1. Introduction

In many behavioral domains, such as facial expression and gesture, sparse structure is prevalent. This sparsity would be well suited for event detection but for one problem. Features typically are confounded by alignment error in space and time. As a consequence, high-dimensional representations such as SIFT and Gabor features have been favored despite their much greater computational cost and potential loss of information. We propose a Kernel Structured Sparsity (KSS) method that can handle both the temporal alignment problem and the structured sparse reconstruction within a common framework, and it can rely on simple features.

2. Face Alignment

We used 2face (www.2face.org), which is a generic 3D face tracker that requires no individual training to track facial landmarks of persons is has never seen before. It locates 3D coordinates of a dense set of facial landmarks. The 3D point distribution model (PDM) describes non-rigid shape variations linearly and composes it with a global rigid transformation, placing the shape in the image frame

\[ x_i = x_i(p) = sR(x_i + \Phi(p)) + t \quad (i = 1, \ldots, M), \]

where \(x_i\) denotes the 3D location of the \(i\)-th landmark and \(p = [x_i, y_i, z_i]\) denotes the parameters of the model, which consist of a global scaling \(s\), angles of rotation in three dimensions, a translation \(t\) and non-rigid transformation \(\Phi\).

3. Time-series Building

We tracked the video sequences with the 2Face tracker and built the time-series from the PCA coefficients of the 3D PDM (parameter \(q\)). Illustrative, this is the compressed representation of the 3D landmark locations without rigid head movements.

4. Global Alignment Kernel

To quantify the similarity of time-series (that form the input of the classifier) we use make use of kernels. Let \(|a|\) denote the length of alignment \(c\). The cost can be defined by means of a local divergence \(\alpha\) that measures the discrepancy between any two points \(\alpha\) and \(\gamma\) of vectors \(a\) and \(b\).

The Global Alignment (GA) kernel assumes that the minimum value of alignments may be sensitive to peculiarities of the time series and intends to take advantage of all alignments weighted exponentially. It is defined as the sum of exponentials and sign changed costs of the individual alignments.

\[ k_{GA}(a, b) = \sum_{c=1}^{|a|} \sum_{d=1}^{|b|} e^{-\alpha (c+1)/2d} \]

Fig. 4. Gram matrices induced by the GA kernel with different parameters.

5. Kernel Structured Sparsity

Using this form, a FISTA (fast iterative shrinkage-thresholding algorithm) optimization can be adapted to the solution. Our experiments were based on the modification of the SLEP package.

We need the following elements for the implementation:
1. The proximal operator of \(f_J\) (it has not changed)
2. \(f_J\) from the cost function
3. The gradient of \(f\)
4. The stopping criterion for FISTA (see supplement).

6. Datasets

Cohn Kanade Extended

Group Formation Task (action units)

6DMG Air-handwriting (gesture)

Acknowledgments: Research reported in this publication was supported in part by the National Institute of Mental Health of the National Institutes of Health under Award Number MK296952 to the Research and Technology Innovation Fund (grant agreement EITIKC_12-3-20140003), the EIT ICT Labs, and the Gatsby Charitable Foundation.