GEOMETRICAL-BASED APPROACH FOR ROBUST HUMAN IMAGE DETECTION

Malek Al-Nawashi[1], Obaida M. Al-Hazaimeh [2] , Mohamad Saraee [3]

[1] School of Computing, Science and Engineering, University of Salford-Manchester, UK;
m.m.al-nawashi@edu.salford.uk
[2] Department of Computer Science, Al- Balqa Applied University, Jordan;
dr_obaida@bau.edu.jo
[3] School of Computing, Science and Engineering, University of Salford-Manchester, UK;
m.saraee@salford.ac.uk

Corresponding author: Obaida M. Al-hazaimeh, Computer Science Dep., AL- BALQA Applied University, AL-HUSON University College, Jordan - Irbid, Al-Huson, 5050, Mobile: 00962775533301, E-mail: dr_obaida@bau.edu.jo

Abstract

In recent years, object detection and classification has been gaining more attention, thus, there are several human object detection algorithms being used to locate and recognize human objects in images. The research of image processing and analyzing based on human shape is one of the hot topic due to the wide applicability in real applications. In this paper, we present a new object classification approach. The new approach will use a simple and robust geometrical model to classify the detected object as human or non-human in the images. In the proposed approach, the object is detected. Then the detected object under different conditions can be accurately classified (i.e. human, non-human) by combining the features that are extracted from the upper portion of the contour and the proposed geometrical model parameters. A software-based simulation using Matlab was performed using INRIA dataset and the obtained results are validated by comparing with five state-of-art approaches in literature and some of the machine learning approaches such as artificial neural networks (ANN), support vector machine (SVM), and random forest (RF). The experimental results show that the proposed object classification approach is efficient and achieved a comparable accuracy to other machine learning approaches and other state-of-art approaches.

Keywords: Human classification, Geometrical model, INRIA, Machine learning, SVM, ANN, Random forest.

1. Introduction

Human object detection is one of the most active research topics in computer vision. Human objects detection can be simply defined as the process of localizing all objects that are human in the images by detecting human features [1]. For a robust detection of human object in images, we need the power of computer vision algorithms to be able to extract common features among

different objects (i.e. Human and Non-human objects). This task has become a quite challenging for researchers in computer vision area due to the fact that different objects tend to have different features which are usually used for object recognition [2-4].

Generally, the human detection algorithm considers as a process based multi-tasking which means two or more processes can be run concurrently such as detect the human, motion and behavior detection (i.e. Normal, and abnormal behavior), recognize the person based on his face (i.e. Face recognition), and counting the number of humans in images [5-8]. Therefore, human detection algorithms may use in a number of different applications such as security purpose in highly sensitive areas (i.e. airports, train stations, supermarkets, etc.) and so many surveillance application systems (i.e. driving assistant) [7]. Despite of all the benefits, human detection algorithm is facing some challenges to detect the human form the images and videos such as dynamically changing in the background and illumination, and varying camera positions. Therefore, appropriate algorithms are needed to manage such these challenges to ensure the quality service of human detection algorithm (i.e. classification accuracy) [3, 9].

Human detection topic in computer vision area has been adopted by many researchers. Most of these researches a high resolution direct face is required or need entire body (i.e. object) to be visible in the images and videos, and a huge database for matching based classification are required. Generally, there are four major categories of object detection methods, i.e. dynamic threshold, flow analysis, background subtraction, and temporal differencing [10]. Moreover, there are four major categories of object classification: shape-based, texture-based, motion-based, and color-based [11].

Machine learning approaches have been widely used in classifying a detected object (i.e. human, non-human) [12]. Machine learning is linked to artificial intelligence (AI), and it is designed using different approaches (i.e. Support Vector Machine model, Artificial Neural Network, Random Forest, etc.) in order to allow a computer to learn by finding statistical regularities (i.e. conditional probabilities) on a set of data. The statistical regularities are used to classify the detected object by means of previous experience, analysis, and self-training [13].

In this paper, we present a geometrical-based approach that can emulate the machine learning algorithms for classifying the detected object as human or non-human in images. The main idea of this approach is using proposed geometrical model parameters to classify the detected object based on some shape features which extracted from the object upper portion. Due the upper portion is always visible and not easy to cover or disappear. Furthermore, the proposed approach can potentially reduce the classification challenges such as illumination, varying camera positions, and human object pose.

This paper will be arranged as follows. Sect. 2, is devoted to providing an extensive overview of the related work on human object detection algorithms and in Sect. 3, the proposed object classification approach is described and in Sect. 4, the experimental results of the proposed approach are presented and in Sect. 5, the evaluation processes of the proposed object classification approach are described and in Sect. 6, the performance analysis between the proposed object classification approach and some of machine approaches are discussed. Finally, the conclusion is presented in Sect. 7.

2. Related works

In this section, we report an extensive literature on the algorithms and methods relevant to motion detection and classification methods, specifically, algorithms for detecting a human object. For accurate detection, the human object must be accurately detected using suitable methods. Many researchers have turned their attention to propose new approaches for human detections, but the new approaches have a number of practical problems, such as dynamically changing in the background and illumination, varying camera positions and human object pose, and clothing and texture parameters [14-16].

Zhong et al. [17] improved the performance of object detection and tracking (i.e. human, car, van, etc.) using a prototype-based deformable template models. The proposed method used a criterion which combines two terms: frame-to-frame deviations of the object shape, and the fidelity of the modeled shape of the input image to track an object in an image sequence. This method is fast and achieves better detection and tracking performance compared to other methods.

Desa et al. [18] proposed a new method for object detection, the proposed object detection method is a combination of two approaches i.e. background subtraction, and temporal differencing. The implementation results of the proposed method show improving the performance in detecting object compared to implement each approach separately. Sugandi et al. [19] utilized a robust approach for moving object detection and tracking based on temporal difference method on low resolution image.

Yuhua et al. [20] presented a new approach for human detect object in video. In this approach a set of parameters was built based on the interest object for the object classification process, and then using a set of sample data which consist of both negative and positive samples been fed to the machine learning classifier. The result shows a high accuracy for object classification, but the proposed approach requires a huge database for matching based classification.

Dalal and Triggs [21] employed a set of features for building a machine learning classifier to propose a robust visual object detection and recognition histogram-based algorithm. The proposed algorithm is done by dividing the image window into small spatial parts (i.e. Cells) and finding the histograms of edge orientations over all the pixels of the cell. The results of the conducted experiments showed a high accuracy for object classification, but the proposed algorithm introduces a more challenging dataset for machine learning classifier.

Al-Nawashi et al. [7] proposed a novel framework for an intelligent surveillance system based on abnormal human activity detection in academic environments. The conducted experiment showed an excellent surveillance system that can simultaneously perform the tracking, semantic scene learning, and abnormality detection in an academic environment with no human intervention.
Xia Lu et al. [22] presented a high accurate approach that can recognize an object automatically based on an efficient and robust algorithm to identify temporal patterns among actions and utilize the identified patterns to represent activities. Liu Ye et al. [23] proposed a tracker sampling approach for generic human motion tracking using both low and high-dimensional trackers.
Cui Jinshi et al. [24] proposed a new framework for human motion tracking based on a fusion formulation which integrates low and high-dimensional tracking approaches into one framework.

Pablo et al. [25] proposed a new algorithm based on contour detector and image segmentation, the contour has been detected using spectral clustering by combining multiple local cues into a globalization framework and then transforming the output of any contour detector into a hierarchical region tree. Malik Jitendra et al. [26] improved the performance of object detection using a general algorithm for partitioning grayscale images into disjoint regions of coherent brightness and texture. The results of the conducted experiments showed an efficient and high accuracy.

Jacques et al. [27] proposed a new approach for gender recognition based on the upper portion contour. In their proposed approach, the Partial Least Squares (i.e. PLS) method is employed to extract some features from the upper portion (i.e. Head) such as gradient, texture and orientation information, and then used a linear Support Vector Machine (i.e. SVM) for classification. Modi et al. [28] developed a new algorithm for human motion recognition using stationary camera. Artificial Neural Network (i.e. ANN) is used in the proposed algorithm for moving object detection and recognition. The results show that human motion can be correctly classified.

3. Proposed approach

The main contribution of this paper is a new human object classification approach (i.e. human, non-human) that will enhance the object detection services (i.e. classification accuracy). The new approach uses a simple technique and a robust geometrical model that can emulate some of the machine learning approaches. The flow diagram of the proposed approach can be described by Figure 1.
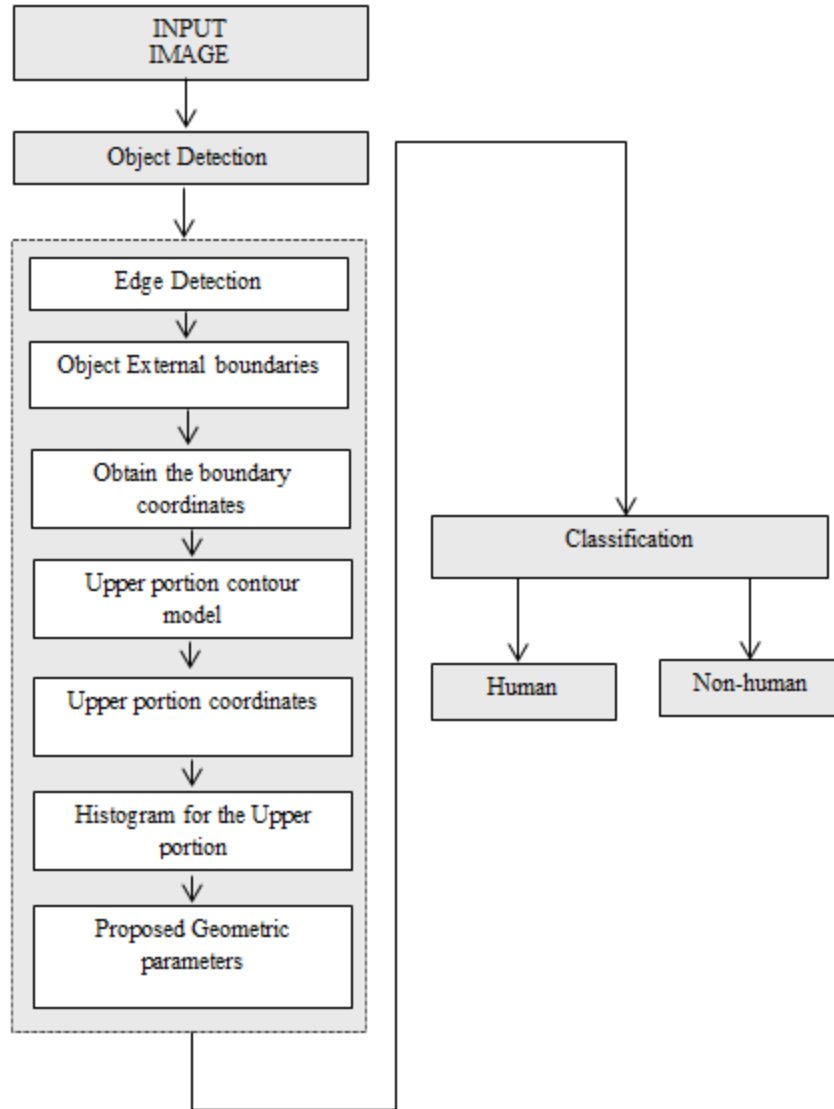
Figure 1: Flow diagram of the proposed approach architecture

To make it clear, the proposed approach is based on a set of parallel and sequential steps, which are partially automated:
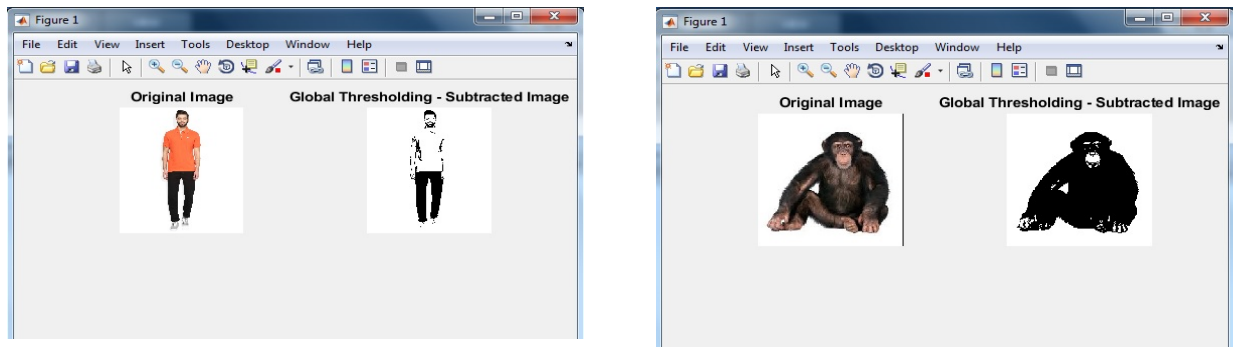
Steps of the proposed approach:	
Step 1:	Background subtraction using histogram-based techniques with global threshold).
Step 2:	Object edge detection using Canny edge detection approach.
Step 3:	Extract the boundary edge using the boundary function.
Step 4:	Obtain [X, Y] coordinates for the boundary contour.
Step 5:	Extract the upper portion of the contour (i.e. Selected object).
Step 6:	Obtain [X, Y] coordinates for the upper portion of the contour.
Step 7:	Represent the obtained X values coordinate in a histogram.
Step 8:	Smooth the histogram using mathematical smooth functions.
Step 9:	Find the upper and lower peaks.
Step 10:	Perform the geometric model parameters.
Step 11:	Classify the object (i.e. Human or non-human).

3.1 Object detection and extraction

The first task of the human detection method is to find the region of interest characterization from the images (i.e. object) [29, 30]. The region of interest (i.e. ROI) refers to the borders of objects [29, 31]. Generally, there are five major categories of object detection methods, i.e. histogram-based techniques, edge-based techniques, region-based techniques, MARKOV random field-based techniques, and hybrid techniques [32]. In this paper, we have performed histogram-based techniques with global threshold to detect and extract the object from images. As shown in Figure 2.

Figure 2: Object detection using histogram-based techniques with global threshold

3.2 Edge detection

Edge detection is an image processing technique for locating the edges within the images (i.e. objects boundaries). The edge within the image can be found by discontinuities in the brightness of the image (i.e. the sharp changes in intensity) [33]. There are several edge detection approaches, i.e. sub-pixel, threshold, canny, and differential [34]. In the proposed approach, we have performed the canny edge detection approach to satisfy the edge detection requirements such as good detection, good localization, and minimal response (i.e. only one response to a certain edge). As shown in Figure 3.
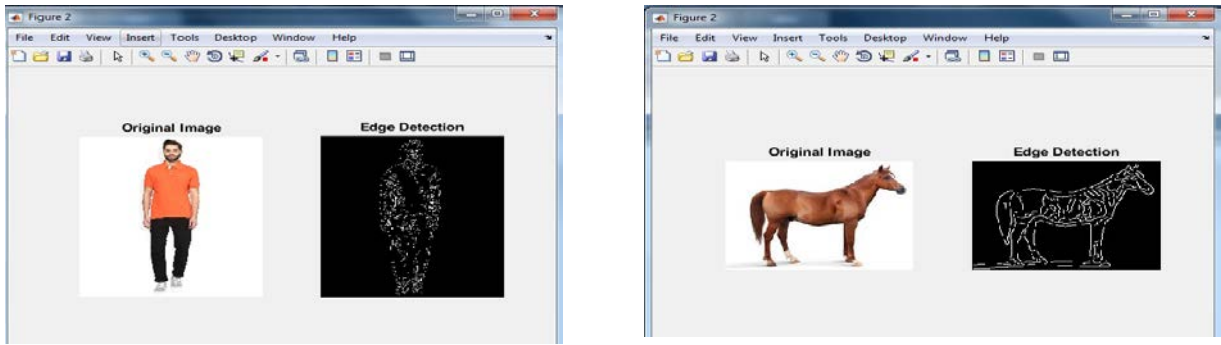
Figure 3: Edge detection using canny edge detection approach

3.3 Boundary edge extraction

After performing a process of edge detection using the canny edge detection approach, the whole object edges (i.e. Internal and external edges) are obtained. To extract the boundary edges only (i.e. External edges), we performed the boundary function $k = boundary(x,y)$ in the proposed approach. The boundary function, returns a vector of point indices representing a single conforming 2-D boundary around the points (x,y), and then obtain the points $(x(k), y(k))$ form the boundary. As shown in Figure 4.
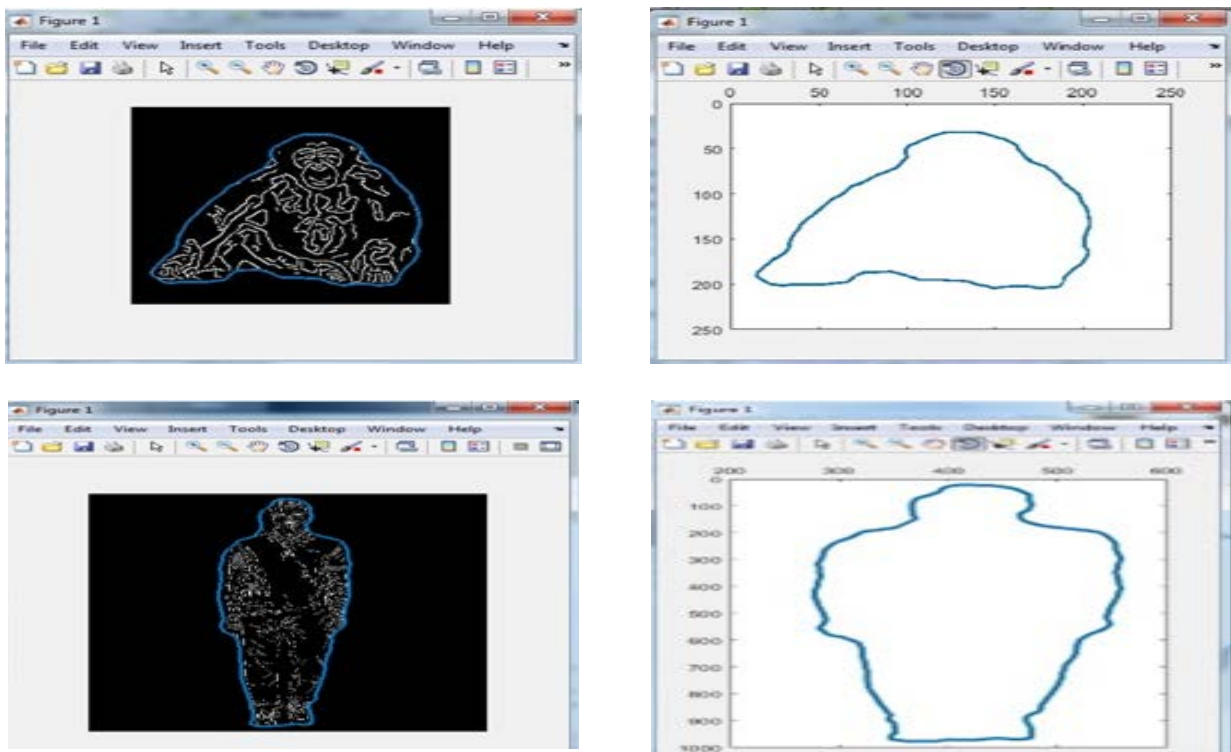

Figure 4: Boundary edge extraction

To make it clear, the boundary function k is a column vector of point indices representing the sequence of points around the boundary, which is a polygon.

3.4   Extract the object upper portion

In this paper, we have focused our study on the upper portion of the object. The upper portion is a very important for human object classification processes because it contains some of human features, i.e. head, neck, shoulders, etc. [27, 35-37]. The upper portion is extracted as shown in Figure 5 based on a set of parallel and sequential steps, which are partially automated:

1. Obtain the row and column projections from the binary image of the detected contour.
2. Smooth the projection curves.
3. Scan the smoothed Row projection to perform the following:
   A. Find the first non-zero pixel to specify the top of the head.
   B. Find the minimum value after the top of the head to specify the neck width.
4. Scan the smoothed column projection to perform the following:
   A. Find the height of the neck which corresponds to the first minimum from the top of the head.
   B. Find the head width which is the maximum value in scanning back from the minimum value and the corresponding height from the head top.
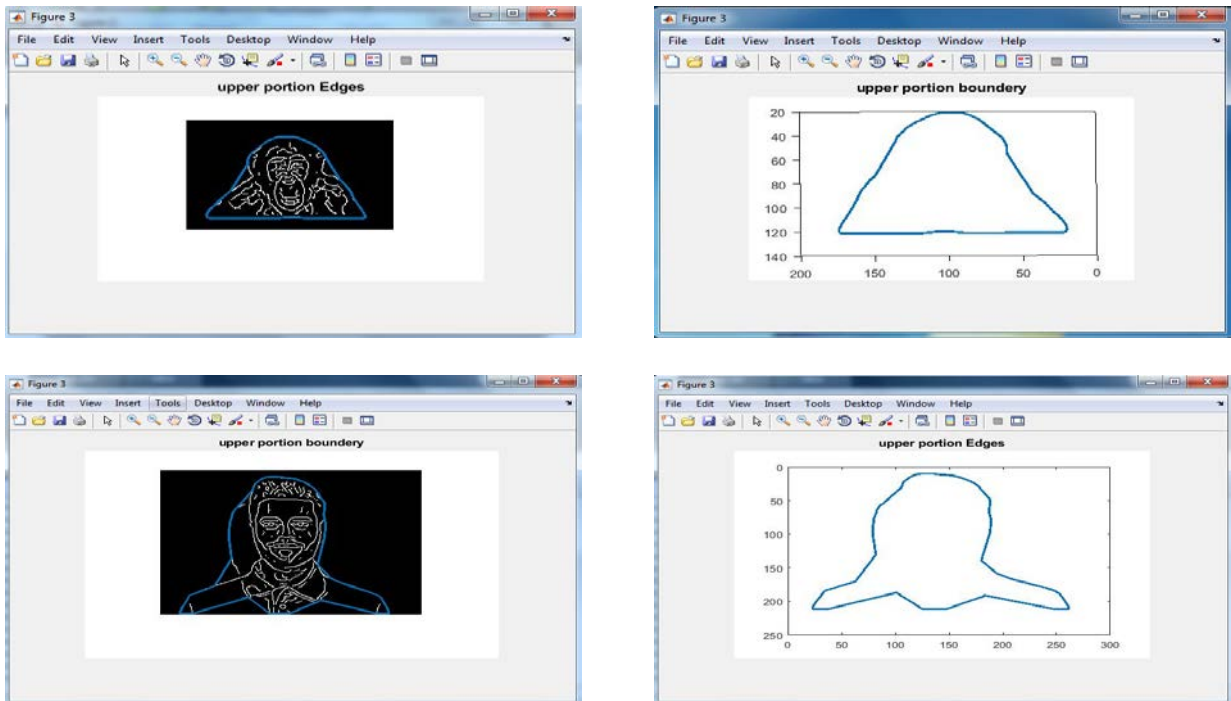5. Determine the shoulder width as 2.5 – 3 times of the head width.

Figure 5: Extract the upper portion

3.5   Geometrical model

After extracting the upper portion of the contour, the *[X, Y]* coordinates for each pixel was obtained sequentially. For accurate object detection and classification, an intensive study and analysis have been done for the values of *[X, Y]* coordinates, and the changing of the coordinate values of the sequence points of the upper portion of different objects (i.e. human, non-human). Finally, we have observed that there is a unique behavior of the flow of *[X]* value (i.e. coordinates of *X*) this behavior is not applicable for non-human object. This unique flow can be simply obtained from head, neck, and shoulder region of the contour. The histogram is used to present the changing of the unique flow behavior of *X* values for different shaped objects (i.e. human, non-human). To avoid the small angles (i.e. odd values) on the histogram, a mathematical smoothing function has been used as [21, 38, 39]. The experiments carried out with our proposed algorithm, the obtained results are shown in Figure 6.
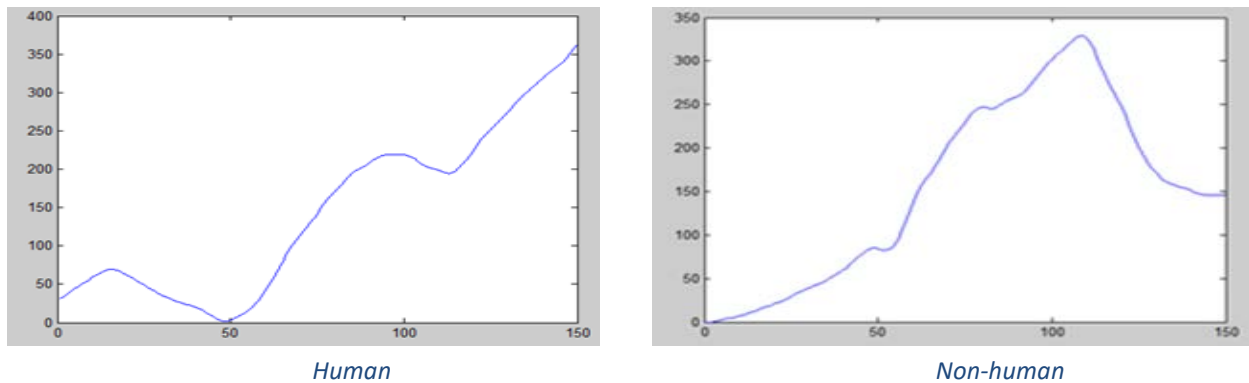

| Human | Non-human |

Figure 6: Histogram for human and non-human objects

As mentioned earlier, an intensive study and analysis have been done for the values of *[X, Y]* coordinates, and we observed that there is a unique behavior of the flow of *[X]*. Therefore, we have designed our object classifier for a human object only based on four parameters. The results of observing and analyzing the histogram of *[X]* values coordinates shows that the numbers of upper peek points are equal to the number of the lower peak points and they are equal to two points, which is not found in the non-human object, this is considered as the first parameter, as shown in Figure 7.
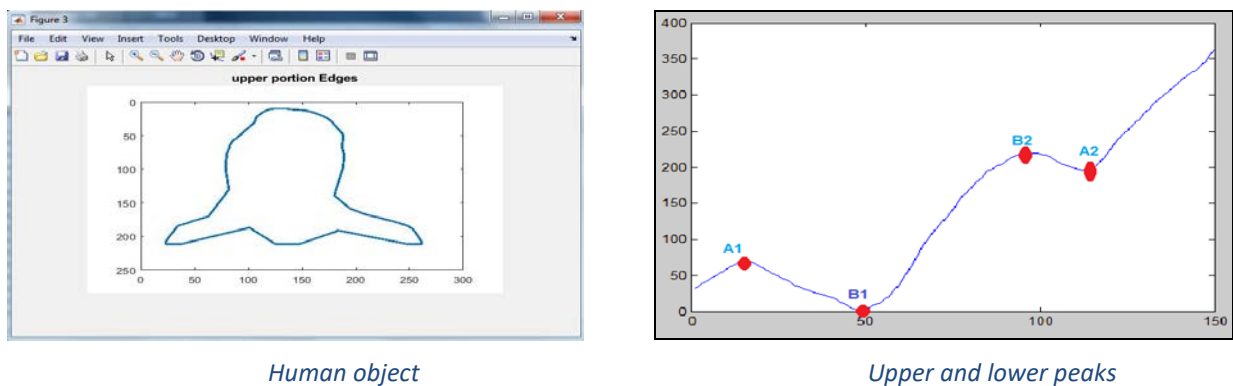

| Human object | Upper and lower peaks |

Figure 7: Upper and lower peaks for human object

The second parameter is built based on the calculations and the measurements which done between the coordinates of the lower and upper peak points as shown in Figure 8, we found that the distance between the first upper peak point and the second lower peak point (D1) is less than the distance between the second upper peak point and the first lower peak point (D2). As well as, the third parameter is based on the calculations and the measurements between the coordinates of the lower and upper peek points as shown in Figure 8, we found that the distance between the two upper peak points (A1, B2) denoted by (D4) is equal to the distance between the two lower peak points (B1, A2) denoted by (i.e.D3). For more accuracy, a small value was added for the difference between the two distances (ts1), which mean that the distance D4 is equal to the distance D3 ± ts1 and it can be written in mathematical words as the absolute value of subtraction between the distance D3 and the distance D4 is equal to ts1.

For the last parameter, we found that the distance between the first lower peak point and the second upper peak point (B1,B2) which denoted by D2 is more than one third of the distance between the start point and the end point of the shoulder (C1, C2) which denoted by (i.e. D5) and less than two third of the distance between the start point and the end point of the shoulder (C1, C2) which denoted by D5 and this is logically true as the fact in anatomical science [35-37].
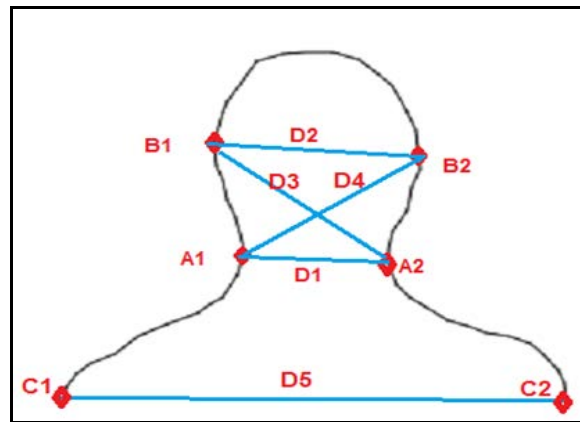

Figure 8: Classifier parameters

To make it clear, the proposed object classification approach is based on a set of parallel and sequential parameters, which are partially automated:

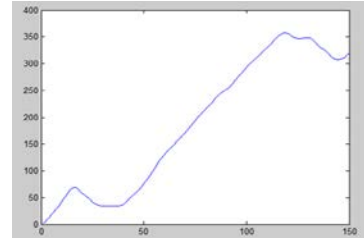| Constrictors of the proposed object classification:	
Parameter 1:	The number of upper peak point = 2, The number of lower peak points = 2
Parameter 2:	The distance D1 < the distance D2
Parameter 3:	The distance D3 = the distance D4 ± (ts1), | D3 − D4| = ts1 where the ts1 is a threshold.
Parameter 4:	The distance D2 > 1/3 of distance D5 And The distance D2 < 2/3 of distance D5, $\frac{1}{3}D5 < D2 < \frac{2}{3}D5$
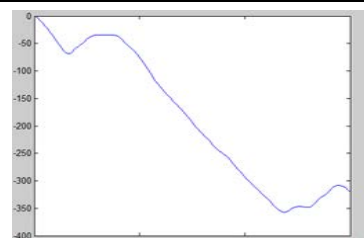
These four parameters simultaneously applied to the human objects only, so that our geometrical parametric model can classify the detected object as human if the result of all of the four parameters

are **True** else, that it will classify as a non-human (see Appendix A, for the detail of the proposed approach).

4. Experimental results

In general, this section concentrates on the performance of the proposed approach, and aims to know the classification accuracy of the proposed object classification approach for human and non-human objects with different pose. The proposed approach achieved a higher accuracy rate in terms of object classification. The performed experiments were implemented through the Matlab application tool on a 1.6 GHz core i5 (IV), 8 GB memory and 750 GB hard disk capacities. INRIA dataset (i.e. 385 digital images) used in our experiment. Table 1, a few of our automatic testing system results are shown based on the proposed approach with a set of color images.

Table 1: Experimental results using the proposed approach

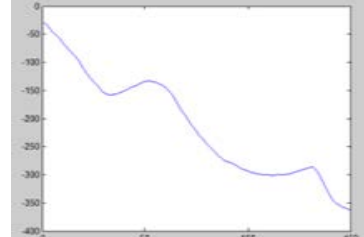| INPUT	Edge detection	Projection of [X] values for (Upper portion contour)	The result of the geometrical parametric model 1 for True; 0 for False				
			Parameter 1	Parameter 2	Parameter 3	Parameter 4	Final Result
			0	0	0	0	Non-Human
			1	1	1	1	Human
			1	1	0	0	Non-human

5. Evaluation

In order to evaluate the classification accuracy of the proposed approach, a comparison between the proposed approach and some of machine learning approaches (i.e. Artificial Neural Networks (ANN), Support Vector Machine (SVM) Model, and Random Forest) are made. The machine learning field was chosen as the domain for evaluation because of the success of data driven artificial intelligence and achieves a high classification accuracy (i.e. human, non-human) [12].

Below, we will first recall the definition of the dataset (i.e. data acquisition) and the main features of the proposed machine learning approach. Then, we will discuss the detailed implementation of the proposed approach and the machine learning approaches. For completeness, a comparison between the proposed geometrical based approach and the machine learning approaches will also be explained.

5.1   Data acquisition

In this paper, a public database for benchmarking human detection from digital images is used. INRIA dataset contains images of humans cropped from a varied set of personal photos [40]. The people are usually standing, but appear in any orientation and against a wide variety of background image including crowds. INRIA contains human images cropped in 64×128 pixels and non-human images of various sizes ($214 \times 320 - 648 \times 486$ pixels) [41]. In this paper, human and non-human images are collected from the INRIA dataset, and then improve the collected images using the image enhancement techniques to have a homogeneous dataset cropped 64×128 pixels.

5.2   Machine learning approach

In this section, we present a machine learning approaches to evaluating a new human object classification approach (i.e. Human, non-human) that will enhance the object detection services (i.e. Classification accuracy). Machine learning-based image classification commonly needs an image dataset (i.e. Human images, and non-human) to split the data into two categories. Then, the optimal 11 features are extracted as shown in Table 2 from each category for training and testing of the classifier using a machine learning toolkit which is called WEKA[42-44]. As shown in Figure 9.
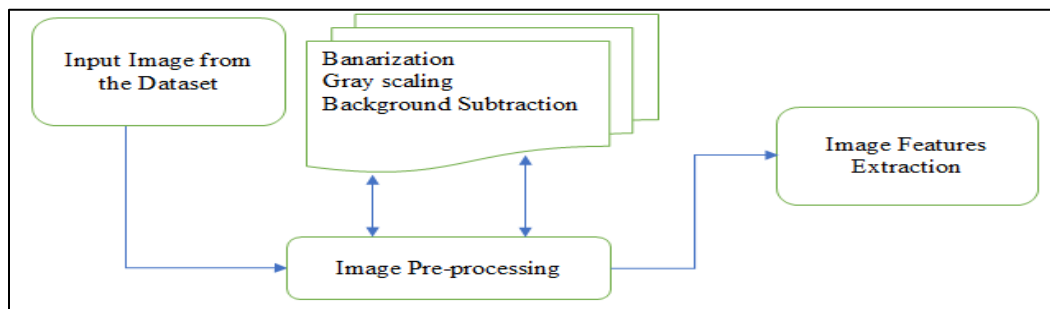

Figure 9: Flow diagram of the proposed approach architecture

Table 2: Features of the machine learning approach

Feature	Formula	

		Description
Circularity	$$S_{CR} = 4\pi \frac{Area}{perimeter^2}$$	The shape of the detected object.
Mean	$$S_M = \bar{b} = \sum_{b=0}^{L-1} b p(b)$$	The average intensity values of the pixels.
Standard Deviation	$$S_D = \sigma_b = \left[\sum_{b=0}^{L-1} ((b - \bar{b}))^2 p(b)\right]^2$$	The deviation or the variance between the pixels in the input image.
Contrast	$$S_C = \sum_i \sum_j (i-j)^2\, p(i,j)$$	The local contrast of the image in terms of gray levels.
Energy	$$S_N = \sum_{b=0}^{L-1} [p(b)]^2$$	The uniformity of the texture.
SKEWNESS	$$S_S = \frac{1}{\sigma_b^3} \sum_{b=0}^{L-1} ((b - \bar{b}))^3 p(b)$$	The asymmetry of the probability distribution of a real-valued random variable (i.e. Positive or negative).
Kurtosis	$$S_K = \frac{1}{\sigma_b^4} \sum_{b=0}^{L-1} ((b - \bar{b}))^4 p(b) - 3$$	The peakedness of the probability distribution of a real-valued random variable.
Entropy	$$S_E = - \sum_{b=0}^{L-1} p(b) \log_2\{p(b)\}$$	The randomness of a gray level distribution.
Homogeneity	$$S_H = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \left(\frac{P_{ij}}{(1 + |i-j|)}\right)$$	The closeness element distribution in GLCM to the diagonal GLCM.
Correlation	$$S_o = \frac{\sum_i \sum_j (ij)\, p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$	The gray level linear dependence between the relative pixels.
Diameter	$$D = 2r$$	Diameter of the edge pixels.

The digital images used in this paper are drawn from the INRIA. A software-based simulation using Matlab was performed to extract the features of the human (i.e. HU) and non-human (i.e. NH). For an example of the extracted features values based on our software simulation. See Figure 10.

Class	Contrast	SKWENESS	Kurtosis	Entropy	Mean	STD	Circulation	Energy	Correlation	Homogeneity	Diameter
NH	0.4096313	1.4647412	3.68919358	2.4586934	0.713589	0.10497	0.170500657	0.45427	0.934805284	0.942542972	77.25822421
NH	1.543092	0.6327101	1.73922325	3.2139312	0.601251	0.26009	0.519347116	0.212383	0.915807358	0.88422887	26.96555684
NH	0.3634939	1.52722	3.88071689	2.3389425	0.72396	0.099062	0.268591532	0.634812	0.890879927	0.940385611	25.34866285
NH	0.8382716	0.255393	1.70289185	4.7602913	0.522904	0.261673	0.974510188	0.185554	0.515996847	0.770823045	7.664083887
NH	0.0697674	0.1292439	1.70976018	4.5640954	0.506209	0.280004	0.815340875	0.501085	0.839330855	0.965116279	8.475885149
NH	1.0309966	0.0943804	1.72407625	5.6132643	0.507041	0.282991	0.764440758	0.085073	0.906028085	0.823847803	58.34654096
NH	0.1013036	0.2145565	1.68338518	4.7371289	0.517438	0.268583	0.763436059	0.451547	0.921573758	0.956244784	94.06135716
NH	0.1221504	0.1518394	1.69952556	5.0986174	0.509636	0.278057	0.759209005	0.424858	0.811680164	0.948126907	79.15356731
HU	0.0759527	0.0015947	1.79877976	5.0321695	0.50063	0.293954	0.788735455	0.373156	0.96935851	0.976240812	85.86579201
HU	0.1463341	0.0060509	1.79229953	5.8292385	0.500485	0.293107	0.687587736	0.236915	0.94951322	0.948412698	75.88524314
HU	0.1518828	0.0338739	1.77452882	5.5845508	0.502124	0.290003	0.60910958	0.27102	0.957502599	0.94182095	83.77675236
HU	0.1796053	0.0135222	1.77979291	5.6617798	0.500324	0.292691	0.614101161	0.246377	0.953511134	0.93569349	81.31811464
HU	0.1754386	-0.001136	1.79884816	5.6168371	0.499668	0.293077	0.711405994	0.23134	0.910003404	0.929006544	35.64520799
HU	0.1124818	0.1390353	1.70625791	5.3618748	0.511466	0.27722	0.44672996	0.256375	0.933455605	0.95308311	94.91928442
HU	0.1018853	0.1390269	1.70471295	5.3245961	0.511393	0.276944	0.390734792	0.276487	0.937241685	0.959387494	106.4762046
HU	0.0463755	0.1656096	1.7094672	4.8644615	0.514252	0.274189	0.682144486	0.475984	0.920752753	0.978265257	43.26714196
HU	0.1343467	0.0092822	1.79246547	5.8303768	0.500543	0.292455	0.69110382	0.264841	0.95075864	0.946510865	55.02971645

Figure 10: Extracted feature's value from human and non-human images

The excel-sheet of 358 images in total, it was converted to comma delimited file (i.e. CSV) to be identified by the machine learning toolkit (i.e. WEKA). In this paper, 11 features (i.e. optimal features) were chosen to be in the classification process. The distributions of instances over the selected features are shown in Figure 11.
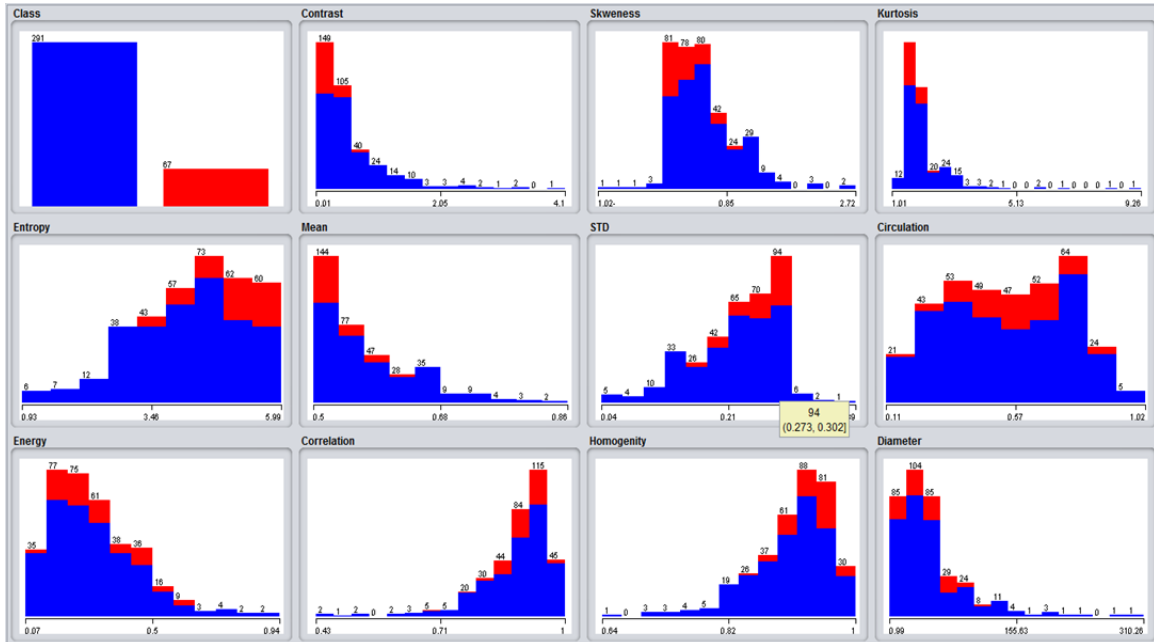

Figure 11: Extracted feature's value from human and non-human images

6. Performance analysis

In this section, we provide some experimental results for analysis the performance (i.e. Classification accuracy) of the proposed object classification approach and some of the machine learning approaches (i.e. Artificial Neural Networks (ANN), Support Vector Machine (SVM) Model, and Random Forest). In this paper, the performance analysis for SVM, ANN, Random Forest, and the proposed object classification approach were conducted in two tests: the testing confusion matrix, and accuracy matrix for all classes.

The confusion matrix for SVM, ANN, and Random Forest for the test set is reported in Table 3, 4 and 5 respectively. The accuracy the confusion matrix is given by:

$$Accuracy = \frac{\sum \text{Diagonal Sample of confusion matrix}}{\text{Total Sample}}$$

Table 3: SVM confusion matrix

a	b	← Classified as		
287	4	a	=	NH
66	**1**	b	=	HU

Table 4: ANN confusion matrix

a	b	← Classified as		
275	16	a	=	NH
21	**46**	b	=	Hu

Table 5: SVM confusion matrix

a	b	← Classified as		
282	9	a	=	NH
27	**40**	b	=	Hu

In the matrix accuracy for SVM, ANN, and Random Forest approach for all classes, TP rate, FP rate, Precision, Recall, F-Measure, MCC, ROC Area, and PRC were calculated and listed in Table 6 , 7, and 8 respectively, and the corresponding classes distribution are shown in Figure 12, 13 and 14 respectively (see Appendix B, for more details).

Table 6: SVM accuracy matrix for all classes

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.986	0.986	0.813	0.986	0.891	0.004	0.501	0.813	NH
0.015	0.014	0.200	0.015	0.028	0.004	0.501	0.187	HU
0.804	**0.803**	**0.698**	**0.804**	**0.730**	**0.004**	**0.501**	**0.696**	←Weighted Average.

Table 7: ANN accuracy matrix for all classes

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.945	0.313	0.929	0.945	0.937	0.651	0.896	0.969	NH
0.687	0.055	0.742	0.687	0.713	0.651	0.896	0.738	HU
0.897	**0.265**	**0.894**	**0.897**	**0.895**	**0.651**	**0.896**	**0.926**	←Weighted Average

Table 8: Random Forest accuracy matrix for all classes

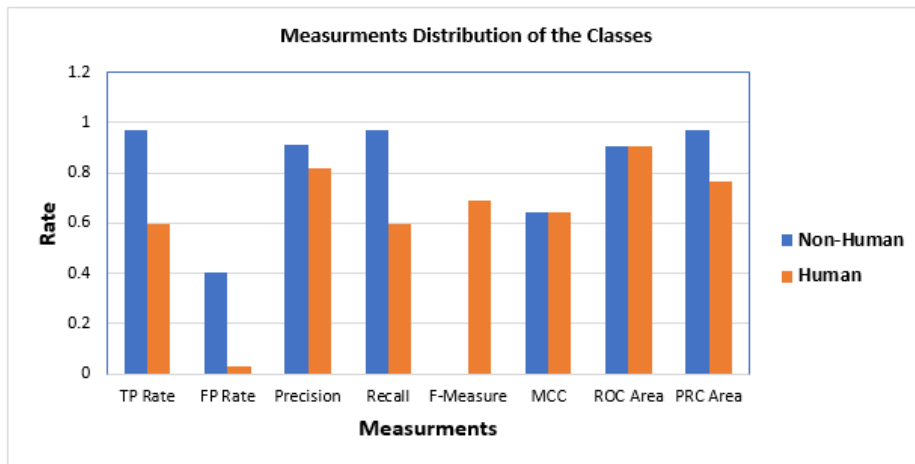| TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.969	0.403	0.913	0.969	0. 940	0.642	0.903	0.968	NH
0.597	0.031	0.816	0.597	0.690	0.642	0.903	0.763	HU
0.899	**0.333**	**0.895**	**0.899**	**0.893**	**0.642**	**0.903**	**0.930**	←Weighted Average
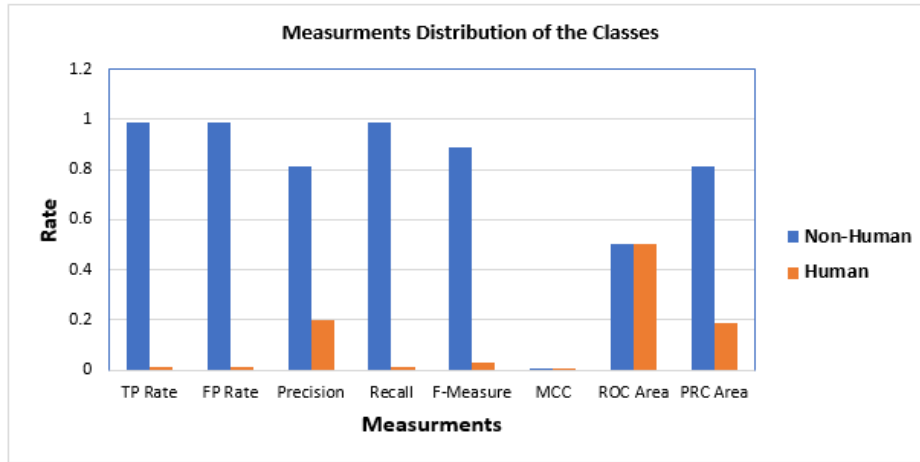
Figure 12: Classes distribution for SVM

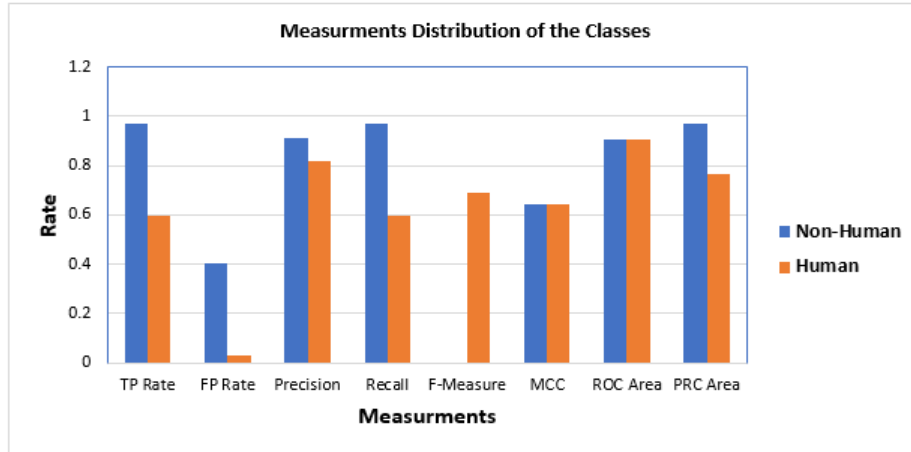*Figure 13: Classes distribution for ANN*

The machine learning approaches (i.e. SVM, ANN, and Random Forest) have introduced and implemented to identify and evaluate the proposed human object classification approach. As the goal of this research, the classification accuracy is a concern to us, and should be kept to a maximum value. Therefore, a comparison between the classification accuracy in our proposed approach and the machine learning approaches are made. As shown in Table 9 and the corresponding flow chart is shown in Figure 15.

Table 9: Classification accuracy

|	**SVM**	**ANN**	**Random Forest**	**Proposed approach**
Correctly Classified Instances	289 in percentage of 80.7263 %	321 in percentage of 89.6648 %	322 in percentage of 89.9441 %	332 in percentage of 92.7374%
Incorrectly Classified Instances	69 in percentage of 19.2737 %	37 in percentage of 10.3352 %	36 in percentage of 10.0559 %	26 in percentage of 7.2625%
Kappa statistic	0.0072	0.6503	0.6314	0.6453
Mean absolute error	0.1927	0.1233	0.1734	0.1583
Root mean squared error	0.439	0.2926	0.2841	0.2318
Relative absolute error	63.0955 %	40.3738 %	56.7494 %	45.8780
Root relative squared error	112.5363 %	75.0073 %	72.8178 %	68.2135%
Total Number of Instances	358	358	358	358
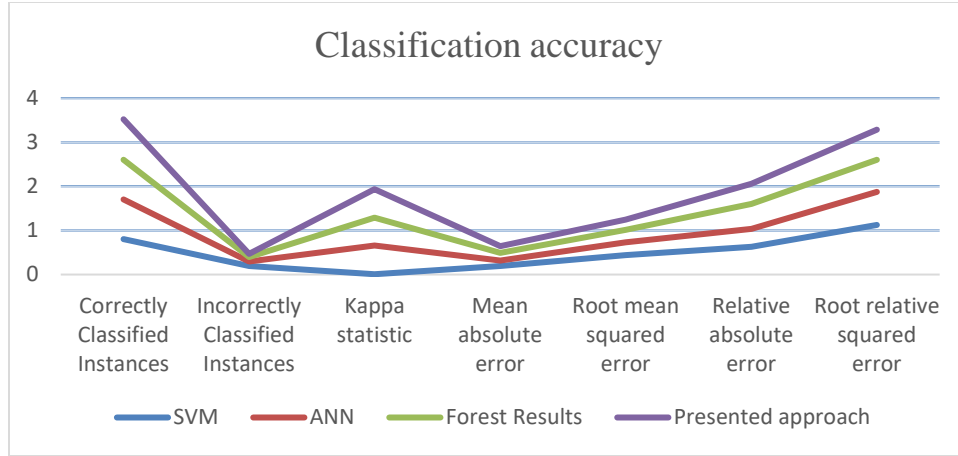
Figure 15: Flow chart of classification accuracy

7. Comparisons with state of the art approaches

In this section, we compare the performance of the proposed object classification approach to some of state-of-art approaches in literature. In our comparisons, five state-of-art approaches are chosen due to their high citation rate. The comparisons were implemented and evaluated in the same environment and conditions. From the experiment results, we have observed the advantages and disadvantages for each approach as shown in Table 10.

Table 10: Comparison between the proposed approach and some of the state-of-the art approaches in literature

Approach	Description	Advantages	Disadvantages
[45]	Human detection by quadratic classification on subspace of extended histogram of gradients.	• Applicable to various environmental conditions and detect the human objects in various scales.	• Requires a pre-defined template or classifier for the human area that must be obtained through training. • Requires significant processing time to detect the human object. • Performance degradation can occur if the intensity of the object is similar to the background.
[46]	Human Detection using learned part alphabet and pose dictionary.		
[47]	Occlusion handling via random subspace classifiers for human detection.		
[39]	Histograms of oriented gradients for human detection.		
[48]	Training deformable object models for human detection based on alignment and clustering.		
Proposed approach	Geometrical-based approach for robust human image detection	• Applicable to various environmental conditions and detect the human objects in various scales. • Simple technique and a robust geometrical model that can emulate some of the machine learning approaches. • Does not require a training procedure to obtain the classifier of human detection.	• Applicable to detect human in still images.

There are several terms that are commonly used to measure the performance (i.e. accuracy) of the human detection approaches such as positive predictive value (PPV), and the computational cost (Time) [39]. To calculate the PPV, we have:

$$PPV = \frac{\#TP}{\#TP + \#FP}$$

Where, *#TP* is the number of true positive and *#FP* is the number of false positive. The PPV value and the computational cost for each approach are calculated and listed in Table 11.

Table 11: Positive predictive value and computational cost for each approach

Approach	Number of True Positive (#TP)	Number of False Positive (#FP)	Positive Predictive Value (PPV)	Computational Performance (Time-*ms*)
[45]	322	12	0.96407	17543
[46]	329	5	0.98502	14298
[47]	316	18	0.94610	18007
[39]	330	4	0.98802	13875
[48]	325	9	0.97305	14176
Proposed approach	332	2	0.99401	13475

In general, a higher value of PPV means a higher accuracy of human detection. The obtained results demonstrate that the proposed object detection and classification approach is efficient and achieved a comparable accuracy and computational cost to other state-of-art approaches.

8. Conclusion

Human object detection and classification in the images is a challenging work. In this paper, we propose a new object classification approach to improve the object classification accuracy. The new approach will use a simple and robust geometrical model to classify the detected object as human or non-human in the images. In the proposed approach, the object is detected and then the detected object under different conditions can be accurately classified (i.e. human, non-human) by combining the features that are extracted from the upper portion of the contour and the geometrical model parameters. The performance evaluation (i.e. classification accuracy) of the proposed approach were conducted in two phases: a comparison between the proposed approach and some of machine learning approaches (i.e. Artificial Neural Networks (ANN), Support Vector Machine (SVM) Model, and Random Forest) phase and a comparison between the proposed approach and five state-of-art approaches in literature phase. The experimental results show the classification accuracy for the machine learning approaches and the proposed approach as the following: SVM is equal to 80.7263 %, ANN is equal to 89.6648 %, random forest is equal to 89.9441 %, and the proposed approach is equal to 92.7374%. Furthermore, the obtained results demonstrate that the proposed object detection and classification approach is efficient and achieved a comparable accuracy and computational cost to other state-of-art approaches. This indicates that the proposed approach is efficient and achieves higher classification accuracy than machine learning approaches and other state-of-art approaches.

Appendix A. Proposed approach

The proposed approach is based on a set of parallel and sequential processes, which are summarized as the following:

1. Background subtraction using histogram-based techniques with global threshold).
2. Object edge detection using CANNY edge detection approach.
3. Extract the boundary edge using the boundary function.

Extract the upper portion of the contour.

4. Extract the boundary *{xi, yi}* coordinate matrix points for the contour which obtained from background subtraction/boundary edge detection.
5. Obtain *Y min*, *Y max*, X min , X max values from the boundaries points obtained in step4
6. Obtain the row and column projections from the binary image of the detected contour.
7. Smooth the projection curves using smooth function *S*.
8. Scan the smoothed Row projection to perform the following:
   8.1. Find the first non-zero pixel to specify the top of the head (*th*).
   8.2. Find the minimum value after the top of the head to specify the neck width (*nw*).
9. Scan the smoothed column projection to perform the following:
   9.1. Find the height of the neck which corresponds to the first minimum from the top of the head (*hn*).
   9.2. Find the head width which is the maximum value in scanning back from the minimum value and the corresponding height from the head top (*hw*).
10. Determine the shoulder width as 2.5 – 3 times of the head width (*hw*).
11. Extract the upper portion of the contour (i.e. selected object).

Establishing the geometric model

12. Obtain *{xi, yi}* coordinates for the upper portion of the contour.
13. Represent the obtained X values coordinate in a histogram.
14. Smooth the histogram using mathematical smooth functions *S*.
15. Find the upper peak points (*up*) and lower peaks points (*lp*) in the histogram obtained from step 14.
    Parameter #1(*P1*): See Figure 8.
16. Find the number of upper peak point and the number of lower peak points.
    16.1. Take a decision
        IF (The number of the upper peak point (*up*) $==$ 2)
        AND
        (The number of the lower peak point (*lp*) $==$ 2) Then, P1$==$1
        Else
        P1$==$ 0.
    Parameter #2(*P2*): See Figure 8.
17. Find the distance (*D1*) between the first upper peak point (*up1*) and the second lower peak point (*lp2*).
18. Find the distance (*D2*) between the second upper peak point (*up2*) and the first lower peak point (*lp1*)
19. Take a decision
    IF (*D1 < D2*) Then, P2$==$1
     Else
    P2$==$ 0.

Parameter #3(P3): See Figure 8.
20. Find the distance ($D4$) between the two upper peak points (*up1, up2*).
21. Find the distance ($D3$) between the two lower peak points (*lp1, lp2*).
22. Take a decision
    IF ($| D3 - D4| = ts1$ where the ts1 is a threshold. Then, P3 $= = 1$
    Else
    P3 $= = 0$.
    Parameter #4(P4): See Figure 8.
23. Find the distance ($D5$) between the start point and the end point of the shoulder (*C1, C2*).
24. Get the distance ($D2$) which obtained in step 18.
25. Take a decision
    IF ($\frac{1}{3}D5 < D2 < \frac{2}{3}D5$) Then, P4$= =1$
    Else
    P4 $= = 0$.
Classification decision
26. **IF (P1=1 And P2=1 And P3=1**
    AND
    P4 $= =1$) Then,
    The detected object is Human
    Else
    The detected object is Non-human
    END

Appendix B. WEKA data analysis

Table 12: Result of WEKA data analysis

WEKA results output	Description / Formula
TP	Rate of true positives (actually positive).
FP	Rate of false positives (actually negative).
Precision	$$\frac{TP}{Predicted\ positive}$$
Recall	$$\frac{TP}{TP + FN}$$
F-Measure	$$2 * \frac{Precision*Recall}{Precision+Recall}$$
MCC	Matthews Correlation Coefficient is used in machine learning as a measure of the quality of binary (two-class) classifications.
ROC Area	Receiver Operating Characteristic: A plot of true positive fraction (= sensitivity) vs. false positive fraction (= 1 – specificity) for all potential cut-offs for a test.
PRC Area	Precision-recall curve. A plot of precision (= PPV) vs. recall (= sensitivity) for all potential cut-offs for a test.

References

[1] A. Broggi, M. Bertozzi, A. Fascioli, and M. Sechi, "Shape-based pedestrian detection," in *Intelligent Vehicles Symposium, 2000. IV 2000. Proceedings of the IEEE*, 2000, pp. 215-220.

[2] M. Heikkila and M. Pietikainen, "A texture-based method for modeling the background and detecting moving objects," *IEEE transactions on pattern analysis and machine intelligence,* vol. 28, pp. 657-662, 2006.

[3] S. Mukherjee and K. Das, "A Novel Equation based Classifier for Detecting Human in Images," *arXiv preprint arXiv:1307.5591,* 2013.

[4] X. Xie and C. Livermore, "A pivot-hinged, multilayer SU-8 micro motion amplifier assembled by a self-aligned approach," in *Micro Electro Mechanical Systems (MEMS), 2016 IEEE 29th International Conference on*, 2016, pp. 75-78.

[5] E. Hjelmås and B. K. Low, "Face detection: A survey," *Computer vision and image understanding,* vol. 83, pp. 236-274, 2001.

[6] K.-W. Wong, K.-M. Lam, and W.-C. Siu, "An efficient algorithm for human face detection and facial feature extraction under different conditions," *Pattern Recognition,* vol. 34, pp. 1993-2004, 2001.

[7] M. Al-Nawashi, O. M. Al-Hazaimeh, and M. Saraee, "A novel framework for intelligent surveillance system based on abnormal human activity detection in academic environments," *Neural Computing and Applications,* pp. 1-8, 2016.

[8] M. A.-H. Obaida, "Combining audio samples and image frames for enhancing video security," *Indian Journal of Science and Technology,* vol. 8, p. 940, 2015.

[9] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Face recognition from a single image per person: A survey," *Pattern recognition,* vol. 39, pp. 1725-1745, 2006.

[10] D. Murray and A. Basu, "Motion tracking with an active camera," *IEEE transactions on pattern analysis and machine intelligence,* vol. 16, pp. 449-459, 1994.

[11] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *International journal of Remote sensing,* vol. 28, pp. 823-870, 2007.

[12] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, *Machine learning: An artificial intelligence approach*: Springer Science & Business Media, 2013.

[13] Z.-Q. Zeng, H.-B. Yu, H.-R. Xu, Y.-Q. Xie, and J. Gao, "Fast training support vector machines using parallel sequential minimal optimization," in *Intelligent System and Knowledge Engineering, 2008. ISKE 2008. 3rd International Conference on*, 2008, pp. 997-1001.

[14] P. Guan, A. Weiss, A. O. Balan, and M. J. Black, "Estimating human shape and pose from a single image," in *Computer Vision, 2009 IEEE 12th International Conference on*, 2009, pp. 1381-1388.

[15] J. Gall, A. Yao, N. Razavi, L. Van Gool, and V. Lempitsky, "Hough forests for object detection, tracking, and action recognition," *IEEE transactions on pattern analysis and machine intelligence,* vol. 33, pp. 2188-2202, 2011.

[16] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Computer vision and image understanding,* vol. 104, pp. 90-126, 2006.

[17] Y. Zhong, A. K. Jain, and M.-P. Dubuisson-Jolly, "Object tracking using deformable templates," *IEEE transactions on pattern analysis and machine intelligence,* vol. 22, pp. 544-549, 2000.

[18] S. M. Desa and Q. A. Salih, "Image subtraction for real time moving object extraction," in *Computer Graphics, Imaging and Visualization, 2004. CGIV 2004. Proceedings. International Conference on*, 2004, pp. 41-45.

[19] B. Sugandi, H. Kim, J. K. Tan, and S. Ishikawa, "Tracking of moving objects by using a low resolution image," in *Innovative Computing, Information and Control, 2007. ICICIC'07. Second International Conference on*, 2007, pp. 408-408.

[20] Y. Zheng, Y. Meng, and Z. Zhu, "Object detection and tracking using Bayes-constrained particle swarm optimization," *Comput. Vis. Res. Prog,* vol. 245, 2008.

[21] N. Dalal, B. Triggs, and C. Schmid, "Human detection using oriented histograms of flow and appearance," in *European conference on computer vision*, 2006, pp. 428-441.

[22] L. Xia, C.-C. Chen, and J. K. Aggarwal, "Human detection using depth information by kinect," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, 2011, pp. 15-22.

[23] Y. Liu, J. Cui, H. Zhao, and H. Zha, "Fusion of low-and high-dimensional approaches by trackers sampling for generic human motion tracking," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, 2012, pp. 898-901.

[24] J. Cui, Y. Liu, Y. Xu, H. Zhao, and H. Zha, "Tracking generic human motion via fusion of low-and high-dimensional approaches," *IEEE transactions on systems, man, and cybernetics: systems,* vol. 43, pp. 996-1002, 2013.

[25] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE transactions on pattern analysis and machine intelligence,* vol. 33, pp. 898-916, 2011.

[26] J. Malik, S. Belongie, T. Leung, and J. Shi, "Contour and texture analysis for image segmentation," *International journal of computer vision,* vol. 43, pp. 7-27, 2001.

[27] J. C. S. Jacques and S. R. Musse, "Improved head-shoulder human contour estimation through clusters of learned shape models," in *Graphics, Patterns and Images (SIBGRAPI), 2015 28th SIBGRAPI Conference on*, 2015, pp. 329-336.

[28] R. V. Modi and T. B. Mehta, "Neural Network based Approach for Recognition Human Motion using Stationary Camera," *International Journal of Computer Applications (0975–8887) Volume,* 2011.

[29] O. M. Al-hazaimeh, "A novel encryption scheme for digital image-based on one dimensional logistic map," *Computer and Information Science,* vol. 7, p. 65, 2014.

[30] O. M. A. Al-Hazaimeh, "Hiding data in images using new random technique," *IJCSI International Journal of Computer Science Issues,* vol. 9, pp. 49-53, 2012.

[31] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk, "Salient region detection and segmentation," *Computer Vision Systems,* pp. 66-75, 2008.

[32] M. Piccardi, "Background subtraction techniques: a review," in *Systems, man and cybernetics, 2004 IEEE international conference on*, 2004, pp. 3099-3104.

[33] S. Lakshmi and D. V. Sankaranarayanan, "A study of edge detection techniques for segmentation computing approaches," *IJCA Special Issue on "Computer Aided Soft Computing Techniques for Imaging and Biomedical Applications" CASCT,* pp. 35-40, 2010.

[34] P. Bao, L. Zhang, and X. Wu, "Canny edge detection enhancement by scale multiplication," *IEEE transactions on pattern analysis and machine intelligence,* vol. 27, pp. 1485-1490, 2005.

[35] H. Jain, A. Subramanian, S. Das, and A. Mittal, "Real-time upper-body human pose estimation using a depth camera," *Computer Vision/Computer Graphics Collaboration Techniques,* pp. 227-238, 2011.

[36] M. Kampmann, "Segmentation of a head into face, ears, neck and hair for knowledge-based analysis-synthesis coding of videophone sequences," in *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*, 1998, pp. 876-880.

[37] H. Li and K. N. Ngan, "Saliency model-based face segmentation and tracking in head-and-shoulder video sequences," *Journal of Visual Communication and Image Representation,* vol. 19, pp. 320-333, 2008.

[38] D.-M. Tsai, "A fast thresholding selection procedure for multimodal and unimodal histograms," *Pattern Recognition Letters,* vol. 16, pp. 653-666, 1995.

[39] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, pp. 886-893.

[40] (15 of April). *INRIA Person Dataset*. Available: http://pascal.inrialpes.fr/data/human/

[41]   T. Watanabe, S. Ito, and K. Yokoi, "Co-occurrence histograms of oriented gradients for pedestrian detection," in *Pacific-Rim Symposium on Image and Video Technology*, 2009, pp. 37-47.

[42]   M. Al-Abri and N. Hilal, "Artificial neural network simulation of combined humic substance coagulation and membrane filtration," *Chemical Engineering Journal,* vol. 141, pp. 27-34, 2008.

[43]   Y. Benayed, D. Fohr, J. P. Haton, and G. Chollet, "Confidence measures for keyword spotting using support vector machines," in *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, 2003, pp. I-I.

[44]   J. Huang, J. Lu, and C. X. Ling, "Comparing naive Bayes, decision trees, and SVM with AUC and accuracy," in *Data Mining, 2003. ICDM 2003. Third IEEE International Conference on*, 2003, pp. 553-556.

[45]   A. Satpathy, X. Jiang, and H.-L. Eng, "Human detection by quadratic classification on subspace of extended histogram of gradients," *IEEE Transactions on Image Processing,* vol. 23, pp. 287-297, 2014.

[46]   C. Yao, X. Bai, W. Liu, and L. J. Latecki, "Human detection using learned part alphabet and pose dictionary," in *European Conference on Computer Vision*, 2014, pp. 251-266.

[47]   J. Marín, D. Vázquez, A. M. López, J. Amores, and L. I. Kuncheva, "Occlusion handling via random subspace classifiers for human detection," *IEEE transactions on cybernetics,* vol. 44, pp. 342-354, 2014.

[48]   B. Drayer and T. Brox, "Training deformable object models for human detection based on alignment and clustering," in *European Conference on Computer Vision*, 2014, pp. 406-420.