Self-Organizing Internal Representation of Ego-Motion and Its application to Motion-Compensated Inter-Frame Subtraction

Takashi Toriu and Thi Thi Zin
Graduate School of Engineering
Osaka City University
Osaka, Japan
{toriu; thithi}@info.eng.osaka-cu.ac.jp

Hiromitsu Hama
R&D Center of 3G Search Engine
Research Center for Industry Innovation
Osaka, Japan
hama@ado.osaka-cu.ac.jp

Abstract— **In a previous paper, we proposed an unsupervised learning algorithm for self-organizing an internal representation of ego-motion. In this paper, we propose a method to predict the image at the next instance using a time evolution operator generated on the basis of the internal representation of ego-motion. In addition, we propose a method of motion-compensated inter-frame subtraction. By subtracting the predicted image at the next instance from the true image, we can obtain an image that has high intensity in the region of the moving object. This method is effective even if the camera itself is in motion. Because this method does not utilize any knowledge of geometric nature during image generation, it is not affected by any image distortion. We show the results of the experiments conducted using a randomly synthesized image and the real image.**

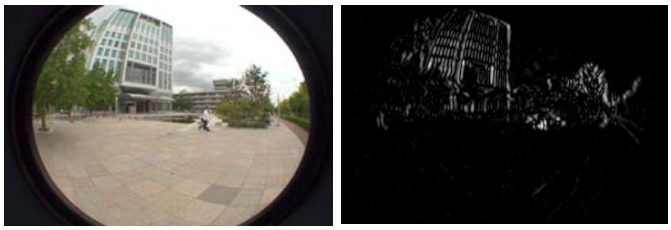
Keywords- Vision; Ego-motion; Unsupervised learning; Internal representation of ego-motion; Motion-compensated Inter-Frame Subtractio

I. INTRODUCTIONE

Existing robot vision systems are designed on the basis of human knowledge of the external world. The knowledge is embedded in a robot as a computer program. Because human knowledge is incomplete, robot behaviour based on this knowledge must also be incomplete. The robot can only solve problems that were selected in advance by the designer.

Recently, cognitive developmental robotics [1, 2] has been put forth as a new paradigm that emphasizes on the automatic development of a robot's knowledge base [3, 4]. It is postulated that robot behaviors should not be fixed, but rather emergent depending on the experiences of the robot. Based on this idea, we have researched on methods of spontaneously generating function of vision from only robot experiences [5-7]. We persisted on the postulate that robot behaviors should not be fixed, but emergent depending on the experiences of the robot [8, 9].

In our previous papers [5-7], we introduced a time evolution operator. In [7], we proposed an unsupervised learning algorithm for self-organizing the internal representation of ego-motion. We showed that motion parameters could be topographically and spontaneously mapped onto a robot's internal parameter space without any knowledge of geometric nature during image generation; therefore, it is not affected by image distortions such as those introduced by omnidirectional or fish-eye cameras. In this paper, we propose a method to predict the image at the next instance using the time evolution operator. In particular, the system first estimates the internal representation of its ego-motion, and then, using the corresponding time evolution operator, it predicts the image at the next instance.



A B

Figure 1. An example of the result of conventional inter-frame subtraction.

Inter-frame subtraction is often used as a method to detect a moving object in an image. This method is effective when the object is moving in a static background because only the region of the moving object has non-zero subtraction. However, when the camera is itself in motion, the inter-frame subtraction method fails to detect the moving object. This is because the background region also has non-zero subtraction. Fig.1 shows an example of the result of conventional inter-frame subtraction. A frame of a video taken by a moving fish-eye camera is shown in A and the result of inter-frame subtraction is shown in B. The walking person in the central region is not distinguished because background region is also enhanced. Solving this problem is an important application of the proposed method for predicting the image at the next instance. By subtracting the predicted image at the next instance from the true image at the next instance, we can obtain an image that has a high intensity in the region of the moving object. This is because the time evolution operator represents ego-motion and

is expected to compensate for the image changes caused by ego-motion.

Historically, the problem of image changes due to ego-motion has been examined as follows. The image changes are generated according to the optical flow arising from ego-motion, and inversely, ego-motion is detected on the basis of the optical flow [10-12]. It is assumed that the optical flow is produced according to a physical law of optics, and the problem of estimating ego-motion is solved using the knowledge of inverse optics. Once ego-motion is estimated, the image at the next instance can be obtained using a physical law of optics. However, generally, the physical process is so complicated that it is only approximated as a simple model. Therefore, when the model is insufficient, the image at the next instance cannot be accurately estimated.

Biological vision is highly flexible. When a person wears glasses with reverse prisms, he/she can adapt to the reversed world after a certain period of adjustment [13,14]. The visual function is considered to be organized on the basis of visual experiences. Furthermore, it has been recognized that the primate visual cortex is self-organized through visual experiences [15,16].

The proposed method is as flexible as biological vision. Because the proposed method does not assume any knowledge of the external world, it can cope with image distortions such as those introduced by reverse prisms, omnidirectional cameras, or fish-eye cameras. In the next section, we formulate the problem of self-organizing the internal representation of ego-motion and propose a method to predict the image at the next instance. Then, in Section III, we outline the algorithm for motion-compensated inter-frame subtraction, and in Section IV, we present the experimental results. Finally, we conclude the paper in Section V.

II. FORMULATION

We introduced a new concept of a time evolution operator in the previous paper [5]. In general, an operator converts a vector into another vector. Let M^d be a time evolution operator associated with an ego-motion specified by symbol d, and let $I(x, y, t)$ be an image at time t. Then, the image at the next instance $I(x, y, t+\Delta t)$ is assumed to be

$$I(x, y, t + \Delta t) = M^d(\Delta t) I(x, y, t) . \quad (1)$$

In the paper [5], it was shown that M^d can be obtained through a learning procedure if many samples of the pair of $I(x, y, t+\Delta t)$ and $I(x, y, t)$ are prepared for the ego-motion d.

In this method, the ego-motion was restricted to a discrete time and eight discrete types of motion. In the paper [6], we advanced the method to cope with continuous time and continuous motion; we introduced an infinitesimal time evolution operator H^d, which satisfies the following relation

$$\frac{\partial I(x, y, t)}{\partial t} = H^d I(x, y, t) \quad (2)$$

The operator H^d for arbitrary motion parameters can be obtained through a learning process using several samples of the pair of $\partial I(x, y, t)/\partial t$ and $I(x, y, t)$. The learning of M^d or H^d is a type of supervised learning; the supervisor must provide the target values of the motion parameters. Once the learning is performed, the system can recall the ego-motion from the pair of $I(x, y, t+\Delta t)$ and $I(x, y, t)$ or the pair of $\partial I(x, y, t)/\partial t$ and $I(x, y, t)$. If the ego-motion d is given, conversely, the system can recall $I(x, y, t+\Delta t)$ or $\partial I(x, y, t)/\partial t$ from the current input $I(x, y, t)$. Thus, we can predict the image at the next instance.

In the paper [7] we proposed an unsupervised learning algorithm to self-organize the internal representation of ego-motion from a number of sample pairs of $\partial I(x, y, t)/\partial t$ and $I(x, y, t)$. We showed that the internal space was topographically isomorphic with the space of real ego-motion parameters. We assumed that the infinitesimal time evolution operator H^d could be represented as

$$H^d = \sum_{k=1}^{K} \lambda^d{}_k H_k , \quad (3)$$

where $H_k (k=1,2,\ldots,K)$ is a set of operator bases, each of which corresponds to an elementary motion and $\lambda^d{}_k (k=1,2,\ldots K)$ are coefficients determined according to the specific motion d. K is a dimension of a motion parameter space. For instance, if the motion is translation (tx, ty) in a two-dimensional plane, K is 2. Before the learning, the system does not have any knowledge about H^d. After the learning is completed, the system generates the operator bases H_k ($k = 1,2,\ldots, K$). The space spanned by these operator bases is the internal space of ego-motion. For each pair of $\partial I(x, y, t)/\partial t$ and $I(x, y, t)$, the system derives such λ_k ($k = 1,2,\ldots, K$) that minimizes the following objective function:

$$J(\lambda_1,\ldots,\lambda_K) = \left(\frac{\partial I(x, y, t)}{\partial t} - \sum_{i=1}^{K} \lambda_k H_k I(x, y, t) \right)^2 . \quad (4)$$

The coefficients λ_k ($k = 1,2,\ldots, K$)s are internally represented motion parameters. These have one-to-one corresponding with the objective motion parameters.

In principle, the unsupervised learning proceeds in the way mentioned above. In practice, however, it is difficult to execute this process because it needs an enormous amount of computation time. In fact, when the number of pixels of the image $I(x, y, t)$ is N x N, the operator H_k is a matrix with a size of N^4, and it needs computation time proportional to N^{12} to obtain H_k; it is too large for practical use. To solve this problem, in the paper [7], we reduced the dimension of the image by principal component analysis. We defined a sensory vector $\mathbf{p}(t)$, which was a low dimensional vector obtained by principal component analysis from the image $I(x, y, t)$. We showed that we could define a time evolution operator H'^d in the same way and the following relationship held:

$$\frac{d\mathbf{p}(t)}{dt} = H'^d \mathbf{p}(t) , \quad (5)$$

and internal representation of ego-motion could be generated by a unsupervised learning in the same way.

Thus, we can predict a sensory vector at the next instance after obtaining time evolution operator, but it does not mean that we can obtain the image at the next instance. To realize

This work is partially supported by SCOPE: Strategic Information and Communications R&D Promotion Program(102307010).

motion-compensated inter-frame subtraction, we need the image at the next instance. In this paper, we propose a new method of unsupervised learning to obtain the time evolution operator in (3), not in (5) with reasonable computation time. Using this method, we can predict the image at the next instance, and therefore, we can realize motion-compensated inter-frame subtraction.

We assume that time evolution operator can be defined using a set of shift invariant linear transformation $h_p(m, n)$ of size F x F and that (2) is represented as follows:

$$\frac{\partial I(x,y,t)}{\partial t} = H^d I(x,y,t) = \sum_{i=-F/2}^{F/2} \sum_{j=-F/2}^{F/2} \sum_{p=1}^{P} v^d{}_p(x,y) h_p(i,j) I(x+i, y+j, t). \quad (6)$$

We regard $h_p(i, j)$ as a F x F digital filter to detect local features of images. Let H_p be an operator defined as

$$H_p I(x,y,t) = \sum_{i=-F/2}^{F/2} \sum_{j=-F/2}^{F/2} h_p(i,j) I(x+i, y+j, t), \quad (7)$$

then, (6) is expressed as follows:

$$\frac{\partial I(x,y,t)}{\partial t} = \sum_{p=1}^{P} v^d{}_p(x,y) H_p I(x,y,t). \quad (8)$$

Comparing (8) and (2), time evolution operator is expressed as follows:

$$H^d = \sum_{p=1}^{P} v^d{}_p(x,y) H_p . \quad (9)$$

To clarify the meaning of (6), (8) and (9), we consider a specific case. We assume that the camera is moving in a static environment and every point (x, y) in the image moves to the point $(x + v^d{}_x dt, y + v^d{}_y dt)$ in accordance with the camera motion, where $v^d{}_x$ and $v^d{}_y$ are optical flows and dt is the time interval. In addition, we assume that the pixel value of the corresponding point does not change, i.e. the following relationship holds:

$$I(x + v^d{}_x dt, y + v^d{}_y dt, t + dt) = I(x,y,t) . \quad (10)$$

In the limit of $dt \to 0$, we obtain

$$\frac{\partial I(x,y,t)}{\partial t} = -v^d{}_x(x,y)\frac{\partial I(x,y,t)}{\partial x} - v^d{}_y(x,y)\frac{\partial I(x,y,t)}{\partial y} = \left\{ v^d{}_x(x,y)\left(-\frac{\partial}{\partial x}\right) + v^d{}_y(x,y)\left(-\frac{\partial}{\partial x}\right)\right\} I(x,y,t) , \quad (11)$$

where we denote $v^d{}_x(x, y)$ and $v^d{}_y(x, y)$ as optical flows because they are depend on position (x, y). If we replace differential operators $-\partial / \partial x$ and $-\partial / \partial y$ with 3 x 3 digital filters

$$h_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}, \quad (12)$$

and

$$h_y = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \quad (13)$$

respectively, we obtain

$$\frac{\partial I(x,y,t)}{\partial t} = \sum_{i=-1}^{1} \sum_{j=-1}^{1} \left(v^d{}_x(x,y) h_x(i,j) + v^d{}_y(x,y) h_y(i,j) \right) I(x+i, y+j, t) \quad (14)$$

Equation (14) is a special case of (6).

We derived (14) using a knowledge how optical flow arises. Because we have a standpoint that we should build a system using a prior knowledge as less as possible, we start from more general formula (6) rather than (10). We call coefficients $v^d{}_p(x, y)$ generalized optical flow because it corresponds to usual optical flow in (11).

As discussed above, time evolution operator can be approximated as a linear sum of a number of operator bases as in (3). From this fact, we conclude that generalized optical flows $v^d{}_p(x, y)$ can be approximated as a linear sum of the basis functions of optical flows $v_{p1}(x, y)$, $v_{p2}(x, y)$, …, $v_{pK}(x, y)$ as follows:

$$v^d{}_p(x,y) = \sum_{k=1}^{K} \lambda^d{}_k v_{pk}(x,y) . \quad (15)$$

The coefficients $\lambda^d{}_1$, $\lambda^d{}_2$, ⋯, $\lambda^d{}_K$ depend on the ego-motion d and three-dimensional structure of the environment. We can consider them as parameters representing the motion and structure.

As in the paper [7], the basis functions of the optical flow are obtained spontaneously by unsupervised learning. First, we prepare M image sequences $I^1(x, y, t)$, $I^2(x, y, t)$, ⋯, $I^M(x, y, t)$. Motion and structure parameters in each image sequence $I^m(x, y, t)$ is assumed to be $d(m)$. For each image sequence, we seek the optimum $v^{d(m)}{}_p(x, y)$, which minimize

$$J[\{v^{d(m)}{}_p\}] = \sum_{t=1}^{T} \left(\frac{\partial I^m(x,y,t)}{\partial t} - \sum_{i=-F/2}^{F/2} \sum_{j=-F/2}^{F/2} \sum_{p=1}^{P} v^{d(m)}{}_p(x,y) h_p(i,j) I(x+i, y+j, t) \right)^2 , \quad (16)$$

where $m = 1, 2, …, M$. We partially differentiate the objective function $J[\{v^{d(m)}{}_p\}]$ by $v^{d(m)}{}_p$ and obtain the following system of linear equations:

$$c^m{}_{tp}(x,y) = \sum_{q=1}^{P} a^m{}_{pq}(x,y) v^{d(m)}{}_q(x,y) \quad (p = 1,2,...,P), \quad (17)$$

where

$$c^m{}_{tp}(x,y) = \sum_{t=1}^{T} \frac{\partial I^m(x,y,t)}{\partial t} \sum_{i,j} h_p(i,j) I(x+i, y+j, t),$$

$$a^m{}_{pq}(x,y) = \sum_{t=1}^{T} \sum_{i,j} h_p(i,j) I(x+i, y+j, t) \sum_{i',j'} h_p(i',j') I(x+i', y+j', t). \quad (18)$$

In this equation, we denote $\sum_{i,j}$ as summation $\sum_{i=-F/2}^{F/2} \sum_{j=-F/2}^{F/2}$. We introduce a P x P matrix A^m whose pth and qth element is $a^m{}_{pq}$ and let $B^m = (A^m)^{-1}$ be inverse of A^m. Then, the solution of the system of linear (13) is obtained as follows:

$$v^{d(m)}{}_p(x,y) = \sum_{q=1}^{P} b^m{}_{pq}(x,y) c^m{}_{tp}(x,y) \quad (p = 1,2,...,P), \quad (19)$$

where $b^m{}_{pq}$ is the pth and qth element of B^m.

Then, we obtain basis functions of optical flow $v_{p1}(x, y)$, $v_{p2}(x, y)$, …, $v_{pK}(x, y)$ from generalized optical flows $v^{d(m)}{}_1(x, y)$, $v^{d(m)}{}_2(x, y)$, …, $v^{d(m)}{}_p(x, y)$, where m = 1, 2, …, M. In fact, we regard each set of $v^{d(m)}{}_1(x, y)$, $v^{d(m)}{}_2(x, y)$, …, $v^{d(m)}{}_p(x, y)$ as a P x N x N dimensional vector $\mathbf{v}^{d(m)}$, and we apply principal component analysis to M vectors of $\mathbf{v}^{d(1)}$, $\mathbf{v}^{d(2)}$, …, $\mathbf{v}^{d(M)}$. The basis functions of optical flow $\mathbf{v}_1$, $\mathbf{v}_2$, …, $\mathbf{v}_K$ are obtained as K principal eigen vectors. Here $\mathbf{v}_k$ is a vector which consists of v_{1k}, v_{2k}, …, v_{Pk}. These basis functions of optical flow generate a space of internal representation of the motion and structure of the environment.

Once a set of basis functions is established, the motion and structure parameters λ_1, λ_2, $\cdots$, λ_K for the arbitrary image sequence $I(x,y,t)$ are obtained as follows. First, we define an objective function as

$$J\left(\lambda^d{}_1, \lambda^d{}_2, ..., \lambda^d{}_K\right) = \sum_{t=1}^{T} \left(\frac{\partial I(x,y,t)}{\partial t} - \sum_{i,j} \sum_{p=1}^{P} v^d{}_p(x,y) h_p(i,j) I(x+i, y+j, t) \right)^2 = \sum_{t=1}^{T} \left(\frac{\partial I(x,y,t)}{\partial t} - \sum_{i,j} \sum_{p=1}^{P} \sum_{k=1}^{K} \lambda^d{}_k v_{pk}(x,y) h_p(i,j) I(x+i, y+j, t) \right)^2. \quad (20)$$

Then, $\lambda^d{}_1$, $\lambda^d{}_2$, $\cdots$, $\lambda^d{}_K$ are obtained by minimizing this objective function. In fact, the objective function $J(\lambda^d{}_1, \lambda^d{}_2, ..., \lambda^d{}_K)$ is differentiated by each $\lambda^d{}_l$ and the result is set to zero. Then, we obtain a system of linear equations for $\lambda^d{}_1$, $\lambda^d{}_2$, $\cdots$, $\lambda^d{}_K$ and obtain the solution by solving it. Next, from (9) and (15), the time evolution operator for the arbitrary image sequence $I(x,y,t)$ is constructed as follows:

$$H^d = \sum_{p=1}^{P} \sum_{k=1}^{K} \lambda^d{}_k v_{pk}(x,y) H_p. \quad (21)$$

Because the time derivative of $I(x, y, t)$ is derived using (2), the image at the next instance is estimated as

$$\hat{I}(x,y,t+dt) = I(x,y,t) + \frac{\partial}{\partial t} I(x,y,t) dt = I(x,y,t) + H^d I(x,y,t) dt. \quad (22)$$

III. Inter-Frame Subtraction

Fig. 2 shows the outline of the algorithm for inter-frame subtraction, which is utilized as a tool for detecting moving objects in an image sequence captured by a moving camera. This algorithm consists of three stages: the first is preliminary learning, the second is unsupervised learning and the third is inter-frame subtraction.

In the preliminary learning stage, F x F digital filters to detect local features are obtained as follows. We do not define local features ad hoc, but we decide them through unsupervised learning by principal component analysis. First, a number of image samples $I^1(x, y)$, $I^2(x, y)$, …, $I^S(x, y)$ are prepared. We convert these images to gray scale images and put a F x F window at randomly selected point on these images. We extract local image from this window and regard it a vector $\mathbf{u}_i$ of dimension F x F. We collect a number of $\mathbf{u}_i$ (i = 1, 2, …, Q) in this way and apply principal component analysis to them to obtain P local feature bases $h_p(i, j)$ and record them in the first dictionary.

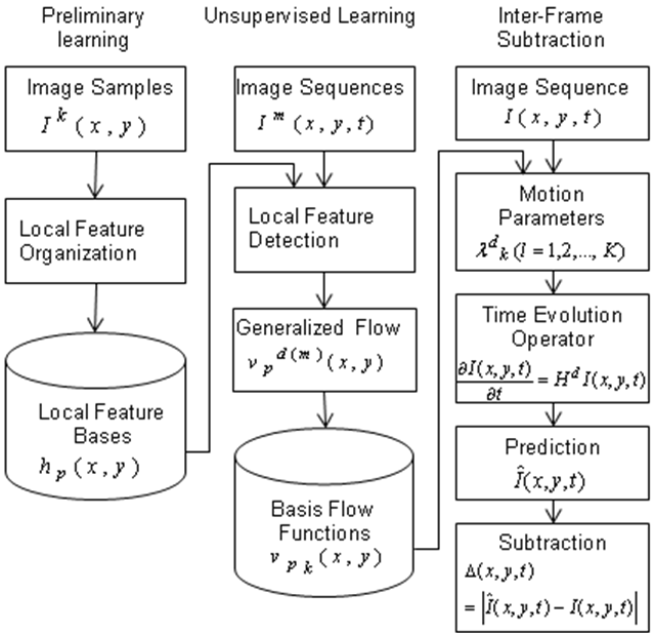


Figure 2. Outline of the algorothm for inter-frame subtraction.

In the unsupervised learning stage, first, a number of image sequences $I^1(x, y, t)$, $I^2(x, y, t)$, …, $I^M(x, y, t)$ are input. Then, using $h_p(i, j)$, we obtain the generalized optical flows $v_p{}^{d(m)}$ (p = 1, 2, …, P) by minimizing the objective function (16) as mentioned in the previous section, where P is the number of local feature bases. Then, we obtain the basis flow functions $v_{pk}(x, y)$ (k = 1, 2, …, K) that generate the internal space of motion and structure parameters by applying principal component analysis to the flow functions, where K is the number of principal eigen vector. The results are recorded in

the second dictionary of the system. In the inter-frame subtraction stage, an image sequence is input, and then, the motion and structure parameters are estimated by minimizing the objective function (20) using the basis flow functions from the dictionary. Next, the time evolution operator H^d is constructed according to (22), and the image at the next instance $\hat{I}(x, y, t)$ is predicted by (21). Then, inter-frame subtraction is performed by subtracting the actual image $I(x, y, t)$ from the predicted image $\hat{I}(x, y, t)$.

IV. EXPERIMENTS

We conducted experiments using synthesized images to validate the proposed method. In these experiments, a textured image was transformed by Gaussian filtering into a randomized image. We assumed that an ideal pinhole camera was moving in front of the backdrop. The region of the backdrop image in the camera field of view was then successively clipped and distorted to produce input images of 256×256 pixels. Fig. 3A and 3B show an undistorted image and a distorted image, respectively.

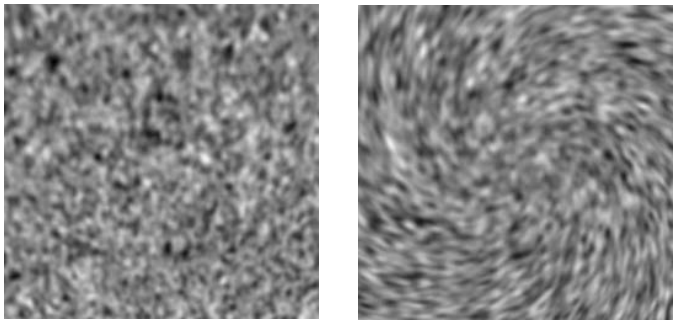


Figure 3. Examples of input images. A is an undistorted image. B is a distorted image.

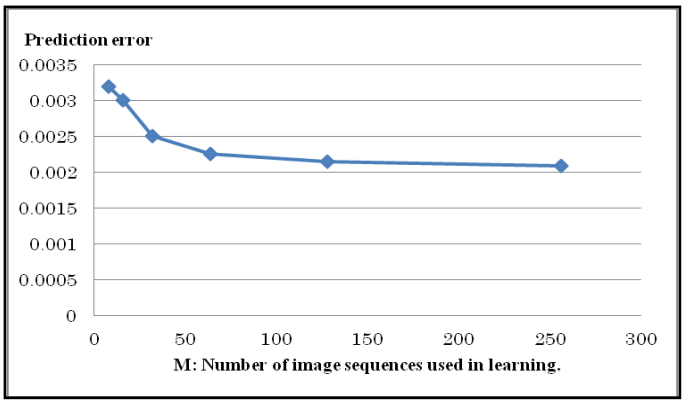


Figure 4. Results of the first experiment.

We conducted three experiments. In the experiments, we predicted the image at the next instance and evaluated the prediction error. The prediction error is the rout mean square error between the pixel values of the true and predicted images. The result of the first experiment is shown in Fig. 4. The horizontal axis M shows the number of image sequences in the learning stage and the vertical axis shows the prediction error normalized by the prediction error in the case of conventional inter-frame subtraction. We show average errors over 100 trials. Each image sequence consists of 64 frames. The dimension of the local feature basis is three and two local bases are used, which are obtained by applying principal component analysis to local images clipped from randomized images. The prediction error slightly decreases as the number of sequences increases.

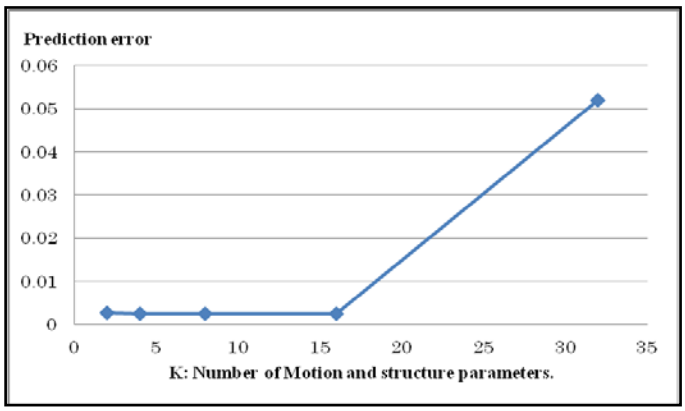


Figure 5. Results of the second experiment.

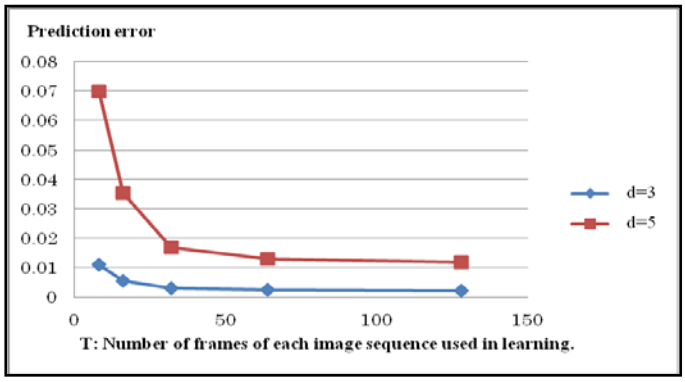


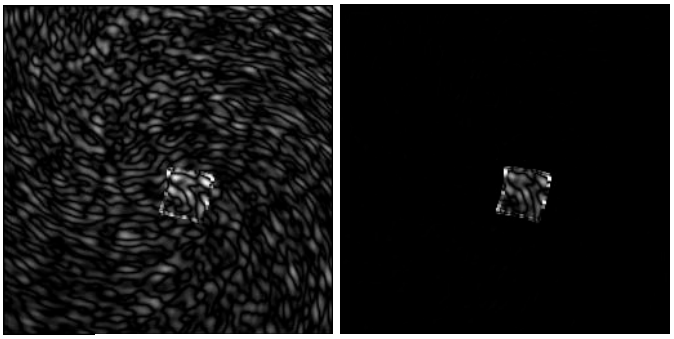
Figure 6. Results of the third experiment.

The result of the second experiment is shown in Fig. 5. The horizontal axis K shows the number of motion and structure parameters. The number of image sequences M = 32. The prediction error becomes high when K = 32. This fact indicates that the number of image sequences used in unsupervised learning stage should be larger than K. Fig. 6 shows the results of the third experiment. The horizontal axis T shows the number of frames of each image sequence used in the learning stage. When T is small, the prediction error is high. T should be larger than 32. The reason why the prediction error is high in small T is discussed in the following. In the learning stage, generalized optical flows $v^{d(m)}{}_1(x, y)$, $v^{d(m)}{}_2(x, y)$, …, $v^{d(m)}{}_p(x, y)$ are determined in each image sequence $I^{(m)}(x, y, t)$ based on the equation

$$\frac{\partial I^{(m)}}{\partial t} = \sum_{i=-F/2}^{F/2} \sum_{j=-F/2}^{F/2} \sum_{p=1}^{P} v^{d(m)}{}_p(x, y) h_p(i, j) I^{(m)}(x+i, y+j, t), \tag{22}$$

where t = 1, 2, …, T. The number of unknown variables $v^{d(m)}{}_p(x, y)$ is P and the number of equations is T. When T is larger

than *P*, this problem is generally well defined. However, when *T* equations are not independent with each other, solution is not determined uniquely. If independency of the equations is low, solutions are expected to be unstable. In our experiment, when the number of frames *T* is small, this kind of problem would arise. This would be the reason why the prediction error is high when *T* is small. We call this problem generalized aperture problem because the same kind of problem in the process of optical flow detection is called aperture problem [16]. In the third experiment, we altered *d*, dimension of features. We compared $d = 3$ and $d = 5$. The prediction error was higher when $d = 5$. The reason may be image is too blurred when the dimension of local feature becomes large.



A B

Figure 7. Results of the motion-compensated inter-frame subtraction.

A B

Figure 8. A result using a real video image.

Furthermore, we conducted an experiment to evaluate the effectiveness of the proposed motion-compensated inter-frame subtraction for detecting a moving object in an image sequence obtained by a moving camera with distortion. Examples of the results using synthesized images are shown in Fig. 7. Image A is obtained by conventional inter-frame subtraction and image B is obtained by the proposed method. In the conventional method, there is significant noise in the background region, but the noise is effectively removed in the proposed method.

The result for real image is shown in Fig. 8. The proposed method is applied to the same image in Fig. 1. The walking person is successfully enhanced. This result shows effectiveness of the proposed method.

V. COMCLUSION

In this paper, we proposed a novel method of motion-compensated inter-frame subtraction for detecting moving objects in an image sequence obtained by a moving camera. This method is based on the internal representation of the motion and structure that is spontaneously generated by unsupervised learning. The image movement is internally represented and based on it, the system compensates for the motion, and the moving objects can be effectively detected by motion-compensated inter-frame subtraction. Because this method does not utilize any knowledge of geometric nature during image generation, it is not affected by image distortions such as those introduced by omnidirectional or fish-eye cameras. In the near future, we plan to consider on generalized aperture problem and to conduct more elaborate experiments using real images in practical situations.

REFERENCES

[1] M. Asada, K. F. MacDorman, H. Ishiguro, and Y. Kuniyoshi, Cognitive developmental robotics as a new paradigm for the design of humanoid robots, *Robotics and Autonomous Systems*, Vol. 37, pp. 185-193, 2001.

[2] Y. Kuniyoshi, Y. Yorozu, Y. Ohmura, K. Terada, T. Otani, A. Nagakubo, and T. Yamamoto, From Humanoid Embodiment to Theory of Mind, in *Embodied Artificial Intelligence, LNAI 3139*, F. Iida et. al. Eds., Springer-Verlag Berlin Heidelberg, pp. 202-218, 2004.

[3] J. Albus, Toward a computational theory of mind, *Journal of Mind Theory*, Vol. 0, No. 1, pp. 1-38, 2008.

[4] S.Boza and R. H. Guerra, A First Approach to Artificial Cognitive Control System implementation Based on the Shared Circuits Model of Sociocognitive Capacities, *ICIC Express Letters*, Vol. 4, No. 5(B), pp. 4167-4176, Nov. 2010.

[5] H. Fukumoto and T. Toriu, Mechanism of Formation of Vision Based on Learning of Correlation between Sensation and Motion, *Int. J. Innovative Computing, Information and Control*, Vol. 5, No. 11(B), pp. 4167-4176, Nov. 2009.

[6] T. Toriu and F. Hirofumi, A Learning Method for Association between Vision and Ego-Motion which is Capable of Adapting to Arbitrary Image Distortion, *23rd Intl. Conf. on Image and Vision Computing*, New Zealand, Nov. 2008.

[7] T. Toriu, Thi Thi Zin, and H. Hama, Unsupervised Learning Algorithm for Self-Organizing Internal Representation of Ego-Motion, ICIC Express Letters, B, Vol. 2, pp. 559-564, 2011.

[8] V. V. Hafner and F. Kaplan, Interpersonal Maps and the Body Correspondence Problem, *Proceedings of the Third International Symposium on Imitation in animals and artefacts*, Y. Demiris K. Dautenhahn, and C. Nehaniv, Eds., Hatfield, UK, pp. 48-53, 2005.

[9] K. F. MacDorman, Grounding Symbols through Sensorimortor Integration, *JRSJ*, Vol. 17, No. 1, pp. 20-24, 1999.

[10] W. H. Warren and D. H. Hannon, Direction of selfmotion is perceived from optical flow, *Nature* Vol. 336, pp. 162-163, 1988.

[11] G. Adiv, Determining three-dimensional motion and structure from optical flow generated by several moving objects, *IEEE Trans. Patt. Anal. Mach. Intell*. Vol.7, pp. 384-401, 1985.

[12] G. M. Straqtton, "Vision without inversion of the retinal image," Psychol. Rev. 4, 314-360, 1897.

[13] Y. Sugita, "Global plasticity in adult visual cortex following reversal of visual input," Nature, 380, 523-526, 1996.

[14] Chr. vonder Malsburg, "Self-Organization of Orientation Sensitive Cells in the Striate Cortex", Kybernetic, Vol. 14, pp. 94-100, 1973.

[15] J. Sirosh and R. Miikkulainen, "Topographic Receptive Fields and Patterned Lateral Interaction in a Self-Organizing Model of the Primary Visual Cortex," Neural Computation, Vol. 9, No. 3, pp. 577-594, 1997.

[16] David J. Heeger, "Optical flow using spatiotemporal filters", International Journal of Computer Vision, Vol.1. No.4. pp279-302, 1988.