

How the Makusafe data set is different

why we shouldn't expect common fall detection schemes to work on Makusafe data

Alden Bradford

Purdue University / The Data Mine

March 28, 2022

Other fall detection schemes

We have based a lot of our strategy on two recent papers, both of which study fall detection using only accelerometer data.

- ▶ (Althobaiti, Katsigiannis, and Ramzan 2020) provides a helpful literature review, describing various strategies people have used for fall detection and the degree of success they had. The article is mainly a description of an experiment where research participants performed motions in a controlled environment, and researchers attempted to build a fall detection strategy based on the measured data.
- ▶ (Šeketa et al. 2021) focuses on using a data segment (window) centered around a high peak acceleration, and seeks to find a good choice of window size.

Our goal here

These papers both had the goal of differentiating between hazardous movements (mainly falls) and nonhazardous movements called "activities of daily living" (ADL). Both reported high F-scores (over 95%), which is a combined measure of positive and negative predictive value. Most of the features we implemented were suggested by these two papers.

So far, our attempts to classify the makusafe data (Makusafe 2022) have been able to get F-scores of only around 20%. In order to understand this discrepancy, we set about trying to reproduce the methods used in (Šeketa et al. 2021), to see if we can figure out why we have not seen the same success.

Other data sets

In (Šeketa et al. 2021), the authors rely on three publicly available datasets:

FallAIID (SALEH and LE BOUQUIN JEANNES 2020)

SisFall (Sucerquia, López, and Vargas-Bonilla 2017)

Erciyes (Özdemir and Barshan 2014)

The FallAIID and SisFall datasets both have records with significantly higher sample rates than (Makusafe 2022). To ensure a fair comparison, we used Fourier analysis to resample all data sets to 25 Hz. The Erciyes dataset was already sampled at 25 Hz. Following the method in (Šeketa et al. 2021) we found all peaks higher than a given threshold (different for each set) and extracted an 8-second window centered around each peak. This gave us several thousand samples, labeled by their motions.

Exploration through factor analysis

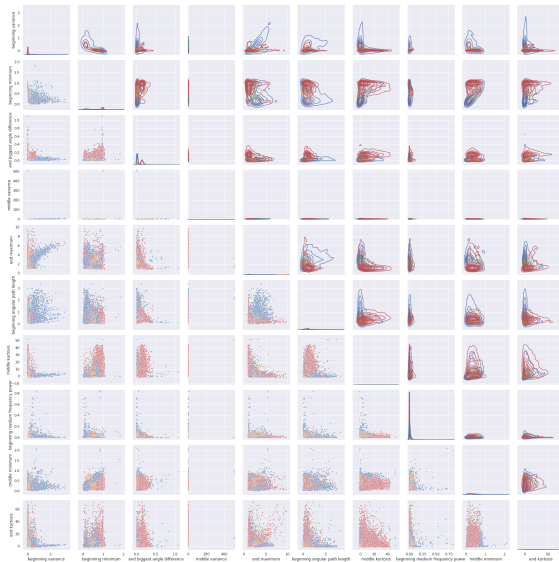
For each sample, we computed all the features recommended in (Šeketa et al. 2021).

By using factor analysis, we were able to select ten features which explain most of the variance in the feature space.

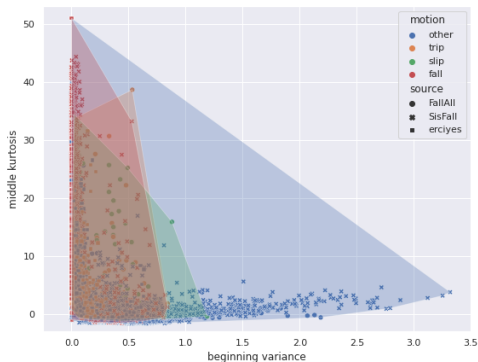
	beginning maximum				beginning minimum				beginning range				beginning mean				beginning standard deviation				beginning variance				beginning skew				beginning kurtosis				beginning total variation				beginning angular path length				beginning biggest angle difference				beginning low frequency power				beginning medium frequency power				beginning high frequency power				middle maximum				middle minimum				middle mean				middle standard deviation				middle variance				middle skew				middle kurtosis				middle total variation				middle angular path length				middle biggest angle difference				middle low frequency power				middle medium frequency power				middle high frequency power				end maximum				end minimum				end range				end mean				end standard deviation				end variance				end skew				end kurtosis				end total variation				end angular path length				end biggest angle difference				end low frequency power				end medium frequency power				end high frequency power			
0	0.95	-0.71	0.92	0.83	0.95	1.00	0.16	-0.23	0.96	0.29	0.30	0.17	0.13	0.96	0.05	-0.33	0.07	0.35	0.33	-0.02	-0.30	-0.32	0.17	-0.33	-0.33	-0.06	-0.06	-0.04	0.67	-0.55	0.60	0.69	0.79	0.67	0.03	-0.23	0.83	-0.28	-0.28	0.06	0.00	0.07																																																																																																																										
1	-0.22	1.00	-0.43	0.22	-0.35	0.10	-0.03	0.30	-0.25	-0.04	-0.01	0.03	-0.24	0.10	0.03	0.34	0.01	0.33	0.02	-0.01	0.15	-0.03	-0.05	0.18	0.18	-0.01	-0.01	-0.01	0.08	0.33	-0.01	0.35	0.01	0.22	-0.04	0.02	0.04	-0.01	-0.01	0.04	-0.02	0.22																																																																																																																										
2	-0.19	0.29	-0.22	-0.08	-0.19	-0.09	-0.10	0.00	-0.17	0.12	0.07	-0.02	-0.10	-0.08	0.10	-0.07	0.10	0.03	0.07	-0.04	0.37	0.28	-0.02	0.79	0.77	-0.02	-0.03	-0.04	-0.21	0.33	-0.28	-0.13	-0.25	-0.11	-0.52	0.28	-0.27	1.00	1.00	-0.02	-0.11	-0.10																																																																																																																										
3	0.02	0.00	0.01	0.02	0.02	0.03	0.00	-0.03	0.02	0.09	0.09	0.01	0.01	0.03	0.97	-0.02	0.30	0.91	1.00	0.13	0.15	0.54	0.03	0.03	1.00	1.00	1.00	0.02	-0.00	0.01	0.02	0.01	0.03	0.01	0.01	0.01	0.02	0.02	0.01	-0.01	0.02																																																																																																																											
4	-0.06	0.04	-0.05	0.04	-0.01	-0.01	-0.03	-0.04	0.03	-0.00	0.02	-0.05	-0.21	-0.02	-0.01	-0.01	0.25	0.08	-0.03	-0.23	-0.37	0.21	-0.22	-0.22	-0.01	-0.02	-0.01	1.00	-0.70	0.90	0.54	0.75	0.57	0.48	0.13	0.15	-0.10	-0.10	0.53	0.56	0.47																																																																																																																											
5	0.13	-0.11	0.13	0.20	0.14	0.11	0.03	-0.06	0.12	1.00	1.00	0.03	-0.02	0.10	0.04	0.02	0.04	0.16	0.07	0.02	0.00	0.01	0.08	0.14	0.14	0.02	0.02	0.02	0.01	0.04	-0.00	0.13	0.04	0.05	-0.07	-0.08	0.04	0.03	0.03	-0.02	-0.05	0.05																																																																																																																										
6	-0.07	0.19	-0.10	-0.10	-0.13	-0.05	0.09	0.04	-0.10	0.03	-0.01	-0.01	0.01	-0.06	0.25	0.06	0.24	-0.09	0.10	-0.09	0.96	1.00	0.03	0.26	0.26	-0.05	-0.05	-0.06	-0.17	0.36	-0.22	-0.16	-0.22	-0.08	0.02	0.03	-0.18	0.15	0.15	-0.02	-0.05	-0.06																																																																																																																										
7	0.67	-0.02	0.53	0.56	0.12	-0.21	0.97	0.31	-0.03	0.04	0.00	0.63	1.00	-0.35	0.18	0.10	0.17	0.97	0.23	-0.12	0.07	-0.08	0.32	-0.12	-0.12	-0.08	-0.10	-0.13	0.09	0.27	0.01	0.38	-0.12	-0.22	0.17	0.09	-0.16	0.02	0.02	0.36	0.11	-0.29																																																																																																																										
8	0.11	0.26	0.02	0.14	-0.09	-0.05	0.27	0.34	-0.08	-0.01	0.01	0.14	0.08	-0.07	-0.17	1.00	-0.23	-0.31	-0.30	0.17	-0.01	0.03	0.47	-0.10	-0.10	0.12	0.14	0.19	0.04	0.20	-0.01	0.06	-0.07	-0.05	0.18	0.08	-0.06	0.02	0.02	0.09	-0.04																																																																																																																											
9	0.15	-0.03	0.12	0.12	0.07	0.11	0.04	-0.03	-0.00	-0.02	-0.03	0.16	0.33	0.11	0.02	0.15	0.01	0.05	-0.03	-0.00	0.03	-0.01	-0.11	0.01	0.01	0.01	0.01	-0.01	0.28	0.26	0.17	-0.37	-0.30	-0.28	0.62	1.00	0.50	0.04	0.04	0.34	0.32	-0.62																																																																																																																										

Feature scatter plot

Each point is an incident, and its color is a motion.



One of those plots caught our eye



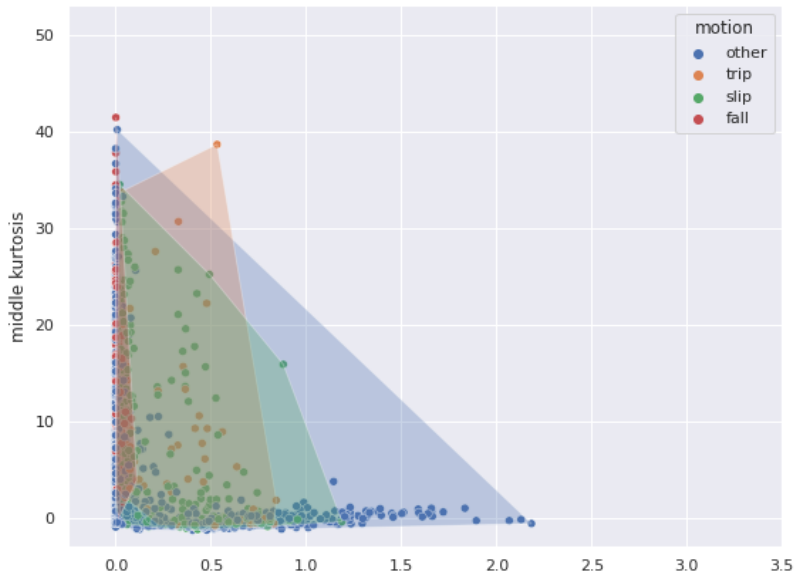
Notice that most of the points with a high variance at the beginning of the sample are ADL, while most of the samples with high kurtosis in the middle are falls.

Kurtosis, Variance

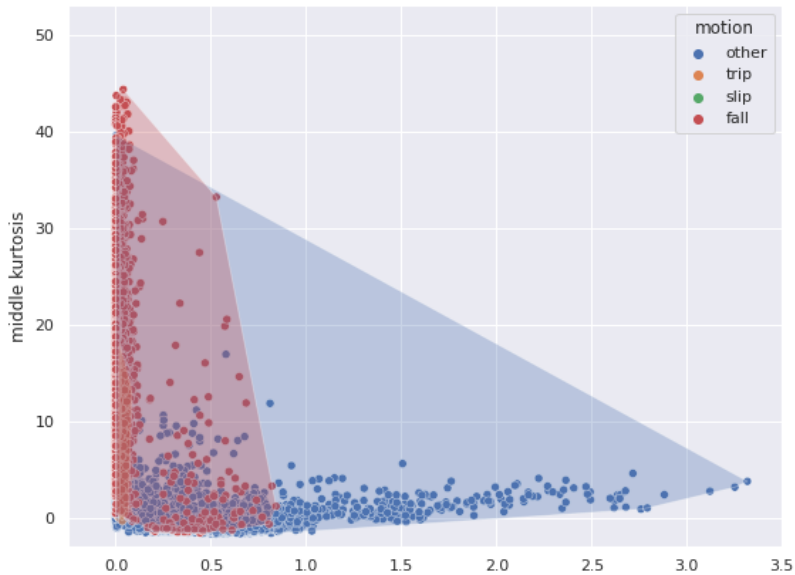
Variance at the beginning measures how much movement was happening before the incident. It makes sense that a lot of motion would increase the chance of a false alarm.

Kurtosis in this case measures how big the central peak was, compared to the typical variation of the sample. It makes sense that a large peak would correspond to a hazardous motion.

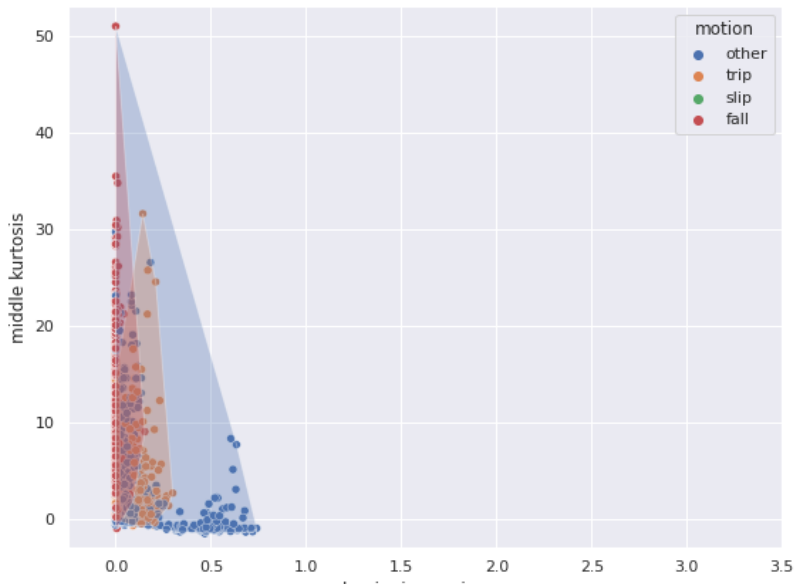
Just the FallAID data



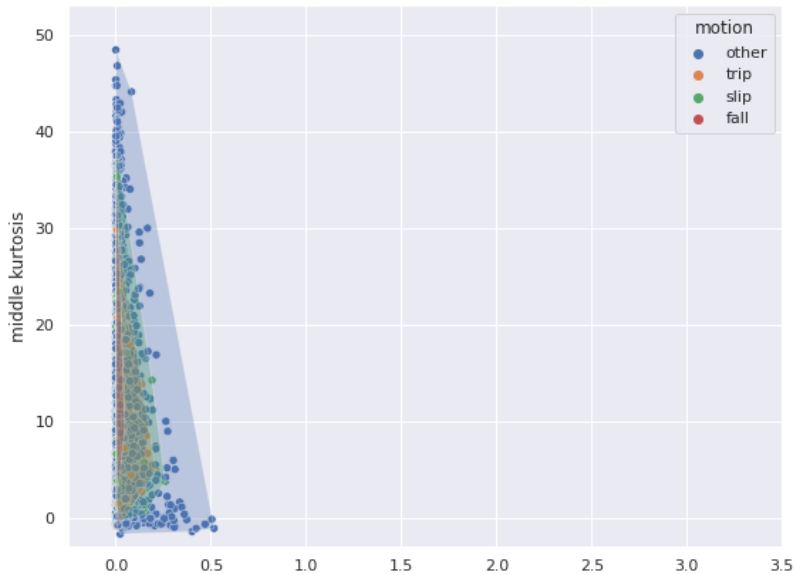
Just the SisFall data



Just the Erciyes data



How does makusafe compare?



Conclusions

- ▶ There is something different about the Makusafe data, something not explained by differences in sampling rate, different trigger thresholds, different mounting locations, or a lack of gyroscope.
- ▶ The hazards are almost trivially easy to separate in the other data sets, using the given features.
- ▶ We should not expect to be able to use the same strategy on the Makusafe data.





Potential causes

- real versus simulated** The other data sets are using simulated falls in a controlled environment. This type of motion may be significantly different from a real fall.
- range of activities** Since most sources are interested in fall detection among the elderly, they consider only ADLs which the elderly are likely to perform. The activities Makusafe customers engage in may not be well-covered by those types of non-hazardous activity.
- differences in definition** It could be that there are incidents in the Makusafe set which would be classified as a fall by the standards of these other papers, but which Makusafe customers would not consider worth reporting.

Future directions

- ▶ We should not expect to be able to apply strategies from other papers to our data.
- ▶ We should reconsider our choice of features – look for a pattern which applies specifically to Makusafe data.
- ▶ We should reconsider what we want our goal to be. It may be unreasonable to think we can find a classifier with a high positive predictive value. Some alternatives are:
 - ▶ find a classifier with a high negative predictive value. Even if we can't pick out the hazards, we may be able to remove the bulk of the non-hazards.
 - ▶ Build an aggregate index of risk – something which can predict what fraction of incidents matching a given profile will be classified as hazards. This could be used to identify dangerous patterns without necessarily being able to identify dangerous incidents.

References I

-  Turke Althobaiti, Stamos Katsigiannis, and Naeem Ramzan. “Triaxial accelerometer-based falls and activities of daily life detection using machine learning”. In: *Sensors* 20.13 (2020), p. 3777.
-  Makusafe. *Makusafe Human Activity Recognition dataset*. 2022. URL: <https://github.com/nikdata/human-activity-recognition>.
-  Ahmet Turan Özdemir and Billur Barshan. “Detecting falls with wearable sensors using machine learning techniques”. In: *Sensors* 14.6 (2014), pp. 10691–10708.
-  Majd SALEH and Régine LE BOUQUIN JEANNES. *FallAIID: A Comprehensive Dataset of Human Falls and Activities of Daily Living*. 2020. DOI: 10.21227/bnya-mn34. URL: <https://dx.doi.org/10.21227/bnya-mn34>.

References II



Goran Šeketa et al. “Event-centered data segmentation in accelerometer-based fall detection algorithms”. In: *Sensors* 21.13 (2021), p. 4335.



Angela Sucerquia, José David López, and Jesús Francisco Vargas-Bonilla. “SisFall: A fall and movement dataset”. In: *Sensors* 17.1 (2017), p. 198.