Contemporary Engineering Sciences, Vol. 10, 2017, no. 28, 1351 - 1363
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ces.2017.711157

Comparison between CNN and Haar Classifiers for Surgical Instrumentation Classification

Paula C. Useche Murillo

Nueva Granada Military University
Bogotá, Colombia

Robinson Jiménez Moreno

Nueva Granada Military University
Bogotá, Colombia

Javier O. Pinzón Arenas

Nueva Granada Military University
Bogotá, Colombia

Abstract

The following paper presents the development, operation and comparison of two methods of object recognition trained for the classification of surgical instrumentation, where a video sequence is used to capture scene information constantly, in order to allow the selection of some of the instruments according to the needs of the doctor. The methods used were Convolutional Neural Networks (CNN) and Haar classifiers, where the first was added a previous element detection stage, and the second one was conditioned to allow it not only to detect elements, but also to classify them. With the CNN an accuracy of 96.4% in the classification of the two categories of the first branch of the tree was reached, while for Haar classifiers 90% accuracy was achieved in the detection of one of the five instruments, whose classifier was the one that presented the best results.

Keywords: Haar Classifiers, Convolutional Neural Network, Object Detection, Surgical Instrumentation Classification

1 Introduction

The Convolutional Neural Networks are artificial intelligence techniques that have been used in the field of pattern recognition since 1998, in the classification of images since 2012 [1], and more recently in the diagnosis of cardiovascular diseases from ECG signals developed by [2] in November of 2017. The versatility offered by these networks to classify different types of categories is due to the fact that they do not have a fixed architecture or depth, but that the user can set and define said characteristics according to the application's requirements, as presented in [3] where the different layers that can be added to the architecture and the function of each of them are explained.

The CNN have come to cover various applications such as face recognition [4], the classification of hand positions as open or closed [5], the classification of characters in a written document [6], the recognition of the face pose [7], and many other applications such as object tracking, scenario classification and speech recognition and language processing as presented in [8].

On the other hand, other techniques of artificial intelligence have been used for the recognition of patterns in images such as those applied in [9], where three different types of classifiers are used for the identification of vehicles in order to allow a car to detect its surroundings and to handle itself autonomously. Among these techniques, there are the cascading classifiers such as the Histogram of Oriented Gradients (HOG), based on the gradient of the image, the Local Binary Patterns (LBP), based on the difference between neighboring pixels of small regions of the image, and Haar, based on Haar type filters. Similarly, in [10] these same classifiers are used for the detection of pedestrians on the road, with training images taken from surveillance cameras from different angles and altitudes.

Cascading classifiers such as Haar type have been widely used in detection of faces [11], objects [12] and other characteristics such as body temperature [13], and the detection of a species of fish underwater [14]. Their detection capabilities are due to the fact that these classifiers work by evaluating an "integral image", obtained from an original image, where fractions of the image considered as background are continuously eliminated, to finally evaluate those aspects that coincide with the elements to be detected, as explained in [15] and as shown in Figure 1, in which a series of classifiers were trained to detect faces in scenes.

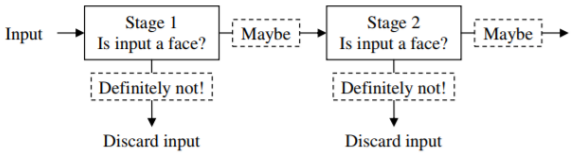


Figure 1. Cascading classifiers for face detection

In Figure 1, each classifier or Stage evaluates the image and generates two outputs: possible face (Maybe) or definitely a background (Definitely not!), and the process is repeated until the last classifier discards the last fractions of the image considered as background and generates face detection as output.

As can be seen, the Haar classifiers have been implemented to a large extent for the detection of various elements in scenes and not for the classification of categories, due to their structure and the functioning of their classifiers, so it was decided to look for ways to add the classifier function and thus allow it to generate both detection and classification.

Consequently, the following work presents a performance comparison between an object classifier by CNNs and an element detection algorithm by Haar classifiers trained to generate the detection and detailed classification of different elements of surgical instrumentation. Said comparison allows to observe and analyze the advantages of a pattern recognition method with respect to the other, and with that select the most convenient for a certain application according to the characteristics required by it. For each test, controlled light conditions and blue colored backgrounds were used to simulate a hospital work environment, constant and invariable.

Below are the 5 sections in which the article was divided, where the first shows the working conditions established for the tests, the second presents the tree of CNNs developed for the classification of surgical instrumentation, the third explains the functioning of the Haar classifiers, their adaptability in order to generate a classification of categories and the classifiers trained for the recognition of each tool, in the fourth, the results obtained by both methods and an analysis of their behavior are presented, and in the last section the conclusions reached are presented according to the results obtained.

2 Methods and materials

To do the comparison between the previously mentioned recognition algorithms, CNN and Haar classifiers, the 5 surgical instruments shown in Figure 2 were taken as the training categories and they were trained with only blue toned backgrounds, square pictures of 480x480 pixels were taken for each tool and two databases were built, one original with 300 training images per tool, and another increased with 2000 images per tool.

The camera was placed at a fixed distance of 31cm from the ground, both to capture the database and to perform the classification tests, and a white light lamp was used to ensure constant light conditions. Each of the working conditions were set in such a way that they resemble a surgical environment, whose light conditions and work surfaces are regulated and invariable.

1. Probe | 2. Dissecting Needle | 3. Splinter Forceps | 4. Big Scissors | 5. Round Scissors

Figure 2. Surgical instruments to classify

The original database consists of 300 images per tool and was used for the training of the CNNs, while the database of 2000 images per category was used for the training of the Haar classifiers as positive images, and was obtained by means of an augmentation of the original database, where rotations were generated every 5° to maximum 15° of rotation in counterclockwise direction, background changes, changes of position, and noise was added to some images. Additionally, a database of 5000 negative images was created, cutting part of the blue background from the original images, and augmenting them in the same way as the positive images.

3 Tree-structured CNN

A tool classifier was designed using four CNNs structured in the form of a tree for the identification of the five surgical instruments shown in Figure 2. The structure of the classification tree used is shown in Figure 3, where in each branch there is a CNN that classifies the input image (dimensions 128x128 pixels in color), in one of the following categories, for instance "Others" or "Scissors", and depending on which category is classified, the image goes to another CNN that generates a classification again, "Probe" or "Thin" for example, and so on until the tree leaves are reached where the type of tool entered is defined.

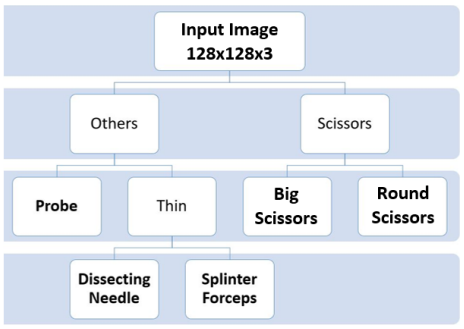


Figure 3. Tree-structured CNN

In Table 1 the parameters of each CNN of the tree are specified, where the light colored cells represent convolution layers followed by ReLified (Rectified Linear Units) layers and the dark cells represent MaxPooling layers, Fw and Fh are the

width and height of the kernel respectively, and the percentage of accuracy of each network individually is shown in the last row of the table.
For the "General" network, a large initial filter was used in the first convolution to highlight the most differential characteristics between both categories, such as the eye rings of the scissors and the thin body of the other instruments. For the "Others" network, large filters were also used to highlight the thin body of the instruments of the "Thin" category and the width of the "Probe" instrument. In the "Scissors" network, on the other hand, small filters were used to focus the training on the details of the tools, given that their overall geometry is very similar. Finally, for the "Thin" network, rectangular filters of wider than high were used to highlight the tips of the "Splinter Forceps" against of the thin body of the "Dissecting Needle Curved".

Table 1. CNN Architectures for the Classification Tree

CNN ARCHITECTURE					
NETWORK	**GENERAL**	**OTHERS**	**SCISSORS**	**THIN**	
DIMENSIONS	Fw, Fh	Fw, Fh	Fw, Fh	Fh	Fw
CLASSIFICATION TREE BY CNN	11F 3s	15	5	2	16
	2F 2s	13	4	2	14
	5	**2F 2s**	**3F 2s**	**2F 2s**	**2F 2s**
	2F 2s	11	3	2	12
	3	9	3	2	10
	3	**3F 2s**	**2F 2s**	**3F 2s**	**2F 2s**
	3	7	3	2	8
		5	3	2	6
			2F 2s		
			3		
			3		
			2F 2s		
Number of Fully Connected	2	2	2	2	2
Accuracy	**96.40%**	**95.00%**	**91%**	**91%**	

In equations 1 to 3 the calculations are shown to obtain the dimensions of the output volume of each layer, where n number the convolution layers and Pooling, W and H indicate the width and height of the image respectively, P the padding, s the stride, D the depth of the layer and K the number of filters of the previous layer. In this case, the volume is reduced in different proportions for each of the trained networks, in order to achieve a more focused extraction of the details of the image ("Scissors" network with small filters), or more global ("General" and "Others" network with large filters).

$$W_{n+1} = \frac{W_n - F_w + 2P}{s} + 1 \tag{1}$$

$$H_{n+1} = \frac{H_n - F_h + 2P}{s} + 1 \tag{2}$$

$$D_{n+1} = K_n \tag{3}$$

Figure 4 shows the confusion matrix for the tree-structured CNN with 50 test images per tool, where the numbers from 1 to 5 represent each of the trained categories, numbered as indicated in Figure 2, the rows indicate the classification generated by the network, the columns the real categories, and the diagonal cells from the upper left corner to the lower right represent the number of correctly classified images, where it can be seen the overall accuracy of the network, of 96%, and the way in which it classifies each image with respect to all the categories.

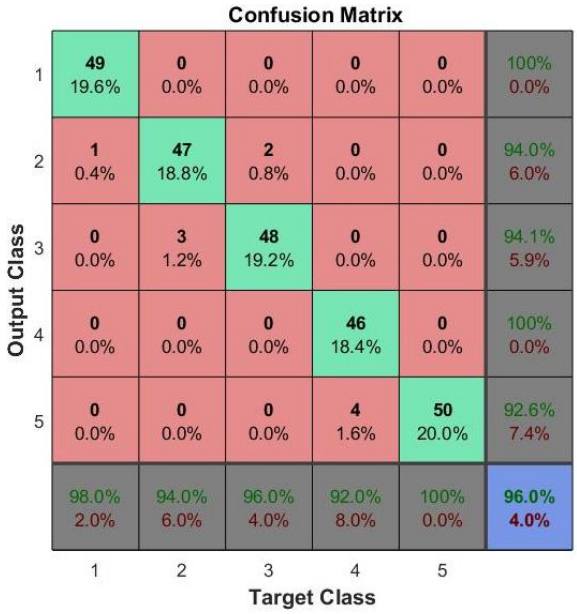


Figure 4. Confusion matrix for the tree-structured CNN

4 Haar classifier

Haar classifiers consist of a series of weak cascading classifiers, where each of them is trained to detect the presence or not of a particular object in an image. In each classifier or stage of the network, a window is slid over the image and the information of said window is evaluated, in case the classifier does not find the desired element, the window is defined as a negative and ignored, the window is repositioned and the object is searched again, otherwise, the window passes to the next stage as a positive to be evaluated again by another classifier. Only that window that manages to pass through all stages as a positive represents a detection [16].

Some of the relevant terms when talking about Haar classifiers are shown in Table 2 and are detailed in [16].

Table 2. Terms for classification by Haar

Term	Description	Symbol
True positive	Successful classification of positive images	
False positive	Misclassification of negatives as positive	
False negative	Misclassification of positives and negatives	
FalseAlarmRate	Percentage of acceptable false positives per stage	Far
NumCascadeStages	Number of cascading stages or classifiers	nCS
NegativeSamplesFactor	Defines the percentage of negative images to be used with respect to the number of positive images.	Nsf
TruePositiveRate	Minimum percentage of positive images to train per stage	Tpr
ObjectTrainingSize	Minimum size of the sliding search window ("w" width, "h" height)	Ots
TotalPositiveSamples	Total number of positive images	Tps
NegativeSamples	Number of negative images per stage	Ns
NumberPositiveSamples	Number of positive images per stage	Nps

In (4) the calculation is shown to obtain the quantity of positive images to be used per stage (Nps) and in (5) the calculation of the number of negative images (Ns). However, during training, it can be used more images than calculated depending on how many negatives and positives generated by the classifiers, which means that the estimated number of stages (nCs) for the classifier is not always completed [16].

$$Nps = \frac{Tps}{1 + (nCS - 1)(1 - Tpr)} \tag{4}$$

$$Ns = Nps * Nsf \tag{5}$$

The main application of the Haar classifiers is to perform the detection of a certain element within a defined environment, however, its structure does not allow classifying categories of elements as if it happens with a CNN. Therefore, 5 Haar classifiers were trained, one for the detection of each one of the surgical instruments, and with each of them the detection and classification of tools was performed.

The parameters nCS, Tpr, Nsf, Far and Ots (in pixels) were varied for the training of Haar type classifiers until reaching those that generated the least number of false positives during video tests. The best classifiers obtained were trained with the parameters of Table 3, where the last row shows the number of stages that the classifier trained before exhausting the training images of the database.

Table 3. Training parameters for Haar classifiers

Parameters	*Probe*	*Big Scissors*	*Round Scissors*	*Dissecting NC*	*Splinter Forceps*
nCS	100	100	100	100	100
Tpr	0.1%	0.5%	0.05%	0.5%	0.05%
Nsf	2	2	2	2	2
Ots [Height Width]	[115 32]	[100 30]	[85 65]	[150 40]	[100 35]
Far	1%	1%	1%	1%	1%
Stages	4	6	3	3	8

During training, it was observed that by proposing a high number of stages (nCS) for each classifier and setting a small percentage of positive images per stage (Tpr), a reduction in the number of false positives was achieved, even though the training stopped before reaching the 10 stages. On the other hand, establishing the Ots for each instrument improved the recognition capabilities of the classifier, since the windows were generated with the expected proportions for each tool, achieving a better capture of the element with a reduction in false positives, but with an increase in training time.

To observe the behavior of each of the classifiers obtained, 24 test images of each instrument were taken and individually evaluated with their respective classifiers, where it was recorded the number of images in which the instrument was detected (Detected), the amount in which only false positives were detected (False Positive) and the number of images without detection (Not Detected), tabulating the results in Table 4, where the last column shows the percentage of accuracy achieved by evaluating the number of successful detections of each instrument with respect to the total number of images in which there was some type of detection.

Table 4. Accuracies of the Haar classifier

	Detected	**False positives**	**Not detected**	**Accuracy**
Probe	9	3	12	75%
Dissecting Needle Curved	18	2	4	90%
Splinter Forceps	8	0	16	100%
Big Scissors	1	3	20	25%
Round Scissors	9	7	8	56.25%

Despite obtaining 100% accuracy of the "Splinter Forceps" tool, the number of images without detection was considerably high, greater than 50% of the total number of test images, so it can be seen a difficulty in detecting said instrument, while in the case of "Dissecting Needle Curved", 90% accuracy was achieved with only 4 images without detection and 2 with false positives, which makes it the best trained classifier with respect to the others.

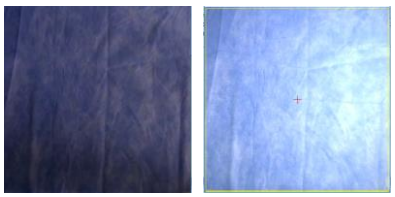
5 Results and analysis

For the CNN training, the images in Figure 2 were taken and cropped in such a way that the tool would occupy the whole image, in order to focus the training of the network exclusively on the tool and not on the background. Subsequently, several CNNs were trained with the images cropped to obtain the networks of Table 1 and the tree of Figure 3 was structured to be able to perform the classification of the surgical instrumentation.

On the other hand, the tool detection algorithm developed allows recognizing objects on a static and invariable background. Its operation is based on a first capture of the work area that allows to obtain information of each one of its pixels,

average them and save them as a background. In the following frames, all the pixels of the image are compared with those stored as background and the presence of a tool is indicated when the difference between a group of 10 pixels or more and the background is higher than a certain threshold. Once the tool is detected, a box is generated on it and a square cut is made to enter only that fraction of the image to the CNN and classify it.

This algorithm allows to recognize multiple elements in a scene, however, it presents problems when there are sudden changes of illumination, because the average of the background changes and begins to be detected as an element, as can be seen in Figure 5b, where a box has been formed around the entire background and the center of the box has been demarcated with a "+", with Figure 5a being the average of the background initially captured.



a. b.

Figure 5. Lighting changes

Under a lighting condition like the one in Figure 5b, the tree-structured CNN was used for the classification of the 5 tools obtaining the results of Figure 6a, where it can be seen a correct classification for 3 of the 5 tools, while with lighting conditions such as those in Figure 5a it was possible to classify correctly the "Probe" tool, as shown in Figure 6b. However, when comparing the recognition dbetween Probe and Dissecting Needle Curved of Figure 6b with Figure 6d, better classification results were found using Haar classifiers than the tree-structured CNN.

a. b. c. d.

Figure 6. Classification by tree-structured CNN and Haar classifiers

On the other hand, Figure 6c shows the classification of the 5 surgical instruments using Haar classifiers, under the lighting conditions of Figure 5b. In the present case, each classifier manages to detect and properly identify four of the five tools, however, false positives are presented on two of them: "Round Scissors" for the eye

rings of the Big Scissors, and "Probe" on the body of the Dissecting Needle Curved. These false positives are generated from the similarity between a section of the instrument with any of the other tools, causing a double classification.

In Figure 7a, it is illustrated an example of false positives generated due to the change of illumination, where, unlike the detection algorithm for CNN, elements can still be recognized without the whole background being captured as an instrument, but the quality of the classification is considerably reduced because of the amount of false positives.

a. b. c.

Figure 7. Classification by Haar classifiers with low lighting and tree-structured CNN with lighting changes

In the same way, light affects the classification for the tree-structured CNN, as shown in Figure 7, where Figure 7b shows the classification with controlled light conditions, and Figure 7c shows the classification of the same tools with poor lighting, where only the Splinter Forceps can be classified correctly. In both cases a previous average of the background was obtained (with and without controlled light) to avoid the problem posed in Figure 5 and with it being able to recognize all the tools in the workspace.

The CNN recognition algorithm allows to detect each of the tools only once, which eliminates the possibility of classifying the same instrument twice, however, the detection depends a lot on the difference of tones between the instrument and the background, so there is a risk of losing information related the tool because of its similarity with the background given that a smaller box is generated around the instrument and a crop with incomplete information is entered into the CNN.

On the other hand, the time it takes the algorithms to recognize and classify the tools varies considerably between both methods, as shown in Figure 8, where the time taken by each algorithm to analyze each frame was plotted for cases in which there are no tools to recognize (Figure 8a and Figure 8c) and when the 5 tools are found in the workspace (Figure 8b and Figure 8d). The horizontal axis represents the frame, and the vertical axis the time used in seconds.

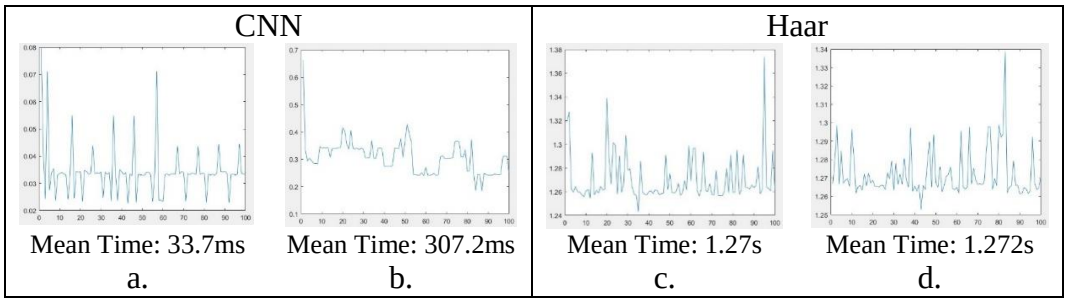


Figure 8. Recognition and classification times for the CNN tree (a, c) and for Haar classifiers (b, d)

As shown in the graphs in Figure 8, CNN demonstrates having better execution times than Haar classifiers, achieving a classification in 300ms and improving its time to 33ms when there is no tool, while Haar classifiers take more than 1s to recognize and classify the tools, even when there are none in the workspace. This difference is due to the fact that Haar classifiers constantly pass a sliding window, of different sizes, throughout the image in search of tools, evaluating multiple times the image for each of the instruments, while CNN only evaluates the tools that are extracted from the recognition phase, without focusing on other parts of the image.

6 Conclusions

It is possible to adapt the Haar classifiers as detection and classification algorithms for five surgical instruments, however, it is necessary to train a classifier for each instrument and continuously test the results with video tests to choose the best one, while with the CNN, confusion matrices are used to observe the classification of each test image with each of the categories and thus select the one with the best classification percentages.

Both the CNNs and the Haar classifiers have drawbacks for the classification of tools against changes in lighting, however, the CNN detection algorithm limits the degree of light variation, given that very abrupt changes cause the entire background to be recognized as an element, while the Haar classifiers continue to recognize instruments, but with an increase in false positives.

For real-time applications it is not really efficient to use Haar classifiers due to the high execution time it requires, while CNN manages to capture information in shorter periods of time thus allowing responding to changes or variations in the workspace over 300ms.

Haar classifiers, compared to CNNs, have the advantage of having the ability to perform detection and classification simultaneously, which eliminates the need to design an additional detection algorithm that extracts the image from the tool and enters it into a classifier, where information losses may arise due to similarities

between the desired object and the background. Additionally, Haar classifiers do not require changes in their structure to improve the quality of recognition, as it does when defining the architecture of a CNN, but a variation in their parameters to change the number of stages and positive and negative training images, making it easier to implement than a CNN.

Acknowledgements. The authors are grateful to the Nueva Granada Military University, which, through its Vice chancellor for research, finances the present project with code IMP-ING-2290 and titled "Prototype of robot assistance for surgery", from which the present work is derived.

References

[1] N. S. Velandia, R. D. H. Beleno and R. J. Moreno, Applications of Deep Neural Networks, *International Journal of Signal System Control and Engineering Application*, **10** (2017), 61-76.

[2] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. *Information Sciences*, **415-416** (2017), 190-198. https://doi.org/10.1016/j.ins.2017.06.027

[3] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks, In *Advances in Neural Information Processing Systems*, 2012, 1097-1105.

[4] S. Lawrence, C. L. Giles, Ah Chung Tsoi, A.D. Back, Face recognition: A convolutional neural-network approach, *IEEE Transactions on Neural Networks*, **8** (1997), no. 1, 98-113. https://doi.org/10.1109/72.554195

[5] J. O. P. Arenas, P. C. U. Murillo and R. J. Moreno, Convolutional neural network architecture for hand gesture recognition, *2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, (2017), 1-4. https://doi.org/10.1109/intercon.2017.8079644

[6] P. Y. Simard, D. Steinkraus, J. C. Platt, Best practices for convolutional neural networks applied to visual document analysis, *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings*, (2003), 958-963. https://doi.org/10.1109/icdar.2003.1227801

[7] X. Yin, X. Liu, Multi-Task Convolutional Neural Network for Pose-Invariant Face Recognition, *IEEE Transactions on Image Processing*, **27** (2018), 964-975. https://doi.org/10.1109/tip.2017.2765830

[8] J. Gu, Z.Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, T. Chen, Recent advances in convolutional neural networks, *Pattern Recognition*, (2017). https://doi.org/10.1016/j.patcog.2017.10.013

[9] Daniel Neumann, T. Langner, F. Ulbrich, D. Spitta, D. Goehring, Online vehicle detection using Haar-like, LBP and HOG feature based image classifiers with stereo vision preselection, *2017 IEEE Intelligent Vehicles Symposium (IV)*, (2017), 773-778. https://doi.org/10.1109/ivs.2017.7995810

[10] W.G. Aguilar, M. A. Luna, J. F. Moya, V. Abad, H. Ruiz, H. Parra, W. Lopez, Cascade Classifiers and Saliency Maps Based People Detection, In: *Augmented Reality, Virtual Reality, and Computer Graphics. AVR 2017*, De Paolis L., Bourdot P., Mongelli A. (eds), Lecture Notes in Computer Science, Vol. 10325, Springer, Cham, 2017 501-510. https://doi.org/10.1007/978-3-319-60928-7_42

[11] P. Viola, M. J. Jones. Robust real-time face detection, *International Journal of Computer Vision*, **57** (2004), no. 2, 137-154. https://doi.org/10.1023/b:visi.0000013087.49260.fb

[12] S. Choudhury, S. P. Chattopadhyay, T. K. Hazra, Vehicle detection and counting using haar feature-based classifier, *2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON)*, (2017), 106-109. https://doi.org/10.1109/iemecon.2017.8079571

[13] C. H. Setjo, Balza Achmad, Faridah, Thermal image human detection using Haar-cascade classifier, *2017 7th International Annual Engineering Seminar (InAES)*, (2017), 1-6. https://doi.org/10.1109/inaes.2017.8068554

[14] B. Benson, J. Cho, D. Goshorn, Ryan Kastner, Field programmable gate array (FPGA) based fish detection using Haar classifiers, American Academy of Underwater Sciences, 2009.

[15] O. H. Jensen, *Implementing the Viola-Jones Face Detection Algorithm*, Master's Thesis, Technical University of Denmark, DTU, DK-2800 Kgs. Lyngby, Denmark, 2008.

[16] MATHWORKS, Train a Cascade Object Detector, (20 October 2017) [Online] Available in: https://es.mathworks.com/help/vision/ug/train-a-cascade-object-detector.html

Received: November 15, 2017; Published: December 10, 2017