ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

A Novel Approach for Human Body Detection in Video Streams Based on Mean Shift Iteration

K.Harish, Prof.B.Anuradha

M.Tech Student, Dept. of ECE, SVU College of Engineering, Tirupati, Andhra Pradesh, India

Professor, Dept. of ECE, SVU College of Engineering, Tirupati, Andhra Pradesh, India

ABSTRACT:A novel methodology for detecting human bodies in dynamic scenario is presented in this paper. We implemented mean shift iteration algorithm for image sequences and videos with complex background to extract human bodies which is applicable for numerous applications like sign recognition, scene understanding, and abnormal action recognition in video streams. We propose a novel methodology which is applicable for video streams with different pose and illumination variations in cluttered environments. This algorithm yields better results compared to state-of-art scenarios.

KEYWORDS: mean shift algorithm; image segmentation; face congnition; sign recognition

I. INTRODUCTION

Human body detection and tracking in videos have received significant attention in the last few years due to the success of Kinect cameras and applications in human computer interaction, surveillance and marker-less motion capture. While there have been successful methods that estimate 2D body pose from a single image, detecting and tracking body configurations in unconstrained video is still a challenging problem. The main challenges are from the large variability of people's clothes, articulated motions[1], occlusions, outliers and changes in illumination. More importantly, existing extensions of 2D methods cannot cope with large pose changes due to camera view change. A common strategy to make these 2D models view-invariant is to gather and label human poses across all possible viewpoints. However, this is impractical, time consuming, and it is unclear how the space of 3D poses[12] can be uniformly sampled. Toaddress these issues, we propose to formulate the problem of human body detection and tracking.

In the existing system, a bottom-up approach for human body segmentation in static images is used. We decompose the problem into three sequential problems: Face detection, upper body extraction, and lower body extraction, since there is a direct pairwise correlation among them. Face detection provides a strong indication about the presence of humans in an image, greatly reduces the search space for the upper body, and provides information about skin color. Face dimensions also aid in determining the dimensions of the rest of the body, according to anthropometric constraints. This information guides the search for the upper body, which in turns leads the search for the lower body. Moreover, upper body extraction provides additional information about the position of the hands, the detection of which is very important for several applications. The basic units upon which calculations are performed are super pixels from multiple levels of image segmentation. In existing system, the extraction is done only in single images. The standing pose of human body is considered for extraction. There is no multiple numbers of poses. The performance accuracy is about 89.53% .

Literature review explained in Section II. proposed method and results in Section III and section IV respectively.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

II. LITERATURE REVIEW

In existing techniques a bottom-up methodology for automatic extraction of human bodies from single images, in the case of almost upright poses in cluttered environments. The position, dimensions, and color of the face are used for the localization of the human body[], construction of the models for the upper and lower body according to anthropometric constraints,and estimation of the skin color. Different levels of segmentation granularity are combined to extract the pose with highest potential. The segments that belong to the human body arise through the joint estimation of the foreground and background during the body part search phases, which alleviates the need for exact shape matching. In order to cope with the highly dimensional pose space, scene complexity, and various human appearances, the majority of existing works require computationally complex training and template matching processes.

An enormous number of methods for the analysis of human action from video proposed in the last ten years. In this Section, we focus mainly on existing procedures relating to this Paper, that is, action detection methods and function trajectorybased Methods. Inspired by object recognition in images, to treat some methods. To prolong actions as spatiotemporal objects and slipping Window-based methods in the video room. A typical procedure It is a video clip of a particular action, such as the use Submission and searching for similar patterns in the video sequences. Without prior knowledge of the temporal or spatial scales of the plot, the search space for the action Template is unaffordable for exhaustive search. For example, you need to look in a 1 broke through billions of partial volumes Video sequence of size 320 ×240 at 30 fps. Reduce the search space, Sapienza et al.

To extract a collection of fixed-size square of a regular spatiotemporal grid, and employ multiple instancelearning technology to find the discriminative cuboid as an action area. Due to the fixedsized cuboid representation cannot their method of action model Variations in both spatial and temporal scales and Perez use a two-stage pipeline, to 'speed up the evidence process.Use a keyframe first generates a classifier Collection of candidate regions, which are then cut by a Space-time action classifier[2]. In order to decouple the search space Small spaces, leave some methods on human Body detection to estimate the spatial limits of the measures. Klaser et al. To obtain that human footprints through acquisition and Tracking human torso, and then to locate actions within the route with a temporal sliding window classifier. Similar, Zhao et al[10]. to use object detectors (eg, human, face, etc.) to receive object tracks[5]. Make a video object of a pipe extracted spatiotemporal coherent features in the track, and comparing such tubes with a specially developed kernel. The performance of this method depends largely on the Quality of human detector. Despite previous successes in Detecting upright pedestrians, it is still an open problem, detect and track human with any position due to the great flexibility of the human body joints.

Yao et al[4]. first BuildAction tracks by recording and tracking of people Body and then scanned using dense 3-D[12] patches in the Action Track for the category and vote the spatiotemporal Center of action. Willems et al. To use FSS measurebased Ranking choose to discriminatory spatiotemporal interest Points, the room bounding box hypotheses for the cast Action. To group the frame by frame hypotheses in recognition, they use the hypothesis with the highest confidence as seed and groups the hypotheses overlapping with the seeds in a Detection. Gilbert etaldeal with the data mining technique to discriminatory connection to discover functions and to generate a probability map for each frame based on their confidence. Then the pixels with probability more than an empirical threshold are recognized as an action area.However, they have not to mention how the frame as areas for action in an associate Detection. Liu et al. to detect discriminatory local features PageRank and an estimate of the action center of gravity these features. As for temporal localization they employ a temporal sliding window. Concentrate most of these methods to devote to the discovery of discriminatory local characteristics, but less Efforts, the efficient and accurate recording of actions based about these features.

III. PROPOSED MITIGATION SCHEME

In our proposed system the extraction of human body is from video and from the scenes of complex poses in motion vectors. In the proposed technique, we intend to deal with more complex poses, without necessarily relying on strong pose prior. Problems like missing extreme regions, such as hair, shoes, and gloves can be solved by incorporation of more masks in the search for these parts, but caution should be taken in keeping the computational complexity from rising excessively.

In this study, we proposed a new technique to detect the human body from motion vectors. Moreover the human detection is accomplished in videosurveillances, security cameras and the human body is detected in complex poses

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

too. We propose a novel framework for automatic segmentation of human bodies in the frames of video streams based on mean shift iteration algorithm.

Mean Shift is a powerful and versatile non- parametric iterative algorithm that can be used for lot of purposes like finding modes, clustering etc. Mean shift considers feature space as a empirical probability density function. If the input is a set of points then Mean shift considers them as sampled from the underlying probability density function. If dense regions (or clusters) are present in the feature space, then they correspond to the mode (or local maxima) of the probability density function. We can also identify clusters associated with the given mode using Mean Shift.

For each data point, Mean shift associates it with the nearby peak ofthe dataset's probability density function. For each data point, mean shift defines a window around it and computes the mean of the data point. Then it shifts the center of thewindow to the mean and repeats the algorithmtill itconverges.After each iteration, we can consider that the window shifts to a denser region of the dataset. The mean shift vector computed with kernel is proportional to the normalized density gradient estimate obtained with the kernel. The mean shift algorithm seeks a mode or local maximum of density of a given distribution. Mean shift can be summed uplike this

1. For each point x

$$p(x) = \frac{1}{n} c \sum k(\| x-xi \|^2)$$

$$\nabla p(x) = \frac{1}{n} c \sum_{i=1}^{n} \nabla k(||x - xi||^2)$$

$$\nabla p(x) = \frac{1}{n} 2c \sum (x-xi)\, k\, (\| x-xi\|^2)$$

2. Choose a search window

g (x) = k(x)

$$g\, (\| x-xi\|^2) = gi$$

$$\nabla p(x) = \frac{1}{n} 2c \sum gi * \left(\frac{\sum xi * gi}{\sum gi} - x \right)$$

$$\nabla p(x) = \frac{1}{n} 2c \sum gi * m(x)$$

3. Compute the mean shift vector

$$m(x) = \frac{\nabla p(x)}{\frac{c}{n} \sum_{i=1}^{n} gi}$$

4. Repeat till convergence

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

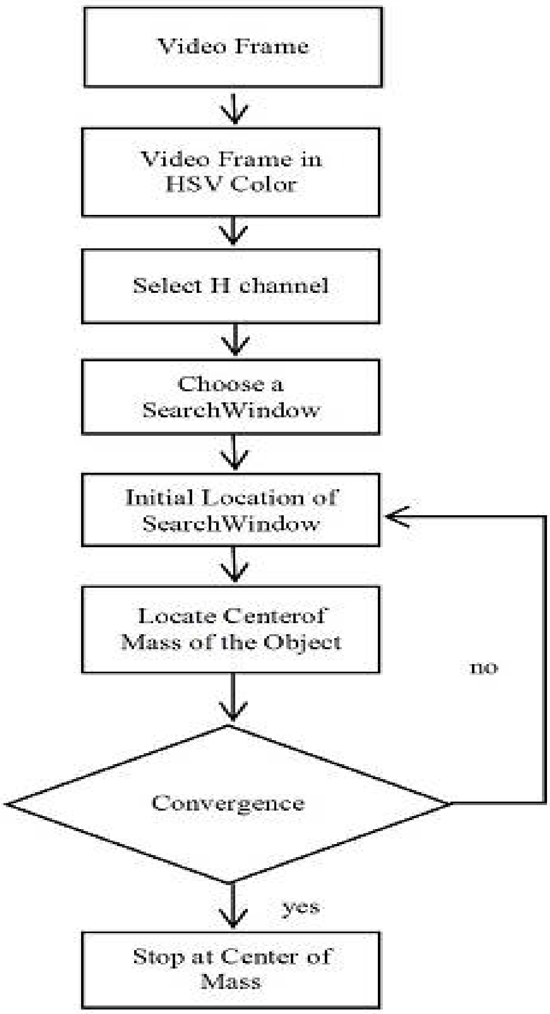

Fig 1 Flow chart for mean shift algorithm

We combine information gathered from different levels of image segmentation, which allows efficient and robust computations upon groups of pixels that are perceptually correlated. Soft anthropometricconstraints permeate the whole process and uncover body regions. Without making any assumptions about the foreground and background, except for theassumptions that sleeves are of similar color to the torso region, and the lower part of the pants is similar to the upper part of the pants, we structure our searching and extraction algorithm based on the premise that colors in body regions appear strongly.

Face detection guides estimation of anthropometric constraints and appearance of skin, while image segmentation provides the image's structural blocks. In this system the extraction of human body is from video and from the scenes of complex poses in motion vectors (Videos).

We intend to deal with more complex poses, without necessarily relying on strong pose prior. Problems like missing extreme regions, such as hair, shoes, and gloves can be solved by incorporation of more masks in the search for these parts. The proposed method gives better results in dynamic environment and in still images than the existing methods.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

IV. SIMULATION RESULTS

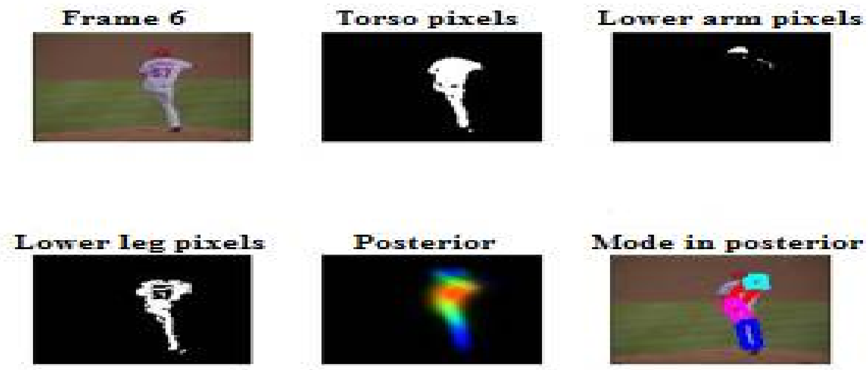

Fig 2:The detection of human body at frame6

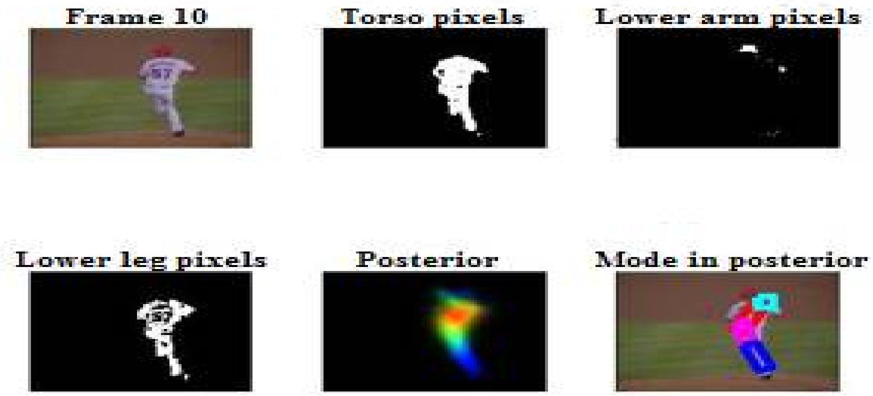

Fig 3: The detection of human body at frame 10

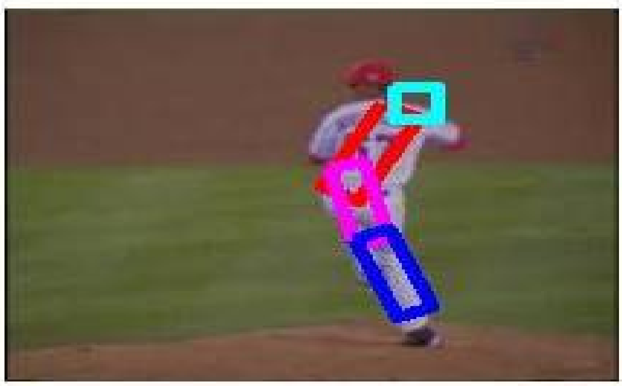

Fig 4: Final Detected Image

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

Comparison Table

	Mean	Standard Deviation
Proposed Method	2.0797	0.040422
Existing Method	0.64257	0.055551

V. CONCLUSION

We presented a novel methodology for extracting human bodies from videos. The main component of the system is the face detection step, where we estimate the rough location of the body, construct a rough anthropometric model, and model the skin's color. The mean shift algorithm employs the detection very accurate compare to the existing methods. Soft anthropometric constraints guide an efficient search for the most visible body parts, namely the upper and lower body, avoiding the need for strong prior knowledge, such as the pose of the body. However, we make some assumptions about the human pose, which restrict it from being applicable to unusual poses and when occlusions are strong. In the future, we intend to deal with more complex poses, without necessarily relying on strong pose prior. Problems like missing extreme regions, such as hair, shoes, and gloves can be solved by incorporation of more masks in the search for these parts, but caution should be taken in keeping the computational complexity from rising excessively. We intend to deal with videos to extract the human bodies with complex poses.

REFERENCES

[1] M. Andriluka, S. Roth, and B. Schiele, "Pictorial structures revisited: People detection and articulated pose estimation," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2009, pp. 1014–1021.
[2] V. Ferrari, M. Marin-Jimenez, and A. Zisserman, "Progressive search space reduction for human pose estimation," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2008, pp. 1–8.
[3] V. Delaitre, I. Laptev, and J. Sivic, "Recognizing human actions in still images: A study of bag-offeatures and part-based representations," in Proc. IEEE Brit. Mach. Vis. Conf., 2010.
[4] B. Yao and L. Fei-Fei, "Grouplet: A structured image representation for recognizing human and object interactions," in Proc. IEEE Conf. Comput.Vis. Pattern Recog., 2010, pp. 9–16.
[5] P. Buehler, M. Everingham, D. P. Huttenlocher, and A. Zisserman, "Long term arm and hand tracking for continuous sign language TV broadcasts,"in Proc. 19th Brit. Mach. Vis. Conf., 2008, pp. 1105–1114.
[6] A. Farhadi and D. Forsyth, "Aligning ASL for statistical translation using a discriminative word model," in Proc. IEEE Comput. Soc. Conf. Comput.Vis. Pattern Recog., 2006, pp. 1471–1476.
[7] L. Zhao and L. S. Davis, "Iterative figure-ground discrimination," in Proc.17th Int. Conf. Pattern Recog., 2004, pp. 67–70.
[8] L. Grady, "Random walks for image segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 11, pp. 1768–1783, Nov. 2006.
[9] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary& region segmentation of objects in ND images," in Proc. IEEE 8th Int.Conf. Comput. Vis., 2001, pp. 105–112.
[10] S. Li, H. Lu, and L. Zhang, "Arbitrary body segmentation in static images,"PatternRecog., vol. 45, no. 9, pp. 3402–3413, 2012.
[11]. Zuffi, S., Romero, J., Schmid, C., Black, and M.J.: Estimating human pose with flowing puppets. In: ICCV. (2013)
[12]. Agarwal, A., Triggs, B.: Recovering 3d human pose from monocular images. IEEE Trans. Pattern Anal. Mach. Intell. 28(1) (2006)