Automatic Digital Biometry Analysis based on Depth Maps

Miguel Reyes[a,b], Albert Clapés[a,b], José Ramírez[c], Juan R. Revilla[c], Sergio Escalera[a,b]

{mreyes,aclapes}@cvc.uab.es, {jramirez,jrevilla}@csc.uic.es, sergio@maia.ub.es

[a] *Dept. Matemàtica Aplicada i Anàlisi, UB, Gran Via de les Corts Catalanes 585, 08007, Barcelona*

[b] *Computer Vision Center, Campus UAB, Edifici O, 08193, Bellaterra, Barcelona*

[c] *Instituto de Fisioterapia Global Mezières, Guillem Tell 27, 08006, Barcelona*

Abstract

World Health Organization estimates that 80% of the world population is affected by back-related disorders during his life. Current practices to analyze musculo-skeletal disorders (MSDs) are expensive, subjective, and invasive. In this work, we propose a novel tool for static body posture analysis and dynamic range of movement estimation of the skeleton joints based on 3D anthropometric information computed from multi-modal data. These data combine RGB information from a video camera and depth from an infrared sensor. Given set of keypoints defined by the user, RGB and depth data are aligned, depth surface is reconstructed, keypoints are matched using a novel point-to-point fitting procedure, and accurate measurements about posture and spinal curvature are computed. Given a set of joints, range of movement measurements is also obtained. Moreover, gesture recognition based on joint movements is performed to look for the correctness in the development of physical exercises. The system shows high precision and reliable measurements, being useful for posture reeducation purposes to prevent MSDs, such as back pain, as well as tracking the posture evolution of patients in physical rehabilitation treatments.

Keywords: Multi-modal Data Fusion, Depth Maps, Posture Analysis, Anthropometric data, Musculo-skeletal disorders, Surface Reconstruction, Gesture Analysis

1. Introduction

World Health Organization has categorized disorders of the musculo-skeletal system as the main cause for absence from occupational work and one of the most important causes of disability in elders in the form of rheumatoid arthritis or osteoporosis. It is estimated that 80% of the world population will suffer from musculo-skeletal disorders or dysfunctions (MSDs) during his life. As a result, MSDs lead to considerable costs for public health systems [1].

The body posture evaluation of a subject manifests, in different degrees, his level of physic-anatomical health given the behavior of bone structures, and especially of the dorsal spine. For instance, common MSDs such as scoliosis, kyphosis, lordosis, arthropathy, or spinal pain show some of their symptoms through body posture. This requires the use of reliable, noninvasive, automatic, and easy to use tools for supporting diagnostic. However, given the articulated nature of the human body, the development of this kind of systems is still an open issue.

Given the difficulty of finding a tool for measuring the posture of the human body at different configurations, digital biometry has become a very useful tool. Digital biometry is defined by the American Society of Anthropometric data as "the technology to obtain reliable information of the physical objects or the environment through the recording of images, its measurement or interpretation". The systems based on this technology are capable to estimate morphological or functional alterations, being a useful resource for health professionals.

The diagnostic evaluation of the anomalies follows through a careful study of musculo-skeletal structure and receptorial aspects. These diagnostic tools are based on monitoring anthropometric relationships with validated accuracy [21, 38, 34, 23, 22, 27, 28, 29]. These kind of tools are minimally invasive and obtain good accuracy results in terms of precision but require an specific scene configuration, being necessary a camera calibration preprocessing due to use of two-dimensional cameras in different planes. Another common handicap of these systems, is their reduced portability to perform a custom analysis for the therapist. These systems have been built highly parameterized for a specific type of analysis. Most of these systems only treat specific areas of the body, primarily the spinal deformities [23, 24, 32, 30, 31]. The solution more frequently applied to measure body posture consists of the installation and alignment of multiple cameras, applying stereo vision methodologies [3, 4]. This kind of system uses to be expensive and require specific

and restricted illumination conditions. The main alternative is accelerometers [5], but these systems also use to be expensive and invasive. Furthermore, the location of accelerometers on the body of a subject is difficult and cannot obtain accurate results because of the spatial measurements of multiaxial articulations. There is another type of diagnostic tools that perform a quantitative clinical analysis based on diagnostic predictions through analysis of anthropometric relations [25, 26]. These systems are able to predict diseases in premature state. Using these tools is highly restricted to a sector of the population, and its success rate is not accurate enough to be used in clinical diagnostics.

Other common solutions are augmented reality systems using computer vision techniques [6]. However, detecting humans in images or videos is a challenging problem because of the high variation of possible configurations of the scenario, such as changes in the point of view, illumination conditions, or background complexity, just to mention a few. In order to treat human pose in uncontrolled scenarios, there is a recent work using range images for object recognition or modeling [7]. This new approach introduced a solution to the problem of intensity and view changes in still images through the representation of 3D structures. The progress and spread of these new methods came slowly since data acquisition devices were expensive and bulky, with cumbersome communication interfaces when conducting experiments. However, Microsoft® has recently launched Kinect™ [8], a cheap multisensory device, based on structured light technology, capable of capturing visual depth information (RGB-Depth technology) to, then, generate real-time depth maps containing discrete range measurements of the physical scene. The device is so compact and portable that it can be easily installed in any environment to analyze scenarios where humans are present. While inexpensive devices such as the Nintendo Wii Balance Board TM, a clinically feasible alternative to a force platform [35], it cannot accurately differentiate joint movements. Many researchers have obtained their first results in the field of human motion capture using this technology. In particular, Shotton et al. [9] presented one of the greatest advances in the extraction of the human body pose from depth maps, which also forms the core of the Kinect™ human recognition framework. Additionally, there have been published several studies supporting its accuracy and portability for clinical applications [36, 37, 33].

In this work, we present a novel system that uses RGB-Depth information to elaborate a semi-automatic postural and spinal analysis, and to automatically estimate the range of movement of the different limbs. Regarding the

postural and spinal analysis, compared to standard alternatives and supported by clinical specialists, the system shows high precision and reliable measurements to be included in the clinical routine. The range of movement estimation and gesture recognition are also reliable for physical rehabilitation purposes.

The rest of the paper is organized as follows: Section 2 introduces the materials and methods, Section 3 presents the system for posture estimation and range of movement analysis, Section 4 presents the results to validate our proposal, and finally, Section 5 concludes the paper.

2. Material and methods

In this section, we provide the details about the sensors, data, and main methods involved in the development of our system.

2.1. Sensors

The data acquisition is done using the Microsoft Kinect™ device. The device bases on a technology that combines a color camera sensor with a depth sensor. The depth sensor consists of an infrared laser projector combined with a monochrome CMOS sensor, which captures three-dimensional video data under any ambient light conditions, but it has a practical range limit of [1.2m - 3.5m] distance. The color video stream uses 8-bit VGA resolution (640×480) with a Bayer color filter, while the monochrome depth sensing video stream is in QVGA resolution (320×240 pixels) with 11-bit depth, which provides 2,048 levels of sensitivity. The device outputs video, to then be processed, at a frame rate of 30Hz.

A specific validation method to evaluate our system using infrared LEDs has been designed to assess the capabilities of Kinect™ as a measurement tool. Each infrared LED comprises a rechargeable button cell IR2032, a microswitch and a cold white color LED, 6000-8000mcd, 3V-20mW (Figure 1).

2.2. Data and groundtruth definition

In order to measure the precision of the proposed methodology in the different modules of the system, we created a novel data set consisting of two different parts.

A single multi-modal frame, comprising a still image and a depth map, is needed to perform an analysis in both static posture and spine curvature analysis. Thus, a battery of 500 single multi-modal frames has been labeled

(a) (b)

Figure 1: (a) Led sensor and infrared filter. (b) Validation of distances and angles.

by three different observers (Figure 2), with an inter observer correlation superior to 99% for all planes (X, Y, Z). Each frame contains a set of angles and distances in order to simulate an analysis protocol for the study of posture, placing twelve infrared led markers on subject's skin. A total of 20 subjects participated in the validation of the method. With the aim to perform an automatic validation of the tests, infrared markers are detected by means of thresholding a HSV infrared-filtered image (filtered at 850 nanometers).

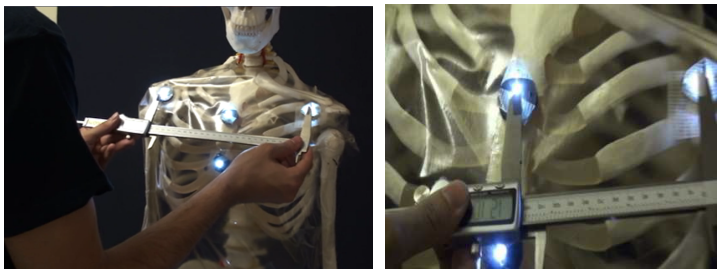


Figure 2: Measuring spatial relations between infrared markers in order to manually label the single multi-modal frames.

Sequences of multi-modal frames containing exercises for physical rehabilitation have been defined to validate range of motion analysis and gesture recognition. For this purpose, a set of 25 sequences of one minute each one has been recorded, in which 5 different subjects appear performing repeatedly and correctly 5 exercises. This part of the data set contains approximately a total of 37300 semi-supervised labeled frames.

2.3. Methods

The postural analysis is done through the examination of 3D anthropometric values. Given a set of markers/keypoints defined by the user, our

proposed method performs the following steps: a) RGB and depth data are aligned, b) noise is removed and depth surface is reconstructed, c) user keypoints and predefined protocols are matched using a novel point-to-point fitting procedure, d) static measurements about posture and spinal curvature are accurately computed. The static measurements about posture consist of spatial relations between those keypoints: pairwise distances, distances relative to a vertical or horizontal axis, and angles among triplets of keypoints or angles between pairs relative to an axis. Regarding the spinal measurements, it is interpolated a curve representing the spine and also clinical spatial relations (distances and angles) among vertebrae are computed.

On the other hand, a skeletal model is obtained from a real-time human pose recognition from single depth images [9]. This model provides spatio-temporal information about the joints of the subject, and so then a dynamic range of movement of his limbs can be robustly estimated. The joints of interest to estimate the range of movement are manually selected by the user or automatically determined by the system depending on the gesture being performed. Dynamic Time Warping is applied over the detected threedimensional position of the joints in order to perform automatic gesture recognition.

3. Calculation

We designed a full functional system devoted to help in the posture reeducation task with the aim of preventing and correcting musculo-skeletal disorders, as well as tracking the posture evolution of patients in physical rehabilitation treatments. The system is composed by three main functionalities: a) static posture analysis (SPA), b) spine curvature analysis (SCA), and c) range of movement analysis, including automatic gesture recognition (RMA). The architecture of the system is shown in Figure 3. First, a pre-processing step to remove noise and reconstruct surfaces is performed. Next, we describe each of these stages.

3.1. Noise removal and surface reconstruction

After aligning RGB and depth data [12, 15], and even though the used depth information is compelling, it is still inherently noisy. Depth measurements often fluctuate and depth maps contain numerous holes where no readings are obtained. In order to obtain a valid and accurate depth map, we perform a depth preprocessing step to eliminate erroneous information

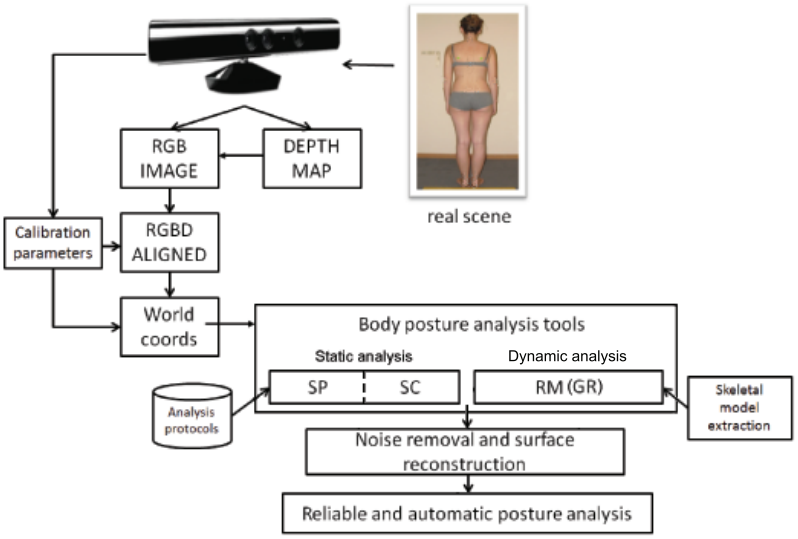


Figure 3: Posture analysis system.

caused by noise and to reconstruct surfaces not well defined. In this sense, posterior estimation of 3D coordinates of markers are successfully computed, and accurate measurements of distances and angles for pose and range of movement are estimated. We perform the following tasks:

Noise removal: For each point, we compute the mean distance from it to all its neighbors. By assuming that the resulted distribution is Gaussian with a mean and a standard deviation, all points whose mean distances are outside an interval defined by the global distances mean and standard deviation are considered as outliers.

Surface reconstruction: The surface reconstruction process rests on an inpainting image processing adaptation for 3D point cloud, based on compactly supported radial basis functions [20]. Radial basis function (RBF) interpolation is an important method for surface reconstruction [19] from 3D scatter points. The 3D point cloud extracted from the scene is encoded as a depth map. Then, the algorithm converts 2D image inpainting problem into implicit surface reconstruction problem from 3D points set. By performing this reconstruction small holes are reconstructed with high precision. Using compactly supported radial basis functions decreases the computational complexity in comparison to other common re-sampling approaches [14], which attempt to recreate the missing parts of the surface with higher order polynomial interpolation among the surrounding data points. Figure 4 shows an

example of this process [1].

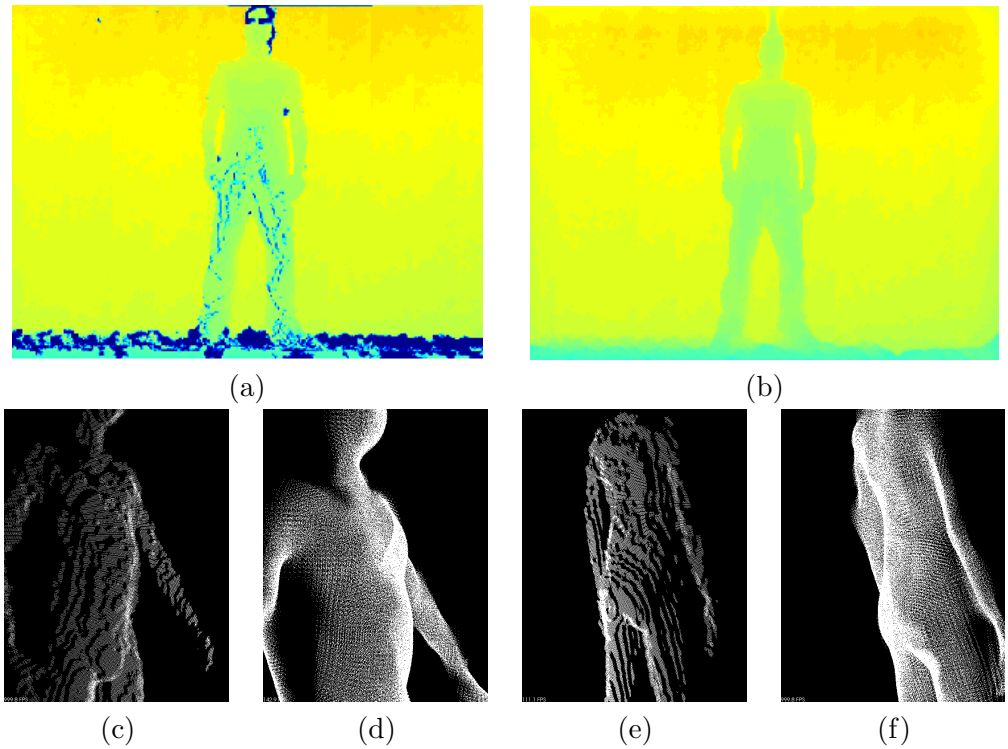


Figure 4: (a) Original depth map. (b) Inpaint filtered and re-sampled. (c)-(d) and (e)-(f) additional non-filtered and filtered surface examples for two depth body acquisition examples, respectively.

Once the data is aligned and depth maps are filtered, the user can access to the main functionalities of the system described below.

3.2. Static posture analysis (SPA)

This module computes and associates a set of three-dimensional angles and distances to keypoints defined by the user. These keypoints correspond to the dermal markers placed on the patient's skin. These dermal markers

[1] We experimentally found that our approach for noise removal and background reconstruction obtained better results than standard approaches based on accumulating temporal images (e.g. 30 frames of a stationary subject) for noise reduction and hole filling.

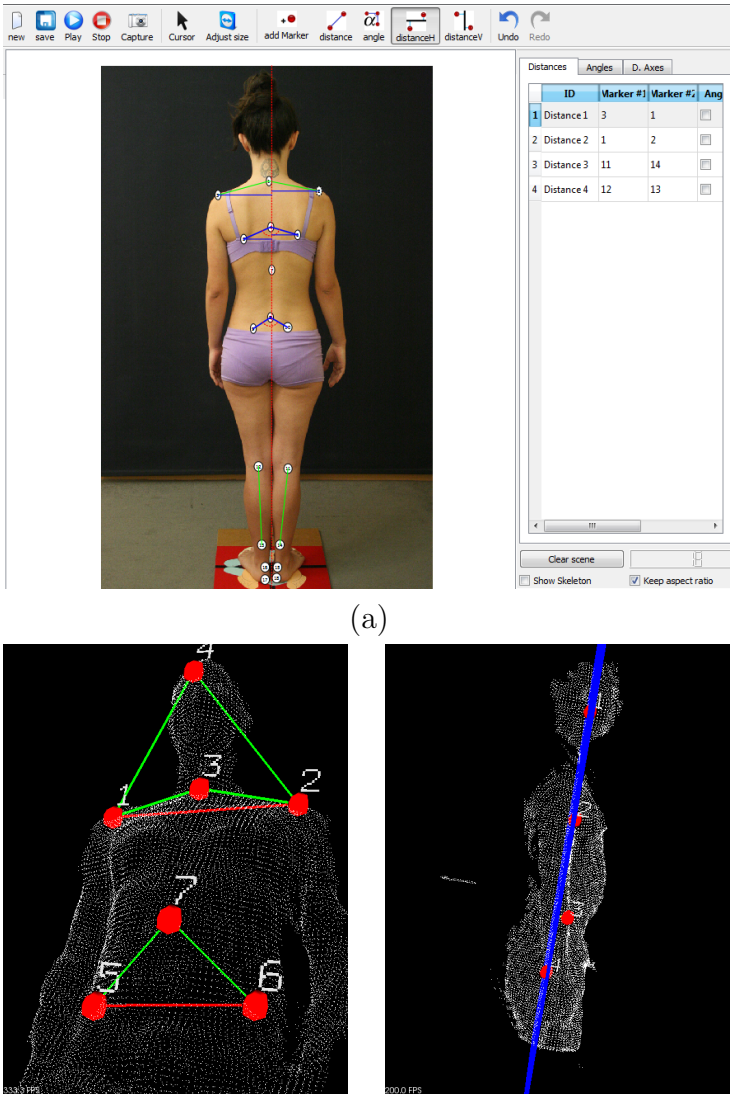
have to be physically placed by a therapist and, then, manually selected interacting with the RGB data displayed in the screen (which internally is aligned with the corresponding depth data, providing real 3D information) by the application to define the set of virtual markers (or keypoints). The keypoints could not be directly placed in the virtual scenario, because they have to correspond to specific body bone structures that need to be located by palpation.

The module also allows the therapist the possibility of designing a protocol of analysis. That is, a predefined set of angular-distance measurements among a set of body keypoints, all of them defined and saved by the user for a posterior automatic matching. The application of these protocols makes possible static posture analysis performed quickly, being highly customized, and done in an automatic way. Thus, when keypoints have been defined, the user can choose the most appropriate protocol and the system will finally provide the measurements automatically. Figure 5 shows an example of a predefined protocol (the set of manual annotated keypoints together with the list of distance and angle relations to be computed) and two examples of 3D views and relations for a set of body landmarks.

In order to obtain an intelligent and automatic estimation of posture measurements, we define a correspondence procedure among manually placed virtual markers and protocol markers. We formulate markers matching as an optimization problem. Suppose a protocol analysis (template) T composed by L markers, $T = \{T_1, T_2, ..., T_L\}$, $T_i = (x_i, y_i, z_i)$, and the current analysis C composed by the same number of markers, $C = \{C_1, C_2, ..., C_L\}$ (predefined template and current set of keypoints defined by the user, respectively). Our goal is to make a one-to-one correspondence so that we minimize the sum of least square distances among assignments as follows:

$$\mathsf{argmin}_{C'} \sum_{i=1}^{L} \|C'_i - T_i\|^2 , \tag{1}$$

where C' is evaluated as each of the possible permutations of the elements of C. For this task, first, we perform a soft pre-alignment between C and T using Iterative Closest Point (ICP) [13], and then, we propose a sub-optimal approximation to the least squares minimization problem. ICP is based on the application of rigid transformations (translation and rotation) in order to align both sequences C and T. This attempts to minimize the error of alignment $E(.)$ between the two marker sequences as follows:



(a)

(b) (c)

Figure 5: Static posture analysis example: (a) Interaction from interface, (b) 3D visual view of selected landmarks and relations, and (c) additional view including reference plane.

$$E(\mathcal{R}, \mathcal{T}) = \sum_{i=1}^{L} \sum_{j=1}^{L} w_{i,j} \|T_i - \mathcal{R}(C_j) - \mathcal{T}\|^2, \tag{2}$$

being $\mathcal{R}$ and $\mathcal{T}$ the rotation matrix and translation vector, respectively. $w_{i,j}$ is assigned 1 if the i-th point of T described the same point in space as the j-th point of C. Otherwise $w_{i,j} = 0$. Two things have to be calculated: first, the corresponding points, and second, the transformation ($\mathcal{R}$, $\mathcal{T}$) that minimizes $E(\mathcal{R}, \mathcal{T})$ on the base of the corresponding points. For this task, we apply Singular Value Decomposition (SVD). At the end of the optimization, the new projection of the elements of C is considered for the final correspondence. Then, Eq. 1 is approximated as follows: Given the symmetric matrix of distances $\mathcal{M}$ of size $L \times L$ which codifies the set of $L \cdot (L-1)/2$ possible distances among all assignments between the elements of C and T, we set a distance threshold $\theta_{\mathcal{M}}$ to define the adjacency matrix A:

$$A(i,j) = \begin{cases} 1 & \text{if} \quad \mathcal{M}(i,j) < \theta_{\mathcal{M}} \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

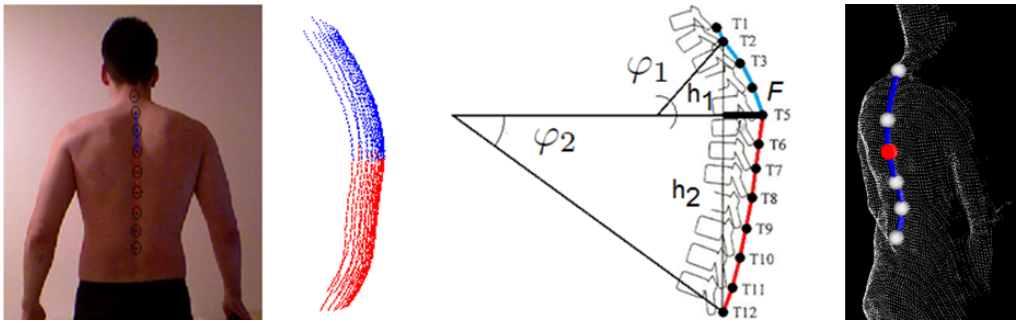
Then, instead of looking for the set of $L!$ possible assignments of elements of C and T that minimizes Eq. 1, only the possible assignments (C_i, T_j) that satisfies $A(i,j) = 1$ are considered, dramatically reducing the complexity of the correspondence procedure[2].

3.3. Spine curvature analysis (SCA)

The objective of this task is to evaluate sagittal spine curvatures (curves of the spine projected on the sagittal plane) by noninvasive graphic estimations in kyphotic and lordotic patients. Kyphosis and lordosis are, respectively, conditions of over-curvature of the thoracic spine (upper back) and the lumbar spine (lower back). The methodology proposed by Leroux et al [10] offers a three-dimensional analysis valid for clinical examinations of those conditions. In order to perform this analysis we proceed as follows. First, the therapist places the virtual markers on the spine (an example of spine interaction and computation are shown in Figure 6(a)). Then, a few markers are selected and the 3D curve that represents the spine is reconstructed by linear interpolation (Figure 6(b)). Finally, the anthropometric kyphosis $\mathcal{K}_a$ and lordosis $\mathcal{L}_a$ are obtained.

The geometric model to compute $\mathcal{K}_a$ is represented in the Figure 6(c). F divides the curve representing the thoracic spine in two asymmetric arcs with

[2]We experimentally found that high values of $\theta_{\mathcal{M}}$ obtain robust results and reduces then computational cost in comparison to other approaches, such as Shape Context [11].



(a) (b) (c) (d)

Figure 6: (a) Sample of analysis. (b) Automatically reconstructed 3D spinal cloud. (c) Geometric model to obtain anthropometric kyphosis and lordosis value. (d) Additional 3D view of the reconstructed spine through the system interface.

different radius. Note that the F component begins at the farthest marker (apex, corresponding to T5) and it ends at the intersection with the T2-to-T12 line. *h1* and *h2* are the distances from T2 to the intersection and the distance from the intersection to T12, respectively. Then, the summation of two angles, φ_1 and φ_2, represents the kyphosis curve value, where:

$$\begin{aligned} \varphi_1 &= 180 - 2 \cdot \arctan\left(\tfrac{h_1}{F}\right), \\ \varphi_2 &= 180 - 2 \cdot \arctan\left(\tfrac{h_2}{F}\right). \end{aligned} \tag{4}$$

$\mathcal{L}_a$ is calculated in a similar way, though the therapist should place the markers in the lumbar spine region.

The capacity analysis of the spine is reinforced by a three-dimensional environment that can be managed by the therapist for a thorough examination (Figure 6(d) and Figure 7).

3.4. Range of movement analysis (RMA)

In order to complement the posture analysis procedure, we compute the range of movement of the different body articulations. This is a facility aimed to assist in diagnoses and physical rehabilitation treatments. For this purpose, we perform user detection using the Random Forests approach with depth features of Shotton et al [9] and, then, compute the skeletal model. This process is performed computing random offsets of depth features as follows:

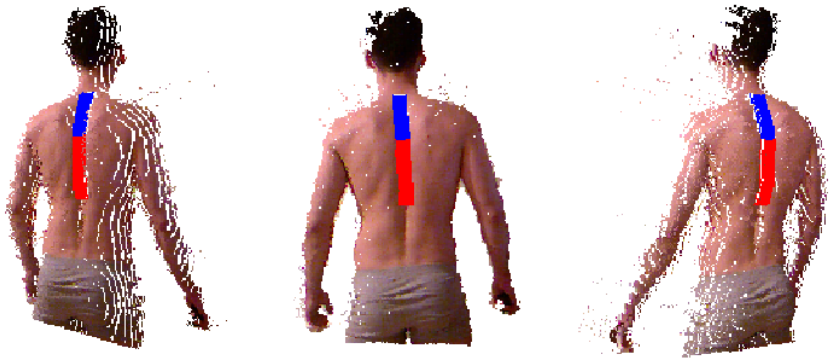


Figure 7: Three-dimensional examination environment managed by the therapist.

$$f_\theta(D, \mathbf{x}) = \mathbf{D}_{\left(\mathbf{x}+\frac{\mathbf{u}}{\mathbf{D_x}}\right)} - \mathbf{D}_{\left(\mathbf{x}+\frac{\mathbf{v}}{\mathbf{D_x}}\right)}, \tag{5}$$

where $\theta = (\mathbf{u}, \mathbf{v})$, and $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{\mathbf{2}}$ is a pair of offsets, depth invariant. Thus, each θ determines two new pixels relative to $\mathbf{x}$, the depth difference of which accounts for the value of $f_\theta(D, \mathbf{x})$. Using this set of random depth features, Random Forests is trained for a set of trees, where each tree consists of split and leaf nodes (the root is also a split node). Finally, we obtain a final pixel probability of body part membership l_i as follows:

$$P\left(l_i|D, \mathbf{x}\right) = \frac{1}{\tau}\sum_{j=1}^{\tau} P_j\left(l_i|D, \mathbf{x}\right), \tag{6}$$

where $P\left(l_i|D, \mathbf{x}\right)$ is the Probability Density Function (PDF) stored at the leaf, reached by the pixel for classification $(D, \mathbf{x})$ and traced through the tree j, $j \in \tau$. Computing the intersection borders among mean shift clusters estimated over the obtained limb labels of the Random Forests procedure, we obtain a three-dimensional skeletal model composed by nineteen joints. The physician then selects joint articulations and automatically obtains their maximum opening and minimum closing values measured in degrees for a certain period of time (Figure 8).

However, to provide a fully-automatic functional range of movement estimation, we had to consider those cases in which a therapist is not present to operate with the system, for instance, when the patient is performing the rehabilitation exercises at home. Obviously, it would be useless obtaining range of movement measurements when an exercise is not well performed.

Figure 8: Skeletal model and example of selected articulations with computed dynamic range of movement (maximum opening and minimum closing values of a joint measured in degrees for a certain period of time).

This problem was solved by including an action/gesture recognition framework to our software.

3.4.1. Gesture recognition

The system is capable to notice if a certain exercise is performed properly or not. For this task, the system includes a set of rehabilitation exercises to ask the patient to perform them. Then, temporal feature vectors computed by the Random Forests approach described before are compared between the patient and the original templates using dynamic programming based on the 3D spatial coordinates of the nineteen joints. The similarity between exercise patterns is computed using Dynamic Time Warping (DTW) [16]. The reference patterns are defined by the therapist and saved in the system database.

The DTW algorithm was defined to match temporal distortions between two models, finding an alignment warping path between two time series $X = \{x_1, ., x_N\}, N \in \mathbb{N}$ and $Y = \{y_1, ., y_M\}, M \in \mathbb{N}$.

If sequences take values from the feature space Φ, to compare two sequences $X, Y \in \Phi$, we need to use a local distance measure, defining a function:

$$d : \Phi \times \Phi \rightarrow \mathbb{R} \geq 0. \tag{7}$$

The algorithm starts by building the local distance matrix $\mathcal{D} \in \mathbb{R}^{N \times M}$ representing all pairwise distances between X and Y:

$$\mathcal{D} \in \mathbb{R}^{N \times M} : d_{i,j} = \|x_i - y_j\|, i \in [1 : N], j \in [1 : M]. \tag{8}$$

Once the local distance matrix is built, the algorithm finds the warping path. The warping path defines a correspondence of an element $x_i \in X$ to $y_j \in Y$ following certain constraints. The warping path (or alignment path) built by DTW is a sequence of points $p = (p_1, .., p_K)$ with $p_l = (p_i, p_j) \in [1 : N] \times [1 : M]$ for $l \in [1 : K]$ which satisfy the following criteria:

Boundary condition: $p_1 = (1, 1)$ and $p_K = (N, M)$.

Monotonicity condition: $n_1 \leq .. \leq n_K$ and $m_1 \leq .. \leq m_K$, preserving the time-ordering of points.

Step size condition: $p_{l+1} - p_l \in \{(1, 1), (1, 0), (0, 1)\}$.

The distance function associated with a warping path computed with respect to the distance matrix (which represents all pairwise distances) is:

$$d_p(X, Y) = \sum_{l=1}^{L} d(x_{n_l}, y_{m_l}). \tag{9}$$

The warping path that supposes the minimum cost of alignment is called optimal warping path, P^*. To find it, we need to test every possible warping path between X and Y, which could be computationally challenging (the number of optimal paths grow exponentially as the lengths of X and Y grow linearly). For this reason, DTW employs a Dynamic Programming - based algorithm that drastically reduces the complexity to $O(MN)$. The Dynamic Programming part of DTW algorithm uses the DTW distance function:

$$DTW(X, Y) = d_{p^*}(X, Y) = min\{d_p(X, Y), p \in P^{N \times M}\}, \tag{10}$$

where $P^{N \times M}$ is the set of all possible warping paths, and builds the accumulated cost matrix $\mathcal{C}$ which is defined as follows:

First row: $\mathcal{C}(1, j) = \sum_{k=1}^{j} d(x_1, y_k), j \in [1, M]$.

First column: $\mathcal{C}(i, 1) = \sum_{k=1}^{i} d(x_k, y_1), i \in [1, N]$.

All other elements: $\mathcal{C}(i, j) = d(x_i, y_j) + min\{\mathcal{C}(i-1, j-1), \mathcal{C}(i-1, j), \mathcal{C}(i, j-1)\}$.

Finally, once accumulated cost matrix is built, the optimal warping path is found applying backtracking from $p_K = (M, N)$ to $p_1 = (1, 1)$ (Figure 9). And, the cost of the sequence matching, accumulated in p_K, is used to measure the correct achievement of the proposed physical exercises.

Moreover, we applied a context-based enhancement to the original DTW, a feature weighting [17], that basically consists of giving more relevance to

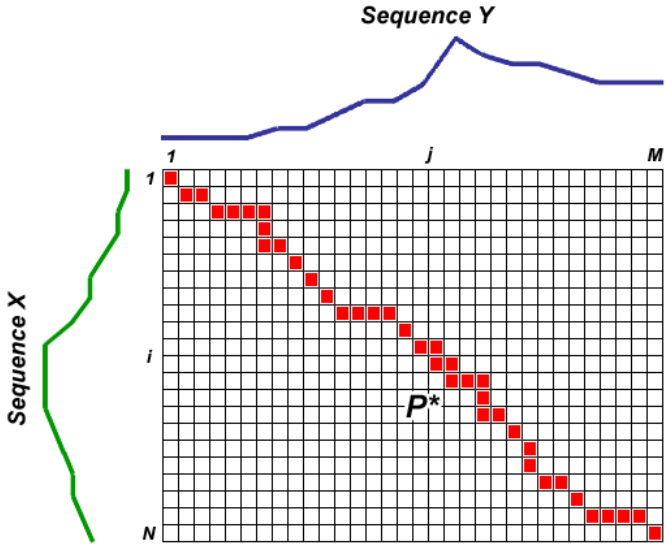


Figure 9: Warping path between two sequences, X and Y. Each cell has associated a cost and the path goes through the cells that minimize the cumulative cost.

the joints that help to improve the inter-class discrimination among gestures.

4. Results

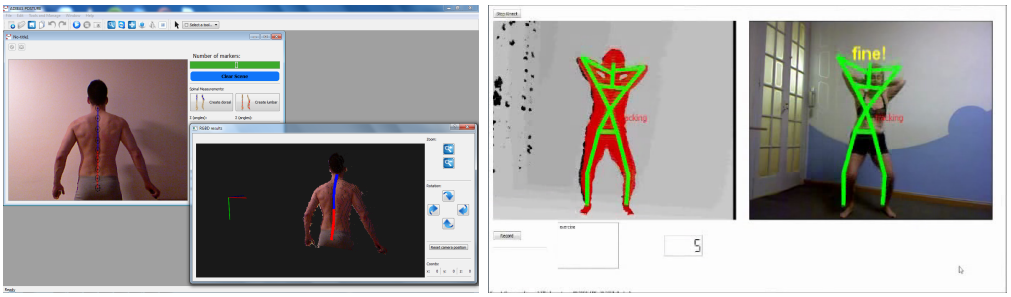
In order to present the results, we discuss the software details and validation procedures.

4.1. Software details

As a result of this work, we obtained a system that combines hardware - the Kinect™ device - and software.

Regarding the software implementation, we used the Kinect™ SDK framework. We also used the Point Cloud Library (PCL) to treat cloud points, and to support a free and three-dimensional visualization we used the Visualization Toolkit library (VTK). The user interface has been developed in the multi-platform Digia Qt technology. The programming language used was C++. The system runs fluently on a standard computer Intel Core i7-3770 CPU 3.90GHz with 8GB RAM.

In Figure 10, it is shown an example of interaction with the system through its graphical user interface.



(a) (b)

Figure 10: (a) Interaction with the system while performing an anatomic spine analysis. (b) A patient is squatting in a physical rehabilitation treatment and receiving feedback from the system.

4.2. System validation

The data described previously, in material and methods section, has been defined and used for the validation of the system.

Results for different distance of the device to the scene are shown in Table 1 and 2 for pose analysis and range of movement analysis, respectively. AAV and '∘' correspond to the average absolute value and degree, respectively. The relative mean errors are computed in % from the computed estimations and the real estimations for the distances and degrees, respectively. In the case of range analysis, the degrees are computed among the tracked joints at the beginning and end of the labeled sequences, meanwhile in the posture analysis the degrees are computed as the angle among labeled landmarks in the same image. These analysis validates the accuracy of the SPA and RMA in millimeters and degrees, respectively. Note the high precision in both tests. In addition, in order to validate the curvature analysis of the spine (SCA), we used a group of 10 patients and performed the Leroux protocol [10], placing nine markers over the spine. The relationship between lateral radiographic and anthropometric measures was assessed with the mean difference. It has used Cobb technique on the lateral radiograph in order to obtain the coefficients of kyphosis and lordosis. The results of the SPA validation are shown in Table 3. Moreover, after discussing with specialists in physiotherapy they agreed that the accuracy of the results is more than sufficient for diagnostic purposes.

The gesture recognition framework included in the range of movement analysis has been validated using the data set of 25 sequences previously

Distance subject-device (m)	1,3	1,9	2,2
AAV (mm)	0,98	1,42	2,1
Relative mean error distance (%)	1,01	1,18	1,71
AAV (∘ angles)	0,51	1,04	1,24
Relative mean error angle (%)	0,46	0,77	1,3

Table 1: Pose estimation.

Distance subject-device (m)	1,3	1,9	2,2
AAV (∘ movement)	2,2	3,8	5,2
Relative mean error angle (%)	0,76	0,88	1,45

Table 2: Range of movement precision.

introduced. The sequences may have different sizes since they can be aligned using DTW. Nevertheless, adjusting the sensitivity for intra and inter-class gesture discrimination is not straightforward. This sensitivity is adjusted by a gesture similarity threshold based on the gesture alignment cost. For this purpose, we computed the value of the gesture alignment cost μ over a leave-one-out cross-validation (LOOCV). We trained for different estimated values of μ. As a validation measurement, we compute the confusion matrix for each test sample of the LOOCV strategy. Being G the number of gestures, all test confusion matrices are added in a performance matrix $\mathcal{P}_G$, and the final accuracy $\mathcal{A}$ per gesture (exercise) is computed as follows,

$$\mathcal{A} = 100 \cdot \frac{Trace(\mathcal{P}_G)}{NC + \sum_{i=1}^{G} \sum_{j=1}^{G} \mathcal{P}_G(i,j)} \tag{11}$$

where NC is the number of samples of the data set not classified as any gesture because achieving an alignment cost over μ. Eventually, μ that gave the better results for our data set was selected as the optimal one. The obtained results applying the feature-weighted DTW (with the optimal μ) on the defined data set, shown in Figure 12, encouraged us to finally complement

	Khyposis range	Lordosis range
AAV (∘)	5	6

Table 3: Validation of spinal analysis.

the RMA module with gesture recognition.

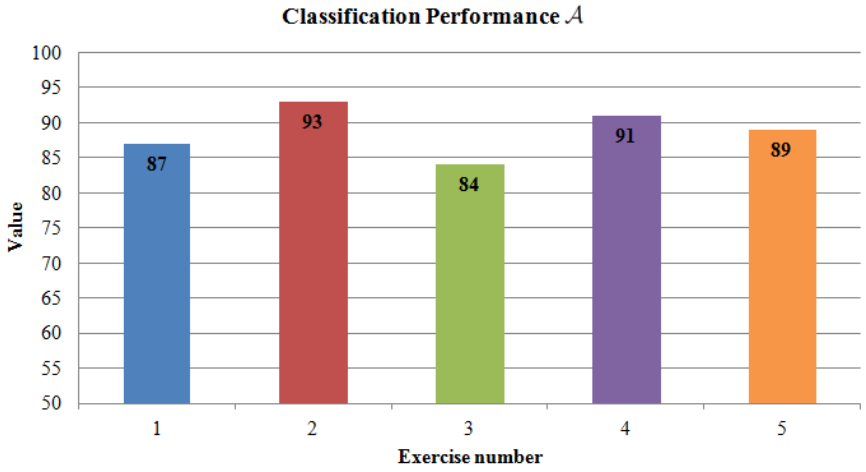


Figure 11: Classification performance $\mathcal{A}$ over the exercises data set for the five exercise categories using feature-weighted DTW.

Finally, in Figure 12 we show the computational complexity in terms of execution time of the different techniques and procedures involved in our system. One can see the fast computation of the different modules that though in some cases does not achieve real time it is suitable for real application purposes. As commented above, the times were computed in a conventional computer Intel Core i7-3770 CPU 3.90GHz with 8GB RAM.

4.3. Applications

The main scenarios of application of our system are posture analysis, physical rehabilitation, and fitness conditioning. In Figure 13, some real examples in which we tested our system are shown.

5. Conclusions

We designed a comprehensive posture analysis tool. This tool is a breakthrough in the field of rehabilitation research. The system allows for semi-automatic static posture and spine curvature analysis, as well as for automatic range of movement estimation and gesture recognition, from 3D anthropometric data acquired by a low cost RGB-Depth camera. The aim of

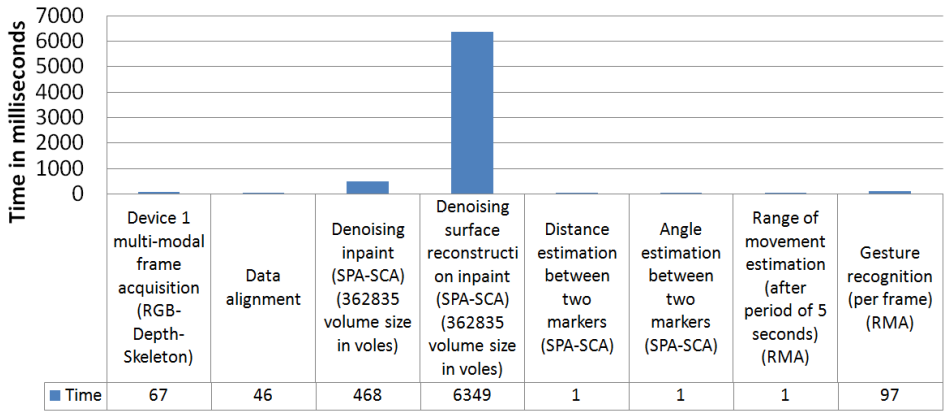


Figure 12: Computational time of the different steps and techniques of our system in milliseconds.

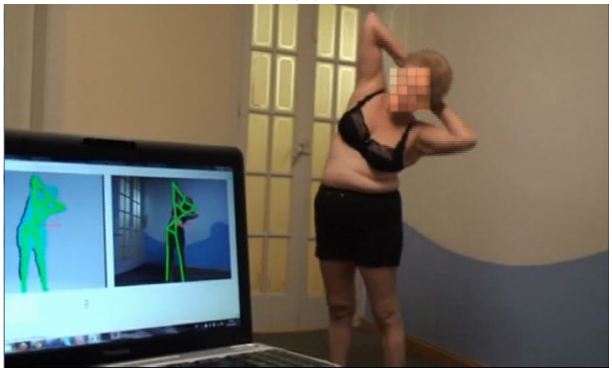
the system is to assist in the prevention and treatment of musculo-skeletal dysfunctions. We integrated several frontier computer vision and artificial intelligence techniques including three-dimensional visual data processing, statistical learning, and time series analysis. Furthermore, the system is meant to be highly adaptable and customizable to the needs of the therapist. The validation study shows high precision and reliable measurements in terms of distance, degree, range of movement estimation, and gesture recognition. Supported by clinical specialists, it is suitable to be included in the clinical routine, including posture reeducation, rehabilitation, and fitness conditioning scenarios.

Glossary

Finally, we provide a list of field-specific terms used in the article:

- *Structured light*: The process of calculating the depth and surface information from the deformation seen in a known projected pattern of pixels (often grids or horizontal bars) to the scene.
- *Musculo-skeletal disorders*: Disorders that can affect the body's muscles, joints, tendons, ligaments, and nerves. Health problems range from discomfort, minor aches and pains, to more serious medical conditions.

(a)



(b)

(c)

Figure 13: The system has been successfully applied in different real case scenarios: (a) posture reeducation, (b) physical rehabilitation, and (c) fitness conditioning.

- *RGB image*: Image represented by the RGB color model, that is an additive color model in which red, green, and blue light are added together to reproduce a broad array of colors.
- *Depth map*: An image that contains information relating to the distance of the surfaces of scene objects from a viewpoint.
- *HSV image*: HSV from hue, saturation, and value. It is a common cylindrical-coordinate representations of points in an RGB color model. It rearranges the geometry of RGB in an attempt to be more intuitive and perceptually relevant than the cartesian (cube) representation.
- *Ground truth data*: Information acquired by field study for the purpose of validation of obtained results.
- *Feature*: In computer vision, denotes a piece of information which is relevant for solving the computational task. It can refer to (a) the result of a general neighborhood operation applied to an image, or (b) specific structure in the image itself (points, edges, etc). Other examples of features are related to motion in image sequences, to shapes, or to properties of such a region.
- *Feature vector*: If two or more different features are extracted resulting in two or more feature descriptors at each image point, then a common practice is to organize the information provided by all these descriptors as the elements of one single vector.
- *Feature space*: The set of all possible feature vectors.
- *Statistical learning*: A framework for machine learning drawing from the fields of statistics and functional analysis. Statistical learning theory deals with the problem of finding a predictive function based on data.
- *Time series analysis*: Comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data.
- *Trace*: In linear algebra, the trace of an square matrix B is defined to be the sum of the elements on the main diagonal (the diagonal from the upper left to the lower right) of B.

Acknowledgements

This work is partly supported by projects IMSERSO-Ministerio de Sanidad 2011 Ref. MEDIMINDER, and RECERCAIXA 2011 Ref. REMEDI.

References

[1] European Agency for Safety and Health at Work. Work-related musculoskeletal disorders. Fact and figures. Accessed 06/04, 2011.

[2] I. Freedman Barak Binyamina, U. Shpunt Alexander Cambridge Ma, I. Machline Meir Ashdod, and I. Arieli Yoel Jerusalem, Depth mapping using projected patterns, Patent US20080240502, 2010.

[3] J. Starck, A. Maki, S. Nobuhara, A. Hilton, and T. Matsuyama, The Multiple-Camera 3-D Production Studio, IEEE Transactions on circuits and systems for video technology, vol 19, no. 6, June 2009.

[4] I. Mikic, M. Trivedi, E. Hunter, and P. Cosman, Articulated Body Posture Estimation from Multi-Camera Voxel Data, CVPR, pp. 455-460, 2001.

[5] K. Parsa, J. Angeles, and A.K. Misra, Pose-and-Twist Estimation of a Rigid Body Using Accelerometers, International Conference on Robotics and Automation - ICRA, vol. 3, pp. 2873-2878, 2001.

[6] A.S. Merians, D. Jack, R. Boian, M. Tremaine, G.C. Burdea, S.V. Adamovich, M. Recce, and H. Poizner, Augmented rehabilitation for patients following stroke, Journal of the American Physical Therapy Association, 2006.

[7] B. Sabata, F. Arman, and J.K. Aggarwal, Segmentation of 3d range images using pyramidal data structures, CVGIP: Image Understanding 57(3), pp. 373-387, 2002.

[8] Microsoft® Corporation. Kinect™ for Windows SDK beta programming guide beta 1 draft version 1.1. 2012.

[9] J. Shotton, A. W. Fitzgibbon, M. Cook, and T. Sharp, Real-time human pose recognition in parts from single depth images, CVPR, pp. 1297-1304, 2011.

[10] M. A. Leroux and K. Zabjek, A noninvasive anthropometric technique for measuring kyphosis and lordosis: application for scoliosis, Spine, PubMed MEDLINE, vol. 25, pp. 1689-1694, 2000.

[11] G. Mori, S. Belongie, and J. Malik, Efficient shape matching using shape contexts, TPAMI, vol. 27, pp. 1832-1837, 2005.

[12] M. Potmesil and I. Chakravarty, A lens and aperture camera model for synthetic image generation, vol. 15, pp. 297-305, ACM. 1981.

[13] R. San Jose Estepar, A. Brun, and C.-F. Westin, Robust generalized total least squares iterative closest point registration, In Seventh International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI04), pp. 234-241, Rennes - Saint Malo, France, 2004.

[14] J. Yao, M. R. Ruggeri, and P. Taddei, Automatic scan registration using 3d linear and planar features, vol. 1, no. 3, pp. 1-22, Secaucus, Springer-Verlag. 2010.

[15] Z. Zhang, A flexible new technique for camera calibration, TPAMI, vol. 22, no. 11, pp. 1330-1334, 2000.

[16] M. Parizeau and R. Plamondon, A comparative analysis of regional correlation, dynamic time warping, and skeletal tree matching for signature verification, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 7, 1990.

[17] M. Reyes, G. Domínguez, S. Escalera, Feature weighting in dynamic time warping for gesture recognition in depth data, 1st IEEE Workshop on Consumer Depth Cameras for Computer Vision, International Conference in Computer Vision, 2011.

[18] P. Senin, Dynamic Time Warping Algorithm Review, No. CSDL-08-04. 2008.

[19] J. C. Carr, R. K. Beatson, J. B. Cherrie, Reconstruction and representation of 3D objects with radial basis functions, Proceedings of SIGGRAPH 2001, pp. 417-424, 2001.

[20] W. Wang and X. Qin, Image Inpainting Algorithm Based on CSRBF Interpolation, International Journal of Information Technology, vol. 12, no. 6, 2006.

[21] M. Betsch, M. Wild, P. Jungbluth, M. Hakimi, J. Windolf, B. Haex, T. Horstmann, and W. Rapp, Reliability and validity of 4D rasterstereography under dynamic conditions, Computers in biology and medicine, vol. 41, pp. 308-12, 2011.

[22] D.E. Harrison, T.J. Janik, R. Cailliet, D.D. Harrison, M.C. Normand, D.L. Perron, and P.A. Oakley, Upright static pelvic posture as rotations and translations in 3-dimensional from three 2-dimensional digital images: validation of a computerized analysis, Journal of manipulative and physiological therapeutics, vol. 31, pp. 137-145, 2008.

[23] J.L. Jaremko, P. Poncet, J. Ronsky, J. Harder, J. Dansereau, H. Labelle, and R.F. Zernicke, Indices of torso asymmetry related to spinal deformity in scoliosis, Clinical Biomechanics, vol. 17, pp. 559-568, 2002.

[24] J.L. Jaremko, P. Poncet, J. Ronsky, J. Harder, J. Dansereau, H. Labelle, and R.F. Zernicke, Estimation of spinal deformity in scoliosis from torso surface cross sections, Spine, vol. 26, pp. 1583-, 2001.

[25] P.O. Ajemba, L. Ramirez, N.G. Durdle, D.L. Hill, and V.J. Raso, A Support Vectors Classifier Approach to Predicting the Risk of Progression of Adolescent Idiopathic Scoliosis, IEEE Transactions on Information Technology in Biomedicine, vol. 9, pp. 276-282, 2005.

[26] H. Wu, J.L. Ronsky, F. Cheriet, J. Küpper, J. Harder, D. Xue, and R.F. Zernicke, Prediction of scoliosis progression with serial three-dimensional spinal curves and the artificial progression surface technique, Medical and Biological Engineering and Computing, vol. 48, pp. 1-11, 2010.

[27] C. Fortin, D.E. Feldman, F. Cheriet, D. Gravel, F. Gauthier, and H. Labelle, Reliability of a quantitative clinical posture assessment tool among persons with idiopathic scoliosis, Physiotherapy, 2011.

[28] C. Fortin, D.E. Feldman, F. Cheriet, and H. Labelle, Validity of a Quantitative Clinical Measurement Tool of Trunk Posture in Idiopathic Scoliosis, Spine, vol. 35, pp. 988-, 2010.

[29] P.O. Ajemba, N.G. Durdle, D.L. Hill, and V. James Raso, Validating an imaging and analysis system for assessing torso deformities, Computers in Biology and Medicine, vol. 38, pp. 294-303, 2008.

[30] P.O. Ajemba, N.G. Durdle, and V. James Raso, Clinical monitoring of torso deformities in scoliosis using structured splines models, Medical and Biological Engineering and Computing, vol. 46, pp. 1201-1208, 2008.

[31] P.O. Ajemba, N.G. Durdle, and V.J. Raso, Characterizing Torso Shape Deformity in Scoliosis Using Structured Splines Models, Biomedical Engineering, IEEE Transactions on, vol. 56, pp. 1652-1662, 2009.

[32] B. Drerup and E. Hierholzer, Back shape measurement using video rasterstereography and three-dimensional reconstruction of spinal shape, Clinical Biomechanics, vol. 9, pp. 28-36, 1994.

[33] Chris C. Martin et al., A Real-time Ergonomic Monitoring System using the Microsoft Kinect, in Proceedings of the 2012 IEEE Systems and Information Engineering Design Symposium, 2012.

[34] E. Ferreira, M. Duarte, E. Maldonado, A. Bersanetti, and A. Marques, Quantitative assessment of postural alignment in young adults based on photographs of anterior, posterior, and lateral views, J Manipulative Physiol Ther. vol 34. pp. 371-80, 2011.

[35] R. Clark, A. Bryant, Y. Pua, P. McCrory, K. Bennell, and M. Hunt, Validity and reliability of the Nintendo Wii Balance Board for assessment of standing balance, Gait and Posture, vol. 31, pp. 307317, 2010.

[36] Kourosh Khoshelham and Sander Oude Elberink, Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications, Sensors, vol. 12, pp. 1437-1454, 2012.

[37] Ross A. Clark, Yong-Hao Pua, Karine Fortin, Callan Ritchie, Kate E. Webster, Linda Denehy, and Adam L. Bryant, Validity of the Microsoft Kinect for assessment of postural control, Gait and Posture, 2012.

[38] V.N. Sarnadskiy, Classification of postural disorders and spinal deformities in the three dimensions according to computer optical topography, Stud Health Technol Inform, vol. 176, 2012.