

# Custom Pictorial Structures for Re-identification

Dong Seon Cheng<sup>1</sup>

cheng.dong.seon@gmail.com

Marco Cristani<sup>1,2</sup>

marco.cristani@univr.it

Michele Stoppa<sup>2</sup>

michele.stoppa@iit.it

Loris Bazzani<sup>1</sup>

loris.bazzani@univr.it

Vittorio Murino<sup>1,2</sup>

http://profs.sci.univr.it/~swan

<sup>1</sup> Dipartimento di Informatica

University of Verona

Italy

<sup>2</sup> Istituto Italiano di Tecnologia

Via Morego, 30

16163 Genova, Italy

Appearance-based person re-identification (re-id) consists in matching images of individuals in different locations over various non-overlapping camera views using only visual cues. Traditionally, appearance-based methods are divided between *learning-based* and *direct* approaches. In the former, a dataset is split into training and test sets [7, 9], with the training individuals used to learn features and/or strategies for combining features to achieve high re-id accuracy, and the test ones used as validation. Direct methods are instead pure feature extractors [2].

In this paper, we present a novel methodology for direct appearance-based person re-id, based on Pictorial Structures (PS) for human body pose estimation. PS essentially rely on two components: one capturing the local appearance of body parts, and the other representing an articulated body structure [5]. Inference in a PS involves finding the MAP spatial configuration of the parts, *i.e.*, the body pose. We build upon the PS framework of [1], where general part detectors localize the body parts, and a kinematic tree prior captures the structural knowledge.

Our proposal takes inspiration from how humans perform re-id. Taking a subset of the VIPeR dataset, we set up a simple re-id experiment where subjects were asked to match test probes to candidate galleries, while being monitored with an eye-tracker system. The obtained fixation maps indicate a tendency to scan salient parts of the body, looking for part-to-part correspondences. We think that encoding and exploiting the human appearance per parts is a convenient strategy for re-id, and PS are the best tool for this task. PS are usually fitted on individual images (see Fig. 1(a)), as independent entities, and we exploit this setting for *single-shot* re-id, which consists in matching pairs of images, a probe and a gallery image for each subject. After fitting a PS on all images, from each localized part we extract an ensemble of features, encoding complementary aspects, such as the chromatic content and the spatial arrangement of colors. The first aspect is captured by HSV histograms, while the second aspect is codified by Maximally Stable Color Regions (MSCR) [6], previously adopted for re-id in [4]. The features of each part are subsequently combined into a single ID signature. Matching between signatures is carried out by standard distance minimization strategies.

On the other hand, *multi-shot* re-id occurs when each subject has multiple images (see Fig. 1(b)), either in the gallery and/or the probe set, which can be exploited to accumulate more visual information and ensure higher re-id accuracy. In this case, we propose a strategy to improve the PS fitting on images of the same subject. This task has received little attention in the literature, like [8], where a large number of consecutive images per person was used, whereas a re-id task often provides only few (2-5) non-consecutive images. Our idea is to learn the local appearance of each part in a given subject so that ad-hoc appearance part detectors can provide more accurate PS fitting. The detectors we used are multidimensional Gaussian filters capturing the appearance of every pixel in each part (see Fig. 1(d)). Moreover, chances of bad learning due to the scarcity of samples per person is mitigated by employing spatial reasoning, *i.e.*, by augmenting the statistics of a pixel with similar neighboring pixel values in a surrounding region, identified through non-parametric Mean Shift segmentation [3] (see Fig. 1(c)). Using these filters on the subject images, we obtain evidence maps for every body part that can be injected in the PS body pose estimation algorithm to obtain a better fitting. Iterating between pose estimation and appearance modeling allows us to refine the fitting at each step.

We have called this new model Custom Pictorial Structure (CPS). Once CPS is fitted on data, features are extracted from each instance as

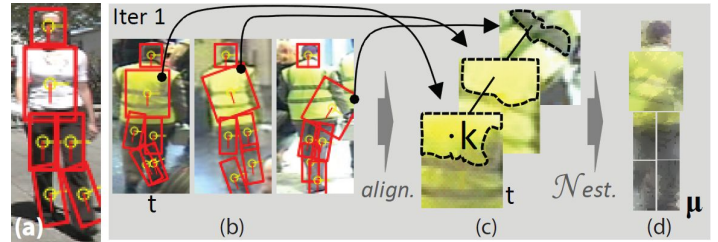


Figure 1: Our approach. Single-shot PS in (a). Multi-shot CPS at iteration 1: (b) initial PS fitting; (c) the parts are aligned and per-pixel statistics is collected employing spatio-temporal reasoning; (d) the ad-hoc part detectors are estimated, whose means  $\mu$  are shown. These detectors are used to improve the PS fitting iteratively (not shown).

in the single-shot case, and the individual signatures are pooled together to obtain a multi-shot ID signature. The matching policy between multi-shot signatures is based on finding, among the individual signatures of the compared individuals, the pair whose distance is minimal.

Summarizing, in this paper we propose:

- the adoption of PS for single-shot re-id, and appropriate per-part feature extractors;
- a novel fitting strategy for PS, specifically suited for few, non-consecutive images of the same subject, and a strategy for matching multi-shot signatures.

Experiments have been carried out on all the available re-id datasets (iLIDS, ETHZ1,2,3, VIPeR) with convincing results in all modalities and outright best performances in the multi-shot case. Moreover, we created a novel dataset starting from the CAVIAR tracking data presenting a unique combination of challenges. We finally evaluate our method against the human capability of re-id on a subset of the VIPeR dataset.

- [1] M. Andriluka, S. Roth, and B. Schiele. Pictorial Structures Revisited: People Detection and Articulated Pose Estimation. In *CVPR*, pages 1014–1021, Miami, USA, June 2009. URL <http://www.d2.mpi-inf.mpg.de/node/381>.
- [2] S. Bak, E. Corvee, F. Bremond, and M. Thonnat. Person Re-identification Using Spatial Covariance Regions of Human Body Parts. In *AVSS*, August 2010.
- [3] D. Comaniciu and P. Meer. Mean Shift: A Robust Approach Toward Feature Space Analysis. *IEEE TPAMI*, 24(5):603–619, 2002.
- [4] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person Re-Identification by Symmetry-Driven Accumulation of Local Features. In *CVPR*, 2010.
- [5] P. F. Felzenszwalb and D. P. Huttenlocher. Pictorial Structures for Object Recognition. *IJCV*, 61(1):55–79, 2005.
- [6] Per-Erik Forssén. Maximally Stable Colour Regions for Recognition and Matching. In *CVPR*, June 2007.
- [7] B. Prosser, W. Zheng, S. Gong, and T. Xiang. Person Re-Identification by Support Vector Ranking. In *BMVC*, 2010.
- [8] D. Ramanan, D. A. Forsyth, and A. Zisserman. Tracking People by Learning Their Appearance. *IEEE TPAMI*, 29(1):65–81, 2007.
- [9] W.R. Schwartz and L.S. Davis. Learning discriminative appearance-based models using partial least squares. In *XXII SIBGRAP 2009*, 2009.