Transferring a Semantic Representation for Person Re-Identification and Search

Zhiyuan Shi, Timothy M. Hospedales, Tao Xiang
Queen Mary, University of London, London E1 4NS, UK

Learning semantic attributes for person re-identification and description-based person search has gained increasing interest due to attributes' great potential as a pose and view-invariant representation. However, existing attribute-centric approaches have thus far underperformed state-of-the-art conventional approaches. This is due to their non-scalable need for extensive domain (camera) specific annotation. In this paper we present a new semantic attribute learning approach for person re-identification and search. Our model is trained on existing fashion photography datasets – either weakly or strongly labelled. It can then be transferred and adapted to provide a powerful semantic description of surveillance person detections, without requiring any surveillance domain supervision. The resulting representation is useful for both unsupervised and supervised person re-identification, achieving state-of-the-art and near state-of-the-art performance respectively. Furthermore, as a semantic representation it allows description-based person search to be integrated within the same framework.

Person re-identification (re-id) and description-based search are crucial tasks in visual surveillance. They underpin many fundamental applications including multi-camera tracking, crowd analysis and forensic search. Both tasks aim to retrieve images of a specific person, but differ in the query used. Person re-identification queries using an image from a different view (e.g., in multi-camera tracking), while person search uses a textual person description (e.g., eyewitness description). Despite extensive research [3, 6], these tasks remain unsolved due to various challenges including the variability of viewpoints, illumination, pose, partial occlusion, low-resolution and motion-blur [2].

	Method	VIPeR				CUHK01				PRID450S			
		r=1	r=5	r=10	r=20	r=1	r=5	r=10	r=20	r=1	r=5	r=10	r=20
single	SDC [9]	25.1	44.9	56.3	70.9	15.1	25.4	31.8	40.9	23.7	38.4	46.1	58.5
	SDALF [1]	19.9	38.9	49.4	65.7	9.9	22.6	30.3	41.0	17.4	30.9	40.8	55.2
	Our unsupervised	**27.7**	**55.3**	**68.3**	**79.7**	**23.3**	**35.8**	**46.6**	**60.7**	**28.5**	**48.9**	**59.6**	**71.3**
fused	SDC_*Final* (eSDC) [9]	26.7	50.7	62.4	76.4	19.7	32.7	40.3	50.6	25.5	40.6	48.4	61.4
	Our unsupervised_*Final*	**29.3**	**52.7**	**66.8**	**79.7**	**22.4**	**35.9**	**47.9**	**64.5**	**29.0**	**49.4**	**58.4**	**69.8**

Table 1: Person re-id performance evaluated by matching accuracy @ rank r (%): unsupervised learning approaches. The best results for single-cue and fused-cue methods are highlighted in bold separately.

	Method	VIPeR				CUHK01				PRID450S			
		r=1	r=5	r=10	r=20	r=1	r=5	r=10	r=20	r=1	r=5	r=10	r=20
single	KML [7]	32.3	65.8	79.7	90.9	24.0	38.9	46.7	55.4	32.4	54.4	62.4	69.6
	KISSME [5]	19.6	48.0	62.2	77.0	8.4	25.1	38.7	50.2	26.5	47.8	57.6	68.5
	SCNCD [8]	**33.7**	62.7	74.8	85.0	-	-	-	-	41.5	66.6	75.9	84.4
	Our supervised	31.1	**68.6**	**82.8**	**94.9**	32.7	51.2	64.4	**76.3**	**43.1**	**70.5**	**78.2**	**86.3**
fused	KML_*Final* [7]	36.1	68.7	80.1	85.6	-	-	-	-	-	-	-	-
	SCNCD_*Final* [8]	37.8	68.6	81.0	90.5	-	-	-	-	41.6	68.9	**79.4**	**87.8**
	Our supervised_*Final*	41.6	71.9	**86.2**	**95.1**	31.5	52.5	**65.8**	**77.6**	**44.9**	**71.7**	77.5	86.7

Table 2: Matching accuracy @ rank r (%): supervised learning approaches on re-id.

In this paper we contribute a new framework that is capable of learning a semantic attribute model from existing fashion datasets, and adapting the resulting model to facilitate person re-identification and search in the surveillance domain. In contrast to most existing approaches to attribute detection which are based on discriminative modelling, we take a generative modelling approach based on the Indian Buffet Process (IBP) [4]. The generative formulation provides key advantages including: joint learning of all attributes; ability to naturally exploit weakly-annotated (image-level) training data; as well as unsupervised domain adaptation through Bayesian priors.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

Figure 1: Visualisation of our model output. Each patch is colour-coded to show the inferred dominant attribute of two types.

Importantly a IBP-based model provides the favourable property of combining attributes factorially in each local patch. This means that our model can differentiate potentially ambiguous situations such as Red-Shirt+Blue-Jeans versus Red-Jeans+Blue-Shirt (See Fig. 2). Moreover, with this representation, attribute combinations that were rare or unseen at training time can be recognised at test time so long as they are individually known (e.g. Shiny-Yellow-Jeans).

Our framework overcomes the significant problem of domain shift between fashion and surveillance data in an unsupervised way by Bayesian adaptation. It can exploit both strongly and weakly annotated source data during training, but is always able to produce a strong (patch-level) attribute prediction during testing (See Fig. 1). The resulting representation is highly person variant while being view-invariant, making it ideal for person re-id, where we obtain state-of-the-art results (See Table 1 and Table 2). Moreover, as the representation is semantic (nameable or describable by a human), we are able to unify description based person search within the same framework, where we also achieve state-of-the-art results.

Figure 2: Person search qualitative results. The top ranked images for each query are shown. Red boxes are false detections.

[1] Michela Farenzena, Loris Bazzani, Alessandro Perina, Vittorio Murino, and Marco Cristani. Person re-identification by symmetry-driven accumulation of local features. In *CVPR*, 2010.

[2] Shaogang Gong, Marco Cristani, Chen Change Loy, and Timothy M. Hospedales. The re-identification challenge. In *Person Re-Identification*. Springer London, 2014.

[3] Shaogang Gong, Marco Cristani, Shuicheng Yan, and Chen Change Loy, editors. *Person Re-Identification*. Springer, 2014.

[4] Thomas L. Griffiths and Zoubin Ghahramani. The indian buffet process: An introduction and review. *JMLR*, 2011.

[5] Martin Koestinger, Martin Hirzer, Paul Wohlhart, Peter M. Roth, and Horst Bischof. Large scale metric learning from equivalence constraints. In *CVPR*, 2012.

[6] Roberto Vezzani, Davide Baltieri, and Rita Cucchiara. People re-identification in surveillance and forensics: a survey. *ACM Computing Surveys*, December 2013.

[7] Fei Xiong, Mengran Gou, Octavia Camps, and Mario Sznaier. Person re-identification using kernel-based metric learning methods. In *ECCV*, 2014.

[8] Yang Yang, Jimei Yang, Junjie Yan, Shengcai Liao, Dong Yi, and StanZ. Li. Salient color names for person re-identification. In *ECCV*, 2014.

[9] Rui Zhao, Wanli Ouyang, and Xiaogang Wang. Unsupervised salience learning for person re-identification. In *CVPR*, 2013.