# Detecting human heads with their orientations

Akihiro Sugimoto*, Mitsuhiro Kimura+ and Takashi Matsuyama+

* *National Institute of Informatics, Tokyo 101-8430, Japan*
+ *Graduate School of Informatics, Kyoto University, Japan*

## Abstract

We propose a two-step method for detecting human heads with their orientations. In the first step, the method employs an ellipse as the contour model of human-head appearances to deal with wide variety of appearances. Our method then evaluates the ellipse to detect possible human heads. In the second step, on the other hand, our method focuses on features inside the ellipse, such as eyes, the mouth or cheeks, to model facial components. The method evaluates not only such components themselves but also their geometric configuration to eliminate false positives in the first step and, at the same time, to estimate face orientations. Our intensive experiments show that our method can correctly and stably detect human heads with their orientations.

*Key Words*: Human head detection, face orientation, tracking, video image analysis.

## 1    Introduction

Automatically detecting and tracking people and their movements is important in many applications such as in– and out-door surveillance, distance learning, or interfaces for human-computer interaction [2, 4, 6, 7, 8, 9]. In particular, the human face is a key object of interest for visual discrimination and identification. A tremendous amount of research has been made for detecting human heads/faces and for recognizing face orientations/expressions (see [3, 23] for surveys). Most existing methods in the literatures, however, focus on only one of these two. Namely, methods to detect human heads/faces (see [1, 12, 19, 20, 24], for example) do not estimate orientations of the detected heads/faces, and methods to recognize face orientations/expressions (see [10, 14, 15, 18, 21], for example) assume that human faces in an image or an image sequence have been already segmented. Recently, a visual object detection framework was proposed and applied to face detection [16, 17]. Though the framework is capable of processing images rapidly with achieving high detection rate, it focuses on rapidly detecting human faces as rectangle regions and does not pay any attention to the contours of their appearances.

To build a fully automated system that recognizes human faces from images, it is essential to develop robust and efficient algorithms to detect human heads and, at the same time, to identify face orientations. Given a single image or a sequence of images, the goal of automatic human-face recognition is to detect human
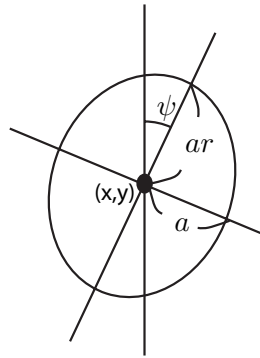
Figure 1: Geometric model of human head appearances.

heads/faces and estimate their orientations regardless of not only their positions, scales, orientations, poses, but also individuals, background changes and lighting conditions.

This paper proposes a two-step method for detecting human heads and, at the same time, for estimating face orientations by a monocular camera. In the both steps, we employ models of the human-head contour and face orientations to enhance robustness and stableness in detection. We also introduce model evaluation with only image-features robust against lighting conditions, i.e., the gradient of intensity and texture.

In the first step, our method employs an ellipse as the contour model of human-head appearances to deal with wide variety of appearances. The ellipse is generated from one ellipsoid based on the camera position with its angle of depression in the environment. Our method then evaluates the ellipse over a given image to detect possible human heads. In evaluation of an ellipse, two other ellipses are generated inside and outside of the ellipse, and the gradient of intensity along the perimeter of the three ellipses is used for accurate detection of human-head appearances.

In the second step, on the other hand, our method focuses on facial components such as eyes, the mouth or cheeks to generate inner models for face-orientation estimation. Based on the camera position with its angle of depression, our method projects the facial components on the ellipsoid onto the ellipse to generate inner models of human-head appearances. Our method then evaluates not only such components themselves but their geometric configuration to eliminate false positives in the first step and, at the same time, to estimate face orientations. Here the Gabor-Wavelets filter, which is verified its robustness and stableness against changes in scale, orientation and illumination, is used for detecting features representing the facial components.

Consequently, our method can correctly and stably detect human heads and estimate face orientations even under environments such as illumination changes or face-orientation changes. Our intensive experiments using a face-image database and real-situation images show the effectiveness of the proposed method.

## 2   Contour model for human-head appearances

The model-based approach is inevitable to enhance stableness against environment changes. This is because features detected from images without any models often generate false positives in recognition.

### 2.1   Human head and its appearances

Human beings have almost the same contour in shape of the head and an ellipse approximates the appearance of the contour. These observations remain invariant against changes in face orientation. We, therefore, model the contour of human-head appearances by the ellipse [1, 12, 19]. An ellipse has five parameters in the image (Fig. 1): the 2D coordinates $(x, y)$ of the ellipse center, the length $a$ of the semiminor axis, the oblateness $r$, and the slant $\psi$ of the ellipse.
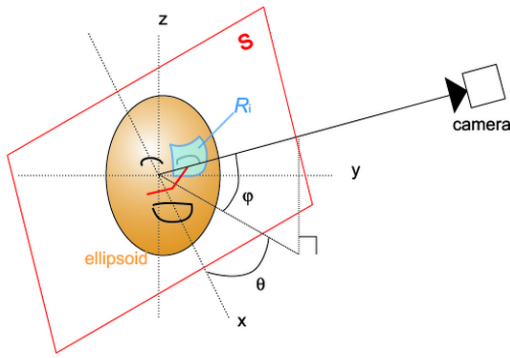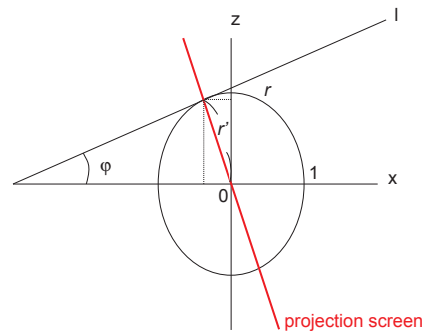
Figure 2: Human-head model.

Figure 3: Projection of the ellipsoid onto the $xz-$plane.

These parameters of the ellipse change depending on the angle of depression of a camera even though human heads are in the same pose. In particular, the change in oblateness is outstanding. To investigate this change in oblateness, we introduce an ellipsoid to the human-head model in 3D. We assume that the ellipsoid is represented in the world coordinates by

$$x^2 + y^2 + \frac{z^2}{r^2} \;=\; 1, \tag{2.1}$$

where $r \geq 1$. We then derive an ellipse as the contour model of human-head appearances depending on the angle of depression of the camera (Fig. 2).

When we set up a camera with any angle of depression, the ellipsoid (2.1) is observed as an ellipse. The length of the semiminor axis of the ellipse is always one. The length of the semimajor axis, on the other hand, is between one and $r$ depending on the angle of depression of the camera.

Now we determine the oblateness, $r'$ ($1 \leq r' \leq r$), of the ellipse observed by a camera with $\varphi$ angle of depression providing that the distance of the camera position from the ellipsoid is large enough. We consider the ellipse obtained through the projection of (2.1) onto the $xz-$plane and its tangential line $\ell$ (Fig. 3). We see that the ellipse, the projection of (2.1) onto the $xz-$plane, is represented by

$$x^2 + \frac{z^2}{r^2} \;=\; 1. \tag{2.2}$$

Let its tangential line with slant $\varphi$ from the $x-$axis be

$$z \;=\; \sin\varphi x + b, \tag{2.3}$$

where $b$ is the $z$-intercept. Combining (2.2) and (2.3), we can compute $b$. We then have the coordinates of their contact point, from which it follows that

$$r' \;=\; \sqrt{\frac{r^4 + \tan^2\varphi}{r^2 + \tan^2\varphi}}.$$

This relates $\varphi$, i.e., the angle of depression of the camera, with $r'$, i.e., the oblateness of the ellipse representing the contour of human-head appearances. We dynamically compute the oblateness of the ellipse from the camera position based on this relationship.

## 2.2 Evaluation of contour model

When we are given an ellipse in the image, how to evaluate the goodness of the ellipse to recognize as a human-head appearance is a difficult problem. Employing image features invariant under changes in environment is indispensable.
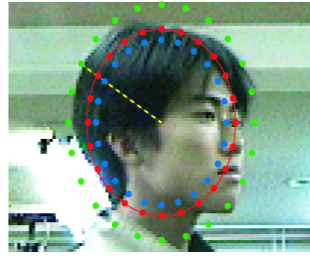
Figure 4: The (red) ellipse representing a human-head appearance.

Color information is widely used there. Color information is robust against scale changes but sensitive to changes in illumination. To overcome this problem, eliminating the luminance from color information and evaluating chromaticity is proposed [1, 13, 19]. The effectiveness of this approach is, however, limited.

We employ, in this paper, the gradient of intensity in evaluating an ellipse to identify whether it is an applicant of human-head appearances. This is because the gradient of intensity is robust against illumination changes.

When we fit an ellipse to the contour of a human-head appearance, we have the following observations (Fig. 4):

- Great gradient magnitude of intensity at the ellipse perimeter.

- Continuous changes in intensity along the ellipse perimeter except for the boundary between hair and skin.

- Continuous changes in intensity from just inside the ellipse.

We thus evaluate a given ellipse in three different ways. One is evaluation on the gradient magnitude of intensity at the perimeter of the ellipse. Another is evaluation on intensity changes along the perimeter of the ellipse and the other is evaluation on intensity changes from the adjacent part inside the ellipse. Introducing these three aspects in evaluation of an ellipse results in more accurately and more robustly obtaining applicants of human-head appearances.

For evaluating an ellipse, we construct two other ellipses (Fig. 5). One is a smaller size ellipse with the identical center and the other is a larger size ellipse with the identical center. In Fig. 5, the red ellipse is to be evaluated and the blue ellipse is the smaller size one and the green is the larger size one. We denote by $orbit(i)$ the intensity of the intersection point of the (red) ellipse to be evaluated and ray $i$ whose end point is the ellipse center. We remark that we have $N$ rays with the same angle-interval and they are sorted by the angle from the horizontal axis in the image. $outer(i)$ and $inner(i)$ are defined in the same way for the cases of the larger size ellipse (green ellipse) and the smaller size ellipse (blue ellipse), respectively.

We now have the following function evaluating the (red) ellipse.

$$f(\boldsymbol{p}) \;\; = \;\; k\frac{1}{N}\sum_{i=1}^{N}\left\{G(i) - O(i) - I(i)\right\}, \tag{2.4}$$

where $\boldsymbol{p}$ is the parameter vector representing the red ellipse and

$$G(i) \;\; = \;\; |outer(i) - orbit(i)|, \tag{2.5}$$
$$O(i) \;\; = \;\; |orbit(i) - orbit(i-1)|, \tag{2.6}$$
$$I(i) \;\; = \;\; |orbit(i) - inner(i)|. \tag{2.7}$$

Note that $k$ is the constant making the value dimensionless [1]. (2.5), (2.6), and (2.7) evaluate the gradient magnitude of intensity at the ellipse perimeter, intensity changes along the ellipse perimeter and intensity changes from just inside the ellipse, respectively. Ellipses having a small value of (2.4) are then regarded as
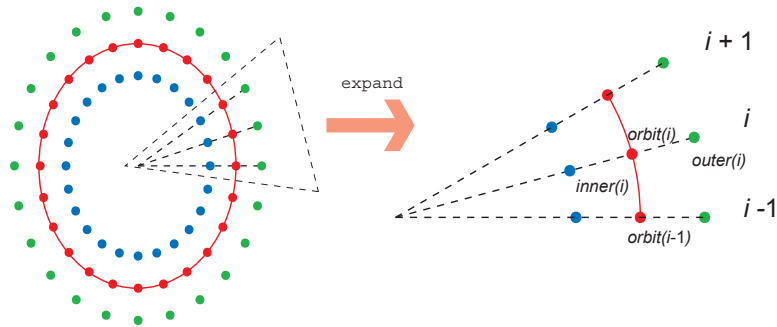
Figure 5: Evaluation of the ellipse (red: the ellipse to be evaluated).



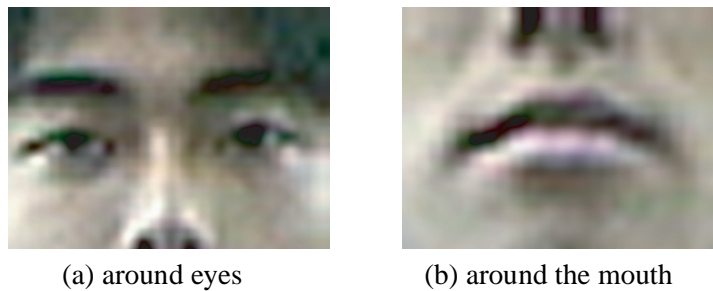(a) around eyes  (b) around the mouth

Figure 6: Textures around facial components.

applicants of human-head appearances. We remark that our ellipse evaluation is effective even if a face region is darker than the surrounding background. This is because our evaluation is based on not intensity itself but the gradient magnitude of intensity.

In the next section, we evaluate the applicants of human-head appearances based on features inherent in the human face to recognize as a human-head appearance and, at the same time, to identify the face orientation.

## 3 Inner models for face orientations

We investigate inside the ellipse in more detail to detect human heads and face orientations providing that applicants of the human heads are already detected as ellipses. In detection, these pre-obtained applicants facilitate determination of parameters such as scale or direction.

### 3.1 Facial components

Eyebrows, eyes, the mouth, the nose and cheeks are the features inherent in the human face. Here we focus on eyes, the mouth and cheeks, and characterize textures around such facial components. We remark that textures are robust against illumination changes.

In oriental countries, we observe around eyes (1) a dark area due to eyebrows, (2) a bright area due to eyelids, and (3) a dark area due to the pupil (see Fig. 6 (a)). These are observations along the vertical direction of the human face and these characterize the texture of an eye area. We also observe that the eye area is symmetrical with respect to the pupil. As for an area around the mouth, on the other hand, we observe (1) a bright area due to the upper lip, (2) a dark area due to the mouth, and (3) a bright area due to the lower lip (see Fig. 6 (b)). In addition, the mouth area is also symmetrical with respect to the vertical center of the face. These observations characterize the texture of a mouth area. We see no complex textures in a cheek area. These observations are almost invariant and stable under changes in illumination, in face-orientation and in scale.

(a) Real-part of a kernel                    (b) Kernels with different orientations
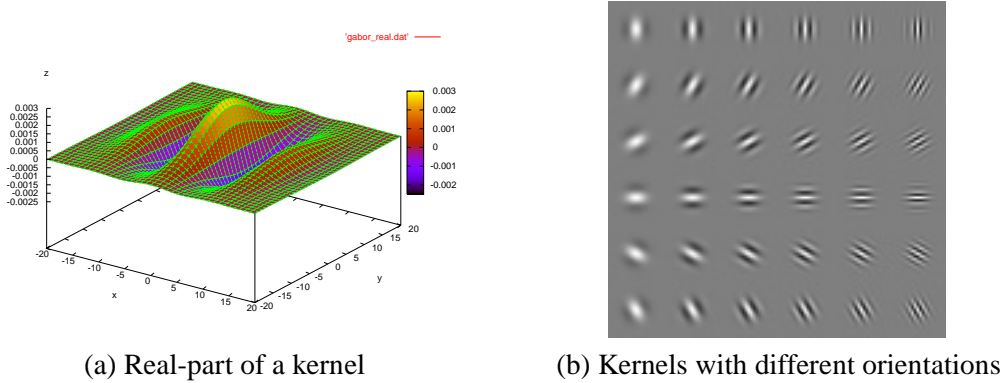
Figure 7: Gabor kernels.

The geometric configuration of the facial components, i.e., the relative position between eyes, the mouth and cheeks is also invariant. Combining the characteristic of textures of the facial components with their geometric configuration enables us to stably recognize human heads/faces.

For each applicant of human-head appearances, we detect the facial components with their geometric configuration to verify whether it is a human-head appearance. We remark that we can easily identify the scale in detecting facial components since we have already obtained applicants of human-head appearances in terms of ellipses.

## 3.2   Detecting facial components using Gabor-Wavelets

In detecting facial feature points described in the previous section, the Gabor-Wavelets filter is most promising in robustness and stableness against illumination changes [5, 11, 14, 15, 22]. We thus use Gabor-Wavelets to extract the facial feature points, eyes, the mouth and cheeks, as a set of multi-scale and multi-orientation coefficients.

Applying the Gabor-Wavelets filter to a point $(x_0, y_0)$ of a given image $f(x, y)$ can be written as a convolution

$$\psi(x_0, y_0, \sigma, \omega, \phi) = \iint \mathrm{d}x\mathrm{d}y f(x, y) G(x - x_0, y - y_0, \sigma, \omega, \phi)$$

with Gabor kernel $G(x, y, \sigma, \omega, \phi)$s where $G$ is formulated [11] by

$$G(x, y, \sigma, \omega, \phi) \quad = \quad \kappa \mathrm{e}^{\frac{-1}{4\pi\sigma^2}(\tilde{x}^2 + \tilde{y}^2)} \mathrm{e}^{\mathrm{j}\omega\tilde{x}}.$$

Here

$$\left[ \begin{array}{c} \tilde{x} \\ \tilde{y} \end{array} \right] = \left[ \begin{array}{cc} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{array} \right] \left[ \begin{array}{c} x \\ y \end{array} \right], \qquad \kappa = \frac{1}{4\pi^2\sigma^2}, \qquad \mathrm{j} = \sqrt{-1}.$$

$\sigma, \omega, \phi$ are the parameters representing the scale, frequency and orientation, respectively. Note that $(\tilde{x}, \tilde{y})$ is obtained by rotating image point $(x, y)$ by $\phi$.

Figure 7 shows an example of a set of Gabor-Wavelets. (a) is the real part of the Gabor kernel with $\sigma = 3.0, \omega = 0.5^\mathrm{o}, \phi = 0^\mathrm{o}$, and (b) is the kernels with the same scale, different orientations and frequencies.

We can selectively apply the Gabor-Wavelets filter to particular locations. In addition, we can easily specify scales, frequencies, and orientations in the application of the Gabor-Wavelets filter. In other words, we can apply the Gabor-Wavelets filter to specific regions in the image, i.e., pre-obtained applicants of human-head appearances, with selective parameters in scale, frequency, and orientation to extract a feature vector. This is
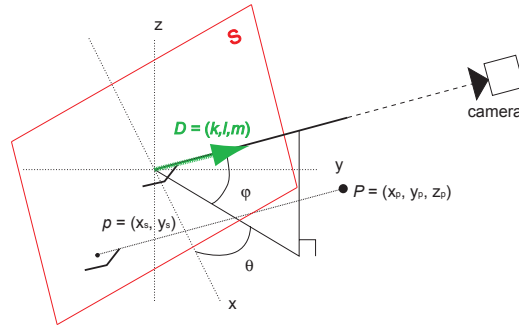
Figure 8: Geometry in deriving an inner model.

because we have already detected applicants of human-head appearances in terms of ellipses (we have already roughly estimated a size of a human-head appearance). This reduces the computational cost in recognizing human-head appearances in the practical sense.

We characterized in Section 3.1 textures around eyes and the mouth along the vertical direction of the human face. To detect these textures we only have to select the parameters in the Gabor-Wavelets filter so that the filter detects the textures along the semimajor axis of the ellipse. Points with maximal values in the response ellipse-region can be eyes and those with minimal values can be a mouth. The area with no singularity, on the other hand, can be cheeks.

### 3.3   Inner models of head appearances with facial components

We generate here inner models of human-head appearances based on the ellipsoid (2.1). As shown in Fig. 2, area $R_i$ on the ellipsoid denoting a facial component such as an eye or a mouth is projected onto the plane when it is viewed from the camera with direction $(\theta, \varphi)$, where $\theta$ is the rotation angle toward the camera from the front of the face and $\varphi$ is the angle of depression of the camera. The projected area then enables us to identify the location of the facial component in the human-head appearance. Hence, we can generate the inner models of human-head appearances. We remark that we can measure $\varphi$ in advance when we set up a camera in the environment.

We consider plane $S$ that goes though the origin and whose normal vector is identical with the viewing line of the camera (Fig. 8). Let $\boldsymbol{D} = (k, l, m)^\top$ be the unit normal vector of plane $S$. $S$ is then expressed by $kx + ly + mz = 0$. It is easy to see that $k, l$ and $m$ are expressed in terms of $\theta$ and $\varphi$:

$$
\begin{aligned}
k &= \cos\varphi\cos\theta, \\
l &= \cos\varphi\sin\theta, \\
m &= \sin\varphi.
\end{aligned}
$$

Letting $\boldsymbol{p}$ be the foot of the perpendicular from a 3D point $\boldsymbol{P}$ onto $S$, we can easily relate $\boldsymbol{P}$ and $\boldsymbol{p}$ by

$$
\boldsymbol{p} = \left( \begin{array}{c} \boldsymbol{X}_S^\top \\ \boldsymbol{Y}_S^\top \end{array} \right) \boldsymbol{P}.
$$

Here $\boldsymbol{X}_S^\top, \boldsymbol{Y}_S^\top$ are the orthogonal unit vectors in 3D representing the coordinates in $S$:

$$
\boldsymbol{X}_S = \frac{1}{\sqrt{1-m^2}} \left( \begin{array}{c} l \\ -k \\ 0 \end{array} \right), \qquad \boldsymbol{Y}_S = \frac{m}{\sqrt{1-m^2}} \left( \begin{array}{c} -k \\ -l \\ \frac{1-m^2}{m} \end{array} \right).
$$

In this way, when depression angle $\varphi$ and rotation angle $\theta$ are specified, we can project a facial area of the ellipsoid onto the image plane to obtain an inner model of the human-head appearance that represents the facial components with their geometric configuration.
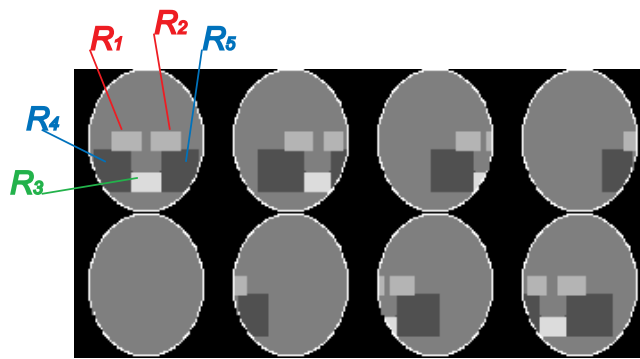
Figure 9: Inner models of human-head appearances with the facial components.

Figure 9 shows the inner models of human-head appearances with $\varphi = 0$ (upper: $\theta = 0^\circ, 30^\circ, 60^\circ, 90^\circ$, lower: $\theta = 180^\circ, 270^\circ, 300^\circ, 330^\circ$). $R_1$ and $R_2$ denote the eye areas. $R_3$ denotes the mouth area, and $R_4$ and $R_5$ denote the cheek areas.

To the response ellipse-region of the Gabor-Wavelets filter, we apply the inner model matching to detect human-head appearances and face orientations. To be more concrete, if we find eyes, a mouth and cheeks in a response ellipse, we then identify that the ellipse is a human-head appearance and that the orientation of the matched inner-model is the face orientation of the appearance. Otherwise, we identify that the ellipse is not a human-head appearance and eliminate the ellipse.

## 4   Algorithm

Based on the discussion above, we describe here the algorithm for detecting human-head appearances with face orientations.

To reduce the computational cost in generating applicants of human-head appearances, we introduce the coarse-to-fine sampling of the parameters representing ellipses. Namely, we first coarsely sample points in the parameter space for the ellipse and then minutely sample the area around the points that are selected based on plausibility of the human-head appearance. Moreover, in the coarse sampling, we fixate parameters depending only on poses of a human head to enhance position identification of the human head. In the fine sampling, we sample all the parameters. The following algorithm effectively detects human heads and, at the same time, estimates their orientations.

**Step 1:**  Capture an image.

**Step 2:**  Search applicants of human heads in the image.

  **2.1:** **(Coarse sampling):** randomly sample position parameters, i.e., $(x, y, a)$, in the parameter space representing the ellipses that are generated from (2.1); let $\{\boldsymbol{p}_i\}$ be the sampled set.

  **2.2:** Evaluate each entry of $\{\boldsymbol{p}_i\}$ by (2.4); let $\{\boldsymbol{p}_{i*}\}$ be the set of samples whose scores of $f$ in (2.4) are less than a given threshold.

  **2.3:** **(Fine sampling):** more minutely sample points in the area around each entry of $\{\boldsymbol{p}_*\}$ (more specifically, more minutely sample parameters $(x, y, a, r, \psi)$ around each entry of $(x_{i*}, y_{i*}, a_{i*})$ where $(x_{i*}, y_{i*}, a_{i*})$ is the position parameters of $\{\boldsymbol{p}_{i*}\}$); let $\{\boldsymbol{p}_j^*\}$ be the sampled set. (Note that $\{\boldsymbol{p}_j^*\}$ is applicants of human-head appearances.)

**Step 3:**  To each entry of $\{\boldsymbol{p}_j^*\}$, generate inner models of human-head appearances.

**Step 4:**  Apply the Gabor-Wavelets filter to each entry of $\{\boldsymbol{p}_j^*\}$ to detect facial feature points.
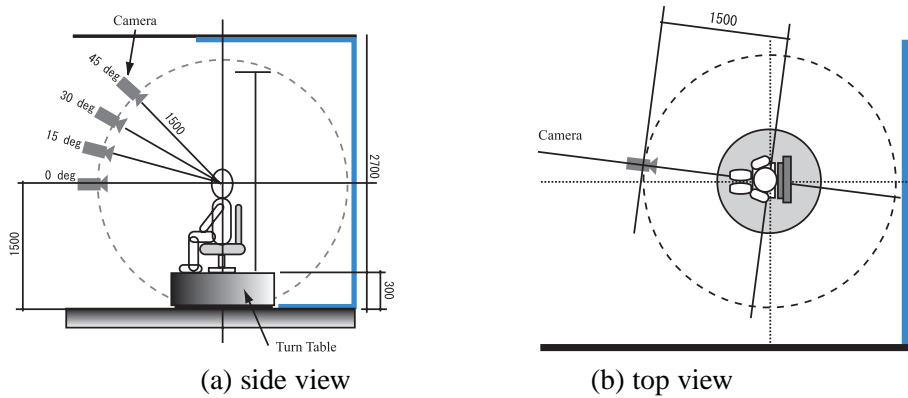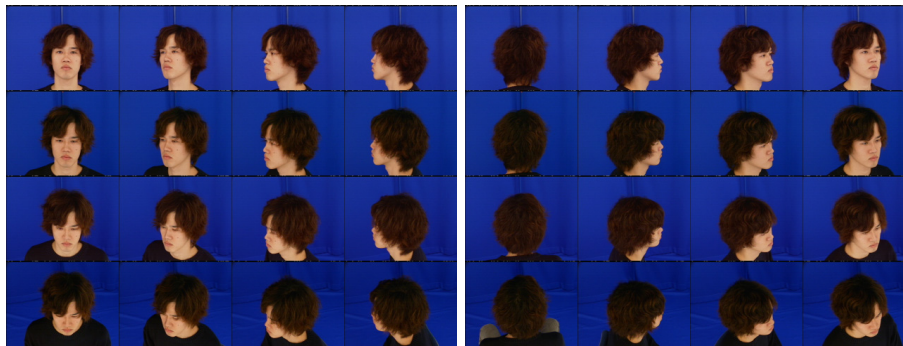
(a) side view　　　　　　　　　　(b) top view

Figure 10: Parameters in obtaining a face-image database.



Figure 11: Sample images of the face-image database (with $0^{\circ}$ of depression).

**Step 5:** To each $p_j^*$, apply the matching with the corresponding inner models.

**Step 6:** If $p_j^*$ matches one of its corresponding inner models with a high score, then recognize $p_j^*$ as a human-head appearance and the face orientation as that of the matched inner-model. If $p_j^*$ does not match any of its corresponding inner models with a high score, then eliminate $p_j^*$.

We remark that iterating the steps above enables us to track human heads with their orientations. Though introducing a transition model of motion to our detection algorithm leads to more effective tracking, it is beyond the scope of this paper.

## 5 Experimental evaluation

### 5.1 Evaluation on face orientations using a face-image database

We first evaluated our algorithm using a face-image database. The database contains face images of 300 persons with the ages ranging uniformly from 15 to 65 years old including men and women. Each person is taken his/her face images from different directions as shown in Fig. 10. To each face image in the database, attached is the ground truth of the direction from which the image is taken.

We used 9600 ($= 32 \times 300$) face images in the database where 32 directions are used in taking images of each person: the angles of depression of the camera were $\varphi = 0^{\circ}$, $15^{\circ}$, $30^{\circ}$, $45^{\circ}$ and the rotation angles with respect to the horizontal axis, i.e., face orientations, were $0^{\circ}$, $30^{\circ}$, $60^{\circ}$, $90^{\circ}$, $180^{\circ}$, $270^{\circ}$, $300^{\circ}$, $330^{\circ}$. Fig. 11 shows samples of the face images of one person in the database.

Table 1: Recognition accuracy for different face orientations (%).

| angles of depression | face orientations | | | | | | | |
|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| | 0º | 30º | 60º | 90º | 180º | 270º | 300º | 330º |
| 0º | 84.7 | 86.3 | 83.7 | 31.0 | 97.0 | 34.7 | 80.0 | 79.3 |
| 15º | 64.7 | 86.3 | 75.3 | 27.7 | 97.7 | 21.0 | 71.7 | 71.7 |
| 30º | 23.7 | 75.3 | 70.3 | 14.0 | 99.0 | 10.0 | 51.0 | 51.7 |
| 45º | 17.7 | 61.7 | 51.0 | 16.0 | 94.7 | 8.3 | 27.0 | 27.0 |

We applied our algorithm to the 9600 images to detect face orientations. Table 1 shows the recognition rates of the estimated face-orientations.

Table 1 shows that face orientations are in general recognized with high scores. We see low accuracy in orientations with $90º$ and $270º$. This is because one eye and one cheek do not appear in the face with such orientations and thus the inner model matching becomes unstable. We also see that accuracy becomes higher as the angle of depression of the camera becomes smaller. The small angle of depression of the camera means that the face is captured from the horizontal direction of the face and that the facial components clearly appear in the image. It is understood that clearly appearing facial components improves the estimation accuracy of face orientations. A large angle of depression, on the other hand, causes great changes not only in human-head appearance but also in face appearance. Handling such great changes with our models has limitation. This is because we generate a contour model and inner models of human-head appearances from only one ellipsoid. On the other hand, we see that face images from the orientation with $180º$, back images of human heads, are recognized stably and accurately independent of the change in angle of depression. This is due to stableness of the Gabor-Wavelets filter in face-feature detection.

## 5.2    Evaluation in the real situation

Secondly, we evaluated the performance of our method in the real situation. We conducted two kinds of experiments here. One focused on human-head detection using real images and the other focused on ellipse evaluation in human-head detection. In both cases, we found the robustness and the efficiency of our method.

### 5.2.1    Human-head detection in the real situation

We here apply our method to real images to see its effectiveness.

We set up a camera with about 2m height and with about $20º$ angle of depression. Then, under changing lighting conditions one person turns round in front of the camera with changing his face orientations and walks for about 10 seconds. The distance from the camera to the person was about 1.5m. To the 100 captured images, we applied our method to detect human-head appearances and face orientations.

Figure 12 shows examples of the captured images with frame numbers. The ellipses detected as a human-head appearance are superimposed on the frames where their colors denote face orientations. We see that appearances of human heads are fairly large in the images.

We observed that human heads are incorrectly detected in several images. In such images, ellipses are detected in the background. This is party because of existing objects in the background; in fact, ellipses are easy to be detected in the background used in this experiment. Nevertheless, human-head appearances are detected correctly and accurately in the rest of the all images.

To quantitatively evaluate the correctness in detection, we first manually fitted an ellipse onto the head appearance in each image to obtain the true ellipse as the reference. We then computed the distance (the position error) between the center of the detected ellipse and that of the true ellipse. We also computed the

Figure 12: Images (with frame numbers) and detected human heads (colors means orientations).

ratio (the size error) of the semiminor length of the detected ellipse to that of the true ellipse. These results are shown in Fig.13. The average and the standard deviation of errors in position were respectively, $8.85$pixels and $22.4$pixels. Those in size were, on the other hand, $0.0701$ and $0.0485$, respectively. Note that difference of the size errors from $1.0$ was employed in this computation.

We see that our method for detecting human heads and face orientations is practical overall in the real situation. An attached video sequence (video-1) demonstrates the performance of our method.

### 5.2.2   Effectiveness of human-head evaluation

We here focus on evaluation of ellipses in human-head detection and show effectiveness of our ellipse evaluation.

We set up a camera with 1.8m height and with about $0^{\circ}$ angle of depression. We generated the situation in which under changing lighting conditions one person walks around in front of the camera with distance between about 2m and 4m for 20 seconds. 200 images were captured during the time. To the captured images, we applied our method to detect human-head appearances and face orientations.

Figure 14 shows examples of the captured images with frame numbers. The ellipses detected as a human-head appearance are superimposed on the frames. Colors of the ellipses denote face orientations. We see that in this experiment appearances of human heads are small in the images.

We again verified that human-head appearances are detected almost correctly and accurately in all the images in spite of changes in illumination. To see the performance of our method, we evaluated the accuracy of the
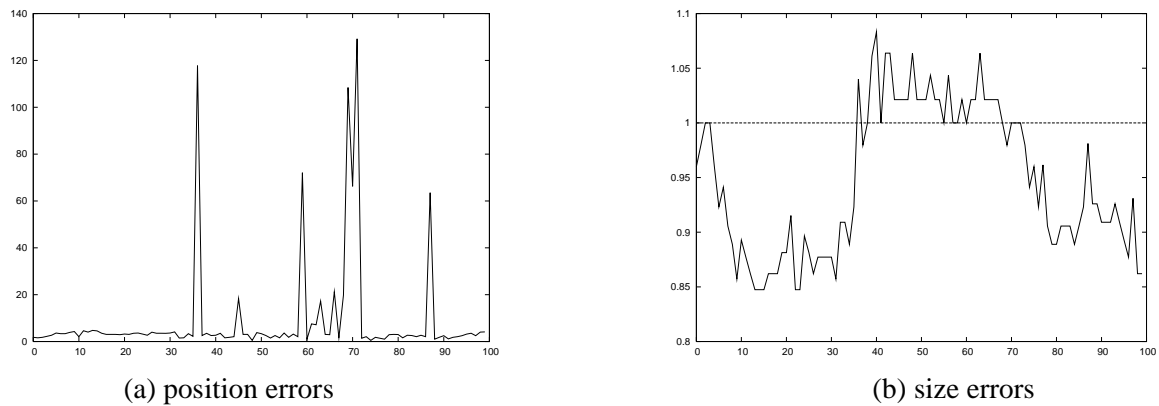
(a) position errors

(b) size errors

Figure 13: Errors in human-head detection.

detected human-head appearances. We first fitted an ellipse onto the head appearance in each image by hand to obtain the true ellipse as the reference. We introduced two kinds of evaluation to the ellipse that was recognized as the human-head appearance: one is the accuracy of the center and the other is the accuracy of the semiminor length. We computed the distance (the position error) between the center of the detected ellipse and that of the true ellipse. We also computed the ratio (the size error) of the semiminor length of the detected ellipse to that of the true ellipse. These results are shown in Figs. 15 and 16. We remark that the same evaluation was applied to the method (called the *simple-evaluation method* (cf. [1, 19])) where the ellipse is evaluated only by (2.5), i.e., the gradient magnitude of intensity at the ellipse perimeter. For the position error and the difference of the size error from 1.0, the average and standard deviation over the image sequence were calculated, which is shown in Table 2.

Figures 15,16 and Table 2 show the effectiveness of our method. Superiority of our method to the simple-evaluation method indicates that introducing the smaller- and larger-size ellipses to ellipse evaluation improves the accuracy in detecting the positions of human-head appearances. An attached video sequence (video-2) demonstrates the performance comparison of our method and the simple-evaluation method (The right-hand side is the result by the proposed method while the left-hand side is that by the simple evaluation method; colors indicates orientations).

# 6   Conclusion

We proposed a two-step method for detecting human heads and estimating face orientations by a monocular camera. In the both steps, we employ models of the human-head contour and face orientations to enhance robustness and stableness in detection. We also introduced model evaluation with only image-features robust against lighting conditions.

Table 2: Errors in detecting the human head.

| error | | our method | simple evaluation |
|---|---|---|---|
| position | mean [pixels] | 3.250 | 6.401 |
| | standard deviation [pixels] | 1.950 | 4.386 |
| size | mean | 0.0762 | 0.0617 |
| | standard deviation | 0.04489 | 0.04877 |

Figure 14: Examples of detected human heads (colors means orientations).

The first step employs an ellipse as the contour model of human-head appearances to deal with wide variety of appearances. The ellipse was generated from one ellipsoid based on a camera position with its angle of depression in the environment. We then evaluated the ellipse over a given image to detect possible human-head appearances where we generated two other ellipses inside and outside of the ellipse to improve accuracy in detection of human-head appearances.

The second step, on the other hand, focuses on facial components such as eyes, the mouth or cheeks to generate inner models for face-orientation estimation. We evaluated not only such components themselves but also their geometric configuration to eliminate false positives in the first step and, at the same time, to estimate face orientations. Here we used the Gabor-Wavelets filter in detecting features representing the facial components because its robustness and stableness against changes in scale, orientation and illumination are verified.

Consequently, our method can correctly and stably detect human heads and estimate face orientations even under changes in face orientation and in illumination. Our intensive experiments showed the effectiveness of the proposed method. Incorporating wider variety of face orientations into our method is left for future work.

## Acknowledgements

## References

[1] S. Birchfield: Elliptical Head Tracking Using Intensity Gradients and Color Histograms, *Proc. of CVPR*, pp. 232–237, 1998.

(a) trajectory of position          (b) position errors
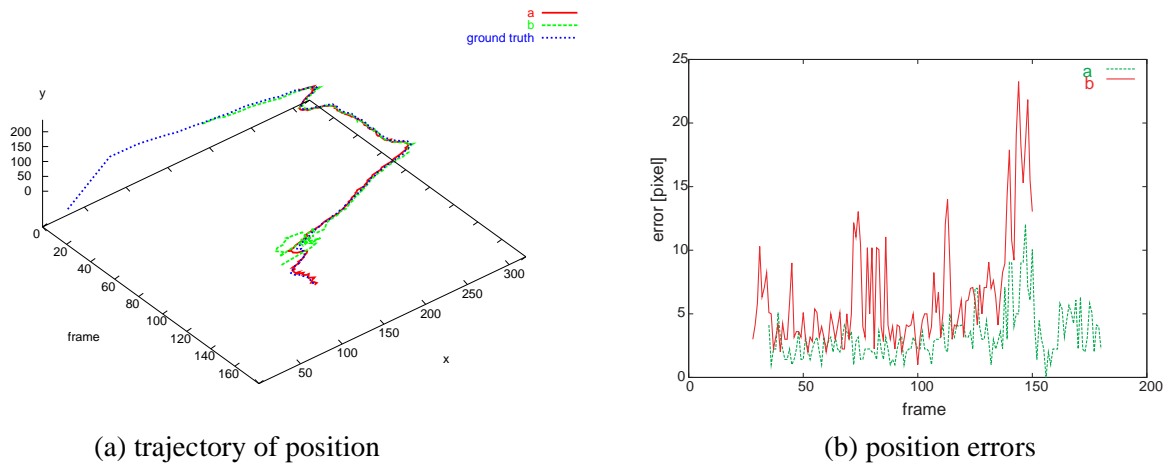
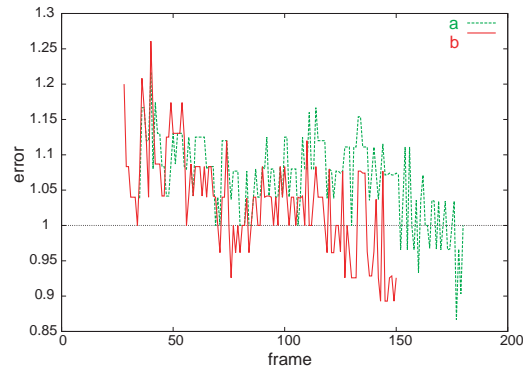Figure 15: Position errors in human-head detection (a: our method, b: simple evaluation).



Figure 16: Size errors in human-head detection (a: our method, b: simple evaluation).

[2] T. J. Cham and J. M. Rehg: A Multiple Hypothesis Approach to Figure Tracking, *Proc. of CVPR*, Vol. 2, pp. 239–245, 1999.

[3] R. Chellappa, C. L. Wilson and S. Sirohey: Human and Machine Recognition of Faces, A Survey, *Proc. of IEEE*, 83 (1995), pp. 705–740.

[4] Y. Cui, S. Samarasekera, Q. Huang and M. Greiffenhagen: Indoor Monitoring via the Collaboration between a Peripheral Sensor and a Foveal Sensor, *Proc. of the IEEE Workshop on Visual Surveillance*, pp. 2–9, 1998.

[5] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman and T. J. Sejnowski: Classifying Facial Actions, *IEEE Trans. on PAMI*, 21 (1999), 10, pp. 974–989.

[6] L. Davis, S. Fejes, D. Harwood, Y. Yacoob, I. Hariatoglu and M. J. Black: Visual Surveillance of Human Activity, *Proc. of the 3rd ACCV*, Vol.2, pp. 267–274, 1998.

[7] D.M. Gavrila: The Visual Analysis of Human Movement: A Survey, *Computer Vision and Image Understanding*, 73 (1999), 1, pp. 82–98.

[8] I. Haritaoglu, D. Harwood and L. S. Davis: W$^4$S: A Real-Time System for Detecting and Tracking People in $2\frac{1}{2}$D, *Proc. of the 5th ECCV*, Vol. 1, pp. 877–892, 1998.

[9] I. Haritaoglu, D. Harwood and L. S. Davis: An Appearance-based Body Model for Multiple People Tracking, *Proc. of the 15th ICPR*, Vol. 4, pp. 184–187, 2000.

[10] T.-K. Kim, H. Kim, W. Hwang, S.-C. Kee and J. Kittler: Independent Component Analysis in a Facial Local Residue Space, *Proc. of CVPR*, 2003.

[11] T. S. Lee: Image Representation Using 2D Gabor Wavelets, *IEEE Trans. on PAMI*, 18 (1996), 10, pp. 959–971.

[12] A. Sugimoto, K. Yachi and T. Matsuyama: Tracking Human Heads Based on Interaction between Hypotheses with Certainty, *Proc. of the 13th Scandinavian Conf. on Image Analysis*, (J. Bigun and T. Gustavsson eds: *Image Analysis*, Lecture Notes in Computer Science, Vol. 2749, Springer), pp. 617–624, 2003.

[13] M. Swain and D. Ballard: Color Indexing, *Int. J. of Computer Vision*, 7 (1991), 1, pp. 11–32.

[14] Y. Tian, T. Kanade and J. F. Cohn: *Recognizing Facial Actions by Combining Geometric Features and Regional Appearance Patterns*, CMU-RI-TR-01-0, Robotics Institute, CMU, 2001.

[15] Y. Tian, T. Kanade and J. F. Cohn: Evaluation of Gabor-Wavelets Based Facial Action Unit Recognition in Image Sequences of Increasing Complexity, *Proc. of the 5th Int. Conf. on Automatic Face and Gesture Recognition*, pp. 229–234, 2002.

[16] P. Viola and M. Jones: Rapid Object Detection using a Boosted Cascade of Simple Features, *Proc. of CVPR*, Vol. I, pp. 511-518, 2001.

[17] P. Viola and M. Jones: Robust Real-Time Face Detection, *Int. J. of Computer Vision*, Vol. 57, No. 2, pp. 137–154, 2004.

[18] Y. Wu and K. Toyama: Wide-Range, Person- and Illumination-Insensitive Head Orientation Estimation, *Proc. of the 4th IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp. 183–188, 2000.

[19] K. Yachi, T. Wada and T. Matsuyama: Human Head Tracking using Adaptive Appearance Models with a Fixed-Viewpoint Pan-Tilt-Zoom Camera *Proc. of the 4th FG*, pp. 150–155, 2000.

[20] Z. Zeng and S. Ma: Head Tracking by Active Particle Filtering, *Proc. of the 5th IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp. 89–94, 2002.

[21] L. Zhang and D. Samaras: Face Recognition under Variable Lighting using Harmonic Image Exemplars, *Proc. of CVPR*, 2003.

[22] Z. Zhang, M. Lyons, M. Schuster and S. Akamatsu: Comparison between Geometry-based and Gabor-Wavelets-based Facial Expression Recognition using Multi-layer Perception, *Proc. Int. Workshop on Automatics Face and Gesture Recognition*, pp. 454–459, 1998.

[23] W. Y. Zhao, R. Chellappa, A. Rosenfeld and P. J. Phillips: *Face Recognition: A Literature Survey*, CAR-TR-984, UMD, 2000.

[24] S. Zhou, V. Krueger and R. Chellappa: Face Recognition from Video: A Condensation Approach, *Proc. of the 5th IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp. 221–226, 2002.