A New Hybrid Texture-Perceptual Descriptor: Application CBIR

Dounia Awad, Vincent Courboulay, Arnaud Revel

A new hybrid texture-perceptual descriptor: application CBIR

Dounia AWAD , Vincent Courboulay, Arnaud Revel

L3I-University of La Rochelle , Av Michel Crepeau 17000 La Rochelle, France

Abstract. Content based image retrieval (CBIR) has been the center of interest for a long time. A lot of research have been done to enhance the performance of these systems. Most of the proposed works focused on improving the image representation(bag-of-features) and classification methods. In this paper, we focus on enhancing the second component of CBIR system: region appearance description method. In this context, we propose a new descriptor describing the spatial frequency property of some perceptual features in the image. This descriptor has the advantage of being lower dimension vs. traditional descriptors as SIFT (60 vs 128), thus computationally more efficient, with only 5% loss in performance using a typical CBIR algorithm on VOC 2007 dataset.

The number of digital images continues to increase, especially with the expansion of social networks: according to Time magazine, more than 130,000 images are uploading each minute on Facebook. Thus, it will be difficult for a human to use this vast collection of images, e.g.: searching manually for images containing objects or persons. Content based image retrieval(CBIR) system is necessary for this kind of tasks.
Content based image retrieval (CBIR) has been the subject of interest in the computer vision community for a long time: a lot of algorithms have been proposed in the last decades. Most of these systems are based on local approaches (illustrated in figure 1). According to [6], local approaches consists in 5 steps: region selection, region appearance description, region appearance encoding, derivation of image features from the set of region appearance codes by spatial pooling, classification . A baseline of CBIR method [6] is represented in figure 1. In the following, we present the different methods used in each step.

region selection As shown in figure 1, a typical CBIR algorithm first scans the image to select the regions of interest. In the literature, there are two concepts to accomplish such task:

- The first concept considers the entire image. Most researchers in this category use dense grid in their algorithms [15] [38]. It consists in dividing images into small regions and extracting information within each region. The main strength of this technique is that is there no loss in data. However, processing all this amount of information is computationally intensive in terms of time and memory: most of the processing time is spent on analyzing useless data. [25].

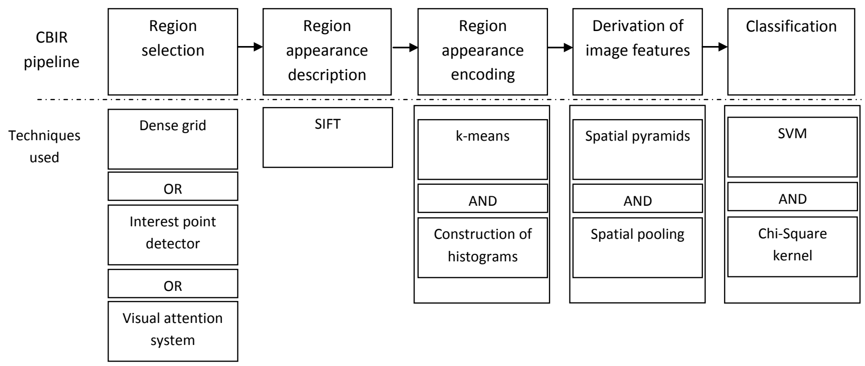

Fig. 1. pipeline of content based image retrieval algorithm

– The second concept, in opposition to the first, considers that not the whole image is necessary for an image retrieval task. The researchers in this category use techniques as:

- Interest point detector [36]: they extract the points of interest basing on simple geometric forms as blobs, corners. They have the advantage of being invariant to scale changes. However, the main limitation of using interest point detector is that they consider that the interest in the image is directly correlated with the presence of geometric structures. This constraint is well known as semantic gap [31].
- Saliency based-region selection: other researchers used more perceptual techniques for selecting regions based on human attention. These systems, known as bottom-up visual attention systems [3], have an objective to select the regions that attract our interest according to some discriminate features [9]: color, orientation, intensity. These features were chosen basing on some psychological and physiological theories [35] [34] [39].

An evaluation of these two concepts based on comparison between interest point detectors and dense grid [38] demonstrated that for some image classification tasks, selecting regions from a regular grid outperforms the use of interest point detectors.

region appearance description The second step of CBIR algorithm consists in describing each selected region by a set of regions appearance attributes, called local descriptors or local features. As a matter of fact, deciding which feature is more fitted to describe a region, is delicate and depends on several factors as the task of the proposed system, the class of the considered object. Generally, this kind of decision is based on compromise between the accuracy of the system and the generality of the chosen feature. Therefore, a lot of descriptors have been proposed. According to their feature basis for describing regions of interest, we can categorize them in 4 families [23]:

1. Descriptor based distribution: they are the most popular descriptors in CBIR domain, as they are usually handcrafted to have some geometric and photometric invariance. The region appearance features are represented by an histogram of gradient locations and orientation or of Haar wavelet responses, e.g: SIFT [20], [4]. This descriptor known as "Scale Invariant Feature Trasnform" was proposed by Lowe in 2004. Since then, this descriptor is widely used by the researchers and it demonstrated an excellent performance in many fields, especially in object recognition domain. Nevertheless, the high dimensionality of SIFT is its main limitation [17]: it prevents researchers from using it in retrieving images in large databases as ImageNet.
2. Differential descriptor: they are one of the first proposed methods to describe the regions of an image. The features in this family are represented by a set of derivatives computed up to given order, e.g: steerable filters [8]. The descriptors generated are vectors of very low dimension, e.g: 9 for steerable filters. According to Schmid [23], although steerable filters was the best efficient descriptor among other low dimensional descriptors, its performance was very low against high dimensional descriptors as SIFT. Actually, they are still used to describe some biomedical images.
3. Textural descriptor: they describe the texture properties of the selected region, the descriptor is computed according to the type of texture of these regions: statistical, structural or random [14]. Thus, the type of the region texture must be known in advance: if it is statistical we can compute the descriptor using cooccurrence Matrix, Markov model, spatial frequency descriptors[41]. In general, they are used in biomedical domain and document processing.
4. Others: other descriptors was proposed to describe features as: color, spectral nature of the image region. Van Gool [11] proposed "Generalized moment invariant" which represents the multi-spectral nature of image region. It consists in computing central moment of high order and degree. However, these moments are sensitive to small geometric and photometric distortions.

Image representation Once these descriptors computed, they can serve as basis to construct the image representation. There are different proposals of image representation in the literature. In this paper, we focus on the most popular image representation in CBIR domain: bag-of-features representation [33]. It consists in encoding an image as a spatial histogram of visual words derived from a given vocabulary. The construction of this histogram starts by learning a k-means [19] visual vocabulary: it consists in partitioning the local descriptors into informative clusters $\mu_1, .., \mu_k$(visual words). Given a set of descriptors $x_1, ..., x_k$ sampled from an image, each local descriptor is assigned to the corresponding visual word as given by equation 1 [42], resulting a non negative vector $f_{hist} \in R^k$ such that $[f_{hist}]_k = |\{i : q_i = k\}|$. This vector represents the histogram encoding the local descriptor for each spatial region.

$$q_{ki} = argmin_k ||x_i - \mu_k||^2 \tag{1}$$

Table 1. Interest detector category

Interest point detector family	Description	The detector the most used in this family
Corner detector	It detects the points which corresponds to rough variation in image intensity	Harris [12] and its variants [22]
Blob detector	It detects the regions which differ from their neighborhood in some properties, e.g: color, intensity	Hessian [2] and its variants
Region detector	Based on segmentation methods	MSER [21]

These spatial regions were obtained using spatial pyramid method [16]: it consists in dividing the image into 1×1, 3×1 (three horizontal stripes), and 2×2 (four quadrants) grids, for a total of 8 regions. Once the encoding is computed for each region, $L1$ normalization is employed. Thus, the image representation is an additive combination of the region encoding [6]. After pooling, a chi-square feature map is applied to the entire image representation. To make this representation suitable for use within Chatfield linear SVM framework [6], $L2$ normalization is used and a linear SVM classifier is applied.

A lot of enhancements have been proposed to improve the performance of CBIR systems. Most propositions focused only on improving the image representation step [6]. We are interested in enhancing the first two components of a CBIR system: many experiment [32] [23] demonstrated that the region selection and appearance description methods can increase or decrease the capacity of a CBIR system on precising which category of object is in the image. In a previous work, we demonstrated that not all the points of interest extracted by interest point detector are pertinent for retrieving task: filtering 60% of points of interest using visual attention system didn't affect the performance of a CBIR system [1].
Inspiring from this work, and according to psychophysical research which has given the evidence that the human brain does a frequency analysis of the image [5] [10], we propose a new hybrid descriptor which describes the spatial frequency of the perceptual features considered in a visual attention system: color, intensity, orientation. This idea is motivated by the fact that a lot of researchers used a combination of different types of descriptors to describe at the same time the color and the orientation features of a region of interest, and thus more computational complexity for their algorithm [7]. In addition, this descriptor can be considered as a first step to solve the limitations of using SIFT in large image databases as it has the advantage of being of lower dimension than SIFT: 60 vs 128, and thus computationally more efficient. In this context, we compare the

impact of using these two descriptors on the performance of the CBIR system shown in figure 1 on VOC 2007 Challenge [7].

This paper is organized as follows. Section 1 presents in details the different proposed techniques for region selection and region appearance description. In section 2, we will present new hybrid descriptor, and its evaluation vs. SIFT for each region selection method presented in section 1.1.

1 Related works

1.1 Region Selection

In the literature, there are two selection strategies: the first one considers the entire image and use dense grid for this. Conversely, the second one selects image regions that are considered as pertinent to predict which object is in the image, and uses techniques as interest point detectors and visual attention systems. In the following, we will present these techniques:

Dense grid as mentioned before, there is no guarantee that the selected regions of interest will yield image representation that is optimal for solving a given image classification [25]. Thus, researchers used dense grid: if the image is completely covered by regions, the whole image can be reconstructed from the set of selected regions, and therefore no information is lost [25]. In our paper, we follow the approach of Bosch et al.[4]: our dense grid consists of dividing the image into small uniform regions with spacing M pixels, here $M = 10$ and the descriptors are computed at each grid points.

Interest point detectors Interest point detectors select the image patterns that differ from its immediate neighborhood and at the same time that are stable and robust to different image variations, e.g: contour, blobs. During the last decades, different detectors have been proposed. According to Schmid [36], we can categorize them into three families (c.f. Table 1).
An evaluation of these different detectors [32] demonstrated that Harris-Laplace is the detector the most efficient and informative. Thus, we decided using Harris-Laplace [22] as the reference detector in our evaluation. As Harris-Laplace is dedicated only to corner-like region, we decided to use another detector Laplacian [18], dedicated to blob-like region and considered as a complementary detector to Harris-Laplace.

Saliency-based detector-visual attention system some researchers used the bottom-up visual attention system as a new saliency-based detector. These systems scan the image for regions that attract our interest. In addition, they are validated against eye movements of human observers. [1] used a visual attention

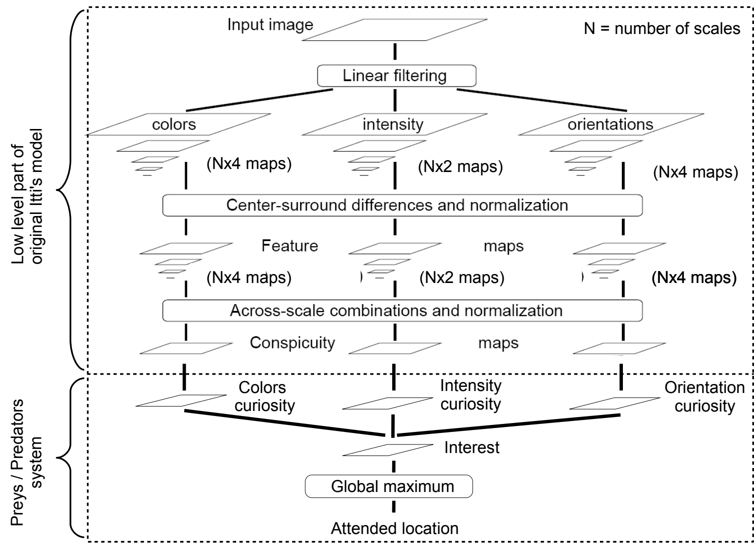

Fig. 2. Perreira Da Silva visual attention system

system as a filtering approach for points of interest extracted by Harris-Laplace and Laplacian and they demonstrated that filtering 60% of these points didn't affect the performance of the reference CBIR system. Inspiring from this work, we decided using Perreira Da Silva visual attention system [27] as reference for saliency based detector.

Perreira Da Silva system Perreira proposed a new competitive and hierarchical system that allows modeling the temporal evolution of the visual focus of the attention and its validation. As shown in figure 2, it is based on the classical algorithm proposed by Itti in [13]: an input image I of width I_w and height I_h is subsampled into multi-resolution pyramids, and each pyramid level $P_{t,\sigma}$ is decomposed into low level features t for Red (R), Green (G), Blue (B), Yellow (Y), Intensity (I_{on}, I_{off}) and local orientations (O_0, O_{45}, O_{90}, O_{135}). From these channels, center surround feature maps f_l for the different features are constructed and normalized. In each channel, maps are summed across scale and normalized:

$$f_l = \bigoplus_{\sigma} P_{t,\sigma} \tag{2}$$

with $\sigma \in \{1, .., N-1\}$, $N = 1 + log_2\,(min\,(I_w, I_h))$.
These maps are summed and normalized once to yield three conspicuity maps. These three conspicuity maps are representative of the three main human perceptual channels: color, intensity and orientation. Perreira Da Silva et al. propose to substitute the second part of Itti's model by an optimal competitive approach: a preys / predators system. [26] shows that despite the non deterministic behavior of preys / predators equations, the system exhibits interesting properties

of stability, reproducibility and reactiveness while allowing a fast and efficient exploration of the scene. We applied the same optimal parameters used by Perreira Da Silva in [26] to evaluate our approach. The output of this algorithm is a saliency map, computed by a temporal average of the focalization computed through a certain period of time t.
In the following, we tackle the descriptor used as reference in our evaluation.

1.2 Region appearance description

After extracting regions of interest, each region is described by a set of scalar values representing one of the image region features. In this paper, we used SIFT [20] as it is one of the most popular and used descriptor in CBIR domain. This can be explained by the fact that it provides descriptors robust to scale changes and small geometric distortions. Therefore, we used original SIFT to describe the points of interest computed by interest points detectors. For dense grid, at each grid point, SIFT descriptor are computed over circular support patches with radii=4, 8, 12 and 16 pixels [4]. After computing the descriptor, the rest of typical CBIR remains the same. In the next section, we present our approach for region appearance description. As mentioned in the introduction, we propose a new hybrid texture-perceptual descriptor that describes the spatial frequency of the three perceptual features: color, orientation, intensity.

2 Our proposition: textural-perceptual descriptor

In this section, we present a new texture-perceptual descriptor that analyze the frequency content of the perceptual features in a multi-resolution pyramids computed in the first phase of Perreira's system (c.f 1.1).
Usually, spatial frequency is analyzed using spatial frequency descriptors. These descriptors are based on image transform matrix as Fourier [5]. Fourier transform is one of the most powerful descriptor in texture analyzing domain. For our descriptor, a transform closely related to Fourier called Hartley transform $H_{k,l}$ is used: it contains the same information that Fourier does. In addition, contrary to Fourier, it has the advantage of being a real function and this offer computational advantages in signal processing application [24].
As shown in figure 3, Hartley transformation matrix of size $M \times M$ is computed as given:

$$H_{k,l} = \sum_{l,k=0}^{M-1} \left(\cos\left(\frac{2\pi lk}{M}\right) + \sin\left(\frac{2\pi lk}{M}\right)\right) \tag{3}$$

where $k, l \in \{0, ..., M-1\}$.
In general, Fourier descriptor is blamed for not efficient in capturing a local features [40]. Several researchers have proposed methods attempting to overcome this drawback. According to Unser, the local texture property of an image region can be characterized by a set of energy measures computed at the output of a filter bank [37]. In this context, Unser proposed an interesting way to exploit the

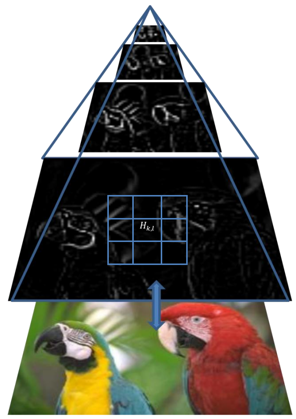

Fig. 3. Example of computing our descriptor features using one of the multi-resolution pyramids

spatial dependencies that characterize the texture of a region, more computationnally efficient called local linear transform [37]: it consists in computing for each point of interest $x_{k,l}$, a local linear property $y_{k,l}$ as given in equation:

$$y_{k,l} = \sum_{l,k=0,..M-1} T_M . x_{k,l} \tag{4}$$

where T_M represents an image transform matrix of size $M \times M$. In our case, T_M represents Hartley transform $H_{k,l}$.
Extending this method, each point of interest $x_{k,l}$ is tracked back in the Perreira's system to multi-resolution feature pyramids computing phase. For each pyramid level $P_{t,\sigma}$, a neighborhood window $w_{x_{k,l}}$ of the same size $M \times M$ as Hartley matrix, is centered around $x_{k,l}$. In our method, $y_{k,l}$ represents texture energy measures and it is defined by:

$$y^c_{k,l} = y_{k,l} = \sum_{l,k=0,..M-1} \left[w_{x_{k,l}} . T^c_M \right]^2 \tag{5}$$

where T^c_M as shown in Fig. 4 is the convolution of each column with each row in the matrix $H_{k,l}$ and $c \in \left\{1, ..., M^2\right\}$.
By combining the $\left(M^2\right)$ channels as given in the equation 6, we obtained $\left(M \times \frac{M+1}{2}\right)$ channels(noted TR) invariant to some rotation transformations [28]:

$$TR_{k,l} = \frac{y_{k,l} + y_{l,k}}{2} \tag{6}$$

To provide better visual perception, a contrast enhancement is applied using equation 7 [29], resulting an histogram f^t_i of dimension $\left(M \times \frac{M+1}{2}\right)$ computed for each multi-resolution feature pyramid level $P_{t,\sigma}$:

$$f_i = \frac{\log(\epsilon) - \log(TR_{k,l} + \epsilon)}{\log(\epsilon) - \log(\epsilon + 1)} \tag{7}$$

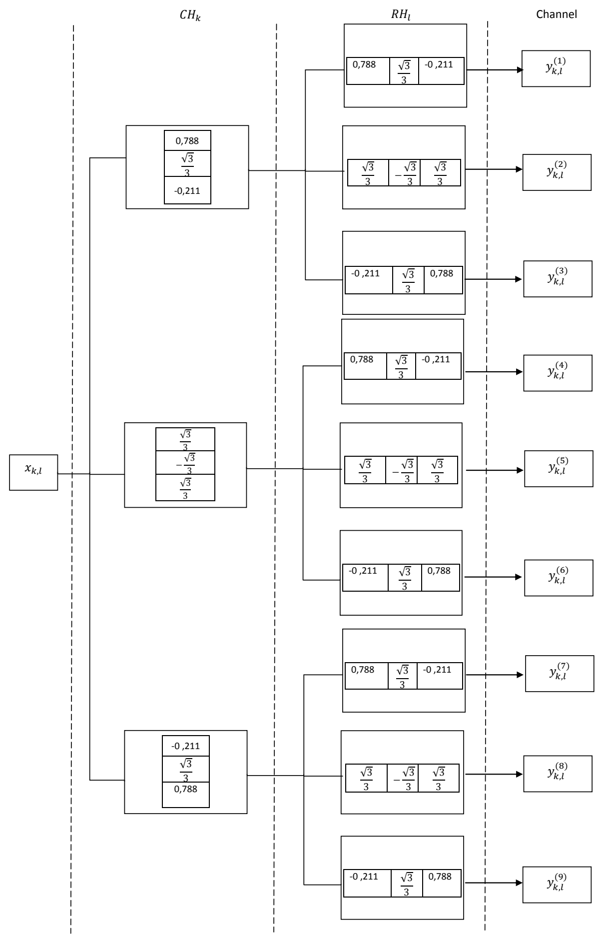

Fig. 4. Computation of a running 3x3 Hartley by successive row and column filtering

where $i \in \left\{1, ..., M \times \frac{M+1}{2}\right\}$ and $\epsilon > 0$ is a suitably small value(we use $\epsilon = 0.05$). Concatenating these histograms for each point of interest $x_{k,l}$, the results in a descriptor of dimension$\left(P \times M \times \frac{M+1}{2}\right)$ with P=number of multi-resolution pyramids. In our evaluation, $M = 3$ and $P = 10$, resulting a descriptor of dimension 60. In the next section, we present the evaluation of this hybrid descriptor vs. SIFT.

2.1 Evaluation of our descriptor vs. SIFT

This section describes the experimental evaluation of the descriptor mentioned above vs. SIFT on VOC 2007 dataset. This dataset contains about 10,000 images split into train, validation, and test sets, and labelled with twenty classes.

Table 2. Impact descriptors for each region selection method on image classification results using Pascal VOC 2007 dataset (%)

Region selection methods	descriptor	aero-plane	bike	bird	boat	bottle	bus	car	cat	chair	cow	dining table	dog	horse	motor-bike	person	potted plant	sheep	sofa	train	tv moni-tor
Dense grid	SIFT	69,58	56,28	40,63	64,17	24,56	60,63	75,06	56,17	49,19	38,90	50,54	36,79	75,69	63,59	81,51	26,55	45,21	46,78	74,36	50,10
Dense grid	our de-scriptor	68,02	44,96	36,90	58,99	21,98	46,57	63,84	43,67	41,53	27,07	51,37	34,26	70,13	54,41	81,29	35,08	46,44	47,93	68,23	45,91
Harris-Laplace, Lapla-cian	SIFT	61,45	48,12	37,76	51,27	21,34	47,30	65,60	46,16	41,45	31,62	35,85	28,37	67,62	50,95	73,06	15,21	29,74	22,51	62,95	37,12
Harris-Laplace, Lapla-cian	our de-scriptor	59,90	36,80	34,03	46,09	18,76	33,24	54,38	33,66	33,79	19,79	36,68	25,84	62,07	41,77	72,83	23,74	30,97	23,66	56,81	32,93
Perreira's system	SIFT	19,17	10,51	17,34	7,51	6,15	9,35	26,55	21,14	13,12	13,33	5,46	21,65	21,64	18,90	50,78	6,22	3,72	7,41	10,28	10,28
Perreira's system	our de-scriptor	44,72	22,28	21,68	14,23	10,99	18,66	44,41	25,62	19,59	8,37	12,96	24,96	36,38	23,72	63,32	20,03	11,03	12,94	32,72	21,36

A 1-vs-rest classifier is learned for each class and the performance is evaluated by Average Precision. As mentioned before, in this experiment we study the impact of each descriptor for each region selection method mentioned in section 1.1: dense grid, Harris Laplace + Laplacian, Perreira's system. The rest of the CBIR system remains the same. The results are shown in table 2.
Observing these results shows that for traditional region selection methods, althougth that the dimension of our descriptor is reduced to the half regardless SIFT, thus gain in run-query time and in memory (c.f table 3), only an average of 5% loss of performance is obtained. Furthermore, For perceptual region selection methods, about 9% gain of performance is obtained using our descriptor regardless SIFT. This results validate our hypothesis concerning the fact our descriptor can be considered as a first step to solve the memory and time consuming by the use of high dimensional SIFT descriptors in object recognition systems. Indeed, this consumption of memory constitute a serious problem in object recognition, especially with large image bases. Furthermore, we suggest that by using our texture-perceptual descriptor, we describe not only the features of region of interest that are are relevant for human perception, but also the features which are relevant for CBIR systems (color, orientation).

Table 3. computational complexity of descriptors with N is the number of points of interest

Region appearance description methods	SIFT	our descriptor
Computational complexity	$O(N^2)$	$O(N)$

3 Conclusion

ln this paper, we proposed a new perceptual-texture descriptor, more computationally efficient which describes different perceptual features for regions of interest and at the same time lower dimension than SIFT descriptor, approximately the half. Furthermore, it performs approximately the same as SIFT (with 5% loss in performance) on different region selection methods. In the future, it would be interesting to build an interactive perceptual CBIR system. We think it can be done by using a suitable image representation as [30] and incorporating relevance feedback step by taking advantage of top-down techniques already implemented in the visual attention system.

References

1. Awad, D., Courboulay, V., Revel, A.: Saliency filtering of sift detectors: Application to cbir. In: Proceedings of the 14th International Conference on Advanced Concepts for Intelligent Vision Systems. pp. 290–300. ACIVS'12, Springer-Verlag, Berlin, Heidelberg (2012)
2. Beaudet, P.R.: Rotationally invariant image operators. In: Proceedings of the 4th International Joint Conference on Pattern Recognition. pp. 579–583. Kyoto, Japan (Nov 1978)
3. Borji, A., Itti, L.: State-of-the-art in Visual Attention Modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence 99 (2012)
4. Bosch, A., Zisserman, A., Munoz, X.: Image classification using random forests and ferns. In: IEEE International Conference on Computer Vision (2007)
5. Campbell, F.W., Robson, J.G.: Application of fourier analysis to the visibility of gratings. J Physiol 197(3), 551–566 (Aug 1968)
6. Chatfield, K., Lempitsky, V., Vedaldi, A., Zisserman, A.: The devil is in the details: an evaluation of recent feature encoding methods. In: British Machine Vision Conference (2011)
7. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. International Journal of Computer Vision 88(2), 303–338 (Jun 2010)
8. Freeman, W.T., Adelson, E.H.: The design and use of steerable filters. IEEE Transactions on Pattern Analysis and Machine Intelligence 13, 891–906 (1991)
9. Frintrop, S.: Computational Visual Attention. In: Computer Analysis of Human Behavior, Advances in Pattern Recognition, pp. 1–34. Springer, a. a. sala edn. (2011)
10. Georgeson, M.A.: Spatial Fourier Analysis and Human Vision, Chapter 2, in Tutorial Essays in Psychology, A Guide to Recent Advances. No. vol. 2 (1979)
11. Gool, L.J.V., Moons, T., Ungureanu, D.: Affine/ photometric invariants for planar intensity patterns. In: Proceedings of the 4th European Conference on Computer Vision-Volume I. London, UK (1996)
12. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proceedings of the 4th Alvey Vision Conference. pp. 147–151 (1988)
13. Itti, L., Koch, C., Niebur, E., Others: A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on pattern analysis and machine intelligence 20(11), 1254–1259 (1998)
14. Jain, A.K.: Fundamentals of Digital Image Processing. Prentice-Hall, Inc., Upper Saddle River, NJ, USA (1989)
15. Jurie, F., Triggs, B.: Creating efficient codebooks for visual recognition. In: Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) vol.1. pp. 604–610. Washington, DC, USA (2005)

16. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR '06) - Vol.2. pp. 2169–2178. Washington, DC, USA
17. Ledwich, L., Williams, S.: Reduced sift features for image retrieval and indoor localisation. In: In Australian Conference on Robotics and Automation (2004)
18. Lindeberg, T.: Feature detection with automatic scale selection. Int. J. Comput. Vision 30(2), 79–116 (Nov 1998)
19. Lloyd, S.: Least squares quantization in pcm. IEEE Trans. Inf. Theor. 28(2), 129–137 (Sep 2006)
20. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60(2), 91–110 (Nov 2004)
21. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: British Machine Vision Conference. vol. 1 (2002)
22. Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors. Int. J. Comput. Vision 60(1), 63–86 (Oct 2004)
23. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. 27(10), 1615–1630 (Oct 2005)
24. Millane, R.: Analytic properties of the hartley transform and their implications. Proceedings of the IEEE 82(3), 413–428 (1994)
25. Nowak, E., Jurie, F., Triggs, B.: Sampling strategies for bag-of-features image classification. In: 9th European Conference on Computer Vision (ECCV '06). vol. 3954, pp. 490–503. Graz, Austria (2006)
26. Perreira Da Silva, M., Courboulay, V., Prigent, A., Estraillier, P.: Evaluation of preys / predators systems for visual attention simulation. In: VISAPP 2010 - International Conference on Computer Vision Theory and Applications. pp. 275–282. INSTICC, Angers (2010)
27. Perreira Da Silva, M., Courboulay, V., Estraillier, P.: Objective validation of a dynamical and plausible computational model of visual attention. In: IEEE European workshop on visual information processing. pp. 223–228. Paris, France (Jul 2011)
28. Rachidi, M., Chappard, C., Marchadier, A., Gadois, C., Lespessailles, E., Benhamou, C.L.: Application of laws masks to bone texture analysis: An innovative image analysis tool in osteoporosis. In: Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on. pp. 1191–1194 (2008)
29. Ritika, Kaur, S.: Article: Contrast enhancement techniques for images - a visual analysis. International Journal of Computer Applications 64(17), 20–25 (February 2013), published by Foundation of Computer Science, New York, USA
30. Sanchez, J., Perronnin, F., Mensink, T., Verbeek, J.: Image Classification with the Fisher Vector: Theory and Practice. International Journal of Computer Vision (Jun 2013)
31. Santini, S., Gupta, A., Jain, R.: Emergent semantics through interaction in image databases. IEEE Trans. on Knowl. and Data Eng. 13(3), 337–351 (May 2001)
32. Schmid, C., Mohr, R., Bauckhage, C.: Evaluation of interest point detectors. Int. J. Comput. Vision 37(2), 151–172 (2000)
33. Sivic, J., Zisserman, A.: Video Google: A text retrieval approach to object matching in videos. In: Proceedings of the International Conference on Computer Vision. vol. 2, pp. 1470–1477 (Oct 2003)
34. Treisman, A., Gormican, S.: Feature analysis in early vision: evidence from search asymmetries. Psychological Review 95(1), 15–48 (1988)

35. Treisman, A., Gelade, G.: A Feature-Integration Theory of Attention. Cognitive Psychology 136(12), 97–136 (1980)
36. Tuytelaars, T., Mikolajczyk, K.: Local invariant feature detectors: a survey. Found. Trends. Comput. Graph. Vis. 3(3), 177–280 (Jul 2008)
37. Unser, M.: Local linear transforms for texture measurements. Signal Process. 11(1), 61–79 (Jul 1986)
38. Winn, J., Criminisi, A., Minka, T.: Object categorization by learned universal visual dictionary. In: Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on. pp. 1800–1807 Vol. 2
39. Wolfe, J.M.: Guided Search 2.0: A revised model of visual search. Psychonomic Bulletin & Review 1(2), 202–238 (1994)
40. Zhang, D., Lu, G.: A comparison of shape retrieval using fourier descriptors and short-time fourier descriptors. In: Shum, H.Y., Liao, M., Chang, S.F. (eds.) Advances in Multimedia Information Processing PCM 2001, Lecture Notes in Computer Science, vol. 2195, pp. 855–860. Springer Berlin Heidelberg (2001)
41. Zhang, J., Tan, T.: Brief review of invariant texture analysis methods. Pattern Recognition 35(3), 735–747 (Mar 2002)
42. Zhou, X., Yu, K., Zhang, T., Huang, T.S.: Image classification using super-vector coding of local image descriptors. In: Proceedings of the 11th European conference on Computer vision (ECCV'10): Part V. Berlin, Heidelberg (2010)