Human Detection using Gradient Maps and Golden Ratio

Feng Su and Gu Fang

School of Computing, Engineering and Mathematics
University of Western Sydney
Locked Bag 1797, Penrith NSW 2751, Australia
t.su@uws.edu.au, g.fang@uws.edu.au

Abstract -

Human detection is an important topic that can be used for many applications, it is mainly found in areas that required surveillance such as airports, casinos, factories, construction and mining sites. In this paper, a novel human detection method is introduced to extract the human figure from an input image without prior information or training. This method firstly uses a head and shoulder detection scheme based on curve detection with scaled gradient magnitude and orientation maps. It is then followed by a human body estimation scheme based on gap detection and golden ratio. Finally, the human figure is extracted through thresholding local gradient magnitude regions and horizontal filling. Tests on various images have shown that this method is capable of detecting and extracting human body figures robustly from different images.

Keywords -

Human Detection; Head and Shoulder Detection; Human Body Estimation; Human Figure Extraction

1 Introduction

Human detection can be defined as a process of detecting the presence of human features from images or videos. Generally, in order to detect and track humans robustly, an algorithm needs to be able to extract common features among different people, so they can be separated from the background. As different people tend to have different features which are usually caused by their variable appearances and postures, this task can be quite challenging.

Human detection is generally used in areas that required surveillance [1, 2], these areas include airports, casinos, factories, construction and mining sites. The basic functions of human detection are used to detect and track the human, while the advanced functions which contain some form of post-processing are used to achieve additional goals. These goals may include people counting [3], face recognition [4] and behaviour recognition [5]. The advanced functions are generally used for high security environments such as airports and casinos, these systems tend to be expensive which equipped with high speed communication lines and powerful processors [6].

Human detection methods can be divided into video based [7] or static image based [8]. The major difference between these two types of methods is that video based approaches could also utilise the motion features such as background subtraction [9] and optical flow [10], this is impracticable with a single image.

The main contribution of this paper is the introduction of a novel human detection method which includes a new method of detecting the head and shoulder, a new method of estimating the size and position of the human body and a new method of extracting an accurate figure of the human body. Also, this human detection method does not require any prior training.

This paper is organised into 5 sections. Introduction and related work are given in Sections 1 and 2, the proposed human detection method is introduced in Section 3, experimental results are provided in Section 4 and conclusions are given in Section 5.

2 Related Work

The most influential human detection method is the Histogram of Oriented Gradients method (HOG) [11]. HOG is a feature descriptor that heavily relies on the extraction of gradient orientations and the use of linear Support Vector Machine classifier [12]. HOG is renowned for its human detection ability by utilising large amount of training images. However, this means the algorithm itself is also heavily limited by the quality and quantity of these training images. This is because, in order to learn a classifier successfully, HOG requires continuously recurring shape events in the given blocks of the training images [13]. HOG generally suffers from speed issues and the training itself is a time consuming task. Many existing methods tend to build their algorithm based on HOG or utilise it as an additional feature for their algorithms [14–20].

One method argues that humans in standing positions tend to have distinguishing colour characteristics [14],

therefore addition colour information is employed with HOG descriptors. This includes colour frequency and co-occurrence features in each channel of Hue-Saturation-Value (HSV) images. It also employed the partial least square regression analysis [15] onto the descriptors. These additional elements improved the accuracy of the HOG, however at the same time the computational cost is greatly increased.

Another method utilises the omega-shape features of humans [16]. This method argues that HOG feature based classifier are generally accurate but very slow to work with, while Haar feature based classifiers [17] are generally fast but suffer from poor accuracies. Thus, by combining these two classifiers, robustness could be achieved. This is done by first employ the Haar feature classifier to exclude obvious negative image patches and then employ HOG for the remaining image patches. This method greatly improved the computation speed with a small drop in accuracy. However, the algorithm only performs detection and tracking of the head, rather than the whole body.

Scale-Invariant Feature Transform (SIFT) can also be used to detect human [18–20]. SIFT [21] is an algorithm to detect and describe local features by extracting distinctive invariant features. SIFT descriptors are similar to HOG in the sense that both descriptors utilize gradient features. The difference is that HOG is computed in dense grids while SIFT is computed in sparse grids with orientation alignment. By combining HOG and SIFT descriptors, higher detection rate can be achieved. However, the efficiency of these methods tends to suffer with online processing, since both descriptors have reasonably high computational costs.

Therefore, this research is aimed to develop a human detection method that capable of extracting the human figures without the need of pre-training.

3 Proposed Method

The proposed human detection method consists of four stages: scaled gradient mapping, head and shoulder detection, human body estimation and human figure extraction. A flowchart of the method is shown in Figure 1.

3.1 Scaled Gradient Mapping

The first step in the scaled gradient mapping stage is to convert the input image to greyscale, then the gradient is computed using two simple kernels, as given in Equations (1) and (2).

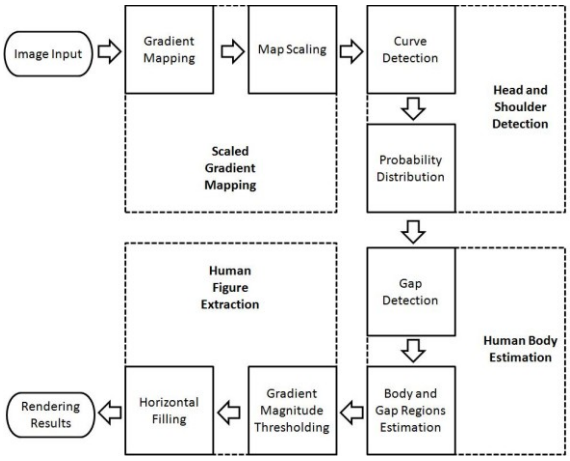


Figure 1. Flowchart of the proposed human detection method

$$G_x = [-1 \quad 0 \quad 1] \tag{1}$$

$$G_y = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \tag{2}$$

where G_x and G_y are the gradient components along x (horizontal direction) and y (vertical direction) axes of the image respectively.

The exact gradient magnitude and orientation values are then computed for every pixel, as given in Equations (3) and (4), then saved into two gradient maps, known as the magnitude map and the orientation map. The origin of the coordinate system is at the upper left corner.

$$G = \sqrt{G_x^2 + G_y^2} \tag{3}$$

$$\theta = \text{atan2}(G_y, G_x) \tag{4}$$

Both gradient maps are scaled into 8 smaller sizes in which the height and the width of the maps are divided by a scale factor S (where $S = \{2, 3, 4, 6, 8, 12, 16, 24\}$), this is done to accommodate various sizes of human in the image. During the scaling procedure, both maps are divided evenly into square blocks with length equal to S. For the magnitude maps, the local maximum magnitude values inside each block are saved. While for the orientation maps, both positive and negative directions of the gradient along x and y axes are used to obtain the orientation for the scaled map, as given in Equation (5).

$$\theta_S = \text{atan2}(G^+_{y,max} - G^-_{y,max}, G^+_{x,max} - G^-_{x,max}) \quad (5)$$

where θ_S is the new orientation value for the scaled orientation map with the scaling factor S, $G^+_{x,max}$ and $G^-_{x,max}$ are the maximum positive and negative x components of the gradient respectively, $G^+_{y,max}$ and $G^-_{y,max}$ are the maximum positive and negative y components of gradient respectively.

3.2 Head and Shoulder Detection

As both shoulders and the top half of the human head have distinctive curvatures, curve detection is used to detect possible head and shoulder in an image using curve templates on the scaled gradient maps. The templates consist of a left and a right curve blocks which are oriented outward with origins lie in the bottom right and bottom left corner respectively. The two curve templates with their orientation values are given in Figure 2.

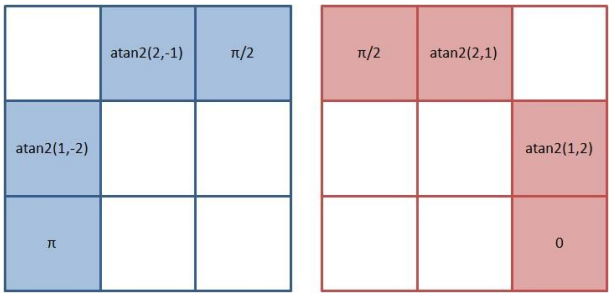


Figure 2. Curve templates (left and right)

All orientation maps are then searched for patches that contain similar orientation values using a constant 3×3 search window. The average difference between the patch and the template relate to its magnitude are calculated and compared to a degree of difference parameter p, as given in Equation (6).

$$G_{max} \frac{\sum_{i=1}^{4}(\theta_i - \mu_i)}{\sum_{i=1}^{4} G_i} \leq p \quad (6)$$

where θ_i and μ_i are the ith orientations in the block of the patch and template respectively, G_i is the ith magnitude in the block of the patch, G_{max} is the maximum magnitude of the image, p is the maximum average difference parameter, default value used is 30° (π/6).

When a region satisfies the maximum average differences allowed, it can be said that there is a high chance of seeing parts of the head or shoulder. The curve detection results of these two templates are then joined together by horizontal filling, that is filling the space between them. This is done only if the pixel distance between the left and right curve is smaller than a width threshold d, this width threshold is determined based on the maximum detectable size of the human figure in the current image, as given in Equation (7).

$$h\frac{\varphi}{2\varphi + 1} \leq d \quad (7)$$

where φ is the golden ratio (≈1.618), h is the maximum height (in pixels) of the image, d is the maximum allowed width (pixel distance) between the left and right curve for horizontal filling.

Golden ratio can be an excellent tool in estimating the proportions of the human body [22, 23], the model of the human body is created based on this to estimate the size and proportion of the human figure, as given in Figure 3. The size of the body is estimated between the top of the head to the knee. This model is also used to determine the width ratio d in Equation (7).

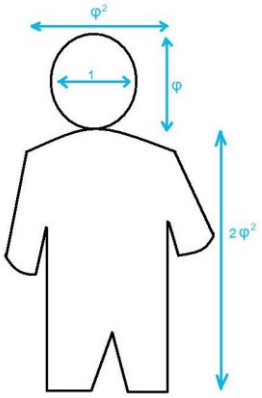


Figure 3. Human body size estimation

After filling the space, depending on the level of background noise, the image may contain several blobs instead of a head and a shoulder blob. To identify the actual head and shoulder blobs, all blobs are filtered through a probability model by using Gaussian probability distribution function. The pair of blobs with the highest probability ratio is then selected. The Gaussian probability distribution function used for the model is given in Equation (8).

$$f(j) = e^{-(j-r)^2/\sigma^2} \quad (8)$$

where j and r are the current and ideal results of a comparison between two blobs, σ is the standard deviation.

The comparisons and their related parameters are listed in Table 1, where w is the width of the head. The overall probability is then calculated, as given in Equation (9).

Table 1. Parameters used for the comparisons of two blobs

Type	Comparisons Between Two Blobs	r	σ
A	centroid difference in x	0	w
B	width ratio	$1/\varphi^2$	$1/(2\varphi^2)$
C	centroid difference in y to width ratio	$1/\varphi$	$1/(2\varphi)$
D	area ratio	$1/\varphi$	$1/(2\varphi)$

$$f(j) = f_A(j)\frac{f_B(j) + f_C(j) + f_D(j)}{3} \quad (9)$$

where A, B, C and D are the four feature comparisons between two blobs, A is given more weight since a small centroid difference in x direction tends to indicate a much higher chance of seeing the head and shoulder blobs.

3.3 Human Body Estimation

Once the head and shoulder blobs are identified, gap detection is then performed to locate possible gaps below the armpit and between the legs. These gaps can be assumed as regions with multiple edges and significant orientation changes. The detection is done by searching all orientation maps for patches that contain similar orientation values to the gap template, as given in Figure 4. The gap template is oriented inward with interception point lies on the centre of the bottom edge.

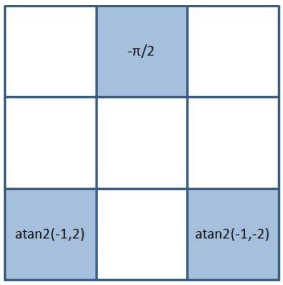


Figure 4. Gap template

Similar to Equation (6) in section 3.2, regions satisfy the maximum average differences p are extracted, where p is increased to 90° (π/2). This is to accommodate the high variance of the orientation values that belong to the gaps. If the width of the extracted gap region is greater than the width of the head, it is also removed.

The human body region estimation is then divided into 11 blocks and applied onto the image. The division of the model is given in Figure 5.

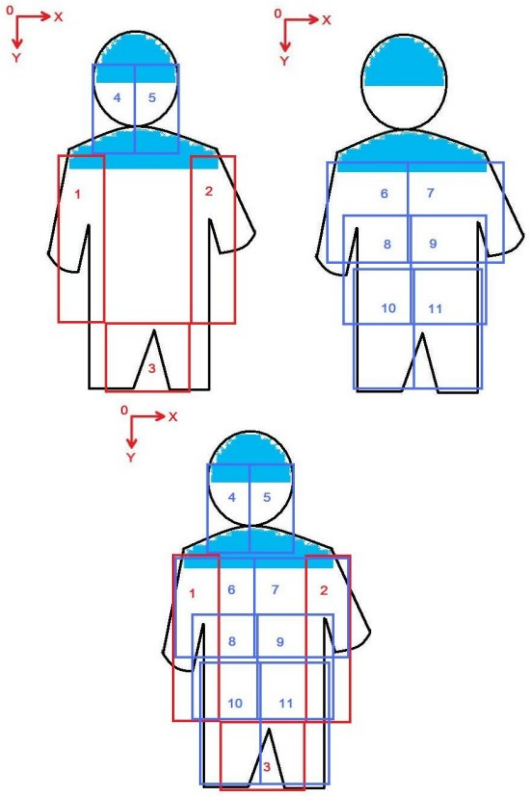


Figure 5. Human body region estimation

In the figure, the filled areas are the head and shoulder blobs, the 3 red blocks are the estimated gap regions, while the 8 blue blocks are the estimated body regions divided from human body size estimation (Figure 3). All the blocks overlaps at least one other regions (blocks 4 and 5 overlaps the head and shoulder blobs), this is done to ensure the maximum coverage and the accuracy in detecting the human body.

The size and position of the blocks are determined based on golden ratio φ and are listed in Table 2, where B_i represents the block number, x_0 and y_0 indicate the coordinates of the top left corner of the blocks, S_x and S_y are the centroid coordinates of the shoulder blob while H_x and H_y are the centroid coordinates of the head blob, w is the width of the head, positive directions for x and y are right and bottom respectively, H_{minx} and S_{miny} are the minimum x position for the head and minimum y postition for the shoulder respectively, G_9 and G_{10} are the outer most gap location in x for region 9 and 10 while G_{11} is the centroid coordinate of gap in x for region 11.

The head and gap regions (blocks 1–5) are determined first and are then used to locate the body regions (blocks 6–11). As regions below the gaps tend to be also gaps, a vertical filling process is further performed for gaps detected in the estimated gap regions (blocks 1–3), in which the lowest positioned gap pixels are filled vertically downwards to their corresponding region boundaries. If no gaps are detected, all body blocks will be aligned to H_x with equal width of $\varphi^2 w/2$.

Table 2. Size and position of the blocks

B_i	Width	Height	x_0	y_0
1	$w\varphi(\varphi - 1)$	$3\varphi^2 w/2$	$H_x - w\varphi(\varphi - ½)$	S_y
2	$w\varphi(\varphi - 1)$	$3\varphi^2 w/2$	$H_x + w\varphi(\varphi - ½)$	S_y
3	φw	$\varphi^2 w/2$	$H_x - \varphi w/2$	$S_{miny} + 3\varphi^2 w/2$
4	$w/2$	$S_y - H_y$	H_{minx}	H_y
5	$w/2$	$S_y - H_y$	H_x	H_y
6	$(G_{10} - G_9)/2$	$S_{miny} - S_y + \varphi^2 w$	G_9	S_y
7	$(G_{10} - G_9)/2$	$S_{miny} - S_y + \varphi^2 w$	$(G_{10} + G_9)/2$	S_y
8	$(G_{10} - G_9 + \varphi^2 w)/4$	$\varphi^2 w$	$(G_{11} - \varphi^2 w/2 + G_9)/2$	$S_{miny} + \varphi^2 w/2$
9	$(G_{10} - G_9 + \varphi^2 w)/4$	$\varphi^2 w$	$(G_9 + G_{10} + 2G_{11})/4$	$S_{miny} + \varphi^2 w/2$
10	$\varphi^2 w/2$	$\varphi^2 w$	$G_{11} - \varphi^2 w/2$	$S_{miny} + \varphi^2 w$
11	$\varphi^2 w/2$	$\varphi^2 w$	G_{11}	$S_{miny} + \varphi^2 w$

3.4 Human Figure Extraction

Gradient magnitude thresholding are then performed for all the blocks in which only the top k percentile of the pixels in the block that belong to the gradient magnitude is remained. This threshold is determined automatically in the algorithm by repeating the human figure extraction stage three times with different thresholds (where k = {10%, 20%, 30%}) and selecting the one that extracted the most pixels for the human figure. It is found empirically that best range for k is between 10% to 30%. This is done to extract important human edge features while accommodate different levels of background noise.

Blocks 4–11 are used to generate the outline of the body, while blocks 1–3 are used to generate the outline of the gap. A minimum body region is also added to the existing body outline. This consists of drawing 4 standing hollow ellipses with their properties listed in Table 3, where E_H, E_U, E_M and E_L represent 4 ellipses (E) consists of head, upper body, middle body and lower body, H_{miny} is minimum y position for the head. The purpose of these ellipses is to ensure a minimum amount of outlines are available, so that the body of the human figure can be filled effectively.

TABLE 3. Properties of ellipse used for the minimum body region

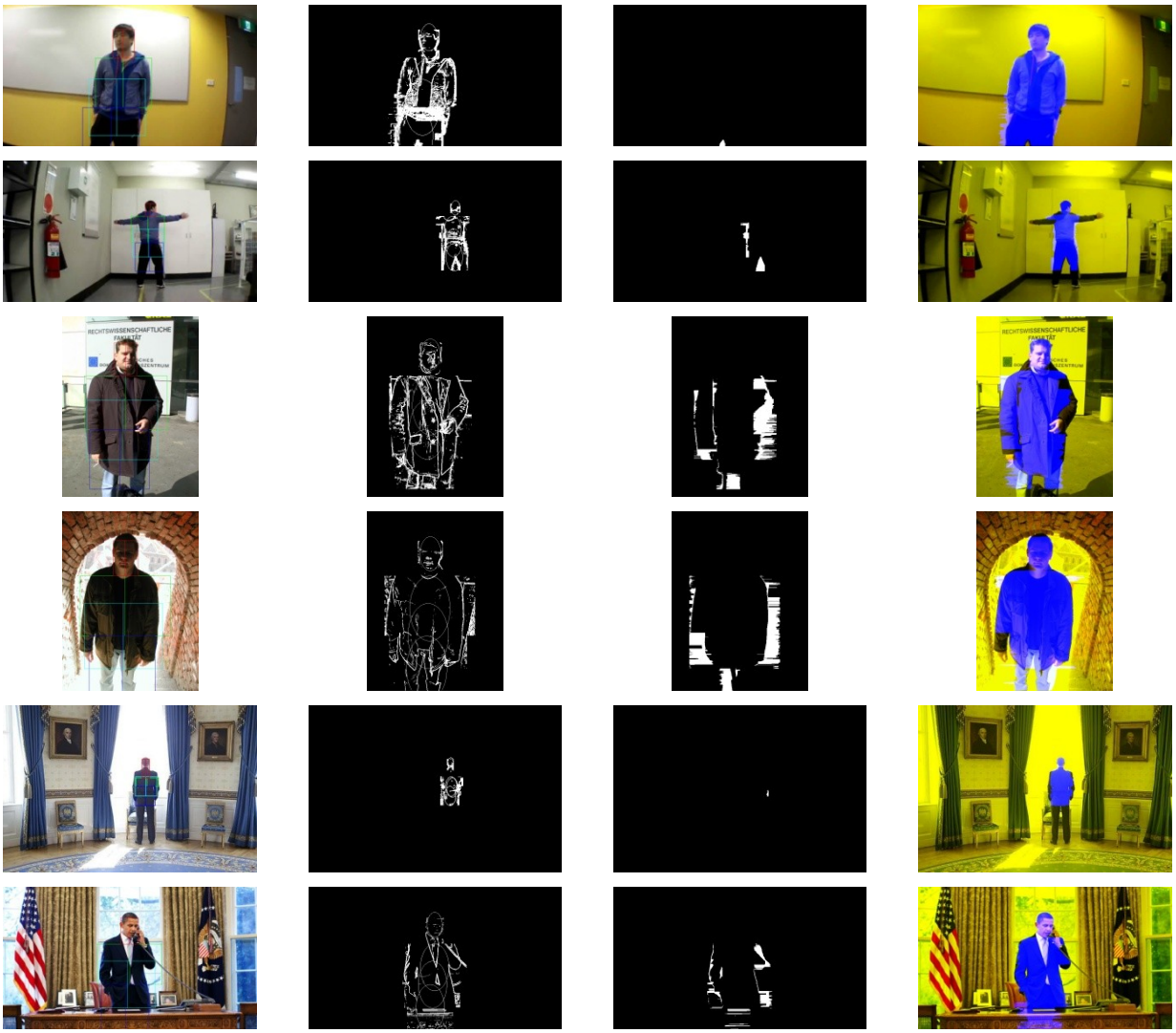
E	Centroid (x)	Centroid (y)	Diameter (x)	Diameter (y)
E_H	H_x	$(S_{miny} + H_{miny})/2$	$(S_{miny} - H_{miny})/\varphi$	$S_{miny} - H_{miny}$
E_U	$(G_{10} + G_9)/2$	$S_y + \varphi^2 w/2$	$\varphi w/2$	$\varphi^2 w/2$
E_M	$(G_9 + G_{10} + 2G_{11})/4$	$S_{miny} + \varphi^2 w$	$\varphi w/2$	$\varphi^2 w/2$
E_L	G_{11}	$S_{miny} + 3\varphi^2 w/2$	$\varphi w/2$	$\varphi^2 w/2$

The final step is the horizontal filling in which the inter-space between the outlines are filled. The blue blocks are filled per block while the red blocks are filled per gap region. The final human figure is generated by adding the head and shoulder blobs detected earlier with the filled body regions (blue) and removing the filled gap regions (red).

4 Experimental Results

Numerous images have been tested to examine the effectiveness of the proposed method, due to space limitations, only some results are shown in Figure 6. The images used are collected from various sources and have different sizes and quality. Six of these images have been presented in the results, the first two images (from the top) are produced by us using a camera of a mobile robot. The next two images (in the middle) are downloaded from INRIA Person Dataset (http://pascal.inrialpes.fr/data/human/) and the last two images are picked from Google images. These results indicate the high effectiveness of our method as the figures of the human body are clearly extracted.

In the figure, the original input images with body estimation regions indicated by the coloured rectangles are depicted in Figure 6(a). The unfilled body outlines generated from the human body regions (blue blocks) are shown in Figure 6(b), while the filled gap outlines generated from the gap regions (red blocks) are shown in Figure 6(c). The final results indicated by the blue coloured regions are presented in Figure 6(d).



(a) Body estimation regions (b) Unfilled body outlines (c) Filled gap outlines (d) Final Results

Figure 6. Human detection results

5 Conclusions

Current research in human detection is lacking of training-less approaches, therefore in this paper, a novel human detection method is introduced using gradient maps and the golden ratio. Compared to existing methods, the advantages of our methods are that firstly rather than an approximated region based human detection, the figure of the human body is completely extracted with high accuracy, secondly the algorithm only need one input image and no training images are required. Currently, the algorithm can only detect a single person with front or back view by selecting the pair of blobs with the highest probability ratio, future works are underway to address this issue.

References

[1] Paul M., Haque S. M. and Chakraborty S. Human detection in surveillance videos and its applications – a review. *Advances in Signal Processing*, 2013(176):1–16, 2013.

[2] Su F., Fang G. and Kwok N. M. Adaptive colour feature identification in image for object tracking. *Mathematical Problems in Engineering*, 2012(509597):1–18, 2012.

[3] Hou Y. L. and Pang G. K. H. People counting and human detection in a challenging situation. *Systems, Man and Cybernetics*, 41(1):24–33, 2011.

[4] Jafri R. and Arabnia H. R. A survey of face recognition techniques. *Journal of Information Processing Systems*, 5(2):41–68, 2009.

[5] Revathi A. R. and Kumar D. A review of human activity recongnition and behavior understanding in video surveillance. *Advanced Computer Science and Information Technology*, pages 375–384, Chennai, India, 2012.

[6] Chowdhury M. S., Kuang Y. C. and Ooi M. P. Fast and accurate human detection for video applications using edgelets. *Computer Applications and Industrial Electronics*, pages 74–79, Kuala Lumpur, Malaysia, 2010.

[7] Ogale N. A. A survey of techniques for human detection from video. *Master Thesis*, University of Maryland, 2006.

[8] Santhanam T., Sumathi C. P. and Gomathi S. A survey of techniques for human detection in static images. *Computational Science*, *Engineering and Information Technology*, pages 328–336, Coimbatore, India, 2012.

[9] Zhang L. J. and Liang Y. L. Motion human detection based on background subtraction. *Education Technology and Computer Science*, pages 284–287, Wuhan, China, 2010.

[10] Figueira D., Moerno P., Bernardino A., Gaspar J. and Victor J. S. Optical flow based detection in mixed human robot environments. *Visual Computing*, pages 223–232, Las Vegas, USA, 2009.

[11] Dalal N. and Triggs B. Histograms of oriented gradients for human detection. *Computer Vision and Pattern Recognition*, pages 886–893, San Diego, USA, 2005.

[12] Cortes C. and Vapnik V. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[13] Dalal N. Finding people in images and videos. *PhD Thesis*, Grenoble Institute of Technology, 2006.

[14] Schwartz W. R., Kembhavi A., Harwood D. And Davis L. S. Human detection using partial least squares analysis. *Computer Vision*, pages 24–31, Kyoto, Japan, 2009.

[15] Wold H. Partial least squares. *Encyclopedia of Statistical Sciences*, volume 6, pages 581–591. Wiley, New York, 1985.

[16] Li M., Zhang Z. X., Huang K. Q. and Tan T. N. Rapid and robust human detection and tracking based on omega-shape features. *Image Processing*, pages 2545–2548, Cairo, Egypt, 2009.

[17] Viola P. and Jones M. Rapid object detection using a boosted cascade of simple features. *Computer Vision and Pattern Recognition*, pages 511–518, Kauai, USA, 2001.

[18] Ammar B., Rokbani N. and Alimi A. M. Learning system for standing human detection. *Computer Science and Automation Engineering*, pages 300–304, Shanghai, China, 2011.

[19] Ammar B., Wali A. and Alimi A. M. Incremental learning approach for human detection and tracking. *Innovations in Information Technology*, pages 128–133, Abu Dhabi, UAE, 2011.

[20] Liu X. H., Jin Z. G. and Gao M. A robust approach for multi-human detection and tracking. *Consumer Electronics*, *Communications and Networks*, pages 832–835, Yichang, China, 2012.

[21] Lowe D. G. Distinctive image features from scale-invariant interest points. *Computer Vision*, 60(2):91–110, 2004.

[22] Spira M. On the golden ratio. *Mathematical Education*, pages 1–16, Seoul, Korea, 2012.

[23] Al-Kazaz Q. N. N. and Aldahham M. Y. A proposal method for selecting smoothing parameter with missing values. *Statistics in Science*, *Business*, *and Engineering*, pages 1–5, Langkawi, Malaysia, 2012.