Proceedings of the ISIM 2000, Oct.4-7, Kyongnam, Korea

Person Identification from Binary Silhouette Image of Full-length Body

K. Hiraoka, S. Morishita, H. Mizoguchi, and T. Mishima
Dept. of Information and Computer Sciences, Saitama University, 255 Shimo-Okubo, Urawa, 338, Japan

Abstract: For human collaborative machines and human interactive systems, function to visually identify person is quite crucial. Conventionally, the visual identification of person is performed via face recognition. And usually identifying programs require ideal conditions on background, brightness, and so on. In the practical situation, these conditions are not so easy and high cost computation is necessary to meet the conditions properly. Hence it is crucially needed to develop some other method which supplements face recognition.

In the present paper, a novel method to visually identify person from binary silhouette image of full-length body is proposed. The experimental result supports the feasibility of the proposed idea.

1 Introduction

Towards the coming aged society, human collaborative machines and/or human interactive systems are expected as a promising area for the intelligent mechanical system of the future. Smart environments, human interactive robots, service robots, artificial pets, and so forth are typical examples. For such systems that treat humans as target, function to visually identify person is quite crucial. If these intelligent mechanical systems could recognize person and behave properly towards him or her, these systems would become more intelligent and useful as human supporting systems. In other words, capability of person identification is a key to add value to these systems. Needless to say, the person identification is also important for security.

Conventionally the visual identification of person is performed via face recognition[1][2]. In this case, gray or color image is used as the input. And usually identifying programs require frontal face images that are properly extracted from scene and does not include background. These face images should have similar size and brightness. In the practical situation, such as the smart environment or the human interactive robot, these conditions are not so easy and high cost computation is necessary to meet the conditions properly.

So far, the authors have studied several methods for face detection and identification. From our experience, we have faced the above problem of the no easy conditions and the high cost computation. And we conclude that some other method of person identification with low computational cost is crucially needed. The method should be pre-processing of the face recognition and augment total performance of the identification system.

Thus, in this paper, the authors propose a novel method to visually identify person from binary silhouette image of full-length body. The proposed method utilizes our online linear discriminant analysis(OLDA)[3][4]. OLDA is developed by the authors as a variant of LDA with adaptability and has been used for the face identification. We intend to combine the proposed method and the face identification to improve total performance of the identification system. To confirm the feasibility and the effectiveness of the proposed method, the authors conduct experiments using 600 silhouette images of three persons. As the result, we find that OLDA can be applied to even the binary silhouette image and performs identification well at over 93% correct identification ratio after learning of 600 images. The experimental result supports the feasibility of the proposed idea.

In the following, section 2 describes our proposed OLDA. In section 3, we discuss application of OLDA to the binary silhouette image of full-length body. Person identification experiments are also explained. Section 4 is the conclusion.

2 Online linear discriminant analysis

In the present study, an online version of linear discriminant analysis is used to identify persons [3][4].

At every time step $t = 1, 2, 3, \cdots$, a new pair $(\boldsymbol{x}(t), c(t))$ is presented, where $\boldsymbol{x}(t)$ is an N-dimensional data vector, $c(t) \in \{1, \cdots, M\}$ is the class of $\boldsymbol{x}(t)$, and M is the number of classes. Based on this pair, auxiliary variables are updated as follows:

$$
\begin{aligned}
t^c(t) &= t^c(t-1) + \delta(c, c(t)), && (1)\\
\bar{\boldsymbol{x}}(t) &= \left(1 - \frac{1}{t}\right)\bar{\boldsymbol{x}}(t-1) + \frac{1}{t}\boldsymbol{x}(t), && (2)\\
\bar{\boldsymbol{x}}^c(t) &= \\
&\begin{cases} \left(1 - \frac{1}{t^c(t)}\right)\bar{\boldsymbol{x}}^c(t-1) + \frac{1}{t^c(t)}\boldsymbol{x}(t) & (c = c(t)), \\ \bar{\boldsymbol{x}}^c(t-1) & (c \neq c(t)), \end{cases} && (3)\\
\boldsymbol{v}^c(t) &= \bar{\boldsymbol{x}}^c(t) - \bar{\boldsymbol{x}}(t), && (4)\\
\boldsymbol{w}(t) &= \boldsymbol{x}(t) - \bar{\boldsymbol{x}}^{c(t)}(t), && (5)\\
B(t) &= \frac{1}{M}\sum_{c=1}^{M} \boldsymbol{v}^c(t)\boldsymbol{v}^c(t)^T, && (6)
\end{aligned}
$$

where $c = 1, \cdots, M$ and $\delta(c, c(t)) = 1\,(c = c(t)),\ 0\,(c \neq c(t))$. Then $N \times L$ discriminant matrix A is updated:

$$
\begin{aligned}
A(t) = A(t-1) + \eta\Big[&B(t)A(t-1) \\
&- \frac{1}{2}B(t)A(t-1)\,A(t-1)^T(\boldsymbol{w}(t)\boldsymbol{w}(t)^T + \epsilon I)A(t-1) \\
&- \frac{1}{2}(\boldsymbol{w}(t)\boldsymbol{w}(t)^T + \epsilon I)A(t-1)\,A(t-1)^T B(t)A(t-1)\Big],
\end{aligned} \quad (7)
$$

where the learning coefficient η and the regularization coefficient ϵ are small positive numbers, and I is the identity matrix. The term $+\epsilon I$ is useful for stabilization of the algorithm [3]. As for tips on efficient calculation, see [4].

In the identification phase, data vector $\boldsymbol{x}$ of unknown class is presented. It is transformed to "feature vector" $\boldsymbol{y} = A(t)^T\boldsymbol{x}$ and compared with $\boldsymbol{y}^c(t) = A(t)^T\bar{\boldsymbol{x}}^c(t)$ $(c = 1, \cdots, M)$.

3 Person identification from silhouette image

We have applied the OLDA method to person identification task from silhouette images. The setting of the experiments is summarized in Table 1.

3.1 Input data

Front-view silhouette images of full-length bodies have been taken for three persons A, B, and C. For each person, 100 images are prepared [1]. Some sample images are shown in Fig. 1. These images are reduced to 10×18 pixels (Fig. 2).

[1] Exactly speaking, 50 of them are front-view and 50 are back-view. They are approximately equivalent since only silhouette is concerned in the present study.

Figure 1: *Some sample images for the experiments. (left: person A, middle: person B, right: person C)*

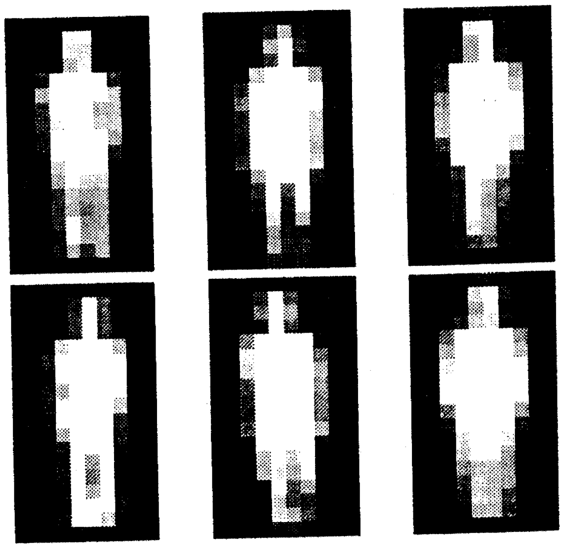

Figure 2: *Reduced images for the experiments. (left: person A, middle: person B, right: person C)*

3.2 Procedures of experiments

Let t be the step of learning. One image is presented at every step.

During $t = 1, \cdots, 200$, only the images of person A are presented (Table 2). In this stage, only

Table 1: *Setting of the experiments*

task	person identification from silhouette images
data vector $\boldsymbol{x}(t)$	silhouette image of full-length body (front view, 256 level gray scale, normalized to $[-1, +1]$)
size of $\boldsymbol{x}(t)$	$N = 10 \times 18 = 180$ (pixels)
number of classes to be identified	$M = 1 \rightarrow 2 \rightarrow 3$ (persons)
number of features	$L = M - 1$ (= number of columns in A)
initial values of elements in A	random values from the uniform distribution on $[-0.001, +0.001]$
learning coefficient	$\eta = 0.05$
regularization coefficient	$\epsilon = 0.001$
procedure of learning	Silhouette images for learning is presented in a random order.
procedure of evaluation	The ratio of the correct identification is evaluated on the same silhouette images as for learning.
number of sample images	100(images per person) $\times$ 3(persons) = 300
number of trials	100 independent trials with different random seeds

Table 2: *Procedures of experiments. During* $t = 1, \cdots, 200$, *only the person A is presented. During* $t = 201, \cdots, 400$, *only the person B is presented and the performance is evaluated on the identification of A and B. During* $t = 401, \cdots, 600$, *only the person C is presented and the performance is evaluated on the identification of A, B and C.*

t	learning	evaluation
$1, \cdots, 200$	A	A
$201, \cdots, 400$	B	A, B
$401, \cdots, 600$	C	A, B, C

the mean vectors $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{x}}^{\mathbf{personA}}$ are updated. At $t = 201$, new class "person B" is added. Accordingly, 180×1 discriminant matrix A and the class mean vector $\bar{\boldsymbol{x}}^{\mathbf{personB}}$ are generated. During $t = 201, \cdots, 400$, only the images of person B are presented. At $t = 401$, new class "person C" is added in the same way. Accordingly, a new column is appended to the discriminant matrix A so that A is enlarged to 180×2. The class mean vector $\bar{\boldsymbol{x}}^{\mathbf{personC}}$ is also generated. During $t = 401, \cdots, 600$, only the images of person C are presented [2].

At every 10 steps, performance of identification is evaluated. In the evaluation, all images of person A ($t = 1, \cdots, 100$), A·B ($t = 201, \cdots, 400$), or A·B·C ($t = 401, \cdots, 600$) are examined and the percentage of correct identification is recorded. Learning is stopped during this evaluation phase. Same images are used for both learning and evaluation.

[2]Since only 100 images are prepared for each person, the same image set is presented two times.

3.3 Experimental results

The result of the simulation is shown in Fig. 3. This result supports the feasibility of the proposed method.

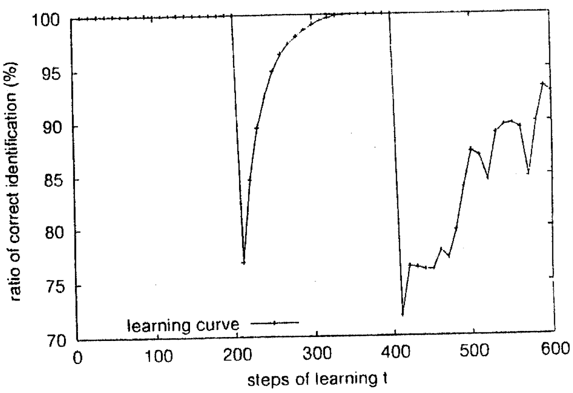

Figure 3: *Mean learning curve for* 100 *independent trials (Horizontal axis: steps of learning. Vertical axis: percentage of correct discrimination). At* $t = 200$ *and* $t = 400$, *new person is added and the performance goes down temporally. And soon, the performance is successfully recovered.*

Fig. 3 is the mean learning curve for 100 independent trials. That is, the same experiment is performed with 100 different random seeds. These random seeds are used to generate the initial value of A. Since OLDA is an iterative algorithm, the performance depends on the initial value.

4 Conclusion

Person identification from silhouette images of full-length bodies is studied in the present paper. Online linear discriminant analysis (OLDA[3]) is applied to this task. And it is experimentally shown that OLDA works feasibly for this task.

A problem on OLDA is setting of the learning coefficient η. If η is too large, the discriminant matrix A can diverge. In the present experiments, such divergence is observed for 10% of 100 trials. An algorithm to prevent it is proposed in [5].

Another point which we must study further is generalization ability and robustness of our method for changes of direction, clothes, and posture. The authors are planning to build more powerful identifiers by use of adaptive background estimation [6] and attention mechanism [7][8].

Acknowledgements

The authors wish to thank Mr. Goto for his help on experiments. This work has been partly supported by CREST of JST (Japan Science and Technology) 279102 and JSPS (Japan Society for the Promotion of Science) 12750203.

References

[1] R, Chellapa, C.L.Wilson, and S. Sirohey, "Human and Machine Recognition of Faces: A Survey", Proc. of the IEEE, Vol.83, No.5, pp.705-741, 1995.

[2] A. Pentland, "Looking at People: Sensing for Ubiquitous and Wearable Computing", IEEE Trans. on PAMI, Vol.22, No.1, pp.107-119, 2000.

[3] K. Hiraoka, et al., "Convergence Analysis of Online Linear Discriminant Analysis", Proceedings of International Joint Conference on Neural Networks (IJCNN), Vol. III, pp. 387–391, 2000.

[4] K. Hiraoka, et al., "Online LDA which can perform both successive learning and incremental learning", Proceedings of The 4th World Multiconference on Systemics, Cybernetics and Informatics (SCI), pp. 96–101, 2000.

[5] S. Morishita, et al., "Study on automatic setting method of learning coefficient in online LDA towards robust convergence", Proceedings of 2000 Information and System Society Conference of IEICE, 2000, *to appear (in Japanese).*

[6] H. Shimai, et. al., "Adaptive background estimation from image sequence by on-line M-estimation and its application to detection of moving objects", Proc. of Infotech Oulu Workshop on Real-Time Image Sequence Analysis, 2000.

[7] K. Hotta, et. al., "Face Matching through Information Theoretical Attention Points and Its Applications to Face Detection and Classification", Proc. of Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp.34-39, 2000.

[8] T. Watanabe, et al., "Autonomous foveating system based on the Pulse-Coupled Neural Network", Proc. of the 1999 International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC'99), Vol. 1, pp. 197–200, 1999.