

SIFT-CCH: Increasing the SIFT distinctness by color co-occurrence histograms

Peer-reviewed author version

ANCUTI, Cosmin & BEKAERT, Philippe (2007) SIFT-CCH: Increasing the SIFT distinctness by color co-occurrence histograms. In: PROCEEDINGS OF THE 5TH INTERNATIONAL SYMPOSIUM ON IMAGE AND SIGNAL PROCESSING AND ANALYSIS. p. 130-135..

DOI: 10.1109/ISPA.2007.4383677

Handle: <http://hdl.handle.net/1942/8035>

SIFT-CCH: Increasing the SIFT distinctness by Color Co-occurrence Histograms

Cosmin Ancuti and Philippe Bekaert
Hasselt University

Expertise Centre for Digital Media
Transnationale Universiteit Limburg- School of Information Technology
Wetenschapspark 2, 3590 Diepenbeek, Belgium
{cosmin.ancuti,philippe.bekaert}@uhasselt.be

Abstract

Describing regions in a distinctive way, in order to find correct correspondences in images of two separated views, represents a complex and essential task of computer vision. Until now, SIFT (Scale Invariant Feature Transform) has been proven to be the most reliable descriptor among the others. One of the main drawbacks of SIFT is its vulnerability to color images, being designed mainly for the gray images.

To overcome this problem and also to increase the overall distinctness of the SIFT in this paper we introduce a new descriptor that combines the SIFT approach with the color co-occurrence histograms (CCH), a concept used extensively in color texture retrieval and object recognition applications. We evaluate the new descriptor in the context of image matching. The experimental results show that our descriptor outperforms the original version, detecting an important number of additional correct matched feature points while the mismatch ratio remains constant.

1. Introduction

Finding reliable corresponding points in images that represent different projection of the same 3D scene or object is a fundamental problem in computer vision. In general, this task is solved by searching images for the invariant feature points (keypoints) and afterwards, describing as distinctive as possible the neighbour regions centered on the detected points.

The keypoints are extracted from the image regions where the signal varies in at least two directions (e.g. corners, blobs). A necessary property of the feature points is to be repeatable identified (as many as possible extracted points under different viewing conditions).

On the other hand, the descriptors of the keypoints surroundings should be high distinctive and robust to noise, geometric and photometric transformations. The relation between detectors and descriptors is crucial and not every combination between detectors and different descriptors give similar results. In general as much as the detected feature points are more invariant, the less information is transported by the descriptor. Common matching procedure often includes computing distances between descriptor vectors like Euclidian or Mahalanobis distance after covariance matrix has been estimated over a large database of images.

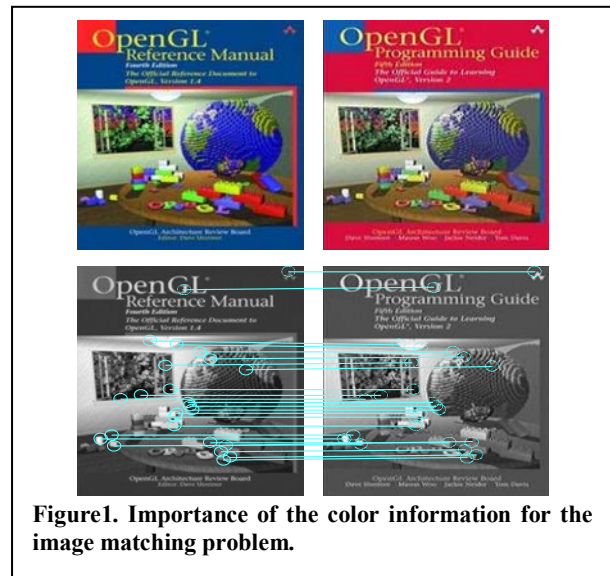


Figure 1. Importance of the color information for the image matching problem.

Different applications like object recognition, image/texture retrieval, 3D reconstruction, camera calibration, robot navigation are relied on correct matching correspondent points.

In [1], SIFT (Scale Invariant Feature Transform) [2],[3] has been proven to be the most discriminative among the others evaluated descriptors. SIFT is relied on extracting scale invariant keypoints using the DoG (Difference of Gaussian) operator. The descriptor part is based on computing the magnitude and orientation of the gradient images, being basically a histogram of gradient orientation computed in the surrounding regions centered on the extracted keypoints.

Despite of its impressive characteristics, SIFT has a major drawback: it does not use the color information. In fact, SIFT has been designed mainly for gray images, ignoring completely the potential information found in the color space. In this paper we overcome this disadvantage of SIFT introducing a new descriptor, SIFT-CCH, which combines the SIFT with color co-occurrence histograms (CCH)[4].

Color information plays a vitally important role in the world as we are seeing. Human visual system can discern thousands of color values, compared to only about tens of gray intensities. To understand the importance of the color for matching problem we present a simple problem. Figure 1 shows the best 25 matched points identified by the SIFT operator for two alike gray images, which in color space represent

entirely different information. This simple example demonstrates the difficulty for a descriptor to distinct color regions that appear very similar in the corresponding gray images when the color is omitted.

In this paper the problem of the color is addressed in an original and relative simple modality, exploiting the color co-occurrence histogram [4] (CCH) characteristics in order to increase the distinctness of the SIFT descriptor. The CCH is an extension of the co-occurrence matrix [5], a common method used in problems like image retrieval, classification and object recognition.

The new descriptor CCH-SIFT, combines the SIFT method with CCH in order to get a more distinctive descriptor that explores efficiently also the color information. The CCH is computed in the neighbour regions centered on the keypoints that are invariant to scale and rotation transformations according to the computed characteristic scales and dominant orientations assigned to the extracted feature points.

In the end, to evaluate our descriptor, various images are tested to compare SIFT-CCH with initial version of the SIFT. The experimental results demonstrate that the new descriptor performs better than the original one, detecting an important number of additional good correspondences in the considered image pairs.

2. Related work

Since powerful local descriptors, relied on extracting invariant feature points, are crucial in several computer vision applications (e.g. image retrieval, camera calibration, robot navigation, object recognition), a wide variety of them have been introduced in the last two decades.

A brief review of the detectors starts with Harris point detector [6] which is still the most widely used basic detector. Unfortunately, the Harris points are not scale invariant and cannot be used with success for wide baseline cases. Lindeberg in [7], detects scale invariant blob like points based on the LoG (Laplacian of Gaussian) operator. The detected feature points are extracted based on the automatic scale selection principle. This principle is fundamental for nearly all the scale invariant feature detectors that have been introduced afterwards. The Harris Laplacian has been presented in [8]. This detector extracts scale invariant feature points searching for the Harris points in the scale space built on the LoG operator. A similar approach is used also for SIFT, where the keypoints are extracted based on the DoG (Difference of Gaussian), a close approximation of the LoG (see Figure 2).

Different techniques, which describe nearby regions of feature points, have been introduced in the literature. The main goal of them is to increase the distinctness of the extracted keypoints neighborhoods. Invariant moments [9] combine central moments of the shape and distribution of the intensity for a considered region. Spin images [10] represents 2D histograms of image regions, where the two dimensions are the distance from the center point and the intensity value. Freeman

introduced steerable filters [11] that guides the derivatives in particular directions. Shape context [12] counts the number of sampled edge points in each bin of a log-polar histogram that extends over a large portion of the image. More recently, SURF [13] uses Hessian matrix and Haar wavelets response combined with the properties of integral images in order to reduce the computational time. SIFT [3] computes a histogram of gradient locations and orientations and has been shown to outperform the other descriptors [1].

Since SIFT is still the most widely used descriptor, it is not surprising that a various versions of it have been proposed in the last years. In PCA-SIFT [14] the Principal Component Analysis is applied to the normalized gradient patch, reducing the length of descriptor vector from 128 to 36. SIFT with global context [15] adds the properties of the shape context descriptor [12] in order to increase the discriminative power when similar texture structure occurs in images. Recently, the CSIFT [16] has been introduced. CSIFT uses the color information based on the color invariance model [17]. Since only the photometrical variation of images (more complex transformation like scale, rotation and affine are disregarded) has been considered this descriptor seems to not be efficient in practical cases.

Color is an important characteristic of the real world that increases the distinctness of the objects. Wyszeki, in [18] presents a detailed study of the color systems. In the last years, color information was used extensively in several computer vision applications and different schemes has been proposed (eg. color histograms [19], color invariant moments [17], co-occurrence matrix [20], color correlogram [21]).

Our approach also uses the color information but in a very different way than CSIFT [16]. The new descriptor combines the SIFT [3] with the color co-occurrence histogram (CCH) [4], with the final scope to increase the overall distinctness and to explore in a simple and effective way the color space information.

3. Feature point detector

The first step in finding good correspondence between images is to detect feature points. We analyzed several scale invariant feature detectors: Harris Laplacian, Hessian Laplacian, edge laplacian (code available on www.robots.ox.ac.uk) and DoG (code available on www.cs.ubc.ca/~lowe/keypoints/). The experiments have disclosed that the most accurate performance of SIFT descriptor is given by the DoG.

Difference of Gaussian (DoG) is a scale invariant detector, used in the initial implementation of the SIFT. Scale invariance of the feature points is relied on the principle of automatic scale selection [7], which is built on the existent relation between images at different resolution levels. In general images contain sharp and

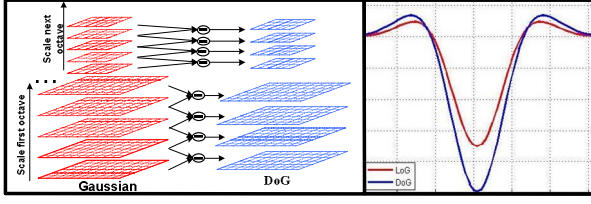


Figure 2. a) Constructing the DoG scale space; b) DoG expression approximates the LoG function

diffuse features, being almost impossible to identify all kinds of features at the same scale level. Lindeberg, in [7] postulates that in absence of other evidence, the selected scale (characteristic scale) is the scale where the function of some combinations of normalized derivatives attain a local maximum. The characteristic scale is independent by the image scale and the ratio between selected scales of two extrema that represent the same image feature is the same with the ratio between image scales.

DoG inherits the forerunner invariant scale feature points, which are based on the LoG (Laplacian of Gaussian) [8]. As can be noticed from the figure 2, this operator is a close approximation of LoG (relation between them can be verified solving the heat diffusion equation).

The detection procedure starts with building an image pyramid convolving the considered image $I(x,y)$ with the gaussian kernel $G(x,y,\sigma)$. Therefore, the scale space can be represented by the expression:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (1)$$

where * represents the convolution operation.

The level of blurring is controlled by the standard deviation σ of the gaussian kernel. The DoG space is obtained subtracting adjacent scale levels of the image:

$$D(x, y, \sigma) = [G(x, y, k\sigma) - G(x, y, \sigma)] * I(x, y) \quad (2)$$

After image $I(x,y)$ has been incrementally convolved with Gaussian kernel creating scale levels separated by a constant k , the image is down sampled with a factor of 2. In consequence each octave has half of the previous octave size. Octaves of the scale space are divided in sub intervals (see Figure 2). Every keypoint points is identified comparing every sample point with its 8 neighbor pixels from the same scale image level and with the 9 pixels from both adjacent scale levels. The sample point is selected only in case when its value is greater or lesser than all the 26 neighbor points.

Next, all selected feature points are verified for the stability. To find more stable locations of the keypoints the $D(x,y,\sigma)$ is fit to a 3D quadratic function using the Taylor expansion of the DoG function centered on the sample point $x=(x,y,\sigma)^T$.

The offset for a more stable position is found solving the following equation:

$$\hat{x} = - \frac{\delta^2 D^{-1}}{\delta x^2} \frac{\delta D}{\delta x} \quad (3)$$

where \hat{x} is the offset vector, and D derivatives are computed using local pixel differences around the point location. Additionally, the keypoints with low contrast and with strong responses along the edges are rejected.

4. SIFT-CCH descriptor

4.1. Scale Invariant Feature Transform (SIFT)

SIFT (Scale Invariant Feature Transform) has been introduced in [2]. An improved version, used in this paper as a reference, has been presented in [3].

Even considerable efforts have been spent to develop new descriptors or to improve the original version, SIFT remains the most attractive and most widely used in practical cases.

The SIFT signature (descriptor) takes into account the image gradient magnitudes and orientations, computed in the region around the feature point location of the image pyramid level determined by the computed characteristic scale of the respective feature point. A 4x4 orientation histogram is built on the 4x4 sub-region of the feature point from a 16x16 region centered on the feature point location. Each histogram has 8 bins corresponding for every 45° (see Figure 3).

Boundary influence is reduced using trilinear interpolation to distribute the value of each gradient sample into adjacent histogram bins. The rotation invariance is obtained by computing a prominent orientation that is assigned to every feature point. After the image gradient orientation and magnitude of every sample point is computed, a 36 bin (a bin for every 10°) orientation histogram is constructed using the gradient orientation in the neighbor patch of the keypoint. The orientation histogram counts the contribution of every point gradient orientation weighted by its gradient magnitude and by a gaussian-weighted circular window from respective region.

4.2. Color Co-occurrence Histogram (CCH)

Even if in the eighties the volumetric description has been considered to have the primordial role [19], color represents an important property of the digital images that manifest increasing consideration in the computer vision applications.

A detailed analysis of the color spaces can be found in [18]. The most used color space is RGB , that is based on the primary components provided by a CCD color camera: red ($R=700nm$), green ($G=546.1nm$) and blue ($B=435.8nm$). A similar color space is XYZ . This representation is relied on the three primaries colors that are not perceived by the human visual system. $L^*a^*b^*$ is another common color system where the luminance (L^*) is combined with a^* (red-green) and b^* (yellow-blue) components. The HIS (Hue Saturation and Intensity) system has been developed to emulate the human color perception. It is perceptually more intuitive but not perceptually uniform. YUV system transports separately the luminance color (Y), the hue (I) and the saturation (Q) being introduced as a television

standard by the NTSC (National Television System Committee). The European correspondent system is YUV used by the PAL and SECAM standards.

Since there is no convergence about which system is more suitable to be used, we follow the indication of Gevers and Smeudels [22]. They evaluated different color systems in the context of object recognition.

Compared to the other color systems, rgb and $I_1I_2I_3$ models appear to be the most appropriate to be used in the color image retrieval applications based on the color metric histogram matching. Due its simplicity, in this paper we are using the rgb representation :

$$r = \frac{R}{R+G+B} \quad g = \frac{G}{R+G+B} \quad b = \frac{B}{R+G+B} \quad (4)$$

Color histogram [22] is a common method to represent texture features used mainly for image retrieval, object recognition and texture classification. Color histograms are efficient and relative easy to compute being robust to translation and rotation transformations and quasi invariant to scale and view angle transformations. The major drawback of them is that the color histograms do not take into consideration the spatial information of the considered region pixels. Therefore, very different images can have similar color distribution. To solve this problem, a simple method is to divide the image into sub-patches and calculate a histogram for each of them. Unfortunately, increasing the number of sub-patches, affects seriously the memory and computational time.

To eliminate this drawback, an alternative is to use the color co-occurrence matrix (CCH) [4]. Besides of the pixel color distribution information the CCH transports also the information of spatial correlation of the neighbor pixels. An equivalent measure that also keeps track of the number of colored pixel pairs that occurs in the space of images at certain distance is the color correlogram [21].

Harlick introduced the co-occurrence matrix in [5] to estimate the spatial gray level dependencies of the pixels. The color co-occurrence histogram is an improved version of the co-occurrence matrix for the color images. Given a color pixel $p_1(r_1, g_1, b_1)$ in the image, the color co-occurrence histogram (CCH) counts the number of occurrences of the color pixel pair (p_1, p_2) , with $p_2(r_2, g_2, b_2)$ an adjacent color pixel located at $d=(\Delta x, \Delta y)$, the relative distance between the pixels (see Figure 3). The color co-occurrence histogram can be seen as a function of the color pixel values and the displacement vector between them.

Considering a color image I of size $M \times N$ and n_Q the number of color quantized levels then the CCH is amounting the number of times when pixel pair (p_1, p_2) reaches the pair of the color levels (i, j) :

$$CCH(x, y, i, j) = \sum_{x=1}^N \sum_{y=1}^M A_i(x, y) \sum_{\Omega} A_j(x + \Delta x, y + \Delta y) \quad (5)$$

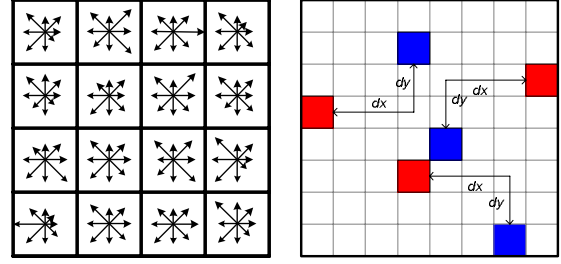


Figure3. a)SIFT 4x4x8 orientation histogram; b) spatial dependencies of the CCH

where Ω includes the number of adjacent pixels at the distance $(\Delta x, \Delta y)$ and A_k is given by the expression:

$$A_k(x, y) = \begin{cases} 1, & p(x, y) = k \\ 0, & otherwise \end{cases} \quad (6)$$

The CCH is equivalent with the second statistical order of the color probability distribution. The first moment is represented by the simple histograms. Actually, if the distance $(\Delta x, \Delta y) = (0, 0)$, then the CCH becomes a simple histogram representation. It has been demonstrated that humans can discriminate textures till 2nd order statistics. The difference in higher order statistics are not perceived by the human visual system, being also expensive to compute and to interpret.

Taken into consideration that a 16 bit color image creates a $2^{16} \times 2^{16}$ dimension of CCH a quantization of the color space need to be applied. The color space (seen as a unitary cube) is quantize uniformly, cutting the color cube into smaller cubes of a length τ (between 0 and 1).

To measure the similar keypoints using CCH we need to adapt the keypoints neighbour patches in order to be scale and rotation invariant. Adapting to scale of the patches is based on the automatic scale selection principle, using the extracted characteristic scale of the feature points. The same idea is used also to attain rotation invariance. Thus, the extracted dominant orientation of the DOG points is used to rotate the neighbour patches around the keypoints.

4.3. Representation and matching procedure

For every extracted feature point a descriptor is attached. The SIFT-CCH descriptor is a two elements vector that combines the SIFT and CCH descriptor vectors:

$$V_{SIFT-CCH} = [\alpha V_{SIFT}, (1 - \alpha) V_{CCH}] \quad (7)$$

where $V_{SIFT-CCH}$ is the resulted vector after the 128-length SIFT vector V_{SIFT} was combined with 16x16 color co-occurrence histogram vector V_{CCH} , with the relative weight factor α .

For matching the SIFT-CCH descriptor we use different metric distance to compare the similarity between the two parts of the SIFT-CCH descriptor.

Therefore, the Euclidean distance is used to measure the dissimilarity between SIFT 128 length descriptors:

$$d_{SIFT}(x, y) = \sqrt{\sum (V_{SIFT}^{(x)} - V_{SIFT}^{(y)})^2} \quad (8)$$

The similarity between two CCH is verified computing their intersection, a measure introduced in [22]:

$$I_{CCH}(x, y) = \sum_{i=1}^{n_Q} \sum_{j=1}^{n_Q} \sum_{k=1}^{n_D} \min[V_{CCH}^{(x)}, V_{CCH}^{(y)}] \quad (9)$$

where n_Q represents the quantized number of color and n_D is the number of distances used to compute CCH.

The intersection measures how well the first CCH accounts for the second one being an attractive measure since the first CCH contains extra background entries they will not affect the quality of the match. A normalized version of the intersection measure is preferred in this paper, dividing the last equation by the sum of all entries of the second CCH.

Following the indications of [4], we use a representation of 16 color levels and 4 distances between pixels.

Our approach to match SIFT-CCH descriptors combines the two metrics presented above. The procedure is relative simple to implement and can be summarized in two steps. First, the Euclidean distances between all SIFT descriptor vectors are computed using Equation 8. A good match is considered only if the distance ratio between the first best matched and the second best match is greater than 80% [3]. The second step computes the distances between CCH (Equation 9) of the keypoints that are not matched with the SIFT distance. To increase the efficiency of our matching procedure in this stage we take in the consideration only the first 10% of the keypoints that has a minimum computed Euclidean distance in the previous step.

5. Experimental results

We evaluate our descriptor taking into account different color images under different transformations (e.g. scale, rotation, small affine). The images are related by the homography (H), considering only planar scenes. To compute H, a first estimate of the homography matrix is obtained using the linear algorithm based on the best matched keypoints [23]. Then, a refined version of the H is computed using the non-linear method [23] which minimizes the Sampson cost function to geometric reprojection error.

In this paper we compare our approach with the original SIFT (www.cs.ubc.ca/~lowe/keypoints/). The figure 4a shows the measure of the incorrect matched keypoints as a function of the total number of matched points. The graphic from figure 4b plots the percent of additional correct matched keypoints found by SIFT-CCH in comparison with SIFT, as a function of the total number of matched points. As can be observed, the percent of additional correct matches is constant positive and is relative independent by the number of matched feature points.

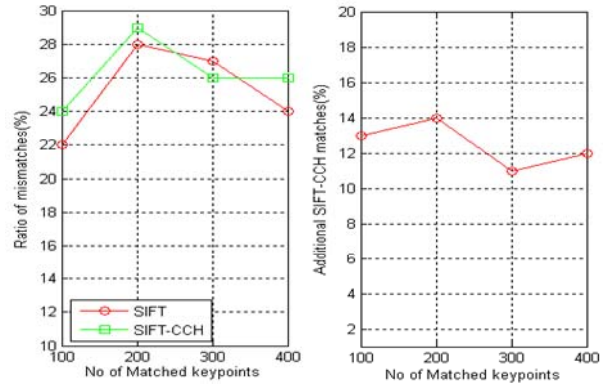


Figure 4. a) Ratio of mismatches; b) additional detected good matches with SIFT-CCH

Also, the mismatched percent remains relative constant. All these lead to the conclusion that our descriptor performs better than the SIFT descriptor.

One of the evaluated image pairs is presented in the Figure 5. The red cross points represents the SIFT good matched points and with green are drawn the additional valid correspondence pairs found with SIFT-CCH.

6. Conclusion

In this paper we introduce a new descriptor that represents an improved version of the SIFT. Our method combines the original SIFT method with the color co-occurrence histogram, a measure that is used extensively for image retrieval and object recognition applications. One of the main goals was to extend in the simple way the SIFT for color images. The method was verified in context of image matching, evaluating the number of correct matched points found in the considered images. The results show that our operator outperforms the original SIFT.

In future work we would like to develop a more effective method to detect keypoints with a higher repeatability score and stability than considered DoG detector. We believe that this will lead to better overall results of the SIFT-CCH descriptor. Additionally, we would like to focus to the affine transformation, trying to improve the performance of our descriptor when more important variation of the view angle affects images.

7. References

- [1] K. Mikolajczyk and C. Schmid, "Scale and Affine Invariant Interest Point Detectors", International Journal of Computer Vision, vol. 1, no. 60, 2004, pp. 63-86.
- [2] D. Lowe, "Object Recognition from Local Scale-Invariant Features", in Proceedings of Seventh International Conference on Computer Vision, 1999, pp. 1150-1157.
- [3] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, vol. 2, no. 60, 2004, pp. 91-110.

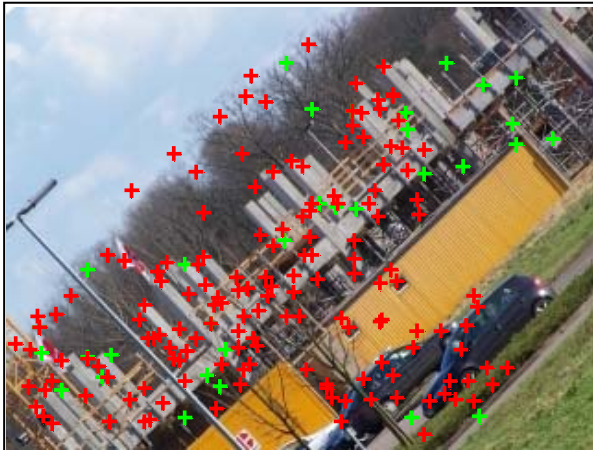


Figure 5. The size of the images is 240x320 , the ratio scale is 1.82 and the difference in orientation is 42°. DoG detector extracts 874/832 keypoints (not all shown). SIFT(red cross points) finds 176 correct and 45 incorrect matches. SIFT-CCH (red+green cross points) finds 204/59 matches.

[4] Peng Chang, J. Krumm, "Object recognition with color cooccurrence histograms", in Proceedings of Conference IEEE Computer Vision and Pattern Recognition vol. 2, 1999, pp. 504.

[5] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification", IEEE Transactions on Systems, Man and Cybernetics, vol. SMC-3, 1973, pp. 610-621.

[6] C.G.Harris and M. Stephens, "A Combined Corner and Edge Detector", in Proceedings of Fourth Alvey Vision Conference, vol. 18, 1988, pp.147-151.

[7] T. Lindeberg, "Feature Detection with Automatic Scale Selection", International Journal of Computer Vision, vol. 30, no. 2, 1998, pp. 77-116.

[8] K. Mikolajczyk and C. Schmid, "A Performance Evaluation of Local Descriptors", International Conference on Computer Vision and Pattern Recognition, 2003, pp. 257-263.

[9] L. Van Gool, , T. Moons and D. Ungureanu, "Affine/Photometric Invariants for Planar Intensity Patterns", in Proceedings of the Fourth European Conference on Computer Vision, vol. 1, 1996, pp. 642-651.

[10] S. Lazebnik, , C. Schmid, and J. Ponce, "Sparse Texture Representation Using Affine-Invariant Neighborhoods", in Proceedings of Computer Vision and Pattern Recognition, 2003, pp. 319-324.

[11] W. Freeman and E. Adelson, "The Design and Use of Steerable Filters", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, no. 9, 1991, pp. 891-906.

[12] S. Belongie, J. Malik and J. Puzicha, "Shape matching and object recognition using shape contexts," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002 24(4):509-522,

[13] H. Bay, and T. Tuytelaars and L.J. Van Gool, "SURF: Speeded Up Robust Features", European Conference on Computer Vision 2006 pp. 404-417

[14] Y. Ke, and R. Sukthankar, "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors", in Proceedings of Conference IEEE Computer Vision and Pattern Recognition, vol. 2, 2004, pp. 506-513.

[16] A.E. Abdel-Hakim, A.A. Farag, "CSIFT: A SIFT Descriptor with Color Invariant Characteristics", in Proceedings of Computer Vision and Pattern Recognition Conference, 2006, pp. 1978- 1983.

[15] E. Mortensen, H. Deng, and L. Shapiro, "A SIFT descriptor with global context", in Proceedings of Conference on Computer Vision and Pattern Recognition, 2005.

[17] J. M. Geusebroek, R. van den Boomgaard, A. W. M. Smeulders and H. Geerts, "Color Invariance", IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(12):1338-1350, 2001.

[18] G. Wyszecki and W. S. Stiles, "Color Science: Concepts and Methods, Quantitative Data and Formulae", John Wiley & sons, second edition, 1982.

[19] M. Swain and D. Ballard, "Color indexing", International Journal of Computer Vision, vol. 7(1), 1991, pp.11-32.

[20] Seong-O Shim, Tae-Sun Choi, "Image Indexing by Modified Color Co-occurrence Matrix", IEEE International Conference on Image Processing, vol. 3, 2003, pp. 493-496.

[21] Jing Huang Kumar, S.R. Mitra, Wei-Jing Zhu, R. Zabih, "Image Indexing Using Color Correlograms", in Proceedings of Conference IEEE Computer Vision and Pattern Recognition, 1997, pp. 768-762

[22] T. Gevers and A.W.M. Smeulders, "Color Based Object Recognition," Pattern Recognition, vol. 32, no. 3, pp. 453-464, 1999.

[23] R. I. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, 2000.