Regular Paper

Unconstrained Person Identification from Ceiling Using Multiview Learning for Tabletop Users

Yasuo Namioka[1,a) Akira Masuda[2,b) Takuya Maekawa[2,c)

Received: December 23, 2016, Accepted: July 4, 2017

Abstract: A novel unconstrained person identification method is presented in this paper. This method is for tabletop systems which exist not only in daily life but also in working environments such as offices and factories. Recent state-of-the-art ubicomp, computer-vision, and CSCW studies have tried to recognize a user's activities and actions on a table using a ceiling-mounted device that overlooks the table, since we can install the ceiling-mounted device in an environment with limited space such as daily life environments and factory environments. Instead of conventional unconstrained person identification methods, such as face identification, we focus on a user's soft biometrics that can be captured from the ceiling such as the shoulder length, shape of the head, and posture of the back to achieve unconstrained person identification by using a ceiling-mounted depth camera. We achieve robust person identification by combining the soft biometrics within a framework of multiview learning. Multiview learning allows us to deal effectively with data consisting of features from multiple sources with different data distributions, i.e., multiple soft biometrics in our case. Our experimental evaluation revealed that our proposed method achieved high identification accuracy of about 94%.

Keywords: person identification, depth sensor, multiview learning

1. Introduction

In recent years there has been a proliferation of tabletop systems due to the progress of ICT technologies. For example, tabletop display systems for browsing information and playing games, such as Microsoft PixelSense have been commercialized. In addition, because tables are readily available in environments that are typical of everyday life, tabletop systems have been developed for supporting and augmenting various daily activities, e.g., recording or supporting discussion, meal preparation, eating, and studying [1], [2], [3]. In addition, recognition of working activities on a table, e.g., medical work and assembly work, has been studied in the pervasive computing and vision research communities [4], [5]. In order to record activities on a table and/or to provide real-time feedback according to a user's action or gesture on a table, recent state-of-the-art ubicomp, computer-vision, and HCI studies have employed a ceiling-mounted camera device that overlooks the table [6], [7] as shown in **Fig. 1**.

Meanwhile, to provide a personalized service to tabletop users, e.g., assembly-work management, personalized lifelogging, CSCW, and personalized recommendation for tabletop displays, person identification technologies for tabletops have been actively studied. Many existing tabletop systems perform person identification by using wireless tags such as RFID tags possessed by users [8]. However, this approach is burdensome for users, making its application difficult in everyday environments.

With a view to achieving unconstrained person identification, many vision and pervasive computing studies have employed face identification using a camera and identification based on skeleton information obtained from a depth sensor, e.g., Microsoft Kinect [9], [10]. However, these approaches have the following problems when applied to tabletop systems.

- The face-based identification methods employ images captured by a camera in front of the user. Therefore, a sensor device should be installed in front of a user, but the types of environments in which it is practicable to install such a sensor device are limited. For example, it is difficult to place a sensor device on a table because the device interferes with work and other activities performed on the table. When a sensor cannot be mounted on a table, it should be mounted on or embedded in a wall in front of a user. However, embedding a sensor device in a wall, e.g., a wall in a kitchen, is difficult.
- If a camera is embedded in or attached to a wall, a user may be occluded by others in the case of multi-user tabletop systems. In addition, a single camera cannot capture the faces of all the users because they sit/stand around a table and thus the face directions of the users differ. Therefore, multiple cameras should be installed.

To achieve unconstrained person identification for tabletop users, we focus on a ceiling-mounted depth camera that overlooks a table, which has already been used in many existing tabletop studies [6], [7] related to recognizing assembly work in a factory, CSCW for tabletop displays, interactive dining tables, recogniz-

[1] Corporate Manufacturing Engineering Center, Toshiba Corporation, Yokohama, Kanagawa 235–0017, Japan
[2] Graduate School of Information Science and Technology, Osaka University, Suita, Osaka 565–0871, Japan
[a) yasuo.namioka@toshiba.co.jp
[b) akira.masuda@ist.osaka-u.ac.jp
[c) maekawa@ist.osaka-u.ac.jp

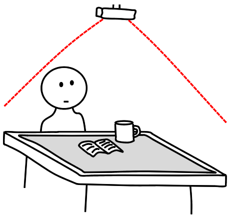


Fig. 1 Our assumed environment where a tabletop user is observed by a ceiling-mounted depth camera.

ing food preparation activities, etc. By using the ceiling depth camera, we can cope with the above problems as follows.

- Because the camera is ceiling-mounted, a user is not occluded by other users.
- One ceiling-mounted camera can capture multiple users simultaneously, thus reducing the installation cost of the system.
- Electrical power can be supplied to the camera from, e.g., a power source of a ceiling light above the table. Moreover, because several tabletop display systems employ a ceiling-mounted projector, electrical power can be easily supplied to the depth camera from the power source of the projector.
- Since the camera is ceiling-mounted, it is unobtrusive, unlike a table-mounted sensor, and does not interfere with the user's activity on the table.

In this paper, we propose the idea of making use of a user's soft biometrics that can be captured from the ceiling, such as the shape of the user's head and the shoulder width, to identify the user using a depth camera installed above the table. Here, soft biometrics is physical or behavioral features that do not identify individuals unlike hard biometrics that uniquely identifies individuals over time such as fingerprint and palm vein. A feature of our identification method is that robust person identification is achieved by combining several soft biometrics that can be captured from the ceiling within the framework of multiview learning. Multiview learning allows us to deal effectively with data consisting of features from multiple sources with different data distributions, i.e., multiple soft biometrics in this case.

In this paper, we assume that a table (or tabletop display system) is installed in an environment of interest, such as a residence, office, factory, or laboratory, and a depth camera is installed above the table. When a user performs a certain activity on the table, e.g., eating at a dining table or browsing web pages using a tabletop display, our method automatically extracts camera biometric features to help identify the user. Specifically, this study focuses on the following features that can be extracted from the ceiling.

- *Skeleton information*: Using the depth camera, we detect the user's joints, such as shoulders and elbows, that can be observed from the ceiling, and then obtain skeleton information of the user from the detected joints. The distance between adjacent joints corresponding to, for example, the shoulder width and the length of the upper arm, can be useful features for person identification because they indicate the user's body height.
- *Shape of body part*: We extract the shape of a body part, such as the head shape, and use it as one of the soft biometrics. For example, the head shape, which is easily captured by a ceiling-mounted camera, varies greatly among users.
- *Dimension of body part*: The dimension of a body part, such as the shoulder indicates the user's physical size and can be a distinguishing feature. For example, the dimension of a fat user's body part and that of a thin user differ greatly.
- *Habitual posture*: We assume that the user performs a certain activity at the table, e.g., eating. When the user performs the activity, the user may adopt a unique habitual posture. When a user eats meals or browses web pages on a table, for example, the user's hunched posture may be a unique feature.

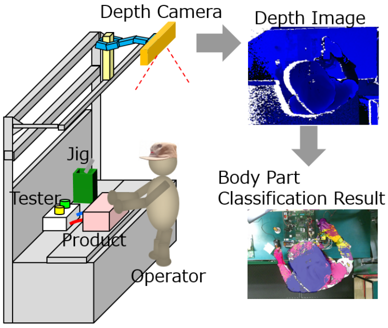


Fig. 2 Application for factory.

Note that each of the above soft biometrics has several disadvantages. For example, the shape of the head slightly changes depending on hairstyle. Therefore, this study employs a multiview-learning-based unified framework for supplementing the disadvantages of the soft biometrics by combining them. Because data distributions of the above soft biometric features are different, we prepare kernels for each feature and combine them based on multiple kernel learning (MKL). We investigate the effect of the combination in the evaluation section.

The identification system with a ceiling depth camera can be used in various environments including residence, office, factory, and laboratory. In the domestic environments, our method can be used to provide/collect health and diet-related information when the system is installed to a dining table. Those information can also be provided/collected by the identification system in office environments installed to a shared table where workers eat lunch. Providing personalized information such as news or announcement is another good application for office environments where tabletop displays are installed. While our experiment in this paper mainly focuses on domestic applications, we introduce applications of our identification method other than domestic applications. Identifying assembly workers in a factory is one promising application since sensor-based monitoring and assistance of assembly work have been actively studied [11], [12]. Many factories have limited spaces to install sensing devices such as Kinect, because many tools and facilities for operation process such as jigs and testers are placed around operators (workers). In this environment, a ceiling-mounted depth camera permits us to capture clear depth images without some occlusion as shown in **Fig. 2**. Therefore a ceiling-mounted camera is an effective solution of the problem. Since the workers do not always work at the same place in many factories, unobtrusive identification methods are required. In addition, monitoring and evaluating work in kitchens in restaurants is another good application because the workers frequently bustle everywhere in the kitchen. Note that, while identification for factories and restaurants are good potential applications, our experimental evaluation mainly focuses on domestic and office applications.

The contributions of this study are: (1) we propose a new unconstrained person identification method for tabletops using a ceiling-mounted depth camera by combining several soft biometric features captured from the ceiling based on multiview learning employing MKL, which can deal effectively with data consisting of features from different sources, (2) we evaluate the method by using real sensor data obtained from 19 participants who performed two different activities, and (3) we confirmed the effectiveness of our method assuming office and domestic environments, showing feasibility of personalized tabletop services such as personalized life logging and news recommender services in these environments.

To the best of our knowledge, this is the first study that investigates the feasibility of person identification for tabletop users by a ceiling-mounted depth camera.

With our identification method, we can achieve such applications as personalized food logging, monitoring and recording assembly work in a line production system, a personalized information service for tabletop displays, and multi-user collaboration for tabletop systems.

2. Related Work

2.1 Sensing Daily and Working Environments

In view of the recent progress of sensing technologies, in many ubicomp studies attempts have been made to sense and recognize various kinds of daily activities in order to support, augment, and record them. In these studies, such sensors as body-worn inertial sensors, RGB-D cameras, and RFID tags attached to everyday objects have been used [13], [14]. In addition, several studies have attempted to capture daily activities performed on a table by using a camera overlooking the table. For example, eating, meal-preparation, medical, and workshop activities have been captured by a camera [5], [15], [16]. In order to provide personalized services in such settings, person identification using a ceiling-mounted sensor is required.

2.2 Device-free Person Identification with Hard Biometrics

While there are some person identification techniques using tags or sensors possessed by a user, such as RFID-based authentication and gesture-based identification, these approaches require the user to always carry a tag or inertial sensor [8], [17]. To reduce the burden on the user, device-free person identification using hard biometrics has been studied and developed. Fingerprint-based identification is the most common approach and used in many identification systems [18]. In addition, other physical attibutes such as hand vein [19] and iris [20] are also applied for identification. While these approaches can uniquely identify users, these approaches require messy actions by users, e.g., placing fingers on a scanner.

2.3 Device-free Person Identification with Soft Biometrics

Because it is difficult to obtain hard biometrics in natural daily life settings, person identification using soft biometrics, which are easier person identification based on soft biometrics, has been actively studied in the ubicomp and HCI research communities. In Ref. [21], for example, a camera embedded in a table captures the shape (contour) of the hand when a user places his hand flat on the tabletop, and the shape is used to identify the user. While this method is device-free, a user is required to perform a specific action. In addition, several studies employ gait information captured by a camera [22]. However, it is difficult to capture gait sensor data in tabletop settings.

While face- and skeleton-based identification is unconstrained [9], [10], these approaches have issues related to occlusion and installation in tabletop settings as mentioned in the introductory section. Also, the face-based identification does not work well when a ceiling-mounted camera is used. In Ref. [23], the authors employ table-edge-mounted RGB-D cameras to capture a user's shoes, and then identify the user by matching the camera image with known shoe images in a database. In Ref. [24], the authors capture a user's sole pattern by using an instrumented floor. In contrast, our study extracts soft biometrics from a ceiling and combines them by using a state-of-the-art statistical approach. As mentioned in the introductory section, person identification using a ceiling-mounted sensor has various advantages in tabletop settings.

Similar to our approach, the authors of Ref. [25] try to capture soft biometrics from a ceiling. A user's height captured by ultrasonic distance sensors mounted above doorways is used for person identification. The authors of Ref. [26] also use a depth camera mounted above doorways to capture the height and silhouette of the body for person identification. In contrast, we capture soft biometrics in tabletop settings and combine several biometric features using multiview learning.

This study employs soft biometrics for person identification for tabletop systems. Here automated soft biometric extraction has a number of applications [27] such as image-tagging and video indexing for photo or video album management, which enables person search with queries specifying physical traits, human computer interaction where personalized avatars can be automatically designed based on physical traits, and health monitoring for early diagnosis of illness, sickness prevention and health maintenance. Such traits include body weight, body mass index, skin abnormalities, and wrinkles.

3. Method

3.1 Overview

Our method assumes that a user with an unknown ID (hereinafter called unknown user) is performing an activity or work on a table as shown in Fig. 1 and the user is identified by using a time series of depth images obtained from a depth camera above the table. When the depth camera detects the user, the camera starts recording depth images and our method uses N_{test} depth images from the start of the recording. Specifically, we estimate a user ID by using each image and then the final result is determined based on the results of the N_{test} images using the majority vote. We also assume that training data for a person identification model are collected under the same setting, i.e., training data can

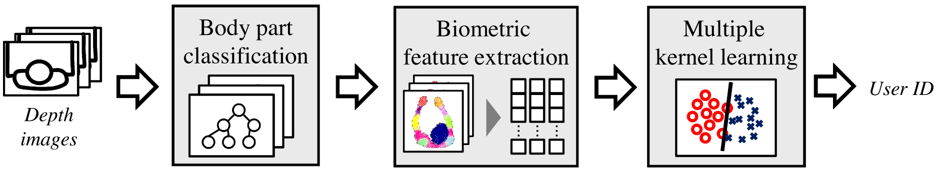


Fig. 3 Overview of process for identifying user by using time-series depth images.

be acquired during daily life. Training data are prepared for each user registered in the system. That is, we assume closed environments such as an office, laboratory, and home where training data for every user are prepared.

Figure 3 shows the procedures of our method. Our method first recognizes body parts such as the head and the right elbow, i.e., body part classification, by using the random forest algorithm [28]. Then our method extracts biometric features of the recognized data such as the distance between joints and the shape of a body part as mentioned in the introductory section. After that, our method identifies the user by using the extracted features based on the multiple kernel learning.

3.2 Soft Biometrics Used in This Study

Before explaining the detailed procedures of our method, we explain rationales of our approach, i.e., what kinds of soft biometrics are used and reasons why we focus on them. As mentioned in the introductory section, this study employs 1) skeleton information, 2) shape of body part, 3) dimension of body part, and 4) habitual posture.

3.2.1 Skeleton Information

The skeleton information has usually been used for person identification. Specifically, the height and the length of limbs, which can be obtained by a depth camera capturing a person from in front of the person using Microsoft Kinect API, are reported to be useful [29]. Since it is difficult to extract this information from the ceiling, we focus on the shoulder and arms, which can be easily captured by the ceiling depth camera. Several anthropometric studies included surveys related to the means and standard deviations for anthropometric parameters such as the shoulder width [30], [31]. For example, Ref. [31] reported that the mean and standard deviation of the shoulder width for 43 adult subjects are 45.90 cm and 3.78 cm, respectively. Also, Ref. [30] reported that the mean and standard deviation of the upper arm length for 4,348 adult males in the U.S. are 39.4 cm and 3.96 cm, respectively. From the means and standard deviations, the anthropometric statistics ($2\sigma/\mu$) are computed, which indicate the variability of the measurements. Because the statistics for the shoulder length and the upper arm length are 0.165 and 0.201, respectively, and larger than the statistics for the height (0.137), these body lengths will be useful soft biometrics for our purpose.

3.2.2 Shape of Body Part

The head is the best body part that the ceiling-mounted camera can capture without occlusion. In Ref. [31], the anthropometric statistics for the head width and depth are reportedly 0.131 and 0.129. (The head depth means the distance between the forehead and the back of the head.) Therefore, this information can serve as well as the height for person identification. However, calculating the head width and depth is difficult because which direction a person faces is unknown and difficult to estimate by using the depth camera. Our idea is to directly compare the shape of the head, i.e., point clouds, of an unknown user and the shape stored in a database using registration techniques for finding a transformation that aligns one point cloud to another.

3.2.3 Dimension of Body Part

Skeleton information provides clues to distinguish between a tall person and a short person. Here we incorporate soft biometrics that are useful to distinguish between a fat person and a thin person because, in Ref. [30], the anthropometric statistics related to the body sizes are reportedly large, e.g., 0.700 for the weight and 0.560 for the waist circumference. In order to capture soft biometrics related to the body sizes using the ceiling-mounted camera, we focus on the dimension of an upper body part such as the bust because it strongly relates to the waist circumference.

3.2.4 Habitual Posture

Several vision-based person identification studies use silhouette [32] to capture the posture of the user. When the user performs a certain activity at the table such as eating, the user may adopt a unique habitual posture, e.g., hunched posture. To capture such characteristics using the ceiling-mounted camera, we focus on the postures of the neck and shoulders, which can be easily captured from the ceiling, since our preliminary investigation revealed that the shape of the back well reflects the posture of a person. To the best of our knowledge, there are no person identification studies that harness 3D posture information of the neck and the shoulders. In the evaluation section, we investigate the usefulness of this feature.

We explain our method illustrated in Fig. 3 in detail. We also explain our proposed approach for extracting sensor data features that well reflect the above soft biometrics.

3.3 Body Part Classification

To extract soft biometric features such as length and shape of body parts, we first recognize body parts using a depth image. We classify each pixel of a depth image into a body-part class (or background class) as depicted in **Fig. 4**. Because Microsoft Kinect API provides skeleton information only when a depth sensor captures a person from in front of him/her, we should implement the body-part classification model by ourselves. While our method is mainly based on the approach used in Kinect API [33], the method should be rotation invariant as regards the user's direction of sitting because the user is free to choose whichever side of the table he/she prefers.

The procedures of the body part classification are simple. We first extract features from each pixel and construct a feature vector concatenating the features. We then classify the vector into an appropriate class. In this study, the vector is classified into

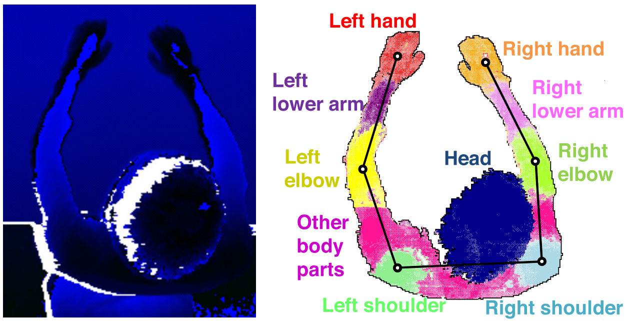


Fig. 4 Example of depth image captured from a ceiling (left), and example of body part classification results (right). The right-hand figure also shows detected skeleton.

head, right hand, left hand, right elbow, left elbow, right shoulder, left shoulder, right lower arm, left lower arm, other body parts, or background class, i.e., an eleven-class classification problem.

3.3.1 Feature Extraction for Body Part Classification

We classify each pixel of a depth image into a body-part class by using shape information around the pixel. Therefore, we extract features that reflect the shape around the pixel. Based on [33], [34], we compute the difference in depth between the pixel of interest and another pixel around the pixel of interest, and the difference will be one of the features.

3.3.2 Preparing Labeled Depth Data

During the training phase, a visual marker is attached to each body part of a participant, and detected markers using RGB images are used to annotate corresponding depth images. Note that the test phase is undertaken without using visual markers. Here, to construct a rotation-invariant classifier, we randomly rotate depth images obtained during the training phase around the Z axis and extract feature vectors used for training the classifier.

3.3.3 Classification with Random Forests

To achieve robust and fast classification, this study uses the random forest algorithm [28]. In the random forest algorithm, a set of features and a set of training instances are randomly selected, and then a decision tree is trained by using the features and instances. This procedure is iterated T times and thus T randomized trees are constructed. Prediction for a test feature vector x' can be made by combining the predictions from all the individual trees. In this study, the probability with which x' belongs to class C_n is simply computed as

$$p(C_n|x') = \frac{n(x', C_n)}{T},$$

where $n(x', C_n)$ is the number of trees that classify x' into C_n.

3.3.4 Skeleton Detection

We first cluster pixels belonging to the head class, i.e., $p(C_{head}|x') > th_c$, by using the x-means algorithm [35]. The largest cluster is regarded as a head cluster. Then, right shoulder, left shoulder, and other body part clusters (See Fig. 4.) are associated with the head cluster. After that, a right (or left) elbow cluster adjacent to an other body part cluster is associated with the other body part cluster. Similar to the above procedures, right lower arm, left lower arm, right hand, and left hand clusters are associated with their adjacent clusters.

Then we compute the joint coordinates of the right shoulder, left shoulder, right elbow, left elbow, right hand, and left hand clusters by simply averaging 3D coordinates of the cluster members. The right portion of Fig. 4 shows the detected joints. Finally, we connect joints as shown in Fig. 4.

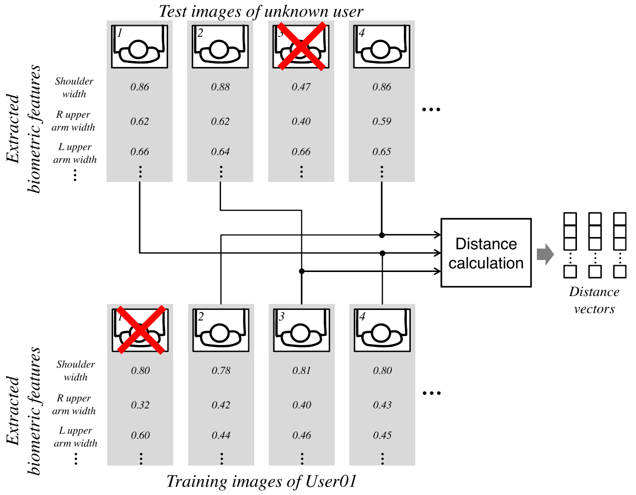


Fig. 5 Comparing sequences of soft biometric features of two users to compute the distance between feature values. Depth images including outlying biometric features are removed.

3.4 Biometric Feature Extraction and Distance Computation

We extract soft biometric features from the detected joints and body part clusters. As shown in the upper portion of **Fig. 5**, we have a series of test images of an unknown user. We extract features such as the shoulder width from each test image. Here, there may exist outlying feature values due to, for example, skeleton detection errors. So, we detect outlying feature values by comparing with feature values of the other test images, and we then remove (ignore) depth images having outlying feature values. In the upper portion of Fig. 5, because the 3rd test image has outlying feature values, it is removed. We also apply the same procedures to training images of each user registered in the system in advance as shown in the lower portion of Fig. 5. (The first image is removed.)

When we judge whether or not an unknown user of the series of test images is identical to a user of a series of training images (for example *User01* in Fig. 5), we compare biometric features of each test image with those of a randomly selected training image of *User01*. For example, the first image of the test images is compared to the fourth image of *User01* in Fig. 5. This is because comparing a test image with each training image is computationally expensive. When we compare the features, we compute the distance between a test feature and its corresponding training feature. For example, when the shoulder width of the 1st test image is 0.86 meters and that of the 4th training image of *User01* is 0.80 meters as shown in Fig. 5, the distance is $|0.86 - 0.80| = 0.06$ meters. We compute the distance for each feature of a pair of a test image and a training image, and construct a distance vector concatenating the distances as shown in the right portion of Fig. 5. The distance vectors are used to judge whether or not the unknown user is identical to *User01*.

We explain biometric features extracted from each depth image and also explain how to compute the distance between two bio-

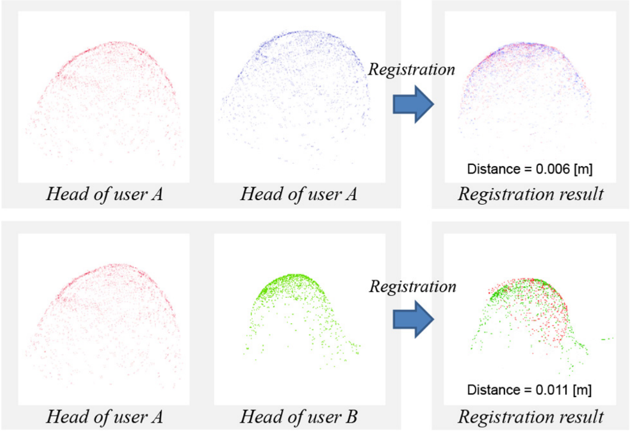


Fig. 6 Lateral view of head point clouds of two participants and registration results of ICP algorithm using them.

metric feature values as follows. Furthermore, we explain how to find outlying feature values in the series.

3.4.1 Skeleton Information

The length of a body part is one of the indicators of the user's body size. From the detected joints, we compute the shoulder width and the lengths of the right lower arm, left lower arm, right upper arm, and left upper arm, which can be observed from the ceiling. When we compute the distance between the length from a test image and that from a training image, we simply use the absolute difference as the distance. Note that, when the hands are occluded, for example, we assume the lengths of the lower arms as missing values.

Due to the skeleton detection errors, the computed length seldom has large errors. So, we find outlying lengths extracted from the series of the test images (or training images) and discard depth images including the outliers that are detected according to the following procedure. In this study, we assume each computed length value as a data point and we employ the mean shift [36]. With the mean shift, we find a small window that includes the maximum number of data points, and data points outside the window become outliers.

3.4.2 Shape of Body Part

We compare the shape of a detected body part of the unknown user and that of a user registered in the system. In this study, we employ the shape of the head because the ceiling camera can easily capture it and the head shape varies depending on users. Specifically, we compute the distance between a point cloud corresponding to the unknown user's head and that corresponding to the registered user's head. In this study, we use the iterative closest point (ICP) algorithm [37] to compute the distance between two shapes (point clouds). **Figure 6** shows examples of the head point cloud registration. The upper portion of the figure shows the registration result of two head point clouds obtained from the same participant. Meanwhile, the lower portion of the figure shows the registration result of head point clouds from two different participants. Because their shapes are different, the ICP algorithm could not find a good transformation.

Note that, due to the limitation of the depth sensing, a depth image sometimes contains missing data (white pixels in the left portion of Fig. 4). The shape of the head cluster in Fig. 4 is irregular and using this shape may degrade the person identification performance. Therefore, we find the outlying head shapes in the series of the depth images and discard them. In this study, we employ agglomerative hierarchical clustering [38] to find outlying head shapes where each head shape is assumed to be a data point in the clustering. The agglomerative hierarchical clustering is a bottom-up approach where each data point starts in its own cluster and a pair of the closest clusters is iteratively merged as one cluster. In the agglomerative hierarchical clustering, we use the single linkage criterion: the distance between two clusters is the minimum distance between any single data point in the first cluster and any single data point in the second cluster. In this study, to find outliers, we merge clusters until we cannot find a pair of clusters having the distance smaller than a threshold. We assume that data points that are not included in the largest cluster are outliers.

Here, the ICP algorithm is computationally expensive. In our current implementation, it takes about 0.2 seconds to compute the distance between two head point clouds. Therefore, we use only the first N_{icp} depth images (head point clouds) of the unknown user to compute the distances. When we want the distance between, for example, the N_{icp} + 1th image of the unknown user and an image of *User01*, we simply re-use the randomly selected distance that has already been computed using images of the unknown user and *User01*.

3.4.3 Dimension of Body Part

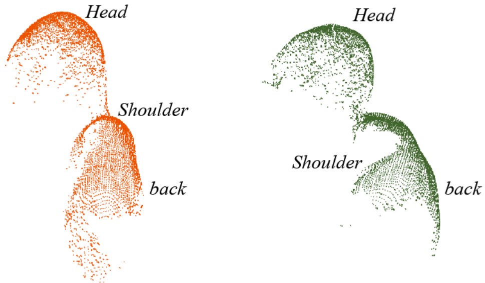
The dimension of a body part is also one of the indicators showing the user's body size. We compute the dimension of a body part in the 3D space by using its corresponding point cloud. In this study, we compute the dimensions of the right shoulder, left shoulder, and bust (sum of right shoulder, left shoulder, and other body part clusters). To compute the dimension, we simply construct a polygon mesh of a body part whose vertices correspond to points included in the point cloud of the body part. Then we compute the dimension of the surface of the constructed polygon mesh by simply summing the dimension of each polygon.

We detect outlying dimension values caused by body part classification errors by using the mean shift in the way described in Section 3.4.1.

3.4.4 Habitual Posture

A posture of a user when performing an activity (e.g., sitting posture) can have a distinguishing feature. In our preliminary investigation, we found that the curvature of the back well reflects the posture of a person. **Figure 7** shows example point clouds of the upper bodies of our participants. The curvature of a person who has a good posture is greatly different from that of a person who has a bad posture. The good posture means a posture with a straight back, and the bad posture means a posture with a curved back as shown in Fig. 7. However, because it is difficult to capture the curvature of the back by using a ceiling-mounted camera in some cases, e.g., when it is occluded by the back of a chair, the present work attempts to capture the posture of the neck and the curvature of the shoulders from the ceiling.

We first explain the posture of the neck. Because the backbone is connected to the neck, we believe that we can capture the curvature of the back from the neck. We assume that the average 3D



(a) User A (b) User B

Fig. 7 Lateral view of upper-body point clouds of two participants. The left portion shows a good posture and the right portion shows a bad posture.

coordinates of the left shoulder joint and the right shoulder joint correspond to the root of the neck. We then assume that the neck corresponds to a line segment connecting the root of the neck and the centroid of the head cluster. We use the angle between the vertical direction and the line segment as one of the soft biometric features, i.e., neck angle. When we compute the distance between the angle from a test image and that from a training image, we simply use the absolute difference as the distance.

Note that the posture of a user can change during an activity. For example, the posture captured when the user is just eating is different from that captured when the user is picking up food located far from him. In this study, we find the most frequent posture (neck angle) that appeared in the activity and we assume that postures other than the most frequent posture are outliers. We find the outlying angle values by using the mean shift in the way described in Section 3.4.1.

We then explain the curvature of the shoulders. As shown in Fig. 7, the shape of the shoulders of a person with a bad posture tends to be curved. In this study, we use a point cloud consisting of data points included in point clouds of both shoulders and points exist between the shoulders as a soft biometric feature. We compute the distance between two point clouds in the way described in Section 3.4.2.

To find the most frequent posture, we also use the agglomerative hierarchical clustering as described in Section 3.4.2. and assume data points belonging to the largest cluster as the frequent postures, i.e., data points (depth images) that do not belong to the largest cluster will be outliers.

3.5 Classification with Multiple Kernel Learning

3.5.1 Overview

We use the distances of biometric features between the unknown user and a user registered in the system to identify the unknown user. **Figure 8** shows the overview. In our method, the extracted biometric features of the unknown user are compared to those of each registered user. (In the upper portion of Fig. 8, the unknown user is compared to *User01*.) As depicted in Fig. 5, distance vectors are constructed concatenating the computed distances, and then the vectors will be inputs of a binary SVM. With this SVM, we can compute the probability with which the unknown user is identified as *User01* by using the margins of the vectors (the signed distances from the SVM hyperplane). We output an ID of a registered user with the largest probability (margin) as the classification result of the distance vector. We compute the final result based on the weighted majority vote of the classification results of the distance vectors. Note that a weight of each vote corresponds to the probability.

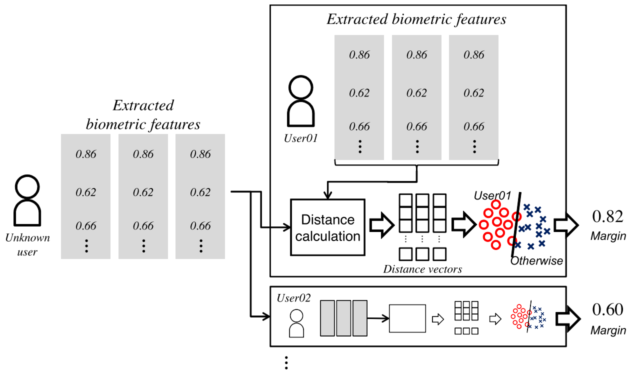


Fig. 8 User identification by using binary SVM prepared for each user registered in the system.

When the SVM of *User01* is trained, the training distance vectors (instances) of the positive class are computed by comparing pairs of *User01*'s biometric features with different time stamps. The training distance vectors of the negative class are computed by comparing *User01*'s biometric features with those of other registered users. Note that the vectors consist of distance values computed from various biometric features such as the skeleton information and shape of a body part. To deal effectively with the vectors consisting of features from different sources, we use a multiple kernel learning (MKL) method in the binary SVM. MKL is one of the multiview learning methods. It employs a linear combination of multiple base kernels and each kernel can describe a different property of the data. So, we prepare a kernel for each different data source, i.e., skeleton information, shape of body part, etc., and combine them.

3.5.2 Multiple Kernels and Person Identification

We use a kernel function to compute the distance between instances to determine a linear decision function in the feature space. Assume that we have N training instances $\{\boldsymbol{x}_i \in \mathcal{X}\}_{i=1}^{N}$. In the kernel-based learning, the decision function, which is used to predict the estimation of unseen test instance $\boldsymbol{x}_\star$, is written as

$$f(\boldsymbol{x}_\star) = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{k}_\star + b,$$

where $\boldsymbol{a}$ and b are the vector of the weights assigned to each training instance and the bias. Also, $\boldsymbol{k}_\star = [k(x_1, \boldsymbol{x}_\star)\dots k(x_N, \boldsymbol{x}_\star)]^{\mathrm{T}}$, where $k(\cdot,\cdot)$ is a kernel function that calculates the distance (similarity) between two instances.

In MKL, we employ the following linear combination of multiple base kernels as the decision function:

$$f(\boldsymbol{x}_\star) = \boldsymbol{a}^{\mathrm{T}}\Big(e_{sk}\boldsymbol{k}_{sk,\star} + e_{sh}\boldsymbol{k}_{sh,\star} + e_d\boldsymbol{k}_{d,\star} + e_h\boldsymbol{k}_{h,\star}\Big) + b, \quad (1)$$

where e_m is the weight of the m-th kernel and $\boldsymbol{k}_{m,\star} = [k_m(\boldsymbol{x}_1, \boldsymbol{x}_\star)\dots k_m(\boldsymbol{x}_N, \boldsymbol{x}_\star)]^{\mathrm{T}}$. Note that $m \in \{sk, sh, d, h\}$, and sk, sh, d and h show skeleton information, shape of a body part, dimension of a body part, and habitual posture, respectively. That is, we prepare kernels for these soft biometric features. (In our actual implementation, we prepare three kernels for each feature: radial basis function, sigmoid, and polynomial.) In each kernel, we configure each kernel's hyperparameters to capture the data distributions of the corresponding biometric feature by using grid search. Refer to Ref. [39] for more detail about the setting of the hyperparameters. Bayesian efficient multiple kernel learning (BEMKL) [40] is used to estimate the parameters in Eq. (1).

4. Evaluation

4.1 Data set

We collected sensor data with Microsoft Kinect v2 mounted about 1.5 meters above a table and sampling rates of 2 Hz. The participants performed two activities while sitting on a chair at the table. Each participant completed three data collection sessions for each activity, and the first 140 seconds of each session was recorded by Kinect. The two activities are 'browsing a web page with a tabletop display' and 'eating meals at a table.' With regard to browsing, we asked the participants to find an interesting news article from a news portal site and read it by using a tabletop display on the table. We selected this activity because we assume an application that provides personalized recommendations, e.g., news search. With regard to eating, we assume a shared table installed in a laboratory or office, and also assume a general personalized information service, e.g., providing personalized news or activity-related information during activities, or a personalized lifelogging application, e.g., recording eating activity for dietary control.

In this research, we investigate person identification when persons use tabletop displays or usual tables in house and office environments. While tabletop displays have not yet been common in houses, we believe that web content browsing will be one of the most common activities on the displays because the activity is now one of the most common activities on tablet computers in houses. In contrast, there are many kinds of activities performed on usual tables, such as studying, reading, taking tea, and so on. However, these activities are too simple and they can make the person identification task easy. Therefore, our evaluation focuses on eating.

19 participants participated in our experiment (i.e., a 19-class classification problem), and the experimental period was about one month. The participants consist of 16 males and 3 females, and their average age is 23.5 years. **Figures 9**, **10**, **11**, and **12** show the distributions of height, weight, BMI, and age of our participants. As shown in the figures, our participants have similar body sizes. Each participant performed three sessions on three different days while wearing his/her clothing during the experimental period. Each participant wore different clothing each day and each participant's clothing was different from that worn by the other participants. In order to confirm the long-term performance of our method, one participant collected data over a period of about 60 days. During the 60 days, the participant collected data on different 8 days. The long-term performance is considered below.

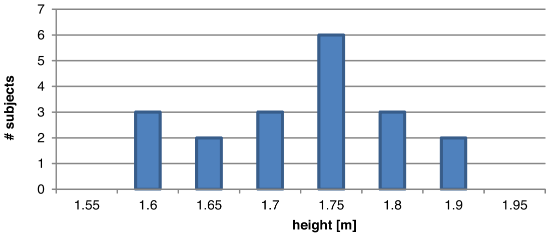


Fig. 9 Distribution of height of subjects.

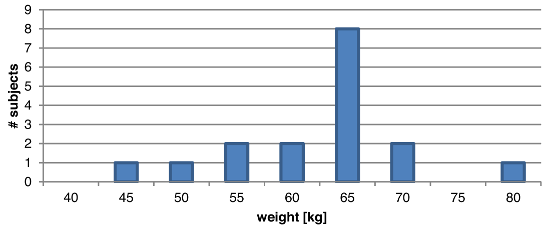


Fig. 10 Distribution of weight of subjects (excluding female subjects).

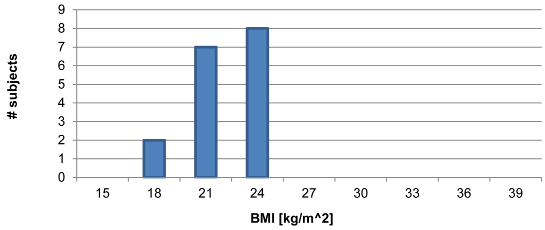


Fig. 11 Distribution of BMI of subjects (excluding female subjects).

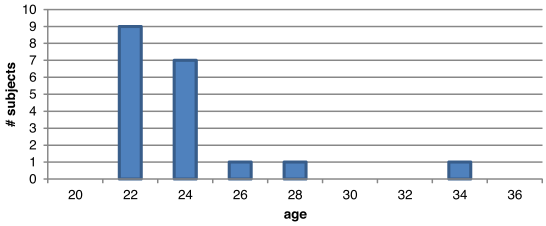


Fig. 12 Distribution of age of subjects.

4.2 Evaluation Methodology

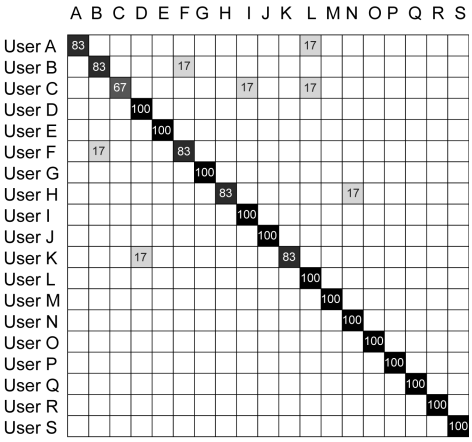
We prepare a person identification model for each activity, independently training a model only on corresponding activity sensor data. The model is trained and tested based on the 'leave-one-session-out' cross-validation approach. That is, we employ testing and training data collected on different days. When evaluating the model, we used the first N_{test} depth images included in the test session. (The sensor sampling rate was about 2 Hz. Note that we use only the first N_{icp} depth images to compute point cloud registration. $N_{test} = N_{icp} = 40$.) The impact of the number of depth images used is considered below.

To investigate the effectiveness of our method, we prepared the following methods.

- *MKL*: This is our proposed method that uses all the biometric features based on MKL.
- *SVM*: This is our proposed method that uses all the biometric features. Note that this method uses multi-class SVM with a linear kernel function instead of MKL.
- *MKL w/o skeleton*: This method does not use the lengths of the body parts obtained from the skeleton as features.
- *MKL w/o shape*: This method does not use the head shape.
- *MKL w/o dimension*: This method does not use the dimensions of the body parts.
- *MKL w/o posture*: This method does not use the habitual posture, i.e., the neck angle and the shoulder shape.

Table 1 Person identification accuracy [%].

	browsing	eating	average
MKL	91.2	96.5	93.9
SVM	91.2	91.2	91.2
MKL w/o skeleton	82.5	91.2	86.8
MKL w/o shape	78.9	87.7	83.3
MKL w/o dimension	91.2	91.2	91.2
MKL w/o posture	82.5	89.5	86.0
MKL only skeleton	59.6	70.2	64.9
MKL only shape	84.2	87.7	86.0
MKL only dimension	50.9	49.1	50.0
MKL only posture	61.4	59.6	60.5



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
User A	83											17							
User B		83				17													
User C			67						17			17							
User D				100															
User E					100														
User F		17				83													
User G							100												
User H								83						17					
User I									100										
User J										100									
User K				17							83								
User L												100							
User M													100						
User N														100					
User O															100				
User P																100			
User Q																	100		
User R																		100	
User S																			100

Fig. 13 Visual confusion matrix of *MKL* (combined result of eating and web browsing).Number in each cell shows the percentages of test instances classified into a corresponding class.

4.3 Results

4.3.1 Identification Performance

Table 1 shows the classification accuracies for the methods. Surprisingly, *MKL* achieved very high accuracy of about 94%. (Random guess ratio is only 5.3% because we have 19 participants.) With regard to eating, our method achieved over 96% accuracy even though our participants had similar body sizes as shown in Fig. 9. Our method achieved almost the same accuracies as the existing state-of-the-art unconstrained identification method for tabletop systems using RGB-D images of shoes [23] (95.8% for 18 participants). Note that the method in Ref. [23] relies on a shoe, which can be occasionally changed. In addition, the method in Ref. [23] requires RGB-D cameras mounted on each edge of a table. In contrast, our method requires only one ceiling-mounted depth camera.

MKL also outperformed *SVM* (about a 2.7% increase). This may be because *MKL* combines multiple kernels tailored for the biometric features. As shown in Table 1, the classification accuracy using the web browsing data was slightly poorer than that using the eating data. This is because some participants used the tabletop display with a slouching posture, and thus their arms were occluded by their bodies.

Figure 13 shows a confusion matrix of *MKL* (combined result of eating and web browsing). Because we have six test sessions for each participant, 17% means wrong identification in only one session (1/6). As shown in the matrix, *User C* is wrongly identified as *User I* and *User L*. The difference of the heights of *User C* and *User I* is only 0.03 meters. Also, the difference of the heights of *User C* and *User L* is 0.01 meters. So, we believe that the wrong estimations were caused by their similar body sizes. In contrast, the difference of the heights of *User A* and *User L* is 0.10 meters. Also, the difference of the heights of *User B* and *User F* is 0.19 meters. We confirmed that the postures of the participants (neck angles) were similar.

As mentioned above, the person identification accuracy of our method is not perfect because our method employs soft biometrics. Therefore, our method is not appropriate for applications that require true security, e.g., applications dealing with computer accounts or banking information. We believe that our method achieved enough accuracy to achieve a general personalized information service, e.g., providing personalized news and tips, and a lifelogging service.

4.3.2 Usefulness of Biometric Features

Table 1 also shows the classification accuracies when we did not use a certain biometric feature. When we did not use the shape feature, the classification accuracy decreased about 10%. So, we conclude that the shape feature greatly contributed to the classification performance. When we did not use the shape feature, our method failed to distinguish between participants with similar heights. (The mean absolute height difference between participants that were not distinguished is 0.058 meters. That for *MKL* is 0.063 meters) Therefore, the feature is useful for distinguishing participants whose body sizes are similar. Note that, because workers wear a hat or helmet in many factories, the head shape information cannot be used for the identification and thus the identification accuracy degrades. However, because works of many factory workers are repetitive and similar sensor data can be observed on different days, identification accuracy can improve due to the contributions of other soft biometrics. Investigating our method in actual factories is our important future work.

As shown in the table, the contributions of the skeleton and posture features were greater than the contribution of the dimension feature. The upper body dimension of a fat participant was greatly different from that of a thin participant, and we assumed that this feature contributes to distinguishing such participants. However, because the upper body was partially occluded depending on the posture of a person, a computed dimension value of the upper body was unstable.

In real environments, the identification performance will greatly drop due to such reasons as wearing a hat, change in hair style, and wearing a heavy down jacket. We can say that Table 1 also indicates the lower bound of the identification accuracy when a certain feature does not completely work. For example, the result of *MKL w/o shape* shows a situation where a user wears a hat or greatly changes his/her hair style. Also, the result of *MKL w/o dimension* shows a situation where a user wears heavy clothes. Moreover, the result of *MKL w/o posture* shows a situation where a user does not perform any activities. While we consider that a user's skeleton information does not greatly change, the result of *MKL w/o skeleton* can indicate a situation where a skeleton detection accuracy significantly drops. When a person holds an object, we confirmed that the estimated length of the arm sometimes has small errors. Therefore, when we cannot use the shape information, e.g., users wear helmets, the identification accuracy can degrade by up to about 10%. In addition, when we cannot

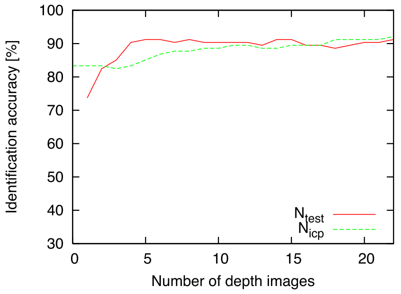


Fig. 14 Transitions of identification accuracies when we changed N_{test} and N_{icp} (average of browsing and eating).

use the skeleton information, e.g., users always hold objects, the identification accuracy can degrade by up to about 8%.

The lower portion of Table 1 shows the classification performance of our method using only the skeleton, head shape, dimension, or habitual posture feature. As shown in the table, it is difficult to identify users by using a single feature. In other words, we could confirm the effectiveness of our multiview learning-based approach, which combines multiple soft biometric features. While using only the head shape feature achieved the best performance, the accuracy was lower than 90%. This may be because the head shape slightly changes depending on hairstyle.

4.3.3 Impact of N_{test}

The person identification is undertaken by using the first N_{test} depth images of test data. **Figure 14** shows the transition of the classification accuracies for *MKL* when we changed N_{test} (N_{test} line). When N_{test} is very small, the classification accuracies are poor. Since our method outputs the final classification result based on the majority vote of the classification results of each test image, small N_{test} is affected by noises. When N_{test} is 4, our method reaches about 90% accuracy. Because the sensor sampling rate was about 2 Hz, $N_{test} = 4$ corresponds to about two seconds.

4.3.4 Impact of N_{icp}

Because it takes about 0.2 seconds to compute the distance between two head point clouds by using the ICP algorithm, this distance computation is the main bottleneck of our identification method. (We should compare a test point cloud to that of each user registered in the system. Also, our method uses the head, left shoulder, right shoulder, and bust point clouds.) Figure 14 shows the transition of the classification accuracies for *MKL* when we changed N_{icp}. ($N_{test} = 40$) As shown in the figure, when N_{icp} is 18, our method reaches about 90% accuracy. Because ICP is used to capture the shape feature, it greatly contributed to the identification accuracy.

4.3.5 Computation Cost

We have measured the computation time of our identification method. When $N_{test} = N_{icp} = 10$, our method outputs the identification result 10.5 seconds after the 10th depth image is captured. ($N_{test} = 10$, $N_{icp} = 1$: 5.6 seconds. $N_{test} = 10$, $N_{icp} = 5$: 7.6 seconds.) It took about 9.3 seconds to compute the point cloud distances with the ICP algorithm. Using the skeleton and dimension features, we plan to reduce the cost related to the ICP algorithm by finding and ignoring registered users whose body sizes are apparently different from the size of a test user.

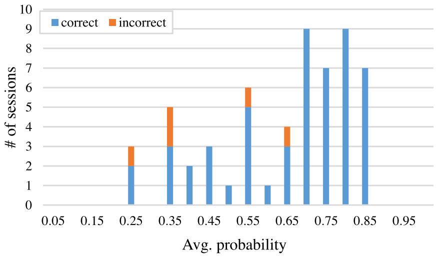


Fig. 15 Histogram of average probability for classified class over N_{test} images (web browsing).

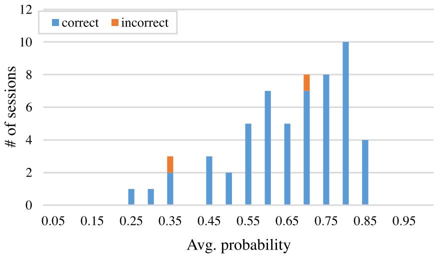


Fig. 16 Histogram of average probability for classified class over N_{test} images (eating).

4.3.6 Dealing with an Unregistered Person

Even though houses and offices are “closed”, these environments can have a visitor. Therefore, here we consider a situation where sensor data from an unregistered person are processed by our identification system. Our method outputs a probability with which a depth image is classified into a registered user class. So, we compute the average probability for the class over N_{test} images, and the class with the largest average is the classified class. When the largest average is smaller than a threshold, we can regard the unknown user as unregistered. **Figures 15** and **16** show histograms of the largest averages for our 114 test sessions. Note that the orange stacked bars indicate incorrectly classified sessions. We can assume that users of the incorrectly classified sessions are unregistered because their classification results do not change regardless of whether or not the users are registered.

With regard to web browsing, when the threshold is 0.65, we can perfectly detect unregistered users in our data. However, about 35% of registered users (sessions) are rejected. With regard to eating, when the threshold is 0.70, we can perfectly detect unregistered users. However, about 58% of registered users are rejected. As explained above, it is difficult to perfectly detect unregistered users without rejecting registered users, and our method is suitable for closed environments.

4.3.7 Long-term Performance

One participant collected data over a period of 63 days. Our method correctly identified the participant during the period. **Figure 17** shows the transition of the probabilities with which the participant is classified into the correct class. During the period, the participant visited a barber on the 13th day. The haircut slightly changes the shape of the head, and the shape feature greatly contributed to the person identification as mentioned above. However, as shown in the figure, the probability on the 14th day was unchanged. Because our method combines several biometric features, it is robust against such noises.

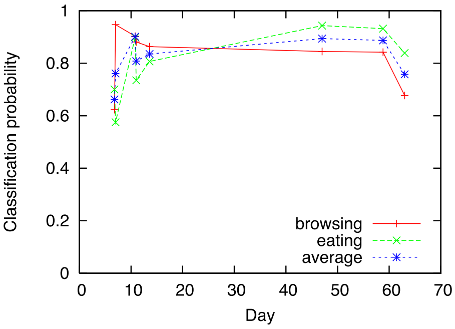


Fig. 17 Transition of the probabilities with which a participant is classified into the correct class. The identification model is trained on the 1st day.

5. Conclusion

We proposed a novel unconstrained person identification method for tabletops. We focus on a user's soft biometrics that can be captured from the ceiling such as the shoulder length, shape of the head, and posture of the back to achieve unconstrained person identification by using a ceiling-mounted depth camera. We achieve robust person identification by combining the soft biometrics within a framework of multiview learning. We evaluated the method by using real sensor data and confirmed the effectiveness of the method. Our method achieved very high accuracy of about 94%. Also, our evaluation revealed that the shape feature greatly contributed to the classification performance. As a part of our future work, we plan to apply our method to real industrial environments such as factories. Since many factory workers wear hats, identifying the workers will be a challenging task. In addition, because our evaluation experiment focuses only on a single activity, i.e., an identification model only for browsing or eating, training our identification model on data including multiple activities is important future work towards practical identification systems.

References

[1] Bianco, S., Ciocca, G. and Napoletano, P.: On the use of MKL for cooking action recognition, *IS&T/SPIE Electronic Imaging*, p.90240G (2014).

[2] Dong, Y., Scisco, J., Wilson, M., Muth, E. and Hoover, A.: Detecting periods of eating during free-living by tracking wrist motion, *IEEE Journal of Biomedical and Health Informatics*, Vol.18, No.4, pp.1253–1260 (2014).

[3] Luff, P., Jirotka, M., Yamashita, N., Kuzuoka, H., Heath, C. and Eden, G.: Embedded interaction: The accomplishment of actions in everyday and video-mediated environments, *ACM Trans. Computer-Human Interaction* (*TOCHI*), Vol.20, No.1, p.6 (2013).

[4] Lukowicz, P., Ward, J.A., Junker, H., Stäger, M., Tröster, G., Atrash, A. and Starner, T.: Recognizing workshop activity using body worn microphones and accelerometers, *Pervasive 2004*, pp.18–32 (2004).

[5] Shi, Y., Huang, Y., Minnen, D., Bobick, A. and Essa, I.: Propagation networks for recognition of partially ordered sequential action, *CVPR 2004*, Vol.2, pp.862–869 (2004).

[6] Lissermann, R., Huber, J., Schmitz, M., Steimle, J. and Mühlhäuser, M.: Permulin: Mixed-focus collaboration on multi-view tabletops, *CHI 2014*, pp.3191–3200 (2014).

[7] Stein, S. and McKenna, S.J.: Combining embedded accelerometers with computer vision for recognizing food preparation activities, *UbiComp 2013*, pp.729–738 (2013).

[8] Roth, V., Schmidt, P. and Güldenring, B.: The IR ring: Authenticating users' touches on a multi-touch display, *UIST 2010*, pp.259–262 (2010).

[9] Abate, A.F., Nappi, M., Riccio, D. and Sabatino, G.: 2D and 3D face recognition: A survey, *Pattern Recogn. Lett.*, Vol.28, No.14, pp.1885–1906 (2007).

[10] Araujo, R.M., Graña, G. and Andersson, V.: Towards skeleton biometric identification using the Microsoft Kinect sensor, *ACM SAC 2013*, pp.21–26 (2013).

[11] Maekawa, T., Nakai, D., Ohara, K. and Namioka, Y.: Toward practical factory activity recognition: Unsupervised understanding of repetitive assembly work in a factory, *UbiComp 2016*, pp.1088–1099 (2016).

[12] Knoch, S., Kerber, F., Pavlov, V. and Ponpathirkoottam, S.: Automatic Capturing and Analysis of Manual Manufacturing Processes with Minimal Setup Effort, *Proc. 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, UbiComp '16*, pp.305–308 (2016).

[13] Maekawa, T. and Watanabe, S.: Unsupervised Activity Recognition with User's Physical Characteristics Data, *International Symposium on Wearable Computers* (*ISWC 2011*), pp.89–96 (2011).

[14] Philipose, M., Fishkin, K.P., Perkowitz, M., Patterson, D.J., Fox, D., Kautz, H. and Hähnel, D.: Inferring activities from interactions with objects, *IEEE Pervasive Computing*, Vol.3, No.4, pp.50–57 (2004).

[15] Wu, J., Osuntogun, A., Choudhury, T., Philipose, M. and Rehg, J.M.: A scalable approach to activity recognition based on object use, *ICCV 2007*, pp.1–8 (2007).

[16] Maekawa, T.: A sensor device for automatic food lifelogging that is embedded in home ceiling light: A preliminary investigation, *PervasiveHealth Workshop on Lifelogging for Pervasive Health* (2013).

[17] Liu, J., Zhong, L., Wickramasuriya, J. and Vasudevan, V.: uWave: Accelerometer-based personalized gesture recognition and its applications, *PerCom 2009*, pp.1–9 (2009).

[18] Sugiura, A. and Koseki, Y.: A user interface using fingerprint recognition: holding commands and data objects on fingers, *UIST 98*, pp.71–79 (1998).

[19] Kumar, A. and Prathyusha, K.V.: Personal Authentication Using Hand Vein Triangulation and Knuckle Shape, *IEEE Trans. Image Processing*, Vol.18, No.9, pp.2127–2136 (2009).

[20] Jain, A.K., Flynn, P.J. and Ross, A.: *Handbook of Biometrics*, Springer (2007).

[21] Schmidt, D., Chong, M.K. and Gellersen, H.: HandsDown: Hand-contour-based user identification for interactive surfaces, *The 6th Nordic Conference on Human-Computer Interaction*, pp.432–441 (2010).

[22] Wang, L., Tan, T., Ning, H. and Hu, W.: Silhouette analysis-based gait recognition for human identification, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.25, No.12, pp.1505–1518 (2003).

[23] Richter, S., Holz, C. and Baudisch, P.: Bootstrapper: Recognizing tabletop users by their shoes, *CHI 2012*, pp.1249–1252 (2012).

[24] Augsten, T., Kaefer, K., Meusel, R., Fetzer, C., Kanitz, D., Stoff, T., Becker, T., Holz, C. and Baudisch, P.: Multitoe: High-precision interaction with back-projected floors based on high-resolution multi-touch input, *UIST 2010*, pp.209–218 (2010).

[25] Srinivasan, V., Stankovic, J. and Whitehouse, K.: Using height sensors for biometric identification in multi-resident homes, *Pervasive2010*, pp.337–354 (2010).

[26] Kouno, D., Shimada, K. and Endo, T.: Person identification using top-view image with depth information, *ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel & Distributed Computing*, pp.140–145 (2012).

[27] Dantcheva, A., Elia, P. and Ross, A.: What else does your biometric data reveal? A survey on soft biometrics, *IEEE Trans. Information Forensics and Security, Institute of Electrical and Electronics Engineers*, Vol.11, No.3, pp.441–467 (2015).

[28] Breiman, L.: Random forests, *Machine learning*, Vol.45, No.1, pp.5–32 (2001).

[29] Preis, J., Kessel, M., Werner, M. and Linnhoff-Popien, C.: Gait recognition with Kinect, *1st International Workshop on Kinect in Pervasive Computing* (2012).

[30] McDowell, M.A. et al.: National Health Statistics Reports, *Anthropometric reference data for children and adults: United States, 2003-2006*, US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics (2008).

[31] Algazi, V.R., Duda, R.O., Thompson, D.M. and Avendano, C.: The cipic hrtf database, *IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, pp.99–102 (2001).

[32] Collins, R.T., Gross, R. and Shi, J.: Silhouette-based human identification from body shape and gait, *5th IEEE International Conference on Automatic Face and Gesture Recognition*, pp.366–371 (2002).

[33] Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M. and Moore, R.: Real-time human pose recognition in parts from single depth images, *Comm. ACM*, Vol.56, No.1, pp.116–124 (2013).

[34] Lepetit, V., Lagger, P. and Fua, P.: Randomized trees for real-time keypoint recognition, *CVPR 2005*, Vol.2, pp.775–781 (2005).

[35] Pelleg, D. and Moore, A.: X-means: Extending k-means with efficient estimation of the number of clusters, *ICML 2000*, Vol.1, pp.727–734 (2000).
[36] Cheng, Y.: Mean shift, mode seeking, and clustering, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.17, No.8, pp.790–799 (1995).
[37] Besl, P.J. and McKay, N.D.: A method for registration of 3-D shapes, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.14, No.2, pp.239–256 (1992).
[38] Johnson, S.C.: Hierarchical clustering schemes, *Psychometrika*, Vol.32, No.3, pp.241–254 (1967).
[39] Hastie, T., Tibshirani, R. and Friedman, J.: Kernel Smoothing Methods, *The Elements of Statistical Learning*, pp.191–218, Springer (2009).
[40] Gönen, M.: Bayesian Efficient Multiple Kernel Learning, *ICML 2012* (2012).

Yasuo Namioka is a chief research scientist at Corporate Manufacturing Engineering Center, Toshiba Corporation, Japan. His research interests include factory data management and analysis for production and quality control. Namioka has a Ph.D. in Engineering from Osaka University. He is a member of IPSJ, DBSJ, JSAI, IEICE, and IEEE.

Akira Masuda was a student at Graduate School of Information Science and Technology, Osaka University, Japan. He is now working for Yahoo Japan. His research interests include sensor data mining and activity recognition.

Takuya Maekawa is an associate professor at Osaka University, Japan. His research interests include ubiquitous and mobile sensing. Maekawa has a Ph.D. in Information Science and Technology from Osaka University.