ON THE INTEROPERABILITY OF LOCAL DESCRIPTORS COMPRESSION

Jie Chen, Ling-Yu Duan ⋆*, Jie Lin, Rongrong Ji, Tiejun Huang, and Wen Gao*

The Institute of Digital Media, School of CS & EE, Peking University, Beijing 100871, China
e-mail:{cjie, lingyu, jielin, rrji, tjhuang, wgao}@pku.edu.cn

ABSTRACT

There are a number of component technologies that are useful for visual search, including format of visual descriptors, descriptor extraction process, as well as indexing, and matching algorithms. As a minimum, the format of descriptors as well as parts of their extraction process should be defined to ensure interoperability. In this paper, we study the problem of *interoperability* among compressed local descriptors at different bit-rates; that is, allowing effective and efficient comparison of compact descriptors, which is fundamentally important to mobile visual search applications. We propose to combine feature transform and multi-stage vector quantization to implement the interoperability of compact local descriptors. First, an orthogonal transform (e.g. Principle component analysis, PCA) is employed to eliminate the correlation between local feature dimensions, which improves the performance of compressed domain descriptor matching with the well-aligned distance computing of sorted important features in transform space. Second, a multi-stage vector quantization (MSVQ) is applied to generate compact codes for local descriptors. At light quantization tables, MSVQ takes advantage of the transform domain features to properly allocate different budgets to each group of transformed feature dimensions, respectively. The interoperability between compressed descriptors at different bit rates can be achieved by the descriptors' fast matching in the orthogonal feature space. In other words, descriptor decoding into the original feature space (SIFT space) is unnecessary, as the distance can be calculated by pre-computed lookup tables. In particular, such efficient matching in transform domain is significant for large-scale visual search. Over a set of benchmark datasets, we have reported superior performance over state-of-the-arts.

1. INTRODUCTION

With the ever growing computational power on mobile devices, recent works have proposed to directly extract compact yet discriminative visual descriptors [2][3][4][9] for low bit rate visual search. Sending compact descriptors may greatly reduce the latency from delivering visual queries in unstable wireless environments. Previous works have focused on either compact global descriptors like [2][9] or compact local descriptors , e.g., CHoG [3] [4]. On the other hand, compact descriptors can significantly reduce the storage of local features at the server end (for geometric consistency check in re-ranking) when dealing with large-scale visual search.

Problem. In wireless environments, descriptor compression at the client end should adapt to the constraints of bandwidth. The descriptors would be generated at different operating points (upper bounds of average descriptor lengths). In addition, descriptor compression may accelerate the matching and retrieval operations from directly operating compressed descriptors, either from a mobile end to a server end, or between mobile ends. Descriptors generated at any one of these operating point should allow retrieval and matching operations with descriptors generated at different operating points. Hence, the issue of *interoperability* arises, aiming to address the challenges of compact visual descriptors, including compactness, scalable length, low complexity, as well as fast comparison or other operations between descriptors at different budgets. In particular, this has also motivated the recent endeavors of the Compact Descriptor for Visual Search (CDVS) MPEG standardization [14][15][16][17][18].

⋆ Corresponding Author

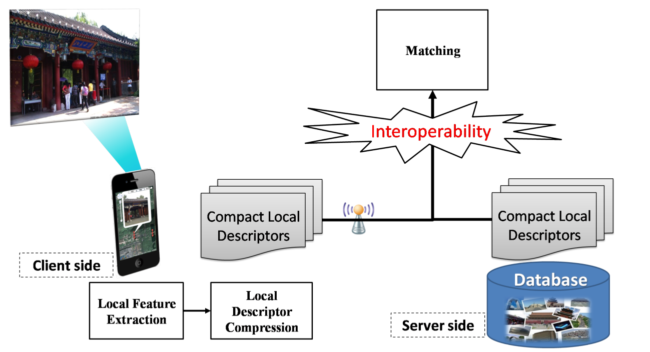

Fig. 1. The proposed scalable local descriptor compression to meet the requirement of interoperability for mobile visual search.

Inspiration. Few existing works in mobile visual search attempt the interoperability issue, which has, however, turned out to be one of key focuses in the ongoing MPEG CDVS standard. Several MPEG proposals have targeted at handling the interoperability from the perspectives of both vector quantization and scalar quantization. For instance, the recent work in [17] introduced an ad-hoc approach to select discriminative feature dimensions in the transformed domain, upon which scalar quantization followed by entropy coding is leveraged for local descriptors compression. Another recent work in [18] performed vector quantization over the raw local descriptors with tree-structured vector quantization. In essential, these works relate to the similarity comparison on features extracted from compressed domain in image retrieval [5]. Towards the interoperability, prospective research efforts aim to maximize the potentials of feature space distance metrics directly from the compressed domain.

Our Approach. In this work, we propose a novel approach to the interoperability between compressed local descriptors. The basic idea is to eliminate the dimension correlation of raw local descriptors using an orthogonal feature transform, e.g. PCA, upon which multi-stage vector quantization (MSVQ) is carried out in the orthogonal feature space. In online search, local descriptors are compressed at various budgets according to available bandwidth, which are directly compared to the compressed local descriptors at the server end, without extra decoding. Through eliminating the correlation between local feature dimensions, the subsequent compact descriptors may improve the performance of transform domain descriptor matching with the well-aligned distance computing of sorted important features in transform space. MSVQ is a memory light approach to compressing dimension reduced features; moreover, the transform domain features can be employed to properly allocate different budgets to each group of transformed dimensions. Accordingly, the interoperability of compressed descriptors at different budgets, can be accomplished by accumulating the distances of product quantized sub-vectors (groups of transformed dimensions) in the orthogonal feature space. Descriptors decoding into the original feature space (say, SIFT space) is unnecessary, as the distance computing can be performed over pre-computed lookup tables.

Related work. Below we brief previous works in compact descriptors. However, the interoperability issue has not been well studied, which differentiates our job from most existing works. The compact descriptors have been widely studied in previous literatures, for instance, reducing the dimension of local descriptors like PCA-SIFT [10], GLOH [11], SURF [1] and MSR descriptors [6], as well as compressing the image-level signatures like miniBoF [7] and Aggregated Local Features [8]. Recent works in mobile visual search [2,3,9] stepped forward to directly extract very compact descriptors at the mobile end to achieve a low bit rate query delivery. For instance, Chandrasekhar *et al.* [3] proposed a Compressed Histogram of Gradients (CHoG) descriptor, which adopts Huffman coding trees to compress a local descriptor into approximately 60 bits. An alternative is to compress the bag-of-features histogram [2,9]. For instance, Chen *et al.* [2] proposed to encode position differences of non-zero bins in bag-of-features, which reported an average length of $\sim 3KB$ per image over a vocabulary of 1 million words.

2. ON THE INTEROPERABILITY

Problem formulation. To deal with the bandwidth variation in mobile visual search, a compact descriptor is typically with variant rates. Intuitively, the scalability can be solved by sorting the local descriptors in a given image based on their "importance", upon which important descriptors are ranked more frontier and vice versa. However, this scheme may discard important local descriptors. Rather than transmitting local descriptors progressively, we prefer to compress local descriptors and the rate may be adapted to the available bandwidth. However, we need to address a key challenge of interoperability between compressed local descriptors at different rates.

Formally speaking, given two d-dimensional local descriptors x_1 and x_2, with different rate constraints R_{x_1} and R_{x_2} respectively, we study the descriptors' interoperability as follows:

$$S_{Inter}(V(R_{x_1}, D_{x_1}), V(R_{x_2}, D_{x_2})) \tag{1}$$

where $S_{Inter}(.,.)$ denotes the interoperation function of compressed local descriptors. V is the encoded representation subject to distortion D. To accomplish a better interoperability, the distortions of descriptors x_1 (or x_2), D_{x_1} (or D_{x_1}), should be minimized subject to the constraints that bitrate R_{x_1} (or R_{x_2}) cannot exceed the thresholds $R_{budget1}$ (or $R_{budget2}$).

Towards effective descriptors matching in the compressed domain, we introduce a distortion function $D(T, Q)$ and formulate a constrained optimization problem:

$$\min_{G,B} \quad D(T, Q(G, B)) \quad subject \quad to \quad R(T, Q(G, B)) \leq R_{budget} \tag{2}$$

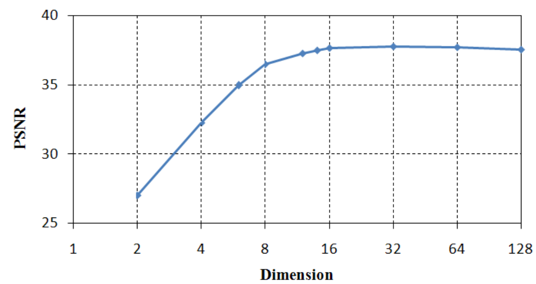

Fig. 2. The PSNR of PCA transformed descriptors versus the PCA transformed dimensions over the MIRFLICKR25000 dataset.

where T denotes a transform of individual local descriptors, Q is a quantizer. G denotes the grouping of transformed dimensions, and B the bit rate allocation in quantizing the sub-vector of each dimension group. $R(T, Q(G, B))$ is the size of a compressed descriptor. Effective local descriptor compression has to reduce the information redundancy in the original feature space. The redundancy typically refers to the uninformative feature dimensions, which are useless for image matching. Actually, the coding of uninformative dimensions would degenerate pairwise matching accuracy or search precision, which is a waste of expensive budget as well. After feature redundancy is removed, quantization (and optional entropy coding) are carried out to produce a compact local descriptor.

3. SCALABLE DESCRIPTOR CODING

We propose the scalable local descriptor compression scheme based on orthogonal feature transform and vector quantization. We aim to reduce the feature dimension correlation of original descriptor space and adapt the quantization in transformed dimensions.

3.1. Orthogonal Feature Transform

While there is a wide variety of transforms available, we adopt Principal Component Analysis (PCA) as an exemplar implementation [1]. PCA is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The first principal component has the largest variance, and each succeeding component has the largest variance possible under the orthogonal constraint. The eigenvectors associated with the energetic eigenvalues of the empirical vector covariance matrix are used to define a matrix $T \in \mathcal{R}^{d \times d}$, transforming a vector $x \in \mathcal{R}^d$ as $x' = Tx$.

Due to the rate constraint R_{budget}, it is expensive to encode all the elements of x'. The intuition is that the important elements ranked higher should be adopted to code when budget is less sufficient. Assume that we select the first $d'(d' \leq d)$ dimensions, the remaining elements would cause an information loss $\varepsilon_T = \| x' - x'(d') \|_2^2$. To meet the requirement of compact descriptors, the vector $x'(d')$ is subsequently encoded with vector quantization.

3.2. Multi-Stage Vector Quantization

We adopt a multi-stage vector quantization to compress each group of transformed dimensions. Comparing to scalar quantization, vector quantization merits in high compression rate as well as the ability to recover the original signal. So we employ vector quantization.

Given a transformed feature x', a partition is given to divide the d-dimensional transformed dimensions into S groups

[1] In principle, any orthogonal transform can be applied in our case, which is indeed not the key focus of our scheme.

$G = \{g_1, ..., g_S\}$ consecutively. For instance, the partition setting of 32, 8, 8, 16, 32 and 32 dimensions for a transformed SIFT descriptor ($d = 128$), are setup to group the most important yet orthogonal dimensions. Subsequently, a set of multi-stage vector quantizers $Q = \{Q_1, ..., Q_S\}$ with bit allocations $B = \{b_1, ..., b_S\}$ is applied within each partition. This is done by training a set of codebooks $\{C_1, ..., C_S\}$ over the transformed dimension groups. Taking the setting of 32, 8, 8, 16, 32 and 32 dimension partition, C_1 corresponds to the quantizer using the first 32 dimensions, C_2 the quantizer using the subsequent 8 dimensions, and so on. Regarding to each Q_s, its codebook C_s is designed in two phrases, namely, "Tree Structure Quantizer" and "Residual Quantizer".

Tree Structure Quantizer. The training of the first-stage quantizer may resort to the state-of-the-art visual vocabulary techniques, such as [13][12] and their variances. In this work, we adopt Hierarchical K-Means clustering to build the initial codebook with dimension M_{1st}. Given the s-th group of transformed dimensions $x'(g_s)$, we quantize the group of transformed dimensions into the nearest centroid w_j ($j \in [1, M_{1st}]$).

Residual Quantizer. In the second stage, we adopt a product quantizer to further quantize the residuals resulting from the codebook at the first stage. More specifically, given the transformed descriptor $x'(g_s)$ and its corresponding quantization vector in the first stage w_j, a residual vector $r(x'(g_s), w_j)$ is then formed as $r(x'(g_s), w_j) = x'(g_s) - w_j$. The product quantizer works as below:

$$q_1(r_1(x'(g_s), w_j)), ...q_{M_{2nd}}(r_{M_{2nd}}(x'(g_s), w_j)) \tag{3}$$

where q_i ($i \in [1, M_{2nd}]$) is the i-th quantizer with codebook size W_i to encode the i-th subvector of residuals r_i. At last, the bit allocation for the quantizer Q_s is denotes as $b_s = \log(M_{1st} + \sum_{i=1}^{M_{2nd}} W_i)$.

Suppose that $x'(d')$ consists of $S'(S' \leq S)$ dimension groups, the resulting quantization error can be obtained by $\varepsilon_Q = \sum_{s=1}^{S'} \| x'(g_s) - Q_s(x'(g_s)) \|_2^2 + \sum_{s=S'}^{S} \| x'(g_s) \|_2^2$.

3.3. The Optimization of G and B

Given a budget constraint R_{budget}, we need to consider the problem of joint optimizing the transformed dimension selection (grouping) and vector quantization. The expected error (distortion) $D(T, Q)$ of a compressed descriptor is the sum of ε_T and ε_Q. The goal is to figure out the optimal transformed dimension grouping G and bit allocation B for each quantizer to minimize $D(T, Q)$. This optimization problem is intractable, as the objective resorts to the learning of G and B iteratively. In practice, we come up with a suboptimal solution. Given a budget, the mean square error $D_e(T, Q)$ is empirically measured on a training descriptor set X as:

$$D_e(T, Q) = \frac{1}{card(X)} \sum_{x \in X} \varepsilon_T + \varepsilon_Q, \tag{4}$$

where card(X) denotes the number of descriptors in X. It derives an objective to optimize the combination setting of G and B. The effectiveness has been empirically proved in our experiments (See Section 4). A good configuration of G and B is listed in Table 1. Meanwhile, referring to Figure 2, the group lengths of dimension grouping G may be justified by the empirical observation of PCA dimensions vs. PCA tranformed descriptors' PSNR as well.

3.4. Interoperation Function

Matching of compressed local descriptors at various bit rates is performed directly in compressed domain. Based on the interoperation function, give the first and second descriptors x_1 and x_2 with a bit rate constraint of d'_1 and d'_2 bits respectively, the matching is performed in the compressed domain by calculating the score in Equation 1. In this work, we use L2 distance

Table 1. Quantization Configuration

	g_1	g_2	g_3	g_4	g_5	g_6
Group Length	32	8	8	16	32	32
Bit Allocation	13	12	12	12	12	12

Table 2. Memory cost of different compression schemes

	Transform	Quantization	Entropy Coding	Total Memory
LBC_DS _SQ_AC	LBC+DS	SQ 256B	AC	256B
TPSVQ	-	TSPVQ 10.5MB	-	10.5MB
Our Approach	PCA 65KB	MSVQ 283KB	-	348KB

$$S_{Inter} = \|M(Q(x'_1(d'_1))), M(Q(x'_2(d'_2)))\|_2 \tag{5}$$

where M denotes the table look-up procedure to compute the distance based on the entry of vector quantizers. If $d'_1 \neq d'_2$, the interoperability is achieved by computing the distance of the overlapped dimension groups $S'_1 \cap S'_2$ of x'_1 and x'_2.

4. QUANTITATIVE VALIDATION

Datasets and Evaluation Protocols. We evaluate the interoperability of our proposed local descriptors compression scheme, in terms of matching/retrieval accuracy, over a group of public available benchmark datasets, referred to as the MPEG CDVS benchmark [14][15][16][17][18]:

(1) Graphics dataset [19] depicts 5 product categories including CDs, DVDs, books, text documents and business cards. There are 1,500 queries and 1,000 reference images. Query images are captured by a mobile phone under varying lighting conditions with background clutter. All images are compressed in JPEG format. 3,000 matching pairs and 30,000 non-matching pairs are involved.

(2) Painting dataset [19] contains 400 queries and 100 reference images for paintings (like history, portraits, landscapes and modern-art), including 364 matching pairs and 3,640 non-matching pairs.

(3) Frame dataset [19] contains 500 video frames, with a range of contents like movies, news reports and sports. There are 400 queries taken by a mobile phone, capturing the screen of laptop, PC and TV, which involves typical specular distortions. 400 matching pairs and 4,000 non-matching pairs are included.

(4) Landmark dataset contains 2,302 queries and 6,367 reference images from 3 benchmarks: 1). the Zurich buildings [20], with 1,115 images of 200 buildings in Zurich city, 115 as queries; 2). the Turin buildings [21] with 1,980 images of 180 landmarks in Turin city, 1,620 as queries; and 3). the PKUbench [22] with 5,574 images of 198 landmarks from PKU campus, 567 as queries. In total, 3,805 matching pairs and 48,675 non-matching pairs are involved.

(5) UKbench dataset [23] contains 2,550 objects, each with 4 images taken from different viewpoints. All the 10,200 images are indexed as reference images and used as queries. 2,550 matching pairs and 25,500 non-matching pairs are formed.

In retrieval experiments, we use a FLICKR1M dataset containing 1 million images as distractors, which is merged with reference datasets to evaluate the scalability in dealing with large-scale image collections. The MIRFLICKR25000 [24] is used as the external dataset to train vector quantization tables (codebooks).

As to the evaluation protocols, we use the mean Average Precision (mAP) to evaluate the retrieval performance. For pairwise matching, we produce the True Positive Rate (TPR) at the False Positive Rate (FPR) of less than 1%.

	41 bits	70 bits	102 bits	134 bits	202 bits
41 bits	0.84	0.84	0.84	0.84	0.84
70 bits	0.84	0.87	0.87	0.87	0.87
102 bits	0.84	0.87	0.879	0.879	0.879
134 bits	0.84	0.87	0.879	0.87	0.87
202 bits	0.84	0.87	0.879	0.87	0.863

(a)

	37 bits	61 bits	81 bits	117 bits	169 bits
37 bits	0.805	0.749	0.755	0.742	0.728
61 bits	0.75	0.847	0.84	0.83	0.828
81 bits	0.749	0.841	0.868	0.846	0.853
117 bits	0.737	0.829	0.847	0.884	0.869
169 bits	0.733	0.825	0.847	0.865	0.89

(b)

	13 bits	37 bits	49 bits	61 bits	73 bits
13 bits	0.882	0.882	0.882	0.882	0.882
37 bits	0.882	0.883	0.883	0.883	0.883
49 bits	0.882	0.883	0.884	0.884	0.884
61 bits	0.882	0.883	0.884	0.891	0.891
73 bits	0.882	0.883	0.884	0.891	0.9

(c)

Fig. 3. The pairwise matching performance over the UKbench dataset for the compressed descriptors at different rates in different methods: a). Baseline(1), b). Baseline(2), c). Baseline(5)

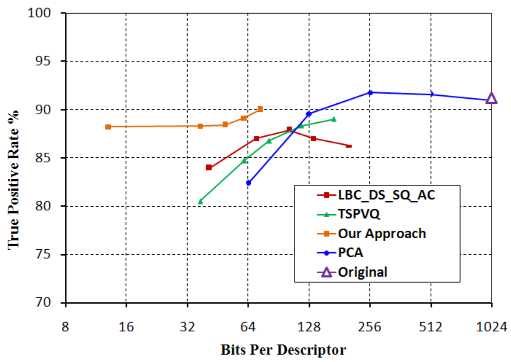

Fig. 4. TPR (at less than 1% FPR) of different methods over UK-Bench dataset in pair-wise matching experiments.

Empirical settings. We use SIFT descriptors as the original (uncompressed) local descriptors. To avoid the influence of local descriptor number on performance, we fix the setting of 300 local descriptors per image. We first apply the PCA transform to project each SIFT feature into a 128-dimensional orthogonal space. Then we segment the transformed features into six consecutively groups with dimensions 32, 8, 8, 16, 32, and 32 respectively, each of which is quantized separately. Table 1 lists the details of bits allocation.

Baselines. We compare five baselines: (1) Linear Bin Combination + Dimension Selection + Scalar Quantization + Arithmetic Coding (*LBC_DS_SQ_AC*, referred to as the H-Mode in MPEG [17]); (2) Tree Structured Product Vector Quantization (TSPVQ) without any orthogonal transform (referred to as the S-Mode in MPEG [18]); (3) PCA with different dimensions of 8,16,32,64,128 without quantization; (4) Original SIFT descriptor (without compression); (5) PCA + Grouping transformed dimensions + Vector Quantization, which is our proposed approach.

Quantitative Comparisons. Figure 4 compares the TPR rate distortions of baselines at different descriptor lengths. We can see that our approach has achieved promising results at much lower bit rates. In practice, given a budget, our approach can accommodate more compressed local descriptors to form a compact descriptor, which could significantly contribute to the pairwise matching. Figure 3 illustrates the performance of inter-operating points matching in the cases of Baseline (1), Baseline (2) and Baseline (5) (ours). Clearly, baseline (2) is with poor interoperability. In Baseline (2), the feature transform is not applied; at different descriptor lengths, separate codebooks are trained to quantize raw descriptors, so that the performance would drop due to the similarity measure across different codebooks. Distinct from Baseline (2), Baselines (1) and (5) employ feature transform, so that more informative feature dimensions are selected for better quantization, so that the loss of similarity measure accuracy from asymmetric quantization based on separate codebooks (no orthogonal transform) in Baseline (2) can be reduced.

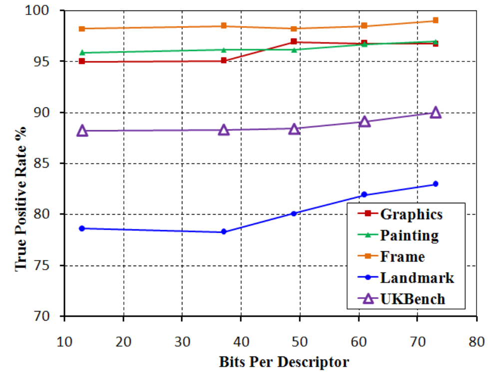

Fig. 5. TPR (at less than 1% FPR) over different datasets in pair-wise image matching experiment.

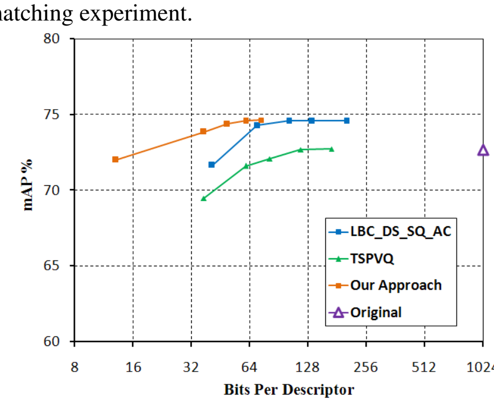

Fig. 6. Retrieval mAP versus bit rates over UKBench dataset.

Comparing to the ad-hoc transform (Baseline 1), our approach employs an orthogonal transform, yielding better interoperability, even at very low bit rates. The matching performance of Baseline (5) outperforms Baseline (1), especially at very low bit rates, which originates from the advantages of vector quantization over scalar quantization. With transform, Figure 5 shows we have achieved consistent promising results including very low bit rates over different datasets. Moreover, our approach (Baseline 5) has achieved the best retrieval performance over the challenging UKBench dataset together with 1 million distractor at lower rates as shown in Figure 6. Due to limited space, the results of all datasets are not listed.

Complexity Analysis. As listd in Table 2, our approach employs MSVQ, which has significantly reduced the size of quantization tables, comparing with the traditional product quantization (Basedline 2). A small memory footprint is important for mobile devices. Compared with Baseline (1), our approach has obtained good performance gains but at the cost of a larger quantization table.

5. CONCLUSIONS

We have attempted to address the novel issue of interoperability in visual search. The interoperability among compressed local descriptors enables effective and efficient comparison of compact descriptors. In particular, the fast descriptor matching directly in transform domain is promising in dealing with large-scale visual search. Future work may involve how to improve the retrieval performance in the context of social network[25].

6. ACKNOWLEDGEMENT

This work was supported in part by the National Basic Research Program of China (2009CB320902), the Chinese Natural Science Foundation under Contract No. 61271311 and No. 61121002, and in part by the Research Fund of ZTE Corporation.

7. REFERENCES

[1] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. *ECCV*. 2006. 2

[2] D. Chen, S. Tsai, and V. Chandrasekhar, *et al.* Tree histogram coding for mobile image matching. *DCC*, 2009. 1, 2

[3] V. Chandrasekhar, G. Takacs, D. Chen, S. Tsai, R. Grzeszczuk, and B. Girod. CHoG: Compressed histogram of gradients a low bit-rate feature descriptor. *CVPR*, 2009. 1, 2

[4] V. Chandrasekhar, G. Takacs, D. Chen, *et al.* Transform coding of image feature descriptors. *VCIP*, 2009. 1

[5] Shih-Fu Chang. Compressed-Domain Content-Based Image and Video Retrieval. *Multimedia Communications and Video Coding*, Oct. 1995. 1

[6] G. Hua, M. Brown, and S. Winder. Discriminant embedding for local image descriptors. *ICCV*, 2007. 2

[7] H. Jgou, M. Douze and C. Schmid. Packing bag-of-features *ICCV*, 2009. 2

[8] H. Jegou, M. Douze, C. Schmid, P. Perez. Aggregating local descriptors into a compact image representation. *CVPR*, 2010. 2

[9] R. Ji, L.-Y. Duan, J. Chen, H. Yao, J. Yuan, Y. Rui, and W. Gao. Location discriminative vocabulary coding for mobile landmark search. *IJCV*, 2011. 1, 2

[10] Y. Ke and R. Sukthankar. PCA-SIFT: A more distinctive rep. for local image descriptors. *CVPR*, 2004. 2

[11] K. Mikolajczyk and C. Schmid. Performance evaluation of local descriptors. *PAMI*, 2005. 2

[12] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. *CVPR*, 2006. 3

[13] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. *ICCV*, 2003. 3

[14] ISO/IEC JTC1/SC29/WG11/N12201 Call for Proposals for Compact Descriptors for Visual Search. 2011. 1, 3

[15] ISO/IEC JTC1/SC29/WG11/N12202 Evaluation Framework for Compact Descriptors for Visual Search. 2011. 1, 3

[16] ISO/IEC JTC1/SC29/WG11/N12735 Description of Core Experiments on Compact descriptors for Visual Search. 2012. 1, 3

[17] ISO/IEC JTC1/SC29/WG11/M25929 CDVS CE2: Local Descriptor Compression Proposal. 2012. 1, 3, 4

[18] ISO/IEC JTC1/SC29/WG11/N12929 Test Model 3: Compact Descriptors for Visual Search. 2012. 1, 3, 4

[19] http://mars01.stanford.edu/mvs 3

[20] http://www.vision.ee.ethz.ch/datasets 3

[21] http://pacific.tilab.com/download/CTurin180.zip 3

[22] http://61.148.212.146:8080/landmark/benchmark 3

[23] http://vis.uky.edu/~stewe/ukbench 3

[24] http://press.liacs.nl/mirflickr 3

[25] S. Liu, P. Cui, H. Luan, W. Zhu, S. Yang and Q. Tian. Social Visual Image Ranking for Web Image Search. Advances in Multimedia Modeling, 2013. 4