A person identification method using facial, clothing and time features

Kazuaki Komatsu, Kazutaka Shimada, Tsutomu Endo

Department of Artificial Intelligence, Kyushu Institute of Technology
680-4 Kawazu Iizuka Fukuoka 820-8502 JAPAN
{k_komatsu, shimada, endo}pluto.ai.kyutech.ac.jp

Abstract. In this paper, we propose a method for personal identification using facial features and context information. The method can overcome a problem of partially occluded images. In the proposal method, facial similarity is calculated by using the CLAFIC method from a face image and the parts in the face, such as the image of eyes. In addition, we focus on clothes of each target person and time information as the context. We apply these features into a scoring method and a machine learning method. In the experiments, we confirmed that proposal method is better than the method using only the face image similarity and the heuristic method based on a scoring.

Keywords: Personal Identification, Context, Facial Feature, Clothes Feature, Time Features.

1. Introduction

Person identification is one of the most important tasks in computer vision. One approach to identify a person using a computer is based on analysis of captured images from a camera. Many researchers have studied image based person identification methods. The most famous approach is to use face information. The face based methods are classified into two approaches: methods based on geometrical features and methods based on an appearance model. As a method based on geometrical features, Kanade [3] has proposed a method using face feature points, such as nose, eyes and so on. On the other hand, the CLAFIC method [10] and the EigenFace method [8] are the most famous approaches of appearance models. These methods need the face image for the identification process. However, the face images are not always captured correctly from a camera. For example, there is a problem with occluded images, such as face images with sunglasses.

To solve this problem, we need to apply other information to the identification process for the improvement of the accuracy. One approach is to utilize context information that each person possesses. Gallagher and Chen [2] have reported the effectiveness of context information in a people recognition system. They applied several features, such as clothing and gender, to the system in the case that it is difficult even for humans to determine individuals, such as brother and sister.

In this paper, we describe a person identification method for occluded images. The method employs three types of features. The first feature set is based on facial information. As mentioned above, facial information is one of the most effective features for the identification process. The second feature set is based on clothing information. Clothing information contains characteristics of individuals. Clothing of each person is one of the most effective features for the identification process because clothing denotes individual personality and does not usually change in a day. Also, we have a natural tendency to wear the same clothing periodically. The third feature set is based on time information. Here we assume that there is a recognizable pattern for human behavior. For example, it is arrival time to the offices or labs. It usually becomes habit-forming. We apply the feature sets to the identification process. For the identification process, we compare two approaches; a scoring based method and a machine learning method.

2. Proposed system

In this section, we explain our person identification method. Figure 1 shows the outline of the method. The target images are frontal shots with a face and a clothing area. We capture images of persons on a chair by a camera on a desk. It combined three types of different features; facial features, clothing and time features.

2.1. Facial feature

2.1.1. Facial parts detection

In this process, we need to detect a face in an image first. For the face detection, we apply the method proposed by [9] and [5]. In addition, we utilize the method proposed by [6] for improvement of the accuracy and the processing speed. By using these method, we extract a face area and face parts from each input. The process is as follows:

1 convert an image into gray scale and detect the face area.

2 apply restrictions for the face part detectors.

- eye restriction: eyes exist in upper half of the face

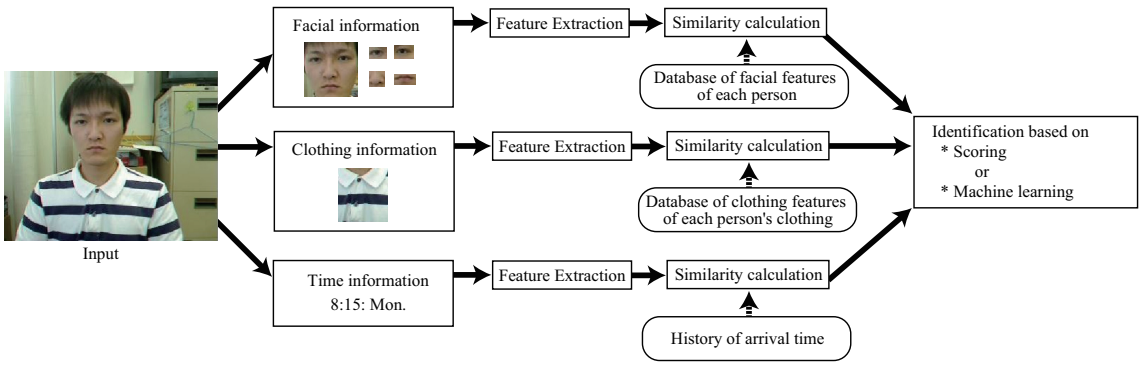

Fig. 1. The outline of the proposed method.

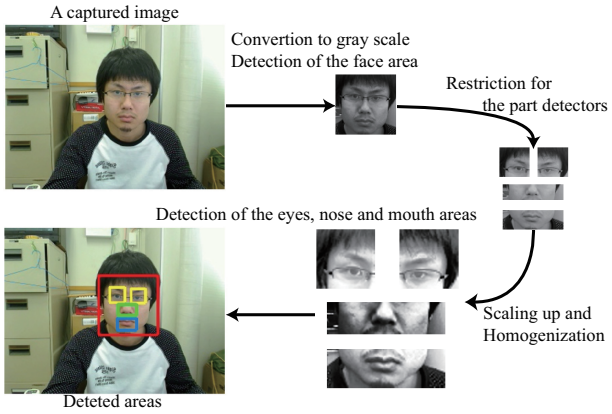

Fig. 2. The outline of the face part detection.

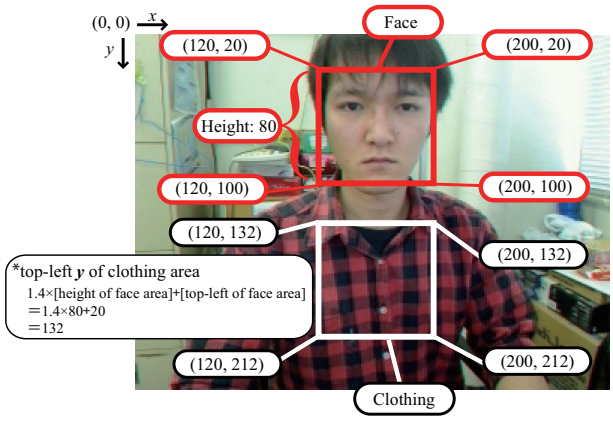

Fig. 3. An example of clothing area detection.

- nose restriction: nose exists in upper side of the face and below the eyes
- mouth restriction: mouth exists below the nose.

If the eyes or nose are not detected, we use the middle area in the three-way split of the face area for nose detection and the bottom area for mouth detection.

3 enlarge each area 4 times and homogenize the areas.

4 extract each part from the areas by using each part detector.

Figure 2 shows the outline of the process.

2.1.2. Facial similarity

We use feature's values from the detected face parts. We compute similarities of the face and face parts by using the CLAFIC method [10]. We generate the correlation matrices for each facial feature, and compute the eigenvector as subspace.

In the identification process, we project features of a target person into the subspace, and compute the sizes of each projection vector. The sizes of the projection vector denote similarities. The facial similarity (S_f) is computed as follows:

$$S_f = S_{face} + S_{lefteye} + S_{righteye} + S_{nose} + S_{mouth} \quad (1)$$

where S_{face}, $S_{lefteye}$, $S_{righteye}$, S_{nose} and S_{mouth} are the similarities from the face, left eye, right eye, nose and mouth parts, respectively.

2.2. Clothing feature

2.2.1. Clothing area detection

For the feature extraction from clothing, our method needs to detect a clothing area in an image. The clothing area detection is based on an assumption that a clothing area exists below the face area detected in Section 2.1. Figure 3 shows an example of the clothing area detection process.

The clothing area is determined on the basis of the size and location of the face. The location of top left of the clothing area (loc_y^c) is computed by

$$loc_y^c = w^c \times height_f + loc_y^f \quad (2)$$

where $height_f$ is the height of the face area. loc_y^f is the top left y of the face area. w^c is a constant factor and

Table 1. Clothing features.

	global	local
color	histogram	mosaic
texture	power spectrum	HLAC

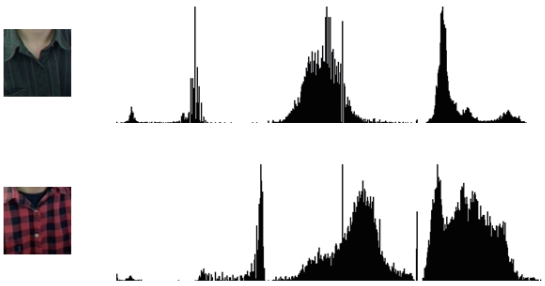

Fig. 4. Color histograms.

the value is 1.4 in this paper. The value is determined experimentally.

2.2.2. Clothing similarity

We apply four clothing features to our method. The four features have different characteristics: global vs. local and color vs. texture. They are classified as Table 1.

- Color histogram
  This is a global feature about color information of clothing. For this feature, we use HSV as the color model. We generate histograms from hue, saturation and brightness values. Figure 4 shows an example of the histograms from clothing.

  We apply Bhattacharyya distance into the similarity calculation. The similarity between two histograms is computed as follows:

$$S = \sum_{u=1}^{m} \sqrt{p_u q_u} \tag{3}$$

  where p and q are normalized histograms. m denotes the number of bins.

  In the identification process, we compute the similarity S_{hist} between histograms of a target person and histograms in the database generated from the training data by using Eq (4).

$$S_{hist} = \sqrt{S_H^2 + S_S^2 + S_V^2} \tag{4}$$

  where S_H, S_S and S_V are the similarity measures of hue, saturation and brightness value, respectively.

- Color mosaic
  This is a local feature about color information of clothing. By using color mosaic, we reduce adverse

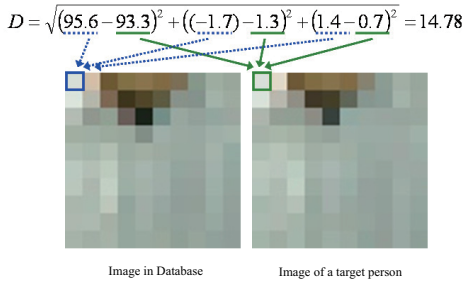

Fig. 5. The similarity calculation between mosaic images.

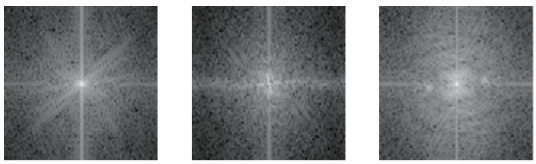

Fig. 6. Power spectrum images.

affects which are caused by wrinkles in the clothing. In this feature, we use a L*a*b* color space.

The similarity measure is based on color difference between images, and computed as follows:

$$D = \sqrt{\Delta L*^2 + \Delta a*^2 + \Delta b*^2} \tag{5}$$

where $\Delta L*$, $\Delta a*$ and $\Delta b*$ are color difference values. Figure 5 shows an example of the similarity calculation. The final similarity S_{mosaic} is computed as follows:

$$S_{mosaic} = \sum_{i=0}^{C} D_i \tag{6}$$

where C is the number of cells in the mosaic image.

- Power spectrum
  This is a global textural feature. The power spectrum is a representation of the magnitude of the various frequency components of an image by Fourier transform. It contains textural information of the clothing.

  We compute a similarity measure from power spectrum images by using the CLAFIC method. The value is $S_{fourier}$.

- Higher Order Local Autocorrelation
  This is a local textural feature. Higher order local autocorrelation is one of the most famous shape features in computer vision [4]. The Nth-order autocorrelation functions with N displacements $a_1, \ldots, a_N$ are computed by

$$x^N(a_1, \ldots, a_N) = \int f(r) f(r+a_1) \ldots f(r+a_N) dr \tag{7}$$

  where r and $f(r)$ are a target pixel and the function. Here N is 2 and the size of a window is 3×3, that is

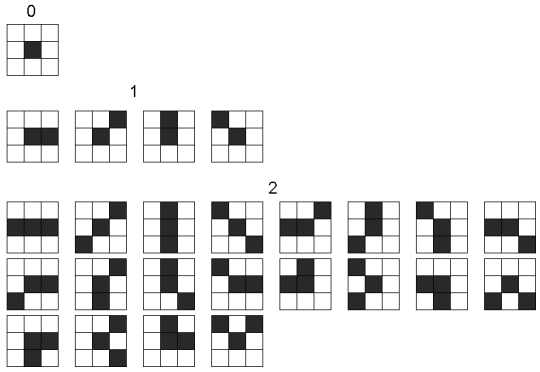

Fig. 7. The mask patterns on HOLA.

the number of mask patterns is 25. Figure 7 shows the mask patterns. The feature value is based on the summation of the mask patterns in each pixel. For the similarity calculation, we apply Mahalanobis distance which is based on correlations between variables. We compute the similarity S_{mask} between the feature of a target person and the feature of each person in the database.

Finally, we integrate the four similarity measures. First, the measures are uniformly scaled because the range of them are different. Then, the similarity of clothing (S_c) is computed as follows:

$$S_c = S_{hist} + S_{mozaic} + S_{fourier} + S_{mask} \tag{8}$$

2.3. Time feature

2.3.1. Time feature extraction

Most people have a routine. Here suppose that we apply our person identification method to an attendance management system as an application. In this situation, there are many recognizable patterns for human behavior. For example, it is arrival time to the offices or labs. Therefore, we add time features to the identification process.

The time features are extracted from metadata in each image file. We extract information on "Date and Time" tag from the Exchangeable image file format (Exif). We introduce time information and a day of the week (e.g. Sunday) of each image.

2.3.2. Time similarity

Time features are computed on the basis of the mode. It is the value that occurs most frequently in training data. Here we quantize time information with hours, e.g., 12:15 is 12. First, we extract the mode ($mode_t$) of each person in the training data. Then, we compute the distance between an input and $mode_t$ of each person in the training data[1]. On the basis of the value, we compute time feature S_{time} as follows:

$$S_{time} = 1 - \frac{|mode_t - time|}{max_t} \tag{9}$$

where $time$ is time information of a target person. max_t is a factor for normalization. The value is the maximum difference[2] between $mode_t$ and each time information in the training data.

The second time feature is based on a day of the week. We also detect the mode about a day of the week ($mode_d$) for each person. Then, we compute the difference between an input and $mode_d$ in the training data. Here we use the minimum distance between days. For example, the distance between Sunday and Monday is 1 and the distance between Tuesday and Saturday is 3. We compute day feature S_{day} as follows:

$$S_{day} = 1 - \frac{|mode_d - day|}{max_d} \tag{10}$$

where day is the day of the week. max_d is a factor for normalization and the value is 3.

Finally, we combine the two similarity measures as follows:

$$S_t = S_{time} + S_{day} \tag{11}$$

2.4. Identification process

In this paper, we compare two methods for the identification process. The 1st method is based on a scoring approach with some parameters. The 2nd method is based on a machine learning approach, AdaBoost with C4.5.

2.4.1. Scoring

The first approach is based on a scoring function. The score of a person i is computed as follows:

$$Score_i = \alpha \times S_f + \beta \times S_c + \gamma \times S_t \tag{12}$$

where S_f, S_c and S_t are computed by Eq. (1) in Section 2.1, Eq. (8) in Section 2.2 and Eq. (11) in Section 2.3 respectively. α, β and γ are weighting parameters for each feature. The values of the parameters are determined experimentally. We set $\alpha = 1.0$, $\beta = 0.5$ and $\gamma = 1.0$. Finally, we select the person which contains the maximum $Score_i$ as the output of the identification process.

2.4.2. Machine learning

The scoring based method contains a problem. It was based on three weighted parameters; α, β and γ. We need to experimentally determine the parameters for the method. They were not always robust values for the identification. We might need to investigate the optimal values again if the dataset is changed. To solve this problem, we apply a machine learning approach to the method.

We construct a classifier based on the facial, clothing and time features and employ the AdaBoost algorithm [1] as the classifier. The AdaBoost algorithm is one of the most famous machine learning techniques. It generates a strong classifier by combining some weak classifiers. We implement the AdaBoost with the open source software

1. Here we treat the distance on time information. For instance, the distance between 1:00 and 23:00 is 2 (not 22).

2. Therefore, it is 12 in this equation.

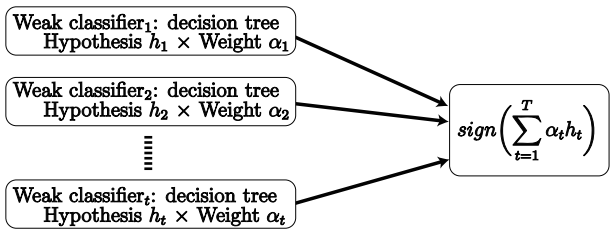

Fig. 8. Adaboost.

Weka[3] and use the C4.5 algorithm [7] as the weak classifiers[4]. C4.5 is also one of the famous machine learning techniques, which is to generate a decision tree. Figure 8 shows the outline of the AdaBoost algorithm. We extend the AdaBoost into multi-label classification with the One-versus-Rest method.

3. Experiment

In this section, we compared the two identification approaches first. In the experiment, we also compared the methods using context with a method using only face features. Then, we discuss the combination of context features.

3.1. Dataset

For the experiment, we simulated (1) faces with sunglasses, (2) faces with a mask and (3) faces with sunglasses and a mask as occluded situation. Figure 9 shows an example of the dataset. In this experiment, we simulated occluded images because all face images in the data set contained all information about a face and the parts. For example, we deleted eye features as hypothetical images about faces with sunglasses.

We captured face images with clothing during 40 days. The number of test subjects was 7 persons. The training data consisted of 875 images (7 persons $\times$ 125 images). The test dataset consisted of 175 images (7 persons $\times$ 25 images). Although the captured dates in the training and test data were different, the clothing in the test data was always contained in the training data. In other words, there was not unknown clothing in the test data.

3.2. Results

Table 2 shows the experimental result with all features, namely facial, clothing and time features. The accuracy was computed by

$$\frac{\text{\# of images identified correctly}}{\text{\# of images}} \times 100.$$

In the table, SG, Mask and SG+Mask denote faces with sunglasses, faces with a mask and faces with sunglasses and a mask, respectively. FaceOnly denotes a method without context features namely clothing and time features. In other words, the method used only nose and mouth features for the situation SG and only right and left eyes features for the situation Mask. For the SG+Mask, the method without context features did not identify any persons because all parts in a face were hidden. The identification process of FaceOnly was based on AdaBoost.

3. http://www.cs.waikato.ac.nz/ml/weka/
4. Actually, it is "J48" in Weka.

Fig. 9. Face and clothing images.

Table 2. The experimental result with all features.

	SG	Mask	SG+Mask	Ave
Scoring	85.1%	**96.6%**	71.4%	84.4%
Adaboost	**96.6%**	92.6%	**97.7 %**	**95.6%**
FaceOnly	69.7%	88.0%	-	-

For the experiment with all features, the machine learning based method outperformed the accuracy in terms of SG and SG+MASK. On the other hand, it generated lower accuracy for Mask as compared with the scoring based method. It was caused by the magnitude of eye features in the dataset. For the Mask situation, the accuracy was high even if the method was based on only eye features (88.0% by FaceOnly in Table 2). In the facial similarity S_f, the scoring method was dominated by the eye features ($S_{lefteye}$ and $S_{righteye}$). The two methods with context information, namely clothing and time features, produced higher accuracies than that without context information (85.1% & 96.6% vs. 69.7% in SG, and 96.6% & 92.6% vs. 88.0% in Mask). Moreover, the method without context was essentially weak for occlusion. In the SG+Mask situation, it could not treat any images. The experimental result shows that context information is robust features for person identification in various situations.

Next we evaluated the combination of the features. Table 3 shows the experimental result with the best features. In the table, F+C denotes the combination of facial features and clothing features. F+C+T denotes the combination of all features.

For the scoring method, time features were poorly-functioning. For the machine learning method based on Adaboost, they decreased the accuracy of the Mask situa-

Table 3. The experimental result with the best features.

	SG	Mask	SG+Mask	Ave
Scoring	91.4% F+C	**98.9%** F+C	81.7% F+C	90.7%
Adaboost	**96.6%** F+C+T	97.1% F+C	**97.7%** F+C+T	**97.1%**

tion. We introduced time and a day of the week information as time features. Only they did not contain enough description for the person identification. We need to consider other time features, such as sojourn time, for the improvement of the accuracy. Besides, the number of test data sets was one of the reasons that time features were not effective. There are numerous variations of arrival time as compared with variations of clothing. To achieve higher accuracy, we need to collect more training data.

The average accuracies of Adaboost outperformed those of the scoring method. The result shows the difficulty of parameter tuning in the scoring method and the robustness of the machine learning based method. On the other hand, placing emphasis on significant features, which are eye features in this experiment, leads to improvement of the accuracy (See the Mask situation). A hybrid approach based on the scoring and Adaboost is one solution for the issue.

4. Discussion and Conclusions

In this paper, we proposed a method for personal identification with context information to overcome a problem of partially occluded images. The method handled three types of features; facial, clothing and time features. The facial features were based on the CLAFIC method. For the clothing features, we utilized four different types of characteristics; global or local and color or texture. The time features consisted of time and a day of the week information.

We compared two methods; a scoring method and a machine learning method based on Adaboost. The average accuracy of the Adaboost outperformed the scoring method (97.1 vs. 90.7 in the best feature set). On the other hand, the scoring method produced higher accuracy than the Adaboost in the Mask situation. Combining the two methods is one future work to achieve higher accuracy. The accuracies of the two methods with context information were dramatically improved as compared with a method without context information. Furthermore, the methods with context could appropriately deal with a problem about the occluded images. These results show the effectiveness of context information for the person identification.

The target occlusion in the experiment was based on sunglasses and a mask on a face. Our method can essentially deal with other occluded images if the images contain a clothing area. The experiment with other situations is one of the important tasks to evaluate the availability of our method.

We focused on identification of a person; in other words, it was a classification task of images. In the person identification, the true rejection rate is often discussed. However, we did not apply any rejection process to our system. The reason why we did not handle rejection was that we had a plan to apply the person identification method to an attendance management system as an application. We think that it does not need strict rejection as compared with security systems. However, the rejection process is essentially important future work, as security systems.

Future work includes (1) applying other clothing an time features, (2) collecting more training data for time features, (3) evaluation with other dataset, e.g., increase of the number of target persons and a dataset in four seasons, and (4) consideration of a rejection process.

References:

[1] Y. Freund and R. E. Schapier. "Experiments with a new boosting algorithm," Proceedings of ICML pp. 148-156, 1996.

[2] A. C. Gallagher and T. Chen. "Using Context to Recognize People in Consumer Images," IPSJ Transactions on Computer Vision and Applications, Vol. 1, pp. 115-126, 2009.

[3] T. Kanade. "Picture processing by computer complex and recognition of human face," Technical report, Kyoto University, Dept. of Information Science, 1973.

[4] T. Kurita, N. Otsu, and T. Sato. "A face recognition method using higher order local autocorrelation and multivariate analysis," Proceedings of 11th IAPR International Conf. on Pattern Recognition, pp. 213-216, 1992.

[5] R. Lienhart, A. Kuranov, and V. Pisarevsky. "Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection," MRL Technical Report, Intel Labs, 2002

[6] D. Motoyoshi, K. Shimada, S. Enokida, T. Ejima and T. Endo. "A Case Study of Speech Activity Detection for an Interactive Robot," Proceedings of 11th meeting on image recognition and understanding (MIRU 2008), pp. 1015-1020, 2008

[7] J. R. Quinlan. "C4.5 Programs for Machine Learning," Morgan Kaufmann Publishers, 1993.

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

[9] P. Viola and M. Jones. "Robust Real-time Object Detection," Second International Workshop on Statistical and Computational Theories of Vision-Modeling, Learning, Computing, and Sampling, pp. 1-25, 2001.

[10] S. Watanabe and N. Pakvasa. "Subspace method in pattern recognition," Proceedings of 1st Int. J. Conf on Pattern Recognition, pp.2–32, 1973.