International Journal of Information Technology and Management (IJIT) Vol. 03, No. 03,May 2019.

AN ALGORITHM FOR HUMAN POSE ESTIMATION FROM CCTV IMAGES

Sami ur Rahman, Sajid Ali Khan, Fakhre Alam

Department of Computer Science, University of Malakand, Chakdara.

ABSTRACT

Close Circuit TV (CCTV) imaging plays a very important role in security and crime control. Continuous observation of the CCTV recording is a tedious job and an alternate solution is the automatic analysis of the scene. Automatic detection and recognition of human from CCTV images is playing important role in crime detection. The detection of different human poses (standing, sitting and lying) in CCTV images is the first step towards automatic scene analysis. In this paper, an algorithm is presented for estimating different human poses from CCTV images. The proposed algorithm works in two steps. In the first step the humans are extracted and in the next step the different poses are calculated using the ratio between the height and width of the human images. The proposed algorithm estimate human poses with 98% accuracy.

KEYWORDS

Pose estimation, CCTV images, Security, Scene Analysis, Crime Control

1. INTRODUCTION

Nowadays, the security situation of the whole world has become very complicated because of the terrorist attacks on schools, colleges, universities and other public and private properties. Normally different security measures are taken to counter security threats. These include deployment of security guards, security alarms installations, metal detector and detection of explosive materials. However, due to the complex nature of security threats, the traditional measures do not work very efficiently. Moreover, these measures can be easily counteracted or deceived.

The new trend in dealing with the robbery and terrorist attacks is the use of technology. Now for security measures, CCTV cameras are used for monitoring. CCTV cameras record the scene which greatly helps in investigations. Although these security cameras are playing an important role in the investigation but these need continuous observation which sometimes become very difficult. The new trend is to use technology for automatic scene analysis and for the automatic detection of security threats. For this purpose, different computer algorithms are used to automatically observe the CCTV footages and detect security threats such as detection of suspected human (thief, terrorist), pose detection, weapon detection, and gunshot detection. These different parameters are automatically detected and then an automatic emergency alarm can be generated. The focus of this paper is to automatically detect human pose in CCTV footage. A pose of a human is the configuration of its body parts. A human can adopt infinite poses however standing; sitting and lying are major poses that a human adopts.

The pose of people at some location can convey valuable information. For example, if the pose of many people in a shopping mall/market suddenly changes from standing to lying pose then it means that something unusual has happened. If someone is found lying down in a street, then it shows a situation that needs attention. In such type of scenarios, the automatic detection of different poses can convey valuable information. The focus of this paper is to automatically detect human pose and to generate an emergency alarm. The idea can also be implemented in health canters and old age houses for monitoring purposes. Rest of the paper is organized as follow: Section two describes the previous work. Overview of the proposed approach is illustrated in Section three. Results are presented in Section four. Section five describes the discussion while conclusion and future work is described in Section six.

2. RELATED WORK

Automatic pose detection is an emerging research area in computer vision. Human pose detection plays a vital role in automatic scene analysis. A lot of research work has been done in this area and several techniques are now available which accurately estimate human pose. Some researchers have used Histogram of Oriented Gradient (HOG) features [1-3] for pose estimation. Their approach is based on matching input image with images from predefined data set. The dataset contain over 1800 annotated human images with a large range of pose variations and backgrounds. In order to estimate human pose in videos, some similar approaches [4-7] were used. All of them detected human body using HOG features and then recognized different body parts and estimated pose from the position of different body parts. Kakadiaris and Metaxas [8] presented a template-based approach for detecting human silhouettes in a specific walking pose. The templates consist of short sequences of 2D silhouettes obtained from motion captured data to incorporate motion information into them and help distinguish actual people who move in a predictable way from static objects whose outlines roughly resemble those of humans. A number of researchers used decision trees [9, 10] to estimate human pose estimation. Some algorithms such as body parts recognition algorithms [11-13] and human body joints recognition algorithms [14-16] were developed for human pose estimation. There were two steps involved in their work: first the technique was used to correctly recognize body parts [11-13] or different joints in human body [14-16]; in the second step they estimated pose from location of these body parts [11-13] or joints [14-16]. A number of techniques [17-19] were used to detect pose of only upright people. They used to compare human body in a frame with INRIA Xmas Motion Acquisition Sequences (IXMAS) Multi-View Human Action Dataset, and the i3DPostMulti-View Human Action and Interaction Dataset detection of human pose. The approaches in [17-19] used only different poses of upright people. A multi-stream multi-task deep network [20-22] was used for joint human detection and head pose estimation in RGB-D videos. A motion based algorithm [23] was used that detects falling of aged people. The ‘falling of human body’ is estimated by the sudden or abrupt change in the human body position. For the detection of human falling they used a technique in which they compare the human pose in current frame with previous frames. Any abrupt change in human pose suggests falling of human body.

So far the work done in this area, the researchers have used training dataset technique which makes their approach slow. More complex dataset will produce more accurate result but affect the time efficiency. Moreover researcher used special hardware for pose estimation, which is slightly expensive. In the view of the above, we suggest a new technique in which there is no need of training dataset and it uses anatomical information from CCTV images for pose estimation. Our system estimate pose using height to width ratio directly from input images and is able to estimate human pose in real-time.

3. PROPOSED ALGORITHM

The proposed approach consists of a series of steps as shown in Figure. 1.

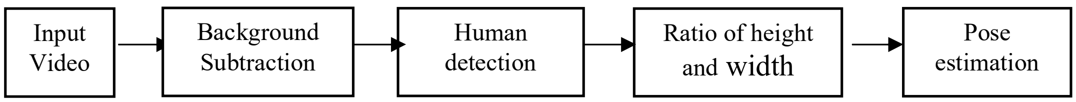
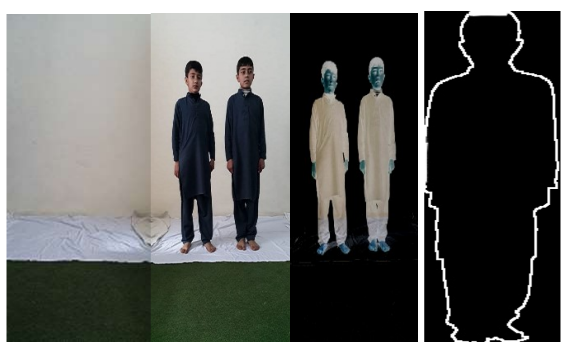


Figure 1. Steps of the proposed approach

3.1 Background Subtraction and Human Detection

The first step in the proposed approach is to accurately detect human body. For this purpose, Gaussian Mixture Model (GMM) is used for background subtraction. After background subtraction the images are enhanced using a series of enhancement filters. After enhancement, the imagers are converted to binary images. In the next step individual humans are segmented as shown in figure 2.



a b c d

Figure 2. Human detection in input image: (a) Background image (b) Input image (c) Background subtracted images d) Segmented Image

3.2 Height and Width Ratio Calculation

The important step in the proposed approach is calculating the height and width of the human body. In our proposed algorithm, first we compute Centre of Gravity (COG) (see Figure 3) of the human body using the following equation. In this equation, coordinates of only those pixels are considered where the pixel value is non-zero.

$$CoG = p(x,y) = \left(\frac{\sum_{i=1}^{n} x_i}{N}, \frac{\sum_{i=1}^{n} y_i}{N} \right) \quad \textbf{(1)}$$

Where $p(x,y)$ is the pixel at CoG, x_i is the pixel coordinate along x-axis and y_i is the pixel coordinate along y-axis and N is the total number of non-zero pixels.

For calculating the height of the object, the Top_y and bottom $Bottom_y$ pixels along y-axis are calculated as follows:

$$Top_y = Max(y_i - CoG_y), \forall y_i > CoG_y \quad \textbf{(2)}$$

Top_y is the pixel with maximum distance from COG along y-axis above CoG_y.

$$Bottom_y = Max(y_i - CoG_y), \forall y_i < CoG_y \quad \textbf{(3)}$$

$Bottom_y$ is the pixel with maximum distance from COG along y-axis below CoG_y.

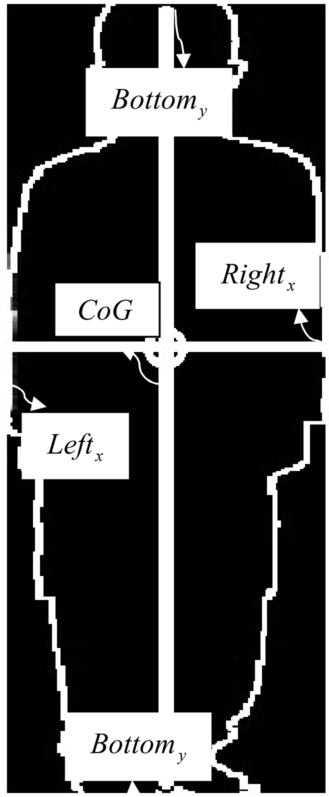


Figure3. Height and width calculation from the binary image

Now the height is calculated as

$$Height = Top_y - bottom_y \tag{4}$$

For calculating the width of the object, the leftmost ($Left_x$) and right most ($Right_x$) pixels along x-axis are calculated as follows:

$$Right_x = Max(x_i - CoG_x), \forall x_i > CoG_x \tag{5}$$

$Right_x$ is the pixel with maximum distance from COG along x-axis towards the right of CoG_x.

$$Left_x = Max(x_i - CoG_x), \forall x_i < CoG_x \tag{6}$$

$Left_x$ is the pixel with maximum distance from COG along x-axis towards the left of CoG_x.

Now the width is calculated as

$$Width = Left_x - Right_x \tag{7}$$

The ratio of the Height and width is calculated as

$$Ratio = Heigth / width \tag{8}$$

3.3 Pose Estimation

In the proposed approach human pose is estimated based on height to width ratio. The experiments were performed from images of different scenarios and the obtained values for different poses are given in table 1.

Table 1. Value for Pose Estimation

Pose	**Height to width ratio**
Standing pose	ratio > 3
Sitting pose	3 > ratio > 0.6
Lying pose	ratio < 0.6

4. Results

We are still working on enhancing our algorithm but here we are presenting the preliminary results. The proposed approach was evaluated using 2000 images. These images contained single person, two persons and more than two persons. The persons in the images had different poses. Our algorithms estimated different poses with 98 % accuracy. We also compared the accuracy of our approach with the state of the art approaches. The comparative results are shown in Figure 4.

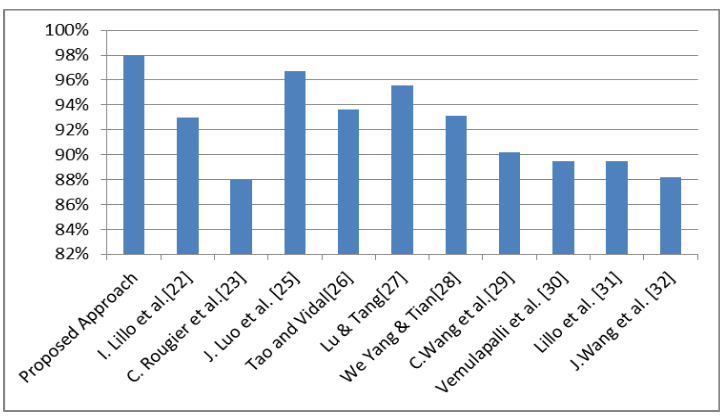


Figure 4. Comparison of the proposed algorithm with the existing approaches.

4. CONCLUSION

This paper presents the preliminary results of our algorithm for pose estimation using CCTV images. The proposed algorithm estimate human poses in real-time. Unlike all existing methods, the proposed approach needs no training data. The algorithm simply calculates the width and height ratio and based on this ratio, estimate the human pose. Future work includes emphasis on pose estimation of overlapped people in complex backgrounds.

REFERENCES

1. N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection," *Computer Vision and Pattern Recognitio*n, San Diego, June 20-25 2005; IEEE, (2005).

2. C. Thurau, V. Hlavac, "Pose primitive based human action recognition in videos or still images" *Computer Vision and Pattern Recognition*, Anchorage, AK, USA, June 23-28 2008; IEEE, pp. 1-8. (2008).

3. X. Wang, T. X. Han, S. Yan, "An HOG-LBP human detector with partial occlusion handling" *12th International Conference on Computer Vision*, Kyoto, Sept 29- Oct 2 2009; IEEE, (2009).

4. Y. Yang, D. Ramanan, "Articulated pose estimation with flexible mixtures-of-parts" *Computer Vision and Pattern Recognition (CVPR*), Colorado Springs USA, June 20-25 2011; IEEE, pp. 1385-1392 (2011).

5. V. Ferrari, M. Marin-Jimenez, A. Zisserman, "Progressive search space reduction for human pose estimation" *IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, June 23-28 2008; IEEE, (2008).

6. S. Johnson, M. Everingham. “Clustered Pose and Nonlinear Appearance Models for Human Pose Estimation” *Proceedings of the British Machine Vision Conference*, September 2010; BMVA Press, 12.1-12.11, (2010).

7. Q. Li, F. He, T. Wang, L. Zhou, S. Xi, "Human Pose Estimation by Exploiting Spatial and Temporal Constraints in Body-Part Configurations" IEEE Access, IEEE, vol. 5, pp. 443-454, (2017).

8. I. A. Kakadiaris, D. Metaxas, "Model-based estimation of 3D human motion with occlusion based on active multi-viewpoint selection" *Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, June 18-20 1996; IEEE, (1996).

9. M. Dimitrijevic, V. Lepetit, P. Fua "Computer Vision and Image Understanding", Elsevier, vol. 104, p. 127-139, (2006).

10. G. Rogez, J. Rihan, S. Ramalingam, C. Orrite, P. H. S. Torr, "Randomized trees for human pose detection*", IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage USA, June 23-28 2008; IEEE, (2008).

11. J. Shotton, Toby Sharp, A. Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, Richard Moore, "Real-time human pose recognition in parts from single depth images" January 2013; ACM, (2013).

12. J. Shotton, A. Jamie, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, R. Moore, "Real-time human pose recognition in parts from single depth images" *Computer Vision and Pattern Recognition CVPR*, Colorado Springs USA, June 20-25 2011; IEEE, (2011).

13. C. Barro´n, Ioannis A. Kakadiaris, "Estimating anthropometry and pose from a single uncalibrated image", Elsevier, 81 (3), pp.269-284. (2001).

14. A. Agarwal, B. Triggs, "Recovering 3D human pose from monocular images," IEEE *Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 44-58, IEEE, (2006)

15. A. Agarwal, B. Triggs, "3D human pose from silhouettes by relevance vector regression" Computer Society Conference on Computer Vision and Pattern Recognition, Washington DC, 27 June-2 July 2004; IEEE, (2004).

16. R. Poppe, "Vision-based human motion analysis: An overview", Elsevier, 108(1-2), 4–18, (2007).

17. M. B. Holte, C. Tran, M. M. Trivedi, T. B. Moeslund, "Human Pose Estimation and Activity Recognition From Multi-View Videos: Comparative Explorations of Recent Developments" *IEEE Journal of Selected Topics in Signal Processing*, 6(5), 538-552, (2012).

18. C. Tran, M. M. Trivedi, "Introducing "XMOB": Extremity Movement Observation Framework for Upper Body Pose Tracking in 3D" 11th IEEE International Symposium on Multimedia, San Diego, CA, Dec 14-16. 2009; (2009).

19. E. Murphy-Chutorian, M. M. Trivedi, "Head Pose Estimation and Augmented Reality Tracking: An Integrated System and Evaluation for Monitoring Driver Awareness" IEEE Transactions on Intelligent Transportation Systems, 11(2), 300-311, June 2010.

20. C. Rougier, J. Meunier, A. St-Arnaud, J. Rousseau, "Fall Detection from Human Shape and Motion History Using Video Surveillance" Advanced Information Networking and Applications Workshops, 2007, AINAW '07. 21st International Conference on, Niagara Falls Ont., May 21-23 2007; (2007).

21. G. Zhang, J. Liu, H. Li, Y. Q. Chen, L. S. Davis, "Joint Human Detection and Head Pose Estimation via Multi-Stream Networks for RGB-D Videos" IEEE Signal Processing Letters, 24(11), 1666 – 1670, (2017).

22. I. Lillo, J. C. Niebles, A. Soto “Sparse composition of body poses and atomic actions for human activity recognition in RGB-D videos, Image and Vision Computing” 59(C), 63-75, (2017).

23. C. Rougier, J. Meunier, A. St-Arnaud and J. Rousseau, "Fall Detection from Human Shape and Motion History Using Video Surveillance," Advanced Information Networking and Applications Workshops, 2007, AINAW '07. 21st International Conference on, Niagara Falls, Ont., 2007,

24. A. Nurhadiyatna, W. Jatmiko, B. Hardjono, A. Wibisono, I. Sina, P. Mursanto, "Background Subtraction Using Gaussian Mixture Model Enhanced by Hole Filling Algorithm (GMMHF)" IEEE International Conference on Systems, Man, and Cybernetics, Manchester, Oct 13-16 2013; IEEE, (2013).

25. J. Luo, W. Wang, H. Qi, Group sparsity and geometry constrained dictionary learning for action recognition from depth maps, in: ICCV, IEEE, 2013.

26. L. Tao, R. Vidal, Moving Poselets : A Discriminative and Interpretable Skeletal Motion Representation for Action Recognition, in: *IEEE International Conference on Computer Vision Workshops*,2015.

27. C. Lu, C.-k. Tang, Range-Sample Depth Feature for Action Recognition, in: *CVPR*,2014.

28. X. Yang, Y. Tian, Super Normal Vector for Activity Recognition Using Depth Sequences,in: *CVPR*,2014.

29. C. Wang, Y. Wang, A. L. Yuille, An approach to pose-based action recognition, in: *CVPR*, 2013.

30. R. Vemulapalli, F. Arrate, R. Chellappa, Human action recognition by representing 3D skeletons as points in a Lsmie group,in: *CVPR*,2014.

31. I. Lillo, A. Soto, J. C. Niebles, Discriminative hierarchical modeling of spatio temporally composable human activities: *CVPR*,2014.

32. J. Wang, Z. Liu, Y. Wu, J. Yuan, Mining actionlet ensemble for action recognition with depth cameras, in: *CVPR*, 2012.