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.

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, 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. The model learns an attribute-centric, parts-based feature representation which differs from and complements existing low-level features for re-identification that rely purely on bottom-up statistics for feature selection.

We are motivated by the operating procedures of human experts and recent research in attribute learning 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. 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. These attributes provide a very different source of information – effectively a separate modality – to the typical low-level features used.

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.

We first extract a low-level colour and texture feature vector denoted $\mathbf{x}$ from each person image I following the method in [3]. This consists of 464-dimensional feature vectors extracted from the image. Each vector uses 8 colour channels (RGB, HSV and YCbCr) and 21 texture filters (Gabor, Schmid) derived from the luminance channel.

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

Given the learned bank of attribute detectors, any person image can now be represented in a semantic attribute space by an N_a dimensional vector: $A(\mathbf{x}) = [p(a_1|\mathbf{x}_i^+),\ldots,p(a_{N_a}|\mathbf{x}_{N_a}^+)]^T$.

We investigate how attributes can be fused to enhance performance and we choose to build on *Symmetry-Driven Accumulation of Local Features* (SDALF), introduced by Farenzena *et al.* [1]. SDALF provides a low-level feature 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 features from which seperate 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:

$$d(I_p,I_q) = \beta_{WH}\cdot d_{WH}(WH(I_p),WH(I_q)) \quad (1)$$
$$+ \beta_{MSCR}\cdot d_{MSCR}(MSCR(I_p),MSCR(I_q)) \quad (2)$$
$$+ \beta_{RHSP}\cdot d_{RHSP}(RHSP(I_p),RHSP(I_q)) \quad (3)$$
$$+ \beta_{ATTR}\cdot d_{ATTR}(ATTR(I_p),ATTR(I_q)). \quad (4)$$

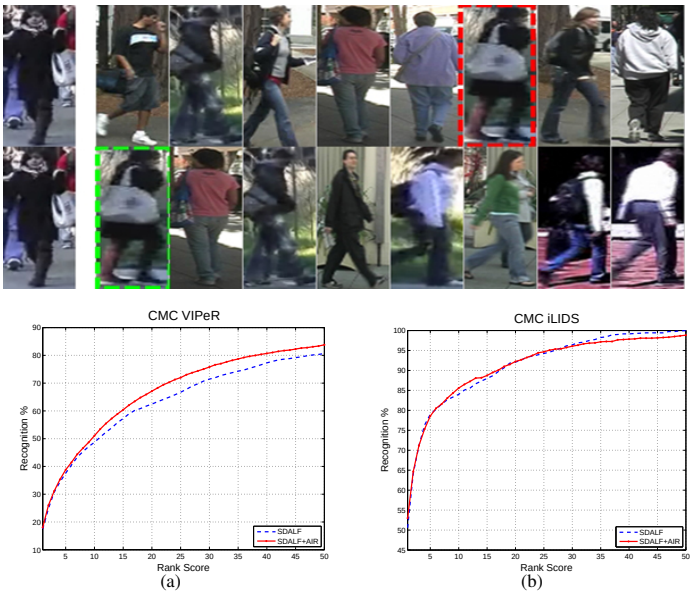

Figure 1: Above: Illustrative result for AIR (green) and SDALF (red); Below: CMC curves for (a) VIPeR ($p = 250$) ; and (b) i-LIDS ($p = 60$).

Here Eqs. (1-3) correspond to the three SDALF distance measures and Eq. (4) fuses our attribute-based distance distance metric. WH, $MSCR$ and $RHSP$ represent the metrics calculated for each of the separate SDALF features using Bhattacharyya.

To compare images' semantic attribute representation, we learn an $L2$-norm distance metric, d_{ATTR}. For diagonal weight matrix W:

$$d_{ATTR}(I_p,I_q) = \left(A(\mathbf{x}_p)-A(\mathbf{x}_q)\right)^T W \left(A(\mathbf{x}_p)-A(\mathbf{x}_q)\right), \quad (5)$$
$$= \sqrt{\sum_i w_i \left(p(a_i|\mathbf{x}_{p,i}^+) - p(a_q|\mathbf{x}_{q,i}^+)\right)^2}. \quad (6)$$

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 selects and weighs attributes to maximally improve the re-identification rate.

The re-identification performance of our complete system is summarised in Figure 1. In each case, our AIR outperforms vanilla SDALF [1], (which in turn decisively outperforms [2]). At the important rank 1 (perfect match), we obtain a relative improvement over SDALF of 3.2% and 4.8% for VIPeR and iLIDS respectively.

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.

[1] 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.

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

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