Real Time Human Gesture Detection using Image Processing

Dr. R. Sathya
Assistant Professoe,
Dept. of IT,
Kongunadu College of Engineering and Technology,
Trichy.

Mrs. S. Sangeetha
Assistant Professoe,
Dept. of IT,
Kongunadu College of Engineering and Technology,
Trichy.

***Abstract*:- Human gesture detection in video is an important topic in computer vision applications such as automated surveillance. Naturalistic and intuitiveness of the hand gesture has been a great motivating factor for the researchers in the area of Human Computer Interaction (HCI) to put their efforts to research and develop the more promising means of interaction between human and computers. This paper presents a novel and efficient framework for traffic personal gesture detection based on *n*-frame cumulative frame difference. The experiments are carried out on the real time traffic personal action dataset using frame differencing.**

Keyword: Gesture Detection, Image Processing, Frame differencing.

1. INTRODUCTION:

Human gesture recognition is a challenging problem that has received considerable attention from the computer vision community in recent years. The research of moving traffic personal hand gesture recognition based on video is traffic in modern society. In a human body parts the hand is the most effective, general purpose interaction tool due to its smart functionality in communication. First investigations about this topic began in the seventies with pioneering studies accomplished by [1].

Most of the current research focus on human action recognition, human behavior analysis, hand action detection, gesture recognition, intelligent monitoring, Humana computer interaction, intelligent transportation, robot visual navigation, precision guidance systems, in addition to medical diagnosis, image compression, 3D reconstruction, video retrieval and other fields. In recent years, a large number of researchers have addressed this problem as evidenced by several survey papers [2, 3, 4, 5]. This research concentrates on traffic personal hand signals. Human gesture recognition for traffic control can be related used for human robot interaction. A human action is done normally with a number of successive actions, which gives an interpretation of the action carried out.

Traffic management on roadways is a challenging task which is increasingly being augmented with automated system. Traffic rules and regulation are devised to assure the smooth flowing of motor vehicles in the road. Moreover, traffic rules and regulation are not only for the driver of the vehicles but at the same times are meant for the pedestrians, cyclist, motor-cyclist and other road users. The proper knowledge of these rules can reduce the number of accident and thus can establish a healthy and organized traffic system in our country.

2. RELATED WORK:

Human gesture recognition is an active topic in computer vision technique. An object detection system generally contains two pivotal parts: feature based image representation, classification of features. In [6] discussed more datasets for human action and activity recognition. In [7] proposed a new tracking method that uses Three Temporal Difference (TTD) and the Gaussian Mixture Model (GMM) approach for object tracking. In [8] utilized a line based pose representation to recognize human actions on Weizmann and KTH datasets.

In [9] proposed representation that keeps most of the shape details and the gait temporal variations. In [10] proposed the gait energy image (GEI), which is the average image of a gait cycle to characterize human walking properties. Experimental results of both synthesized and real database testified that the frame difference energy image [FDEI] is a feasible gait representation. When the noise at different moments is uncorrelated and identically distributed, GEI was found to be less sensitive to silhouette noise in individual frames. The performance of GEI is notable, but this representation loses detailed information and does not contain temporal variation. The gait history image (GHI) [11] and gait moment image (GMI) [12] were developed based on GEI. In [13] trajectory gradients are computed and summarized an action is represented as a set of subspaces and a mean shape.

In [14] block based human model for real time monitoring of a large number of visual events and states related to human activity analysis, are used as components of a library to describe more complex activities in important areas such as surveillance. Activity recognition approach is proposed in [15] extracted motion information from the difference image based on Region of Interest (ROI) using 18-Dimensional features called Block Intensity Vector (BIV). The experiments are carried out on the KTH dataset using SVM classifier. In [16] local self-similarity (LSS) is a descriptor that capture locally internal geometric layout of self similarities within an image region while accounting for small local affine deformation. Most of the recognition system uses the data sets like KTH, Weizmann. Some other data sets were used by the action recognition system discussed in [17].

3. INDIAN TRAFFIC SIGNALS:

In a human traffic control environment, drivers must follow the directions given by the traffic police officer in the form of human body gestures. The traffic control commands are categorized into three types such as, stop all vehicles in every road direction, stop all vehicles in front of and behind the traffic police officer and stop all vehicles on the right of and behind the traffic police officer. Each traffic hand signal is a combination of the arms directions.

Twelve Indian traffic hand signal can be constructed from these control command types. The twelve traffic police hand signals listed as follows, to start one side vehicles, to stop vehicles coming from front, to stop vehicles approaching from back, to stop vehicles approaching simultaneously from front and back, to stop vehicles approaching simultaneously from right and left, to start vehicle approaching from left, to start vehicles coming from right, to change sign, to start one side vehicles, to start vehicles on T-point, to give VIP salute, to manage vehicles on T-point. There are two possible solutions to this recognition: active way or passive way. The passive way is to use body sensors to recognize the traffic police gestures. The active way is to use cameras on unmanned vehicles to recognize the traffic hand signals. This method is called vision based approaches.

4. PROPOSED APPROACH:

Real time traffic personal actions are used for experimental purpose. The video is processed at 25 frames per second. Smoothing is done by Gaussian convolution with a kernel of size 3 X 3 and variance sigma = 0.5. It is essential to preprocess all video sequences to remove noise for very well features extraction and classification. The video sequence is converted into frames in .jpg format.

4.1. Frame Differencing:

Motion information in a video sequence is extracted by pixel-wise differencing of consecutive frames. Fig. 1 shows the two consecutive frames and their motion information. Motion information T_k or difference image is calculated using

$$T_k(i,j) = \begin{cases} 1, & \text{if } D_k(i,j) > t; \\ 0, & \text{otherwise;} \end{cases} \quad \text{---------- (1)}$$

$$D_k(i,j) = | I_k(i,j) - I_{k+1}(i,j) |$$
$$1 \le i \le w,\ 1 \le j \le h \quad \text{---------- (2)}$$

Where $I_k(i,j)$ is the intensity value of the pixel (i,j) in the k^{th} frame, t is the threshold, w and h are the width and height of the image respectively. The value of $t=30$ is used in the experiments.

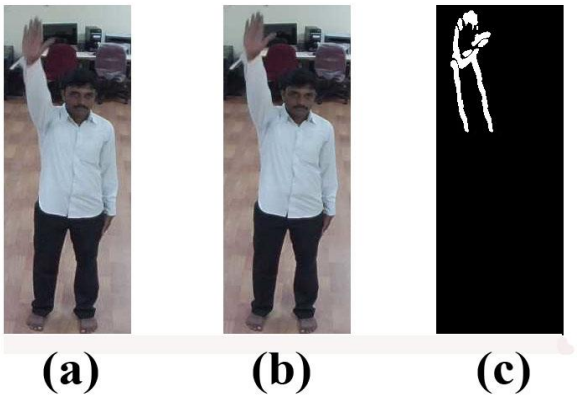

Fig. 1. (a), (b) Two consecutive frames. (c) Motion information of (a) and (b).

4.2. n – Frame Cumulative Differencing:

For identifying the region showing maximum intensity, n-frame cumulative differencing is applied, as seen in Fig. 2. Fig. 3 (a) shows 3-frame cumulative difference image. Fig. 3 (b) shows 4-frame cumulative difference image. Fig. 3 (c) shows 5-frame cumulative difference image. Fig. 3 (d) shows 7-frame cumulative difference image. Fig. 4 (e) shows 10-frame cumulative difference image.

$$Dk(x,y) = It(x,y) - It+1(x,y)$$
$$1 \le x \le w,\ 1 \le y \le h \quad \text{---------- (3)}$$

Where D_k is the difference image obtained by subtracting by two consecutive frames I_t and I_{t+1}. $I_{(x,y)}$ is the pixel (x,y), w and h are width and height of the image respectively.

Consecutive difference images are calculated as follows:

$$D_n(x,y) = I_p(x,y) - I_{p+1}(x,y)$$
$$D_{n+1}(x,y) = I_{p+I}(x,y) - I_{p+2}(x,y)$$
$$D_{n+2}(x,y) = I_{p+2}(x,y) - I_{p+3}(x,y)$$
---------- (4)
.....
$$D_{n+k}(x,y) = I_{p+k}(x,y) - I_{p+k+1}(x,y)$$

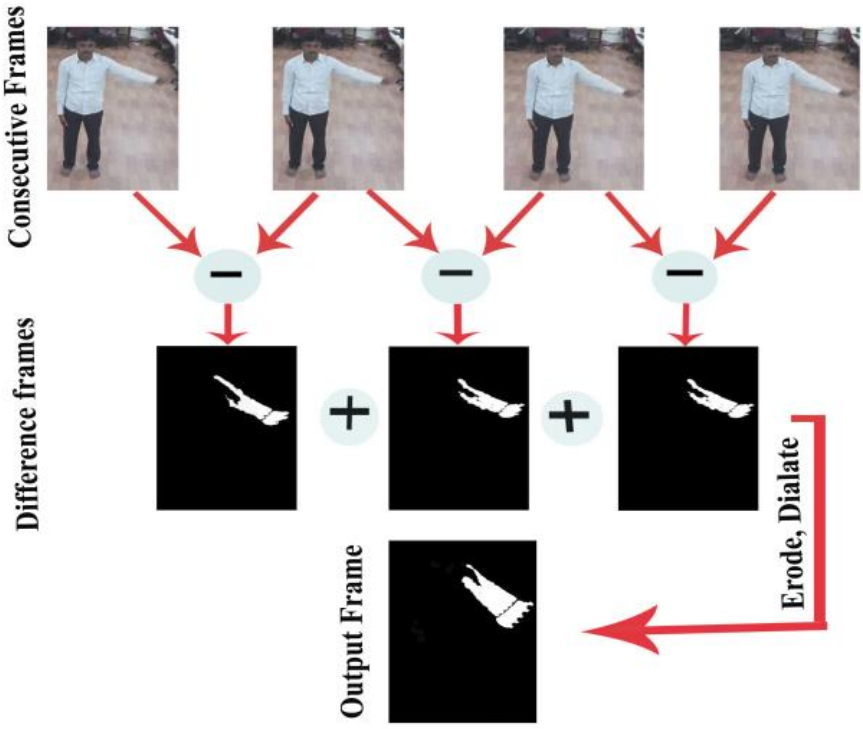

Fig. 2. n - Frame cumulative difference.

Fig. 3. Cumulative Difference.

4.3 ROI Extraction:

In order to assure a fast ROIs generation process and suppress as many resulting negative ROIs as possible the approaches uses efficient and accurate ROI generation scheme in [18]. Background subtraction is popularly applied in general to extract the foreground regions. The same is utilized in this approach for extracting the foreground image, by subtracting a reference background image.

Once the foreground image is extracted, the next step is to identify the ROI for further analysis. For ROI extraction, the approach used in [19] is utilized. The height of the bounding box $H_{(t)}$ for ROI extraction is calculated using $Height_{(t)} = H_{(t)}/H_{(max)}$, where $H_{(t)}$ is the height of the bounding box in the video frame at time 't', H(max) is the maximum value that $H_{(t)}$ has for the entire video sequence. Width of the bounding box is fixed similarity using $Width_{(t)} = W_{(t)} / W_{(max)}$. Finally, ROI is extracted as $ROI = Height_{(t)} / Weight_{(t)}$.

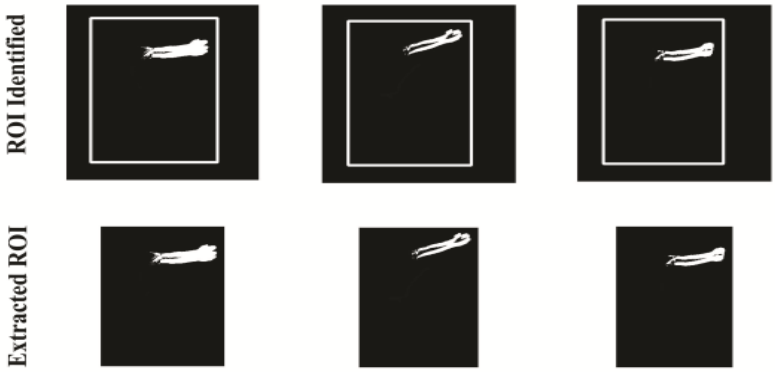

Fig. 4. ROI identified in first row, second row represents extracted ROI.

The bounding boxes extracted for various frames in the video sequence are shown in Fig. 4. For the purpose of the uniformity, the ROI region is considered to be of size 60 X 40 for all actions without any loss in information.

5. CONCLUSION AND FUTURE WORK

This work presented for traffic personal gesture detection for traffic surveillance using n-frame cumulative difference. Indian traffic personal performs 12 actions are taken performing the experiment. The ROI extracted from the various cumulative frame difference images are used. In future work, intend to enhance the flexibility of this approach by using traffic personal gestures to recognize and using various algorithm in complex environment scenes.

REFERENCE:

[1] Johansson G. Visual perception of biological motion and a model for its analysis. Attention, Perception, & Psychophysics, 1973: 14: 201-211: doi:http://dx.doi.org/10.3758/BF03212378.
[2] Poppe R. A survey on vision-based human action recognition. Image and Vision Computing, 2010: 28 (6): 976-990.
[3] Daniel Weinland, Remi Ronfard, Edmond Boyer. A survey of vision-based methods for action representation, segmentation and recognition. Computer Vision and Image Understanding, 2011: 115: 224-241.
[4] Alexandros Andrac Chaaraoui, Pau Climent-acrez, Francisco Flarez-Revuelta. A review on vision techniques applied to Human Behaviour Analysis for Ambient-Assisted Living. Expert Systems with Applications, 2012: 39: 10873-10888.
[5] Sathya. R and Kalaiselvi Geetha. M. Human action recognition to understand hand signals for traffic surveillance. Elixir International Journal of Computer Science and Engineering, 2013: 63: 18149-18156.
[6] Jose M. Chaquet a, Enrique J. Carmona, Antonio Ferna¡ndez-Caballero. A survey of video datasets for human action and activity recognition. Computer Vision and Image Understanding, 2013: 117: 633-659.
[7] Ssu-Wei Chen, Luke K. Wang, Jen-Hong Lan. Moving Object tracking Based on Background Subtraction Combined Temporal Difference. International Conference on Emerging Trends in Computer and Image Processing, 2011.
[8] Sermetcan Baysal , Panar Duygulu. A line based pose representation for human action recognition. Signal Processing: Image Communication, 2013: 28: 458-471.
[9] Changhong Chen, Jimin Liang, Heng Zhao, Haihong Hu and Jie Tian. Frame difference energy image for gait recognition with incomplete silhouettes. Pattern Recognition Letters, 2009: 977-984.
[10] Han, J., Bhanu, B. Individual recognition using Gait Energy Image. IEEE Trans.Pattern Anal. Machine Intell, 2006: 28 (2): 316-322.
[11] Liu, J., Zheng, N. Gait History Image: A novel temporal template for gait recognition. In: IEEE Internat. Conf. on Multimedia and Expo, 2007: 663-666.
[12] Ma, Q., Wang, S., Nie, D., Qiu, J. Recognizing humans based on Gait Moment Image. In: Eighth ACIS Internat. Conf. on SNPD, 2007: 606-610.
[13] Lei, J., Ren, X., & Fox, D. Fine-grained kitchen activity recognition using rgb-d. In Proceedings of the 2012 ACM conference on ubiquitous computing UbiComp New York, NY, USA: ACM, 2012: 208-211.
[14] Encarnacion Folgado, Mariano Rincon, Enrique J Carmona and Margarita Bachiller, A block based model for monitoring for human activity, Neurocomputing, 2011: 1283-1289.
[15] J. Arunnehru and M. Kalaiselvi Geetha, Automatic Activity Recognition for Video Surveillance, International Journal of Computer Applications, 2013: 75(9).
[16] E. Shechtman and M. Irani, Matching local self-similarities across images and videos. In: IEEE Conference on Computer Vision and Pattern Recognition, 2007: 1-8.
[17] R. Sathya and M. Kalaiselvi geetha, Vision based Traffic Police Hand Signal Recognition in Surveillance Video - A Survey, International Journal of Computer Applications, 2013: 81(9): 1-10.
[18] Nazli Ikizler and Pinar Duygulu, Histogram of oriented rectangles: A new pose descriptor for human action recognition. Image and Vision Computing, 2009: 1515–1526.
[19] Qiong Liu, Jiajun Zhuang and Jun Ma. Robust and fast pedestrian detection method for far infrared automotive driving assistance systems, Infrared Physics and Technology, 2013: 288–299.
[20] Wallraven, C., Caputo, B. and Graf, A. Recognition with local features: Kernel receipe. In Proceedings of ICCV, 2003: 257–246.
[21] Wolf, L. and Shashua, A. Kernel Principal angles for classification machines with applications to image sequences interpretation. In Proceedings of CVPR, 2003: 635–640.
[22] Nello Cristianini and John Shawe-Taylor. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press, 2000.
[23] Tom Mitchell. Machine Learning. McGraw-Hill Computer science series. 1997.
[24] Vapnik, V. Statistical Learning Theory. Wiley, NY, 1998.
[25] C.-C.Chang and C.J. Lin. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 2011: 1–27.