

A Multi-Stage Approach for Fast Person Re-identification

Bahram Lavi^(✉), Giorgio Fumera, and Fabio Roli

Department of Electrical and Electronic Engineering,
University of Cagliari Piazza d'Armi, 09123 Cagliari, Italy
{lavi.bahram,fumera,roli}@diee.unica.it
<http://pralab.diee.unica.it>

Abstract. One of the goals of person re-identification systems is to support video-surveillance operators and forensic investigators to find an individual of interest in videos taken by a network of non-overlapping cameras. This is attained by sorting images of previously observed individuals for decreasing values of their similarity with the query individual. Several appearance-based descriptors have been proposed so far, together with ad hoc similarity measures, mostly aimed at improving ranking quality. We address instead the issue of the processing time required to compute the similarity values, and propose a multi-stage ranking approach to attain a trade-off with ranking quality, for any *given* descriptor. We give a preliminary evaluation of our approach on the benchmark VIPeR data set, using different state-of-the-art descriptors.

1 Introduction

Person re-identification is a computer vision task consisting in recognizing an individual who had previously been observed over a network of video surveillance cameras with non-overlapping fields of view [1]. One of its applications is to support surveillance operators and forensic investigators in retrieving videos where an individual of interest appears, using an image of that individual as a query (*probe*). To this aim, the video frames or tracks of all the individuals recorded by the camera network (*template gallery*) are sorted by decreasing similarity to the probe, allowing the operator to find the occurrences (if any) of the individual of interest, ideally, in the top positions. This task is challenging in typically unconstrained surveillance settings, due to low image resolution, unconstrained pose, illumination changes, and occlusions, which do not allow to exploit strong biometrics like face. Clothing appearance is therefore one of the most widely used cues, although cues like gait and anthropometric measures have also been investigated. Most of the existing person re-identification techniques are based on a specific descriptor of clothing appearance (typically including color and texture), and a specific similarity measure between a pair of descriptors which can be either manually defined or learnt from data [1, 3, 4, 6, 12]. Their focus is to improve recognition accuracy, i.e., ranking quality.

In this work we address the complementary issue of the processing time required to compute the similarity measure (*matching score*). Many of the similarity measures defined so far are indeed rather complex, and require a relatively high processing time (e.g., [3, 16]). Moreover, in real-world application scenarios the template gallery can be very large, and even when a single matching score is fast to compute (e.g., the Euclidean distance between fixed-length feature vectors [12]), computing it for all templates is time-consuming. This issue has been addressed so far only by a few works [2, 9, 15].

Inspired by the multi-stage approach for generic classification problems of [14], and by the object detection approach based on a cascade of classifiers of [19], we propose a multi-stage ranking approach specific to person re-identification, aimed at attaining a trade-off between ranking quality and processing time. Both the approaches of [14, 19] consist in a cascade of classifiers, where each stage uses features that are increasingly more discriminant but also more costly [14] or slower to compute [19]. The goal of [14] is to assign an input instance (e.g., a medical image) to one of the classes (e.g., the outcome of a diagnosis) with a predefined level of confidence, using features (e.g., medical exams) with the lowest possible cost; if a classifier but the last one does not reach the desired confidence level, it *rejects* the input instance (i.e., withholds making a decision), and sends it to the next stage. This approach has later been exploited to attain a trade-off between classification accuracy and processing time, e.g., in handwritten digit classification [8, 17, 18]. The similar approach of [19] focuses on designing fast object detectors: its goal is to detect background regions of the input image as quickly as possible, using classifiers based on features fast to compute, and to focus the attention on regions more likely to contain the object of interest, using classifiers based on more discriminant features that also require a higher processing time.

The above approaches cannot be directly applied to person re-identification, which is a *ranking* problem, not a classification one. In this paper we adapt it to person re-identification, to attain a trade-off between ranking quality and processing time, for a *given* descriptor and similarity measure. To this aim, we build a multi-stage re-identification system in which the chosen descriptor is used in the last stage, whereas “reduced” versions of the *same* descriptor are used in previous stages, characterized by a decreasing processing time and a lower recognition accuracy; the first stage ranks *all* templates, whereas each subsequent stage re-ranks a subset of the top-ranked templates by the previous stage. After summarizing in Sect. 2 related re-identification approaches, in Sect. 3 we describe our approach and discuss possible design criteria. We then give in Sect. 4 a preliminary evaluation of the attainable trade-off between recognition accuracy and processing cost, using the benchmark VIPeR data set and four state-of-the-art descriptors.

2 Related Work

As mentioned in Sect. 1, only a few works have addressed so far the issue of the processing time required to compute the matching scores in person

re-identification systems [2, 9, 15]. In particular, only in [2] the proposed solution is a multi-stage system: the first stage selects a subset of templates using a descriptor with a low processing time for computing matching scores (a bag-of-words feature representation and an indexing scheme based on inverted lists was proposed). The second stage ranks only the selected templates using a different, more complex mean Riemann covariance descriptor. Differently from our approach, only two stages are used in [2], based on *different* and *specific* descriptors; and only a subset of templates is ranked, possibly missing the correct one. We point out that the achieved reduction of processing time, with respect to ranking all templates using the second-stage descriptor, was not reported. In [15] we proposed a dissimilarity-based approach for *generic* descriptors made up of bags of local features, possibly extracted from different body parts, aimed at reducing the processing time for computing the matching scores. It converts any such descriptor into a fixed-size vector of dissimilarity values between the input image and a set of representative bags of local features (“prototypes”) extracted from the template gallery; the matching score can then be quickly computed, e.g., as the Euclidean distance. The method of [9] aims at reducing the processing time in the specific multi-shot setting (when several images per individual are available), and for specific descriptors based on local feature matching, e.g., interest points. It first filters irrelevant interest points, then uses a sparse representation for the remaining ones, before computing the matching scores.

Other authors proposed multi-stage systems for improving ranking quality, without considering processing time [5, 7, 11, 13, 20]. In the two-stage system of [5] a manually designed descriptor is used in the first stage, which returns the operator the 50 top-ranked templates; if they do not include the probe individual, a classifier is trained to discriminate the latter from other identities, and is then used to re-rank the remaining templates. In [13] person re-identification is addressed as a content-based image retrieval task with relevance feedback, with the aim of increasing recall, assuming that several instances of a probe can be present in the template gallery. In each stage (i.e., iteration of relevance feedback) only the top-ranked templates are shown to the operator, and his feedback is used to adapt the similarity measure to the probe at hand. A similar strategy was proposed in [11]: at each stage only the top-ranked templates are presented to the operator, who is asked to select an individual with a different identity and a very different appearance than the probe. A post-rank function is then learnt, exploiting this feedback and the probe image, and the remaining templates are re-ranked in the next stage. A similar, two-stage approach was proposed in [20]: after presenting to the operator the top-ranked templates from the first stage, the operator is asked to label some pairs of locally similar and dissimilar regions in the probe and template images; this feedback is exploited to re-rank the templates. Another two-stage approach was proposed in [7]: a small subset of the top-ranked templates by of a given first-stage descriptor is re-ranked by the second stage, using a manifold-based method that exploits three specific low-level features.

3 Proposed Approach

Let D denote a given descriptor, \mathbf{T} and \mathbf{P} the descriptors of a template and probe image, respectively, $m(\cdot, \cdot)$ the similarity measure between two descriptors, and $G = \{\mathbf{T}_1, \dots, \mathbf{T}_n\}$ the template gallery. For a given \mathbf{P} , a standard re-identification system computes the matching scores $m(\mathbf{P}, \mathbf{T}_i)$, $i = 1, \dots, n$, and sorts the template images by decreasing values of the score. Ranking quality is typically evaluated using the cumulative matching characteristic (CMC) curve, i.e., the probability (recognition rate) that the correct identity is within the first r ranks, $r = 1, \dots, n$. By definition, the CMC curve increases with r , and equals 1 for $r = n$. If t is the processing time for a single matching score, the time for computing all the scores on G is $n \times t$.

To attain a trade-off between recognition accuracy and processing cost, for a given descriptor and similarity measure, the solution we investigate is a multi-stage architecture (see Fig. 1) based on the following rationale. Consider first two descriptors D_1 and D_2 with similarity measures m_1 and m_2 and processing cost t_1 and t_2 ; assume that D_1 is less accurate than D_2 , i.e., its CMC curve lies below the one of D_2 , as in Fig. 2 (left). The CMC curve of a less accurate descriptor approaches the one of a more accurate one as r increases, and the difference drops below a given threshold Δ after some rank $r_1 < n$ (see Fig. 2, left). This means that D_1 and D_2 exhibit almost the same recognition accuracy for $r > r_1$. If $t_1 < t_2$, one can attain a similar recognition accuracy as D_2 , with a lower processing cost, by a two-stage ranking procedure: D_1 is used first to rank all n templates; the top- r_1 ones are then re-ranked using D_2 . The corresponding processing time is $T = n \times t_1 + r_1 \times t_2$. If all the n templates are ranked by D_2 , instead, the processing time is $T_2 = n \times t_2$. In order for $T < T_2$, r_1 must satisfy:

$$r_1 < n(1 - t_1/t_2) . \tag{1}$$

If (1) does not hold, one can attain $T < T_2$ by re-ranking in the second stage a lower number of templates than r_1 , at the expense of a lower accuracy.

The above approach can be extended to a higher number of stages $N > 2$, using descriptors D_1, \dots, D_N with increasing accuracy and processing time, $t_1 <$

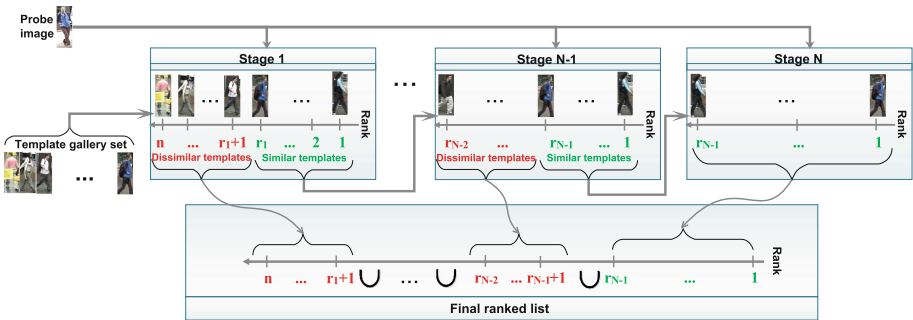


Fig. 1. Overview of the proposed multi-stage ranking approach.

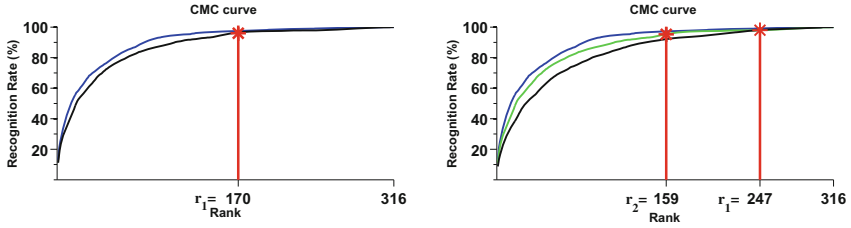


Fig. 2. An example of the criterion used in this work for selecting the number of templates to be ranked in each stage, for a template gallery of size $n = 316$. Left (two-stage system): CMC curve of the first- (black) and second-stage descriptor (blue), and the rank r_1 after which the difference between the two CMC curves is lower than $\Delta = 1\%$. Right: CMC curves of the three stages of a three-stage system, and the corresponding values of r_1 and r_2 ; note that r_2 has been obtained from the CMC curves of the second and third stages, computed on a template gallery of size $r_1 < n$. (Color figure online)

$t_2 < \dots < t_N$. Let n_i be the number of matching scores computed by the i -th stage, with $n_1 = n$; since each stage computes a higher number of scores than the next one ($n_1 > n_2 > \dots > n_N$), the overall processing time is:

$$T = \sum_{i=1}^N n_i \times t_i . \quad (2)$$

Let $T_N = n \times t_n$ be the processing time to rank all templates using the most accurate descriptor D_N . For the multi-stage system to attain an accuracy as much similar as possible to the one of D_N , with $T < T_N$, the values n_i ($i > 1$) must be chosen by generalizing the above criterion. More precisely, for $i = 2, \dots, N$: (i) find the lowest rank r_{i-1} such that the CMC curves of D_{i-1} and D_N , computed on a template gallery of size n_{i-1} , are closer than a given threshold Δ ; (ii) to attain an overall accuracy similar to the one of D_N , choose $n_i = r_{i-1}$ (see Fig. 2, right). If this choice leads to $T \geq T_N$, then T can be decreased by choosing lower values of n_i , $i > 1$, at the expense of a lower recognition accuracy.

To attain a trade-off between accuracy and processing time for a *given* descriptor D and similarity measure m , the above multi-stage architecture can be implemented by using D in the last stage, i.e., $D_N = D$, and defining D_{N-1} , D_{N-2} , \dots , D_1 as increasingly *simpler* versions of the *same* descriptor D , i.e., versions exhibiting a decreasing recognition accuracy and $t_{N-1} > t_{N-2} > \dots > t_1$. This is the solution we empirically investigate in the rest of this paper. The definition of simpler versions of a given descriptor depends on the specific descriptor at hand. As a simple example, if a descriptor includes color histograms and a distance measure between them, one could reduce the number of bins. In the next section we shall give concrete examples on four different descriptors.

4 Experimental Analysis

We evaluate our approach on two- and three-stage systems, using a benchmark data set and four different appearance descriptors.

4.1 Experimental Setting

Data Set. VIPeR [4] is a benchmark, challenging dataset made up of two images for each of 632 individuals, acquired from two different camera views, with significant pose and illumination changes. As in [3], we repeated our experiments on ten different subsets of 316 individuals each; for each individual we use one image as template and one as probe; we then report the average CMC curve.

Descriptors. SDALF [3]¹ subdivides the body into left and right torso and legs. Three kinds of features are extracted from each part: maximally stable color regions, i.e., elliptical regions (blobs) exhibiting distinct color patterns (their number depends on the specific image), with a minimum size of 15 pixels; a $16 \times 16 \times 4$ -bins weighted HSV color histogram (wHSV); and recurrent high-structured patches (RHSP) to characterize texture. A specific similarity measure is defined for each feature; the matching score is computed as their linear combination. We did not use RHSP due to its relatively lower performance. We increased the minimum MSCR blob size to 65 and 45 for the first and second stage, respectively (which reduces the number of blobs), and reduced the corresponding number of wHSV histogram bins to $3 \times 3 \times 2$ and to $8 \times 8 \times 3$.

gBiCov is based on biologically-inspired features (BIF) [12],² which are obtained by Gabor filters with different scales over the HSV color channels. The resulting images are subdivided into overlapping regions of 16×16 pixels; each region is represented by a covariance descriptor that encodes shape, location and color information. A feature vector is then obtained by concatenating BIF features and covariance descriptors; PCA is finally applied to reduce its dimensionality. We obtained faster versions of gBiCov by increasing the region size to 32×64 and 16×32 pixels for the first and second stage, respectively.

LOMO [10]³ extracts an $8 \times 8 \times 8$ -bins HSV color histogram and two scales of the Scale Invariant Local Ternary Pattern histogram (characterizing texture) from overlapping windows of 10×10 pixels; only one histogram is retained from all windows at the same horizontal location, obtained as the maximum value among all the corresponding bins. These histograms are concatenated with the ones computed on a down-sampled image. A metric learning method is used to define the similarity measure. We increased the window size to 20×20 and 15×15 for the first and second stage, respectively, and decreased the corresponding number of bins of the HSV histogram to $3 \times 3 \times 2$ and $4 \times 4 \times 3$.

MCM (Multiple Component Matching) [16]⁴ subdivides body into torso and legs, and extracts 80 rectangular, randomly positioned image patches from each

¹ Source code: <http://www.lorisbazzani.info/sdalf.html>.

² Source code: <http://vipl.ict.ac.cn/members/bpma>.

³ Source code: http://www.cbsr.ia.ac.cn/users/sclicao/projects/lomo_xqda/.

⁴ The source code is available upon request.

part. Each patch is described by a $24 \times 12 \times 4$ -bins HSV histogram. The similarity measure is the average k -th Hausdorff distance between the set of patches of each pair of corresponding body parts. In our experiments we reduced the number of patches to 10 and 20 for the first and second stage, respectively, and the corresponding number of bins of the HSV histogram to $3 \times 3 \times 2$ and $12 \times 6 \times 2$.

4.2 Experimental Results

We carried out our experiments using an Intel Core i5 2.6 GHz CPU. For each descriptor we designed a two- and a three-stage system. We used the same version of a given descriptor both in the first stage of two-stage systems and in the second stage of three-stage systems. For each descriptor, the average time for computing a single matching score in each stage is reported in Table 1; note that for MCM the first- and second-stage versions have a much lower processing time than the original one, with respect to the other descriptors: this is due to the use of the Hausdorff distance, which makes the processing time proportional to the *square* of the number of image patches (see above). The number of matching scores n_i computed by the i -th stage ($i > 1$), for each descriptor, is reported in Table 2. These values were computed using the criterion described in Sect. 3, setting a threshold $\Delta = 1\%$ and $\Delta = 0.5\%$ for two- and three-stage systems, respectively. We point out that this criterion aims at keeping recognition accuracy as high as possible (possibly identical to that of the original descriptor), while reducing processing time. We also remind the reader that n_1 always equals the total number of templates, which is $n = 316$ in our experiments. The average CMC curves are reported in Figs. 3 (two-stage systems) and 4 (three-stage systems). Inside each plot we also report a comparison between the CMC curve of multi-stage systems obtained from different values of Δ . The ratio of the corresponding processing time with respect to the one of the original, most accurate descriptor, is reported in Table 2.

Figures 3 and 4 show that the CMC curves of multi-stage systems are nearly identical to the ones of the corresponding original descriptor; some differences are visible in two- and three-stage systems for the SDALF and MCM descriptors, only for ranks higher than 50. The reduction in processing time was however not high: Table 2 shows that it was 25% to 32% for two-stage systems, and 17% to 27% for three-stage systems, depending on the descriptor. Reducing the number of matching scores computed by each stage, through the use of a higher

Table 1. Average processing time (in msec.) for computing one matching score in each stage of two- and three-stage systems, for each of the four descriptors.

		SDALF	gBiCov	LOMO	MCM
Two-stage	1st stage	2.08	0.0057	0.0023	0.060
Three-stage	1st stage	1.60	0.0015	0.0017	0.051
	2nd stage	2.08	0.0057	0.0023	0.060
Original descriptor (last stage)		9.44	0.0400	0.0370	27.400

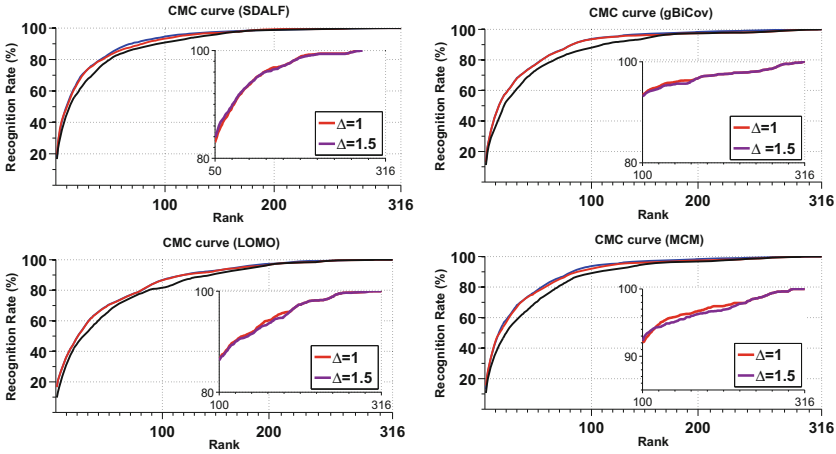


Fig. 3. CMC curves of two-stage systems, for $\Delta = 1\%$ (see text for the details). Blue: original descriptor; black: first-stage; red: two-stage system. The inner plots show a comparison between the CMC curves of two-stage systems obtained for $\Delta = 1\%$ and $\Delta = 1.5\%$: they differ only for the highest ranks shown in these plots. Figure is best viewed in color. (Color figure online)

threshold Δ , affected only slightly the ranking accuracy of multi-stage systems, and only for the highest ranks (see the CMC curves inside the boxes in Figs. 3 and 4). On the other hand, this provided a significant reduction in processing time, especially on three-stage systems, where it ranged from 29% to 53% for $\Delta = 1.5\%$.

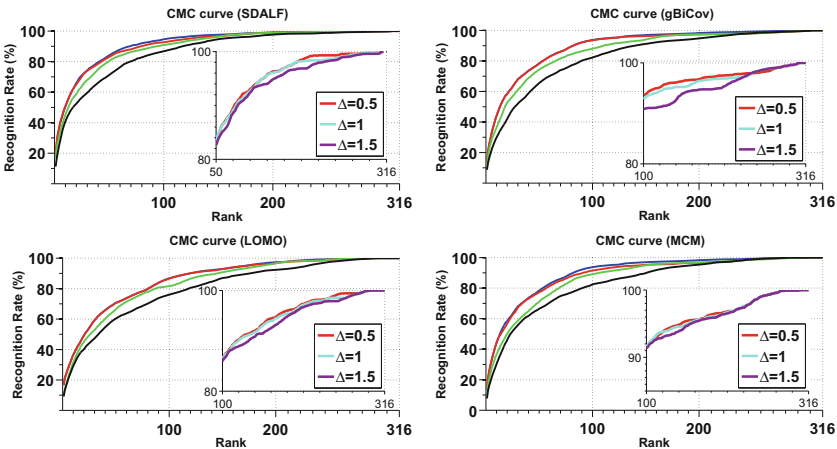


Fig. 4. CMC curves of three-stage systems, for $\Delta = 0.5\%$ (see text for the details). Blue: original descriptor; black: first-stage; green: second stage; red: three-stage system. Inner plots: comparison between the CMC curves of three-stage systems obtained for $\Delta = 0.5\%$, 1% , and 1.5% , which differ only in the highest ranks shown in these plots. Figure is best viewed in color. (Color figure online)

The above results provide evidence that the proposed multi-stage ranking approach is capable of improving the trade-off between recognition accuracy and processing time of a given descriptor. The attainable improvement depends on the specific descriptor, i.e., on its similarity measure and on the parameters that can be modified to obtain faster versions of it. This is clearly visible from Table 2; in particular, using the same number of stages and the same criterion to choose the number of matching scores computed by each stage, we consistently attained the lowest reduction in processing time using SDALF. We point out that our experiments were not aimed at finding the best set of parameters, and their best values, to optimize the trade-off between recognition accuracy and processing time of each descriptor. Accordingly, we believe that a more focused choice could provide a higher reduction of processing time than the one attained in our experiments, without affecting recognition accuracy.

Table 2. Number of matching scores computed by each stage but the first one in the two- and three-stage systems, for each of the four descriptors, and for the different values of Δ considered in the experiments. The ratio of the corresponding processing time, with respect to the original descriptor, is also reported.

	Two-stage systems				Three-stage systems								
	$\Delta = 1$		$\Delta = 1.5$		$\Delta = 0.5$			$\Delta = 1$			$\Delta = 1.5$		
	n_2	Time	n_2	Time	n_2	n_3	Time	n_2	n_3	Time	n_2	n_3	Time
SDALF	150	0.7	138	0.66	271	150	0.83	214	151	0.8	192	130	0.71
gBiCov	170	0.68	163	0.66	283	178	0.73	247	160	0.66	209	133	0.55
LOMO	196	0.68	186	0.65	282	198	0.73	265	192	0.71	254	154	0.58
MCM	236	0.75	162	0.51	248	235	0.75	239	170	0.54	230	148	0.47

5 Conclusion

We proposed a multi-stage ranking approach for person re-identification systems, aimed at attaining a trade-off between ranking quality and processing time of a given appearance descriptor. Our approach focuses on practical application scenarios characterized by a very large template gallery to be ranked in response to a query by a human operator, and/or by a similarity measure exhibiting a high processing time. A first empirical evidence on the benchmark VIPeR data set, using four different descriptors, showed that the proposed approach is capable of reducing processing time with respect to the original descriptor, attaining at the same time almost the same ranking quality. The observed reduction in processing time is not high, although it can be improved by suitably tuning the parameters of the descriptor at hand. In practice, it could also be difficult to accurately estimate the corresponding optimal values of the number of templates to be ranked by each stage (but the first one), as they depend on the size of template

gallery. We are currently investigating design criteria focused instead on strict requirements on processing time, for application scenarios where a reduction in ranking quality is acceptable.

References

1. Bedagkar-Gala, A., Shah, S.K.: A survey of approaches and trends in person re-identification. *Image Vis. Comput.* **32**(4), 270–286 (2014)
2. Dutra, C.R., Schwartz, W.R., Souza, T., Alves, R., Oliveira, L.: Re-identifying people based on indexing structure and manifold appearance modeling. In: SIBGRAP-Conférence on Graphics, Patterns and Images, pp. 218–225 (2013)
3. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: International Conference on Computer Vision and Pattern Recognition, pp. 2360–2367 (2010)
4. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008. LNCS, vol. 5302, pp. 262–275. Springer, Heidelberg (2008). doi:[10.1007/978-3-540-88682-2_21](https://doi.org/10.1007/978-3-540-88682-2_21)
5. Hirzer, M., Beleznaï, C., Roth, P.M., Bischof, H.: Person re-identification by descriptive and discriminative classification. In: Heyden, A., Kahl, F. (eds.) SCIA 2011. LNCS, vol. 6688, pp. 91–102. Springer, Heidelberg (2011). doi:[10.1007/978-3-642-21227-7_9](https://doi.org/10.1007/978-3-642-21227-7_9)
6. Hirzer, M., Roth, P.M., Bischof, H.: Person re-identification by efficient impostor-based metric learning. In: International Conference on Advanced Video and Signal-Based Surveillance, pp. 203–208 (2012)
7. Huang, S., Gu, Y., Yang, J., Shi, P.: Reranking of person re-identification by manifold-based approach. In: International Conference on Image Processing, pp. 4253–4257 (2015)
8. Kaynak, C., Alpaydin, E.: Multistage cascading of multiple classifiers: one man’s noise is another man’s data. In: International Conference on Machine Learning, pp. 455–462 (2000)
9. Khedher, M.I., El Yacoubi, M.A.: Two-stage filtering scheme for sparse representation based interest point matching for person re-identification. In: Battiato, S., Blanc-Talon, J., Gallo, G., Philips, W., Popescu, D., Scheunders, P. (eds.) ACIVS 2015. LNCS, vol. 9386, pp. 345–356. Springer, Heidelberg (2015). doi:[10.1007/978-3-319-25903-1_30](https://doi.org/10.1007/978-3-319-25903-1_30)
10. Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning. In: International Conference on Computer Vision and Pattern Recognition, pp. 2197–2206 (2015)
11. Liu, C., Loy, C.C., Gong, S., Wang, G.: POP: person re-identification post-rank optimisation. In: International Conference on Computer Vision, pp. 441–448 (2013)
12. Ma, B., Su, Y., Jurie, F.: Covariance descriptor based on bio-inspired features for person re-identification and face verification. *Image Vis. Comput.* **32**(6), 379–390 (2014)
13. Metternich, M.J., Worring, M.: Track based relevance feedback for tracing persons in surveillance videos. *Comput. Vis. Image Underst.* **117**(3), 229–237 (2013)
14. Pudil, P., Novovicova, J., Blaha, S., Kittler, J.: Multistage pattern recognition with reject option. In: International Conference on Pattern Recognition, vol. II, pp. 92–95 (1992)

15. Satta, R., Fumera, G., Roli, F.: Fast person re-identification based on dissimilarity representations. *Pattern Recogn. Lett.* **33**(14), 1838–1848 (2012)
16. Satta, R., Fumera, G., Roli, F., Cristani, M., Murino, V.: A multiple component matching framework for person re-identification. In: Maino, G., Foresti, G.L. (eds.) *ICIAP 2011*. LNCS, vol. 6979, pp. 140–149. Springer, Heidelberg (2011). doi:[10.1007/978-3-642-24088-1_15](https://doi.org/10.1007/978-3-642-24088-1_15)
17. Sperduti, A.: Theoretical and experimental analysis of a two-stage system for classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 893–904 (2002)
18. Trapeznikov, K., Saligrama, V., Castañón, D.: Multi-stage classifier design. *Mach. Learn.* **92**(2–3), 479–502 (2013)
19. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: *International Conference on Computer Vision and Pattern Recognition*, vol. 1, p. 511 (2001)
20. Wang, Z., Hu, R., Liang, C., Leng, Q., Sun, K.: Region-based interactive ranking optimization for person re-identification. In: Ooi, W.T., Snoek, C.G.M., Tan, H.K., Ho, C.-K., Huet, B., Ngo, C.-W. (eds.) *PCM 2014*. LNCS, vol. 8879, pp. 1–10. Springer, Heidelberg (2014). doi:[10.1007/978-3-319-13168-9_1](https://doi.org/10.1007/978-3-319-13168-9_1)