ON THE DESIGN OF A NOVEL JPEG QUANTIZATION TABLE FOR IMPROVED FEATURE DETECTION PERFORMANCE

Jianshu Chao, Hu Chen, and Eckehard Steinbach*

Institute for Media Technology, Technische Universität München, Munich, Germany

ABSTRACT

Keypoint or interest point detection is the first step in many computer vision algorithms. The detection performance of the state-of-the-art detectors is, however, strongly influenced by compression artifacts, especially at low bit rates. In this paper, we design a novel quantization table for the widely-used JPEG compression standard which leads to improved feature detection performance. After analyzing several popular scale-space based detectors, we propose a novel quantization table which is based on the observed impact of scale-space processing on the DCT basis functions. Experimental results show that the novel quantization table outperforms the JPEG default quantization table in terms of feature repeatability, number of correspondences, matching score, and number of correct matches.

Index Terms— JPEG, Quantization table, Feature detectors, Scale-space

1. INTRODUCTION

The detection of salient image features is a fundamental step in many computer vision applications, such as content-based image retrieval (CBIR), object recognition, mobile visual search, and so on. Typically, there are two steps in the feature extraction process. First, the salient interest points or regions are detected by a detector and then, distinctive vectors called descriptors are extracted from a patch around the keypoints. The descriptor from one image can be compared with descriptors extracted from other images by calculating a similarity score (normally Euclidean Distance). In this paper, we are interested in improving the performance of feature detectors which operate on JPEG-compressed images.

Due to limited storage or communication capacity, e.g., in digital camera or mobile visual search applications, images are typically compressed after their acquisition. Various standards for image compression have been proposed and adopted for multimedia applications, e.g., JPEG and JPEG 2000. JPEG 2000 has better compression performance, however, due to its complexity and low coding speed it is not widely used until today. Therefore JPEG is still the most widely used standard in consumer electronic products. In the computer vision community, image features are hence normally extracted from JPEG images. However, the feature extraction performance quickly degrades in the presence of JPEG compression artifacts as shown in [1]. To address this issue, feature-preserving image compression has been proposed. In our previous work [2], the macroblocks in images are compressed using variable quantization which is supported by the JPEG standard extension syntax [3]. The extension syntax is, however, not widely supported and the encoding process is computationally expensive. Therefore, a simpler method is proposed in this paper.

*This work has been supported by a PhD grant from the China Scholarship Council for Jianshu Chao.

JPEG compression is based on a default quantization table; however, it should be noted that the standard does not mandate a specific quantization table. Thus, alternative tables can be defined by the users. There are some works on designing novel quantization tables targeting different applications. [4] proposes a quantization table which is based on the human visual system, outperforming the baseline JPEG and many other schemes in terms of rate-distortion performance. An optimized quantization table scheme for individual images is proposed in [5]. In [6] and [7], the recognition performance is improved in iris recognition and face recognition systems by compressing images with novel quantization tables. The authors in [8] propose to encode canonical patches with a gradient-preserving quantization matrix, which achieves lower gradient distortion and better descriptor matching performance. [9] proposes a novel distortion measure and employs an evolutionary algorithm to find a better quantization table for feature extraction. In this paper our goal is to design a novel quantization table for improved feature detection performance for standard feature detector approaches such as Hessian-Laplacian, Harris-Laplacian [10, 11], SURF [12] and DoG [13]. The proposed approach is fully standard compatible and the encoded images can be decoded by any baseline JPEG decoder. At the same time the detection performance is improved. In addition, the computational complexity of both the encoder and decoder does not increase compared to other approaches in the literature.

The remainder of this paper is organized as follows. In Section 2, the evaluation criteria, the data set and the evaluated detectors are presented. In Section 3, we first analyze the scale selection process of each detector. Based on this analysis, the novel quantization table is proposed. Section 4 presents the results which demonstrate that the proposed quantization table leads to improved detector performance when compared to JPEG encoding with the default quantization table and the recently proposed table in [9]. Section 5 concludes the paper.

2. EVALUATION FRAMEWORK

2.1. Criteria

In the literature, various criteria have been studied for evaluating feature detectors [14, 1, 15, 16]. The criteria and the test data set provided by the authors in [1] are extensively used in many current feature evaluation works. [1] introduces the *overlap error* which is defined as the error of region-to-region ellipses from a pair of affine region detectors. The repeatability score is calculated as the ratio between the number of keypoint correspondences which have less than 40% overlap error and the smaller number of keypoints in the pair of images. Then, descriptors for each keypoint are calculated and compared based on their Euclidean distance. If two keypoints are repeatable and the descriptor distance is closest at the same time, they are deemed as a correctly matched pair. The matching score is defined as the ratio between the number of correct matches and

the smaller number of detected regions. The open source software VLBenchmarks [17] implements this evaluation framework, and is used in our experiments.

2.2. Data set

The data set provided by [1] consists of eight image sets, *graf, wall, boat, bark, bikes, trees, ubc* and *leuven*. In our experiment the first image from each set (eight in total) are used. Since features are detected normally only for the Y component, first the color images are transformed to YCbCr domain and the Y components are stored as uncompressed gray scale images. The software from the Independent JPEG Group (version 8d) [18] is used to compress the gray scale image data set. Different quantziation tables can be tested by using the parameter *-qtables*. The JPEG quality values are set to 4, 8, 12, 16 and 20 both for the default table and our proposed table, resulting in five compressed images for each test image. The quality values are set to 12, 24, 36, 48 and 60 for the proposed table in [9] in order to generate images with similar bit rates as ours. Then, the repeatability score, the number of correspondences, the matching score and the number of correct matches are calculated between the uncompressed image and the corresponding JPEG encoded images. Finally, the average scores and bit rates from eight gray images with the same quality are computed.

2.3. Detectors

Detected features should have a well-defined location in the image, be scale-invariant and be robust to viewpoint and illumination changes. Various detectors have been proposed in the literature. The authors in [1] compare several popular detectors showing that there is no detector which is superior to other detectors for all image conditions. The experiments illustrate that Hessian-Affine, Harris-Affine [19, 20] and MSER [21] detectors seem to be the top three for most types of image changes. In our experiments only JPEG compression artifacts are considered, therefore, we examine the Hessian-Laplacian and Harris-Laplacian [10, 11] detectors instead of Hessian-Affine and Harris-Affine. The executable for the Hessian-Laplacian and Harris-Laplacian detectors is provided by the authors of [1]. Besides, two additional widely-used detectors, DoG (SIFT) [13] and SURF [12], are also compared. Table 1 shows some noteworthy parameters for each detector and gives the source of the implementation that is used in this paper. The parameter *PeakThresh* 7.65 results in the same threshold value 0.03 as in the classic paper [13], and the parameter *FirstOctave* 0 yields less features in order to be comparable to other features. Other parameters are set by default in the corresponding source codes. For calculating the matching score and the number of correct matches, the descriptors of Hessian-Laplacian, Harris-Laplacian and MSER are computed using the SIFT descriptor implementation provided by [1]. The DoG and SURF implementations already output corresponding descriptors.

Table 1. Selected parameters for the studied detectors.

Detectors	Parameters	Values	Source
Hessian-Laplacian	threshold	500	Oxford [1]
Harris-Laplacian	threshold	1000	Oxford [1]
MSER	es	1	Oxford [1]
DoG	PeakThresh	7.65	VLBen.
	FirstOctave	0	[17]
SURF	threshold	1000	ETH [12]

3. QUANTIZATION TABLE DESIGN

3.1. Scale-space representation

The scale-space based detectors normally proceed as follows: 1) computation of multi-scale images, 2) local extrema detection, 3) location and scale refinement, and 4) weak keypoint removal. As a first step towards the design of a novel quantization table we compare the operators used for scale selection of the scale-space based detectors. Inspired by biological vision, the scale-space theory has been extensively studied by the computer vision community in the context of feature detection tasks [22, 11, 23, 24]. The scale-space representation is a family of smoothed images at different scales. Typically, given an image signal f, its scale-space representation L is computed by convolving the image f with the Gaussian kernel [11, 23]:

$$G(x, y; \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

leading to

$$L(x, y; \sigma) = G(x, y; \sigma) * f(x, y) \tag{2}$$

The standard deviation σ is varied to obtain multi-scale images. Based on the concept of Gaussian smoothed scale-space representation, different detectors use different convolution kernels. The operation is performed by convolving the image signal f with Gaussian derivatives in order to detect corners, edges or blobs. The produced multi-scale image family is called scale-space derivatives [23]. Then various types of linear or nonlinear combinations of scale-space derivatives are used to detect keypoints. The different detectors considered in this work perform this as follows:

Laplacian of Gaussian (Hessian-Laplacian and Harris-Laplacian)

$$\nabla^2_{norm} L = \sigma^2 (L_{xx} + L_{yy}) \tag{3}$$

Approximate determinant of the Hessian (SURF)

$$\begin{aligned} det(\mathcal{H}_{approx}) &= \widetilde{L_{xx}}\widetilde{L_{yy}} - (0.9 \cdot \widetilde{L_{xy}})^2 \\ &\approx L_{xx} L_{yy} - (L_{xy})^2 \end{aligned} \tag{4}$$

Difference-of-Gaussian (DoG)

$$\begin{aligned} DoG &= L(x, y; k\sigma) - L(x, y; \sigma) \\ &= (G(x, y; k\sigma) - G(x, y; \sigma)) * f(x, y) \\ &\approx (k-1)\nabla^2_{norm} L \end{aligned} \tag{5}$$

where

$$L_{xx} = \frac{\partial^2 G(x, y; \sigma)}{\partial x^2} * f(x, y) \tag{6}$$

$$L_{yy} = \frac{\partial^2 G(x, y; \sigma)}{\partial y^2} * f(x, y) \tag{7}$$

$$L_{xy} = \frac{\partial^2 G(x, y; \sigma)}{\partial x \partial y} * f(x, y) \tag{8}$$

and $\widetilde{L_{xx}}$, $\widetilde{L_{yy}}$, $\widetilde{L_{xy}}$ are box-filtered approximations thereof [12]. In brief, the scale-space derivatives can be represented by the common formula [23]:

$$L_{x^\alpha y^\beta}(x, y; \sigma) = (\partial_{x^\alpha y^\beta} G(x, y; \sigma)) * f(x, y) \tag{9}$$

where α and β vary from 0 to 2, e.g. α=2 and β=0 for L_{xx}; α=1 and β=1 for L_{xy}. Equivalently, due to the associative property it can also be computed by directly differentiating the scale-space representation L.

$$\begin{aligned} L_{x^\alpha y^\beta}(x, y; \sigma) &= \partial_{x^\alpha y^\beta}(G(x, y; \sigma) * f(x, y)) \\ &= \partial_{x^\alpha y^\beta} L(x, y; \sigma) \end{aligned} \tag{10}$$

In contrast to these scale-space based interest point detectors, MSER detects keypoints without any smoothing process involved. It finds stable and connected intensity regions with a watershed segmentation algorithm. During this process, the fine and large structures with various sizes can be automatically detected without producing multi-scale images.

Table 2. The energies from $G_{1.2} * F_{(u,v)}$.

0.6984	0.4652	0.2654	0.1050	0.0297	0.0065	0.0013	0.0002
0.4652	0.3098	0.1768	0.0700	0.0198	0.0043	0.0008	0.0001
0.2654	0.1768	0.1008	0.0399	0.0113	0.0025	0.0005	0.0001
0.1050	0.0700	0.0399	0.0158	0.0045	0.0010	0.0002	0.0000
0.0297	0.0198	0.0113	0.0045	0.0013	0.0003	0.0001	0.0000
0.0065	0.0043	0.0025	0.0010	0.0003	0.0001	0.0000	0.0000
0.0013	0.0008	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000
0.0002	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000

3.2. Design approach

In the first step, as shown in the above formulae, the detectors except MSER use different combinations of the scale-space derivatives. For one specific scale, the image could be convolved with a band-pass filter, e.g. the Laplacian of Gaussian. During the detection process, however, multiple scales need to be calculated. No matter what combination of scale-space derivatives the detectors use, the operation that is shared by all detectors is that the images are first smoothed by a low-pass Gaussian filter as illustrated in Equation (10). When the σ increases, the filter shifts to low frequency in the frequency domain. So the cut-off frequency is defined by the standard deviation σ_0 of the initial scale layer. The larger frequencies beyond that frequency are useless for feature detection.

Thus, we evaluate the impact of the Gaussian filter on the discrete cosine transform (DCT) coefficients for the sake of designing a novel quantization table. The DCT transforms an image block (typically 8 by 8 pixels) into the frequency domain, and the DCT coefficients correspond to different spectral sub-bands. In the spatial domain the image block can be represented by a linear combination of the DCT basis images times the DCT coefficients as the weights [25].

$$f = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C_{(u,v)} F_{(u,v)} \tag{11}$$

where $C_{(u,v)}$ are the coefficients and $F_{(u,v)}$ are the DCT basis functions. Thus,

$$\begin{aligned} G_\sigma * f &= G_\sigma * (\sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C_{(u,v)} F_{(u,v)}) \\ &= \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C_{(u,v)} (G_\sigma * F_{(u,v)}) \end{aligned} \tag{12}$$

We use the energy of $G_\sigma * F_{(u,v)}$ as an importance score for the coefficient $C_{(u,v)}$. The energy of $A_{M\times M} = G_\sigma * F_{(u,v)}$ is computed as

$$\mathcal{E}(A) = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} A_{ij}^2 \tag{13}$$

When the energy of $G_\sigma * F_{(u,v)}$ is zero, this means that this DCT basis function has no contribution after smoothing. Thus, it can be quantized using a coarse quantizer. Contrarily, large values relate to high importance of this DCT basis function, which should be quantized by a fine quantizer. As a result, our novel quantizers can be represented by

$$Q_{(u,v)} = min\{round[s \cdot \frac{1}{\mathcal{E}(G_\sigma * F_{(u,v)})}],\ 255\} \tag{14}$$

where $\mathcal{E}$ is the energy calculation function. s is used to make the first AC quantizer $Q_{(0,1)}$ equal to the one in the default table (value 11), since standard JPEG applies Differential Pulse Code Modulation (DPCM) coding for DC coefficients and Run Length Coding (RLC) for AC coefficients. By using the same first AC quantizer, the resulting image has a similar bit rate compared to when using the default quantization table. In addition all the values of the quantization table are limited to at most 255.

3.3. Novel quantization table

Each detector uses a similar initial scale σ_0, with which the original image is convolved with different operators, e.g. σ_0=1.2 for Hessian-Laplacian, Harris-Laplacian and SURF; σ_0=1.5199 for DoG in [17]. Therefore, we design the quantizers $Q_{(u,v)}$ according to $G_{1.2}$. The energy matrix of each coefficient is shown in Table 2 and the corresponding quantization table is presented in Table 3. From the table it can be seen that 1) it is symmetric, 2) it is suitable for standard zig-zag scan and RLC, 3) many of the high frequencies seem to have little relevance for feature detection. Next we report the results from our experiments which show that the proposed quantization table leads to improved feature detection performance compared to the default quantziation table of JPEG.

Table 3. The proposed quantization table.

7	11	19	49	172	255	255	255
11	17	29	73	255	255	255	255
19	29	51	128	255	255	255	255
49	73	128	255	255	255	255	255
172	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255
255	255	255	255	255	255	255	255

4. EXPERIMENTAL RESULTS

The work in [9] is related to our work, so the performance of their proposed quantization table is also compared. Fig.1 shows the performance comparison among the default quantization table, our novel quantization table and the quantization table proposed in [9]. It can be observed from Fig.1 that our proposed quantization table performs the best. The performance of all the scale-space based detectors is improved with our novel quantization table. For example, the number of correct matches is improved by 5% to 12.5% at a bit rate of 0.35 bpp for the scale-space based detectors. The MSER detector, however, has a similar performance as for the default quantization table, which is expected. Since it does not contain a Gaussian smoothing process and the compression artifacts generate many spurious features, its performance can not be improved by our approach.

Our quantzation table is designed for a base σ_0=1.2, however, this value could probably be changed in other detectors. For example, if a smaller σ_0 is applied in scale selection, the cut-off frequency is higher and many high frequencies could be helpful for improving detection. If a larger σ_0 is used, the quantizers will have more large

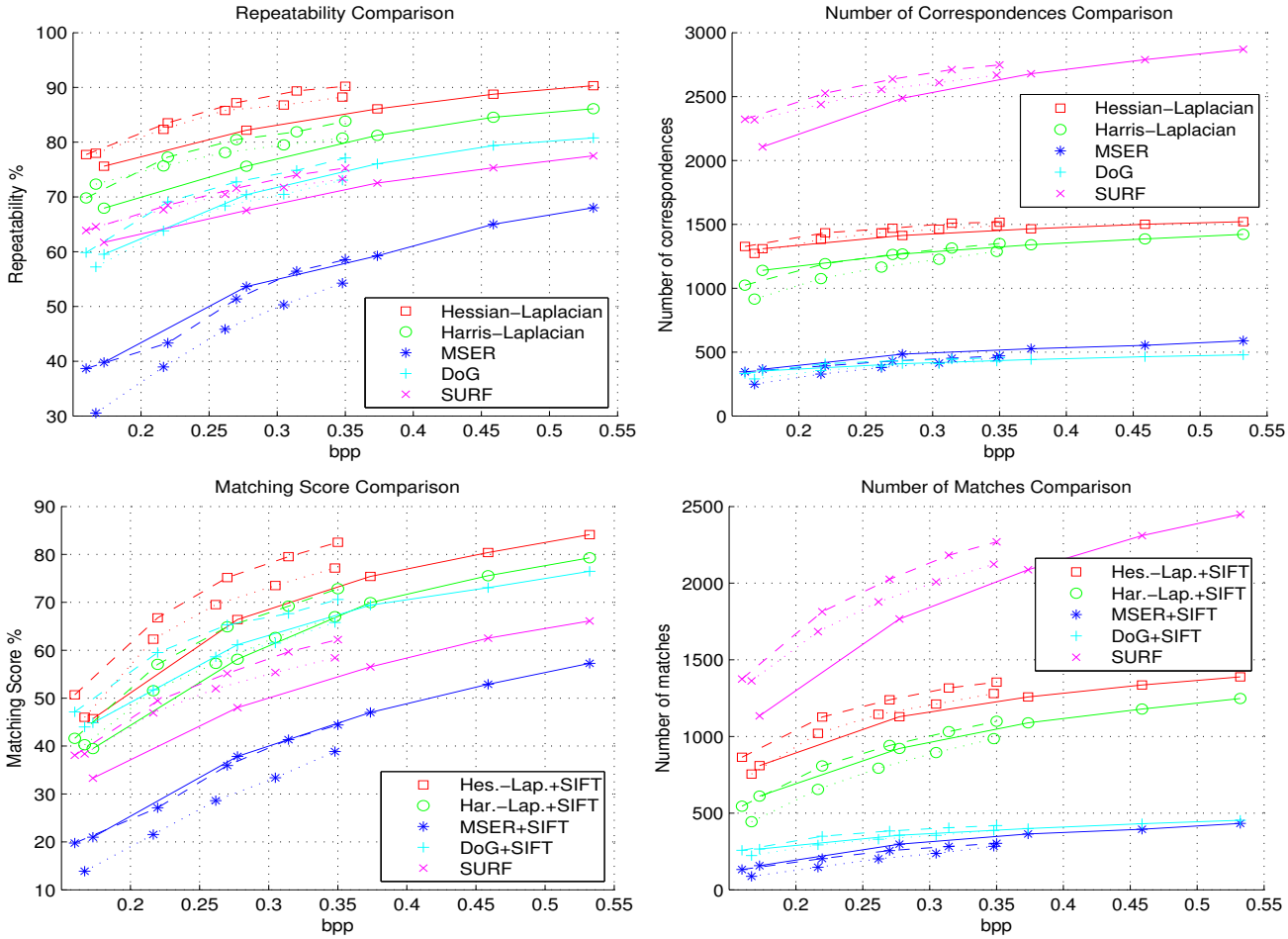

Fig. 1. Comparison for the default quantization table (solid curves), our proposed quantization table (dashed curves) and the quantization table in [9] (dotted curves).

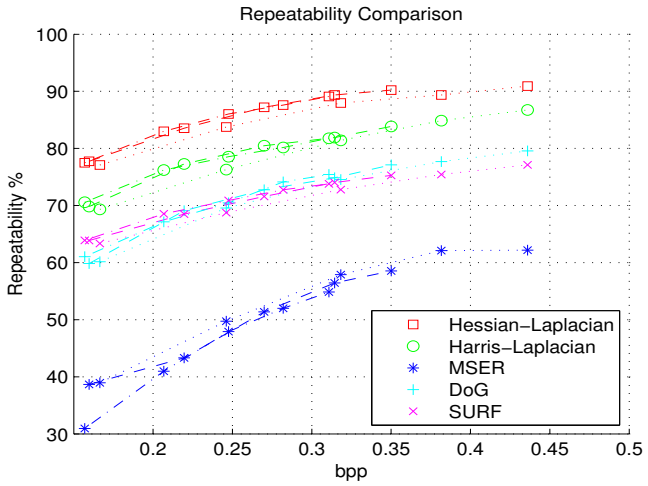

Fig. 2. Repeatability comparison for the proposed quantization tables with σ_0=0.8 (dotted curves), σ_0=1.2 (dashed curves) and σ_0=1.6 (dash-dotted curves).

values. Therefore, we next use σ_0 = 0.8, 1.2 and 1.6 to generate different quantization tables and the detector repeatability scores are compared for them. Fig. 2 demonstrates that the table obtained for σ_0=0.8 performs the worst while the repeatability scores of σ_0 = 1.2 and 1.6 are similar for the tested scale-space based detectors. This is because the table generated for σ_0=0.8 leads to more bits due to its finer quantizers for high frequencies. In contrast, the table for σ_0=1.6 leads to lower bit rate. On the other hand, the σ_0 is actually not the smallest scale of the detected features, e.g. DoG and SURF detect features starting from the second layer in the scale space. We also find that the results are worse if we use too large σ_0, which eliminates too much image information. The detection process is complex and devised to be robust to errors and the standard JPEG encoding syntax targets at best possible rate-distortion performance. Many aspects affect the rate-repeatability curves, thus, the proposed quantization table is not necessarily the best one for one specific detector.

5. CONCLUSION

In this paper, first we review the properties of scale-space based feature detectors and then propose a novel technique for designing the JPEG quantization table. JPEG images with the proposed quantization table lead to higher repeatability score, number of correspondences, matching score and number of matches compared to JPEG images with the default quantization table. The MSER detector shows, as expected, no improvement when applied to JPEG images with our novel quantization table. Improving the MSER feature detection performance will be addressed in our future work.

6. REFERENCES

[1] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L.V. Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1-2, pp. 43–72, Nov. 2005.

[2] J. Chao and E. Steinbach, "Preserving SIFT features in JPEG-encoded images," *Proc. IEEE International Conference on Image Processing, Brussels, Belgium*, Sept. 2011.

[3] ITU-T Rec. T.84, *Information Technology - Digital Compression and Coding of Continuous-Tone Still Images: Extensions*, 1996.

[4] L. Chang, C. Wang, and S. Lee, "Designing JPEG quantization tables based on human visual system," in *Proc. IEEE International Conference on Image Processing*, 1999, pp. 376–380.

[5] A. B. Watson, "DCT quantization matrices visually optimized for individual images," in *Proc. AIAA Computing in Aerospace*, 1993, pp. 286–291.

[6] G. Jeong, C. Kim, H. Ahn, and B. Ahn, "JPEG quantization table design for face images and its application to face recognition," *IEICE - Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E89-A, no. 11, pp. 2990–2993, Nov. 2006.

[7] M. Konrad and H. Stögner, "Evolutionary optimization of JPEG quantization tables for compressing iris polar images in iris recognition systems," *International Symposium on Image and Signal Processing and Analysis*, pp. 534–539, Sept. 2009.

[8] M. Makar, H. Lakshman, V. Chandrasekhar, and B. Girod, "Gradient preserving quantization," in *Proc. IEEE International Conference on Image Processing*, Sept. 2012.

[9] L. Duan, X. Liu, J. Chen, T. Huang, and W. Gao, "Optimizing JPEG quantization table for low bit rate mobile visual search," in *Proc. Visual Communications and Image Processing*, San Diego, CA, USA, Nov. 2012, pp. 1–6.

[10] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points," in *Proc. International Conference on Computer Vision*, Vancouver, Canada, 2001, pp. 525–531.

[11] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, no. 2, pp. 79–116, 1998.

[12] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features," in *Proc. European Conf. Computer Vision (ECCV)*, Graz, Austria, May 2006, pp. 404–417.

[13] D.G. Lowe, "Distinctive image feature from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[14] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *International Journal of Computer Vision*, vol. 37, no. 2, June 2000.

[15] P. Moreels and P. Perona, "Evaluation of features detectors and descriptors based on 3d objects," *International Journal of Computer Vision*, vol. 73, no. 3, July 2007.

[16] A. Haja, B. Jähne, and S. Abraham, "Localization accuracy of region detectors," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, June 2008, pp. 1–8.

[17] K. Lenc, V. Gulshan, and A. Vedaldi, "Vlbenchmarks," http://www.vlfeat.org/benchmarks/, 2012.

[18] "Independent JPEG Group," http://www.ijg.org/.

[19] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *Proc. European Conference on Computer Vision*, Copenhagen, Denmark, 2002.

[20] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International Journal on Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004.

[21] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *Proc. British Machine Vision Conf. (BMVC)*, Cardiff, UK, Sept. 2002, pp. 384–396.

[22] R. Young, "The gaussian derivative model for spatial vision: I. retinal mechanisms," *Spatial Vision*, vol. 2, pp. 273–293, 1987.

[23] T. Lindeberg, "Scale-space," *Encyclopedia of Computer Science and Engineering*, vol. 4, pp. 2495–2504, 2009.

[24] T. Lindeberg, "Scale selection properties of generalized scale-space intererest point detectors," *Journal of Mathematical Imaging and Vision*, vol. 4, Sept. 2012.

[25] S. Khayam, "The discrete cosine transform (DCT): Theory and application," 2003, Tutorial, Department of Electrical & Computing Engineering, Michigan State University.