



Supervised evolutionary learning: Use of gradient histogram and particle swarm algorithm to detection and tracking pedestrian in sequence of infrared images

Aprendizaje evolutivo supervisado: Uso de histograma de gradiente y algoritmo de enjambre de partículas para detección y seguimiento de peatones en secuencia de imágenes infrarrojas

Author:

 Seyed Mohammad Shahrokhi^{1*},
 Mohammadreza Amini²

INVESTIGACIÓN CIENTÍFICA

How to cite this article:

Shahrokhi S, Amini M. Supervised evolutionary learning: Use of gradient histogram and particle swarm algorithm to detection and tracking pedestrian in sequence of infrared images, *Innovaciencia* 2021; 9(1), 1-17.
DOI: <https://doi.org/10.15649/2346075X.2319>

Fecha de recepción:

Received: 12 July 2021
Accepted: 03 November 2021
Published: 01 December 2021

Keywords:

Pedestrian Identification; Infrared Imaging; Maximal Learning Neural Network; Gradient Histogram; Particle Swarm Optimization.

ABSTRACT

Recently, tracking and pedestrian detection from various images have become one of the major issues in the field of image processing and statistical identification. In this regard, using evolutionary learning based approaches to improve performance in different contexts can greatly influence the appropriate response. There are problems with pedestrian tracking/identification, such as low accuracy for detection, high processing time, and uncertainty in response to answers. Researchers are looking for new processing models that can accurately monitor one's position on the move. In this study, a hybrid algorithm for the automatic detection of pedestrian position is presented. It is worth noting that this method, contrary to the analysis of visible images, examines pedestrians' thermal and infrared components while walking and combines a neural network with maximum learning capability, wavelet kernel (Wavelet transform), and particle swarm optimization (PSO) to find parameters of learner model. Gradient histograms have a high effect on extracting features in infrared images. As well, the neural network algorithm can achieve its goal (pedestrian detection and tracking) by maximizing learning. The proposed method, despite the possibility of maximum learning, has a high speed in education, and results of various data sets in this field have been analyzed. The result indicates a negligible error in observing the infrared sequence of pedestrian movements, and it is suggested to use neural networks because of their precision and trying to boost the selection of their hyperparameters based on evolutionary algorithms.



*Corresponding author: Seyedmohammad Shahrokhi Email: sm.shahrokhi75@gmail.com

^{1*} Department of Electrical Engineering, Borujerd Branch, Islamic Azad University, Borujerd, Iran, Email: Sm.Shahrokhi75@Gmail.Com

² Department of Electrical Engineering, Borujerd Branch, Islamic Azad University, Borujerd, Iran, Email: Mr.Amini@Iaub.Ac.Ir

1. INTRODUCTION

It is needed to know that infrared rays are part of invisible spectra of light, which include electromagnetic radiation at wavelengths higher than normal and visible light [1, 2]. The spectrum of this light ranges from red to visible light; it is mostly in the range of a few hundred nanometers (430 THz) to 1 mm (300 GHz). On other hand, some consider it to have a spectrum of up to 1000 nm [4, 5]. Most of the heat radiation from objects at room temperature (the environment) is infrared [6, 7]. Infrared radiation has various applications in various sciences, most notably infrared detectors, night vision devices, telescopes, food heaters, airplanes, and driving applications [8]. One of the main uses of these waves is in the imaging process, in which they see lighter objects or objects that have lower temperatures. This property can be employed to analyze the moving human body above ambient temperature [9]. Recent research shows that this feature has been extensively used in pedestrian tracking despite problems such as failure to properly detect humans [10]. In short, in infrared imagery, the human form is brighter than fixed objects in the environment (such as trees, streets, etc.). Unfortunately, it would be difficult to distinguish pedestrians in question and separate them into visible images in this process. For example, things like animals, transformers, cars, power boxes, especially in the library, will complicate human tracking. Another problem is that a large amount of information is stored for sending information to track people elsewhere. It is also difficult to track the process if the target located in video frames is not large enough. Other problems in this field are the inability to trace multipurpose targets in infrared video frames that have received little attention from researchers. In this regard, having an efficient algorithm in tracking humans or pedestrians by a few pixels in a sequence of thermal or infrared images. Since it is important to utilize the imagery in human video surveillance, one or more of main objectives may be pursued, in this method categorizing steps as following (1) protection, protection, and safety, (2) control, management, and supervision (3) training and research [11].

Currently, human video surveillance systems are installed and used to protect and monitor important, high-risk, places (museums, shops, prisons, airports, subway and railway stations, hospitals, and schools). It will be easier to identify and prosecute individuals with this system [16].

2. SIMILAR METHODS

Machine learning methods have a wide range of applications [17-19]. In the present study, different goals have been studied in this regard. There are several methods for identifying humans or pedestrians that can be generalized based on the type of tracking techniques used. These include motion-based methods [9 and 12], brightness [11 and 13], neural networks [14], AdaBoost [3], Support vector machines - SVMs [17, 20], and finally, based on the separation technique. The separation techniques include Kalman filter, time graph, particle filter, Gaussian Mixture Model, mathematical maximum algorithm, and the like. In 2013, Elguebaly et al. [1] investigated the detection of targets in infrared images based on the Gaussian mixture model.

In 2014, Teutsch et al. [3] studied the detection of the pedestrian in infrared image Hotspot classification. On the other hand, Soundrapandiyan et al., 2015 tracked the objectives through image threshold and background subtraction algorithm in sample infrared images. Rajkumar et al., 2015 also employed local thresholds to find pedestrians. In 2015, Berg and his colleagues tracked objects based on different temperatures, which was assumed to be a baseline indicator [9]. Recently, extensive research has been carried out that demonstrates the importance of the subject, and most researchers are looking for novel methods, high accuracy, and good functional ability [20–26].

Conventional and non-conventional probability density functions can be estimated when a set of features is selected. Finally, the Likelihood Ratio Test (LRT) can identify how one moves. It should be kept in mind that the histogram ratio method has been suggested for identifying certain specific situations during human motion. Supervised methods have drawbacks that can be said: In the training data, pedestrian movements vary with different conditions and under different constraints and it is difficult to obtain visualizations in general and some methods are not generalizable. Lorca et al. (2012) applied the stereo method, which has advantages over mono methods in determining the relative distance between the pedestrian and the vehicle. Alahi et al. In 2013 [Alahi article] used their research in urban areas with low-quality cameras to identify pedestrians.

In their research, Armanfard et al [29] developed a method of Texture-Edge Descriptor. They aimed to find areas of the image that were more likely for a pedestrian on moving.

Negari et al. 2014 [30] investigated the pedestrian identification operation in three stages. At the pre-processing stage, they first identified animated areas of the image using a background subtraction algorithm based on surface lines. They then created the active feature space along with the Histogram of Oriented Gradients (HOG). In the second step, the pedestrians were able to prove the presence or absence of active feature space. In the third step, the proof of this hypothesis was analyzed by linear Support Vector Machines (SVMs). Combining these three stages of detection and tracking would reduce the amount of false-positive errors. Although previous approaches are effective in identifying and tracking pedestrians, the existence of various gaps/challenges in their identification process has led to the suggestion that other models address these challenges. In the present study, using a neural network with maximum learning capability for detection, it is suggested to use a wavelet kernel because its parameters are found by the particle swarm algorithm. Prior to that, the frameset features are found by the Histogram of Oriented Gradients (HOG), and the process of tracking infrared video sequences is made possible by Gaussian Mixture Model. In the third part of the present study, the proposed method is presented. Section 4: The results of implementing the proposed method will be presented separately. Section 5: Conclusion of discussions will have presented.

3. PROPOSED METHOD

The proposed method is based on two main phases and one final step : (1) tracking by a Gaussian mixture model; (2) Identification and detection that is the combination of feature extraction with Histogram of oriented gradients and using neural networks with maximum optimized learning in classification. Figure 1 clearly shows it.

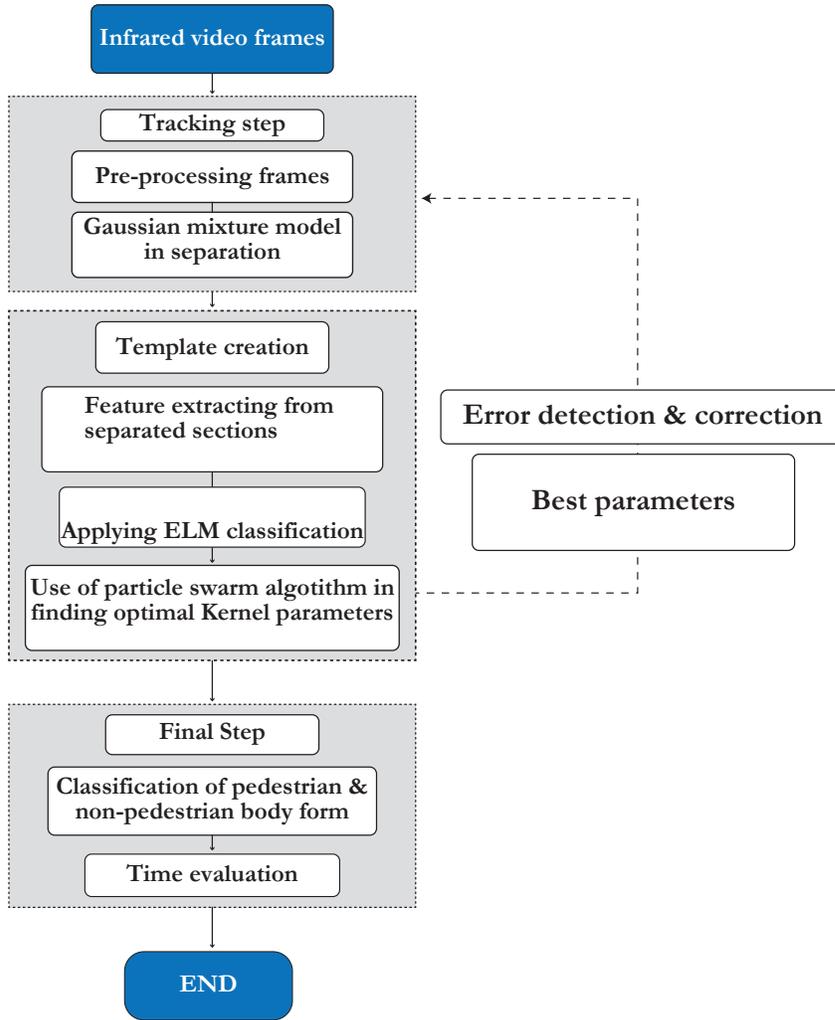


Fig 1. Steps of the proposed solution

3.1. TRACKING STEP

For tracking pedestrians, we remove the elements added to an image by applying a gentle filter. This process removes excess elements with namely density-based pixel aggregation (DBSPA). Furthermore, the Otsu threshold model is employed in this process to optimize initial separation and pre-processing. This model works mainly on different images. For example, medical images where disease information (date, name, etc.) is recorded on the image will be deleted.

The second step is to apply Gaussian Mixture Model (GMM), which is a parametric probability density function and shows the weighted mean for Gaussian component densities. This model uses the recording of some parameters at each iteration to maximum likelihood estimation. It is therefore similar to the K-means algorithm. The relevant parameters are approximated employing data training and the use of estimation algorithms (MAP or EM). Since the mixture distribution is a linear combination of probability density functions of its components, so [31]:

$$p(x) = w_1 p_1(x) + w_2 p_2(x) + \dots + w_n p_n(x) \quad (1)$$

In this equation, for $i = 1, \dots, n$ n is the number of mixture components p_i and w_i are Gaussian mixture model weights or coefficients. According to Equation (2), the mixture components in Gaussian mixture distribution are Gaussian density functions:

$$p(x) = w_1 N(x | \mu_1, \Sigma_1) + w_2 N(x | \mu_2, \Sigma_2) + \dots + w_n N(x | \mu_n, \Sigma_n) \quad (2)$$

The full Gaussian mixture model is parameterized based on covariance matrices, density weights of all components, and mean vectors. The GMM parameters will be in accordance with Equation (3):

$$\lambda = \{w_i, \mu_i, \Sigma_i\} \quad i = 1, \dots, n \quad (3)$$

In the above equation, μ_i is mean vector & Σ_i is a covariance matrix, and weights of mixture model apply

to $\sum_{i=1}^n w_i = 1$ equation. If a set of unspecified X data exists, the Gaussian mixture model is defined so that it can match X distribution. The approximations to maximum likelihood can be obtained by using a special case of the mathematical expectation-maximization algorithm [31]. The EM algorithm is dependent on hidden variable λ where the maximum likelihood is obtained by using X set training:

$$p(X | \lambda) = \prod_{t=1}^T p(x_t | \lambda) \quad (4)$$

To maximum likelihood between the samples and Gaussian distribution based on these equations, the model parameters are repeatedly modified. EM algorithm consists of two steps:

A) Expectation maximization algorithm where GMM parameters for each sample of d-dimensional $x \in \{X\}_{t=1, \dots, T}$ data are obtained by deductive probability & for i-th component by using equation (5):

$$P(i | x_t, \lambda) = \frac{w_i g(x_t | \mu_i, \Sigma_i)}{\sum_{k=1}^n w_k g(x_t | \mu_k, \Sigma_k)} \quad (5)$$

That $g(x_t | \mu_i, \Sigma_i)$ is given in accordance with equation (6)

$$g(x_t | \mu_i, \Sigma_i) = \frac{1}{\sqrt{(2\pi)^d \times \Sigma_i}} \exp\left\{-\frac{1}{2}(x_t - \mu_i)' \Sigma_i^{-1} (x_t - \mu_i)\right\} \quad (6)$$

B) Maximum likelihood where the parameters are estimated according to deductive probability calculated in previous step. On other hand, GMM parameters are updated according to equations (7) to (9):

$$\bar{w}_i = \frac{1}{T} \sum_{t=1}^T P(i | x_t, \lambda) \quad (7)$$

$$\bar{\mu}_i = \frac{\sum_{t=1}^T P(i|x_t, \lambda) x_t}{\sum_{t=1}^T P(i|x_t, \lambda)} \quad (8)$$

$$\bar{\sigma}_i^2 = \frac{\sum_{t=1}^T P(i|x_t, \lambda) x_t^2}{\sum_{t=1}^T P(i|x_t, \lambda)} - \bar{\mu}_i^2 \quad (9)$$

The algorithm loops are repeated until a boundary of convergence is created.

3.2. TEMPLATE CREATION

The algorithm's construction step consists of three phases. These steps include (1) Feature extraction from segment separated by the HOG method, (2) applying the feature vector to neural network classification by maximal learning with wavelet kernel (3) using particle swarm algorithm by particle swarm optimization algorithm.

3.2.1. FEATURE EXTRACTION

In the HOG method, even if we do not know the exact position of the gradient or corresponding edges, the distribution of local gradients or edge orientations can well describe the image. This feature shows nearby image directions. Each of these directions is called a cell. Depending on whether the gradient is asymptomatic or not, the interval is divided into n equal parts by 0-180 degrees or 0-360 degrees. In which, n is the number of gradient directions or all histogram intervals. Each of these distances forms a histogram channel. In each cell, the direction-gradient histogram is calculated for pixels within the cell. The histograms are then combined to form a data block and normalized to resist light changes. In this respect, the normalized histogram is called the direction gradient. The blocks can mostly be overlapped. However, feature vector lifetime increases, but performance accuracy will improve. For color images, the gradient for each color channel is calculated separately. The largest value for each pixel is calculated as the gradient vector in each cell.

3.2.2. CLASSIFICATION

Neural networks are mainly referred to as maximal learning machines and were described by Huang, 2014. The neural network is another popular classifier in machine learning. One of the most important applications of neural networks in a process simulation is that an independent relationship can be inferred between its input and output. Maximum learning machines have some kind of mechanism that can be the link between machine learning and biological learning and neural networks as a supervised bioinspired algorithm which can be used for general learning (i.e., classification or regression) purposes. It's mostly one-step learning and one for the linear model, but when the problem is nonlinear and complex, then we'll need kernels. The most important maximal learning machines associated with kernels include linear kernels, polynomials, radial basis function, and wavelet kernels. Among kernels, the wavelet model typically has a better response and operates faster [32]. The machine outputs with maximum learning potential of m neurons and activation function f are shown below:

$$o_j = \sum_{i=1}^m \beta_i f(l_i x_r + b_i) \quad (5)$$

This algorithm performs well and fast. As well, to minimize training error in machine algorithms with maximal learning, training error and output weights should be minimized simultaneously. Therefore, this process will be effective in increasing the overall performance of neural networks:

$$\begin{aligned} & \text{Min } \|AS - C\|, \\ & \|S\| \end{aligned} \quad (6)$$

That can write:

$$S = A^T \left(\frac{1}{E} + AA^T \right)^{-1} C \quad (7)$$

Where E indicates the adjustment factor, A is the hidden layer output matrix, and C is the expected output matrix of samples. Therefore, the output function of the maximum learning algorithm is described as follows:

$$u(r) = v(r) A^T \left(\frac{1}{E} + AA^T \right)^{-1} C \quad (8)$$

The wavelet kernel function is suitable for simulations and better performance. It can be expressed as:

$$b(r, g) = \cos\left(w \frac{\|r - g\|}{x}\right) \exp\left(-\frac{\|r - g\|^2}{y}\right) \quad (9)$$

In this equation to use this kernel, three parameters r , g , and w , must be precisely determined. The most common method for finding the best fit values of three parameters r , g , and w is to fix one of the parameters and search in the defined interval for other parameters. Determining these parameters will greatly improve the accuracy of the process.

3.2.3. OPTIMIZATION

Particle Swarm Optimization Algorithm is a population-based parallel search algorithm that operates with a group including random solutions (particles). Each particle is multidimensional (depending on the problem) with two vectors $V_i [t]$ and $X_i [t]$ which represent the current position and velocity at the moment of t associated with particle i , respectively. At each stage of population movement, the location of each particle is replaced by two values of best personal experience and best group experience. In each iteration, the algorithm updates particle based on equations (10) and (11) after finding a new velocity and position:

$$V_i [t + 1] = w \times v_i [t] + C_1 r_1 (X_i^{i, best} [t] + X_i [t]) + C_2 r_2 (X_i^{g, best} [t] - X_i [t]) \quad (10)$$

$$X_i[t + 1] = X_i[t] + V_i[t + 1] \quad (11)$$

In which, W represents the inertia coefficient that is linearly in range (1-0.25). Furthermore, the coefficients of C_1 and C_2 , level of personal and group experience in the interval (1-0.5), are selected for either 1.5 or 1.2. Two numbers $r1$ and $r2$ are random in the range (1-0). The final velocity value of each particle is restricted to an interval to prevent the algorithm from diverging, where V is a member of maximum & minimum amplitudes. Equation (11) updates the vector of particle current position due to its new velocity. The binary algorithm update equation is changed in two steps. Step One: Use Equation (10) to change velocity. Step Two: Applying sigmoid function in Equation (12) to limit velocity between 0 and 1:

$$\begin{aligned} & \text{If } (\text{rand}(\) < \mathbf{S}(V_{ij}[t + 1])) \text{ Then } X_{ij}[t + 1] = 1 \\ & \text{Else} \\ & X_{ij}[t + 1] = 0 \text{ and } s(v_{ij}[t + 1]) = \frac{1}{1 + e^{-v_{ij}[t + 1]}} \end{aligned} \quad (12)$$

In an evolutionary algorithm to select the best values and parameters, the algorithm continues until there is no boundary of convergence or until it isn't limited to reproducible generation. fitness function according to relation (13) is as follows:

$$\text{Fitness} = \alpha \times (1 - G_{\text{mean}}) + (1 - \alpha) \times (r \times g) \quad (13)$$

In the previous equation, α denotes the magnitude or coefficient of error obtained from a class prediction by classification. Besides, r and g represent neural network wavelet kernel parameters, respectively, with maximum learning. Moreover, G_{mean} is also derived from its classification function, which is directly related to its accuracy. This cost function aims to minimize the amount of error caused by classification.

4. RESULT

The obtained results of the simulation include the tracking and classification section which are discussed separately. The Gaussian mixture model is used in the tracking section. In the second part, the identification is performed using the HOG method and neural networks with optimization maximal learning. All video was randomly converted into 20 Clips (AVI format, resolution: 320×240 pixels, at a rate of 15 frames per second). It includes pedestrians and thermal spectrum objects in 10 clips, and all pedestrian-like clips were inserted in another 10 clips. The total time of these clips is about 30 minutes, and about 20,000 frames were selected. Some videos that were filmed in an environment (pedestrians or objects similar to a human body) were counted against each other. In total, from 20,000 video frames from infrared images, about 53,489 thermal objects can be seen, that almost 32412 of them is related to pedestrian movement and the rest related to other objects such as cars, trees, animals, and devices that show self-heating. In the first step, the objects that were correctly identified as thermal objects in the image have been tracked. In the second step, the intention is to separate the pedestrians from non-pedestrians. There are various features for pedestrian detection that we extracted from proposed features by the Extractor that basically includes information from foreground pixels, edges, and other key points after tracking. It should be noted that since entire features

were obtained from all frames, color and location features were cited as low-level features such as local features including contrast, homogeneity, size, shape, and edge. Figure 2 shows the results of the tracking step. Figures 3 and 4 illustrate the performance of the classification algorithm in separating detection steps.

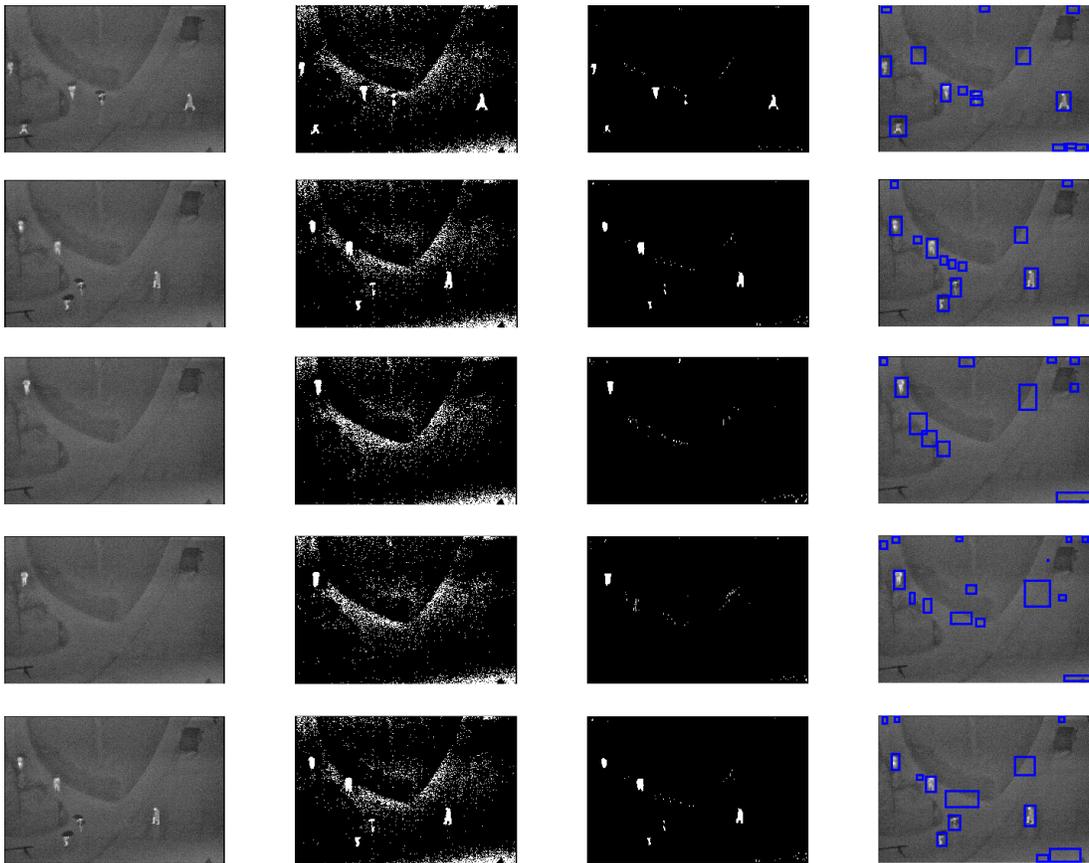


Fig. 2. To read (left to right):

Column 1: includes infrared imaging of pedestrian motion.

Column 2: shows threshold actions.

Column 3: Displays Separation based on pre-processing by applying the filter.

Column 4: Describes tracking of various thermal objects.



Fig. 3. Tracking in infrared image collection

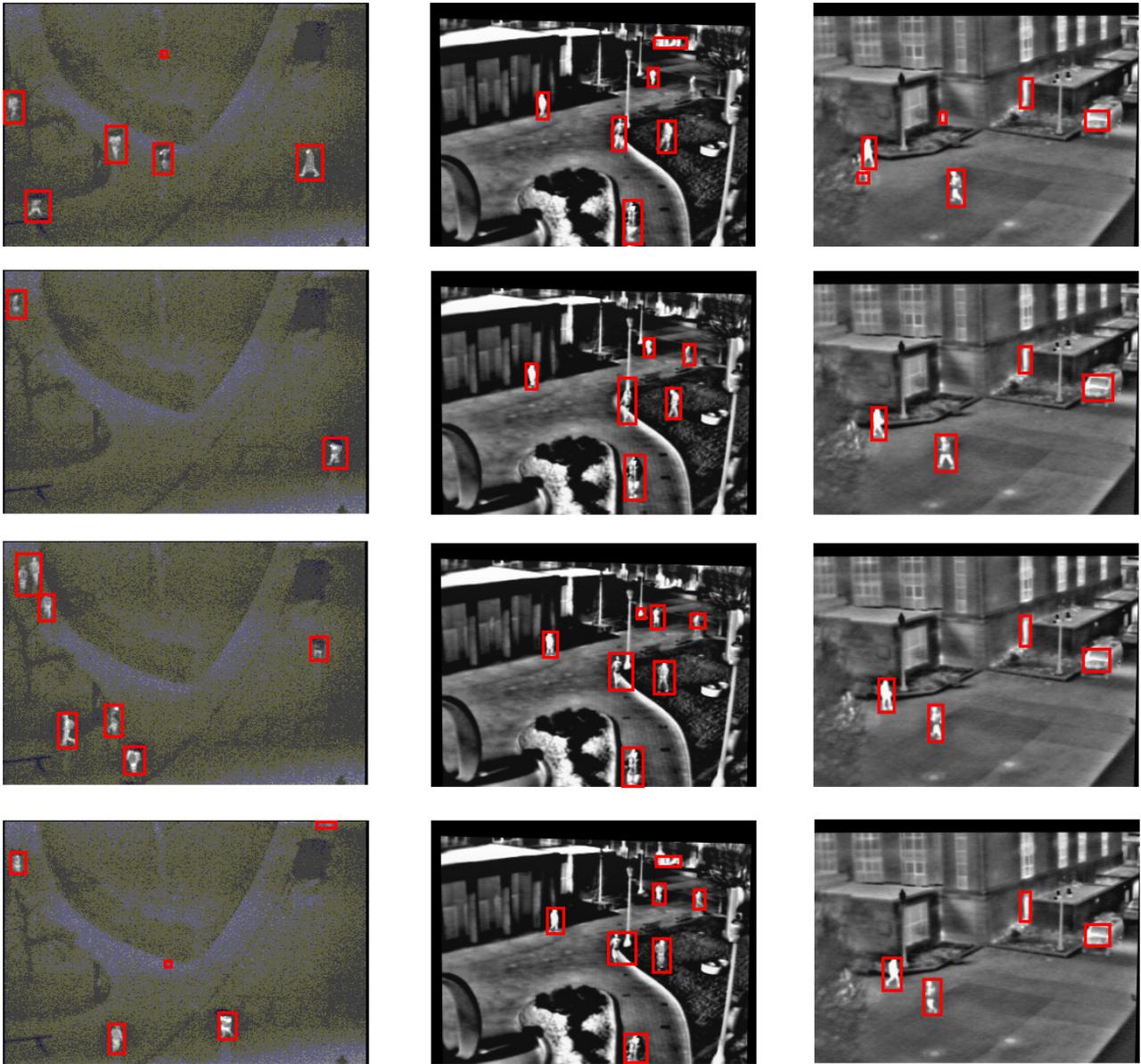


Fig. 4. Implementing the second stage of simulation in target classification in infrared images for different frames.

In the first phase, all objects that have been detected in the image are detected correctly, but in the second part, it is intended to be isolated from non - passers-by. There are many different features for pedestrian detection, which we extracted by the proposed analyzer and originally contained information from the foreground, edge, and other key points following the detection. The global features were also received from the entire frame and indicated features such as color and location along with the local characteristics such as contrast, homogeneity, size, shape, and edge as low-level features. in Figures (3) and (4), we see the performance representation of the classifier algorithm in the breakdown of detection steps.

Table 1 shows classification results. In this table, wavelet kernels, radial basis function (RBF), and PSO and GA algorithms are compared together. This operation was repeated 10 times and the K-fold value with K equal to 10 was tested. Finally, the results were computed as a vague matrix.

Table 1: The obtained results of model implementation compared to similar methods in column 1. These results were obtained in the experimental step by dividing data by CV equal 10 composed moderate method.

	ELM-Wavelet-PSO			ELM-RBF-PSO			ELM-Wavelet-GA		
Fold 1	Confusion Matrix	precision		Confusion Matrix	precision		Confusion Matrix	precision	
	0.0564	0.9463	0.9439	0.0643	0.9357	0.9150	0.0763	0.9237	0.9224
	0.9416	0.0583		0.8944	0.1056		0.9211	0.0789	
Fold 2	Confusion Matrix	precision		Confusion Matrix	precision		Confusion Matrix	precision	
	0.0662	0.9337	0.9337	0.1146	0.8854	0.8760	0.0754	0.9246	0.9320
	0.9337	0.0526		0.8667	0.1333		0.9394	0.0606	
Fold 3	Confusion Matrix	precision		Confusion Matrix	precision		Confusion Matrix	precision	
	0.0991	0.9009	0.9101	0.1477	0.8523	0.8869	0.0510	0.9490	0.9172
	0.9193	0.0806		0.9216	0.0784		0.8854	0.1146	
Fold 4	Confusion Matrix	precision		Confusion Matrix	precision		Confusion Matrix	precision	
	0.0857	0.9142	0.9285	0.0603	0.9397	0.9131	0.0468	0.9532	0.9377
	0.9428	0.0571		0.8865	0.1135		0.9223	0.0777	
Fold 5	Confusion Matrix	precision		Confusion Matrix	precision		Confusion Matrix	precision	
	0.0471	0.9529	0.9454	0.1104	0.8896	0.9120	0.0641	0.9359	0.9211
	0.9379	0.0621		0.9345	0.0655		0.9064	0.0936	
Fold 6	Confusion Matrix	precision		Confusion Matrix	precision		Confusion Matrix	precision	
	0.0615	0.9384	0.9441	0.0536	0.9464	0.9204	0.0825	0.9175	0.9337
	0.9499	0.0501		0.8944	0.1056		0.9500	0.0500	
Fold 7	Confusion Matrix	precision		Confusion Matrix	precision		Confusion Matrix	precision	
	0.0771	0.9229	0.9399	0.0593	0.9407	0.9129	0.0438	0.9562	0.9316
	0.9569	0.0431		0.8852	0.1148		0.9070	0.0930	
Fold 8	Confusion Matrix	precision		Confusion Matrix	precision		Confusion Matrix	precision	
	0.0716	0.9283	0.9400	0.0364	0.9636	0.9157	0.0773	0.9227	0.9159
	0.9518	0.0481		0.8778	0.1222		0.9091	0.0909	
Fold 9	Confusion Matrix	precision		Confusion Matrix	precision		Confusion Matrix	precision	
	0.0538	0.9462	0.9399	0.0663	0.9337	0.9319	0.0890	0.9110	0.8875
	0.9336	0.0663		0.9300	0.0700		0.8641	0.1359	
Fold 10	Confusion Matrix	precision		Confusion Matrix	precision		Confusion Matrix	precision	
	0.0457	0.9542	0.9444	0.1172	0.8828	0.8982	0.0503	0.9497	0.9247
	0.9347	0.0652		0.9137	0.0863		0.8997	0.1003	
average	0.9402	0.9338	0.9369	0.9005	0.9169	0.9074	0.9104	0.9343	0.9223

5. CONCLUSION

Recent studies and investigations into the detection and tracking of pedestrians indicate that researchers have paid particular attention to this issue in various researches. In the present article, we consider detection and tracking to study. For this aim, we used the Gaussian mixture model algorithm to estimate motion and motion tracking. In the next step, after detecting pedestrians from human-like objects employing the histogram of the oriented gradients and neural networks or maximal learning, we calculated the presence or absence of the pedestrian with an error of less than 8%. In the future, we will review the implementation process for more data. Applying this model can be effective in pedestrian detection in public places. The proposed method in the paper often makes the separation process based on the filter model and classification in decision making.

Despite the high precision, the possibility of visually impaired pedestrians in the form of recognition, as well as the filter implementation, has some disadvantages. because of the high sensitivity to image quality which causes the desired accuracy of the output and suggests that this problem can be achieved by using an efficient technique like wavelet transform in image quality improvement.

REFERENCES

1. **Zou, H., Sun, H., & Ji, K.** (2012, December). Real-time infrared pedestrian detection via sparse representation. In *Computer Vision in Remote Sensing (CVRS), 2012 International Conference on* (pp. 195-198). IEEE.
2. **Wang, J. T., Chen, D. B., Chen, H. Y., & Yang, J. Y.** (2012). On pedestrian detection and tracking in infrared videos. *Pattern Recognition Letters*, 33(6), 775-785. <https://doi.org/10.1016/j.patrec.2011.12.011>
3. **Teutsch, M., & Müller, T.** (2013, May). Hot spot detection and classification in LWIR videos for person recognition. In *SPIE Defense, Security, and Sensing* (pp. 87440F- 87440F). International Society for Optics and Photonics. <https://doi.org/10.1117/12.2015754>
4. **Elguebaly, T., & Bouguila, N.** (2013). Finite asymmetric generalized Gaussian mixture models learning for infrared object detection. *Computer Vision and Image Understanding*, 117(12), 1659-1671. <https://doi.org/10.1016/j.cviu.2013.07.007>
5. **Teutsch, M., Muller, T., Huber, M., & Beyerer, J.** (2014). Low resolution person detection with a moving thermal infrared camera by hot spot classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 209-216). <https://doi.org/10.1109/CVPRW.2014.40>
6. **Akhoulfi, M. A., Porcher, C., & Bendada, A.** (2014, June). Fusion of thermal infrared and visible spectrum for robust pedestrian tracking. In *SPIE Defense+ Security* (pp. 90760O-90760O). International Society for Optics and Photonics. <https://doi.org/10.1117/12.2063306>
7. **Soundrapandiyam, R., & Mouli, P. C.** (2015). Adaptive Pedestrian Detection in Infrared Images Using Background Subtraction and Local Thresholding. *Procedia Computer Science*, 58, 706-713. <https://doi.org/10.1016/j.procs.2015.08.091>
8. **Rajkumar, S., & Mouli, P. C.** (2015, February). Pedestrian detection in infrared images using local thresholding. In *Electronics and Communication Systems (ICECS), 2015 2nd International Conference on* (pp. 259-263). IEEE. <https://doi.org/10.1109/ECS.2015.7124904>
9. **Berg, A., Ahlberg, J., & Felsberg, M.** (2015, August). A thermal Object Tracking benchmark. In *Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on* (pp. 1-6). IEEE. <https://doi.org/10.1109/AVSS.2015.7301772>
10. **Hwang, S., Park, J., Kim, N., Choi, Y., & So Kweon, I.** (2015). Multispectral pedestrian detection: Benchmark dataset and baseline. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1037-1045). <https://doi.org/10.1109/CVPR.2015.7298706>
11. **Pawłowski, P., Piniarski, K., & Dąbrowski, A.** (2015, September). Pedestrian detection in low resolution night vision images. In *Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2015* (pp. 185-190). IEEE. <https://doi.org/10.1109/SPA.2015.7365157>
12. **Yang, C., Liu, H., Liao, S., & Wang, S.** (2015). Pedestrian Detection in Thermal Infrared Image Using Extreme Learning Machine. In *Proceedings of ELM-2014 Volume 2* (pp. 31-40). Springer International Publishing. https://doi.org/10.1007/978-3-319-14066-7_4
13. **Zhao, X., He, Z., Zhang, S., & Liang, D.** (2015). Robust pedestrian detection in thermal infrared imagery using a shape distribution histogram feature and modified sparse representation classification. *Pattern Recognition*, 48(6), 1947-1960. <https://doi.org/10.1016/j.patcog.2014.12.013>
14. **Zhuang, J., & Liu, Q.** (2016). Transferred IR pedestrian detector toward distinct scenarios adaptation. *Neural Computing and Applications*, 27(3), 557-569. <https://doi.org/10.1007/s00521-015-1877-0>

15. **Ma, Y., Wu, X., Yu, G., Xu, Y., & Wang, Y.** (2016). Pedestrian Detection and Tracking from Low-Resolution Unmanned Aerial Vehicle Thermal Imagery. *Sensors*, 16(4), 446. <https://doi.org/10.3390/s16040446>
16. **Dai, C., Zheng, Y., & Li, X.** (2007). Pedestrian detection and tracking in infrared imagery using shape and appearance. *Computer Vision and Image Understanding*, 106(2), 288-299. <https://doi.org/10.1016/j.cviu.2006.08.009>
17. **Seryasat, O. R., Honarvar, F., & Rahmani, A.** (2010, October). Multi-fault diagnosis of ball bearing based on features extracted from time-domain and multi-class support vector machine (MSVM). In 2010 IEEE International Conference on Systems, Man and Cybernetics (pp. 4300-4303). IEEE. <https://doi.org/10.1109/ICSMC.2010.5642390>
18. **Seryasat, O. R., & Haddadnia, J.** (2018). Evaluation of a new ensemble learning framework for mass classification in mammograms. *Clinical breast cancer*, 18(3), e407-e420. <https://doi.org/10.1016/j.clbc.2017.05.009>
19. **Seryasat, Omid Rahmani, and Javad Haddadnia.** "Evaluation of a new ensemble learning framework for mass classification in mammograms." *Clinical breast cancer* 18.3 (2018): e407-e420. <https://doi.org/10.1016/j.clbc.2017.05.009>
20. **Biswas, S. K., & Milanfar, P.** (2017). Linear support tensor machine with LSK channels: Pedestrian detection in thermal infrared images. *IEEE transactions on image processing*, 26(9), 4229-4242. <https://doi.org/10.1109/TIP.2017.2705426>
21. **Cai, Y., Liu, Z., Wang, H., & Sun, X.** (2017). Saliency-based pedestrian detection in far infrared images. *IEEE Access*, 5, 5013-5019. <https://doi.org/10.1109/ACCESS.2017.2695721>
22. **Ma, M.** (2019). Infrared pedestrian detection algorithm based on multimedia image recombination and matrix restoration. *Multimedia Tools and Applications*, 1-16.
23. **Kwak, J. Y., Ko, B. C., & Nam, J. Y.** (2017). Pedestrian tracking using online boosted random ferns learning in far-infrared imagery for safe driving at night. *IEEE Transactions on Intelligent Transportation Systems*, 18(1), 69-81. <https://doi.org/10.1109/ITITS.2016.2569159>
24. **Bai, X., Wang, Y., Liu, H., & Guo, S.** (2018). Symmetry information based fuzzy clustering for infrared pedestrian segmentation. *IEEE Transactions on Fuzzy Systems*, 26(4), 1946-1959. <https://doi.org/10.1109/TFUZZ.2017.2756827>
25. **Shen, G., Zhu, L., Jihan, L. O. U., Shen, S., Liu, Z., & Tang, L.** (2019). Infrared multi-pedestrian tracking in vertical view via Siamese Convolution Network (December 2018). *IEEE Access*. <https://doi.org/10.1109/ACCESS.2019.2892469>
26. **Lahmyed, R., El Ansari, M., & Ellahyani, A.** (2018). A new thermal infrared and visible spectrum images-based pedestrian detection system. *Multimedia Tools and Applications*, 1-25. <https://doi.org/10.1007/s11042-018-6974-5>
27. **Hsu, W. Y.** (2018). Automatic pedestrian detection in partially occluded single image. *Integrated Computer-Aided Engineering*, 25(4), 369-379. <https://doi.org/10.3233/ICA-170573>
28. **Llorca, D. F., Sotelo, M. A., Hellín, A. M., Orellana, A., Gavilán, M., Daza, I. G., & Lorente, A. G.** (2012). Stereo regions-of-interest selection for pedestrian protection: A survey. *Transportation research part C: emerging technologies*, 25, 226-237. <https://doi.org/10.1016/j.trc.2012.06.006>
29. **Armanfard, N., Komeili, M., & Kabir, E.** (2012). TED: A texture-edge descriptor for pedestrian detection in video sequences. *Pattern Recognition*, 45(3), 983-992. <https://doi.org/10.1016/j.patcog.2011.08.010>
30. **Negri, P., Goussies, N., & Lotito, P.** (2014). Detecting pedestrians on a movement feature space. *Pattern recognition*, 47(1), 56-71. <https://doi.org/10.1016/j.patcog.2013.05.020>
31. **Rezaee, Kh, Haddadnia, J., Delbari, A.** Elderly drop monitoring system based on Gaussian hybrid model and body anatomical changes in video images, *Journal of Machine Vision and Image Processing*, 1(2), 76-77
32. **Huang, G. B., Zhou, H., Ding, X., & Zhang, R.** (2011). Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(2), 513-529. <https://doi.org/10.1109/TSMCB.2011.2168604>