Workshop on Virtual Reality Interaction and Physical Simulation VRIPHYS (2014)
J. Bender, C. Duriez, F. Jaillet, and G. Zachmann (Editors)

Virtual Fitting Pipeline: Body Dimension Recognition, Cloth Modeling, and On-Body Simulation

D. Siegmund, T. Samartzidis, N. Damer, A. Nouak, C. Busch[1]

{dirk.siegmund, timotheos.samartzidis, naser.damer, alexander.nouak}@igd.fraunhofer.de, christoph.busch@h-da.de
Fraunhofer Institute for Computer Graphics Research Germany *IGD*
[1] Hochschule Darmstadt - University of Applied Sciences

Abstract

This paper describes a solution for 3D clothes simulation on human avatars. The proposed approach consists of three parts, the collection of anthropometric human body dimensions, cloths scanning, and the simulation on 3D avatars. The simulation and human machine interaction has been designed for application in a passive In-Shop advertisement system. All parts have been evaluated and adapted under the aim of developing a low-cost automated scanning and post-production system. Human body dimension recognition was achieved by using a landmark detection based approach using both two 2D and 3D cameras for front and profile images. The human silhouettes extraction solution based on 2D images is expected to be more robust to multi-textured background surfaces than existing solutions. Eight measurements corresponding to the norm of body dimensions defined in the standard EN-13402 were used to reconstruct a 3D model of the human body. The performance is evaluated against the ground-truth of our newly acquired database. For 3D scanning of clothes, different scanning methods have been evaluated under apparel, quality and cost aspects. The chosen approach uses state of the art consumer products and describes how they can be combined to develop an automated system. The scanned cloths can be later simulated on the human avatars, which are created based on estimation of human body dimensions. This work concludes with software design suggestions for a consumer oriented solution such as a virtual fitting room using body metrics. A number of future challenges and an outlook for possible solutions are also discussed.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation

1. Introduction

This work outlines a system created to give consumers in fashion retail an idea of how an item of clothing will look on them before trying it on. In the form of a short video, items of clothing are projected virtually onto an avatar of the user. The virtual projection of a real consumer enables retailers to advertise more styles more quickly and vividly than with analogue media. That advertising strategy is similar to the so-called "Behavior Targeting" to which customers are especially susceptible [Bea10]. This form of digital presentation allows manufacturers and retailers to present their clothing in a way that reflects the brand itself as well as the feeling evoked by each item individually. The color, shape and movement pattern of the "avatars" can be adjusted according to the brand and target group, providing the user with an advertisement-like experience. Another reason for this mode of operation is connected with the self-image of the user, the changing world of advertising and customer behavior, which are increasingly oriented towards the "shopping experience". The touched fields of research by this concept are body dimension recognition, garment scanning and their simulation on digital avatars.

When infra-red based sensors entered the mass market in 2008 research on garment scanning and body dimension recognition became more intense from academia and industrial groups. Applications which resulted from this research include 2D try-on "flat photo image" applications and 360° rotational accurate sizing and fit applications for the on-site retail market and home use. The concept of the proposed system is based, in large part, on research used for dealing with virtual try on and tailor made applications. Existing commercial applications use different approaches in the fields

© The Eurographics Association 2014.

DOI: 10.2312/vriphys.20141228

delivered by EUROGRAPHICS DIGITAL LIBRARY www.eg.org diglib.eg.org

body dimension recognition, cloths scanning, and simulation of cloth. The following sections focus individually on every field of that research.

2. State of the Art Applications

Some examples for augmented and virtual reality fitting systems were demonstrated by Fitnect [Fit14], Bodymetrics [Bod14] and Cyberfit [Cyb14]. Fitnect uses a Microsoft Kinect camera to recognize the body size and adjust 2D garment with the movement of an user. The low resolution of the Kinect camera results in a not particularly sharp image, causing the intersection of the real image and the projection to become obvious. The 2D garment animation of augmented reality solutions is generally not as good as in 3D systems, which include modeling, rendering and animation. Bodymetrics uses a Kinect camera for body dimension recognition but in a virtual reality application context. The emphasis lies in a clothing simulation on a users virtual representation. The online simulation does not take into account the physical characteristics of a garment. . Cyberfit is an in-shop system using six Kinect cameras to capture body measurements and sewing pattern in order to simulate an item of clothing. The projection of clothing on the customer's body takes place without the simulation of drapery. As most manufacturers in that market don not interchange sewing pattern that approach did not fit the requirements of an interoperable application. In contrast to our proposed virtual reality solution Bodymetrics does not use Motion-Captures, instead the user interacts live with his mirror image using skeleton tracking.

3. System Design

This work is focused on the advertisement market and requires therefore a good simulation performance, easy setup and fast body size recognition. The clothes scanning the process must be automated and integrated seamlessly into the pipeline of current visualization techniques. It provides a modular capturing studio, which can be built by synchronizing multiple DSLR-cameras and developing a smooth workflow for digitizing apparel. We managed to design a process, which makes it possible to reconstruct, digitize, simulate and include the apparel in an interactive application nearly without human intervention. The diagram in Figure 1 schematically describes the software components used. The overall system consists of four parts. The main application is developed by using the Unreal Development Kit [Unr13] which is responsible for the visualization, the character creation, authentication and the processing of information input from the user. The second constituent of the system is the body dimension recognition component, which is integrated into the main application and accesses, in turn, independently the cameras. Another component is the MakeHuman API [Mak14], which communicates directly with the body

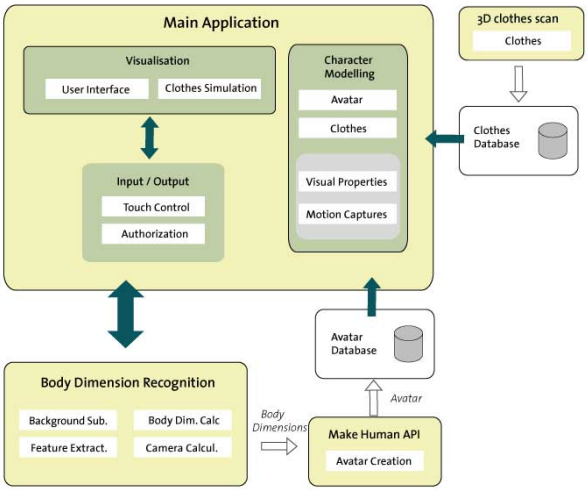


Figure 1: System Architecture

measurement recognition system. The MakeHuman API receives five body measurements in a standardized format according to EN- 13402 [EN-14]: the European standard. The last component of the application is responsible for the creation of the 3D models of the items of clothing. Avatars and items of clothing are saved in separate databases, which are then connected to one other within the main system.

4. Body Dimension Recognition

In recent years, there has been active research on replacing the process of taking anthropometric measurements with automated systems. In several works [KH12] [LW11], the writers see the results of an image-based 2D anthropometric measurement comparable to 3D environments. As the system is meant to be an In-Shop advertisement installation Blackshatter, X-Ray and Structured-Light based Scanner would be inconvenient for the identified target group. The use of ordinary webcams with a resolution 1920x1080 promises, therefore, a cheap and easily scalable system. Kohlschuetter [KH12] mentioned camera calibration and object extraction as potential sources of problems in unclosed scenarios. Reasons for that are limited space, multi-textured surfaces and artificial, direct and/or indirect light coming from different angles. This work therefore compares the extraction results of a 2D approach with the results of a 3D camera setup. For 3D human silhouette extraction the Microsoft Kinect is used, which provides a VGA camera and a depth sensor for measuring the distance to an object. The optical sensors of the Kinect provide images with a resolution of 640x480 pixels and a field of vision of horizontally 57° and vertically at 43°. This results in requiring a minimum distance of 1.2 - 3m in order to cover a human by height and width in vertical alignment. In contrast the HP HD Pro 2D

© The Eurographics Association 2014.

C920 Webcam provides an angle of 78 ° with a minimal distance of 1.8m in vertical alignment in order to cover humans with 2m height. The Kinect uses a diffractive optical element and an infrared laser diode to generate an irregular pattern of dots from which the depth information is calculated. It incorporates a color and a two megapixel gray scale chip with an IR filter, which is used to determine the disparities between the emitted light dots and their observed position. Via stereo triangulation between a near-infrared camera image and a near-infrared laser source generated image, the depth of an object in the scene is identified. An approach using at least two images (frontal and lateral) is more plausible as results for silhouette-based shape reconstruction using just frontal images as [Sze93] [NW97] show.

4.1. Camera Calibration

Camera calibration is necessary due to camera distortion caused by the pinhole design of most cameras without lenses [VCWL12]. As accuracy is one of the most important factors, it is important to have the distortion as close as possible to the ideal pin-hole projection model. Furthermore, it is important for the algorithm to determine the exact relationship between the camera's natural units (pixels) and real world units. Camera resectioning can correct distortion by using an image of the camera. In this process, the parameters of a pinhole camera are estimated, which approximate the properties of the employed camera. Usually, pinhole camera parameters are represented in a 3x4 matrix. Both the Kinect camera and the webcam are suspected to show distortion. The Kinect camera shows furthermore an offset between the IR-sensor image and the RGB image of 8 pixel [ROS14]. A calibration procedure that uses a planar target will be used to create the needed values. A markerless setup was chosen to establish an inconspicuous installation. An asymmetrical circle pattern was used in order to calculate the matrices parameters through basic geometrical equations. The base position of the human was specified by a marker on the ground. The point image coordinates can be automatically recovered to the known real world coordinates. Nevertheless, the calibration becomes invalid when the camera position or the footprints are changed. As a result, the following reprojection errors became apparent as shown in Table 1.

	Original	Calibrated
Webcam	6.08	0.13
Depth (Kinect)	0.35	0.18
RGB (Kinect)	0.55	0.14

Table 1: Camera Reprojection Errors

4.2. Background Substraction

As the scenario does not allow the installation of greenscreening to use keying, a process combining several image processing algorithms in order to extract a human silhouette from background will be introduced. Object detection in static video applications is usually performed using techniques such as background subtraction, optical flow and temporal differencing. The evaluated background subtraction method was presented by Zivkovic [Ziv04] [ZvdH06] and evaluated by Vacavant [VCWL12] and uses texture properties of the background observed over a period of frames. It is assumed that every pixel's intensity value in the video can be modelled using a Gaussian mixture model [KB01]. A simple heuristic determines which intensities are most probably part of the background, other pixels belong to the foreground. Foreground pixels are grouped using 2D connected component analysis [SG99]. With those accumulated texture properties a background image is constructed. For every new frame and pixel, the heuristic makes a decision about whether it should be in the background or in the foreground.

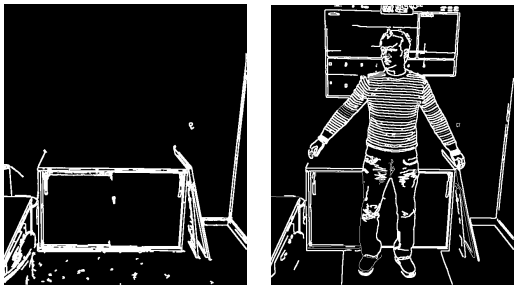
Several factors like subjects replacing each other rapidly or changes in light cause the foreground to be falsely detected or not detected at all. Due to the fact that the application will be used in an in-shop scenario, it is expected that the subject will change frequently and quickly. The constantly accumulated background image did not give good results in this required scenario. Accumulated static background frames also do not work very well on image regions with low contrast.

The background subtraction using the original Mixture of Gaussian (MOG) algorithm [BBV08] showed therefore a poor performance on low contrast images. The Canny algorithm [Can86] is a edge detector [Can14], that provides a low error rate and performs well when detecting edges. Under the described condition, a low threshold gave the best results.

The returned image got binarised by setting a threshold. Applying the MOG background subtraction to the binarized contours did not result in a good foreground substraction. A simple background subtraction between the Canny output containing only the background [Figure 2a] and an image containing both subject and background [Figure 2b] was therefore choosen. The substract computes a matrix-matrix or matrix-scalar difference between the two image representing matrices. As both images are binarised, corresponding white pixels are converted to black.

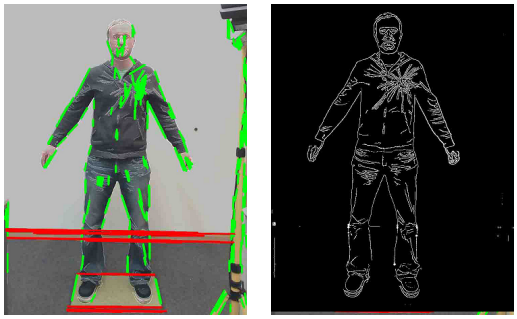
The result after subtraction showed still noise and unwanted objects. This is caused by the Canny algorithm applying different detectors to each edge. That problem could be addressed by accumulating the Canny output over several images (30) using a bitwise operator and applying morphological operators erosion and dilation after subtraction. To extract the outline of the subject the topological structured analysis polygon approximation was used [SA85]. The largest contour has been drawn to a new image and filled by using a flood and fill algorithm. The now unclosed contour is caused by the subtraction process identifying the inter-

© The Eurographics Association 2014.



(a) (b)

Figure 2: (a) Background (b) Background with subject



(a) (b)

Figure 3: (a) Frontal View (b) Literal View

secting lines between leg and background objects as background. This is a principal problem of the approach using Canny and subtraction. The non-uniform background of the subject resulted in even fewer closed contours. This problem could be solved as follows: As outlines of human beings, when standing upright, are mostly vertical, horizontal lines of the background can be marked for further preprocessing. Furthermore, horizontal lines of a human standing upright are not very long (head, shoulders etc.). On the image containing the subject, vertical lines of a certain length are therefore selected. The intersection points between both lines will be calculated and points wrongly erased during subtraction will be reconstructed. Figure 3a illustrates the detected horizontal and vertical lines from both viewing angles. In order not to draw the intersection point onto positions, which do not belong to the human body, the image got masked before applying these points. To be able to select horizontal and vertical lines, Hough Line Transform is used. Figure 3b shows the merged Canny output contour with the intersection points. With the recreated points, the contour extraction returns a better closed silhouette (shown in Figure 4).

For extraction of a silhouette with the Kinect camera skeleton model capabilities have been used. The stream selects any single person as the foreground combining depth and RGB image of the camera. The non-uniform background

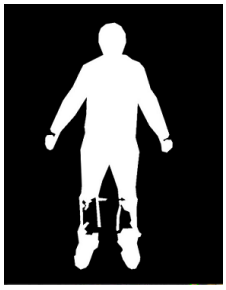


Figure 4: Merged Silhouette

of the extracted contour is a result of the interpolation of the depth sensor image.

4.3. Landmark Detection and Feature Extraction

There are two ways to extract feature points from a 2D silhouette. The approach by Lin [LW11], based on a chain code, requires good quality contours from subjects who are almost completely undressed. The contours from fully dressed people are not of good quality. Therefore the approach proposed by Kohlschuetter seems more promising [KH12]. The feature point extraction can be accomplished exactly like in his approach. Regarding the requirements one more feature point need to be specified in order to take all needed measurements. This is the lowest left point on the side image that is considered to be the lowest point of the human body. Tables 2 and 3 give an overview about body regions and the feature points that can be found there.

Body Region	Measurement	No. Feature Point
Chest	Chest width	F1, F7-F10
Hips	Hips width	F1, F7-F8
Waist	Waist width	F6-F8

Table 2: Measurements Front View

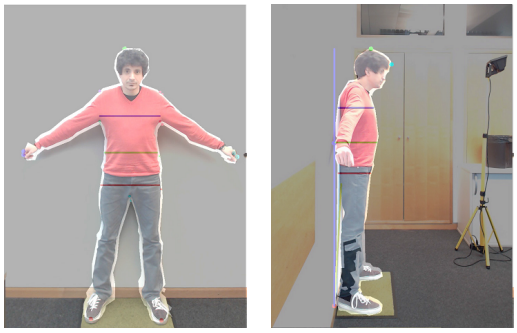
Body Region	Measurement	No. Feature Point
Chest	Chest width	F1, F7-F1, S3-S4
Hips	Hips width	F1, F7-F8, S3-S4
Waist	Waist width	F6-F8, S3-S4
Lower Body	Inner leg lenght	S2, F6
Body	Body Height	S1-S2

Table 3: Measurementes Side View

As the extracted contour of a video stream cannot always be assumed to be correct, a quality check is needed. The quality check begins after calculating the vertical extremes in the front image. The number of white pixels in the image can be compared to a threshold in relation to the distance

© The Eurographics Association 2014.

between the lower and upper extremes with a threshold. Following the Kohlschuetter approach, body dimensions got calculated. Two more measurements were calculated in addition to those as he blanked the legs out in his work. The body height is the distance between the lowest left point of the side image and the highest point of the front image. The inside leg measurement is, in this approach, the distance between the crotch in the front image and the lowest left point of the side image.



(a) (b)

Figure 5: (a) Frontal View (b) Literal View

Figures 5a and 5b illustrate the detected regions on the 2D silhouette images of measured subjects as an overlay between the silhouette and the RGB image. The encountered feature points are marked as colored dots. The white border between the human body and the background is the border of the extracted silhouette.

4.4. Results

The result of the measurement process using the webcam contains samples of four male and one female subject of different ages and figures. All subjects got measured with regular office clothes (without jackets). Every subject was guided trough the automated dimension recognition process. The lighting conditions have been different between every measurement process. Table 4 (on the next page) gives an overview over the calculated error rate grouped by measure category. This value represents the ratio of deviation to ground truth measurements. Because of the small amount of collected samples the deviation does alternate a lot between the single values. The body height recognition and the inner legs length showed the best performance (0,02566m and 0,05469m). The worst performance showed the hips and waist measurement category. Table 5 presents a comparison between the two different camera types. Unfortunately there where only three subjects with data for both approaches. The table shows the average error rate over all measurement categories for the three subjects. The results proves that the performance of the Microsoft Kinect camera is comparable to the results of the webcam. One possible reason can be found in the low resolution of both Kinect cameras. It does also sees that the Kinect depth camera considers rather pixel to be background than foreground. Another possible reason is, that errors during the measurement process occurred (e.g. pose variation).

Figure 6: Reconstructed 3D model

5. Clothes Scanning Methodologies

5.1. State of the art

There are different reconstruction and scanning techniques available. Laser scanning is used in professional reconstruction jobs including reverse engineering. Together with structured light, these two methods provide precise enough data to even reconstruct industrial parts used in engineering applications. [Tan11] Time of flight cameras recently were used in the first successful commercial product the Kinect for X-Box One. They have the ability to compute depth through the time light needs to travel a certain distance. This can be done multiple times during a second, which leads to real-time tracking abilities, motion capture and three dimensional reconstruction.

Recently light-fields are also used for reconstruction. [ESG99] They pose a new way of dealing with photographic images in contrast to classical structure from motion, which is applied as the only solution capable to capture efficiently high detail models of textured subjects. Additionally structure from motion and photogrammetry is currently the only method which enables really qualitative moving full body capture. [PS14] Research continues to grow in this field even recently enabling to capture material properties like specularity strength, color, normals and roughness and diffuse maps with normals. [TFG*13] And of course structure from motion is not limited to the visible light spectrum, which makes it also very attractive for scientific applications.

In this work we used the software Agisoft Photoscan [Agi14] which uses photogrammetry to reconstruct three dimensional meshes and textures. Agisoft photoscan seems to have many similarities with the open source software VisualSFM, however, the results are unmatched in quality and

© The Eurographics Association 2014.

Sample/ER	Chest	Hips	Waist	Inner Legs	Height
Subject 1	0,05941	0,15534	0,13187	0,0274	0,02809
Subject 2	0,07258	0,09322	0,07937	0,07813	0,03371
Subject 3	0,14423	0,0198	0,14019	0,06494	0,02186
Subject 4	0,09302	0,11382	0,05797	0,04819	0,01093
Subject 5	0,07921	0,07767	0,0989	0,05479	0,03371
Average	0,08969	0,09197	0,10166	0,05469	0,02566

Table 4: Error Rate in m by Measurement Category (Webcam)

ER/Sample	Subject 1	Subject 2	Subject 3	Average
Average Webcam	0,08042	0,0714	0,0782	0,07667
Average Kinect	0,08743	0,08141	0,08009	0,08298

Table 5: Comparison of the Average of Error Rate between Webcam and Kinect measurement

detail. As in VisualSFM a sparse reconstruction is made by matching the different images. After that a pairwise matching algorithm creates depth maps which can be used for reconstruction. The depth maps are aligned and a super dense point cloud is computed. The point cloud is meshed and textured in a similar fashion to VisualSFM. However, the quality and speed of the calculations is remarkable [Figure 9]. In the following sections we describe how the setup has to be implemented.

5.2. Lighting

The setup consists of multiple light sources with soft boxes. Three 100 W Daylight lamps were used to create a soft light without hard shadows but more light sources can be used. The subject, in this case a dress, was placed on the dress form in the middle of the light sources. The light sources were placed in 120 degree steps around the object to form 3 point lighting. Additionally fluorescent lighting was used from the top to provide additional brightness in the top regions. All lights were described as 'usable for photography with digital single lens reflex cameras'. They had a light temperature of 5700 K and the light generated filled the whole visible light spectrum. Ideally the light setup would consist of a completely white background which is lit uniformly from the back to eliminate any shadowing. A uniform indirect lighting seems to work best for image capturing and processing. Additionally the light setup would be static and not movable to simulate a complete white 360 degree high dynamic range image lighting used in current rendering software.

5.3. Camera Setup

The camera setup consists of multiple cameras which are synchronized to shoot at the same moment. Usually a camera setup of 50 cameras, or preferably more, is used. The cameras are located in a cylindrical order around the object. All Cameras have to be focused on the object and in the best case they use prime lenses. [Agi14] For an industrial implementation where multiple clothes have to be scanned per day, the full setup could prove useful and be profitable. To acquire the images with the camera, the 'manual' mode has to be used. The most important properties while acquiring images are: The focus has to be always on the subject and the subject should appear sharp in all images.

5.4. Methods of image pre-compositing

In general, the generation of masks for each image in such a way, that the only pixels relevant for reconstruction are part of the clothing, proved to be better. Especially at the generation of textures and the automation of open seams. For instance in the neck area, where the model has a very distinct smooth edge, the isolated subject model was open while the other was watertight. By providing the masks, the program was able to discard these faces automatically. [Figure 7]

5.5. Sparse reconstruction

The matching algorithm places features in three dimensional space, which produces a sparse point cloud. To find those features the image must have a certain entropy. Additionally the subject should not have big glossy and reflective surface because features are not consistent in those areas of the image. Semi-transparent surfaces also generate inconsistent features. In general reflective, refractive and transparent surfaces have to be avoided because the SIFT feature descriptors used produces features which are not consistent and therefore do not have any matches.

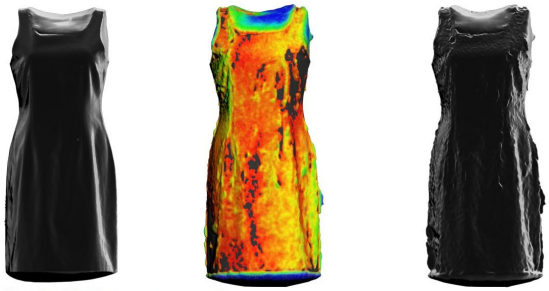
5.6. Dense calculation

In VisualSFM CMVS and PMVS2 [FCSS10] are used to calculate the dense point cloud. Agisoft Photoscan seems to

© The Eurographics Association 2014.

Figure 7: The seams are exactly at the position of the cloth seam, with a distance of about 10*cm* inwards, in a model made out of masked images (mask is marked green).

work similar however the dense point cloud and meshing work better and more accurate than in VisualSFM as seen in Figure 8. The real dress is seen in Figure 9.



```
Hausdorff Distance computed
Sampled 897466 pts (rng: 0) on mesh1.obj searched closest on
mesh2.obj
min : 0.000000 max 0.174333 mean : 0.010903 RMS : 0.017281
Values w.r.t. BBox Diag (4.014784)
min : 0.000000 max 0.043423 mean : 0.002716 RMS : 0.004304
```

Figure 8: Left: Base mesh quality (Agisoft);Middle: Hausdorff distance measurement (Quality Range: Coloration is min. (red) 0.00 max (blue) 0.05 in world units with respect to the bounding box diagonal). A rough estimation of the maximum distance is provided by measuring the bounding box. This results in a distance of $1,21cm \pm 0.4cm$. However these values may vary strongly depending on the subject and reconstruction parameters.

5.7. Model creation

Agisoft Photoscan cleans isolated points out of the point cloud in a similar fashion the erode algorithm in Meshlab

Figure 9: Reference photo of the dress

is working [Vis]. This produces a high quality mesh which can be used to project the textures.

5.8. Texture generation

The texture is projected through two methods: One is a mosaic method with hard edges between the different image projections but very high quality and sharp results. This can be used if the subject is always sharp in the images and results in a very high fidelity and details in the final textures. The second method is called "average" and produces smooth edges between texture projections. However it can not reach the sharpness of mosaic texture blending because the projected images are alpha blended, which results in overlays of details and tearing.

6. Clothes Simulation

6.1. Setup for cloth simulation (Real-time) using Unreal Development Kit

The clothes simulation with the NVidia Apex PhysX plugins and the Unreal Development Kit was not straight forward. Since the topic is not covered extensively regarding setup, implementation and especially debugging. However, since the supporting companies have shown the process, we assume a user can reconstruct the process.

6.1.1. Skin Cloth

The cloth is skinned with the help of skinning geometries. Those volumes are meshes that are an approximation of the body part they have to represent. This process is nearly automatic through the use of some scripts which cut the skinned character mesh into simple pieces. The workflow is as follows: The skinned character mesh was divided into submeshes by using the script "Cut the Puppet" [Bül13]. Those

© The Eurographics Association 2014.

sub-meshes can be optimized to be suitable for the collision simulation using a pro optimize batch script [Tho14]. Additionally a ragdoll object has to be created and assigned to the character mesh. This ragdoll object has to include all bones from the hierarchy of the cloth mesh.

6.1.2. Prerequisites

After skinning the mesh, the Apex modifier is applied. The scene with the character and all prerequisites has to be exported as an FBX-format and as Apex Project format or in this case, APX-format. After the import in the Unreal Development Kit the cloth simulation can be assigned into the slot zero in the "Skeletal Mesh/Clothing Assets". In the tab "Skeletal Mesh/Clothing Lod Map" the deepest node of the clothing Asset and assignment the ID of the Asset have to be performed. Note here that the ID has to be unique when exported out of 3DS Max. The corresponding property in 3DS Max is "Material ID". After this, the animation can be added and the simulation can be viewed in the viewport.

6.2. Result

The test system uses one NVidia GTX 470 graphics card, which was able to simulate 28.000 vertices at 20fps. The result can be viewed in: [Sam14a].

A set of over 10 apparel was acquired throughout the scanning process ranging from shoes, boots, dresses to t-shirts and hats. The main focus was set on dresses because of their complexity in animation and simulation. The results met the requirements and surpassed the expectations because the detail in the textures was so high that even print and zoomed versions of the clothes could be used in journals from different angles without retouching.

The authors are aware that they used only armless clothing. This was mainly because of financial and complexity purposes in lighting and acquiring images. However since full body scans are made, a full camera setup or a more controlled environment would make this possible. Additional results and supporting material can be found in: [Sam14b].

7. Conclusion

This work presented a full pipeline solution for 3D clothes simulation on human avatars. The solution consisted of three main sub-approaches. It started with the body dimension recognition and clothes scanning. Both parts where merged in the final step of 3D avatar simulation. The implementation and performance of the three different sub-solutions where presented and discussed in details.

A possible explanation for the better performance of the body dimension recognition for the categories body height and inner legs length compared to the other measurements in founded on the circumference formulas. While body height and legs inner length are just distances between two landmarks all other measurements are interpreted as ellipses. A revised design of the used formulas can be topic of future work. The high error rate of both approaches does have different reasons. It could be discovered that the background subtraction in the webcam approach is tendious error-prone. The different parameters such as light or low contrast between foreground and background does influence the result a lot. In the Kinect approach the resolution is too low to sustainably obtain good extracted silhouettes. In both approaches, pose variations can lead to errors in the detected landmarks. The algorithm should be developed further to detect wrongly detected landmarks. A standardization of the pose of human body by using 3D scanners could eliminate this problem. The calculated circumferences for hips, waist and chest do often exceed the real measurement. If this can be proven with more data, a standard deviation could enhance the results. Some measurement categories have not been developed during this work (e.g. under brust girth). In future work, the anthropomorphic differences between males and females should be included into the measurement process. A classification of people wearing shoes like high heels or skirts can result in better body dimension measurements.

For segmentation of the subject in clothes scanning methodologies, as described in section 5.5 there are different methods possible. However, the method proposed by Campbell et al. [CVHC11] could be well suited. Additionally acquiring point clouds through the shape of the subject and fusing those results with the results of the dense reconstruction could provide a reliable method to improve the calculation of difficult areas with low textural information which is a bottleneck for SIFT feature extraction. Another method involves projecting irregular binary pixel patterns onto the subject from four different projectors or more. This method would need two sets of images one with projected patterns for shape acquisition and another set from the same camera positions for texture data. This could also eliminate the problem of semi-transparent cloth material.

© The Eurographics Association 2014.

References

[Agi14] AGISOFT LLC: agisoft.ru, 02 2014. URL: http://downloads.agisoft.ru/pdf/Image%20Capture%20Tips%20-%20Full%20Body%20Capture.pdf. 5, 6

[BBV08] BOUWMANS T., BAF F. E., VACHON B.: Background modeling using mixture of gaussians for foreground detection Ű a survey. In *Recent Patents on Computer Science* (2008), pp. 219–237. 3

[Bea10] BEALES H.: The value of behavioral targeting. *Network Advertising Initiative* (2010). 1

[Bül13] BÜLTER C.: buelter3d, 01 2013. URL: http://buelter.freeunix.net/?p=472. 7

[Bod14] BODYMETRICS: Bodymetrics, 2014. URL: http://www.bodymetrics.com/. 2

[Can86] CANNY J.: A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell. 8*, 6 (June 1986), 679–698. 3

[Can14] CANNY EDGE DETECTOR: Canny edge detector, 2014. URL: http://docs.opencv.org/. 3

[CVHC11] CAMPBELL N. D., VOGIATZIS G., HERNÁNDEZ C., CIPOLLA R.: Automatic object segmentation from calibrated images. In *Visual Media Production (CVMP), 2011 Conference for* (2011), IEEE, pp. 126–137. 8

[Cyb14] CYBERFIT: Cyberfit, 2014. URL: http://tinyurl.com/kwsrlms. 2

[EN-14] EN-13402: En-13402, 2014. URL: http://www.textilnorm.din.de/. 2

[ESG99] EISERT P., STEINBACH E., GIROD B.: 3-D shape reconstruction from light fields using voxel back-projection. In *3-D shape reconstruction from light fields using voxel back-projection* (1999). 5

[FCSS10] FURUKAWA Y., CURLESS B., SEITZ S. M., SZELISKI R.: Towards internet-scale multi-view stereo. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on* (2010), IEEE, pp. 1434–1441. 6

[Fit14] FITNECT: Fitnect, 2014. URL: http://www.fitnect.hu/. 2

[KB01] KAEWTRAKULPONG P., BOWDEN R.: An improved adaptive background mixture model for realtime tracking with shadow detection, 2001. 3

[KH12] KOHLSCHÜTTER T., HEROUT P.: Automatic human body parts detection in a 2d anthropometric system. In *ISVC (2)* (2012), Bebis G., Boyle R., Parvin B., Koracin D., Fowlkes C., Wang S., Choi M.-H., Mantler S., Schulze J. P., Acevedo D., Mueller K., Papka M. E., (Eds.), vol. 7432 of *Lecture Notes in Computer Science*, Springer, pp. 536–544. 2, 4

[LW11] LIN Y.-L., WANG M.-J. J.: Automated body feature extraction from 2d images. *Expert Syst. Appl. 38*, 3 (Mar. 2011), 2585–2591. 2, 4

[Mak14] MAKEHUMAN TEAM: Makehuman, 2014. URL: http://www.makehuman.org/. 2

[NW97] NIEM W., WINGBERMÜHLE J.: Automatic Reconstruction of 3D Objects Using a Mobile Monoscopic Camera. In *Image and Vision Computing* (1997), pp. 173–180. 3

[PS14] PERRY-SMITH L.: Infinite realities®, 02 2014. URL: http://www.youtube.com/watch?v=PqcAJlSbKKE. 5

[ROS14] ROS.ORG: Kinect calibration, 2014. URL: http://wiki.ros.org/kinect_calibration/technical/. 3

[SA85] SUZUKI S., ABE K.: Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing 30*, 1 (1985), 32–46. 3

[Sam14a] SAMARTZIDIS T.: Supporting material - three dimensional scanning of clothes, for simulation and presentation purposes in a virtual fitting room, 02 2014. URL: http://www.youtube.com/watch?v=07DzS4l0qYg. 8

[Sam14b] SAMARTZIDIS T.: Supporting material - three dimensional scanning of clothes, for simulation and presentation purposes in a virtual fitting room, 02 2014. URL: http://www.youtube.com/watch?v=L4nqLADGAHo. 8

[SG99] STAUFFER C., GRIMSON W. E. L.: Adaptive background mixture models for real-time tracking. In *CVPR* (1999), IEEE Computer Society, pp. 2246–2252. 3

[Sze93] SZELISKI R.: Rapid octree construction from image sequences. *CVGIP: Image Underst. 58*, 1 (July 1993), 23–32. URL: http://dx.doi.org/10.1006/ciun.1993.1029, doi:10.1006/ciun.1993.1029. 3

[Tan11] TANG P.: 3d3solutions.com, 10 2011. URL: http://tinyurl.com/kzvntdc. 5

[TFG*13] TUNWATTANAPONG B., FYFFE G., GRAHAM P., BUSCH J., YU X., GHOSH A., DEBEVEC P.: Acquiring reflectance and shape from continuous spherical harmonic illumination. *ACM Trans. Graph. 32*, 4 (July 2013), 109:1–109:12. 5

[Tho14] THOMPSON G.: Gtvfx msh prooptimizeselection. www.gtvfx.com, 01 2014. URL: http://www.scriptspot.com/3ds-max/scripts/gtvfx-msh-prooptimizeselection. 8

[Unr13] UNREAL ENGINE: Unreal development kit 2013, Sept. 2013. URL: https://www.unrealengine.com/products/udk. 2

[VCWL12] VACAVANT A., CHATEAU T., WILHELM A., LEQUIÈVRE L.: A benchmark dataset for foreground/background extraction. In *ACCV 2012 / Background Models Challenge* (2012). 3

[Vis] VISUAL COMPUTING LAB ISTI - CNR: Meshlab. http://meshlab.sourceforge.net/. 7

[Ziv04] ZIVKOVIC Z.: Improved adaptive gaussian mixture model for background subtraction. In *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 2 - Volume 02* (Washington, DC, USA, 2004), ICPR '04, IEEE Computer Society, pp. 28–31. 3

[ZvdH06] ZIVKOVIC Z., VAN DER HEIJDEN F.: Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recogn. Lett. 27*, 7 (May 2006), 773–780. 3

© The Eurographics Association 2014.