IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 24, 2019, accepted November 14, 2019, date of publication November 26, 2019,
date of current version December 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2955915

A Dense Stereovision System for 3D Body Imaging

MING YAO[1] **AND BUGAO XU**[1,2]

[1]Department of Biomedical Engineering, University of Texas at Austin, Austin, TX 78712 USA (e-mail: mingyao@utexas.edu)
[2]Department of Merchandising and Digital Retailing, University of North Texas, Denton, TX 76201 USA (e-mail: Bugao.Xu@unt.edu)

Corresponding author: Bugao Xu (bugao.xu@unt.edu)

ABSTRACT Three-dimensional (3D) body imaging has become an important mean for applications based on body measurement, such as apparel customization and obesity assessment. In this article, we present a 3D body imaging system built upon stereovision technology. The system utilizes paired, high-resolution single-lens reflex (SLR) cameras to image the front and back body surfaces of a person, and robust and efficient stereo matching algorithms to reconstruct the 3D surface of the body with high-density data clouds. The system's accuracy and repeatability have been evaluated on mannequins and human subjects in comparison with other measurement methods. It was found that the geometrical measurements from reconstructed 3D body models, including body circumferences and volumes, were highly repeatable and consistent with the manual and other 3D instrumental measurements ($CV < 0.1\%$, and $R^2 > 0.99$).

INDEX TERMS Stereovision, camera calibration, 3D registration, body imaging.

I. INTRODUCTION

A 3D whole-body imaging device is an ideal tool for body dimension measurement [1]. Such a device, commonly referred to as a body scanner, captures the surface profiles of a person's exterior surface through non-contact optical techniques, and reconstructs a digital model representing the shape of the scanned person. Total and regional body volumes, as well as circumferences, depths and breadths at various body landmarks can be readily obtained from the 3D digital model, providing essential information needed in many applications, such as fitness evaluation, obesity assessment, apparel customization, and so on.

Popular technologies utilized in 3D imaging involve laserline triangulation [2], coded structured light [3], [4], and stereovision [5], [6]. A recent review of popular 3D imaging techniques can be found in [7]. A laserline scanner usually requires a mechanical device to move the laser projector to scan the object, and thus its imaging time is associated with the scanning speed. Unlike laserline scanning, both coded structured light and stereovision are static means that require no mechanical motion. The former uses an active projection to create multiple sets of stripe patterns in sequence for depth computation, while the latter utilizes cameras to take pictures of a scene from multiple perspectives and recovers depth information through stereo matching. The capability of fast scene capture makes stereovision technology a superior solution for body imaging, because it eliminates measurement errors originated from involuntary body movement during image capture. A stereovision system is passive in nature without a need for artificial lighting, and can be built with economical off-the-shelf cameras.

Thanks to the prevalence of high resolution digital cameras, the resolution of structured-light scanners has been greatly improved. The Mephisto EX (4DDynamics, Belgium) utilizes an HDTV machine vision camera with a resolution of 1920 × 1080 pixels at the 8-bit color depth as the main geometry camera [1]. An optional Canon Digital SLR (DSLR) camera is used along with the geometry camera to capture texture maps. A total of four scanner units deployed at four corners are needed for whole body coverage. A major disadvantage associated with a structured light scanner is that multiple units cannot be used simultaneously, since the light patterns from multiple sensor units interfere with each other. Thus, the multiple units have to be activated sequentially, which prolongs the image acquisition time. In order to reduce measurement errors caused by subject's movement during the scanning, a motion compensating function is included in its 3D model construction software.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhen Ren.

Based on the PrimeSense (Israel) technology, Microsoft™ released a low-cost depth sensor called Kinect in 2010. Since then, Kinect has been extensively explored in the field of body scanning. A Kinect sensor uses an infrared light pattern and achieves depth estimation through light coding [8]. [TC]^2 (Cary, NC) released a Kinect-based body scanner KX-16 which was a first whole body scanner claimed under $10,000. KX-16 uses 16 Kinect sensors for whole-body coverage. A similar system is available from Size Stream (Cary, NC), which can be configured with a choice of either 14 or 20 Kinect sensors. Other Kinect-based body scanners include Styku (Los Angeles, CA) and Bodymetrics (London, UK), both of which utilize fewer sensor unites. Because Kinect's internal calibration does not correct depth distortion introduced by the lens geometry [9], [10], Kinect has a limited accuracy which does not satisfy the needs of many applications.

Apart from active lighting, a passive 3D imaging technology-stereovision has been maturing over the past decade. The robustness and accuracy of stereovision systems are greatly benefited from the availability of high-resolution digital cameras. Canfield Scientific, Inc (Parsippany, NJ) developed a family of VECTRA® 3D imaging systems for face and partial body surface acquisition. The VECTRA H1 uses a camera with a split-optical path stereo lens for facial imaging, while the VECTRA XT uses three pairs of stereo sets for frontal upper torso imaging. A stereovision system requires a well-illuminated environment for the best image quality. The VECTRA systems are built with light panels to eliminate specular reflection on skin surfaces. A whole-body stereovision system requires multiple stereo units to be placed around the body for a full coverage. A twelve-camera body imaging system was reported in [11], in which cameras were placed on four posts around a scanned subject, and LED lightings were used to improve image quality. However, the reported system relied on a commercial software for 3D computation, thus details about the 3D computation was not discussed in the report. The Infinite-Realities (London, UK), a 3D scanning and character creation studio, was reported to have 115 Canon DSLR cameras deployed and synchronized around a scanned subject. Although a better resolution can be achieved, the cost prohibits it from large scale deployment.

Our previous work on 3D anthropometry reported how to construct a practical stereovision system for body composition assessment [12]. In that reported system, a total of four pairs of video cameras (640 × 480 pixels) were used and placed 3.6 m apart for the whole body coverage. A digital projector was used with each stereo unit to provide artificial textures on the scanned subject to facilitate stereo matching.

The study to be reported in this article was a continued effort in improving the capability of a stereovision system by utilizing high-resolution, consumer-grade cameras, and state-of-the-art stereo matching algorithms. With high-resolution images, active lighting devices were no longer needed because the richness of surface (skin or garment) textures provided adequate matching primitives for reliable stereo computation. The cameras can be placed closer to a subject due to a larger viewing angle, effectively reducing the system footprint while maintaining the same view coverage. The stereo matching algorithm used in this new system implemented a multi-scale, coarse-to-fine strategy, virtual interface concept and hybrid approach to combat the time complexity and the matching ambiguity when processing paired high-resolution images.

II. SYSTEM DESIGN

Image quality is crucial for accurate depth computation. Stereo images must be taken in a controlled lighting condition in which the scene is well illuminated and is free from specular reflection. The image should be corrected to remove lens distortion, and rectified to satisfy the epipolar geometry. This section presents the setup of our stereovision system and the calibration technique we developed in order to capture high quality pictures for dense stereo matching.

A. SETUP

Our hardware setup is illustrated in FIGURE 1. To form a stereo unit, two digital SLR cameras were fastened on an aluminum plate in a way that the optical axes of these two cameras were roughly in parallel. The baseline of the stereo unit was set to be 150 mm. Two identical stereo units were mounted on a stainless steel stand to cover one side of a body. The stand was placed 1.1 m away from the target to cover a height up to 1.9 m. Another stand with the same setup was placed at the same distance on the back side, giving the system a footprint of 2.4×1.5 m^2. The main reason why only two stands were used was the simplicity of the system, which in turn impacted its dimension, cost and maintenance. It also helped to increase the flexibility of the system for fast and easy field deployment.

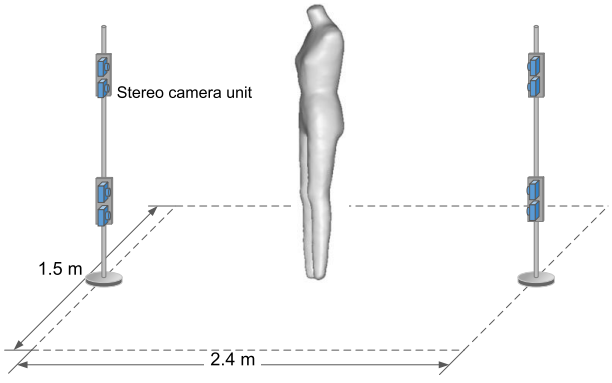


FIGURE 1. Schematic illustration of the system setup. The stereovision system consists of four stereo units, and has eight cameras in total.

The cameras used in our system were Canon EOS Rebel T3i (Canon Inc., Japan) cameras with an 18-megapixel sensor and an 18-mm focal length lens. Camera's built-in flash with a diffuser was used for image capture in a controlled environment where there is no strong ambient light (e.g., top lighting or window lighting). To avoid the interference of

multiple flashes firings at the same time, a 300 ms delay was added between the two flashes of each stereo pair, but was no delay introduced between the front units and back units. The total time of whole-body imaging was around one second. All cameras were connected to a computer via USB cables. It took about five seconds to transfer all eight stereo images back to the computer from one single shot. The total hardware cost of the entire sytsem was approximately $5000. TABLE 1 breaks down the total cost to each individual component.

TABLE 1. Cost breakdown of the proposed 3D body imaging system.

Items	Quantity	Unit price ($)
Canon DSLR Cameras	8	550
Posts and stands	2	250
USB cables	8	5

B. CALIBRATION

System calibration involved two stages: camera calibration and 3D registration. The camera calibration determined the intrinsic and extrinsic parameters of two cameras in a stereo unit, while the 3D registration calculated the pose of each stereo unit in a user-defined world coordinate system. With 3D registration, the 3D surface computed from each stereo unit can be merged into a common coordinate system.

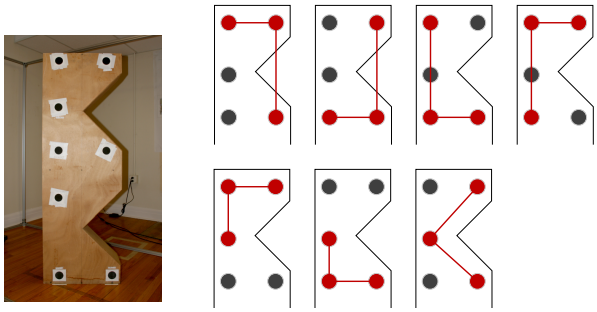
1) CAMERA AND STEREO CALIBRATION

Our camera calibration followed the technique originally proposed by Zhang [13], in which a planar chessboard pattern was used to establish feature point correspondences. The pattern included 17 $\times$ 11 blocks, and the size of each block was 40 $\times$ 40 mm. The fabrication error of the pattern was controlled under 0.2 mm. The typical calibration errors were between 0.4 and 0.8 pixel. Once each camera was calibrated individually, the stereo pair could be calibrated. All the camera calibration procedures were implemented with OpenCV.

2) 3D REGISTRATION

The 3D registration was a process to compute the transformation between the reference camera's 3D coordinate system to a common world coordinate system. This is usually done by using a 3D target with feature points that can be easily measured. Since the registration target was visible to all stereo units, the transformation matrix, containing the rotation, translation and scaling coefficients, of each stereo unit can be computed to position the units in the world coordinate system. Because this transformation did not change the Euclidean distance between the feature points, it involved rotation and translation only (no scaling) and the unit acted as a rigid body model. As demonstrated by Horn's solution to the absolute orientation problem [14], three non-collinear (i.e., coplanar) points were needed to determine this transformation.

Because locating feature points through image processing techniques was prone to errors, the registration could be inaccurate if the transformation matrix was determined only by three coplanar feature points. To increase the accuracy of the registration, we adopted a solution by running Horn's method multiple times over combinations of more feature points. As shown in FIGURE 2a, a registration target with eight circles was designed to allow the top five circles to be visible to the top stereo unit, which covered the upper body, and the bottom five circles visible to the lower stereo unit. The two circles lying in the middle were shared by both the upper and lower stereo units. The centers of the circles, i.e., the feature points, were determined via image segmentation and center-of-weight calculation. Three out of the five feature points in a unit could be selected to calculate the transformation matrix, giving 7 combinations with more global representations that allowed multiple matrixes to be averaged for an optimized solution. Some combinations of the feature points are illustrated in FIGURE 2b.



(a) The target for 3D registration.

(b) The combinations of three feature points (in red) selected out of five for each absolute orientation computation.

FIGURE 2. The 3D registration target and the feature points attached on the surface of the target.

III. DEPTH COMPUTATION

Even at a high resolution, the contrast of human skin texture was not on par with artificial patterns found in an active system. Thus, a major challenge in this work was to resolve the stereo matching ambiguities for dense depth computation. A practical stereo-matching algorithm has to deal with matching ambiguity resulted from various situations, such as inconsistent lighting, sensor noise, homogeneous or repetitive texture, and unmatchable pixels due to occlusion. A robust stereo matching strategy must be able to accommodate for all these circumstances.

With eight cameras used in our system, there were roughly 144 megapixels to be processed to generate a 3D model. High-resolution stereo matching had to deal with excessive computations and heavy memory consumption. The time complexity for an algorithm is $O(W \times H \times D)$, where W and H are the width and the height of an image, and D is the range of disparity. This also applies to the memory consumption in the algorithm. We implemented a novel strategy to improve both time and memory efficiencies, as illustrated in the following sections.

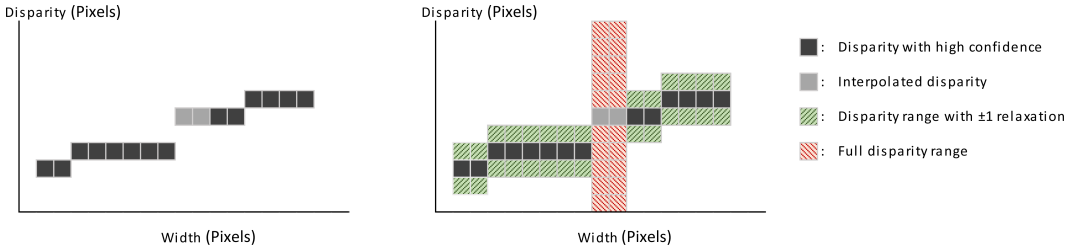


FIGURE 3. Generating disparity search ranges from the guide computed from a previous resolution scale. ***Left***: the disparities of one row of elements within a disparity map. ***Right***: the search ranges at each pixel location based on the confidence from a previous match.

A. MULTI-SCALE MATCHING

A multi-scale, coarse-to-fine strategy was implemented to address the stated challenges by assigning the matching result obtained from a previous (lower) resolution scale to be an initial guess in the next (higher) resolution scale. This strategy prevented unnecessary search for a possible match in the whole disparity space, and thus could greatly reduce the time complexity. It was also helpful for minimizing matching ambiguity because both localized textures and texture gradients at each feature point were taken into account. As a result, it enforced a non-localized optimization during the matching.

Within this multi-scale framework, an image pyramid was constructed by successive Gaussian filtering and down-sampling by a factor of two from the original images. The number of scales was configurable in the stereo matching. The criterion for selecting the number of scales was that the images at the lowest resolution still retained major textures of a body surface. In our experiments, it was chosen to be four.

Given a pyramid of stereo images, the matching started from the top of the pyramid, and was referred to as *initial matching* or *coarse match*, in which large-scale features were matched and a disparity map in the low resolution was generated. In the coarse match, a full disparity range search was performed for every pixel in the image, allowing a 3D surface to be formed at any depth within the predetermined depth-of-interest.

The disparity map computed from a lower resolution scale was then used as a guide in the next higher resolution scale to constrain the disparity search range. Thus, the subsequent matches at higher resolution scales were referred to as *guided matches*. The lower-resolution map (or, the guide) was first up-sampled by a factor of two with nearest-neighbor interpolation, and then the value of each element in the map was scaled by a factor of two. To constrain the search range for the new match, we took a strategy to differentiate pixels that were matched with a high confidence from pixels that were originally mismatched but interpolated in the previous scale. FIGURE 3 illustrates the concept of generating the disparity search range from a disparity map of the previous resolution scale. For high-confidence pixels (e.g., passed left-right check in the previous scale), their new disparity values in the current scale should be close to their estimates with errors only from the up-sampling step. Thus, a ± 1 relaxation was applied to these pixels. For those pixels that failed the left-right check in the previous scale, their disparity values were interpolated from their neighbors whose texture information was similar. Even though we applied various constraints as described in Section III-C.4, it was still possible that an interpolated disparity deviated from its true value. For these pixels, the full disparity range corresponding to the depth of the scene was assigned. In our experiments, the percentage of interpolated pixels varied across different scales. We noticed that it was common to have 2-3% of the total foreground pixels interpolated at lower resolution scale, while less than 1% of pixels interpolated at higher resolution scale. This can be explained that more robust matching can be achieved at higher resolution because of better texture.

In the coarse match, a 3D matching cost volume was computed with its base matching the size of image and its height corresponding to the disparity range. The complexity of the coarse match in big O notation is $\mathrm{O}(W_j H_j D_j)$, in which j denotes the resolution scale, W_j and H_j are the width and height of stereo images at scale j, and D_j is the disparity range at scale j. In the subsequent matches, disparity searching was only performed on the fixed ranges, i.e., [estimate $-$ 1, estimate $+$ 1] for pixels matched with a high confidence. Since these pixels were the majority of a 3D surface, the complexity of the subsequent matches decreased to $\mathrm{O}(W_j H_j)$, reducing computation by an order of magnitude. The cost volume in the subsequent matching was irregular in shape due to the variable disparity range at each pixel location. The memory footprint for the cost volume was also saved in an order of magnitude. The workflow of our multi-scale stereo matching framework is illustrated in FIGURE 4.

At each resolution scale, our stereo matching algorithm followed a four-step procedure presented in Section III-C in more details. Once the matching at the highest resolution scale was done, a sub-pixel enhancement process based on quadratic polynomial interpolation was performed to reduce the errors caused by discrete disparity steps.

B. VIRTUAL INTERFACE

To further simplify computation and to reduce matching errors, we adopted the concept of a *virtual interface* [15] for foreground and background segmentation in the disparity

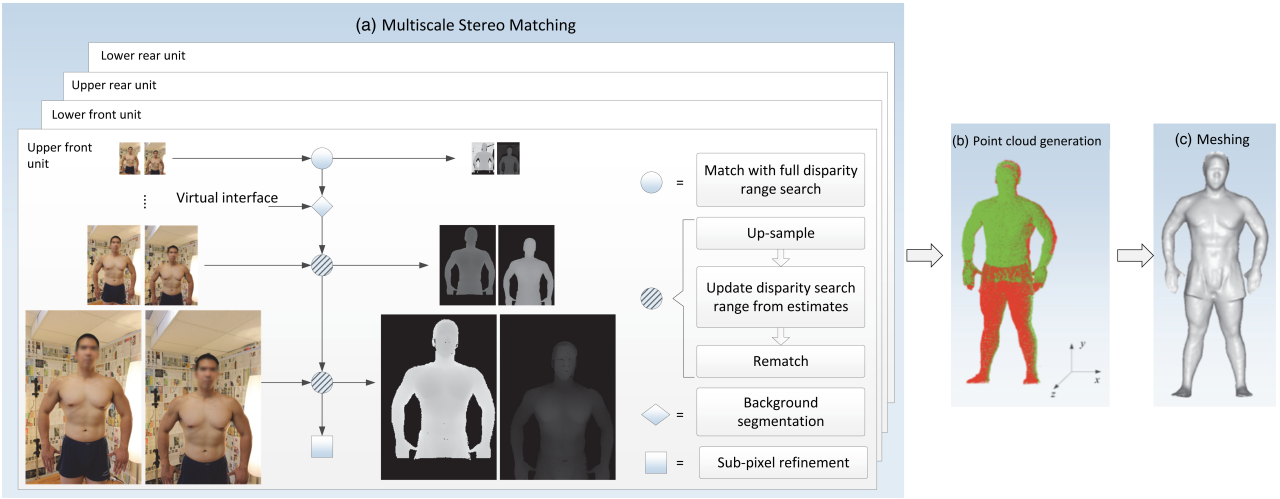


FIGURE 4. The workflow of the multi-scale stereo matching and 3D model generation. (a) A pyramid of images at different scales are shown on the left, and their corresponding disparity maps are shown on the right; (b) The dense 3D point clouds generated from the four stereo units; and (c) The final 3D mesh after mesh simplification and subdivision.

space. A virtual interface is a combination of surfaces in the disparity space that correspond to surfaces in the 3D space and segment a target from the rest of the environment. With the virtual interface, we computed and refined the depth maps that contained only the foreground, and avoided unnecessary computation on the background. To construct the virtual interface, we defined four virtual planes that were placed at the front, rear, top and bottom of the space where a person would be standing (FIGURE 5). The procedure to compute the virtual interface was detailed in [15], which addressed the virtual interface for the left image only. In order to be able to perform left-right check in the coarse match, we extended the virtual interface to the right image as well in the procedures provided below.

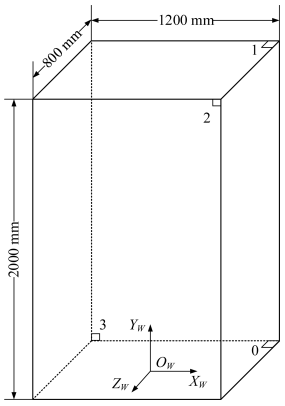


FIGURE 5. The virtual interface that defines the 3D region of interest. Four virtual planes are utilized: front (Plane 2), rear (Plane 3), top (Plane 1) and bottom (Plane 0).

The origin of the world coordinate system, O_W, is at the zero elevation of the floor, and is centered at the spot where a target stands. The positive Z_W-axis points to the frontal stereo units. To divide the 3D space into foreground and background, three of the four planes were applied in each stereo unit. For example, the bottom, top and rear planes were used for the frontal stereo units, and the bottom, top and frontal planes were used for rear stereo units.

FIGURE 6 shows a 3D plane, Π, being viewed by a pair of rectified stereo cameras configured in a parallel-axis setup. O_l and O_r are focal points, with their baseline distance being b. The normal of the plane Π is $\boldsymbol{n} = \begin{bmatrix} n_x & n_y & n_z \end{bmatrix}^T$. Without loss of generality, the plane is defined in the left camera's coordinate system with the normal being $\boldsymbol{n}$ and the perpendicular distance from the origin O_l being s. Let $\boldsymbol{X}_l$ and $\boldsymbol{X}_r$ be the left and right camera coordinates of an arbitrary point P on Π. Thus, $\boldsymbol{X}_l$ and $\boldsymbol{X}_r$ satisfies

$$\boldsymbol{X}_r = \boldsymbol{H}\boldsymbol{X}_l, \tag{1}$$

with

$$\boldsymbol{H} = \boldsymbol{R} + \frac{1}{s}\boldsymbol{t}\boldsymbol{n}^T. \tag{2}$$

$\boldsymbol{R}$ and $\boldsymbol{t}$ are the relative rotation and translation of the right camera with respect to the left camera. $\boldsymbol{H}$ is the homography

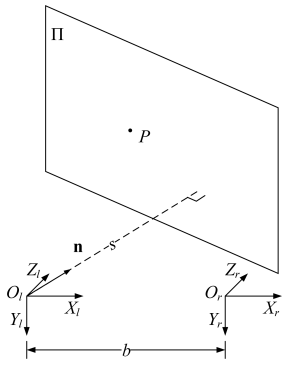
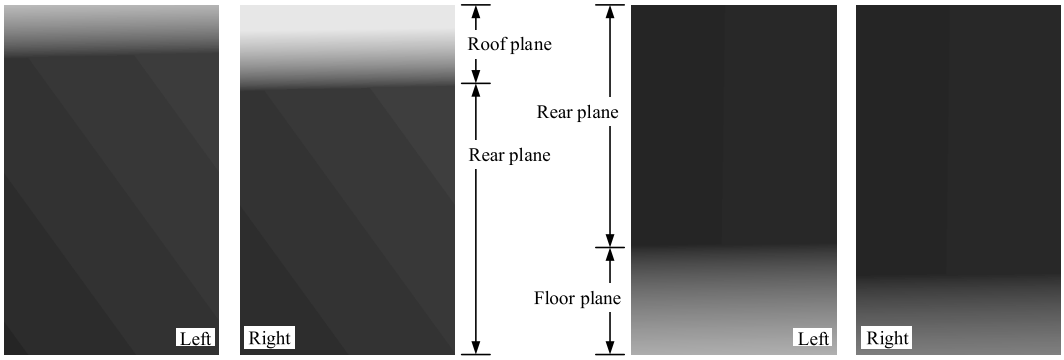


FIGURE 6. The homography that is induced by a 3D plane observed by a pair of stereo cameras.



(a) Upper frontal stereo unit. (b) Lower frontal stereo unit.

FIGURE 7. The background disparity maps computed for the two frontal stereo units. A light pixel indicates a near range, and a dark pixel indicates a far range. The roof plane and the rear plane are visible to the upper unit, while the floor plane and the rear plane are visible to the lower unit.

related with Π. Specifically, for rectified stereo geometry, $\boldsymbol{R} = \boldsymbol{I}, \boldsymbol{t} = [-b\ 0\ 0]^T$, and thus we have

$$\boldsymbol{H} = \begin{bmatrix} 1 - \frac{b}{s}n_x & -\frac{b}{s}n_y & -\frac{b}{s}n_z \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{3}$$

and its inverse

$$\boldsymbol{H}^{-1} = \begin{bmatrix} \frac{1}{1 - \frac{b}{s}n_x} & \frac{\frac{b}{s}n_y}{1 - \frac{b}{s}n_x} & \frac{\frac{b}{s}n_z}{1 - \frac{b}{s}n_x} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4}$$

Denote $\tilde{\boldsymbol{x}}_l = [x_l\ y_l\ f]^T$ and $\tilde{\boldsymbol{x}}_r = [x_r\ y_r\ f]^T$, which are the homogeneous coordinates of the images of point P in the left and right image planes, respectively. f is the focal distance. Then according to the perspective projection, we have $\lambda_l \tilde{\boldsymbol{x}}_l = \boldsymbol{X}_l$ and $\lambda_r \tilde{\boldsymbol{x}}_r = \boldsymbol{X}_r$, where λ_l and λ_r are scalar values. In addition, $\lambda_l = \lambda_r$ stands for the parallel-axis stereo geometry. Then by replacing $\boldsymbol{X}_l$ and $\boldsymbol{X}_r$ in (1), we obtain

$$\tilde{\boldsymbol{x}}_r = \boldsymbol{H}\tilde{\boldsymbol{x}}_l, \tag{5}$$

and

$$\tilde{\boldsymbol{x}}_l = \boldsymbol{H}^{-1}\tilde{\boldsymbol{x}}_r. \tag{6}$$

By combining (3) and (5) and rearrange, we can compute the disparity by

$$d_l = x_r - x_l = -\frac{b}{s}\boldsymbol{n}^t\tilde{\boldsymbol{x}}_l, \tag{7}$$

for the left reference, and similarly

$$d_r = x_l - x_r = \frac{1}{1 - \frac{b}{s}n_x}\begin{bmatrix} 1 & \frac{b}{s}n_y & \frac{b}{s}n_z \end{bmatrix}\begin{bmatrix} x_r \\ y_r \\ f \end{bmatrix} - x_r, \tag{8}$$

for the right reference.

In practice, it is easier to define the plane Π in the global world coordinate system, so it is necessary to transform it into the camera's coordinate system for background segmentation. We can obtain that

$$\boldsymbol{n} = \hat{\boldsymbol{n}}^T\boldsymbol{R}^*, \tag{9}$$

and

$$s = \hat{s} - \hat{\boldsymbol{n}}^T\boldsymbol{t}^*, \tag{10}$$

in which $\hat{\boldsymbol{n}}$ is the plane normal defined in the world coordinate system, and $\hat{s}$ is its distance to the world coordinate system origin. $\boldsymbol{R}^*$ and $\boldsymbol{t}^*$ are the camera coordinate systems' rotation and translation with respect to the world coordinate system and obtained through 3D registration.

TABLE 2 shows the four planes that defined the virtual interfaces for foreground and background segmentation. The floor plane had been slightly lifted off the ground by 2 mm to separate the body from the ground. Examples of the computed background disparity maps are shown in FIGURE 7. The grayscale values of these maps were scaled up to enhance the variations within each map. Pixels of a light gray indicate they were close to the stereo unit, while pixels of a dark gray indicate they were far away. From these disparity maps, it can be seen that the rear plane was visible to both the upper and lower stereo units, while the roof plane was only visible to the upper unit and the floor plane was only visible to the lower unit. The right camera in the upper unit covered more roof plane than the left camera, because the right camera was mounted higher in elevation than the left camera. The same

TABLE 2. Planes of virtual interface. Plane parameters are defined in the world coordinate system.

	$\hat{\boldsymbol{n}}$	$\hat{s}$ (mm)
Plane 0 (floor)	$[0\ \ 1\ \ 0]^T$	2
Plane 1 (roof)	$[0\ \ 1\ \ 0]^T$	2000
Plane 2 (front)	$[0\ \ 0\ \ 1]^T$	400
Plane 3 (rear)	$[0\ \ 0\ \ -1]^T$	400

phenomenon was observed on floor plane in the lower unit. An interesting feature revealed by these background disparity maps is that the variation of surface depth of the rear plan shows a diagonal gradient pattern in the upper unit, while the pattern in the lower unit is uniform. This was caused by the fact that our upper stereo unit had slight rotations around both the Y- and Z-axis with respect to the world coordinate system, but the lower unit had nearly zero rotation around the Z-axis.

C. DENSE STEREO MATCHING

Based on the stereo matching algorithms presented by Scharstein and Szeliski [16], our stereo matching consisted of four steps: matching cost computation, cost aggregation, disparity computation, and disparity refinement.

1) COST COMPUTATION

As a general practice, an important consideration in selecting a cost function is the size and shape of the matching window which is placed at each pixel. The robustness of match increases with window size. However, the implicit assumption, that disparities are constant within the matching window, usually does not hold. As a result, large windows may lead to matching errors or blurring surfaces. To take a trade-off, we applied a hybrid cost function that consisted of three terms: cost of normalized cross-correlation $C_{\mathrm{NCC}}(\boldsymbol{p}, d)$, cost of background-suppressed absolute difference in colors $C_{\mathrm{AD}}(\boldsymbol{p}, d)$, and cost of census transform $C_{\mathrm{CT}}(\boldsymbol{p}, d)$. The combined cost function is in the form of

$$C(\boldsymbol{p}, d) = \rho(C_{\mathrm{NCC}}, \lambda_{\mathrm{NCC}}) + \rho(C_{\mathrm{AD}}, \lambda_{\mathrm{AD}}) + \rho(C_{\mathrm{CT}}, \lambda_{\mathrm{CT}}), \tag{11}$$

where $\rho(C_{[\cdot]}, \lambda_{[\cdot]})$ is a robust function on variable $C_{[\cdot]}$:

$$\rho(C_{[\cdot]}, \lambda_{[\cdot]}) = 1 - \exp\left[-\frac{C_{[\cdot]}(\boldsymbol{p}, d)}{\lambda_{[\cdot]}}\right]. \tag{12}$$

Rather than using a matching window that is fixed in size and shape, our NCC cost was defined on an adaptive neighborhood region, in which only pixels that belonged to the same texture were selected in the NCC computation [17]. The cost of AD was a pixel-wise RGB color difference. It was leveraged by an edge-preserving bilateral filtering and was more robust to noise than its original form. The cost of census [18] had superior performance at texture boundaries, so we also included it in our cost function. The purpose of (12) is twofold: firstly, it maps different cost measures to the range [0, 1], such that (11) would not be severely biased by any one of the measures; secondly, it allows customizable control over outliers with the parameter λ. This computation was done for every pixel $\boldsymbol{p}$ of the reference image and at every possible disparity d. $C(\boldsymbol{p}, d)$ is referred to as the matching cost volume.

2) COST AGGREGATION

Pixelwise cost calculation is generally unreliable, and a wrong match may have a lower cost than a correct one, due to noise, weak texture, or surface reflection. Therefore, additional constraints should be added to enhance smoothness by penalizing changes of neighboring disparities. Our cost aggregation strategy adopted the method originally proposed by Hirschmüller [19] that utilized multiple paths around a pixel to approximate global optimization. By refining the parameters along the aggregation path, this method produced aggregated results comparable to the adaptive weight method [20], [21] with much less computation. This cost aggregation resulted in a 3D cost volume, $S(\boldsymbol{p}, d)$, in which each voxel represented the *aggregated* cost at $(\boldsymbol{p}, d)$.

3) DISPARITY COMPUTATION

Cost aggregation minimized a global energy function so that an optimal disparity map could be found by selecting a disparity value for each pixel that yielded the minimal cost at that location. This procedure was implemented with a winner-takes-all (WTA) strategy. The disparity map, D_{b}, which corresponds to the *base* image (i.e. the Left image in our case), I_{b}, was determined by selecting disparity d for each pixel $\boldsymbol{p}$ that had the lowest cost, that is

$$D_{\mathrm{b}}(\boldsymbol{p}) = \underset{d}{\operatorname{argmin}}\, S(\boldsymbol{p}, d). \tag{13}$$

The disparity map, D_{m}, which corresponds to the *match* image (i.e. the Right image), I_{m}, was generated in the similar way by taking the *match* image, I_{m}, as the reference and matched to I_{b}.

D_{b} and D_{m} were then used to determine occlusions and mismatches by performing a *left-right* consistency check. In the left-right check, if $\boldsymbol{p}$ in the base image I_{b} matched to $\boldsymbol{q}$ in the match image I_{m}, then $\boldsymbol{q}$ must also match to $\boldsymbol{p}$, that is, $D_{\mathrm{b}}(\boldsymbol{p}) = -D_{\mathrm{m}}(\boldsymbol{q})$ in the mathematical form. To take different foreshortening into account, we tolerated a disparity mismatch of up to one disparity step in our implementation. A disparity was set to be invalid (i.e., $D_{\mathrm{inv}} = 0$, which represents infinite distance in the 3D space), if $D_{\mathrm{b}}(\boldsymbol{p})$ and $-D_{\mathrm{m}}(\boldsymbol{q})$ differed by more than one:

$$D(\boldsymbol{p}) = \begin{cases} D_{\mathrm{b}}(\boldsymbol{p}), & \text{if } |D_{\mathrm{b}}(\boldsymbol{p}) + D_{\mathrm{m}}(\boldsymbol{q})| \geq 1 \\ D_{\mathrm{inv}}, & \text{otherwise} \end{cases} \tag{14}$$

The consistency check enforced the uniqueness constraint by permitting one-to-one match only. The disparity computation and consistency check required visiting each pixel at each disparity for a constant number of times, thus was linear in complexity. With the consistency check, our procedures in generating a validated disparity map was complete. A summary of the processes is given in FIGURE 8.

4) DISPARITY REFINEMENT

Even with the left-right consistency check, a disparity map could still include small areas of wrong disparities. The disparity refinement was the last stage of the stereo matching pipeline to correct any errors with such constraints as surface smoothness and color consistency. These constraints were enforced by certain rules in validating a disparity value:

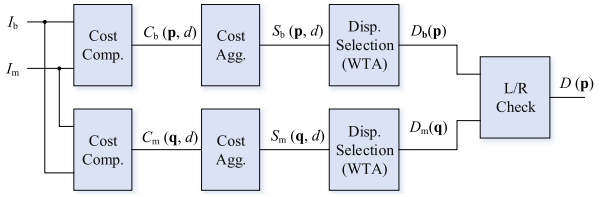


FIGURE 8. Processing steps for matching cost computation, aggregation, and disparity computation.

1) disparity discontinuity cannot occur within textureless regions, and 2) the surface of a textureless area can be approximated by a plane on which some visible texture may present elsewhere on the same plane. Firstly, the disparity refinement procedure identified the occlusion regions on a disparity map, because the occluded pixels and mismatched pixels needed to be handled differently. Then, the mismatched areas and the occluded areas underwent an iterative region voting and depth-consistent extrapolating [18], which allowed reliable disparity values from a neighborhood region to propagate into problematic areas. Finally, disparity edges were checked and made consistent to the texture map.

D. SURFACE RECONSTRUCTION

Once a disparity map is computed, a dense 3D point cloud can be generated with calibrated camera parameters. We utilized a sub-division surface algorithm [22] for 3D surface reconstruction. The basic idea of the method can be described in three steps. First, the original 3D points were re-sampled to reduce the density for 3D meshing. The explicit neighborhood information of the re-sampled data was then used to create an initial dense mesh. Secondly, the initial dense mesh was simplified by collapsing triangle edges. It produced an estimate of the control mesh. Finally, the control mesh was optimized by fitting its sub-division surfaces to the original data. The final mesh would be the reconstructed body model. In the surface reconstruction process, the upper and lower meshes from the same side of the subject were blended together at the overlapped region between the waist and hip lines, smoothening the transition between the upper and lower surfaces captured by two different stereo units. The gaps along the side of a body model due to the occlusion to cameras were closed by connecting the edges of the surface point cloud.

The original raw point cloud computed from the disparity map was at a resolution of 0.5 mm/vertex. Given that human body surface is mostly smooth and continuous, it would be unnecessary to sample the point cloud for 3D surface reconstruction at the original resolution. In addition, it would be computationally cost-prohibitive to work with the raw point cloud on the PC computer used in the project and would prolong the total processing time. Through our experiments, by down-sampling the point cloud to 2-3 mm/vertex, we did not notice a significant decrease in the final volumetric and circumference measurements. Thus, the 2mm sampling interval was utilized for all the 3D body models. The final result of the surface reconstruction was a closed triangle mesh approximating the original 3D point cloud. The number of vertices on the surface mesh was greatly reduced from the original point cloud, and they represented a simplified form of a 3D body model, which was more suitable for efficient measurements.

FIGURE 9b shows a collection of reconstructed body models for circumference and volume measurements. It can be observed that surfaces are smooth due to the re-sampling and sub-division mesh simplification. Up to this point, the complete workflow for 3D computation has been presented. FIGURE 10 sumarizes all the procedures used to generate body dimensional measurement from stereo images.

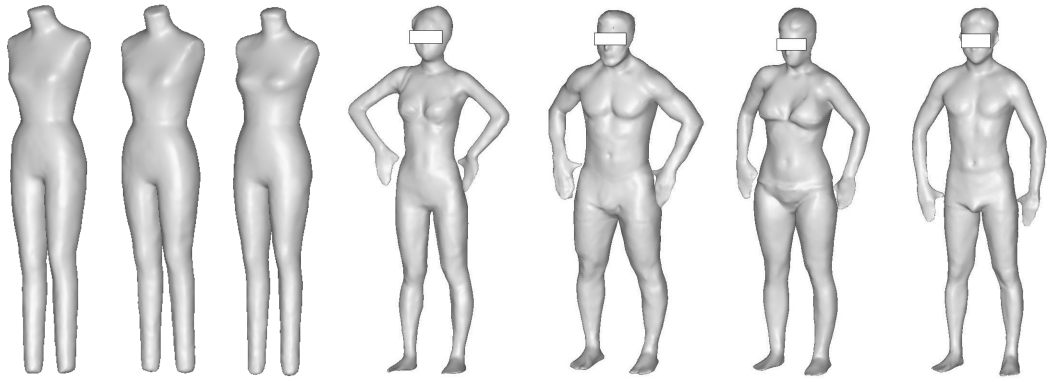
E. COMPUTATIONAL COMPLEXITY ANYLYSIS

Stereo matching is a computationally intensive task, and the complexity of a stereo matching algorithm may be an essential metric in quality and performance evaluation. Here, the complexity of each of the four components in the stereo matching framework is summarized in TABLE 3.

Direct Computation of NCC is computationally intensive due to its convolutional nature. Assume the average size of the support region is S, the computational complexity to match two images with image size of M and disparity range D is $\mathrm{O}(MSD)$. We accelerated the NCC over the adaptive matching region with the orthogonal integral image technique [18], which decreased the computational complexity to $\mathrm{O}(MD)$. In the census transform, a 63-bit binary code was computed on a 9×7 support window, and thus its complexity is $\mathrm{O}(MSD)$. The cost of color absolute difference is a pixel-wise computation and its complexity is proportional to the size of image M. However, since we applied an edge-preserving bilateral filtering in the preprocess, the overall complexity became $\mathrm{O}(MSD)$, in which S represents a 15×15 support region.

In the cost aggregation, which followed the multipath method [19], we avoided the 2D global energy minimization problem by computing the aggregated cost only along 16 pre-defined directions passing through each pixel. This reduced the complexity to $\mathrm{O}(CMD)$, in which C was a constant representing the total number of paths. The disparity computation had the $\mathrm{O}(MD)$ complexity, since it was the winner-takes-all approach. In the disparity refinement step, a histogram of disparity values in a support region was computed, and the most voted value was picked to be the new disparity. The iterative region voting was an $\mathrm{O}(CSM)$ procedure, in which C was the count of iterations and S was the support size. The procedure was set to repeat six times in our case. The depth-consistent extrapolation was an $\mathrm{O}(H)$ procedure, and its complexity was dependent on the size of disparity map holes denoted as H.

We benchmarked the performance of our stereo computation on a hex-core computer. It took 2.5 minutes to process a pair of stereo images of 5184×3456 pixels, with our multi-scale matching framework and the left-right



(a) Mannequins of size 8, 10 and 12 from left to right.

(b) Reconstructed body models of subjects with various body shapes and sizes.

FIGURE 9. Examples of reconstructed 3D body models.

TABLE 3. Complexity analysis of stereo matching.

	Cost computation					Disparity refinement	
Operations	NCC	Census trans.	Color AD	Cost agg.	Disp. comp.	Region voting	Depth-const. extraplation
Complexity	$\mathrm{O}(MD)$	$\mathrm{O}(MSD)$	$\mathrm{O}(MSD)$	$\mathrm{O}(CMD)$	$\mathrm{O}(D)$	$\mathrm{O}(CSM)$	$\mathrm{O}(H)$
Notes		S=9×7 px	S=15×15 px	C=16		C=6	

Symbols: M denotes the size of an image, S represents the support size, D is the range of disparity, C is a constant, H is the size of a hole in the disparity map.

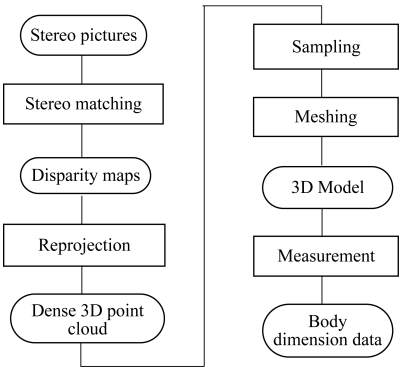


FIGURE 10. Flowchat of data in generating 3D body models out of stereo images.

consistency check. The computation time was recorded with a code level parallel acceleration technique through OpenMP. We parallelized operations including adaptive support region computation, bilateral filtering, census transform with Hamming distance, fast NCC, absolute color distance and path-wise cost aggregation. We observed a 6.5× speedup for the bilateral filtering (the most parallel-friendly operation), and a 1.7× gain on the cost aggregation (the least parallel-friendly operation). The overall speedup of the parallel computing was 2.5× of non-parallel computing on the hex-core computer.

IV. SYSTEM EVALUATION

A. EXPERIMENTS

To evaluate the accuracy and repeatability of the developed body imaging system, both industrial mannequins, whose precise body dimensions are known, and human subjects were used as objects to be imaged by the system. For each object, a 3D body model was generated by converting the 3D point cloud into a triangular mesh, from which surface circumferences at different body landmarks and the whole body volume could be readily measured.

1) MEASUREMENTS ON MANAQUINES

Three standard mannequins of different sizes (Wolf Form Company, Englewood, NJ) for apparel draping design were used for evaluation. The manufacturer-defined sizes of these mannequins are 8, 10 and 12. The reconstructed 3D models of these mannequins are shown in FIGURE 9a. Models were displayed in the same scale to show their size differences. A MyoTape body tape measure (AccuFitness LLC, Greenwood Village, CO) was used for circumferential measurements. Each mannequin was imaged five times with a repositioning each time. Chest, waist and hip circumferences, and total body volume were measured on 3D models automatically. The coefficient of variance of each measurement was computed to estimate the repeatability of such measurement. The accuracies of circumference and volume measurements were evaluated by comparing them to those obtained with the tape measure and a handheld 3D scanner, Go!SCAN (Creaform Inc., Quebec, Canada). Go!SCAN is an active lighting device and has a working depth range about 20–50 cm.

2) MEASUREMENTS OF HUMAN SUBJECTS

Twenty adults (twelve males and eight females) were recruited as human subjects for this study. A subject was

asked to wear tight-fit underwear for body imaging. During the imaging, he/she stood still in a posture with the legs slightly spread and the arms abducted from the torso. Ten sets of stereo images were captured for each subject. Since these body pictures were personally identifiable, they were encoded with a special procedure so that pixel colors in each picture were scrambled before they were saved onto the computer. Subjects' body dimensions, including chest circumference, waist circumference and hip circumference were measured and compared to MyoTape measurements.

B. STATISTICAL ANALYSIS

The repeatability of the system was evaluated by computing the coefficient of variance (CV) and the intra-class correlation coefficient (ICC) from the results of one-way random effects ANOVA. Based on the between- and within-group mean errors available from ANOVA, CV was computed as the ratio of within-group standard deviation (SDw) to the global mean, and is presented in percentage format. The ICC was determined as follows

$$ICC = \frac{MS_b - MS_w}{MS_b + (n-1) \times MS_w}, \tag{15}$$

where MS_b and MS_w are the between- and within-group mean square errors (MS), respectively. n is the number of samples per group. The comparisons of measurements using tape and stereo imaging were performed with t-tests and linear regression analysis. The comparisons were conducted on human subjects but not on the mannequins, due to limited number of samples (only three different sizes of mannequins were available).

C. RESULTS

1) EVALUATION ON MANAQUINES

The results of the repeatability test on the three mannequins were computed from ANOVA in which repeated scans of each mannequin were treated as multiple tests in one group. Each mannequin was imaged five times, and thus five tests were available for each group. TABLE 4 shows the results from the ANOVA analysis, i.e., the within-subject standard deviation (SD_w) and CV. It should be noted that the between-subject mean square errors (MS_b) and the P-value reported by ANOVA were ignored for this evaluation, because we had already known that significant differences existed among the three groups.

TABLE 4. Repeatability test on mannequins of three different sizes.

Circ. / Vol.	Mean	MS_w	SD_w	CV (%)
Chest (mm)	902.8	5.2	2.3	0.25
Waist (mm)	672.2	1.9	1.4	0.20
Hip (mm)	941.6	2.0	1.4	0.15
Volume (L)	50.789	0.0022	0.047	0.09

Symbols: MS_w, within-subject mean square error (MS); SD_w, within-subject standard deviation; CV, coefficient of variance.

The CVs for our multi-group analysis were presented as percentage values. The CVs were $\leq 0.2\%$ for waist and hip circumferences, and $< 0.1\%$ for volume. A low CV value indicates small variation in measurements. The CV increased to 0.25% for chest circumference due to the rapid variation of the circumference sampled at different vertical locations that were above or below chest line. A small change at the vertical location close to chest line could result in a larger circumference than in the waist and hip regions. This pattern has been observed on both mannequins and female human subjects, because their waist circumferences are significantly shorter than their chest circumferences.

As to measurement precision, TABLE 5 shows the comparison of circumference measurements between the stereovision and manual tape methods on the size-12 mannequin. The measurement data are presented in the format of (Mean$\pm$ SEM). The standard error of the mean (SEM) was computed by the standard deviation divided by the square root of samples, which was five for both stereo imaging and tape measure. The comparison results were generated by a paired two-sample t-test with one variable being the stereo measurement and the other being the tape measurement. Because all P-values were > 0.05, the measurements between these two methods were not considered to be significantly different. However, since the P-values were still relatively low, noticeable differences could be expected. The "Difference" column indicates that the system may result in slightly higher measurements than the tape measure.

TABLE 5. Circumferences of the size-12 mannequin measured by stereo imaging and tape.

Circ.	Stereo	Tape	Difference	P-value
Chest (mm)	925.6 ± 0.9	923.6 ± 0.4	2.0 ± 0.8	0.09
Waist (mm)	702.2 ± 0.8	699.1 ± 0.5	3.0 ± 1.1	0.06
Hip (mm)	972.3 ± 0.5	965.2 ± 0.5	2.1 ± 1.0	0.13

Measurement data are presented in (Mean $\pm$ SEM) format. The mannequin was imaged and tape measured five times. The P-values were obtained from paired two sample t-tests.

TABLE 6 shows the whole body volumes of the three mannequins measured by the stereovision and the handheld scanner. The mannequins were imaged five times by the stereovision, but they were only scanned once by the Go!SCAN because a great amount of difficulty was experienced in

TABLE 6. Whole body volumes of the three mannequins measured by stereo imaging and Go!SCAN.

Mannequins	Stereo imaging	Go!SCAN	Difference
Size 8 (L)	48.324 ± 0.018	48.077	0.247
Size 10 (L)	51.418 ± 0.026	51.138	0.280
Size 12 (L)	52.626 ± 0.009	52.349	0.277

Volumes measured by stereo imaging are presented in (Mean $\pm$ SEM) format. The mannequins were hand-scanned by Go!SCAN only one time, due to the inability to close body surface mesh on small body parts, e.g., the end of legs.

getting the complete surface of the mannequins with the Go!SCAN. It had trouble aligning body patches in some regions, such as the end of the legs, when it performed real-time data fusion. Manual editing was needed to complete non-scanned portions of body meshes. As a result, certain errors should be expected in Go!SCAN's measurements. The difference in body volume measurements between these two methods were about 0.3 L for a 50 ± 2 L body. The ratio of the difference with respect to the measurement value was around 0.6%. A t-test was not performed on these two methods because of the limited number of scans from Go!SCAN.

2) EVALUATION ON HUMAN SUBJECTS

The repeatability of the circumference measurements and body volume measurements is shown in TABLE 7. All the ICCs were > 0.99, and the CVs were $< 1.0\%$. The highest precision was recorded in the body volume which had the lowest CV. This was mainly because there was no body landmark identification in calculating the whole body volume from a 3D model. However, locating the chest, waist and hip lines could differ in different attemps. Compared to the repeatability tests of circumferences on the mannequins, the CVs of the human subject measurements were higher. This was a sign that a higher variation existed in the measurements from the multiple scans of the same human subject, partially because finding the landmark (chest, waist, etc.) precisely on a human subject was harder than on a mannequin. Subject's body moving and breathing could add errors to the measurements. Breathing affected the chest circumference measurement most, as noticed in TABLE 7.

TABLE 7. Repeatability test on 20 human subjects.

Circ. / Vol.	Mean	MS_w	MS_b	SD_w	CV	ICC
Chest (mm)	914.3	31.0	27255.6	5.6	0.61	0.9966
Waist (mm)	767.2	12.6	26409.2	3.5	0.46	0.9986
Hip (mm)	956.6	20.9	18916.2	4.5	0.48	0.9967
Raw vol. (L)	62.408	0.047	718.290	0.218	0.35	0.9998

Symbols: MS_w, within-subject mean square error (MSE); MS_b, between-subject MSE; SD_w, within-subject SD; CV, coefficient of variance; ICC, intra-class correlation coefficient.

The accuracy of the stereovision measurements with respect to the tape measurements is shown in TABLE 8. The P-value was computed by paired t-tests. The P-value for chest circumference was > 0.05, indicating that there was no significant differences between the two sets of measurements at a significance level of 0.05. However, the P-values for waist and hip circumferences were < 0.05 for the same reason as given above.

TABLE 8. Comparison of circumferences measured by stereo imaging and tape on human subjects.

Circ.	Stereo	Tape	Difference	P
Chest (mm)	914.3 ± 20.8	912.3 ± 20.4	2.0 ± 0.9	0.053
Waist (mm)	767.2 ± 20.4	764.6 ± 20.4	2.6 ± 0.9	0.012
Hip (mm)	934.4 ± 14.2	930.4 ± 13.8	4.0 ± 1.5	0.017

Measurement data are presented in (Mean $\pm$ SEM) format. The P-values were from paired-sample t-tests.

It should be noticed that the biases on circumference measurements were all positive values for human subjects. This is caused by the difference in measuring approaches between 3D body models and real bodies. On a 3D body model, the measurement was taken by fitting a curve precisely on the surface of a 3D mesh, and by tracing the exact geometrical changes along the landmark line. On the human subject, on the other hand, the tape measure usually could not touch the concave surface areas when it was tensioned. Thus, a tape measurement would be shorter than the fitted curve on the corresponding 3D body model.

V. CONCLUSION

In this article, we reported the work in developing a dense stereo-matching algorithm for a more affordable, reliable and deployable system for body imaging. Eight off-the-shelf digital SLR cameras were used to construct four stereo units, and each unit was calibrated and registered with a specially-designed target, which significantly reduced the time needed to deploy such a system for field testing. To take advantage of high-resolution stereo images for depth calculation, a multi-scale stereo matching framework was developed to accelerate depth search while retaining the quality of the depth map. The overall performance of the system was evaluated with both mannequins and human subjects. Body dimension measurements, such as chest, waist, and hip circumferences were found to be accurate and reliable.

REFERENCES

[1] H. A. M. Daanen and F. B. T. Haar, "3D whole body scanners revisited," *Displays*, vol. 34, no. 4, pp. 270–275, 2014.

[2] J. Wang, D. Gallagher, J. C. Thornton, W. Yu, M. Horlick, and F. X. Pi-Sunyer, "Validation of a 3-dimensional photonic scanner for the measurement of body volumes, dimensions, and percentage body fat," *Amer. J. Clin. Nutrition*, vol. 83, no. 4, pp. 809–816, 2006.

[3] I. Ishii, K. Yamamoto, K. Doi, and T. Tsuji, "High-speed 3D image acquisition using coded structured light projection," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct./Nov. 2007, pp. 925–930.

[4] P. Treleaven, "Sizing us up," *IEEE Spectr.*, vol. 41, no. 4, pp. 28–31, Apr. 2004.

[5] B. Xu, W. Yu, M. Yao, M. R. Pepper, and J. H. Freeland-Graves, "Three-dimensional surface imaging system for assessing human obesity," *Opt. Eng.*, vol. 48, no. 10, 2009, Art. no. 107204, doi: 10.1117/1.3250191.

[6] T. Beeler, B. Bickel, P. Beardsley, B. Sumner, and M. Gross, "High-quality single-shot capture of facial geometry," *ACM Trans. Graph.*, vol. 29, no. 4, p. 40, 2010.

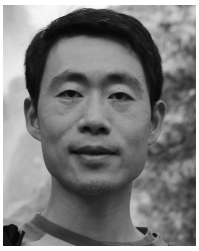
[7] M. Arbutina, D. Dragan, S. Mihic, and Z. Anisic, "Review of 3D body scanning systems," *Acta Technica Corviniensis Bull. Eng.*, vol. 10, no. 1, p. 17, 2017.

[8] P. Zanuttigh, G. Marin, C. D. Mutto, F. Dominio, L. Minto, and G. M. Cortelazzo, "Operating principles of structured light depth cameras," in *Time-of-Flight Structured Light Depth Cameras*. Cham, Switzerland: Springer, 2016, pp. 43–79.

[9] D. Herrera, J. Kannala, and J. Heikkilä, "Accurate and practical calibration of a depth and color camera pair," in *Computer Analysis of Images and Patterns*. Cham, Switzerland: Springer, 2011, pp. 437–445.

[10] C. Zhang and Z. Zhang, "Calibration between depth and color sensors for commodity depth cameras," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2011, pp. 1–6.

[11] M. Pesce, L. M. Galantucci, and F. Lavecchia, "A 12-camera body scanning system based on close-range photogrammetry for precise applications," *Virtual Phys. Prototyping*, vol. 11, no. 1, pp. 49–56, 2016.
[12] W. Yu, "Development of a three-dimensional anthropometry system for human body composition assessment," Ph.D. dissertation, Dept. Biomed. Eng., Univ. Texas, Austin, TX, USA, 2008.
[13] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.
[14] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *J. Opt. Soc. Amer, A*, vol. 4, no. 4, pp. 629–642, 1987.
[15] W. Yu and B. Xu, "A portable stereo vision system for whole body surface imaging," *Image Vis. Comput.*, vol. 28, no. 4, pp. 605–613, Apr. 2010.
[16] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, nos. 1–3, pp. 7–42, Apr. 2002.
[17] K. Zhang, J. Lu, G. Lafruit, R. Lauwereins, and L. V. Gool, "Robust stereo matching with fast normalized cross-correlation over shape-adaptive regions," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 2357–2360.
[18] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang, "On building an accurate stereo matching system on graphics hardware," in *Proc. GPUCV*, Nov. 2011, pp. 467–474.
[19] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.
[20] K.-J. Yoon and I. S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 650–656, Apr. 2006.
[21] S. Mattoccia, S. Giardino, and A. Gambini, "Accurate and efficient cost aggregation strategy for stereo correspondence based on approximated joint bilateral filtering," in *Proc. Asian Conf. Comput. Vis*. Cham, Switzerland: Springer, 2010, pp. 371–380.
[22] J. Warren and H. Weimer, *Subdivision Methods for Geometric Design: A Constructive Approach*. Amsterdam, The Netherlands: Elsevier, 2001.



MING YAO received the B.S. and M.S. degrees in communication engineering from Donghua University, Shanghai, China, in 2003 and 2006, respectively, and the Ph.D. degree in biomedical engineering from The University of Texas at Austin, TX, USA, in 2015. Since 2015, he has been a Computer Scientist in software development with Adobe Inc., San Jose, CA, USA. His research interests include image processing, computer vision, and computer graphics.

BUGAO XU received the Ph.D. degree from the University of Maryland at College Park, in 1992. He joined the Faculty of the University of Texas at Austin, in 1993. Since 2016, he has been a Professor and the Chair with the Department of Merchandising and Digital Retailing, and a Professor with the Department of Computer Science and Engineering, University of North Texas. His research interests include high-speed imaging systems, image and video processing, and 3D imaging and modeling.