Research Subject

Modeling Visual Recognition and Online Linear Discriminant Analysis (Yoshizawa Group)

(1) Goal and summary

Aiming development of brain-oriented recognition and decision technologies, our group has studied three subjects.

The first subject is modeling of brain information processing; its goal is understanding of information processing mechanism in the brain with neuron level and neural network level. As a model of robust information processing, which is a major weak point of today's computers, we investigate implementation of the brain.

The second subject is the study on recognition technology. Only from a two-dimensional movie, we can recognize three dimensional positions and motions of a person in it. Though the mechanism of that ability is not known completely, we have a conjecture that a model of human body is built in the brain and it is applied to two-dimensional images so that natural motions can be estimated. According to this strategy, we have tried estimation of three-dimensional motions based on positions of characteristic body points which are detected from two-dimensional movies. Confirmation of feasibility and drawbacks of this method have been investigated experimentally through the present study.

The third subject is the study on discrimination technology. The brain can learn new things and it has adaptability to change of situations during the learning. It can also extract useful components for its decision from high-dimensional input data. In order to develop discrimination technologies which have the above nature, we have constructed online linear discriminant analysis (OLDA) and several other techniques. In addition to theoretical investigation of them, we have also built several vision systems as applications of them, face recognition system, arbitrary pattern recognition system, human walking pitch extraction system based on visual tracking of person's heel, and so on.

Modeling of brain information processing As an analysis of neuron level, spike-timing-dependent synaptic plasticity (STDP) is studied. An additive rule of STDP automatically achieves synaptic competition and activity regulation, where synaptic balance is moderately regulated to control the post synaptic activity. On the other hand, a multiplicative STDP rule can not achieve the synaptic competition nor the synaptic regulation. It is not clear what type of STDP rules can achieve synaptic regulation and how it is related to synaptic competition. Here we clarify the mechanisms of synaptic regulation and propose multiplicative STDP rules which can achieve synaptic regulation. .

As an analysis of neural network level, binocular fusion is studied. When two eyes view radically different patterns, only one pattern which is dominant is perceived at a time. This perceptual dominance is termed binocular rivalry, which may provide a new tool for investigating the mechanism of binocular fusion. Here, we show that this phenomenon is explained by using a network which consists of locally connected neurons with orientation preference. Also, we investigated the property of the connection which is required for realization of the qualitative results shown by the psycho-physical experiments on the binocular rivalry. .

Recognition technology The second subject is the study on recognition technology. Only from a two-dimensional movie, we can recognize three dimensional positions and motions of a person in it. If we develop such a system which can obtain motion information from movies, it must contribute much to realization of intelligent robots which cooperate with human. Though the mechanism of that ability is not known completely, we have a conjecture that a model of human body is built in the brain and it is applied to two-dimensional images so that natural motions can be estimated. According to this strategy, we have tried estimation of three-dimensional motions based on positions of characteristic body points which are detected from two-dimensional movies. Confirmation of feasibility and drawbacks of this method have been investigated experimentally through the present study.

The use of only one video camera is assumed in our study. The process of motion estimation from a movie consists of two stages:

1. Detection of characteristic body points such as joints from the input movie via image processing

2. Estimation of three-dimensional motion and rotation angles of joints based on fitting of the three-dimensional internal model of human body to two-dimensional detected characteristic points.

The former stage, detection of characteristic points via image processing, is generally a hard problem. Since the amount of information from the input movie is large, hard works must be repeated unless we specify which information with what accuracy is necessary for the latter stage. From this point of view, we examine the latter stage, estimation of three-dimensional motion from two-dimensional information, so that we obtain knowledge on required information and accuracy.

As the model of human body, we use a joint structure model which is an extension of stick models. This model consists of 14 characteristic points. Each joint has restriction on admissible joint angles. It is supposed that such a model in three-dimensional space is projected onto two-dimensional image. The procedures of model fitting is performed as follows:

1. The scale of the object is specified based on two projected triangles which correspond to the shoulder and the hip. In this step, three-dimensional rotation degrees of each triangle is limited to four possibilities.

2. Based on the above specified scale, rotation angles of other joints are limited to two possibilities for each joint. In this stage, 2048 possibilities of combination of joint angles are obtained.

3. Inadmissible possibilities are examined and removed according to the admissible range of the angle for each joint.

4. With the assumption of continuous motion, the most smooth motions among remaining possibilities are tracked. Since possibilities are drastically limited in some frames depending on the posture and the camera angles, number of remaining possibilities decreases as time goes on.

5. The most smooth possibility of motion among tracked possibilities is selected and it is reported as the final estimation.

In the present study, we evaluate the influence of additive noise to two-dimensional coordinates of characteristic points to restriction process on possibilities of three-dimensional motion. In this evaluation, we pay attention to the following criteria.

Largest distance between succeeding frames: The sum of three-dimensional distance of each characteristic points between succeeding frames is defined as “distance between frames”. The maximum value of distance between frames through the movie is defined as “largest distance” here. It represents discontinuity of the estimated motion.

Number of remaining possibilities: This criterion is the number of finally remaining possibilities in the last frame of the movie.

Rate of impossible frames: This criterion is the rate of frames in which no possibility remains. Such situation can be caused by severely distorted images with large noise.

Trivially, quantitative values of these criteria varies according to presented movies. Instead, we pay attention to qualitative dependence on the noise level. As samples of the experiments, several scenes are selected from cinema films and two-dimensional coordinates of characteristic points are extracted by hand. In addition to the initial noise in hand extraction, artificial Gaussian noise of specified power is injected into the extracted coordinates. Then, the previously mentioned methods of model fitting is applied to the noisy coordinates, and dependencies to the noise level is investigated.

In experiments, we observed the following behaviors.

- Largest distance between succeeding frames does not change significantly according to the noise level.

- In contrast, rate of impossible frames increases according to the noise level.

Namely, naturalness of estimated motion does not have significant dependence on the amount of noises unless the noises are so large that no candidate of three-dimensional posture remains, for the samples in the experiments.

Discriminant technology(1) — Online linear discriminant analysis Online linear discriminant analysis (OLDA) is proposed and analyzed both theoretically and experimentally. In addition, recognition systems are developed and evaluated.

Simple methods have some advantages. They are easy to be implemented. Their behaviors are easy to be understood since they tend to have rigid theoretical analysis. Their performance can be superior to that of complex methods as far as tasks are in some classes since complex methods are often troubled by problems of slow learning and overfitting. For example, Webb says that linear schemes should be considered before more complex methods.

Fisher’s linear discriminant analysis (LDA) is a simple method of pattern classification. Its function is also referred to as dimensionality reduction, feature extraction,

Table 1: Comparison between PCA and LDA

	PCA	LDA
Input	real vectors	real vectors with their class labels
Aim	reconstruction of original data	preservation of class separability
Calculation	eigenvalue problem	generalized eigenvalue problem

multivariate data projection, and so on. In spite that LDA is an old method, it is still used practically in broad areas, e.g. image recognition.

In practical uses, we often face problems: (1) whole data are not given at once, (2) in addition, properties of data change through time, or (3) dimension of data is too large. To overcome them, we desire a method which has the following nature: (1) it can work when data are given one by one, (2) it has adaptability, and (3) its update procedure is computationally inexpensive, and it does not need to memorize whole the past data. We call such an algorithm as "online algorithm." The main contribution of the present paper is proposal of an online LDA algorithm.

For principal component analysis (PCA), which is another simple method of dimensionality reduction and feature extraction, online PCA algorithms have been studied and established well. Differences between PCA and LDA are summarized as follows (Table (1)):

Input: Real vectors are presented as input data for PCA. In addition to them, their class labels are required for LDA.

Aim: By PCA, we obtain the optimal linear transformation in the sense that we can reconstruct the original high-dimensional data from the reduced low-dimensional data with least square error. By LDA, we obtain reduced low-dimensional data such that (1) class separability is preserved and (2) Euclidean metric works appropriately as a measure of difference between reduced data.

Calculation: In order to perform PCA, we calculate major eigenvalues and corresponding eigenvectors of the variance matrix. In order to perform LDA, we calculate major *generalized eigenvalues* and corresponding *generalized eigenvectors* of between-class variance matrix under within-class variance matrix as the metric.

We have heuristically found a matrix dynamics

$$\begin{aligned}\frac{d}{dt}A(t) = BA(t) &- \frac{1}{2}BA(t)A(t)^T WA(t) \\ &- \frac{1}{2}WA(t)A(t)^T BA(t)\end{aligned} \tag{1}$$

and constructed an OLDA algorithm from it by stochastic approximation method. We later noticed that this dynamics can be expressed by a potential function and we have presented theoretical foundation of this algorithm based on mathematical analysis of the potential function. An example of the flow and the potential function of the dynamics (1)

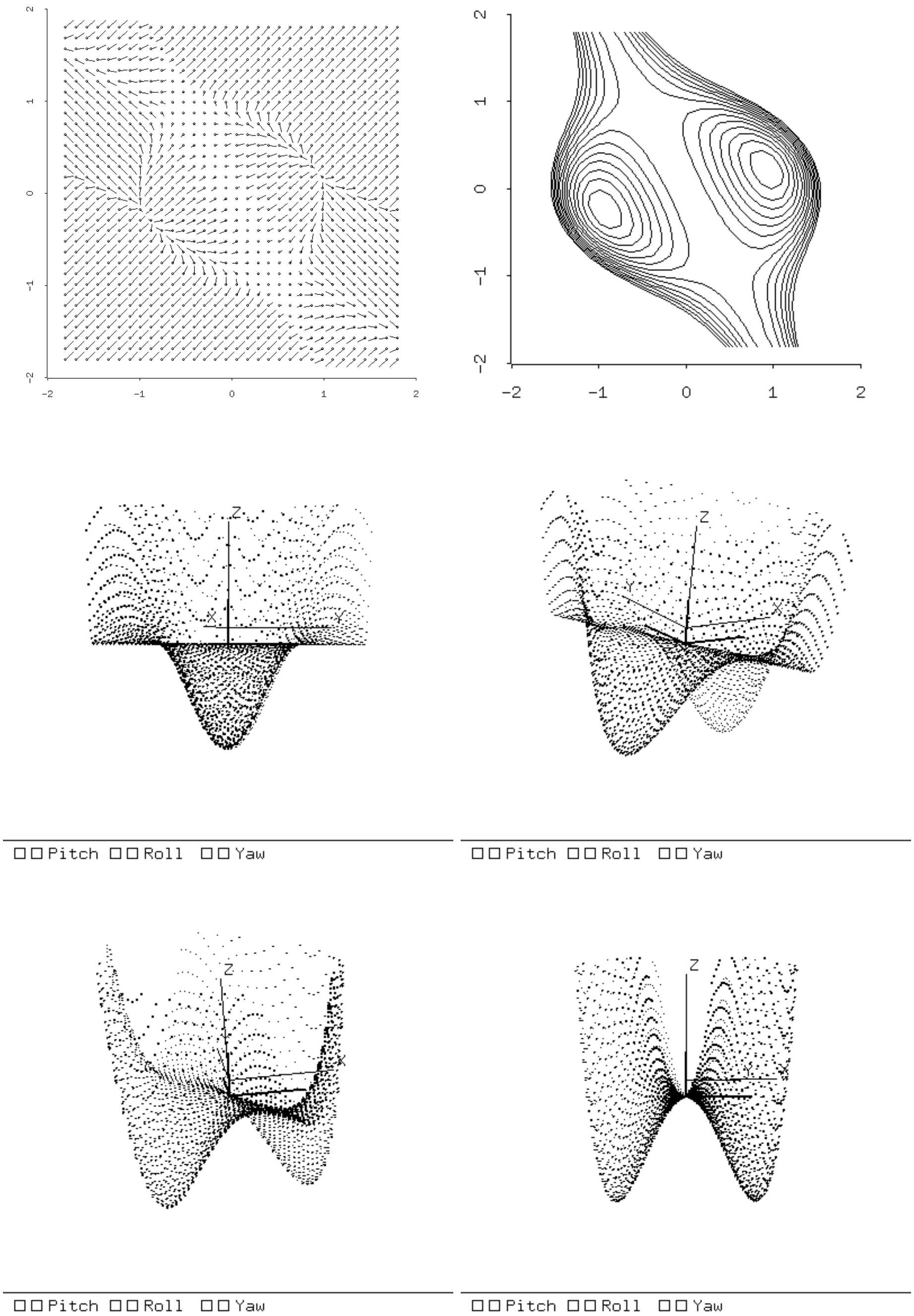

Figure 1: An example of the flow (top left) and the potential function (top right) of the dynamics (1). The middle and the lower figures are the pictures of the potential function from various angles.

is graphically shown in Figure 1 for the case of $N = 2$, $L = 1$. The convergence of the matrix dynamics for online LDA is analyzed. In particular, all fixed points are identified and their stability is determined.

Note that naive discretization of the dynamics (1) causes updating of large scale matrices. We have pointed out that it can be avoided by well-designed calculation process. At every time step $t = 1, 2, 3, \cdots$, a new pair $(\boldsymbol{x}(t), c(t))$ is presented, where $\boldsymbol{x}(t)$ is a data vector, $c(t) \in \{1, \cdots, M\}$ is the class of $\boldsymbol{x}(t)$, and M is the number of classes. The number L of features is less than M. Based on this pair, auxiliary variables are updated as follows:

$$t^c(t) = t^c(t-1) + \delta(c, c(t)), \quad \bar{\boldsymbol{x}}(t) = \left(1 - \frac{1}{t}\right) \bar{\boldsymbol{x}}(t-1) + \frac{1}{t}\boldsymbol{x}(t), \tag{2}$$

$$\bar{\boldsymbol{x}}^c(t) = \begin{cases} \left(1 - \frac{1}{t^c(t)}\right) \bar{\boldsymbol{x}}^c(t-1) + \frac{1}{t^c(t)}\boldsymbol{x}(t) & (c = c(t)), \\ \bar{\boldsymbol{x}}^c(t-1) & (c \neq c(t)), \end{cases} \tag{3}$$

$$\boldsymbol{v}^c(t) = \bar{\boldsymbol{x}}^c(t) - \bar{\boldsymbol{x}}(t), \quad \boldsymbol{w}(t) = \boldsymbol{x}(t) - \bar{\boldsymbol{x}}^{c(t)}(t), \tag{4}$$

$$\boldsymbol{y}^c(t) = A(t-1)^T \boldsymbol{v}^c(t), \quad \boldsymbol{z}(t) = A(t-1)^T \boldsymbol{w}(t), \tag{5}$$

$$F(t) = \frac{1}{M}\sum_{c=1}^{M} \boldsymbol{v}^c(t)\boldsymbol{y}^c(t)^T, \quad \boldsymbol{g}(t) = \frac{1}{M}\sum_{c=1}^{M} \boldsymbol{y}^c(t)\left(\boldsymbol{y}^c(t)^T \boldsymbol{z}(t)\right), \tag{6}$$

where $c = 1, \cdots, M$. Then the discriminant matrix A is updated as

$$A(t) = A(t-1) + \eta\Big(F(t) - \frac{1}{2}F(t)\boldsymbol{z}(t)\boldsymbol{z}(t)^T - \frac{1}{2}\boldsymbol{w}(t)\boldsymbol{g}(t)^T\Big), \tag{7}$$

where $\eta > 0$ is the learning coefficient. Note that the variables $\boldsymbol{y}^c, \boldsymbol{z}, F, \boldsymbol{g}$ are introduced instead of calculating $B(t) = (1/M)\sum_{c=1}^{M} \boldsymbol{v}^c(t)\boldsymbol{v}^c(t)^T$ itself so that $N \times N$ matrices are not needed to update A. In addition, $W(t) = (1/t)\sum_{\tau=1}^{t} \boldsymbol{w}(\tau)\boldsymbol{w}(\tau)^T$ is replaced with the instantaneous value $\boldsymbol{w}(t)\boldsymbol{w}(t)^T$.

As for the initial values, $t^c(0) = 0$, $\bar{\boldsymbol{x}}(0)$ and $\bar{\boldsymbol{x}}^c(0)$ are arbitrary vectors, and $A(0)$ is an arbitrary matrix which satisfies rank$A(0) = L$. In order to estimate the class for the presented data vector $\boldsymbol{x}$ by use of the obtained discriminant matrix A, we calculate $\boldsymbol{y} = A^T\boldsymbol{x}$ and $\boldsymbol{y}^c = A^T\boldsymbol{x}^c$ for each class c. Then we select the most likely c which minimizes the square error

$$\|\boldsymbol{y} - \boldsymbol{y}^c\|^2. \tag{8}$$

In practice, we have experienced divergence of variables during successive updating by the influence that

1. number of samples is finite, and
2. the original continuous-time dynamics is approximated by a discrete-time dynamics.

To prevent it, we have proposed relaxation method so that updating process is stabilized. In addition, the potential function of the dynamics has local minima and therefore the dynamics can be trapped to the local solution. We have proposed to select the initial matrix $A(0)$ near the zero matrix O based on the knowledge of the possible range of local solutions which is proved in mathematical analysis. Though the strict theoretical foundation of that proposal is insufficient, convergence to local solutions never occurred in our experiments of face image recognition.

Iterative updating algorithms like the above one can have a problem on selection of the learning coefficient in general. Indeed, divergence of the discriminant matrix are observed for some initial matrices. Thus, we have proposed an automatic tuning of the learning coefficient, and the effect of this tuning is experimentally confirmed. When we use conventional OLDA, we set learning coefficient η constant. However, there is no reason to set η so. Updating η in accordance with a 'scale' of A, we let η pertinence value. For this purpose, first, we note the fact that the following equation holds when discrimination matrix A has converged to a desirable value

$$E[A^T \boldsymbol{w}(t)\boldsymbol{w}(t)^T A] = I. \tag{9}$$

In particular,

$$\frac{1}{L}E[\boldsymbol{w}^T(t)AA^T\boldsymbol{w}(t)] = 1 \tag{10}$$

is implied from (9). Secondly, as a measure of 'scale' of A, we introduce a variable $\rho(t)$ which is an approximation of left side of (10). Expectation $E[\,\cdot\,]$ in left side of (10) is replaced with weighted time-average in $\rho(t)$. This $\rho(t)$ is updated by

$$\rho(t+1) = \rho(t) + \eta_\rho \left(\frac{1}{L}\boldsymbol{w}(t)^T A(t)A(t)^T \boldsymbol{w}(t) - \rho(t) \right), \tag{11}$$

where η_ρ is a small positive number. The initial value of ρ is set $\rho(0) = 1$. Finally, at every time t, we set learning coefficient η as

$$\eta(t) = \frac{\eta_0}{\rho(t)}, \tag{12}$$

where η_0 is a small positive number. We will refer to the parameter η_0 as base learning coefficient. With normalization of this η_0 by $\rho(t)$ (a measure of 'scale' of A), we can let η be proper at all times. Note that conventional OLDA is a special case of the presented algorithm. Indeed, ρ is not updated if $\eta_\rho = 0$. Hence, η remains constant in this case. Especially if $\rho(0) = 1$, it is clear from Eq.(12) that $\eta(t) = \eta_0$. In the above sense, η_0 corresponds to conventional parameter η. In order to verify the usability of the presented method, we did some experiments applying the method with some tasks and variations of OLDA. We examine the results of these experiments focusing three points: (a) dependence on conventional parameter η_0, (b) dependence on new parameter η_ρ, and (c) increase or decrease of discrimination ratio. First, we mention dependence on conventional parameter η_0. On η_0, it have been observed expansion of the range of allowable value. Namely, the presented method gives more robustness than conventional OLDA. Next, we mention dependence on new parameter η_ρ. If we set η_0 too large, discrimination ratio tends to decrease. However, setting of η_ρ is not strict compared with setting of η_0 in conventional method. Finally, we mention increase or decrease of discrimination ratio. On conventional OLDA, of course, in the case where the discriminant matrix A

have diverged, the discrimination ratio is enhanced by the presented method. However, in the case where A has not diverged, in the worst case, the discrimination ratio is decreased from 82% to 59%. We consider that the reason for this decreasing is too small η regulated by the presented method. Although it is an undesirable side effect, we can compromise on this decreasing of discrimination ratio because the presented method have the merit(robustness) which covers it.

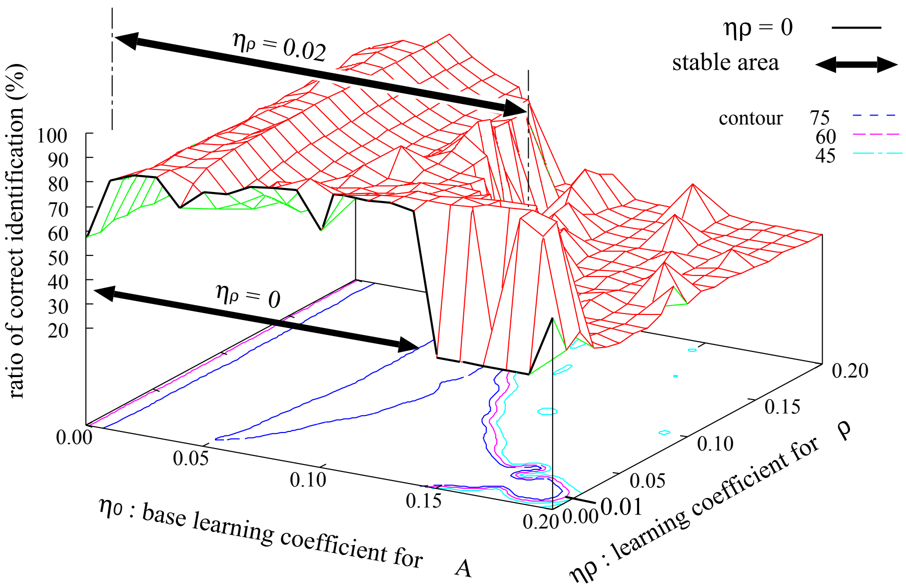

Figure 2: Average ratio of correct classification for the case of 3 classes ($t = 300$)

From the basic form (1), we have considered several variations. The following variations can be combined independently,

Asymmetric type: As a direct extension of the popular dynamics in online PCA, we obtain

$$\frac{d}{dt}A(t) = BA(t) - WA(t)A(t)^T BA(t). \tag{13}$$

Though the dynamics (13) has the merit that no local solutions exists, the convergence of the OLDA algorithm based on (13) is slower compared with the OLDA based on the symmetric dynamics (1). Interpolation of Eqs.(1) and (13),

$$\begin{aligned}\frac{d}{dt}A(t) = BA(t) &- \alpha BA(t)A(t)^T WA(t) \\ &- (1-\alpha)WA(t)A(t)^T BA(t)\end{aligned} \tag{14}$$

is also examined.

Sanger type: By this method, not only the generalized principal space but also the generalized principal components themselves can be extracted. In contrast to Sanger type, we call the basic one (13) as Oja type.

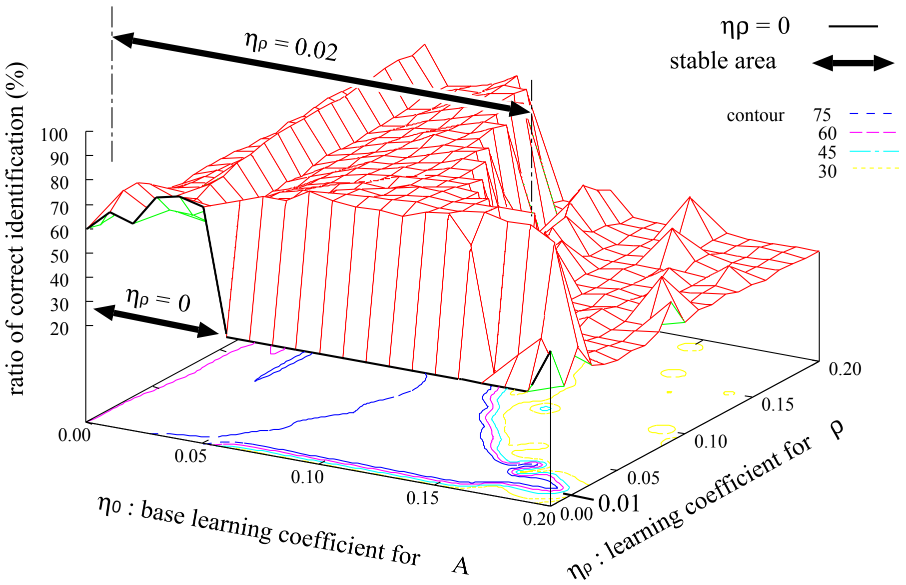

Figure 3: Average ratio of correct classification for the case of 4 classes ($t = 400$)

Hamahira type: This is the OLDA algorithm based on the Hamahira's variation

$$\begin{aligned} \frac{d}{dt}A(t) = BA(t) &- \frac{1}{2}A(t)A(t)^T BWA(t) \\ &- \frac{1}{2}A(t)A(t)^T WBA(t). \end{aligned} \tag{15}$$

We compared combinations of these variations experimentally for the task of face image recognition, and obtained the following results.

- Symmetric type achieved better classification performance than asymmetric type.
- Significant difference is not observed in comparison between Oja type and Sanger type.
- Hamahira type has the merit of faster convergence, while its final performance after sufficient training is worse than basic type.

In addition to the above academic outcomes, we have implemented practical systems based on OLDA and obtained the following technical outcomes.

Face recognition system: We have combined OLDA with face detection system and developed face a recognition system. It can accept online registration and deletion in real time on the fly.

Person identification from full-body silhouette images: In order to assist weak points of person identification from face images, such as illumination dependency, we examined person identification from full-body silhouette images. .

Arbitrary pattern recognition system: As an improvement of the above face recognition system, we extended it so that arbitrary pattern can be registered and deleted on the fly. (Figure 4) These operations can be instructed interactively with a mouse.

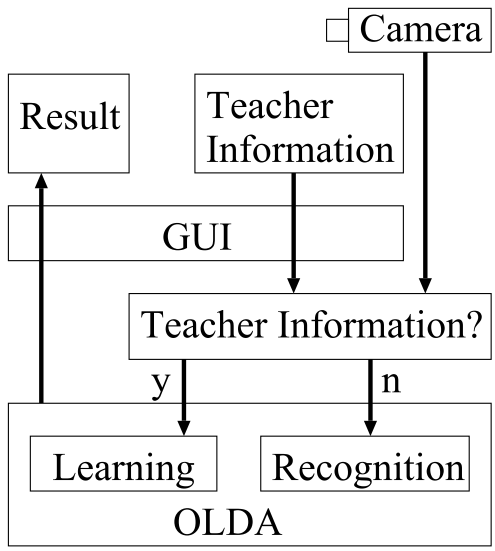

Figure 4: Arbitrary pattern recognition system based on OLDA

Human walking pitch extraction system: We have extracted human walking pitch based on visual tracking of person's heel. Arbitrary pattern recognition system based on OLDA is used for registration and discrimination of the heel (Figure 6). In addition, neural oscillator is entrained by the extracted pitch and synchronization is confirmed in simulation.

Discriminant technology(2) — Dimensionality reduction For pattern classification on high-dimensional data, such as images, the dimensionality reduction as a preprocessing is effective. By dimensionality reduction, we can (1) reduce storage capacity or amount of calculation, and (2) avoid "the curse of dimensionality" and improve classification performance. Popular tools for dimensionality reduction are Principal Component Analysis (PCA), and Linear Discriminant Analysis (LDA). However, these tools have weaknesses. PCA does not use class labels which sample data have. LDA has a restriction on the number of dimensions after reduction. These cause shortage of information for pattern classification. To overcome them, we propose a new dimensionality reduction technique based on an information-theoretic measure for distance between distributions. It takes the class labels into consideration and still it does not have restriction on number of dimensions after reduction. Improvement of classification performance has been confirmed experimentally.

In addition, learning of multi-attributes is studied; it is interesting because of the following reasons.

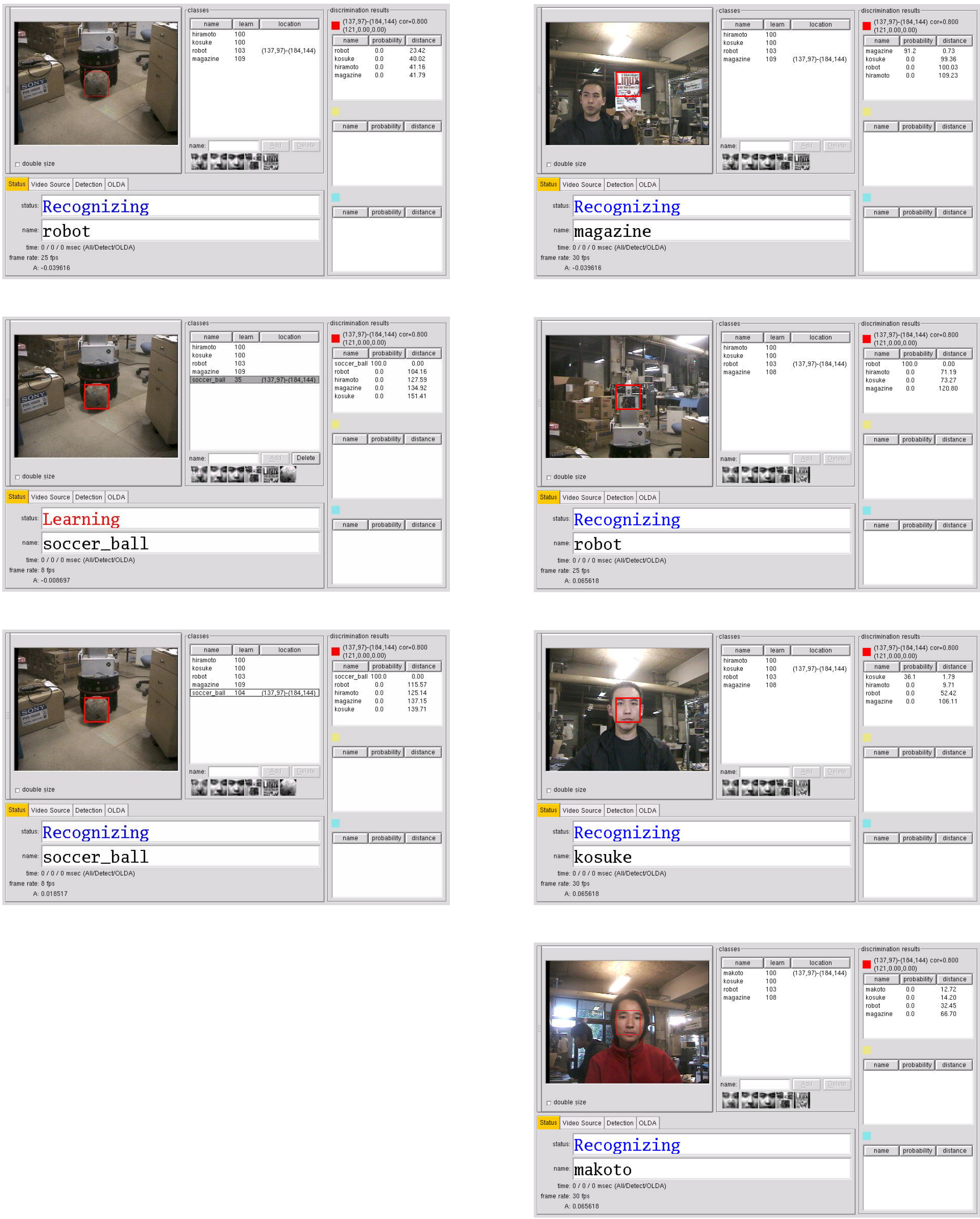

Figure 5: Sample screen shots of the recognition system

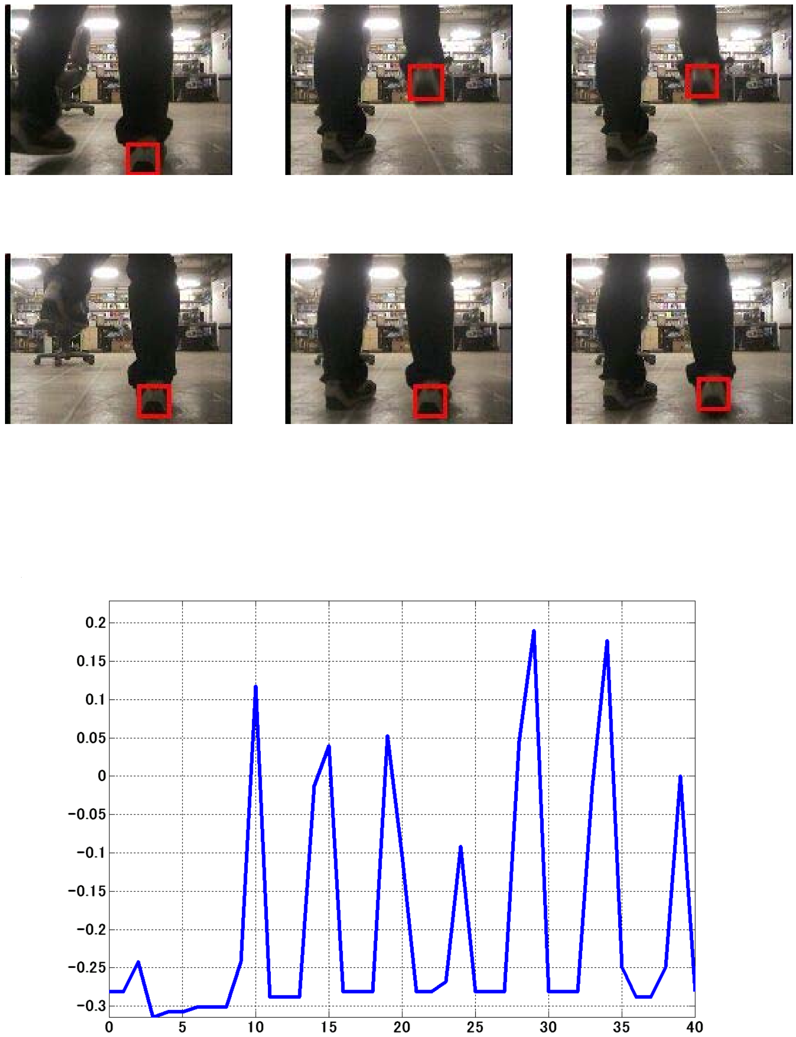

Figure 6: Heel tracking and walking pitch extraction (horizontal axis: frame, vertical axis: heel position (radian))

1. Real-world objects often have two or more significant attributes. For example, face images have attributes of persons, expressions, and so on.

2. Even if you are interested in only one of those attributes, additional informations on auxiliary attributes can help recognition of the main one.

From this motivation, we have proposed several approaches.

1. Estimation of missing data based on internal representation which is acquired by a layered neural network

2. Unsupervised classification via competitive units

3. Supervised classification with an EM-like algorithm which is obtained from information-theoretic consideration; its main idea is mutual suggestion of hints between a pair

of classifiers.

(2) Results and their importance

In the study on modeling of brain information processing, we have clarified the mechanisms of synaptic regulation and proposed multiplicative STDP rules which can achieve synaptic regulation. We have also shown that binocular rivalry is explained by using a network which consists of locally connected neurons with orientation preference. Also, we have investigated the property of the connection which is required for realization of the qualitative results shown by the psycho-physical experiments on the binocular rivalry. These outcomes are expected to contribute understanding of information processing in the brain.

In the study on recognition technology, three dimensional motion estimation from two-dimensional movie is examined. We have tried estimation of three-dimensional motions based on positions of characteristic body points which are detected from two-dimensional movies. In experiments on qualitative properties for the case that noises are added to the detected positions, an interesting behavior is observed; unless the noises are so large that no candidate of three-dimensional posture remains, naturalness of estimated motion has little dependence on the amount of noises. Our result suggests probability of three-dimensional motion capture based only on two-dimensional in contrast to today's motion capture systems which requires special sensors, special markers, special rooms, and so on.

In the study on discrimination technology, we have constructed online linear discriminant analysis (OLDA) and related techniques. They have adaptability and can follow a changing situation. In addition to theoretical investigation of them, we have also built several vision systems as applications of them, face recognition system, arbitrary pattern recognition system, human walking pitch extraction system based on visual tracking of person's heel, and so on.

References

[1] Y.Sakai: Neuronal Integration Mechanisms Have Little Effect on Spike Autocorrelations of Cortical Neurons, Neural Networks, Vol.14, No.9, pp.1145–1152, 2001.

[2] T.Muraoka, Y.Sakai, S.Yoshizawa: Motion Capturing System from Video Images, Proc. of 2001 Annual Conference of Japanese Neural Network Society, pp.115–116, 2001.

[3] K.Hiraoka, S.Yoshizawa, K.Hidai, M.Hamahira, H.Mizoguchi, and T.Mishima: Convergence Analysis of Online Linear Discriminant Analysis, International Joint Conference on Neural Networks (IJCNN), III-387-391, 2000.

[4] K.Hiraoka, M.Hamahira, K.Hidai, H.Mizoguchi, T.Mishima, and S.Yoshizawa: Fast algorithm for online linear discriminant analysis, IEICE Transactions on Fundamental of Electronics, Communications and Computer Science, Vol. E84-A, No. 6, pp.1431–1434, 2001.

[5] K.Okabe, K.Hidai, M.Hiramoto, H.Mizoguchi, K.Hiraoka, T.Shigehara, T.Mishima, and S.Yoshizawa: OLDA based Arbitrary Pattern Recognition System with Interactive Trainability, Proceedings of SI2002, pp.159–160, 2002.