Person Re-Identification by Support Vector Ranking

Bryan Prosser
bryan@eecs.qmul.ac.uk
Wei-Shi Zheng
jason@eecs.qmul.ac.uk
Shaogang Gong
sgg@eecs.qmul.ac.uk
Tao Xiang
txiang@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.

Problem: This paper addresses the problem of multi-camera person re-identification as a ranking problem using a scaleable combination model of Boosting and Support Vector Machines.

Related Work: Most existing work has concentrated on compiling feature sets as a template to describe an individual, followed by template matching using a direct distance measure chosen independently from the data [2, 4, 6]. Gray and Tao [3] proposed to use Adaboost to search through a large feature set for those features that are more relevant (more discriminative). Regardless of the choice of features and distance measures, re-identification by these approaches are difficult because there is often too much of an overlap between feature distributions of different objects.

Our Approach: In this work, we present a novel reformulation of the person re-identification problem. While previous approaches have looked at this problem as a classification of correct vs incorrect match, we propose an approach based on the information retrieval concept of document ranking [1]. The main difference between this approach and previous person re-identification techniques is that we are not concerned with comparing direct distance scores between correct and incorrect matches. Instead, we are only interested in the relative ranking of these scores that reflects the relevance of each likely match to the query image.

There are two main approaches for ranking: Boosting and Support Vector Machines (SVMs). Boosting involves picking weak rankers in each individual feature dimension, which is likely to lead to very weak rankers thus reducing matching effectiveness. In contrast, ranking SVMs seek to learn a ranking function in a higher dimensional feature space holistically (rather than individual feature dimension) where true matches and wrong matches can be made more separable than in the original feature space. However, they can also be computationally and spatially intensive when dealing with large numbers of negative samples. In order to overcome these problems, we explore an Ensemble of Primal SVMs (PRSVM) [1] that reduces the computation and memory cost by using several SVMs trained on subsets of the data while incorporating the SVM parameter C into the framework.

Model Overview: Given a dataset $X = \{(x_i, y_i)\}_{i=1}^m$ where x_i is a multi-dimensional feature vector representing the appearance of a person captured in one view, y_i is its label and m is the number of training samples (images of people). Each query feature vector x_i, has relevant feature vectors, $x_{i,j}^+$, and related irrelevant feature vectors $x_{i,j'}^-$. The goal of ranking any paired image relevance is to learn a ranking function δ for all pairs of $(x_i, x_{i,j}^+)$ and $(x_i, x_{i,j'}^-)$ such that the relevance ranking score $\delta(x_i, x_{i,j}^+)$ is larger than $\delta(x_i, x_{i,j'}^-)$. In our work δ is computed by a linear function w:

$$\delta(x_i, x_{i,j}) = w^\top |x_i - x_{i,j}|, \tag{1}$$

Rather than learning a batch mode RankSVM, which has a high memory cost and requires iterating over parameter values, we aim to learn a set of weak RankSVMs each computed on a small set of data and then combine them to build a stronger ranker using ensemble learning. The strong ranker w_{opt} is constructed by a set of weak rankers w_i as follows:

$$w_{opt} = \sum_i^N \alpha_i \cdot w_i, \tag{2}$$

where the weight vector α_i is learned iteratively using a boosting-based approach detailed in the paper.

In order to construct the weak rankers we divide a data set Z into n smaller subsets and each weak ranker is learned based on that group of data $\tilde{Z}_i$. Each subset $\tilde{Z}_i$, is comprised of a combination of a set group Z_i and some randomly chosen samples from the remaining set $Z - Z_i$. This overlap reduces the chances of the weak rankers having too poor a performance, while the process of using the subsets reduces the memory consumption. Additionally, for each Z_i we create a weak ranker for each value of the SVM importance weight C; that is if there are s candidate values of parameter C, then $N = s \cdot n$ weak rankers are computed. This makes selection of the parameter C in the primal-based RankSVM unified into the ensemble learning framework, without using any additional cross-validation that requires reforming training samples.

Figure 1: Resulting ranked images on the VIPeR dataset [3].

For each $\tilde{Z}_i$, we compute a weak ranker w_i by using a primal-based RankSVM of Chapelle and Keerthi [1], which is tractable given a moderate size dataset. To compute RankSVM, we first calculate a set of relevant and the related irrelevant absolute difference vectors in $\tilde{Z}_i$, denoted by $P_i = \{(\hat{x}_{i,s}^+, \hat{x}_{i,s}^-)\}$. Then, for some positive parameter C, the primal-based RankSVM solves the squared hinge loss function based on criterion of

$$w = \arg\min_w \frac{1}{2}\|w\|^2 + C\sum_{s=1}^{|P|} \ell\left(0, 1 - w^\top\left(\hat{x}_s^+ - \hat{x}_s^-\right)\right)^2, \tag{3}$$

Combining the PRSVM with our ensemble approach makes use of the computation gains of the PRSVM while reducing the memory useage by training with overlapping subsets of the data and allows us to incorporate the importance weight C, removing the need to run expensive cross-validation loops.

For validating our model, we test and compare a selection of non-learning, learning and ranking methods on both the VIPeR dataset [3] and the i-LIDS dataset [5]. We show that a ranking based approach to person re-identification gives significant improvement over existing re-identification techniques. We also show that the proposed Ensemble RankSVM is able to achieve comparable results to conventional RankSVM whilst being computationally much more efficient thus having superior scalability. A sample of these results can be seen in Figure 1.

[1] O. Chapelle and S. Keerthi. Efficient algorithms for ranking with svms. *Information Retrieval*, 18, 2010. to appear.

[2] N. Gheissari, T. Sebastian, P. Tu, and J. Rittscher. Person reidentification using spatiotemporal appearance. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 1528–1535, 2006.

[3] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *European Conference on Computer Vision*, Marseille, France, 2008.

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

[5] UK Home Office. i-lids dataset: Multiple camera tracking scenario, 2010. `http://www.ilids.co.uk`.

[6] B. Prosser, S. Gong, and T. Xiang. Multi-camera matching under illumination change over time. In *ECCV Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications*, Marseille, France, 2008.