Work Follow-up

Vivien

August 18, 2017

1 Features Engineering

1.1 Face Parameterization

From the landmarks, we extract a new parameterization that make sense, it keep the following quantity:

- head pose: yaw, roll, pitch, x, y, depth
- muscular action: raising eyebrows, pushing eyebrows aside, opening eyes, mouth width, mouth opening, lips protrusion, mouth smile

The quantity suffer of some instability mainly the difficulty to normalize the face regarding rotational head pose that can corrupt some features. In the past, we tried to invert a slightly richer parameterization to look at the information lost.

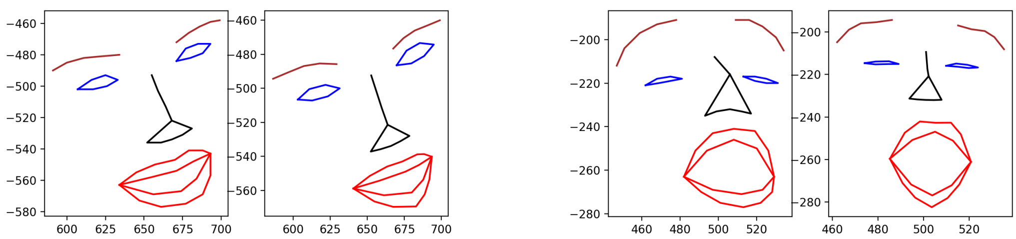


Figure 1: Comparisons between original (left) and reconstructed face (right)

1.2 Signal Processing

The signal extracted are high dimensional, noisy with frame missing. We would like to extract a simpler understanding of a signal as:

$$f(t) = H_f(t).\left(\sum_i a_i \phi_i(t) + \varepsilon(t)\right)$$

where ϕ represent simple things happening on the signal, ε model the noise, and $H_f \in \{0,1\}^{[0,T]}$ is modeling the missing frames.

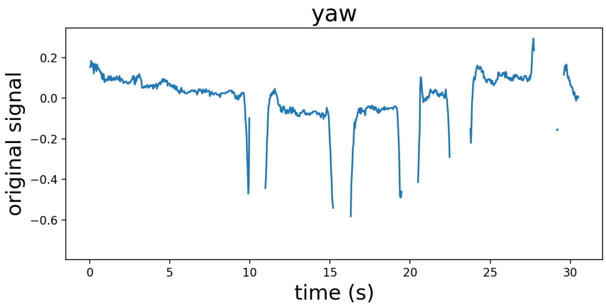


Figure 2: Example of signal

After trying decomposition with wavelets, or richer dictionary, the best information seems to be obtained by finding maximum in a scale space, compute after convolution against some Gaussian.

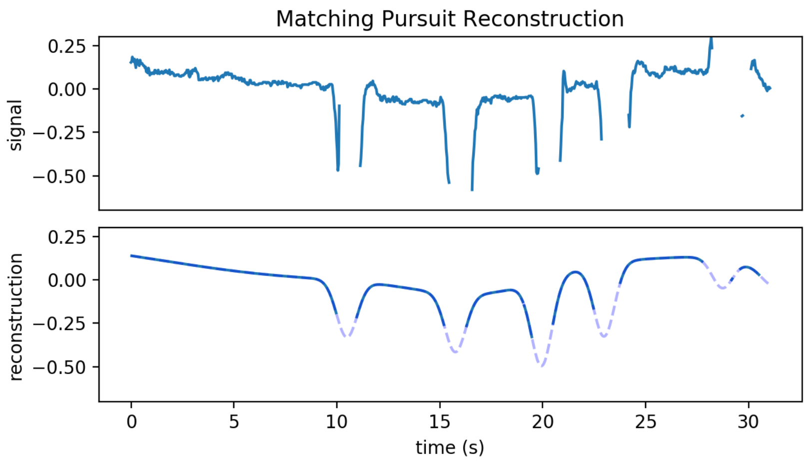


Figure 3: Reconstruction with a multi-scaled pursuit

1.3 First Layer

To describe the evolution of a feature in time, one can look at increment, which can be done by filtering with Gaussian derivative filter. From previous work, we decide to introduce a first layer on the raw face parameterization, consisting in the response at different scale against two filter: a Gaussian one to recognize event, and a Gaussian derivative to achieve higher order information such as speed. This two filters could be put in relation with wavelet construction based on an averaging and a detailed filter.

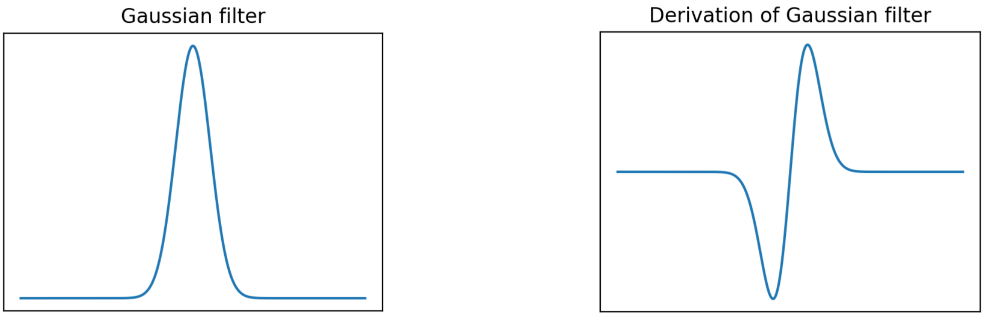


Figure 4: Filters of a first convolution layer

1.4 Dynamic Characterization & Second layer

One can see our data as stochastic process that we have to classify, to describe those one can look at moment such as variance, or density estimator such as histogram.

To achieve local information we would like to achieve local histogram or local variance. For the variance, we take a averaging filter w and look at

$$var = x^2 * w - (x * w)^2$$

we add a pooling layer after to only kept this averaging centered on few events. For the histogram, it look the same, but considering:

$$histo = \left(\mathbf{1}_{a_i < x < a_{i+1}} * w\right)_i$$

So we end up with a new representation of the type $T_2.W_2.T_1.W_1x$, where W_1 correspond to the first layer, T_1 to the non-linearity $x * 2$ or $\mathbf{1}_{a<x}$, W_2 to local averaging and T_2 pooling. The good part of this representation is that we clearly understand what all the coordinates correspond to.

2 Understandable Learning

In the past, we have been looking at blind machine learning. From a psychologist perspective, the diagnosis should be easily understandable from few full of sense features. Thus, let's try to reduce the number features on which to based our classifier.

A lot of classifier remains on the Euclidean structure (linear classifier, ℓ^2 nearest neighbors), if so one doesn't have to describe the all features space but only the one spanned by the data. Thus we design a simple way to reduce the data dimension. It mainly consists in removing features until achieving linear independence. Yet, one should assure that the features that remain aren't random. For this purpose, we first order the features regarding how much they discriminate autistic people from normal kid, before applying a Gram-Schmidt process where we also discard features that are too much correlated.

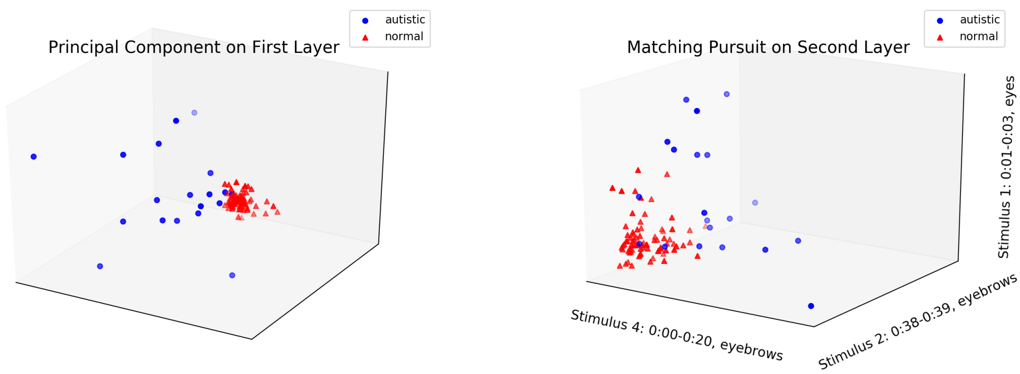


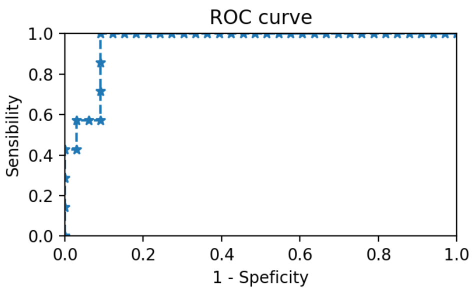
Figure 5: Points after the feature reduction step

3 Further ideas

There is a lot of other ideas to develop. On the one hand, the statistical learning part is still under construction. On the other hand, one could rethink the features engineering part to go further into pattern recognition framework. Let's review those ideas.

3.1 Statistical Guarantee or Psychological Warrant

Because of the small size of the database one cannot try too much method without worrying about ending with overfitting. Thus statistics can't really give a warrant to our findings. Yet, if the size of the dataset is not suited for machine learning, it is enough to test statistical alternative, or explore ideas expressed by psychologists, that could after our findings give a warrant on those.



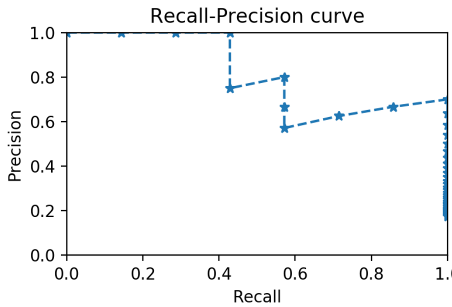


Figure 6: Results beating the MCHAT (rbf Svm on joint-histogram). Yet not stable regarding training/testing splitting. Careful reader will notice that 7 autistics people were used for validation.

3.2 Modeling normal behaviors

Autism could be seen as presenting weird behaviors. Not all the autistic are supposed to present the same behavior at the same moment. Yet one could take few moments where the autistic class present several subject outside of a normal range, or a normal behaviors model (learnt or not), and try to build on top a aggregated classifier.

3.3 Pattern Recognition

How to combine facial features locally to recognize or cluster some feelings expression.

This could be done through dictionary learning and bag of word description, or filtering and CNN learning. Note that databases can be find online to learn how to recognize facial expression on a single frame. This idea could be seen, as, rather than augmenting the dataset size, augment the supervision, through a labelling of feelings, or reduce the data description complexity.