Towards Person Identification and Re-identification with Attributes

Ryan Layne, Timothy M. Hospedales, and Shaogang Gong

School of EECS, Queen Mary University of London, UK

Abstract. Visual identification of an individual in a crowded environment observed by a distributed camera network is critical to a variety of tasks including commercial space management, border control, and crime prevention. Automatic re-identification of a human from public space CCTV video is challenging due to spatiotemporal visual feature variations and strong visual similarity in people's appearance, compounded by low-resolution and poor quality video data. Relying on re-identification using a probe image is limiting, as a linguistic description of an individual's profile may often be the only available cues. In this work, we show how mid-level semantic attributes can be used synergistically with low-level features for both identification and re-identification. Specifically, we learn an attribute-centric representation to describe people, and a metric for comparing attribute profiles to disambiguate individuals. This differs from existing approaches to re-identification which rely purely on bottom-up statistics of low-level features: it allows improved robustness to view and lighting; and can be used for identification as well as re-identification. Experiments demonstrate the flexibility and effectiveness of our approach compared to existing feature representations when applied to benchmark datasets.

1 Introduction

Person re-identification, or *inter-camera entity association*, is the task of recognising an individual in diverse scenes obtained from non-overlapping cameras. In particular, for long-term people monitoring over space and time, when an individual disappears from one view they need be differentiated from numerous possible targets and re-identified in another view, potentially under a different viewing angle and lighting condition and subject to variable degrees of occlusion.

Relying on manual re-identification in large camera networks is prohibitively costly and inaccurate. Operators are often assigned more cameras to monitor than is optimal and manual matching can also be prone to attentive gaps [1]. Moreover, human performance is subjectively determined by individual operator's experience therefore is often difficult to transfer and also subject to operator bias [2]. For these reasons, there has been extensive work in the computer vision community on automated re-identification. These efforts have primarily focused on developing feature representations which are discriminative yet invariant to view angle and lighting [3], and improved learning methods to better discriminate identity [4]. Nevertheless, despite extensive research, automated re-identification

A. Fusiello et al. (Eds.): ECCV 2012 Ws/Demos, Part I, LNCS 7583, pp. 402–412, 2012.

is still a largely unsolved problem. This is due to the underlying challenge that most features are still either insufficiently discriminative for cross-view entity association, especially with low resolution images, or insufficiently robust to view angle and lighting changes.

Contemporary approaches to re-identification typically exploit low-level features [5, 6, 3], because they can be relatively easily measured. In this paper, we take inspiration from the operating procedures of human experts [7–9], and recent research in attribute learning [10] to introduce a new class of mid-level *attribute* features. When performing person re-identification, human experts seek and rely upon matching appearance or functional attributes that are unambiguous in interpretation, such as hair-style, shoe-type or clothing-style [7]. This attribute-centric representation is also used when a description is provided verbally (e.g., by an eye-witness) to an operator. We term this process attribute-profile identification, or zero-shot re-identification. Many of these mid-level attributes can be measured reasonably reliably with modern computer-vision techniques. This provides both a mechanism for attribute-profile identification as well as a valuable new class of features for re-identification. Crucially, attributes and low-level features provide very different types of information – effectively separate modalities. We will show how, with appropriate data fusion, attributes and low-level features can provide powerful re-identification as well as attribute-profile identification capabilities.

1.1 Related Work and Contributions

Re-identification. Contemporary approaches to re-identification typically exploit low-level features such as colour, texture, spatial structure [3], or combinations thereof [6, 11]. Once a suitable representation has been obtained, nearest-neighbour [3] or learning-based matching algorithms such as ranking [6] may be used for re-identification. In each case, a distance metric (e.g., Euclidean or Bhattacharyya) must be chosen to measure the similarity between two samples. It is also be possible to discriminatively optimise the distance metric [4]. Other complementary aspects of the problem have also been pursued to improve performance, such as improving robustness by combining multiple frames worth of features along a tracklet [11] and learning the topology or activity correllations of the camera network [12] to cut down the matching space.

Attributes. Attribute based modelling has recently been exploited to good effect in object [10] and action [13] recognition. To put this in context, in contrast to low-level features, or high-level classes / identities, attributes are the mid-level *description* of a class or instance. There are various unsupervised (e.g., PCA or topic-models) or supervised (e.g., neural network) modelling approaches which produce data-driven mid-level representations. These techniques aim to project the data onto a basis set defined by the assumptions of the particular model (e.g., maximisation of variance, likelihood, or sparsity). In contrast, attribute learning focuses on representing data instances by projecting them onto a basis set defined by domain-specific axes which are semantically meaningful to humans.

Semantic attribute representations have various benefits: (i) If data is sparse (as in re-identification, which can be seen as one-shot learning) they can be more powerful than low level features [10, 14, 13] because they provide a form of transfer learning since attributes can be learned from a larger dataset apriori; (ii) they can be used in conjunction with raw data for greater effectiveness [13] and (iii) they are a suitable representation for direct human interaction, therefore allowing searches to be specified or constrained by attributes [10, 14, 15].

Attributes in Identification and Surveillance. One view of attributes is as a type of transferrable context [16] in that they provide auxiliary information about an instance to aid in (re)-identification. Here they are related to the study of soft-biometrics, which aims to enhance biometric identification performance with ancillary information [17, 18]. Alternatively they can be used for semantic attribute-profile identification (zero-shot learning [10]) in which early research has aimed to retrieve people matching a verbal attribute description from a camera network [8]. However, this has so far only been illustrated on relatively simple data with a small set of equally-reliable facial attributes. We will illustrate

that one of the central issues for exploiting attributes for general automated (re)-identification is dealing with their unequal and variable informativeness and reliability of measurement from raw data.

Contributions. In this paper, we move towards leveraging semantic mid-level attributes for automated person identification and re-identification. Specifically, we make four main contributions: (i) We introduce and evaluate an ontology of useful attributes which can be relatively easily measured using computer vision methods from the set of attributes used by human experts . (ii) We show how to learn an attribute-space distance metric to leverage attributes for re-identification. (iii) We evaluate the resulting approach and improve state of the art re-identification performance on standard benchmark datasets. (iv) We show how attributes and raw-data can also be used together for zero-shot re-identification.

2 Quantifying Attributes for Re-identification

In this section, we first describe our space of defined attributes (Section 2.1), then how to train detectors for each attribute (Section 2.2). Finally, we show how to learn a distance-metric for attribute space, and fuse these attributes with raw low-level features for re-identification (Section 2.3).

2.1 Attributes

Based on the operational procedures of human experts [7], we define the following space of $N_a = 15$ binary attributes for our study: *shorts, skirt, sandals, backpack, jeans, logo, v-neck, open-outerwear, stripes, sunglasses, headphones, long-hair, short-hair, gender, carrying-object.* Twelve of these are related to attire, and three are soft biometrics. Figure 1 shows an example of each attribute[1].

[1] We provide our annotations here: `http://www.eecs.qmul.ac.uk/~rlayne/`

Fig. 1. Example positive images for each attribute in our ontology. From left to right: *shorts, sandals, backpack, open-outerwear, sunglasses, skirt, carrying-object, v-neck, stripes, gender, headphones, short-hair, long-hair, logo, jeans.*

2.2 Attribute Detection

Low-level Feature Extraction. To detect attributes, we first extract an 2784-dimensional low-level colour and texture feature vector denoted $\mathbf{x}$ from each person image I following the method in [6]. This consists of 464-dimensional feature vectors extracted from six equal sized horizontal strips from the image. Each strip uses 8 colour channels (RGB, HSV and YCbCr) and 21 texture filters (Gabor, Schmid) derived from the luminance channel. We use the same parameter choices for γ, λ, θ and σ^2 as [6] for Gabor filter extraction, and for τ and σ for Schmid extraction. Finally, we use a bin size of 16 to describe each channel.

We train Support Vector Machines (SVM) to detect attributes.

We use Maji et al.'s implementation [19] of LIBSVM and investigate Linear, RBF, χ^2 and Intersection kernels. We select the Intersection kernel as it compares closely with χ^2 but can be trained much faster For each attribute, we perform cross validation to select SVM slack parameter C from $C \in [-10, 5]$. SVM scores are probability mapped, so each attribute detector i outputs a posterior $p(a_i|\mathbf{x})$.

Attribute Training and Representation. The prevalence of each attribute (e.g., jeans, sunglasses) varies dramatically so some attributes have a limited number of positive examples. To avoid bias due to imbalanced data, we train each attribute detector with all the positive examples, and obtain a matching number of negative examples by regularly subsampling the rest of the data.

Given the learned bank of attribute detectors, any person image can now be represented in a semantic attribute space by stacking the posteriors from each attribute detector into a N_a dimensional vector: $A(\mathbf{x}) = [p(a_1|\mathbf{x}), \dots, p(a_{N_a}|\mathbf{x})]^T$.

2.3 Re-identification

Model and Fusion. In order to use our attributes for re-identification, we choose a base re-identification method, and investigate how attributes can be fused to enhance performance. In particular we choose to build on *Symmetry-Driven Accumulation of Local Features* (SDALF), introduced by Farenzena et al. [3]. SDALF provides a low-level feature and Nearest Neighbour (NN) matching strategy giving state-of-the-art performance for a non-learning NN approach, and can be fused with additional sources of information.

Farenzena et al. introduces a state of the art distance metric d_{SDALF} to compare person images I_p and I_q. Within this nearest neighbour strategy, we can integrate our attribute-based distance d_{ATTR} as follows:

$$d(I_p, I_q) = (1 - \beta_{ATTR}) \cdot d_{SDALF}(SDALF(I_p), SDALF(I_q)) \quad (1)$$

$$+ \beta_{ATTR} \cdot d_{ATTR}(ATTR(I_p), ATTR(I_q)). \quad (2)$$

Here Eq. (1) corresponds the SDALF distance and Eq. (2) fuses our attribute-based distance metric. For our attribute representation, we will learn a Mahalanobis $L2$ distance metric d_{ATTR}, detailed next.

Attribute Metric Learning. Since attributes are unequal due to variability in number of training samples, how reliably they are measured, and how informative they are, we need to decide how to weight the attributes. To address this, we exploit the information theoretic distance metric learning strategy from [20]. We define the distance (Eq. (2)) between attribute profiles $A(\mathbf{x})$ as the following Mahalanobis distance, paramaterized by positive definite matrix Λ:

$$d_{ATTR}(I_p, I_q; \Lambda) = (A(\mathbf{x}_p) - A(\mathbf{x}_q))^T \Lambda (A(\mathbf{x}_p) - A(\mathbf{x}_q)). \quad (3)$$

A distance metric paramaterized by Λ can be represented by the corresponding multi-variate Gaussian $p(\mathbf{x}; \Lambda, \mu) \propto \exp(-d_\Lambda(\mathbf{x}, \mu)/2)$. The Kullback-Leibler divergence $\mathcal{KLD}(\Lambda||\Lambda)$ between two such Gaussians thus provides a well-founded measure of the similarity between two Mahalanobis distance metrics. Building on this measure of similarity between distance metrics, choosing a distance metric to optimise the separability of person images via attributes can be expressed via the following large-margin constraint satisfaction problem [20]:

$$\min_{\Lambda} \mathcal{KLD}\left(p(\mathbf{x}; \Lambda_0)||p(\mathbf{x}; \Lambda)\right) \; s.t. \quad (4)$$

$$d_A(\mathbf{x}_i, \mathbf{x}_j) \le u \qquad if \qquad (i,j) \in S,$$

$$d_A(\mathbf{x}_i, \mathbf{x}_j) \ge l \qquad if \qquad (i,j) \in D,$$

where Λ_0 is a regulariser representing a simple identity-matrix metric, $(i, j) \in S$ indicates instances i and j are images of the same person, and $(i, j) \in D$ indicates images of different people. The matrix Λ obtained by optimising Eq. (4) provides the optimal distance metric via Eq. (3) and hence Eq. (1).

2.4 Attribute-Profile Identification / Zero-Shot Re-identification

In addition to re-identification based on a probe image, we can also directly identify a person given solely their semantic attribute description (aka zero-shot learning [10] or attribute search [8]). Given an attribute description in the form of a binary vector $\mathbf{a}$, we can attempt to find this person by NN matching $\mathbf{a}$ against the attribute profiles $A(\mathbf{x}_i)$ of each person i in the dataset.

Surprisingly, in a multi-camera context, we can also use raw-data to improve attribute-profile identification [21]. The intuition is if searching for a given profile

$\mathbf{a}$ in view A, we also can use the match from another view B (or multiple matches from A if available) to obtain an estimated appearance/low level feature $\hat{\mathbf{x}}$ as additional context. This then provides an additional source of information from view B which can be used together with $\mathbf{a}$ in the full framework (Eq. (1)). Of course the matching within view B is imperfect, so we take $\hat{\mathbf{x}}$ as $\hat{\mathbf{x}} = \frac{1}{K}\sum_l \mathbf{x}^B_{\mathbf{a},l}$, averaging over the top K matches $\mathbf{x}^B_{\mathbf{a},l}$ to prototype $\mathbf{a}$ in view B.

3 Experiments and Discussion

Datasets We select three challenging datasets with which to validate our model, VIPeR [5], i-LIDS pedestrians [22] and ETHZ [23]. **VIPeR** contains 632 pedestrian image pairs from two cameras with different viewpoint, pose and lighting. Images are scaled to 128x48 pixels. We follow [5, 3] in considering Cam B as the gallery set and Cam A as the probe set. Performance is evaluated by matching each test image in Cam A against the Cam B gallery. **i-LIDS** [22] contains 479 images of 119 pedestrians captured from non-overlapping cameras observing a busy airport hall. In addition to pose and illumination variations, images are also subject to occlusion. Images are scaled to 128x64 pixels. We follow [3] in randomly selecting one image for each pedestrian to build a gallery, while the others form the probe set, averaging results over 10 trials. **ETHZ** was developed using a mobile camera and contains high variations in person appearance; but low pose variation. As in [3], images are normalised to 64x32 pixels and we test on Seq. 1 only which consists of 83 persons with 4,857 detections. Since the number of people here is too small to split separate metric training and testing sets we report figures for vanilla attributes instead.

Conditions. For each dataset, we select a portion for training, while re-identification performance is reported on the held out test portion. There are two phases to training: attribute detector learning (Section 2.2) and attribute distance metric learning (Section 2.3). Because VIPeR contains the largest amount of data, and the most diversity in attributes, we train the attribute detectors for all experiments on this dataset. This is important because it highlights the value of attributes as a source of transferrable information [16].

For the metric learning, we learned on the training portion of each dataset. We quantify re-identification performance in the standard way [5, 3]: recognition rate is visualised with Cumulative Matching Characteristic (CMC) curves, which indicate the probability of the correct match appearing in the top n.

We compare the following re-identification methods: **SDALF** [3] using code provided by the authors (note that SDALF is already shown to decisively outperform [24]); **Attr** vanilla attribute based re-identification (euclidean distance); **AccMI** attribute based re-identification with a weighting given by product of accuracy and mutual information with identity [25]; **MLA** attribute based-reidentification with discriminatively learned distance metric; **SDALF+MLA** SDALF fused with MLA (weight β_{ATTR} determined by optimisation on training set).

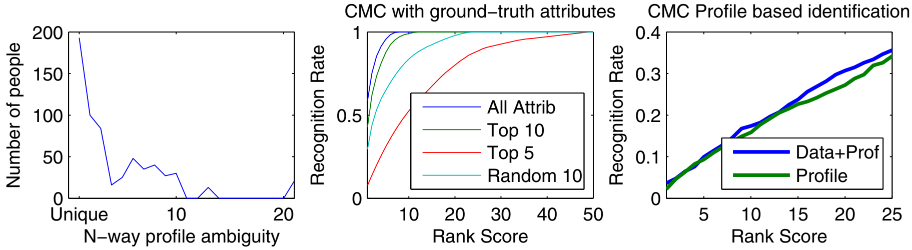


Fig. 2. (a) Most people are uniquely identifiable by attributes. (b). VIPeR reindentification with perfect attribute classifiers (p=632). (c) Attribute search / zero-shot re-identification (p=168).

Table 1. Attribute Detection Performance

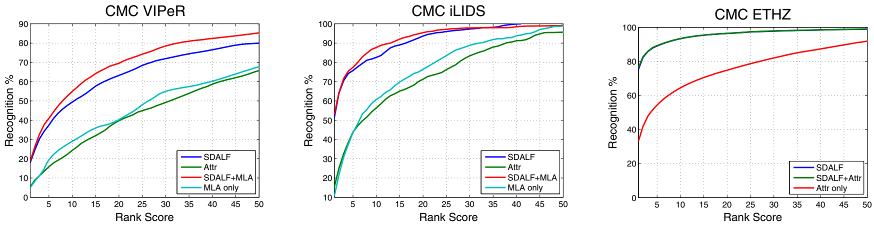
Attribute	Abs	Mean	Attribute	Abs	Mean	Attribute	Abs	Mean
shorts	0.79	0.74	sandals	0.64	0.58	backpacks	0.66	0.52
jeans	0.76	0.73	carrying	0.75	0.50	logo	0.59	0.58
vnecks	0.44	0.53	openouter	0.64	0.56	stripes	0.41	0.47
sunglasses	0.66	0.60	headphones	0.74	0.58	shorthair	0.52	0.52
longhair	0.65	0.55	male	0.68	0.68	skirt	0.67	0.76

3.1 Attribute Analysis

We first analyse the potential of our attribute ontology with regards to the VIPeR dataset. Fig. 2(a) shows a histogram of the number of individuals against degree of attribute profile uniqueness or ambiguity. Clearly the majority of people can be uniquely or almost uniquely identified by their profile, while there are a small number of people with a very generic profile. The CMC curve (for gallery size p=632) that would be obtained assuming perfect attribute classifiers is shown in Fig. 2(b). This impressive result highlights the potential for attribute-based re-identification. Also shown are the results with top 5 or 10 attributes (sorted by mutual information with identity), and a random 10 attributes. This shows that: (i) as few as 10 attributes are sufficient if they are good (high MI) and perfectly detectable, while 5 is too few; and (ii) attributes with high MI are significantly more useful than low MI (always present or absent) attributes.

Attribute Detection. Attribute detection in VIPeR achieves an average accuracy of 64%, with 11 detectors performing greater than 60% (Table 1). This highlights the issue of inequality of attributes and the importance learning a good distance metric to focus on the most reliable and discriminative attributes.

Zero-shot identification. We next evaluate the novel task of identification based solely on the manual attribute profile of a target (instead of the standard re-identification approach of providing a probe image) using VIPeR. This corresponds to the task of identifying an individual in a surveilled space based on an, e.g. radioed, textual description of their attributes. This is challenging both



(a) (b) (c)

Fig. 3. Averaged CMC curves for (a) VIPeR (gallery size $p = 250$), (b) iLIDS (gallery size $p = 60$) and (c) ETHZ1 (gallery size $p = 80$)

.

(a) VIPeR (b) i-LIDS

Fig. 4. Examples where MLA (green) increases the re-id rank of the correct match vs SDALF (red)

because of profile ambiguity (Fig. 2(a)) and limited accuracy of the attribute detectors (Table 1). Performing zero-shot identification enhanced with data from other camera (Section 2.4), we raise the CMC nAUC from 67% to 68% and double the Rank 1 match rate from 2% to 4% (Fig. 2(c)).

3.2 Re-identification

Quantitative Evaluation The re-identification performance of all models is summarised in Figure 3 and Table 2. In each case, optimisation with of the distance metric improves re-identification over vanilla attributes (MLA vs Attr). Optimised attributes in conjunction SDALF outperforms vanilla SDALF (SDALF+MLA vs SDALF). Importantly, at the most valuable low rank $r = 1$

Table 2. Breakdown of re-identification rates at specified ranks, and area under CMC

		Attr	MLA	SDALF	SDALF+MLA			Attr	MLA	SDALF	SDALF+Attr
ILIDS	R1	15.42	10.75	51.33	52.92	ETHZ1	R1	32.92	35.27	75.86	75.86
	R5	43.83	43.58	75.75	77.42		R5	53.73	55.90	88.82	88.82
	R10	56.42	60.75	82.58	87.75		R10	63.43	65.64	92.77	92.77
	R25	76.25	83.50	96.00	97.17		R25	78.27	79.70	96.76	96.78
	nAUC	77.42	80.58	92.96	93.73		nAUC	83.64	84.48	96.84	96.93
		Attr	MLA	SDALF	SDALF+MLA	AccMI					
VIPER	R1	5.40	5.06	18.02	18.78	6.40					
	R5	15.80	19.24	37.38	40.94	15.86					
	R10	24.34	29.06	49.38	54.94	26.02					
	R25	44.94	48.06	68.44	74.28	46.86					
	nAUC	80.72	82.90	86.56	90.15	81.00					

(perfect match), SDALF has re-identification rates of 18.0%, 51.3% and 75.9% while our full method has rates of 18.8%, 52.9% and 76% (for VIPeR, iLIDS and ETH respectively; gallery size $p = 250$, $p = 60$, $p = 80$). We note that a simpler attribute weighting baseline based on accuracy and mutual information with identity (AccMI,[25]) does not improve much on vanilla unweighted attributes. Moreover, this method requires ground-truth for attributes (so we can only test it on VIPeR), which is not a limitation shared by our approach.

Some examples of re-identification using MLA and SDALF are shown in Figure 4 (a) and (b) for VIPeR and i-LIDS respectively. These illustrate how attributes can complement low-level features. In the first examples for VIPeR and iLIDS the detectors for *backpacks* and *carrying* respectively push the true match up the rankings compared to SDALF.

4 Conclusions

We have shown how state-of-the-art low-level feature representations for automated re-identification can be further improved by taking advantage of a mid-level attribute representation reflecting semantic cues used by human experts [7]. Existing approaches to re-identification [3, 6, 5] focus on high-dimensional low-level features which are assumed invariant to view and lighting. However, their simple nature and invariance also limits their discriminative power for identity. In contrast, attributes provide a low-dimensional mid-level representation which makes no invariance assumptions (Variability in appearance of each attribute is learned by the classifier). Importantly, although individual attributes vary in robustness and informativeness, attributes provide a strong cue for identity. Their low-dimensional nature means they are also amenable to discriminatively learning a good full-covariance distance metric in order to take into account inter-attribute correlations. In developing a separate cue-modality, our approach is potentially complementary to most existing approaches, whether focused on low-level features [3], or learning methods [4].

The proposed attribute-centric model provides an important contribution and novel research direction for practical re-identification: by providing a complementary and informative mid-level cue, as well as opening up new applications such as zero-shot identification within the same framework. As a novel application, consider how semantic attributes could potentially be used to constrain or relax re-identification, for example by specifying invariance to whether or not the target has removed or added a hat.

The most promising direction for future research is improving the attribute-detector performance, as evidenced by the excellent results in Fig. 2(b) using ground-truth attributes. The more limited empirical performance is due to lack of training data, which could be addressed by transfer learning to bring attribute detectors trained on large databases (e.g., web-crawls) to re-identification.

Acknowledgements. Ryan Layne is supported by a EPSRC CASE studentship supported by UK MOD SA/SD. The authors also wish to thank Toby Nortcliffe of the Home Office CAST for insights on human expertise.

References

1. Keval, H.: CCTV Control Room Collaboration and Communication: Does it Work? In: Proceedings of Human Centred Technology Workshop, pp. 11–12 (2006)
2. Williams, D.: Effective CCTV and the challenge of constructing legitimate suspicion using remote visual images. Journal of Investigative Psychology and Offender Profiling 4(2), 97–107 (2007)
3. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: Computer Vision and Pattern Recognition (2010)
4. Zheng, W.S., Gong, S., Xiang, T.: Person re-identification by probabilistic relative distance comparison. In: Computer Vision and Pattern Recognition (2011)
5. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking. In: Performance Evaluation of Tracking and Surveillance (2007)
6. Prosser, B., Zheng, W.S., Gong, S., Xiang, T.: Person Re-Identification by Support Vector Ranking. In: British Machine Vision Conference, pp. 21.1–21.11 (2010)
7. Nortcliffe, T.: People Analysis CCTV Investigator Handbook. Home Office Centre of Applied Science and Technology (2011)
8. Vaquero, D.A., Feris, R.S., Tran, D., Brown, L., Hampapur, A., Turk, M.: Attribute-based people search in surveillance environments. In: Workshop on the Applications of Computer Vision, pp. 1–8 (2009)
9. Cheng, D.S., Cristani, M., Stoppa, M., Bazzani, L.: Custom Pictorial Structures for Re-identification. In: British Machine Vision Conference (2011)
10. Lampert, C., Nickisch, H., Harmeling, S.: Learning to detect unseen object classes by between-class attribute transfer. In: Computer Vision and Pattern Recognition (2009)
11. Bazzani, L., Cristani, M., Perina, A., Farenzena, M., Murino, V.: Multiple-shot Person Re-identification by HPE signature. In: International Conference on Pattern Recognition, pp. 1413–1416 (2010)
12. Loy, C.C., Xiang, T., Gong, S.: Time-Delayed Correlation Analysis for Multi-Camera Activity Understanding. International Journal of Computer Vision 90(1), 106–129 (2010)
13. Liu, J., Kuipers, B.: Recognizing human actions by attributes. In: CVPR (2011)

14. Siddiquie, B., Feris, R.S., Davis, L.S.: Image ranking and retrieval based on multi-attribute queries. In: Computer Vision and Pattern Recognition (2011)
15. Kumar, N., Berg, A., Belhumeur, P.: Describable visual attributes for face verification and image search. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(10), 1962–1977 (2011)
16. Zheng, W.S., Gong, S., Xiang, T.: Quantifying and Transferring Contextual Information in Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 1(8), 1–14 (2011)
17. Jain, A.K., Dass, S.: Soft biometric traits for personal recognition systems. In: Biometric Authentication, pp. 1–7 (2004)
18. Dantcheva, A., Velardo, C., D'Angelo, A., Dugelay, J.L.: Bag of soft biometrics for person identification. Multimedia Tools and Applications 51(2), 739–777 (2010)
19. Maji, S., Berkeley, U.C., Berg, A.C.A., Malik, J.: Classification using intersection kernel support vector machines is efficient. In: Computer Vision and Pattern Recognition (2008)
20. Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-Theoretic Metric Learning. In: International Conference on Machine Learning (2007)
21. Fu, Y., Hospedales, T.M., Xiang, T., Gong, S.: Attribute Learning for Understanding Unstructured Social Activity. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part IV. LNCS, vol. 7575, pp. 521–534. Springer, Heidelberg (2012)
22. Zheng, W.S., Gong, S., Xiang, T.: Associating Groups of People. In: British Machine Vision Conference (2009)
23. Schwartz, W.R., Davis, L.S.: Learning discriminative appearance-based models using partial least squares. In: Brazilian Symposium on Computer Graphics and Image Processing (2009)
24. Gray, D., Tao, H.: Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 262–275. Springer, Heidelberg (2008)
25. Yang, Y., Pedersen, J.O.: A comparative study on feature selection in text categorization. In: International Conference on Machine Learning (1997)