Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition

Yuan An

In the last article, spatio-temporal is introduced [5]. Today, dimensionality reduction is talked about in brief.

Dimensionality Reduction

Dimensionality reduction can be used to address several affect recognition challenges such as illumination variation, registration errors and identity bias. Components that reduce dimensionality may operate across multiple layers, such as early preprocessing (*e.g.* downsampling input image, applying masks) and intrinsic representation layers. In this section, the dimensionality reduction techniques are grouped into three classes, namely pooling, feature selection and feature extraction methods.

Pooling

Pooling, an example defined specifically for appearance representations, reduces dimensionality over local blocks of the representation by describing the features within the block jointly. This description discards the location of adjacent feature and thereby increase the tolerance against registration errors. Such function of pooling have a biological motivation as they mimic parts of mammals' vision systems [3].

Pooling is usually applied on multiple small neighborhoods across the image. There exists a variety of pooling techniques, such as binning features over local histograms, sampling the minimum or maximum value within a neighborhood or computing the sum or average of the features across the neighborhood. Sensitivity to illumination variations is generally addressed by normalizing the output of pooling (*e.g.* subtracting the local mean, or performing unit-norm normalization).

Pooling is usually considered as an intrinsic layer of the representation [4]. Representation such as the low-level histogram representations are defined to be dependent exclusively on a certain type of pooling.

Feature Selection

Feature selection aims at refining the facial representation by selecting a subset of its features, and optionally weighting the selected features. This process may be designed to have a semantic interpretation, such as discovering spatial or spatio-temporal regions of interest. Such applications of feature selection may reduce identity bias, as they are expected to discover the regions that are informative in terms of expressions rather than identity. Alternatively, the feature selection process may be designed to reduce dimensionality in a rather straightforward manner, without emphasis on the physical correspondence of the selected feature.

Feature selection can be performed with a range of techniques. A simple form is selecting and weighting certain spatial regions manually [6]. Most systems rely on data-driven feature selection and the most popular paradigm is boosting. Boosting refers to a set of generic techniques, which are designed for prediction (classification/regression) [2]. Many affect recognizer neglect the prediction role of boosting techniques and use them only for feature selection.

Feature Extraction

Feature extraction methods extract novel features (*e.g.* holistic features) from the initial representations. They map an input representation onto a lower dimensional space to discover a latent structure from the representation. This transformation can be non-adaptive or adaptive (learnt from training data).

The most popular non-adaptive transformation is the discrete cosine transformation (DCT) whereas the most popular adaptive transformation is PCA. PCA computes a linear transformation that aims at extrating decorrelated features out of possibly correlated feature. Under controlled headpose and imaging conditions, these features capture the statistical structure of expressions efficiently [1]. PCA is used by many systems including the winner of the AVEC continuous challenge.

References

[1] A. J. Calder, A. Burton, P. Miller, A. W. Young, and S. Akamatsu. A principal component analysis of facial expressions. *Vision Research*, 41(9):1179–1208, 2001. 1

[2] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. *The Annals of Statistics*, 28(2):337–407, 2000. 1

[3] H. D. H. and W. T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1):106–154, 1962. 1

[4] Y. LeCun, K. Kavukcuoglu, and C. Farabet. Convolutional networks and applications in vision. In *IEEE International Symposium on Circuits and Systems*, pages 253–256, 2010. 1

[5] E. Sariyanidi, H. Gunes, and A. Cavallaro. Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(6):1113–1133, 2015. 1

[6] C. Shan, S. Gong, and P. W. McOwan. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*, 11(10):9573–9588, 2011. 1