Which Reference View is Effective for Gait Identification Using a View Transformation Model?

Yasushi Makihara   Ryusuke Sagawa   Yasuhiro Mukaigawa   Tomio Echigo
Yasushi Yagi
Department of Intelligent Media, The Institute of Scientific and Industrial Research, Osaka University
567-0047, 8-1 Mihogaoka, Ibaraki, Osaka, JAPAN
{makihara, sagawa, mukaigaw, echigo, yagi}@am.sanken.osaka-u.ac.jp

Abstract

Gait identification is a promising method of individual identification at a distance from a camera and identification of those who observed from various views or those who going to various directions is required in particular for actual use. In this paper, we discuss a selection of reference views for the various-view gait identification using a view transformation model (VTM). In the gait identification process, we first extract frequency-domain gait features from gait silhouette sequences, and then obtain the various-view gait features by transforming a few reference features with the VTM. We made experiments using 736 sequences from 20 subjects of 24 view directions. We evaluate the performance for each single reference and for each combination of two references. In addition, we inspect the relation between the performance and the number of references.

1. Introduction

Gait identification has recently gained considerable attention because gait is a promising cue for surveillance systems to ascertain identity at a distance from a camera. Many approaches of gait recognition are proposed as model-based ones [15][13][12][1] and appearance-based ones[9][2][5][8], however, most of these approaches are view-dependent and limited to near fronto-parallel views.

Yu et al. [16] discussed the effects of view angle variation on gait identification and reported the performance drop when view difference was large. However, they evaluate the performance without view transformation, that is, they directly match a gallery (training) set and a probe (test) set from different views.

To cope with the view changes, Han et al. [3] used overlapped range of walking views for two different-view sequences of straight-walk. Kale et al.[4] proposed a method with perspective projection of a sagittal plane. However, these two methods does not works well when view difference is large. Spencer et al.[11] proposed reconstruction of articulated motion under the canonical (side) view. However, model-based identification methods sometimes suffer from mis-correspondence of feature points. Shakhnarovich et al. [10] proposed a visual hull-based method, the methods, however, needs multiple-view synchronized images for all subjects.

Makihara et al. [6] extended a view transformation model (VTM) [14] to the frequency domain and showed that various-view gait identification was achieved using a few reference views. However, they did not discuss which reference view is effective on gait identification performance.

Therefore, we discuss the selection of reference views for the various-view gait identification using the VTM. Concretely speaking, we discuss the following three issues.

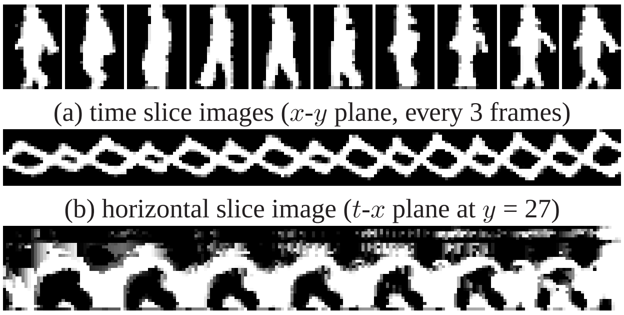
1. Which reference view is effective when a single reference is used?
2. Which combination is effective when two references are used?
3. How many references are necessary to obtain enough performance?

The outline of this paper is as follows. First, we describe extraction and matching of gait feature in the frequency domain in section 2. Next, adaptation to view direction changes is addressed with the formulation of the VTM in section 3. Finally, we present experimental results and analyses of reference views for gait identification using the VTM in section 4, and give our conclusions and future works in section 5.

2. Extraction and matching of gait feature

2.1. Construction of gait silhouette volume

The first step is constructing a gait silhouette volume (GSV). First, gait silhouettes are simply extracted by back-



(a) time slice images (x-y plane, every 3 frames)

(b) horizontal slice image (t-x plane at y = 27)

(c) vertical slice image (t-y plane at x = 6)

Figure 1. An example of GSV

ground subtraction of temperature images captured by a infrared-ray camera. Second, we obtain the height and the center of a silhouette region for each frame. Third, we scale the silhouette so that the height can be just 30 pixels, and so that the aspect ratio of each region can be kept. Forth, we register the silhouettes so that its center can correspond to the image center. Finally, we produce a spatio-temporal silhouette volume, that is, GSV by piling up the silhouette images on the temporal axis.

We show an example of a constructed GSV in Fig. 1 as time slice (x-y plane), horizontal slice (t-x plane), and vertical slice (t-y plane) images. We can confirm existence of gait periodicity from Fig. 1(b), (c).

2.2. Frequency-domain feature extraction

The second step is frequency-domain feature extraction from the constructed GSV. First, we detect gait period N_{gait} by maximizing the normalized autocorrelation of the GSV for the temporal axis. Here, we set the domain of gait period to be [20, 40] empirically for the natural gait period.

Next, we pick up the subsequences $\{\boldsymbol{S}_i\}(i = 1, 2, ..., N_{sub})$ for every N_{gait} frames from a total sequence $\boldsymbol{S}$. Note that the frame range of the ith subsequence $\boldsymbol{S}_i$ is $[iN_{gait}, (i+1)N_{gait} - 1]$. Then the Discrete Fourier Transformation (DFT) for the temporal axis is applied for each subsequence, and amplitude spectra are subsequently calculated as

$$G_i(x, y, k) = \sum_{n=iN_{gait}}^{(i+1)N_{gait}-1} g(x, y, n) e^{-j\omega_0 kn} \quad (1)$$

$$A_i(x, y, k) = \frac{1}{N_{gait}} |G_i(x, y, k)|. \quad (2)$$

where $g_{gsv}(x, y, n)$ is the silhouette value at position (x, y) at the nth frame, ω_0 is a base angular frequency for the gait period N_{gait}, $G_i(x, y, k)$ is the DFT of GSV for k-times the base frequency, and $A_i(x, y, k)$ is an amplitude spectrum for $G_i(x, y, k)$. In this paper, we choose direct-current elements ($k = 0$) (averaged silhouette) and low-frequency elements ($k = 1, 2$) as gait features experimentally. As a result, the dimension N_A of amplitude spectra $A_i(x, y, k)(k = 0, 1, 2)$ sums up to $20 \times 30 \times 3 = 1800$.

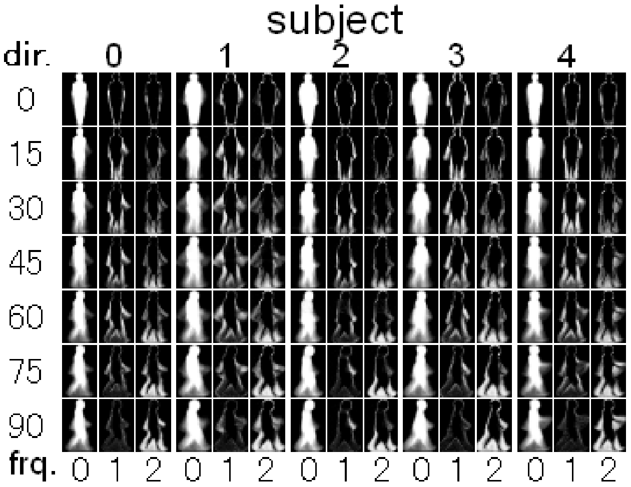


Figure 2. Gait feature for each subject from each view direction (every 15 deg)

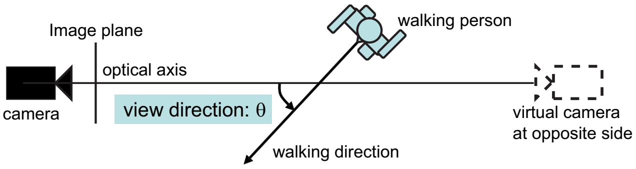


Figure 3. Definition of view direction θ (top view)

Figure 2 shows extracted amplitude spectra of multiple subjects from various view directions. The view direction is defined as the angle formed by an optical axis and a walking direction, as shown in Fig. 3, and in this paper the unit of the view direction is a degree. Amplitude spectra vary widely among view directions for each subject, and to some extent they also have individual variations for each view direction. Moreover, we can see that all the subjects have similar common tendencies for amplitude spectra variations across view direction changes. This fact indicates a real possibility that the variations across view direction changes are expressed with the VTM independently of individual variations.

2.3. Matching measure

We first define a matching measure between two subsequences. Let $\boldsymbol{a}(\boldsymbol{S}_i)$ be N_A dimensional feature vector composed of elements of the amplitude spectra $A_i(x, y, k)$ for subsequence $\boldsymbol{S}_i$. The matching measure $d(\boldsymbol{S}_i, \boldsymbol{S}_j)$ is simply chosen as the Euclidean distance:

$$d(\boldsymbol{S}_i, \boldsymbol{S}_j) = ||\boldsymbol{a}(\boldsymbol{S}_i) - \boldsymbol{a}(\boldsymbol{S}_j)||. \quad (3)$$

Next, we define a matching measure between two total sequences. Let $\boldsymbol{S_P}$ and $\boldsymbol{S_G}$ be total sequences for probe and gallery, respectively, and let $\{\boldsymbol{S_P}_i\}(i = 1, 2, \ldots)$ and $\{\boldsymbol{S_G}_j\}(j = 1, 2, \ldots)$ be their subsequences, respectively. Gallery subsequences $\{\boldsymbol{S_G}_j\}$ have variations in general and probe subsequences $\{\boldsymbol{S_P}_i\}$ may contain outliers. A measure candidate $D(\boldsymbol{S_P}, \boldsymbol{S_G})$ to cope with them is the median value of the minimum distances of each probe subsequence $\boldsymbol{S_P}_i$ and gallery subsequences $\{\boldsymbol{S_G}_j\}(j = 1, 2, \ldots)$:

$$D(\boldsymbol{S_P}, \boldsymbol{S_G}) = \mathrm{Median}_i[\min_j\{d(\boldsymbol{S_P}_i, \boldsymbol{S_G}_j)\}. \quad (4)$$

3. View transformation model

3.1. Formulation of VTM

We briefly describe the formulation of a VTM in a way similar to that in [14]. Note that we apply the model to the frequency-domain feature extracted from GSV while that in [14] directly applied it to a static image.

We first quantize view directions into K directions. Let $\boldsymbol{a}^m_{\theta_k}$ be a N_A dimensional feature vector for the kth view direction of the mth subject. Supposing that the feature vectors for K view directions of M subjects are obtained as a training set, we can construct a matrix whose row indicates view direction changes and whose column indicates each subject; and so can decompose it by Singular Value Decomposition (SVD) as

$$\begin{bmatrix} \boldsymbol{a}^1_{\theta_1} & \cdots & \boldsymbol{a}^M_{\theta_1} \\ \vdots & \ddots & \vdots \\ \boldsymbol{a}^1_{\theta_K} & \cdots & \boldsymbol{a}^M_{\theta_K} \end{bmatrix} = USV^T = \begin{bmatrix} P_{\theta_1} \\ \vdots \\ P_{\theta_K} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}^1 \cdots \boldsymbol{v}^M \end{bmatrix}. \quad (5)$$

where U is the $KN_A \times M$ orthogonal matrix, V is the $M \times M$ orthogonal matrix, S is the $M \times M$ diagonal matrix composed of singular values, P_{θ_k} is the $N_A \times M$ submatrix of US, and $\boldsymbol{v}^m$ is the M dimensional column vector.

The vector $\boldsymbol{v}^m$ is an intrinsic feature vector of the mth subject and is independent of view directions. The submatrix P_{θ_k} is a projection matrix from the intrinsic vector $\boldsymbol{v}$ to the feature vector for view direction θ_k, and is common for all subjects, that is, it is independent of the subject. Thus, the feature vector $\boldsymbol{a}^m_{\theta_i}$ for the view direction θ_i of the mth subject is represented as

$$\boldsymbol{a}^m_{\theta_i} = P_{\theta_i}\boldsymbol{v}^m. \quad (6)$$

Then, feature vector transformation from reference view direction θ_{ref} to θ_i is easily obtained as

$$\hat{\boldsymbol{a}}^m_{\theta_i} = P_{\theta_i} P^+_{\theta_{ref}} \boldsymbol{a}^m_{\theta_{ref}}. \quad (7)$$

where $P^+_{\theta_{ref}}$ is the pseudo inverse matrix of $P_{\theta_{ref}}$. In practical use, transformation from one view direction may be insufficient because motions orthogonal to the image plane are degenerated in the silhouette image. For example, it is difficult for even us humans to estimate a feature $\boldsymbol{a}^m_{90}$ from $\boldsymbol{a}^m_0$ (see Fig. 2 for example). Therefore, when features for more than one view direction (let them be $\theta_{ref}(1), \ldots, \theta_{ref}(k)$) are obtained, we can more precisely transform a feature for the view direction θ_i as

$$\hat{\boldsymbol{a}}^m_{\theta_i} = P_{\theta_i} \begin{bmatrix} P_{\theta_{ref}(1)} \\ \vdots \\ P_{\theta_{ref}(k)} \end{bmatrix}^+ \begin{bmatrix} \boldsymbol{a}^m_{\theta_{ref}(1)} \\ \vdots \\ \boldsymbol{a}^m_{\theta_{ref}(k)} \end{bmatrix}. \quad (8)$$

In the above formulation, there are no constraints for view transformation, but each body point such as head, hands, and knees appears at the same height, respectively, for all view directions because of the height scaling as described in 2.1. Therefore, we constrain transformation from a height y_i to another height $y_j (\neq y_i)$ and define the above transformation separately at each height y_i.

3.2. Reference addition with geometrical model

Here, we describe addition of virtual reference based on the geometrical model [16] to make view transformation more precise.

When a target subject is observed at a distance from a camera and weak perspective projection is assumed, the silhouette image observed with a virtual camera at the opposite side from the view direction[1] θ as shown in Fig. 3 (let the image be $I_{opp}(\theta)$), becomes a mirror image of the original silhouette image from view direction θ (let it be $I(\theta)$). In addition, it is clear that $I_{opp}(\theta)$ is the same as $I(\theta + 180)$. Hence, $I(\theta + 180)$ becomes a mirror image of $I(\theta)$.

Moreover, when a left-right symmetry of gait motion is assumed, $I(360-\theta)$ becomes a mirror image of $I(\theta)$. When both of the weak perspective projection and the symmetry are assumed, $I(180 - \theta)$ becomes the same image of $I(\theta)$.

Hence, once a gait feature for reference direction θ_{ref} is obtained, we can virtually add the same feature for direction $(180 - \theta_{ref})$ and a mirror feature for direction $(\theta_{ref} + 180)$ and $(360 - \theta_{ref})$. Because the addition of the virtual references is based on the assumptions, the actually observed feature is prior to the virtually added features when features for multiple reference directions are observed and an additional reference direction is the same as another observed reference direction.

4. Experimental results and analyses

4.1. Dataset and evaluation method

We use a total of 736 gait sequences from 20 subjects for the experiments. The sequences include 24 view directions

[1] Note that the view direction θ is defined for the actual camera and that it is used in common for both the actual and the virtual cameras.

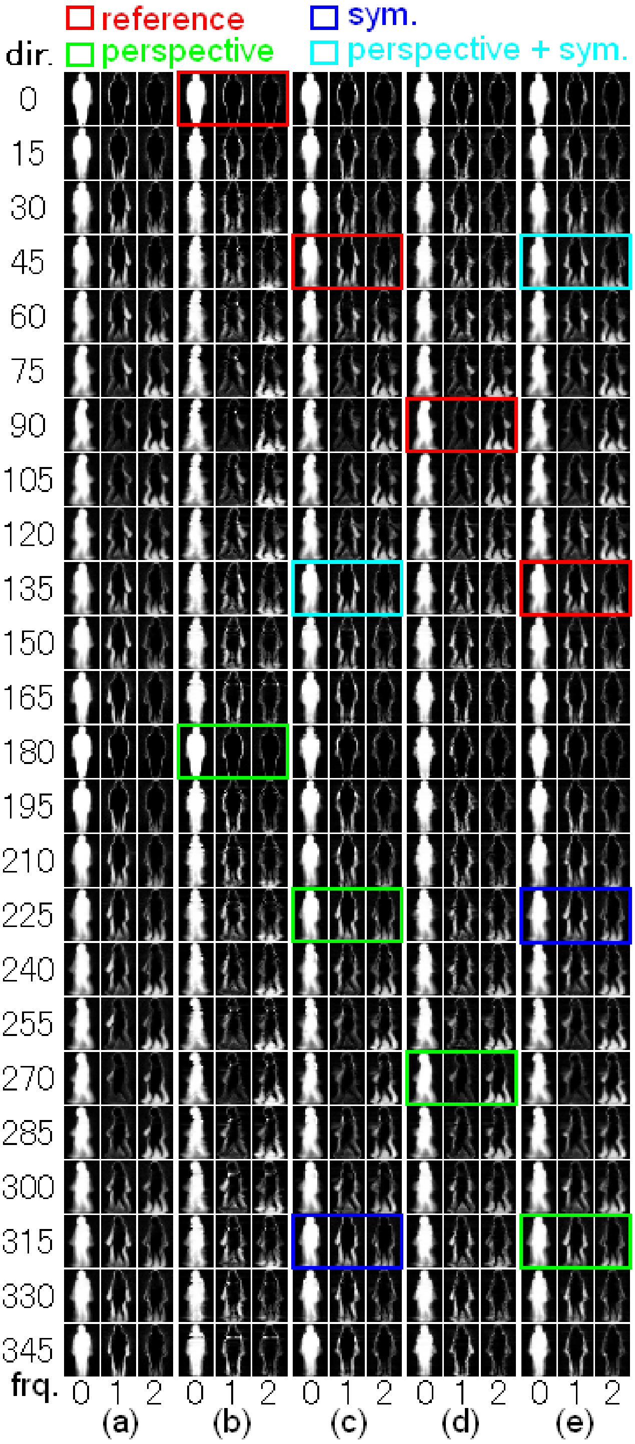
at every 15 degrees. Each sequence consists of from 10 to 20 steps of straight walks. The training set for the VTM is composed of 240 sequences of 10 subjects from 24 view directions. Then, we prepare reference sets of $20k$ sequences from 20 subjects of k view directions (let them be θ_{ref} for a single reference and be $\theta_{ref}(1), \cdots, \theta_{ref}(k)$ for multiple references). Here, to reduce the number of combination of multiple references for simplicity, we use 8 reference directions at every 45 deg, that is, directions 0, 45, 90, 135, 180, 225, 270, and 315 deg. A gallery set for each view direction is obtained by view transformation with eq. (7) for a single reference or eq. (8) for multiple references. A probe (test) set is composed of the other sequences except for those of subjects included in the training set for the VTM, and each sequence is indexed in advance with the view direction as θ_{probe} because view direction estimation is easily done using a walking person's velocity in the image or by view direction classification with averaged features for each view direction.

We also prepare for a gallery set with no transformation (let it be NT) for comparison. In this method, a gallery feature for each view direction is replaced with a reference feature minimizing an angle formed by the gallery view direction (let it be $\theta_{gallery}$) and the reference view direction. For example, when reference features for $\theta_{ref}(1) = 0$ and $\theta_{ref}(2) = 90$ are obtained, gallery features for $45 \leq \theta_{gallery} \leq 135$ or $225 \leq \theta_{gallery} \leq 315$ are replaced with the reference feature for $\theta_{ref}(2)$, and the others are replaced with that for $\theta_{ref}(1)$.

We constructed a matching test using the dataset. A probe is assigned verification when eq. (4) is above a certain threshold value, and a Receiver Operating Characteristics (ROC) [7] curve is obtained by plotting pairs of verification rate and false alarm rate for various threshold values. The tests are repeated for different 20 training sets for the VTM and the averaged performance is evaluated by verification rate at 10% false alarm for the ROC curve.

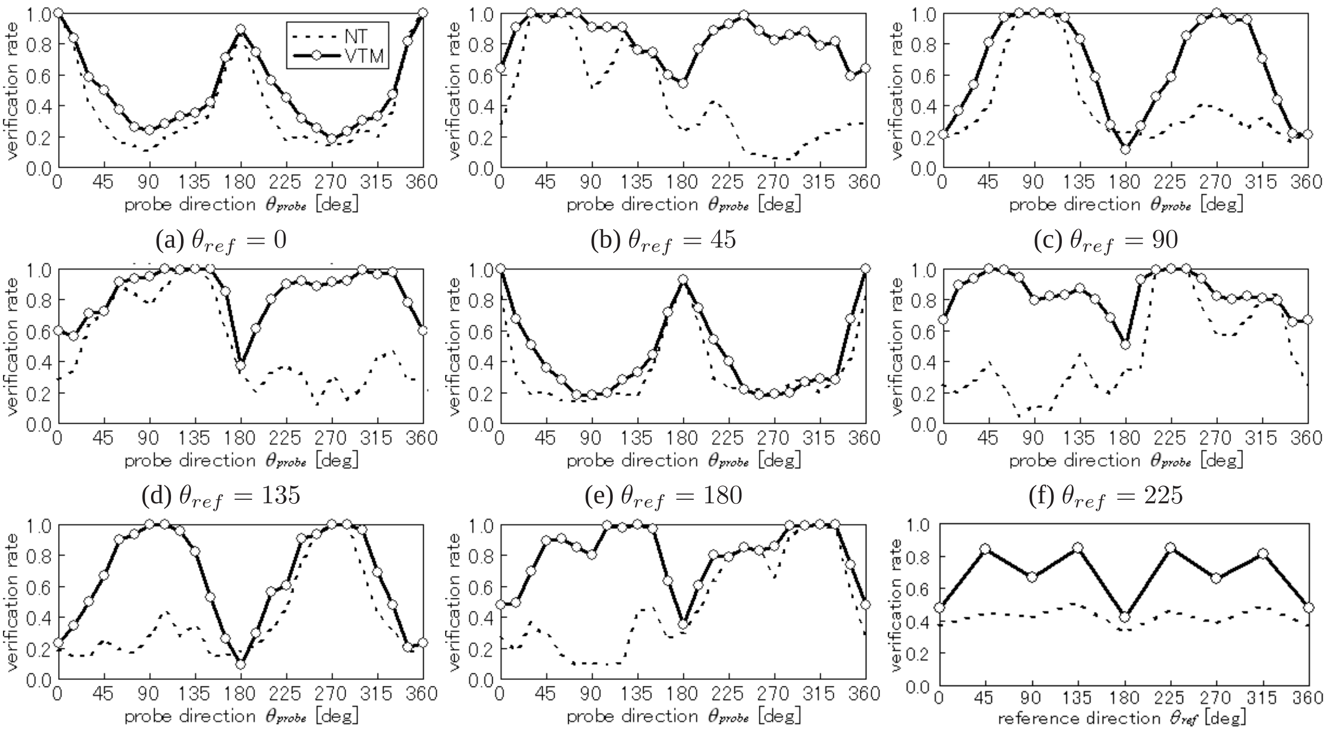
4.2. Single reference

We first show transformed features with the VTM using a single reference in Fig. 4. We can see the transformed features are similar to the original feature (Fig. 4(a)) for each direction as a whole. For 0-deg reference (Fig. 4(b)), a front-back swing motion of arm and leg is orthogonal to the image plane and degenerated, then transformed features near side view ($\theta = 90, 270$) are degraded compared with those from the other references. For 90-deg reference (Fig. 4(d)), although the front-back swing motion is observed the best, the width of body is orthogonal to the image plane and degenerated. Therefore transformed features near front view ($\theta = 0, 180$) are degraded compared with the others. On the other hand, for 45-deg reference (Fig. 4(c)), features for orthogonal views ($\theta = 135, 315$) are added by the assumption of left-right gait symmetry. Hence, transformed features are relative fine compared with the others. For 135-



(a): Original feature, (b)-(d): Transformed feature from 0, 45, 90, and 135 deg by the VTM respectively.

Figure 4. Transformed feature from a single reference



(a) $\theta_{ref} = 0$ (b) $\theta_{ref} = 45$ (c) $\theta_{ref} = 90$

(d) $\theta_{ref} = 135$ (e) $\theta_{ref} = 180$ (f) $\theta_{ref} = 225$

(g) $\theta_{ref} = 270$ (h) $\theta_{ref} = 315$ (i) average on prove view direction

Figure 5. Verification rate at 10% false alarm for single reference

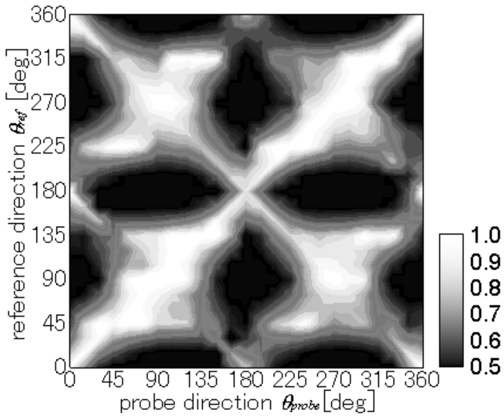


Figure 6. Verification rate for all combinations of probe view θ_{probe} and reference view θ_{ref}

deg reference (Fig. 4(e)), the tendency is the same as the 45-deg reference.

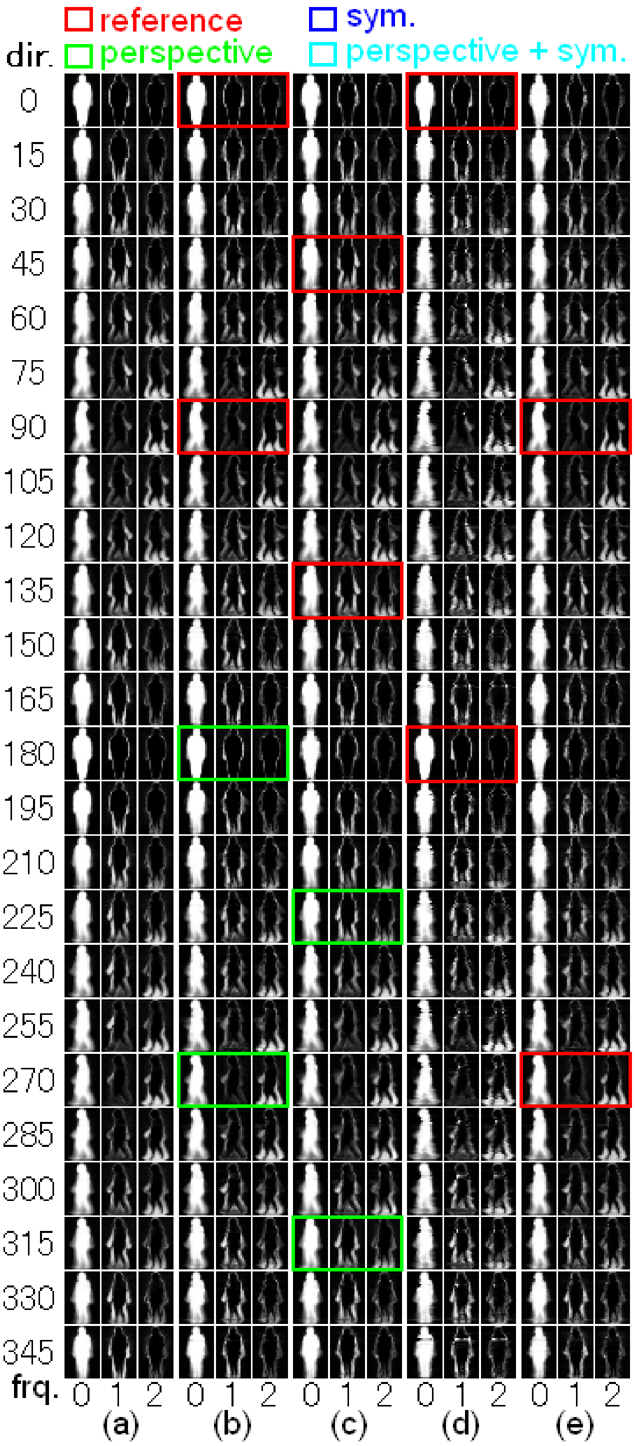
Next, we show verification rates for each reference view direction θ_{ref} in Fig. 5(a)-(h). It is clear that the probes with the same view direction as the reference have very high performances for all references (for example, $\theta_{probe} = 0$ in Fig. 5(a) and $\theta_{probe} = 90$ in Fig. 5(c)). In addition, the probe with the same view direction as the added references have also relatively high performance (for example, $\theta_{probe} = 180$ in Fig. 5(a) and $\theta_{probe} = 270$ in Fig. 5(c)).

On the other hand, the probes with the orthogonal view direction to the reference have poor performance (for example, $\theta_{probe} = 90$ in Fig. 5(a) and $\theta_{probe} = 0$ in Fig. 5(c)). As an exception, for angled view such as $\theta_{probe} = 45, 135$, references of orthogonal view are added, therefore the performance to the orthogonal probe are also relatively high.

Figure 6 shows the performance summary by VTM for all combinations of probe view θ_{probe} and reference view θ_{ref} included in Fig. 5(a)-(h). We can see that the performance is high on the following lines.

- $\theta_{probe} = \theta_{ref}$: the same view direction.
- $\theta_{probe} = \theta_{ref} + 180$: the added feature by the weak perspective assumption.
- $\theta_{probe} = 360 - \theta_{ref}$: the added feature by the left-right symmetry assumption.
- $\theta_{probe} = 180 - \theta_{ref}$: the added feature by the weak perspective and the left-right symmetry assumption.

Figure 5(i) shows averaged performance on probe view direction. The performance of NT for each reference is similar to each other and is inferior to that of VTM. On the other hand, the performance of VTM for angled view reference ($\theta = 45, 135, 225, 315$) is relatively high.



(a): Original feature, (b)-(d): Transformed feature from {0,90}, {45,135}, {0,180}, and {90,270} by the VTM respectively.

Figure 7. Transformed feature from two references

4.3. Combination of two references

We first show transformed features with the VTM using some combinations of two references in Fig. 7. We denote a combination of two reference as $\{\theta_{ref}(1), \theta_{ref}(2)\}$ in this section. For orthogonal combinations (Fig. 7(b)(c)), degenerated information for orthogonal direction to the image plane such as front-back motion for 0-deg feature and width of body for 90-deg feature is compensated with each other. Therefore we can see the transformed features are more similar to the original feature (Fig. 7(a)) for each direction than those for a single view direction (Fig. 4). On the other hand, for 180-deg opposite combinations (Fig. 7(d)(e)), the transformed features are little improved compared with those in case of the single reference. This is because the degenerated information for orthogonal plane is not compensated and because one of the two references has been already added by the weak perspective assumption even in case of the single reference.

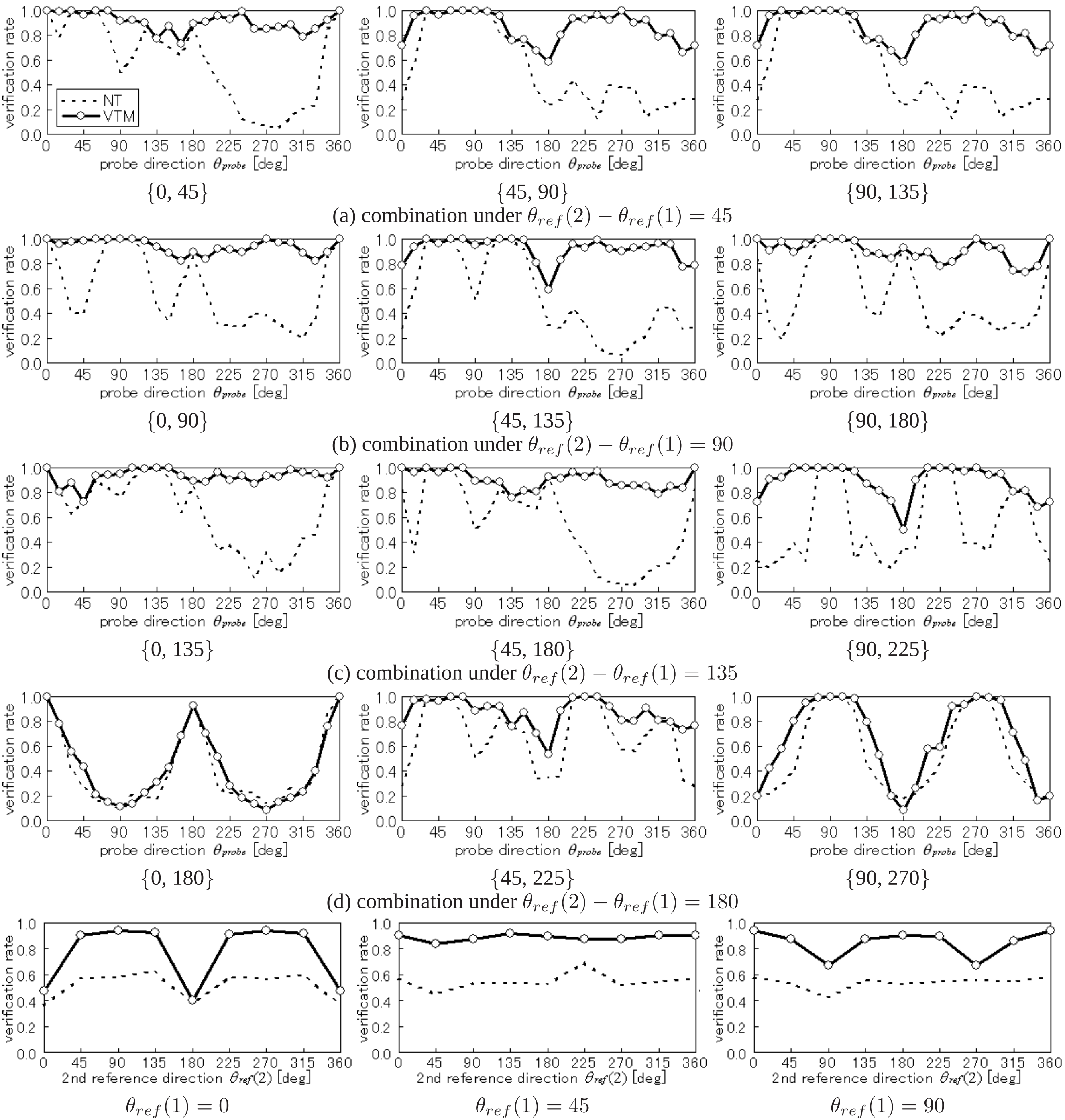
Next, we show verification rates for some combinations of two reference view directions in Fig. 8(a)-(d). As a result, we can see the performance for orthogonal combination (Fig. 8(b)) is high and that for intermediate combinations between orthogonal and parallel (Fig. 8(a)(c)) is also relatively high. On the other hand, that for 180-deg opposite combination (Fig. 8(d)) is little improved as compared with single reference.

Figure 8(e) shows averaged performance on probe view direction. In this figure, repeated reference (for example, $\theta_{ref}(1) = \theta_{ref}(2) = 0$ indicates the same meaning of single reference $\theta_{ref} = 0$). In case of $\theta_{ref}(1) = 45$, the performance of single reference is enough high, then the performance difference by combination is not outstanding. On the other hand, in case of $\theta_{ref}(1) = 0$ and 90, the performance difference is outstanding, that is, the performance is high for orthogonal combination and is little improved for parallel combination.

To confirm this tendency, we show the averaged performance on probe for combinations of two references by VTM in Fig. 9. We can see high performance line $\theta_{ref}(2) - \theta_{ref}(1) = 90 + 180n, n = -2, -1, 0, 1)$ corresponding to the orthogonal combination, and low performance line $(\theta_{ref}(2) - \theta_{ref}(1) = 180n, n = -1, 0, 1)$ corresponding to the parallel combination. Therefore, the orthogonal combination is suitable for two references for VTM.

4.4. The number of references

Here, we discuss the effect of the number of references (let it be N_{ref}) for VTM. We show the averaged verification rate for N_{ref} in Fig. 10. In this figure, ”best” and ”worst” indicates the performance using the best and the worst combination of reference views respectively and ”average” in-



{0, 45} {45, 90} {90, 135}

(a) combination under $\theta_{ref}(2) - \theta_{ref}(1) = 45$

{0, 90} {45, 135} {90, 180}

(b) combination under $\theta_{ref}(2) - \theta_{ref}(1) = 90$

{0, 135} {45, 180} {90, 225}

(c) combination under $\theta_{ref}(2) - \theta_{ref}(1) = 135$

{0, 180} {45, 225} {90, 270}

(d) combination under $\theta_{ref}(2) - \theta_{ref}(1) = 180$

$\theta_{ref}(1) = 0$ $\theta_{ref}(1) = 45$ $\theta_{ref}(1) = 90$

(e) average on probe view direction

Figure 8. Verification rate at 10% false alarm rate for two references

dicates the averaged performance on all the combinations. In case of the best, we obtain more than 90% verification rate if N_{ref} is more than or equal to 2. In addition, even in case of the worst, we obtain more than 80% verification rate if N_{ref} is more than or equal to 3. As a result, we can see a few references are enough to transform features precisely and to achieve high verification rate for various-directions walks.

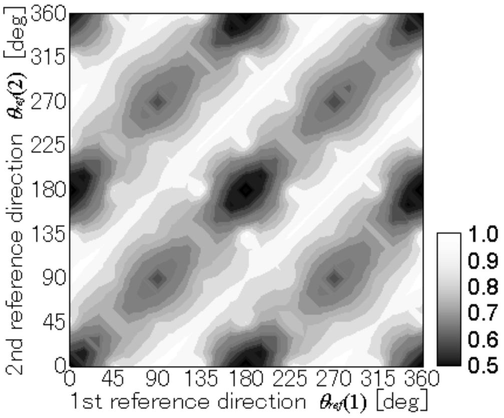


Figure 9. Averaged performance on probes for combinations of two references

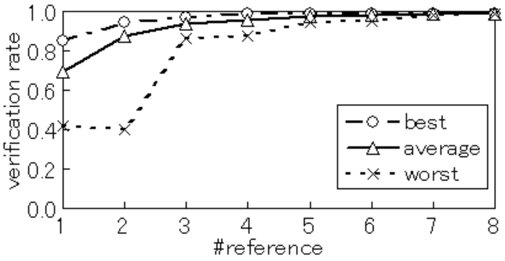


Figure 10. Averaged verification rate for the number of references

5. Conclusion

In this paper, we discussed the selection of reference views for various-view gait identification using a view transformation model (VTM). As a result of experiments, it was cleared that angled reference view such as 45 deg and 135 deg are effective for a single reference, and that combinations of orthogonal references such as a combination of 0 deg and 90 deg are effective for two references. In addition, we confirmed that a few references are enough to achieve high performance on various-view gait identification.

Future works are as follows.

- Experiments for a general database, such as the HumanID Gait Challenge Problem Datasets [9].
- Adaptation to camera tilt changes

References

[1] A. Bobick and A. Johnson. Gait recognition using static activity-specific parameters. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 423–430, 2001. 1

[2] J. Han and B. Bhanu. Individual recognition using gait energy image. *Trans. on Pattern Analysis and Machine Intelligence*, 28(2):316– 322, 2006. 1

[3] J. Han, B. Bhanu, and A. Roy-Chowdhury. A study on view-insensitive gait recognition. In *Proc. of IEEE Int. Conf. on Image Processing*, volume 3, pages 297–300, Sep. 2005. 1

[4] A. Kale, A. Roy-Chowdhury, and R. Chellappa. Towards a view invariant gait recognition algorithm. In *Proc. of IEEE Conf. on Advanced Video and Signal Based Surveillance*, pages 143–150, 2003. 1

[5] T. Kobayashi and N. Otsu. Action and simultaneous multiple-person identification using cubic higher-order local auto-correlation. In *Proc. of the 17th Int. Conf. on Pattern Recognition*, volume 3, pages 741–744, Aug. 2004. 1

[6] Y. Makihara, R. Sagawa, M. Yasuhiro, T. Echigo, and Y. Yagi. Gait recognition using a view transformation model in the frequency domain. In *Proc. of European Conf. on Computer Vision*, May 2006 (to appear). 1

[7] P. Phillips, H. Moon, S. Rizvi, and P. Rauss. The feret evaluation methodology for face-recognition algorithms. *Trans. of Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000. 4

[8] R. Sagawa, Y. Makihara, T. Echigo, and Y. Yagi. Matching gait image sequences in the frequency domain for tracking people at a distance. In *Proc. 7th Asian Conference on Computer Vision*, volume 2, pages 141–150, Jan. 2006. 1

[9] S. Sarkar, J. Phillips, Z. Liu, I. Vega, P. Grother, and K. Bowyer. The humanid gait challenge problem: Data sets, performance, and analysis. *Trans. of Pattern Analysis and Machine Intelligence*, 27(2):162–177, 2005. 1, 8

[10] G. Shakhnarovich, L. Lee, and T. Darrell. Integrated face and gait recognition from multiple views. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 439–446, 2001. 1

[11] N. Spencer and J. Carter. Towards pose invariant gait reconstruction. In *Proc. of IEEE Int. Conf. on Image Processing*, volume 3, pages 261–264, Sep. 2005. 1

[12] R. Tanawongsuwan and A. Bobick. Modelling the effects of walking speed on appearance-based gait recognition. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 783–790, 2004. 1

[13] R. Urtasun and P. Fua. 3d tracking for gait characterization and recognition. In *Proc. of the 6th IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 17–22, 2004. 1

[14] A. Utsumi and N. Tetsutani. Adaptation of appearance model for human tracking using geometrical pixel value distributions. In *Proc. of the 6th Asian Conf. on Computer Vision*, volume 2, pages 794–799, 2004. 1, 3

[15] C. Yam, M. Nixon, and J. Carter. Automated person recognition by walking and running via model-based approaches. *Pattern Recognition*, 37(5):1057–1072, 2004. 1

[16] S. Yu, D. Tan, and T. Tan. Modelling the effect of view angle variation on appearance-based gait recognition. In *Proc. of 7th Asian Conf. on Computer Vision*, volume 1, pages 807–816, Jan. 2006. 1, 3