2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Smart Query Expansion Scheme for CDVS Based on Illumination and Key Features

Tao Lu[1 2], Chuang Zhu[2], Huizhu Jia[2*], Lingyu Duan[2], Li Tao[2], Jiawen Song[2], Xiaodong Xie[2] and Wen Gao[2]

[1] School of Electronic and Computer Engineering
Shenzhen Graduate School, Peking University
No.2199 Lishui Road, Shenzhen 518055, China

[2] National Engineering Laboratory for Video Technology
Dept. EECS, Peking University
No.5 Yiheyuan Road, Beijing 100871, China

{lutao, czhu, hzjia, lingyu, chntaoli, jiawens, donxie, wgao}@pku.edu.cn

***Abstract*—Given a query image, retrieving images depicting the same object in a large scale database is becoming an urgent and challenging task. Recently, Compact Description for Visual Search (CDVS) is drafted by the ISO/IEC Moving Pictures Experts Group (MPEG) to support image retrieval applications, and it has been published as an international standard. Unfortunately, with regard to applications with hugely mutative illumination, perspective and noisy background, CDVS suffers from an inevitable performance loss. In this paper, firstly we introduce the query expansion to address performance loss caused by the scene complexity in CDVS. Secondly, a query expansion instance selection method based on illumination is proposed, which achieves better performance. Thirdly, we adopt a key feature matching score based weighted strategy in basic query expansion to improve retrieval performance. We evaluate our proposed methods on the Oxford (5K images) dataset and a reality traffic vehicle dataset (12K images), and the result shows that the proposed methods boost mean average precision (MAP) by 7% ~ 10% in Oxford dataset and 7% ~17% in vehicle dataset.**

Keywords—query expansion; compact description for visual search; matching; image retrieval; illumination

I. INTRODUCTION

In recent years, with the popularity of digital image devices in various areas, the demand for directly visually searching based on image has become stronger and stronger. To address the computation complexity, memory load and bandwidth limitation in visual search, the MPEG drafted the Compact Description for Visual Search [1] standard, in which image compact descriptor consists of global descriptor (GD) and local descriptor (LD) generated from selected SIFT points. CDVS is proved to achieve remarkable improvements in increasing compactness of image descriptor and reducing the computation complexity and memory demand while obtaining high MAP. However, in feature points based visual searching system, a large range of scale changes, great affine transformation and quantization distortion in descriptor extraction can cause some target objects missed in retrieval. Besides, in some complex application scenes, such as traffic vehicle retrieval, the illumination may change greatly from day to night, which leads to much image retrieval performance loss because of the huge illumination difference. All these issues prevent image retrieval from practical applications.

To address these issues, approaches including matching graph, rank fusion and query expansion have been proposed. The matching graph [2] [3] is constructed offline on the database side, which connects all related images. On the query side, the graph can provide a set of related database images. In [4], ordered retrieval sets given by multiple retrieval methods are modeled as graphs and two graph based methods, Graph-PageRank and Graph-density, are proposed to fuse different retrieval sets. Yanzhi Chen *et al.* [5] introduce a group-query based rank fusion method, in which a boosted classifier is trained to merge and re-rank the individual results from different queries. Query expansion (QE) is firstly introduced into Bag of Words (BoW) image retrieval system based on the assumption the spatially consistent images depict the same object. Experiments show the query expansion is of high performance and worthy of further study.

In this paper, we focus on query expansion, in which the spatially verified images are used to issue new queries. In [6], the proposed automatic query expansion (AQE) has been shown to bring a significant performance boost. In [7], Chum *et al.* propose Incremental Spatial Re-ranking (iSP) to count the matched features in previously verified images for more effective spatial verification. In [8], negative data in the first query are taken into consideration and a linear SVM classifier is trained for discriminatively query expansion (DQE). DQE is extended in [9] using pairwise instead of pointwise learning in re-ranking stage to preserve the sub-ranking order. In [10], the contextual query expansion method based on common visual patterns (CVPs) is introduced, in which two contextual query expansions on visual word-level and image-level are explored to improve retrieval performance. All the query expansion strategies show great performance improvement but are faced with a critical problem: how to select the most effective re-query image from verified results. As shown in Fig. 1, the

This work was supported in part by the National Science Foundation of China (61421062, 91538111), National Key Technology Research and Development Program of the Ministry of Science and Technology (2014BAK10B00), the Major National Scientific Equipment Project (2013YQ030967), the National Basic Research Program (973 Program): 2015CB351806 and China Postdoctoral Science Foundation (2016M590020)

*the corresponding author, Huizhu Jia is with Peking University, also with Cooperative Medianet Innovation and Beida(Binhai) Information Research

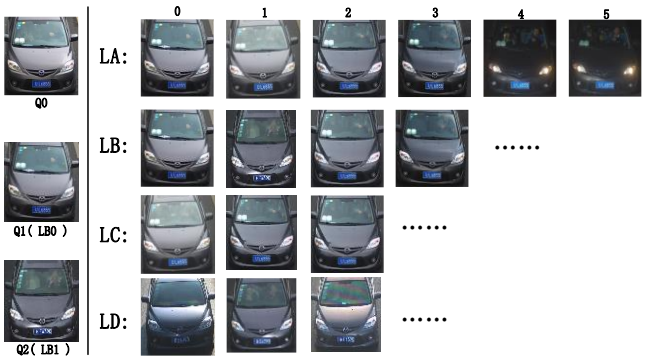

Fig. 1. An example in basic query expansion

origin query image Q0 has 4 related images in LA. Origin query returns 4 results in LB ranked according to spatial consistency, in which the LB0 is related and LB1 are unrelated. In QE, all 3 results in LC returned by Q1 (LB0) are related but already contained in LB. Another 3 images in LD are retrieved by Q2 (LB1), but all of them depict different object from Q0. In short, the most ideal image for query expansion should be related and have certain difference with origin query image. Therefore, an effective query expansion image selection method is valuable.

In this paper, a proposed basic query expansion (BQE) is firstly introduced into CDVS image retrieval system. And secondly an expansion image selection method based on illumination (EISBI) is proposed to address illumination difference, following a pairwise strong spatial verification to remove unrelated images in origin retrieval results. Thirdly we propose a key feature matching score based weight strategy (MSBW), which is inspired by the robust feature selection method based on self-matching score in [11]. The remaining sections are organized as follows. The image retrieval system based on CDVS is introduced in section II and the proposed methods are described in Section III. In Section IV, the experiment method, evaluation criterion and comparison results are detailed. Finally, we conclude this paper in Section V.

II. IMAGE RETRIEVAL SYSTEM BASED ON CDVS

To meet practical image retrieval system demands, CDVS adopts many proposals, including 6 different query descriptor lengths (512B, 1K, 2K, 4K, 8K and 16K) for different scenarios. We build the image retrieval system based on CDVS retrieval framework, including CDVS bitstream extraction and

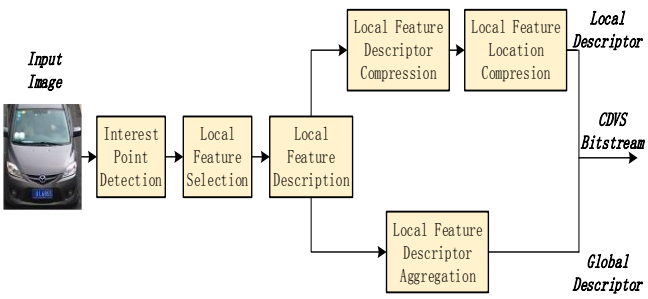

Fig. 2. CDVS Bitstream Extraction

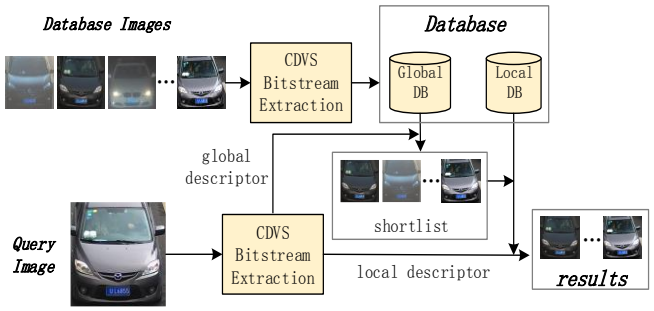

Fig. 3. Image Retrieval Using CDVS Bitstream

image retrieval using CDVS bitstream.

A. CDVS Bitstream Extarction

The CDVS standard defines the compressed bitstream syntax, and the CDVS bitstream extraction includes distinctive local feature detection and compressed descriptor generation. The six key procedures illustrated in Fig. 2.

Firstly, in CDVS, the result of Laplacian of Gaussian (LoG) filtering is approximated by the adopted low-degree polynomial (ALP) to detect key points. Secondly, a subset of features are selected based on the relevance measure. Thirdly, the popular SIFT [12] descriptor characterizes the interest points with a 128-dimension gradient histogram. Then, in local feature compression process, CDVS adopts a transform coding scheme followed by ternary scalar quantization and entropy coding. Local feature location compression module produces the compressed local feature descriptor and location data, which is called local descriptor (LD). Meanwhile, the selected local features are aggregated into global descriptor (GD) by Gaussian Mixture Model (GMM). At last, the LD and GD are packaged into CDVS bitstream.

B. Image Retrieval Using CDVS Bitstream

The CDVS standard also specifies the retrieval framework. As show in Fig. 3, a collection of local and global descriptors are aggregated into the database respectively. The query local and global descriptor is extracted from query image. The image retrieval module consists of global descriptor query, spatially Geometric Consistency Check (GCC) and re-ranking based on local descriptor.

Firstly, each global descriptor in database is compared with query global descriptor, and a number of highly ranked images, such as 500, are selected in the shortlist according to hamming

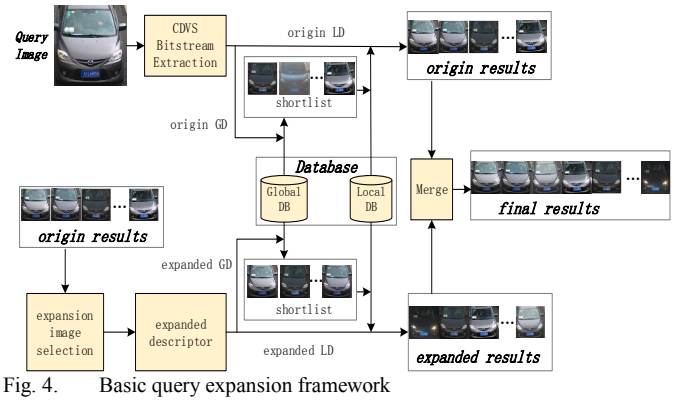

Fig. 4. Basic query expansion framework

distance. Secondly, an exhaustive pairwise comparison strategy tries to find all matched local feature pairs between the query image and each candidate image in the shortlist. For each pair of matched features, a matching score is given to evaluate the matching reliability. Then, CDVS adopts DISTRACT to achieve fast GCC and the matched point pairs are divided into inliers and outliers. At last, the matching score summation of inliers is used to re-rank the shortlist, and the final retrieval results are given in the sorted order.

III. Proposed Query Expansion for CDVS

In this paper, the proposed basic query expansion with is introduced into CDVS and the number of expanded re-query images is researched. Then two novel methods, including EISBI and MSBW, are proposed to address the existing problems in basic query expansion as described above.

A. Proposed Basic Query Expansion for CDVS

As show in Fig. 4, the proposed basic query expansion framework in CDVS consists of 4 modules, including the origin query, the expansion image selection, the expanded query and retrieval results fusion. The origin query is firstly conducted according to the standard CDVS retrieval framework using the origin input image as query instance, and the origin retrieval result list is returned. In the expansion image selection step, the reliable verified images are selected from origin list and their expanded descriptors, including local descriptors and global descriptors are obtained from the database. Then in expanded query step, the selected images are used to issue new queries and new expanded retrieval result lists are returned. At last, the origin result list and expanded result lists are merged together according the fusion strategy and the final sorted images are returned as the retrieval results.

Notably, compared to basic CDVS retrieval system, new problems need to be considered in basic query expansion because of the 3 new modules. Firstly, in expansion image selection step, it is an important problem that how to select expansion images and how many images should be selected, which is critical for query expansion performance improvement. Then, how to retrieve using the selected images should be seriously handle, since expanded query is expected to return the related but not already returned images. Lastly, the efficient query result lists fusion method is also worth to be further studied.

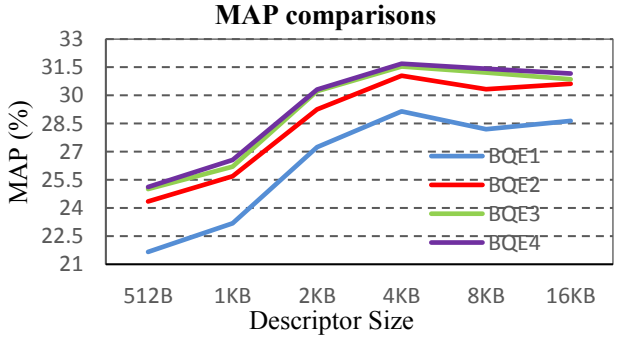

Fig. 5. MAP comparisons that different number of images are requeried in QBE

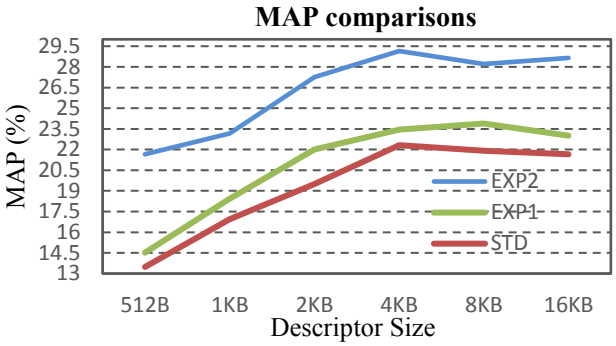

Fig. 6. MAP comparisons between different spatial verifications in expanded query

In our proposed basic query expansion, some experiments and measures are adopted to address the problems above. Firstly in expansion image selection, the top-k ranked images are selected according to the matching similarity score from the spatial verified results. Experiment results suggest that top-3 is an optimal choice. As shown in Fig. 5 (BQE1 ~ BQE4 mean the top-1 ~ 4 images are selected), more performance improvement is obtained by selecting more highly ranked images for re-query, but the MAP gain will be negligible when lower ranked image (4th or lower) is used for query expansion. Some more reliable expansion image selection methods will be discussed in part B. Secondly, the query expansion is intended to find the ignored targets in the origin query. Therefore in query expansion the candidate images must be verified against the expanded local descriptor rather than the origin local descriptor, because the ignored targets may be removed once again in spatial verification using the origin local descriptor. As shown in Fig. 6, STD means the standard CDVS retrieval, EXP1 and EXP2 stand for the query expansion using the origin local descriptor and selected expansion local descriptor. Compared to EXP2, EXP1 improves the MAP performance much less. And also a more effective strategy based on key features will be introduced into expanded query step in part C. Thirdly, when fusing the origin query result list and expanded query result lists, we resort the verified images according to the spatial verification matching score. If one target appears in multi lists, the highest score is adopted. The final resorted list is returned as result.

B. Expansion Image Selection Based on Illumination

The query expansion is an effective improvement, because new issued queries based on highly ranked verified results can obtain other related images. But as shown in Fig. 1, the origin query Q0 and the highest ranked image LB0 is quite the same, when LB0 is used to issue a new query Q1 in query expansion, all the returned images in LC are contained in LB. The expanded re-query with Q1 gets no any related results and makes no sense. But the positive related images (LA4 and LA5) are captured in quite low illumination, they fail to be retrieved by both Q0 and the top 3 highly ranked images (LB0, LB1 and LB2) in the basic query expansion. Besides, an incorrectly verified image (LB1) is in high rank or even the first rank. Experiments show, there are approximately 9% unrelated images sorted in the first rank and about 11% of the top 3 ranked results are unrelated in large scale image retrieval, the

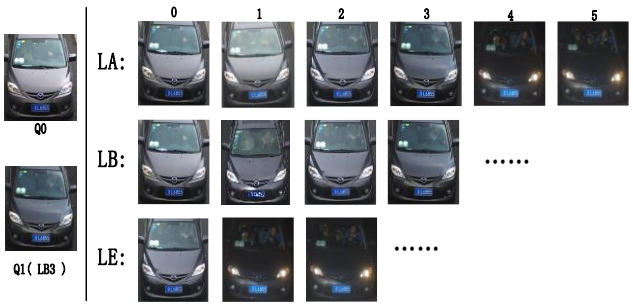

Fig. 7. Improvement of proposed expansion image selection method

numbers even reach 17% and 20% in 512B mode. As seen, the query expansion using unrelated image (Q2, also LB1) will find other unrelated results (LD0 ~ LD2).

To address this problem, the verified results with matching score greater than a given threshold are taken into consideration in query expansion. Then a pairwise strong spatial verification is applied to remove the unrelated retrieved image and the passed images form a candidate image list. And finally the candidate verified image with biggest illumination difference from origin query image is used to re-query. Fig. 1 as an example, LB3 is selected for query expansion. As shown in Fig. 7, in which the new query is issued using LB3 with the lowest IM value, the results of query expansion are listed in LE. The related image LA4 and LA5 are returned in LE by LB3. In practice, during the CDVS bitstream extraction for each image, an illumination measure IM is calculated based on the patches around the selected feature points, the IM is treated as an expression of the image illumination. In the query expansion, the image with the biggest IM value difference with origin query image are selected from the candidate list. The new query is issued by the selected candidate image. The expansion image selection is processed as follow.

For each feature, FIM acts as a measure to feature illumination according to (1). Y_i is the Y component ,also called gray value, of one pixel in the feature point patch (a 5x5 square centered at the feature point in our experiment), and N_{patch} is the pixel number in the patch.

$$FIM = \frac{1}{N_{patch}} \sum_{i}^{N_{patch}} Y_i \quad (1)$$

Eq. (2) averages the FIM of all selected local feature to get IM to measure illumination of the whole image. The $N_{features}$ means the number of selected features in Local Feature Selection.

$$IM = \frac{1}{N_{features}} \sum_{i}^{N_{features}} FIM_i \quad (2)$$

In CDVS bitstream extraction, the IM (0~255) is quantized into 32 levels as (3), and only 5 bits are needed to encode the illumination, which can achieve an equivalent performance with an 8-bits IM.

$$IM = \ IM \gg 3 \quad (3)$$

In query expansion, the originally verified images with matching score exceeding TH_{score} form the candidate re-query image set M_{can}. In (4), M_p is the verified image with origin query image and MS_p is the matching score for M_p.

$$M_{can} = \{M_p | MS_p > TH_{score} \ and \ M_p \ is \ verified\} \quad (4)$$

For each image in M_{can}, matching and spatial verification are applied with each other image in M_{can} during the pairwise strong spatial verification, and the number of matched images is defined as MC. Then the images with MC less than TH_{MC} (half of the image number in M_{can}) are removed from M_{can} as (5).

$$M_{can} = \{M_p | MC_p \geq TH_{MC} \ and \ M_p \in M_{can}\} \quad (5)$$

The image that has largest IM value difference with origin query image is selected from M_{can} for query expansion as (6), in which the IM_p means IM of an image in M_{can}, and IM_{Q_0} corresponds to the origin query image.

$$p = argmax\left(abs(IM_p, IM_{Q_0})\right) \quad (6)$$

The selected image IM_p is considered reliably related with origin query image, and also more capable to be matched with the related images that are not retrieved by the origin query. Therefore IM_p is a reasonable choice to issue new query in query expansion.

Fig. 8. An illustration of matching score based weight strategy

C. Matching Score Based Weight Strategy

The feature selection is an essential technology in descriptor extraction. The self-matching score based feature selection method [11] achieves a better performance compared to the relevance measure in CDVS. Inspired by the idea of [11], we proposed the matching score based weight strategy in query expansion. As described in [11], our experiment shows that the matched features are more likely to be matched with features in other target image, which means the matched features are more discriminative. As shown in Fig. 8, the related image B is returned by the origin query image A at high rank and used to issue a new basic query expansion. Another two images C and D are matched with image B, in which image C is related and image D is unrelated. There are totally 55 feature points in image B. In matching image A and with image B, 34 pairs of features are matched up (show as yellow circles and matching lines) and the remaining 21 green feature points in image B are not matched with any features point in image A. The 34 matched features in image B are called KFS (key feature set).

Fig. 9. Example Images in Vehicle Dataset

In the basic query expansion, 25 features in image B are matched with features in related image C, and 18 of the matched features are contained in KFS. But in matching B and the unrelated image D, only 2 out of 10 matched features in B belong to the KFS. Therefore it is a reasonable conclusion that the matched key features in re-query image are more likely to be matched in query expansion with the positive results, and in contrast the unmatched features are less discriminative and hard to be matched. The key feature matching score is used to measure the feature discrimination in expansion query.

Firstly, all the features in re-query image are weighted as W_{f_ori} (=1.0). Then in the query expansion, the weights of features in re-query image are adjusted according to the feature matching score with the origin image. Taking the top right image B in Fig. 8 as example, the features in KFS (show as yellow circle) are considered more distinctive and large weight is given to features in KFS. The adjustment is operated as (7) followed by the normalization (8).

$$W_f = W_{f\_ori} + MS_f \tag{7}$$

$$W_f = \frac{W_f \times N_{feature}}{\sum_f^{N_{feature}} W_f} \tag{8}$$

$N_{feature}$ is the number of features in re-query image. MS_f is the matching score (ranging from 0 to 1.0) of a feature in re-query image and its weight is given by W_f. Specially, MS_f of feature that is not included in KFS is assigned with 0.

Then in the query expansion, the matching score of the re-query image and candidate image is calculated according to (9) using the adjusted weight. The MS is the matching score between two images, $N_{feature}$ is the number of matched inliers.

$$MS = \sum_f^{N_{feature}} MS_f * W_f \tag{9}$$

Finally, the weighted matching scores MS of all the candidates in the shortlist are used to re-rank and merge the newly verified results into the origin retrieval images.

IV. EXPERMENTS

A. Experiments Sets And Evaluation Method

We evaluate the basic query expansion in CDVS and our proposed two methods on the Oxford 5K dataset [13] and a vehicle dataset captured on reality traffic road. The retrieval performance is expressed by MAP.

The Oxford dataset contains 5062 high resolution building images (1024 x 768) collected from Flickr by searching for particular Oxford landmarks. There are 11 different landmarks and each is represented by 5 queries. There are total 55 query images and the remaining 5007 images are database images.

Besides, we collect a set of traffic images from actual road video cameras and 13813 vehicles are detected and intercepted, including 1519 kinds of different makes or models. Each query image has at least 5 ground truths. In the vehicle dataset, there are 1519 query images and 12294 database images. The traffic images are captured both in the daytime and the nighttime. The images depicting the same vehicle may be quite different in illumination. Some example images in the dataset are shown in Fig. 9, each row depicts the same vehicle.

The MAP is calculated by averaging the Average Precision (AP) of all queries. For each query, AP means the area under the precision-recall curve, in which the precision is the retrieved related images number out of all retrieved images number and the recall is the number of retrieved related images to the number of all related images. The proposed method is

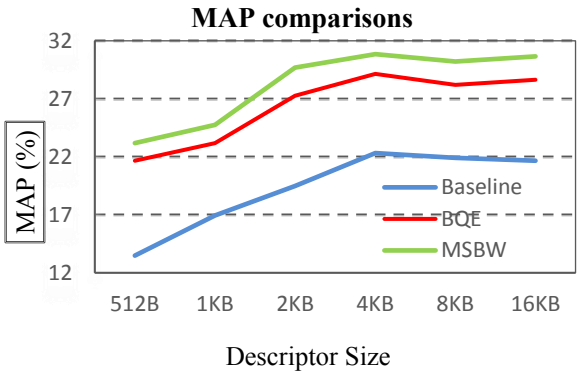

Fig. 10. MAP comparisons of the proposed methods in Oxford dataset

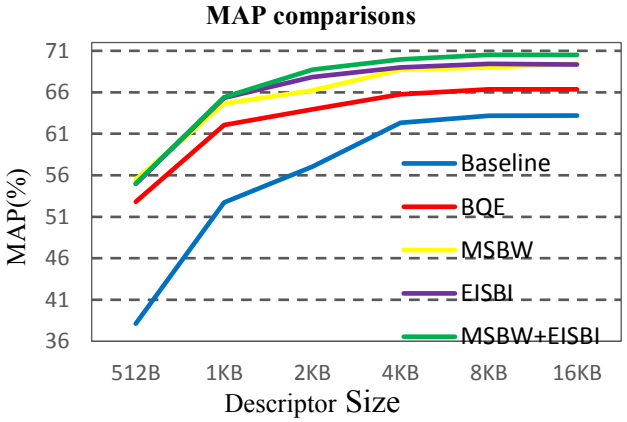

Fig. 11. MAP comparisons that different number of images are re-queried in QBE

TABLE I. MAP PERFORMANCE COMPARISONS

BR	MAP (%)		
	Baseline	BQE	MSBW
512B	13.48	21.66	23.18
1KB	16.95	23.18	24.74
2KB	19.49	27.25	29.69
4KB	22.32	29.15	30.85
8KB	21.91	28.20	30.23
16KB	21.66	28.46	30.66

TABLE II. MAP PERFORMANCE COMPARISONS

BR	MAP (%)				
	Baseline	BQE	MSBW	EISBI	MSBW+EISBI
512B	38.13	52.81	55.50	54.97	55.01
1KB	52.71	62.05	64.64	65.32	65.42
2KB	57.05	63.96	66.24	67.85	68.72
4KB	62.32	65.77	68.72	69.01	69.98
8KB	63.17	66.36	68.72	69.43	70.52
16KB	63.20	66.36	69.47	69.34	70.49

integrated into TM 11.0 [14] (CDVS Reference Software Test Model Framework), and the retrieval performance is treated as Baseline.

B. Experiments Results

a) Oxford dataset: The basic query expansion using the top ranked verified image and matching score based weight strategy are tested on the Oxford dataset. The MAP performance of BQE and MSBW compared to Baseline are shown in Table I and plotted in Fig. 10. We can see that at least 6.5% performance improvement is achieved by introducing the query expansion into CDVS. Especially in low bit rate mode (512B ~ 2KB), there is almost 9% MAP gain. And the MSBW improves MAP by another 1.5% ~2.5% compared to BQE.

b) Vehicle dataset: Vehicle dataset is collected with huge illumination changes. Except for BQE and MSBW, the expansion image selection based on illumination is also compared in the vehicle dataset.

The MAP and performance comparisons are summarized in Table II and shown in Fig. 11. For BQE, the performance is improved by 3% ~ 15% MAP increase for different descriptor size. MSWB achieves another 2% ~ 3.5% MAP gains compared to BQE. EISBI behaves slightly better than MSWB when the descriptor size is between 1KB and 8KB. Also the two proposed methods can be activated together into BQE for CDVS, which will introduce 7% ~ 17% MAP improvement compared to the basic CDVS image retrieval.

V. CONCLUSION

The CDVS image retrieval system suffers a performance drawback in some complex application scenes. In this paper, firstly the proposed basic query expansion is introduced into CDVS. Secondly, a pairwise strong spatial verification is applied to make sure of the related verified images, and we evaluate the illumination of image by the IM value, then an expansion image selection method based on image IM value is introduced to solve the problem of illumination change. Thirdly, we define the key features based on the assumption that matched features are more distinctive, then a key feature matching score based weight strategy is adopted, in which larger weights are given to the matched key features in expanded images based on the matching scores. The weight is used to calculate image matching score. The proposed methods are integrated into CDVS TM11.0 and experiments show that in Oxford dataset, We boost the retrieval performance by 6.5% ~ 9% MAP gain. In vehicle dataset, total 7% ~ 17% MAP improvement is achieved.

REFERENCES

[1] Duan L Y, Chandrasekhar V, Chen J, et al. Overview of the MPEG-CDVS Standard[J]. Image Processing, IEEE Transactions on, 2016, 25(1): 179-194

[2] Qin D, Gammeter S, Bossard L, et al. Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors[C]. Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011: 777-784.

[3] Philbin J, Zisserman A. Object mining using a matching graph on very large image collections[C]. Computer Vision, Graphics & Image Processing, 2008. ICVGIP'08. Sixth Indian Conference on. IEEE, 2008: 738-745.

[4] Zhang S, Yang M, Cour T, et al. Query specific rank fusion for image retrieval[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2015, 37(4): 803-815.

[5] Chen Y, Li X, Dick A, et al. Boosting object retrieval with group queries[J]. Signal Processing Letters, IEEE, 2012, 19(11): 765-768.

[6] Chum O, Philbin J, Sivic J, et al. Total recall: Automatic query expansion with a generative feature model for object retrieval[C]. Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. IEEE, 2007: 1-8.

[7] Chum O, Mikulik A, Perdoch M, et al. Total recall II: Query expansion revisited[C]. Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011: 889-896.

[8] Arandjelović R, Zisserman A. Three things everyone should know to improve object retrieval[C]. Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012: 2911-2918.

[9] Liu H, Xu X, Uchiyama H, et al. Query expansion with pairwise learning in object retrieval challenge[C]. Frontiers of Computer Vision (FCV), 2015 21st Korea-Japan Joint Workshop on. IEEE, 2015: 1-5.

[10] Xie H, Zhang Y, Tan J, et al. Contextual query expansion for image retrieval[J]. Multimedia, IEEE Transactions on, 2014, 16(4): 1104-1114.

[11] Xin X, Li Z, Ma Z, et al. Robust feature selection with self-matching score[C]. Image Processing (ICIP), 2013 20th IEEE International Conference on. IEEE, 2013: 4363-4366.

[12] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60(2): 91-110.

[13] Philbin J, Chum O, Isard M, et al. Object retrieval with large vocabularies and fast spatial matching[C]. Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. IEEE, 2007: 1-8.

[14] Test Model 11: Evaluation framework for Compact Descriptor for Visual Search, document ISO/IECJTC1/SC29/WG11/N14680, Jul. 2014