7

Validation of the instrumented chair part two: To identify the phases of the sit-to-stand test in older people

Results presented in this chapter have been presented at the 26th National Conference on Communications (NCC) in Kharagpur, co-sponsored by the IEEE. The version presented here is similar to that publication with the addition of information that links studies in the thesis. This work was completed by the PhD scholar with standard input from the supervision team. The citation for this work is presented below:

Shukla, B.K, Jain, H., Vijay, V., Yadav, S.K., Hewson, D.J. Sensor fusion based assessment framework for home-based monitoring of elderly.Published in IEEE,National Conference on Communications (NCC) 2020

7.1 PART1- A FUSION-BASED APPROACH TO IDENTIFY THE PHASES OF THE SIT-TO-STAND TEST IN OLDER PEOPLE

7.2 INTRODUCTION

As mentioned earlier that the development of an automatic assessment system for clinical tests like the 5STS and TUG have been used extensively in research recently with application towards elderly monitoring. The field has been explored in two different sensor domains - using visual sensors like Kinect camera and using electronic sensors such as gyroscopes and load cells. While both of these domains have individually been seen performing well in assessing clinical tests, both are lacking in some respects. Based on the head position of the human skeleton acquired using Kinect sensors systems, the individual cycles of 5STS were able to be identified Ejupi *et al.* [2016], with these cycles then used to discriminate between fallers and non-fallers. Though the technique performed well and was found suitable for an unsupervised in-home examination of elderly, it is prone to errors caused due to some drawbacks of the Kinect such as the framework not being suitable for scenarios that involve cluttered backgrounds and occluded postures.

Some recent works have introduced algorithms that used videos from RGB cameras to assess the quality of the 5STS movement. In one such study, a three-camera system was used to extract 3- D features based on the distance between body parts, with these distances then used to create body centroids Allin and Mihailidis [2009]. The locations of these body segments were subsequently followed using ellipsoid tracking. The STS performance was quantified using the rise time of the STS, which was shown to be highly correlated with the Berg Balance Score, which is a widely used clinical balance test. Although the system worked well, individual body parts required manually labelling for all participants for at least one image to ensure that the system was able to learn the color information for people tested. Furthermore, the system required three cameras to be used, rendering it difficult to use outside a laboratory setting. The STS has also been automatically evaluated using pose-based descriptors from volumetric image data Pehlivan and Duygulu [2011]. This system extracted features from the descriptors. Classification of the movement was then carried out using the nearest neighbor method, which was successfully able to identify the STS from among other movements. In another study, STS time was derived from 3-D modelling of the body in voxel space using an algorithm to fit an ellipse to the image, followed by the use of image features to determine the body's orientation Banerjee *et al.* [2013]. The most accurate segmentation method in this latter study was that obtained using voxel height and the ellipse fit. This system was, therefore, proposed for use as part of a monitoring system to detect falls in community-dwelling older people. However, a drawback of the system is the requirement to use two cameras to calculate the voxels. In addition, the background subtraction system proposed is highly dependent on the background in which the recording takes place, with problems encountered when cluttered backgrounds are present, which would cause false silhouette extractions leading to a sub- optimal solution Aggarwal and Cai [1999].

In contrast to the camera-based systems identified above, other techniques require subjects to wear the sensors. For instance wearable sensors including diverse technology such as triaxial accelerometers have been used to distinguish between different stationary postures such as sitting and standing, or to identify movement such as walking and running, while transitions between positions such as sitting and standing (the STS) and standing and lying (falling) have also been reported Kang *et al.* [2010]. In another study, accelerometers were used to differentiate between Parkinson's disease and normal subjects based on the STS movement performed during the TUG Weiss *et al.* [2010]. However, although sensor-based tests have produced good results, the requirement for the user to wear the sensors whenever the test is undertaken could be inconvenient in some cultures.

An alternative to body-worn sensors is to incorporate them into everyday devices. For instance, previous work has evaluated the STS with sensors mounted on the chair used for the test Wilson and Nicol [2005]. Since this study almost 15 years ago, little work has been performed using a sensor-equipped chair. As a part of this paper, we propose two modalities for evaluating STS performance, with these methods subsequently combined using data fusion techniques:

1) *RGB-based assessment* : RGB cameras are widely available in many devices such as mobile phones. We propose the use of a single camera system such as those used recently to identify different body postures and exercise movements Jain and Harit [2016, 2017] The camera-based system works by extracting human poses from each frame in a video are extracted using deep pose library (discussed in chapter 6) and the pose signals of elderly subjects are then matched to the standard STS performance of an expert to identify the different phases of the STS.

2) *Instrumented Chairs with Load Cells* : An instrumented chair as discussed in chapter 5 is able to identify different phases of STS by analyzing pattern of variations in the center of pressure, as shown in Figure 7.1.

It would appear likely that the RGB-based system assesses the position of the subject solely based on pose features, meaning that contact with objects such as the chairs would not be represented well, thus leading to some errors where the position of interest is identified based on contact occurring with the chair, such as seat on and seat off positions. On the other hand, the instrumented chair would be expected to have difficulty identifying phases when the subject is not in contact with the chair, such as the standing phase of the STS test.

One solution to this problem would be to develop an assessment framework that can fuse the capabilities of both domains. Such a system would still be easy to use at home by including RGB based pose monitoring and could be useful in geriatric assessment. As a part of this framework we introduce one such framework and test it for the 5STS, which is widely used test in geriatrics. The 5STs consists of a sequence of the six transition phases shown in Figure 7.1 performed five times, with the movement described in detail in the next section. The test requires the overall STS time to be measured, with the additional individual phase times of potential benefit to enhance the evaluation of the performance of the older person tested. Such a test would require an efficient approach to segment the videos into the respective STS phases.

In the next section we discuss approaches to segment the videos into individual phases of STS

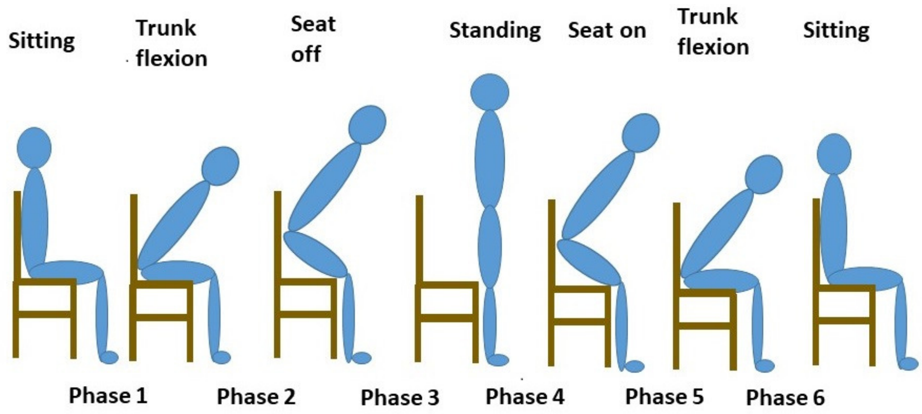

Figure 7.1 : Sit-to-Stand Phases

test using the two sensor modalities discussed above. This is then followed by a discussion of STS phase segmentation technique before we introduce the fusion framework with a novel sensor-fused segmented dynamic time warping technique. Finally, in Section 4 we present our experimental setup and show how the fusion of both camera and sensor-based methods can help improve assessment accuracy.

7.3 PHASES OF SIT-TO-STAND TEST

A single STS is defined by six phases (Figure 7.1), beginning with a stable seated position, before the subject bends forward using trunk flexion (Phase 1), then begins to leave the chair, known as seat off (Phase 2).The subject then stands up to a fully standing position (Phase 3) before beginning to sit down again, with this phase stopping when the subject regains contact with the chair (Phase 4). The person then sits back on chair but remains in trunk flexion (Phase 5), before returning to an upright seated position (Phase 6). It appears likely that the measurement of the time taken to perform each of these phases of the STS could be an important addition to the standard test. The instrumented STS would enable the functional ability to be more fully explored including an assessment of the power developed during the movement, assessed during the standing- up phases of the test, while irregular times for the individual STS tests would also be of interest. If an older person performs well in the first STS and then decreases their execution speed towards the end due to low muscle strength and endurance, this would be important to detect. Thus, the individual phase analysis and overall STS time measurement is important. We discuss approaches of phase detection in the following section using the two sensor modalities, followed by a fusion technique.

7.4 PROPOSED APPROACH

In this section we discuss the phase detection algorithms from two approaches: 1) RGB-based template matching approach; 2) instrumented chair based phase detection. Following the individual techniques, a fusion approach to fuse the phase boundary results of the two modalities for improved phase detection.

7.4.1 RGB-based Template Matching Approach

Our first approach towards phase detection of 5STS involves the use of a single camera system. We first extract poses from the individual frames of the video using a Stacked Hourglass deep pose

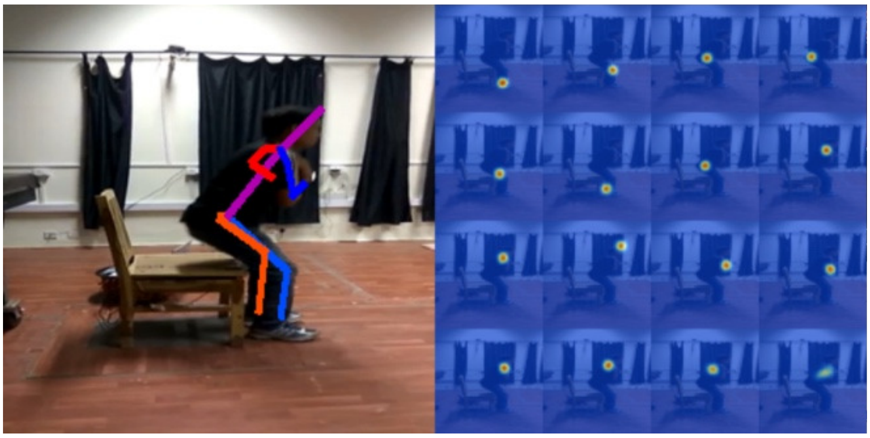

Figure 7.2 : Pose Feature Extracted from Stacked Hourglass Deep Pose library Newell *et al.* [2016] with 16 joint points : head, neck, left shoulder, left elbow, left wrist, right shoulder right elbow, right writ, pelvis, left hip, left knee, left ankle, right hip, right knee, right ankle

library Newell *et al.* [2016]. This provides us with 16 joint points as shown in Figure 7.2.n expert is then used identify the different phases of the STS performance from a video recording.

Given a test action pose sequence of an elderly subject T and an expert STS sequence E, our system aims to find the size phases of STS cycle for the test elderly sequence Q given the boundaries of the phases of an expert STS performance E. This requires aligning the two sequences and reporting the frames of the test sequence that match to the boundaries of the phases in the expert sequence.

Dynamic Time Warping is a widely used exemplar based sequence matching approach. It is a nonlinear time warping scheme finds the best warping function between two input signals such that the distance of the warps path is minimum. It is invariant to some degree of time variation between the two sequences. The technique is described by some constraints that reduces the search space. These are -

- **monotonicity constraint** - that prevents the warping path from going back in time axis
- **boundary conditions** - that limits the warping path to start from the first time instance and end at the last time instance for both the test and the expert sequences.

Given a test sequence Q composed of poses $\{q_1, q_2, q_3, ..., q_M\}$ and an expert action video sequence E containing poses $\{e_1, e_2, e_3, ..., e_N\}$, a DTW table of size $M \times N$ is created and the boundaries are set as infinity. For $1 \leq i \leq M$ and $1 \leq j \leq N$, each grid (i, j) is filled with a minimum warping distance defined by

$$d_w(i,j) = min\begin{pmatrix} d_w(i-1,j-1) \\ d_w(i,j-1) \\ d_w(i-1,j) \end{pmatrix} + cost(i,j)$$

where $cost$ is the cosine distance between two poses. The DTW method backtracks from the end grid (M, N) to the start grid (1, 1) and construct the entire alignment path which is invariant to temporal transformation. Thus we inherently get the boundaries of phases of the test STS sequence by finding the matched frames.

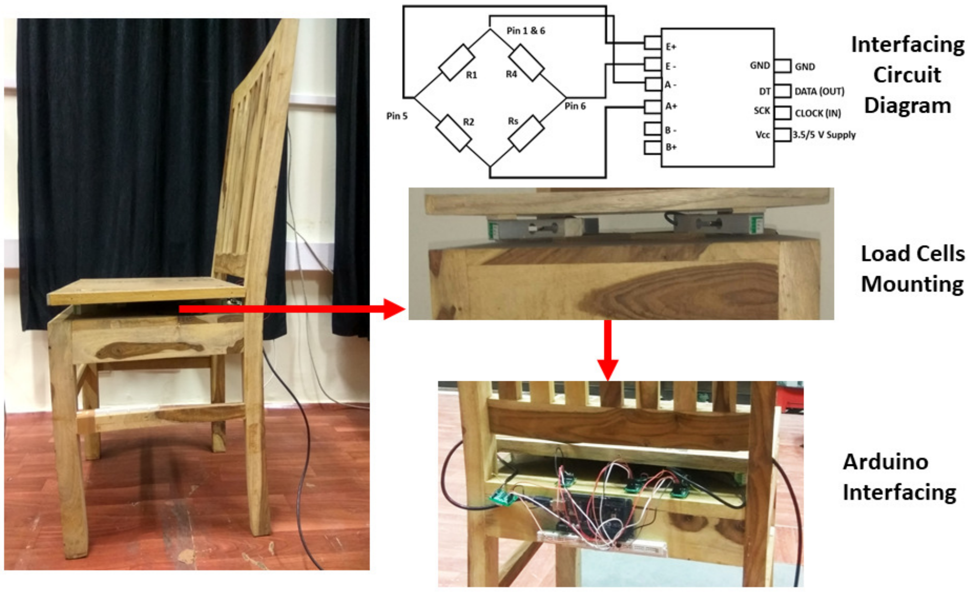

Figure 7.3 : Instrumented Chair

7.4.2 Instrumented Chair based Phase Detection

A portable instrumented chair was designed to measure forces and movement during the STS as discussed in chapter 5. (Figure 7.3). The sum of the four load cells was used to calculate total ground reaction force (Fz) during the STS, while the barycenter of the four individual Fz was used to obtain instantaneous position of the center of pressure (CoP). The start of each STS was identified using the CoP signal, with a movement of 5cm towards taken as the start of Phase 1, while Phase 2 was considered to begin when the CoP reached a maximal forward position without a corresponding decrease in total Fz. Thereafter, Phase 3 began when Fz began to decrease, with a level below 5N taken to be seat-off. Phase 4 could not be calculated for the chair owing to the lack of contact with the participant, meaning that the next phase obtained was the start of Phase 5, when seat-on was considered to the point at which Fz increased above 5N. Phase 5 ended when total mass exceeded body mass, with the end of Phase 6 occurring when CoP returned to the starting position. This procedure was repeated for each of the five individual STS movements.

7.4.3 Sensor fusion based Segmented Dynamic Time Warping Technique

The RGB-based phase detection works on the principal of template matching. For such a system, events such as seat off and seat on cannot be identified accurately. This is because the pose just before or after the seat-off event are approximately the same as the pose at the seat-off event. This is also the case with the seat-on event. In such cases there are high chances of misalignment while the two sequences are matched.

Such events can easily be identified by an instrumented chair where the load cell readings decrease to zero as soon as a person performs a seat off and conversely increases above zero when seat-on occurs. However, the instrumented chair is not without its own limitations in its current format. Although it can accurately identify events that involve direct contact with the chair, it cannot identify phases where the person has lost contact with the chair. All such phases would thus be marked as the standing phase, meaning Phases 3 and Phase 4 in Figure 7.1would be combined into a single standing phase. However, these two phases could be identified using the RGB-based system.

Thus, we can see that the two modalities complement each other and that their fusion could

Figure 7.4 : Sample frames of elderly

lead to better phase-detection results. We propose a Sensor-fusion based Segmented Dynamic Time Warping Technique, which works in two steps performed one after the other.

As a first step, we leverage the capability of the instrumented chair in identifying the seat-on and seat-off events and segment the 5STS into five individual segments from seat-off to seat-on. This is then followed by matching a template applied to the individual segments of the performance to identify the remaining key events. In doing so, all events can be more accurately identified than using the two techniques individually. We discuss the performance of the three techniques in the next section.

7.5 EXPERIMENTS

We begin by describing the dataset used in the experiments, followed by the results of the techniques applied.

7.5.1 Dataset

Performance of the three techniques was compared using data collected from a sample of 15 older people subjects including both males and females. The subjects were aged 66.9 ± 6.8 years, weighed 65.0 ± 12.1 kg, were of height $1.58 \pm 0.07m$, and had a BMI of $26.0 \pm 4.3kg/m2$.

Each participant performed one trial of the 5STS test using a self-selected comfortable speed. An example of an older participant performing the STS is shown in Figure 7.4. For the template matching

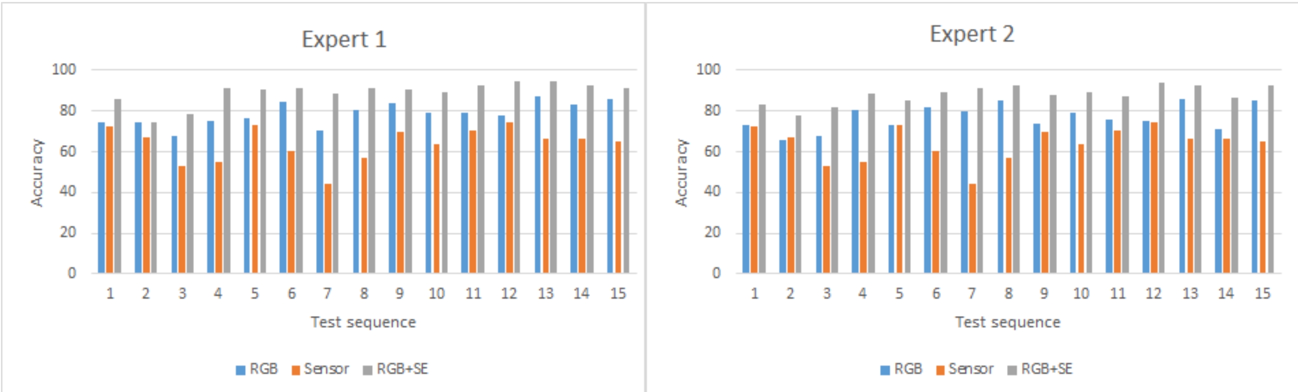

Figure 7.5 : Comparison of Segmentation Accuracy of three techniques for 15 elderly subjects

Table 7.1 : Comparison Of Segmentation Accuracy between Techniques; Values are bootstrapped means and 95% confidence intervals

	Slow Expert	Fast Expert
RGB	75.5 (72.3-79.2)	77.4 (74.1 - 80.3)
Chair	72.8 (70.9 - 74.6)	
Fusion	89.4 (87.8 - 91.1)	90.0 (88.4 - 91.4)

approach, we chose two templates from an expert performer aged 30 years. The two templates used varying speeds to demonstrate the capability of our template-matching approach. Ethical approval was obtained for testing from the Asian Centre for Medical Education, Research & Innovation approved the study (ACMERI/18/001), with subjects giving their signed and informed consent to participate in the study.

7.5.2 Evaluation Metrics

The ground truth for the boundaries of the events for all the subjects was provided by the expert while they looked at the video frames. We normalized the sensor boundaries to that of the RGB frames to make the results comparable to the ground truth that was devised using the RGB frames. The percentage of frames or data points that fell within a given phase was determined for each phase of the 5STS for all participants, with averages taken for each phase across all of the five STS movements The percentage of frames correctly classified according to the ground truth defines the performance of the three techniques. The Shapiro-Wilk test revealed that all data was non-normal due to the ceiling effect of the accuracy approaching 100% for some techniques, therefore bootstrapping was used for all statistical analyses. Analysis of variance was used to compare the accuracy of each technique, with $p < 0.05$ taken for statistical significance. Bonferroni adjustments were used for post-hoc comparisons between the three techniques.

7.6 RESULTS

The accuracy of each method is shown in Table 7.1 for the three techniques. There was a significant overall effect of technique for both slow and fast templates. When each template was compared between the techniques, no differences were observed between the RGB and the chair, however the fusion of the two systems performed significantly better than either of the techniques

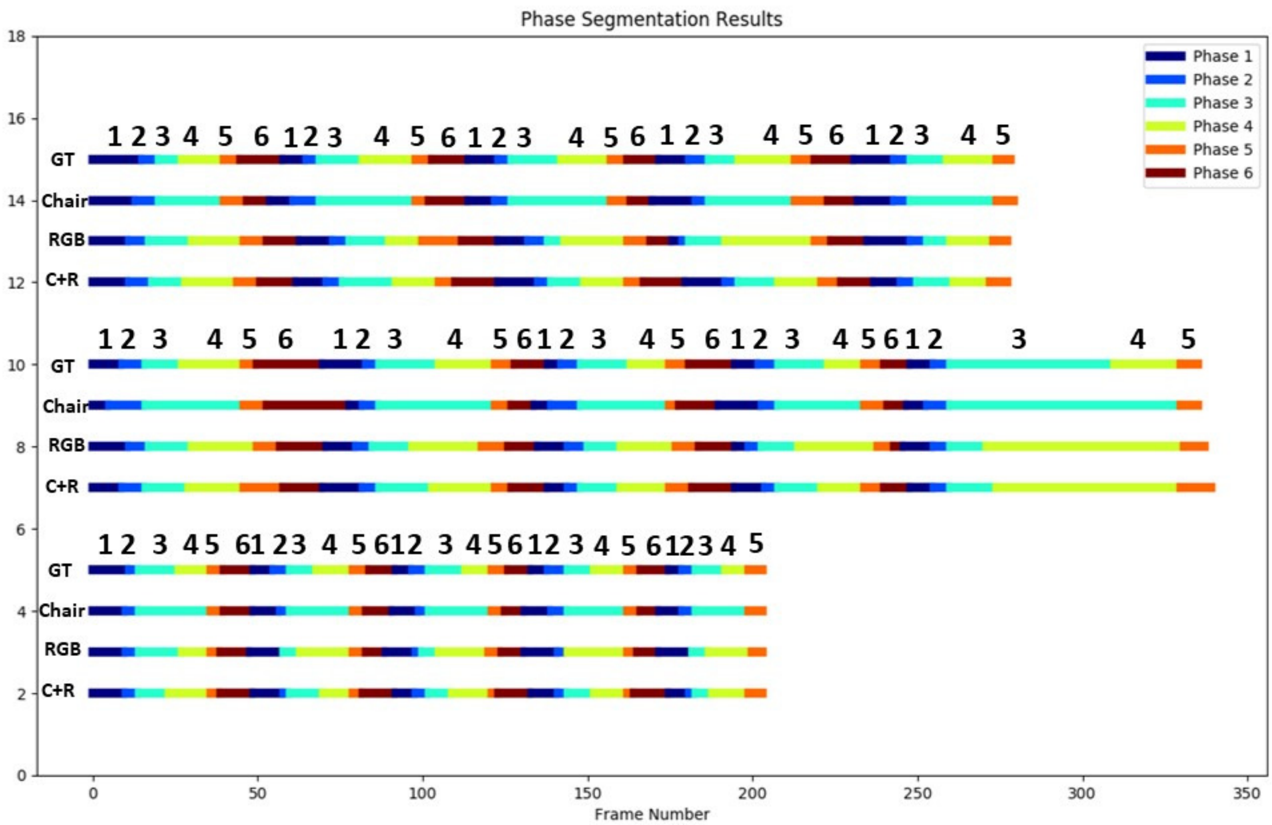

Figure 7.6 : Qualitative comparison of the three techniques towards individual phase detection of Sit-to-Stand Test; Different colors illustrate the 6 phases; GT: Ground Truth, C+R - fusion of chair and RGB

individually.

Individual accuracy data for each subject for is shown in Figure 7.5 for the proposed techniques and the two template speeds. It can be seen that the average accuracy of the fusion approach was greater for all participants, however, there was no consistent pattern in the differences in accuracy between the RGB and chair methods.

A qualitative illustration of the phase detection using the three techniques for three test cases is shown in Figure 7.6. It can be seen that the chair does not identify the fourth phase as this phase begins while participants were standing. In contrast, the fusion based technique is close to the ground truth for all phases.

7.7 CONCLUSION

As a part of this work we developed a method to identify six phases of the STS movement. We proposed a segmentation technique using information from two sensor modalities: 1) An RGB-based camera approach; and 2) an instrumented chair with four load cells. It is seen that although these two techniques perform well individually, they each have their own limitations and advantages. However, when the data from the two techniques undergoes fusion, the two techniques complement each other, with one's limitation rectified by the strength of the other. We propose the novel Sensor fusion based Segmented Dynamic Time Warping Technique to leverage the complementary capabilities of the two sensors. The fusion technique is seen to outperform both the sensors individually for 15 older subjects and hence offers a better approach for clinical assessment. Next part of this chapter will focus on the addition of extra sensors to the chair in order to identify the phases when the person is standing up. These sensors could include other sensors such as infra-red or ultrasonic sensors to detect the position of the person being tested.

7.8 PART2-DEVELOPMENT OF AN INSTRUMENTED CHAIR TO IDENTIFY THE PHASES OF THE SIT-TO-STAND MOVEMENT

In previous work (Chapter 6), a novel iSTS device was developed, which consisted of an instrumented chair. The chair was used to detect movement of the center of pressure (CoP) and ground reaction forces of the person performing the STS movement, with good results obtained for total STS time and STS velocity. Instrumented chair detected Sitting, seat off, seat on phases of STS However, the chair was unable to identify all the phases of the STS as signals could not be recorded when contact was lost between the chair and the person being tested.

In the previous section a fusion approach using the sensors of the chair and an RGB camera was successful, however this requires the simultaneous use of two systems. An alternative to the fusion approach would be to integrate additional sensors into the chair to identify the phases when the person is not in contact with chair. For instance, ultrasonic sensors can be used to detect the distance to an object, such as a person. Accordingly, the aim of this study is to integrate ultrasonic sensors into a chair equipped with load cells in order to accurately identify the phases of the STS movement.

7.8.1 Instrumented Chair new version

In the new version of the chair, an ultrasonic (US) distance sensor Figure 7.7 (HC-SR04, Adraxx, New Delhi, India) was fitted on the back of the chair at a height of 35 cm from the seat with a downwards angle of 30 degrees. The US sensor sends eight 40kHz square waves, with returning signals reflected from the object in front used to determine the distance of the object. The US sensor has a range of 0.02-4.5 m with a precision of 3mm. Similar to the previous study, signals from the load cells and US sensor were acquired using an Arduino micro-controller board (Arduino Mega 2560, Arduino LLC,

Figure 7.7 : Chair with Ultrasonic sensor

Somerville, MA, USA) at 40 Hz using a custom-built software program that was written in Python. The analogue signals from the load cells are converted to digital signals using an analogue to digital converter (ADC) (HX711 Avia Semiconductors, Xiamen, China).

For the purposes of this study, two phases were chosen to represent the sit-to-stand part of the STS, with another two phases for the sitting down part of the STS. These phases can be seen in Figure 7.8. The two transition points of trunk flexion were removed from the study as trunk flexion was mostly seen where the seat off events were observed with a frame difference of 1 or 2.

The method used to identify the four phases of the STS by the chair is shown in Figure 7.9. The antero-posterior (AP) CoP signals were normalized such that the maximal position towards the back of the chair was taken to be 0, while a maximal forward position was taken to be 1, while the time when the person was not in contact with the chair also set to be 1. The start of the STS occurs when the person is sitting, which was taken to be the moment when AP CoP was a maximal distance from the front of the chair. The end of Phase 1 was taken to occur when Fz decreased below 5N, which indicated seat-off. The rising phase then occurred, with the end of Phase 2 taken to be when a maximal distance occurred for the US sensor. Phase 3 ended when Fz increased above 5N to indicate seat-on, while Phase 4 ended when AP CoP returned to the starting position. The timings of all phases were obtained for all individual STS movements.

7.8.2 Experimentation

A comparison between the different systems (RGB, Chair vesrion 1, Chair Version 2, Fusion technique) was made using data from 10 subjects (g males, 4 females) aged 25.9 ± 3.5 years. The subjects

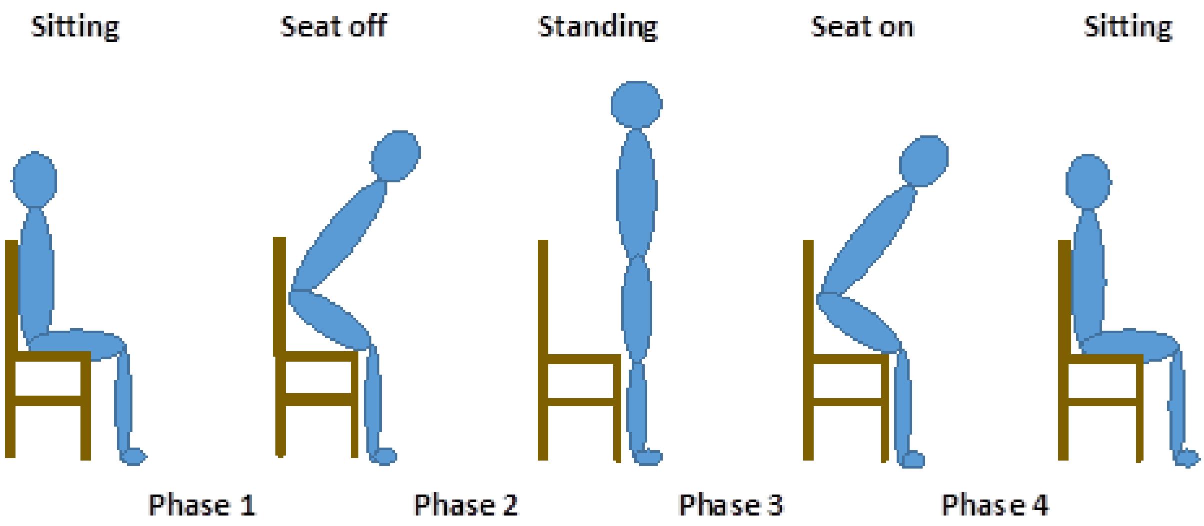

Figure 7.8 : Phases of the Sit-to-Stand movement

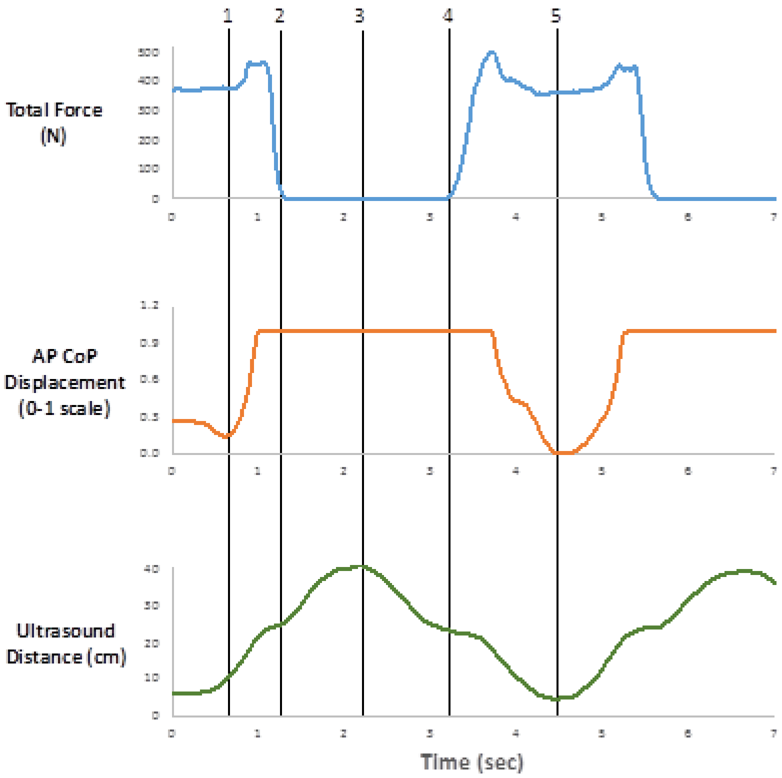

Figure 7.9 : Phases of the Sit-to-Stand movement

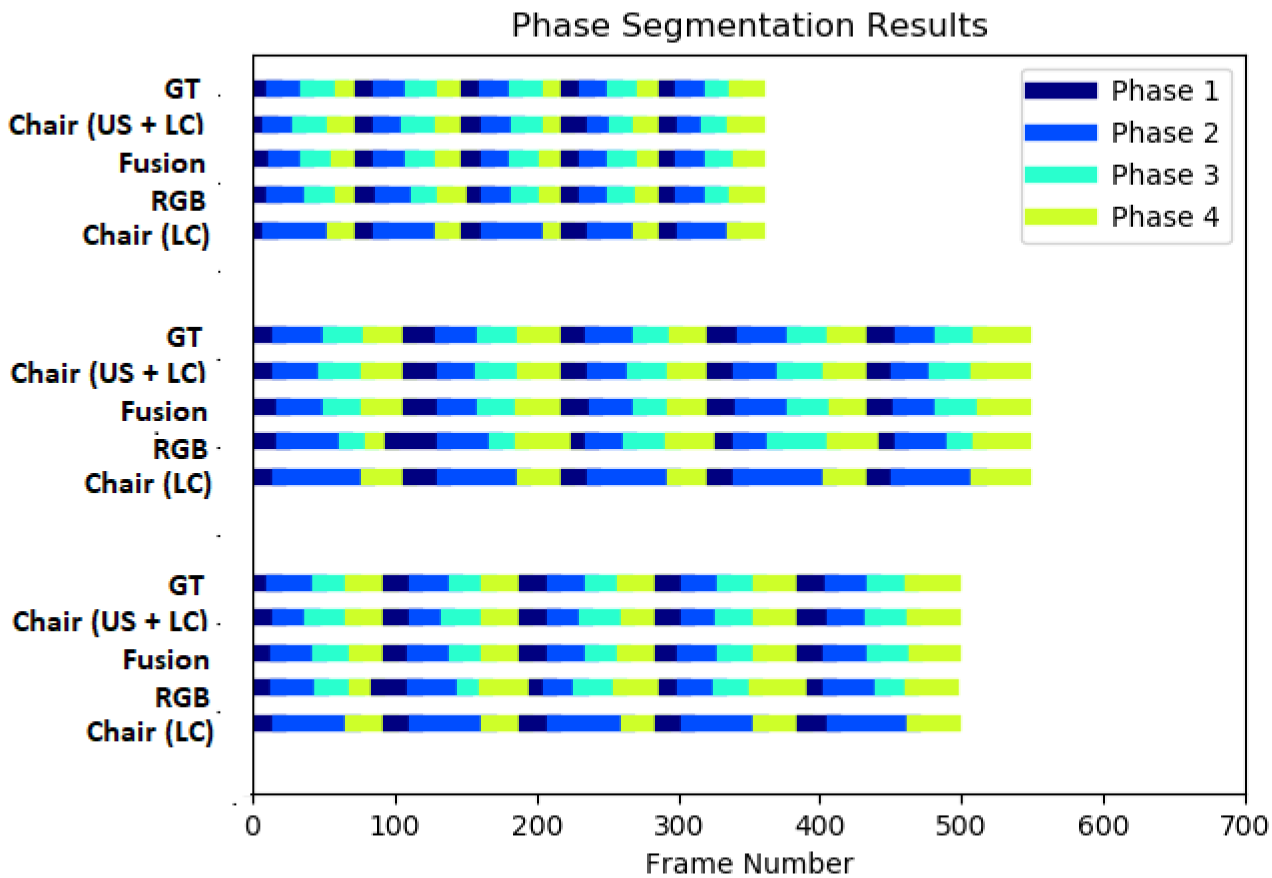

Figure 7.10 : Qualitative comparison of the three techniques for phase detection of the STS. GT: Ground Truth, US: Ultrasound; LC: Load Cell.

weighed 62.0 ± 12.2 kg, were of height 1.66 ± 0.06 m, and had a BMI of 22.5 ± 4.3 kg.m2. Each person performed the 5STS twice, at two self-selected fast and slow speeds. All subjects provided signed informed consent, with ethical approval granted from the Asian Centre for Medical Education, Research & Innovation (ACMERI/18/001).

Ground truth for each test was provided by an expert, who identified the frames corresponding to the boundaries of each of the four phases of the STS for the five movements performed in each test. The percentage of each frames or data points within the boundaries of each phase was calculated for all participants for each of the three methods. The ceiling effect of a 100% accuracy meant that data was not normally distributed, therefore bootstrapping was used for statistical analyses.

7.8.3 Results

A qualitative comparison of the detection of the individual phases for three trials is shown in Figure 7.10. It can be seen that the chair is unable to detect the end of Phase 2 when the person being tested is no longer in contact with the chair.

The accuracy of the segmentation for all methods is shown in Figure 7.11 for all trials. It can be seen that there was substantial variation between subjects, particularly for the RGB method, which on some occasions obtained poor results of less than 20% classification accuracy. Likewise, the accuracy of the chair using only the load cells was limited due to an inability to detect the end of Phase 2. However,

Table 7.2 : Comparison of segmentation accuracy for the three systems. Data are bootstrapped means and 95% confidence intervals.

Method	Accuracy
Chair Load cells	69.8 (68.1, 71.3)
RGB Camera	78.0 (64.4, 86.6)
Fusion RGB Camera and Chair Load Cells	85.5 (78.4, 90.6)
Chair Load Cells and Ultrasound Sensor	89.0 ± 86.3, 91.3)

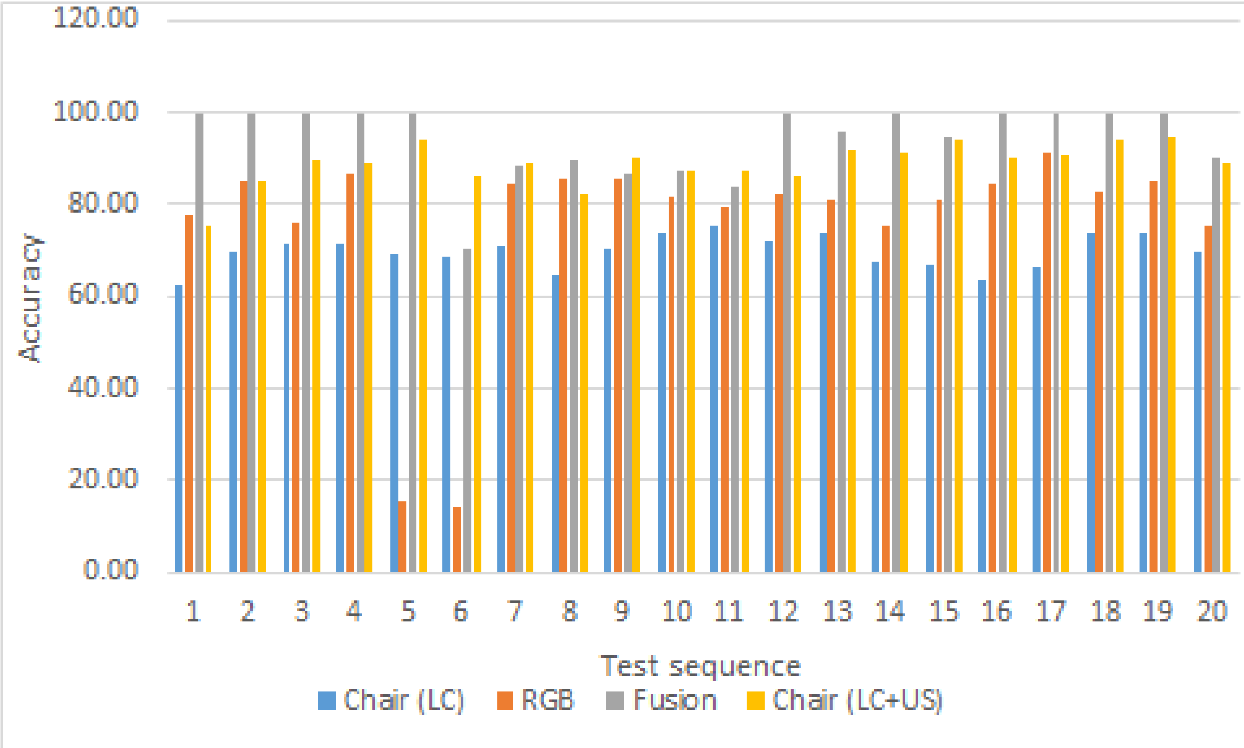

Figure 7.11 : Comparison of Segmentation Accuracy of three techniques for 15 elderly subjects for 4 phases

when fusion was used for RGB and Load Cells or when all chair sensors were used, more consistent and higher segmentation accuracy was obtained.

The accuracy of the different segmentation methods is shown in Table 7.2. When the methods were compared, overall there was a significant difference effect of the method on the results (F=7.68, p=0.000). When the individual methods were compared, there was no significant difference between the RGB and Chair Load Cell methods (mean difference 5.2; 95% confidence interval 2.7, -16.3). The iSTS method had the highest segmentation accuracy, with significantly better results than the load cells alone (19.2; 16.0, 22.2) and the RGB camera alone (11.0; 2.6, 21.7). There was no significant difference between the fusion method of the chair with load cells and ultrasound sensor (3.4; -2.3, 10.4).

7.9 DISCUSSION

Instrumented chair containing load cells and ultrasonic sensors performs comparable with fusion of RGB camera and the original chair having only load cells. The best values were obtained for the instrumented chair, although both systems performed better than any single system alone. The comparable performance of the fully instrumented chair is encouraging, with a single system offering advantages over the fusion method as expert analysis is not needed to develop templates as part of the analysis.

...