Recognition Using Face And Body

Project group 9
Bartosz Biesaga

30 I 2024

1 Existing biometric authentication using face and body features

Multimodal biometric authentication systems (systems that use multiple biometric features to identify individual) are used in place of unimodal systems in aim to improve system's security and may be used in aim to improve performance, when one of the features is handicapped. Some of the used combinations are face and signature, voice and face, iris and fingerprint. Face is a very popular choice when it comes to recognition, but what about body? The only way we found that person's body was used for recognition was their gait (their way of walking).

2 How we combined face and silhouette features for employee recognition

We decided to try a different approach and use features extracted from person's pose, more precisely we created a feature vector containing distances between nearby joints (wrist-elbow, elbow-shoulder, shoulder-neck, shoulder-hip, neck-head, hip-knee, knee-ankle, hip-hip, shoulder-shoulder) divided by mean distance between shoulder and corresponding hip as a way to standardize features no matter how far a person is on a photo.

To combine such vector with features extracted from face we initially wanted to calculate weighted sum of respective distances, but eventually decided that for a new image we will look for best face and pose match from employee database which cosine distances are smaller than chosen thresholds (0.7 for face and 0.04 for silhouette). If there is such face in database then we recognize given person as best face match from database. Otherwise if there is such pose in database then we recognize given person as best pose match from database. If neither face nor pose matches then we recognize a person as non-employee. In this approach we favor face embeddings over pose embeddings since they are way more characteristic and unique, but also utilize pose embeddings with rigorous threshold to help employee recognition in bad conditions when face recognition fails.

Initially we tried L2 distance, but sometimes it gave smaller distance for images of two different persons than for different images of same person, which wasn't the case for

cosine distance. Also since cosine distance yields way smaller values it was more intuitive to work with and choose thresholds for.

3 Testing our method

To test our approach we decided to create a small database of 10 persons – 5 man and 5 woman – for each we extracted face and pose features from 2 images. Then we created a test sample of 24 images – 2 for each person from database, but one with blurred face to test recognition using only pose features in combined approach and 4 images of people that are not registered in database. We present our results in tables below. First row contains labels of persons identity and last column tells us how someone was recognized (face/pose) and label of as who they were recognized. Leftmost column contains labels of people from test set, underscore at the start means that they don't appear in database and ,,_blur" at the end means that only their pose was used for recognition. Remaining columns contain information about distances between embeddings extracted from test image and embeddings saved in database divided by chosen threshold, each ratio separated with semicolon. If there are 4 values in column, 2 face embedding distance ratios come first and then 2 for pose. With such ratios, value above 1 means that distance was greater than threshold and below 1 means that it was smaller.

3.1 Using only face for recognition

Table 1: Results for only face recognition.

	z	**nr**	**lc**	**s**	**az**	**ad**	**cm**	**mm**	**al**	**r**	**recognized**
_rd	1.31;1.35;	1.15;1.24;	1.58;1.42;	1.21;1.30;	1.02;1.13;	1.61;1.52;	1.61;1.53;	1.30;1.32;	1.28;1.39;	1.28;1.43;	not in database
lc	1.04;1.18;	1.11;1.24;	1.00;0.87;	1.51;1.21;	1.34;1.46;	1.53;1.34;	1.40;1.49;	1.18;1.07;	1.44;1.36;	1.27;1.30;	face: lc
al	1.49;1.34;	1.17;1.19;	1.34;1.44;	1.41;1.37;	1.11;1.14;	1.30;1.42;	1.33;1.38;	1.53;1.31;	0.76;1.14;	1.44;1.42;	face: al
s	1.35;1.36;	1.40;1.31;	1.34;1.34;	0.37;0.78;	1.25;1.24;	1.47;1.27;	1.23;1.32;	1.40;1.37;	1.35;1.50;	1.30;1.30;	face: s
_eo	1.16;1.16;	1.28;1.31;	1.08;1.05;	1.02;1.15;	1.27;1.32;	1.37;1.24;	1.39;1.33;	1.07;1.11;	1.31;1.43;	0.94;1.17;	face: r
r	1.00;1.30;	1.42;1.22;	1.43;1.35;	1.32;1.08;	1.37;1.35;	1.32;1.25;	1.38;1.39;	1.20;1.05;	1.35;1.40;	0.97;1.26;	face: r
_bl	1.32;1.16;	1.58;1.56;	1.40;1.50;	1.53;1.47;	1.32;1.42;	1.31;1.35;	1.23;1.30;	1.32;1.17;	1.43;1.47;	1.53;1.26;	not in database
_bc	1.43;1.53;	1.35;1.41;	1.35;1.42;	1.15;1.35;	1.35;1.40;	1.47;1.39;	1.02;1.18;	1.46;1.50;	1.41;1.46;	1.49;1.47;	not in database
cm	1.20;1.38;	1.25;1.32;	1.25;1.28;	1.17;1.39;	1.42;1.45;	1.21;1.28;	0.42;0.34;	1.36;1.24;	1.28;1.25;	1.47;1.36;	face: cm
z	1.25;0.82;	1.30;1.27;	1.24;1.26;	1.48;1.52;	1.32;1.38;	1.44;1.49;	1.32;1.33;	1.32;1.28;	1.44;1.36;	1.33;1.29;	face: z
ad	1.47;1.31;	1.48;1.41;	1.39;1.32;	1.32;1.40;	1.26;1.34;	1.00;1.07;	1.26;1.32;	1.54;1.36;	1.21;1.44;	1.53;1.48;	not in database
mm	1.36;1.17;	1.12;1.23;	1.52;1.30;	1.36;1.36;	1.23;1.33;	1.27;1.22;	1.36;1.43;	0.76;0.84;	1.40;1.55;	1.32;1.46;	face: mm
az	1.25;1.17;	1.22;1.26;	1.50;1.46;	1.28;1.35;	0.58;0.69;	1.18;1.34;	1.36;1.23;	1.23;1.32;	1.23;1.38;	1.36;1.32;	face: az
nr	1.34;1.10;	0.41;0.57;	1.38;1.33;	1.23;1.38;	1.04;1.11;	1.48;1.40;	1.21;1.33;	1.25;1.21;	1.40;1.40;	1.28;1.21;	face: nr

When we used only face for recognition and threshold equal to 0.7, we got quite good results with only one person from database barely not getting recognized (their best ratio was slightly above 1.0) and one person outside of database getting recognized as someone from database.

3.2 Using only pose for recognition

Table 2: Results for only pose recognition.

	z	nr	lc	s	az	ad	cm	mm	al	r	recognized
_rd	0.09;0.13;	0.19;0.24;	0.07;0.06;	0.07;0.07;	0.18;0.12;	0.15;0.07;	0.07;0.13;	0.13;0.08;	0.13;0.28;	0.15;0.10;	pose: lc
lc	0.10;0.13;	0.24;0.30;	0.08;0.12;	0.18;0.10;	0.26;0.18;	0.11;0.12;	0.19;0.25;	0.20;0.19;	0.20;0.23;	0.08;0.09;	pose: lc
al	0.16;0.23;	0.12;0.19;	0.14;0.17;	0.12;0.09;	0.13;0.11;	0.21;0.06;	0.13;0.21;	0.27;0.17;	0.11;0.28;	0.24;0.18;	pose: ad
s	0.25;0.32;	0.12;0.13;	0.25;0.17;	0.02;0.20;	0.10;0.08;	0.22;0.08;	0.24;0.34;	0.42;0.11;	0.09;0.15;	0.39;0.27;	pose: s
_eo	0.13;0.18;	0.17;0.19;	0.12;0.08;	0.02;0.09;	0.16;0.10;	0.16;0.07;	0.13;0.19;	0.22;0.05;	0.11;0.22;	0.24;0.15;	pose: s
r	0.15;0.21;	0.26;0.32;	0.16;0.19;	0.13;0.17;	0.24;0.15;	0.25;0.09;	0.20;0.24;	0.34;0.28;	0.26;0.37;	0.27;0.22;	pose: ad
_bl	0.19;0.23;	0.40;0.43;	0.08;0.08;	0.26;0.21;	0.42;0.33;	0.25;0.26;	0.07;0.09;	0.06;0.17;	0.30;0.43;	0.09;0.03;	pose: r
_bc	0.28;0.37;	0.04;0.06;	0.22;0.16;	0.08;0.21;	0.05;0.08;	0.22;0.08;	0.19;0.30;	0.37;0.10;	0.02;0.11;	0.32;0.21;	pose: al
cm	0.17;0.21;	0.22;0.29;	0.07;0.10;	0.12;0.15;	0.21;0.17;	0.28;0.10;	0.05;0.08;	0.15;0.15;	0.19;0.39;	0.16;0.12;	pose: cm
z	0.10;0.08;	0.61;0.68;	0.12;0.16;	0.25;0.17;	0.58;0.38;	0.29;0.29;	0.18;0.15;	0.11;0.30;	0.51;0.67;	0.18;0.17;	pose: z
ad	0.17;0.25;	0.08;0.11;	0.19;0.16;	0.12;0.10;	0.10;0.08;	0.14;0.07;	0.18;0.29;	0.29;0.12;	0.05;0.20;	0.28;0.19;	pose: al
mm	0.18;0.21;	0.29;0.37;	0.05;0.12;	0.22;0.17;	0.30;0.25;	0.27;0.17;	0.11;0.12;	0.14;0.21;	0.24;0.37;	0.05;0.07;	pose: lc
az	0.29;0.38;	0.02;0.04;	0.32;0.28;	0.16;0.20;	0.03;0.06;	0.21;0.07;	0.31;0.46;	0.52;0.23;	0.04;0.17;	0.42;0.34;	pose: nr
nr	0.22;0.29;	0.11;0.16;	0.11;0.10;	0.09;0.15;	0.12;0.14;	0.25;0.08;	0.09;0.15;	0.21;0.06;	0.06;0.20;	0.20;0.13;	pose: mm

When we used just pose features results got catastrophically worse. Not only every single test subject got recognized as someone from database, which gives us 4 wrong results already, but also 6 samples of people that are in database got recognized as someone else from it, which leaves us with just 5 correct recognitions. We can easily conclude that recognition using only pose features is neither safe nor accurate.

3.3 Combined approach

Since combined approach really is just a two stage recognition where we first try to recognize using face and when we don't find good enough face match then we try to recognize using pose, we end up with always recognizing person from given image as someone from database. It happens, because pose features are not very distinctive between different individuals. So in this approach we have some alright stuff since we use face embeddings first and when we find good enough match they work quite well just like it was shown in section 3.1, but we also get a lot of bad stuff from recognition using only pose, which introduces a lot of error when it comes to people from outside of database and those from database but without face available.

Table 3: Results for combined approach.

	z	nr	lc	s	az	ad	cm	mm	al	r	recognized
_rd	1.31;1.35;0.09;0.13;	1.15;1.24;0.19;0.24;	1.58;1.42;0.07;0.06;	1.21;1.30;0.07;0.07;	1.02;1.13;0.18;0.12;	1.61;1.52;0.15;0.07;	1.61;1.53;0.07;0.13;	1.30;1.32;0.13;0.08;	1.28;1.39;0.13;0.28;	1.28;1.43;0.15;0.10;	pose: lc
lc	1.04;1.18;0.10;0.13;	1.11;1.24;0.24;0.30;	1.00;0.87;0.08;0.12;	1.51;1.21;0.18;0.10;	1.34;1.46;0.26;0.18;	1.53;1.34;0.11;0.12;	1.40;1.49;0.19;0.25;	1.18;1.07;0.20;0.19;	1.44;1.36;0.20;0.23;	1.27;1.30;0.08;0.09;	face: lc
al	1.49;1.34;0.16;0.23;	1.17;1.19;0.12;0.19;	1.34;1.44;0.14;0.17;	1.41;1.37;0.12;0.09;	1.11;1.14;0.13;0.11;	1.30;1.42;0.21;0.06;	1.33;1.38;0.13;0.21;	1.53;1.31;0.27;0.17;	0.76;1.14;0.11;0.28;	1.44;1.42;0.24;0.18;	face: al
s	1.35;1.36;0.25;0.32;	1.40;1.31;0.12;0.13;	1.34;1.34;0.25;0.17;	0.37;0.78;0.02;0.20;	1.25;1.24;0.10;0.08;	1.47;1.27;0.22;0.08;	1.23;1.32;0.24;0.34;	1.40;1.37;0.42;0.11;	1.35;1.50;0.09;0.15;	1.30;1.30;0.39;0.27;	face: s
z_blur	2.00;2.00;0.11;0.10;	2.00;2.00;0.63;0.70;	2.00;2.00;0.13;0.17;	2.00;2.00;0.27;0.18;	2.00;2.00;0.60;0.40;	2.00;2.00;0.32;0.31;	2.00;2.00;0.18;0.14;	2.00;2.00;0.11;0.32;	2.00;2.00;0.53;0.71;	2.00;2.00;0.19;0.18;	pose: z
_eo	1.16;1.16;0.13;0.18;	1.28;1.31;0.17;0.19;	1.08;1.05;0.12;0.08;	1.02;1.15;0.02;0.09;	1.27;1.32;0.16;0.10;	1.37;1.24;0.16;0.07;	1.39;1.33;0.13;0.19;	1.07;1.11;0.22;0.05;	1.31;1.43;0.11;0.22;	0.94;1.17;0.24;0.15;	face: r
r	1.00;1.30;0.15;0.21;	1.42;1.22;0.26;0.32;	1.43;1.35;0.16;0.19;	1.32;1.08;0.13;0.17;	1.37;1.35;0.24;0.15;	1.32;1.25;0.25;0.09;	1.38;1.39;0.20;0.24;	1.20;1.05;0.34;0.28;	1.35;1.40;0.26;0.37;	0.97;1.26;0.27;0.22;	face: r
mm_blur	2.00;2.00;0.18;0.21;	2.00;2.00;0.29;0.37;	2.00;2.00;0.05;0.12;	2.00;2.00;0.23;0.18;	2.00;2.00;0.30;0.26;	2.00;2.00;0.27;0.18;	2.00;2.00;0.10;0.12;	2.00;2.00;0.14;0.21;	2.00;2.00;0.24;0.37;	2.00;2.00;0.05;0.07;	pose: lc
r_blur	2.00;2.00;0.16;0.21;	2.00;2.00;0.25;0.33;	2.00;2.00;0.16;0.19;	2.00;2.00;0.13;0.17;	2.00;2.00;0.24;0.15;	2.00;2.00;0.26;0.09;	2.00;2.00;0.20;0.24;	2.00;2.00;0.34;0.28;	2.00;2.00;0.26;0.37;	2.00;2.00;0.27;0.22;	pose: ad
az_blur	2.00;2.00;0.27;0.37;	2.00;2.00;0.02;0.04;	2.00;2.00;0.30;0.27;	2.00;2.00;0.16;0.19;	2.00;2.00;0.03;0.06;	2.00;2.00;0.20;0.06;	2.00;2.00;0.30;0.45;	2.00;2.00;0.50;0.23;	2.00;2.00;0.03;0.17;	2.00;2.00;0.41;0.32;	pose: nr
nr_blur	2.00;2.00;0.21;0.27;	2.00;2.00;0.15;0.21;	2.00;2.00;0.09;0.08;	2.00;2.00;0.08;0.15;	2.00;2.00;0.16;0.16;	2.00;2.00;0.26;0.09;	2.00;2.00;0.07;0.11;	2.00;2.00;0.17;0.06;	2.00;2.00;0.10;0.24;	2.00;2.00;0.18;0.11;	pose: mm
ad_blur	2.00;2.00;0.15;0.22;	2.00;2.00;0.12;0.15;	2.00;2.00;0.17;0.15;	2.00;2.00;0.12;0.08;	2.00;2.00;0.12;0.10;	2.00;2.00;0.15;0.07;	2.00;2.00;0.16;0.26;	2.00;2.00;0.24;0.12;	2.00;2.00;0.07;0.24;	2.00;2.00;0.26;0.18;	pose: ad
_bl	1.32;1.16;0.19;0.23;	1.58;1.56;0.40;0.43;	1.40;1.50;0.08;0.08;	1.53;1.47;0.26;0.21;	1.32;1.42;0.42;0.33;	1.31;1.35;0.25;0.26;	1.23;1.30;0.07;0.09;	1.32;1.17;0.06;0.17;	1.43;1.47;0.30;0.43;	1.53;1.26;0.09;0.03;	pose: r
_bc	1.43;1.53;0.28;0.37;	1.35;1.41;0.04;0.06;	1.35;1.42;0.22;0.16;	1.15;1.35;0.08;0.21;	1.35;1.40;0.05;0.08;	1.47;1.39;0.22;0.08;	1.02;1.18;0.19;0.30;	1.46;1.50;0.37;0.10;	1.41;1.46;0.02;0.11;	1.49;1.47;0.32;0.21;	pose: al
cm_blur	2.00;2.00;0.17;0.21;	2.00;2.00;0.24;0.31;	2.00;2.00;0.07;0.10;	2.00;2.00;0.13;0.16;	2.00;2.00;0.23;0.18;	2.00;2.00;0.29;0.11;	2.00;2.00;0.05;0.07;	2.00;2.00;0.15;0.16;	2.00;2.00;0.20;0.41;	2.00;2.00;0.16;0.12;	pose: cm
cm	1.20;1.38;0.17;0.21;	1.25;1.32;0.22;0.29;	1.25;1.28;0.07;0.10;	1.17;1.39;0.12;0.15;	1.42;1.45;0.21;0.17;	1.21;1.28;0.28;0.10;	0.42;0.34;0.05;0.08;	1.36;1.24;0.15;0.15;	1.28;1.25;0.19;0.39;	1.47;1.36;0.16;0.12;	face: cm
z	1.25;0.82;0.10;0.08;	1.30;1.27;0.61;0.68;	1.24;1.26;0.12;0.16;	1.48;1.52;0.25;0.17;	1.32;1.38;0.58;0.38;	1.44;1.49;0.29;0.29;	1.32;1.33;0.18;0.15;	1.32;1.28;0.11;0.30;	1.44;1.36;0.51;0.67;	1.33;1.29;0.18;0.17;	face: z
lc_blur	2.00;2.00;0.10;0.11;	2.00;2.00;0.28;0.33;	2.00;2.00;0.07;0.12;	2.00;2.00;0.19;0.10;	2.00;2.00;0.29;0.20;	2.00;2.00;0.13;0.14;	2.00;2.00;0.18;0.24;	2.00;2.00;0.18;0.19;	2.00;2.00;0.22;0.27;	2.00;2.00;0.07;0.08;	pose: r
s_blur	2.00;2.00;0.25;0.32;	2.00;2.00;0.12;0.13;	2.00;2.00;0.24;0.17;	2.00;2.00;0.02;0.20;	2.00;2.00;0.10;0.08;	2.00;2.00;0.22;0.09;	2.00;2.00;0.24;0.33;	2.00;2.00;0.41;0.11;	2.00;2.00;0.09;0.15;	2.00;2.00;0.38;0.26;	pose: s
ad	1.47;1.31;0.17;0.25;	1.48;1.41;0.08;0.11;	1.39;1.32;0.19;0.16;	1.32;1.40;0.12;0.10;	1.26;1.34;0.10;0.08;	1.00;1.07;0.14;0.07;	1.26;1.32;0.18;0.29;	1.54;1.36;0.29;0.12;	1.21;1.44;0.05;0.20;	1.53;1.48;0.28;0.19;	pose: al
al_blur	2.00;2.00;0.16;0.24;	2.00;2.00;0.12;0.20;	2.00;2.00;0.15;0.19;	2.00;2.00;0.13;0.10;	2.00;2.00;0.13;0.10;	2.00;2.00;0.22;0.06;	2.00;2.00;0.14;0.22;	2.00;2.00;0.29;0.19;	2.00;2.00;0.12;0.29;	2.00;2.00;0.25;0.20;	pose: ad
mm	1.36;1.17;0.18;0.21;	1.12;1.23;0.29;0.37;	1.52;1.30;0.05;0.12;	1.36;1.36;0.22;0.17;	1.23;1.33;0.30;0.25;	1.27;1.22;0.27;0.17;	1.36;1.43;0.11;0.12;	0.76;0.84;0.14;0.21;	1.40;1.55;0.24;0.37;	1.32;1.46;0.05;0.07;	face: mm
az	1.25;1.17;0.29;0.38;	1.22;1.26;0.02;0.04;	1.50;1.46;0.32;0.28;	1.28;1.35;0.16;0.20;	0.58;0.69;0.03;0.06;	1.18;1.34;0.21;0.07;	1.36;1.23;0.31;0.46;	1.23;1.32;0.52;0.23;	1.23;1.38;0.04;0.17;	1.36;1.32;0.42;0.34;	face: az
nr	1.34;1.10;0.22;0.29;	0.41;0.57;0.11;0.16;	1.38;1.33;0.11;0.10;	1.23;1.38;0.09;0.15;	1.04;1.11;0.12;0.14;	1.48;1.40;0.25;0.08;	1.21;1.33;0.09;0.15;	1.25;1.21;0.21;0.06;	1.40;1.40;0.06;0.20;	1.28;1.21;0.20;0.13;	face: nr

4 Conclusions

When it comes to multimodal biometric recognition, it looks like using person's pose in proposed way causes more harm than good. Another approach that we have considered – calculating weighted sum of distances between embeddings of face and pose features – could turn out to be less harmful, but since used pose features are neither unique nor characteristic they would probably always negatively impact system's performance. Modifying how exactly they are used doesn't seem to have any advantage over changing biometric feature to something more distinctive between individuals.