INTEGRATION OF POSES TO ENHANCE THE SHAPE OF THE OBJECT TRACKING FROM A SINGLE VIEW VIDEO

Hemavathy J[1], Shahul Hameed. C[2], Ani Bernish. S[3]

[1]Asst.Prof, Information Technology, Panimalar Engineering College, Tamil Nadu, India
[2]Asst.Prof, Information Technology, Panimalar Engineering College, Tamil Nadu, India
[3]Asst.Prof, Information Technology, Panimalar Engineering College, Tamil Nadu, India

Abstract
In computer vision, tracking human pose has received a growing attention in recent years. The existing methods used multi-view videos and camera calibrations to enhance the shape of the object in 3D view. In this paper, tracking and partial reconstruction of the shape of the object from a single view video is identified. The goal of the proposed integrated method is to detect the movement of a person more accurately in 2D view. The integrated method is a combination of Silhouette based pose estimation and Scene flow based pose estimation. The silhouette based pose estimation is used to enhance the shape of the object for 3D reconstruction and scene flow based pose estimation is used to capture the size as well as the stability of the object. By integrating these two poses, the accurate shape of the object has been calculated from a single view video.

---***---

1. INTRODUCTION

Object reconstruction in three dimensional views is an emerging topic in computer vision. This is mainly used in computer graphics, animation, biomedical, surveillance and for many other analysis purposes. 3D object reconstruction is the construction of 3D shape of the object from a video, which is having the two dimensional information about the 3D objects in sequence of frames. For object tracking, different types of removable physical markers are used on the moving object in the recorded video. Using the transformation of markers or sensors, the movement or motion trajectory of the object is tracked. They are very expensive and used only to track the object which is already having the sensors or markers in the body.

In Adaboost edge detection technique [14], multi-view 3D human pose estimation is still a difficult task in uninhibited environments. Several associated approaches are still dependent on silhouettes of the topic obtained by background subtraction. Background subtraction is complex if the cameras are moving or the background is vibrant. Scene flow [15] is the three dimensional action field of position in the world coordinate; just as optical flow is the two dimensional action field of position in an image. It presents a structure for the calculation of dense, non-rigid scene flow from optical flow. This method does include the knowledge of structure computed independently at the next time instant and also the optical flow which results in less accuracy.

The proposed method is marker less and we have used single view camera to capture the movement of a person. To reconstruct the shape of the object from the 2D sequence of images without the camera calibration, estimate the initial pose and its registration is the main task. The rest of paper is organized as follows. We discuss the related work in section II, Section III explains in detail about the proposed work, results and evaluation metrics are discussed in Section IV and Section V provides the conclusion of the paper.

2. RELATED WORKS

3D shape reconstruction is the developed in computer vision field, which is mainly used for visual effects and the analysis purposes. To construct the 3D objects from a 2D video sequence, pose estimation, silhouette extraction and 3D scene flow estimation are the important tasks. The various pose estimation methods are proposed in [1], [2], [3], and [4].

Yogesh Rathi et al. [1] and Jehoon Lee et al. [2] have proposed particle filtering method for pose estimation. [1] Suggested to use particle filtering in static images for segmentation and pose registration problems. But it is producing major occlusion effects between objects. [2] Track the rigid body object in a 2D image sequence and to estimate the pose in 3D space. Different 3D model reconstruction methods are proposed in [3], [4] and [11]. Soon-Yong Park [3] et al comes with the suggestion for automatic three-dimensional model reconstruction technique, which acquire the complete and closed 3D models for real object.

Youngbae Hwang et al. [11] developed a method for 3D reconstruction using Silhouette Extraction which suggests three refinements to get the accurate silhouettes of the moving objects. Christoph Vogel et al. [5] has given one novel method for 3D scene flow estimation, its regularizing 3D scene flow computation by penalizing deviations from local rigidity of motion and integrated into an energy minimization framework. Jung-Ho Ahn [6] et al, Matt Olson

et al. [7], Bruce Gooch [8], Alexandre Alahi et al. [10], Youngbae Hwang et al. [11] and Robert T. Collins et al. [12] proposed various methods for silhouette extraction. [6], [8], [10] and [12] implemented the simple background subtraction algorithms to compute the silhouettes of the moving objects. [7] Using octree data structure to capture the silhouettes.

3. PROPOSED WORK

Initial pose estimation is the task of registering the shape of the object being tracked. It is used to adjust the position and orientation of the object. The proposed integrated method is used to capture the accurate shape of the object. Depending upon the accurate result from the pose estimation method the 3D shape can be reconstructed from a single view video. In this paper, the proposed method is used to estimate the accurate object from the human motion. First method, Scene Flow based Pose Estimation method is used to capture the scale and stability of the object and the second method, Silhouette based pose registration is used to enhance the shape of the object from the 2D video. By integrating these methods, the shape of the object can be registered more accurately. Fig.1 shows the overview of the proposed method. Initial pose of the moving object is registered first, which is computed by the active contour method. This initial pose will be used for comparison with further movement of the object.

3.1. Scene Flow Based Pose Estimation Method

Scene flow is the structure of evident movement of objects, edges and surfaces. Optical flow is the motion points of an object in a two dimensional space coordinates. Scene flow is the motion points of moving object in three dimensional world coordinates. It is used to estimate the motion of moving objects from the series of frames. It is computed between two consecutive frames in all the video frames. Scene flow composes of two components namely Normal Flow and Parallel Flow. The direct method without any constraints is called as normal flow. We have used hierarchical version of Lucas-Kanade optical flow algorithm for parallel flow. This standard algorithm is a predictor for efficient and robust stereo. Optical flow vector is the temporal rate of change of the image plane coordinates (u, v) at a particular point (x, y) is determined by the spatial-temporal variations of the intensity pattern I(x, y). The input frames are shifted to the world coordinates from the special coordinates and the gradient function is computed by using the gradient formula for the input frames.

$$E_{edge} = -|\nabla I(x, y)|^2 \qquad (1)$$

The linear system is updated with its displacement values, before value updating the error frame (T(x, y) - I(x, y)) and Hessian matrix are generating. The transformation of the moving object is then calculated using (2). The two frames are separated by a small time increment Δt and the movement vector (u, v) obtained   by comparing the two consecutive images. $I_x(x,y)$ is the increase in brightness per pixel at position (x, y) in x direction. $I_y(x,y)$ is the increase

in brightness per pixel at position (x, y) in y direction. The optical flow is defined as motion field in 2D dimensional view. Scene flow is defined as motion field of 3D dimensional.

$$\sum_{x,y} I_x \left((T(x,y)fi \ - \&I(x,y)) @ \sum_{x,y} I_y \left((T(x,y) \ - \&I(x,y)) \right) \right]$$

.... (2)

By combining the nearest neighbour pixel's information Lucas-Kanade algorithm solves the optical flow equation. Lucas-Kanade optical flow algorithm divides the frame or image into smaller blocks and assuming the constant velocity in each block.
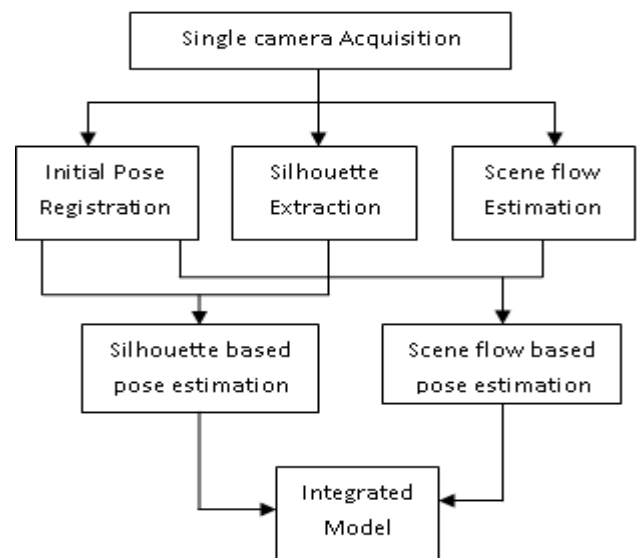

Fig-1.: Architectural Diagram of Integrated Method

3.2. Silhouette based Pose Estimation Method

A silhouette may be the image of a person, an object or a scene and it represented as a solid shape of a single colour. Normally silhouette edges are represent in black colour to match the outline of the subject. Foreground detection [8] methods are used to extract the silhouettes from the video. The mixture of Gaussian models both background and the foreground from consecutive frames like {{X_1, X_2, ... X_t} = I (x, y, i): $1 \le i \le t$} any time t, at a particular pixel (x, y). Model the values of the particular pixel as a mixture of Gaussians. The probability of observing the current pixel value is,

$$P(X_t) = \sum_{i=1}^{k} \omega_{i,t} * \eta \left(X_t, \mu_{i,t}, \Sigma_{i,t} \right)$$

(3)

Where, $\omega_{i,t}$ is the weight of the ith Gaussian in the mixture at the time of t. $\mu_{i,t}$ is the mean value of the ith Gaussian in the mixture, at time t. $\Sigma_{i,t}$ is the covariance matrix of the ith Gaussian in the mixture, at time t. η is the Gaussian probability density function. Weights $\omega_{i,t}$ are updated or

normalized for every new frame. Every new pixel value X_t is checked against the existing k Gaussian distributions until a match is found. The match is defined as a pixel value within 2.5 standard deviation of a distribution.

Detect which Gaussians may correspond to background colours based on the persistence and the variance of each of the Gaussians.

Pixels that do not match with the background pixel value, Gaussians are grouped using connected components, represented as the foreground. The extracted silhouette sequences of frames are integrated with the active contour model using gradient function.

3.3. Integrated Model

In this section, the resulting sequence is produced by the integration of scene flow based and silhouette based pose estimation. The moving object details from both the methods are integrated based on time by using the Boolean operator Logical OR. Silhouette extraction is used to extract the shape of the moving object. It starts with the background estimation and subtraction. Background subtraction is used to calculate the foreground moving objects. In video, it combines the process of extraction and separation of the foreground objects. Then the video frames are segmented from the foreground objects.

The input video stream sequence is denoted by I(x,y). Initial pose of the moving object is denoted as $P_{I,t}(x,y)$ at the time of t. generally the moving object is mentioned as (obj $t_i(x,y)$) at the time of t and in the spatial cooardinate (x,y).

The output of Scene flow based pose estimation is mentioned as (obj$_1$ $t_i(x,y)$) at the time t as well as the output of Silhouette based pose estimation of the moving object is denoted as (obj $t_i(x,y)$) at the time of t and in the position (x,y) in fig. 2. In proposed method, without any smoothing function the evident tracked the object without any noisy data. It produces more information than the scene flow based pose estimation like the size and shape of the tracked object.

1.	Start
2.	Read the input video as frame by frame I(x, y) .
3.	Register the initial pose $P_{I,t}(x,y)$ of the moving object (obj $t_i(x,y)$) at the time t.
4.	Compute the Scene Flow based pose estimation for the moving object (obj$_1$ $t_i(x,y)$)
5.	Calculate the silhouette based pose estimation of the moving object (obj2 $t_i(x,y)$)
6.	Repeat
	a. Integrate (obj$_1$ $t_i(x,y)$) and (obj2 $t_i(x,y)$) over time using Logical OR operator.
7.	Till the end of sequence.
8.	End

Fig - 2: Algorithm for Integrated Method

4. RESULTS AND DISCUSSION

The proposed method has been implemented in MATLAB 7.5 tool on Windows 7 Platform. A common video 'Denis walk' is taken as the input video. The input video is in 'avi' format. Proposed experiments and results are carried out as shown Fig. 3, Fig. 4, Fig. 5, Fig. 6 and Fig. 7
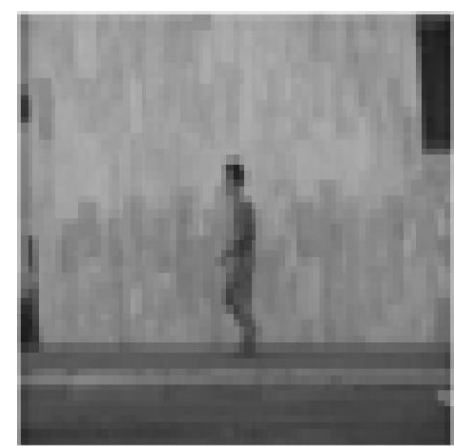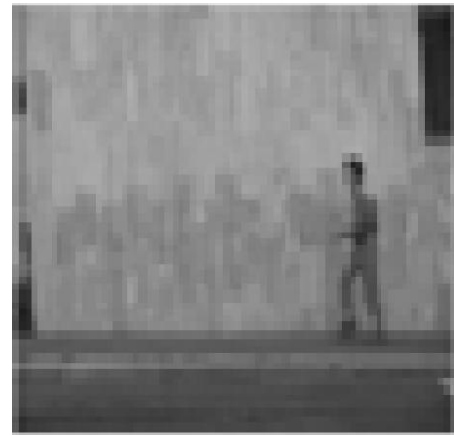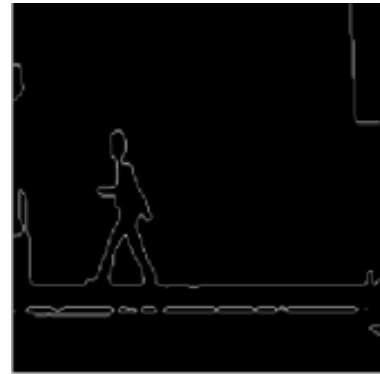
Fig -3: Sample Frames from the Input video

Fig.-4: Selective frames after Pose Estimation
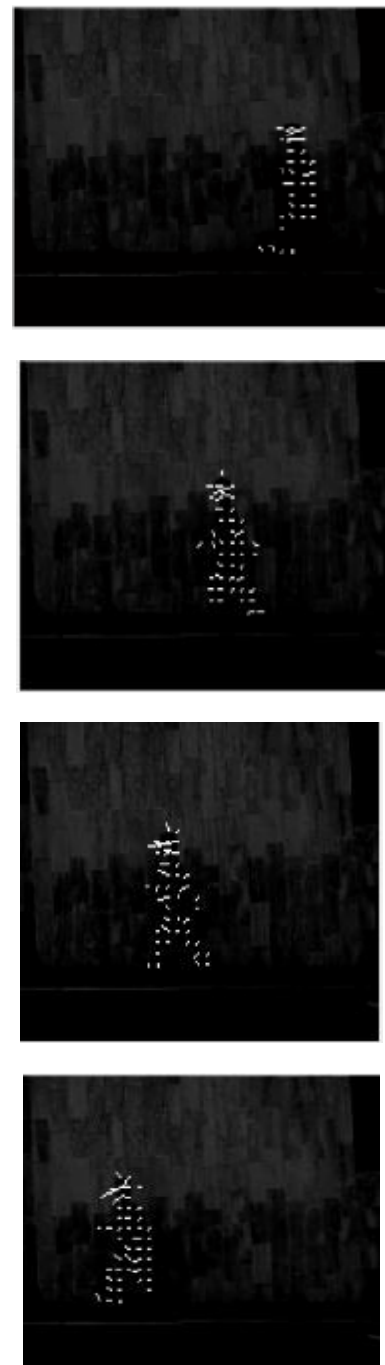
Fig.-5: Output Frames of Scene Flow based Pose Estimation

Fig-7: Integrated Output

Fig.-6: Output frames of Silhouette based Pose Estimation.

Table –1: Comparison of Execution Time

Methods	Execution time(sec)
Silhouette based pose Estimation	10.486
Scene flow based pose Estimation	52.1525
Integrated model	131.799

The execution time of silhouette based pose estimation, scene flow based pose estimation and integrated model are presented in Table I. The Integrated system is able to track the movement of the person more accurately when compared to the silhouette based pose estimation and scene flow based pose estimation. Fig 7 represents the output of the proposed system.

4.1. Comparative Analysis

The evaluation of proposed methods for the moving object datasets are calculated using the following metrics (4).

$$Accuracy = \frac{Number\ of\ TP + Number\ of\ TN}{TP + TN + FP + FN}$$

(4)

True positive, True negative, false positive and false negative is denoted as TP, TN, FP, FN respectively.

Table -2: Comparative Analysis of Methods

Evaluation Metrics	Silhouette based Pose estimation	Scene flow based pose estimation	Integrated Model
Accuracy	24%	51%	52%

The accuracy of the proposed method is compared with the existing method and the results are shown in Table 2. The object's pixel wise accuracy is calculated between pose estimation and the silhouette based pose estimation. It produces 24% of accuracy. The scene flow based accuracy is calculated between pose estimation and scene flow based pose estimation. It produces 51% of accuracy, but it is having some noisy output. The integrated method is again compared with the pose estimation results by pixel wise. The proposed Integration model produces 52% of accuracy without any smoothing or filtering technique.

5. CONCLUSION AND FUTURE WORK

Normally pose estimation methods are mainly used for 3D reconstruction of 2D objects. Reconstruction of 3D shape from a single view video is a difficult task because it doesn't provide enough information as much as the multi view videos. The proposed method provides more exact information about the shape and size of the objects using silhouette based pose estimation and scene flow based pose estimation. It can be applied to multiple moving object videos and dynamic background outdoor videos for future enhancement. The performance analysis has concluded that the integrated system tracks the movement of the person more accurately when compared to the silhouette based pose estimation and scene flow based pose estimation. The future work of this paper is to reconstruct the shape of the object in 3D view.

REFERENCES

[1] S.C.Bilir, Y.Yemez,"Non-rigid 3D shape Tracking from Multiview video", Computer Vision and image Understanding 116(2012)1121-1134.

[2] Ugur Gudukby,Lbrahim Demir,Yigithan Dedeoglu,

[3] "Motion Capture and human Pose Reconstruction from a single-view Video Sequence", Digital Signal Processing 23(2013)1441-1450.

[4] ShoheiNobuhara , Takashi Matsuyama ,"Heterogeneous Deformation Model for 3D Shape and Motion Recovery from Multi-Viewpoint Images", Proceedings of the 2nd International Symposium on 3D dataProcessing, Visualization and Transmission(2004).

[5] E. de Aguiar, C. Theobalt, C. Stoll, H.-P. Seidel, "Marker-less deformable mesh tracking for human shape and motion capture", in: Proc. Computer Vision and Pattern Recognition (CVPR), 2007.

[6] J. Starck, A. Hilton, "Surface capture for performance-based animation", IEEE Comput. Graph. Appl. 27 (3) (2007) 21–31.

[7] T. Matsuyama, X. Wu, T. Takai, S. Nobuhara, "Real-time 3D shape reconstruction, dynamic 3D mesh deformation, and high fidelity visualization for 3D video", Comput. Vis. Image Understand. 96 (3) (2004) 393–434.

[8] K.Varanasi, A. Zaharescu, E. Boyer, R. Horaud, "Temporal surface tracking using mesh evolution", in: Proc. European Conference on Computer Vision (ECCV), 2008, pp. 30.

[9] A Alahi, L Bagnato, D Matti," Foreground silhouette extraction robust to sudden changes of background appearance Image Processing (ICIP 2012.

[10] Park, M.Subbarao, " Pose Estimation and Integration of Complete 3D Model Reconstruction" Computer Vision, 2002,(WACV) 2002.

[11] Y Rathi, N Vaswani, A Tannenbaum,,"Particle Filtering for Geometric Active Contours with Application to Tracking Moving and Deforming Objects" Computer Vision, 2005.

[12] Youngbae Hwang, Jun-Sik Kim, Inso Kweon," Silhouette Extraction for Visual Hull Reconstruction ", Conference on Machine Vision Applications(MVA2005 IAPR) May 16-18, 2005, Japan.

[13] C.H. Esteban, F. Schmitt, "Silhouette and stereo fusion for 3D object modeling", Comput. Vis. Image Understand. 96 (3) (2004) 367–392.

[14] Y. Yemez, Y. Sahillioglu, "Shape from silhouette using topology-adaptive mesh deformation", Pattern Recognition. Lett. 30 (2009) 1198–1207.

[15] Hunsung Kim,Ryuleki Sakamoto,Itaru Kitahara,tomji Toriyama and Kiyoshi Kogure,"Robust Silhouette Extraction Technique using Background subtraction" (2007) 552-557.

[16] Shih-Chia Huang," An advanced motion detection algorithm with video quality analysis for video surveillance systems", IEEE Transactions On Circuits And Systems For Video Technology, Vol. 21, No. 1, January 2011.

[17] C.Sminchisesu,A.Telea, " Human Pose Estimation from Silhouettes a Consistent Approach using Distance Level Sets", International Conference in Central Europe on Computer Graphics 2002, Visualization and Computer Vision.

BIOGRAPHY

Mrs.Hemavathy received M.E degree in Information Technology from Madras Institute of Technology, Chennai, TamilNadu, India in 2013. Presently she is working Panimalar Engineering College, Chennai , India.