ICIC Express Letters
Volume **12**, Number **12**, December **2018**

ICIC International ©2018 ISSN 1881-803X
pp. 1205–1212

PROPOSED APPROACH FOR AUTOMATIC UNDERWATER OBJECT CLASSIFICATION

Mostafa Satea Mohammed[1], Hatem Awad Khater[1]
Yasser Fouad Hassan[1] and Ashraf Elsayed[1,2]

[1]Department of Mathematics and Computer Sciences
Faculty of Science
Alexandria University
Baghdad, Al Bab Al Gadid WA Mansha, Qism Moharram Bek, Alexandria Governorate, Egypt
{ mustafasatea894; hatem.a.khater }@gmail.com; { y.fouad; ashraf.elsayed }@alexu.edu.eg

[2]Zewail City of Science and Technology
Ahmed Zewail Road, October Gardens, 6th of October City, Giza 12588, Egypt

Received June 2018; accepted September 2018

Abstract. *This paper introduces an approach for automatic underwater object classification which consists of six main steps: noise removal and improve the appearances of the input image, then automatically detect and segment the objects based on salient object detection techniques from the image, extract the object features by combining the results of two feature descriptors, learning, classify the object and finally, measure the approach performance with five factors. This paper provides two major contributions: first, combine SIFT after normalization with PCA and GLCM; second, the classification step is based on salient object detection. There are multi models included within the proposed approach and this paper introduces comparisons between the obtained results.*
Keywords: Underwater image, Salient object detection, Multi-class SVM, Features combination, Supervised classification

1. **Introduction.** The main problem in the underwater imaging system is the characteristics of the environment that reflect negatively on the nature of images as well as the loss of some of the boundaries of objects or overlap with the background, which leads to the difficulty of identification and the discovery of features and classification correctly. The aim of this study is to design an underwater system that is able to detect, recognize and classify objects from underwater images based on the especially need for providing an automatic system for underwater image processing. Many researches had provided classification approaches for underwater classification. Marcos et al. [1,2] used feed-forward, back propagation neural network for classification, and they used Local Binary Pattern (LBP) as texture descriptor and Normalized Chromaticity Coordinates (NCC) or mean Hue Saturation Value (HSV) as color descriptor. Johnson-Roberson et al. [3,4] used Support Vector Machine (SVM) to classify both acoustic features and visual features separately with assigned weight which are specified empirically. Shihavuddin et al. [5] used Complete Local Binary Pattern (CLBP), Gray Level Co-occurrence Matrix (GLCM), Gabor filter response, opponent angle and hue channel color histograms as feature descriptor. For classification they used either K-Nearest Neighbor (KNN), Support Vector Machine (SVM) or Probability Density Weighted Mean Distance (PDWMD).

2. **Methodology.** The proposed approach suggests multi-technique for each step, these techniques work separately, except the feature descriptors which were combined as a contribution.

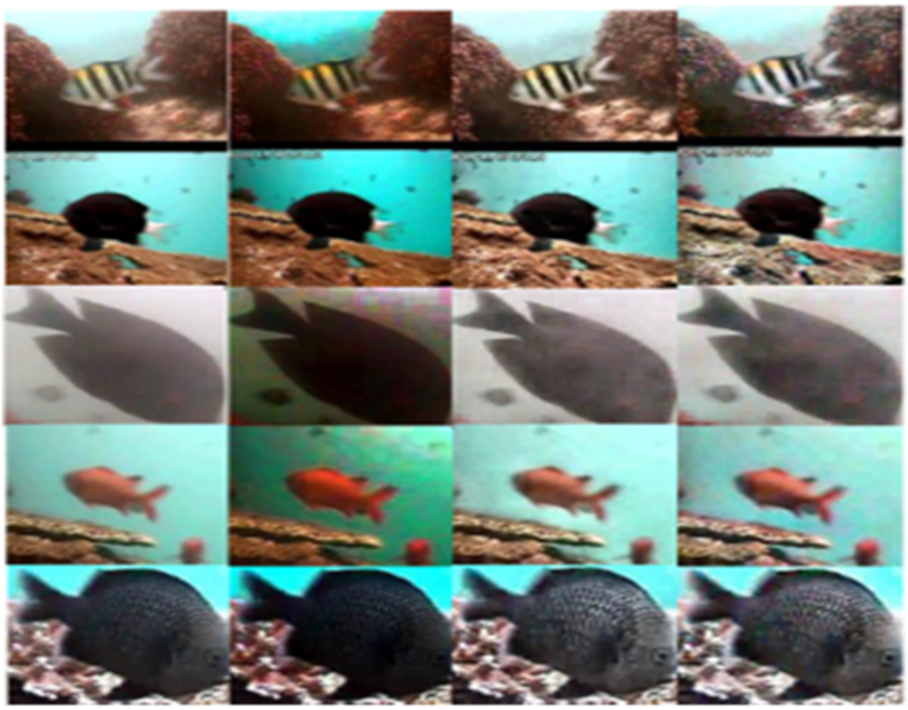
DOI: 10.24507/icicel.12.12.1205

2.1. Image enhancement.

2.1.1. *Dark Channel Prior (DCP).* The dark channel prior technique is built on the theory that says in the most of non-sky patches, at least one color channel has very low intensity at some pixels [6]. The low intensity in the dark channel has been often because of three components: colorful items, dark items and gloom [7]. The dark channel prior is effective for a variety of haze images while it may be unhelpful when the scene objects are inherently similar to the atmospheric light and no shadow is cast on them. Dark channel prior technique can be used to remove haze in underwater images because underwater images are similar to the haze images as they are all degraded by medium, also they do not conform the failure condition [8].

2.1.2. *Contrast Limited Adaptive Histogram Equalization (CLAHE).* CLAHE is used to improve the image contrast by dividing the image to several non-overlapping small regions of almost equal sizes. After dividing step, histogram of each region is calculated. Then obtain clip limit for clipping histograms according to desire limit for contrast expansion. Next, each histogram is redistributed where the height value does not pass the clip limit. Finally, Cumulative Distribution Function (CDF) [6] of the resultant contrast limited histograms is determined for grayscale mapping. CLAHE is working on different color spaces, separately RGB and HSV [9].

2.1.3. *Mixture Contrast Limited Adaptive Histogram Equalization (CLAHE-Mix).* This technique was developed for underwater image enhancement and it is operated of CLAHE on RGB and HSV color spaces and both results are mixed together using Euclidean norm [10]. The main goal of CLAHE-Mix is to reduce the undesired artifacts as well as brightness produced by CLAHE and enhance the image contrast and at the same time maintain the natural look of underwater image [11]. Figure 1 presents the results of applying the selected enhancement techniques on the underwater images: (a) test images, (b) DCP, (c) CLAHE and (d) CLAHE-Mix.



(a) (b) (c) (d)

FIGURE 1. The result of applying enhancement techniques on underwater images

2.2. **Object detection and segmentation.**

2.2.1. *High-Dimensional Color Transform (HDCT).* It is an automatic technique for detecting salient regions of an image based on the representation of the saliency map of an image as a linear combination of high-dimensional color space where salient regions and backgrounds can be clearly separated [12]. Figure 2 presents the flowchart of high-dimensional color transform.

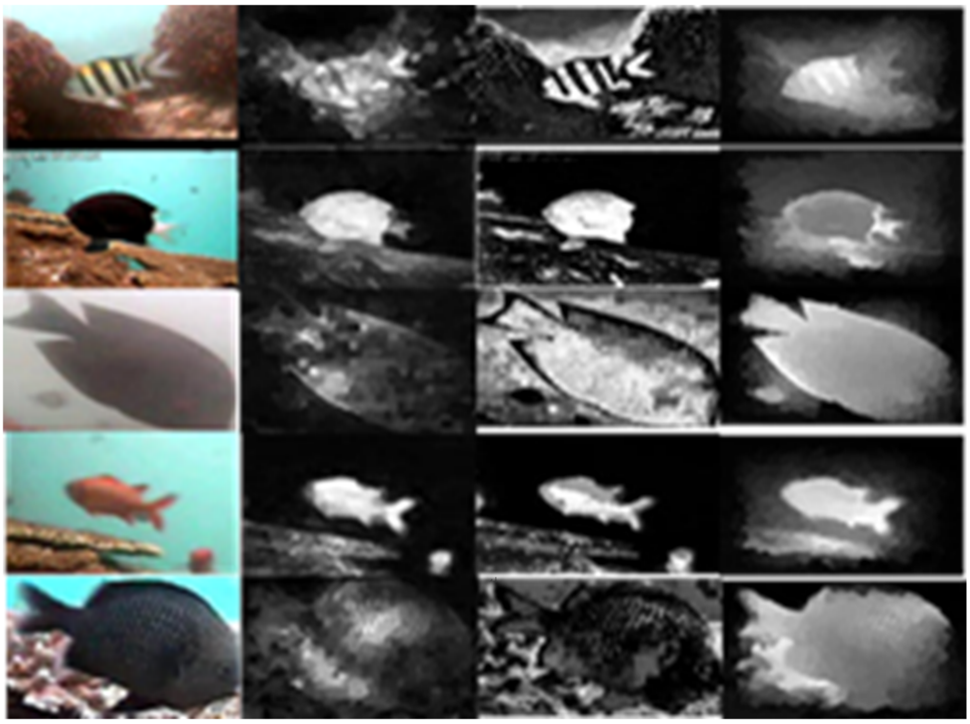
Figure 2. The flowchart of HDCT

2.2.2. *Region Principal Color based saliency detection (RPC).* RPC is a bottom-up saliency detection, which is introduced to detect salient objects in natural images, it is based on a regional principal color contrast model, which incorporates low-level and medium-level visual cues. RPC procedure consists of the following steps, start from reducing the number of pixel color to be further contrasted, build color histograms by using a quantized image, obtain global color saliency based on pixel color statistics, then segment the quantized image to regions and represent the saliency of each region as it is principle color's saliency. Finally, measure two categories of spatial relationships to produce the full resolution saliency map [13].

2.2.3. *Saliency detection via Graph-Based Manifold Ranking (GBMR).* This technique is a bottom-up saliency detection, which uses different procedures to detect salient objects, it does not consider the contrast between the salient objects and their regions, and actually, it considers both foreground and background cues by ranking the similarity of the image elements (pixels and regions) with foreground or background cues via graph-based manifold ranking [14]. Salient detection via graph-based manifold ranking procedure consists of two steps. Firstly, exploit the boundary prior [15,16] by using the nodes on each side of image as labeled background queries. Then compute the saliency of nodes based on their relevance to those queries as background labels. The labeled map is integrated to generate a saliency map. Secondly, apply binary segmentation on the saliency map obtained from the first step, and take the labelled foreground nodes as salient queries. The saliency of each node is computed based on its relevance to foreground queries for the final map [14]. All the results of the previous techniques were improved by using K-mean algorithm to determine the objects more accurately. Figure 3 presents the results of applying the selected saliency detection techniques on the underwater images: (a) test images, (b) HDCT, (c) RPC and (d) GBMR.

2.3. **Features extraction.**

2.3.1. *Scale-Invariant Feature Transform (SIFT).* SIFT is used to extract distinctive invariant features from images that can be used to perform reliable matching between different views of an object [17]. These features are invariant to scale, rotation and illumination conditions [17]. SIFT features have an advantage which is robust against distortion and addition of noise [18]. The output of SIFT is high dimensionality matrix and needs high resources such as memory space and computation time, and therefore must apply normalization method on this matrix. Principal Component Analysis (PCA)



(a) (b) (c) (d)

Figure 3. The result of applying saliency detection techniques on underwater images

is used to perform the normalization step. PCA is the most popular statistical method, which is extracted of a lower dimensional space by analyzing the covariance structure of multivariate statistical observation [19]. PCA is basically receiving $n \times m$ matrix, called M, where n is the actual number of dimensions and m is the number of feature vectors. The first step is to obtain mn which is a mean vector for each dimension. Then, mn is subtracted from every feature vector. Later, calculate $\mathrm{M} \times \mathrm{M}^{\mathrm{T}}$ covariance matrix. Subsequently, calculate the n eigenvalues with their corresponding n-dimensional eigenvectors. Finally, as a higher eigenvalue represents a higher quantity of information, PM which is kernel PCA matrix can be obtained by ordering eigenvectors according to the values of its corresponding eigenvalue from higher to lower. Each row in PM represents an eigenvector [20].

2.3.2. *Gray Level Co-occurrence Matrices (GLCM).* GLCM is a technique of extracting second order statistical texture feature which is widely used in many texture analysis applications and remains an important feature extraction method in the field of texture analysis [21]. GLCM is a matrix, its rows and columns are equal to the number of gray level G in the image. The matrix element $P(i, j|\Delta x, \Delta y)$ is the relative frequency with two pixels separated by pixel distance $(\Delta x, \Delta y)$, occurring within a given neighborhood, one with an intensity i and the other with intensity j [22]. The matrix element $P(i, j|d, \theta)$ contains the second order statistical probability values for changes between gray level i and j at particular displacement distance d and at particular angle θ [22]. The proposed approach suggests a combination of the results of SIFT after normalization with PCA and the results of GLCM, The combination is done by the following procedure.

1) The results of SIFT are $m \times m$ matrix for each image.
2) This matrix is normalized with PCA to be $m \times 2$ matrix.
3) Convert the $m \times 2$ matrix to be vector.
4) Add the features resulted from GLCM at the end of the vector.
5) The vectors are grouped to build the "feature group".

This combination will make the classification step take the advantages of each feature descriptor and will improve the overall behavior for the underwater system.

2.4. **Learning and object classification.** These two steps are discussed together because they are connected strongly and they are different according to the selected classifier.

2.4.1. *Multi-class Support Vector Machine (SVM).* The proposed approach used One-Versus-Rest (1VR) SVM schema [23] because of the attributes of this schema which is better dealing with a high number of classes in the dataset. 1VR schema builds SVM for each class. The training phase is done by assigning the samples of a class which are positive and all the other classes are negative, during the test phase, the class label is determined by the binary classifier that gives maximum output value. A major problem of the one-versus-rest approach is the imbalanced training set. 1VR used radial basis function as the kernel function because it is general purpose kernel function. Multi-class support vector machine is introduced and has a wide range of usage in application such as optical character recognition, intrusion detection, speech recognition and bioinformatics [24].

2.4.2. *Pattern Recognition Neural Network (PRNN).* The pattern recognition neural network is a supervised network using feedforward with backpropagation algorithm which is based on the concept of improving the network performance by reduction of error of output data. This network is trained to update weights and bias according to the scaled conjugate gradient method [25], the training stops when any of these conditions occurs [26].

- The maximum number of repetitions is reached.
- The maximum amount of time is exceeded.
- Performance is minimized to the goal.
- The performance gradient falls below the minimum performance gradient.
- Validation performance has increased more than the maximum validation failure-times since the last time is decreased.

2.4.3. *K-Nearest Neighbor (KNN).* KNN is a popular classification method in data mining and statistics because of its simple implementation and significant classification performance [27]. KNN is supervised non-parametric lazy learning algorithm, non-parametric means it does not make any assumption on the data distribution, while lazy learning means it does not make any data generalization, so training step does not exist or is very limited and all the training data are needed during the testing step [28]. KNN depends on the whole training data as a reference set to predict a new instance, it works by following this procedure: it finds the group of the k closest instances in the training set to the test instances. From these k neighbor instances a decision is made based on the predominance of a particular class. As a consequence, both the distance metric used to compute the closeness of the instances and the number of neighbors considered are key elements in this method. In order to find the best value for these parameters, a cross-validation procedure can be followed using the available training data [29].

3. **Experimental Results.** The dataset which is used for testing the proposed approach is used in Fish4Knowledge project [30], this dataset contains more than (20.000) fish images dividing into (15) classes, and this dataset is randomly divided into 15% for testing and 85% for training. The proposed approach was implemented with MATLAB 2016, according to the multi-technique for each step there is an essential need for testing all the techniques with each other, so there are 27 models introduced by the proposed approach, each model is built by using one technique from each step, these models are divided equally to 9 models that uses SVM, 9 models that use PRNN and 9 models that use KNN but it is important to mention that the 27 models use the same feature

group which is generated by combining the results of SIFT with PCA and the results of GLCM. Table 1 presents the best results for the 9 models that use SVM, the best result for accuracy is 77.668 which can be obtained by using CLAHE and HDCT and feature group, while the worst result is 65.461 which can be found when using CLAHE-Mix and HDCT and feature group. The minimum time for execution is 48.360 seconds which can be found by using CLAHE and GBMR and feature group, while the maximum time for execution is 65.333 second which can be found by using CLAHE-Mix and HDCT technique and feature group.

Table 1. The best results with SVM

No.	Techniques	ACC	Recall	Precision	Time
1	DCP + HDCT (Model 1)	77.358	0.769	0.769	59.287
2	DCP + RPC (Model 2)	77.215	0.769	0.769	58.125
3	DCP + GBMR (Model 3)	69.925	0.692	0.692	61.845
4	CLAHE + HDCT (Model 4)	77.668	0.769	0.769	59.985
5	CLAHE + RPC (Model 5)	77.070	0.692	0.692	56.033
6	CLAHE + GBMR (Model 6)	71.295	0.692	0.692	48.360
7	CLAHE-Mix + HDCT (Model 7)	65.461	0.615	0.615	65.333
8	CLAHE-Mix + RPC (Model 8)	77.611	0.769	0.769	51.150
9	CLAHE-Mix + GBMR (Model 9)	77.318	0.769	0.769	52.545

Table 2 presents the best results for the 9 models that use PRNN, the best result for accuracy is 96.273 which can be obtained by using CLAHE and RPC and feature group, while the worst result is 90.334 which can be found when using DCP and HDCT and feature group. The minimum time for execution is 136.710 seconds which can be found by using CLAHE-Mix and RPC and feature group, while the maximum time for execution is 457.093 seconds which can be found by using DCP and GBMR and feature group.

Table 2. The best results with PRNN

No.	Techniques	ACC	Recall	Precision	Time	HLS
1	DCP + HDCT (Model 10)	90.334	0.907	0.907	221.805	50
2	DCP + RPC (Model 11)	92.712	0.930	0.930	291.323	70
3	DCP + GBMR (Model 12)	91.571	0.919	0.919	457.095	50
4	CLAHE + HDCT (Model 13)	90.353	0.907	0.907	241.103	70
5	CLAHE + RPC (Model 14)	96.273	0.953	0.953	213.435	40
6	CLAHE + GBMR (Model 15)	91.591	0.919	0.919	314.573	60
7	CLAHE-Mix + HDCT (Model 16)	91.691	0.919	0.919	261.563	70
8	CLAHE-Mix + RPC (Model 17)	96.250	0.965	0.965	136.710	20
9	CLAHE-Mix + GBMR (Model 18)	92.615	0.930	0.930	378.510	80

Table 3 presents the best results for the 9 models that use KNN, the best result for accuracy is 92.063 which can be obtained by using DCP and RPC and feature group, while the worst result is 76 which can be found when using CLAHE-Mix and HDCT and feature group. The minimum time for execution is 6.975 seconds which can be found by using CLAHE-Mix and GBMR and feature group, while the maximum time for execution is 8.138 seconds which can be found by using RPC with any enhancement techniques and feature group.

The results show that PRNN is better than SVM and KNN in accuracy, and this is because of the following reasons: it is a non-parametric classifier, it is a universal functional approximator with arbitrary accuracy, it is a data driven self-adaptive technique efficiently handling noisy inputs and computation rate is high. The best model in accuracy

Table 3. The best results with KNN

No.	Techniques	ACC	Recall	Precision	Time
1	DCP + HDCT (Model 19)	77.244	0.769	0.769	7.440
2	DCP + RPC (Model 20)	92.063	0.923	0.923	8.138
3	DCP + GBMR (Model 21)	76.562	0.769	0.769	7.208
4	CLAHE + HDCT (Model 22)	83.471	0.833	0.833	7.440
5	CLAHE + RPC (Model 23)	91.736	0.923	0.923	8.138
6	CLAHE + GBMR (Model 24)	76.744	0.769	0.769	7.208
7	CLAHE-Mix + HDCT (Model 25)	76	0.769	0.769	7.440
8	CLAHE-Mix + RPC (Model 26)	84.615	0.857	0.857	8.138
9	CLAHE-Mix + GBMR (Model 27)	76.377	0.769	0.769	6.975

is built by using CLAHE and RPC and feature group and PRNN. The results show that KNN is better than SVM and PRNN in execution time, and this is because KNN does not have training step and does not need to update weights like PRNN. The best model in execution time is built by using CLAHE-Mix and GBMR and feature group and KNN.

4. **Conclusions and Future Work.** This study provides a fully automatic underwater object classification system with multi-model, this system uses multi-technique for each step, so there are 27 models built by combining one technique from each step. There are major differences between them, and those differences are clearly shown in the obtained results. Underwater object classification depends on the results of enhancement, segmentation, efficiency features, attributes of the selected classifiers and the characteristics of the dataset. The future work could be an underwater object classification that is able to deal with multi-object in the same image and/or overlapped objects.

REFERENCES

[1] M. S. A. Marcos, L. David, E. Peñaflor, V. Ticzon and M. Soriano, Automated benthic counting of living and non-living components in Ngedarrak Reef, Palau via subsurface underwater video, *Environmental Monitoring and Assessment*, vol.145, pp.177-184, 2008.

[2] M. Soriano, S. Marcos, C. Saloma, M. Quibilan and P. Alino, Image classification of coral reef components from underwater color video, *Oceans*, pp.1008-1013, 2001.

[3] M. Johnson-Roberson, S. Kumar, O. Pizarro and S. Willams, Stereoscopic imaging for coral segmentation and classification, *Oceans*, pp.1-6, 2006.

[4] M. Johnson-Roberson, S. Kumar and S. Willams, Segmentation and classification of coral for oceanographic surveys: A semi-supervised machine learning approach, *Oceans*, pp.1-6, 2007.

[5] A. Shihavuddin, N. Gracias, R. Garcia, A. C. Gleason and B. Gintert, Image-based coral reef classification and thematic mapping, *Remote Sensing*, vol.5, pp.1809-1841, 2013.

[6] K. He, J. Sun and X. Tang, Single image haze removal using dark channel prior, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.33, pp.2341-2353, 2011.

[7] J. K. Babool and S. Singh, A survey of old technique of single underwater image enhancement, *Imperial Journal of Interdisciplinary Research*, vol.2, 2016.

[8] Y. Zhu, L. Chang, J. Dai, H. Zheng and B. Zheng, Automatic object detection and segmentation from underwater images via saliency-based region merging, *Oceans*, pp.1-4, 2016.

[9] A. M. Reza, Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement, *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*, vol.38, pp.35-44, 2004.

[10] M. S. Hitam, E. A. Awalludin, W. N. J. H. W. Yussof and Z. Bachok, Mixture contrast limited adaptive histogram equalization for underwater image enhancement, *International Conference on Computer Applications Technology (ICCAT)*, pp.1-5, 2013.

[11] M. K. Saggu and S. Singh, Haze removal from underwater images using mixture contrast limited adaptive histogram equalization, *International Journal for Scientific Research & Development*, vol.3, 2015.

[12] J. Kim, D. Han, Y.-W. Tai and J. Kim, Salient region detection via high-dimensional color transform and local spatial support, *IEEE Trans. Image Processing*, vol.25, pp.9-23, 2016.

[13] J. Lou, M. Ren and H. Wang, Regional principal color based saliency detection, *PloS One*, vol.9, 2014.

[14] C. Yang, L. Zhang, H. Lu, X. Ruan and M.-H. Yang, Saliency detection via graph-based manifold ranking, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.3166-3173, 2013.

[15] L. Grady, M.-P. Jolly and A. Seitz, Segmentation from a box, *IEEE International Conference on Computer Vision (ICCV)*, pp.367-374, 2011.

[16] V. Lempitsky, P. Kohli, C. Rother and T. Sharp, Image segmentation with a bounding box prior, *IEEE the 12th International Conference on Computer Vision*, pp.277-284, 2009.

[17] D. G. Lowe, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, vol.60, pp.91-110, 2004.

[18] N. Najva and K. E. Bijoy, SIFT and tensor based object detection and classification in videos using deep neural networks, *Procedia Computer Science*, vol.93, pp.351-358, 2016.

[19] D. Engel, L. Hüttenberger and B. Hamann, A survey of dimension reduction methods for high-dimensional data analysis and visualization, *OASIcs-OpenAccess Series in Informatics*, 2012.

[20] R. E. G. Valenzuela, W. R. Schwartz and H. Pedrini, Dimensionality reduction through PCA over SIFT and SURF descriptors, *IEEE the 11th International Conference on Cybernetic Intelligent Systems (CIS)*, pp.58-63, 2012.

[21] B. Kannan, A. Unnikrishnan and S. V. Bino, *Grey Level Co-Occurrence Matrices: Generalisation and Some New Features*, 2012.

[22] S. A. Medjahed, A comparative study of feature extraction methods in images classification, *International Journal of Image, Graphics and Signal Processing*, vol.7, 2015.

[23] V. N. Vapnik, An overview of statistical learning theory, *IEEE Trans. Neural Networks*, vol.10, pp.988-999, 1999.

[24] Z. Wang and X. Xue, Multi-class support vector machine, *Support Vector Machines Applications*, pp.23-48, 2014.

[25] M. F. Møller, A scaled conjugate gradient algorithm for fast supervised learning, *Neural Networks*, vol.6, pp.525-533, 1993.

[26] *https://ch.mathworks.com/help/nnet/ref/trainscg.html/*.

[27] S. Zhang, X. Li, M. Zong, X. Zhu and R. Wang, Efficient KNN classification with different numbers of nearest neighbors, *IEEE Trans. Neural Networks and Learning Systems*, 2017.

[28] *https://saravananthirumuruganathan.wordpress.com/2010/05/17/a-detailed-introduction-to-k-nearest-neighbor-knn-algorithm/*.

[29] M. Galar, J. Derrac, D. Peralta, I. Triguero, D. Paternain, C. Lopez-Molina, S. García, J. M. Benítez, M. Pagola and E. Barrenechea, A survey of fingerprint classification part II: experimental analysis and ensemble proposal, *Knowledge-Based Systems*, vol.81, pp.98-116, 2015.

[30] *http://groups.inf.ed.ac.uk/f4k/*.