Person re-identification under the problem of path selection

Xiaolong MA[1,2]; Zhiguo JIANG[2]; and Jianming HU[3]

[1] China Academy of Electronics and Information Technology, 100041 Beijing, China; E-Mail: go0up@163.com
[2] School of Astronautics, Beihang University, 100191 Beijing, China; E-Mail: jiangzg@buaa.edu.cn
[3] Department of Automation, Tsinghua University, 100084 Beijing, China; E-Mail: hujm@mail.tsinghua.edu.cn

ABSTRACT

In this paper, we introduce a novel person re-identification method under the problem of path selection. Unlike other supervised person re-identification algorithms, our method can identify persons in the video without artificial markers. Our method divides each image sequences into several slices and selects the most distinguished ones, which can improve the performance. More crucially, this model is unsupervised and hence readily scalable to real-world large scale ReID settings and more suitable to previous path selection, i.e. no need of exhaustively collecting a large number of cross-view pairwise labels for each camera pair as required by most existing ReID models, which need supervised training. Experimental results show that our method can outperform other methods and achieve excellent results.

INTRODUCTION

At present, when the public transportation is equipped, the location of workplaces is mainly considered, while the location of workers' residence is hardly considered. Take the bus transportation as an example, the bus route is designed to link several important places, such as CBD, hospital, and industrial park. However, the busy degree between different places varied. The current bus route design hasn't considered the busy degree between different connections. Every existing connection on the same bus route is assigned to the same weight, even though their importance is diverse. In fact, the difference between workplaces leads to diverse busy degree of connections. The existing practices by the government are increasing the number of buses near the hot spots. Even if the problem that some residences lack buses and some have more buses than needed is settled, the problem that these bus routes are fixed while the workers' transportation varies still exists. Let's not argue the number of workers whose transportation changes over time. The location information of every worker is needed for personalized travel. The previous transportation condition, which can be used to analyze workers' travel intentions, is needed for the prediction of the next travel route. Workers' travel intentions and current traffic conditions are the key points for path selection.

Path selection based on person re-identification is an efficient strategy. The person identity information can be devoted to personalized positioning and path planning. Person re-identification is typical, concise and fast biometrics identification, which aims to match people across a pair of non-overlapping camera views at different locations, and this task is inherently challenging (Shaogang et al., 2014). Moreover, it is safe, reliable and exact to cognize the identity and easy to cooperate with computers in security and surveillance scenes, which is convenient to form large-scale automated systems. Compared with other biometrics identification technologies, person re-identification is more robust to occlusion and expedient to collect information, especially in the far range.

Most existing approaches (Chunxiao et al., 2014; Rui et al., 2014) perform ReID by extracting spatial visual appearance (texture and color) features from one or multiple person images. However, it is intrinsically limited to employ visual appearance alone due to the inherent visual ambiguity and unreliability caused by appearance similarity among different people and appearance variations of the same person from unknown and significant cross-view changes in viewpoint, illumination, background clutter and occlusion. On the other hand, there are plenty of cameras allocated in the intersection of the urban road to monitor the traffic flow, which supply abundant infrastructure, while the video data is not called into full play. It is essential to extract the key information from present video data for intelligent path selection.

In this work, we formulate an intelligent person ReID method by exploiting incomplete and noisy person sequences captured from busy public scenes. The main contributions of this work are: (1) We derive a sequential localized space-time feature representation of person image sequences. This representation facilitates the extraction and exploitation of the space-time information encoded in the original unregulated sequences, without the need of sequence fragmentation. (2) We formulate a novel sequence based person ReID model for extracting space-time identity-discriminative information from ambiguous and partial observations captured in person sequences. This is achieved by introducing a new Time Shifted Dynamic Time Warping (TS-DTW) model capable of simultaneously performing alignment, data selection and cumulating localized space-time information randomly distributed over the whole sequence in a unified manner. We show the effectiveness of the proposed model on two benchmark sequence ReID datasets (PRID2011 (Martin et al., 2011) and iLIDS-VID (Taiqing et al., 2014)) under both the closed-world and more realistic open-world settings (Liao et al., 2014).

RELATED WORK

Gait recognition (Sruti et al., 2015; Christoforos et al., 2016) has been extensively exploited for people identification using space-time features, e.g. Gait Energy Image (GEI) (Han et al., 2006). These methods assume that image sequences are aligned, as well as having complete gait cycles and accurate gait phase estimation. However, these constraints are often invalid in person ReID context as shown in Figures 2 and 3. Different from gait recognition methods, the proposed TS-DTW can relax these stringent assumptions by automatically searching for the optimal alignment configuration and selecting discriminative space-time data.

Dynamic Time Warping (DTW) (Rakthanmanon et al., 2012) is one common sequence matching algorithm widely used in speech recognition. Conventional DTW assumes that the two sequences have the same number of temporal cycles (phases) and are aligned at the starting and ending points. It is non-trivial to achieve reliable human parsing/pose detection (Kanaujia et al., 2007), especially with noisy image sequences from crowded public scenes. To address the problem, we design a new matching model TS-DTW especially for matching unregulated person sequences with automatic sequence alignment and discriminative space-time feature selection distributed over each entire sequence.

Most existing ReID methods (Bhuiyan et al., 2014; Zheng et al., 2016) only consider one-shot image per person per view. There has been some effort made on multi-shot ReID. Xu et al. (Xu et al., 2013) trained a robust appearance model by enhancing local image region/patch spatial feature representation. To that end, the most related work is the recently introduced fragment selection based DVR ranking model (Taiqing et al., 2014). Compared with DVR, the proposed TS-DTW is an unsupervised model readily scalable to large scale camera networks. Whilst DVR is supervised therefore requires labeled pairwise sequences for each camera pair, this significantly restricts DVR's usefulness in practical systems of large scale settings.

METHODOLOGY

We consider person ReID by matching a probe sequence $F \in P$ from camera view B with a gallery sequence set G from camera A in an unsupervised way. Each image sequence contains a set of consecutive frames I: $F = (I_1,\ldots,I_t)$, where t is varying as in typical surveillance videos, tracked person image sequences do not have guaranteed uniform length (arbitrary number of frames), nor number of walking cycles and starting/ending phases. To this end, let us first describe how to represent image sequences before we formulate a novel Time Shifted Dynamic Time Warping (TSDTW) model for person ReID by sequence matching without the need of pairwise labeled data for model training.

Sequential Localized Space-Time Representation

We evenly divide each individual sequence into multiple short slices with a small number L of image frames. Each slice encodes some localized space-time information about the walking characteristics of the corresponding person. As a result, an image sequence F can be converted into a space-time slice sequence $S = \{s_1,\ldots,s_m\}$. This localized slice-based sequence representation has three advantages over the bag-of-fragments model (Taiqing et al., 2014): (1) It keeps the original sequential data form, whilst the latter only considers each fragment of a sequence as an isolated instance without temporal ordering among fragments. (2) Alignment between sequences (e.g. start/end with the same walking phase) is made more robust due to the existence of a large number of short localized slices corresponding to various walking phases. (3) It provides more flexible opportunities for exploring localized space-time information irregularly distributed across the original image sequences.

Space-time feature representation: To represent the space-time information of localized slices, HOG3D histogram feature (Klaser et al., 2008) is utilized due to their superior expressive capability shown in activity recognition (Shi et al., 2013) and more recently person ReID (Taiqing et al., 2014). We set the slice length L to 4 so that reasonably reliable HOG3D features can be computed. Specifically, we spatially decompose every slice into 2×5 even cells according to human body parts, e.g. head, torso, and legs. There is 50% overlap between any two adjacent cells for increasing robustness against tracking errors. The space-time gradient histogram extracted from each cell is then concatenated to form a feature vector x for the slice s. Finally, we obtain a HOG3D feature sequence $X = \{x_1,\ldots,x_m\}$ for a slice sequence $S = \{s_1,\ldots,s_m\}$.

Time Shifted Dynamic Time Warping

We wish to extend the conventional DTW algorithm by taking into account additional time shift alignment and warping between two sequences. The proposed Time Shifted Dynamic Time Warping (TS-DTW) model is an extended Dynamic Time Warping (DTW) algorithm, particularly designed for solving the misalignment problem and selective sequence matching for person ReID. DTW aims to measure the distance/similarity between two temporal sequences/series by searching for the optimal nonlinear warp path. Specifically, given two cross-view slice sequences $S^p=\{sp\ 1, \ldots, sp\ n, \ldots\}\in P$ (probe) and $S^g=\{sg\ 1, \ldots, sg\ m, \ldots\}\in G$ (gallery), we define a warp path as：

$$W = [w_1,\ldots,w_L] W = [w_1,\ldots,w_L] \qquad (1)$$

where the l-th entry $w_l = (n_l,m_l)$ means that the n_l-th slice from S^p and m_l-th slice from S^g are matched. The warp path length holds: $\max(|S^p|,|S^g|) \leq L < |S^p|+|S^g|$. $|\cdot|$ denotes set size. We then define the sequence matching distance $D(S^p,S^g)$ between S^p and S^g as:

$$D(S^p,S^g) = \sum_{l=1}^{L} \mathrm{dist}(s^p_{n_l}, s^g_{m_l})$$

$$D(S^p,S^g) = \sum_{l=1}^{L} \mathrm{dist}(s^p_{n_l}, s^g_{m_l}) \qquad (2)$$

with dist(·,·) as the distance metric between two slices, e.g. L1. The objective of DTW is to find the optimal warp path W^* such that

$$W \ = \mathrm{argmin}_{w\in\Im} D(S^p,S^g) W \ = \mathrm{argmin}_{w\in\Im} D(S^p,S^g) \qquad (3)$$

where $\Im\Im$ is the set of all possible warp paths. This optimization can be realized using dynamic programming (Müller, 2007) subject to three constraints: (1) Bounding constraint: $w_1 = (1,1)$ and $w_L = (|S^p|,|S^g|)$; (2) Monotonicity constraint: $n_1 \leq n_2 \leq \cdots \leq n_L$ and $m_1 \leq m_2 \leq \cdots \leq m_L$; (3) Step size constraint: $w_{l+1} - w_l \in (1,0), (0,1), (1,1)$ for $l \in [1:L-1]$.

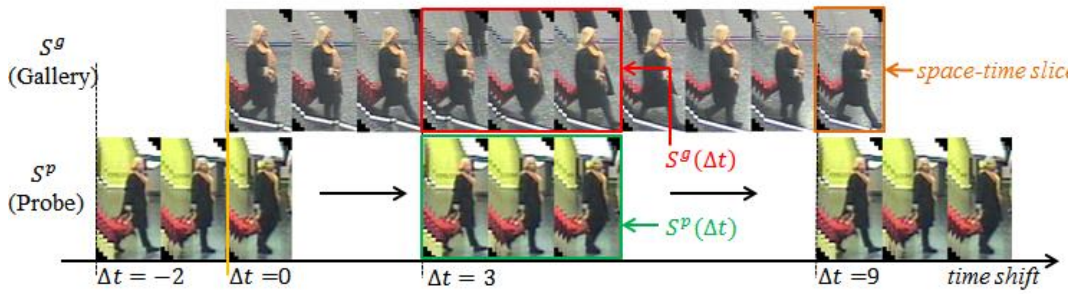

Figure 1. Illustration of the Time Shifted Dynamic Time Warping model.

As indicated in the bounding constraint, DTW assumes that starting and ending data points/slices of the two sequences are aligned (in zero-order sense). However, this is most invalid in person image sequences available for ReID as aforementioned. We relax this assumption by extending DTW with additional process of searching for the optimal alignment and data matching configuration. This shares the principle of cross correlation analysis models (Loy et al., 2010). To match cross-view person (slice) sequences beyond zero-order, we define the TS-DTW matching distance $D^{ts}(S^p, S^g, \Delta t)$ as a function of time−shift Δt:

$$D^{\mathrm{ts}}(S^p, S^g, \Delta t) = \frac{D(S^p(\Delta t), S^g(\Delta t))}{|S^p(\Delta t)|}$$
$$D^{\mathrm{ts}}(S^p, S^g, \Delta t) = \frac{D(S^p(\Delta t), S^g(\Delta t))}{|S^p(\Delta t)|}$$
(4)

where $S^p(\Delta t)$ and $S^g(\Delta t)$ are the aligned/overlapped segments sub-sampled from the original unaligned S^p and S^g under a shift amount $\Delta t \in T = [1-|S^p|, |S^g|-1]$ (Figure 1). The denominator $|S^p(\Delta t)|$ is a normalization factor for making the sub-sampled segment pairs with different amount of slices (or walking cycles) comparable. The optimal TS-DTW sequence matching distance can be obtained as

$$D^{\mathrm{ts}}(S^p, S^g, \Delta t) = \min_{\Delta t \in T}(D^{\mathrm{ts}}(S^p, S^g, \Delta t))$$
$$D^{\mathrm{ts}}(S^p, S^g, \Delta t) = \min_{\Delta t \in T}(D^{\mathrm{ts}}(S^p, S^g, \Delta t))$$
(5)

By this newly introduced time shifting operation over T, all possible alignment configurations between the two sequences are automatically attempted. Importantly, data selection over both sequences is also performed simultaneously, and the two processes are complementary to each other. This complementary interaction between the two processes allows us to more accurately match incomplete and noisy person sequences for person ReID in an unsupervised manner, as shown in our experiments.

EXPERIMENTS

Figure 2. Sequence examples from PRID2011 (left) and iLIDS-VID (right).

Datasets: Two benchmark image sequence based ReID datasets were used, PRID2011 (Martin et al., 2011) and iLIDS-VID (Taiqing et al., 2014) (Figure 2), for evaluating the performance of the proposed approach.

Figure 3. Examples of GEI features (right in each pair) from iLIDS-VID (left).

Baseline models: We compared the proposed TS-DTW model with related state-of-the-art methods as follows: (1) Gait Recognition (Martín-Félez et al., 2012): The state-of-the-art gait recognition model using Gait Energy Image (GEI) feature (Han et al., 2006) and the ranking SVM (Chapelle et al., 2010) model. We can only evaluate on iLIDS-VID because the original videos of the whole image frame are not included in PRID2011 which are required by the silhouette extraction algorithm (Sobral, 2013). (2) SDALF (Farenzena et al., 2010): A classic hand-crafted visual appearance ReID feature. Both single and multiple shot variates are considered. (3) Saliency (Zhao et al., 2013): The state-of-the-art spacial appearance based ReID method. (4) SS-ColLBP (Hirzer et al., 2012)&RankSVM (Chapelle et al., 2010): A ranking SVM (Chapelle et al., 2010) based ReID model with one of the most effective features Colour&LBP (Hirzer et al., 2012). (5) MS-ColLBP&RSVM (Taiqing et al., 2014): A multishot extension of the above method as in (Hirzer et al., 2012). (6) Dynamic Time Warping (DTW) (Berndt et al., 1994): The widely used sequence matching algorithm. Its input is the whole space-time slice sequence. (7) DTW-EQ: Similar to DTW (Berndt et al., 1994) but using the sequences with the same length as model input by cutting off the tail part of longer sequence. (8) L1/L2-norm: The basic distance metrics. Their input is the same as DTW-EQ. (9) DVR (Taiqing et al., 2014): The state-of-the-art sequence based person ReID model which achieves the best performance.

Closed-World ReID Evaluation

Settings: We followed the data partition setting as (Taiqing et al., 2014). Specifically, for either PRID2011 or iLIDS-VID, we split the entire dataset into two partitions: one half for training, and the other half for testing. Note that TS-DTW does not utilise the training partition since it is unsupervised. We evaluated a total of 10 folds and reported the averaged results with the conventional Cumulated Matching Characteristics (CMC) curves as the evaluation metric.

Table 1. Comparing unsupervised methods. (SS: Single-Shot; MS: Multi-Shot)

Dataset	**PRID2011**				**iLIDS-VID**			
Rank *R*	***R*=1**	***R*=5**	***R*=10**	***R*=20**	***R*=1**	***R*=5**	***R*=10**	***R*=20**
SS-SDALF	4.9	21.5	30.9	45.2	5.1	14.9	20.7	31.3
MS-SDALF	5.2	20.7	32.0	47.9	6.3	18.8	27.1	37.3
L1-norm	16.3	34.4	48.2	66.3	15.6	38.9	51.0	64.5
L2-norm	17.4	35.7	49.1	66.2	15.7	36.7	49.4	61.3
Salience	25.8	43.6	52.6	62.0	10.2	24.8	35.5	52.9
DTW	19.9	41.2	53.6	65.8	15.9	32.1	41.5	55.5
DTW-EQ	20.9	43.3	55.5	73.9	15.6	39.9	52.1	65.4
TS-DTW (ours)	**28.2**	**56.4**	**68.4**	**83.3**	**21.3**	**46.0**	**61.9**	**73.1**

Comparing unsupervised methods: Comparisons of unsupervised methods are shown in Table 1. It is evident that our TS-DTW model achieves the best people recognition accuracy on both datasets, which suggests the effectiveness of our model in matching unregulated sequences for person ReID by automatically aligning sequences and selecting and cumulating localized space-time information randomly distributed in the whole sequences. It is observed that the hand-crafted ReID appearance feature SDALF is much worse than L1/L2-norm on both datasets. This may be due to: (1) manually designing person-discriminative visual appearance features is limited in coping with noisy data; (2) adaptive (person-specific) and localized saliency selection is beneficial. The state-of-the-art appearance Saliency model is capable of more effectively utilising spatial appearance information. By applying the conventional DTW on the original person sequences, it gives similar or poorer ReID rate than L1/L2-norm. This may be because of (1) the misalignment problem, i.e. different starting and/or ending walking poses; (2) more critically, no data selection over sequences. This is partly solved by DTW-EQ, which enforces the two matched sequences to have a similar number of walking cycles although not necessarily aligned, which provides improved ReID rate. But it is still significantly inferior to TS-DTW due to the lack of effective alignment and data selection. This suggests the effectiveness of our time shifting based selection mechanism for ReID.

Table 2. Comparing supervised methods. (SS: Single-Shot; MS: Multi-Shot)

Dataset	**PRID2011**				**iLIDS-VID**			
Rank *R*	***R*=1**	***R*=5**	***R*=10**	***R*=20**	***R*=1**	***R*=5**	***R*=10**	***R*=20**
Gait Recognition	-	-	-	-	2.8	13.1	21.3	34.5
SS-ColLBP&RSVM	22.4	41.8	51.0	64.7	9.1	22.6	33.2	45.5
MS-ColLBP&RSVM	34.3	56.0	65.5	77.3	23.2	44.2	54.1	68.8
DVR	28.9	55.3	65.5	82.8	23.3	42.4	55.3	68.4
TS-DTW (ours)	**28.2**	**56.4**	**68.4**	**83.3**	**21.3**	**46.0**	**61.9**	**73.1**

Comparing supervised methods: We also compared TS-DTW against supervised ReID models which utilize additional information from cross-view pairwise labeled data for model training. The results are reported in Table 2. Among all competitors, the state-of-the-art gait recognition method yields the poorest person ReID results (on iLIDS-VID). The potential reason is that the GEI features computed from unregulated person sequences can be easily contaminated by random background clutter and frequent occlusions, e.g. due to random passers-by (Figure 3). In contrast, TS-DTW mitigates this problem by its alignment and data selection processes. It can be observed from the results of SS-ColLBP&RSVM and MS-ColLBP&RSVM that more reliable visual features are extracted when more images are made available, particularly on iLIDS-VID. It is also evident that our unsupervised TS-DTW is very competitive against the state-of-the-art supervised models DVR 2 and MSColLBP&RSVM. In particular, compared with the fragment based selection strategy of DVR, TS-DTW is shown to be similarly effective or even better, particularly on iLIDS-VID which is more challenging in terms of background clutter and occlusion. This shows that DVR's selection mechanism may suffer more from noisier data where noise distribution is complex and it is less likely to generate optimal fragments. Whilst TS-DTW is more robust in that its selection is not fragment-wise but localized-slice-grained. Note that TS-DTW do not require pairwise labeled data learning across views. This can be a significant benefit to real-world ReID applications since the assumption on the availability of pairwise labeled data makes these supervised methods non-scalable to large scale camera networks with many pairs of camera views.

Open-World ReID Evaluation

In addition to the closed-world ReID scenario, we further evaluated a more realistic setting called open-world (Liao et al., 2014). Specifically, its key difference from the closed-world setting is that a probe person $i \in P$ is not assumed to exist necessarily in the gallery G under the openworld setting. This setting is more plausible to real-world ReID settings since we generally have no prior knowledge about whether one person (in gallery) re-appears in certain (probe) camera views in most application cases. That is, P and G may only be partially overlapped in different camera views.

Settings: A similar data partition as the closed-world case, with the only difference that the gallery set of the testing partition is reduced by one third random selected people (these considered as imposters), i.e. 60 gallery people on PRID2011 and 100 on iLIDS-VID. Similarly, we reported the averaged results over 10 folds of experiment.

Evaluation metric: Two separate steps are involved in performance evaluation under the open-world setting (Liao et al., 2014): (1) Detection - decide if a probe person $p \in P$ exists in the gallery or not; (2) Identification - compute the truly matched rates over only accepted target people.

Table 3. Comparing open-world ReID at Rank-1.

Dataset	**PRID2011**				**iLIDS-VID**			
FAR *F*	**1**	**10**	**50**	**100**	**1**	**10**	**50**	**100**
L1-norm	3.5	7.7	13.2	19.0	2.0	4.6	12.0	18.4
MS-ColLBP&RSVM	4.3	6.7	24.3	**39.8**	1.1	4.8	15.6	25.9
MS-SDALF	0.5	1.0	4.5	6.3	0.2	0.5	3.3	8.4
Salience	**5.2**	9.7	20.8	28.3	1.4	4.2	8.3	12.4
DTW-EQ	3.7	9.2	18.3	24.3	2.0	4.5	12.1	18.6
DVR	2.7	6.5	22.5	32.5	2.9	6.9	**20.5**	**28.8**
TS-DTW	4.3	**10.3**	**25.0**	34.2	**3.7**	**8.8**	17.0	26.0

The open-world ReID results are compared among the most competitive ReID methods in Table 3. The top ranks are typically considered as very important in ReID, particularly for Rank-1. In general, the proposed TS-DTW model gains the best or the second best ReID performance over all FARs in comparison to either supervised or unsupervised models. This observation is in line with the comparison in the closed-world setting. This further shows the advantage of our model over all competitors, using either sequence matching or spatial appearance matching. It is also evident that the localized saliency based appearance matching is very effective to open-world ReID, particularly at low FARs.

CONCLUSION

In this work, we presented an intelligent person ReID framework by formulating a novel Time Shifted Dynamic Time Warping model for simultaneously sequence alignment, data selection and localized space-time feature matching between pairs of inherently incomplete and noisy image sequences from two different camera views. By using a sequential localized space-time representation of person sequences, the proposed method facilitates more effective extraction of space-time information available in person image sequence data. More importantly, our model is unsupervised and does not require exhaustive cross-view pairwise data annotation for every camera pair in model building. Extensive comparative evaluations under both the closed-world and open-world ReID settings demonstrate clearly the advantages of the proposed model over a wide range of existing and state-of-the-art gait recognition, sequence matching, supervised and unsupervised ReID methods.

ACKNOWLEDGEMENT

This work is supported by National Key R&D Program in China (2016YFB0100906), National Science and Technology Support plan Project under Grant No. 2014BAG03B01 and National Natural Science Foundation of China under Grant No. 61673232, 61650301 and Beijing Manicipal Science and Technology Program(D15110900280000, D171100006417003).

REFERENCES

Berndt, D. J., and Clifford, J. (1994). "Using Dynamic Time Warping to Find Patterns in Time Series." *KDD workshop*, 10(16), 359-370.

Bhuiyan, A., Perina, A., and Murino, V. (2014). "Person re-identification by discriminatively selecting parts and features." *European Conference on Computer Vision*, 147–161.

Chapelle, O., and Keerthi, S. S. (2010). "Efficient algorithms for ranking with SVMs." *Information Retrieval*, 13(3), 201–215.

Charalambous, C. C., and Bharath, A. A. (2016). "A data augmentation methodology for training machine/deep learning gait recognition algorithms." *Proceedings of the British Machine Vision Conference*, York, U.K.

Choudhury, S. D., and Tjahjadi, T. (2015). "Robust view-invariant multiscale gait recognition." *Pattern Recognition*, 48(3), 798–811, 2015.

Farenzena, M., Bazzani, L., Perina, A., Murino, V., and Cristani, M. (2010). "Person re-identification by symmetry-driven accumulation of local features." *Computer Vision and Pattern Recognition*, 2360–2367.

Gong, S., Cristani, M., Yan, S., and Loy, C. C. (2014). *Person re-identification*. Springer.

Han, J., and Bhanu, B. (2006). "Individual recognition using gait energy image." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28, 316–322.

Hirzer, M., Beleznai, C., Roth, P., and Bischof, H. (2011). "Person re-identification by descriptive and discriminative classification." *Image Analysis*, 91–102.

Hirzer, M., Roth, P. M., Köstinger, M., and Bischof, H. (2012). "Relaxed pairwise learned metric for person re-identification." *European Conference on Computer Vision*, 780–793.

Kanaujia, A., Sminchisescu, C., and Metaxas, D. (2007). "Semi-supervised hierarchical models for 3d human pose reconstruction". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1–8.

Klaser, A., Marszałek, M., and Schmid, C. (2008). "A spatio-temporal descriptor based on 3d-gradients." *British Machine Vision Conference*, 275, 1–10.

Liao, S., Mo, Z., Zhu, J., Hu, Y., and Li, S. Z. (2014). “Open-set person re-identification.” *arXiv preprint arXiv*, 1408. 0872.

Liu, C., Gong, S., and Loy, C. C. (2014). “On-the-fly feature importance mining for person re-identification.” *Pattern Recognition*, 47, 1602–1615.

Loy, C. C., Xiang, T., and Gong, S. (2010). “Time-delayed correlation analysis for multi-camera activity understanding.” *International Journal of Computer Vision*, 90(1), 106–129.

Martín-Félez, R., and Xiang, T. (2012). “Gait recognition by ranking.” *European Conference on Computer Vision*, Springer Berlin Heidelberg, 328–341.

Müller, M. (2007). “Dynamic time warping.” *Information retrieval for music and motion*, 69–84.

Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., and Keogh, E. (2012). “Searching and mining trillions of time series subsequences under dynamic time warping.” *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 262–270.

Shi, F., Laganiere, R., Petriu, E., and Zhen, H. (2013). “Lpm for fast action recognition with large number of classes.” *IEEE International Conference on Computer Vision Workshop on Action Recognition with a Large Number of Classes*.

Sobral, A. (2013). “BGSLibrary: An opencv c++ background subtraction library.” *IX WVC−Workshop de Visao Computacional*.

Wang, T., Gong, S., Zhu, X., and Wang, S. (2014). “Person reidentification by video ranking.” *European Conference on Computer Vision*, 688–703.

Xu, Y., Lin, L., Zheng, W. S., and Liu, X. (2013). “Human re-identification by matching compositional template with cluster sampling.” *Proceedings of the IEEE International Conference on Computer Vision*, 3152–3159.

Zhao, R., Ouyang, W., and Wang, X. (2013). “Unsupervised salience learning for person re-identification.” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3586-3593.

Zhao, R., Ouyang, W., and Wang, X. (2014). “Learning mid-level filters for person re-identification.” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 144–151.

Zheng, W. S., Gong, S., and Xiang, T. (2016). “Towards open-world person re-identification by one-shot group-based verification.” *IEEE transactions on pattern analysis and machine intelligence*, 38(3), 591–606.