotation in depth, with robustness against wide variations in the illumination condition. Although the conventional view-based methods cannot cope with such a wide range of poses, the proposed method uses the 3-D appearance model and thus obtains templates in any variable pose. The pose is estimated by minimizing the error between the target image and an image reproduced by the 3-D appearance model (Figure 6 (left)). By taking a rough initial estimate, the pose is iteratively updated so that the error is decreased by the steepest descent.

Since numeric differentiation requires repeating costly 3-D rendering and illumination reproduction, the cost of calculating the gradients of the error with respect to the current pose is high. The proposed method offers an efficient way: 'Ideal optical flow' is determined by using template matching between the sub-blocks of the image reproduced with the current pose and the target image. The optimal pose update is determined so as to produce the flow.

The performances of our proposed method were evaluated with test images that include variations in pose from frontal to profiles and under wide variations of illuminations. The estimation error was less than 2.3 degrees even when the initial error was as large as 10 degrees (Figure 6 (right)). It was also shown that the proposed method is efficient enough to implement a real-time face tracking system.

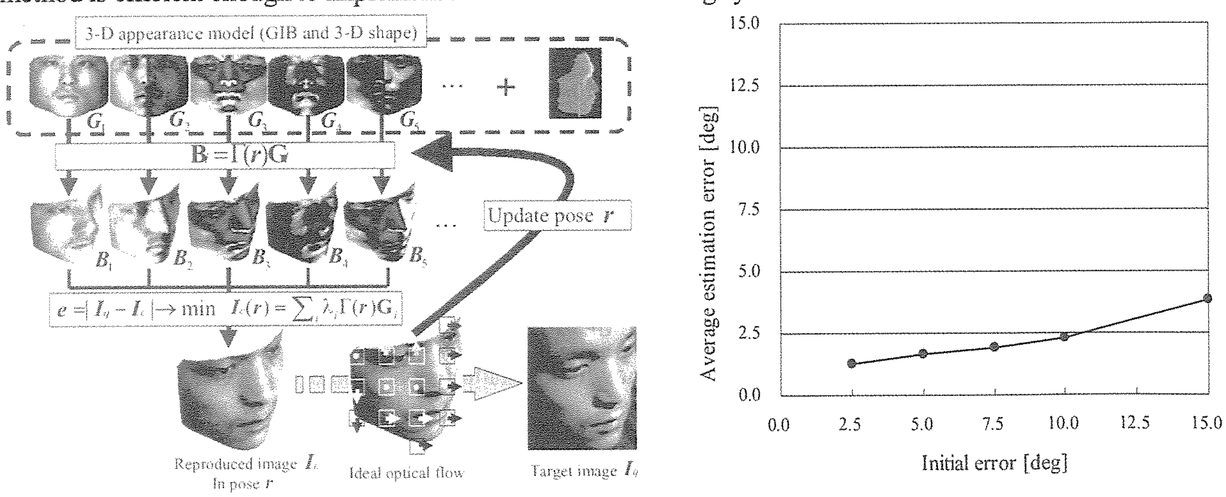

Figure 6: Flow of pose estimation by aligning 3-D appearance model to target image (left), and pose estimation accuracies with respect to errors in initial estimates (right).

8. Evaluation of 2D-3D Face Recognition System

The performance of the proposed 2D-3D face recognition system was evaluated using the extensive database. In the experiment, 14,000 test images (including 70 wide variations in pose and illuminations per subject, shown in Figure 4) were matched to 200 enrolled subjects (Figure 7 (left)). The conventional standard methods failed to identify most of the test images even those in the frontal pose. In contrast, the proposed methods achieved a highly accurate matching rate of 93.8% for all the test images. The efficiency of the 3-D appearance model was also verified. Although the conventional view-based methods require impractically large amount of data to cope with the pose variations in the test images, the proposed methods successfully worked even though they used only ten GIBs and 3-D shape data (Figure 7(right)).

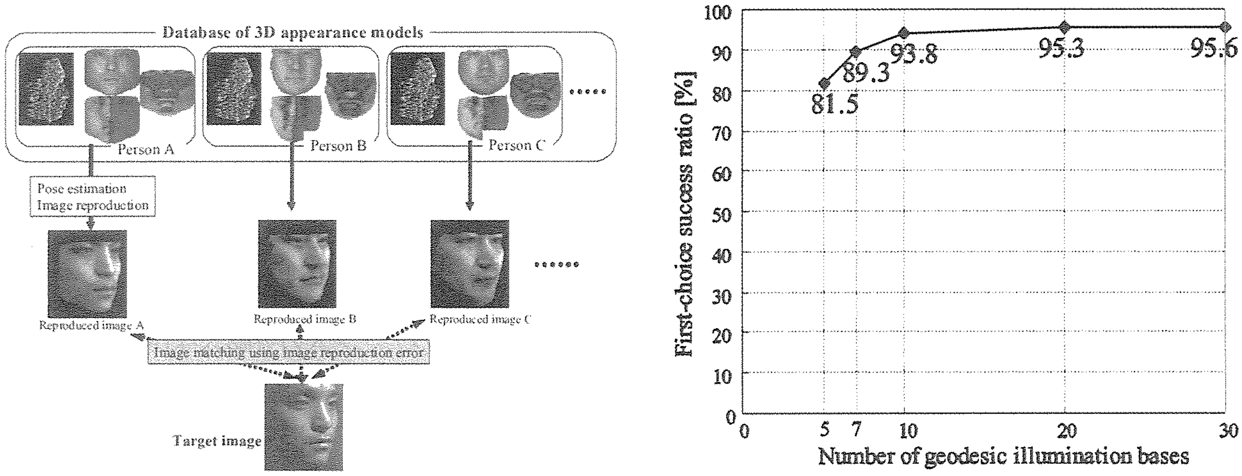

Figure 7: Outline of recognition by matching target image to database of 3-D appearance models (left), and recognition performances for test images (shown in Figure 4) achieved by using various numbers of GIBs (right).

9. Conclusions

The proposed methods achieved face recognition with robustness against wide variations of pose and illumination condition in which the target image is taken. The proposed 3-D measurement methods offer accurate and dense 3-D measurements with a sufficiently short time measurement. The 3-D appearance model reproduces appearance in any pose and illumination conditions with a compact data. The proposed pose estimation method aligns the 3-D appearance model with the target image to reproduce the matching templates in the same environment. The performances were verified by the experiments using the extensive database that consists of images taken under large variations in pose and illumination conditions. The robustness against these variations is the key to realize autonomous recognition, which is required in the new applications such as video surveillance, presence managements, and vision of service robots. Moreover, the proposed 3-D measurement methods are hopeful to extend the applications of 3-D models of human body to medics, clothing industries, and graphics arts, and so on. The proposed pose estimation method is also expected to enlarge the applications of face tracking in the areas of the human-computer interfaces. The method is prominent in great robustness against the variations of the environments and applicability to a large range of pose variation.

論 文 審 査 結 果 の 要 旨

本論文は，顔画像を用いて一般環境における自律的な個人識別を実現するための基本的な技術課題を明らかにし，その解決策としての３次元計測と画像認識の技術を論じたもので，全９章からなる。

第１章は序論であり，従来の顔画像認識とその応用の流れを総括し，広範な環境変化，特に姿勢と照明の変化に対する頑強性およびそれに必要な登録データの簡便な計測が課題であることを論じている。

第２章では，従来技術の限界と解決すべき課題を示し，その解として，顔の３次元計測を行い，顔画像の環境変動を３次元的にモデル化する独自の顔画像認識システムを提案している。独自の画像認識の技術に加え，この３次元計測技術も含めた開発を具体的に提案している点で評価できる。

第３章では，顔画像認識に必要な高精度３次元形状計測を，一般人が静止可能な短時間内で実現する３次元形状計測手法を提案し，実装している。複数カメラ間の幾何拘束を利用することで，少ない投影パターン数での高精度な３次元形状計測を可能とした点で，その有用性を評価できる。

第４章では，３次元計測装置を実装する際に生じた投影パターン波形の理論との誤差に対して，計測精度が影響されない３次元計測手法をさらに提案している。これは実用化において重要かつ有用な成果であり，透視変換不変量である複比を利用して誤差を打ち消す点は独自なものと評価できる。

第５章では，顔画像認識の性能評価を行うための，多様な環境変動を網羅した大量の画像，および，３次元形状データの収集について述べている。大量のデータ収集と分析用のデータ入力等膨大な作業を実行し，従来研究で不十分であった実際的な性能評価を可能とした点で，重要な成果である。

第６章では，姿勢と照明環境の変動による顔の見え方の変化を少ないデータ量で効率的に記述する３次元アピアランスモデルを提案している。照明基底を３次元物体表面上で計算し姿勢に応じて投影することにより，任意の姿勢における照明基底を生成でき，未知の照明条件を高精度に再現可能なことを示しており，独自性と有用性を評価できる。

第７章では，認識対象画像の顔の姿勢を，高精度かつ効率的に推定する手法を提案している。３次元モデルを用いることで任意の姿勢変動が推定可能であり，また，この推定に用いる誤差関数の任意姿勢での微分値を効率的に計算する手法を示した点で独自性がある。この姿勢推定による前処理によって顔の認識性能の高い頑強性を実現するとともに，顔追跡などのマン・マシンインタフェース分野にも応用可能な高速処理を実現した点で，有用性を評価できる。

第８章では，ここまでの各提案手法を統合したシステムの性能評価を行っている。大規模な実験データを用い，従来手法との比較で大きな環境変動に対する飛躍的な性能向上を実証した点で，重要な成果である。

第９章は結論である。

以上要するに本論文は，顔画像認識技術が一般社会において広く実用化されるための基本課題を明らかにし，その重要課題である姿勢と照明の変化に対し頑強な認識システムを実現するための有効な解と実用的な性能評価結果を示すものであり，システム情報科学の発展に貢献するところが少なくない。

よって，本論文は博士(情報科学)の学位論文として合格と認める。