Correlation based Identity Filter: An Efficient Framework For Person Search

Wei-Hong Li[1], Yafang Mao[2], Ancong Wu[1] and Wei-Shi Zheng[2]

[1] School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou, China

[2] School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China

liweih3, maoyf5, wuancong@mail2.sysu.edu.cn and zhwshi@mail.sysu.edu.cn

Abstract. Person search, which addresses the problem of re-identifying a specific query person in whole candidate images without bounding boxes in real-world scenarios, is a new topic in computer vision with many meaningful applications and has attracted much attention. However, it is inherently challenging because the annotations of pedestrian bounding boxes are unavailable and we have to identify/find the target person from the whole gallery images. The existence of many visually similar people and dramatic appearance changes of the same person arising from the great cross-camera variation such as illumination, viewpoint, occlusions and background clutter also leads to the failure of searching a query person. In this work, we designed a Correlation based Identity Filter (CIF) framework for re-identifying the query person directly from the whole image with high efficiency. A regression model is learnt for obtaining a correlation filter/template for a given query person, which can help to alleviate the accumulated error caused by doing detection and re-identification separately. The filter is light and can be obtained and applied to search the query person with high speed with the utilization of Block-Circulant Decomposition (BCD) and Discrete Fourier Transform (DFT) techniques. Extensive experiments illustrate that our method has the important practical benefit of searching a specific person with a light weight and high efficiency and achieves better accuracy than doing detection and re-identification separately.

Keywords: Person Search, Correlation based Identity Filter, Regression, Pedestrian Detection, Person Re-identification.

1 Introduction

Person re-identification (re-id) aims to re-identify the same person across disjoint camera views in multi-camera system. It has drawn intensive attention in the computer vision society in recent decades with many important applications in video surveillance and multimedia, such as criminals detection and cross-camera tracking. However, person re-id by visual matching is particularly challenging because of the existence of many visually similar people and appearance variances of the same people resulting from great cross-camera variation such as viewpoints, poses, illumination, occlusions, resolutions and background clutter.

In current person re-id literature, numerous benchmarks and algorithms have been proposed in recent years and the performance on these benchmarks have been improved substantially. However, in the development of a person re-id systems for real-world applications, a challenge remained unsolved. The person re-id system with practice applications mainly consists of two components, including pedestrian detection and person re-id. Bulk of existent person re-id benchmarks and models only focus on the matching task with assumption that all persons in images have been perfectly detected and cropped manually. More sepecifically, these works proposed models to re-identify a query person by matching it with cropped pedestrians in the gallery instead of seaching from the whole image. In real-world scenarios, perfectly annotated pedestrian bounding boxes are unavailable in surveillance system. To manually annotate pedestrian is extremely costly and impractical, and existent pedestrian detectors inevitably produce false alarms, misdetections, and misalignments which would compromise the person re-id performance. This makes it unable to directly apply current person re-id algorithms to real-world tasks.

In order to overcome these challenges, developing a method that can jointly detect the pedestrian and match the query person, known as person search, is a new branch of person re-id and has been investigated by a few work. Existent person search methods are dominated by end-to-end deep CNN networks [18,16] which consist of two parts: 1) generating candidate pedestrain boxes in images by region proposal network and extracting features of these boxes by ROI pooling; 2) matching the query image and these boxes by calculating pairwise euclidean distances.

In contrast to existent person re-id methods, person search algorithms have to jointly detect a pedestrian and match the query person. Some inevitable false alarms, misdetections and misalignments would harm the person search performance. Therefore, in this work, we reformulate the person search as a regression task so that the model which is learnt for obtaining a correlation filter/template for a given query person, can help to alleviate the accumulated error caused by doing detection and re-identication separately. And we can develop a Correlation based Identity Filter (CIF) framework that can efficiently re-identify the query person from the whole gallery image. The regression score represented as a confidence map is utilized to measure the similariy between a candidate and the query person, and to reflect the potential of a candidate being a pedestrian. This is also unique as compared with these person search methods.

More specifically, for a given query person, a great amount of training samples, which contains samples of the query person, other identities and background, are collected by dense sampling. We further generate labels for these samples so that the regression model we learn on such training data is capable of re-identify the query person over a whole image and distinguishing a pedestrian from the background. In practice, the regression weight can be fastly solved by adopting the Block-Circulant Decomposition technique, which is equivalent to obtaining a correlation filter/template for a given query person. In testing procedure, when a gallery image is presented, we extract samples with various scales at all locations. And we evaluate the potential of a candidate matched with the query person and meanwhile being a pedestrian by the learnt regression weight, which can be seen as a spatial filtering operation over the gallery images.

Since the regression weight is light and the Discrete Fourier Transform technique can be employed to accelerate the spatial filtering operation, our CIF can fastly search the query person in the whole gallery images.

In summary, we make following contributions. 1) We develop a Correlation based Identity Filter (CIF) framework for person search. In our development, we consider reformulating the person search which needs to jointly handle two tasks (i.e., pedestrian detection and person re-id) as a regression task. For an image of the given query person, the regression model we learn can be used for matching the query person with the candidates at all locations and meanwhile judging whether the candidates is a pedestrian or not. 2) Since we extract the training and testing samples by the cycle shift, the learning process is equivalent to obtaining a correlation filter and then can be accelerated by the Block-Circulant Decomposition technique. This makes our framework flexibly and fastly learn the correlation based identity filter for the given query person. Due to the fact that the regression weight is light and we employ the Discrete Fourier Transform technique to accelerate the spatial filtering operation, our CIF can fastly search the query person in the whole gallery images even though we have to match numerous candidate samples with the query person.

2 Related Work

Person Re-identification. A large number of models for person re-identification problem have been proposed in literature, including feature learning [14,19,9], distance metric learning [20,6,8,12,9,11], and deep learning [7,1,17]. These methods focus on improving matching accuracy with assumption that the pedestrian images are well detected and cropped.

Person Search. Person search is a new topic in computer vision and has been investigated by only a few works [16,18]. Existent works [16,18] mainly proposed a kind of end-to-end deep CNN networks which are able to jointly handle two tasks: 1) generating candidate pedestrain boxes in images by region proposal network and extracting features of these boxes by ROI pooling; 2) matching the query image and these boxes by calculating pairwise euclidean distances. Instead of focusing on designing a good CNN network for better feature representation for person search, in this work, we consider formulating a shallow framework that detects and matches the query person on the whole gallery images jointly by regression. In particular, we proposed to learn a correlation based identity filter that is capable of re-identifying the query person on a whole image and distinguishing a pedestrian from the background, which differs from other person search methods. Although our framework has to be retrained when a new query is given, the weight of our CIF is light and can be learnt fastly with the use of Block-circulant Decomposition and Discrete Fourier Transform techniques.

Correlation Filter. Recently, Corrlation Filter have shown its effectiveness and efficiency in object tracking [4] and pedestrian detection [3], and has attracted increasing attention. J. F. Henriques et al. [4] proposed the kernel correlation filter for high-speed object tracking, which is the fastest object tracker so far. J. Valmadre et al. [15] developed an end-to-end trainable correlation filter network for object tracking which obtained improved performance on several tracking benchmarks. J. F. Henriques et al.

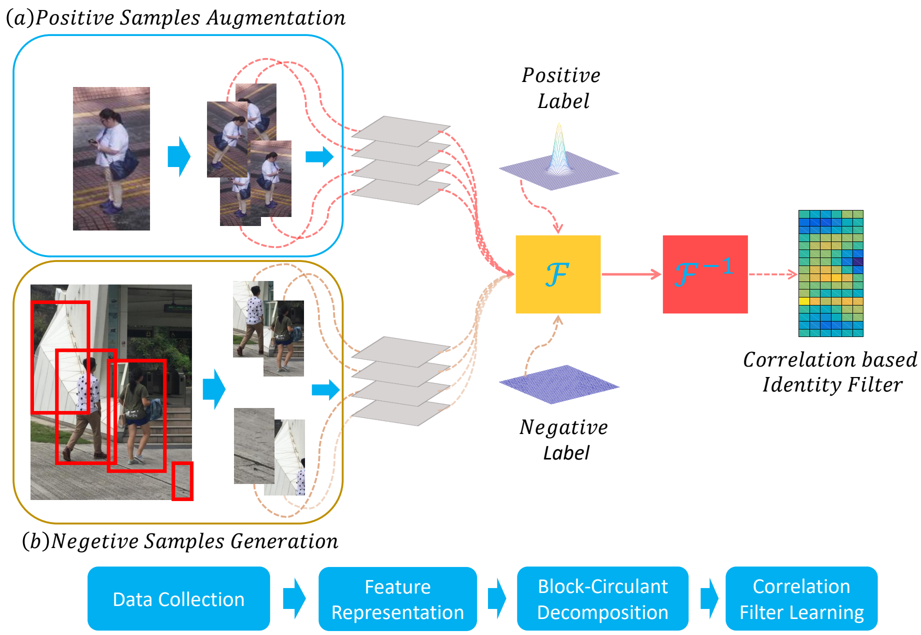


Fig. 1. An illustration of the training procedure of our CIF. In this figure, the yellow rectangle represents the Discrete Fourier transform (DFT) and the red one means the Inverse Discrete Fourier transform (IDFT). (Better viewed in color).

[3] proposed to adapt the correlation filter for efficient pedestrian beyond hard negative mining. In contrast to these work, we formulate the person search that is to re-identify the query person on the whole gallery images as a regression task, and then correlation filter, the effective and efficient technique, can be well adopted for person search.

3 Approach

In this work, we present a Correlation based Identity Filter (CIF), a new and efficient framework, for person search. Given an image of one query person, we first collect a large training set which consists a lots of training samples and their labels. We further learn a regression model over this training set, which is equivalent to learning a correlation based identity filter which encodes the identity information of the query person and the invariant information of the pedestrian. Finally, the regression model is employed to re-identify the query person over the whole gallery images, which can be seen as a spatial filtering operation by the identity filter on the images. An illustration of the training procedure and the one of the testing pipeline is shown in Figure 1 and Figure 2, respectively.

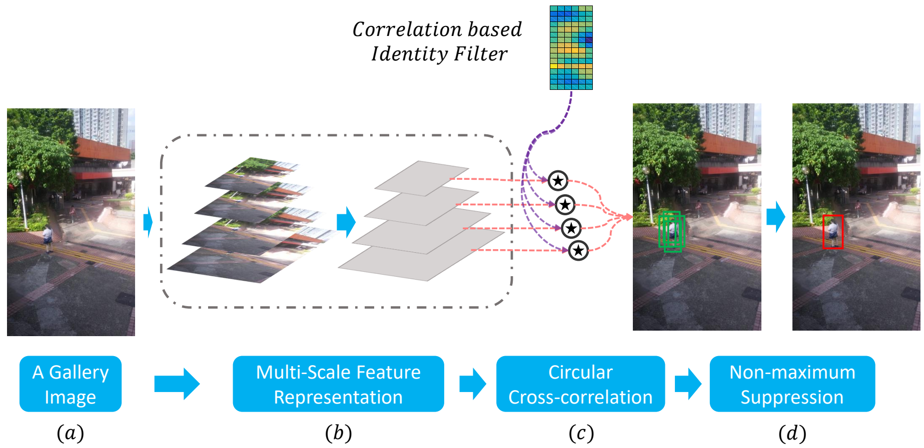


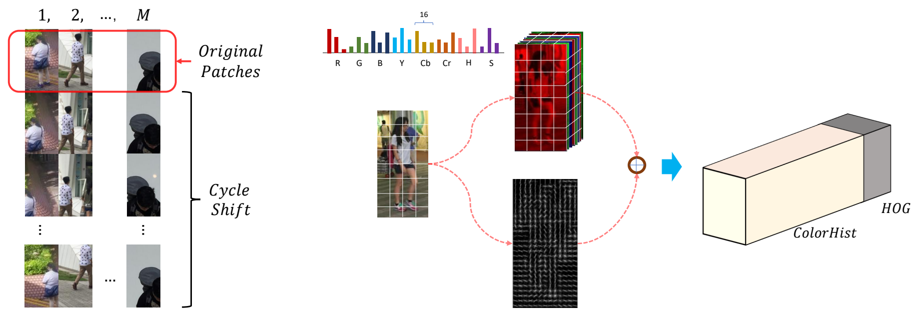
Fig. 2. An illustration of the testing procedure of our CIF. In this figure, the $\star$ denotes the circular cross-correlation.

3.1 Cropping Original Patches and Dense Sampling

Instead of utilizing the patch inside the annotated bounding box of all pedestrians as the training samples, in this work, we consider cropping dense samples from the surrounding area of each person. Since dense sampling results in a vast quantity of hard samples, it can help to alleviate the drifting problem in pedestrian detection [3] and enhance the ability of discriminate different identities.

Cropping Original Patches. First of all, instead of randomly extracting training samples over the whole training images, we crop an image patch which centers on the center of the bounding box of the person. And the sizes of these patches are larger than the sizes of their bounding box. We further rotate the patch of the query person with different angles so as to alleviate the impact of rare images of the query person. These image patches containing persons are utilized to captured the discriminative information between the query person and other identities. Moreover, in order to distinguish the difference between a pedestrian and the background clutter, as shown in Figure 1 (b), some background patches are also cropped randomly. For convenience of description, we denote each of cropped patches as $\mathbf{x}_i \in \mathcal{R}^d (i = 1, 2, \cdots, M)$ and called original patches.

Dense Sampling by Cycle Shift. From each of original patches, we extract L samples by cycle shift (Figure 3(a)) which can be efficiently handled by Discrete Fourier Transform (DFT) techniques proposed in [4]. For notational simplicity, the samples generate from the patch $\mathbf{x}_i$ are denoted by a sample matrix $\tilde{\mathbf{X}}_i = \mathcal{F}^{-1}(diag(\mathcal{F}(\mathbf{x}_i))) \in \mathcal{R}^{d \times L}$, where $\mathcal{F}(\cdot)$ denotes the forward DFT, $\mathcal{F}^{-1}(\cdot)$ is the Inverse DFT and $diag(\cdot)$ is a diagonalization operator. And then we generate labels $\tilde{\mathbf{Y}}_i \in \mathcal{R}^L$, which is a column vector, for these samples. In particular, the samples where the query person is at the center of the sample are annotated by 1 and others are labelled by 0. Therefore, we collect the shifted samples of all cropped patches as a large scale training sample set and denote them as a large sample matrix $\tilde{\mathbf{X}} = [\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, ..., \tilde{\mathbf{X}}_M] \in \mathcal{R}^{d \times N}$, where each column



(a) Sampling via cycle shift. (b) Feature Representation

Fig. 3. Sampling and feature representation. (Better viewed in color).

$\tilde{\mathbf{x}}_i$ is a $d-$dimensional vector representing a sample that obtained by cycle shift and $N = M \times L$. And a label vector $\tilde{\mathbf{Y}} = [\tilde{\mathbf{Y}}_1^T, \tilde{\mathbf{Y}}_2^T, \cdots, \tilde{\mathbf{Y}}_M^T]^T \in \mathcal{R}^N$, where each element $\tilde{\mathbf{y}}_i$ that is a label corresponding to $\mathbf{x}_i$, is also generated.

3.2 Learning Identity Filter

Further, we consider the person search as a regression task, and then we fomulate a function $f(\tilde{\mathbf{x}}) = \mathbf{w}^T\tilde{\mathbf{x}}$ to measure the potential of a candidate or a sample can be matched with the query person and meanwhile is a pedestrian. In other words, for each query person, we aim at learning a sole identity filter $\mathbf{w}$ that can not only distinguish a pedestrian from an image but also identify whether a candidate is the query person or not.

Concretely, based on the previous collected data, we seek a solution $\mathbf{w}$ that minimizes the squared error over samples $\tilde{\mathbf{x}}_i$ and their generated regression label $\tilde{\mathbf{y}}_i$,

$$\min_{\mathbf{w}} \sum_{i=1}^{N} L(f(\tilde{\mathbf{x}}_i), \tilde{\mathbf{y}}_i) + \lambda ||\mathbf{w}||^2, \tag{1}$$

where $L(f(\tilde{\mathbf{x}}_i), \tilde{\mathbf{y}}_i)$ is a loss function. In this work, we exploit the squared epsilon-insensitive loss:

$$L(f(\tilde{\mathbf{x}}_i, \tilde{\mathbf{y}}_i)) = max(0, |\mathbf{w}^T\tilde{\mathbf{x}}_i - \tilde{\mathbf{y}}_i| - \epsilon)^2. \tag{2}$$

The above formulation Eq. 1 then can be expressed in its dual form:

$$\min_{\boldsymbol{\alpha}} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \boldsymbol{\alpha}_i \boldsymbol{\alpha}_j \tilde{\mathbf{x}}_i^T \tilde{\mathbf{x}}_j + \sum_{i=1}^{N} (\frac{1}{2}\lambda \boldsymbol{\alpha}_i^2 - \boldsymbol{\alpha}_i \tilde{\mathbf{y}}_i + \epsilon |\boldsymbol{\alpha}_i|). \tag{3}$$

We can rewrite the Eq. 3 as:

$$\min_{\boldsymbol{\alpha}} \frac{1}{2} \boldsymbol{\alpha}^T \mathbf{G} \boldsymbol{\alpha} + \frac{\lambda}{2} \boldsymbol{\alpha}^T \boldsymbol{\alpha} - \boldsymbol{\alpha}^T \tilde{\mathbf{Y}} + \epsilon |\boldsymbol{\alpha}|, \tag{4}$$

where $\mathbf{G} = \tilde{\mathbf{X}}^T\tilde{\mathbf{X}}$.

The relationship between the solution of Eq. 1 and the one of Eq. 4 is $\mathbf{w} = \tilde{\mathbf{X}}\boldsymbol{\alpha}$. Although the covariance matrix $\mathbf{G}$ is extremely large, in this work, we exploit the Block-circulant Decomposition proposed in [3] to fastly obtained the matrix $\mathbf{G}$ so that $\mathbf{w}$ can be learnt with high efficiency.

3.3 Search The Query Person

When a gallery image $\mathbf{z}$ is presented, we apply the previous learnt correlation based identity filter on the whole image to search the query person. First of all, we construct a pyramid of multi-scale image representation (Figure 2(b)) to account for large scale change of person appearance such as the scale of the person. Then the learnt correlation based identity filter of the query person is applied on the feature of one layer of the pyramid and produce confidence map where each element reflects the similarity of the query person and a candidate and the propability of a candidate is a pedestrian by circular cross-correlation.

More specifically, we denote the testing image as $\mathbf{z}$ and we can generate a large amount of candidate by dense sampling and denote these samples as a sample $\tilde{\mathbf{Z}} \in \mathcal{R}^{d\times K}$, where each column $\tilde{\mathbf{z}}_i$ is a candidate. Therefore, the potential that a candidate is the query person and meanwhile is a pedestrian can be calculated by:

$$f(\tilde{\mathbf{Z}}) = \mathbf{w}^T\tilde{\mathbf{Z}}. \tag{5}$$

Fortunately, although dense samples are extracted, the above equation can be calculated fastly by circular cross-correlation:

$$f(\tilde{\mathbf{Z}}) = \mathbf{w}^T \star \mathbf{z} = \mathcal{F}^{-1}(diag(\mathcal{F}(\mathbf{w})^*) \odot diag(\mathcal{F}(\mathbf{z}))), \tag{6}$$

where $\odot$ is a element-wise operator, $\star$ denotes the circular cross-correlation and $\mathcal{F}(\mathbf{w})^*$ is a complex-conjugate.

Intuitively, evaluating $f(\tilde{\mathbf{Z}})$ at all locations can be seen as a spatial filtering operation over the image patch $\mathbf{z}$. Finally, the candidate with highest confidence map score is selected as a pedestrian that is most likely to be the query person.

4 Experiment

In this section, extensive experiments are conducted on CUHK-SYSU person search dataset [18] for evaluation of our proposed method and some baseline.

4.1 Dataset and Evaluation Protocols

We evaluate our method and other baselines on CUHK-SYSU person search dataset proposed by [18]. This dataset contains 18, 184 images where 96, 143 pedestrians bounding boxes are annotated. The test set consists of 6,978 images of 2900 persons, and for each of the test person, one image was selected as the query. Besides, different queries have different galleries, and a set of protocols were defined in [18] with the gallery size ranging from 50 to 4000. we followed the standard evaluation settings [18] for performing a fair comparison with existent methods. The mean Average Precision (mAP) and Rank$-k$ matching rate of different methods are reported to show the effectness and efficiency of our methods.

4.2 Implementation Details

Feature Representation. Color Histogram (ColorHist) is a kind of widely used feature in person re-id, while Histogram of Oriented Gradient (HOG) is one of best hand-craft feature for pedestrian detection. Since we aim at learning a filter that is able to detect a person that is matched with the query person, in this work, we proposed to represented each training image patches and the gallery images by the combination feature of ColorHist and HOG (Figure 3(b)). For a given data, the feature is then represented as a $s_1 \times s_2 \times 159$ cube, where the value of s_1 and s_2 rely on the size of the image.

Table 1. Comparison results of different methods .

Gallery Size	Detector	Method	mAP (%)	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)	Rank-20 (%)
50	SSD	Euclidean	39.33	40.83	56.76	63.86	71.03
		KISSME	36.71	38.07	52.55	59.45	66.03
		XQDA	36.99	38.45	53.07	59.86	66.66
	GT	Euclidean	48.46	51.24	65.24	70.76	77.59
		KISSME	46.95	49.76	63.14	69.48	75.24
		XQDA	47.43	50.48	63.31	69.55	75.69
		CIF (Ours)	45.41	52.38	61.21	63.55	65.31
100	SSD	Euclidean	35.01	37.07	51.55	57.76	64.66
		KISSME	32.37	34.62	47.76	52.83	59.69
		XQDA	32.75	35.17	48.07	53.34	60.17
	GT	Euclidean	43.69	47.24	59.76	64.59	70.69
		KISSME	42.33	46.28	57.28	63.34	69.59
		XQDA	42.58	46.38	57.59	63.55	69.76
		CIF (Ours)	43.35	50.48	59.07	61.31	63.59

Compared methods. Since the practice person search algorithms have to process two tasks: 1) pedestrian detection; 2) person re-id. In this work, we modified SSD [13], a fast state-of-the-art object detector, for pedestian detection and train it on Caltech Pedestrian Detection Benchmark [2]. The modified SSD obtain a comparable performance on Caltech Benchmark and is adopted as a baseline detector. Besides, the ground-truth annotation pedestrian bounding boxes are also utilized as a ground-truth detector, called GT. For evaluation of our method, we compared three different person re-id methods: 1) Euclidean which match the query person using the Euclidean distance; 2) KISSME

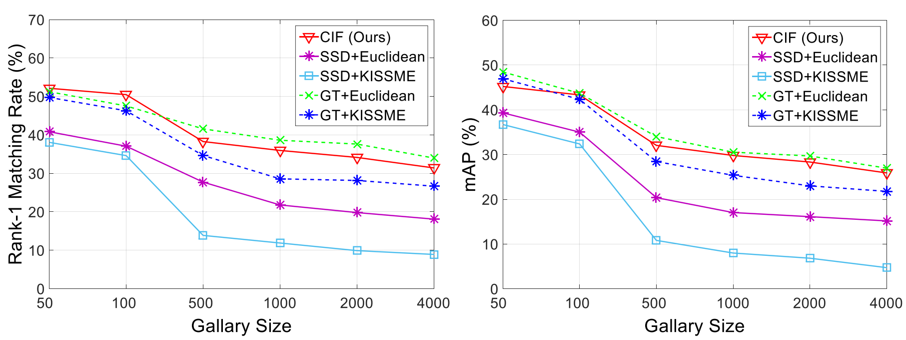
[5]; 3) XQDA [10]. Therefore, 6 methods are formed: 1) SSD + Euclidean; 2) SSD + KISSME; 3) SSD + XQDA; 4) GT + Euclidean; 5) GT + KISSME; 6) GT + XQDA. All experiments were implemented on an Intel E5-2650 v3 2.30GHz CPU with 256 GB RAM.

4.3 Results

Compared with other methods. We report comparison results of our CIF and other methods in Table. 1. The results shows that our CIF surpass the three methods whose detector are SSD, and can ourperform or approximate the performance of those methods which based on the ground-truth annotation bounding boxes. Firstly, this indicates that our CIF is able to detect a person that is matched to the query person with high probability. Compared to the methods which match the query person with candidates provided by ground-truth bounding boxes, our proposed method is able to distinguish a pedestrian from the background.

Table 2. Results of our CIF on CUHK-SYSU person search dataset.

Gallary Size	mAP (%)	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)	Rank-20 (%)
50	45.41	52.38	61.21	63.55	65.31
100	43.35	50.48	59.07	61.31	63.59
500	32.05	38.28	46.48	49.21	51.79
1000	29.80	35.97	43.83	46.90	49.24
2000	28.33	34.15	41.70	44.48	47.22
4000	25.90	31.41	40.50	42.57	43.56



(a) (b)

Fig. 4. Comparison of different methods with various gallery size (Better viewed in color).

Evaluation with various gallery sizes. We evaluate our proposed method and other methods on the CUHK-SYSU person search dataset with different gallery size and

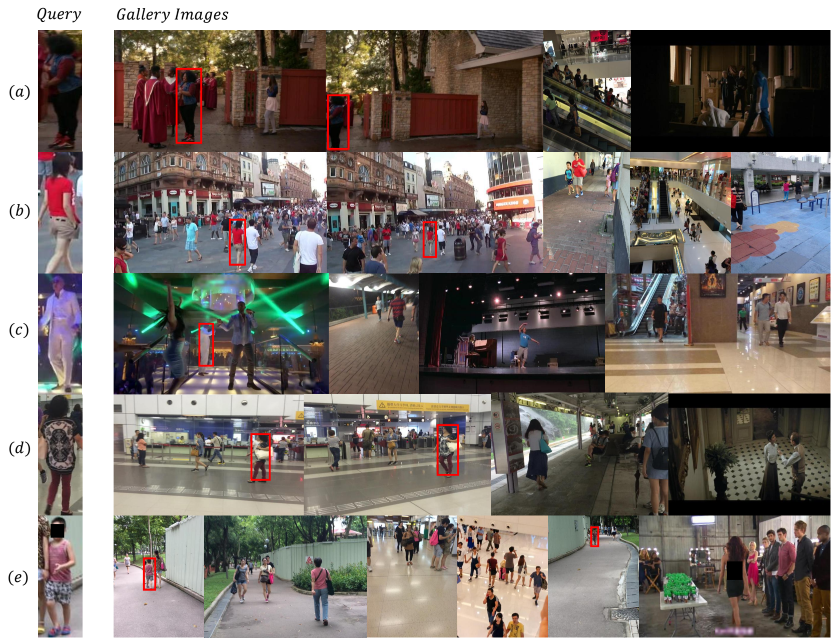


Fig. 5. Some visual result of our CIF. For each query person, one query image and its rank$-k$ images are shown in each row.

the results are shown in Table. 2 and Figure 4. It is evidence that the performance of all methods including our proposed method decreases when the gallery size increases. However, CIF also outperform methods which match the query person with the candidates detected by SSD and can surpass or approximate the ones based on ground-truth annotation bounding boxes.

Visual Results. We also shows some visual experimental results in Figure 5. It is evidence that our proposed method is capable of re-identifying the query person from the gallery image although some inherent challenges such as the change of appearance (Figure 5(a), (d)), background clutter (Figure 5(b), (c)) and the visual similarity of different persons (Figure 5(e)) exist.

Searching speed.

For evaluating the speed of person search when applying all methods, we measured the mean frames per second (Mean FPS). All methods are conduct on the original gallery images without resize operation. In particular, our CIF can runs at 5.3 frames per second with CPU while other methods (i.e., SSD + Euclidean, SSD + KISSME and SSD + XQDA) operate at only 0.17 FPS. Besides, CIF takes only 30 seconds to learn on training data for each query person before being applied on test images for searching for the target people. This results illustrate that our CIF can fastly find the query person from a whole image.

5 Conclusion

The main contribution is to first cast the person search as a regression task and develop a Correlation based Identity Filter framework for fast person search. A regression model is learnt for obtaining a correlation filter/template for a given query person, which can help to alleviate the accumulated error caused by doing detection and re-identification separately. Finally, the regression model is employed to re-identify the query person over the whole gallery images with high efficiency, which can be seen as a spatial filtering operation by the identity filter on the images. This all forms the proposed CIF. Extensive evaluations reported CUHK-SYSU person search dataset have shown the efficient performance of CIF.

References

1. Ahmed, E., Jones, M., Marks, T.K.: An improved deep learning architecture for person re-identification. In: CVPR (2015)
2. Dollar, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: An evaluation of the state of the art. TPAMI 34(4), 743–761 (2012)
3. Henriques, J.F., Carreira, J., Caseiro, R., Batista, J.: Beyond hard negative mining: Efficient detector learning via block-circulant decomposition. In: ICCV (2013)
4. Henriques, J.F., Caseiro, R., Martins, P., Batista, J.: High-speed tracking with kernelized correlation filters. TPAMI 37(3), 583–596 (2015)
5. Koestinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. In: CVPR (2012)
6. Köstinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. In: CVPR (2012)
7. Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: Deep filter pairing neural network for person re-identification. In: CVPR (2014)
8. Li, Z., Chang, S., Liang, F., Huang, T.S., Cao, L., Smith, J.R.: Learning locally-adaptive decision functions for person verification. In: CVPR (2013)
9. Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning. In: CVPR (2015)
10. Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning. In: CVPR (2015)
11. Liao, S., Li, S.Z.: Efficient psd constrained asymmetric metric learning for person re-identification. In: ICCV (2015)
12. Lisanti, G., Masi, I., Del Bimbo, A.: Matching people across camera views using kernel canonical correlation analysis. In: ICDSC (2014)
13. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: Ssd: Single shot multibox detector. In: ECCV. pp. 21–37. Springer (2016)
14. Ma, B., Su, Y., Jurie, F.: Covariance descriptor based on bio-inspired features for person re-identification and face verification. Image and Vision Computing (2014)
15. Valmadre, J., Bertinetto, L., Henriques, J.F., Vedaldi, A., Torr, P.H.: End-to-end representation learning for correlation filter based tracking. CVPR (2017)
16. Xiao, J., Xie, Y., Tillo, T., Huang, K., Wei, Y., Feng, J.: Ian: The individual aggregation network for person search. arXiv preprint arXiv:1705.05552 (2017)
17. Xiao, T., Li, H., Ouyang, W., Wang, X.: Learning deep feature representations with domain guided dropout for person re-identification. In: CVPR (2016)

18. Xiao, T., Li, S., Wang, B., Lin, L., Wang, X.: Joint detection and identification feature learning for person search. arXiv preprint arXiv:1604.01850 (2017)
19. Yang, Y., Yang, J., Yan, J., Liao, S., Yi, D., Li, S.Z.: Salient color names for person re-identification. In: ECCV (2014)
20. Zheng, W.S., Gong, S., Xiang, T.: Person re-identification by probabilistic relative distance comparison. In: CVPR (2011)