Context-based Matching Refinement for Person Search

Byeong-Ju Han*, Jae-Won Yang*, Oggyu Lee*, and Jae-Young Sim*†
* Department of Electrical Engineering, UNIST, Ulsan, Korea
† Graduate School of Artificial Intelligence, UNIST, Ulsan, Korea
E-mails:{bjhan, jwyang, oglee, jysim}@unist.ac.kr

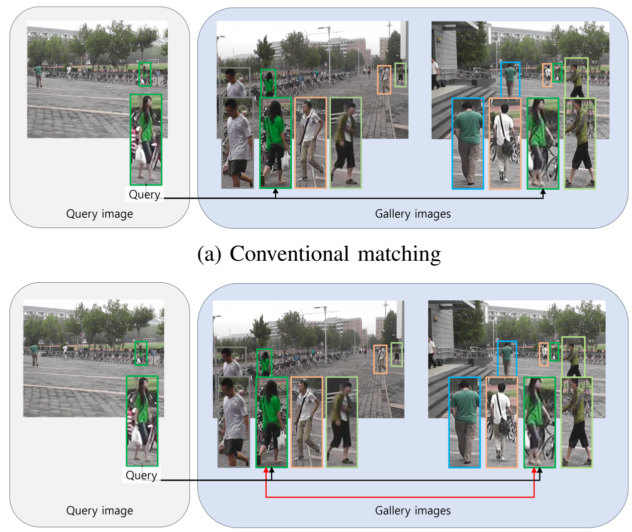
***Abstract*—Person search is the combined task of person detection and re-identification that aims to find target persons having the same identity to a given query person. Existing methods of person search mainly focus to extract useful features of persons, but do not use the relationship among the persons in gallery images during the test phase. In this paper, we propose a context-based matching refinement scheme for person search by utilizing the contextual relationship of persons in gallery images. We compute the matching context such that a target person has a high context value with respect to a query person, when it has relatively high similarities to the positive samples of the query person. Experimental results show that the proposed method significantly outperforms the existing state-of-the-art person search methods on PRW dataset.**

I. Introduction

With the advent of diverse surveillance cameras, a large number of pedestrian images are being provided everyday. Finding target persons from a large number of pedestrian images is an important problem in many computer vision applications. In recent years, person search has drawn much attention that aims to find target persons in pedestrian images matching to a given query person. Person search is composed of two problems of person detection and person re-identification, where the person detection is the task to find the bounding boxes of persons in images, and person re-identification is the task to find the detected person images with the same identity to a query person. Therefore, compared to person detection and person re-identification tasks, person search is a relatively challenging task but have more impacts in diverse applications.

There have been many research works for person search, however, researchers have usually focused on how to extract useful features for persons during the training phase while often neglecting the context information in gallery images. Most of the existing person search methods simply obtain the matching results according to the similarities of person appearance features. Moreover, they do not embed the person appearance completely due to the lack of training data in a variety of human poses and camera view directions. Since the matching scores usually depend on the appearance features of persons, existing person search methods suffer from the outliers.

In this paper, we propose a matching refinement algorithm for person search in the test phase. We first investigate the context information in terms of the similarities among the



(a) Conventional matching

(b) Proposed matching

Fig. 1: Comparison between the conventional person matching and the proposed context-based matching.

candidate persons in gallery images. Then we refine the initial matching between a query and target persons by augmenting the initial feature similarity using the per-computed context information in gallery images. Fig. 1 compares the conventional matching process and the proposed one for person search. Whereas the conventional person search methods compare a query person to each person detected in gallery images independently, the proposed algorithm compare all of the detected persons in the query images and the gallery images to each other and exploit their relationship as context cues to improve the matching performance.

II. Related Works

A. End-to-End Methods

Deep neural networks for person search have been developed where the person detection and person re-identification tasks are jointly trained in an end-to-end manner. Xiao et al. [1] first introduced an online instance matching concept to effectively train person search network using labeled and unlabeled person identities together. Zheng et al. [10] performed diverse experiments using state-of-art person detection and recognition methods, and proposed a confidence weight similarity metric to incorporate the detection scores into the similarity measures of person identities. Chen et al. [2] utilized a hierarchical relationship between person detection and re-identification to reduce the conflict of the two tasks and guide the network to discriminate the features of persons from the background. Chen et al. [4] also decomposed person embeddings into the norm and angle to reduce the conflict of person detection and re-identification tasks. Tian et al. [3] introduced key-point detection and instance segmentation as a multi-task manner to jointly train the network to be more robust to occlusion and pose variation. Zhong et al. [5] alleviated the occlusion problem by refining the occluded bounding boxes and aligning the partial body parts for matching. Shi et al. [6] utilized unlabeled persons with incomplete annotation by assigning temporal pseudo labels to unlabeled persons with similar appearances to labeled identities. Munjal et al. [8] proposed a query-guided region proposal network where the query images are used to re-weight the features from gallery images.

B. Two-Step Methods

In two-step approach, person detection and re-identification networks are separately trained to avoid the conflicting property between two tasks. Chen et al. [12] predicted the mask map highlighting the person regions to give more attention to the features of persons while suppressing background features. Han et al. [14] refined the initially detected bounding boxes of persons according to a re-identification loss to include more informative regions. Dong et al. [13] used the query person images to guide the proposal network to filter out the persons not similar to the query person and provide reliable matching candidates for person re-identification. Wang et al. [15] solved the inconsistency problem between the person detection and re-identification by utilizing a mixed training set for the person re-identification which contains the person images cropped by not only the ground truth bounding boxes but also the detected bounding boxes.

C. Context-based Person Search Methods

Few existing methods successfully enhanced the person search performance by using the context information of images. Yan et al. [17] interpreted the relationship between two persons on a graph where the nodes and edges are defined by the detected persons and their similarities. Dai et al. [7] applied the uniqueness prior that all persons in the same image have different identities to assign pseudo labels to the unlabeled persons due to incomplete annotation for training. Li et al. [9] applied a bipartite matching algorithm to enforce the persons appeared in the same query image not to match the same person in the gallery according to the uniqueness prior during a test. Hantao et al. [16] designed a new similarity measure between a query person and a target person in the gallery which provides a low value when the target person has a high similarity with any neighboring person to the query person.

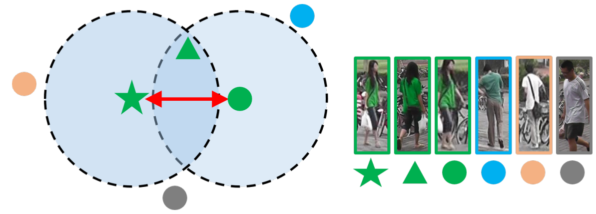


Fig. 2: Concept of matching context. The test query is denoted by the star, and the candidate persons in gallery images are denoted by the circles, respectively, in the feature space. The dotted circles represent a distance threshold for matching. The person in triangle is used as a context cue to match the green circle to the query person.

III. Proposed Methods

Whereas the existing methods that utilize the context prior observed in a single query image or a pair of query and gallery images, the proposed method estimates a new context priority embedded across the gallery images. We investigate the matching context in term of probabilistic similarities among the detected persons in the gallery images, and utilize them to refine the initial matching scores of the query person to the candidate persons in the gallery. Fig. 2 visualizes the concept of context-based matching refinement that the target persons yielding high similarities to the positive samples of a given query person are assigned increased similarity to the query person.

A. Matching Context in Gallery Images

We first compute the initial similarities $s(g_i, g_j)$ between the i-th and j-th persons in the gallery set $\mathcal{G}$ using the cosine similarity. For each person g_i, we generate a positive set $\mathcal{P}(g_i)$ including the persons of top-k similarities to g_i that are highly regarded to have the same identity to g_i. We empirically set as $k = 3$.

We then compute the matching context $\rho(g_i, g_j)$ between g_i and g_j, defined as

$$\rho(g_i, g_j) = \sum_{g_k \in \mathcal{P}(g_i)} s(g_j, g_k) \cdot p(g_k|g_i), \quad (1)$$

where

$$p(g_k|g_i) = \frac{\exp(s(g_i, g_k))}{\sum_{g_l \in \mathcal{P}(g_i)} \exp(s(g_i, g_l))}. \quad (2)$$

Note that $\rho(g_i, g_j)$ yields a high value up to 1 when g_j provides high similarities with the positive samples of g_i. It means that g_j is contextually correlated to g_i.

B. Matching Refinement

Based on the estimated matching context between the persons in the gallery, we compute an augmented similarity $\hat{s}(q_i, g_j)$ for a given query person q_i and $g_j \in \mathcal{G}$ as

$$\hat{s}(q_i, g_j) = s(q_i, g_j) + \alpha \sum_{g_k \in \mathcal{P}(q_i)} \rho(g_k, g_j) \cdot p(g_k|q_i). \quad (3)$$

		CUHK		PRW	
		mAP	Top-1	mAP	Top-1
NAE	Original	91.4	92.4	39.9	67.5
	Proposed	91.4	92.4	43.7	68.1
HOIM	Original	90.2	91.2	36.1	64.7
	Proposed	90.2	91.2	40.4	65.4
SeqNet	Original	93.7	94.5	43.5	68.5
	Proposed	93.8	94.3	49.2	71.4

TABLE I: Comparison of quantitative performance.

The contextual priority weight α is empirically set to 5.0. Note that the augmented similarity $\hat{s}(q_i, g_j)$ reflects the appearance similarity between q_i and g_j as well as the matching context priority between g_j and each positive sample $g_k \in \mathcal{P}(q_i)$ considered to have the same identity to the query q_i.

IV. Experimental Results

A. Datasets

We used CUHK-SYSU [1] and PRW [10] dataset to evaluate the performance of person search. CUHK-SYSU dataset consists of 18,184 images with total 8,432 different labeled person identities and 96,143 annotated bounding boxes. The images are captured at real streets and extracted from movies. This dataset uses various gallery sizes where the existing person search methods use the default gallery size of 100. PRW dataset consists of 11,816 video frames captured by fixed cameras located at six different positions. It has 932 different labeled person identities and 43,110 annotated bounding boxes. Both datasets are split into training and test sets, and the proposed method works in the test sets that include the query and gallery images. Fig. 3 visualizes the probability distributions of the number of persons with a same identity in CUHK-SYSU and PRW datasets, respectively. We see that PRW dataset has relatively larger numbers of persons in a same identity compared to CUHK-SYSU dataset.

B. Evaluation Metrics

The mean average precision (mAP) and top-k score have been used as the standard metrics to evaluate the quantitative performance of person search. mAP reflects both of the precision and recall scores of matching. Top-k score represents at least one of the estimated top k persons matching to the query is a true matching.

C. Performance Comparison

We compare the performance of the proposed person search method with that of the three recent state-of-the-arts methods: NAE [4], HOIM [2], and SeqNet [9]. As shown in Table I,

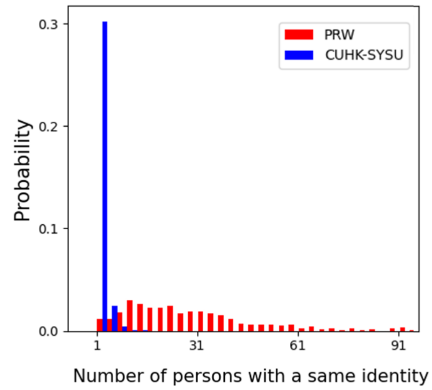
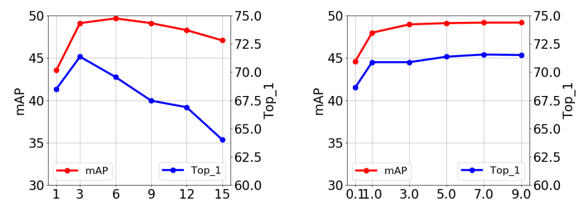


Fig. 3: Probability distributions of the number of persons having a same identity in CUHK-SYSU and PRW datasets.



(a) (b)

Fig. 4: Performance of person search with varying hyper-parameters: (a) k and (b) α.

the proposed method significantly improves the performance of the state-of-the-art methods on PRW dataset, however, no performance gain is observed on CUHK-SYSU dataset. As plotted in Fig. 3, CUHK-SYSU dataset yields relatively smaller numbers of persons with a same identity and therefore provides less context information compared to PRW dataset.

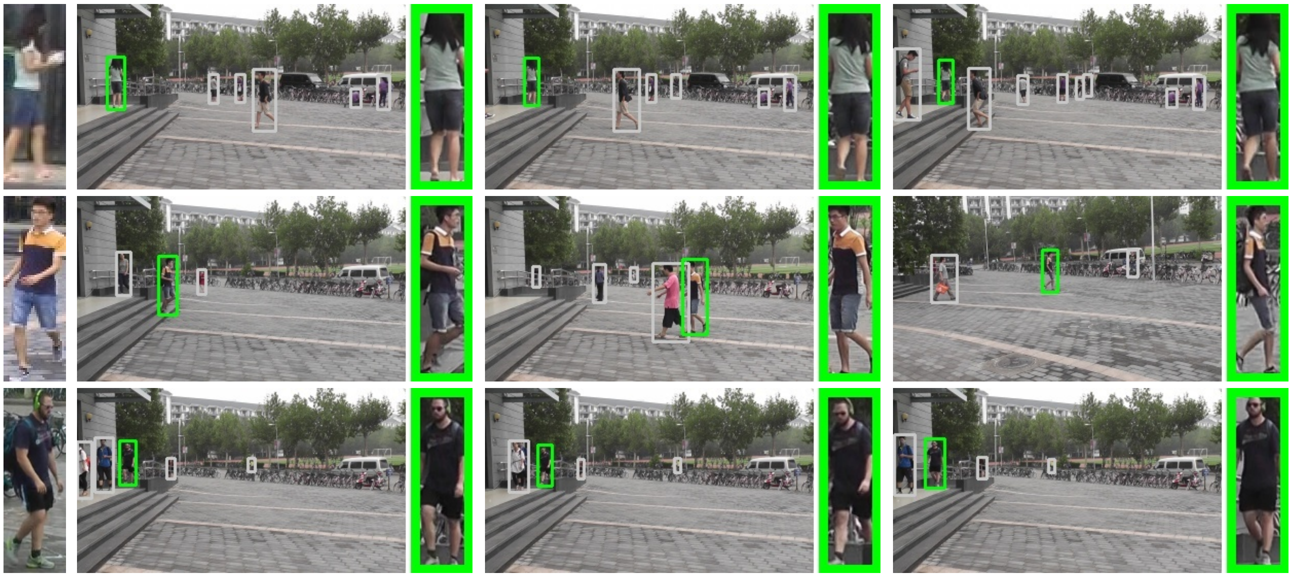
Fig. 5 visualizes the matching results of the proposed method on PRW test dataset, which shows that the proposed method successfully finds the persons in gallery images matched to the query persons.

D. Performance with Hyper-Parameters

We conduct the ablation study using the multi-view gallery of PRW dataset to select the two hyper-parameters of the proposed method: k and α.

The hyper-parameter k determines the initial positive set which mainly affects the matching context computation and the matching refinement. As shown in Fig. 4 (a), we have the best performance when using $k = 3$. The small values of k make the proposed method extract the matching context information from only too similar samples. In contrary, the high values of k make the proposed method include false negative samples to be considered as positive samples.

Fig 4 (b) visualizes the impact of the ratio of matching context weight α to update the similarity between the query



(a) (b) (c) (d)

Fig. 5: Qualitative results of person search on PRW testset. (a) Test queries. (b d) The results of top-3 best matching persons to the test queries, respectivelyi. The detected persons are localized by the bounding boxes where the true positives are highlighted in green.

person and detected gallery persons. The best performance is achieved when $\alpha = 5.0$.

V. Conclusion

We proposed a novel matching refinement algorithm for person search using context information in gallery images. We compared all the detected persons in the gallery images to each other and used their relationship as an effective context cue to augment the initially computed similarity between the query persons and target persons. Experimental results demonstrated that the proposed method is applicable to existing state-of-the-art person search methods and further improves the matching performance significantly in terms of mAP and top-1 on PRW dataset.

Acknowledgment

This work was supported in part by Institute for Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.20170006670021001, Information-Coordination Technique Enabling Augmented Reality with Mobile Objects), and in part by IITP grant funded by MSIT (No.20200013360011001, Artificial Intelligence graduate school support(UNIST)).

References

[1] T. Xiao, S. Li, B. Wang, L. Lin, X. Wang , Joint Detection and Identification Feature Learning for Person Search, *in Proc. IEEE CVPR*, July, 2017, pp.3376-3385.

[2] D. Chen, S. Zhang, W. Ouyang, J. Yang, B. SchieleJ, Hi, Hierarchical online instance matching for person search, *in Proc. AAAI*, April, 2020, pp.10518-10525

[3] K. Tian, H. Huang, Y. Ye, J. Lin, G. Huang, End-To-End Thorough Body Perception for Person Search, *in Proc. AAAI*, April, 2020, pp.12079-12086

[4] D. Chen, S. Zhang, J. Yang, B. Schiele, Norm-Aware Embedding for Efficient Person Search, *in Proc. IEEE CVPR*, June, 2020, pp.12615-12624.

[5] Y. Zhong, X. Wang, S. Zhang, Robust Partial Matching for Person Search in the Wild, *in Proc. IEEE CVPR*, June, 2020, pp.6826-6834.

[6] W. Shi, H. Liu, F. Meng, W. Huang, Instance Enhancing Loss: Deep Identity-Sensitive Feature Embedding for Person Search, *in Proc. IEEE ICIP*, Sept, 2018, pp.4018-4112.

[7] J. Dai, P. Zhang, H. Lu,H. Wang, Dynamic imposter based online instance matching for person search, *Pattern Recognition*, April, 2020, 100:107120.

[8] B. Munjal, S. Amin, F. Tombari, F. Galasso, Query-guided End-to-End Person Search, *in Proc. IEEE CVPR*, June, 2019, pp.6826-6834.

[9] Z. Li, D. Miao, Sequential End-to-end Network for Efficient Person Search, *in Proc. AAAI*, May, 2021, pp.2011-2019.

[10] L. Zheng, H. Zhang, S. Sun, M. Chanraker, Y. Yang, Q. Tian, Person Re-identification in the Wild, *in Proc. IEEE CVPR*, July, 2017, pp.3346-3355.

[11] S. Ren, K. He, R. Girshick,J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, *in Proc. Neurips*, 2015, vol. 28.

[12] D. Chen, S. Zhang, W. Ouyang,J. Yang,Y. Tai, Person Search via A Mask-Guided Two-Stream CNN Model, *in Proc. ECCV*, Sept, 2018, pp.734-750.

[13] W. Dong, Z. Zhang, C. Song, T. Tan, ,Instance Guided Proposal Network for Person Search, *in Proc. IEEE CVPR*, June, 2020, pp.2584-1594.

[14] C. Han et al, Re-ID Driven Localization Refinement for Person Search, *in Proc. IEEE ICCV*, Nov, 2019, pp.9813-9822.

[15] C. Wang, B. Ma, H.Chang, S. Shan, X. Chen, TCTS: A Task-Consistent Two-Stage Framework for Person Search, *in Proc. IEEE CVPR*, June, 2020, pp.11952-11961.

[16] Y. Hantao, X. Changsheng, Joint person objectness and repulsion for person search. *in proc. IEEE Transactions on Image Processing*, 2021. pp.30, 685-696.

[17] Y. Yan, Q. Zhang, B. Ni, W. Zhang, M. Xu, X. Yang, Learning context graph for person search, *in proc. IEEE CVPR*, June, 2019, pp. 2158-2167.