Detection of Abnormal Human Action Using Image Sequences

Hirokazu Seki and Yoichi Hori
Department of Electrical Engineering,
Faculty of Engineering, the University of Tokyo,
2-11-16 Yayoi, Bunkyo, Tokyo, 113-8656, Japan
Phone: +81-3-5841-7683, Fax: +81-3-5841-7687
E-Mail: `seki@hori.t.u-tokyo.ac.jp`

Abstract

We propose the system to learn aged people's usual actions automatically in about one week and to detect his/her abnormal action by observing the room with camera, because Japan will be a serious aging society rapidly. For the method of detection, we use 2-step eigenspace representation which is excellent in compression of image data and calculation of the correlation among images. At the learning stage, we use Self-Organizing Map(SOM) to realize automatic learning. To detect abnormalities such as the speed and direction of actions, we use Parametric Eigenspace Method(PEM). Some results show the effectiveness of the proposed method.

Key words: aged society, aged people monitoring system, motion analysis, eigenspace method, parametric eigenspace method, Self-organizing map

1 Introduction

Japan is now a serious aging society and this problem will be more serious in the near future. Therefore, engineering support for aged people is required. As one of such support, we propose the system to detect aged people's abnormal action by observing the room with camera. If he/she acts differently from usual behavior, the proposed system detects it, and immediately reports to his family or medical institution. Fig.1 shows an example of care system for aged people, in which the proposed system is useful.

First, the system automatically learns his/her normal actions and behavior pattern in about one week, after that, detects his/her abnormal actions based on the learned result. The proposed system has the following features:

(1) The system automatically learns actions without knowledge of the structure of his/her room such as position of the entrance, bed or table, and knowledge of the regular behavior pattern of human being.

(2) The system don't understand the meaning of his/her actions such as going to the bathroom, eating or sleeping on the bed. The system only classifies actions, for example, A, B, C, $\cdots$.

(3) The system occationally judges normal action abnormal by mistake, however it doesn't matter because it gives some opportunities to communicate with his/her family or medical institution.

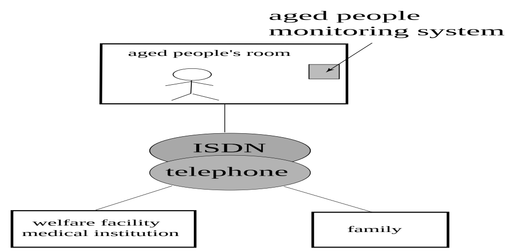


Fig. 1: Example of care system for aged people

If such system has been realized, it is also possible to apply to other systems, for example, observation of robots in the factory and traffic.

2 Analysis of human action using image sequence

2.1 The methods of human motion analysis

Many methods for action recognition from image sequence have been proposed, for example, DP matching[1], HMM[2], and Temporal Templates[3].

We want to realize the simpler and real time system using a CCD camera or a camera of TV telephone, therefore, we analyze image sequences using eigenspace representation without a reconstruction of 3D structure.

2.2 Classification of abnormalities detected with camera

Abnormalities of aged people's actions which can be detected by image information are defined as follows:

(Ab-1) different action from reference

If he/she acts differently from the learned action, the system should judge it abnormal, for example, falling down suddenly or struggling on the bed.

(Ab-2) speed of action

In general, aged people act slowly, therefore, if he/she acts very quickly or slowly, for example, slips off the step or other people enter the room, the system should judge it abnormal.

(Ab-3) direction of action

This is an important element in observation of robots in the factory and traffic.

(Ab-4) behavior pattern

For example, if he/she goes out at midnight or sleeps all day, the system should judge it abnormal. That is, the system should learn his/her usual behavior pattern and detect the different behavior pattern based on it.

After the learning stage, the system detects the above four abnormalities (Ab-1)~(Ab-4).

2.3 Extraction of the distance among images

In this paper, we consider the images of abnormal actions have low correlation to the learned images. To extract the distance between images, we use eigenspace method[4, 5] which is excellent in calculation of the correlation between images and compression of image data.

2.3.1 Eigenspace method

We will now mention the eigenspace method used both to classify actions and to detect abnormalities.

An normalized image data at time t is represented as $\boldsymbol{y}(t)$. The covariance matrix of image data set $\boldsymbol{y}(t)$ is represented by

$$\boldsymbol{Q} = \sum_{t=1}^{T} (\boldsymbol{y}(t) - \boldsymbol{c})(\boldsymbol{y}(t) - \boldsymbol{c})^T \tag{1}$$

where $\boldsymbol{c}$ is the mean vector for $\boldsymbol{y}(t)$. k eigenvectors $\boldsymbol{e}_1$,$\boldsymbol{e}_2$, $\cdots$,$\boldsymbol{e}_k$ $(\lambda_1 \geq \cdots \geq \lambda_k \geq \cdots \geq \lambda_K)$ are determined by solving eigenvalue problem:

$$\lambda_j \boldsymbol{e}_j = \boldsymbol{Q}\boldsymbol{e}_j \tag{2}$$

The k-dimensional subspace spanned by these eigenvectors is called the eigenspace. Then, one image vector is projected onto the eigenspace by

$$\boldsymbol{z}(t) = [\boldsymbol{e}_1, \cdots, \boldsymbol{e}_k]^T (\boldsymbol{y}(t) - \boldsymbol{c}) \tag{3}$$

The closer the projections are in the eigenspace, the more highly correlated are the images. Thus, we can detect abnormality based on the distance in the eigenspace.

2.3.2 Parametric Eigenspace Method(PEM)

An image can be mapped to a point in the eigenspace, therefore a sequential movement can be represented as a smooth locus in the eigenspace. This is called Parametric Eigenspace Method(PEM).

2.4 Abnormality detection process of proposed system

We use parametric eigenspace method both to classify actions and detect abnormal aged people's actions (Ab-1)~(Ab-4).

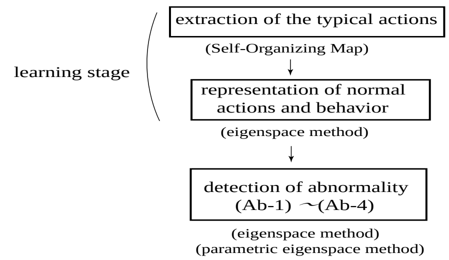


Fig. 2: Abnormality detection process

3 Learning method of normal action and behavior pattern

First, the system must learn his/her normal actions and behavior pattern automatically in about one week. This is an unsupervised learning.

When the eigenspace to classify actions is constituted from the whole image data obtained in about one week, the dimension of the covariance matrix $\boldsymbol{Q}$ is too large to calculate the eigenvectors. Even if it is possible, we must classify a large number of projections without knowing the number of class. To solve this problem, we use the Self-Organizing Map.

3.1 Extraction of typical actions using Self-Organizing Map

In the case of aged people monitoring system, his/her actions such as sleeping, eating, and watching TV may occur within the each fixed region in the image. That is, some similar images according to actions frequently appear, and we extract these images using SOM.

Self-Organizing Map(SOM)[6] is the unsupervised, competitive learning method, which maps the high dimensional input data onto the low dimensional network of cells keeping the topological neighborhood relation.

At each learning step, the "winner" cell c for which matches best with input $\boldsymbol{x}(t)$ is found by

$$\|\boldsymbol{x}(t) - \boldsymbol{w}_c\| = \min_i \|\boldsymbol{x}(t) - \boldsymbol{w}_i\| \tag{4}$$

where $\boldsymbol{w}_i$ is the weight vector of cell i whose dimension is equal to that of the input vector $\boldsymbol{x}(t)$.

Next, all the cells within a neighborhood set $N_c(t)$ which is centered around cell c are updated by

$$\boldsymbol{w}_i := \begin{cases} \boldsymbol{w}_i + \alpha(t)(\boldsymbol{x}(t) - \boldsymbol{w}_i) & \text{if } c \in N_c(t) \\ \boldsymbol{w}_i & \text{if } c \notin N_c(t) \end{cases} \quad (5)$$

where $\alpha(t)$ is a monotonically decreasing scalar-valued gain $0 < \alpha(t) < 1$, and $N_c(t)$ shrinks monotonically with time. We use the two-dimensional network because of visual understanding.

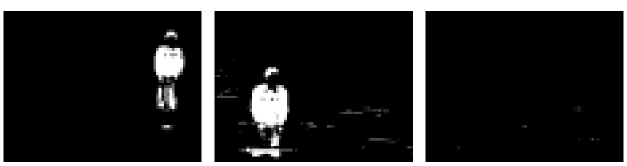
To classify these cells, we use the cluster map[7]. Distance $d(i,j)$ is a measure of the similarity to the neighboring cells.

$$d(i,j) = \frac{1}{|D(i,j)|} \sum_{(u,v)\in D(i,j)} (\boldsymbol{w}_{i,j} - \boldsymbol{w}_{u,v})^T (\boldsymbol{w}_{i,j} - \boldsymbol{w}_{u,v}) \quad (6)$$

where $D(i,j)$ is the neighborhood set of cell (i,j). Thus, by labeling the cells which have small-valued $d(i,j)$, the sets of cell having the similar weight vectors each other are extracted as clusters.

When the learning data contain many similar data, the neighborhood weight vectors of SOM come very similar. On the other hand, when the data having few similar data are learned, the neighborhood weight vectors of SOM come to very different.

We experiment with 240 images(Fig.3) in which a human frequently stays in the right corner(A) and the left corner(B), and goes outside the image(C). After repeated learning with these images, $d(i,j)$ of the two-dimentional map and clustering result by labeling in ascending order of $d(i,j)$ are obtained as Fig.4, Fig.5.



A B C

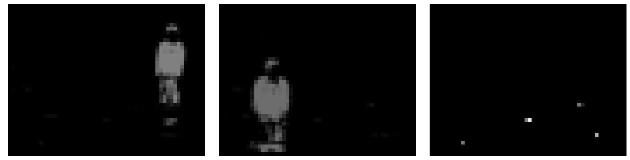
Fig. 3: Three kinds of typical images

Fig. 4: Cluster map

			2	2	2		
			2	2			
1			2	2			
1							
1							
		3	3	3	3	3	3

Fig. 5: Labeling

Fig.6 show the weight vectors of cells (8,8), (1,4) and (4,1) which have small-valued $d(i,j)$ respectively.



(a) cell (8,8) (b) cell (1,4) (c) cell (4,1)

Fig. 6: Weight vectors

3.2 Classification of action using eigenspace method

Fig.7 shows the eigenspace constituted from the sixteen weight vectors labeled A, B and C. Thus, we can constitute the eigenspace in which three kinds of images appearing frequently are classified properly.

Each projected point onto the eigenspace has the label(A,B,C) given at clustering on the cluster map.

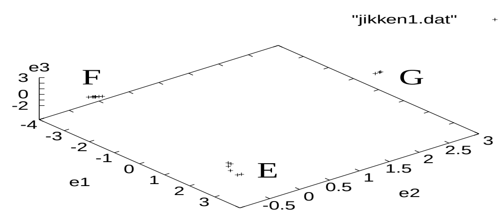


Fig. 7: Three-dimensional Eigenspace constituted from the 16 weight vectors

An input image $\boldsymbol{y}(t)$ is projected onto this eigenspace by

$$\boldsymbol{z}(t) = [\boldsymbol{e}_1, \cdots, \boldsymbol{e}_k]^T(\boldsymbol{y}(t) - \boldsymbol{c}) \quad (7)$$

Fig.8 shows the projections of all the reference image data onto the eigenspace Fig.7.

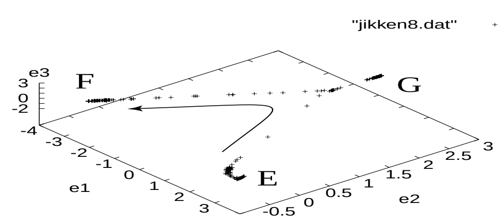


Fig. 8: Projections of all the reference image data onto the eigenspace Fig.8

The reference point closest to the projected point $\boldsymbol{z}(t)$ is found by

$$d_1 = \min_i \|\boldsymbol{z}(t) - \boldsymbol{f}_i\| \quad (8)$$

Thus, an input image is classified using the label of the closest reference point $\boldsymbol{f}_i$, we can recognize his action as A,B,C,$\cdots$.

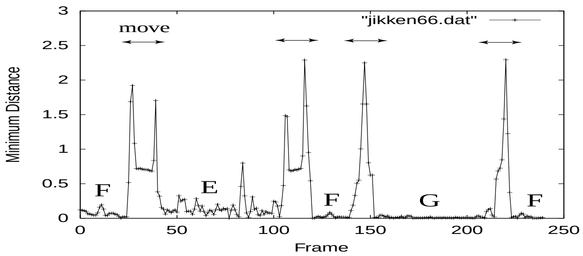


Fig. 9: Minimum Distance

When the minimum distances are large, human moves in the room. For example, in the Fig.9, he moves from B to A in $t = 26 \sim 40$, when the minimum distances are large. Similarly, we recognize his movement from A to B in $t = 105 \sim 119$, from B to C in $t = 142 \sim 151$ and from C to B in $t = 215 \sim 222$.

3.3 2-step eigenspace method

We want to separately analize his actions such as sleeping, eating or moving. Therefore, we constitute the several eigenspace according to the actions.

In addition, eigenspace method uses the global information of images, therefore, very little local information in the image is reflected in the eigenspace.

For these reasons, we use 2-step eigenspace method as shown in Fig.10. First, the "universal" eigenspace which can classify actions is constituted, and the input image is classified according to the action A, B, C, $\cdots$. Next, the sequensial data in the local image region in which the action occurs are analized in the eigenspace of the corresponding action.

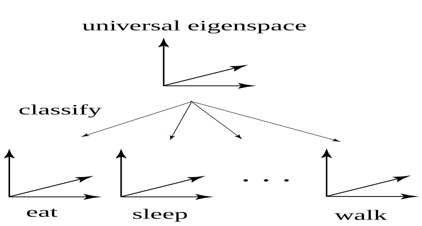


Fig. 10: 2-step eigenspace method

That is, the universal eigenspace and several eigenspaces according to actions are constituted in the learning of normal actions and behavior pattern.

3.4 Determination of the local regions to be focused

To constitute the eigenspaces according to actions, we determine the local regions to be focused using weight vectors of SOM. The human region should be centered in the focusing region, and so we calculate the center pixel of the human region and gradually widen the window surrounding the human. Fig.12 shows the focusing region determined from Fig.11.

Fig. 11: weight vector of cell (8,8)

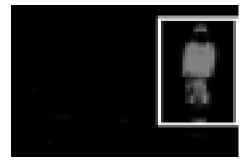


Fig. 12: region to be focused

3.5 Representation of normal action and behavior pattern

We represent normal aged people's actions as the set of projection point in the eigenspace. The details are as follows.

[**The actions at the same place**]

$\rightarrow$ the projected points in the eigenspace according to actions

[**The moving action**]

$\rightarrow$ the locus in the eigenspace accoding to actions (Parametric Eigenspace Method)

[**Behavior pattern**]

$\rightarrow$ the time label appended to the projected points in the universal eigenspace

4 Detection of abnormal action using eigenspace method

After the learning of normal actions, abnormality detection using the 2-step eigenspace method starts.

4.1 Detection of different actions from reference (Ab-1)

As stated above, the closeness in the eigenspace represents the correlation among images, therefore, we can detect the different action from reference as lower correlation images by the distance in the eigenspace. We search the reference point closest to the projected point $\boldsymbol{z}^{(j)}(t)$ of an input image onto the eigenspace of the jth action by

$$d_2^2 = \min_i \|\boldsymbol{z}^{(j)}(t) - \boldsymbol{f}_i^{(j)}\|^2 \tag{9}$$

where $\boldsymbol{f}_i^{(j)}$ are the reference points in the eigenspace of the jth action. If d_2 is larger than the threshold, the system judges the action abnormal.

When the eigenspace is constituted from the reference images(30 frames) sitting on a chair such as Fig.13, the input image sequence(30 frames, fall down from $t = 10$) such as Fig.14($t = 15$) and Fig.15($t = 19$) are represented in the eigenspace as Fig.15.

Fig. 13: Reference pattern

Fig. 14: Input pattern($t = 15$)

Fig. 15: Input pattern($t = 19$)

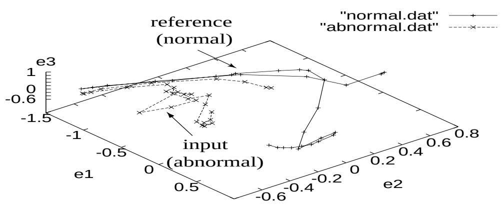


Fig. 16: The locus in three-dimensional eigenspace (reference pattern and input pattern)

Fig.17 shows the minimum distances in three-dimensional eigenspace searched by (9). The minimum distance comes to be large after $t = 10$ when the human collapses from the chair. Thus, lower correlation images of abnormal action can be detected by the distance.

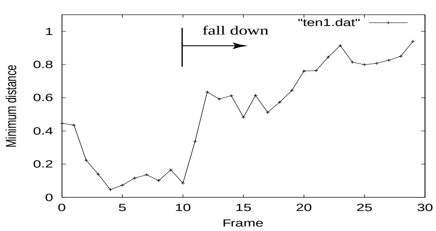


Fig. 17: The minimum distance in three-dimensional eigenspace

4.2 Detection of the speed and direction of action (Ab-2,Ab-3)

To detect the speed and direction of action, we compare betweeen the loci by (10) using Parametric Eigenspace Method(PEM)[4, 5]. The locus closest to that of the input image sequence $\boldsymbol{z}^{(j)}(t)$ is found by

$$d_3^2 = \min_{a,b} \sum_{t=1}^{T} \|\boldsymbol{z}^{(j)}(t) - \boldsymbol{f}^{(j)}(at+b)\|^2 \qquad (10)$$

$$\tilde{a} = \arg\min_{a} \sum_{t=1}^{T} \|\boldsymbol{z}^{(j)}(t) - \boldsymbol{f}^{(j)}(at+b)\|^2 \qquad (11)$$

where a is the time stretch factor, and b is the time shift factor. Then, the absolute value of $\tilde{a}$ represents the speed of action and the sign of $\tilde{a}$ represents the direction of action. That is,

if	$\tilde{a} < -1$,	opposite direction,	quickly
if	$-1 \le \tilde{a} < 0$,	opposite direction,	slowly
if	$0 \le \tilde{a} < 1$,	same direction,	slowly
if	$1 \le \tilde{a}$,	same direction,	quickly

Fig.18 shows one of the reference images(30 frames) walking from the right side to the left side, and Fig.19 shows the locus in the eigenspace. Two kinds of image sequences(14 frames) are inputed. They are (I)walk more quickly in the same direction, (II)walk more slowly in the opposite direction. Each locus in the eigenspace is shown in Fig.20.

Fig. 18: Example of reference image

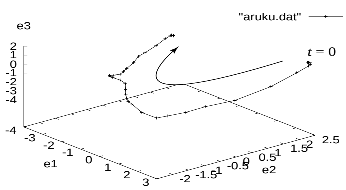
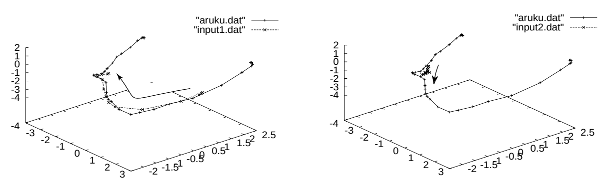


Fig. 19: The locus in three-dimensional eigenspace (reference)



(I) quickly, same direction

(II) slowly, opposite direction

Fig. 20: The locus in three-dimensional eigenspace

$\tilde{a}$ and d_3 when the closest locus is searched using (10) are given as follows. The dimensions of the eigenspace are $3 \sim 5$. In the case of (I), $1 \le \tilde{a}$. In the case of (II), $-1 \le \tilde{a} < 0$. This result shows the speed and direction of action are estimated by $\tilde{a}$.

	(I)			(II)		
k	$\tilde{a}$	$\tilde{b}$	d_3	$\tilde{a}$	$\tilde{b}$	d_3
3	1.39	6.57	3.20	-0.17	27.26	1.15
4	1.45	6.44	3.55	-0.33	27.16	2.10
5	1.48	6.34	3.93	-0.30	27.12	2.62

4.3 Detection of abnormal behavior pattern (Ab-4)

To detect abnormal behavior pattern such as going out in the midnight or sleeping all day, we define the new distance by the equation (12) made of the distance in the universal eigenspace d_e and the difference d_t between the learning time and the input time.

$$d_4 = \min_i \sqrt{c_1 d_e^2 + c_2 d_t^2} \quad (12)$$

$$d_e^2 = \|\boldsymbol{z}(t) - \boldsymbol{f}_i\|^2, \quad d_t^2 = (t_i - t)^2 \quad (13)$$

where c_1 and c_2 are the weight to d_e^2 and d_t^2, and t_i is the time appended to the projected points in the universal eigenspace. Fig.21 shows the learned behavior pattern.

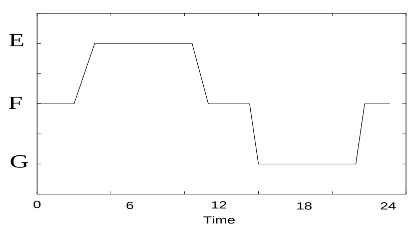


Fig. 21: Reference behavior pattern

If he stays in A all day, the distances d_4 ($c_1 = 1, c_2 = 0.3$) are given as Fig.22. At the time he should be in B or C, d_4 come to be large. Thus, we can detect abnormal behavior pattern using this new distance.

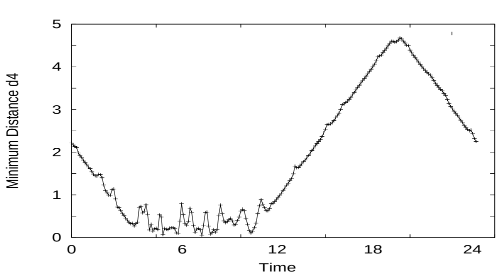


Fig. 22: The distance d_4 of equation (12)

5 Conclusion

In this paper, we have proposed the abnormality detection method for the aged people monitoring system. We have applied eigenspace method which is excellent in calculation of the correlation among images both to classification of actions and abnormality detection.

We realized automatic learning by constituting the universal eigenspace using Self-Organizing Map. To avoid the explosion of calculation and clustering, we extract the images appearing frequently using SOM, and constitute the universal eigenspace in which we can fully classify actions. Next, a human action is analized in each eigenspace according to the action using eigenspace method. Sequential images of the action are represented by a locus in the eigenspace. We can detect the different action based on the distance of projected points onto the eigenspace, and the speed and direction of action by the introduced two parameters to characterize the loci using Parametric Eigenspace Method.

Some simulation results show the effectiveness of the proposed method. The proposed system is developed mainly for monitoring aged people, however, it can be applied to various systems, for example, observation of inpatients in hospital, traffic, robots in the factory, and security monitoring in buildings and stores.

References

[1] T. Darrell and A. Pentland, "Space-Time Gesture", In *CVPR'93*, pp.355-340 (1993).

[2] M. Brand, M. Oliver, and A. Pentland, "Coupled hidden markov models for complex action recognition", In *CVPR'97*, pp.994-999 (1997).

[3] J. W. Davis and A. F. Bobick,"The Representation and Recognition of Human Movement Using Temporal Templates", In *CVPR'97*, pp.928-934 (1997).

[4] H. Murase and S. K. Nayar, "Visual Learning and Recognition of 3-D Objects from Appearance", In *IJCV'95*, Vol. 14, pp.5-24 (1995).

[5] H. Murase and R. Sakai, "Moving object recognition in eigenspace representation:Gait analysis and lip reading", *Pattern Recognition Letters*, Vol. 17, pp.155-162 (1996).

[6] T. Kohonen, "The Self-Organizing Map", *Proc. IEEE*, Vol. 78, No. 9, pp.1464-1480 (1990).

[7] M. Tanaka, Y. Furukawa and T. Tanino, "Clustering by Using Self Organizing Map (in Japanese)", *IEICE Transactions in Japanese*, (D-II), Vol. J79-D-II, No. 2, pp.301-304 (1996).