Using Individuality to Track Individuals: Clustering Individual Trajectories in Crowds using Local Appearance and Frequency Trait

Daisuke Sugimura
The University of Tokyo
Tokyo 153–8505, Japan
sugimura@iis.u-tokyo.ac.jp

Kris M. Kitani
The University of Electro-Communications
Tokyo 182–8585, Japan
kitani@is.uec.ac.jp

Takahiro Okabe
The University of Tokyo
Tokyo 153–8505, Japan
takahiro@iis.u-tokyo.ac.jp

Yoichi Sato
The University of Tokyo
Tokyo 153–8505, Japan
ysato@iis.u-tokyo.ac.jp

Akihiro Sugimoto
National Institute of Informatics
Tokyo 101–8430, Japan
sugimoto@nii.ac.jp

Abstract

In this work, we propose a method for tracking individuals in crowds. Our method is based on a trajectory-based clustering approach that groups trajectories of image features that belong to the same person. The key novelty of our method is to make use of a person's individuality, that is, the gait features and the temporal consistency of local appearance to track each individual in a crowd. Gait features in the frequency domain have been shown to be an effective biometric cue in discriminating between individuals, and our method uses such features for tracking people in crowds for the first time. Unlike existing trajectory-based tracking methods, our method evaluates the dissimilarity of trajectories with respect to a group of three adjacent trajectories. In this way, we incorporate the temporal consistency of local patch appearance to differentiate trajectories of multiple people moving in close proximity. Our experiments show that the use of gait features and the temporal consistency of local appearance contributes to significant performance improvement in tracking people in crowded scenes.

1. Introduction

Tracking individuals in a crowded scene (Figure 1) presents new challenges that must be addressed to achieve robust and accurate tracking. In particular, it is difficult to track people who are moving in similar directions, with similar speed and frequent partial occlusions.

Due to partial occlusion (example in Figure 1), it is very unlikely that the entire body is observed over an entire video sequence. This makes it difficult to acquire meaningful boundaries of foreground objects due to the presence of many overlapping people. Therefore, standard appearance-based techniques and model-based techniques, such as background subtraction-based blob detection and time series filtering, that use whole appearance are an inadequate means of tracking people in crowds.

Figure 1. An example of a crowded scene.

Recently, in contrast to standard appearance-based and model-based techniques, approaches that make use of the motion of local feature points have been proposed [2, 10, 13]. These approaches assume that feature points that belong to the same person are likely to have similar motion. As such, tracking people is achieved by clustering local feature trajectories based on the similarity of motion between trajectories and the spatial proximity of local feature trajectories. In general, using local feature trajectories makes a system more robust to partial occlusion, as long as some portion of the person is observed.

In crowds, however, the motion trajectories of neighboring individuals tend to be very similar because they are constrained by the motion of the crowd. Consequently, tracking methods based on motion similarity and spatial proximity

2009 IEEE 12th International Conference on Computer Vision (ICCV)
978-1-4244-4419-9/09/$25.00 ©2009 IEEE

have difficulty in discriminating between individuals walking in close proximity.

To overcome this problem, we present a novel approach to tracking that characterizes the *individuality* (i.e., unique spatio-temporal traits) of a person by using: (1) gait features in the frequency domain and (2) the temporal variation of local appearance patches. By utilizing a trajectory clustering-based tracking framework, we leverage a person's individuality to efficiently track individuals in a crowd.

The use of temporal gait characteristics in the frequency domain is more discriminative than using the distance between motion trajectories, because two people walking together in the same direction can be differentiated by simply observing the frequency of an individual's stride. In fact, the discriminative power of gait frequency features has made it a standard measure for identifying people in biometrics [11].

We also measure the change (or consistency) of local appearance over time to differentiate between neighboring individuals even when their gait frequencies are very similar. We extract a sequence of local triangular patches (encompassed by three feature trajectories) and measure the variation in appearance of the local patch over time. The assumption is that the appearance of a small patch on a single person is expected to be consistent over time, while patches that straddle two people will vary heavily over time.

Herein lies the two-fold contribution of this work. (1) We use the gait features in the frequency domain to help identify individuals in a crowd. To our best knowledge, this work is the first to use gait features in the context of tracking people in a crowd. (2) We use the temporal consistency of local appearance patches to discover regions that belong to the same individual. In contrast to previous trajectory-based methods [2, 10, 13], our method utilizes mid-level visual features constructed from a group of local features (i.e., a sequence of image patches created by three adjacent trajectories) instead of individual low-level features.

2. Related works

A majority of previous work uses models of body parts to enable the tracking algorithm to deal with frequent partial occlusions. Zhao [19] addressed minor partial occlusions by fitting a part-based model of the human body against partial observations to track the movement of a human body. Several learning schemes have also been proposed that use part-based detectors to track partially occluded people [9, 14, 18]. These approaches were shown to be effective in detecting pedestrians in cluttered outdoor scenes. However, these methods also require that models be learned in advance.

In Dong *et al*.'s work [5], multiple pedestrian detection was achieved by estimating the number and positions of people from a large foreground blob by using a shape model of the human body. Khan *et al*. [7] proposed a homography-based method using multiple cameras to fit a predefined human body model to the foreground blob. However, when dealing with crowded situations it may be difficult to acquire meaningful foreground blobs.

To deal with the uncertainty involved with detecting people, a number of particle filtering algorithms [6, 8, 12] have also been proposed. However, the performance of these systems is degraded when a significant portion of the target is not visible due to heavy occlusion. Work using optical flow [3, 15] has also been proposed to compute the motion of local features to make the system more robust to partial occlusions. However, like other methods they do not take into account the rich information encoded in feature trajectories taken over a longer duration of time.

Ali *et al*. tracked individuals by using transition probabilities based on floor fields which describe the macroscopic movement of people in a crowded scene [1]. Methods using floor fields usually target extremely dense crowds (e.g., 100's - 1000s of people running a marathon) where the individuals are very small and motion is restricted by dominant motion flow. In contrast, we target crowds under 50 people, like crowds observed at a busy intersection or a crowded subway platform.

In other approaches, Brostow *et al*. tracked people in crowds by clustering trajectories of local image feature points under a Bayesian framework [2]. Specifically, the spatial proximity between a pair of feature trajectories are used to compute a prior probability, while the likelihood is calculated using the *motion coherency* between feature trajectories. Their experiments showed robust tracking of people in a crowded subway station. Rabaud *et al*. [13] also segmented people in crowds by using motion coherency and identified individuals using the assumption that features from a single person will have the same affine motion. Li *et al*. implemented a similar framework as Brostow but also incorporated a mechanism to learn the likelihood function of the motion coherency metric [10]. However, when we deal with dense crowds of people, the use of motion coherency (similarity between motion trajectories) is not sufficient because individuals are constrained to move together.

Therefore, in contrast to previous work we introduce two new metrics, namely the gait frequency features and the temporal consistency of local appearance patches, to enable our system to deal with dense crowds where motion trajectories of individuals are similar.

3. Clustering feature trajectories to track individuals

We track individuals in a crowd by clustering a graph representing a set of feature trajectories extracted from each

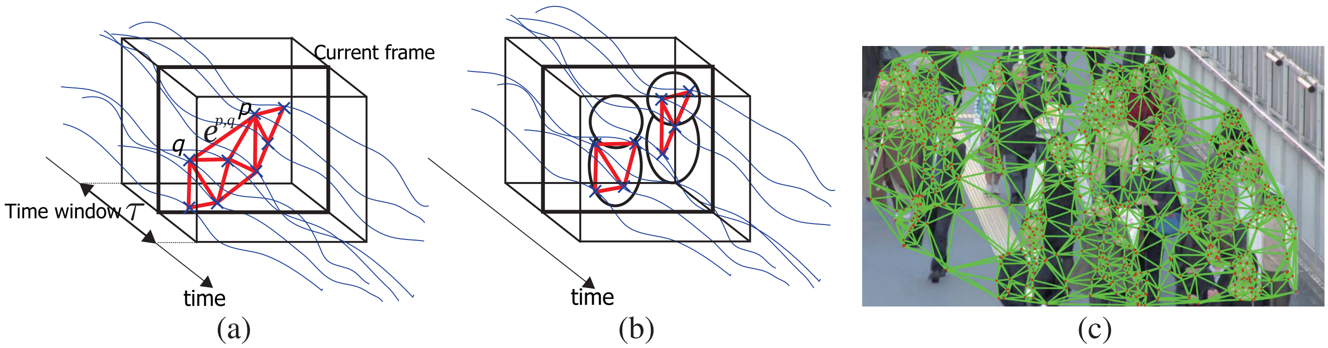

Figure 2. Graph clustering: (a) The initial graph is constructed by using the set of feature trajectories in the time window τ. (b) The final result by pruning the edges with its edge weight, where the circles represents clusters (individuals). (c) An example of the initial connected graph in a real sequence. Each red node represents a feature trajectory and the blue lines are the set of edges in the graph.

time window of a video sequence (Figure 2).

First, the KLT tracker is used to generate a set of trajectories of feature points [16, 17] within a specified time window of a video sequence. New feature points are also generated for every frame within the time window to ensure that a sufficient number of trajectories are extracted from the sequence, as was done in [13]. The set of trajectories is then reduced by eliminating trajectories that are relatively static (background features), trajectories with unreasonably high velocity and trajectories with short duration.

The initial graph is constructed from the resulting set of trajectories. Each node of the graph corresponds to a separate trajectory, and the edges connecting the nodes are obtained via Delaunay triangulation [4]. An example of the graph is shown in Figure 2-(c).

An edge connecting two nodes p and q is assigned a weight representing the dissimilarity between the corresponding trajectories based on the four metrics: (1) gait features, (2) temporal consistency of local patch appearance, (3) spatial proximity and (4) coherency of motion (similar to [2]). Assuming that we are given these weights, the total edge weight is computed as the product of each individual weight.

$$e^{p,q} = e^{p,q}_{freq} \cdot e^{p,q}_{app} \cdot e^{p,q}_{prox} \cdot e^{p,q}_{coh}\ , \qquad (1)$$

where the right hand side of (1) denotes the weight based on the gait feature $e^{p,q}_{freq}$, the local appearance weight $e^{p,q}_{app}$, the the spatial proximity $e^{p,q}_{prox}$ and coherency of motion $e^{p,q}_{coh}$, respectively. The details of these metrics are given in section 4. Note that, if one of the individual weights is zero, the dissimilarity $e^{p,q}$ always becomes zero regardless of the dissimilarities based on other metrics. To avoid this, each individual weight is replaced with a predetermined value $e_{\min}$ if the weight is smaller than a certain value.

The graph is then clustered into connected sub-graphs, each of which corresponds to each individual moving in the scene. This is done by simply pruning a set of edges that have high edge weights, i.e., high dissimilarity scores, by using a threshold th_p. Figure 2 (a) and (b) show an example of a graph before and after pruning. Currently our method processes a set of feature trajectories bounded by a finite time window τ, which spans equally forwards and backwards in time with respect to the current frame. The time window is shifted by $\tau/4$ to process a video sequence, which means that the clustering is performed every $\tau/4$ frames.

4. Discovering individuality via edge weights

The edge weight between two trajectories (nodes) are computed based on gait features, local appearance, spatial proximity and motion coherency, to determine whether they belong to the same person or not. We explain these measures in this section and show how these measurements can characterize the unique traits of individuals being tracked.

4.1. Gait features in the frequency domain

Gait features in the frequency domain are highly discriminative and can be effective in characterizing the individuality of a person in a crowd. In our method, we use the term "gait" in a broad sense to refer to the general manner in which a person walks. In particular, we utilize that periodic motion of an individual along with the vertical axis as the gait feature.

To determine the gait feature for a given trajectory, we first fit a line to the trajectory via linear regression (depicted in Figure 3) and then extract the periodic component $y_p(t)$ of the trajectory. We then apply the Fast Fourier Transform (FFT) to the residual periodic signal $y_p(t)$ to find the amplitude spectra and phase of the trajectory.

The dissimilarity of a pair of trajectories is evaluated using both the dissimilarity between the amplitude spectra and the dissimilarity between the peaks for each frequency's phase. The Euclidean distance, which is commonly used in biometrics for identification of people (e.g. [11]), is used to measure the dissimilarity of the amplitude spectra. The

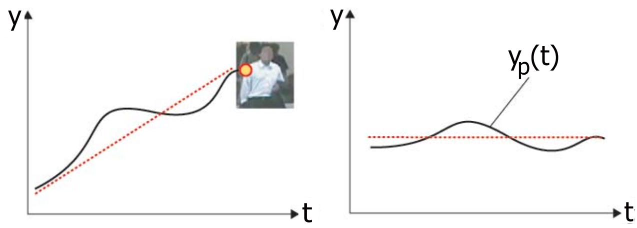

Figure 3. Pre-processing for measuring the gait features in frequency domain. Line fitting for removing linear motion component of feature trajectory (left image). Extracted periodic motion $y_p(t)$ (right image).

final dissimilarity metric of a pair of trajectories for the gait features $e^{p,q}_{freq}$ is given as

$$e^{p,q}_{freq} = \sqrt{\sum_{k=0}^{\tau/2} [a^p_k - a^q_k]^2 \cdot (|\phi^p - \phi^q|)} \ , \tag{2}$$

where a^p_k and ϕ^p denote the k-th component of the amplitude spectrum and the phase of the peak frequency with the largest amplitude of the p-th feature trajectory, respectively. Note that we only need to consider a single side of the frequency band for a real signal.

4.2. Temporal variation of local appearance

While gait features are highly effective for characterizing pedestrians in a crowd, gait features for neighboring pedestrians can still become similar when the crowd is dense. Therefore, we introduce another metric based on the temporal variation of local appearance to deal with such cases.

In our method, the temporal variation of local appearance is measured to quantify the change in appearance of a sequence of small triangular patches bounded by three adjacent trajectories. A visualization of a sequence of triangular patches is given in Figure 4. A sequence of patches corresponding to a region straddling two people is expected to have higher variation when compared to a sequence of patches that are extracted from one individual. Here we use the temporal variation of a hue-saturation color histogram across the sequence of patches. To be specific, the RMS of the Bhattacharyya distance between the color histogram $\mathbf{h}(t)$ and the average color histogram $\bar{\mathbf{h}}$ of the patch within the time window, is used to define the dissimilarity $e^{p,q}_{app}$ for a sequence of patches as

$$e^{p,q}_{app} = \sqrt{\frac{1}{\tau} \sum_{t=t_s}^{t_s+\tau-1} d^2_{hist}(\mathbf{h}(t), \bar{\mathbf{h}})} \ , \tag{3}$$

where t_s is the index of the first frame within the time window, and $d_{hist}(\cdot,\cdot)$ is the Bhattacharyya distance between two histograms. Since each edge (the weight between two trajectories) can potentially be shared by up to two adjacent patches, the largest dissimilarity score among those two patches is used as the dissimilarity score $e^{p,q}_{app}$ for that edge.

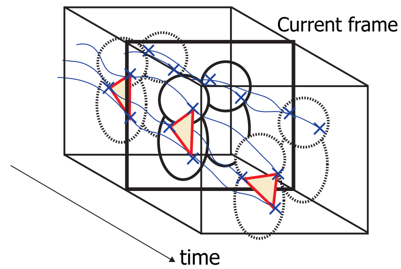

Figure 4. A sequence of triangular patches used to measure the variation of appearance over time. Each vertex of a triangular patch is created by the trajectory of three adjacent features.

4.3. Spatial proximity and motion coherency

We also make use of spatial proximity and coherent motion, borrowing from previous work with crowds [2], to evaluate the dissimilarity between pedestrian trajectories. Since, a pair of feature trajectories are likely to remain in close proximity if they belong to the same person, we use the maximum displacement between a pair of feature trajectories within a time window as the dissimilarity measure $e^{p,q}_{prox}$. We also use the standard deviation of the distance between two trajectories over the time window, to measure the dissimilarity with respect to motion coherency $e^{p,q}_{coh}$.

5. Experimental results

In order to demonstrate both the effectiveness and robustness of our method, we tested our approach using both synthetic and real video sequences of crowds. To show the improvements made by our proposed system, we make a comparison with a baseline system that uses a similar setup as [2] (i.e., motion coherency and spatial proximity). Our experiments were run on a Windows PC with an Intel Core 2 Quad 2.66GHz CPU and 3 GB of RAM.

5.1. Baseline comparison

5.1.1 Effect of using gait features

To examine the effectiveness of the gait features, we tested our tracking algorithm on a scene in which two synthetic targets move in close proximity in the same direction. Each target is made to bob up and down at different frequencies. Since we are analyzing the advantage of using the gait features, we do not use the temporal variance of local appearance for these experiments.

Figure 5 shows the improvement in tracking caused by the use of the gait features. In Figure 5-(a) we can see that

(a) (b)

Figure 5. Effect of using gait features. (a) Two targets are properly tracked by using the gait features. (b) Neighboring targets are mistakenly tracked as one target without the use of gait features.

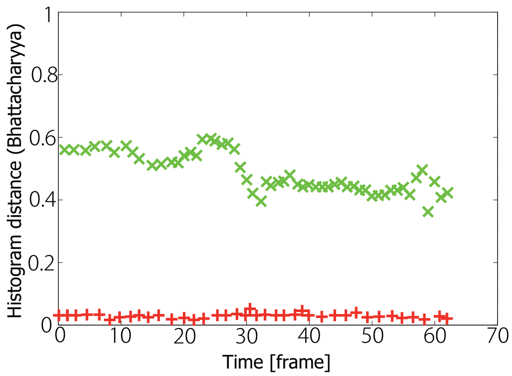

Figure 6. Histogram distance of two local patches over time. The red crosses represent a patch located on one target in contrast to the green x's which represent a patch which straddles two targets. There is more variation in appearance when the patch straddles two targets.

the use of the gait features allows our method to successfully track the two targets. In contrast, it can be observed in Figure 5-(b) that without the use of gait features, the baseline algorithm tracks the two targets as one large moving target.

5.1.2 Effect of using temporal consistency of local patches

Next we simulate the extreme case where two targets move in the same direction with exactly the same gait frequency to show the effectiveness of using the temporal consistency of local appearance patches. In this experiment, the baseline system used gait features in addition with motion coherence and spatial proximity to observe the contribution of using temporal consistency of local appearance patches to the tracking performance.

Figure 6 shows the histogram distances for two patches, where one patch lies on a single target (series of red crosses) and the other patch straddles the two targets (series of green x's). We observe from the results that the temporal consistency of appearance is a highly discriminative indicator of whether or not the patch (and therefore the features) lie on the same target.

5.2. Real crowd sequences

Now we test the performance of our method using 4 different video sequences of real crowds of people. 3 of the sequences we tested, which we call sequence (A), (B) and (C), are taken at different locations and the difficulty of the tracking task increases from sequence (A) to (C). Sequence (A) is a relatively sparse crowd with about 14 pedestrians. Sequence (B) contains a denser crowd of about 22 people and includes significant partial occlusion. The most challenging sequence (C) contains about 30 people crossing an intersection from different directions with heavy occlusion. In addition to the 3 sequences, we also ran tests using the UCSD dataset [13]. The specifications of each video sequence and the parameters used for each experiment are given in Table 1 and Table 2, respectively. We set the thresholds for pruning edges th_p as half of the median value computed from the set of the edge weights $e^{p,q}$. We ensure that a typical gait cycle (about 1 sec.) is contained in the time window (τ) by setting the time widow to cover at least one second, as in [11].

We evaluated the tracking results using the recall and precision rates, which are given in Table 3 and in Figure 7. The number of true positives is the number of people who are correctly detected (one cluster detected for one person). When a person is not detected (no clusters detected on a person) this is counted as a false negative. We defined two kinds of false positives (failure modes) as: (I) multiple clusters are detected on a single person and (II) multiple people are clustered as one person. Once an individual has been observed for at least τ frames, we begin calculating the recall and precision rates. This is because our tracker requires several frames to initialize tracking.

Our use of appearance consistency and gait features allows our method to attain an average recall rate of 56 %. In contrast, the recall rate of the baseline system is about 33%. We reason that the baseline tracker using only motion coherency and spatial proximity could not handle these sequences because the individuals are very close to each other and move in the same direction. Figures 8, 9, 10 and 11 compare several key frames from our approach against the baseline system. It is observed that our system is able to track more individuals with added stability over time.

Table 4 shows the false alarm rate for each sequence. We can see that false negative rates are higher than false positive rates in these examples. This is primarily due to the fact that the non-textured clothing does not generate reliable features enough for tracking and also makes it difficult to compute reliable optical flow (Figure 12-(a)). As a result, the system is prevented from tracking such individuals. We would like to suggest that the use of color-based local feature detectors can increase the number of reliable features generated for tracking.

We can also observe that the type (I) false positive ac-

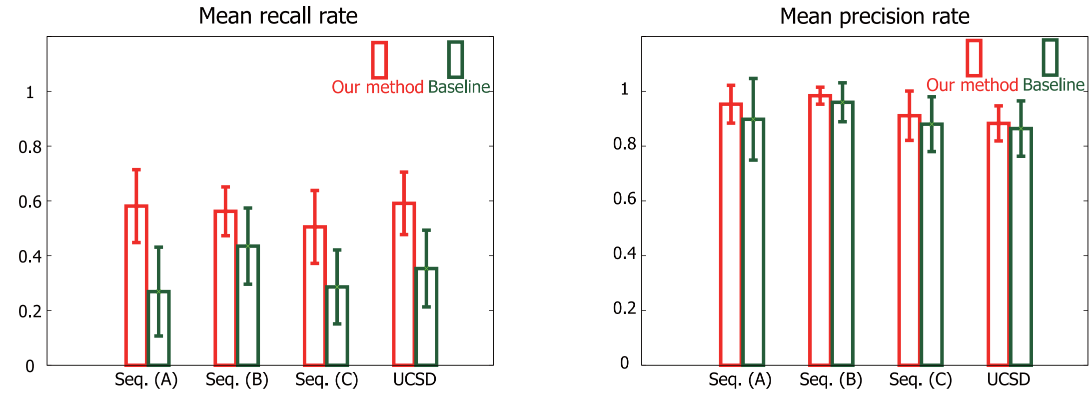

Figure 7. Tracking performance via recall and precision rate: the left bar graph shows the average of recall rate for each sequence, and the right bar graph shows the average of precision rate for each sequence. It can be observed that tracking performance is improved on each sequence.

Table 3. Tracking results: N denotes the average of the number of people in the FOV. μ and σ represents the mean and the standard deviation, respectively. t_c denotes the average of computation time for clustering processing.

	N	Our system					Baseline				
		Recall rate		Precision rate			Recall rate		Precision rate		
		μ	σ	μ	σ	t_c [sec]	μ	σ	μ	σ	t_c [sec]
Seq. (A)	14.0	0.581	0.133	0.953	0.069	30.45	0.269	0.162	0.898	0.149	4.16
Seq. (B)	22.0	0.562	0.089	0.984	0.031	28.75	0.435	0.139	0.960	0.071	4.33
Seq. (C)	29.4	0.505	0.133	0.911	0.090	10.09	0.286	0.135	0.880	0.100	2.23
UCSD	21.3	0.591	0.114	0.883	0.064	6.75	0.353	0.140	0.864	0.101	3.33

Table 1. Characteristics of the videos used in the experiments.

	resolution	frame rate	frames
Seq. (A)	1280x720	60	700
Seq. (B)	1280x720	60	700
Seq. (C)	800x600	30	520
UCSD	720x480	30	1540

Table 2. Parameters used for our experiments with real video.

	τ	Our system		Baseline	
		$e_{\min}$	th_p	$e_{\min}$	th_p
Seq. (A)	64	0.1	0.27	0.1	64.8
Seq. (B)	64	0.1	1.38	0.1	560.0
Seq. (C)	40	0.1	0.68	0.1	66.6
UCSD	40	0.1	1.11	0.1	207.7

counts for the largest portion of the false positive counts. It arises from the fact that the motion coherency of the trajectories that belong to the same person may represent different properties. In fact, there are several motions generated by only one individual, such as swinging hands, turning of the head and etc. This intra-person variation of the motion coherency makes it difficult to track a single person as one (rigid) object.

In addition, we observe that one cluster appears on the individual's bag and another cluster appears on the body (Figure 12-(b)). Since the frequency characteristics are different between the trajectories of the body and those of the bag, this individual is recognized by the system as two different individuals.

Table 4. False alarm rate. FN shows the mean of the number of false negatives and FP shows the mean of the number of false positives ((I): multiple clusters are detected on a single person and (II): multiple people are clustered as one person).

	Our system			Baseline		
	FN	FP		FN	FP	
		(I)	(II)		(I)	(II)
Seq. (A)	5.70	0.32	0.09	9.80	0.50	0.05
Seq. (B)	9.50	0.22	0.00	12.95	0.10	0.30
Seq. (C)	12.90	1.05	0.22	19.25	1.03	0.08
UCSD	7.33	1.35	0.33	12.94	1.04	0.19

We would like to address the issues of intra-pedestrian variation and carried objects in our future work.

6. Conclusions

We have proposed a method for tracking individuals in crowds by extracting a person's unique spatio-temporal features. In particular, a person's individuality was represented using a person's gait features and the temporal consistency of local patch appearance. Through our experiments, it was shown that the use of the gait features was effective in dis-

Figure 8. Tracking results on sequence (A) (left: our system, right: baseline).

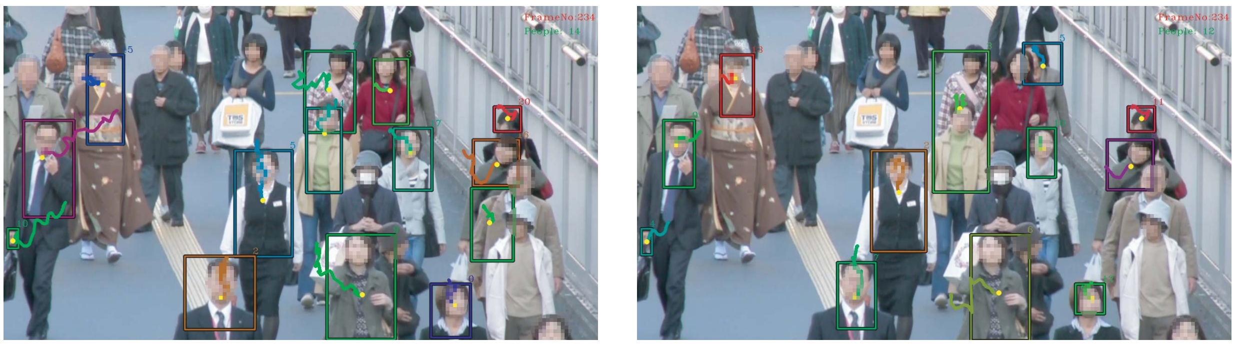

Figure 9. Tracking results on sequence (B) (left: our system, right: baseline).

(a) (b)

Figure 12. Examples of failure cases. (a) Tracking fails due to the lack of reliable feature trajectories around non-textured regions. (b) Multiple clusters appear on one person caused by the difference of frequency characteristics between the individual's bag (or hand) and the body.

criminating between individuals in a crowd. Also by monitoring the visual variance of local patches over time, our system was able to accurately track individuals in crowds despite their similarity of motion and partial occlusion. Our experiments with both synthesis data and real video sequences of crowded scenes showed that our approach is both effective and robust for tracking individuals in crowds.

7. Acknowledgments

Daisuke Sugimura is supported by Grant-in-Aid for JSPS Fellows. This research is supported in part by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for 20300061.

References

[1] S. Ali and M. Shah. Floor fields for tracking in high density crowd scenes. In *Proc. of IEEE European Conference on Computer Vision*, volume 2, pages 1–14, 2008.

[2] G. Brostow and R. Cipolla. Unsupervised bayesian detection of independent motion in crowds. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition*, volume 1, pages 594–601, 2006.

[3] J. R. Casas, A. P. Sitjes, and P. P. Folch. Mutual feedback scheme for face detection and tracking aimed at density estimation in demonstrations. *IEEE Vision, Image and Signal Processing*, 152(3):334–346, 2005.

[4] L. P. Chew. Constrained delaunay triangulations. In *Proc. of the Third Annual Symposium on Computational Geometry*, pages 215–222, 1987.

[5] L. Dong, V. Parameswaran, V. Ramesh, and I. Zoghlami. Fast crowd segmentation using shape indexing. In *Proc. of IEEE International Conference on Computer Vision*, pages 1–8, 2007.

[6] M. Isard and J. MacCormick. Bramble: A bayesian multiple-blob tracker. In *Proc. of IEEE International Conference on Computer Vision*, pages 34–41, 2001.

[7] S. M. Khan and M. Shah. A multiview approach to tracking people in crowded scenes using a planar homography

Figure 10. Tracking results on sequence (C) (left: our method, right: baseline).

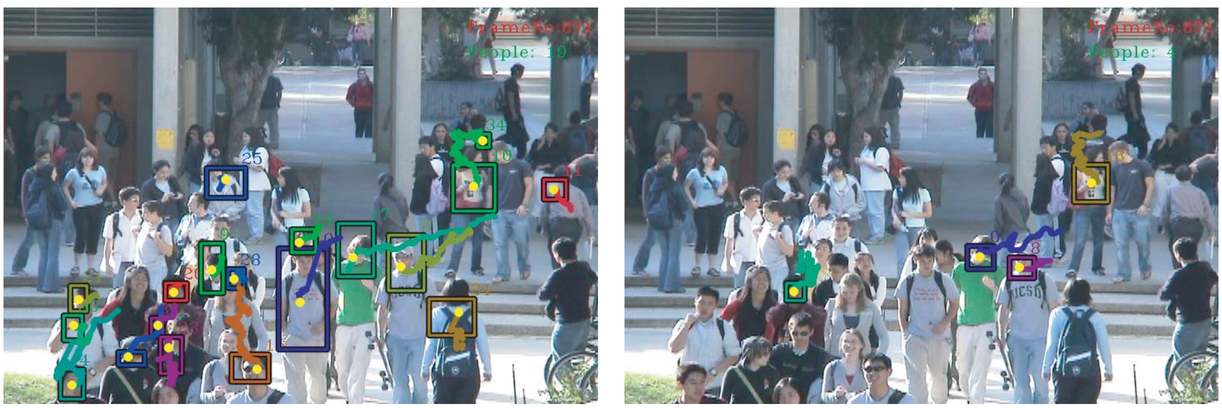

Figure 11. Tracking results on UCSD data [13] (left: our method, right: baseline).

constraint. In *Proc. of European Conference on Computer Vision*, volume 4, pages 133–146, 2006.

[8] Z. Khan, T. R. Balch, and F. Dellaert. An mcmc-based particle filter for tracking multiple interacting targets. In *Proc. of European Conference on Computer Vision*, volume 4, pages 279–290, 2004.

[9] B. Leibe, E. Seemann, and B. Schiele. Pedestrian detection in crowded scenes. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition*, volume 1, pages 878–885, 2005.

[10] Y. Li and H. Ai. Fast detection of independent motion in crowds guided by supervised learning. In *Proc. of IEEE International Conference on Image Processing*, volume 3, pages 341–344, 2007.

[11] 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 European Conference on Computer Vision*, volume 3, pages 151–163, 2006.

[12] K. Okuma, A. Taleghani, N. Freitas, J. Little, and D. Lowe. A boosted particle filter: multitarget detection and tracking. In *Proc. of European Conference on Computer Vision*, volume 1, pages 28–39, 2004.

[13] V. Rabaud and S. Belongie. Counting crowded moving objects. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition*, volume 1, pages 705–711, 2006.

[14] M. D. Rodriguez and M. Shah. Detecting and segmenting humans in crowded scenes. In *Proc. of ACM International Conference on Multimedia*, pages 353–356, 2007.

[15] O. Sidla, Y. Lypetskyy, N. Brandle, and S. Seer. Pedestrian detection and tracking for counting applications in crowdes situations. In *Proc. of IEEE International Conference on Video and Signal Based Surveillance*, 2006.

[16] C. Tomasi and T. Kanade. Detection and tracking of point features. *Technical Report CMU-CS-91-132*, 1991.

[17] C. Tomasi and J. Shi. Good features to track. In *Proc. of IEEE CS Conference on Computer Vision and Pattern Recognition*, pages 593–600, 1994.

[18] B. Wu and R. Nevatia. Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors. *International Journal on Computer Vision*, 75(2):247–266, 2007.

[19] T. Zhao and R. Nevatia. Tracking multiple humans in complex situations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1208–1221, 2004.