Re-id: Hunting Attributes in the Wild

Ryan Layne
r.d.c.layne@qmul.ac.uk
Timothy M. Hospedales
t.hospedales@qmul.ac.uk
Shaogang Gong
s.gong@qmul.ac.uk

Computer Vision Group
Queen Mary University of London
London E1 4NS.
http://qml.io/vision

Re-identification research breaks down into two main areas; developing effective representations that are discriminative for identity whilst invariant to lighting and viewpoint change [2] and development of learning methods trained to discriminate identities [1, 3]. Feature-centric approaches [2, 4] suffer from the problem that it is extremely challenging to obtain features that are discriminative enough to distinguish people reliably, while simultaneously being invariant to all the practical covariates such as motion blur, clutter, view angle and pose change, lighting and occlusion. In contrast, learning approaches [3] better use a given set of features, by discriminatively training models to maximise re-identification performance, for example metric learning [3] and support vector machines (SVM) [1].

In this paper we address these issues by automatically constructing a bottom-up attribute ontology, and learning an effective representation by large-scale mining noisy but abundant content from social photo sharing sites. We discover attributes automatically by clustering photo tag and comment data. These clusters are used to train a large bank of detectors, resulting in a number of visually detectable attributes[1]. This process is significantly more scalable than manually annotating data per surveillance site for attribute learning and the greater volume and diversity of data used to train these automatically discovered attributes results in a more generalisable attribute representation than conventional approaches on surveillance datasets. We validate our contribution and obtain excellent results on a set of four of the most challenging re-identification datasets.

We first apply self-tuned spectral clustering based on the BOW tf-idf metatext representations with a vocabulary of $\approx 5,000$ bigrams. We calculate the similarity between the frequency of the unigrams and bigrams rather than using the Levenshtein distance on the second gram within each bigram. Our intuition is that in our case it is the co-occurrence of the grams that is semantically relevent, not the similarity to other bigrams. Spectral clustering performs well regardless of the spatial arrangement of the underlying clusters, making it suitable for our needs. We extract bounding boxes of people from this extremely varied collection of photos; after conservatively thresholding person detection confidence, we are left with 69,000 person crops with corresponding meta-text features. We train an independent LDA model for each of the $N_a = 200$ discovered attribute clusters. Finally we build a representation for any person's image X in an internet-attribute semantic-space by stacking the positive-class posteriors from each detector into a N_a dimensional vector: $IA(X)$. To compensate for the differences in image quality between internet and surveillance data, we align the two datasets, using domain adaptation.

The attributes obtained thus far are trained directly from discovered text clusters so there is variability in their reliability and their usefulness for re-identification. We therefore address learning a linear weighting $\mathbf{w}$ to rescale the attributes IA such that they are weighted according to their maximum utility for re-identification.

[1] This is in contrast to expert defined ontology, which while intuitive to experts, may correspond to properties not possible to detect reliably with current vision techniques.

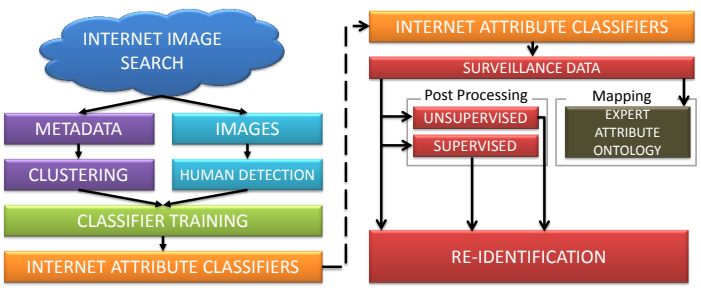

Figure 1: Schematic overview of our pipeline; Post-Processing modules such as distance-metric learning or domain-adaptation can be applied depending on the level of supervision available in order to boost "rank 1" or overall system performance as needed

Figure 2: Our FUSIA internet attribute (IA) representation provides a distributed representation of conventional expert-defined attributes such as "red shirt" (right), meaning that it can be mapped to them to allow query in terms of existing expert attribute ontologies (EAO) built for other surveillance data (SD).

We wish to enforce both a strong early-rank score, and good overall performance. To achieve this, we maximise the *product* of the CMC curve values $\hat{p}(k)$ at each rank k

$$\hat{P}_{\mathbf{w}}(k) = CMC_{\mathbf{w}}(k) = \frac{1}{n}\sum_{p=1}^{n} \mathbf{1}\left(k_p \leq k\right) \quad (1)$$

where k_p is the distribution of the ranks based on NN re-identification using $L1$ distances $D(IA_p, IA_g)$ between each attribute encoded probe $IA_p \in \mathcal{P}$ and all gallery member, $IA_g \in \mathcal{G}, g = 1,...,n$. Specifically we use greedy search to select the weight $\mathbf{w}$ that maximises the following metric when used to scale each dimension/attribute a:

$$\min_{\mathbf{w}} \prod_{k=1}^{n} \hat{P}_{\mathbf{w}}(k) \quad (2)$$

Finally, we integrate our representation with metrics based on other low-level features. Specifically, we fuse BR-SVM [1] (trained on ELF features), SDALF [2] and our weighted internet attributes after further discriminative training [3]. The resulting pseudo-metric's fusion weights *beta* can be trivially selected with standard optimisation methods:

$$D(X_p, X_g) = d_{KISS}(IA(X_p), IA(X_g)) \quad (3)$$
$$+ \beta_{SDALF} \cdot d_{SDALF}(X_p, X_g) \quad (4)$$
$$+ \beta_{BRELF} \cdot d_{BRELF}(X_p, X_g). \quad (5)$$

We perform nearest-neighbour re-identification with the above metric, obtaining state of the art re-identification performance (Figure 3). Our FUSIA representation also provides a distributed representation of conventional expert-defined attributes. It can be mapped to them, thus allowing queries in terms of existing attribute ontologies (Figure 2).

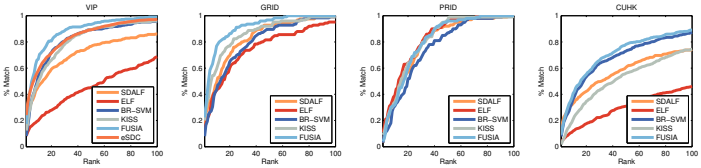

Figure 3: Overall re-identification performance of our FUSIA representation versus alternatives (CMC Curves)

[1] T. Avraham, I. Gurvich, M. Lindenbaum, and S. Markovitch. Learning Implicit Transfer for Person Re-identification. In *European Conference on Computer Vision, International Workshop on Re-identification*, 2012.

[2] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[3] M. Hirzer, P. M. Roth, M. Köstinger, and H. Bischof. Relaxed pairwise learned metric for person re-identification. In *European Conference on Computer Vision*, 2012.

[4] R. Layne, T. M. Hospedales, and S. Gong. Attributes-based Re-Identification. In S. Gong, M. Cristani, S. Yan, and C. C. Loy, editors, *Person Re-Identification*. Springer London, 2013.