Trajectories Normalization for Viewpoint Invariant Gait Recognition

Frédéric Jean Robert Bergevin
Computer Vision and Systems Laboratory
Laval University, QC, Canada
{fjean, bergevin}@gel.ulaval.ca

Alexandra Branzan Albu
Dept. of Electrical and Computer Engineering
University of Victoria, BC, Canada
aalbu@ece.uvic.ca

Abstract

This paper proposes a method to obtain fronto-parallel (side-view) body part trajectories for a walk observed from an arbitrary view. The method is based on homography transformations computed for each gait half-cycle detected in the walk. Each homography maps the body part trajectories to a simulated side view of the walk. The proposed method is stable as the resulting normalized trajectories are not influenced by missing or omitted parts of the raw trajectories. Experiments confirm that normalized trajectories are in agreement with the ones that would be obtained from a side view.

1. Introduction

Using gait as a biometric in automatic surveillance systems is being seriously considered since it may lead to non-intrusive recognition of people at a distance. Gait is better modeled using images of the walk from a fronto-parallel (side) view. This is because important components of the motion for a walk are most visible in planes parallel to the direction of walk. Indeed, gait characteristics such as stride length can be directly extracted from images if people are observed at a distance from a fronto-parallel view (weak perspective projection) [8]. However, one may not assume that a person will always be observed from such a view in an unconstrained environment. Moreover, a person will be observed from another viewpoint each time her direction of walk is changed. Gait characteristics extracted from arbitrary views (and walk directions) are dependent on those views [7]. Motions observed in the image are indeed deformed by the perspective projection effect. A meaningful gait comparison can then only be performed with characteristics extracted from identical views.

Some methods were proposed to circumvent the view-dependence problem of the extracted gait characteristics. For instance, [6] deals with the perspective projection by using knowledge about the observed person and the camera parameters. Other methods use multiple cameras [1, 5]. However, in many real surveillance scenarios, neither the use of camera calibration nor the use of multiple views are possible. The approach proposed in [4] considers instead constraints on the lower limb motion to automatically compute the camera projection matrix. This enables a perspective correction of the observed limb trajectories. However, the limb positions are found in images using intrusive reflective markers.

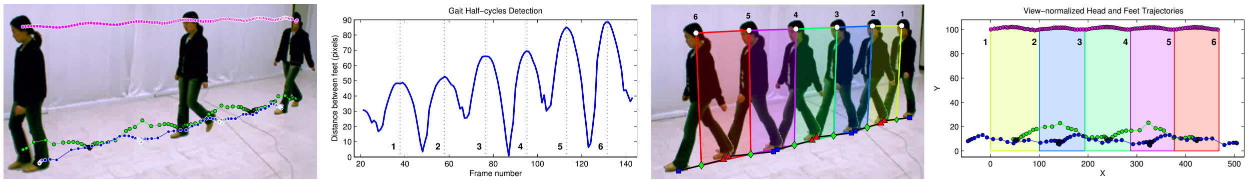
The method in [3] computes a transformation that is used to make body part trajectories appear as if they were observed from a fronto-parallel view. It is based on estimation and homographic transformation of observed walking planes. This segmentation is performed after gathering information about the direction of the walking trajectory. Consequently, the method can only be used to perform offline view normalization of body part trajectories.

The method proposed in this paper also computes homographic transformation of walking planes but on a different basis that is suitable for real-time gait modeling and recognition. The proposed view normalization is said to be stable since the resulting trajectories are not influenced by missing or removed parts of the initially observed trajectories. It is non-intrusive and fully automatic since it requires no a priori knowledge of the observed person or the camera parameters.

The remainder of the paper is organized as follows. The proposed normalization method is presented in Section 2. Experimental results are presented in Section 3. Finally, Section 4 draws conclusions and outlines the main directions of future work.

2. Proposed Approach

An overview of the proposed method is shown in Figure 1. The three raw body part trajectories in Fig-



(a) Body parts tracking (b) Gait half-cycles detection (c) Planes computation (d) View normalization

Figure 1. Overview of the proposed method.

ure 1(a) are obtained from silhouettes extracted at each frame of a monocular video sequence [2]. In this paper, only the head and the feet trajectories are considered since they are the only trajectories required by the proposed method. Other body part trajectories could be view-normalized as well. The temporal variation of the inter-feet distance is then analyzed in order to determine the moments at which the feet are furthest apart (maximum distance moments), as shown in Figure 1(b). Here, there are six of these moments. Time intervals between two consecutive moments define a gait half-cycle. A view normalization is to be performed in each of these half-cycles. Figure 1(c) shows the estimated "walking plane" in each gait half-cycle, with image pixels for extracted silhouettes at the maximum distance moments. A walking plane can be considered here as the $2D$ projection of the $3D$ sagittal plane, that is the plane along the direction of the walk passing through the center of mass of a person. The four positions defining a walking plane are estimated by the head and the middle feet positions on the floor at the beginning and the end of a gait half-cycle. If a person were observed from a fronto-parallel view, those planes would appear as rectangles. A "normalized" version of the planes from an arbitrary view has right angles, as shown in Figure 1(d). A homography between an observed plane and its normalized version can then be computed and used in order to "normalize" the body part trajectories within the corresponding gait half-cycle.

The following sections present a detailed description of each step of the proposed method.

2.1. Preprocessing

The tracking of the head and the feet is performed with the method presented in [2]. This method can be used in real-time and provides at each frame t, for $t \in [1, T]$, a mean position (mass center) of the head $\mathbf{p}_{\mathrm{h}}(t)$, and the feet $\mathbf{p}_{\mathrm{f1}}(t)$, $\mathbf{p}_{\mathrm{f2}}(t)$ (see Figure 1(a)). It also provides the observed contact points of each foot on the floor, $\mathbf{p}_{\mathrm{c1}}(t)$ and $\mathbf{p}_{\mathrm{c2}}(t)$ (red triangles and blue squares in Figure 1(c)). The method analyses previously extracted silhouettes in order to find the positions of the body parts. Optical flow is used to handle feet occlusions and thus maintain feet correspondence between frames.

2.2. Gait Half-Cycles Detection

A gait half-cycle is defined as the interval between the two consecutive moments where the feet are furthest apart from each other. These maximum distance moments are detected by finding the maximums of a function $d(t)$ representing the apparent distance between the feet over time (blue curve in Figure 1(b)):

$$d(t) = G(t) * \|\mathbf{p}_{\mathrm{f1}}(t) - \mathbf{p}_{\mathrm{f2}}(t)\|, \tag{1}$$

where $\|\cdot\|$ is the Euclidean norm. $G(t)$ is a gaussian kernel used for filtering the function $d(t)$, with $\sigma = 1$ and a kernel width of 5. A maximum in the function $d(t)$ is detected at time t^* if

$$t^* = \arg\max_t \{d(t)\}\,;\ t \in [t^* - a, t^* + a]\,, \tag{2}$$

where $a = \lfloor W/2 \rfloor$ and W denotes the size of a window centered at t^*. W is defined as $W = \lfloor f_s/2f_g \rfloor + 1$ if $\lfloor f_s/2f_g \rfloor$ is even, or $W = \lfloor f_s/2f_g \rfloor$ otherwise. The parameter f_s is the camera acquisition frame rate, in frames per second. The parameter f_g is the maximum gait cadence (in Hz) to be considered, and represents the maximum number of complete gait cycles performed in one second. In this paper, $f_s = 30$ fps and f_g is set to 2 Hz, which allows for detecting the maximum distance moments for typical speed walks.

The maximum distance moments t_i^*, $i = 1 \ldots N$, detected in $d(t)$ are then refined to moments γ_i with sub-frame accuracy using parabolic interpolation:

$$\gamma_i = t_i^* + \frac{d(t_i^* - 1) - d(t_i^* + 1)}{2\{d(t_i^* + 1) - 2d(t_i^*) + d(t_i^* - 1)\}}. \tag{3}$$

These refined moments are then used to define the gait half-cycles as intervals $\mathcal{C}_i = [\gamma_i, \gamma_{i+1}]$, for $i = 1 \ldots N-1$. The detected moments are shown as vertical dotted lines in Figure 1(b). Sub-frame accuracy may reduce the variance observed in the half-cycles duration, especially in the case of lower frame rates. It also simplifies the assigning of frames on the boundary of two consecutive half-cycles to one of them.

2.3. Walking Planes Computation

The walking planes are defined in the images such that they represent locally the direction of walk of the observed person. One way to achieve this is to define a plane for each half-cycle using representative positions at the beginning and the end of the half-cycle. Here, these positions are chosen as the head and middle feet position on the floor $\mathbf{p}_{\mathrm{m}}(t) = \{\mathbf{p}_{\mathrm{c1}}(t) + \mathbf{p}_{\mathrm{c2}}(t)\}/2$, as shown in Figure 1(c). The latter is an approximation of the projection of the head position on the floor. A plane Π_i is then defined as the set of four points corresponding to the top and bottom positions at the beginning and the end of the half-cycle:

$$\Pi_i : \{\mathbf{p}_{\mathrm{h}}(\gamma_i), \mathbf{p}_{\mathrm{m}}(\gamma_i), \mathbf{p}_{\mathrm{h}}(\gamma_{i+1}), \mathbf{p}_{\mathrm{m}}(\gamma_{i+1})\}\,. \quad (4)$$

These positions are linearly interpolated at a non-integer time value γ as follows:

$$\mathbf{p}(\gamma) = \mathbf{p}(\lfloor\gamma\rfloor) + (\gamma - \lfloor\gamma\rfloor)\{\mathbf{p}(\lceil\gamma\rceil) - \mathbf{p}(\lfloor\gamma\rfloor)\}\,. \quad (5)$$

It is important to notice that defining the walking planes on a half-cycle basis makes them independent of each other. Consequently, the method provides stable results for each complete half-cycle

2.4. View Normalization

In order to perform the view normalization of the body part trajectories, a normalized version of the planes is defined. These normalized versions of the planes represent their shape in a fronto-parallel view, which are rectangles with the same heights and right angles, as shown in Figure 1(d). The normalized version of the planes are defined as

$$\widetilde{\Pi}_i : \left\{ \begin{bmatrix} \alpha_{i-1} \\ \beta \end{bmatrix}, \begin{bmatrix} \alpha_{i-1} \\ 0 \end{bmatrix}, \begin{bmatrix} \alpha_i \\ \beta \end{bmatrix}, \begin{bmatrix} \alpha_i \\ 0 \end{bmatrix} \right\}, \quad (6)$$

where β is an arbitrary height (here $\beta = 100$ units). The horizontal positions at the beginning and the end of the corresponding half-cycle are defined as $\alpha_i = \tau \sum_{j=1}^{i}(\gamma_{j+1} - \gamma_j)$, with the special case $\alpha_0 = 0$. This makes consecutive normalized planes connected, which is necessary to obtain continuous normalized trajectories. The normalized planes' width are also defined to be proportional to the duration of their corresponding gait half-cycle. It is assumed in this paper that the velocity of the performed walk is constant. Consequently, the parameter τ, which is expressed in units per frame, can be set to an arbitrary value (here $\tau = 5$).

A homography $\mathbf{H}_i$ that maps a walking plane Π_i to its normalized version $\widetilde{\Pi}_i$ is then computed. This homography is used on the observed body part trajectories $\mathbf{p}_{\mathrm{b}}(t)$, with $\mathrm{b} \in \{\mathrm{h}, \mathrm{f1}, \mathrm{f2}\}$. A normalized version $\widetilde{\mathbf{p}}_{\mathrm{b}}(t)$ within half-cycle $\mathcal{C}_i$ is then obtained as follows:

$$\lambda \begin{bmatrix} \widetilde{\mathbf{p}}_{\mathrm{b}}(t) \\ 1 \end{bmatrix} = \mathbf{H}_i \begin{bmatrix} \mathbf{p}_{\mathrm{b}}(t) \\ 1 \end{bmatrix};\ t \in [\lceil\gamma_i\rceil, \lfloor\gamma_{i+1}\rfloor], \quad (7)$$

where λ is a scale factor, and $i = 1 \dots N-1$. One can see that each body part trajectory is view-normalized independently within each gait half-cycle using the corresponding homography.

3. Experimental Results

Two experiments are conducted in order to assess the performance of the proposed method. The first experiment evaluates the obtained view-normalized trajectories by performing a comparison with trajectories from a fronto-parallel view. This will show that the obtained trajectories effectively appear to be observed from a fronto-parallel view. In the second experiment, the stability of the proposed method is demonstrated and compared with results obtained with another view-normalization method. Both experiments are performed using a database composed of video sequences of ten human subjects walking in straight line. The subjects were observed simultaneously by four synchronized cameras. Each camera optical axis crosses the walking trajectory at a different angle: $90°$ (fronto-parallel view), $75°$, $60°$, and $45°$. The subjects walked once back and forth in front of the cameras, which leads to a database of 80 video sequences (10 subject $\times$ 4 views $\times$ 2 intervals).

Body part trajectories obtained from an arbitrary view and from a fronto-parallel view can be compared by performing an optimal alignment (rigid transformation) between them. The parts of the walk observed at the same time by both cameras are known from camera synchronization, as well as the point correspondence between the views. In the first experiment, the alignment process is first performed on original (raw) trajectories observed from arbitrary and fronto-parallel views. The alignment process is also performed on the normalized trajectories obtained with the method proposed in this paper. Figure 2 shows the alignment obtained with the raw and normalized head trajectories for the $90°$ and $45°$ views. One may see that normalized trajectories from two different views are better aligned than raw trajectories. This means that gait characteristics extracted from those normalized trajectories would be more suitable for view-invariant gait recognition.

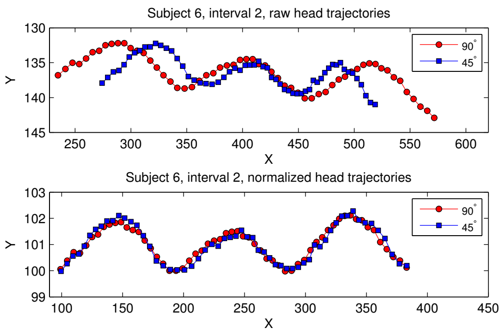


Figure 2. Alignment examples

Table 1 shows the root mean square errors statistics obtained over all subjects. Here, the head and feet trajectories from the $75°$, $60°$ and $45°$ views are aligned with the ones from the fronto-parallel view. Values are shown for alignment of pairs of raw trajectories and two pairs of normalized trajectories, one obtained with the method presented in [3] (old), and the other obtained with the method presented in this paper (new). Prior to computing the alignment error, each trajectory is rescaled independently in x and y in order to fit inside a 1×1 box. Values in Table 1 are expressed in these scale-independent units. The values are much lower for normalized trajectories than raw trajectories, especially when the difference between the view angles is greater. This confirms that normalization effectively reproduces a side view of the walk. In most cases the method proposed in this paper leads to slightly better values than the method presented in [3]. One reason may be that the new method uses planes that better represent the direction of motion.

The stability of the new method is confirmed by masking in turn the first and the last quarter of the observed trajectories. A normalized trajectory is computed from each partial trajectory and aligned with its corresponding part of the complete normalized trajectory. Because of the nature of the new method, its partial and complete normalized trajectories match perfectly. This is not the case with the partial and complete normalized trajectories obtained with the method in [3]. The mean scale-independent error obtained on 240 trajectories (10 subjects $\times$ 4 views $\times$ 2 intervals $\times$ 3 body parts) is 0.0038 units (standard deviation 0.0046). The proposed method is better suited to be used in a real-time system where consistent normalization needs to be performed in real-time at each new frame.

Statistic	Type	90deg - 75deg			90deg - 60deg			90deg - 45deg		
		Head	Foot1	Foot2	Head	Foot1	Foot2	Head	Foot1	Foot2
mean	raw	1.17	1.94	2.15	2.30	3.00	3.33	3.60	4.62	5.00
	old	1.05	2.00	2.25	1.44	2.66	2.74	1.57	3.37	3.52
	new	0.87	1.98	2.05	1.02	2.59	2.49	1.30	3.52	3.29
median	raw	1.09	1.89	1.82	2.24	2.96	3.25	3.59	4.26	4.76
	old	0.96	1.97	2.21	1.45	2.61	2.73	1.55	3.12	2.97
	new	0.82	1.88	1.82	1.00	2.57	2.28	1.27	3.19	2.75
std	raw	0.25	0.59	0.74	0.28	0.70	0.67	0.43	1.22	1.12
	old	0.38	0.55	0.81	0.43	0.87	0.89	0.41	1.24	1.39
	new	0.22	0.63	0.80	0.23	0.78	0.88	0.23	1.27	1.43
min	raw	0.70	1.03	1.13	1.81	2.03	2.33	2.91	3.35	3.57
	old	0.49	1.04	1.08	0.63	1.50	1.59	0.91	2.16	1.81
	new	0.58	0.71	1.12	0.63	1.60	1.60	0.95	1.76	1.78
max	raw	1.77	3.31	3.93	2.85	5.38	5.05	4.54	8.25	7.50
	old	2.24	3.16	3.85	2.14	5.01	4.86	2.50	7.95	6.29
	new	1.38	3.01	3.95	1.42	5.02	4.94	1.74	7.93	6.02

Table 1. Statistics on alignment ($\times 10^2$)

4. Conclusion

A method to obtain a simulated fronto-parallel view of $2D$ body part trajectories has been presented. Walking planes are computed from head and feet trajectories and used to compute normalizing homography transformations. The method is fully automatic and does not need a priori information on the observed people or the camera calibration parameters. Moreover, the stability of the method makes it suitable for use in a real-time system. Future work will be focused on gait characteristics extraction and gait recognition.

Acknowledgments: This work is supported by Natural Sciences and Engineering Research Council of Canada, Fonds Québécois de la Recherche sur la Nature et les Technologies, and by Precarn inc.

References

[1] F. Cuzzolin. Using bilinear models for view-invariant action and identity recognition. In *Proc. of CVPR*, volume 2, pages 1701– 1708, 2006.

[2] F. Jean, R. Bergevin, and A. B. Albu. Body tracking in human walk from monocular video sequences. In *Proc. of Computer and Robot Vision*, pages 144 – 151, 2005.

[3] F. Jean, R. Bergevin, and A. B. Albu. Computing view-normalized body parts trajectories. In *Proc. of Computer and Robot Vision*, pages 89–96, 2007.

[4] M. Goffredo et al. *Human Perambulation as a Self Calibrating Biometric*, volume LNCS 4778, pages 139–153. 2007.

[5] A. Tyagi, J. W. Davis, and M. Keck. Multiview fusion for canonical view generation based on homography constraints. In *Proc. of ACM Workshop on Video surveillance and sensor networks*, pages 61–70, 2006.

[6] Y. Makihara et al. Adaptation to walking direction changes for gait identification. In *Proc. of ICPR*, volume 2, pages 96–99, 2006.

[7] S. Yu, L. Wang, W. Hu, and T. Tan. Gait analysis for human identification in frequency domain. In *Proc. of Conf. on Image and Graphics*, pages 282–285, 2004.

[8] R. Zhang, C. Vogler, and D. Metaxas. Human gait recognition. In *Proc. of CVPR*, volume 1, pages 18–25, 2004.