

Comparative study of path normalizations for path prediction

Yuji shinomura † Tomotaka Harano † Toru Tamaki †
Toshiyuki Amano ‡ Kazufumi Kaneda † Seiichi Uchida ††

†:Hiroshima University,
shino@eml.hiroshima-u.ac.jp
hara@eml.hiroshima-u.ac.jp
tamaki@hiroshima-u.ac.jp
kin@hiroshima-u.ac.jp
‡:NAIST, amano@is.naist.jp
††:Kyushu University, uchida@is.kyushu-u.ac.jp

Abstract This paper discusses two methods on normalization of a sample path for predicting paths of a pedestrian by using eigenspace. A path of a person is defined as a sequence of successive coordinates of the person over frames and represented the path as a vector with $2M$ elements of M number of coordinates. A problem of their prediction is that the method is based on subspace. To make a subspace from sample paths, all paths need to be normalized and resampled such that a path vector has the same number of elements. Because different sample paths have different number of frames. In this paper, we apply a normalization method using DP matching and discuss results of two predictions: resampling and DP.

1 Introduction

For the development of surveillance camera system and study on human behavior, human behavior recognition in movie are studied in computer vision. Especially, in area of human recognition and tracking, there are various method, such as robust feature for human recognition [1], particle filter [6], 3D tracking using multiple stereo cameras [2], classification based on learning data [10]. What is the next step? It is Prediction of human behavior. It is useful not only for a fast search in human recognition but also for preprocessing for human behavior. In general, prediction is done based on observed, for example, Kalman filter and AR model as linear model. However, these model are not appropriate for non-linear model like human action that the environment restricts human action.

To consider the environment of human action, prediction methods based on learning sample human action have been developed. Nakajima et al. [5] and Mori et al. [4] proposed motion prediction based on Eigen-gestures. Yamamoto et al. [11] proposed research the prediction of human walking path predicts human walking path in the future (Fig. 1b). A human walking path is defined as a sequence of successive coordinates of the person over

frames and represented as a vector with $2M$ elements of M number of coordinates (Fig. 1). First, this method makes eigenspace from learning sample paths. Next, the future walking path is predicted from the eigenspace. This method attempts to predict based on learning human walking pattern under the environment such as buildings, staircases, and entrances.

Fig. 1b shows a result of human walking path prediction by [11]. Actual path (white line) does not correspond to prediction path (black line). This result is does not good. Reasons may be normalizing learning path, lack of eigenvectors, and small the number of learning path, but the effect of these cause have not investigated yet.

In this paper, we investigate the effect of the normalization. [11] normalized learning paths by resampling but this normalization doesn't consider the influence of nonlinear walking velocity between frames. To consider this influence, we apply normalization method using DP (Dynamic Programming) and compare the difference of results of two predictions resampling and DP.

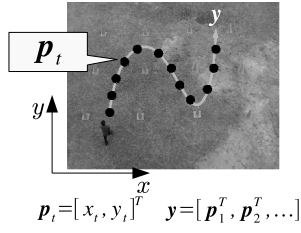


Fig. 1: Definition of a walking path.

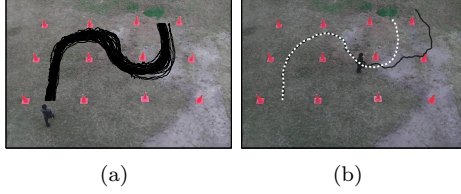


Fig. 2: (a) Learning and (b) Prediction. White line is an actual path. Black line is a prediction path.

2 Predicting Paths of a Pedestrian by using Eigenspace

We explain about the prediction method based on Eigenspace [9, 11]. This method has 2 processes (learning and prediction).

2.1 Making Eigenspace based on learning sample path

At learning process, we learn N number of walking paths of a the same course (Fig. 1a). But, different sample paths have different number of frames. Yamamoto et al. [11] normalized sample paths to make vectors with the same number of elements. The paths are normalized in length defined as a sum of Euclidean distances between two successive coordinates. Fig. 3 shows the following normalization. First, coordinates of each sample path is sparsely downsampled with linear interpolation to reduce noise (Fig. 3b). Second, all paths are cut to the shortest length (Fig. 3c), and resampled so that all paths have the same length and M number of coordinates (Fig. 3d).

After this normalization, N normalized paths make eigenspace E_N . We subtract $\mathbf{m} = \frac{1}{N} \sum_{i=1}^N \mathbf{y}_i$ from learning paths \mathbf{y}_i and get N eigenvectors \mathbf{e}_i . here, E_N is a matrix that stores eigenvectors in its columns.

$$E_N = [\mathbf{e}_1, \dots, \mathbf{e}_N] \quad (1)$$

\mathbf{e}_i is a $2M$ dimensional vector as

$$\mathbf{e}_i = [\mathbf{e}_{i1}^T, \mathbf{e}_{i2}^T, \dots, \mathbf{e}_{iM}^T]^T \in \mathbf{R}^{2M}. \quad (2)$$

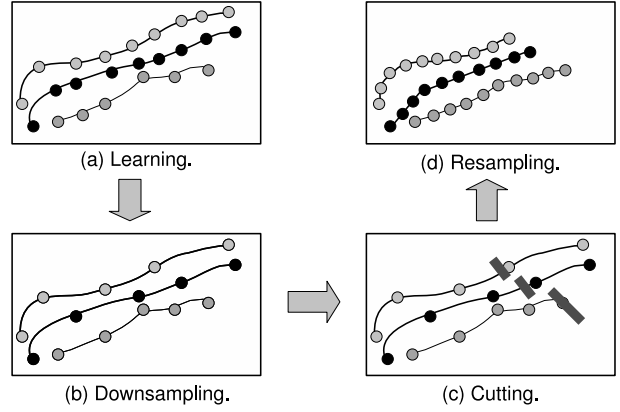


Fig. 3: Normalization.

2.2 Prediction based on Eigenspace

At prediction process, we track new pedestrian and predict the future walking path. The pedestrian is tracked from 1 frame to t th frame. This tracked path has the t number of observed coordinates. This path is normalized by the same normalization used for learning sample paths. After normalization, the tracked path is resampled as s coordinates and we call this path \mathbf{y}'

$$\mathbf{y}' = (\mathbf{p}_1^T, \dots, \mathbf{p}_s^T)^T \in \mathbf{R}^{2s}, \quad s \leq M. \quad (3)$$

Here, the path between $s+1$ th coordinate \mathbf{p}_{s+1} and M th coordinate \mathbf{p}_M is not observed. Hence, we set $\mathbf{p}_{s'} = \mathbf{0} = (0, 0)^T$, (for $s' = s+1, s+2, \dots, M$), and define \mathbf{y}'' as the path of $2M$ dimension.

$$\begin{aligned} \mathbf{y}'' &= (\mathbf{p}_1^T, \dots, \mathbf{p}_s^T, \underbrace{\mathbf{0}^T, \dots, \mathbf{0}^T}_{(M-s)})^T \\ &= (\mathbf{y}'^T, \underbrace{\mathbf{0}^T, \dots, \mathbf{0}^T}_{(M-s)})^T \in \mathbf{R}^{2M}. \end{aligned} \quad (4)$$

Then \mathbf{y}'' is projected onto the eigenspace as point \mathbf{a} , and \mathbf{a} is represented as follows

$$\mathbf{a} = E'^T \mathbf{y}'', \quad E' = \text{diag}(\underbrace{1, \dots, 1}_{2s}, \underbrace{0, \dots, 0}_{2(M-s)}) E \quad (5)$$

From property of eigenspace E' in Eq.(5). Therefore,

$$E'^T E' \mathbf{a} = E'^T \mathbf{y}'' \quad (6)$$

$$\mathbf{y}^* = E(E'^T E')^{-1} E'^T \mathbf{y}'' \quad (7)$$

Here, we say \mathbf{y}^* as the predicted path. Then, \mathbf{y}^* is an inverse projection of \mathbf{a} .

Note that $\text{rank}(E'^T E') = N$ or $\det(E'^T E') \neq 0$ should be held so that the linear system doesn't become underdetermined. This means $2s > N$, hence

the prediction can be done after several positions of a pedestrian are observed.

3 Normalization of sample path by resampling

Yamamoto et al. [11] supposed that a pedestrian walks at constant velocity in a movie. Thus, it is supposed that the coordinates of path in the image are plotted as the same distance between two successive coordinates. However, the distances between two successive coordinates are not same due to the differences of walking velocity and the error of human position by backgroundsubtraction.

The advantage of normalization by resampling is to exclude high-frequency component in walking path. Thus, rough shapes of learning sample paths in the image are learned.

On the other hand, the resampling has the disadvantages. First, learning sample paths for making eigenspace are different from observed paths because normalized paths are made by interpolation and resampling. Second, this method does not consider the influence of nonlinear relationship of walking velocity between frames. Hence, we need to consider above problems and apply a new method of normalization that considers the nonlinear relationship.

4 Normalization based on DP matching algorithm

To overcome the problem above, we apply DP matching algorithm [7, 8, 3] for normalization. DP matching makes correspondence between input and reference patterns so as to minimize the error of the patterns. In this paper, DP matching algorithm cancels the nonlinear time difference among sample paths and compensates the nonlinear shape difference.

First, input pattern is defined as $X = \{\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_I\}$ and reference pattern is defined as $Y = \{\mathbf{y}_1, \dots, \mathbf{y}_j, \dots, \mathbf{y}_J\}$. We can consider these patterns as walking paths (Fig. 4a). DP matching optimizes so that \mathbf{y}_j corresponds to \mathbf{x}_i which minimizes the cost function $J = u_j (i = 1, \dots, I)$. Euclidean distance between \mathbf{y}_j and \mathbf{x}_i is set follows:

$$d_i(u_i) = \|\mathbf{x}_i - \mathbf{y}_j\| \quad (8)$$

and, we set a cost function F

$$\min_{u_1, u_2, \dots, u_I} F = \sum_{i=1}^I d_i(u_i) \quad (9)$$

$$\text{subject to } 0 \leq u_i - u_{i-1} \leq 1 \quad (10)$$

$$u_i = 1, u_I = J \quad (11)$$

Eq.(10) is a limit condition to assure $X \geq Y$ (Fig. 4), and Eq.(11) is a boundary condition of start and end positions.

The advantage of normalization by DP matching algorithm is that we choose coordinates from observed path to make normalized path.

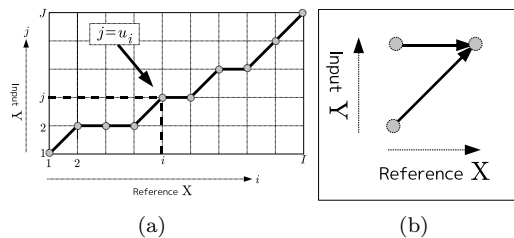


Fig. 4: (a)DP matching. (b) limit condition.

5 Experiment of comparison with different predictions

Now, we show results of prediction experiments to compare normalizations by resampling and normalization by DP.

5.1 Experimental setting

We explain an environment of experiments how to get learning sample paths. In the experiment, a video camera was fixed to a tripod, from a building, we recorded a pedestrian walking path in a movie in mpeg format (640×480 pixel). We implemented the proposed method, and evaluated using the real image sequences. Fig. 5 shows the walking course. Then 30 paths were obtained by off-line process. A walking path is obtained by Backgroundsubtraction and the center of gravity is extracted as the coordinates.

5.2 Normalization by resampling

We applied the normalization by resampling for learning sample paths. Each path was downsampled to 50 coordinates and resampled to 300 coordinates. In this experiment, the length of paths was 603.192, and the distance between two coordinates in normalized paths was 2.017. A tracked path was resampled so the distance become 2.017.

5.3 Normalization by DP matching

We applied the normalization by DP for sample paths. First, we chose the longest path as a reference pattern (548 coordinates). DP was applied for

the other 29 paths as input pattern. Also, DP was applied to a tracked path.

5.4 Prediction results based on two normalization

We compared prediction results based on each eigenspace by making each normalization. Two prediction results were compared at the same frame number: 50, 150, 250, 300 th frame. Fig. 10 shows the prediction result based on resampling, and Fig. 10 shows the prediction result based on DP.

First, we compared paths predicted at different time. The prediction error of prediction by resampling is large in the early frames (Fig. 10a, 10b). But, the more time passed, the closer the prediction approaches the actual path (Fig. 10c, 10d). The cause of error in early frames is the lack of observed coordinates. 78 dimension is known in Fig. 10a and 240 dimension is known in Fig. 10b in 600 dimensions of the eigenspace. These observed dimension is not enough to predict. The predictions by DP (Fig. 11), in early frames are closer to actual path than normalization by resampling. Thus, this method can predict the feature of learning sample paths from eigenspace even the lack of observed coordinates.

Next, we compared the shape of prediction path. The shape of prediction path by resampling are similar to actual path in Fig. 10 by resampling and prediction path were smooth shape (Fig. 10c, 10d) so we consider that the eigenspace learned smooth of path as the feature of learning sample path. On one hand, the shape of prediction path by DP is not smooth. Because normalization by DP did not remove the noise by downsampling and resampling, eigenspace learned the noise of the paths.

5.5 Quantitative evaluation

We evaluated quantitatively prediction paths. We define evaluation function SSD (Sum of squared difference between prediction path and actual path each frames which is not observed) and $Ave\ of\ SSD$ (average of SSD). \mathbf{p}_i^* is defined as i th coordinates

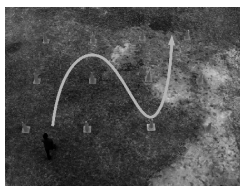


Fig. 5: Walking course.

in the prediction path \mathbf{y}^* .

$$SSD = \sum_{i=M-s}^M \|\mathbf{p}_i^* - \mathbf{p}_i\|^2 \quad (12)$$

$$Ave\ of\ SSD = \frac{SSD}{M-s} \quad (13)$$

Fig. 6 shows SSD of prediction by resampling. Fig. 7 shows SSD of prediction by DP. SSD by DP is smaller than by resampling at each frame. However, SSD by resampling reduces in proportion to increasing the frame number, whereas SSD by DP does not decrease monotonically.

Next, discuss $Ave\ of\ SSD$. Fig. 8 shows $Ave\ of\ SSD$ by resampling. Fig. 9 shows $Ave\ of\ SSD$ by DP. $Ave\ of\ SSD$ by resampling is smaller than by DP at each frame. However, $Ave\ of\ SSD$ by DP does not decrease monotonically. From two quantitative evaluations, we conformed the DP matching is more effective than resampling in SSD and $Ave\ of\ SSD$.

6 Conclusions

In this paper, we applied DP matching algorithm for normalization of learning sample paths and evaluated normalization methods by resampling and DP. We evaluated qualitatively shape of prediction paths and quantitatively SSD and $Ave\ of\ SSD$ of predicted paths. Evaluation indicates that DP method is more proper normalization method than resampling, but SSD by DP does not decrease monotonically so we need to investigate the regulation of effect by DP. Furthermore, we will investigate normalization at the same number of normalized coordinates, and the influence of reducing the number of eigenvectors.

References

- [1] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. *Proc. of CVPR2005*, pp. 886–893, 2005.
- [2] Y. Iwashita, R. Kurazume, T. Tsuji, K. Hara, and T. Hasegawa. 3d tracking of multiple moving objects using fast level set method. *Journal of the Robotics Society of Japan*, Vol. 23, No. 7, pp. 813–820, 2005.
- [3] H. Mitoma, S. Uchida, and H. Sakoe. Online character recognition based on elastic matching and quadratic discrimination. *Proc. Int. Conf. Doc. Anal. Recog.*, Vol. 1, pp. 36–40, 2005.

- [4] Akihiro Mori, Seiichi Uchida, Ryo Kurazume, Rin ichiro Taniguchi, Tsutomu Hasegawa, and Hiroaki Sakoe. Early recognition and prediction of gestures. *Proc. of ICPR2006*, Vol. 3, pp. 560–563, 2006.
- [5] Masato Nakajima, Seiichi Uchida, Akihiro Mori, Ryo Kurazume, Rin-ichiro Taniguchi, Tsutomu Hasegawa, and Hiroaki Sakoe. Motion prediction based on eigen-gestures. *Proc. of the 1st First Korea-Japan Joint Workshop on Pattern Recognition (KJPR2006)*, pp. 61–66, 2006.
- [6] K. Nummiaro, E. Koller-Meier, and L. Van Gool. An adaptive color-based particle filter. *Journal of Image and Vision Computing*, Vol. 21, No. 1, pp. 99–100, 2003.
- [7] H. Sakoe and S. Chiba. A dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. on Acoust. Speech and Sinal Proc.*, Vol. ASPP-26, pp. 43–49, 1978.
- [8] S. Seki, K. Takahashi, and R. Oka. Gesture recognition from motion image by spotting algorithm. *Proc. of ACCV1993*, Vol. 2, pp. 759–762, 1993.
- [9] Y. Shinomura, T. Tamaki, T. Amano, and K. Kaneda. Upgrading eigenspace-based prediction using null space and its application to path prediction. *Proc. of Subspace2007*, pp. 24–31, 2007.
- [10] X. Wang, K. Tieu, and E. Grimson. Learning semantic scene models by trajectory analysis. *Proc. of ECCV2006*, Vol. 3, No. 110-123, 2006.
- [11] R. Yamamoto, T. Tamaki, and M. Yamamoto. Predicting paths of a pedestrian by using eigenspace. *Proc. of Vision Engineering Workshop*, pp. 146–152, 2004. (in Japanese).

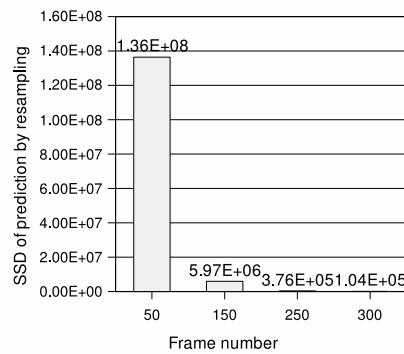


Fig. 6: SSD of Prediction by resampling .

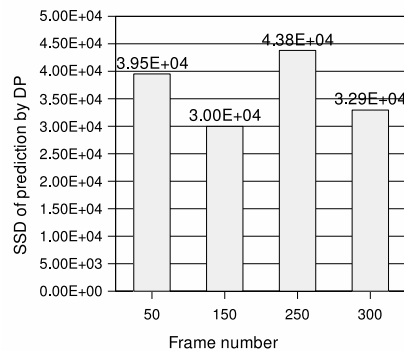


Fig. 7: SSD of Prediction by DP .

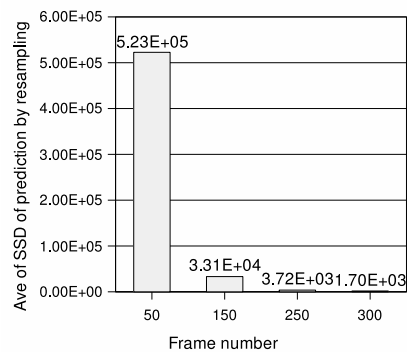


Fig. 8: Ave of SSD of Prediction by resampling .

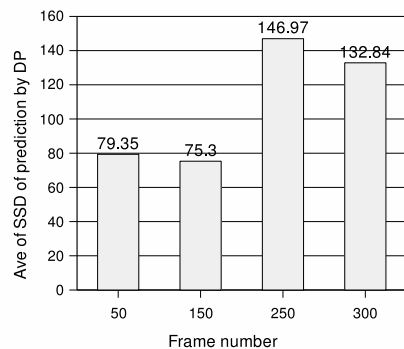
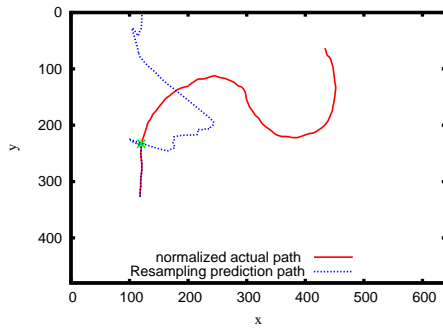
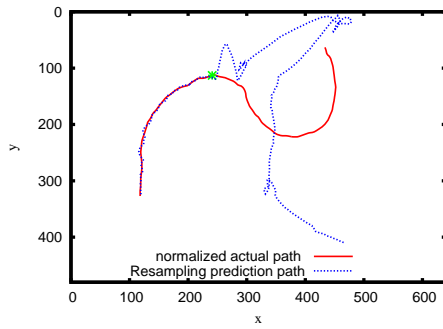


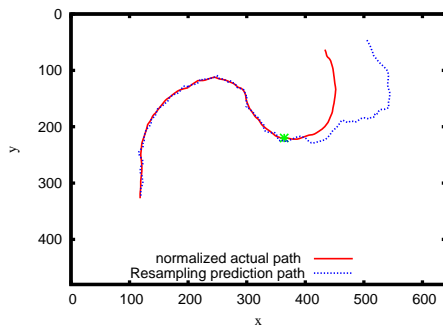
Fig. 9: Ave of SSD of Prediction by DP .



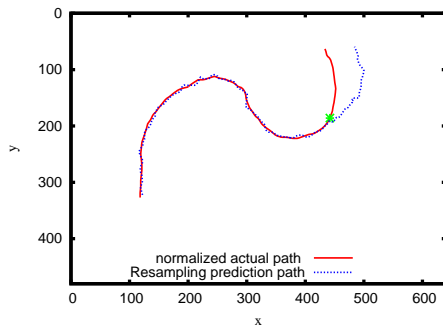
(a) 50th : 39 coordinates are known



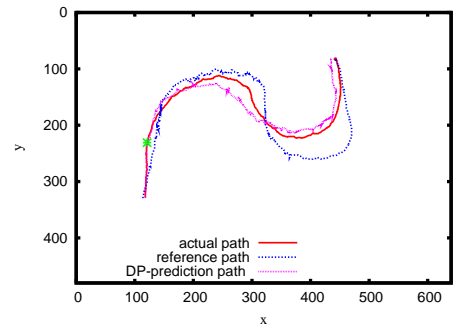
(b) 150th : 120 coordinates are known



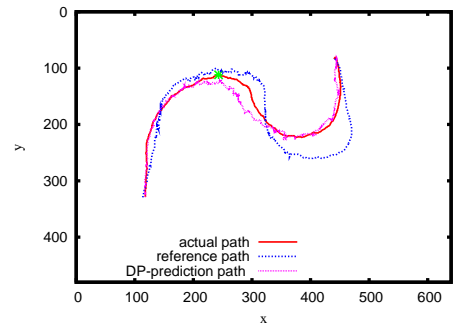
(c) 250th : 199 coordinates are known



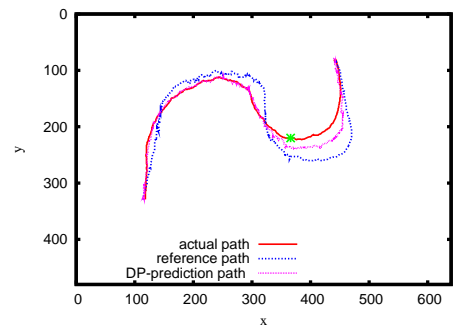
(d) 300th : 239 coordinates are known



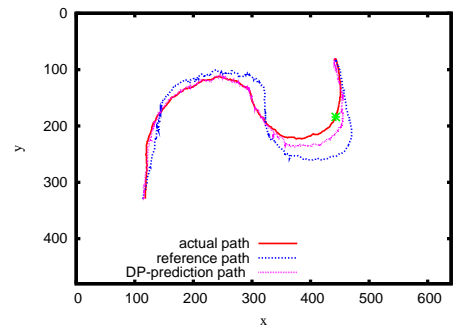
(e) 50th : 50 coordinates are known



(f) 150th : 150 coordinates are known



(g) 250th : 250 coordinates are known



(h) 300th : 300 coordinates are known

Fig. 10: Prediction result based on normalization by resampling. (a), (b), (c), (d) at each frame. * is the current person position.

Fig. 11: Prediction result based on normalization by resampling. (a), (b), (c), (d) at each frame. * is the current person position.