Object Recognition

Practical III – Human Behaviour Analysis

Author: Erol Kazancli

1. Introduction

In this practical work, Dynamic Time Warping algorithm is used to calculate distances between different gesture types to determine what type of a gesture a given gesture sequence is. Several experiments are done with different approaches. Some heuristics are followed which are explained for all the experiments.

2. Implementation

I implemented the DTW algorithm based on the explanations in the Wikipedia page for DTW using the Euclidean distance as the cost function. I used infinity values for the initial values of the first row and column of the DTW matrix and took the right lower corner value as the distance value between two gestures. The following files are added to the provided code for my experiments:

DataPreparation.m
Training.m
Testing.m
findDistanceDTW.m
findEuclideanDistance.m
translateSeq.m

3. Experiments

3.1 Experiments With a Single Person

For my experiments I chose some training data from a person type for all gesture types. Before adding each gesture data, I aligned them to a specific size using the imresize matlab function. I chose 150 as this specific sequence size since the median looked to be around this value. As a result each gesture had a size of 150 * 80. You can see how this dataset is prepared in the Datapreparation.m matlab file. The training dataset is stored in the aligned.mat data file to be used in the experiments.

For one gesture type, the distances between each pair of the gestures of the same gesture type are calculated to find an upper limit for distance for each gesture type.
The max values found for the training dataset used is as follows:

	1	2	3	4	5	6	7	8	9	10	11	12
max	184	260	204	199	126	182	223	188	197	151	188	284

Table 1 - Upper Limits found for single person experiments

After calculating the max values several test data are chosen from the data not used in the training belonging to the same person. Several samples from different gesture types are tested. I used two approaches to compare the performances. The first is taking the average of the distances to each data in the training data. The second is taking the minimum distance for each gesture type in the training data.

The following table shows the average distance values of the test data of gesture type 1 to the training data for each gesture type. The columns are gesture types, the rows are different samples of gesture type 1 and the values are the DTW average distances to each class.

	1	2	3	4	5	6	7	8	9	10	11	12
s.1	193	338	205	209	170	214	235	253	163	201	166	262
s.2	142	333	215	142	187	171	243	235	141	210	196	247
s.3	140	360	228	154	193	178	265	256	144	218	187	267
s.4	136	327	188	145	168	171	225	223	140	197	176	235
s.5	153	333	211	135	197	179	231	241	160	220	217	236
s.6	247	414	307	211	256	275	318	360	262	318	327	350
s.7	170	370	254	174	214	224	270	283	192	256	231	283
s.8	111	331	215	157	214	171	243	232	139	195	169	235
s.9	118	339	226	154	223	174	253	235	143	210	174	244
s.10	123	364	247	175	231	195	275	256	160	224	182	266

Table 2 - Average distances to classes for gesture type 1

From the results above we can see that in most cases the gesture type are found correctly but in some cases the distance to the 4th, 5th or 9th gesture type turned out to be very close or even less than this distance, which were false matches.

If we also set an upper limit using our previously found maximum values using the training dataset, we fail to find a match in some cases, like the case of 1st and 6th sample in this case. They are simply rejected and we conclude that they do not belong to any of the gestures. This feature may be especially useful in cases when the given sample does not belong to any of the gestures. But it also causes some false negatives as we see here.

The following table shows the minimum distance values of the test samples to the training dataset for each gesture type.

	1	2	3	4	5	6	7	8	9	10	11	12
s.1	131	284	175	130	144	148	212	223	118	181	114	238
s.2	98	282	185	125	155	141	222	209	103	193	140	215
s.3	95	310	199	137	169	154	244	242	102	202	135	236
s.4	94	275	163	120	139	140	205	206	97	188	128	208
s.5	92	286	185	116	164	157	212	213	118	200	151	191
s.6	188	363	228	153	220	241	271	334	225	287	257	276
s.7	97	305	194	140	177	195	239	261	153	238	173	241
s.8	75	279	173	135	188	142	225	219	106	178	108	214
s.9	80	286	187	131	196	142	237	217	107	197	112	225
s.10	92	310	211	158	212	166	259	241	124	207	120	243

Table 3 - Minimum Distances to Classes for gesture type 1

We can see comparing the minimum distances looks safer to find the correct gesture.

The following tables are for the gesture type 2, the average distances and the minimum distances respectively.

	1	2	3	4	5	6	7	8	9	10	11	12
s.1	412	313	275	351	288	349	297	330	361	347	379	331
s.2	357	210	276	303	286	314	301	296	330	302	361	283
s.3	403	233	327	325	305	365	322	389	372	356	424	361
s.4	377	230	273	334	296	330	307	303	343	313	367	302
s.5	366	196	282	337	322	331	301	286	346	300	357	281
s.6	344	192	265	329	311	316	290	275	322	277	337	260
s.7	348	180	282	314	306	319	288	301	330	288	355	280
s.8	381	218	297	350	328	348	316	321	357	319	380	308
s.9	405	254	308	345	316	357	334	345	371	348	397	330
s.10	411	257	317	332	303	369	326	380	378	358	424	356
s.11	406	249	290	334	281	351	313	356	358	346	394	338

Table 4 - Average Distances to Classes for Gesture 2

For this gesture type the boundaries look clearer. In most cases the correct gesture was found and the difference compared to other gestures are significantly lower increasing the confidence degree.

In this gesture type all the samples except the first one pass the max condition.

	1	2	3	4	5	6	7	8	9	10	11	12
s.1	381	173	231	305	246	309	261	303	330	324	342	281
s.2	321	138	251	278	268	284	282	269	311	281	327	258
s.3	355	188	242	284	263	313	268	370	350	334	386	306
s.4	346	147	248	296	272	305	284	278	293	292	328	274
s.5	322	125	270	299	298	303	276	266	316	276	324	257
s.6	303	119	243	282	288	283	264	253	271	259	306	245
s.7	302	129	256	278	286	272	262	277	301	270	325	267
s.8	342	126	273	315	299	318	301	290	302	302	340	289
s.9	368	166	271	314	298	315	320	318	352	329	369	287
s.10	356	187	259	315	273	332	302	355	344	337	388	310
s.11	366	188	264	293	270	293	293	339	340	332	369	273

Table 5 - Minimum distances to classes for gesture 2

With the minimum distances, the distinctions are even clearer. But using the minimum distances we need to be careful with the upper limit condition, which can be more easily fulfilled in this case. So we expect more false positives in cases when the given sample does not belong to any of the gestures with the minimum distance approach. Therefore if we are using the minimum distance for our comparisons then we need to develop another technique to find the upper limit for the distance.

3.1 Experiments With Multiple People

I made experiments with a test dataset from a different person with the gesture type 10. These are the average and minimum distances to each training gesture dataset.

	1	2	3	4	5	6	7	8	9	10	11	12
s.1	292	276	239	283	248	290	290	254	281	252	317	249
s.2	301	303	246	288	256	286	310	263	278	247	314	266
s.3	305	308	253	292	258	287	317	267	279	254	316	271
s.4	300	300	247	290	252	285	310	261	276	252	311	266
s.5	302	294	244	291	250	286	304	262	278	248	309	263
s.6	304	297	246	294	252	288	309	265	280	252	310	267
s.7	302	298	246	294	251	287	310	261	278	252	309	265
s.8	305	300	254	294	259	290	316	264	282	261	321	266
s.9	308	300	256	296	261	292	318	265	284	265	324	269
s.10	295	293	246	281	254	281	306	256	274	254	316	258

Table 6 - Average Distances to classes for gesture 10

	1	2	3	4	5	6	7	8	9	10	11	12
s.1	165	203	153	170	161	153	147	177	164	110	204	168
s.2	202	228	164	172	170	172	169	194	177	117	198	191
s.3	202	235	164	166	167	166	172	193	174	125	199	192
s.4	197	229	162	166	167	163	167	186	169	121	191	187
s.5	194	224	164	166	172	158	165	186	168	113	190	180
s.6	196	225	163	165	168	159	166	187	171	117	189	185
s.7	195	227	163	161	166	160	169	184	168	117	189	184
s.8	198	229	163	167	167	164	171	181	172	132	204	187
s.9	201	230	164	172	169	165	172	182	174	135	207	188
s.10	187	222	161	169	169	157	165	177	166	126	201	179

Table 7 - Minimum distances to classes for gesture 10

As can be seen from the figures, the borders now get more vague since the training and test data belong to different persons. The distances are expected to be more between gestures of different persons even though they belong to the same gesture type, because every person has a slightly different way of using gestures. However if we consider the minimum distance as the criterion, then the accuracy approaches to 1 in this experiment. Using the minimum distance or a voting mechanism between all the distances may be a better idea in this case to make a better prediction. A better and safer idea could be taking the averages of the n minimum distances to each class and averaging them, which would take into account the most similar cases in each class.

To decrease the variation between the same gestures across different people I made another experiment this time translating the initial location of all gestures to (0,0). This turned out to be a good idea since as you can see from the table below the accuracy increased significantly for the gesture 10 with this method. In this experiment I used the gesture samples from 4 people (people of code 19, 21,22,28) for all gesture types and tested with gestures from a different person(person with code 06).

	1	2	3	4	5	6	7	8	9	10	11	12
s.1	225	259	107	157	88	182	128	222	161	114	168	253
s.2	211	264	111	141	130	157	147	209	148	95	159	255
s.3	218	268	114	146	126	163	153	210	151	104	164	257
s.4	215	268	111	144	133	161	161	190	150	122	164	258
s.5	209	271	111	136	131	155	155	198	143	118	159	256
s.6	211	267	107	137	123	155	148	207	143	105	157	255
s.7	204	273	110	131	128	150	163	189	139	122	155	258
s.8	212	269	110	141	133	159	169	183	149	133	165	254
s.9	217	263	113	144	134	162	169	173	151	138	167	254
s.10	214	264	108	142	132	159	162	178	150	131	166	253

Table 8 - Average distances to classes for gesture 10 with translated data

But as you might have noticed this was a hard case. The samples from the gesture 10 of this person was hard to identify. The samples from other gesture types showed better results which you will see in the confusion matrix later.

The identification of the upper limit in the case with different people becomes more difficult because the inter class variation has increased and the intra-class variations are somewhat blended into each other. This is due to the fact that a gesture made by two different persons might be quite different, blurring the borders between classes. Here are the maximum and average interclass distances.

	1	2	3	4	5	6	7	8	9	10	11	12
max	370	381	400	471	403	414	478	329	448	368	357	393

	1	2	3	4	5	6	7	8	9	10	11	12
avg	175	205	185	199	210	213	213	188	202	183	190	186

4. Results

The following is the confusion matrix using the average distances to all classes and without using the upper limit values found in my last experiments. As can be seen there are some hard cases like 10 as I pointed out before. However, considering the experiments were done with gestures from several people the results look good in most cases.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	**8**	0	0	0	**2**	0	0	0	0	0	0	0
2	0	**9**	**1**	0	0	0	0	0	0	0	0	0
3	0	0	**10**	0	0	0	0	0	0	0	0	0
4	0	0	0	**9**	**1**	0	0	0	0	0	0	0
5	0	0	0	0	**10**	0	0	0	0	0	0	0
6	0	0	0	**4**	**1**	**5**	0	0	0	0	0	0
7	0	0	0	0	0	0	**10**	0	0	0	0	0
8	0	0	0	0	**1**	0	0	**9**	0	0	0	0
9	0	0	0	**3**	**1**	0	0	0	**6**	0	0	0
10	0	0	**6**	0	**1**	0	0	0	0	**3**	0	0
11	0	0	0	0	0	0	0	0	**7**	1	**2**	0
12	0	0	0	**1**	**1**	0	0	0	**1**	0	0	**7**

Table 9 - Confusion matrix

As can be seen from the confusion matrix the accuracy in some cases approach 0.9 while in other gesture types like 10, 11 they are very low. I did not include the results with the minimum distances but in this case the accuracy increases to nearly one without taking into account the upper limit numbers. The detailed average distance values for each gesture can be found at the end of this section.

I was unable to use the max values found for imposing an upper limit since they were too high belonging to outliers. For better results in this respect we could use methods like the n'th max value or the average of n max values or a point between the average and the max…

You can also find below the detailed distance findings for each gesture to each class below, which I based my findings above. The rows represent a sample in the test dataset from the given gesture type and the columns are the average distances to each class in the training set.

Gesture 1:

	1	2	3	4	5	6	7	8	9	10	11	12
s.1	224	281	137	167	105	197	150	233	170	150	162	266
s.2	93	381	207	137	244	158	223	256	129	219	144	302
s.3	85	390	215	127	257	147	232	256	123	224	152	306
s.4	81	367	193	116	231	138	207	246	112	200	138	292
s.5	85	383	203	127	248	149	223	256	123	215	147	303
s.6	85	368	195	120	231	141	209	247	112	202	141	293
s.7	91	393	214	136	262	157	237	263	134	228	155	310
s.8	84	379	202	125	244	146	219	254	120	211	147	301
s.9	86	379	200	124	243	147	220	253	122	212	145	300
s.10	224	281	137	167	105	197	150	233	170	150	162	266

Gesture 2:

	1	2	3	4	5	6	7	8	9	10	11	12
s.1	262	347	189	204	194	234	200	285	229	211	244	320
s.2	475	226	361	412	320	423	373	426	412	363	405	406
s.3	425	230	331	359	310	367	356	348	368	340	364	400
s.4	401	197	304	336	279	347	312	361	342	298	338	363
s.5	379	166	280	310	264	318	294	347	320	270	322	337
s.6	386	180	292	318	264	328	306	341	325	278	325	363
s.7	411	196	308	344	283	354	323	366	349	299	345	383
s.8	390	191	306	328	286	337	319	350	338	297	336	364
s.9	395	182	291	332	269	344	310	351	337	295	333	353
s.10	387	178	282	323	257	337	299	349	327	282	324	331

Gesture 3:

	1	2	3	4	5	6	7	8	9	10	11	12
s.1	212	258	84	143	96	170	103	220	152	111	160	245
s.2	215	289	72	168	146	187	135	243	178	147	187	270
s.3	211	276	63	160	132	181	122	235	169	132	178	262
s.4	215	275	81	160	128	182	125	234	170	139	174	260
s.5	228	283	72	182	142	202	124	259	190	143	194	273
s.6	219	282	73	175	141	196	122	255	182	143	190	272
s.7	213	276	63	161	128	182	121	234	169	129	177	260
s.8	210	270	58	158	125	180	118	229	165	125	173	260
s.9	212	273	60	164	130	186	117	236	170	132	178	263
s.10	212	258	84	143	96	170	103	220	152	111	160	245

Gesture 4:

	1	2	3	4	5	6	7	8	9	10	11	12
s.1	206	271	114	140	78	167	128	215	145	131	152	250
s.2	118	331	171	64	203	96	188	220	81	167	126	267
s.3	119	332	173	67	202	99	188	217	83	169	129	268
s.4	132	338	179	83	205	112	198	209	96	183	140	272
s.5	130	343	185	87	213	116	205	211	100	191	142	273
s.6	118	343	173	73	212	105	195	225	89	175	130	273
s.7	122	340	172	71	212	99	194	227	89	172	133	272
s.8	118	321	159	65	182	96	170	215	78	156	122	261
s.9	117	336	173	67	206	97	188	225	81	169	128	269
s.10	116	323	161	63	189	93	174	218	76	156	121	262

Gesture 5:

	1	2	3	4	5	6	7	8	9	10	11	12
s.1	299	254	177	235	76	261	172	270	233	185	211	292
s.2	266	249	152	203	63	230	153	246	200	155	184	276
s.3	296	259	182	236	82	262	176	271	233	186	210	296
s.4	302	259	188	240	79	265	179	275	238	190	214	298
s.5	304	255	194	241	79	266	180	276	239	191	214	296
s.6	278	265	170	219	78	246	167	260	215	171	197	284
s.7	307	261	193	245	78	270	181	279	243	195	219	298
s.8	311	253	198	244	78	269	183	276	244	196	220	297
s.9	317	253	203	250	79	275	189	278	251	200	224	301
s.10	302	252	187	237	73	263	175	271	237	188	212	294

Gesture 6:

	1	2	3	4	5	6	7	8	9	10	11	12
s.1	291	299	164	230	118	251	180	282	228	184	224	302
s.2	159	342	183	99	225	100	205	235	116	180	162	280
s.3	168	352	200	108	238	109	224	242	125	184	173	289
s.4	164	366	214	117	243	117	224	264	127	195	180	296
s.5	157	355	195	102	235	101	219	234	118	193	165	289
s.6	175	351	195	120	235	114	216	254	132	185	176	293
s.7	157	367	202	105	247	103	225	244	123	203	176	295
s.8	185	317	178	114	195	117	179	249	125	146	169	267
s.9	164	331	166	99	211	100	193	229	118	170	161	275
s.10	173	351	189	115	227	112	212	234	127	182	171	288

Gesture 7:

	1	2	3	4	5	6	7	8	9	10	11	12
s.1	228	272	117	163	116	189	95	241	170	126	181	257
s.2	237	289	137	178	131	203	98	259	183	144	193	269
s.3	242	290	143	183	137	207	105	263	187	149	198	273
s.4	236	292	134	179	136	203	93	261	184	145	195	268
s.5	247	296	143	189	139	214	100	271	194	149	205	277
s.6	252	303	153	195	151	218	108	274	203	157	214	265
s.7	236	292	141	177	138	202	96	261	182	145	193	265
s.8	249	304	154	194	145	217	101	274	196	154	205	276
s.9	235	298	143	177	142	200	94	259	181	148	195	267
s.10	254	306	162	200	151	222	106	279	203	157	212	275

Gesture 8:

	1	2	3	4	5	6	7	8	9	10	11	12
s.1	249	304	160	186	142	213	176	220	196	179	203	284
s.2	370	413	351	325	353	329	377	214	336	369	353	366
s.3	374	403	361	324	361	325	371	249	336	357	354	355
s.4	374	398	358	323	355	325	373	236	334	358	352	364
s.5	392	405	374	343	374	344	392	254	354	376	369	381
s.6	395	411	371	349	374	352	395	238	358	385	371	389
s.7	365	356	318	314	314	320	340	215	324	337	331	342
s.8	391	406	362	350	367	355	386	235	361	385	372	375
s.9	383	411	360	344	365	348	382	238	355	381	367	359
s.10	378	386	343	334	344	339	369	222	343	367	353	365

Gesture 9:

	1	2	3	4	5	6	7	8	9	10	11	12
s.1	296	273	176	231	91	258	183	267	226	180	210	299
s.2	145	356	191	115	217	139	207	254	127	181	163	280
s.3	126	337	170	93	190	124	187	210	96	182	138	273
s.4	136	331	168	100	180	130	182	213	105	181	144	270
s.5	109	339	180	88	198	116	185	231	79	174	127	275
s.6	107	344	183	93	202	120	190	230	81	178	128	279
s.7	109	360	198	99	222	126	207	237	88	193	136	283
s.8	132	342	181	107	194	135	194	215	107	190	149	280
s.9	130	330	174	103	179	132	181	220	98	176	140	275
s.10	113	338	178	94	193	122	185	229	82	174	129	276

Gesture 10:

	1	2	3	4	5	6	7	8	9	10	11	12
s.1	225	259	107	157	88	182	128	222	161	114	168	253
s.2	211	264	111	141	130	157	147	209	148	95	159	255
s.3	218	268	114	146	126	163	153	210	151	104	164	257
s.4	215	268	111	144	133	161	161	190	150	122	164	258
s.5	209	271	111	136	131	155	155	198	143	118	159	256
s.6	211	267	107	137	123	155	148	207	143	105	157	255
s.7	204	273	110	131	128	150	163	189	139	122	155	258
s.8	212	269	110	141	133	159	169	183	149	133	165	254
s.9	217	263	113	144	134	162	169	173	151	138	167	254
s.10	214	264	108	142	132	159	162	178	150	131	166	253

Gesture 11:

	1	2	3	4	5	6	7	8	9	10	11	12
s.1	345	309	232	285	136	310	228	320	278	224	255	355
s.2	157	342	190	142	188	170	204	247	134	193	143	272
s.3	147	318	164	124	160	154	179	238	115	163	118	280
s.4	156	316	166	135	163	165	177	233	130	176	130	272
s.5	171	317	171	141	163	171	194	236	133	175	137	267
s.6	157	307	162	131	152	161	176	232	122	160	122	282
s.7	158	318	165	137	160	168	181	243	131	166	131	282
s.8	144	308	157	121	150	152	171	229	113	160	114	269
s.9	152	317	172	126	158	158	186	232	121	168	123	269
s.10	141	314	165	118	159	148	178	226	110	162	114	269

Gesture 12:

	1	2	3	4	5	6	7	8	9	10	11	12
s.1	229	284	135	162	126	189	145	227	172	151	180	264
s.2	340	386	331	300	335	304	335	298	312	335	335	245
s.3	384	440	382	353	392	357	386	359	364	385	385	261
s.4	398	447	398	367	410	368	407	370	378	398	396	290
s.5	472	470	450	437	460	440	459	409	449	455	459	343
s.6	343	365	303	298	298	306	314	282	303	314	317	236
s.7	356	367	331	311	332	315	336	317	318	331	342	243
s.8	308	370	313	270	324	273	321	287	283	311	309	275
s.9	372	413	362	341	376	345	368	342	351	371	371	278
s.10	141	314	165	118	159	148	178	226	110	162	114	269

5. Conclusion

From the results found we can see that DTW algorithm works pretty well, even across samples from different people if the preprocessing for the gestures are done correctly. For the best results, I did an alignment between all gestures to a common size which is 150 * 80. This alignment may be optimized by choosing another heuristic such as taking the median length of all gestures. I also translated all the gestures such that the initial values are all the same, which is (0,0). This decreased the variations among the same gestures and made the inter and intra class distance difference starker, which increased the confidence. I also took a maximum distance for each class using the training set, comparing the distance between each pairs of the same gesture. This maximum found was useful with the same-person gestures but when the gestures came from different persons the borders between classes blurred and the max did not work. To find an upper limit in this case, better heuristics should be used such as taking out the outliers, taking the n'th maximim, taking the average of n max values or taking a number between average and the max. In my testing to find the correct class I used two methods, one comparing the average distances to each class and the other comparing the minimum distances. I chose the smallest distance class. With the average distance method the accuracy approached 0.9 in most cases. With the minimum distance comparison, the accuracy increased to nearly 1 but in this case finding an upper limit got harder and taking the maximum of the distances of each pair in the training set for this upper limit did not work. So with this method we risk a high number of false positives in case the given sample does not belong to any of the gestures, unless a better heuristic to find an upper limit for the distance is found.

6. How to run the code

- Include the data files under './MicrosoftGestureDataset-RC/data'
- Use datapreparation.m file to prepare the training data set. It currently uses the data from the people with code (19, 21, 22, 28). If you want to limit the training dataset to fewer people and increase the speed then change these values. This file takes the gestures and aligns them and stores them in the aligned.mat file to be used later.
- Using the aligned.mat file prepared before, use the training.m file to find the max and average values between same gesture pairs.
- Using the aligned.mat file prepared, use the testing.m file to find the average and minimum distances to the training gesture classes. This file currently uses the testing data from the person 06. For each gesture analysis the name of the data file in the code should be changed.

References:

. Dynamic time Warping, https://en.wikipedia.org/wiki/Dynamic_time_warping