Literature Review III

Viswanath Pulle - 01690370

Primary Paper - 3D Human Model Reconstruction from Sparse Uncalibrated Views
Secondary Paper - Scanning 3D Full Human Bodies Using Kinects

3-D technology and virtual reality has evolved significantly, a lot of initiatives have been taken to incorporate these advances effortlessly in our life by capturing 3D models of humans. Most research makes use of regular and depth cameras to capture these 3D models. There are certain drawbacks using this technology. For example, Depth cameras like Kinect require a lot of complicated procedures and can be used only for very short distances. In the case of regular cameras they rely majorly on Multiview stereo(MVS) algorithms which require calibrated cameras and plenty of images to process the depth of a body.
The **Primary paper** tackles this problem by researching a technique that reconstructs high quality 3D models of humans wearing clothes from images taken by sparse uncalibrated cameras.

The authors referenced the **Secondary paper** (Full Human Bodies Using Kinects) - to understand how depth cameras are used in generating 3D model. They have used 3 Kinects to capture the entire human body, One each for the upper, lower and the middle to avoid overlapping. The body needs to be quite close to Kinects as it can be used for only short distances and this limitation is overcome in the two-stage algorithm developed in the primary paper. The reason they have chosen Kinects as it can be easily acquired by the general public and is relatively cheap. They have overcome that Kinect faced of producing low quality images by proposing a two-phase approach that uses three Kinects. This technique generates good quality models when taken from close range. The Kinects are adjusted such that the ones that capture the upper and lower are on one side and the Kinect that captures the middle is on the other. This prevents overlapping and interference in the images captured. The first phase of the algorithm is to capture the non-rigid registration of the scanned data. The method that they use is similar to the one used by the primary to capture the model by using a rough template as markers for the generating the 3D model. This template is deformed based on the features and the corresponding color maps generated by the Kinects. The second phase performs a graph algorithm to complete the global alignment with the loop closure constraint. We can see the primary has relied on this paper on how the good quality 3D models can be generated using depth cameras and have relied on the idea to use a deformable template to create its rough model.

There are several issues which need to be addressed as the existing algorithms can't be used to describe the features for this new techniques. The reason is that, they require a large number of images for reconstruction and also the previous algorithms required calibrated cameras which use Structure from Motion(SVM) to generate the point clouds which are also not very accurate. Also, the main problem that arises is that the MVS fails to detect the intricate patterns in clothes

and cannot extract the information required for the geometric reconstruction in 3D.The primary paper resolves the reconstruction by employing a two-stage algorithm.
In the first stage of the algorithm we initially use a template model for a human before we start generating the point clouds to fit in, the non-grid dense correspondences (NRDC) algorithm which generated the dense correspondences based on the camera calibration is also used. To compensate for the NRDC low accuracy we reference the water tight model to produce a coarse model of the human.
In the second stage of the algorithm we capture the intrinsic details which are referred to as wrinkles in the human model and reconstruct it. Clothes are hard to capture, So the team has adopted a hierarchical density-based clustering algorithm to recreate the patterns on the human model. This algorithm does only requires about a dozen images to reconstruct the human model.

The primary paper also gives a systematic overview of how the algorithm works with an experiment. We begin the process by taking n images and numbering each image as it comes through. For the first stage we need to create the coarse model this done by first running every pair of adjacent images through NRDC and calculate their dense correspondences. Each pixel in one image is checked with the other and is only returned if it's above a confidence threshold. Once each set of correspondences have been calculated and propagated they each form a 2D skeleton which are bundled and projected using a SFM solver algorithm.
The output of the SFM algorithm is a 3D Skelton which is noisy and in complete. They cleaned up the image by enforcing spatial consistencies of pixel correspondence. This is done by removing two related points if their Euclidian distance is greater than a threshold value, after this we apply the density constrain. After we clean up the model we pass it through SFM once again. We generate the final output of the first state in the algorithm by using the deformable template model. We deform the human model template such that it merges with our output of the SFM processing. The projection of the image in the mesh does not happen exactly because of the imprecise calibration but their system fixes this by deforming the silhouettes of the image to match the projections of the mesh.

The first stage mainly focuses on getting the outermost surface of the model. The second stage of the algorithm is to geometrically reconstruct the contours of the clothes in the models. They have incorporated a shape-from-shading (SFS) algorithm to gather the details for reconstruction. Even colors with different shades are collected and put into different clusters. Hierarchical density clustering was implemented to cluster the different shades of geometrical patterns. This algorithms requires only two inputs to compute the number of clusters. Each cluster is certain shade which can be used for more accurate shading extraction. We run this algorithm on each image and then we merge clusters from different views if their mean color fall below a certain threshold. We then compare the clusters and the hue of the largest cluster is taken as the hue for the pixel. Finally the shading is taken as the ratio of the original image to pixel chosen. This method is used for the clothes and the hair and skin is handled by different algorithms. The skin is handled by skin-color-detection algorithm and the hair by a separate cluster at the top of the segmented human figure. The geometric in the clothes is replicated using the shading patterns. For the face of the coarse model the face frontal views is detected by the camera and then back projected onto the model, this is done using the estimated camera projection matrix.

The researchers of this two-stage algorithm have also implemented the algorithm. A cellphone camera and 15 images each of boy1 and boy2 were taken. The models created looked very promising even though there were still certain challenges that required to be overcome such as if the human gives a difficult pose, it then becomes difficult to replicate it accurately. They have also the compared previously used point-cloud MVS (PCMVS) algorithm with the PMVS algorithm that they have used in their technique. The PCMVS algorithm requires a calibrated camera and took 20 views to reconstruct a particular human model that was chosen. The PMVS algorithm took only 16 views and did not require the calibrated camera that was used by the PCMVS to produce the same human model. Based on their research they have gone against the norm which says that more the number of images the accurate the model will be. They ran several tests varying the number of input images and found that 15 uniformly sample views around the person will be sufficient. The runtime of the algorithm (C++) when run on a 3.4GZ Intel Quad core processor was 33 minutes to generate a single human model. The algorithm that they created still has a lot of scope for improvement as there are still a few limitation that persisted. The algorithm is not completely automatic and require a few skeletal images to be provided to it to perform the reconstruction. Another limitation is certain parts that require detailed reconstruction such as the hands and feet do not come off accurately which show that still more work is required on this algorithm.

In Conclusion, the primary paper proposes an algorithm that solves the problem for reconstruction of 3d models for humans. The secondary paper provided a great base for the authors to further their research and develop the highly efficient algorithm.
The results prove that it is very efficient compared to the existing algorithms. So, I think this would create a platform for more advanced research to be performed on 3d modelling using sparse uncalibrated cameras rather than depending on Kinect camera's technology.

References:

Primary Paper:
Han, Xiaoguang & K. Wong, Kwan-Yee & Yu, Yizhou. (2016). 3D Human Model Reconstruction from Sparse Uncalibrated Views. IEEE Computer Graphics and Applications. 36. 1-1. 10.1109/MCG.2016.68.

Secondary Paper:
J. Tong, J. Zhou, L. Liu, Z. Pan and H. Yan, "Scanning 3D Full Human Bodies Using Kinects," in IEEE Transactions on Visualization and Computer Graphics, vol. 18, no. 4, pp. 643-650, April 2012.
doi: 10.1109/TVCG.2012.56