Part-based Collaborative Spatio-temporal Feature Learning for Cloth-changing Gait Recognition

Lingxiang Yao[1], Worapan Kusakunniran[2], Qiang Wu[1], Jian Zhang[1], Jingsong Xu[1]

[1]School of Electrical and Data Engineering, University of Technology Sydney

[2]Faculty of Information and Communication Technology, Mahidol University

Introduction

- Gait provides a non-contact way of recognizing a target person in a distance without any co-operation, which enables gait to be widely used in security surveillance and forensic authentication.
- The result of image/video-based gait recognition is influenced by exterior factors. Among these factors, clothing changes can be regarded as one of the most challenging factors for gait recognition.
- A part-based collaborative spatiotemporal feature learning method is proposed in this paper for cloth-changing gait recognition. Spatial and temporal features are separately generated from the $H-W$ and $T-W$ views through 2D convolutions. A collaborative spatio-temporal gait feature is produced by assembling these two features together.
- The proposed method can obtain the state-of-the-art result for cloth-changing gait recognition on two most generally known gait datasets, CASIA Gait Dataset B and OU-ISIR Treadmill Gait Dataset B.

Methodology (1)

Method Overview

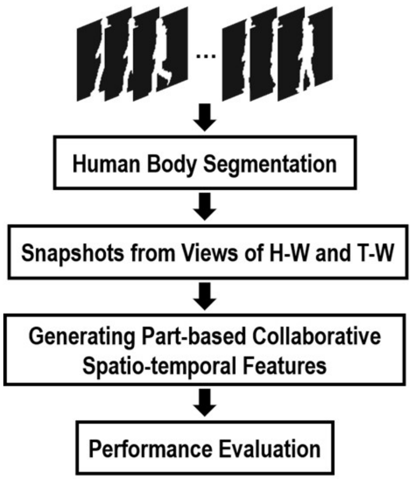

Fig. 1 Flowchart of the proposed method.

- For each sequence, the human body of each frame is first segmented into two regions, the affected and unaffected regions (Fig.2). After that, the snapshots of the segmented unaffected regions are produced by projecting these regions from the $H-W$ and $T-W$ views (Fig.3).. Our proposed part-based collaborative spatial-temporal features are developed from these snapshots through a well-designed network (Fig.4). The operations of segmentation and projection can be treated as the preparatory process and have been organically integrated into our proposed network.

Human Body Segmentation

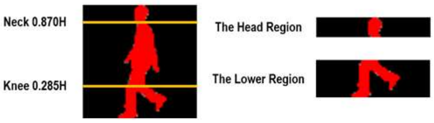

Fig. 2 Segmentation of human bodies.

- Technically, human bodies can be separated according to the known anatomical properties. For a body height, we can segment the human body based on some vertical positions, e.g., neck, waist, pelvis, and knees.
- For cloth-changing gait recognition, the upper bodies are highly influenced by clothing changes. Therefore, in our method we just concentrate on two regions, the head region covering from the top of one's head to one's neck and the lower region covering from their knees to the ground. In addition, the segmented two regions in our method are actually a little wider than they ought to be.

Methodology (2)

Snapshots from View of T −W

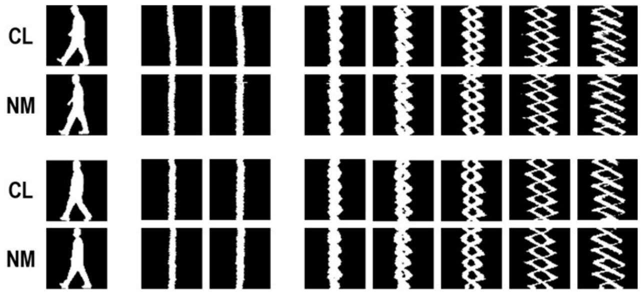

Fig. 3 Samples of snapshots from the $T-W$.

- As the most familiar viewing angle, view of $H-W$ provides a direct manner of depicting human gaits in the spatial domain. Each frame sequence captured by one camera can be treated as a snapshot collection generated from this view.
- View of $T-W$ is a rare view in the real world. It mainly captures the displacement of a horizontal section across a period. Collecting gait snapshots from this view is much similar as generating gait signals using wearable devices. Both methods can offer a motion portrait about how signals change as time goes by.
- Also, many successful practices in spatial feature learning can be easily introduced to the temporal domain, which contributes to improving the expression ability of our generated temporal features.

Part-Based Collaborative Feature Learning

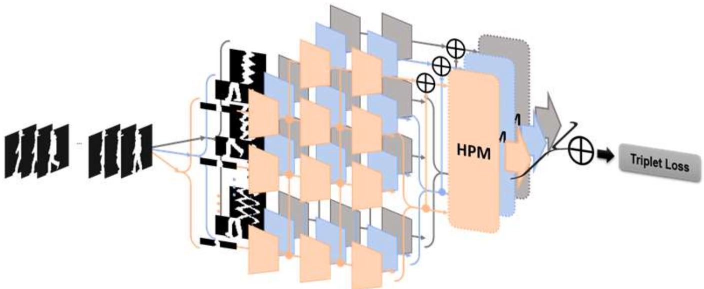

Fig. 4 Structure of the proposed network.

- In our method, the part-based collaborative spatiotemporal features are produced by inputting the snapshots from the $H-W$ ad $T-W$ views, because they depict gaits from two different perspectives and there exists a connection between them.

Experiments

- The robustness and effectiveness of the proposed method are verified by relevant experiments on CASIA Gait Dataset B and OU-ISIR Treadmill Gait Dataset B.
- More specific details can be found in our paper.

Table.1 Averaged rank-1 accuracy on CASIA-B under different experimental settings, excluding identical-view cases.

Gallery NM#1-4		0° -180°											
Probe CL#1-2		0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	Mean
ST	GaitSet[1] (30frames)	29.4	43.1	49.5	48.7	42.3	40.3	44.9	47.4	43.0	35.7	25.6	40.9
	Ours (30frames)	**34.0**	**47.1**	**51.0**	**54.0**	**52.9**	**48.9**	**49.8**	**50.3**	**48.2**	**41.4**	**30.5**	**46.2**
	GaitSet[1] (64frames)	37.4	50.1	54.2	52.0	49.6	44.9	47.9	48.6	46.6	40.0	29.3	45.5
	Ours (64frames)	**38.1**	**52.3**	**57.9**	**59.1**	**56.2**	**51.3**	**53.8**	**56.6**	**56.3**	**48.0**	**31.2**	**51.0**
MT	AE [2]	18.7	21.0	25.0	25.1	25.0	26.3	28.7	30.0	23.6	23.4	19.0	24.2
	MGAN [3]	23.1	34.5	36.3	33.3	32.9	32.7	34.2	37.6	33.7	26.7	21.0	31.5
	GaitSet[1] (30frames)	52.0	66.0	72.8	69.3	63.1	61.2	63.5	66.5	67.5	60.0	45.9	62.5
	Ours (30frames)	**59.2**	**74.7**	**77.4**	**74.5**	**69.5**	**66.3**	**69.8**	**74.4**	**73.6**	**69.2**	**52.5**	**69.2**
	GaitSet[1] (64frames)	63.8	72.5	78.0	76.8	67.3	64.4	67.1	71.2	71.7	68.3	52.7	68.5
	Ours (64frames)	**61.8**	**77.6**	**83.1**	**80.4**	**74.3**	**70.5**	**75.7**	**80.8**	**81.1**	**74.9**	**54.9**	**73.4**
LT	CNN-LB [4]	37.7	57.2	66.6	61.1	55.2	54.6	55.2	59.1	58.9	48.8	39.4	54.0
	GaitNet [5]	42.1	-	-	70.7	-	**70.6**	-	69.4	-	-	-	63.2
	GaitSet[1] (30frames)	61.4	75.4	80.7	77.3	72.1	70.1	71.5	73.5	73.5	68.4	50.0	70.4
	Ours (30frames)	**64.2**	**80.9**	**83.0**	**79.5**	**74.3**	69.1	**74.8**	**78.5**	**81.0**	**77.0**	**60.3**	**74.8**
	GaitSet[1] (64frames)	69.3	82.4	83.3	78.7	74.3	70.5	74.9	78.0	77.6	74.7	60.8	75.0
	Ours (64frames)	**71.8**	**86.6**	**87.7**	**83.2**	**78.3**	**75.4**	**81.0**	**85.2**	**84.9**	**82.0**	**64.1**	**80.0**

Contact

Worapan Kusakunniran

Faculty of Information and Communication Technology, Mahidol University

Email: worapan.kun@mahidol.edu

Website: https://sites.google.com/a/mahidol.edu/worapan-kusakunniran/home

Phone: (66) 02 441 0909, (66) 02 354 4333

References

1. H. Chao, Y. He, J. Zhang, and J. Feng, Gaitset: Regarding gait as a set for cross-view gait recognition.
2. S. Yu, H. Chen, Q. Wang, L. Shen, and Y. Huang, Invariant feature extraction for gait recognition using only one uniform model.
3. Y. He, J. Zhang, H. Shan, and L. Wang, Multi-task gans for view-specific feature learning in gait recognition.
4. Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan, A comprehensive study on cross-view gait based human identification with deep cnns.
5. Z. Zhang, L. Tran, X. Yin, Y. Atoum, X. Liu, J. Wan, and N. Wang, Gait recognition via disentangled representation learning.

25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 2021