Video from Nearly Still: an Application to Low Frame-rate Gait Recognition

Naoki Akae, Al Mansur, Yasushi Makihara and Yasushi Yagi
The Institute of Scientific and Industrial Research
Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka, 567-0047 Japan
`{akae,mansur,makihara,yagi}@am.sanken.osaka-u.ac.jp`

Abstract

In this paper, we propose a temporal super resolution approach for quasi-periodic image sequence such as human gait. The proposed method effectively combines example-based and reconstruction-based temporal super resolution approaches. A periodic image sequence is expressed as a manifold parameterized by a phase and a standard manifold is learned from multiple high frame-rate sequences in the training stage. In the test stage, an initial phase for each frame of an input low frame-rate image sequence is estimated based on the standard manifold at first, and the manifold reconstruction and the phase estimation are then iterated to generate better high frame-rate images in the energy minimization framework that ensures the fitness to both the input images and the standard manifold. The proposed method is applied to low frame-rate gait recognition and experiments with real data of 100 subjects demonstrate a significant improvement by the proposed method, particularly for quite low frame-rate videos (e.g., 1 fps).

1. Introduction

Human gait is a useful biometric feature that can be efficiently recognized at a distance and can be applied to uncooperative subjects. With the recent developments of computer vision techniques, it became easier to extract and process gait feature and, as a result, vision-based gait recognition methods has recently gained considerable attention from the human identification field.

However, in real application, construction of a robust gait recognition system is still challenging because of the intra-class variation of gait features due to change in walking speed [13, 28], viewpoint [12, 31], clothing [8], and elapsed time [23, 15]. In addition, gait recognition from a low-frame-rate video is another major problem as the sparsity of the observed gait phases (or stances) makes the existing gait recognition methods perform poorly. This problem is frequently encountered when a video captured by a CCTV camera is used for gait recognition where the video is recorded at a quite low frame-rate (e.g., 1 to 5 fps) due to low transmission bandwidth and limited storage capacity.

While most of the gait recognition methods use normal frame-rate videos for recognition, few works [16, 3] attempted to solve the problem of using low frame-rate videos for gait recognition. Although Mori *et al*. [16] uses low frame-rate videos for probe sequence, normal frame-rate videos are used as gallery sequence. Therefore, this method cannot be used for cross-camera person matching where both the probe and gallery sequences are low frame-rate. Another line of research related to the low frame-rate problem is temporal interpolation. In [3], a level-set approach to temporal interpolation is proposed and applied to gait recognition in low frame-rate videos. In another work, Prismal *et al*. [22] used linear interpolation of the moment values for temporal interpolation. However, these methods are applicable for mid-rate videos (e.g., more than 6 fps) and does not perform well for quite low frame-rate videos (e.g., 1 or 2 fps).

Methods proposed for temporal Super Resolution (SR) [11, 25, 26, 2] are closely related to low frame-rate problem. Using multiple-camera setup, Shechtman *et al*. [25] presented an algorithm for space-time SR by multiple low-resolution videos of the same dynamic scene. However, the multiple-camera setup is unavailable in most of surveillance scenes. By exploiting the self-similarity, Shimano *et al*. [26] proposed a video temporal SR method from a single image sequence. However, this algorithm is applied for increasing the temporal resolution of an input video at most twice and will not work in a quite low frame-rate with articulated motion such as human gait.

Using an energy minimization framework, Makihara *et al*. [11] proposed a temporal SR method from a single quasi-periodic image sequence. Based on phase registration, low frame-rate sequence observed from multiple periods is exploited to construct one period of a high frame-rate sequence (referred to as manifold later). However, due to the error in initial phase estimation using linear phase evolution prior, this method suffers from wagon wheel effect (backward motion) in quite low frame-rate cases (1 or 2 fps).

Besides this, it suffers from the problem of stroboscopic effect when the sampling interval coincides with the gait period and the observed image sequence appears almost still, in other words, most of the gait stances are not observed, which leads to manifold reconstruction failure.

To solve the problem of the wagon wheel effect, Akae *et al*. [2] utilized an exemplar of high frame-rate image sequence to improve the initial phase estimation in the framework similar to [11] and applied the synthesized high frame-rate sequence to low frame-rate gait recognition. However, it still suffers from the stroboscopic effect and hence its gait recognition performance is not so much improved compared with the previous methods.

In this paper, we propose a unified framework of example-based and reconstruction-based periodic temporal SR to overcome the stroboscopic effect. The prior knowledge of gait dynamics trained from multiple training subjects is represented by a standard manifold and is used as an example-based cue, while multiple periods of observations of the low frame-rate sequence are used as a reconstruction-based cue. The proposed method can construct a high frame-rate sequence even when the observed image sequence appears almost still due to the stroboscopic effect because the missing phases are recovered based on the standard manifold. In addition, we apply the high frame-rate sequence obtained by the proposed method to low frame-rate gait recognition and achieved a superior performance over the previous methods.

This paper is organized as follows: After addressing the related work in section 2, section 3 describes the preprocessing for silhouette-based gait recognition. Section 4 describes the proposed method of periodic temporal SR procedure. Experimental results using real gait image sequences are presented in section 5. Finally, section 6 concludes the paper and identifies future research direction.

2. Related work

Temporal super resolution: Temporal SR approaches could be classified into two classes: Example-based approaches and reconstruction-based approaches.

Following an example-based approach, space-time SR can be achieved from a single video recording of general dynamic scenes by exploiting the fact that small space-time patches of a natural video have the tendency to recur itself inside the same video at multiple spatio-temporal scales [24], or by exploiting self-similar appearance of patches across different temporal resolution [26]. However, the temporal SR rate handled in these methods is quite limited (e.g., twice).

Reconstruction-based approaches increase both the spatial and the temporal resolution, by combining information from multiple videos of the same dynamic scene, at sub-pixel and sub-frame accuracy [1, 25]. In [1], N cameras were used for N-times temporal SR which requires few tens of cameras in our problem setting of low frame-rate gait recognition and, is impractical for surveillance application. Instead of simultaneous recording by multiple video cameras, sequential recording of the multiple repetitions of a cyclic motion by single camera could also be used [11, 27, 2]. Assuming that a single snapshot contains a group of the same animals at different phases of their cyclic motion, Xu *et al*. [32] infer the motion cycle of animals from a still picture. However, such an assumption is unsuitable for human gait motion as each subject has unique style of walking and this property is utilized in gait recognition. In another similar work, Ding *et al*. [5] proposed a method of reconstructing missing portions in a set of video frames by minimizing the rank of a matrix directly constructed from the available frames. This method assumes that the missing portion of the video is small compared to the observed part and, therefore, may not work in our problem setting.

Manifold Learning: Low dimensional manifold representation is a useful and common practice for analyzing human gait, pose tracking, or motion synthesis. Cheng *et al*. [4] proposed a person identification system by learning a low-dimensional embedding by Gaussian Process Latent Variable Model (GP-LVM), and modeling the temporal dynamics using Hidden Markov Models (HMMs). Lee *et al*. [9] introduced a model that combines a kinematics manifold and visual manifold for 3D tracking. Through learning a decomposable generative model, Elgammal *et al*. [6] proposed a method to separate style and content of human gait using manifolds representing human gait motion. Scaled Gaussian Process Latent Variable Model is used in [30] to learn prior models for 3D people tracking. However, all of these approaches employ a normal frame-rate sequence for manifold learning.

Gait Recognition: Traditionally, both period-based and frame-based features are used for human gait recognition. Period-based features [10, 12, 29] are usually computed from multiple images within one gait period. As reported in [19], averaged silhouette [10] (also known as Gait Energy Image, GEI [7]) works the best compared to the other frame-based features. Other types of period-based gait features include amplitude spectra [12], Fourier descriptors [33], and Gabor features [29]. However, as multiple images within a period is required, these features (including the averaged silhouette) fails to produce an acceptable recognition rate when the number of images within one gait period becomes very small, as observed in low frame-rate videos.

3. Preprocessing

In this section, preprocessing part of our silhouette-based gait recognition is briefly discussed. At the first step, given

Figure 1. Normalized gait silhouette sequence

a gait image sequence, a silhouette for each frame is extracted by background subtraction-based graph-cut segmentation [14]. Next, based on the height and the center of gravity of each silhouette, a normalized gait silhouette sequence of pre-determined size is generated by image size-normalization and registration (Fig. 1).

Principal Component Analysis (PCA) is then applied to the normalized gait silhouette sequence and each normalized gait silhouette is projected into PCA space for dimension reduction. The dimension of the PCA space is decided so that 99% of the total variance is preserved. The normalized gait silhouette sequence is expressed as a manifold parameterized by *phase* in parametric eigen space [17]. As shown in Fig. 2, the gait silhouette and gait silhouette sequence are expressed as a point and a trajectory in the PCA space, respectively.

4. Periodic temporal super resolution

4.1. Definition of quasi-periodic image sequence

In this paper, an input gait silhouette sequence is regarded as a quasi-periodic image sequence. At time t, an image drawn from the periodic image sequence with period P is denoted by the vector $\boldsymbol{x}(t)$, and satisfies

$$\boldsymbol{x}(t + P) = \boldsymbol{x}(t) \ \ \forall t. \tag{1}$$

Dimensionless time parameters, *phase s* and *relative phase $\tilde{s}$*, are then introduced as

$$s = s_P(t) = \frac{t}{P} \tag{2}$$

$$\tilde{s} = s - \lfloor s \rfloor, \tag{3}$$

where $s_P(\cdot)$ is the phase evolution function and $\lfloor \cdot \rfloor$ is a floor function. The periodic image sequence can now be represented in the phase domain as $\boldsymbol{x}_s(s) = \boldsymbol{x}(s_P^{-1}(s))$.

The periodic image sequence constructs a manifold with respect to the relative phase $\tilde{s} \in [0, 1]$ and satisfies $\boldsymbol{x}_s(1) = \boldsymbol{x}_s(0)$. An input gait silhouette sequence is composed of N^{in} discretely observed images $\boldsymbol{X}^{in} = \{\boldsymbol{x}_i^{in}\} (i = 0, \ldots, N^{in} - 1)$ which are denoted as

$$\boldsymbol{x}_i^{in} = \boldsymbol{x}_s(s_{P,i}) \tag{4}$$

$$s_{P,i} = s_0 + \frac{i}{fP}, \tag{5}$$

where $\boldsymbol{s}_P = \{s_{P,i}\}$ is the phase sequence for the input image sequence, s_0 is the phase for the first frame, and f is the frame-rate. Due to fluctuations of the sampling rate of the network camera and/or target motion, the input image

sequence is often degraded to a *quasi-periodic image sequence* $\boldsymbol{X}_Q^{in} = \{\boldsymbol{x}_{Q,i}^{in}\}$ and can be defined as

$$\boldsymbol{x}_{Q,i}^{in} = \boldsymbol{x}_s(s_{Q,i}) \tag{6}$$

$$s_{Q,i} = s_{P,i} + \Delta s_i, \tag{7}$$

where $\boldsymbol{s}_Q = \{s_{Q,i}\}$ is a quasi-periodic phase sequence.

In other words, the temporal SR problem can be stated as a simultaneous estimation problem of a periodic manifold $\boldsymbol{x}_s$ and a phase sequence $\boldsymbol{s}_Q$ from an input quasi-periodic image sequence $\boldsymbol{X}_Q^{in}$.

4.2. Manifold representation

In this work, the periodic manifold $\boldsymbol{x}_s$ is represented by a parametric eigenspace method [17] and a cubic N-spline function parameterized by the phase in the eigenspace is used to construct the manifold. Let us consider N^{cp} control points $\{\boldsymbol{y}_j^{cp}\}$ in the M dimensional eigenspace with corresponding phases $\{s_j^{cp}(= j/N^{cp})\}, \ (j = 0, \ldots, N^{cp} - 1)$. At the j-th interval $[s_j^{cp}, s_{j+1}^{cp}]$, a spline parameter vector for a k power-term coefficient is defined as $\boldsymbol{a}_{j,k}^{sp} \in \mathbb{R}^M (k = 0, 1, 2, 3)$. Let $A_j^{sp} = [\boldsymbol{a}_{j,0}^{sp}, \ldots, \boldsymbol{a}_{j,3}^{sp}]^T \in \mathbb{R}^{4 \times M}$ and $A^{sp} = [A_0^{spT}, \ldots, A_{N^{cp}-1}^{sp}{}^T]^T \in \mathbb{R}^{4N^{cp} \times M}$ are the sub-matrix (at the j-th interval) and the total spline matrix, respectively. A point at relative phase $\tilde{s}$ at the j-th interval can be then interpolated in the eigenspace as

$$\hat{\boldsymbol{y}}(\tilde{s}) = A^{spT} \boldsymbol{w}(\tilde{s}) \tag{8}$$

$$\boldsymbol{w}(\tilde{s}) = [0, \ldots, 0, 1, w, w^2, w^3, 0, \ldots, 0]^T \tag{9}$$

$$w = \frac{\tilde{s} - s_j^{cp}}{s_{j+1}^{cp} - s_j^{cp}}, \ (s_j^{cp} \le \tilde{s} \le s_{j+1}^{cp}) \tag{10}$$

where $\boldsymbol{w}(\tilde{s})$ is an interpolation coefficient vector whose $4j$-th to $(4j+3)$-th components are $[1, w, w^2, w^3]$, and w is the interpolation ratio between the control points.

Relationship between the control points matrix $Y^{cp} = [\boldsymbol{y}_0^{cp}, \ldots, \boldsymbol{y}_{N^{cp}-1}^{cp}]^T$ and the spline parameter matrix A^{sp} is derived from the C2-continuous boundary conditions [11] as

$$A^{sp} = DY^{cp}, \tag{11}$$

where $D \in \mathbb{R}^{4N^{cp} \times N^{cp}}$ is a coefficient matrix. Therefore, given the control points Y^{cp}, interpolation $\hat{\boldsymbol{y}}(\tilde{s})$ for a relative phase $\tilde{s}$ is obtained by Eqs. (8) and (11). As discussed in the following sections, the reconstruction problem of the periodic manifold $\boldsymbol{x}_s(s; Y^{cp})$ can now be reformulated as an estimation problem of the control points Y^{cp}.

4.3. Standard manifold construction

A standard manifold is constructed by using image sequences with normal frame-rate from N^{tr} training subjects, i.e., non-recognition targets. A common M dimensional eigenspace of normalized gait silhouettes among all the
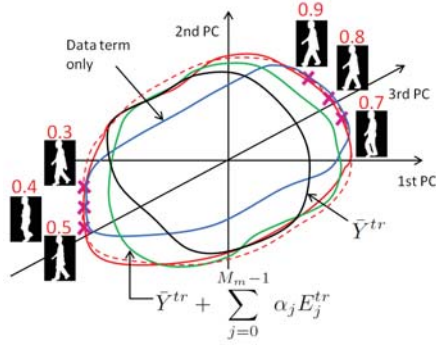
Figure 2. Effect of the data term and manifold fitness. Low frame-rate input frames with corresponding phases are shown as red crosses in the silhouette eigenspace (only three principal components are shown for simplicity). In case of stroboscopic effect, because only a few phases of a gait cycle are observed, the input frames are embedded on some isolated parts of the true manifold (red, dotted).

training subjects is first computed (referred to as silhouette eigenspace later) and a manifold for each training subject is then obtained by the periodic temporal SR technique [2]. Let the matrix Y_i^{tr} represents the manifold control points of i-th training subject and note that the manifold control points are phase-synchronized each other because the same exemplar is exploited for initial phase estimation in [2].

After unfolding the control point matrix Y_i^{tr} into a vector $\boldsymbol{z}_i^{tr} \in \mathbb{R}^{N^{cp}M}$ by column stacking, PCA is again applied to reduce the dimension (referred to as the manifold eigenspace later). Given the mean vector and M_m eigen vectors of the control point vectors as a result of PCA, the mean matrix $\bar{Y}^{tr}$ and eigen matrices $\{E_j^{tr}\}(j = 0, \cdots, M_m - 1)$ is also obtained by converting the control point vectors into matrices. Now, an estimated control point matrix by the standard manifold is expressed as

$$\hat{Y}^{tr} = \bar{Y}^{tr} + \sum_{j=0}^{M_m-1} \alpha_j E_j^{tr}, \qquad (12)$$

where $\boldsymbol{\alpha} = [\alpha_0, \ldots, \alpha_{M_m-1}]^T$ is standard manifold co-efficients, i.e., the weights for eigen matrices $\{E_j^{tr}\}$.

4.4. Energy minimization framework

Similar to the other temporal SR approaches [25, 26], we used an energy minimization approach. Let an quasi-periodic input image sequence in the eigenspace is expressed as $Y_Q^{in} = \{\boldsymbol{y}_{Q,i}^{in}\}$ where the accompanying phase sequence $\boldsymbol{s}_Q = \{s_{Q,i}\}$ is unknown. The interpolation coefficient vector for the i-th phase $s_{Q,i}$ is defined as $\boldsymbol{w}(s_{Q,i})$, in the same way as Eq. (9). Given a set of control points, Y^{cp}, approximation in the silhouette eigenspace at the i-th phase $s_{Q,i}$ is

$$\hat{\boldsymbol{y}}(Y^{cp}, s_{Q,i}) = A^{spT} \boldsymbol{w}(s_{Q,i}) = Y^{cpT} D^T \boldsymbol{w}(s_{Q,i}), \quad (13)$$

The energy function is then constructed by considering the four aspects: (1) data fitness between the interpolation $\hat{\boldsymbol{y}}(Y^{cp}, s_{Q,i})$ and the input $\boldsymbol{y}_{Q,i}^{in}$, (2) fitness between the manifold Y^{cp} and the estimated standard manifold $\hat{Y}^{tr}$, (3) smoothness of the periodic manifold $\boldsymbol{y}_s(s; Y^{cp})$, and (4) smoothness of the phase evolution $\boldsymbol{s}_Q$ based on the linear phase evolution prior. The actual form of the function is

$$E(Y^{cp}, \boldsymbol{\alpha}, \boldsymbol{s}_Q) =$$
$$\frac{1}{N^{in}} \sum_{i=0}^{N^{in}-1} ||Y^{cpT} D^T \boldsymbol{w}(s_{Q,i}) - \boldsymbol{y}_{Q,i}^{in}||^2 +$$
$$\lambda_t \frac{1}{N^{cp}} ||Y^{cp} - (\bar{Y}^{tr} + \sum_{j=0}^{M_m-1} \alpha_j E_j^{tr})||^2 +$$
$$\lambda_m \frac{1}{N^{cp}} ||BY^{cp}||^2 +$$
$$\lambda_s \frac{1}{N^{in}} \sum_{i=1}^{N^{in}-1} \left(s_{Q,i+1} - s_{Q,i} - \frac{1}{P'}\right)^2, \qquad (14)$$

where the first, second, third, and fourth terms are the data term, the standard manifold fitness term, smoothness term for the manifold, and the smoothness term for phase evolution, respectively. B is a coefficient matrix for calculating the manifold curvature, and $P'(= fP)$ is the global period (in terms of frames) assuming linear phase evolution.

The effect of the data term and the standard manifold fitness term could be easily understood from Fig. 2. If the energy minimization is done without the standard manifold fitness term, the manifold reconstruction relies on the input frames and the smoothness constraints only. Therefore, the resultant manifold (blue) differs largely from the true manifold, particularly in phase missing part. On the other hand, if the manifold is strictly constrained to the standard manifold while the input data lies outside of the standard manifold representation, the resultant manifold (green) will be still far from the true manifold, particularly near the phases of the input data, which implies to discard the meaningful subject specific information.

Therefore, by considering all the four terms of the energy function, the manifold is attracted to the input data to preserve such meaningful individualities, while it is also attracted to the standard manifold in part of missing phases. Consequently, the reconstructed manifold (red, solid) becomes much closer to the true manifold because both subject specific information and global dynamics prior is utilized.

4.5. Iterative solution

The energy function $E(Y^{cp}, \boldsymbol{\alpha}, \boldsymbol{s}_Q)$ is a quadratic form with respect to the manifold control points Y^{cp} and the standard manifold coefficients $\boldsymbol{\alpha}$, therefore, under the fixed phase $\boldsymbol{s}_Q$, Y^{cp} and $\boldsymbol{\alpha}$ are obtained by a linear solution.
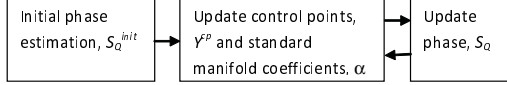
Figure 3. Iterative solution

However, the phase s_Q is a complex form, as the spline curves are switched piecewise based on phase s_Q and the interpolation ratio w appears as a sixth-order polynomial in the data term. To solve this highly nonlinear optimization problem, iterative solutions are employed as shown in Fig. 3.

Following [2], initial phase, $s_Q^{init} = \{s_{Q,i}^{init}\}$ is estimated using continuous Dynamic Programming (DP) [18] and global period P' is then computed from averaged phase evolution speed. Computed s_Q^{init} is redefined as s_Q^0 and then fed to the energy minimization framework. The manifold control points $Y^{cp,r}$ and standard manifold coefficients α^r at r-th iteration are calculated using a linear solution under the fixed phase sequence s_Q^r. Phase sequence s_Q^{r+1} at $r + 1$-th iteration under the fixed manifold control points $Y^{cp,r}$ and standard model coefficients α^r is then solved via Taylor expansion and convex quadratic programming under the monotonic increase constraint. Manifold control points $Y^{cp,r}$, the standard manifold coefficients α^r, and the phase sequence s_Q^r are alternately updated based on the previous solution (see Fig. 3), and these steps are iterated until convergence.

5. Experiments

5.1. Dataset and setup

The proposed method is evaluated with real gait sequence from the OU-ISIR Gait Database [20]. 200 real gait sequences from 100 subjects were used for evaluation. In this dataset, each subject walked on a speed-controlled treadmill and two gait sequences were captured: one for the probe and the other for the gallery. The normalized gait silhouette size, frame-rate, and recording time for each sequence are 22×32 pixels, 60 fps, and 6 sec respectively. Low frame-rate gait sequences were constructed by down sampling the original 60 fps sequences at a specified interval. We perform three experiments: temporal SR, low frame-rate gait recognition using the temporal SR itself, and sensitivity analysis of the proposed method with parameter change. We used following parameter values as default: the number of training manifolds $N^{tr} = 85$, dimension of silhouette eigenspace $M = 112$, and the parameters in the energy minimization framework $\lambda_t = 1$, $\lambda_m = 5$, $\lambda_s = 1$, and $N^{cp} = 100$.

5.2. Temporal super resolution

In this section, we compare the proposed periodic temporal SR results with level-set morphing (Morph) [3], and

Figure 4. Results of temporal SR. 1st row: input low frame-rate image sequence (1 fps), 2nd row: Morph [3], 3rd row: TSR [2], 4th row: Proposed method, and bottom row: ground truth (60 fps) image sequence (every 4 frames)
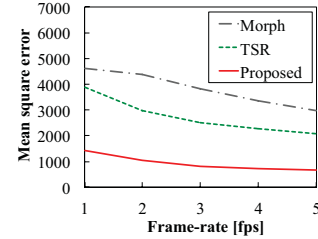

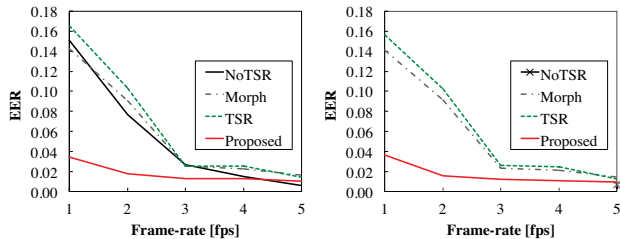Figure 5. Mean squared error of reconstructed images

temporal SR using exemplar image sequence and morphing (TSR) [2]. Figure 4 shows results of temporal SR from an 1-fps image sequence (the first row). The sampling rate of this low frame sequence is such that we could observe the same phase of a gait cycle repeatedly (stroboscopic effect). Both the level-set morphing (the second row) and TSR (the third row) cannot successfully recover a high frame-rate image sequence because the two key frames are too far apart (a single support phase and the next single support phase) and fails to interpolate intermediate gait stances (double support phases). The proposed method (the fourth row) successfully reconstructs a natural gait silhouette sequence with neither a wagon wheel effect nor stroboscopic effect, and hence looks quite similar to the ground truth normal frame-rate image sequence (the bottom row). We also report the mean squared error E for the reconstructed frames computed as:

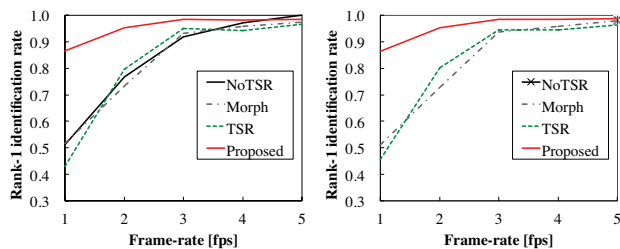$$E = \frac{1}{NWH} \sum_{k=0}^{N-1} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} [I(i,j,k) - \hat{I}(i,j,k)]^2, \quad (15)$$

where W, H, N, I, and $\hat{I}$ are the width, height, the number of images, ground truth image, and reconstructed image, respectively. As seen from Fig. 5, reconstruction error in the proposed method is much lower than TSR [2] and Morph [3].

5.3. Gait recognition

In this experiment, we apply the reconstructed high frame-rate sequences for human gait recognition, and compare the recognition performance with original low frame-rate sequence (NoTSR), morphing-based reconstruction

(a) GEI   (b) Frequency domain feature
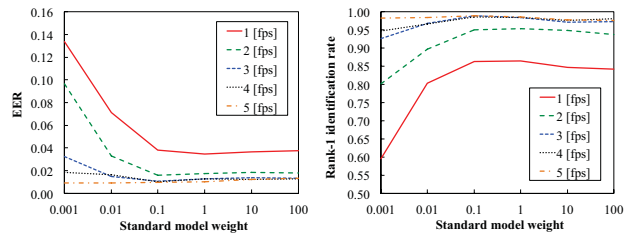Figure 6. EER with different features

(a) GEI   (b) Frequency domain feature
Figure 7. Rank-1 identification rate with different features

(a) EER   (b) Rank-1 identification rate
Figure 8. Influence of standard manifold weight on EER and Rank-1 identification rate

6. Conclusions

In this paper, we introduced a unified framework of example-based and reconstruction-based approaches to the temporal SR for low frame-rate videos. By incorporating the standard manifold obtained from the training subjects into the energy minimization framework, the periodic image sequence is successfully reconstructed even from nearly still input images due to the stroboscopic effect. Experimental results using a real gait image sequence demonstrated the effectiveness of the proposed method in terms of the recognition performance.

The proposed method assumes the same motion among the periods, which is a requirement for reconstruction-based periodic temporal SR, however, it is violated in some cases (e.g., gait fluctuations, walking speed changes, viewpoint transitions). As future works, we investigate how to extend the standard manifold to cope with these situations.

Acknowledgement

References

[1] A. Agrawal, M. Gupta, A. Veeraraghavan, and S. Narasimhan. Optimal coded sampling for temporal super-resolution. In *Proc. of of IEEE computer society conference on Computer Vision and Pattern Recognition 2010*, pages 599 –606, San Francisco, CA, USA, Jun. 2010. 2

[2] N. Akae, Y. Makihara, and Y. Yagi. Gait recognition using periodic temporal super resolution for low frame-rate videos. In *Proc. of the International Joint Conference on Biometrics*, pages 107–120, Washington, D.C., USA, Oct. 2011. 1, 2, 4, 5

[3] M. S. Al-Huseiny, S. Mahmoodi, and M. S. Nixon. Gait learning-based regenerative model: A level set approach. In *The 20th Int. Conf. on Pattern Recognition*, pages 2644–2647, Istanbul, Turkey, Aug. 2010. 1, 5

(Morph) [3] and periodic temporal SR (TSR) [2]. Two different gait features are used in these experiments: averaged silhouette [10] also known as Gait Energy Image (GEI) [7] and frequency-domain feature [12]. The performance measures used are Equal Error Rate (EER) [21] and rank-1 identification rate.

The EER and rank-1 identification rates of the proposed method along with benchmark methods are shown in Fig. 6 and Fig. 7, respectively. The horizontal axis represents frame-rates of both the probe and gallery sequences. The proposed method outperforms all of the benchmark methods, and the improvement is much more significant particularly at quite low frame-rates. Note that NoTSR results are unavailable for frequency-domain feature for frame-rates up to 4 fps due to sampling theorem.

5.4. Parameter sensitivity

In this section, we report experiments to show the effect of parameter values on the performance of the proposed method. Figure 8 shows the effect of changing the weight of the standard manifold fitness term λ_t (see Eq. 14). When λ_t is small compared to the coefficients of other terms of the energy function, impact of the standard manifolds are negligible. As a result, we cannot successfully reconstruct the part of the manifold where no input frames are available, and consequently, recognition performance drops (e.g., 0.001 or 0.01). On the other hand, we also notice that for wide range of λ_t values (e.g., 0.1 to 100), performance remains almost unchanged, which means the standard manifold coefficient is insensitive to some extent.

[4] M.-H. Cheng, M.-F. Ho, and C.-L. Huang. Gait analysis for human identification through manifold learning and hmm. *Pattern Recognition*, 41(8):2541–2553, 2008. 2

[5] T. Ding, M. Sznaier, and O. I. Camps. A rank minimization approach to video inpainting. In *Proc. of the IEEE Int. Conf. on Computer Vision*, Rio de Janeiro, Brazil, Oct. 2007. 2

[6] A. Elgammal and C. S. Lee. Separating style and content on a nonlinear manifold. In *Proc. of the 2004 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 478–485, Washington, DC, USA, Jun. 2004. 2

[7] J. Han and B. Bhanu. Individual recognition using gait energy image. *Trans. on Pattern Analysis and Machine Intelligence*, 28(2):316– 322, 2006. 2, 6

[8] M. A. Hossain, Y. Makihara, J. Wang, and Y. Yagi. Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *Pattern Recognition*, 43(6):2281–2291, Jun. 2010. 1

[9] C. S. Lee and A. Elgammal. Coupled visual and kinematic manifold models for tracking. *International Journal of Computer Vision*, 87(1-2), 2010. 2

[10] Z. Liu and S. Sarkar. Simplest representation yet for gait recognition: Averaged silhouette. In *Proc. of the 17th Int. Conf. on Pattern Recognition*, volume 1, pages 211–214, Aug. 2004. 2, 6

[11] Y. Makihara, A. Mori, and Y. Yagi. Temporal super resolution from a single quasi-periodic image sequence based on phase registration. In *Proc. of the 10th Asian Conf. on Computer Vision*, pages 107–120, Queenstown, New Zealand, Nov. 2010. 1, 2, 3

[12] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. Gait recognition using a view transformation model in the frequency domain. In *Proc. of the 9th European Conf. on Computer Vision*, pages 151–163, Graz, Austria, May 2006. 1, 2, 6

[13] Y. Makihara, A. Tsuji, and Y. Yagi. Silhouette transformation based on walking speed for gait identification. In *Proc. of the 23rd IEEE Conf. on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, Jun 2010. 1

[14] Y. Makihara and Y. Yagi. Silhouette extraction based on iterative spatio-temporal local color transformation and graph-cut segmentation. In *Proc. of the 19th Int. Conf. on Pattern Recognition*, Tampa, Florida USA, Dec. 2008. 3

[15] D. Matovski, M. Nixon, S. Mahmoodi, and J. Carter. The effect of time on the performance of gait biometrics. In *Proc. of the 4th IEEE Int. Conf. on Biometrics: Theory Applications and Systems*, pages 1–6, Washington D.C., USA, Sep. 2010. 1

[16] A. Mori, Y. Makihara, and Y. Yagi. Gait recognition using period-based phase synchronization for low frame-rate videos. In *Proc. of the 20th Int. Conf. on Pattern Recognition*, pages 2194–2197, Istanbul, Turkey, Aug. 2010. 1

[17] H. Murase and S. K. Nayar. Parametric eigenspace representation for visual learning and recognition. In *Proc. of SPIE, 2031*, 1993. 3

[18] R. Oka. Spotting method for classification of real world data. *Computer Journal*, 41(8):559–565, 1998. 5

[19] M. Okumura, H. Iwama, Y. Makihara, and Y. Yagi. Performance evaluation of vision-based gait recognition using a very large-scale gait database. In *Proc. IEEE 4th Int. Conf. on Biometrics: Theory, Applications and Systems*, pages 1–6, Washington, D.C., USA, Sep. 2010. 2

[20] Ou-isir gait database. http://www.am.sanken.osaka-u.ac.jp/GaitDB/index.html. 5

[21] 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. 6

[22] S. P. Prismall, M. S. Nixon, and J. N. Carter. Novel temporal views of moving objects for gait biometrics. In *Audio Visual Biometric Person Authentication*, pages 725–733, 2003. 1

[23] 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

[24] O. Shahar, A. Faktor, and M. Irani. Space-time super-resolution from a single video. In *Proc. of the IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Colorado Springs, USA, June 2011. 2

[25] E. Shechtman, Y. Caspi, and M. Irani. Space-time super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):531–545, Apr. 2005. 1, 2, 4

[26] M. Shimano, T. Okabe, I. Sato, and Y. Sato. Video temporal super-resolution based on self-similarity. In *Proc. of the 10th Asian Conf. on Computer Vision*, volume 1, pages 93–106, Queeenstown, New zealand, Nov. 2010. 1, 2, 4

[27] M. Singh, A. Basu, and M. K. Mandal. Event dynamics based temporal registration. *IEEE Trans. on Multimedia*, 9(5):1004–1015, 2007. 2

[28] R. Tanawongsuwan and A. Bobick. Performance analysis of time-distance gait parameters under different speeds. In *In 4th International Conference on Audio- and Video-based Biometric Person Authentication*, 2003. 1

[29] D. Tao, X. Li, X. Wu, and S. Maybank. Human carrying status in visual surveillance. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 1670–1677, New York, USA, Jun. 2006. 2

[30] R. Urtasun, D. J. Fleet, A. Hertzmann, and P. Fua. Priors for people tracking from small training sets. In *Proc. of of IEEE computer society International conference on Computer Vision 2005*, pages 403–410, Beijing, China, Oct. 2005. 2

[31] Y. Wang, S. Yu, Y. Wang, and T. Tan. Gait recognition based on fusion of multi-view gait sequences. In *Proc. of the IAPR Int. Conf. on Biometrics 2006*, pages 605–611, Jan. 2006. 1

[32] X. Xu, L. Wan, X. Liu, T.-T. Wong, L. Wang, and C.-S. Leung. Animating animal motion from still. *ACM Trans. Graphics*, 27(5), 2008. 2

[33] G. Zhao, R. Chen, G. Chen, and H. Li. Recognition of human periodic movements from unstructured information using a motion-based frequency domain approach. *Image and Vision Computing*, 24:795–809, 2006. 2