The Scientific Journal of Cihan University – Slemani **PP: 18-32**
Volume (4), Issue (1), June 2020
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Multi Moving Object detection system using Modified K-means and Deep Learning Algorithms algorithm

Azhee Wria Muhamad[1], Luqman Mohammed Mustafa[2], Ari Sabir Arif[3]

[1,2,3] Computer Science Department, College of Basic Education, University of Sulaimani, Sulaimani, Iraq

E-mail :azhee.muhamad@univsul.edu.iq[1], Luqman.mustafa@univsul.edu.iq[2], ari.arif@univsul.edu.iq[3]

Abstract:

We present a multi moving object detection system for surveillance systems that detect multiple moving objects in varying lighting conditions. This work proposes a modified k-means algorithm and deep neural networks for feature extraction and recognizing and tracking of the objects. Traditional k-means algorithm is time-consuming, so we modify the k-means algorithm for speed and optimized feature extraction. The simulations performed as a part of the study indicates better accuracy precision and the F1 score (F-score or F-measure an amount of a test's accuracy. It accounts both the precision p and the recall (r) of the test to compute the score: (p) is the number of correct positive product divided by the number of all positive results come back by the classifier, and (r) is the number of true positive results separated by the number of all relevant sample) scores for the proposed method. The performance of the present approach was compared to that of the other works, and the results indicate that the proposed system offered in this paper can be considered a suitable in terms of quality, accuracy, and speed.

Keywords: Deep learning, Object Detection, K-means, Video surveillance

الملخص:

نقدم نظام كشف الأجسام المتعددة الحركة لأنظمة المراقبة التي تكتشف الأجسام المتحركة المتعددة في ظروف الإضاءة المختلفة. يقترح هذا العمل خوارزمية K-Mean الوسائل والشبكات العصبية العميقة لاستخراج المعالم والتعرف على الأشياء وتتبعها. تستغرق خوارزمية k-mean التقليدية وقتًا طويلاً ، لذلك تقترح بتعديل خوارزمية k-mean للسرعة واستخراج الميزة الأمثل. تشير عمليات المحاكاة التي تم إجراؤها كجزء من الدراسة إلى دقة أفضل F1- (F-scoreأو F-measureمقدار من دقة الاختبار ، وهي تحسب الدقة p والاستدعاء (r) للاختبار لحساب النتيجة: (p) هو رقم المنتج الإيجابي الصحيح مقسومًا على عدد النتائج الإيجابية التي تم إرجاعها بواسطة المصنف ، و (r) هو عدد النتائج الإيجابية الحقيقية مفصولة برقم كل العينة ذات الصلة) الطريقة المقترحة. تمت مقارنة أداء النهج الحالي بأداء الأعمال الأخرى ، وتشير النتائج إلى أن النظام المقترح المقدم في هذه الورقة يمكن اعتباره مناسبًا من حيث الجودة والدقة والسرعة.

پوختە:

ئێمە ئەم توێژینەوەیە پێشکەش دەکەین کە تایبەتە بە سیستەمی دەرخستنی تەنە فرەجوڵەکان بۆ چاودێری کردنی ئەو سیستەمانەی کە کەشفی ئەو تەنە فرەجوڵانە دەکەن کە جیاوازن لە حاڵەتەکانی رووناکی دا.ئەم کارە پێشنیاری گۆرنکاری لە بابەتی K-Mean ئەلگۆریزم و بنەما سەرەکیەکانی شەبەکات بۆ سیفاتەکانی دەرکردن و ناسینەوە و دۆزینەوەی بابەتەکان .لە رێگا کۆنکەی K-Mean ئەلگۆریزم کات بەفیڕۆ دەدات.هەربۆیە ئێمە گۆرنکاری دەکەین لە K-Mean ئەلگۆریزم بۆ خێرایی و باشکردنی سیفاتی دەرکردن. پرۆسەکان جێبەجێ کراون وەک بەشێک لە م توێژینەوەیە کە ئەمەش ئاماژە دەدات بە باشتر لە وورددبینکردن و وەڵە نمرەی F1 (نمرەی Fیان پێوانی F بڕێکە لە وەردبینی تاقیکردنەوەیەک . ئەمەش حساب دەکرێت بۆ

DOI: http://dx.doi.org/10.25098/4.1.23

The Scientific Journal of Cihan University – Slemani **PP: 18-32**
Volume (4), Issue (1), June 2020
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

هەردوو وردبینی (p) و بیرهێنانەوەی (r) لە تاقیکردنەوەدا بۆ حساب کردنی نمرەکە: p ژمارەیەکە لە ڕاستی ئەرێنی ئە بەر هەمەی کەدابەشکراوەبە سەر ئەو ژمارەیە کە لە هەموو دەرئەنجامەکەن ئەرێنی دێت بەدوایی دابەشکارەکە وە (r) ئە ژمارەیە لە دەرئەنجامە ڕاستەکە ئەرێنیەکەن جیادەکرێتەوە بە ژمارەیەک لە هەموو نمرەکانی نمونە پەیوەندی دارەکان تایبەت بە ڕێگا ی پێشکراو،ئەدایی ڕێگای ئە تۆێژینەوەیە بەراورد کارە بە تۆێژینەوەکانی تر. وە دەرئەنجامەکان دەری ئەخەن کە سیستەمی پێشنیارکراو پێشکەش کراو لە تۆێژینەوەیەدا دەتوانرێت حسابکات بەشێوازێکی زۆرگونجاو لەڕویی جۆریەتی و وردبینی و خێرایی.

1. Introduction:

Video surveillance and tracking play a dominant role in today's era for monitoring suspicious activity and anomaly detection (like fires or unattended objects in shopping malls and railway stations). Since the data generated in such applications are huge in volume, it is absolutely necessary to implement as Artificial intelligent.
(AI) based automatic surveillance systems to such solutions. These are capable of detecting multiple objects and tracking anomalous events, making them highly sought-after solutions. Within this scope, the term "fully automated system" refers to a combination of smart algorithms that run in real-time, and effective digital perception. There is no human intervention in fully automated systems, where the aforementioned smart algorithms perform low level tasks (like motion-detection and tracking) and high level tasks (like object and anomalous activity detection).

It is, a daunting task to detect, classify, and track multiple objects in a fully-automated environment, especially considering the wide range of applications. These applications include security measures, surveillance systems, traffic control mechanisms, and human-computer interaction. Even with all the software and hardware developments, real time tracking still represents a challenge in its own regard and still is an active research subject to its numerous issues like object occlusions, possible random or irrelevant movements, varying background complexity, and different lighting conditions.

In their paper issued in 2015 [1], Hakan Bilen, Marco Pedersol and Tinne Tuytelaars discuss using weakly supervised multi-object detection is challenging process, when the training the function in voles learning at the same time, the model appearance and the object position in all picture and provide for fixing the error is to consider the location of each object is as a hidden variable and decrease the loss produced by such hidden variable during learning [1]

Marc van et al proposed a method named Vibe that estimates the background from the moving images. Overall, video surveillance has been reduced to a three-step process for numerous applications: (i) motion detection, (ii) object detection, and (iii) tracking [2]. It covers stochastic replacement, spatial diffusion, and no chronological handling. Paula viola and Michael Jones. Proposed a Viola-Jones object detection framework to provide competitively object detection rates, it was trained to detect a variety of object class [3].

Traditional moving object detection is performed using SIFT [4], Kalman filter [5] and Background subtraction. [6] Newell, J. Li, X. Liang, S. Shen. Proposed a method for supervising convolutional neural networks (CNN) for the task of detection and grouping of object and solution vision problems can be put in this manner, including multi-person pose estimation, instance segmentation, and multi-target tracking and used in human pose estimation and camera frames and Infrared

The Scientific Journal of Cihan University – Slemani **PP: 18-32**
Volume (4), Issue (1), June 2020
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

frames. [7] SouYoung Jin, Aruni Roy-Chowdhury, proposed a technique dependent on the affiliation procedure by the neural system. [8]

We propose a novel method for multi object detection with modified K-Means and Deep neural networks. Further deep learning provides a better classification algorithm for the Moving objects. This method is suitable and accurate for fixed cameras. We validate the effectiveness of the proposed system with other methods in terms of miss rate and accuracy. In this paper, the research is presented as the following:

Section 2 introduces the suggested method,

Section 3 discusses the results obtained from the experiments performed by the method, and Section 4 summarizes the conclusions.

The contribution of this paper is as follows.

1. Firstly, how to detect multi object movements. Then, we determine what kind of objects and its class using a deep learning approach.
2. Secondly, we apply our proposed algorithm to detect all kinds of objects and cutting edge deep learning and high speed.

The entered video setting is apportioned into pint-sized masses. A grouping decision is prepared by whether the mass refers to the variations or not included in each tablet. To conclude, the inconstant structures from the appearance tablets are removed. Then and there the pictures tablets are categorized into two classifications as a focal point and circumstantial grounded in this perspective. The circumstantial ground prototype and the sub-divider need to be suitable while the contextual ground is time-changing. The contextual adjustment is hazardous. For instance, if an individual sits for a lengthy period or snoozes in the place the suitable contextual archetypal will deliberate the individual as a part of the contextual.

To elucidate this delinquent, the updated understanding acquired from entity tracing with the inferior prospective organization is united. Therefore, it can be utilized for updating background. The forefront possibly will still encompass substances, so as to resolve this issue; a resolution progression will be advanced depended on an adjusted k-average set of rules to disconnect these substances from the contextual perspective. In the next units, the recommended background deduction structure will be investigated in detail. [1]

DOI: http://dx.doi.org/10.25098/4.1.23

The Scientific Journal of Cihan University – Slemani **PP: 18-32**
Volume (4), Issue (1), June 2020
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

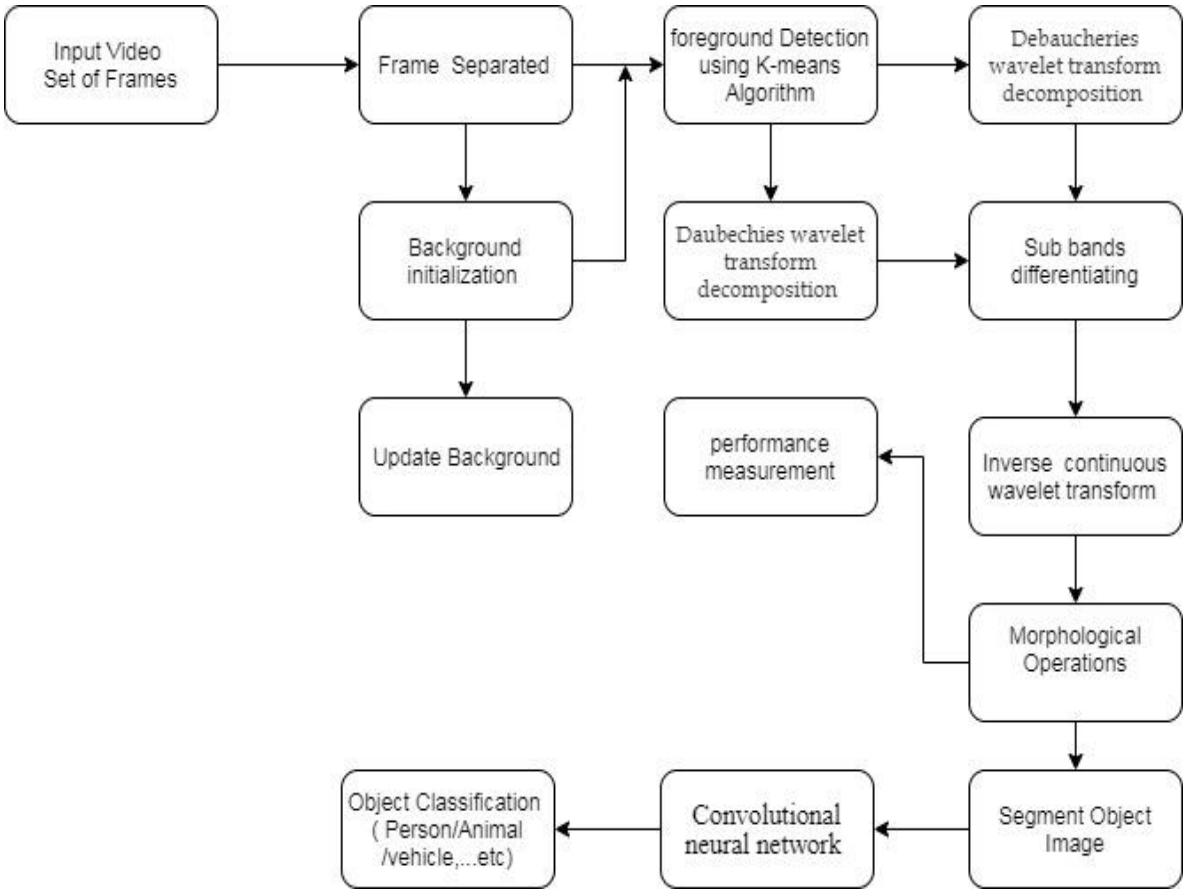


Figure 1. Data flow diagram of the proposed method

2. Proposed Method

Figure 1, represents the method suggested in this paper, whereas the proposed Convolutional neural network (CNN) architecture is described in Table 1.

2.1 Foreground detection using the "modified k-means algorithm":

Enhanced K-means (S, k), target, object, as proved in a movie series

$$S = \{x1, x2, x3, .. xn\} \quad (1)$$

Generate initial centroids as per the below steps

In this study, the Euclidian distances are used. In a given situation the distance between two vectors (for example v1 as $X = (X_1, X_2, \dots, X_n)$ and v2 as $Y = (Y_1, Y_2, \dots, Y_n)$ is expressed with the following: [7]

$$\boldsymbol{d(X,Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}} \quad (2)$$

DOI: http://dx.doi.org/10.25098/4.1.23

The Scientific Journal of Cihan University – Slemani **PP: 18-32**
Volume (4), Issue (1), June 2020
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

The below formula is used to describe the distances between data points (X and Y in this example)

$$\boldsymbol{d(X,V) = min(d(X,Y), Y\ EV)} \quad (3)$$

Let U be a data-point set, a dataset containing n objects, In the case where the number of data points is n within the population of U, which is the target for partitioning into k classes, they are set to 1 ($m = 1$), after which the following algorithm is employed:

For each data point in U, the distance to other data points; get the data points with the shortest distance between them, create a new data point set called Am ($I < m < k$) consisting of the two obtained data points, and remove these two data points from the population U.

the closest data point still in the U to the Am, add it to the Am and remove it from U.

Keep repeating step (2) until Am has I elements (data-points) where K clusters and j groups can be obtained from. Then;

Compute

$$\boldsymbol{(a\ X\ n/k)/2\ (0 < a \leq 1)}; \quad (4)$$

In case $m < k$, m will equal $m + 1$; and in such a case, get two different data-point pairs which have the shortest distance amongst the data points still in U, create another data-point set called Am, remove these two new data points from U, and repeat return to (2).

Once no more data points remain in U, get the sum of vectors for each of the Am $I < m < k$), then divide the sum by data-point count in Am, at this point, every single data-point set will return a vector, each of which can be selected as the initial centroids.

Once the centroid is calculated, repeat the process, but this time with a standard k-means algorithm, beginning from step 2. At this point, the value given to the a should be different with regards to the different types of data. In some cases, when the a is given too small value, it is possible that only centroids that are in the same and that have similar data-points will be obtained. On the other hand, if the a is given too big of value, then it is not possible centroids will be obtained from regions that contain similar data-point values. The tests performed as part of this paper have revealed that a value of 0.75 returns optimal clustering.

Input: The cluster count

$$\boldsymbol{K'(K' > K)}$$

In addition to a dataset that contains n objects (Xi).

Output: A set consisting of k number of clusters (Cj), which minimize the squared-error criterion

DOI: http://dx.doi.org/10.25098/4.1.23

The Scientific Journal of Cihan University – Slemani **PP: 18-32**
Volume (4), Issue (1), June 2020
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Begin

Multiple sub-samples {S1, S2, . . ., Sj};

$$for\ m = 1\ to\ j\ do$$

K-mean (Sm, K'); ***//executing k-means,***

Table. 1 Proposed Convolutional neural network (CNN) Architecture

S.No	Name	Type	Activations
1	Image input-227x227x3 image with 'zerocenter normalization	Image Input	227x227x3
2	conv1-96 11x11x3 convolution with stride [4 4] and padding [0 0 0 0]	Convolution	55x55x96
3	relu 1 –ReLU	ReLU	55x55x96
4	norm 1-cross channel normalization with 5 channels per element	Cross Channel Normalization	55x55x96
6	conv2-256 5x5x48 convolution with stride [1 1] and padding [2 2 2 2]	Convolution	27x27x96
7	relu2 –ReLU	ReLU	27x27x96
8	norm2-Cross channel normalization with 5 channels per element	Cross Channel Normalization	27x27x96
9	conv3-384 3x3x256 convolutions with a stride [1 1] and padding [1 1 1 1]	Convolution	13x13x384
10	relu3-ReLU	ReLU	13x13x384
11	conv4-384 3x3x192 convolutions with a stride [1 1] and padding [1 1 1 1]	Convolution	13x13x384
12	conv3-256 3x3x192 convolutions with stride [1 1] and padding [1 1 1 1]	Convolution	13x13x256
13	pool5-3x3 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	6x6x256
14	fc6-4096 fully connected layer	Fully Connected	1x1x4096
15	relu6-ReLU	ReLU	1x1x4096
16	fc7 -4096 fully connected layer	Fully Connected	1x1x4096
17	relu7-ReLU	ReLU	1x1x4096
18	special_2-64 fully connected layer	Fully Connected	1x1x64
19	Relu-ReLU	ReLU	1x1x64
20	fc8_2 -5 fully connected layer	Fully Connected	1x1x5
21	Softmax –Soax	Softmax	1x1x5
22	classoutput –crossentropyex	Classification Output	-

The Scientific Journal of Cihan University – Slemani **PP: 18-32**
Volume (4), Issue (1), June 2020
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

$$J_c(m) = \sum_{j=1}^{K} \sum_{X_i \in C_j} |X_i - Z_j|^2 \qquad (5)$$

Get the min {J} to act as the refined initial points

$Z_j, j \in [1, k']$ (6)

K-means(S, K');

//Execute k-means algorithm once more, this time with the chosen initials

//Produce K' methods.

Repeat

Combine two nearby clusters into a single one, and recalculate the new centre after the merger. Continue until the clusters counts fall down to k

//Merge

$(K' \rightarrow K)$

End

This suggested algorithm was found to work exceptionally well in samples with small data set counts, where a significantly lower number of iterations are needed. Large datasets have any significant initial sampling cost, and the main costs occur in clusters of the whole dataset for the initial centres when creating the K* clusters and **(O (nd))** when merging the K' clusters to create K Clusters **O (nd (k'-k))** this means that the total computational complexity of the suggested algorithm is **O (ndk').**

The total range of clusters K and a dataset containing n objects (Xi) a set of K clusters Cj that minimizes the squared error criterion.

$Begin$

$m = 1$

$Initialize\ K\ Prototypes\ Z_j, j \in [1, K]$

//Arbitrarily choose K Objects as centres

Repeat

$for\ i = 1\ to\ n\ do$

Begin

$for\ j = 1\ to\ K\ do$

$compute\ D(X_i, Z_j) = |X_i - Z_j|;$

If $D(X_i, Z_j) = min\{D(X_i, Z_j)\}$

DOI: http://dx.doi.org/10.25098/4.1.23

The Scientific Journal of Cihan University – Slemani **PP: 18-32**
Volume (4), Issue (1), June 2020
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Then $X_i \in C_j$

End

If $m = 1$ then

$$J_c(m) = \sum_{j=1}^{K} \sum_{X_i \in C_j} |X_i - Z_j|^2 \qquad (7)$$

$$m = m + 1 \qquad (8)$$

$for\ j = 1\ to\ K\ do$

$$Z_j = \frac{1}{n_j} \sum_{i=1}^{nj} X_i^j J_c \qquad (9)$$

$Calculate\ the\ mean\ value\ for\ each\ clusters$

$$J_c(m) = \sum_{j=1}^{K} \sum_{X_i \in C_j} |X_i - Z_j|^2 \qquad (10)$$

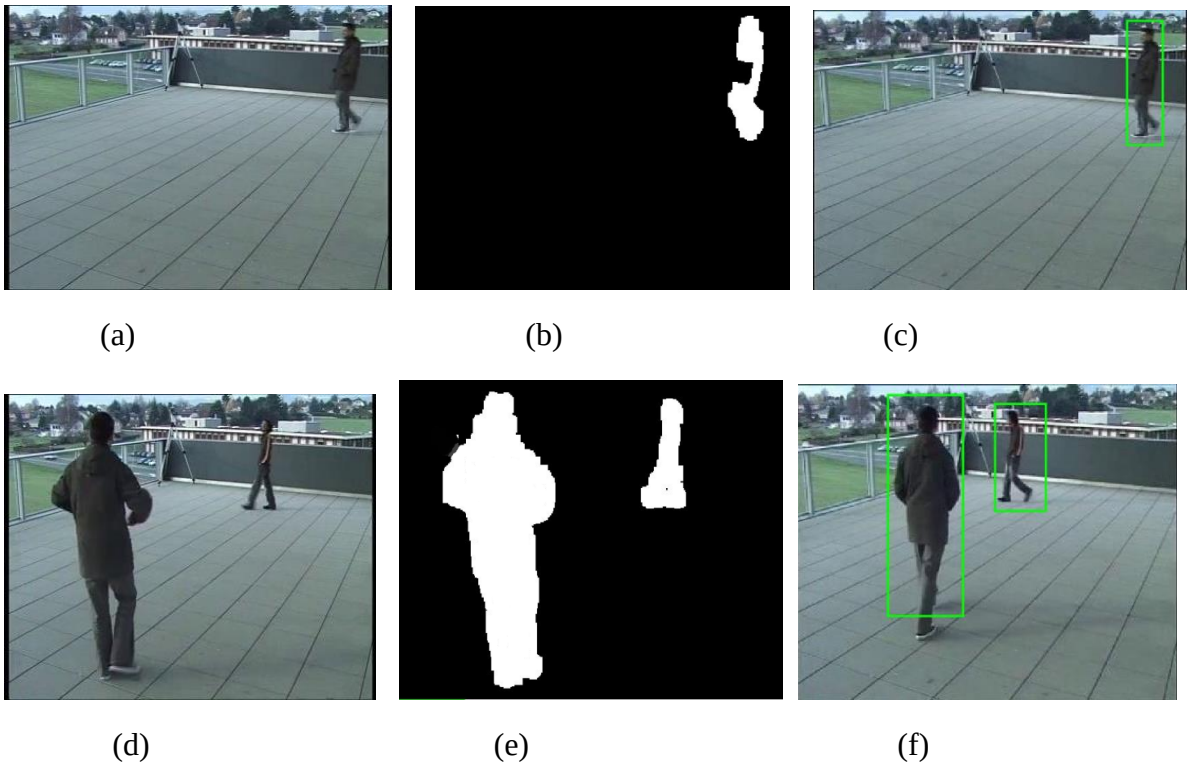
2.2 Proposed Convolutional neural network (CNN) Architecture

A convolutional neural network is an approach to deep learning where different network layers (like input, convolution, pooling (max and average), and linear rectified units are brought together in unison. The total number of layers involved can be set to the amount required, based on the size of the input. Not all of the layers in the network have to be used at any given point either. The idea behind the CNN is that it is capable of organizing itself without any required intervention, and the depth of the network increases its learning capacity [4] [6]. That being said, the increased network depth also comes with the increased computation time requirements, and despite the advancements in hardware and software development, this is still a limiting factor.

In the present study, the layers that were involved in the network were carefully selected to achieve maximum output with the minimum layer count. This was made easier by the fact that CNN requires no pre-processing or handcrafted feature extraction. In CNN, required feathers are extracted where the raw pixel data from the image is mapped inconsequentially, which are then classified by the layers set forth. It is a matter of design to set the network parameters to achieve

The Scientific Journal of Cihan University – Slemani **PP: 18-32**
Volume (4), Issue (1), June 2020
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

high performance for the classification stage. The following section provides information regarding each CNN unit.



(a) (b) (c)

(d) (e) (f)

Figure .2 Proposed Convolutional neural network (CNN) Architecture

The Image (a) and (d) frame from complementary metal-oxide semiconductor (CMOS) camera. (b) And (e) foreground detected by K-means algorithm. (c) And (f) frame with pedestrian obtained from the complementary metal-oxide semiconductor (CMOS).

2.2.1 Layer 1: image input

The first layer of the framework is the image input layer and is mandatory for all networks. It is where the main image is input to the network, and both 2D and 3D objects can be used as input. The only initialization required is regarding the dimensions of the image at this layer [8].

2.2.2 Layer 2: Convolutio

[9] Convolution process is complete, a feature map will be obtained, which will represent the input for the following layer.

2.2.3 Layer 3: Batch normalization

This particular layer is all about normalizing the data and thereby increasing both the learning capabilities of the system and the speed of his algorithm. Once normalized, the data can be freely transferred between various layers as a standardized sample, which increases the learning rate and drastically improves learning speed.

DOI: http://dx.doi.org/10.25098/4.1.23

The Scientific Journal of Cihan University – Slemani **PP: 18-32**
Volume (4), Issue (1), June 2020
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

2.2.4 Rectified Linear Unit (ReLu)

This unit is a sub-unit of the convolution layer where the operation known as the threading is executed for each of the elements present in the input. This operation sets any value below zero to zero in order to ensure data redundancy while preserving the important elements of the image. This operation does not change the size of the layer.

2.2.5 Layer 4: Max pooling

In this layer, using a stride value selected by the user, the "max pooling" operation is performed on each feature map in order to reduce their size. The max value is represented over the window, and when applied, the contents of the window are replaced with that number. This layer reduces the size of the output it produces; compared to the input it receives [9].

2.2.6 Layer 5: Fully connected layer

In the neural network, this layer receives the output of the previous layer and connects it to every single neuron

Present in this layer. The output provided by this layer represents the class count for the classification process [9].

2.2.7 Layer 6: Soft-max layer

The soft-max layer ensures the input data is shrunk to the range between 0 and 1, which preserves the data but eliminates certain outliers within the image. Furthermore, the classification layer is where the "loss function" is present, besides the expected output label and size values. Figure 2 displays the general outlines of the CNN architecture developed for the study.

3. Experimental Results:

With the object tracking method suggested presently, it is possible to evaluate the various illumination conditions for scenes with multiple objects and potential occlusions. For the experiments performed as part of the study, the commonly used Wallflower dataset was used, besides certain video sequences from PETS 2001 dataset where people move through the image, a lobby video from I2R dataset, and a video taken change detection.net where various pedestrians are walking, besides a highway. These tests were performed to evaluate the differences between the suggested algorithm and the Frame differencing, Temporal differencing, Optical Flow [3].

The Scientific Journal of Cihan University – Slemani **PP: 18-32**
Volume (4), Issue (1), June 2020
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Table 2: Comparative analysis of the proposed method with other methods

Video Sequence	Algorithms	Accuracy	F1-Score	Sensitivity	Specificity
Wallflower	Proposed	0.92	0.89	0.80	0.79
	SIFT	0.83	0.85	0.75	0.73
	Optical Flow	0.76	0.81	0.68	0.68
	Kalman Filter	0.8	0.78	0.62	0.60
PETS2001	Proposed	0.86	0.90	0.87	0.86
	SIFT	0.85	0.86	0.76	0.75
	Optical Flow	0.71	0.77	0.63	0.61
	Kalman Filter	0.76	0.81	0.68	0.66
Time of Day	Proposed	0.85	0.90	0.77	0.75
	SIFT	0.8	0.78	0.65	0.63
	Optical Flow	0.7	0.73	0.57	0.56
	Kalman Filter	0.74	0.79	0.65	0.64
Pedestrians	Proposed	0.95	0.94	0.89	0.88
	SIFT	0.94	0.91	0.84	0.83
	Optical Flow	0.71	0.77	0.63	0.62
	Kalman Filter	0.74	0.79	0.65	0.65

Each pixel in a contextual detraction technique's Grouping has been identified to be: correct positive or true positive abbreviated as (TN) for a properly categorized focal point pixel, incorrect affirmative named (FP) for a contextual perspective pixel that was erroneously subdivided as forefront, factual negative (TN) for an appropriately sub-classed circumstantial pixel, and incorrect negative (FN) for a front pixel that was mistakenly subdivided as contextual perspective or namely background. Through the computation of TP, TN, FP, and FN the dissimilar approaches may be appraised. Next, formulas 11,12, 13 and 14 have been applied to estimate the Specificity, Sensitivity, Accuracy and F1-Score. Specificity, Sensitivity, Accuracy and F1-Score were delineated in equations 11,12, 13 and 14 consecutively.

$$Specificity = \frac{Tn}{Fp+Tn} \qquad (11)$$

DOI: http://dx.doi.org/10.25098/4.1.23

The Scientific Journal of Cihan University – Slemani **PP: 18-32**
Volume (4), Issue (1), June 2020
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

It has been observed in the video structure sensitivity and specificity can be considered as the statistical processes to measure the performance of a binary classification test. The amount of authentic positives which are accurately recognized is quantified by sensitivity.

$$Recall\ or\ Sensitivity\ = \frac{Tp}{Tp+Fn} \qquad (12)$$

The percentage of negatives which are acceptably acknowledged is standardized by specificity.

$$Accuracy = \frac{Tp+Tn}{Tp+Tn+Fp+Fn} \qquad (13)$$

The two procedures are meticulously connected to the theories of category I and category II inaccuracies. The research was directed in this technique for detraction of contextual perspective through applying altered approaches

Furthermore, the findings from three experimentations of suggested adapted K-means technique of background detraction are illustrated in Figure 2 and Table 1, which demonstrate the numerical outcomes of countless structures of the projected scheme as compared to other background subtraction.

$$F1 = \frac{2(Recall)(Precision)}{Recall + Precision} \qquad (14)$$

In recommending K-means procedure more acceptable results are achieved and the four features of Accuracy, F1-Score, Sensitivity and Specificity in background deduction are more satisfied in comparison with the other background algorithm. [14]

The parametric analysis of the suggested approach is summarized in Table 2. The test results show that the suggested method has an average accuracy of over 98% of multiple-people tracking in a frame (Table 2)

The suggested method was compared with other methods proposed by SIFT[11], Kalman filter[12], and optical flow[13].

This proposes method is more suitable for detection objects in each frame and has a lower false positive per each frame to detect objects, therefore the performance of the proposed technique assessed based on the comparison of our method to the other conventional method as shown in figure 3.

The Scientific Journal of Cihan University – Slemani **PP: 18-32**
Volume (4), Issue (1), June 2020
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

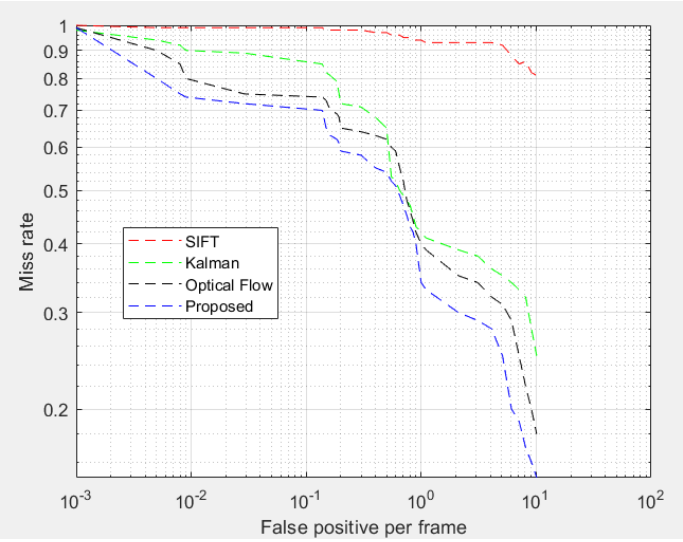


Figure 3. Performance measurement of the proposed method with other three existing methods

The proposed method is more suitable for hardware implementation and has lower resource utilization, thereby employing parallelism in algorithms, along with a high frame rate. The computation time in frames per second (fps) for various algorithms is listed in Table 3. The proposed method for same image resolution (320-240) has a higher frame rate (fps) that is suitable for real-time applications.

Table 3 **Comparative resolution and frames**/second, the proposed method with other methods

S.I.NO	***Method***	***Resolution***	***Frames/Second(fps)***
1	Proposed Method	320 × 240	62
2	SIFT	320 ×240	56
3	Kalman filter	320 ×240	52
4	Optical flow	320 ×240	52

5. Conclusion:

We present an integrated approach based on K-Means and deep learning algorithm for moving multi-object detection. The experiments demonstrated that the system has more accurate in detecting multi objects compared to other widely used methods. The suggested method has been evaluated on various datasets and the results clearly indicate that it is better in terms of accuracy, precision, and recall. In the future, the same algorithm can be improved even further to adapt to moving cameras. To elucidate this problem, the updated knowledge acquired from entity tracking with a lower-ranking feature-based organization is fused. Therefore, it can be utilized for updating background. The forefront may also still contain objects, so as to resolve this issue; a decision process will be advanced depended on a modified k-means algorithm to disconnect these objects from the background. The proposed background subtraction scheme has been investigated in details.

DOI: http://dx.doi.org/10.25098/4.1.23

References:

[1] Hakan Bilen, Marco Pedersoli and Tinne Tuytelaars. "Weakly Supervised Object Detection with Convex Clustering." The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 1081-1089.

[2] Barnich, Olivier; Van Droogenbroeck, Marc (2009). "ViBe: A powerful random technique to estimate the background in video sequences". 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 945–948.

[3] P. Viola; M. Jones, Rapid object detection using a boosted cascade of simple features, Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001.

[4] M. C. Roh, and J. Y. Lee, "Refining faster-RCNN for accurate object detection," IAPR Int. Conference. MVA, vol.15, pp.514–517, 2017.

[5] Karthikeyan PANJAPPAGOUNDER RAJAMANICKAM and Sakthivel PERIYASAMY, "Entropy Based Illumination-Invariant Foreground Detection." IEICE TRANS. INF. & SYST VOL.E102–D, NO.7 (JULY 2019): pp. 1436.

[6] Niranjil Kumar, Sureshkumar, "Background Subtraction Based on Threshold Detection using Modified K_Means Algorithm," Proceedings of the 2013 International Conference on Pattern Recognition, Information and Mobile Engineering, p. 381, 21 February 2013.

[7] J. Li, X. Liang, S. Shen, T. Xu and S. Yan, "Scale-aware fast R_CNN for Pedestrian Detection," CORR, vol. abs/ 1510.08160, p. 25, June 2016.

[8] SouYoung Jin, Aruni Roy-Chowdhury, Huaizu Jiang, AshishSingh, Aditya Prasad, Deep Chakraborty, and Erik Learned-Miller, "Unsupervised Hard Example Mining from Video for Improved Object Detection," European Conference on Computer Vision (ECCV), 2018.

[9] P. K. Abhishek Kumar Chauhan, "Moving Object Tracking using Gaussian Mixture Model and Optical Flow," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 4, April 2013.

[10] L. Zhang, L. Lin, X. Liang, and K. He, "Is Faster R-CNN Ding Well for Pedestrian Detection?" IEEE International Conference on Computer Vision, Vols. abs/ 1607.07032,27, 27 Jul 2016.

[11] Zhao, L., Thorpe, C.E., "Stereo- and neural network-based pedestrian detection," IEEE Transaction on Intelligent Transportation Systems 1, vol. 2, Sep 2000.

[12] Yang C. Li B. and Xu G., "Particle filter based multi-pedestrian tracking by HOG and HOF,"

IEEE Int. Conf. Inform. Sci. and Tech., 2014.

[13] W. Lan, J. Dang, Y. Wang, and S. Wang, "Pedestrian Detection Based on YOLO network model," IEEE Int. Conf. Mechatronics and Automation, 2018.

[14] R. Girshick, "Fast-RCNN," Proceeding of the 2015 IEEE International Conference on Computer Vision, pp. 1440-1448, December 2015.

DOI: http://dx.doi.org/10.25098/4.1.23