# Design of a Passive System for Human Body Reconstruction in the Fashion Industry

Gianfranco FORLANI[a], Carla NARDINOCCHI*[b], Riccardo RONCELLA[a]

[a] Department Civil Engineering, University of Parma, Italy
[b] Department I. C. E. A., University of Rome 'Sapienza', Italy

## Abstract

The fashion industry is looking with interest at the progress in 3D human body modelling, to streamline and optimize the production of customized, on-demand clothes. The accuracy required for length measurements of a 3D body model is in the order of magnitude of a few millimetres. The paper presents the first results of a study for the development of a body scanning system for the apparel and the fashion industry based on photogrammetric techniques. With the objective to anticipate at least some of the practical aspects of the study, an investigation into the accuracy and completeness of the body model reconstruction as a function of camera network configuration, type of image matching algorithm and camera orientation has been performed.

**Keywords:** photogrammetry, 3D body modelling, scanning system, fashion industry

## 1. Introduction

The interest of the apparel manufacturing industry in 3D body modelling is increasing since the 1990s. In fact, after the optimization of military uniform deliveries, the industry is now interested in the market of customized civil clothing. The necessary measurements will be acquired from a 3D body model or 3D avatar. Moreover, the generation of a 3D avatar will give manufacturers the opportunity to show their customer in advance the appearance and fit of the clothes. As for classical tailor work, also for a 3D body avatar the required accuracy of body measurements is in the order of magnitude of a few millimetres.

Methods for the generation of a 3D body model can be divided in active and passive; the laser scanner (LR) and the white light scanner (WLS) methods belong to the first category, while close range photogrammetry (CRP) methods belong to the latter. WLS methods are derived from industrial applications and used in this context mainly because they are able to address the lack of a well defined texture of the skin. On the other side, photogrammetric methods, which are based on the recognition of the same points in two or more images of the body, need to produce an artificial texture of the skin in order to improve the accuracy and reliability of the measurement. A review of the recent advances in 3D full body scanning with application to the fashion and apparel industry can be found in [3].

Several commercial WLS systems are available on the market, which generate a 3D body scan for apparel, medical or other purposes. Use of photogrammetry to this aim is reported in [10]: full 3D body reconstruction is obtained by orientation and image matching of a single image sequence taken around the body; the system however is not specifically aimed at manufacturing applications. On the other hand, photogrammetry has been widely and successfully applied to face reconstruction [4,2] as well as in medical applications (see [5] for a review).

As in any photogrammetric survey for automated surface reconstruction, image orientation and image matching are the fundamental issues. Since image acquisition will most obviously take place in a dedicated fitting room, image resection or a bundle block adjustment can be used for image orientation; approaches based on structure from motion [8] would be less efficient in such cases. Indeed, any accurate reconstruction method of the human body must account for intrinsic body sway movements; this implies very short image acquisition times that can be achieved only with multiple synchronized cameras.

In the actual image matching process, a combination of feature-based and signal-based techniques might be used, the former to provide an initial approximation of the surface, the latter to densify it. But certainly the main problem a photogrammetric system needs to address is the uniform texture of human skin. As shown in [4], in order to improve the accuracy and reliability of the measured points, the matching process must be assisted by producing an artificial texture on the skin, for example projecting a random pattern.

This paper presents some preliminary results of a feasibility study for a low or medium cost photogrammetric system for 3D human body model generation. The system performance is being investigated using a virtual fit-room and synthetically generated image data, which allow to evaluate point wise the reconstruction error. The benefits expected from the system are: high user acceptance (i.e. a very user-friendly self-service system similar to the well-known passport photo cabinets); limited space requirements on the retail shop floor, high measurement accuracy; fast image capture.

The paper outline is the following: in the second paragraph of the paper a description of the system concepts will be given and all system components, such as hardware, software and fit-room will be introduced. In the third paragraph preliminary performance test on 3D model generation will be described, which are based on synthetically generated image data acquired in a virtual fit room using 3D Studio Max. The results will be discussed in the fourth paragraph where a comparison of the generated 3D models with the reference one will be presented. Finally, in the last paragraph conclusions and future developments will be discussed.

## 2. Design of a passive system for 3D body reconstruction

### 2.1 System concept

Our objective in developing a human body scanning system for the apparel and the fashion industry based on photogrammetric techniques is to obtain a length measurement accuracy of body parts around 5 mm and to keep the overall hardware costs below 15000

The main characteristics of the system should be the following:
- use of low-cost digital cameras with resolution of 1.5 - 2.0 Mpix;
- use of synchronized or quasi-synchronized cameras to allow short image acquisition times;
- image acquisition in a specially designed fitting-room;
- image texture generated by having the human model dress with a tight suite;
- threefold image coverage of each body part.

Complete 3D reconstruction in near real time, although an ultimate goal, is not yet being pursued; however, a quick assessment of the image acquisition quality to give the green-light to the customer is foreseen.

The target accuracy of 5 mm for body length measurements should be translated into an expected object point accuracy, i.e. in the main parameter driving the setup of any photogrammetric survey. For the time being, it has been decided not to try to find out a mathematical relationship between the two figures using for instance variance propagation. By using the virtual measurement environment described in section 3 we expect to be able to gain some empirical estimates; in the current very preliminary stage, a target value of 1 mm has been chosen for point accuracy, that seems reasonably appropriate.

### 2.2 System components

The main components of the photogrammetric system for human body reconstruction are the fit room, the hardware (cameras and texture generation device) and the software for image orientation, image matching and point cloud modelling.

#### 2.2.1. The fit-room

Due to the cost of retail shop floor space, a limited space occupation is mandatory. For this reason a very small fit-room has been considered. In the first step of this study a booth of 2.0x2.0 meters has been chosen, although smaller booth dimensions up to 1.5x1.5 meters are also foreseen. Photogrammetric systems have the advantage of improved accuracy when reducing object distance, even if they require, on the other hand, a larger number of cameras to cover the entire body. To accommodate for cameras, cables, hub, pc and whatever other components necessary, a fit room of circular section and 1.0 meters radius could be inserted in a surrounding booth of square section. Circular targets with diameter of 1.5 centimeters are placed on the box floor and walls in known position.

#### 2.2.2 Hardware

Lens focal length, sensor size and sensor spatial resolution are the parameters to fix for the selection of commercial cameras which would provide an object resolution below 1 millimeter from a distance of 1 meter.

There are a wide range of digital cameras available on the market. As a first step, basic camera models have been selected to obtain a 'minimum cost' system, without ruling out the use of higher resolution, high S/N ratios cameras in future developments to improve accuracy if necessary.

Cameras with a spatial resolution of 1,600 x 1,200 pixels are foreseen; to contain costs, B/W cameras will be used, since color is not expected to improve results, being the illumination and the texture generation controlled. To contain the number of cameras, a field of view of at least 60° degrees have been selected, which is achievable with sensor classes of 1/2" or 1/3" and focal lengths in the range of 4 to 6 millimeters. Cameras using the cheaper CMOS technology and USB interface have been chosen to reduce hardware costs. Finally, video cameras have been preferred to still-cameras, due to wide availability of built-in digital camera control.

There are additional factors to take into consideration when designing a large camera array like the one necessary in this project. The data acquisition process has to be efficient and safe (no data should be lost); so particular attention has to be given to the performance of data cables, hubs and PCs as well. Moreover, synchronized, or at least almost synchronized, image capture is essential for consistent 3D body reconstruction. Hardware-triggered synchronization would be the best solution for an image capture alignment among multiple cameras; nevertheless, also software–triggered synchronization can also be considered a valid solution, being applied to the image capture of a quasi-static object. Data transmission rates are not a constraint for our application as only a single or a few frames of the video stream are actually acquired.

Considering the required accuracy, the use of calibrated cameras is mandatory. A rigorous camera calibration process has to be performed before arranging the cameras in the booth. Calibration aspects will not be further discussed here.

At this stage of the study two competing solutions are being considered for texture generation of the body: using pattern projectors or providing the customer with a specially designed full-body suit, such as those used in many sports, with a suitably printed pattern.

Using projectors is definitely more expensive: at least 8 would be necessary, unless very wide angle lenses are available; psychologically, wearing a suit is better than being photographed in underwear (some may give parasite reflections); a light environment is better than a dark one for most people. Using the full-body suit on the other hand require selection of an appropriate size and careful check that no folds remains; the suit should also be disposable, which adds to system running costs.

No specific requirements for the distribution and types of light points in the fit-room have been defined yet, though their ideal location is at ceiling and floor level, to avoid interference with cameras.

### 2.2.3 Software

As far as software is concerned, four modules are envisaged: orientation, matching, modelling and measuring.

The orientation module is based on space resection algorithms and automatic target detection. Though the camera should remain undisturbed once installed, to check the overall stability it is foreseen that the images will be oriented at every use of the system. To speed up the extraction and recognition of control points, coded targets might be used; otherwise, being the scene static, the approximate image locations of each target can be stored for each camera (as well as the approximate image orientation parameters) during system set up. Using template matching the initial location of the image coordinates will be improved and transferred to the image resection algorithm.

The image matching module can apply either least squares matching (LSM) [6] or multiphoto geometrically constrained matching (MGCM) [7] as a core algorithm. The matching procedure is embedded in a multi-resolution loop and uses a parallel strategy for dense matching. Images are epipolar resampled to speed up computations. An approximate object surface is necessary to the dense matching algorithm.

In most case, the gcp and the tie points generated by image orientation are dense enough to this aim, especially if the orientation has been performed using automated methods. However in this case such points are not measured on the Region of Interest (ROI); therefore a procedure often used on structure from motion techniques is applied: image features are extracted and matched using the SURF operator [1] and the known epipolar geometry.

The modelling module will import the point clouds from the different pairs or from n-tuple of images; using the appropriate tools, robust smoothing can be applied to reduce noise and outliers. The different patches are then combined applying averaging techniques in the overlap regions. Registration can be improved using Iterative Conjugate Points (ICP) or other surface matching techniques. For the time

being, this stage is performed manually but, due to the need for short processing times a full automation is desirable.

The measuring module will import the 3D model from the previous one and will provide the basic tools to perform interactively the necessary measurements.

## 3. Preliminary performance tests based on synthetically generated data

With the objective to anticipate some practical aspects of the study, an investigation into the accuracy and completeness of the body model reconstruction has been performed, as a function of camera network configuration, of the type of image matching algorithm and of camera orientation.

To this purpose, a 3D model of a human body has been inserted into a virtual environment under 3D Studio Max™. This software allows to define and place cameras in the scene, drape texture on objects, control illumination and generate synthetic images with different noise patterns for each camera. Using such images in the surface reconstruction process, the reconstruction error can be evaluated point-wise as the distance from the reference model surface.

### 3.1 Creation of a virtual fit-room-

A square booth of 2.0 meters x 2.0 meters illuminated with 8 points of scattered light (*omnilight*) located in the 8 corners of the booth has been created. Along the walls and on the pavement, a regular grid with 20 centimeters spacing of 3D target points of known position have been depicted.

After some internet search for suitable models, a triangulated, synthetically generated 3D model [5] has been selected. Actually, some characteristics (see figure 1) make it not ideal for 3D body reconstruction. As far as the body pose is concerned, the arms will prevent the complete reconstruction of the body trunk; to some extent, occlusions will prevent full reconstruction of arms and legs. Moreover, the resolution of the model is not the highest (about 53.000 triangles) and therefore, as it is apparent from the enlargement in figure 1, the body surface shows a tessellation with many well defined edges, i.e. a much less smoothed one compare to the real thing.
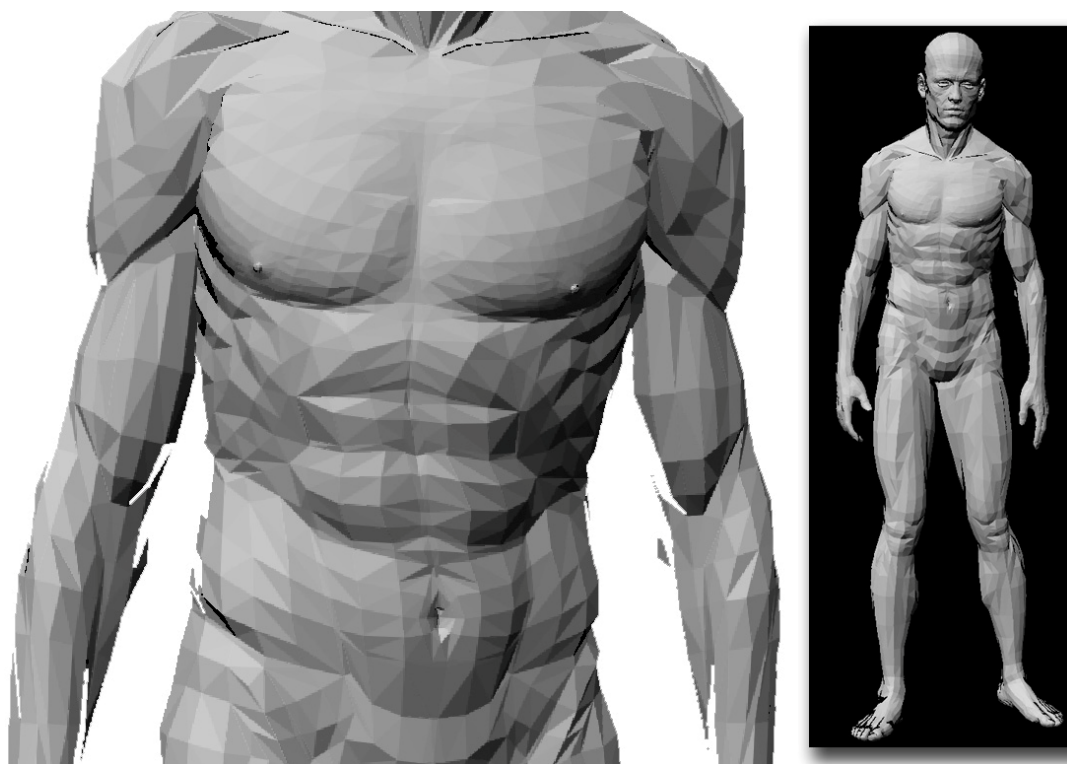


Figure 1. The synthetic human body model used in the 3D reconstruction tests.

Overall, though, the model has been considered sufficient for testing the method and therefore imported in 3D Studio Max. From the wide choice of patterns available, a cellular pattern has been chosen for the 3D body texture with an average cell size of 0.5 cm. Figure 2 shows a portion of the pattern.
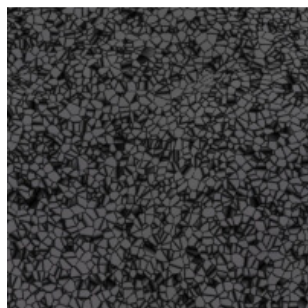
*Figure 2. Cellular pattern selected from 3D Studio Max; the average cell size is about 0.5 cm.*

Finally, a number of virtual cameras, defined by their spatial resolution, sensor size and focal length have been inserted in selected positions of the booth as explained in the next paragraph.

## 3.2 Camera network

The camera network design had as primary objective to produce with the smallest number of cameras at least a triple coverage of the object. Figure 3 shows three orthographic views of the camera network of the virtual fit-room: from two sides and from top. The light points in the corners are shown in yellow color and the cameras in blue color along the sides. The cameras (see figure 3) have been placed in triplets, at different elevation at the center of each wall of the booth. In each triplet, the baseline is about 15 centimeters; since all camera axes point to the body vertical axis, the left and right camera of each triplet are rotated by 10° degrees with respect to the central one.

Overall there are 10 triplets of cameras; more specifically, at the front and back side there are three triplets, one at a height of 0.3 meters, one at 0.9 meters and one at 1.8 meters, the last one with a 10° degrees inclination below the horizontal; on the lateral sides the triplets are placed at 0.3 meters and 1.7 meters respectively, both tilted of 10° respectively above and below the horizontal. The cameras on the left and right sides have a portrait orientation while the ones on the front and back side have a landscape orientation.
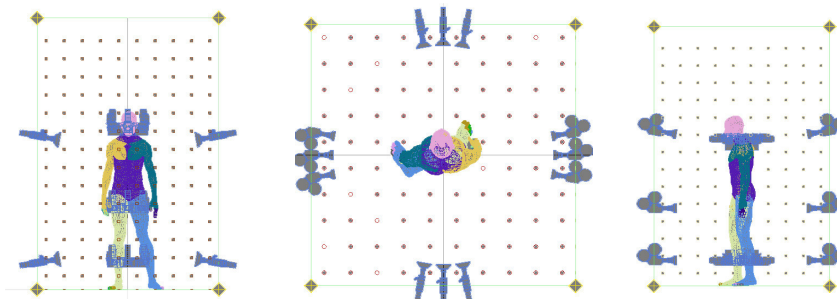


*Figure 3. Front, top and side view of the virtual fit-room.*

In the following table (Table 1) the camera parameters are shown.

*Table 1. Camera parameters set up in the 3D Studio Max environment*

| Camera Parameters | |
|---|---|
| Sensor size (mm) | 4.48 x 3.36 |
| Focal Length (mm) | 4.0 |
| Field of View (deg) | 59.5 x 46.4 |
| Spatial resolution (pixel) | 1,600 x 1,200 |
| Ground resolution on the body (mm) | 0.7 |

## 3.3 Image generation

To avoid further sources of error at this stage the images have been generated without introduction of any kind of noise or distortion.
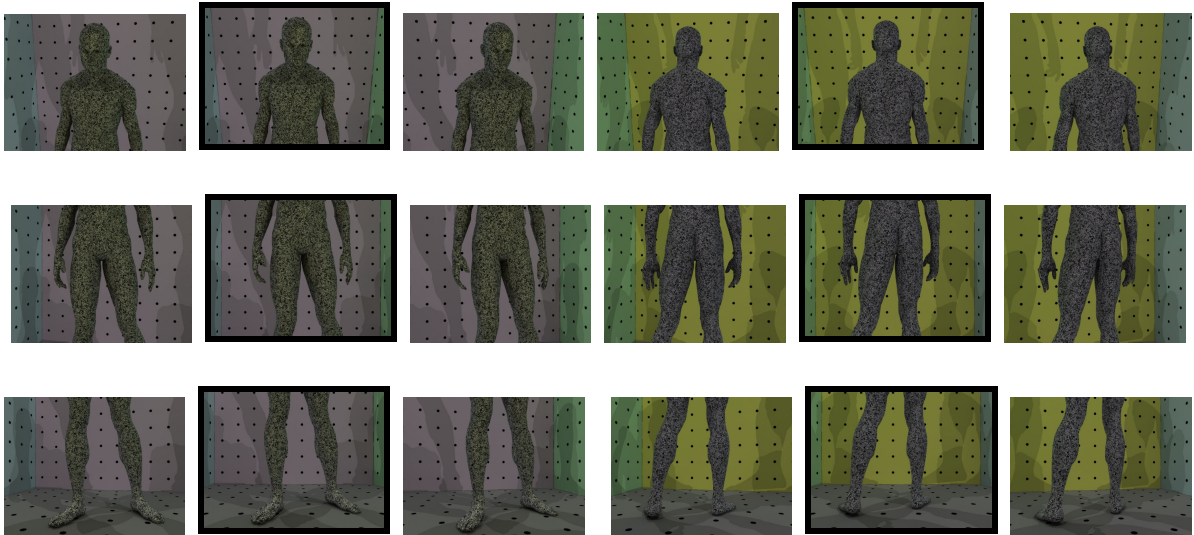
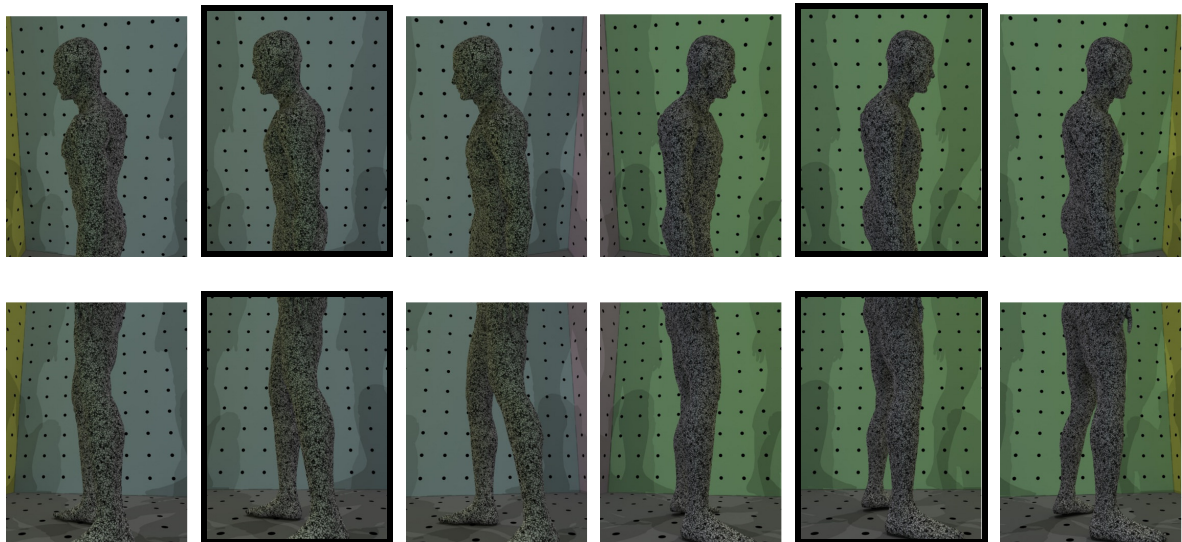*Figure 4. Images of front and back side*



*Figure 5: Images of left and right sides.*

Rendering has been applied at every camera view. Figures 4 and 5 show the generated images. The views of the cameras located along the central axis of each side are shown with a black border.

A light version of the system with fewer cameras has also been foreseen: the idea is to substitute the triplets of cameras with a single camera mounted on a frame that rotate around the booth center rapidly acquiring the three images. The mechanism would provide a strong reduction of the number of cameras (from 30 to 10) at the price of longer image acquisition times.

### 3.4 3D Model Generation

The 3D body model generation has been performed using both the LSM and MGCM algorithms. As far as LSM is concerned, the overall model has been obtained by joining 20 partial models (patches); in each triplet the central image was taken as template and matched with both the left and right, producing two separate model patches. As far as MGCM is concerned, only the body front has been processed. In this case, a single model patch has been produced from each triplet of images.

As far as the matching parameters are concerned, the acceptance threshold (i.e. the normalized correlation coefficient – NCC) has been kept in most cases to 0.9, i.e. only rather "safe" matches have

been accepted. The grid step on the reference image, that affects the spatial object resolution of the output model, has been kept to two pixels (i.e. the resolution of the image has not been fully exploited). Tests have been performed to find out the optimum value of the template size, by increasing it from 7x7 pixels to 17x17 pixels by steps of two pixels. Only the images of the body trunk front have been used to this aim, where preliminary test showed that while the upper part was generally well covered by good matches, the lower had an uneven distribution of successful matches, with several "holes". To compare the results, the point clouds obtained from each template size have been triangulated in a 3D modelling program and compared to the reference surface. Figure 6 shows the color maps of the absolute reconstruction error for three different templates, superimposed to the triangulated surface. Red correspond to errors of 1 mm while green is about half a mm and blue is less than a tenth of a mm.
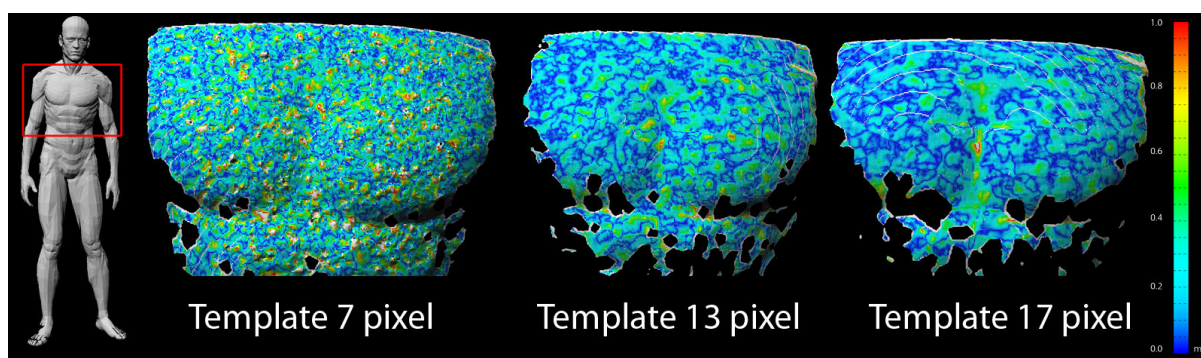


*Fig. 6. Accuracy of 3D reconstruction with different template sizes.*

As it can be expected, smaller templates produce a too rough surface while larger ones tend to smooth too much; more matches are accepted with small templates compare to larger one: this leads to less data gaps in the point cloud. As far as accuracy is concerned, the results are almost in the same range from 9x9 to 15x15 pixels, while they are worse for 7x7 and 17x17 pixels. A template size of 13x13 pixels has been eventually used with the LSM, while both a 9x9 and a 13x13 have been used with the MGCM.

As anticipated in section 2.2, the image matching procedure requires seed points that are provided by a feature based matching routine based on SURF descriptors. The procedure, recently implemented, performed satisfactorily in most image pairs; only on the ribs and on the upper part of the back very few points were detected. This made it necessary to run the matching several times, in the same area, each time adding the new seed points to increase the area covered. It has also been necessary to decrease the acceptance threshold for the NCC to 0.80 to achieve a point density similar to the one obtained in other body parts.

## 4. Comparison of the generated 3D Models to the reference

Discrepancies of every generated patches with respect to the reference model have been computed in order to derive the accuracy of the reconstruction. Assuming for sake of simplicity a normal case imaging geometry, error-free exterior orientation parameters and an image measurement accuracy of 0.25 pixel, the expected point accuracy of stereo matching restitution should be better than 1 mm, given the camera parameters and the 15 cm baselength.

### 4.1 Reconstruction by LSM
Figure 7 shows the complete model from the 20 patches, overlaid with a color error map; each patch has been first triangulated and then registered by ICP to the reference model. The accuracy has been determined by projecting the point cloud of each model patch on the reference. As can be seen from the color bar, the absolute error is for about half of the model in the order of 0.2-0.3 mm and for the majority below 1 mm. The rms error for each patch ranges from 0.3 to 0.5 mm. Some body parts are missing because of occlusions but also because of failure of the matching algorithm to find enough correlation. This can be expected for some body parts with high curvature but is less clear why this also happened in the upper part of the torso. Connection between the front and the side models is not always satisfactory.
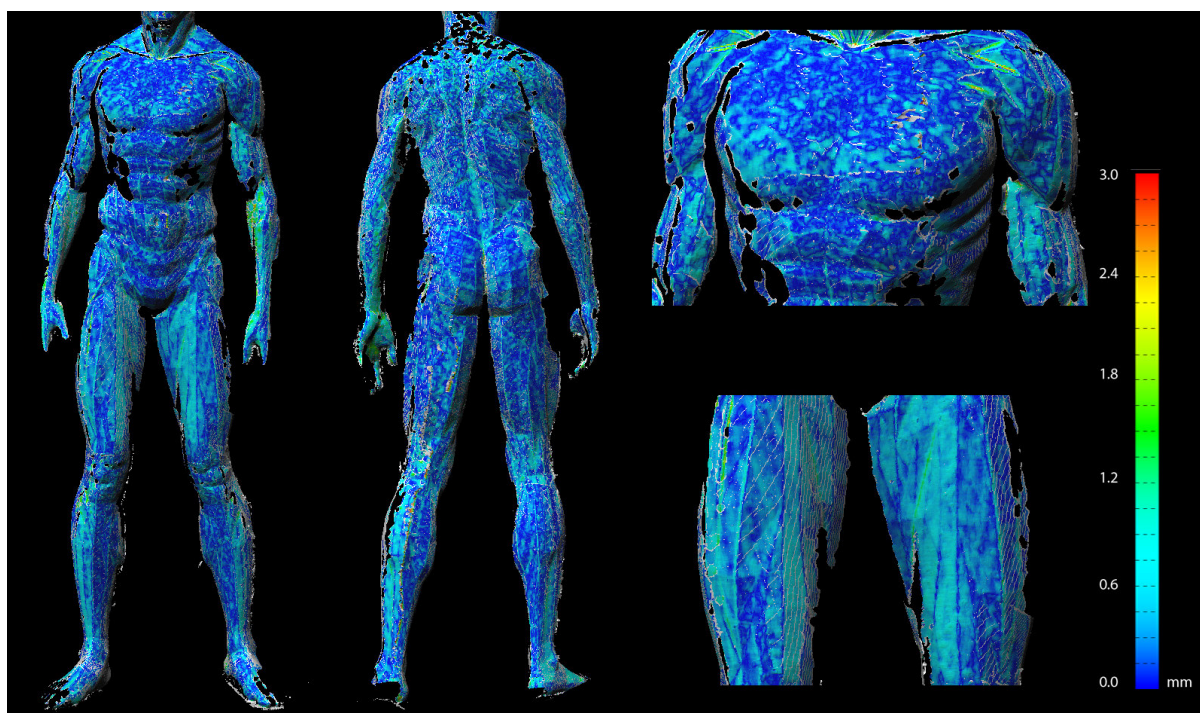
*Fig. 7. Accuracy of 3D reconstruction with LSM.*

## 4.2 Reconstruction by MGCM

The body front has also been reconstructed using the MGCM algorithm. Due to negligible overlap between different triplets, the full potential of the technique cannot be exploited, but it was interesting to find out how this approach performs with respect to LSM. From previous experiences we noticed that a denser seed point lattice is necessary with respect to LSM. Therefore the points cloud obtained by LSM was downsampled to 1 pt/cm$^2$ and used to provide seed points.

Figure 8 shows two reconstructed models of the front overlaid with the absolute error color map: the higher produced with a template size of 13 pixels, the lower with a template size of 9 pixels. With respect to the LSM, the completeness of the model is somewhat less satisfactory (a 0.9 threshold for NCC has been used), but the average accuracy is better (notice that the color bar ranges from 0 to 0.5 mm). It is also apparent that errors are concentrated (though not exclusively) along the major edges of the model (see also figure 1). Because part of this error might be due to the smoothing effect of the template size, the reconstruction has been performed with two templates size. With the 13 pixels template the rms is 0.2 mm (i.e. pretty much the same as for LSM); with the 9 pixels template the rms error is 0.15 mm, a significant improvement that has clearly been gained on the edges. On the other hand, where the original body surface is smooth, the reconstructed one looks better for the 13 pixels with respect to the 9 pixels case.
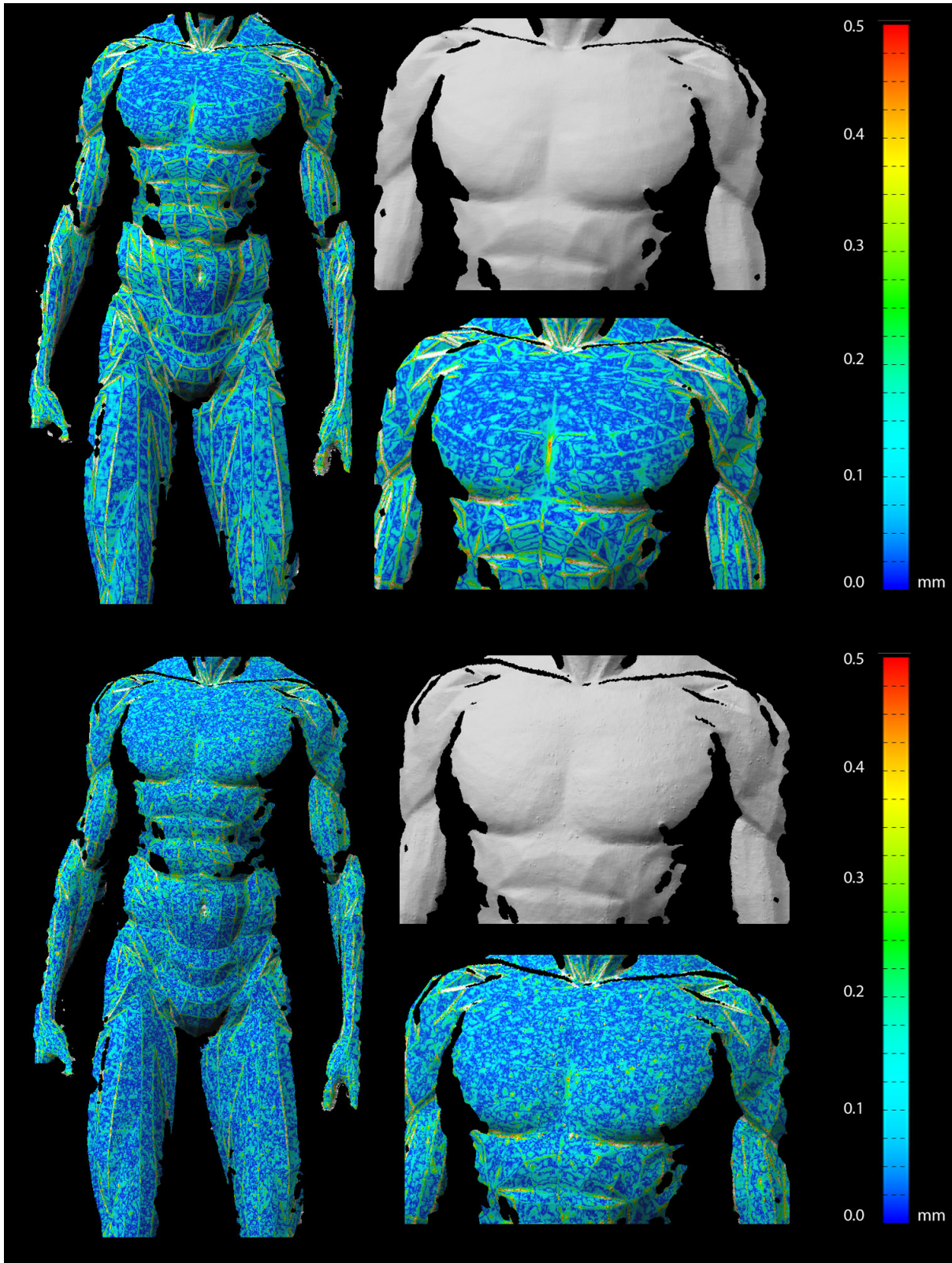
*Fig. 8. Accuracy of 3D reconstruction with MGCM using different template sizes.*

## 4.3 Empirical estimate of orientation errors on 3D reconstruction

To gain some hint of the effect of the exterior orientation accuracy on the quality of the 3D model, the body front has been reconstructed using the true exterior orientation parameters, i.e. the values imposed under 3D StudioMax when the synthetic images have been generated. As with the complete model, six model patches were produced from the three front triplets. Each model patch has been

compared to the reference model. The root mean square error turned out to be pretty much the same for each model patch, at about 0.23 mm for all. Figure 9 shows the front body reconstructed model overlaid with the color error map of the absolute error (the color scale is the same as of figure 6). It is apparent that a more complete reconstruction is achieved and also that the overall accuracy is better.
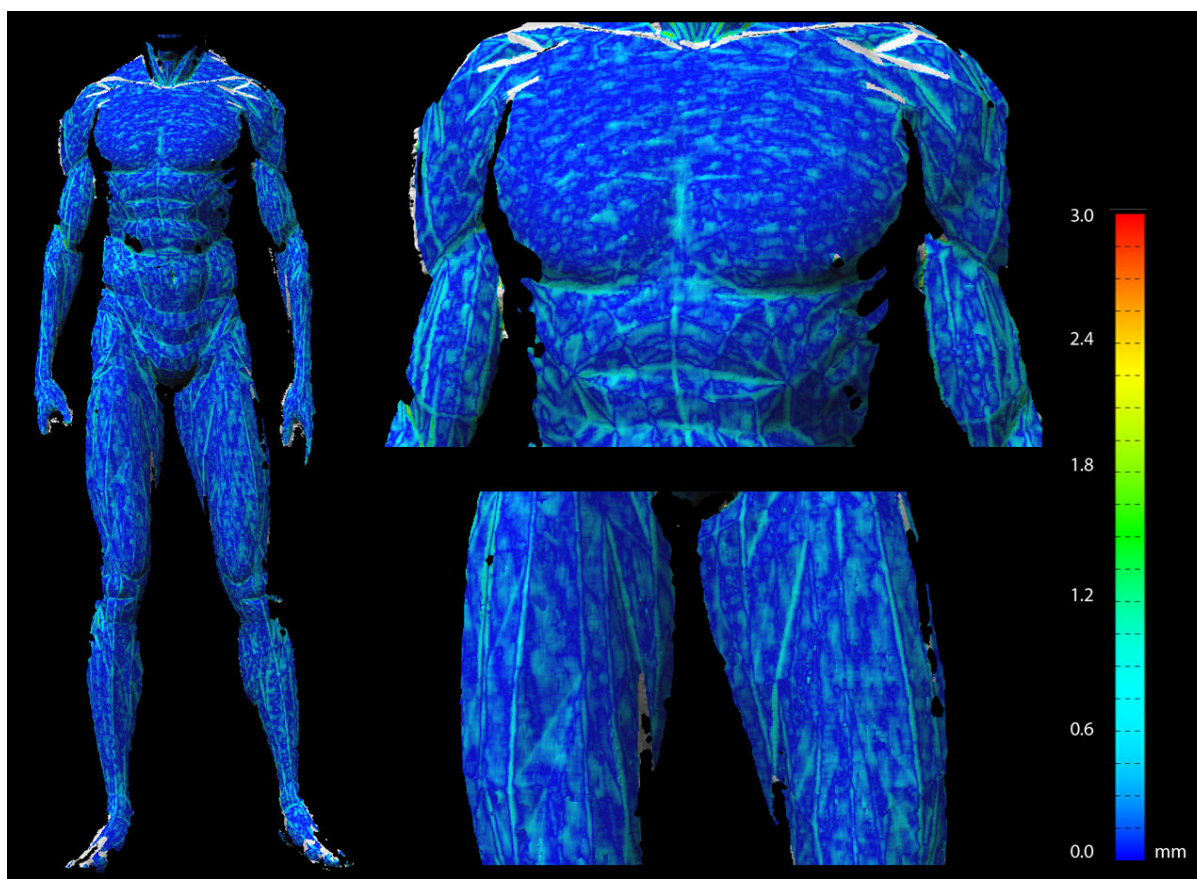


*Fig. 9. Accuracy of 3D reconstruction with error-free image orientation parameters.*

## 5. Conclusions and future development

A series of tests carried out to assess a feasibility project for the photogrammetric reconstruction of a human body has been performed using images generated in a virtual environment. The results are very preliminary and can only used as a guide to further improvement. Although the results in ideal case indicate that there is some spare accuracy potential, more simulations have to be carried out to find out the effect of residual interior and exterior orientation errors and problems with illumination. It will also be important to find out whether the camera set up is satisfactory or more camera need to be added or the imaging geometry needs changes. Shortly, preliminary tests with real images will be carried out once a suitable body suit is available.

## References

1. Bay, H., Ess, A., Tuytelaars,T., Van Gool L. (2008). "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346--359, 2008
2. Breuer P., K. I. Kim, W. Kienzle, V. Blanz, B. Schoelkopf, 2007. „Automatic 3D Face Reconstruction from Single Images or Video", Technical Report No. 160, Max Planck Institute for Biological Cybernetics.
3. D'Apuzzo, N., (2009): "Recent advances in 3d full body scanning with applications to fashion and apparel", In: Gruen, A., Kahmen, H. (Eds.), Optical 3-D Measurement Techniques IX, Vienna, Austria, 2009.

4. D'Apuzzo N., 2003. "Surface Measurement and Tracking of Human Body Parts from Multi Station Video Sequences". Ph.D. Thesis, Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland, Mitt. No.81, ISBN 3-906467-44-9, 2003.

5. Free 3D Model. (accessed 2010) http://artist-3d.com/free_3d_models/dnm/

6. Gruen A., 1985. "Adaptive least squares correlation: a powerful image matching technique." South African Journal of Photogrammetry, Remote Sensing and Cartography, Vol. 14(3), pp. 175-187.

7. Gruen A., Baltsavias E. P., 1988. "Geometrically Constrained Multiphoto Matching." Photogrammetric Engineering and Remote Sensing, Vol. 54, No. 5, pp. 633 – 641.

8. Hartley, R., Zisserman, A., 2000. "Multiple View Geometry in computer vision." Cambridge University Press, Cambridge, pp. 1-496.

9. Patias P., 2002. "Medical imaging challenges photogrammetry". In: ISPRS Journal of Photogrammetry & Remote Sensing 56 (2002) pp. 295– 310.

10. Remondino, F., (2002): "Human Body Reconstruction from Image Sequences", In Luc Van Gool (Eds.), Pattern Recognition (DAGM 2002), Lecture Notes in Computer Science 2449, Springer 2002, pp. 50-57, Zurich, Switzerland, September 2002.