PERSON RE-IDENTIFICATION BY MANIFOLD RANKING

Chen Change Loy[1], Chunxiao Liu[2], Shaogang Gong[3]*

[1] Dept. of Information Engineering, The Chinese University of Hong Kong
[2] Dept. of Electronic Engineering, Tsinghua University, China
[3] School of EECS, Queen Mary University of London, UK

ABSTRACT

Existing person re-identification methods conventionally rely on labelled pairwise data to learn a task-specific distance metric for ranking. The value of unlabelled gallery instances is generally overlooked. In this study, we show that it is possible to propagate the query information along the unlabelled data manifold in an unsupervised way to obtain robust ranking results. In addition, we demonstrate that the performance of existing supervised metric learning methods can be significantly boosted once integrated into the proposed manifold ranking-based framework. Extensive evaluation is conducted on three benchmark datasets.

Index Terms— person re-identification, manifold, ranking, distance metric learning, video surveillance

1. INTRODUCTION

For person re-identification (Re-ID), one is given a probe query image to match against a set of gallery candidates captured in distributed locations at different times. The aim is to associate the same individual and discard irrelevant ones. This task is critical in surveillance applications but known to be an intrinsically hard problem due to various sources of variations, such as changes in illumination, background clutter, and occlusion [1].

State-of-the-art methods [2, 3, 4, 5] treat re-identification as a learning-to-rank problem [6]. Typically, supervised metric learning [7] is adopted to learn a task-specific distance function, so as to discard non-informative features selectively. Learning such a function often requires a large quantity of paired training data where the correspondences of pairs of images are known and labelled. The lack of paired instances, due to either non-existence or limited annotations, limits severely the scalability of the model in real-world scenario.

The rich information and structure inherent in the *unlabelled gallery data* are generally disregarded in existing person re-identification studies. The intuition of exploiting manifold of unlabelled data has been extensively explored in many
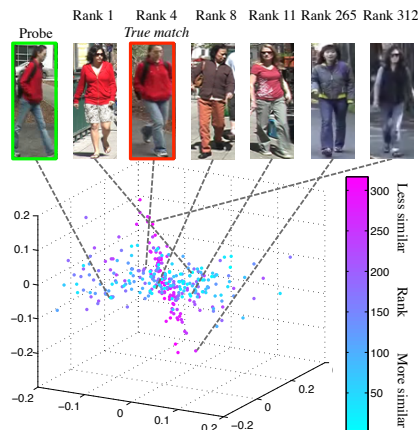
Fig. 1. A two-dimensional manifold embedding of gallery images with two clear clusters – images visually similar to the probe mostly reside on the blue circular region, whilst dissimilar images can be found in the pink ellipsoid region. One can imagine that visual ranking would be more effective in this manifold space.

other fields, such as document summarisation [8], image retrieval [9, 10, 11], and automatic photo annotation [12], but not specifically in the task of re-identification. Instances in gallery though generally represent different individuals, they do share common visual characteristics, of which the smooth intrinsic geometry can be exploited to facilitate the ranking task (see Fig. 1).

To this end, we propose a novel manifold ranking (MRank) or unsupervised metric learning based framework for re-identification. Unlike existing methods [2, 3, 4, 5] that consider only exhaustively labelled pairwise data, MRank exploits the manifold structure collectively revealed by a large quantity of gallery instances to facilitate the diffusion of query information. Intuitively, a higher rank will be assigned to gallery instances situated near to the probe sample in the manifold space, whilst locally nearby instances are encouraged to have similar ranks.

The contributions of this study: (i) we investigate the importance of using unlabelled gallery data for rank diffusion/sharing, which has been largely neglected by previous re-identification studies; (ii) we systematically formulate and

validate existing MRank models for the re-identification task. No previous re-identification studies have reported such an analysis; (iii) we show that existing supervised metric learning methods [2, 3] can benefit from the unsupervised MRank framework to gain improved performance.

2. RE-IDENTIFICATION BY MANIFOLD RANKING

We study two manfold ranking (MRank) models, namely normalised graph Laplacian (MRank-L_n) [13] and unnormalised iterated graph Laplacian (MRank-L_u) [14] for the re-identification task. The MRank-L_n was first presented in [13] and have been widely explored in various retrieval tasks [9, 10, 11]. Nonetheless, this approach is sensitive to the setting of parameter. We discuss this issue in detail from the perspective of re-identification and subsequently present the more robust MRank-L_u [14].

We first provide the problem setup and mathematical notations. We consider single-shot person re-identification where only one image is available for each person in a testing gallery. For each image, we extract a d-dimensional feature vector, denoted by $\mathbf{x} = (x_1, \ldots, x_d)^\mathsf{T} \in \mathbb{R}^d$. Given a probe instance $\mathbf{x}^p$, we assign it with a positive label $+1$, and 'spread' this query information to n unlabelled gallery instances $\{\mathbf{x}_i^g\}_{i=1}^n$, each of which has an initial label 0. The vector of initial label assignment is formally defined by $\mathbf{y} = (y_1, \ldots, y_{n+1})^\mathsf{T}$, with $y_i = 1$ if $\mathbf{x}_i = \mathbf{x}^p$, and $y_i = 0$ otherwise. We define a combined set $\mathcal{V} = \mathbf{x}^p \bigcup \{\mathbf{x}_i^g\}_{i=1}^n$, and denote a ranking function as $f : \mathcal{V} \to \mathbb{R}$. With the function we aim to estimate a ranking score vector $\mathbf{c} = (c_1, \ldots, c_{n+1})^\mathsf{T}$, so that each instance $\mathbf{x}_i^g$ has a score c_i.

Neighbourhood graph construction: The first and common step in MRank is to approximate the unknown manifold by a k-nearest neighbour (kNN) graph $G = \langle \mathcal{V}, E \rangle$ defined on $\mathcal{V}$. The edges E are weighted by a pairwise affinity matrix $A \in \mathbb{R}^{(n+1) \times (n+1)}$ defined by

$$A_{ij} = \exp\left(-\text{dist}^2(\mathbf{x}_i, \mathbf{x}_j)/\sigma^2\right), \tag{1}$$

for $i \neq j$ and $A_{ii} = 0$. Note that the diagonal elements of A are set to zero to avoid self-reinforcement during the label propagation. The σ is a scaling parameter.

In an unsupervised scenario, the distance metric, dist : $\mathcal{V} \times \mathcal{V} \to \mathbb{R}$ is usually defined as the Euclidean distance. Nevertheless, if the correspondences of pairs of images are known and labelled, one can learn a task-specific distance metric [2, 3, 4]. Using the supervised metric effectively 'distort' the implicit embedding space approximated in MRank. We will show in Sec. 3 that using supervised metric along with MRank yields improved performance than just employing supervised metric alone.

Constructing a graph G from scratch each time a new probe instance is observed can be computationally infeasible, particularly so in computing the affinity matrix, which

involves a complexity of $\text{O}(n^2)$. To avoid the computational overhead, one can pre-compute the affinity matrix of gallery instances off-line $A^g \in \mathbb{R}^{n \times n}$. When a new probe is observed, we then compute $d_{p-g} = \text{dist}\left(\mathbf{x}^p, \mathbf{x}_j^g\right)$ and increase the dimension of affinity matrix from A^g to $A = \begin{bmatrix} 0 & d_{p-g} \\ d_{p-g}^\mathsf{T} & A^g \end{bmatrix}$. The computational complexity is now linear to the gallery size[1].

Manifold ranking: Given A, we can then estimate the normalised graph Laplacian, L_n and the unnormalised graph Laplacian L_u as follows

$$L_n = I - D^{-1/2} A D^{-1/2} = I - S, \tag{2}$$
$$L_u = D - A, \tag{3}$$

where D represents a diagonal matrix with $D_{ii} = \sum_j A_{ij}$. Consequently the closed-form solutions of the two MRank models are given as

$$\text{MRank}-L_n : \quad \mathbf{c} = (\beta I + L_n)^{-1} \mathbf{y}, \tag{4}$$
$$\text{MRank}-L_u : \quad \mathbf{c} = \left[(\beta I + L_u)^{-1}\right]^m \mathbf{y}, \tag{5}$$

where $\beta \geq 0$ is a parameter common to both models, whilst $m \geq 0$ is specific to MRank-L_u.

First we give details on MRank-L_n. Let $\alpha = 1/(1+\beta) \in [0, 1)$. The fundamental idea is to iterate the estimation of

$$f(t+1) = \alpha S f(t) + (1-\alpha)\mathbf{y}. \tag{6}$$

In each iteration t, each instance receives propagated labels from its neighbours (first term), and retains its initial label assignment (second term). The balance of these two terms is controlled by α (thus the β). The iteration process is repeated until convergence (see [15] for proof) to obtain the limit of the sequence $\{f(t)\}$ as the ranking score vector $\mathbf{c}$.

A closed-form solution can be obtained by finding the limit of $f(t)$ when it converges to $\mathbf{c}$. Hence, by replacing $f(t+1)$ and $f(t)$ in Eqn. (6) with $\mathbf{c}$, we have

$$\mathbf{c} = \alpha S \mathbf{c} + (1-\alpha)\mathbf{y} \tag{7}$$
$$(I - \alpha S)\mathbf{c} = (1-\alpha)\mathbf{y} \tag{8}$$
$$\mathbf{c} = (1-\alpha)(I - \alpha S)^{-1}\mathbf{y} \tag{9}$$

where I is an identity matrix. For ranking, we can omit the global factor $(1 - \alpha)$ without changing the results. By replacing $\alpha = 1/(1 + \beta)$ and $L_n = I - S$ back into Eqn. (9), we obtain the closed-form solution in Eqn. (4), $\mathbf{c} = \frac{1}{1+\beta} (I + \beta I - S)^{-1} \mathbf{y} \equiv (\beta I + L_n)^{-1} \mathbf{y}$.

The MRank-L_n is sensitive to the value of β due to the use of L_n. Specifically, as pointed out in [14], the ranking function exhibits the following expression if β is too small

$$\mathbf{c} \equiv \nu_1(\mathbf{x})/\beta + \epsilon, \tag{10}$$

---

[1] One could expand A continuously through including all observed probes to possibly obtain a better manifold. To avoid tractability issue due to indefinite size of A, approximation such as [11] is needed. In our study, we always start from A^g each time a probe is observed.

where ϵ is a small discriminative term critical for ranking. $\nu_1(\mathbf{x})$ is the first eigenvector of L_n, which is non-constant and determined by the density $p(\mathbf{x})$. In the task of person re-identification, $p(\mathbf{x})$ is not uniform given the vast variety of individual appearances distributed with uneven frequency. As a result, the $\nu_1(\mathbf{x})$ is a function unrelated to the relevance ordering of instances, and the resulting $\mathbf{c}$ will also be uninformative. We will show in Sec. 4 that small β value leads to poor performance in MRank-L_n.

In comparison to MRank-L_n, the second model MRank-L_u is more robust to β. The reason is that L_u has a constant $\nu_1(\mathbf{x}) = \left(\frac{1}{\sqrt{1+n}}, \ldots, \frac{1}{\sqrt{1+n}}\right)^{\mathsf{T}}$, thus the $\nu_1(\mathbf{x})/\beta$ will not affect the final ranking score. In MRank-L_u, the term $(\beta I + L_u)^{-1}$ is raised to the power of m to address the diverging problem [14], i.e. where all instances have 0 rank in the limit of infinite unlabelled instances.

In Sec. 3, we will extensively compare both MRank-L_n and MRank-L_u for the re-identification task. A summary of MRank is given in Alg. 1.

Algorithm 1: Manifold Ranking (MRank) for Re-ID

Input: A probe instance $\mathbf{x}^p$, gallery set $\{\mathbf{x}_i^g\}_{i=1}^n$;
Output: Ranking score $\mathbf{c}$;
1 **begin** Initialisation
2   Pre-compute affinity matrix $A^g \in \mathbb{R}^{n \times n}$ for gallery set;
3   Define initial label assignment as $\mathbf{y} = (y_1, \ldots, y_{n+1})^{\mathsf{T}}$, with $y_i = 1$ if $\mathbf{x}_i = \mathbf{x}^p$, and $y_i = 0$ otherwise;
4 **begin** Manifold Ranking
5   Receive probe instance $\mathbf{x}^p$;
6   Compute dist $\left(\mathbf{x}^p, \mathbf{x}_j^g\right)$;
7   Expand affinity matrix from A^g to A (Eqn. (1));
8   Compute the graph Laplacian L_n or L_u (Eqn. (2) or Eqn. (3));
9   Compute ranking score $\mathbf{c}$ (Eqn. (4) or Eqn. (5));

---

3. RESULTS

Datasets: Three datasets including i-LIDS [16], VIPeR [17], and QMUL underGround Re-IDentification (GRID) [18, 19] were used for evaluation (see Fig. 2). Among the three datasets, i-LIDS and VIPeR are widely used for benchmarking. The challenging GRID dataset[2] is captured in a busy underground station, with severe inter-object occlusion and large viewpoint variations.

Features: Similar to [2, 3, 5], we partitioned an image equally into six horizontal stripes, and extracted a mixture of colour (RGB, HSV and YCbCr) and texture histograms (8 Gabor filters and 13 Schmid filters), forming a 2784-dimensional feature vector for each image.

Implementation details: There are a total of three free parameters in MRank-L_n, i.e. the k in the neighbourhood graph construction, the σ in the Gaussian kernel (Eqn. (1)), and the

Fig. 2. Example frames of datasets. Note the large appearance variations between the probes and true matches.

β in Eqn. (4). MRank-L_u has an additional parameter m. In all the experiments we fixed $k = 15$ and quantified σ automatically using the self-tuning method[3] [20]. The value of β was obtained through cross-validation on the training set. We set $m = 2$ throughout our experiments.

Evaluation settings: We followed the same evaluation procedures outlined in [3, 5]. Specifically, the matching performance was measured using the averaged cumulative match characteristic (CMC) curve over 10 trials. The number of training/testing person paired images in i-LIDS, VIPeR, and GRID were set to 69/50, 316/316, and 125/(125 paired images + 775 non-paired images), respectively.

Comparative evaluations: We compared the MRank-L_n and MRank-L_u with ℓ_1-norm [21], ℓ_2-norm [22], RankSVM [2], and PRDC [3]. Recall that MRank can be initialised with supervised distance metrics – we denote these variants as MRank-L_n (dist) or MRank-L_u (dist). The CMC performance comparisons between the MRank and conventional methods are presented in Table 1.

In general, the MRank without supervision outperforms both ℓ_1-norm and ℓ_2-norm based methods. It is worth pointing out that the performance of MRank (ℓ_2-norm) is generally better than MRank (ℓ_1-norm). This can be due to the nature of ℓ_2-norm in 'punishing' much more heavily on larger distances than ℓ_1-norm does, which allows the former to better capture the local neighbourhood structure in gallery.

Importantly, we observe higher recognition rates yielded by MRank as compared to those obtained using task-specific distance metrics such as RankSVM and PRDC alone, with an average percentage change improvement of 14.1% at rank-1 recognition rate on all three datasets. Figure 3 depicts some examples for further insights. One can observe that both the MRank-L_n and MRank-L_u initialised with PRDC gain noticeably better retrieval results as compared to the original PRDC without manifold ranking. In the second example, PRDC fails to retrieve the true match, whist MRank not only associates the right individual, but also shows superior capability in retrieving more relevant candidates, as evidenced by the comparatively higher visual consistency across the top-ranked instances. The better performance of MRank is ex-

Method	i-LIDS ($p = 50$)					VIPeR ($p = 316$)					GRID ($p = 900$)				
	r = 1	r = 5	r = 10	r = 15	r = 20	r = 1	r = 5	r = 10	r = 15	r = 20	r = 1	r = 5	r = 10	r = 15	r = 20
ℓ_1-norm [21]	29.60	**54.80**	67.60	74.60	81.00	**9.43**	**20.03**	**27.06**	**30.95**	**34.68**	4.40	11.68	16.24	19.12	24.80
MRank-L_n (ℓ_1-norm)	**31.40**	54.40	**68.40**	75.60	**83.60**	8.48	18.70	24.40	28.83	32.66	**7.12**	12.32	17.68	20.64	**25.36**
MRank-L_u (ℓ_1-norm)	30.60	53.40	68.20	**76.00**	82.80	8.35	17.06	22.47	26.33	30.76	6.00	**13.28**	**17.92**	**21.12**	24.00
ℓ_2-norm [22]	28.20	54.00	66.20	72.40	79.40	10.95	23.92	31.39	38.86	**44.11**	4.88	14.24	20.32	22.40	26.24
MRank-L_n (ℓ_2-norm)	**31.40**	55.60	**67.60**	**77.40**	82.20	**11.42**	**24.27**	**33.73**	**38.92**	**44.11**	5.76	14.96	**21.76**	**25.12**	**30.96**
MRank-L_u (ℓ_2-norm)	31.00	**56.00**	67.40	77.00	81.20	10.57	24.24	33.42	38.83	43.42	5.76	**15.44**	21.28	24.96	28.40
RankSVM [2]	42.60	67.60	78.80	86.00	92.00	14.87	37.12	50.19	58.48	65.66	10.24	24.56	33.28	39.44	43.68
MRank-L_n (RankSVM)	**42.80**	**70.40**	**81.80**	86.40	**92.40**	19.27	42.41	55.00	63.86	70.06	**12.24**	**27.84**	36.32	**42.24**	**46.56**
MRank-L_u (RankSVM)	41.80	69.60	81.40	**87.00**	91.40	**19.34**	**42.47**	**55.51**	**64.11**	**70.44**	11.44	27.60	**36.40**	**42.24**	46.24
PRDC [3]	44.80	68.00	77.60	84.20	88.20	16.01	37.09	51.27	59.43	65.95	9.68	22.00	32.96	38.96	44.32
MRank-L_n (PRDC)	47.80	**71.60**	**80.60**	85.00	**90.60**	**19.37**	**42.78**	**54.78**	**63.77**	**69.62**	10.88	24.96	**35.84**	41.44	46.40
MRank-L_u (PRDC)	49.00	70.60	**80.60**	**85.60**	**90.60**	18.45	41.74	53.67	62.72	69.27	**11.12**	**26.08**	35.76	**41.76**	**46.56**

Table 1. Performance comparisons between MRank variants and conventional methods without manifold ranking. We use MRank (dist) when we initialise MRank with a specific distance metric. The value p is the number of person in a test set. The MRank achieves an overall percentage change improvement of 14.1% at rank-1 recognition rate over the-state-of-the-art RankSVM [2] and PRDC [3] methods.
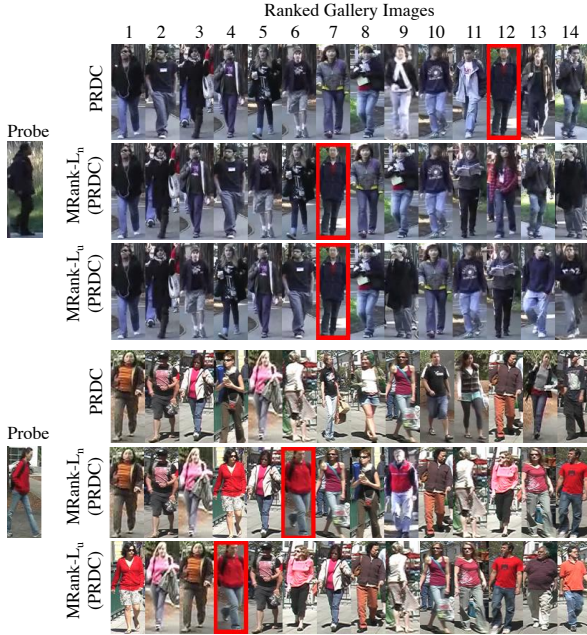

Fig. 3. Example results by manifold ranking (MRank), compared against PRDC [3] without manifold ranking. The leftmost image is the probe. The true match within the ordered gallery candidates is highlighted with red border.

pected since query propagation takes place in a structured manifold, on which the potential and less-plausible candidates have already been separated into distinct clusters (see Fig. 1).

Sensitivity of β: Here we show the results of β sensitivity test on the normalised graph Laplacian L_n and unnormalised graph Laplacian L_u. Owing to space limit, we only show the results on VIPeR with PRDC to initialise MRank. It is evident from Fig. 4 that L_u is less sensitive to β in comparison to L_n. The results support our algorithmic analysis in Sec. 2. In MRank-L_n, one can practically circumvent the selection of β by setting its value to zero, and perform a pseudo-inverse to
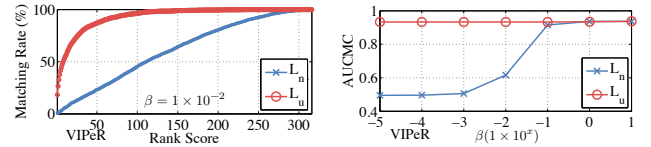

Fig. 4. β sensitivity test on the unnormalised Laplacian (L_u) and normalised Laplacian (L_n). PRDC was used to initialise MRank. (Left) CMC curves with $\beta = 10^{-2}$. (Right) Area under the CMC (AUCMC) with β varied from 10^{-5} to 10.

Method	VIPeR ($p = 316$)				
	r = 1	r = 5	r = 10	r = 15	r = 20
PRDC	16.01	37.09	51.27	59.43	65.95
MRank-L_n (PRDC), β=0	18.20	41.04	52.85	61.23	67.66

Table 2. At a small expense in performance degradation, the L_n-based manifold ranking can be made more stable against β (by setting its value to zero) and using pseudo-inverse to compute the ranking function.

remove the first troublesome eigenvector[4] [14]. The outcome of this strategy can be seen in Table 2. Specifically, by setting $\beta = 0$, MRank-L_n avoids the selection of β, whilst still gaining superior performance to the original PRDC.

In general, MRank-L_u is advantageous to MRank-L_n for re-identification task from the β stability standpoint.

4. CONCLUSION

We have systematically formulated and validated two different graph Laplacian-based methods for the re-identification task. Extensive experiments on three benchmark datasets have demonstrated that the learning of gallery manifold is critical and beneficial to the ranking task, and can be best exploited to improve the ranking performance of existing supervised-based techniques.

---

[4]E.g. using pinv in Matlab with small tolerance for eigenvalue cutoff.

5. REFERENCES

[1] S. Gong, C. C. Loy, and T. Xiang, "Security and surveillance," *Visual Analysis of Humans*, pp. 455–472, 2011.

[2] B. Prosser, W. Zheng, S. Gong, and T. Xiang, "Person re-identification by support vector ranking," in *BMVC*, 2010, pp. 21.1–21.11.

[3] W. Zheng, S. Gong, and T. Xiang, "Re-identification by relative distance comparison," *TPAMI*, 2012.

[4] M. Hirzer, P. Roth, M. Köstinger, and H. Bischof, "Relaxed pairwise learned metric for person re-identification," in *ECCV*, 2012, pp. 780–793.

[5] C. Liu, S. Gong, C. C. Loy, and X. Lin, "Person re-identification: What features are important ?," in *ECCV Workshop on Person Re-identification*, 2012.

[6] T.Y. Liu, *Learning to rank for information retrieval*, vol. 13, Springer, 2011.

[7] L. Yang and R. Jin, "Distance metric learning: A comprehensive survey," Tech. Rep., Michigan State Universiy, 2006.

[8] X. Wan, J. Yang, and J. Xiao, "Manifold-ranking based topic-focused multi-document summarization," in *IJCAI*, 2007, pp. 2903–2908.

[9] J. He, M. Li, H.J. Zhang, H. Tong, and C. Zhang, "Generalized manifold-ranking-based image retrieval," *TIP*, vol. 15, no. 10, pp. 3170–3177, 2006.

[10] Y. Huang, Q. Liu, S. Zhang, and D.N. Metaxas, "Image retrieval via probabilistic hypergraph ranking," in *CVPR*, 2010, pp. 3376–3383.

[11] B. Xu, J. Bu, C. Chen, D. Cai, X. He, W. Liu, and J. Luo, "Efficient manifold ranking for image retrieval," in *ACM SIGIR*, 2011, pp. 525–534.

[12] J. Jia, N. Yu, and X.S. Hua, "Annotating personal albums via web mining," in *ACM MM*, 2008, pp. 459–468.

[13] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf, "Ranking on data manifolds," in *NIPS*, 2004, vol. 16, pp. 169–176.

[14] X. Zhou, M. Belkin, and N. Srebro, "An iterated graph laplacian approach for ranking on manifolds," in *SIGKDD*, 2011, pp. 877–885.

[15] D. Zhou, O. Bousquet, T.N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," *NIPS*, vol. 16, pp. 321–328, 2004.

[16] W. Zheng, S. Gong, and T. Xiang, "Associating groups of people," in *BMVC*, 2009.

[17] D. Gray and H. Tao, "Viewpoint invariant pedestrain recognition with an ensemble of localized features," in *ECCV*, 2008, pp. 262–275.

[18] C. C. Loy, T. Xiang, and S. Gong, "Time-delayed correlation analysis for multi-camera activity understanding," *IJCV*, vol. 90, no. 1, pp. 106–129, 2010.

[19] C. C. Loy, T. Xiang, and S. Gong, "Multi-camera activity correlation analysis," in *CVPR*, 2009, pp. 1988–1995.

[20] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," in *NIPS*, 2004, pp. 1601–1608.

[21] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu, "Shape and appearance context modeling," in *ICCV*, 2007.

[22] W. Hu, M. Hu, X. Zhou, T. Tan, J. Lou, and S. Maybank, "Principal axis-based correspondence between multiple cameras for people tracking," *TPAMI*, vol. 28, no. 4, pp. 663–671, 2006.