コンピュータビジョンと
イメージメディア
（２００１．７．５）
128－6

ビューに基づいた追跡による face-to-faceのインタラクションの実現

李 仕剛
岩手大学工学部情報システム工学科
〒 020-8551 岩手県盛岡市上田四丁目3番5号
li@cis.iwate-u.ac.jp

あらまし 移動ロボットと人間との対話はface-to-faceで行うのが自然である．本論文は，移動ロボットが人間の顔を発見し，人間の正面に移動する手法を提案する．まず，肌色に基づいて，顔の候補領域を発見する．次に，学習した顔の向きのビューと比較して，顔の領域を決め，その向きを推定する．最後に，能動的視覚の手法を利用して，視覚フィードバックにより，ロボットが顔を追跡しながら，人間の顔の正面に移動する．

Realizing Face-to-face Interaction by View-based Tracking

Shigang Li
Faculty of Engineering
Iwate University
Morioka, 020-8551, Japan
li@cis.iwate-u.ac.jp

Abstract In this paper, we propose a method of realizing face-to-face interaction by view-based tracking between a human and a mobile robot. Although individuals can be recognized easily by observing frontal faces, it is difficult to do it from profiles of face. To cope with this problem, our moblie robot first finds face-like areas based upon the skin color distribution which is made from real image samples. Second, a facial pattern in the image is found and its direction relative to the robot is estimated in terms of the memorized views. Finally, the mobile robot moves to the front of the face according to the face direction while tracking the face by using visual feedback based upon active vision approach.

1 Introduction

To construct a friendly interface between intelligent robots and humans, it is better to do communication between them by face-to-face. Thus, a mobile robot needs to not only find a face of humans, but also move to the front of the face if the face has a viewing direction initially. There are tow problems we have to cope with. One is to find a human face even if it has viewing direction relative to the front. The other is to let the mobile robot move to the front of a human.

Until now many researches have been done on face recognition. In [12], face images are projected onto a feature space that spans the significant variations among known face images. The significant features are known as "eigenfaces", because they are eigenvectors (principal components) of the set of faces. Upright and frontal views of faces in gray scale images can be detected by a neural network-based algorithim [10]. Frontal faces rotated within an image plane can also be detected by employing multiple networks [11]. However, it is rather difficult to find profiles or semi-profiles of faces. One aspect of 3-D object recognition that has been examined in detail by a number of different studies is the dependence of recognition performance on the viewing direction. The dependence on viewing is present even when the object is seen in both the training and subsequent testing under conditions that facilitate the recovery of 3-D shape, using stereo, shading, and motion [4]. This implies that in view-based recognition a view has its corresponding range for recognition [13]. Thus, multiple views which have their recognition range respectively are needed to be stored for recognition. A view-based multiple-observer eigenspace technique is proposed for use in face recognition under variable pose [9].

Many researches have been reported on face tracking also [6] [7]. The work described in [3] and [1] is concerned with the classification of the expression into states such as happiness, anger, fear, sadness and disgust. A system which tracks a person's facial features in real time has been completed [7]. Basd on this face tracker, the system can distinguish different gestures such as "yes", "no" and "maybe". However, the tracking often fails when the viewing direction from the front is large so that some facial features are occluded. However, a camera is fixed in almost all of these systems. Although the camera can make a pan or tilt rotation, it cannot moves from one place to another in space. These method cannot be employed to robot human communication directly.

To realize face-to-face communication between a robot and a human, it is necessary for a mobile robot to track a face to its front. In this papaer, we propose a method of view-based tracking of a human face by a mobile robot. Our mobile robot first finds face-like areas using the skin color distribution which is made from real image samples. Second, a facial pattern in the image is found and its direction relative to the robot is estimated approximately from the stored several views with different face directions. Then the mobile robot moves to the front of the face according to the face direction while tracking the face by using visual feedback based upon active vision approach.

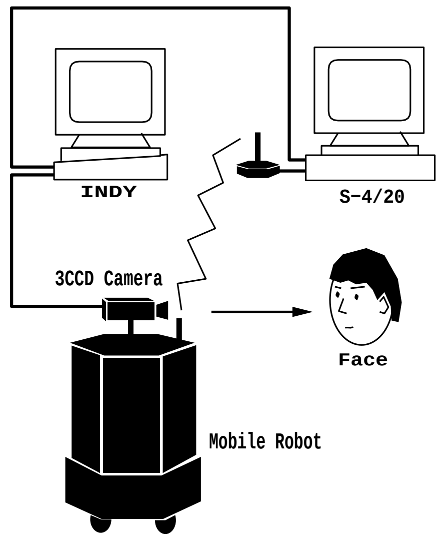


図 1: Our system.

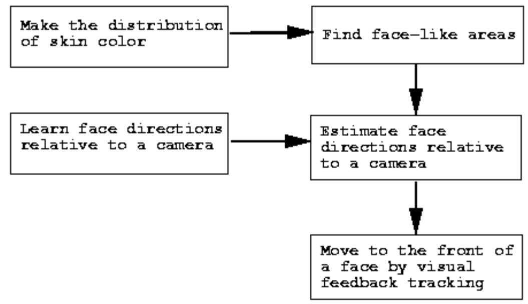


図 2: The schema of our method for recognizing human faces.

2 Scheme of our algorithm

We assume that the robot moves in an indoor environment and makes communication with a human as shown in Fig.1. The scheme used is shown in Figure 2.

There are two phases, learning phase and recognizing phase, in our method. First, a moblie robot finds face-like areas based upon the skin color distribution which is made from real image samples. Second, a facial pattern in the image is found and its direction relative to the robot is estimated. The robot takes facial images at $0, \pm 45, \pm 90$ degrees in advance. All of these patterns at different directions are analyzed with Principal Component Analysis. The obtained eigenvectors are used as features for the description of facial patterns with different directions. When a new image is taken, the pattern at different positions with different scales in an image is projected into the space constituted by the eigenvectors, and the direction of a face relative to the robot is estimated by comparing with memorized ones. Then, the mobile robot moves to the front of the face according to the face direction while tracking the face by using visual feedback. The tracking is carried out based upon an active vision method. A fixation point is selected as a features point which has the largest covariance in facial patterns. Since the robot moves on a flat floor, a circular path is planned and the robot tracks the face so as the fixation point always appear in the central of image (x = 0) and have the same height (y = constant).

3 Finding facial area using skin color distribution

As the first step, we use skin color to find the possible region of faces. We use the uniform color space proposed by Farnsworth [5] to represent colors in an images. Colors are represented by two number u and v.

$$u = 4R/(R+15G+3B) \quad v = 6G/(R+15G+3B)$$

As in [2], the skin color is represented by a distribution in color space. This distribution represents the frequency that each color in the color space appears at the re-

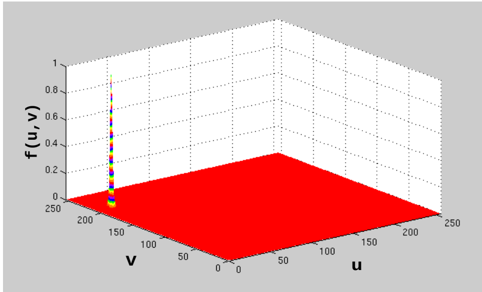


図 3: The distribution of skin colors.

gions of human skin in images. The distribution obtained from tens of examples is shown in Fig.3. If the value of a color in the distribution is 1.0, it means the color is a skin color with the highest possibility.

For a new input image, every pixel in the image is evaluated to determine how well the color of pixels looks like skin color based upon the computed skin distribution function. The regiones having a high similarity with skin color are extracted as candinates of face regions.

4 Estimating face direction

After extracting the candinates of face regions, we find the face and estimate the face direction from these regions. Since the face direction to the camera mounted on the robot is unknown at first, facial viewes taken at $0, \pm45, \pm90$ degree relative to frontal face are stored and an new input image is compared with the stored ones to estimating the face direction. The processing is

- Acquire an initial set of face images (the training set) at $0, \pm45, \pm90$ degree relative to frontal face. Fig.5 shows the samples of a person in our experiment.

(a)Input image.

(b)extracted regions having skin colors

図 4: Finding candinates of face regions by using skin colors.

- Calculate the eigenfaces from the training set using principal components analysis, keeping only the M images that correspond to the highest eigenvalues. These M images define the face space as the same with [12]. Fig.6 shows the six eigenfaces with the biggest eigenvalues. The 21 significant eigenvectors are used in our experiment.

- Calculate the corresponding distribution in M-dimensional weight space for each space of known direction, by projecting the face images onto this M-dimensional space.

Having initialized the system, the following steps are then used to estimate the face direction:

- Calculate a set of weights based on the input image and the M eigenvectors by

図 5: Face samples of a person at $0, \pm 45, \pm 90$ degree relative the frontal face.

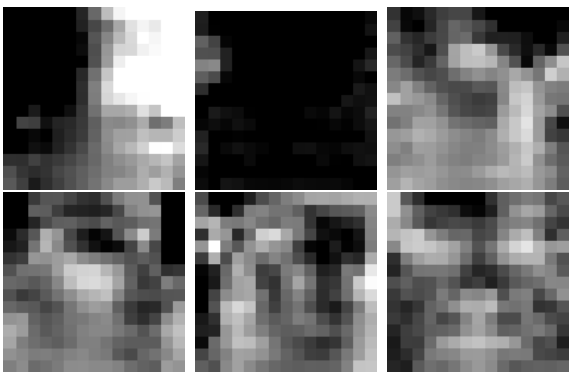


図 6: Six eigenfaces with the biggest eigenvalues.

projecting the input image onto each of the eigenvectors.

- Determine if the images is face at all by checking to see if the image is sufficiently close to the M eigenvector space.

- If it is close enough to the M eigenvector space, determine the face direction of weight pattern by comparing the distance with known ones.

5 Tracking a face to its front by visual feedback

5.1 Tracking a face along a circular path

After determining the rough direction of a face, the mobile robot moves to the front of the face by tracking the face along a circular path based upon the active vision paradigm. Unlike the conventional systems in which a camera is mounted on a manipulator, we use a mobile robot to realize the active vision paradigm [8].

Controlling a mobile robot to move along a circular path by fixating a point is not easy due to the disturbances from environments. We suppose:

- The robot moves in an inddor environment. The movement of the robot is indicated by 2-D coordinate (X, Y) and a rotation θ.

- The optical axis of the camera is mounted on the robot and perpenticular to the moving direction of the robot.

In order to control a mobile robot to move along a circular path, we need to know the radius of the circle. The distance from the camera to the fixation point can be computed by motion stereo method. If the radius of the circle is R and the speed of the robot is V, the angular velocity of the robot can be computed as

$$W = \frac{V}{R}. \tag{1}$$

Since there are many disturbances in the real world, the robot often deviates from the planned trajectory. To cope with this problem, we use the visual feedback.

If the robot moves along a circular path, the following tow conditions must be satisfied.

- The motion direction of the robot is orthographic with the direction of the fixation point.
- The distance of the camera from the fixation point is constant.

From the above conditions, the following visual cues can be obtained.

- From the first condition, the fixation point should appear in the vertical axis in the image.
- From the second condition, the fixation point have a constant height in the image.

By observing the change of the fixation point in the image, we control the robot to move along a circular path to the front of a face.

5.2 Control method

The mobile robot used in our expriment is Nomad200. What we need to control are its speed and angular velocity.

Here we use a camera-centered coordinate. The origin of the coordinate system is set at the focus of the camera while $X-axis$ is aligned with optical axis. Suppose the velocity and the angular velocity are V and W between t and $t+\Delta t$, respectively, we have

$$r = \frac{V}{W}, \ \theta = W\Delta t \tag{2}$$

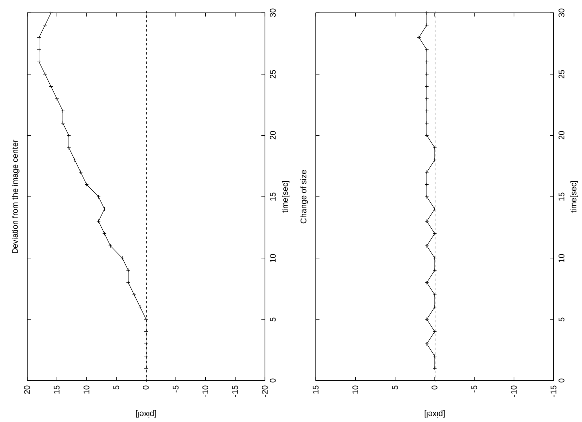
where r is the radius of the circular path and θ is the rotation angle. The mobile robot is controlled in terms of the angular velocity $\frac{V}{R}$ from equation (1).

6 Experiment

We completed our algorithm in a real mobile robot system. A robot called Nomad200 made by Nomadic Technologies is used. A camera , 3CCD Color Video Camera Module: XC-003, is mounted on the robot. The robot is controlled with wireless communication by S-4/20 workstation. The image is processed by a INDY workstation. A person sat in the front of the robot with a face direction of about 45 degree reltive to the camera. Fig.9 (a) shows the found face region using skin color cue. Fig.9 (b) shows the found face with about 45 degree relative to the frontal face, which is indicated by an rectangle. The feature point with the largest variance in the face pattern is selected as a fixation point as show in Fig.10. Fig.7 shows the experiment result when the visual feedback is not used, in which the deviation from the center of images increases with the camera motion. Fig.8 shows the experiment result when the visual feedback is used, in which the deviation from the center of images changes within a small range and the height of fixation is almost kept as a constant. The process of the robot tracking the human face is shown in Fig.11. The frontal face after the tracking is shown in Fig.12.

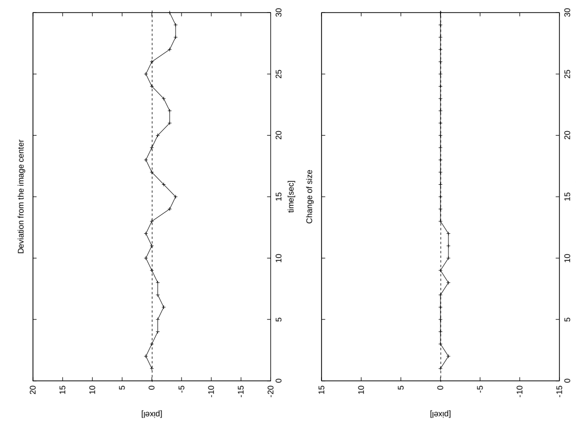
7 Summary and Conclusions

In this paper, we propose a method fo view-based tracking of a human face by a moible. In order to achieve a face-to-face interaction between a human and a robot, it is necessary to let the robot moves to the front of humans' faces. To cope with this problem, our moblie robot first finds face,



(a) (b)

図 7: Experimental results without visual feedback. (a) Deviation from the center of images. (b) Length change of the fixation point.



(a) (b)

図 8: Experimental results using visual feedback. (a) Deviation from the center of images. (b) Length change of the fixation point.

(a) (b)

図 9: (a) Finding a face region using skin color cue. (b) Finding a face with about 45 degree relative to the frontal face.

estimates its direction relative to the frontal face, and moves to the front of the face according to the face direction while tracking the face by using visual feedback.

This is a preliminary work. Many future works remain to be unsolved. The change of illumination influences the face direction estimation in our experiment. How to select robust feature points as fixation points used in the face tracking is another theme.

Acknowledgments

The author would like to thank Mr. Masanori SHIBATA for his programming this work.

参考文献

[1] Black and Yacoob, "Tracking and recognizing rigid and non-rigid facial motions using parametric models of image motion," *Proc. of ICCV*, pp.374-381, 1995.

[2] Q. Chen, H. Wu and M. Yachida, "Face detection by Fuzzy Pattern Matching," *Proc. of ICCV*, pp.591-596, 1995.

[3] T. Darrel and A.P. Pentland, "Attention-driven expression and gesture analysis in an interactive environment," *Proc. of Int. Workshop on Automatic Face-and-Gesture Recognition*, pp. 51-60, 1995.

[4] S. Edelman and H.H. Bulthoff, "Orientation dependence in the recognition of familiar and novel views of three-

図 10: The feature point with the largest variance in the face pattern is selected as a fixation point.

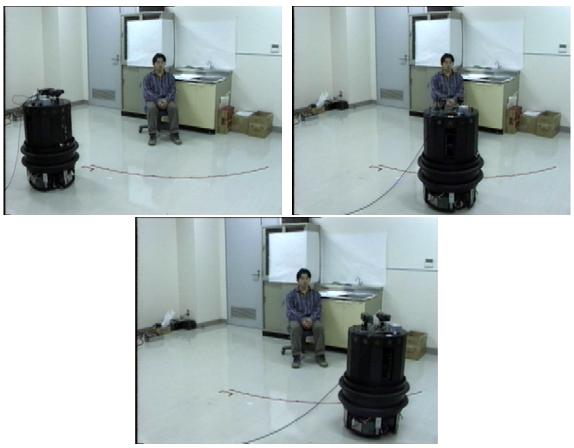


図 11: The process of the robot tracking the human face.

dimensional objects," *Vision Research*, Vol. 32, pp. 2385-2400, 1992.

[5] G. Wyszecki and W.S. Stiles, "Color Science," *John Wiley & Sons, Inc.*

[6] G.D. Hager and K. Toyama, "Xvision: Combining image warping and geometric constraints for fast vision tracking," *Proc. of ECCV*, pp. 507-517, 1996.

[7] J. Heinzmann and A. Zelinsky, "Robust real-time face tracking and gesture recognition," *Proc. of Int. Joint Conf. on Artificial Intelligence*, Vol. 1, pp. 1525-1530, 1992.

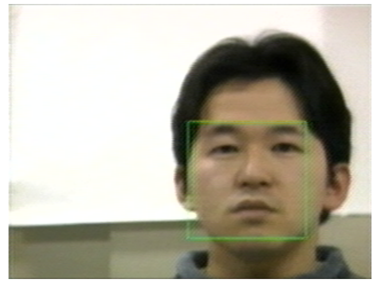


図 12: The front face view acquired by move the camera.

[8] S. Li, I. Miyawaki, H. Ishiguro and S. Tsuji, "Finding of 3D Structure by an Active-vision-based Mobile Robot," *Proc. of IEEE 1992 International Conference on Robotics & Automation*, pp. 1812-1817, 1992.

[9] A. Pentland, B. Moghadam and T. Statner, "View-based and modular eigenspaces for face recognition," *Proc. of CVPR* , pp. 84-91, 1994.

[10] H. A. Rowley, S. Baluja and T. Kanade, "Neural network-based face detection," *IEEE Trans. on PAMI*, Vol. 20, No.1, pp. 23-38, 1998.

[11] H. A. Rowley, S. Baluja and T. Kanade, "Rotation invariant neural network-based face detection," *Proc. of IEEE Int. Conf. on CVPR*, pp. 38-44, 1998.

[12] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, Vol. 3, No.1, pp. 71-86, 1991.

[13] S. Ullamn, "High-level vision," *The MIT Press, Cambridge, Massachusetts*, 1996.