Annals of Biomedical Engineering, Vol. 46, No. 8, August 2018 (© 2018) pp. 1206–1215
https://doi.org/10.1007/s10439-018-2033-7

3D Reconstruction of Scoliotic Spines from Stereoradiography and Depth Imaging

BENJAMIN GROISSER,[1] RON KIMMEL,[2] GUY FELDMAN,[3] NIMROD ROZEN,[3] and ALON WOLF[1]

[1]Department of Mechanical Engineering, Technion Israel Institute of Technology, 32000 Haifa, Israel; [2]Department of Computer Science, Technion Israel Institute of Technology, 32000 Haifa, Israel; and [3]Department of Orthopedics, Emek Medical Center, Yitshak Rabin Boulevard 21, 1834111 Afula, Israel

(*Received 25 December 2017; accepted 18 April 2018; published online 23 April 2018*)

Associate Editor Michael R. Torry oversaw the review of this article.

Abstract—Spine shape can be reconstructed from stereoradiography, but often requires specialized infrastructure or fails to account for subject posture. In this paper a protocol is presented for stereo reconstructions that integrates surface recordings with radiography and naturally accounts for variations in patient posture. Low cost depth cameras are added to an existing radiographic system to capture patient pose. A statistical model of human body shape is learned from public datasets and registered to depth scans, providing 3D correspondence across images for stereo reconstruction of radiographic landmarks. A radiographic phantom was used to validate these methods *in vitro* with RMS 3D landmark reconstruction error of 2.0 mm. Surfaces were automatically and reliably registered, with SD 12 mm translation disparity and SD .5° rotation. The proposed method is suitable for 3D radiographic reconstructions and may be beneficial in compensating for involuntary patient motion.

Keywords—Depth camera, SCAPE, Cobb angle, Stereo calibration, Stereoradiography.

INTRODUCTION

Assessment and treatment of scoliosis benefits from three dimensional spine reconstructions, as several relevant clinical indices can only be evaluated in three dimensions. Such reconstructions can be generated from Computed Tomography or Magnetic Resonance Imaging, but these scanning technologies require the patient to be supine, while clinical assessment of scoliosis is performed in load-bearing posture.[21] As such, methods for reconstructing 3D spine models from biplanar radiographs have been under development for several decades with a particular focus on posterior–anterior (PA) and lateral (LAT) imaging planes.[3]

Several stereoradiography protocols have been developed over the years to reconstruct 3D structures from planar radiographs. In all cases, stereo calibration is required to reconstruct corresponding points. Historically, this has been accomplished by including a calibration target in the radiographic field of view from which calibration parameters can be deduced. The direct linear transform (DLT) has been a popular means of calibration, as parameters are abstract and can be found with simple closed-form optimization.[15,27] However, DLT suffers from extrapolation errors requiring that the patient be surrounded by the calibration target.[36] In practice, this necessitates either a bulky cage structure[11] or a rigid vest with radiopaque markers that obscure parts of patient anatomy[12] or both. An advantage of these external structures is that they mitigate changes in pose between images; the cage can serve to stabilize the patient and the vest moves with the torso, allowing direct measurement of patient position.

An alternative to DLT calibration is explicit computation of geometric parameters, treating the radiographic system as a pinhole camera. For this approach extrinsic parameters have been initialized with a calibration target before optimization *via* reprojection error of landmarks on the spine itself.[22] More recently, it has been demonstrated that most parameters can be either measured or estimated *a priori*, potentially making online calibration unnecessary.[28,29] However, these results have only been demonstrated on synthetic data or with elaborate support structures (turntable and rigid vest). Furthermore, such statistical methods are inherently incapable of correcting for pose deformations, i.e., articulated changes in spine shape

Address correspondence to Benjamin Groisser, Department of Mechanical Engineering, Technion Israel Institute of Technology, 32000 Haifa, Israel. Electronic mail: bgroisser@campus.technion.ac.il

between radiographs. *Nota bene*: in the interest of clarity, changes in "position" shall indicate rigid body transformations while "pose" refers to subject posture, e.g. bending/twisting.

To address this issue, hybrid scan protocols have been developed incorporating an optical scanner using Moiré photography[9] in tandem with radiography. This system has been demonstrated to reduce reconstruction error *in vitro* based on optical measurements of a phantom back surface.[24] However, the system is based on a simple topographic reconstruction that relies heavily on a single projected curve to correct for position, and may be perturbed by local topographic features.

Meanwhile, the last decade has seen explosive developments in computer vision and geometric modeling. The SCAPE model,[4] introduced in 2005, has become the foundation of an increasingly sophisticated array of body models.[10,31,35] This paper seeks to improve on the hybrid scan strategy by leveraging these recently developed computer vision methods.

Here a calibration protocol is presented using off the shelf depth cameras and validated surface registration techniques to better capture patient shape and pose. To that end a BlendSCAPE[20] model is trained from publically available body scan datasets to learn statistical models of human shape and pose-dependent deformations. Training models on high quality datasets allows accurate interpretation of noisy and occluded surface data.[4,26] The result is dense correspondence between surface scans which can be used to compute extrinsic geometric parameters for stereo reconstruction. These generative models have the potential to be robust to measurement noise and local topographic variations between poses, which may provide increased accuracy in 3D reconstructions.

MATERIALS AND METHODS

Experimental Setup and Materials

Radiography for this study was performed at Elisha Hospital in Haifa, Israel on a Fuji-SCAN 3000 Computed Radiography X-ray producing 1760 × 2140 pixel grayscale images. Two circular radiopaque markers were fixed to the cassette holder at the edge of the imaged area. All images were aligned to these markers by affine transformation, thereby controlling for variations in cassette placement within the housing.

The emitter position was fixed for the entire study perpendicular to the cassette tray at a distance of about 2 m. Two Kinect for Xbox One cameras were positioned to give a clear view of a subject's dorsal surface in posterior–anterior (PA) and Lateral (LAT) imaging stance respectively (Fig. 1).

Each depth camera was controlled by a separate computer using the Kinect SDK to capture bursts of 10 frames, which were then averaged. Color, 3D position, and infrared (IR) intensity were captured for each frame at 512 × 424 pixel resolution with depth-based background subtraction. To avoid interpolated points at depth discontinuities, depth images were filtered with a Canny edge detector and edge pixels discarded. Color data was used only to mask out the green pole supporting the mannequin.

The radiographic phantom used was a zinc-coated full column spine model (C1 to sacrum) with plastic flex-and-hold internal support made by Sawbones.

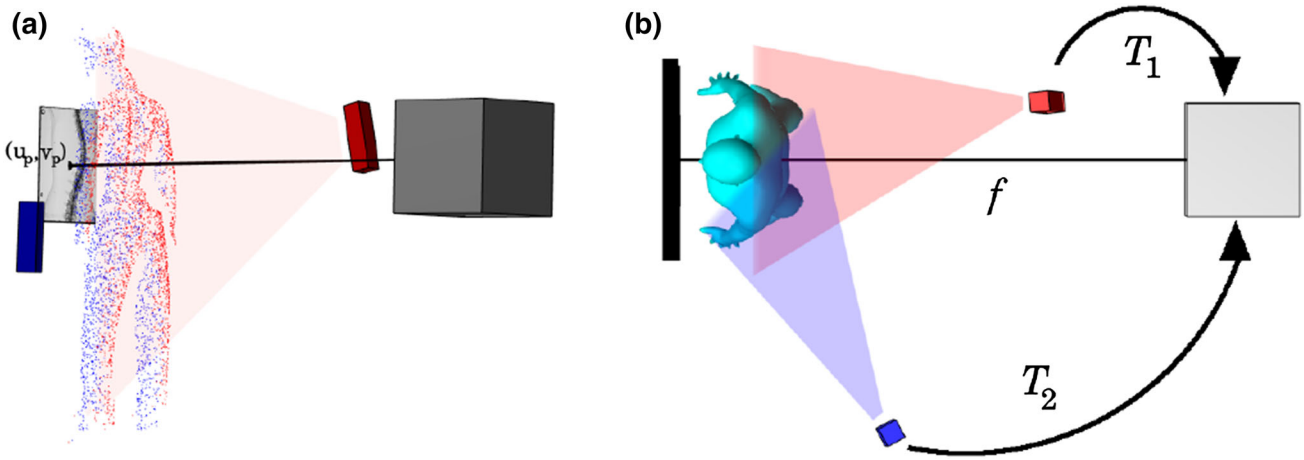

FIGURE 1. Scan room setup with two depth cameras. (a) Perspective view, rendered to scale showing X-ray emitter (gray cube) and depth cameras (red and blue). Also shown are point clouds (down sampled for visibility) and radiograph with principal point (u_p; v_p). (b) Top–down view with focal distance f labeled. T_1 and T_2 represent rigid transformations from local camera space to 3D radiographic coordinates.

This flexible model was mounted in a hollow mannequin; sacrum and cervical spine were fastened while thoracic and lumbar sections remained accessible and adjustable. Five markers were placed on points of the dorsal torso of the mannequin approximating vertebral prominans, posterior superior iliac spines, and inferior angle of the scapulas.

Digital surface meshes were prepared for each bone in the radiographic phantom. The model was CT scanned at 1 × 1 × 1 mm voxel resolution and, after manual cleanup, meshed with the Marching Cubes algorithm yielding meshes of approximately 250 vertices per vertebra. Six landmarks were manually labeled on the mesh surface of each bone: superior and inferior points on left and right pedicles as well as vertebral endplates.

Nine distinct deformations of the Sawbones phantom were imaged twice from both PA and LAT perspectives for a total of 36 radiographs. Kinect depth scans were collected concurrent with each radiographic image. Each of the nine spine curves was also photogrammetrically scanned (see "Validation" section).

All data processing was performed in MATLAB.[34]

Calibration

In contrast to the DLT technique, here intrinsic and extrinsic parameters are decomposed. Assuming a regular, oriented grid for the digitized image the projection of a point from global 3D space to homogeneous radiographic image coordinates can be expressed

$$\begin{bmatrix} wu \\ wv \\ w \end{bmatrix} = \begin{bmatrix} f/s & 0 & u_p & 0 \\ 0 & f/s & v_p & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R_{i_1} & R_{i_2} & R_{i_3} & T_{i_x} \\ R_{i_4} & R_{i_5} & R_{i_6} & T_{i_y} \\ R_{i_7} & R_{i_9} & R_{i_9} & T_{i_z} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (1)$$

where $[x\ y\ z]$ is the point in world coordinates, R_i and $\mathbf{t}_i$ are rotation and translation for camera i in world space, f is the distance from the emitter to the imaging screen, s is the sampling pitch of the sensor, $(u_p; v_p)$ is the principal point in image coordinates, u and v are the pixel coordinates in image space and w is a scaling factor.[22] Calibration consists of optimizing both intrinsic and extrinsic parameters.

Rigid body transformations can be decomposed into six degrees of freedom: a three-element rotation vector $\mathbf{w}$ that can be mapped to a 3 × 3 rotation matrix (using Rodrigues' formula[14]) and a three-element translation vector $\mathbf{t}$. Labeling the depth cameras 1 and 2 as shown in Fig. (1) there are a total of 15 calibration parameters: w_{1x}, w_{1y}, w_{1z}, t_{1x}, t_{1y}, t_{1z}, w_{2x}, w_{2y}, w_{2z}, t_{2x}, t_{2y}, t_{2z}, s, u_p, v_p. In this study these parameters were optimized in two steps: first, by finding the transformation from camera 2 coordinates to camera 1 coordinates, then solving for the remaining nine parameters with a local optimizer.

A calibration target was constructed consisting of a checkerboard pattern affixed to an acrylic plate with embedded radiopaque markers at known coordinates. Simultaneous images of the checkerboard were taken with both depth cameras in 10 mutually visible positions. Corners were automatically detected on the IR image with built-in MATLAB routines and the corresponding 3D position obtained. Rotation and translation between cameras was established with Singular Value Decomposition (SVD).[5]

Concurrently with depth imaging, ten radiographs of the checkerboard target were collected. The 3D locations of the radiopaque markers were projected from camera space into radiographic image coordinates using Eq. (1). Initial parameters, including the position of camera 1 in global coordinates, were estimated based on measurements of the imaging room. The remaining nine calibration parameters were optimized by minimizing marker reprojection error with MATLAB's nonlinear least squares solver LSQNONLIN.

Body Modeling

For this study a BlendSCAPE model was learned using the published coregistration method.[20] The SCAPE model and its derivatives use a triangulated mesh to represent human surfaces. Triangle edges are expressed as three element vectors and deformed by a series of transformations: articulated joint motion $R(\theta)$, body shape D for each subject p, and pose-dependent deformations $Q(\theta)$. For each triangle f these edge deformations can be modeled as

$$T_f = R_f(\theta) D_f(p) Q_f(\theta) T_f^* \quad (2)$$

where T^* represents the template mesh. Posed vertex locations are then found by solving the Procrustes equation to minimize reconstructed edge error. Details on SCAPE and coregistration can be found in previously published work.[4,20,30] Briefly, this model allows human surfaces to be represented with a small number of parameters. Body shape (D) and pose (θ) are decomposed so that all subjects are modeled as undergoing similar pose-dependent deformations (Q). In practice this approximation is tolerable, especially in upright posture with minimal pose variation as in this clinical application.

To learn the model parameters, several publicly available datasets of full-body scans were assembled: the SCAPE dataset,[4] the MPI scans published by

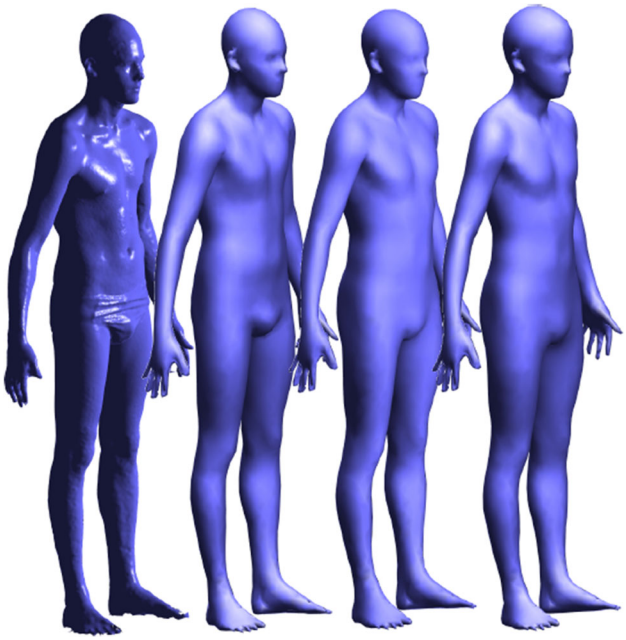

FIGURE 2. Low dimensional body modeling. Sample subject from the FAUST dataset shows how high- dimensional surface scans can be represented with parametric models. Far left: high resolution surface scan (about 175,000 points). Middle left: BlendSCAPE model after coregistration (90,504 shape parameters). Middle right: PCA representation with five principal components. Far right: semantic encoding using four anthropomorphic parameters. Reproduced with Permission from Ref. 8.

Hasler *et al.*,[19] and the FAUST dataset.[8] A 5030-vertex template mesh was generated in standard "T" pose, then registered (following protocol in "Surface Registration" section, omitting the genetic search) to a single scan from each dataset. This provided the initial mesh correspondence for each dataset. Poses for scans not provided with a registered mesh (FAUST test scans) were manually initialized (Fig. 2).

In total 890 scans of 121 subjects were coregistered to learn the BlendSCAPE parameters. Subjects' shape parameters D were used to form a low-dimensional PCA space of human shape.[4] The first five principal components were used as a basis to compute a linear regression[2] to satisfy

$$\mathbf{M}[f_1 \ldots f_l \; 1]^T = \beta \tag{3}$$

where the linear mapping $\mathbf{M}$ converts semantic features [**f** 1] to PCA weights β, which can then be used to recover shape parameters D. Semantic features used in this protocol were height, waist circumference, cube root of volume, and sex (manually labeled). Taking inspiration from multi-scale registration methods[7] a low resolution mesh (629 vertices) was generated in Blender[13] corresponding to the full resolution mesh, and another BlendSCAPE model trained for this template.

Surface Registration

Surface registrations proceed in stages. Joint angles θ are initialized as a modified "A" pose (relaxed arms) while subject shape D is initialized with semantic features. Gross initial rigid body registration is facilitated by automatically detected surface markers. A subset of joint angles are then optimized with a genetic algorithm to minimize

$$\begin{aligned}\arg\min_{\theta_{\mathrm{GA}}} &\frac{1}{|S|}\sum_{x_s\in \mathrm{T}} \rho\left(\min_{x_t\in \mathrm{T}}\|x_\mathrm{s}-x_\mathrm{t}\|\right)\\ &+\frac{1}{|V|}\sum_{x_v\in \mathrm{P}}\min_{x_p\in \mathrm{P}}\|x_\mathrm{p}-x_\mathrm{v}\|^2\end{aligned} \tag{4}$$

where S is the scan, T is the registered mesh, V are vertices on the soles of the feet, and P is the plane of the floor. θ_{GA} is an 18-element pose vector containing global rotation and translation as well as three joint angles for each shoulder and thigh segment. The first error term penalizes the Geman-McClure[18] distance $\rho_{(x)} = x^2/(x^2 + g^2)$ between scan points and the posed mesh while the second keeps the mesh in proper position relative to the floor. Robust distance g was set to 0.4m to include all relevant scan points. The genetic algorithm is computationally expensive and is not run to convergence, merely to find acceptable local optima. Next two nonlinear optimizations are performed over pose θ and body shape β using automatically generated gradients:

$$\begin{aligned}\arg\min_{\theta,\beta} &\frac{1}{|S|}\sum_{x_s\in \mathrm{S}} \rho\left(\min_{x_t\in T}\|x_\mathrm{s}-x_\mathrm{t}\|\right)\\ &+\frac{1}{|V|}\sum_{x_v\in V}\min_{x_p\in \mathrm{P}}\|x_\mathrm{p}-x_\mathrm{v}\|^2\\ &+\lambda_\theta\sum_{\theta_j\in\theta}\left\|\theta_j-\theta_{A_j}\right\|^2+\lambda_\beta\sum_{\beta_t\in\beta}\left\|\beta_\mathrm{t}/\sigma_\mathrm{t}\right\|^2\end{aligned} \tag{5}$$

Here, in addition to the scan term and floor offset term, penalties are included for deviations from reference pose θ_{A} and Mahalanobis distance from mean body shape. Equation (5) is optimized first using the low resolution BlendSCAPE model; resulting parameters then initialize another search with the full resolution model. For both search resolutions $\lambda_\theta = 0.1$, and $\lambda_\beta = 0.01$ while the robust distance was reduced from $g = 0.2$ m to $g = 0.1$ m. Finally, the deformed mesh is used to initialize a nonrigid mesh registration

$$\begin{aligned}\arg\min_{v} &\frac{1}{|S|}\sum_{x_S\in S} \rho\left(\min_{x_t\in \mathrm{T}}\|x_\mathrm{s}-x_\mathrm{t}\|\right)+\frac{\lambda_L}{L}\sum_{l\in L}\|x_\mathrm{l}-x_{\mathrm{T_l}}\|^2\\ &+\lambda_C\sum_{f\in F}a_\mathrm{f}\left\|T_\mathrm{f}-R_\mathrm{f}(\theta)D_\mathrm{f}Q_\mathrm{f}(\theta)T_\mathrm{f}^*\right\|_\mathrm{F}^2\end{aligned} \tag{6}$$

Here vertex locations $\mathbf{v}$ are optimized using analytically computed gradients to minimize the robust surface-to-scan distance, the squared distance between markers L and the corresponding points on the deformed mesh x_T, as well as squared Frobinius distance between model edges and registered mesh edges scaled by face area a_f. Constants were dynamically scaled during registration[30] converging at $g = 0.03$ m, $\lambda_L = 2$, $\lambda_C = 5$. This heavy regularization towards the model helps reduce the influence of noise from low-cost depth cameras, and works because fine detail is not required for this application. With regards to notation, in the last term T_f represents two edge vectors of the face f.[33] Landmark locations on the mesh T_l is also iteratively updated to account for placement error.[25]

Similar to updated landmarks, adaptive landmarking was used to align the center line of the mesh with the "symmetry line" of the scan. For each scan an "asymmetry map" was prepared according to Hierholzer and Drerup and the 3D curve of least asymmetry found,[16] which correlates closely with vertebral processes in scoliotic subjects.[17] For each midline vertex, the closest point on this curve was used as a landmark, updated iteratively.

The final result of this multi-resolution search is a registered mesh for each subject scan, regularized by statistical body models and anchored to manual and automatic landmarks in the region of interest. It should be noted that manual landmarks are primarily important for search initialization. The final registration is driven by topographical features which are robust to skin motion and minor variations in posture[17] (Fig. 3).

Spine Reconstruction

Stereocorresponding points were marked in each radiograph from T4 to L4 inclusive (vertebrae visible in all radiographs). Six landmarks were used for each vertebra:[6] superior and inferior points on left and right pedicles as well as vertebral endplates.

Each point in radiographic image coordinates corresponds to a line in 3D space. Reconstruction of stereocorresponding points is possible using Eq. (1) if the transformations T_1 and T_2 are known for the subject during PA and LAT scans respectively. Stereoreconstructions were performed using two protocols:

a. A transformation from LAT to PA position was found for each vertebra in the reconstruction, using the mesh registrations described in "Surface Registration" section. The six landmarks on the PA radiograph were averaged and back projected into 3D space. Vertices on the dorsal surface of the registered mesh within 10 cm of the projected line were used to find a rigid registration between LAT and PA scans.

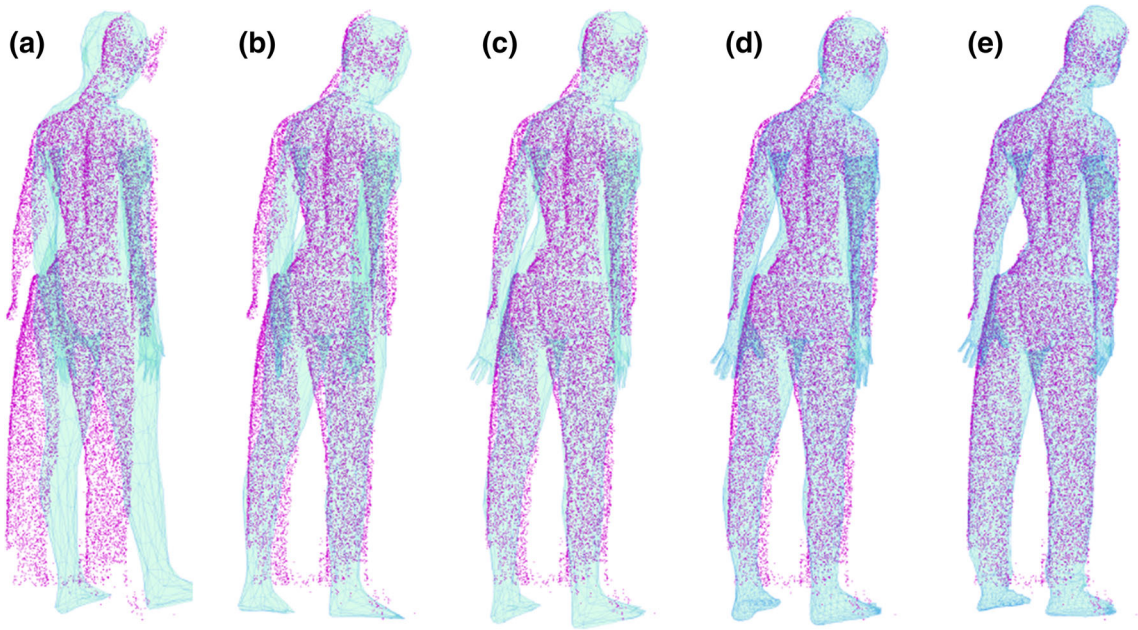

FIGURE 3. Coarse-to-fine registration of template mesh (blue) to scan points (magenta, down sampled for visibility). (a) Initialization with semantic shape parameters and rigid landmark alignment (b) Genetic Algorithm optimization of reduced pose parameters θ_{GA} (c) Gradient descent pose and shape optimization using low-polygon mesh (d) Gradient descent with full-resolution mesh (e) Nonrigid registration regularized with BlendSCAPE model.

b. The average translation **t** and z-axis rotation w_z for the dorsal vertices were found across all PA scans, and then across all LAT scans. The same intrinsic parameters (focal point, focal distance) were used as from method (a). This reconstruction provided a comparison with previously published methods that use a statistical approximation of patient position.[28]

Using the mesh spine model derived from CT, full spine reconstructions were generated from landmark reconstructions. Each bone in the mesh was aligned to minimize squared landmark distances. Full spine reconstructions were used for 3D visualization and to generate synthetic radiographs.

Validation

In order to validate spine reconstructions, a ground truth spine registration was obtained from an optical surface scan of the Sawbones model. For each spine configuration, 30–40 RGB images of the phantom were captured with a Canon T5 DSLR. Photogrammetry software (PhotoScan[1]) was used to reconstruct point clouds of the model surface. The meshed CT model was then registered to the point cloud using piecewise rigid transformations i.e., each bone is given six degrees of freedom. After landmark-based initialization, the registration proceeds with iterative closest point matching to optimize translation T and rotation R for each bone

$$\arg\min_{\mathrm{T,R}} \frac{1}{|S|}\sum_{x_S\in S}\rho\left(\min_{x_m\in \mathrm{M}}\|x_\mathrm{s}-x_\mathrm{m}\|\right) + \lambda_N\sum_{j=2}^{K}\left\|\left(T_j-T_{j-1}\right)-\left(T_j^*-T_{j-1}^*\right)\right\|^2 \quad (7)$$

The first term is the robust distance from the scan S to the model surface M, while the second term regularizes the registration by penalizing deviations from the relative template pose T^* (neutral spine) over K bones. The weight term λ_N is brought to zero as the registration converges. Radiographic reconstructions can then be compared against this registered model according to various clinical indices (Fig. 4).

Point-to-point landmark error was found by rigidly aligning all 72 reconstructed 3D landmarks with matching landmarks from the ground truth spine before computing Euclidean distances. Synthetic PA radiographs were also prepared by densely sampling the registered spine mesh model and projecting points into radiographic image space, whereon traditional Cobb angles were measured by a clinician. Computer Cobb angle[32] was computed by interpolating a continuous curve through vertebral bodies and finding the maximal angle between lines normal to the projected curve. Finally, spine length was computed by summing distances between vertebral bodies (from T4 to L4).

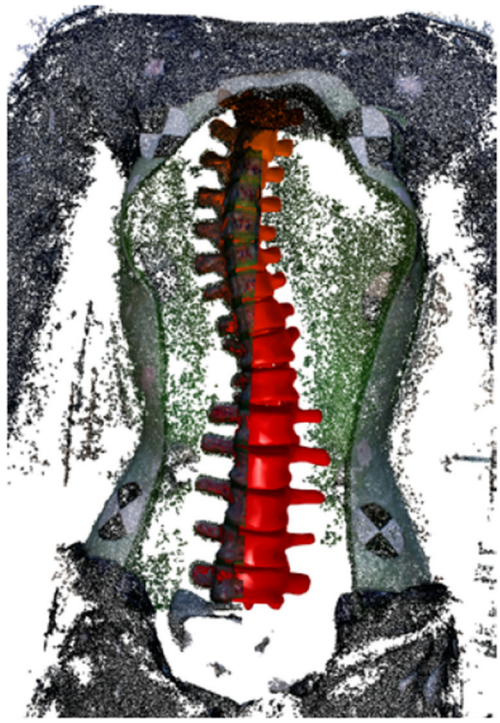

FIGURE 4. Ground truth model registration. Photogrammetric point cloud shown with registered spine mesh model (red). Scan points down sampled and one half of spine model exposed for visibility.

RESULTS

Surface registrations were visually inspected and found to be universally good in the area of the torso. One mesh was misaligned in the lower extremities but did not appear to impact the area of interest. No ground truth measure was available to evaluate mesh registrations, but torso shape was consistent across registrations, as expected for a semi-rigid mannequin. Patch registrations (used in the proposed reconstruction protocol) were compared for each PA-LAT pair of surface scans: mean Standard Deviation (SD) for translation offset $||\mathbf{t}||$ was 12 mm and mean SD for rotation vector $||\mathbf{w}||$ was 0.5°. Average values were found for translation and rotation parameters in PA and LAT position with the following SD: $\mathbf{t}_{PA} = [16;\ 21;\ 1.2]$, $\mathbf{w}_{PA} = [1.2;\ 0.56;\ 7.2]$, $\mathbf{t}_{LAT} = [24;\ 17;\ 3.3]$, $\mathbf{w}_{LAT} = [1.1;\ 0.94;\ 9.0]$, where $\mathbf{t}_i$ is measured in millimeters and $\mathbf{w}_i$ is in degrees (Fig. 5).

Point-to-point 3D reconstruction error had a Root Mean Squared (RMS) error of 2.0 mm (0.8 mm SD) across all landmarks and all trials. Repeatability was

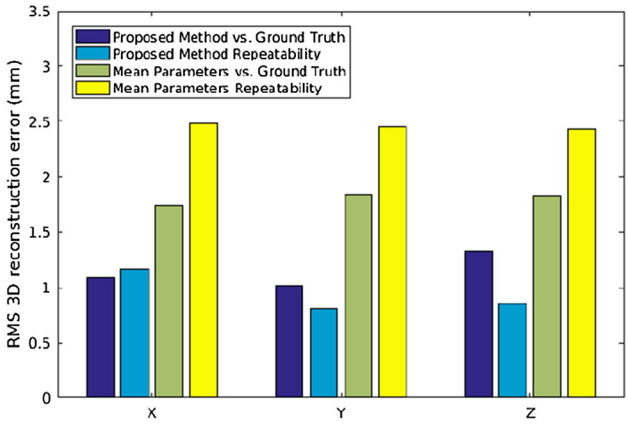

FIGURE 5. Landmark reconstruction error after rigid alignment. Comparisons to ground truth included all landmarks across 18 scans ($n = 1296$) while repeatability was computed for all pairs of reconstructions for each spine curve ($n = 648$). Error was computed in PA scan position (X = posterior–anterior, Y = left–right, Z = cranial–caudal).

TABLE 1. Differences between clinical indices and ground truth values.

	Ground truth error		
Clinical index	RMS	SD	*p*
Cobb angle (°)	7.1	7.0	0.19
Splined cobb PA (°)	1.2	1.2	0.25
Splined cobb LAT (°)	1.0	0.8	0.003
Spine length (mm)	4.0	2.0	$< 10^{-6}$

Cobb Angle was manually measured on PA radiograph and synthetic PA radiograph. Splined Cobb measurements were performed after projecting onto coronal (PA) and sagittal (LAT) planes. For each angular measurement, the maximal curve on the ground truth spine was compared against, whether lumbar or thoracic. Comparisons were performed with Student's *t* test, $n = 18$ for all indices.

measured by aligning the two reconstructions taken for each spine curve, resulting in a landmark displacement of RMS 1.7mm (0.8 mm SD). Corresponding values for reconstructions using average extrinsic parameters were RMS 3.1 mm (1.5 mm SD) displacement from ground truth, RMS 4.3 mm (2.0 mm SD) repeatability. These measurements are decomposed into $[x; y; z]$ components for presentation in Fig. 5 (Table 1).

Several spine level clinical indices were also measureradiographs due to the internal support structured and compared with the ground truth reconstruction, summarized in Table 1. Coronal measurements (traditional and computer Cobb) were not statistically different from ground truth, while measurements dependent upon sagittal markings (computer Cobb, spine length) showed systematic bias. This bias also appeared for mean parameter reconstructions (data not shown), indicating the source of the error was likely landmark selection in the LAT radiographs (see Discussion). RMS error in traditional Cobb angles were of similar magnitude to intra-subject variability: repeated measurements of the same curve had a standard deviation of 5.5° for real radiographs and 6.3° for synthetic radiographs ($n = 9$) (Fig. 6).

DISCUSSION

In this paper an inter-modal scanning system is described for stereoradiographic spine reconstructions. Rather than computing stereo parameters *via* a radiographic calibration target included in the field of view, a set of depth cameras are added to the scanning system. At the cost of some additional infrastructure, this surface information provides an anchor for merging radiographs.

One attractive feature of surface-based correspondences is the potential to measure and correct for changes in subject posture between scans. The BlendSCAPE model, while over-engineered for the case of an immobile mannequin, naturally accounts for variations in pose and body shape. Transformation parameters **t** and **w** were reliably extracted from surface scans, with 12 mm SD and 0.5° SD respectively. These values can be read as the "remove and replace" error, as a rigid mannequin with invariant surface topography was used in this study.

This said, this SCAPE based registration protocol was specifically designed to be robust to human subjects in a clinical environment. A minimum number of landmarks are used, and primarily for initialization, reducing the impact of operator error. The "symmetry line" as a topographical feature helps regularize the registration without skin-based markers that may move relative to the underlying spine structure.

Despite the success of surface registrations in finding extrinsic parameters, simply using the mean parameters produced reconstructions that were not dramatically inferior. Although landmark error was significantly reduced using the proposed methods ($p < 0.05$ from paired Student's *t*-test), the error dropped by only about 1mm despite the high variability in mannequin position (8° SD in *z*-axis rotation).

Careful examination of the data revealed that using mean parameters is remarkably robust to variations in subject positioning, provided that rotation is strictly limited to the z axis. This was exactly the case for the mannequin used in this study, and is consistent with previously published work, where patient positioning is carefully controlled with a rotating platform and

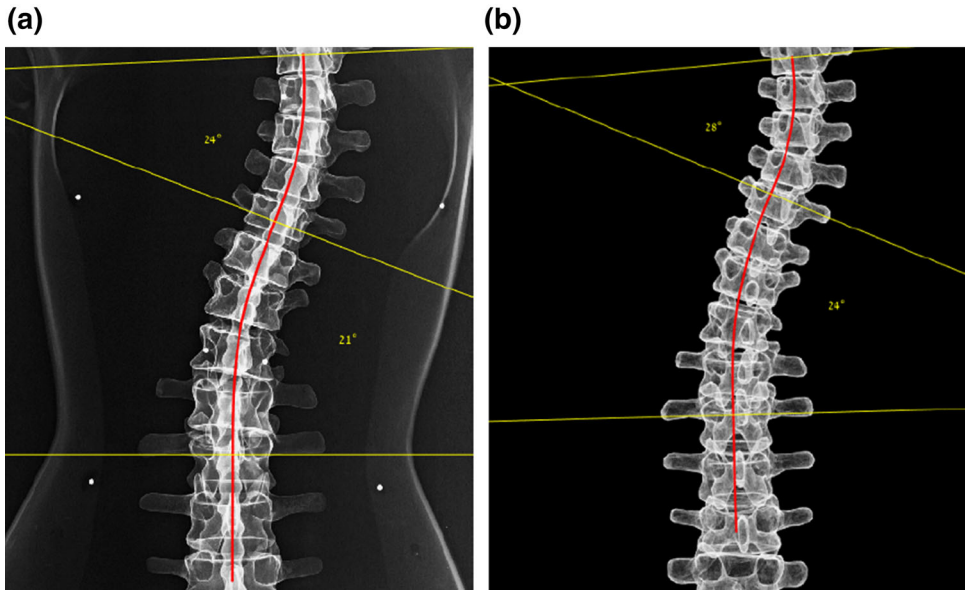

FIGURE 6. Splined spine line and computer Cobb angles. Measurements were performed after splining vertebral bodies in 3D and projecting into radiographic image coordinates. (a) Real radiograph with computer Cobb angles (b) synthetic radiograph generated from proposed reconstruction methods. These images were also used to compare manually measured Cobb angles.

rigid brace.[28] Synthetic experiments have showed that even 2° off-axis rotation can significantly degrade reconstruction accuracy.[29] It would be of great interest to know how much variation in patient posture is to be expected without a turntable, which might be another excellent application for the 3D scanning protocol presented here.

It is also worth noting that while point-to-point reconstruction error is similar to previously published results,[6,22,28] coronal metrics showed higher error and statistical bias. One explanation could be the inter-modal comparison to a ground truth spine model derived from an optical scan. However, if this were the dominant source of error it would be expected that repeatability tests would show significantly lower reconstruction error. In fact, only a modest reduction in error is observed for the proposed method, and a marked increase when using average extrinsic parameters. Instead, the likely source of increased baseline error is difficulty identifying pedicle landmarks on LAT radiographs due to the internal support structure (visible in Fig. 6). While other studies have found reconstruction error to be systematically lower for pedicle landmarks,[23] here repeatability was significantly better for endplate landmarks (RMS 1.5 vs. 1.7 mm, $p < 0.0001$ paired Student's t test).

Taken together, these results demonstrate the potential for surface based stereoradiographic reconstructions using depth cameras. However, much work remains to confirm the utility of these methods in clinical practice. The first step must be to validate the protocol on nonrigid scan targets, as human subjects will deform slightly between scans. Since the SCAPE model was designed to model the entire range of human motion, there is every reason to hope that these techniques will be capable of compensating for minor variations in patient posture–but this must be demonstrated empirically.

It should be noted that the statistical body models used in these algorithms were trained on healthy adults; models trained on scoliotic subjects may be beneficial for clinical application. However, the non-rigid registration parameters presented here already allow for highly plastic shape matching; although the mannequin deviated greatly from the training data (exaggerated neck, bloated legs) the registered surface hewed closely to the scan (~ 2 mm point-to-surface distance). As such, better surface registrations–and consequently stereo parameters–are more likely to be realized with higher quality surface scans, as accuracy and range continue to improve rapidly in low-cost depth cameras.

The 3D reconstructions generated using this method can be visualized directly by a physician for qualitative assessment or used to generate a quantitative clinical metrics. For example, to visualize the Cobb angle in Fig. 6 the stereocorresponding points were fitted to vertebral surfaces and projected into image space. Similarly, metrics such as axial rotation or spine balance can be extracted from the fitted vertebral models. Reconstructions of the same patient from different

times or in different postures could be quantitatively compared or tracked.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

[1]Agisoft, L. Photoscan http://www.agisoft.com/.

[2]Allen, B., B. Curless, and Z. Popović. The space of human body shapes: reconstruction and parameterization from range scans. *ACM Trans. Gr.* 22:587–594, 2003.

[3]Andre, B., J. Dansereau, H. Labelle, B. André, J. Dansereau, and H. Labelle. Optimized vertical stereo base radiographic setup for the clinical three-dimensional reconstruction of the human spine. *J. Biomech.* 27:1023–1035, 1994.

[4]Anguelov, D., P. Srinivasan, and D. Koller. Scape: shape completion and animation of people. *ACM Trans. Gr.* 24:408–416, 2005.

[5]Arun, K. S., T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3D point sets. *IEEE Trans. Pattern Anal. Mach. Intell.* 9:698–700, 1987.

[6]Aubin, C. E., J. Dansereau, F. Parent, H. Labelle, and J. Guise. Morphometric evaluations of personalised 3D reconstructions and geometric models of the human spine. *Med. Biol. Eng. Comput.* 35:611–618, 1997.

[7]Bogo, F., M. J. Black, M. Loper, and J. Romero. Detailed full-body reconstructions of moving people from monocular RGB-D sequences. Proceedings of the Iccv pp. 2300–2308, 2015.

[8]Bogo, F., J. Romero, M. Loper, and M. J. Black. FAUST: Dataset and evaluation for 3D mesh registration. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition pp. 3794-3801, 2014.

[9]Breque, C., J. C. Dupre, and F. Bremand. Calibration of a system of projection moiré for relief measuring: Biomechanical applications. *Opt. Lasers Eng.* 41:241–260, 2004.

[10]Chen, Y., Z. Liu, and Z. Zhang. Tensor-based human body modeling. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition pp. 105–12, 2013.

[11]Cheriet, F., J. Dansereau, Y. Petit, C. É. Aubin, H. Labelle, and J. A. Guise. Towards the self-calibration of a multiview radiographic imaging system for the 3D reconstruction of the human spine and rib cage. *Vision Interface* 13:761, 1999.

[12]Cheriet, F., C. Laporte, S. Kadoury, H. Labelle, and J. Dansereau. A novel system for the 3-D reconstruction of the human spine and rib cage from biplanar X-ray images. *IEEE Trans. Biomed. Eng.* 54:1356–1358, 2007.

[13]Community, B. O. Blender - a 3D modeling and rendering package http://www.blender.org.

[14]Dai, J. S. Euler-Rodrigues formula variations, quaternion conjugation and intrinsic connections. *Mech. Mach. Theory* 92:144–152, 2015.

[15]Dansereau, J., and I. A. F. Stokes. Measurements of the three-dimensional shape of the rib cage. *J. Biomech.* 21:893–901, 1988.

[16]Drerup, B., and E. Hierholzer. Automatic localization of anatomical landmarks on the back surface and construction of a body-fixed coordinate system. *J. Biomech.* 20:967–970, 1987.

[17]Drerup, B., and E. Hierholzer. Back shape measurement using video rastereography and 3-dimensional reconstruction of spinal shape. *Clin. Biomech.* 9:28–36, 1994.

[18]Geman, S., D. E. McClure, and D. Geman. A nonlinear filter for film restoration and other problems in image processing. *CVGIP* 54:281–289, 1992.

[19]Hasler, N., C. Stoll, M. Sunkel, B. Rosenhahn, and H. P. Seidel. A statistical model of human pose and body shape. *Comput. Gr. Forum* 28:337–346, 2009.

[20]Hirshberg, D. A., M. Loper, E. Rachlin, and M. J. Black. Coregistration: Simultaneous alignment and modeling of articulated 3D shape. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 7577 LNCS, pp. 242–255, 2012.

[21]Janicki, J., and B. Alman. Scoliosis: Review of diagnosis and treatment. *Paediatr. Child Health* 12:771–776, 2007.

[22]Kadoury, S., F. Cheriet, C. Laporte, and H. Labelle. A versatile 3D reconstruction system of the spine and pelvis for clinical assessment of spinal deformities. *Med. Biol. Eng. Comput.* 45:591–602, 2007.

[23]Labelle, H., J. Dansereau, C. Belleeur, and J. C. Jéquier. Variability of geometric measurements from three-dimensional reconstructions of scoliotic spines and rib cages. *Eur. Spine J.* 4:88–94, 1995.

[24]Legaye, J., P. Saunier, R. Dumas, and C. Vallee. Correction for patient sway in radiographic biplanar imaging for three-dimensional reconstruction of the spine: In vitro study of a new method. *Acta Radiol.* 50:781–790, 2009.

[25]Loper, M. Human shape estimation using statistical body models. Ph.D. thesis, Tubingen, 2017.

[26]Loper, M., N. Mahmood, and J. Romero. SMPL: a skinned multi-person linear model. *ACM Trans. Gr.* 34:248, 2015.

[27]Mitulescu, A., I. Semaan, J. A. De Guise, P. Leborgne, C. Adamsbaum, and W. Skalli. Validation of the non-stereo corresponding points stereoradiographic 3D reconstruction technique. *Med. Biol. Eng. Comput.* 39:152–158, 2001.

[28]Moura, D. C., and J. G. Barbosa. Real-scale 3D models of the scoliotic spine from biplanar radiography without calibration objects. *Comput. Med. Imaging Gr.* 38:580–585, 2014.

[29]Moura, D. C., J. G. Barbosa, A. M. Reis, and J. M. R. S. Tavares. A flexible approach for the calibration of biplanar radiography of the spine on conventional radiological systems. *CMES* 60:115–137, 2010.

[30]Pishchulin, L., S. Member, S. Wuhrer, T. Helten, C. Theobalt, and B. Schiele. Building statistical shape spaces for 3D human modeling. *Comput. Vision Pattern Recogn.* 67:1–10, 2015.

[31]Pons-Moll, G., J. Romero, N. Mahmood, and M. J. Black. Dyna: A model of dynamic human shape in motion. *ACM Trans. Gr.* 34:120, 2015.

[32]Stokes, I. A. F., D. Shuma-Hartswick, and M. S. Moreland. Spine and back-shape changes in scoliosis. *Acta Orthop.* 59:128–133, 1988.

[33]Sumner, R. W., and J. Popović. Deformation transfer for triangle meshes. *ACM Transactions on Graphics* 23:399, 2004.

[34]The MathWorks Inc. MATLAB.

[35]Tsoli, A., N. Mahmood, and M. J. Black. Breathing life into shape: capturing, modeling and animating 3d human breathing. *ACM Trans. Gr.* 33:1–11, 2014.

[36]Wood, G. A., and R. N. Marshall. The accuracy of DLT extrapolation in 3-dimensional film analysis. *J. Biomech.* 19:781–800, 1986.