Advanced 3D movement analysis algorithms for robust functional capacity assessment

Asma Hassani[1]; Alexandre Kubicki[2]; France Mourey[3,4]; Fan Yang[1]

[1]Le2i UMR6306, CNRS, Arts et Métiers, Univ. Bourgogne Franche-Comté, France;
[2]IFMS Montbeliard, France;
[3]Institut National de la Santé et de la Recherche Médicale (INSERM) U1093, Cognition Action et Plasticité Sensori-Motrice, BP 27877, campus universitaire, Université de Bourgogne, Dijon, France;
[4]Faculté de Médecine, Université de Bourgogne, Dijon, France

Keywords

Patient self-care home care and e-health, clinical informatics, sitting posture recognition, skin detection, 3D real-time video processing

Summary

Objectives: We developed a novel system for in home functional capacities assessment in frail older adults by analyzing the Timed Up and Go movements. This system aims to follow the older people evolution, potentially allowing a forward detection of motor decompensation in order to trigger the implementation of rehabilitation. However, the pre-experimentations conducted on the ground, in different environments, revealed some problems which were related to KinectTM operation. Hence, the aim of this actual study is to develop methods to resolve these problems.
Methods: Using the KinectTM sensor, we analyze the Timed Up and Go test movements by measuring nine spatio-temporal parameters, identified from the literature. We propose a video processing chain to improve the robustness of the analysis of the various test phases: automatic detection of the sitting posture, patient detection and three body joints extraction. We introduce a realistic database and a set of descriptors for sitting posture recognition. In addition, a new method for skin detection is implemented to facilitate the patient extraction and head detection. 94 experiments were conducted to assess the robustness of the sitting posture detection and the three joints extraction regarding condition changes.
Results: The results showed good performance of the proposed video processing chain: the global error of the sitting posture detection was 0.67%. The success rate of the trunk angle calculation was 96.42%. These results show the reliability of the proposed chain, which increases the robustness of the automatic analysis of the Timed Up and Go.
Conclusions: The system shows good measurements reliability and generates a note reflecting the patient functional level that showed a good correlation with 4 clinical tests commonly used. We suggest that it is interesting to use this system to detect impairment of motor planning processes.

Correspondece to
Asma Hassani
Université de Bourgogne, Laboratoire LE2I, Bâtiment I3M,
64 rue de Sully,
21000, Dijon.
Tel.: +33 3 80 39 36 08
Fax: +33 3 80 39 59 10
Email: hassani.asma@gmail.com

Appl Clin Inform 2017; 8: 454–469
https://doi.org/10.4338/ACI-2016-11-RA-0199
received: November 29, 2016
accepted: February 11, 2017
published: May 10, 2017
Citation: Hassani A, Kubicki A, Mourey F, Yang F. Advanced 3D movement analysis algorithms for robust functional capacity assessment. Appl Clin Inform 2017; 8: 454–469
https://doi.org/10.4338/ACI-2016-11-RA-0199

© Schattauer 2017

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

1. Introduction

Geriatric rehabilitation constitutes a major public health issue. The biomechanical deficits such as the loss of muscle mass and muscle power, are considered to play a crucial role in the frailty process [1]. However, motor planning impairments could also lead to declines in physical functions [2]. In this context of functional capacities impairment, physiotherapists and other health professionals are interested in motor function in order to maintain or improve it as much as possible. The functional abilities assessment is mainly conducted by health professionals in care units. However, home care is most often the wish of older adults to maintain their independence and thus to help improve the quality of life [3]. It seems then interesting to move this analysis as much as possible to the frail elderly patients' homes.

The Timed Up and Go (TUG) is a quick and simple clinical test that allows a qualitative analysis of the patient's stability during the various test phases [4–6] and has been shown to predict falls risk in the elderly [7–9]. Moreover, it requires no special equipment or training and the risk of musculo-articular injury is low when performing the TUG [10]. This test consists in standing up from a chair, walking a distance of 3m, turning, and walking back to the chair and sitting down. It includes the Sit-To-Stand and Back-To-Sit transfers that are important activities of independent daily living and were studied in older adults to assess aging effects on motor planning processes through their kinematic features [11–13]. For instance, Mourey et al [11] noted an age-related slowdown when performing the Back-To-Sit, which was attributed to more cautious behavior related to lack of visual information and probably to a difficulty in dealing with the gravity effects. In the study by Dubost et al [12], trunk angles in normal older adults were smaller than in young subjects during the Back-To-Sit. This lack of trunk tilt was explained by a non-optimal behavior related to changes in motor planning processes and that aims to decrease the risk of anterior disequilibrium. The results of both these studies showed that older adults had a greater difficulty to perform the Back-To-Sit. For these reasons, we chose to reproduce an automatic analysis of the TUG, which is satisfactory for our purposes: to allow functional capacities assessment, to be safe, easy and can be made without the direct participation of a health professional.

Thus, for making the patient involved in his own care to optimize his following rehabilitation at home and maintain his functional independence, we developed a low-cost, robust and home-based system for real-time 3D TUG movements analysis to assess functional capacities in the elderly. The system includes the Kinect™ sensor to track the patient's 3D skeleton without placing markers on his body. Indeed, Kinect™ has been used in several applications such as students' physical rehabilitation in schools [14], the evaluation postural control [15], fall detection [16], static lifting movements' assessment [17] and face recognition [18]. This sensor can also integrate complex and continually adaptive exercises requiring specific movements and track the extent to which these movements are performed [19]. In addition, Kinect™ is a low-cost and portable device that combines an RGB camera, a depth sensor and a multi-array microphone. It provides inexpensive depth sensing for a large variety of emerging applications in computer vision, augmented reality and robotics.

Three experiments allowed the TUG analysis of young subjects, frail and non-frail aged adults using the proposed system in heterogeneous environments: patients' homes [2], laboratory [20] and geriatric day-hospital [20]. As a first step, the aim was to verify the adaptation of the system in different places and with different subjects. They showed a good measurement reliability of the identified parameters. Then, we measured the influence of age-related frailty effects on motor planning processes through the kinematic features of the Sit-To-Stand and Back-To-Sit transfers in order to weight the different parameters related to the functional level of the subject and thus assign a motor control note during the automatic analysis of the TUG. The results showed that frail patients with the lowest functional level reached the lower trunk angle during the Back-To-Sit. In addition of motor control parameters, the most discriminating criterion between frail aged adults and young subjects was the TUG duration. Based on the different results of these experiments, we introduced a motor control note that reflects the patient functional level and correlates with 4 clinical tests commonly used [21].

However, the field tests also revealed some malfunction of Kinect™. For instance, when the subject stands up or sits down with a large trunk tilt, the Kinect™ cannot correctly identify the center of mass and shoulders. Thus, we propose a video processing chain applied to the color image stream

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

and the depth map provided by Kinect™ for improving the robustness and accuracy of the system: automatic detection of the sitting posture, patient detection and three body joints extraction.

2. The Timed Up and Go description and experimental setup

The TUG is a clinical measure of balance and mobility in the elderly and in neurological populations [22]. The time taken to complete the test allows predicting the risk of falling [7, 23]. The average TUG duration is 25.8s in non-fallers, 33.2s in subjects having fallen once and 35.9s in multiple fallers [24]. Moreover, a score between 13.51s and 35.57s is consistent with a frail subject and a score less than or equal to 13.5s correlates with the locomotor independence state [8]. The TUG movements allow estimating nine spatio-temporal parameters that were identified in the literature as relevant for balance assessment: a) movement duration, b) trunk angle, c) ratio, d) shoulder path curvature and e) TUG duration. The first four parameters were calculated for each Sit-To-Stand and Back-To-Sit.

Our system includes the Kinect™ sensor for detecting the subjects' movements. This sensor produces accurate results, especially when tracking shoulder movements (segment lengths and angle estimation) [25, 26]. It was placed at a height of 50-60cm from the ground and at a distance of 2-2.5m from the chair with a tilt angle of 20° (▶Figure 1). No markers or wearable sensors were attached to the participant body. The Kinect™ skeleton data are used for the real-time calculation of the balance assessment parameters, which starts (ends) when the sitting posture is recognized. Extracted features correspond to the shoulder displacement kinematics during the Sit-To-Stand and Back-To-Sit and the TUG duration that corresponds to the time interval between the moment when the forward phase starts and the moment when the backward phase ends. The shoulder movement duration during Sit-To-Stand corresponds to the time interval between the moment when the shoulder depth component (anterior–posterior axis) exceeds 8.5% of its initial position, corresponding to the lift-off of the buttocks from the seat and the moment when the head vertical component reaches or exceeds 94% of the person's size (i.e., when the maximum hip, trunk, knee extension and maximum head flexion velocity are reached). The thresholds were experimentally determined. In the Back-To-Sit, it is defined as the time interval between the moment when the shoulder vertical component drops its peak value and the moment when the hip vertical components reach their minimum values and the trunk angle reaches its limit. The movement duration was measured in seconds.

The trunk angle corresponds to the maximal trunk angle reached by participants during each transfer. These maximal trunk angles, computed in the sagittal plane, were measured in degrees between the trunk axis and the vertical axis passing through the center of mass of the body. Concerning the ratio, it represents the ratio between the shoulder vertical and horizontal movements durations. As regards the shoulder path curvature, shoulder paths during forward and backward displacements were similar and almost straight, therefore the curvatures of path for upward and downward displacements were only calculated [27]. Curvature is defined as:

$$cur = Dmax/L$$

where L corresponds to a straight line passing between the initial and the final position of shoulder displacement and Dmax means the maximal perpendicular distance measured from the actual path to the straight line.

In the next section we will present the processing chain proposed to overcome the problems encountered during the various experiments that are related to Kinect™.

3. Overview of the video processing system

Experiments were performed in a geriatric day-hospital to test the system in a real environment, its installation requirements and its adaptation to the different types of patients. Through these experiments, some limits and constraints related to Kinect™ have been identified. Indeed, when the trunk inclination is greater than 70°, Kinect™ cannot correctly identify the center of mass and shoulders. Also, when wearing loose clothing or the subject suffers from a significant genu valgus, it is sometimes difficult to detect the correct positions of the joints constituting the skeleton. Similarly, if the

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

person uses a cane or in case of the close presence of a caregiver, the sensor cannot properly dissociate the subject. These problems have an impact on measures to perform. Therefore, we propose a video processing chain to resolve them, which consists of a sitting posture detection method and an algorithm for 3 body joints extraction.

▶Figure 2 shows an overview of the operating process of the system. The sitting posture detection method is used to trigger the TUG analysis and detect its end. It is based on the Support Vector Machine (SVM) classification method. Regarding the joints extraction method, it is applied only during the Sit-To-Stand and Back-To-Sit when the 3D skeleton produced by Kinect™ is poorly detected. The skeleton is poorly detected if:

$$\begin{cases} (H_X, H_Y, H_Z) = (0,0,0) \text{and } (S_X, S_Y, S_Z) = (0,0,0) \\ \text{or} \\ Hd_Y < y_{th} \\ \text{or} \\ \text{trunk angle} > \text{angle}_{th} \end{cases}$$

where H, S and Hd represent the mass center, the shoulder center and the head center, respectively. y_{th} and angle_{th} are threshold values of the vertical component of the head and the trunk angle, respectively that were determined empirically. We first carry out a patient detection and then compute the center of mass. Finally, a method for detecting the head and shoulders is applied to extract the positions of their centers. Thanks to these points, we can track shoulder movements during the two transfers and hence calculate the spatio-temporal parameters.

4. Sitting posture detection

The human movements' interpretation and analysis can be performed by using 3D parameters such as joint angles and positions, which require a 3D tracking of the entire body or some of its parts. The proposed method consists in representing the sitting posture based on a set of characteristics extracted from 3D skeleton joints. Indeed, this posture is a rest position in which the body rests on the bottom, with the trunk vertically or with a slight body bending forward or backward and also characterized by knees flexion. Based on these characteristics, a total of 16 features has been extracted for each frame for representing the sitting posture (▶Figure 3):

- The trunk angle θ,
- The angle between trunk and leg β,
- The distance between head and hip center D_{HdH},
- The differences D_{HKL} and D_{HKR} between the distance between hip center and knee at the y-axis and that at x and z axes, for the left and right body,
- The distance between shoulder center and hip center at x-axis D_{SHx},
- The distance between shoulder center and hip center at z-axis D_{SHz},
- The 3D coordinates of the head, the shoulder center and hip center: H_{dx}, Hd_y, Hd_z, S_x, S_y, S_z, H_x, H_y and H_z.

The SVM classifier with the radial basis function kernel has been used to classify two postures: sitting and non-sitting. To get closer to a real operation of the system under realistic conditions, the training data are retrieved from 15 individuals performing the TUG in different environments, various illuminations and different conditions. Performance evaluation of different combinations is based on the calculation of sensitivity, specificity, recall, precision and the global error.

We built our own dataset for sitting posture recognition in several conditions and environments to train the classifier and evaluate descriptors. We acquired a total number of 1611 training vectors containing sitting position and several motions such as body transfers and walking. The experiences were performed by taking account of the main difficulties of realistic TUG test execution by older adults at home environment. In order to test the sitting posture detection method, 12 participants wearing different clothes including ample clothes performed various TUG tests in different conditions: 2 persons in the Kinect™ field of view, a great trunk tilt, lower limbs completely glued and variable illumination and environment (home, laboratory). They aged twenty six to fifty years.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

5. Person tracking method presentation

5.1 Related works

The purpose of this step is to locate the target person in the scene and extract three 3D points corresponding to three human body joints: the mass center, the head center and the center of the line between both shoulders (shoulder center). There are several approaches for people detection and tracking in the literature of computer vision and robotics, which can be classified into two broad categories: motion-based analysis approaches [28-30] and appearance-based approaches [31, 32]. For the first category, the motion detection consists in segmenting the moving regions to locate moving objects in a sequence. These approaches can be classified into three categories: background subtraction, methods based on optical flow calculation and those based on temporal difference.

The background subtraction method consists in carrying out the difference between the current image and a background image that has been modeled previously. The quality of the extracted regions depends on that of the background image modeling. This method requires a reference image that is difficult to obtain and should be updated during the sequence to take into account possible changes such as moving objects and the illumination change. Thus, the background subtraction difficulty lies not only in the subtraction but also in background maintenance [33, 34]. These methods can be very effective in scenes where the background is well known and whose appearance does not change much over time.

With respect to the methods based on the optical flow, they consist in calculating at time t the displacement d of point $p = (x, y)$. The optical flow calculation is particularly useful when the camera is moving, but its estimation is both expensive in terms of calculation time and very sensitive to high amplitude movements. In addition, estimates are generally noisy at the borders of moving objects and difficult to obtain in large homogeneous regions. It also assumes that the differences of images can be explained as a consequence of a movement, while they can also be related to changes in the characteristics of objects, backgrounds and lighting.

Regarding the temporal difference, it consists in detecting the movement region based on the differences of successive images. This method can detect moving objects with low computational cost. However, the simultaneous extraction of fast objects and slow objects is usually impossible and therefore, it is difficult in this case to find a compromise between the number of missed targets and false detections.

Based-appearance approaches can be global or local. Global approaches, such as Principal Component Analysis, involve taking a single decision for the entire image. Regarding local approaches, we distinguish between methods based on extraction points or areas of interest and methods based on regular path of the image. The spatio-temporal interest points are widely used in the recognition of human actions and movements. In [35], the author proposed a method of spatio-temporal detection of local areas where there is a strong spatio-temporal joint variation. This represents an extension of the method of detecting the spatial points of interest of Harris and Stephens [36] and Fostner et al [37]. Indeed, the interest points correspond to a strong local spatial variation (edges, corners, textures …).

Among the methods based on regular path of the image, the most popular method is that proposed by Viola and Jones [38], which is particularly characterized by its speed. It is based on Haar features to locate faces and uses integral images to calculate the characteristics. Training and feature selection are performed by the AdaBoost in cascade: at each stage of the cascade, the search area is increasingly reduced by eliminating a large portion of the areas not containing faces and then classifiers become more complex.

5.2 Skin detection

Skin detection is an essential step in the person tracking algorithm since it reduces the search area of subjects in the image and facilitates the head detection. There are several methods to distinguish the skin regions from the rest of the scene and build a skin color model [39, 40]. In this study, we adopted a method called Explicit Skin Cluster that consists in explicitly defining the boundaries of the skin area in an appropriate color space. The advantage of methods using the pixel tone is the

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

simplicity of skin detection rules used, leading to a fast classification and that they require no prior training. However, their major problem is the difficulty in empirically determining a color space and the relevant decision rules that provide a high recognition rate. Although the color of the skin can vary significantly, recent research shows that the main difference is in the intensity rather than chrominance [41]. Various color spaces are used to label the pixels as color skin pixels such as RGB [42], HSV [43] and YCbCr [44]. However, we must choose the most robust space to adapt different conditions such as the distance between the subject and the sensor.

In this study, we take in account of stern constraints for robust skin region extraction facing some conditions: brightness change, distance from the sensor and similarity of clothes or background colors and skin color (▶Figure 4). We conducted a combination of two color spaces: $\tilde{H}_\Delta\tilde{S}_\Delta\tilde{L}_\Delta$ of the Color Logarithmic Image Processing (CoLIP) framework [45] and RGB. $\tilde{H}_\Delta$, $\tilde{S}_\Delta$ **and** $\tilde{L}_\Delta$ represent the hue, the saturation and the lightness, respectively. The segmentation procedure consists in finding the pixels that meet the following constraints:

$$\begin{cases} (\tilde{H}_\Delta \in [0.54, 0.87] \; and \; \tilde{S}_\Delta \in [0.01, 0.35]) \\ and \; (R > 72 \; and \; R > G \; and \; R > B \; and \; G > 41 \; and \; B > 19 \; and \; R - G > 15) \\ \quad and \; \neg(|R - G| < 31 \; and \; B < 50) \end{cases}$$

The idea is to combine hue and saturation components of CoLIP model with the R, G and B components of the RGB space to have a variable domain for skin color and thus improve the robustness of the detection method. Indeed, the hue is in relation to saturation which itself depends on the luminance: when the luminance is close to 0 or 1, the dynamics of saturation decreases and the hue contains information that is increasingly irrelevant. Hue physically corresponds to the dominant wavelength of a color stimulus. Saturation is the colorfulness of an object relative to its own brightness and measures the color purity. Thus, the combination of hue and saturation defines a fixed area of skin color. The objective of the R, G and B is to privilege some colors and to neglect others.

A median filter is applied to the image in CoLIP space ($\tilde{a}, \widetilde{rg}, \widetilde{yb}$) in order to avoid the noise caused by the image acquisition conditions. The proposed approach is based not only on color information as traditional skin detection methods, but also on the depth and area of the regions detected as skin (▶Figure 5).

5.3 Joints extraction

In this study, we focus on appearance-based approaches. The motion information is not used since it does not allow extracting simultaneously fast and slow objects. The aim is to distinguish the older person from the background during the completion of a clinical test. The way to perform the Sit-To-Stand and Back-To-Sit transfers varies from person to another, depending on their functional abilities. For example, it may be fast when the person drops the chair and slow when it has trouble getting up. Since color is very important information to better understand and interpret a scene and an essential element for the person detection in the image, the patient detection algorithm is based on the combination of the color image, represented in the CoLIP space (an appearance-based method) and the depth map provided by Kinect™ (see previous subsection). We chose the CoLIP framework that enables better segmentation after making a series of comparisons with the color spaces $L^*a^*b^*$ and HSL. We used its antagonist representation, represented by a logarithmic achromatic tone $\tilde{a}$ and two logarithmic chromatic antagonist tones denoted $\widetilde{rg}$ (red-green opposition) and $\widetilde{yb}$ (yellow-blue opposition) and its Hue/saturation/lightness representation $\tilde{H}_\Delta\tilde{S}_\Delta\tilde{L}_\Delta$ [46]. In the experience conducted, the CoLIP framework is more robust to changes in lighting.

The depth is also very important information. It reduces the person's search box, especially in our case since we know about the range of the distance between the patient and the sensor. We therefore use the depth map associating with each pixel of the color image, the distance between the object represented by this pixel and the sensor in mm. A morphological opening is applied to the depth map to smooth the depth values and reduce the noise due to its acquisition. The scene to be analyzed contains information about the human body, but also its environment. To keep only the human body information, a threshold is applied.

The patient detection algorithm comprises four main steps as follows (▶Figure 6):

1. Subtraction of a part of the background,
2. Subjects detection,

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

3. If the subjects number is 2, patient extraction,
4. Restoration of the human body missing regions.

In the first step, a threshold according to the objects depths in the image is made. The deleted part and the remaining part are denoted *DP* and *RP*, respectively. Then, we seek the regions belonging to *DP* and connected to *RP* according to these 2 criteria:

$$\begin{cases} d_{\tilde{H}\Delta\tilde{S}\Delta\tilde{L}\Delta}(c_i, c_j) < m1 \\ |\tilde{a}(i) - \tilde{a}(j)| \leq m2 \end{cases}$$

where $i \in RP$, $j \in DP$ and ã is the logarithmic achromatic tone. *m1* and *m2* are two empirical values. $d_{\tilde{H}\Delta\tilde{S}\Delta\tilde{L}\Delta}$ represents the distance between two colors c_i and c_j and is defined as follows:

$$d_{\tilde{H}\Delta\tilde{S}\Delta\tilde{L}\Delta}(c_i, c_j) = \sqrt{\left(c_i^{\tilde{L}\Delta} - c_j^{\tilde{L}\Delta}\right)^2 + \left(c_i^{\tilde{S}\Delta}\right)^2 + \left(c_j^{\tilde{S}\Delta}\right)^2 - 2c_i^{\tilde{S}\Delta} c_j^{\tilde{S}\Delta} \cos(c_i^{\tilde{H}\Delta} \div c_j^{\tilde{H}\Delta})}$$

The ÷ denotes the angular difference between two hues. Let h_i and h_j 2 hues ∈ [0°, 360°]. The difference between these two values [47] is defined as:

$$h_i \div h_j = \begin{cases} |h_i - h_j| \; if \; |h_i - h_j| \leq 180° \\ 360° - |h_i - h_j| \; if \; |h_i - h_j| > 180° \end{cases}$$

In other words, we seek the adaptive neighborhoods of each pixel x belonging to *DP*: with each point x of the image f is associated a set of adaptive neighborhoods belonging to the spatial support $D \subseteq R^2$ of f. A neighborhood V^h of x is a connected and homogeneous set with respect to an analysis criterion h. h is a combination of the brightness and the distance between the color of x and that of a neighboring pixel.

In the second step, the subjects' detection is based on the skin regions depth: an automatic thresholding is applied to the remaining regions after background subtraction (▶Figure 7). This thresholding depends on the maximum skin depth (max_d) and its minimum depth (min_d). Thus, the subjects corresponds to the connected regions whose depth ∈ [$min_d - m_0$, $max_d + m_0$] where m_0 is a tolerance value. Next, another thresholding is applied depending both on the surfaces and the dimensions of the objects.

The third step of the algorithm is to extract the patient as follows:

- If the number of connected objects is 2, the patient is the person to the left,
- Else (either a single person or 2 persons):
  - Cut the object into 2 parts according to its gravity center,
  - Calculate the distance d between the two peaks corresponding to 2 parts,
  - If $d < d_0$, there is one person, else, the patient corresponds to the left part, where d_0 is a threshold value.

In the fourth step, the purpose is to recover the missing body parts. It consists in finding the regions that are connected to the body under the following constraints:

$$\begin{cases} \tilde{H}_\Delta(i) \div \tilde{H}_\Delta(j) < h_0 \\ |\tilde{a}(i) - \tilde{a}(j)| \leq a_0 \\ |p(i)\text{-}p(j)| \leq p_0 \end{cases}$$

where p is the pixel depth. h_0, a_0 and p_0 are thresholds. Morphological filters to plug the holes, remove small items and classics morphological operations (dilation, erosion) are applied to obtain cleaned binary images that can be labeled.

After detecting the patient, we extract 3D points represented through *(X, Y, Z)* where *X* and *Y* are the point coordinates in the image and Z is its depth relative to the Kinect™ sensor. The mass center extraction is relatively simple. It is to seek the gravity center of the body region.

The head detection method consists in extracting the region that is defined as skin and satisfies the following criteria:

- C_1: region surface ∈ [s_c- s_1, $s_c + s_2$] where s_c represents 9% of the total body surface and s_1 and s_2 are 2 tolerance values.
- C_2: region size ∈ [$t1_c$, $t2_c$] where $t1_c$ and $t2_c$ are 1/10 and 1/7 of the body size, respectively.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

- C_3: region surface ∈ $[s_{ce} - sc_1, s_{ce} + sc_2]$ where s_{ce} is the surface of the circle whose diameter = region size. sc_1 and sc_2 are 2 tolerance values.

The C_1 criterion is based on the Wallace rule of nines that assigns 9% of the total body surface area for the head and neck. The C_2 criterion was defined based on anthropometry. Regarding the C_3 criterion, based on the assumption that at any camera angle where the head contour is visible, the head is assumed to be nearly a full circle, we calculate the surface of the circle whose diameter is the object height and we compare this surface with that of the object. Thus, the head position corresponds to the center of the detected object.

Based on anthropometric values corresponding to the body segments lengths [48], we delineated the shoulder area by the interval $[d_1/3.5, 1.65d_1/3.5]$ where d_1 is the distance between the top of the head and the mass center. The shoulder center corresponds to the center of this region.

6. Experiment evaluations

We firstly evaluated the sitting position detection process at the classifier output level by computing the precision, the specificity, the accuracy, the recall and the classification error rate. We based on the results of experiments conducted by the 12 subjects. The test data and the training data were composed of 6504 vectors and 1611 vectors of 16 attributes, respectively. The results obtained are presented in the ▶Table 1 and show the efficiency of the classifier to separate the two classes: sitting and non-sitting posture (error rate is 0.67%), which results subsequently, in a precision in the measurement of the duration of the TUG, the Sit-To-Stand and Back-To-Stand transfers.

Then, we evaluated the joints extraction method reliability by calculating the trunk angle. With this processing chain, 94 experiments were performed under the various conditions mentioned previously. We compared the trunk angles calculated according to the 3D skeleton provided by Kinect™ to those calculated based on 3D points resulting from the proposed method. ▶Figure 8 presents some joints extraction results. We applied this method on 84 images where trunk angles calculated using the Kinect™ skeleton were wrong. The success rate of the proposed method is 96.42%.

7. Conclusions and prospects

We developed a real-time 3D TUG test movement analysis system for in-home functional abilities assessment in older adults, using the Kinect™ sensor. This system allows to assign a motor control note indicating the motor frailty. However, field experiments revealed some limitations associated with Kinect™. Thus, we proposed a video processing chain in order to increase the robustness of this sensor and then that of the analysis of the various TUG phases. We developed a new method for detecting the sitting posture and evaluated its robustness using a realistic database. It showed good efficiency: the global error is 0.67% and seems acceptable to real applications of sitting posture detection.

We also implemented a robust method for detecting the skin region. This is an important step of the extraction algorithm of the 3D points: head center, shoulder center and center of mass. These three joints are used to track the patient, including shoulder movements, while performing the TUG. Patient detection is based on the combination of an appearance-based method and the depth information. Evaluating this method by the trunk angle calculation, the success rate is 96.42%.

Thanks to these results, we suggest that the proposed system allows the automatic functional capacities assessment in older adults with good measurement reliability. In addition, the motor control note biomarker can allow a forward detection of a motor decompensation and thus to optimize the process of rehabilitation and to follow the evolution of a frail patient status.

As prospects, it could be interesting to test the system on a large population of elderly people at home. This study could show whether the system allows a reliable assessment of motor function under real conditions at home. The aim would be to propose this system at home to follow the patient's evolution after hospitalization for example. It is therefore interesting to undertake tests taking place over a long period so as to allow performing a longitudinal follow-up, in time, of the patient's

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

functional abilities. This could help to verify the ability of the system to detect changes in their functional level. In addition, these experiments are needed to confirm the acceptability of the system on a larger scale.

On the other hand, we want to improve the automatic extraction module of the skin regions. We think to model and deepen the results of skin detection using other available databases. This approach can be integrated with other interesting applications such as the identification of persons and facial emotion recognition.

Finally, it should also establish an optimized ergonomic human-machine interface to facilitate its use by the elderly and the health professional. It is also possible to make the system generic by adapting it to other sensors.

Question

What is the most appropriate method for tracking patient's movements?
A) the background subtraction
B) methods based on optical flow calculation
C) methods based on temporal difference
D) appearance-based methods.

The correct answer is the appearance-based methods. Indeed, motion-based approaches (A, B and C) of people tracking consist mostly in determining which pixels are moving by the difference between successive images or the background subtraction. This evaluation leads to image segmentation, generally into two regions (pixels in movement and motionless pixels). From this binary image, we can extract a number of geometrical characteristics allowing the recognition of the shape or the action. However, these approaches have problems such as they are expensive in terms of calculation time, the need of a reference image or they do not allow extracting simultaneously fast and slow objects, while the TUG movements are variable that depend on the functional capacities of the patient: they can be slow or fast. In addition, the test can be carried out in different, more or less complex environments. For these reasons, we applied an appearance-based methods combined with the depth information for patient detection.

Clinical Relevance Statement

This study presents an innovative system for automatic and real-time analysis of the clinical test Timed Up and Go. It introduces a new method for the sitting posture detection that enables a robust analysis of the TUG. This system allows the automatic functional capacities assessment in older adults with good measurement reliability.

Conflict of Interest

There were no known conflicts of interest amongst any of the authors of this paper.

Human Subject Research Approval

Not applicable.

Acknowledgments

This work is a part of the STREAM project (system of telehealth and rehabilitation for autonomy and home care) conducted under the program of investments for the future that was launched by the French Government in March 2012.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

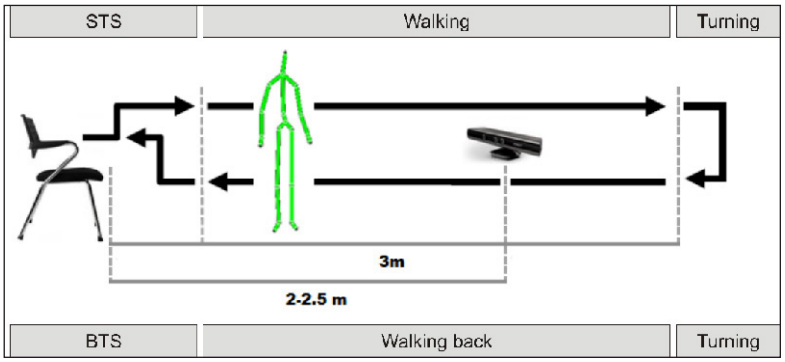

Fig. 1 Overview of the experimental setup of the automatic analysis of the TUG. Abbreviations: STS: Sit-To-Stand; BTS: Back-To-Sit.

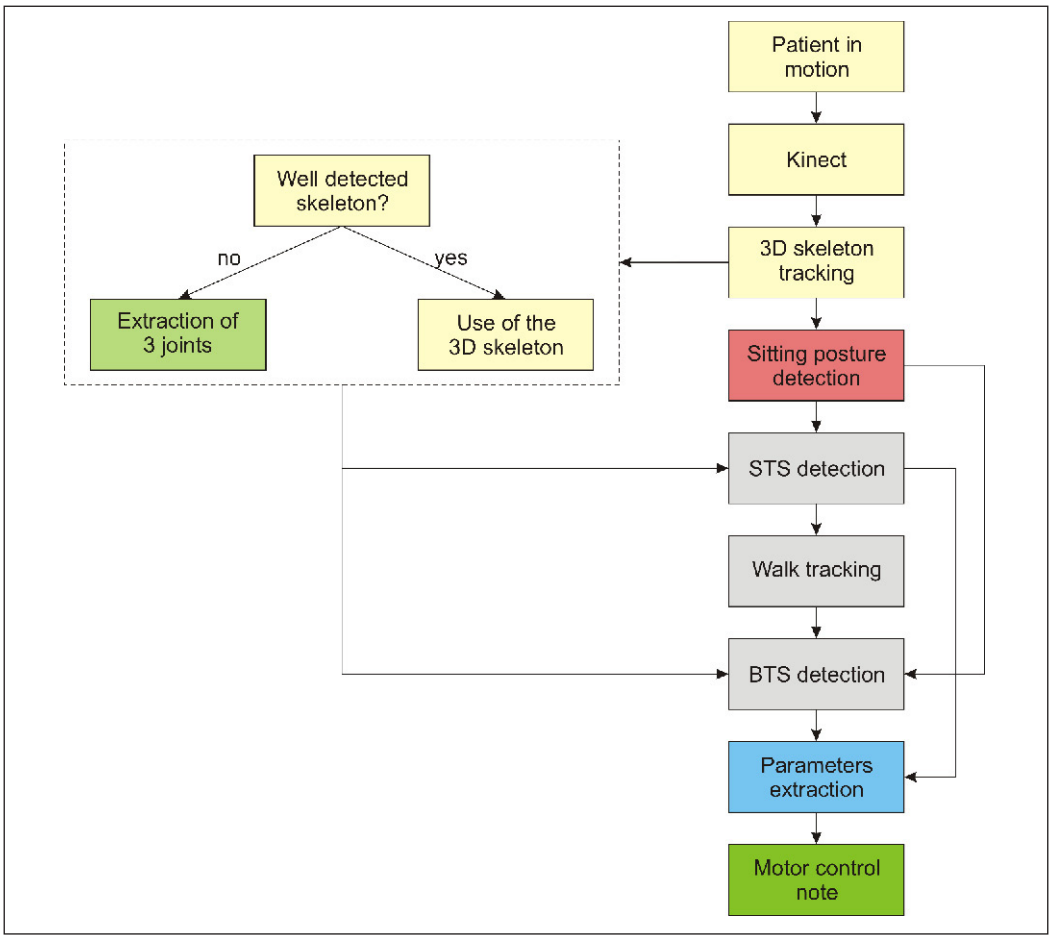

Fig. 2 Global TUG movements' analysis diagram using the proposed video processing chain.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

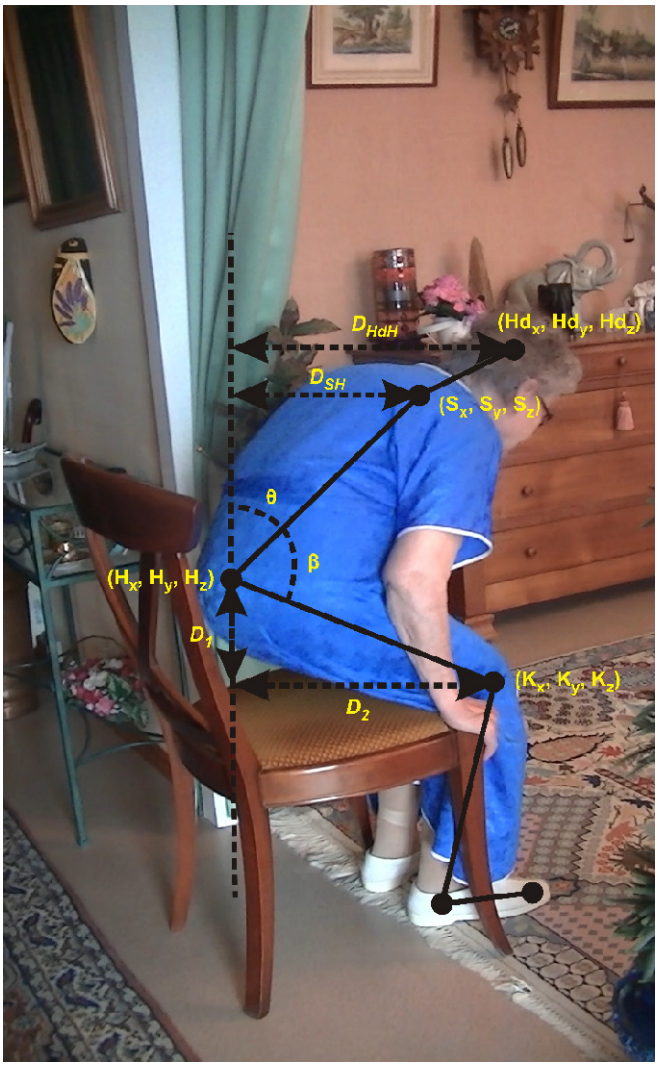

Fig. 3 Characteristics of the sitting posture.

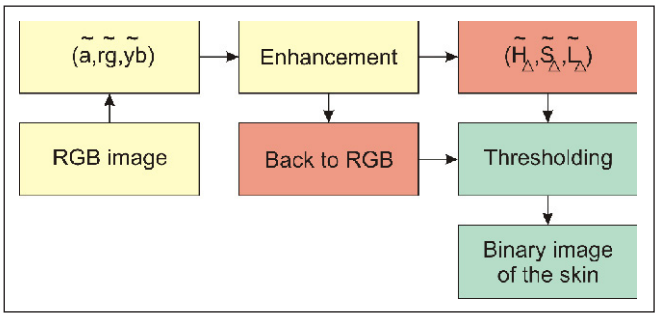

Fig. 4 Overview of skin detection method.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

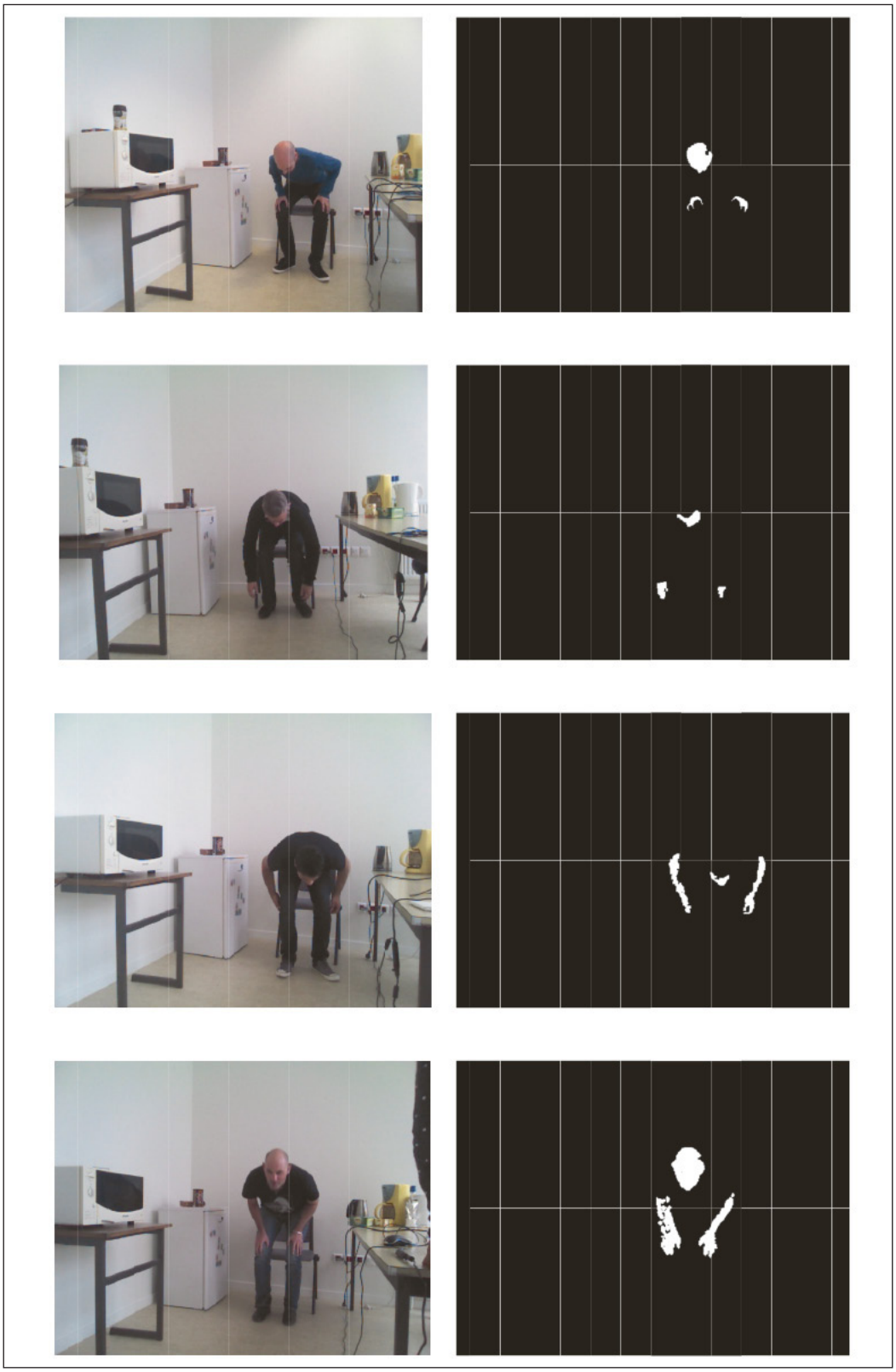

Fig. 5 Final results of skin detection after depth and area filtering.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Fig. 6 Patient detection: case of two persons. (a) Initial image. (b) Background subtraction. (c) Subjects detection. (d) Patient detection and restoration of the missing information.

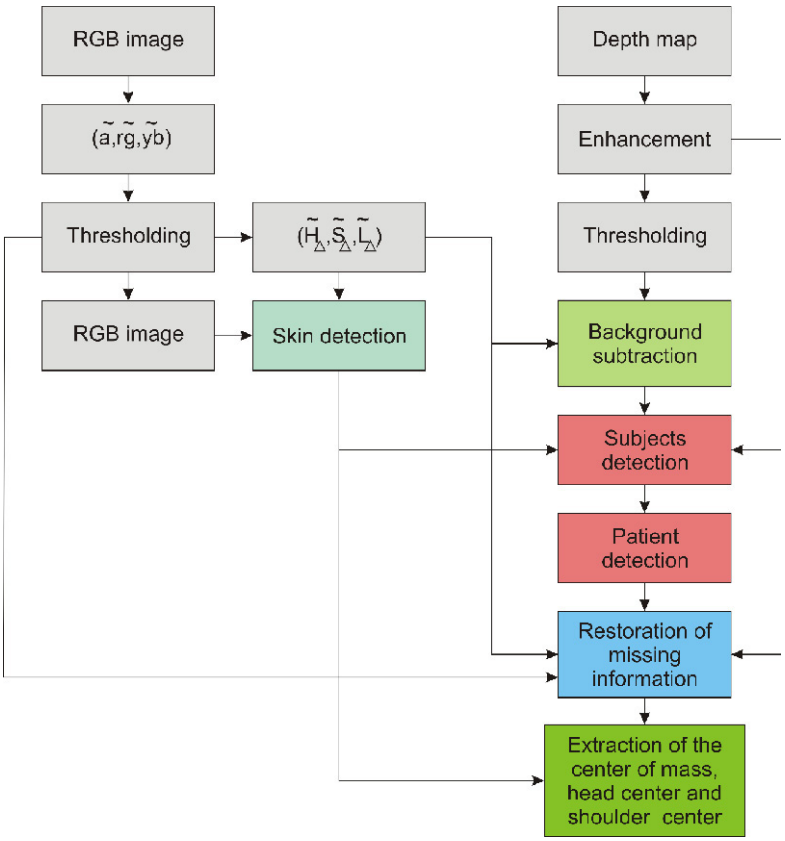

Fig. 7 The extraction process of 3D points corresponding to three joints in the body: the head center, the shoulder center and the center of mass.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

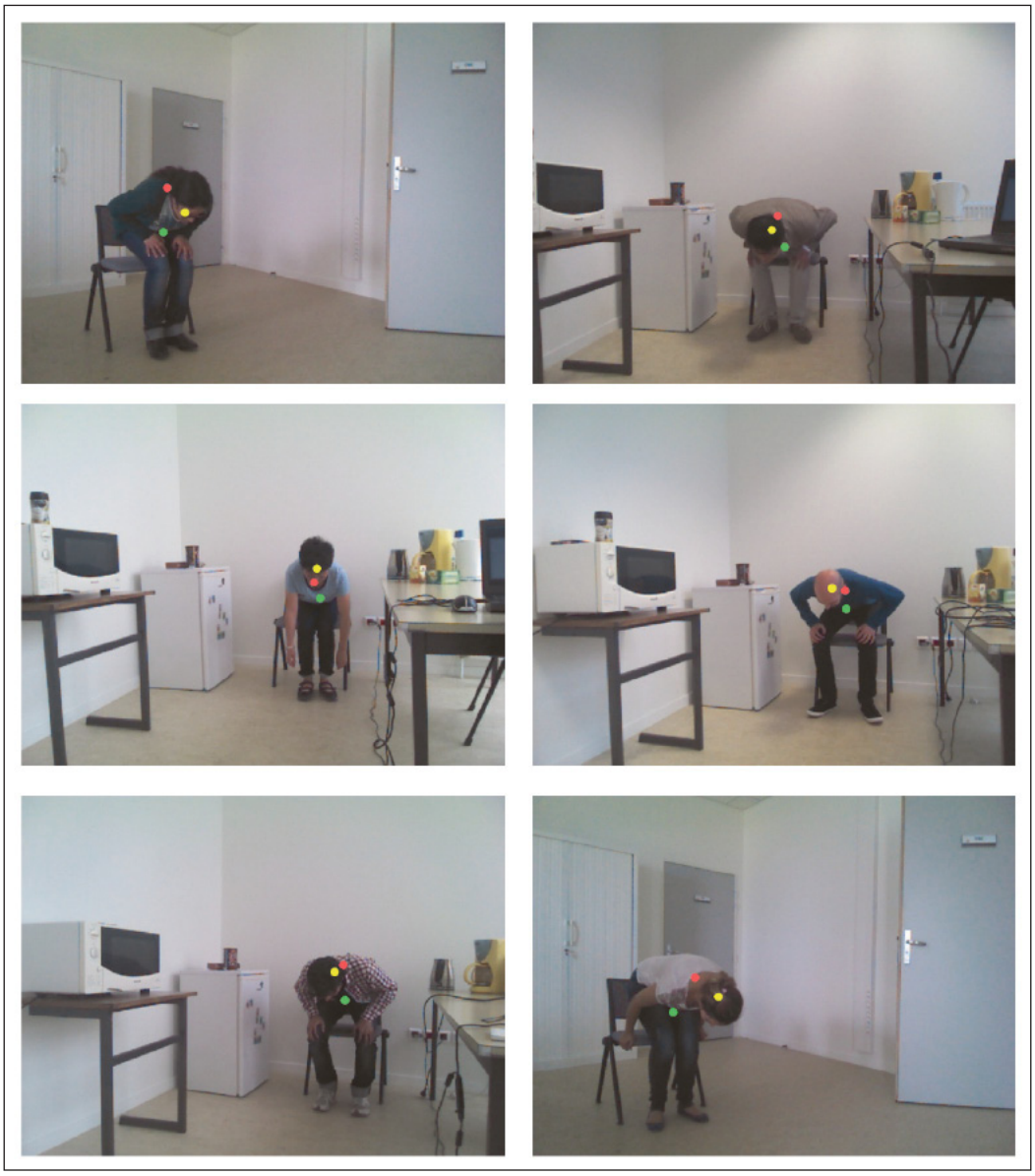

Fig. 8 Detection of the shoulder center (green circle), the head center (yellow circle) and the center of mass (pink circle); and comparison between the angle α1 calculated according to the Kinect skeleton and α2 calculated according to 3D extracted points using the proposed method. (a) α1=47.47°; α2=71.02°. (b) α1=16.77°; α2=99.53°. (c) α1=28.31°; α2=82.02°. (d) α1=51.47°; α2=84.80°. (e) α1=51.64°; α2=84.61°. (f) α1=49.67°; α2=104.63°.

Table 1 Sitting posture detection performance (%).

Specificity	Accuracy	Precision	Recall	Error rate
99.64	99.32	99.54	98.92	0.67

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

References

1. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, Seeman T, Tracy R, Kop WJ, Burke G, McBurnie MA, Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001; 56(3): M146–M156.
2. Hassani A, Kubicki A, Brost V, Mourey F, Yang F. Kinematic analysis of motor strategies in frail aged adults during the Timed Up and Go: how to spot the motor frailty? Clinical Interventions in Aging 2015; 10: 505–513.
3. Willner V, Schneider C, Feichtenschlager M. eHealth 2015 Special Issue: Effects of an Assistance Service on the Quality of Life of Elderly Users. Appl Clin Inform 2015; 6(3): 429–442.
4. Frenken T, Vester B, Brell M, Hein A. aTUG: fully-automated timed up and go assessment using ambient sensor technologies. 2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) 2011; 55-62.
5. Higashi Y, Yamakoshi k, Fujimoto T, Sekine M, Tamura T. Quantitative evaluation of movement using the timed up-and-go test. IEEE Engineering in Medicine and Biology Magazine 2008; 27(4): 38–46.
6. McGrath D, Greene BR, Doheny EP, McKeown DJ, De Vito G, Caulfield B. Reliability of quantitative TUG measures of mobility for use in falls risk assessment. 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2011; 466–469.
7. Whitney S, Marchetti G, Schade A, Wrisley D. The sensitivity and specificity of the timed „Up & go" and the dynamic gait index for self-reported falls in persons with vestibular disorders. Journal of vestibular research: equilibrium & orientation 2004; 14(5): 39–409.
8. Shumway-Cook A, Brauer S, Woollacott M. Predicting the probability for falls in community-dwelling older adults using the timed up & go test. Physical therapy 2000; 80(9): 896–903.
9. Beauchet O, Fantino B, Allali G, Muir SW, Montero-Odasso M, Annweiler, C. Timed Up and Go test and risk of falls in older adults: a systematic review. J Nutr Health Aging 2011; 15(10): 933–938.
10. Greene BR, Kenny RA. Assessment of Cognitive Decline Through Quantitative Analysis of the Timed Up and Go Test. IEEE Transactions on Biomedical Engineering 2012; 59(4): 988–995.
11. Mourey F, Pozzo T, Rouhier-Marcer I, Didier J. A kinematic comparison between elderly and young subjects standing up from and sitting down in a chair. Age and ageing 1998; 27(2): 137–146.
12. Dubost V, Beauchet O, Manckoundia P, Herrmann F, Mourey F. Decreased trunk angular displacement during sitting down: an early feature of aging. Physical therapy 2005; 85(5): 404-412.
13. Mourey F, Grishin A, d'Athis P, Pozzo T, Stapley P. Standing up from a chair as a dynamic equilibrium task: a comparison between young and elderly subjects. J Gerontol A Biol Sci Med Sci 2000; 55: 425–431.
14. Chang YJ, Chen SF, Huang JD. A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. Research in Developmental Disabilities 2011; 32(6): 2566-2570.
15. Clark RA, Pua YH, Fortin K, Ritchie C, Webster KE, Denehy L, Bryant AL. Validity of the microsoft kinect for assessment of postural control. Gait & Posture 2012; 36(3): 372-377.
16. Nghiem AT, Auvinet E, Meunier J. Head detection using kinect camera and its application to fall detection. 2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA) 2012; 164–169.
17. Kuipers DA, Wartena BO, Dijkstra BH, Terlouw G, van T Veer JT, van Dijk HW, Prins JT, Pierie JP. iLift: A health behavior change support system for lifting and transfer techniques to prevent lower-back injuries in healthcare. International Journal of Medical Informatics 2012; 96: 11-23.
18. Hayat M, Bennamoun M, El-Sallam AA. An RGB-D based image set classifcation for robust face recognition from Kinect data. Neurocomputing 2016; 171: 889-900.
19. Skjæret N, Nawaz A, Morat T, Schoene D, Helbostad JL, Vereijken B. Exercise and rehabilitation delivered through exergames in older adults: An integrative review of technologies, safety and efficacy. International Journal of Medical Informatics 2016; 85 (1): 1-16.
20. Hassani A, Kubicki A, Brost V, Yang F. Preliminary study on the design of a low-cost movement analysis system: reliability measurement of timed up and go test. VISAPP 2014; 33–38.
21. Hassani A, Real-time 3D movements analysis for a medical device intended for maintaining functional independence in aged adults at home. Ph.D. thesis, University of Burgundy (March 2016).
22. Persson CU, Danielsson A, Sunnerhagen KS, Grimby-Ekman A, Hansson P-O. Timed Up & Go as a measure for longitudinal change in mobility after stroke – Postural Stroke Study in Gothenburg (POSTGOT). Journal of NeuroEngineering and Rehabilitation 2014; 11: 83.
23. Nocera JR, Stegemöller EL, Malaty IA, Okun MS, Marsiske M, Hass CJ, National Parkinson Foundation Quality Improvement Initiative Investigators. Using the Timed Up & Go Test in a Clinical Setting to Predict Falling in Parkinson's Disease. Archives of physical medicine and rehabilitation 2013; 94(7): 1300–1305.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

24. Warzee E, Elbouz L, Seidel L, Petermans J. Evaluation de la marche chez 115 sujets hospitalisées dans le service de gériatrie du CHU de Liège. Annales de Gérontologie 2010; 3(2): 111–119.
25. Bonnechère B, Jansen B, Salvia P, Bouzahouene H, Omelina L, Cornelis J, Rooze M, Van Sint Jan S. What are the current limits of the kinect sensor? 9th Intl Conf. Disability, Virtual Reality & Associated Technologies (ICDVRAT) 2012; 287-294.
26. Clark RA, Bower KJ, Mentiplay BF, Paterson K, Pua YH. Concurrent validity of the Microsoft Kinect for assessment of spatiotemporal gait variables. Journal of Biomechanics 2013; 46(15): 2722–2725.
27. Manckoundia P, Mourey F, Pfitzenmeyer P, Papaxanthis C. Comparison of motor strategies in sit-to-stand and back-to-sit motions between healthy and Alzheimer's disease elderly subjects. Neuroscience 2006; 137(2): 385–392.
28. Liu CL, Lee CH, Lin PM. A fall detection system using k-nearest neighbor classifier. Expert systems with applications 2010; 37(10): 7174-7181.
29. Rougier C, Meunier J, St-Arnaud A, Rousseau J. Robust video surveillance for fall detection based on human shape deformation. IEEE Transactions on Circuits and Systems for Video Technology 2011; 21(5): 611–622.
30. Liao YT, Huang CL, Hsu SC. Slip and fall event detection using Bayesian belief network. Pattern recognition 2012; 45(1): 24–32.
31. Malagón-Borja L, Fuentes O. Object detection using image reconstruction with PCA. Image and Vision Computing 2009; 27(1–2): 2–9.
32. Wang J, Barreto A, Wang L, Chen Y, Rishe N, Andrian J, Adjouadi M. Multilinear principal component analysis for face recognition with fewer features. Neurocomputing 2010; 73(10–12): 1550–1555.
33. Toyama K, Krumm J, Brumitt B, Meyers B. Wallflower: Principles and Practice of Background Maintenance. The Proceedings of the Seventh IEEE International Conference on Computer Vision 1999; 1: 255–261.
34. Chiranjeevi P, Sengupta S. Spatially correlated background subtraction, based on adaptive background maintenance. Journal of Visual Communication and Image Representation 2012; 23(6): 948–957.
35. Laptev I. On Space-Time Interest Points. International Journal of Computer Vision 2005; 64(2–3): 107-123.
36. Harris C, Stephens M. A combined corner and edge detector. Fourth Alvey Vision Conference 1988; 147–151.
37. Föstner MA, Gülch E. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centers of Circular Features. ISPRS Intercommission Workshop 1987.
38. Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001; 1: I-511-I-518.
39. Kang S, Choi B, Jo D. Faces detection method based on skin color modeling. Journal of Systems Architecture 2016; 64: 100–109.
40. Huang L, Ji W, Wei Z, Chen BW, Yan CC, Nie J, Yin J, Jiang B. Robust skin detection in real-world images. Journal of Visual Communication and Image Representation 2015; 29: 147–152.
41. Nallaperumal K, Ravi S, Babu C, Selvakumar R, Fred A, Seldev C, Vinsley SS. Skin Detection Using Color Pixel Classi_cation with Application to Face Detection: A Comparative Study. International Conference on Conference on Computational Intelligence and Multimedia Applications 2007; 3:436-441.
42. Kovac J, Peer P, Solina F. 2D versus 3D colour space face detection. 4th EURASIP Conference focused on Video/Image Processing and Multimedia Communications 2003; 2: 449–454.
43. Shaik KB, Ganesan P, Kalist V, Sathish BS, Merlin Mary Jenitha J. Comparative Study of Skin Color Detection and Segmentation in HSV and YCbCr Color Space. Procedia Computer Science 2015; 57: 4–48.
44. Kukharev G, Nowosielski A. Visitor identification – elaborating real time face recognition system. 12th Winther School on Computer Graphics (WSCG) 2004; 157–164.
45. Gouinaud H, Gavet Y, Debayle J, Pinoli JC. Color correction in the framework of Color Logarithmic Image Processing. 2011 7th International Symposium on Image and Signal Processing and Analysis (ISPA) 2011; 129–133.
46. Gouinaud H. Traitement logarithmique d'images couleur. Ph.D. thesis, Ecole Nationale Supérieure des Mines de Saint-Etienne 2013.
47. Angulo J. Morphological colour operators in totally ordered lattices based on distances: Application to image filtering, enhancement and analysis. Computer Vision and Image Understanding 2007; 107(12): 56–73.
48. Drillis R, Contini R. Body segment parameters. DHEW 1166–03. New York University, School of Engineering and Science, New York.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.