Pattern Recognition Letters 000 (2018) 1–9

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Bayesian query expansion for multi-camera person re-identification

Yutian Lin[a,*], Zhedong Zheng[a], Hong Zhang[b], Chenqiang Gao[c], Yi Yang[a]

[a] Center for Artificial Intelligence, University of Technology Sydney, Australia
[b] College of Computer Science and Technology, Wuhan University of Science and Technology, China
[c] Chongqing Key Laboratory of Signal and Information Processing, Chongqing University of Posts and Telecommunications, China

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Person re-identification
Re-rank
Query expansion

ABSTRACT

Person re-identification (re-ID) is challenging because pedestrians may exhibit distinct appearance under different cameras. Given a query image, previous methods usually output the person retrieval results directly, which may perform badly due to the limited information provided by the single query image. To mine more query information, we add an expansion step to post-process the initial ranking list. The intuition is that a true match in the gallery may be difficult to be found by the query alone, but it can be easily retrieved by other true matches in the initial ranking list. In this paper, we propose the Bayesian Query Expansion (BQE) method to generate a new query with information from the initial ranking list. The Bayesian model is used to predict true matches in the gallery. We apply pooling on the features of these "true matches" to get a single vector, *i.e.*, the expanded new query, with which the retrieval process is performed again to obtain the final results. We evaluate BQE with various feature extraction methods and distance metric learning methods on four large-scale re-ID datasets. We observe consistent improvement over all the baselines and report competitive performances compared with the state-of-the-art results.

1. Introduction

Multi-camera person re-identification (re-ID) has become a focus in the recent vision research [7,42,48,65]. Its goal is to find the aimed person in other non-overlapping camera views. It is critical in practical applications such as searching a suspect in surveillance videos captured by large camera networks in a city, or finding a lost kid by searching surveillance videos captured in a theme park. Person re-ID is challenging, partially due to the distinct characteristics of different cameras, such as view point, illuminations, *etc*. This paper proposes a re-ranking method that targets at reducing the impact of camera differences by introducing a post-processing method to improve the precision and recall of the pedestrian retrieval system.

Let us imagine that we want to find a missing child in a city with thousands of surveillance cameras. Given the image of this child, we first obtain a rank list by a conventional re-ID algorithm. It is possible that some true matches receive top ranks, and we find that they are captured by cameras in George Street. Other shots of the child are not similar to the initial query and are not retrieved. But it may be possible that the retrieved images in George Street are similar to some of those missed images in, for example, the Main Street. We can then fuse the original query with the retrieved images to do another retrieval process, so that hard positives can be found by the new query. This story serves as the main motivation of this paper: by expanding the query with other necessary information discovered in the initial ranking list, we are capable of finding the other challenging true matches in a multi-camera system.

In a different view from previous works on the feature or metric learning [6,43], we aim to improve the existing re-ID algorithms by query expansion. We investigate some large-scale multi-camera datasets, such as Market-1501 [57], and find that the same identity's appearance captured by different cameras usually undergoes large variance. On the contrary, the appearance of each pedestrian under the same camera is usually similar. As shown in Fig. 1, query expansion can help to improve the original re-ID performance. Inspired by the above considerations, we propose a Bayesian query expansion re-ranking algorithm to improve the performance of the existing re-ID methods. In a nutshell, after investigating the initial rank list, a new query is constructed based on the top returned images. Since the top-ranked images may be false matches, we develop a Bayesian framework to identify if a returned candidate is a true match; only the candidates with high probability scores will be used for query expansion. In this manner, the expanded query feature will contain more discriminative cues that may be absent

* Corresponding author.
E-mail address: yutian.lin@student.uts.edu.au (Y. Lin).

https://doi.org/10.1016/j.patrec.2018.06.009
0167-8655/

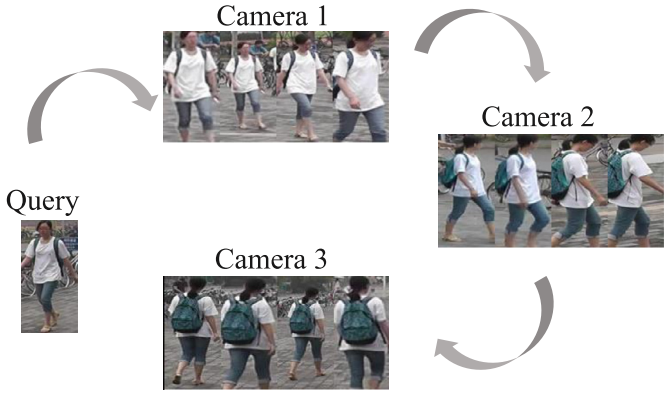

Fig. 1. An example of how query expansion works. Given a query image, true matches captured by camera 1 somehow can be easily found. With the additional discriminative cues (the backpack) of retrieved images captured by camera 1, true matches captured by camera 2 can be found. Similarly, we can then obtain the true matches captured by camera 3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in the initial query and will be used to search the system for a second time to improve re-ID recall. The pipeline of our method is shown in Fig. 2.

To evaluate the performance of the proposed method, we perform experiments on four large-scale datasets including Market-1501 [57], DukeMTMC-reID [62], MARS [56], and CUHK03 [18]. We show that BQE effectively improves the performance of the baseline systems, and that it has comparable accuracy with several competing re-ranking methods while enjoying efficient offline computations and robustness of parameter changes. Using the expanded queries, we are able to achieve very competitive results to the state-of-the-art methods.

Our contributions are summarized as follows:

(1) We propose a Bayesian Query Expansion (BQE) method for re-ID re-ranking. BQE generates a new query with information from the initial gallery ranking list, which is used to retrieve the gallery images again.
(2) The Bayesian model is trained by the distances between images in the training data and is used to calculate the probability of images in the gallery being a “true match”. We apply pooling on the features of the “true matches” to get a vector for query expansion.
(3) Our method effectively improves the person re-ID performance on several datasets, including Market-1501, CUHK03, DukeMTMC-reID and MARS. We also achieve the state-of-the-art accuracy on Market1501.

2. Related work

2.1. Hand-crafted and deep learning person re-ID systems

To address the re-ID problem, a number of approaches focus on developing robust features. In these studies, many hand-crafted features have been developed, such as color and texture histograms [6,9,19]. Zhao et al. [54,55] propose a feature that combined SIFT feature with color histogram. In [33] local feature is combined with texture, and in [40,58] local feature is combined with SIFT based on the bag-of-word (BoW) structure. Liao et al. propose Local Maximal Occurrence (LOMO) descriptor, which analyzes the horizontal occurrence of local features, and maximizes the occurrence to make a stable representation. In [26], Matsukawa et al. propose a hierarchical Gaussian descriptor, which calculates both mean and covariance information of pixel features in each patch and region hierarchy.

As deep learning arisen in recent years, researchers designed convolutional neural network (CNN) models specifically for re-ID task. Feature representations learned by CNN have achieved good performance [1,15,18,43]. Li et al. [18] train the network with pairs of pedestrian images, where the verification model with patch-matching layer is adopted. Xiao et al. [43] train a classification model from multiple domains and propose a domain guided dropout. Zheng et al. [61] adopt a joint classification and verification model and use two pairs of images for training. In [39], a Special Dense Convolutional Neural Network (SD-CNN) is used to extract the feature. Specifically, they apply joint Bayesian to measure the similarity of pedestrian image pairs.

On the other hand, a large number of metric learning algorithms [28,60] have been proposed to address the re-ID task. The general idea of metric learning is to maximize the inter-class distance and minimize the intra-class distance. The Euclidean distance is often used to assess the similarity of features. Another commonly used measure is Mahalanobis distance. In [13], Kostinger et al. propose KISSME based on Mahalanobis distance to deal with general pairwise constraints. Liao et al. [19] propose XQDA, which could be regarded as an extension of KISSME, in which a discriminant subspace was further learned together with a metric. In [21], a regularized Bayesian metric learning (RBML) method is proposed, which models and regulates the eigen-spectrums of the within-class and between-class covariance matrices in a parametric manner.

In our experiments, BoW and CNN features are used to validate our algorithm. As for metric learning, Euclidean distance, XQDA and KISSME are used and compared.

2.2. Re-ranking methods in multimedia retrieval

Re-ranking methods have widely applied in multimedia retrieval. Some of the algorithms are with human interaction [30,45], others are implemented without any extra information [10,23,63]. For the second type, the performance of the re-ranking methods highly depends on the query and the initial ranking list [27].

Many works of query expansion are proposed in the field of image retrieval [4,5] or text retrieval [37,44]. In such works, the top ranked images or documents from the original rank list are used to generate a new query that can be used to obtain a new ranking list. Nevertheless, the effects of all query expansion methods are highly affected by the initial matching results.

Another re-ranking method Pseudo relevance feedback(PRF) [23,24,49] shows a similar motivation, which assumes the top ranked sample as “pseudo relevant” to address the re-ranking problem. In [11], Jain et al. use pseudo relevant in learning method to classify the remaining samples into relevant or irrelevant classes. In [46], pseudo relevant is used as extent query to retrieve the ranking list. Lee et al. [14] propose a cluster based re-sampling method to select better pseudo-relevant documents based on the relevance model. In [47], the lowest ranked images are used as negative examples, and the initial query is used as positive example to train classifiers. Finally the images are ranked by the confidence scores. In [24], a set of positive pairs in the initial ranking list are selected and used to learns a re-ranking model by Ranking SVM.

2.3. Re-ranking methods in re-ID

Comparing with retrieval, fewer studies have investigated the re-ranking method in the field of person re-ID, some of which are presented with human in the loop and others fully automatic.

Liu et al. [22] present a post-rank optimization (POP) model which constructs an incremental affinity graph and the negative selections are propagated to their neighbors to refine their search.

Please cite this article as: Y. Lin et al., Bayesian query expansion for multi-camera person re-identification, Pattern Recognition Letters (2018), https://doi.org/10.1016/j.patrec.2018.06.009

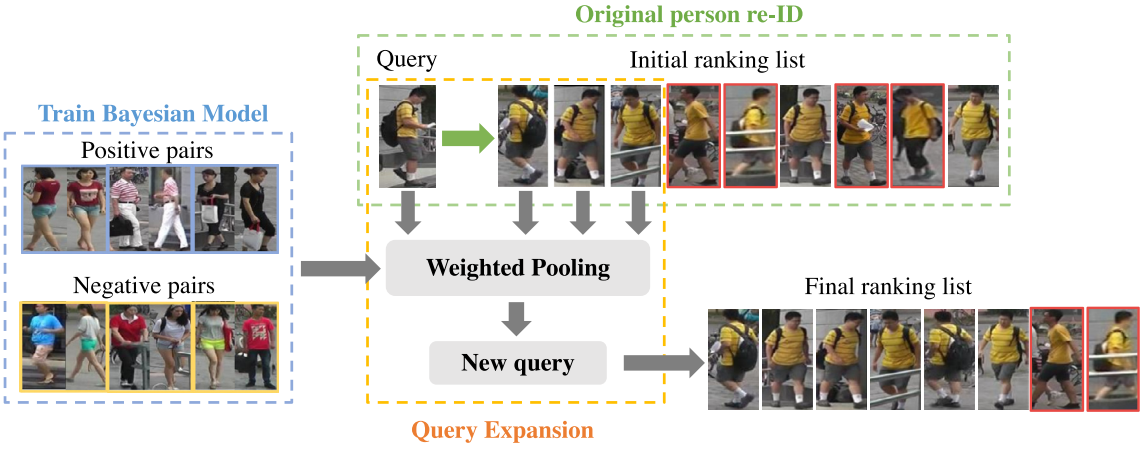

Fig. 2. Overview of the proposed approach for person re-identification. The Bayesian model is trained on the distances of relevant image pairs and irrelevant image pairs. Given a query, an initial ranking list is first obtained by a normal retrieval method. The features of top-ranked candidates are then pooled with weights calculated by Bayesian model to generate a new query. Finally, the new query is used for query expansion. The red bounding boxes denote the false matches in the ranking list. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Further, Wang et al. [38] propose a Human Verification Incremental Learning model (HVIL), which constructs an incrementally optimized ranking function and is updated in real-time.

Other researches pay attention to automatic re-ranking method. Ma et al. [25] learn an adaptive function specific for each query, which combines the base score function and query match estimated by modeling query variations. In [8], the content and context information is analyzed. And then the visual ambiguities common are removed to re-rank the initial ranking list. Leng et al. [16] calculate the matching rates of common k-nearest neighbors between every two bidirectional ranking lists as context and content similarity, and the rates are used to revise the initial query result. Li et al. [17] propose a common Near-Neighbor analysis, which analyzes the pair of samples of neighbors using both relative and direct information for re-ranking. In [50], the similarity of top retrieved images and last retrieved images are calculated by different baseline methods, then similarity ranking aggregation and dissimilarity ranking aggregation are used to optimize the ranking result. In [64], Zhong et al. encode the top-k retrieved images as the k-reciprocal feature and use it for re-ranking under the Jaccard distance.

3. Proposed method

3.1. Problem formulation

In this section, we present the Bayesian query expansion framework. The overview pipeline of our method is shown in Fig. 2. Briefly, our system consists of a Bayesian model (see Section 3.2) and the query expansion process (see Section 3.3).

In more detail, the dataset is divided into three parts: query, gallery and training data. During the offline procedure, a Bayesian posterior estimator is firstly trained on the training data. Given a distance metric, the Bayesian model can predict the probability of a candidate being a true match. During online retrieval, an initial ranking list is obtained after computing the similarities between the query and the gallery images. Based on the ranking list, the probability of each candidate being a true match is computed by the Bayesian model. Then, features of images in the initial ranking list with high probability are pooled together with the original query, such that a new query is issued to perform another round of retrieval. Note that, after the new round of retrieval, the query expansion process can be leveraged again, resulting in an iterative algorithm.

Formally, each image is represented by a d-dimensional feature vector, denoted by $x \in \mathbb{R}^d$. Let $T = \{x_i^t | i = 1, 2, \ldots M\}$ be the training set, and $G = \{x_i^g | i = 1, 2, \ldots N\}$ the gallery set. Then the identity label of training image x_i^t, query image q and gallery image x_i^g are denoted as l_i^t, l^q and l_i^g. Let $d(q, x_i^g)$ denote the distance between a query image q and a gallery image x_i^g. Then the initial ranking list is denoted as $R(q) = [x_1^g, x_2^g, \ldots x_n^g]$, where $d(q, x_i^g) < d(q, x_{i+1}^g)$. Our goal is to re-rank the initial ranking list based on an offline-trained Bayesian model.

3.2. Bayesian model

In spirit, the Bayesian model characterizes the matching score distribution of the true matches and the false matches. The model is created on the training set and deployed during testing to estimate the probability of a top-ranked image being a true match to the query.

Let us formulate the person re-ID re-ranking problem in a more formal way. For each image x in the ranking list returned from a person re-ID system, there is a distance computed by the learned metric. Since images with small distance to the query will be listed at the top, we wonder if we can re-rank the image list using the top images to improve the performance. Selecting the candidate images is crucial, because false matches will have opposite effect on the performance. The problem is, can we estimate the probability of an image being a true match when given a distance between the query and the image, *i.e.,* P(x is a true match | d(x,q))? As shown in Fig. 3, images of the same or different identities usually have obvious different range of distance, so that the distance can help us distinguish the candidates. In this paper, the Bayesian model is used to estimate the probability of relevance of an image in the ranking list.

To be specific, for a query q and a gallery image x_i^g, we propose to compute the probability that the two images belong to the same identity given the distance between the two images, *i.e.*, $P(l^q = l_i^g | d(q, x_i^g))$.

By Bayes' theorem, the probability can be rewritten as follows,

$$P(l^q = l_i^g | d(q, x_i^g)) = \frac{P(d(q, x_i^g) | l^q = l_i^g) P(l^q = l_i^g)}{P(d(q, x_i^g))}, \quad (1)$$

where $P(d(q, x_i^g))$ can be calculated by,

$$P(d(q, x_i^g)) = P(d(q, x_i^g) | l^q = l_i^g) P(l^q = l_i^g) \quad (2)$$

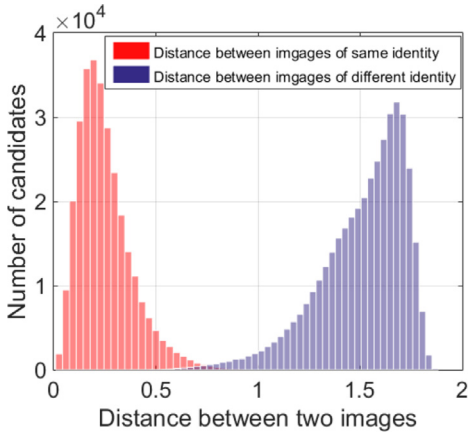

Fig. 3. Histogram of Euclidean distance on the training set of the Market-1501 dataset. The red bars in the left represent the distance between the images of the same identity, while the blue bars in the right represent the distance between images of different identities. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$+P(d(q,x_i^g)|l^q \neq l_i^g)P(l^q \neq l_i^g). \quad (2)$$

We can estimate the probability using the training data. Here, we directly use $P(l_i^t = l_j^t)$ and $P(l_i^t \neq l_j^t)$ to approximate the value of $P(l^q = l_i^g)$ and $P(l^q \neq l_i^g)$. To calculate $P(d(q,x_i^g)|l^q = l_i^g)$, and $P(d(q,x_i^g)|l^q \neq l_i^g)$, we calculate the distances between every images in the training data, and use the range of distance to replace the exact value of distance. We divide the range of distance ($d(q,x_i^g)$ values) into M intervals. Then the number of candidates within each interval is counted. As shown in Fig. 3, each bar represents an interval. Assume that $d(q,x_i^g)$ falls in [0.2,0.3], then $P(d(q,x_i^g)|l^q = l_i^g)$ can be calculated by the number of candidates divided by the frequency of the red bar in this interval. $P(d(q,x_i^g)|l^q \neq l_i^g)$ can be calculated in a similar way. The number of intervals M is chosen based on the size of the dataset. Note that if a distance in the test phase is larger than the upper bound (or smaller than the lower bound) in the training phase, we use the result of the upper (or lower bound).

3.3. Query expansion

For query expansion, a new query is issued to re-rank the candidates. Here only K candidates with high probability are pooled as a new query, where $K \leq$ *the number of true matches* and $K \ll n$. The value of K is evaluated in Section 4.3. The strategies of feature pooling are various [2].

There are two simple strategies for query expansion: average query expansion (AQE) and max query expansion (MQE). For these two methods, average pooling and max pooling are used to fuse the feature of the query image and the top ranked candidates, respectively. For AQE, the expanded query is calculated as:

$$q_{new} = \frac{\sum_{i=1}^{K} x_i^g + q}{K+1}. \quad (3)$$

The shortage of these strategies is that the effectiveness strongly relies on the quality of the initial ranking list and the value of parameter K. When the initial ranking list is not satisfying or the K is large, false matches will be used to construct the new query, which would affect the precision.

To overcome the shortage, we assign different weight to each candidate when doing feature pooling. Given a query, the probability of each candidate being a true match of the query is computed in Section 3.2. Then the expanded probe q_{new} of the initial query q is computed by pooling the top K images and query q with the probabilities. Here we simply use average pooling with weight, where the weight is exactly the probability. The formula is as follows:

$$q_{new} = \frac{\sum_{i=1}^{K} P(l^q = l_i^g) * x_i^g + q}{\sum_{i=1}^{K} P(l^q = l_i^g) + 1} \quad (4)$$

Finally this new query is used to calculate the distance and re-rank the initial ranking list.

More iterations. We assume that the expanded query will lead to a better ranking list, which can produce a better query. Thus we can conduct the procedure of producing ranking list, feature pooling, and query expansion repeatedly. By repeating BQE, the effect will be strengthened. We denote T as the number of iterations, and this parameter is evaluated in Section 4.3.

3.4. Complexity analysis

Suppose the size of the training set and the gallery set is M and N, respectively. The Bayesian model is computed offline with complexity $\mathcal{O}(M^2)$. For query expansion procedure, we need to compute the probability and construct the new query. The time complexity of generating a new query is $\mathcal{O}(K)$, where K is the number of pooled images. Since parameter $K <$ number of true matches, and $K \ll N$, the complexity can be constrained to $\mathcal{O}(1)$. Then the pairwise distance is computed with complexity $\mathcal{O}(N)$. As a result, for one query, the computation complexity is $\mathcal{O}(N)$.

4. Experiment

4.1. Datasets and settings

The Market-1501 dataset [57] is one of the largest person re-ID dataset, which contains 32,668 gallery images, 3368 query images captured by 6 cameras. It includes 500K irrelevant images as distractor set, which makes the dataset even more challenging. This dataset is captured in a noisy campus environment. Following the experimental protocol in [57], the dataset is split into 751 identities for training and 750 identities for testing.

The DukeMTMC-reID dataset [62] is a subset of the DukeMTMC dataset [29]. It contains 1812 identities captured by 8 cameras. A number of 1404 identities appear in more than two cameras, and the rest 408 IDs are distractor images. Using the evaluation protocol specified in [62], the training and testing set both contain 702 IDs, with 16,522 training images and 17,661 gallery images, respectively.

The MARS dataset [56] is the largest video re-ID dataset, which contains 1261 individuals and around 20,000 video sequences captured by six cameras. The MARS dataset is divided into train and test sets, containing 631 and 630 identities respectively. Each identity has 13.2 tracklets on average. For the MARS dataset, we use the multi-shot protocol [56] in our experiment.

The CUHK03 dataset [18] contains 13,164 images of 1360 pedestrians captured by six cameras. Each identity appears in two disjoint camera views. Note that the original evaluation protocol of CUHK03 has 20 train/test splits. To maintain consistency with other datasets, we use the train/test protocol proposed in [64]: 7365 images of 767 identities are used as the training set. 5332 images and 1400 images of the remaining 700 identities are used as the gallery set and the query set, respectively. We only conduct experiments on the DPM-detected images.

Evaluation metrics. Generally, the Cumulative Matching curve (CMC) is used to evaluate the performance of re-ID problem, which represents the probability of finding the true match in the top-n images. In this paper, we report the rank-1, 5, 10 and 20 accuracy on CMC. Meanwhile, following the existing re-ID works [57,62], we also use mAP as the metric of the retrieval performance. For each

Please cite this article as: Y. Lin et al., Bayesian query expansion for multi-camera person re-identification, Pattern Recognition Letters (2018), https://doi.org/10.1016/j.patrec.2018.06.009

Table 1
BQE results on Market-1501. The best and second highest results are highlighted in bolditalic and bold.

Methods	Rank-1	Rank-5	Rank-20	mAP
DADM [31]	39.4	–	–	19.6
MBC [34]	45.56	67	82	26.11
SML [12]	45.16	68.12	84	–
DLDA [41]	48.15	–	–	29.94
SL [3]	51.9	–	–	26.35
DNS [51]	55.43	–	–	29.87
LSTM [36]	61.6	–	–	35.3
S-CNN [35]	65.88	–	–	39.55
2Stream [61]*	79.51	90.91	96.23	59.87
GAN [62]*	79.33	–	–	55.95
APR+EU	84.29	93.20	97.00	64.67
APR+EU+BQE	***85.24***	**93.46**	**97.32**	**69.79**
APR+EU+BQEI	**84.91**	***93.47***	96.07	***70.74***
APR+Kissme	83.90	93.14	97.00	63.34
APR+Kissme+BQE	84.89	93.25	97.35	68.60
APR+XQDA	82.27	92.40	96.80	63.05
APR+XQDA+BQE	83.40	92.71	***97.43***	67.06
IDE+EU	73.69	88.15	94.83	51.48
IDE+EU+BQE	74.88	88.19	93.37	56.91
IDE+XQDA	72.35	86.78	94.32	50.19
IDE+XQDA+BQE	73.42	85.71	94.49	54.08
IDE+kissme	73.49	88.07	95.25	50.85
IDE+kissme+BQE	74.56	87.52	94.46	55.03
Bow+EU	34.03	50.80	65.91	13.15
Bow+EU+BQE	34.14	50.82	65.93	13.36
Bow+XQDA	41.93	63.26	79.87	21.90
Bow+XQDA+BQE	42.97	63.58	81.02	23.93
Bow+Kissme	41.26	60.38	78.33	20.74
Bow+Kissme+BQE	42.55	60.73	81.22	22.39

query, its average precision is computed from its precision-recall curve. MAP is then computed as the mean value of average precisions across all queries. The underlying meaning is that CMC reflects retrieval precision, while mAP reflects the recall.

Implementation details.

To demonstrate the robustness of this method, we adopt various feature extraction methods and metric learning approaches as baselines.

- ID-discriminative Embedding (IDE) [59]. The IDE extractor is trained on classification model using ResNet-50. It generate a 2048-dim vector for each image.
- Attribute-Person Re-identification Embedding (APR) [20]. The descriptor is trained on classification model using both ID and attribute labels. The network is also trained on ResNet-50 model. For each image, a 2048-dim vector is extracted.
- Bag-of-Words (BoW) descriptor [57]. In the BoW model, local features are aggregated into a global feature vector. For each image, a 5600-dim descriptor is computed.

We also employ three distance learning methods, including Euclidean distance, KISSME and XQDA.

For all the experiments, we set the parameter $K = 5$ on Market-1501, DukeMTMC-reID and MARS. We set $K = 3$ on CUHK03, because the true matches of this dataset are fewer. For all the four datasets, the number of intervals M is set to be 100. When conduct BQE for iterations, the parameter T is set to be 3.

4.2. Evaluation of BQE

Comparison with the baselines.

We evaluate if the BQE method outperforms the baselines. Results on the four datasets, *i.e.,* Market-1501, Duke, MARS and CUHK03, are reported in Tables 1–4, respectively. Especially we show the baseline and BQE results on the four datasets using the same setting in Fig. 4. We have several observations from these results.

Table 2
BQE results on DukeMTMC-reID. The best and second highest results are highlighted in bolditalic and bold.

Methods	Rank-1	Rank-5	Rank-20	mAP
BoW+kissme [57]	25.13	–	–	12.17
LOMO+XQDA [19]	30.75	–	–	17.04
GAN (R, 702) [62]	67.68	–	–	47.13
SVDNet [32]	***76.7***	–	–	56.8
APR+EU	70.69	84.78	91.78	51.89
APR+EU+BQE	74.95	***85.95***	***92.77***	**59.48**
APR+EU+BQEI	**76.48**	**85.81**	91.49	***60.97***
APR+Kissme	70.78	84.15	91.15	50.62
APR+Kissme+BQE	75.49	85.456	91.60	55.79
APR+XQDA	71.18	84.02	91.24	51.21
APR+XQDA+BQE	74.77	85.32	91.60	58.43
IDE+EU	61.71	76.75	85.60	41.21
IDE+EU+BQE	70.37	81.73	89.22	53.03
IDE+Kissme	66.87	80.92	88.46	44.95
IDE+Kissme+BQE	69.12	80.25	88.06	48.60
IDE+XQDA	67.05	79.75	88.06	45.33
IDE+XQDA+BQE	70.33	81.28	88.82	51.72

Table 3
BQE results on MARS. The best highest results are highlighted in bolditalic.

Methods	Rank-1	Rank-5	Rank-20	mAP
HistLBP+XQDA [56]	18.6	33.0	45.9	8.0
gBiCov+XQDA	9.2	19.8	33.5	3.7
LOMO+XQDA	30.7	46.6	60.9	16.4
BoW+Kissme	30.6	46.2	59.2	15.5
IDE+EU [56]	56.47	73.09	83.23	38.82
IDE+EU+BQE	59.32	77.18	87.19	42.75
IDE+Kissme [56]	63.54	79.35	88.4	45.20
IDE+Kissme+BQE	66.07	81.23	89.78	47.54
IDE+XQDA [56]	65.59	81.75	90.10	46.84
IDE+XQDA+BQE	66.46	***82.25***	***90.39***	51.38
IDE+XQDA+BQEI	***68.42***	81.59	88.82	***52.73***

Table 4
BQE results on CUHK03. The best results are highlighted in bolditalic.

Methods	Rank-1	Rank-5	Rank-20	mAP
k-reciprocal [64]	34.7	–	–	37.4
IDE+EU	21.35	37.50	57.42	19.75
IDE+EU+BQE	24.00	39.85	59.21	21.91
IDE+Kissme	27.88	46.57	65.71	26.80
IDE+Kissme+BQE	30.71	50.71	68.93	30.49
IDE+XQDA	29.50	50.00	71.35	28.14
IDE+XQDA+BQE	33.85	54.00	***74.78***	32.07
IDE+XQDA+BQEI	***34.78***	50.92	73.01	34.46

First, our method exceeds the baselines on four datasets. On the Market-1501 dataset, our method consistently improves the rank-1 accuracy and mAP with all features (IDE, APR and Bow model). For example, when using IDE [20] and Euclidean distance, the improvements are +0.95% in rank-1 accuracy and +5.12% in mAP. Moreover, experiments conducted with three metrics all show better results than the baselines, which prove the effectiveness of our method on different distance metrics.

On the DukeMTMC-reID dataset, we test BQE with IDE, APR and three matrix learning methods. In all the settings, the performance of BQE is superior to the baselines. For example, when using APR [20] and Euclidean distance, our performance get the most obvious improvement. The performance gain is +4.26% in rank-1 accuracy and +7.59% in mAP.

Consistent findings also hold for the MARS dataset and the CUHK03 dataset, *i.e.*, when IDE and three matrix learning methods are used, BQE improves both rank-1 accuracy and mAP over the baselines.

Please cite this article as: Y. Lin et al., Bayesian query expansion for multi-camera person re-identification, Pattern Recognition Letters (2018), https://doi.org/10.1016/j.patrec.2018.06.009

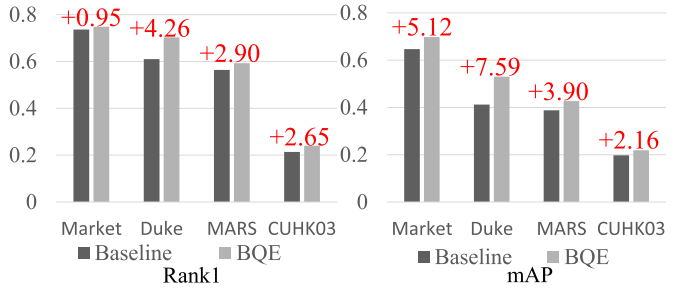

Fig. 4. Comparison between the baselines and BQE on four datasets. All the experiments are implemented using IDE and Euclidean distance.

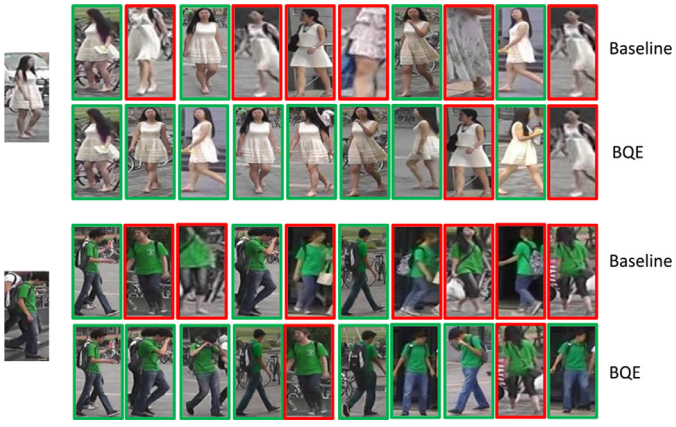

Fig. 5. Sample results on the Market-1501 dataset. The green bounding boxes and red bounding boxes denote the true matches and the false ones, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

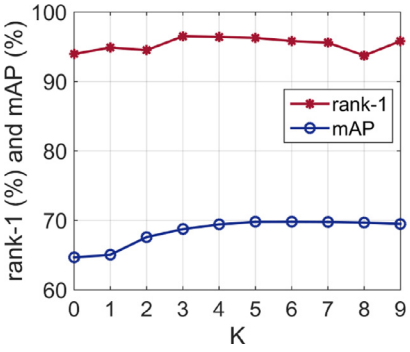

Fig. 6. mAP and CMC vs. K value changes on the Market-1501 dataset. The experiments are implemented using APR and Euclidean distance.

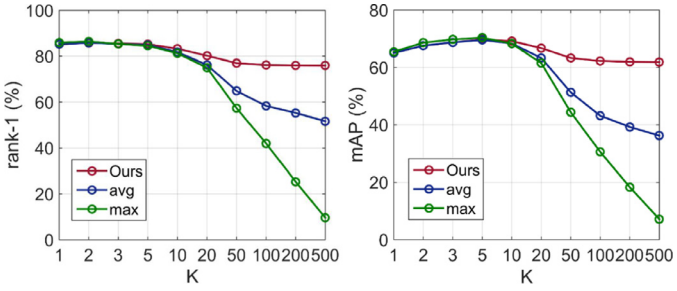

Fig. 7. mAP and CMC of three methods vs. K value changes on the Market-1501 dataset. The experiments are implemented using APR and Euclidean distance.

Second, the improvement of mAP is larger than that of rank-1 accuracy. For example, on the Market-1501 dataset, when using CNN features ([59] or [20]), mAP increases 4% $\sim$ 5%, while rank-1 increases 1% $\sim$ 2%. The results on the DukeMTMC-reID dataset are similar, *i.e.*, the improvement of mAP and rank-1 accuracy is about 5% $\sim$ 11% and 3% $\sim$ 9%, respectively. We speculate that when rank-1 is a false match, noise will be introduced to the new query, thus it's difficult to have a true match on rank-1 in the new ranking list. However, diversity will be introduced to the new query as well, and some true matches will have a higher rank in the new ranking list, so mAP has a relatively higher improvement.

Two sample re-ID results on the Market-1501 dataset are shown in Fig. 5. It is clear that more true matches are found using BQE. Our method improves the baseline algorithm to a great extent.

Comparison with the state-of-the-art methods. On the Market-1501 dataset, we obtain rank-1 of 85.24%, mAP of 69.79% with the APR. We achieve the best rank-1 accuracy and mAP among the competing methods. On the DukeMTMC-reID dataset, we achieve rank-1 of 76.48% and mAP of 60.97% with APR and Euclidean distance when repeat BQE for 3 iterations. We achieve the best mAP among the competing methods, and the second best in rank-1 accuracy (the highest rank-1 accuracy is reported by Sun et al. [32]). On the CUHK03 dataset, our method yields the best rank-1 accuracy result of 34.78% and the second best mAP of 34.46% (the highest mAP is reported by Zhong et al. [64]).

Effective of more iterations. We conduct BQE method for three iterations with the setting that produces the best performance. On these four datasets, BQE with more iterations (BQEI) shows superior results. On the Market-1501 dataset, BQEI achieves a better mAP of 70.74% (+0.95%) and a relatively lower rank-1 accuracy of 84.91% (−0.33%). On the DukeMTMC-reID dataset, compared with the BQE method, rank-1 increases from 74.95% to 76.48%, mAP increases from 59.48% to 60.97%. On the MARS dataset, an improvement of 1.98% on rank-1 accuracy and an improvement of 1.35% on mAP are observed. On the CUHK03 dataset, rank-1 accuracy and mAP of BQEI are 34.78% and 34.46%, respectively. An improvement of 0.93% on rank-1 and an improvement of 2.39% on mAP are observed.

4.3. Algorithm analysis

Number of top ranked candidates. An important parameter to be considered is K, which defines how many candidates will be pooled to construct a new query together with the original query. As we can imagine, when K is 0, the new query is equal to the original query. When K becomes larger, more images in the ranking list are used to generate the new query. To assign a suitable K value, we compare the result under different K values for different datasets. As for Market-1501 benchmark, the curve is shown in Fig. 6. At first, both recall and precision are improved. When K increases, mAP raises slowly and then remains the same value, rank-1 accuracy drops slowly and has a fluctuation. We set $K = 5$ to get a satisfying result.

Stability study. Fig. 7 presents the performance of three query expansion methods when have different K value. Notice that when K is small, the three algorithms produce similar results on both rank-1 accuracy and mAP. As K increases, the performance of BQE drops a little and remain a relative high accuracy. On the contrary, for AQE and MQE, both rank-1 accuracy and mAP drop rapidly.

We further investigate the weight of top 500 images in Fig. 8. We observe that the weight of top ranked images drops rapidly, and remains nearly zero when $K > 50$. The Bayesian model play an important role on control the impact of the top ranked images to form the expanded query.

Number of iterations. The impact of the number of iteration T is shown in Fig. 9. On the one hand, when T increases, mAP increases slowly and then remains a relatively stable level. The rea-

Please cite this article as: Y. Lin et al., Bayesian query expansion for multi-camera person re-identification, Pattern Recognition Letters (2018), https://doi.org/10.1016/j.patrec.2018.06.009

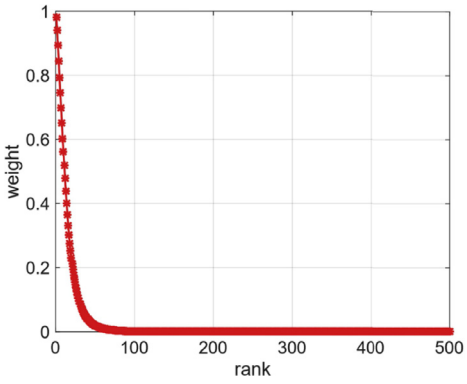

Fig. 8. Pooling weight of top 500 ranked images on Market-1501.

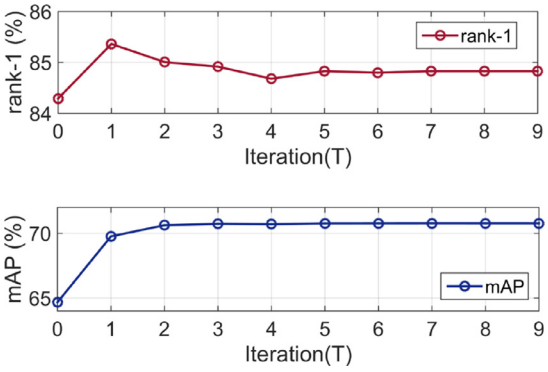

Fig. 9. Rank-1 and mAP vs. T value changes on the Market-1501 dataset. The experiments are implemented using APR and Euclidean distance.

son is that as more images are merged into the query, the diversity of the query vector is enhanced, thus improving the retrieval recall.

On the other hand, although mAP is improved, more iterations of query expansion do not exactly lead to a higher rank-1 accuracy. We observe a best rank-1 accuracy in the first iteration, and then it decreases to about 84.9%. For further iterations, rank-1 accuracy remains, which is still higher than the baseline. The main reason for this phenomenon is that some false matches might be ranked on the top, and are pooled into the new query several times during the iterations. That being said, on DukeMTMC-reID (Table 2), MARS (Table 3) and CUHK03 (Table 4), both rank-1 and accuracy improves with more iterations.

Overall speaking, mAP usually benefits from more iterations of BQE; rank-1 accuracy can be improved as well, but may be compromised in some cases due to the deteriorated query.

4.4. Comparison with other re-ranking methods

In this section we implement and compare some query expansion methods such as AQE and MQE as we discussed in Section 3.3. Other kinds of re-ranking methods are also adopted and compared with BQE and the baseline.

- SVM query expansion. An SVM classifier is firstly trained on training data, then it replaces the Bayesian model in predicting labels of candidates in the initial ranking list.
- Graph based re-ranking method [53]. The top candidates' nearest neighborhoods are used as queries to get the ranking lists. Then a weighted undirected graph is built using the top candidates from those ranking lists. Here, the Jaccard similarity coefficient between two neighbors is used as weight.
- K-reciprocal encoding [64]. This method calculates a k-reciprocal feature for re-ranking under the Jaccard distance. The final distance is computed by combining the original distance and the Jaccard distance. On the contrary, we propose a new query to compute the final distance. The time complexity of this method is $\mathcal{O}(NlogN)$, while ours is $\mathcal{O}(N)$. The two methods are complementary to each other. In Table 5 we show that we further improve the performance when build BQE upon the K-reciprocal encoding [64].

Table 5
Re-ranking methods comparison on the Market-1501 dataset. Our method is more robust to parameters than average/max pooling. Our method also requires much less off-line computation and thus more flexible against database updates than [52,64]. The best and second highest results are highlighted in bolditalic and bold.

Methods	rank-1	mAP
APR	84.29	64.67
APR+AQE	84.56	70.40
APR+MQE	84.88	69.60
APR+SVM	85.12	67.33
APR+Graph[52]	84.71	67.98
APR+k-reci [64]	***85.87***	***77.27***
APR+BQE	85.24	69.79
APR+BQEI	84.91	70.74
APR+Graph+BQE	85.42	70.35
APR+k-reci+BQE	***86.57***	**76.95**

The CMC and mAP results of different re-ranking methods are reported in Table 5. Here all the results are calculated using APR under Euclidean distance. We observe that all the re-ranking methods consistently improves the rank-1 accuracy and mAP over the original re-ID baseline. Among the query expansion methods (the second row to the six row in Table 5), the proposed BQE method achieves the best result in rank-1 accuracy and BQE with three iterations achieves the best result in mAP. Other query expansion methods also beat the baseline to some extent.

The graph based method [52] exhibits a similar result to the query expansion methods. We also notice that the K-reciprocal encoding method achieves best result with 85.87% on rank-1 and 77.27% on mAP. However, the shortcoming of this method is that it needs much more offline training time, a process that needs to be re-computed every time the gallery is updated. We use BQE behind some other re-ranking methods, and the results are shown below. We also use BQE after these re-ranking method, and observe that when BQE is conducted after other graph based method, the rank-1 accuracy improves from 84.71% to 85.42%, and mAP improves from 67.89% to 70.30%. When we perform BQE after k-reciprocal method, rank-1 accuracy improves from 85.87% to 86.57.

5. Conclusion

We introduce a Bayesian query expansion algorithm to improve the performance of existing person Re-ID approaches. The pairwise similarity scores between images from the same and different identities are used to train the Bayesian model. For each query, the top ranked candidates in the initial ranking list are selected, and the features are pooled with the query using Bayesian model. A new query is then obtained and can be used to produce a new ranking list. The experiments show that our approach consistently improves the performance of baselines and is very robust to feature representation and metric learning methods. Our results are competitive with the state-of-the-art methods on four large-scale re-ID datasets.

Please cite this article as: Y. Lin et al., Bayesian query expansion for multi-camera person re-identification, Pattern Recognition Letters (2018), https://doi.org/10.1016/j.patrec.2018.06.009

In the future, we will further investigate effective and efficiency re-ranking methods. A promising direction is to integrate human labor in the re-ranking process [22,38].

References

[1] X. Chang, Y. Yang, Semisupervised feature analysis by mining correlations among multiple tasks, IEEE Trans. Neural Netw. Learn. Syst. 28 (10) (2017) 2294–2305.
[2] X. Chang, Y.-L. Yu, Y. Yang, E.P. Xing, Semantic pooling for complex event analysis in untrimmed videos, IEEE Trans. Pattern Anal. Mach. Intell. 39 (8) (2017) 1617–1632.
[3] D. Chen, Z. Yuan, B. Chen, N. Zheng, Similarity learning with spatial constraints for person re-identification, in: The IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[4] O. Chum, A. Mikulik, M. Perdoch, J. Matas, Total recall ii: query expansion revisited, in: The IEEE Conference on Computer Vision and Pattern Recognition, 2011.
[5] O. Chum, J. Philbin, J. Sivic, M. Isard, A. Zisserman, Total recall: automatic query expansion with a generative feature model for object retrieval, in: IEEE International Conference on Computer Vision, 2007.
[6] M. Farenzena, L. Bazzani, A. Perina, V. Murino, M. Cristani, Person re-identification by symmetry-driven accumulation of local features, in: The IEEE Conference on Computer Vision and Pattern Recognition, 2010.
[7] C. Gao, D. Meng, Y. Yang, Y. Wang, X. Zhou, A.G. Hauptmann, Infrared patch-image model for small target detection in a single image, IEEE Trans. Image Process. 22 (12) (2013) 4996–5009.
[8] J. Garcia, N. Martinel, C. Micheloni, A. Gardel, Person re-identification ranking optimisation by discriminant context information analysis, in: The IEEE International Conference on Computer Vision, 2015, pp. 1305–1313.
[9] D. Gray, H. Tao, Viewpoint invariant pedestrian recognition with an ensemble of localized features, in: European Conference on Computer Vision, 2008.
[10] J. He, C. Zhang, N. Zhao, H. Tong, Boosting web image search by co-ranking, in: Acoustics, Speech, and Signal Processing, 2, 2005, pp. ii–409.
[11] V. Jain, M. Varma, Learning to re-rank: query-dependent image re-ranking using click data, in: International Conference on World Wide Web, 2011, pp. 277–286.
[12] C. Jose, F. Fleuret, Scalable metric learning via weighted approximate rank component analysis, in: European Conference on Computer Vision, 2016.
[13] M. Köstinger, M. Hirzer, P. Wohlhart, P.M. Roth, H. Bischof, Large scale metric learning from equivalence constraints, in: The IEEE Conference on Computer Vision and Pattern Recognition, 2012.
[14] K.S. Lee, W.B. Croft, J. Allan, A cluster-based resampling method for pseudo-relevance feedback, in: IEEE International Conference on Advanced Video and Signal-based Surveillance, 2008, pp. 235–242.
[15] Y.-G. Lee, S.-C. Chen, J.-N. Hwang, Y.-P. Hung, An ensemble of invariant features for person reidentification, IEEE Trans. Circuits Syst. Video Technol. 27 (3) (2017) 470–483.
[16] Q. Leng, R. Hu, C. Liang, Y. Wang, J. Chen, Person re-identification with content and context re-ranking, Multimedia Tools Appl. 74 (17) (2015) 6989–7014.
[17] W. Li, Y. Wu, M. Mukunoki, M. Minoh, Common-near-neighbor analysis for person re-identification, in: Image Processing (ICIP), 2012 19th IEEE International Conference on, 2012, pp. 1621–1624.
[18] W. Li, R. Zhao, T. Xiao, X. Wang, Deepreid: deep filter pairing neural network for person re-identification, in: The IEEE Conference on Computer Vision and Pattern Recognition, 2014.
[19] S. Liao, Y. Hu, X. Zhu, S.Z. Li, Person re-identification by local maximal occurrence representation and metric learning, in: The IEEE Conference on Computer Vision and Pattern Recognition, 2015.
[20] Y. Lin, L. Zheng, Z. Zheng, Y. Wu, Y. Yang, Improving person re-identification by attribute and identity learning, arXiv:1703.07220(2017).
[21] V.E. Liong, J. Lu, Y. Ge, Regularized bayesian metric learning for person re-identification, in: European Conference on Computer Vision, 2014, pp. 209–224.
[22] C. Liu, C. Change Loy, S. Gong, G. Wang, Pop: person re-identification post-rank optimisation, in: The IEEE Conference on International Conference on Computer Vision, 2013.
[23] Y. Liu, T. Mei, Optimizing visual search reranking via pairwise learning, IEEE Trans. Multimedia 13 (2) (2011) 280–291.
[24] Y. Liu, T. Mei, X.-S. Hua, J. Tang, X. Wu, S. Li, Learning to video search rerank via pseudo preference feedback, in: Multimedia and Expo, 2008 IEEE International Conference on, 2008, pp. 297–300.
[25] A.J. Ma, P. Li, Query based adaptive re-ranking for person re-identification, in: Asian Conference on Computer Vision, 2014.
[26] T. Matsukawa, T. Okabe, E. Suzuki, Y. Sato, Hierarchical gaussian descriptor for person re-identification, in: The IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[27] T. Mei, Y. Rui, S. Li, Q. Tian, Multimedia search reranking: a literature survey, ACM Comput. Surv. (CSUR) 46 (3) (2014) 38.
[28] S. Pedagadi, J. Orwell, S. Velastin, B. Boghossian, Local fisher discriminant analysis for pedestrian re-identification, in: The IEEE Conference on Computer Vision and Pattern Recognition, 2013.
[29] E. Ristani, F. Solera, R. Zou, R. Cucchiara, C. Tomasi, Performance measures and a data set for multi-target, multi-camera tracking, in: European Conference on Computer Vision Workshop, 2016.
[30] U. Rohini, V. Varma, A novel approach for re-ranking of search results using collaborative filtering, in: Computing: Theory and Applications, 2007. ICCTA'07. International Conference on, 2007, pp. 491–496.
[31] C. Su, S. Zhang, J. Xing, W. Gao, Q. Tian, Deep attributes driven multi-camera person re-identification, in: European Conference on Computer Vision, 2016.
[32] Y. Sun, L. Zheng, W. Deng, S. Wang, Svdnet for pedestrian retrieval, in: The IEEE International Conference on Computer Vision, 2017.
[33] Q. Tian, N. Sebe, M.S. Lew, E. Loupias, T.S. Huang, Image retrieval using wavelet-based salient points, J. Electron. Imaging 10 (4) (2001) 835–850.
[34] E. Ustinova, Y. Ganin, V. Lempitsky, Multi-region bilinear convolutional neural networks for person re-identification, in: IEEE International Conference on Advanced Video and Signal-based Surveillance, 2017.
[35] R.R. Varior, M. Haloi, G. Wang, Gated siamese convolutional neural network architecture for human re-identification, in: European Conference on Computer Vision, 2016.
[36] R.R. Varior, B. Shuai, J. Lu, D. Xu, G. Wang, A siamese long short-term memory architecture for human re-identification, in: European Conference on Computer Vision, 2016.
[37] E.M. Voorhees, Query expansion using lexical-semantic relations, in: IEEE International Conference on Advanced Video and Signal-based Surveillance, 1994.
[38] H. Wang, S. Gong, X. Zhu, T. Xiang, Human-in-the-loop person re-identification, in: European Conference on Computer Vision, 2016.
[39] S. Wang, L. Duan, N. Yang, J. Dong, Person re-identification with deep dense feature representation and joint bayesian, in: Image Processing (ICIP), 2017 IEEE International Conference on, 2017, pp. 3560–3564.
[40] C. Wengert, M. Douze, H. Jégou, Bag-of-colors for improved image search, in: Proceedings of the 19th ACM International Conference on Multimedia, 2011, pp. 1437–1440.
[41] L. Wu, C. Shen, A. van den Hengel, Deep linear discriminant analysis on fisher networks: a hybrid architecture for person re-identification, Pattern Recognit. (2017).
[42] Y. Wu, Y. Lin, X. Dong, Y. Yan, W. Ouyang, Y. Yang, Exploit the unknown gradually: one-shot video-based person re-identification by stepwise learning, in: The IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[43] T. Xiao, H. Li, W. Ouyang, X. Wang, Learning deep feature representations with domain guided dropout for person re-identification, in: The IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[44] J. Xu, W.B. Croft, Query expansion using local and global document analysis, in: IEEE International Conference on Advanced Video and Signal-based Surveillance, ACM, 1996.
[45] T. Yamamoto, S. Nakamura, K. Tanaka, Rerank-by-example: Efficient browsing of web search results, in: International Conference on Database and Expert Systems Applications, 2007, pp. 801–810.
[46] R. Yan, A. Hauprmann, Query expansion using probabilistic local feedback with application to multimedia retrieval, in: Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management, 2007, pp. 361–370.
[47] R. Yan, A. Hauptmann, R. Jin, Multimedia search with pseudo-relevance feedback, in: International Conference on Image and Video Retrieval, 2003, pp. 238–247.
[48] Y. Yan, F. Nie, W. Li, C. Gao, Y. Yang, D. Xu, Image classification by cross-media active learning with privileged information, IEEE Trans. Multimedia 18 (12) (2016) 2494–2502.
[49] Y. Yang, F. Nie, D. Xu, J. Luo, Y. Zhuang, Y. Pan, A multimedia retrieval framework based on semi-supervised ranking and relevance feedback, IEEE Trans. Pattern Anal. Mach. Intell. 34 (4) (2012) 723–742.
[50] M. Ye, C. Liang, Y. Yu, Z. Wang, Q. Leng, C. Xiao, J. Chen, R. Hu, Person reidentification via ranking aggregation of similarity pulling and dissimilarity pushing, IEEE Trans. Multimedia 18 (12) (2016) 2553–2566.
[51] L. Zhang, T. Xiang, S. Gong, Learning a discriminative null space for person re-identification, in: IEEE International Conference on Computer Vision, 2016.
[52] S. Zhang, M. Yang, T. Cour, K. Yu, D.N. Metaxas, Query specific fusion for image retrieval, in: European Conference on Computer Vision, 2012, pp. 660–673.
[53] S. Zhang, M. Yang, T. Cour, K. Yu, D.N. Metaxas, Query specific rank fusion for image retrieval, IEEE Trans. Pattern Anal. Mach. Intell. (2015).
[54] R. Zhao, W. Ouyang, X. Wang, Person re-identification by salience matching, in: IEEE International Conference on Computer Vision, 2013.
[55] R. Zhao, W. Ouyang, X. Wang, Learning mid-level filters for person re-identification, in: The IEEE Conference on Computer Vision and Pattern Recognition, 2014.
[56] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, Q. Tian, Mars: a video benchmark for large-scale person re-identification, in: European Conference on Computer Vision, 2016.
[57] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian, Scalable person re-identification: a benchmark, in: The IEEE Conference on Computer Vision and Pattern Recognition, 2015.
[58] L. Zheng, S. Wang, Z. Liu, Q. Tian, Packing and padding: Coupled multi-index for accurate image retrieval, in: IEEE Conference on computer vision and pattern recognition, 2014.
[59] L. Zheng, Y. Yang, A.G. Hauptmann, Person re-identification: past, present and future, arXiv:1610.02984 (2016).
[60] W.-S. Zheng, S. Gong, T. Xiang, Person re-identification by probabilistic relative distance comparison, in: The IEEE Conference on Computer Vision and Pattern Recognition, 2011.
[61] Z. Zheng, L. Zheng, Y. Yang, A discriminatively learned cnn embedding for per-

Please cite this article as: Y. Lin et al., Bayesian query expansion for multi-camera person re-identification, Pattern Recognition Letters (2018), https://doi.org/10.1016/j.patrec.2018.06.009

son reidentification, ACM Trans. Multimedia Comput. Commun. Appl. (TOMM) 14 (1) (2017) 13.
[62] Z. Zheng, L. Zheng, Y. Yang, Unlabeled samples generated by gan improve the person re-identification baseline in vitro, in: The IEEE Conference on International Conference on Computer Vision, 2017.
[63] Z. Zhong, M. Lei, D. Cao, J. Fan, S. Li, Class-specific object proposals re-ranking for object detection in automatic driving, Neurocomputing 242 (2017) 187–194.
[64] Z. Zhong, L. Zheng, D. Cao, S. Li, Re-ranking person re-identification with k-reciprocal encoding, in: The IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3652–3661.
[65] Z. Zhong, L. Zheng, Z. Zheng, S. Li, Y. Yang, Camera style adaptation for person re-identification, in: Computer Vision and Pattern Recognition (CVPR), 2018 IEEE Conference on, 2018.

Please cite this article as: Y. Lin et al., Bayesian query expansion for multi-camera person re-identification, Pattern Recognition Letters (2018), https://doi.org/10.1016/j.patrec.2018.06.009