Exploring Clinical Correlations in Centroid-Based Gait Metrics from Depth Data Collected in the Home

Robert Wallace
EECS Department
University of Missouri
Columbia, MO, USA

Carmen Abbott
School of Health Professions
University of Missouri
Columbia, MO, USA

Marjorie Skubic
EECS Department
University of Missouri
Columbia, MO, USA

ABSTRACT

A longitudinal study in the home setting using inexpensive depth cameras was done over 34 months to investigate the ability to predict clinical events. Previous work developed a set of metrics based upon the movement of the centroid computed from segmented depth data [14]. A predictive analysis method is developed allowing the identification of significant changes in the subject's gait. These changes are compared to the subject's clinical events and correlated with standard Fall Risk Assessments (FRA). The method developed here allows the proper clustering of all purposeful walks in the residence to isolate the subject from visitors, and identification of significant changes using a set of metrics unique to each subject. Correct detection of events and non-events ranged between 75% and 94% across a set of 7 residents. These predicted events were also found to correlate strongly with established monthly FRAs.

Author Keywords

Gait; Depth data; Fall Risk; Clustering

ACM Classification Keywords

Algorithms

INTRODUCTION

More than one third of older adults fall each year [6][10]. According to the United States Centers for Disease Control and Prevention, in 2012, the direct medical cost of falls among older adults, adjusted for inflation, was over $30 billion [17]. Falls have significant impact on morbidity and mortality of older adults, and are the leading cause of both non-fatal and fatal injuries in those 65 and older [2].

To better monitor the population of older adults in order to identify an increased risk of falling prior to a fall or between fall risk assessments, a set of metrics was developed that has been used to characterize a subject's walks in a naturalistic home environment in real time. Ultimately, these metrics can be integrated into an automated monitoring system to provide gait assessment between fall risk assessments, as well as used to alert clinicians to sudden changes in condition due to stroke, disease progression, and other clinical conditions. Due to the home setting, we focused on the inexpensive depth camera in the Microsoft Kinect.

PervasiveHealth '17, May 23–26, 2017, Barcelona, Spain

ACM ISBN 978-1-4503-6363-1/17/05…$15.00
https://doi.org/10.1145/3154862.3154890

BACKGROUND

Gait Analysis

Two methods currently being used to measure gait parameters using inexpensive depth cameras involve the use of skeletal data, and depth data with the feet in view. A validated method of determining gait parameters based upon skeletal data was developed by Xu *et al.*[16] Stone *et al.* used depth data with the feet in view to compute gait parameters [11][12]. One disadvantage of these approaches, when used in a typically cluttered residence, is the likelihood of occlusion. The approach used in this study focuses on the centroid of the subject's image and not the feet. The centroid is much less likely to be occluded by furniture or other contents of the subject's home.

Fall Risk Assessments

In this study, performance against five fall risk assessments was evaluated. The FRA's were administered monthly for the duration of the study. The "Timed Up and Go" (TUG) test measures the time required to stand, walk ten feet, turn, return and sit [9]. The Berg Balance Scale (BBS) is a test of 14 different items and is used to monitor fall risk principally relying on assessment of the patient's balance [1]. The Short Physical Performance Battery (SPPB) test [4] is commonly used to assess lower extremity strength and has been shown to have some correlation to recurrent falls [13]. Habitual gait speed (HGS) has been studied as a means to detect a significant change in a person's walking speed [3][7] as a predictor of falls. The Functional Reach Assessment (FReach) guides the patient through a series of reach tasks designed to gauge not only the flexibility, but also the patient's ability to balance.

METHODS

Clustering for Resident Separation

With approval from the University IRB, Microsoft Kinect cameras were placed into the apartments of residents in an assisted living facility. Data collection and pre-processing were done as described in [14] for all 11 residents. The residents were tracked and segmented to obtain a point cloud from which the centroid values are determined. Walks were captured continuously for at least 4 months and for as many as 34 months. Walks were limited to purposeful walks, which are defined as relatively straight walks at least 1.22 meters long with a speed at least 12.7 cm/s and duration of at least 1 second [11]. Sixty metrics were calculated for each walk and stored. Algorithms and validation results for the metrics are described in [14].

The raw data contains all walks recorded in the residence, which may include walks made by any visitors, staff or even another resident in a multi-resident apartment. To isolate the walks made by the resident being studied, the set of walks was broken into chunks of 14 days and clustered. By clustering the data using smaller chunks we minimize inaccuracies due to long term trends resulting from changes in the subject's health.

By studying the Improved Visual Assessment of Cluster Tendencies (iVAT) images [5] for each metric, the eight metrics that looked the most promising across all residents were identified. The metrics used are shown in Table 1. The four most promising metrics were "required" for each attempt while some combination of zero to four of the remaining were chosen to maximize the Davies-Bouldin index for that chunk.

To improve clustering, the data were transformed using principle component analysis (PCA) to maximize the orthogonality between features. The PCA transformed data were then clustered using the possibilistic fuzzy c-means algorithm (PFCM) [8]. Each of the 16 combinations of metrics was clustered in this manner; the clustering with the best Davies-Bouldin index was selected.

Instead of using the raw metrics, the rate of change of the normalized metric was used. We wanted a more dramatic change to be flagged as an event. This is done by first computing a baseline mean, standard error about the mean, and Minimum Detectable Change (MDC) over the first 30 days. For the remaining points, the number of MDCs that the point deviates from the baseline mean is computed. The daily mean of the slopes of the MDC curve for each walk during the day are computed and used in the next section.

Algorithm	***Metrics Used***
Clustering	Step Time*, Height*, Speed*, Avg. Instantaneous Speed*, Trunk Sway, Step Ratio, Step Length, Average XY Entropy
Pred. Anal.	Step Length, 10' Walk, Speed, Stride Time, Left & Right Step Time, Trunk Sway, Step Time, Vertical Asymmetry, Peak to Peak in X and Z directions, Step Ratio, and Average Entropy in X, Y, Z, and XY directions.

Table 1: Initial set of metrics which are optimized during clustering and predictive analysis. For each resident, the set is reduced to the smallest set that will achieve the best Davies-Bouldin (clustering) or True Positive / False Positive score (Predictive Analysis). Metrics marked with an asterisk are required for clustering

Predictive Analysis

It was hypothesized that outliers in each cluster would correspond to significant changes in gait metrics and, consequently, higher risk of a clinical event (falls, emergency department visit, etc.). It was felt that a sudden change in one or more metrics might portend a clinical event. These changes, if significant, would fall outside the resident's "normal" values and be detectable as outliers. The One Class Possibilistic C-Means (OCPCM) was used for outlier detection due to its stability with respect to outliers. Removal of an outlier will impact the center to a much lesser extent than removal of a true member of the cluster [15].

Outlier Identification

A detailed algorithm is shown in Algorithm 1. The approach is to repeatedly run the OCPCM algorithm, removing the point with the lowest possibility after each run. As more points are removed, the movement of the cluster center will be more dramatic. At some point, the increase in movement will level off as points within the cluster are removed.

Identifying Events

To identify the best set of outliers, we first need to select the optimal set of metrics. A detailed algorithm is shown in Algorithm 2. Each metric is removed from the set, outliers are identified, and scored. The best set is passed recursively into the same routine where an attempt is made to remove another metric. The recursion ends when scores are not improved. The best set of metrics is then run through algorithm 1 and the outliers are labeled as predicted events.

```
1.  With full dataset, run OCPCM algorithm to obtain list of possibilities
2.  Store location of cluster center
3.  Repeat until 50% of data has been removed.
4a.    Remove point with lowest possibility and re-run OCPCM
4b.    Store distance center moved due to point deletion
5.  Find mean of center movements for second half of removed points
6.  Identify the first point that is > 75% of this value.
7.  Label all points with a lower possibility as outliers.
```

Algorithm 1: Outlier identification using OCPCM

```
1.  Start with set of Predictive Analysis metrics (Table 1)
2.  Run Algorithm 1 on full dataset; compute True Positive score (TP)
    and False Positive score (FP)
Recursive Entry Point (pass data, TP score, and FP score):
3.  Repeat for all metrics in dataset.
3a.         Remove one metric from dataset.
3b.         Run Algorithm 1 and compute TP and FP
4.  Select metric set with best TP and FP
5.  IF (TP improves) OR (TP stays same, AND FP improves)
    THEN recursively call passing new dataset and scores.
    ELSE return passed dataset to caller.
```

Algorithm 2: Find best set of metrics to identify outliers

Scoring Results

Four different scores are computed to gauge performance of these algorithms. A true positive is scored when an outlier can be paired with a clinical event (recorded fall, emergency department visit, or hospitalization) within 7 days. Once a true positive is scored, no other outliers will be considered during the seven-day time window. A false positive is scored when an outlier is not followed by a clinical event.

Two other scores were used to gauge performance. The first is the percentage of correctly classified points. This is simply the sum of the true positives and the true negatives (number of points lying in the cluster that do not correspond to clinical events). The other additional score is the number of clinical events that have outliers within seven days. This score represents a change in metrics following a clinical event.

Correlation to FRA

For each resident, the set of optimal metrics and an aggregate of these metrics were evaluated against the resident's monthly fall risk assessments. To generate the aggregate, the normalized optimal metrics are averaged correcting for those metrics for which a decrease indicates a worsening physical state. As the assessments were performed monthly, they were compared against the average of each metric over 14 days centered on the FRA date. Finally, Pearson Correlation was done to evaluate the null hypothesis that the FRA data and the metric being considered are independent. The value of ρ and p are reported.

RESULTS

Clustering for Resident Separation

While some limited ability to separate resident's data was seen when using the raw data to cluster, transforming the data using PCA significantly improved the ability of the PFCM algorithm to isolate the resident's walks from the walks of others. Figure 1 shows the separation of one 14-day chunk of data for a two resident apartment. In the figure, the clustering of the PCA transformed data showed two distinct clusters split primarily based upon component #1 along a nearly vertical line. When the clustering is viewed using metrics such as total height and the average time for a single step, the clustering is still apparent. The entire set of 11 residents were clustered using this approach and all 11 yielded similarly well-clustered results.

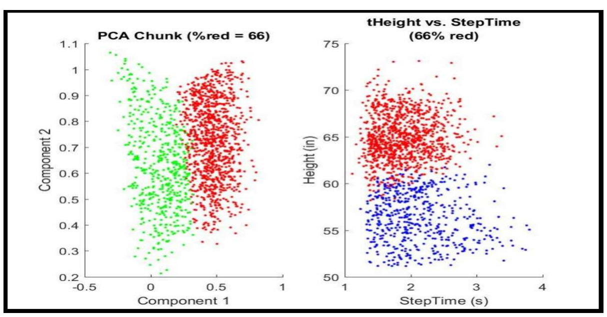

Figure 1: Clustering on PCA transformed data. The plot on the left shows the first two PCA transformed components after clustering using PFCM. The plot on the right shows total height plotted against stride time. More than two features were used to cluster the data, however only two were shown to clearly illustrate clustering.

While each resident clustered well, the set of metrics used for each resident was different. Resident 1, for example, generated the best results with 8 metrics. Resident 2 required 9 metrics, resident 3 needed 15 metrics to achieve the best results, while resident 4 needed 14 metrics. For all residents considered, the Entropy in the X and Z directions, step time (average time for all steps in the walk), and left step time (average time for all steps *by the left leg* in the walk) were the metrics most often used to cluster.

Predictive Analysis

The procedure described was run on data from 11 residents. Seven residents had corresponding clinical events while four did not. The results are shown in Table 2. For many residents (63%), a large number of metrics was needed to achieve the best results. Resident #2 had the lowest true positive rate. This resident, however, had multiple falls recorded each day over the latter 2/3 of the study period. In this case where falling is normal, one would not expect "outliers" to predict a higher likelihood of falling as falling is the new normal. Of the remaining residents, 3 were in the second quartile, 2 in the third, and 1 in the fourth. All 11 had a false positive rate less than 25% while more than half were at or below 10%. The residents showing N/A for true positive and change did not have any recorded clinical events over the study period.

	M/F	#Metrics	TruePos	FalsePos	Correct	Change
1	M	7	60%	16%	83%	80%
2	F	10	18%	11%	80%	28%
3	F	8	25%	6%	93%	75%
4	F	14	33%	25%	75%	83%
5	F	9	80%	7%	93%	60%
6	F	16	N / A	6%	94%	N / A
7	F	15	50%	22%	78%	50%
8	F	16	N / A	7%	93%	N / A
9	M	16	N / A	7%	93%	N / A
10	F	16	N / A	8%	92%	N / A
11	M	15	26%	17%	81%	N / A

Table 2: Scores for eleven residents. M/F shows resident's gender. #Metrics indicates the number of metrics used for this resident. The "Correct" column shows the percentage of correctly identified points (TruePositive + True Negative). "Change" shows the percentage of clinical events which are followed by a significant change in metrics within seven days.

Correlation to Fall Risk Assessments

Lastly, for each of the 5 subjects that did record clinical events, and for whom fall risks assessments were performed, the events were correlated to the fall risk assessments as described above. The results are shown in Table 3. All five residents had at least one metric that was strongly correlated to one of the FRA while three residents had more than three metrics that exhibited strong correlation to an FRA. As can be seen elsewhere in this study, the optimal set of metrics and FRA are unique to each resident, though either BBS or TUG are present in all five of these residents.

DISCUSSION

A recurring theme of these metrics, both in this study, and in previous studies, is the uniqueness of the combinations for each resident. The relationship is likely a result of the subject's medical condition and diagnosis. In the clustering, event identification, and correlation components, a search was done for the optimum set of metrics. The results in Table

2 show a different number of metrics for many subjects from the entire set of 16 down to as few as 7.

Clustering produced clean results once PCA was used. This is as expected as the PCA transformation maximized the orthogonality between each feature helping to create a new set of features which could easily be separated by the PFCM.

Subj	Metric	FRA	ρ (p)
1	10' Walk	TUG	0.71 (0.00)
	Left Step Time	BBS	0.70 (0.00)
	Aggregate	BBS, TUG	0.69 (0.00)
3	Aggregate	BBS	-0.61 (0.01)
	Entropy X	BBS	-0.60 (0.01)
	Entropy Z	BBS	-0.57 (0.02)
4	Stride Time	FReach	0.54 (0.02)
	Aggregate	HGS	-0.43 (0.06)
	10' Walk	BBS	0.42 (0.07)
5	Left Step Time	TUG	0.51 (0.05)
	Left Step Time	SPPB	0.51 (0.06)
	Left Step Time	HGS	0.49 (0.07)
7	Step Ratio	TUG	-1.00 (0.00)
	Entropy Y	BBS	-1.00 (0.00)
	Avg. Speed	FReach	-1.00 (0.02)

Table 3: Strongest correlations between metrics and FRA's for each resident with reported clinical events. Subjects 2 and 11 are not included as no FRA assessments were captured for these residents.

Predictive analysis also had good results. Forty-two percent of the residents yielded a score of 50% or higher with one at 80%. False positives for all residents were at or below 25% with a number of residents at or below 8%.

Beyond the inherent variability in the physical conditions, medical history, and diagnoses of the residents, there are a few explanations for these results. First, the threshold separating outliers is set to a fixed 75% of the mean value of the non-outliers. As expected, adjusting this threshold will improve the true positive rate at the cost of the false positive rate. A better approach may be to train a model based upon the resident's clinical event history. In future work, we will explore this strategy. A second source of error can result from using reported clinical events. Some events may go unreported. It is easy to imagine a change in the resident's medical state, such as a mild stroke that causes stumbles or minor falls, which may not be reported to clinical staff. While these should be categorized as true positives, they would be classified as false positives since the outlier did not result in a reported clinical event. Integrating this type of system with a fall detection system might improve results under these circumstances.

Outliers would also be likely if the clustering had errors. If, after clustering, a walk by a visitor was included, that walk would increase the likelihood that the daily average would change significantly enough to cause an outlier for the day. As the resident's clinical condition did not change, this would be classified as a false positive.

CONCLUSIONS

This study shows that it is possible to predict significant clinical events using centroid generated gait metrics from in-home depth data. While there is some potential for improvement, perhaps using a dynamic threshold to classify outliers, or integrating this system with a fall detection system, the results clearly indicate some ability to predict clinical events. Furthermore, results have also shown correlation of these metrics to established fall risk assessment instruments providing support for the use of the metrics to assess physical ability.

REFERENCES

1. Katherine Berg. 1989. Measuring balance in the elderly: preliminary development of an instrument. *Physiotherapy Canada* 41, 6: 304–311.
2. G Bergen, MR Stevens, and ER Burns. 2016. Falls and Fall Injuries Among Adults Aged ≥ 65 Years - United States, 2014. *MMWR Morb Mortal Wkly Rep* 65: 993–998.
3. Richard W. Bohannon. 1997. Comfortable and maximum walking speed of adults aged 20—79 years: reference values and determinants. *Age and ageing* 26, 1: 15–19.
4. Jack M. Guralnik, Luigi Ferrucci, Eleanor M. Simonsick, Marcel E. Salive, and Robert B. Wallace. 1995. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. *New England Journal of Medicine* 332, 9: 556–562.
5. Timothy C. Havens and James C. Bezdek. 2012. An efficient formulation of the improved visual assessment of cluster tendency (iVAT) algorithm. *IEEE Transactions on Knowledge and Data Engineering* 24, 5: 813–822.
6. Pekka Kannus, Jari Parkkari, Seppo Koskinen, Seppo Niemi, Mika Palvanen, Markku Järvinen, and Ilkka Vuori. 1999. Fall-induced injuries and deaths among older adults. *Jama* 281, 20: 1895–1899.
7. H.-K. Kuo, Suzanne G. Leveille, Y.-H. Yu, and William P. Milberg. 2006. Cognitive function, habitual gait speed, and late-life disability in the National Health and Nutrition Examination Survey (NHANES) 1999–2002. *Gerontology* 53, 2: 102–110.
8. N.R. Pal, K. Pal, J.M. Keller, and J.C. Bezdek. 2005. A possibilistic fuzzy c-means clustering algorithm. *IEEE Transactions on Fuzzy Systems* 13, 4: 517–530.
9. D Podsiadlo and S Richardson. 1991. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society* 39, 2: 142—148.
10. P. Scuffham, Stephen Chaplin, and Rosa Legood. 2003. Incidence and costs of unintentional falls in older people in the United Kingdom. *Journal of epidemiology and community health* 57, 9: 740–744.
11. Erik E. Stone and Marjorie Skubic. 2013. Unobtrusive, Continuous, In-Home Gait Measurement Using the Microsoft Kinect. *IEEE Transactions on Biomedical Engineering* 60, 10: 2925–2932.
12. Erik Stone, Marjorie Skubic, Marilyn Rantz, Carmen Abbott, and Steve Miller. 2015. Average in-home gait speed: Investigation of a new metric for mobility and fall risk assessment of elders. *Gait & Posture* 41, 1: 57–62.
13. N Veronese, F Bolzetta, E Toffanello, S Zambon, M De Rui, E Perissinotto, A Coin, M Corti, G Baggio, G Crepaldi, G Sergi, and E Manzato. 2014. Association Between Short Physical Performance Battery and Falls in Older People: The Progetto Veneto Anziani Study. *Rejuvenation Research* 17, 3: 276–284.
14. R. Wallace, C. Abbott, C. Gibson-Horn, M. Rantz, and M. Skubic. Metrics from In-Home Sensor Data to Assess the Effect of Clinical Therapies. *Smart Health Journal* in press.
15. Shuang Wang. 2011. Change Detetction for Eldercare Using Passive Sensing. University of Missouri–Columbia.
16. X. Xu, R. McGorry, L Chou, J Lin, and C Chang. 2015. Accuracy of the Microsoft Kinect^{TM} for measuring gait parameters during treadmill walking. *Gait & Posture* 42, 2: 145–151.
17. Cost of Fall Injuries in Older Persons in the United States, 2005. *Centers for Disease Control and Prevention*. Retrieved Nov. 22, 2014