Person Re-identification by Attributes

Ryan Layne
rlayne@eecs.qmul.ac.uk
Timothy Hospedales
tmh@eecs.qmul.ac.uk
Shaogang Gong
sgg@eecs.qmul.ac.uk

Queen Mary Vision Laboratory,
School of Electronic Engineering and
Computer Science,
Queen Mary, University of London,
London, E1 4NS, U.K.

Abstract

Visually identifying a target individual reliably in a crowded environment observed by a distributed camera network is critical to a variety of tasks in managing business information, border control, and crime prevention. Automatic re-identification of a human candidate from public space CCTV video is challenging due to spatiotemporal visual feature variations and strong visual similarity between different people, compounded by low-resolution and poor quality video data. In this work, we propose a novel method for re-identification that learns a selection and weighting of mid-level semantic attributes to describe people. Specifically, the model learns an attribute-centric, parts-based feature representation. This differs from and complements existing low-level features for re-identification that rely purely on bottom-up statistics for feature selection, which are limited in discriminating and identifying reliably visual appearances of target people appearing in different camera views under certain degrees of occlusion due to crowdedness. Our experiments demonstrate the effectiveness of our approach compared to existing feature representations when applied to benchmarking 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 human re-identification in large camera networks is prohibitively costly and inaccurate. Human operators are often assigned more cameras to monitor than is optimal and a manual matching process can also be prone to attentive gaps [11]. Moreover, human performance is subjectively determined by individual operator's experience therefore is often difficult to transfer and also subject to operator bias [23, 25]. 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 [4], and improved learning methods to better discriminate identity [28]. Nevertheless, despite extensive research, automated re-identification 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, such as colour [17], texture, spatial structure [4], or combinations thereof [6, 20], because they can be relatively easily and reliably measured. High-level features such as ethnicity, gender, age or indeed identity itself, would be the most useful. However, such "soft-biometrics" are exceptionally difficult to measure reliably in surveillance video, where individuals are at "stand-off" distances and in unknown poses. In this paper, we take inspiration from the operating procedures of human experts [19, 24], and recent research in attribute learning [13] to introduce a new class of mid-level *attribute* features. When performing person re-identification, human experts tend to seek and rely upon matching attributes appearance or functional attributes that are unambiguous in interpretation, such as hair-style, shoe-type or clothing-style [19]. Some of these mid-level attributes can be measured reasonably reliably with modern computer-vision techniques, and hence provide a valuable additional class of features which has thus far not been exploited for re-identification. Crucially, these attributes provide a very different source of information – effectively a separate modality – to the typical low-level features used. We will show how with appropriate data fusion, they can be used in a complementary way to existing low-level features to provide powerful re-identification capabilities.

1.1 Related Work and Contributions

Re-identification Contemporary approaches to re-identification typically exploit low-level features such as colour [17], texture, spatial structure [4], or combinations thereof [6, 20]. Once a suitable representation has been obtained, nearest-neighbour [4] or learning-based matching algorithms such as ranking [20] may be used for re-identification. In each case, a distance metric must be chosen to measure the similarity between two samples, for example Euclidean, $L1$-Norm or Bhattacharyya. Alternatively, the distance metric may also be discriminatively optimised [27]. Various 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 [1] and learning the topology of the camera network [15] or activity correllations [16] to cut down on the matching space.

Attributes Attribute based modelling has recently been exploited to good effect in object [13] and action [14] recognition as well as in the classification of images and video. 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., multi-layer neural network) modelling approaches which produce or exploit data-driven statistical 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, independence, or sparsity). In contrast, attribute learning approaches focus on representing data instances by projecting them onto a basis set defined by domain-specific axes which are semantically meaningful to humans, preferrably unambiguous and more robust to subjective interpretation.

Semantic attribute representations have various benefits: They are often more powerful than using low-level features directly for high-level tasks such as classification [13]; they provide a form of transfer/multi-task learning because an attribute can be learned from multiple classes or instances exhibiting that attribute [8]; they can be used in conjunction with raw data for greater effectiveness [8, 14]; and finally they are a suitable representation for

direct human interaction, therefore allowing searches to be specified or constrained by attributes [12, 13, 22]. This final property can facilitate man-in-the-loop active learning when available data for model training is sparse and biased.

One view of attributes is as a type of transferrable context [27] in that they provide auxiliary information about an instance to aid in identification. Attributes are also related to the study of soft-biometrics, which aims to enhance biometric identification performance with ancillary information [8, 9]. In a surveillance context, the semantic nature of the attribute representation has been used in initial research on attribute-based person search, that aims to retrieve instances of all people matching a verbal attribute description from a camera network [24]. 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 reliability of measurement from raw data.

Contributions In this paper, we make a first step towards leveraging semantically defined mid-level attributes for automated person re-identification. Specifically, we make three main contributions: (i) We introduce and evaluate an ontology of useful attributes from the subset of attributes used by human experts which can also be relatively easily measured by bottom-up low-level features computed using established computer vision methods. (ii) We show how to select those attributes that are most effective for re-identification and how to fuse the attribute-level information with standard low-level features. (iii) We show how the resulting synergistic approach – Attribute Interpreted Re-identification (AIR) – obtains state of the art re-identification performance on two standard benchmark datasets.

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 select, weight 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 [19], 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.

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 [20]. 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 [20] 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) [21] to detect attributes. We use Maji *et al.*'s implementation [18] of Chang *et al.*'s LIBSVM [2] and investigate Linear, RBF, χ^2 and

Figure 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.*

Intersection kernels. We select the Intersection kernel as it compares closely with χ^2 but can be trained much faster [1]. For each attribute, we perform cross validation to select values for SVM slack parameter C from the set $C \in \{-2,\ldots,5\}$ with increments of $\varepsilon = 1$. The SVM scores are probability mapped, so each attribute detector i outputs a posterior $p(a_i|\mathbf{x})$ for that attribute.

Spatial Feature Selection Since some attributes (e.g., shorts) are highly unlikely to appear outside of their expected spatial location, we can improve attribute detection performance by performing coarse spatial feature selection to determine the set of strips which are informative for each attribute. Specifically, we exhaustively evaluate all $2^n - 1$ combinations of the horizontal strips, selecting the combination which achieves the highest mean accuracy in cross-validation. We denote the selected features for attribute i as $\mathbf{x}_i^+$.

Imbalanced Attribute Training The prevalence of each attribute (e.g., jeans, sunglasses) in a given dataset tends to vary dramatically and some attributes have a limited number of positive examples in an absolute sense as a result. To avoid bias due to imbalanced data, we train each attribute detector with all the positive training examples of that attribute, and obtain negative examples by subsampling the rest of the data at regular intervals ε. Where the number of negative training instances T_n exceeds the number of positive instances T_p, we scale the majority class down sampling at a greater interval by setting ε as

$$\varepsilon = \begin{cases} 2 & \text{if } T_p < T_n \\ 2\|\frac{T_p}{T_n}\| & \text{otherwise} \end{cases} \tag{1}$$

Mid-level Attribute Representation 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: $A(\mathbf{x}) = [p(a_1|\mathbf{x}_i^+),\ldots,p(a_{N_a}|\mathbf{x}_{N_a}^+)]^T$.

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.* [4]. SDALF provides a low-level feature

[1] Our experiments on model performance vs. training time resulted in the intersection kernel giving 61.37% mean accuracy with 42.4 seconds training, as compared to the χ^2 kernel (60.87% with 121.1s), the radial basis function kernel (53.29% with 332.4s) and the linear kernel (59.51% with 40.2s) respectively.

and Nearest Neighbour (NN) matching strategy giving state-of-the-art performance for a non-learning NN approach, as well as a compatible fusion method capable of admitting additional sources of information.

Farenzena *et al*. introduce three sets of features from which seperate Bhattacharyya distance metrics can be constructed. These distance metrics are combined in order to obtain the distance d between two particular person images I_p and I_q. Within this nearest neighbour strategy, we can integrate our attribute-based distance d_{ATTR} as follows:

$$\begin{aligned} d(I_p,I_q) &= \beta_{WH}\cdot d_{WH}(WH(I_p),WH(I_q)) && (2)\\ &+ \beta_{MSCR}\cdot d_{MSCR}(MSCR(I_p),MSCR(I_q)) && (3)\\ &+ \beta_{RHSP}\cdot d_{RHSP}(RHSP(I_p),RHSP(I_q)) && (4)\\ &+ \beta_{ATTR}\cdot d_{ATTR}(ATTR(I_p),ATTR(I_q)). && (5) \end{aligned}$$

Here Eqs. (2-4) correspond to the three SDALF distance measures and Eq. (5) fuses our attribute-based distance distance metric. WH, $MSCR$ and $RHSP$ represent the metrics calculated for each of the separate SDALF features using Bhattacharyya.

Attribute Selection and Weighting Since all attributes are not equal due to variability in how reliably they are measured, how prevalent they are in the imbalanced data, and how informative they are about identity, we need to decide which attributes to include from the full set and how to weigh them. For our attribute representation, we will learn a weighted $L2$-norm distance metric, d_{ATTR} defined as a weighted sum of the $L2$-distances between the images for each attribute i.

$$\begin{aligned} d_{ATTR}(I_p,I_q) &= (A(\mathbf{x}_p)-A(\mathbf{x}_q))^T W (A(\mathbf{x}_p)-A(\mathbf{x}_q)), && (6)\\ &= \sqrt{\sum_i w_i \left(p(a_i|\mathbf{x}^+_{p,i}) - p(a_q|\mathbf{x}^+_{q,i})\right)^2}. && (7) \end{aligned}$$

Searching the N_a dimensional space of weights directly to determine attribute selection and weighting is computationally intractable. We therefore employ a greedy search which initialises all $w_i = 1$ and then greedily adjusts and weighs attributes to maximally improve the re-identification rates as detailed in Algorithm 1. Finally, fixing the attribute weights $\mathbf{w}^*$, we optimise the overall attribute-vs-SDALF weight β_{ATTR} (Eq. (5)) by cross-validation.

Algorithm 1 Greedy optimisation for attribute selection and weighting

1. For each available attribute i.
   (a) Evaluate re-id performance for weights $w_i \in [0.1, 3]$.
2. Select attribute i^* and weight w_i^* which maximally improve re-identification rate.
3. If no attributes improve re-identification, terminate.
4. Else fix i^* with w_i^* and repeat.

3 Experiments and Discussion

We validate our AIR method on two public datasets. We quantify re-identification performance in the standard way [4, 6]: recognition rate is visualised with Cumulative Matching

Characteristic (CMC) curves and re-identification rate via Synthetic Recognition Rate (SRR) curves. CMC curves indicate the likelihood of the correct match appearing in the top n^{th} ranked matches, whilst SRR represents the probability of the m best matches being correct.

3.1 Datasets and Conditions

We selected two challenging datasets with which to validate our model, the VIPeR dataset introduced by Gray *et al.*, [6], and pedestrian images [26] from the i-LIDS dataset[2].

VIPeR VIPeR is comprised of 632 pedestrian image pairs from two cameras with different viewpoint, pose and lighting conditions. The images are uniformly scaled to 128x48 pixel size. We follow [4, 6] 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 i-LIDS [26] contains 479 images of 119 pedestrians captured from non-overlapping cameras observing a busy airport hall. In addition to pose and illumination variations, images in this dataset are also subject to occlusion unlike VIPeR. Images are scaled to 128x64 pixel size. We follow [4] in randomly selecting one image for each pedestrian to build a gallery set, while the others form the probe set. This is repeated 10 times and the results averaged.

Training For each dataset, we select half 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 selection and weighting (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 [27]. In creating the attribute training set, we include data from both cameras, sampling the set regularly so that we do not bias the detectors towards seasonal trends present in the data. Unlike low-level features which are designed to be viewpoint invariant, the appearance of attributes can be highly dependant on the viewpoint as well as the pose of the person. We therefore annotated view in VIPeR into three broad classes: front, back, side. When selecting positive examples for the training set of each attribute detector, we ensured that instances from each angle were included. In this way the attribute detectors learn some view invariance to the view of the attributes. For the attribute-weighting, we learned on the training portion of each dataset [3].

3.2 Attribute Detection

Performance on VIPeR Attribute detection in VIPeR achieves an average accuracy of 59.33%, with 4 detectors performing greater than 60%, and the remaining detector accuracy slightly greater than 50%. Table 1 details the results. This result highlights the issue of inequality of attributes and the importance of individual attribute selection and weighting.

Interestingly, some attributes which seem visually subtle can be detected reasonably well. Gender, for example, is detected with 68% accuracy, and Headphones with 58% accuracy due to the small but stereotyped appearance of the audio cable near the neck. The under-performing detectors tended to be for attributes more sensitive to variations in pose. For example, v-necks are detectable only from the front and logos are usually smaller and less visible from side and back poses.

[2] http://www.homeoffice.gov.uk/science-research/hosdb/i-lids/

[3] We provide our annotations here: http://www.eecs.qmul.ac.uk/~rlayne/#bmvc_attrs

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

Table 1: Attribute Detection Performance.

Figure 2: Strip selection results for attribute detection in VIPeR. From left to right: *shorts, carrying, backpacks, logos, stripes*.

Part selection As discussed in Section 2.2, we perform feature selection to choose the spatial parts (strips) used for detecting each attribute. Figure 2 illustrates the largely intuitive parts associated with some attributes. This result shows that not all features support each attribute, and that automated feature selection for spatially localised attributes is possible. Importantly, by reducing the feature space for each attribute, this also helps to alleviate the challenge of leaning from sparse data for each attribute.

3.3 Re-identification

Quantitative Evaluation The re-identification performance of our complete system is summarised in Figure 3. In each case, our AIR outperforms vanilla SDALF [4], (which in turn decisively outperforms [7]). Importantly, at the most valuable low rank $r = 1$ (perfect match), SDALF has re-identification rates of 18.2% and 48.6% while AIR has re-identification rates of 16.5% and 52.1% for VIPeR and iLIDS respectively (gallery size $p = 250$ and $p = 100$). At rank 1, SDALF outperforms AIR, however from rank 5 onwards AIR systematically improves upon SDALF with rates of 38.44% and 78.33% to SDALF's 37.92% and 76.0%. This corresponds to improvements of 0.52% and 2.38%.

Some examples of re-identification using AIR 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 and third VIPeR examples, the detectors for *backpacks* and *jeans* respectively push the true match up the rankings compared to SDALF; while the same holds for the *carrying* and *backpack* detectors for the first and third i-LIDS examples.

Selection and weighting To provide insight into our contribution, Figure 3 (a) and (d) show the attribute weightings estimated by our model. AIR automatically selects all 15 attributes for the VIPeR dataset, and 14 attributes for i-LIDS. In VIPeR, carrying, backpacks, stripes and headphones are determined to be the most informative attributes by a significant margin; while in i-LIDS v-necks and headphones are also assigned significant weight; long-hair is deselected.

To illustrate the importance of attribute selection and weighting, we break down our results as shown in Table 2. Appropriate selection and weighting of attributes is crucial for

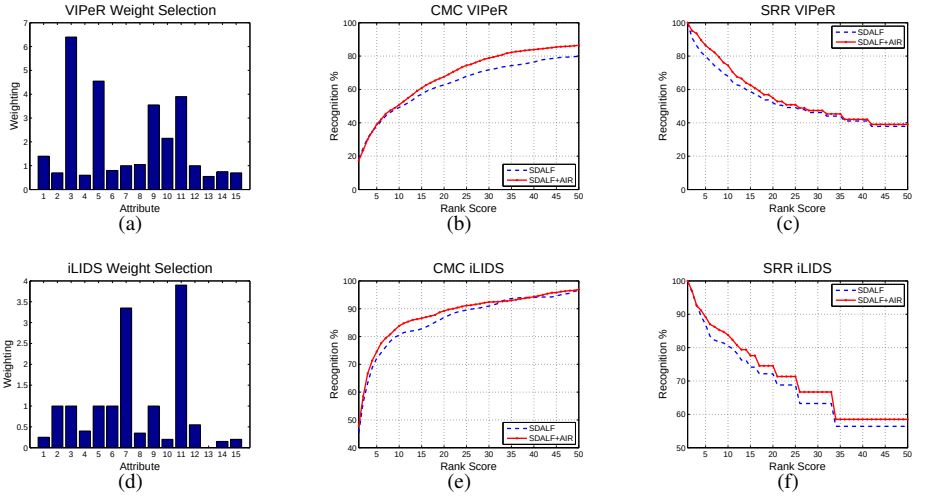

Figure 3: (a) and (d) Final weightings selected by our greedy search in VIPeR and i-LIDS respectively; (b-c), CMC and SRR curves for VIPeR (gallery size $p = 250$, $t = 10$ runs); (e-f), CMC and SRR curves for i-LIDS (gallery size $p = 100$, $t = 20$ runs). At Rank 1, AIR has re-identification rate of 16.52% and 52.13% for VIPeR and iLIDS compared to 18.20% and 48.62% for SDALF.

(a) VIPeR (b) i-LIDS

Figure 4: Examples where AIR (green) provides improvement in re-id rank vs SDALF (red).

enabling constructive fusion with SDALF: At rank 5, weighted attributes achieve a 101% and 103% margin over fusing unweighted attributes for VIPeR and i-LIDS respectively; at rank 10, weighting increases by 103% and 104% and at rank 25, 109% and 102%.

		AIR	W. AIR	SDALF	SDALF+AIR	SDALF+W. AIR
VIPER	Rank 1	5.56	4.84	**18.28**	17.00	17.40
	Rank 5	15.76	17.44	37.88	36.48	**39.04**
	Rank 10	24.72	29.24	49.16	50.76	**50.84**
	Rank 25	45.16	50.60	67.96	72.52	**74.44**
	Rank 50	65.68	68.64	79.80	84.88	**86.44**
	nAUC	80.97	83.38	86.50	89.94	**90.75**
ILIDS	Rank 1	11.65	16.45	45.30	**47.95**	47.80
	Rank 5	30.25	35.60	72.15	74.30	**74.55**
	Rank 10	46.80	53.90	80.45	82.90	**83.80**
	Rank 25	63.25	68.35	89.50	89.15	**91.15**
	Rank 50	84.80	93.10	96.60	96.00	**97.00**
	nAUC	77.14	82.23	92.70	92.68	**93.27**

Table 2: Effects of attribute selection and weighting on re-identification rates, given as percentages. From left to right: unweighted AIR, weighted AIR, vanilla SDALF, SDALF fused with unweighted AIR, SDALF fused with weighted AIR.

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 [19]. Existing approaches to re-identification [4, 6, 20] 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 and can be measured reasonably reliably by a suitably trained detector. Importantly, although individual attributes vary in robustness, the combined and weighted attribute profile provides a strong discriminative cue for identity. This can complement and improve low-level feature based representations significantly. In developing a separate cue-modality and weighting strategy, our AIR is potentially complementary to most existing approaches, whether focused on features [4], or learning methods [28].

The proposed attribute-centric re-identification model provides an important contribution and novel research direction for practical re-identification: both by providing a complementary and informative mid-level cue, as well as by opening up completely new applications via the interpretable semantic representation. As a novel application, consider how semantic attributes could potentially be used to constrain or permute a search for a particular person, for example by specifying invariance to whether or not they have removed or added a hat.

Future Work Future improvements to our work include developing a fuller ontology of attributes, more powerful attribute detectors and improvements to our greedy heuristic for attribute search and weighting. More interesting challenges include addressing the sparsity of attribute training data by transfer learning better attribute detectors from large databases

(e.g., web crawl results) to the smaller re-identification benchmark datasets; learning explicitly viewpoint aware models of attributes; and inferring the most discriminative set and weighting of attributes at the level of each individual query instead of a given dataset.

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] Loris Bazzani, Marco Cristani, Alessandro Perina, Michela Farenzena, and Vittorio Murino. Multiple-shot Person Re-identification by HPE signature. In *International Conference on Pattern Recognition*. 2010.

[2] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3), 2011.

[3] Antitza Dantcheva, Carmelo Velardo, Angela D'Angelo, and Jean-Luc Dugelay. Bag of soft biometrics for person identification. *Multimedia Tools and Applications*, 51(2):739–777, October 2010.

[4] Michela Farenzena, Loris Bazzani, Alessandro Perina, Vittorio Murino, and Marco Cristani. Person re-identification by symmetry-driven accumulation of local features. In *IEEE Conference on Computer Vision and Pattern Recognition*. 2010.

[5] Niloofar Gheissari, Thomas B Sebastian, Peter H Tu, Jens Rittscher, and Richard Hartley. Person Reidentification Using Spatiotemporal Appearance. *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.

[6] Doug Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *IEEE International Workshop on Performance Evaluation for Tracking and Surveillance*, volume 3, 2007.

[7] Douglas Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. *European Conference on Computer Vision*, pages 262–275, 2008.

[8] Sung Ju Hwang, Fei Sha, and Kristen Grauman. Sharing features between objects and their attributes. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[9] Anil K Jain, Sarat C Dass, and Karthik Nandakumar. Soft biometric traits for personal recognition systems. In *Proceedings of International Conference on Biometric Authentication*, 2004.

[10] Qifa Ke and Takeo Kanade. Robust L1 norm factorization in the presence of outliers and missing data by alternative convex programming. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

[11] Hina Keval. CCTV Control Room Collaboration and Communication: Does it Work? In *Proceedings of Human Centred Technology Workshop*, 2006.

[12] Neeraj Kumar, A Berg, and P Belhumeur. Describable visual attributes for face verification and image search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10):1962–1977, 2011.

[13] Cristoph H. Lampert, Hannes Nickisch, and Stefan Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. June 2009.

[14] Jingen Liu and Benjamin Kuipers. Recognizing human actions by attributes. *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[15] Chen Change Loy, Tao Xiang, and Shaogang Gong. Multi-camera activity correlation analysis. In *IEEE Conference on Computer Vision and Pattern Recognition*. 2009.

[16] Chen Change Loy, Tao Xiang, and Shaogang Gong. Time-Delayed Correlation Analysis for Multi-Camera Activity Understanding. *International Journal of Computer Vision*, 90(1):106–129, May 2010.

[17] Christopher Madden, Eric Dahai Cheng, and Massimo Piccardi. Tracking people across disjoint camera views by an illumination-tolerant appearance representation. *Machine Vision and Applications*, 18(3-4):233–247, March 2007.

[18] Subhransu Maji, AC Berg, and Jitendra Malik. Classification using intersection kernel support vector machines is efficient. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[19] Toby Nortcliffe. *People Analysis CCTV Investigator Handbook*. Home Office Centre of Applied Science and Technology, 2011.

[20] Bryan Prosser, Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Person Re-Identification by Support Vector Ranking. In *British Machine Vision Conference*, 2010.

[21] Bernhard Schölkopf and Alexander J. Smola. *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT Press, 2002.

[22] Behjat Siddiquie, Rogerio Feris, and Larry Davis. Image ranking and retrieval based on multi-attribute queries. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[23] Gavin J.D. Smith. Behind the screens: Examining constructions of deviance and informal practices among CCTV control room operators in the UK. *Surveillance and Society*, 2:376–395, 2004.

[24] Daniel A. Vaquero, Rogerio S. Feris, Duan Tran, Lisa Brown, Arun Hampapur, and Matthew Turk. Attribute-based people search in surveillance environments. *Workshop on the Applications of Computer Vision*, pages 1–8, December 2009.

[25] David Williams. 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.

[26] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Associating groups of people. In *British Machine Vision Conference*, 2009.

[27] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Quantifying and transferring contextual information in object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99, 2011.

[28] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Person re-identification by probabilistic relative distance comparison. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.