Neural Network based Approach for Recognition Human Motion using Stationary Camera

Rachana V. Modi
PG Student
Nirma University
Ahmedabad,India

Tejas B. Mehta
Assistant professor
Nirma University
Ahmedabad, India

ABSTRACT

Video surveillance is currently one of the most active research topics in the computer vision community. During motion, the surveillance system can detect moving objects and identify them as humans, animals, vehicles. This strong interest is driven by a wide spectrum of promising applications in surveillance system such as Military security, Public and commercial security, etc. The model includes detection, feature extraction and recognition of people from image sequences involving humans. In proposed system frame differencing and Neural Network is used for moving object detection and recognition of human motion respectively. Experimental results show that human motion can be correctly classified.

General Terms

Motion detection and recognition

Keywords

Human Motion Recognition, Neural Network.

1. INTRODUCTION

Recognition of Human Motion (RHM) system can be applied for video surveillance system which is currently useful for Military security (such as Patrolling national borders and Measuring flow of refugees), Public and commercial security (such as Monitoring of banks, department stores, airports, museums, stations, private properties and parking lots for crime prevention and detection). The task to classify and identify objects in the video is surprisingly difficult for human operator. Recognition of Human Motion (RHM) is automated video surveillance system which includes detection and recognition of people [1].

The previous researches on recognition of moving object have been presented. A view-based method for recognizing 3D objects from 2D images [2] is exploited. An aspect graph structure is implemented to generate aspects using a notion of similarity between views. Another approach was shape-based moving object recognition [3] in which a statistical motion detection algorithm is used to identify moving objects. Fourier descriptors [4] are computed as feature vectors to describe the shape of object. A case based classification algorithm is used for recognition of object. Next, Motion-based recognition method used for extract motion symmetry of moving objects for gait recognition [5]. For classification k-Nearest Neighbor rule [6] is adopted. Another approach was specific feature based method, Recurrent Motion Image (RMI) is a specific feature vector. It is proposed to estimate repetitive motion behavior of moving objects. Thus, moving objects can be classified as single person, group of persons or vehicle based on their RMI classifier. Another is a hybrid classification system [7] can be used to recognize moving objects based on motion and appearance features simultaneously. Support vector machine (SVM) as a classifier [8] is used for classification of moving object.

Section 1 describes general introduction of system. Section 2 presents introduction of neural network. Section 3 briefly presents proposed model for Recognition of Human Motion (RHM) system. Section 4 describes gives experimental results of proposed system. Section 5 discusses constraints and future work while conclusion is mentioned in section 6.
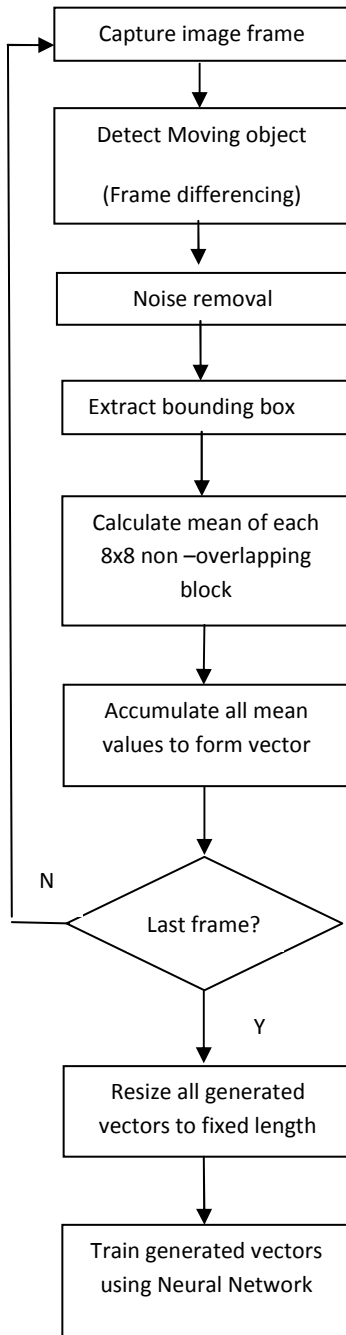
2. INTRODUCTION TO NEURAL NETWORK

Neural network works as the human brain. A neural network is an interconnected network of many artificial neurons. These artificial neurons are objects used to simulate the neurons in the human brain. Neural networks are used widely in machine learning applications. They are usually used for classification problems [9].

The recent vast research activities in neural classification have established that neural networks are a promising alternative to various classification methods. It is very important to understand how a neural network works. Neural network is a set of connected input/output units, where each connection has a weight associated with it. Neural network learns by adjusting the weights so as to be able to correctly classify the training data and hence, after testing phase, to classify unknown data.

The weighted outputs of these units are fed into hidden layer. The weighted outputs of the last hidden layer are inputs to units making up the output layer. The units in the hidden layers and output layer are sometimes referred to as output units. A network containing two hidden layers is called a three-layer neural network.

3. PROPOSED SYSTEM

Proposed system is divided into two categories, namely preprocessing and employing neural network to classify the features Extracted. The overall functioning of proposed system is represented graphically as shown in figure 1.

```
            ┌─────────────────────────┐
            │  Capture image frame    │
            └─────────────────────────┘
                        │
            ┌─────────────────────────┐
            │  Detect Moving object   │
            │  (Frame differencing)   │
            └─────────────────────────┘
                        │
            ┌─────────────────────────┐
            │     Noise removal       │
            └─────────────────────────┘
                        │
            ┌─────────────────────────┐
            │  Extract bounding box   │
            └─────────────────────────┘
                        │
            ┌─────────────────────────┐
            │  Calculate mean of each │
            │  8x8 non –overlapping   │
            │         block           │
            └─────────────────────────┘
                        │
            ┌─────────────────────────┐
            │  Accumulate all mean    │
            │  values to form vector  │
            └─────────────────────────┘
                        │
                   ◇ Last frame? ◇  ── N
                        │ Y
            ┌─────────────────────────┐
            │  Resize all generated   │
            │  vectors to fixed length│
            └─────────────────────────┘
                        │
            ┌─────────────────────────┐
            │  Train generated vectors│
            │  using Neural Network   │
            └─────────────────────────┘
```

Fig 1: Flow of proposed system

3.1 Preprocessing

Object detection is the first step of the pre level processing which detects foreground objects in the sequence of frames. There are some methods all ready exist to detect the foreground objects such as simple background subtraction, frame differencing etc. In Simple Background subtraction method identifies moving objects by selecting the parts of the image which differ significantly from a reference image, result of that

shown in figure 5.1(c). Another method is Frame differencing in which it takes the difference between two consecutive frames and identifies moving objects. By comparing two methods system get better result in frame differencing. For detecting the foreground region system use Frame Differencing Background Subtraction Method, which really works in outdoor environment as well as indoor environment [10]. The Frame differencing identifies moving objects from the consecutive frames that differs significantly from the previous frame.

The method basically employs the image subtraction operator. The image subtraction operator [11] takes two images as input and produces as output a third image whose pixel values are simply those of the first image minus the corresponding pixel values from the previous image. The subtraction of two images is performed straightforwardly in a single pass.

The Algorithmic steps are given as below.

1. Let the current frame be I_{curr}.

2. Retrieve the previous frame I_{prev} from an array of image variables which are used for temporary storage of frames. The array is programmed to behave as a queue with only three elements at any given point of execution.

3. Perform frame differencing operation on the current frame I_{curr} and previous frame I_{prev} where the resultant image is represented as output Image I_{out} as given below.

$$I_{out} = I_{curr} - I_{prev}$$

4. Perform binary threshold operation on resulting image. If value of pixels in input image is greater than value of threshold then the pixel value is set to 1 while remaining pixels are set to 0. Based on observations, threshold value is set to 0.2.

This approach gives fairly well results which shown in figure 5.1(d). After pre-level processing stage we know object is under motion and it is really helpful for further processing

3.2 Employing Network for pattern recognition

Classification is one of the most active research and application areas of neural networks. Neural network is used as classifier for recognizing human. First step is to extract the features for training with neural network. Feature Extraction plays a major role to detect the moving objects in sequence of frames for dimensionality reduction of data, so for that bounding box is used to visualize only moving object from the whole frame. To generate bounding box, pixel values are found out from the first hit of the intensity values from top, bottom, left and right. By using this dimension values a rectangular bounding box is plotted within the limits of the values produced which explains a clear view on bounding box [12]. Using this process we get bounding box around the moving object which is shown in figure 3.

Bounding boxed image still contain too many pixels which in turn forms the larger input vector for neural network training. In order to reduce input vector length, boxed image is divided into 8x8 non –overlapping blocks. Mean value of each block is calculated. All such mean values form a reduced length vector.

Figure 4 shows the image blocks replaced by its mean values. Such vector represents a unique pattern for the various types of objects. It is also observed from figure 3 that we have different sized image retrieved after forming bounding box. This results into variable length vectors. In order to train our dataset using neural network, all vectors need to be converted into fixed length vectors. This process is carried out for various other frames containing the different type of objects in motion. Inputs of image vector are fed simultaneously into the input layer of neural network.

4. EXPERRIMENTAL RESULTS

The videos used for this study were taken from the side view of the people and vehicles. Color videos were taken with a codec and Sony digital camera with on uncompressed AVI formats. All video were taken at a 320 x 240 pixel resolution at 25 frames / second. For all experiments in this work, the algorithms were implemented in MATLAB 7.8 (R2009a). Sample foreground region detection is shown in figure 2. This Observation suggests that frame differencing gets better result than normal background subtraction. After detecting foreground object, its corresponding bounding box is extracted as shown in figure 3. It is also observed that size of bounding boxed image varies from one frame to another. As discussed earlier values are generalized by taking the average of values for each block of bounding box. The resultant image is shown in figure 4.

(a) **(b)** **(c)**
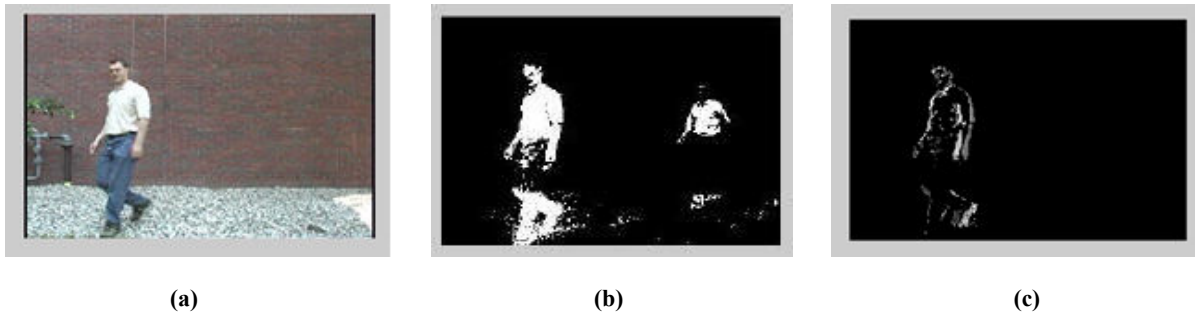
Fig 2: (a) Original image (b) Background subtraction (c) Frame differencing

Frame 1 Frame 5 Frame 25 Frame 29 Frame 30

Fig 3: Extracted Bounding box from various image frames

Fig 4: Averaging values for each block

Fig 5: Walking pattern

To generate more accurate results for human motion identification, half bottom part of the frames are used. The half part of the frame shows the leg pattern of human. When human is in walking condition, then patterns of legs are shown in figure 5. The Pattern of human walking is demonstrating repetitive motion. After some time interval it will be same. Other side we show the pattern of car is constantly same there is no change in pattern. There is an optional case, when car is

turned. Using time interval, three different movements of walking pattern is considered. These three patterns are merged to form a vector which describes the walking pattern of human. This vector can further be given to any classifier. This process is carried out for various other frames containing the different type of objects in motion. The inputs of image vector are fed simultaneously into the input layer of neural network [13].

Table 1 Classification result for case 1

Phase	Total Samples	Correctly Classified Sample		Miss-Classified Sample	MSE	%E
		Human	Not Human			
Training Phase	316	250	18	48	$1.15435e^{-1}$	$15.18987e^{-0}$
Validation Phase	68	52	1	15	$1.53194e^{-1}$	$22.05882e^{-0}$
Testing Phase	68	60	0	8	$1.03171e^{-1}$	$11.76470e^{-0}$

Table 2 Classification result for case 2

Phase	Total Samples	Correctly Classified Sample		Miss-Classified Sample	MSE	%E
		Human	Not Human			
Training Phase	880	637	144	99	$8.89052e^{-2}$	$11.25000e^{-0}$
Validation Phase	126	89	18	19	$1.21109e^{-1}$	$15.07936e^{-0}$
Testing Phase	252	177	29	46	$1.39620e^{-1}$	$18.25396e^{-0}$

Table 3. Table captions should be placed above the table

Case 1			Case 2		
362	69	84.0%	903	152	85.6%
80.1%	15.3%	16.0%	71.8%	12.1%	14.4%
2	19	90.5%	12	191	94.1%
0.1%	4.2%	9.5%	1.0%	15.2%	5.9%
99.5%	21.6%	84.3%	98.7%	55.7%	87.0%
0.5%	78.4%	15.7%	1.3%	44.3%	13.0%

While experimenting we have considered two cases. In Case 1 shows the results when entire extracted bounding boxed image is taken as neural network input.. Case 2 shows the results obtained when three bottom half images are given to neural network input. Table 1 and Table 2 shows the proportion of the samples used in training, validation and testing phase along with total number of correctly classified and misclassified samples for case 1 and case 2 respectively. Confusion matrixes for both cases are shown in table 3.

5. CONSTRAINTS AND FUTURE SCOPE

Proposed system cannot work under moving background condition. Further preprocessing is required to separate foreground object under various conditions. The system does not handle occlusion condition. For future enhancement the accuracy need to be improved and to detect object under various conditions such as moving background, occlusion etc.

6. CONCLUSION

In this paper, proposed technique is developed for video surveillance system using stationary camera. We proposed an approach using frame difference as a preprocessing stage. From the results frame difference is better than normal background subtraction method. The results from the extracted and post processed frames show that classifier such as neural network can be applied for recognizing the human object in motion to some extent.

7. REFERENCES

[1] L. Wang, W. Hu, and T. Tan, 2003 "Recent developments in human motion analysis", Pattern Recognition, Vol. 36,No. 3, pp.585-601

[2] C.M. Cyr and B.B. Kimia, "A Similarity-Based Aspect-Graph Approach to 3D Object Recognition," Int'l J. Computer Vision, vol. 57, no. 1, 2004, pp. 5-22

[3] D. Toth and T. Aach, "Detection and Recognition of Moving Objects Using Statistical Motion Detection and Fourier Descriptors," Proc. 12th Int'l Conf. Image Analysis and Processing (ICIAP 03), 2003, pp. 430-435.

[4] J.C. Russ, The Image Processing Handbook, 5th ed., CRC Press, Boca Raton, 2006, p. 589.

[5] J.B. Hayfron-Acquah, M.S. Nixon, and J.N. Carter, "Recognizing Human and Animal Movement by Symmetry," Proc. IEEE Int. Conf. Image Processing (ICIP 01), 2001, pp. 290-293.

[6] J.C. Russ, The Image Processing Handbook, 5th ed., CRC Press, Boca Raton, 2006, p. 619

[7] Y. Bogomolov, G. Dror, S. Lapchev, E. Rivlin, and M. Rudzsky, "Classification of Moving Targets Based on Motion and Appearance," Proc. British Machine Vision Conf. (BMVC 03), 2003, pp. 429-438.

[8] D.A. Forsyth and J. Ponce, Computer Vision: A Modern Approach, Prentice Hall, 2002, p. 615.

[9] James A. Anderson , Introduction to Neural Network , PHI, 1999

[10] Jain, R., and H. H. Nagel, "On the analysis of accumulative difference pictures from image sequences of real world scenes," IEEE Trans. PAMI, vol. PAMI-1, pp. 206-214, Apr. 1979.

[11] Rafael C. Gonzalez, Richard E. Woods, "Digital Image processing", Second Edition. Prentice Hall, 2002

[12] Y. T. Zhou, V. Venkateswar, and R. Chellappa, "Edge detection and linear feature extraction using a 2-D random field model", IEEE Trans. Pattern Analysis, Intell,11(1):84-95, Mach.1989

[13] Ripley BD. Pattern recognition and neural networks. Cambridge: Cambridge University Press; 1996