CODING MODE DECISION ALGORITHM FOR BINARY DESCRIPTOR CODING

Pedro Monteiro and João Ascenso

Instituto Superior Técnico - Instituto de Telecomunicações

ABSTRACT

In visual sensor networks, local feature descriptors can be computed at the sensing nodes, which work collaboratively on the data obtained to make an efficient visual analysis. In fact, with a minimal amount of computational effort, the detection and extraction of local features, such as binary descriptors, can provide a reliable and compact image representation. In this paper, it is proposed to extract and code binary descriptors to meet the energy and bandwidth constraints at each sensing node. The major contribution is a binary descriptor coding technique that exploits the correlation using two different coding modes: Intra, which exploits the correlation between the elements that compose a descriptor; and Inter, which exploits the correlation between descriptors of the same image. The experimental results show bitrate savings up to 35% without any impact in the performance efficiency of the image retrieval task.

***Index Terms*—**Local features, feature coding, binary descriptors, mode decision, visual sensor networks

1. INTRODUCTION

Nowadays, visual sensor networks (VSN) have emerged as an important research topic, due to its unique challenges in terms of performance, energy consumption and resource allocation [1]. Typically, a VSN can be characterized by several wireless low power sensing nodes that are battery operated, capable of acquiring visual data and performing visual processing tasks. Visual sensor networks are considered useful for several applications, such as object recognition, habitat monitoring and visual surveillance. In a VSN sensing nodes may acquire visual data, compress and transmit the pixel-level representation in bandwidth-limited networks to a central location where the visual content (either still images or videos) is analyzed. This paradigm is called compress-then-analyze (CTA) [2]. However, it requires significant energy and bandwidth resources at the VSN nodes, such as powerful video encoders and energy demanding transmission techniques to cope with the huge amount of data that is generated. Considering that the sensing nodes have several resource constraints, such as limited CPU power, memory for data storage and energy consumption, it is necessary to develop new solutions to optimize the visual data transmission. Thus, a novel paradigm has emerged, where local feature descriptors are computed at geographically distributed sensing nodes, which work collaboratively to make an efficient visual analysis. This paradigm is called analyze-then-compress (ATC) and reverts the order by which the compression and analysis tasks are done in the CTA paradigm [2]. In ATC, visual analysis tasks are performed on a succinct representation of the image to avoid the coding and transmission of the pixel-level representation as in CTA; thus, the VSN bandwidth and energy requirements can be met. In the ATC paradigm, data compression of local descriptors is performed as in the recent field of mobile visual search [3], where local features are compressed at mobile terminals to assure robust visual analysis under bandwidth and low latency constraints. In VSNs, the amount of data (local features) to transmit must be as low as possible which is vital to keep the energy demand of the sensing nodes low (radio can be kept inactive when no transmission occurs).

In the literature, several state-of-the-art visual descriptors are available. Two of them, the Scale-Invariant Feature Transform (SIFT) [4] and the Speeded Up Robust Features (SURF) [5] are very powerful but have rather high bandwidth and complexity requirements [6]. To address these concerns, binary descriptors have been proposed as an important alternative to SIFT and SURF descriptors, since they require low computational effort and, therefore, low energy consumption. To obtain a low bitrate representation of binary descriptors, the statistical correlation between descriptors extracted from the same image (Inter) and the statistical correlation between the descriptor elements (Intra) should be exploited. In this paper, a binary descriptor coding solution with two coding modes is proposed: an Intra mode, where each descriptor is coded individually without any reference to other descriptors, and an Inter mode, where each descriptor is differentially predicted from other descriptors (spatial correlation). While for the Intra mode, an order between descriptor elements must be found to maximize their correlation, for the Inter mode, it is necessary to find a prediction path which corresponds to the reference descriptors (Intra or Inter) used to code each Inter descriptor. The proposed solution is able to reduce the bitrate needed for transmission by selecting the 'best' coding mode for each descriptor while maintaining the same overall accuracy. The rest of this paper is organized as follows: in Section 2, related work is reviewed. After, Sections 3 and 4 propose the novel Intra coding mode and descriptor coding framework with mode decision and Section 5 presents and discusses the experimental results. Finally, Section 6 concludes the paper.

2. REVIEWING STATE-OF-THE-ART

The main objective of local features is to provide a representation of the image that is invariant to occlusions, scaling, orientation, deformations (e.g. affine) and illumination changes. The local feature extraction pipeline consists in the following steps: i) detection, where a set of distinctive key-

The project GreenEyes acknowledges the financial support of the Future Emerging Technologies (FET) program within the Seventh Framework Program for Research of the European Commission, under FET-Open grant number: 296676.

points is found and a region (patch) associated around each keypoint is established, ii) extraction, where a descriptor is built using the pixel-level patch data. One popular choice for feature extraction are the SIFT and SURF descriptors. SIFT relies on the calculation of a set of gradient orientation histograms localized in different parts of the descriptor patch and SURF uses Haar wavelet filters over a predefined set of sample points. However, the recent binary descriptors Binary Robust Independent Elementary Features (BRIEF) [7], Binary Robust Invariant Scalable Keypoint (BRISK) [8] and Fast Retina Keypoint (FREAK) [9] require much less computation and storage resources. These binary descriptors rely on a set of difference intensity tests (sampling pattern) over a smoothed patch and are very fast to extract and match, thus suited for resource-constrained applications. The output values of the intensity tests (0/1) correspond to the descriptor elements that are aggregated into a single descriptor vector. In BRIEF [7], a pre-defined sampling pattern following a bi-dimensional Gaussian distribution is used while for BRISK [8] a sampling pattern with concentric circles is adopted; the FREAK descriptor [9] is inspired by the human visual system characteristics using a retinal sampling pattern and a training procedure to find the best intensity tests. Also, BRISK and FREAK have enhanced BRIEF by being invariant to rotation and scale, thus improving the overall efficiency.

In addition, several techniques to code real-value and binary descriptors are available in the literature. In [10], a coding scheme to compress SIFT-like gradient orientation histograms is proposed. These histograms are calculated over spatial and gradient bins, quantized, and entropy coded. For Intra binary descriptor coding, [11] proposes a method to find a permutation of the intensity tests according to their correlation and a method to find a subset of the descriptors elements that maximizes the descriptor discriminative power while also achieving bitrate savings. Techniques to code binary descriptors by exploiting the Inter correlation between descriptors were also proposed, first using a simple prediction (XOR) based scheme [12] and then with a more advanced clustering solution [13]. However, it is not available a contribution that jointly exploits both Intra-descriptor and Inter-descriptor correlation. Thus, inspired by predictive video coding, this paper proposes a method to perform Intra/Inter coding mode decision along with a novel Intra coding algorithm (inspired by [11]) that is described next.

3. INTRA CODING MODE: OFFLINE STAGE

The sampling pattern defines a set of intensity tests (typically 512) that are performed with a predefined random or distance based order (from short to higher distances). However, the order in which intensity tests are done does not considered the correlation between the outcome of the intensity tests, the descriptor elements (DE). Thus, to exploit the correlation between DEs, a predefined intensity test order must be calculated in advance. This order must maximize the correlation between adjacent descriptor elements to predict a DE from its previous already computed DE. Also, this order must be generic, since adaptation is not possible for each descriptor (the decoder must know in advance the order used at the encoder). The solution proposed here is to learn offline, which DEs are more correlated to others, by using a high number of descriptors extracted from a varied and representative dataset. The proposed algorithm to compute the optimal order proceeds as follows:

1. **Initialization**: Using one of the state-of-the-art binary descriptor extraction algorithms (BRIEF, BRISK or FREAK), a large set of binary descriptors is extracted from a dataset of images. Thus, a matrix $D_{i,k}$ of binary descriptors is obtained: each *i-th* row contains the descriptor for the *i-th* keypoint and each *k-th* column represents the intensity test according to the sampling pattern.
2. **Initial prediction order**: Then, a prediction order is obtained, by computing the residual error (bitwise XOR represented through $\oplus$) between the DE $D_{i,k}$ and all possible DEs $D_{i,m}$ not previously selected. The idea is to find columns of the descriptor matrix D that are similar (smallest residual error), through an iterative algorithm that uses all binary descriptors obtained in the previous step. The optimal order in terms of residual error is given by:
$$m^* = \arg \min_{m \in]k,M]} \left\{ \sum_{i=1}^{N} D_{i,k} \oplus D_{i,m} \right\} \quad (1)$$
where N represents the number of descriptors extracted, M the number of descriptors elements, m^* represents the intensity test selected and k the previously selected intensity test. Equation (1) is iteratively repeated until all intensity tests are selected, i.e. for all values of k (from 1 to M). The first intensity test selected corresponds to the DE that has the highest number of 1s considering all binary descriptors; the idea is to concentrate a high amount of 1s at the beginning of each binary descriptor. Nevertheless, it was experimentally verified that this choice does not affect much the coding efficiency and other alternative solutions are possible, e.g. highest number of 0s.
3. **Final prediction order**: Then, the intensity test order previously obtained is refined. First, a pair of intensity tests is selected as in the previous step and it is evaluated if their positions should be exchanged or not. However, the evaluation criterion is not based on the residual error but rather on the entropy (an estimate of the rate) of all binary descriptors, thus leading to a set of correlated DEs that can be coded more efficiency.

Now, when the binary descriptor is built, the intensity test order established above is used; this means that the Intra coding mode can be efficiently applied. Regarding the Inter mode, each descriptor is differentially encoded (XOR based) online using other previously coded descriptors as reference since the descriptors can be reordered in any arbitrary way.

4. INTRA/INTER BINARY DESCRIPTOR CODING

The computation of low bitrate descriptors is rather important for several applications that are limited by bandwidth, storage or latency, such as mobile visual search and visual sensor networks. Thus, the objective of the proposed algorithm is to lower the data rate needed to stream binary features by selecting the best coding mode (Intra or Inter) for each extracted descriptor. Also, it is important to find prediction paths, which are defined as one Intra descriptor and a set of Inter descriptors, which are differentially coded using as reference the previously coded descriptor. The architecture of the proposed binary descriptor encoder for the sensing nodes is presented in Fig. 1.

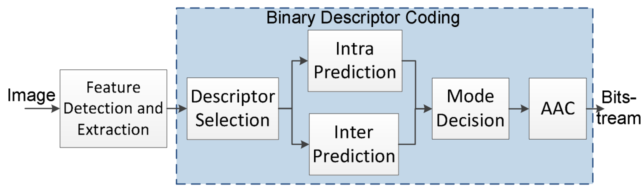

Fig. 1 - Binary descriptor encoder architecture.

As shown in Fig. 1, the descriptors are first extracted with any state-of-the-art detection and extraction methods, such as the BRIEF, BRISK and FREAK algorithms, obtaining the matrix D as previously defined. At the beginning, all descriptors can be used as reference for Inter coding. The proposed binary descriptor coding solution proceeds as follows:

1. **Descriptors Selection:** First, the pair of descriptors that has the minimum entropy $H(D_i \oplus D_k)$ is selected, where D_i, D_k represent all DEs of two distinct binary descriptors *i* and *k*. To be computationally efficient, this entropy calculation between all descriptors is computed only one time. The pair of descriptors is selected from the lowest to the highest entropy in an effort to obtain the best two descriptors (one Inter and one Intra) from a coding efficiency perspective.
2. **Intra and Inter Prediction:** Then, the following residuals are computed:

$$Z_i = D_{i,m-1} \oplus D_{i,m}, \qquad m = 1, \ldots, M \tag{2}$$

$$Z_k = D_{k,m-1} \oplus D_{k,m}, \qquad m = 1, \ldots, M \tag{3}$$

$$Z_{i-k} = D_i \oplus D_k \tag{4}$$

where Z_i, Z_k corresponds to the Intra mode residual of the *i-th* and *k-th* descriptors, respectively, and Z_{i-k} corresponds to the Inter mode residual. For Intra descriptors, a DPCM-like scheme is applied to the DEs of the binary descriptor while for Inter descriptors a DPCM-like scheme is applied between the two selected descriptors which should be well correlated.

4. **Mode Decision:** The best mode for each descriptor is evaluated based on a rate criterion. The distortion is not evaluated since the descriptors are losslessly encoded, i.e. the binary descriptors before encoding and after decoding are the same. Consider that the rate of encoding both descriptors as Intra is $R_I = H(Z_i) + H(Z_k)$, the rate of encoding the *i-th* descriptor as Intra and the *k-th* as Inter is: $R_{P1} = H(Z_i) + H(Z_{i-k})$ and the rate of encoding the *k-th* descriptor as Intra and the *i-th* as Inter is: $R_{P2} = H(Z_k) + H(Z_{i-k})$. Thus, the minimum rate among R_I, R_{P1} and R_{P2} establishes which is the best coding mode for the two selected descriptors:

$$\begin{cases} D_i, D_j \in \{\text{Intra}\} & R_I < R_{P1} \wedge R_I < R_{P2} \\ D_i \in \{\text{Intra}\}, D_k \in \{\text{Inter}\} & R_{P1} < R_I \wedge R_{P1} < R_{P2} \\ D_k \in \{\text{Intra}\}, D_i \in \{\text{Inter}\} & R_{P2} < R_I \wedge R_{P2} < R_{P1} \end{cases} \tag{5}$$

Afterwards, the algorithm goes back to step 1 where other two descriptors are selected. In the next iterations, the descriptors that can be used as reference are: all Intra descriptors and the last Inter descriptor of a prediction path. All descriptors that are Inter coded and are already used as reference cannot be selected. This rule allows the definition of linear prediction paths for which a descriptor can be used as reference just once (as an IP…P GOP in video coding with one reference frame), avoiding the creation of complex prediction trees, which require extra signaling. In addition, when a descriptor of an existing prediction path is selected, its mode cannot be changed and its rate is considered to be zero; only the Intra coding rate ($H(Z_i)$) and the Inter coding rate ($H(Z_{i-j})$) of the isolated descriptor is considered. The proposed algorithm stops when all combinations between the descriptors marked as available for reference are tested. Then, for all prediction paths, it is evaluated which one of the two terminal descriptors should be used as Intra, by considering the rate of all Intra and Inter descriptors in the prediction path.

3. **Adaptive Arithmetic Coding (AAC):** Depending on the coding mode selected, the corresponding Intra or Inter residual is entropy coded with a binary adaptive arithmetic codec. In addition, a coding mode flag (1bit) to specify the mode of each descriptor is also arithmetic coded and multiplexed with the coded residual, obtaining a bitstream that is ready to be transmitted.

5. PERFORMANCE EVALUATION

In this section, the binary descriptor coding solution proposed in Section 4 is evaluated with the intensity test order for Intra coding as defined in Section 3. Rate-efficiency results are obtained by measuring the bitrate needed to represent the compressed descriptors and by matching the decoded (query) descriptors against a database of images. To evaluate the efficiency, an image retrieval pipeline was implemented, however, other visual tasks could also be used, e.g. object detection, tracking and recognition. To perform image retrieval, pair-wise matching between the decoded descriptors from the query images and the database descriptors is first performed. Then, wrong matches between the query and the database descriptors are removed using the well-known ratio test and RANdom SAmple Consensus (RANSAC). This procedure returns a ranked list of database images given a certain query image.

5.1 Test Conditions

To obtain the intensity test order for Intra coding, a dataset of 1117 images from the Oxford Buildings Dataset [14] was used, notably All Souls, Ashmolean, Radcliffe Camera and

Christ Church. All images have a spatial resolution of 1024×768 and, for each image, 300 descriptors were extracted with the BRIEF, BRISK and FREAK algorithms. To evaluate the coding efficiency and the rate-efficiency performance, two datasets were used: ZuBuD [15] and Torino [16]. The ZuBuD dataset contains 1005 images of 201 building in the city of Zurich, whereas the Torino dataset contains 1260 images of 180 buildings of the Torino city. For the ZuBuD dataset, a separate set of query images was available and used for this work while for the Torino dataset one image from each building was used to define the query images; the remaining ones are the database images. For all images, the spatial resolution is 640×480, except for the ZuBuD query images that have a spatial resolution of 320×240.

To evaluate the proposed algorithm, BRISK and FREAK descriptors were used on both datasets, whereas the BRIEF was only used on the Torino dataset since for the ZuBuD dataset the query and database images have different spatial resolutions, and the BRIEF descriptor is not scale invariant. For the keypoints extraction, the SURF detector was always used to guarantee that the same patches are obtained independently of the descriptor extractor used. Only in this way is possible to evaluate the techniques among themselves. The SURF keypoint detector and the BRISK, FREAK and BRIEF descriptor extractor implementations correspond to the ones available in the OpenCV 2.4.5 library.

To evaluate the rate-efficiency performance, 8 operational points were defined. These rate-efficiency points represent several bitrate budgets and are defined through the number of coded descriptors N, which range from 15 to 400 descriptors. To select which descriptors are coded, the extracted descriptors are first sorted based on the Hessian response, a reliable metric widely used in the past [6]; then, the top N descriptors are selected for coding. The success (efficiency) of the image retrieval task is evaluated through the commonly used mean average precision (MAP) metric, described in [6]. MAP is calculated by taking the mean of average precision (AP), where AP is calculated for each query by averaging the precision at each point a correct image is retrieved. The MAP value varies between 0 and 1 and represents an aggregate and comparable measure of the image retrieval performance. The bitrate corresponds to the average bitrate of the descriptors of all query images: 115 for the ZuBuD dataset and 180 for the Torino dataset.

5.2. Experimental Results

The rate-accuracy results for the ZuBuD and Torino datasets are presented in Table 1 for the following solutions: 1) *BDE-Intra:* all binary descriptors are Intra coded; 2) *BDE-Inter*: all binary descriptors are Inter coded; 3) *BDE-Intra/Inter*: the complete coding framework is used where the coding mode decision establishes if Intra or Inter prediction is performed; 4) *BDE-Raw*: uncompressed binary descriptors are transmitted. In Table 1, the bitrate savings ΔBDE-ΔIntra/Inter, ΔBDE-Intra and ΔBDE-Inter of the proposed coding solutions regarding the BDE-Raw (PCM coded) are presented in percentage. Note that the proposed coding solutions have no impact on the efficiency since the binary descriptors are losslessly encoded. From the experimental results shown it is possible to conclude that:

- **Influence of the number of coded descriptors:** As expected, for BDE-Inter, the bitrate savings increase when the number of coded descriptors increases since a more correlated set of descriptors is obtained, leading to efficiency improvements. Regarding BDE-Intra, the number of coded descriptors does not influence the bitrate savings, since only the correlation between DEs is exploited. For BDE-Intra/Inter, bitrate savings increase for the BRIEF descriptor, since the Inter mode is highly efficient.
- **Influence of the binary descriptor extraction algorithm:** The bitrate gains are related to the binary descriptor extraction algorithm. For example, the proposed BDE-Inter/Intra obtains savings from 26.2% to 29.4% for BRISK and savings from 25.6% up to 30.6% for FREAK, but higher bitrate savings were achieved for the BRIEF descriptor (up to 35.5%). This behavior can be explained by the different binary descriptor sampling patterns.
- **BDE-Intra versus BDE-Inter:** In most cases, the bitrate savings obtained for BDE-Intra mode are higher when compared to BDE-Inter mode, especially when few descriptors are coded. However, BDE-Inter is able to slightly outperform BDE-Intra for the BRIEF descriptor when more than 30 descriptors are coded and for the FREAK descriptor when more than 300 descriptors are coded (for the Torino dataset). This highlights that the spatial correlation between descriptors is not high and thus, the Intra mode is rather efficient.
- **BDE-Intra/Inter versus BDE-Intra and BDE-Inter:** As expected, the proposed BDE-Intra/Inter solution outperforms both BDE-Intra and BDE-Inter, thus showing that the evaluation and selection of the best coding mode for each descriptor improves significantly the rate-accuracy performance.

Furthermore, it is also possible to observe that the bitrate savings are greater for the Torino dataset when compared to the ZuBuD dataset; this shows that when the spatial resolution increases (from 320×240 to 640×480) the correlation between descriptor elements and descriptors is higher, for both FREAK and BRISK descriptors.

Table 1. Rate-efficiency performance for the FREAK, BRISK and BRIEF descriptors using the ZuBuD and Torino datasets.

Number of Descriptors	FREAK descriptor – ZuBuD dataset				
	MAP	ΔBDE-Raw [bits/img]	ΔBDE-Intra/Inter	ΔBDE-Intra	ΔBDE-Inter
15	0.11	7680	25.7	23.9	14.3
30	0.19	15360	25.6	23.5	15.8
50	0.33	25600	25.5	23.2	17.0
100	0.48	50764	25.9	23.1	19.1
150	0.54	72557	26.4	23.2	20.4
200	0.59	85428	26.7	23.2	21.0
300	0.66	89173	26.8	23.2	21.2
400	0.71	89173	26.8	23.2	21.2

Number of Descriptors	BRISK descriptor - ZuBuD dataset				
	MAP	ΔBDE-Raw [bits/img]	ΔBDE-Intra/Inter	ΔBDE-Intra	ΔBDE-Inter
15	0.18	7680	26.3	24.8	10.2
30	0.40	15360	26.3	24.6	12.4
50	0.58	25600	26.5	24.5	13.6
100	0.66	51200	26.6	24.5	15.2
150	0.71	76800	26.8	24.6	16.1
200	0.75	102044	26.9	24.7	16.7
300	0.77	145818	27.1	24.9	17.5
400	0.79	170465	27.1	24.9	17.9

Number of Descriptors	FREAK descriptor - Torino dataset				
	MAP	ΔBDE-Raw [bits/img]	ΔBDE-Intra/Inter	ΔBDE-Intra	ΔBDE-Inter
15	0.18	7680	29.1	25.9	17.1
30	0.34	15360	29.5	25.7	19.8
50	0.44	25600	30.0	25.7	21.7
100	0.60	51200	30.1	25.4	23.2
150	0.67	76800	30.0	25.2	24.0
200	0.71	102400	30.2	25.1	24.7
300	0.75	153600	30.4	25.0	25.6
400	0.78	204456	30.6	24.9	26.3

Number of Descriptors	BRISK descriptor - Torino datset				
	MAP	ΔBDE-Raw [bits/img]	ΔBDE-Intra/Inter	ΔBDE-Intra	ΔBDE-Inter
15	0.25	7680	27.9	25.7	12.1
30	0.43	15360	28.6	25.7	14.7
50	0.54	25600	29.0	25.9	16.3
100	0.68	51200	29.3	26.0	18.1
150	0.73	76800	29.4	26.0	19.0
200	0.77	102400	29.4	26.0	19.5
300	0.80	153600	29.4	26.0	20.3
400	0.82	204800	29.5	26.1	20.9

Number of Descriptors	BRIEF descriptor - Torino dataset				
	MAP	ΔBDE-Raw [bits/img]	ΔBDE-Intra/Inter	ΔBDE-Intra	ΔBDE-Inter
15	0.05	3840	32.2	25.6	21.0
30	0.09	7680	33.7	25.2	25.1
50	0.13	12800	34.5	25.0	27.3
100	0.21	25600	34.9	24.5	29.3
150	0.25	38400	35.1	24.1	30.1
200	0.28	51200	35.1	23.9	30.6
300	0.31	76800	35.3	23.5	31.3
400	0.33	102400	35.5	23.2	31.8

6. FINAL REMARKS AND FUTURE WORK

In this paper, a mode decision algorithm for coding visual features extracted from images is proposed. This algorithm selects between Intra and Inter prediction modes, which exploit the correlation between descriptor elements and between descriptors, respectively. As future work, it is proposed to remove some descriptor elements (lossy coding) after the prediction, thus further improving the descriptor coding performance, lowering the bandwidth requirements for the transmission of binary descriptors in a visual sensor network.

REFERENCES

[1] S. Soro and W. Heinzelman, “A Survey of Visual Sensor Networks,” Advances in Multimedia, vol. 2009, Article ID 640386, 2009.

[2] A. Redondi, L. Baroffio, M. Cesana, M. Tagliasacchi, “Compress-then-Analyze vs. Analyse-then-Compress: Two Paradigms for Image Analysis in Visual Sensor Networks,”, IEEE International Workshop on Multimedia Signal Processing, Pula, Italy, September 2013.

[3] B. Girod, V. Chandrasekhar, D.M. Chen, N.-M. Cheung, R. Grzeszczuk, Y. Reznik, G. Takacs, S.S. Tsai, R. Vedantham, “Mobile Visual Search,” IEEE Signal Processing Magazine, IEEE , vol. 28, no. 4, pp. 61-76, July 2011.

[4] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, vol. 60, no. 2, November 2004.

[5] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "SURF: Speeded Up Robust Features," Computer Vision and Image Understanding, vol. 110, no. 3, June 2008.

[6] A. Canclini, R. Cilla, A. Redondi, J. Ascenso, M. Cesana, M. Tagliasacchi, “Evaluation of visual feature detectors and descriptors for low-complexity devices”, IEEE/EURASIP Digital Signal Processing Conference, Santorini, Greece, 2013.

[7] M. Clonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary Robust Independent Elementary Features," in European Conference on Computer Vision, Crete, Greece, September 2010.

[8] S. Leutenegger, M. Chli, and R. Siegwart, "BRISK: Binary Robust Invariant Scalable Keypoints," in IEEE International Conference on Computer Vision, Barcelona, Spain, November 2011.

[9] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast Retina Keypoint," in IEEE Conference on Computer Vision and Pattern Recognition, Providence, Rhode Island, USA, June 2012.

[10] V. Chandrasekhar et al., "CHoG: Compressed Histogram of Gradients A Low Bit-rate Feature Descriptor," in IEEE Conference on Computer Vision and Pattern Recognition, Miami, Florida, USA, June 2009.

[11] A. Redondi, L. Baroffio, J. Ascenso, M. Cesana, and M. Tagliasacchi, "Rate-Accuracy Optimization of Binary Descriptors," in IEEE International Conference on Image Processing, Melbourne, Australia, September 2013.

[12] J. Ascenso and F. Pereira, "Lossless Compression of Binary Image Descriptors for Visual Sensor Networks," in International Conference on Digital Signal Processing, Santorini, Greece, July 2013.

[13] P. Monteiro and J. Ascenso, "Clustering Based Binary Descriptor Coding for Efficient Transmission in Visual Sensor Networks," in Picture Coding Symposium, San Jose, California, USA, December 2013.

[14] J. Philbin, R. Arandjelovic, and A. Zisserman. The Oxford Buildings Dataset, [Online], November 2012, http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/

[15] H. Shao, T. Svoboda, and L. V. Gool, "ZuBuD - Zurich Buildings Database for Image Based Recognition," Computer Vision Lab, Swiss Federal Institute of Technology, Technical Report 206, April 2003.

[16] Requirements Subgroup, "Evaluation Framework for Compact Descriptors for Visual Search," ISO/IEC JTC1/SC29/WG11/N12202, July 2011.