Data Mining Final Project

Francisco R. Ortega

Professor: Dr. Tao Li

FALL 2009

1.Introduction

In the data mining class one of the aspects of interest were classifications. For the final project, the decision to use images for facial expressions recognition and specific classifiers (PCA and SVM,) were the motivation for the topic. In specific, what methods can be use for facial expression recognition? In Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection has a solid discussion on methods using PCA and FCA (Fisher Component Analysis.) [1].

2.Background

Principal Component Analysis.

PCA has the following important objectives [2]:

- Extract relevant information
- Data reduction/compression
- Dataset simplification

Principal Component Analysis is defined in equation 1. Equation 2 shows the criterion function, which is minimized when, vectors $e_1,e_2,\ldots.en$ are the eigenvector of the matrix, having the maximum value[2][4].

$$\vec{x} = \vec{m} + \sum_{i}^{d'} a_i e_i \qquad \text{(Eq. 1)}$$

$$J_{d'} = \sum_{k=1}^{n} \left\| \left(\left(\vec{m} + \sum_{i=1}^{d'} a_{ki} e_i \right) - x_k \right) \right\|^2 \qquad \text{(Eq. 2)}$$

Fisher Linear Discriminat

PCA provides a method to represent data but Fisher Linear Discriminant provides a way to discriminate between different classes. One way of looking at this is that PCA looks for a direction to represent the data and FCA looks for a direction to discriminate the data [2].

In general, the generalization of FLD can be seen in equations 3 and 4. This assumes that dimension is greater that the number of classes [3]. The idea is to select W so the scatter ration between the classes and within the classes are maximized as shown in equation 5[1]

$$S_B = \sum_{k=1}^{c} N_i (m_i - m)(m_i - m)^T \qquad \text{(Eq. 3)}$$

$$S_W = \sum_{k=1}^{c} \sum_{x_i \in \vec{X}_j} \left((\vec{m}_i - \vec{m})(\vec{m}_i - \vec{m})^T \right) \qquad \text{(Eq. 4)}$$

$$Wopt = \max_w \left| \frac{W^T S_B W}{W^T S_W W} \right| \qquad \text{(Eq. 5)}$$

Support Vector Machine

An optimized solution for SVM, called Lagrange dual function is shown in equation 6. The advantages is that h can be computed using inner products, hence a less expensive computational cost [3]. Figure 1 shows a separable case with SVM as opposed to the one in the right.

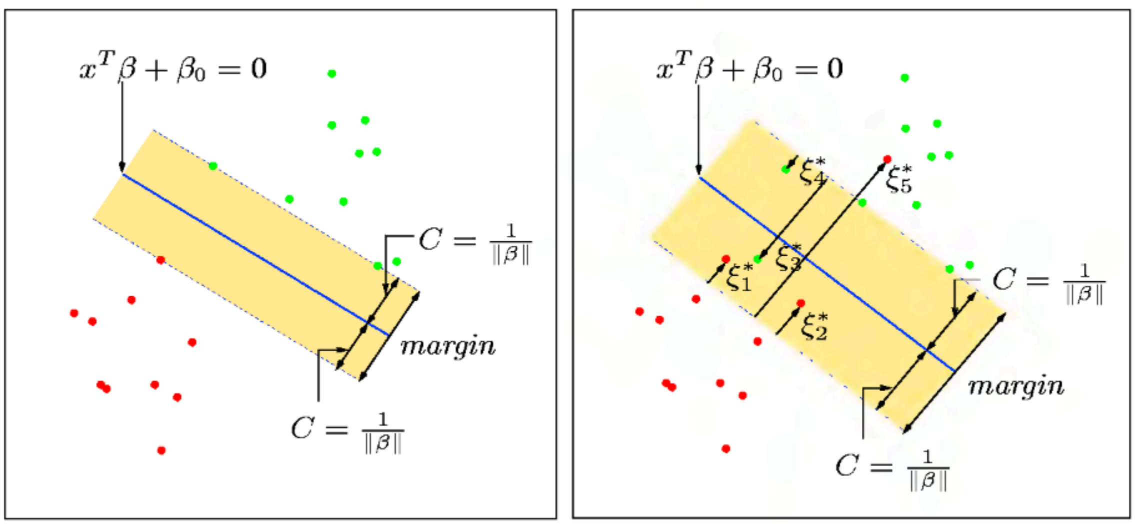


Fig 1 from: The Elements of Statistical Learning [3]

$L_d = \sum_{k=1}^{N} \alpha_k - \frac{1}{2} \sum_{k=1}^{N} \sum_{m=1}^{N} \alpha_k \alpha_m y_k y_m \langle h(x_k), h(x_m) \rangle$	(Eq. 6)
$f(x) = \sum_{k=1}^{N} \alpha_k y_k \langle h(x), h(x_k) \rangle + \beta_0$	(Eq. 7)
$\hat{f}(x) = \sum_{k=1}^{N} \hat{\alpha}_k y_k \langle h(x), h(x_k) \rangle + \hat{\beta}_0$	(Eq. 8.)

3. Methods

JAFFE training data set was used. This dataset contained 11 persons, with 7 expressions (Happy, Sad, Surprise, Neutral, Fear, Anger, Sad) and each of that expression with 3 shots. All the images are in grayscales. A matlab procedure was used to automatically import all the data into a 3-dimensional array <person,expression,takenumber>. In addition to this, an additional array was created with the exact same number of inputs with a modified picture, which was applied Sobel operator for edge detection technique. Figure 2 and 3 show the difference between both.

The following steps to take are the classification techniques. Due to time, the program was not able to be complete and correct. However, in [1] a modification is proposed to the FLD. The W factor/weight makes the difference in how FLD works. Equation 5 gets replace by equation 11 where each of the W factors is shown in equation 9 and 10. One of the reasons for this proposed modification by [1] it's because S_w in imaging is always zero (singular.)[1] Even if this was near singular, then S_w will dominate W_{opt} in equation 5. The reason of this singularity, it's because the classes in S_w it's much smaller than the pixels in each image [1]. Finally, graph 1 shows a way the difference between PCA and FLD, hence their contribution to combined certain aspects of PCA to FLD.

$$W_{fld} = \max_w \left| \frac{W^T W_{pca}^T S_B W_{pca}^T W}{W^T W_{pca}^T S_W W_{pca} W} \right| \qquad \text{(Eq. 9)}$$

$$W_{pca} = \max_w \left| W^T S_t W \right| \qquad \text{(Eq. 10)}$$

$$W_{opt}^T = W_{fld}^T W_{pca}^T \qquad \text{(Eq. 11)}$$

In addition to their contribution, my original idea was to combine PCA with SVM. Given that FLD is shown in [1] to work better for the recognition, my idea was shifted to use equation 11 with a few other methods before and after, which are listed below:

1. Upload images as 8-bit grayscale into Matrix I, where I has a tuple of <personidentifier,expressionidentifier,takenumber>
2. Upload subset of images already classified into X and Y. X will contain the attributes and Y will contain the targets (classes.)
3. Choose image filter or operator (i.e. sobel.), which will create X_b and I_b
4. Select appropriate W and run modified FLD (equation 11.)
5. Save results
6. Use SVM with results from step 4. (Also use Y to train.)

In addition, some additional changes could be done. For example, it's possible to detect motion. Therefore, if this would be video as opposed to pictures, then using motion detection algorithms, the challenge would be to detect the area of interest (near the mouth) and use that space to train the SVM before step 6. However, subtle movements in the upper part of the face maybe of interest; however, due to resolution, it's acceptable to ask if this would have any implication.

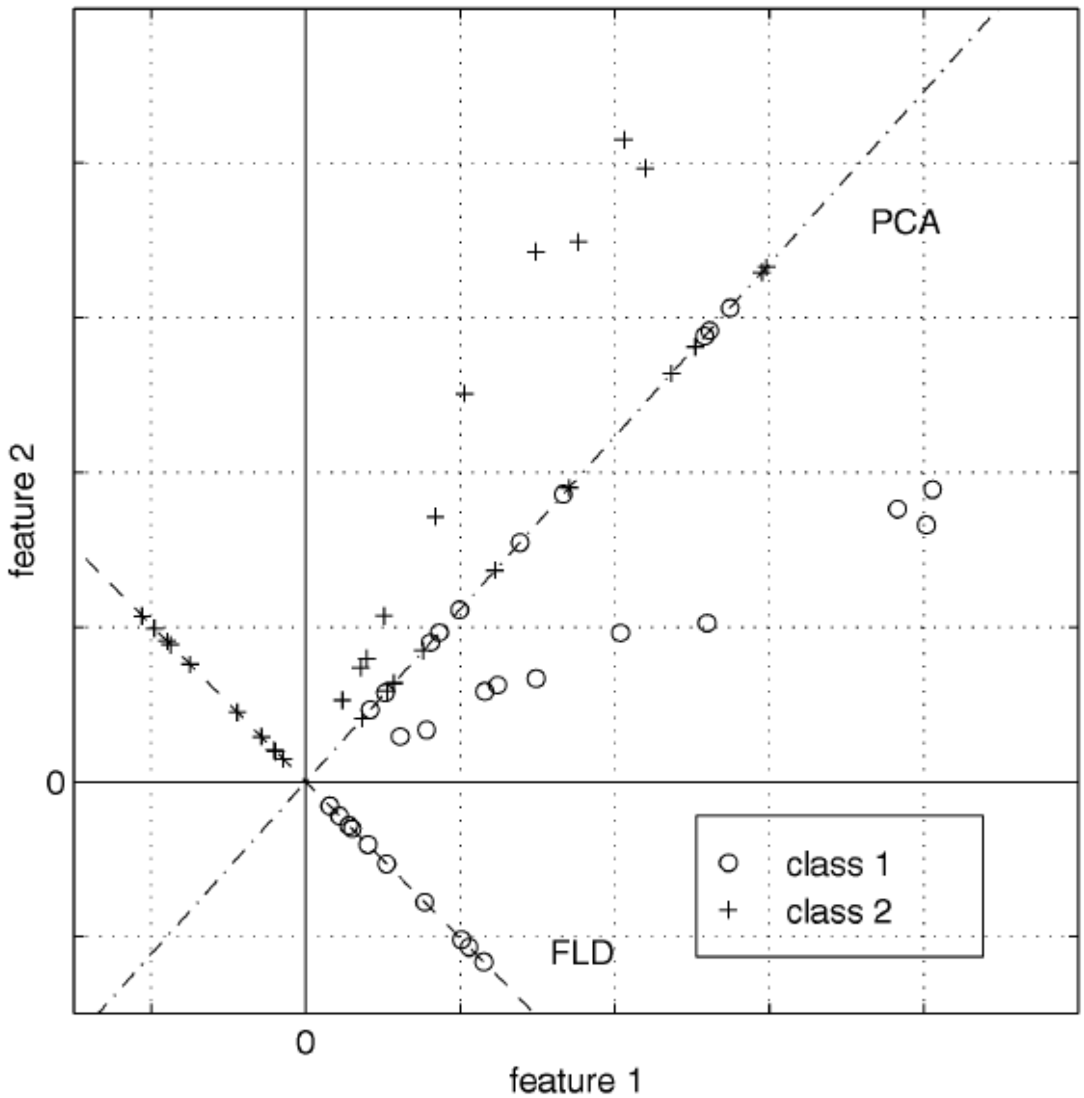
Candide is based on Ekman's Facial Action Coding System (FACS.) [Rydfalk, 1987] Facs allows to control facial expression, and therefore, facial recognition. Candide-3 is composed of 12 shape units as shown in table x, to allow different head shapes within the limitation of the amount of vertices. It also contains action units (AU) is a movement in the face (e.g. jaw drop, mouth stretch, lower lip depressor...) and an action unit vector (AUV) is the vector that contains a set of AU. [Ahlberg, 2001]. Figure 4 shows the skeleton for Candide-3.

For additional information about candide-3 and FAPS (MPEG-4) that are not covered in this paper as well the original candide see, see reports: [Ahlberg, 2001], [Rydfalk, 1987].

Finally, a valid argument by [Pantic et el 2009] survey, which deals with facial expression recognition, is about data correctness with facial expressions. Is it the posed images the same as spontaneous expressions? Even though there have been studies to be able to make a difference, I still considered this as an open question.


Figure 2


Figure 3



Graph 1. [1]

Figure 4 – Candide-3

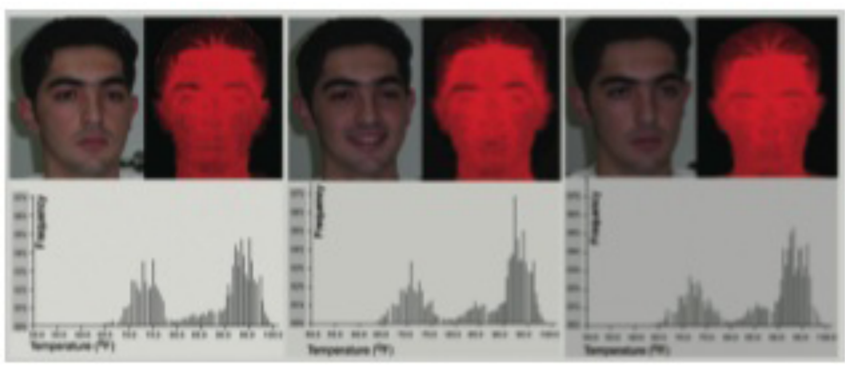


Figure 5 – Infrared Recognition [6]

4. Future Work

As in figure 4, additional techniques are being used to detected facial expressions. In this figure, the use of infrared cameras and sobel operators is used to help the data acquisition. For more information about the state of the art, please see [Pantic et el. 2009].

5. Conclusion

The use of Support Vector Machines in conjunction with PCA/LDA still is to be determined. It's possible that the used of both may not bring the correct results. My current interest in classification will yield additional testing until I can find a suitable way that improves the current state of the art.

Note: If the reference is not in the bibliography, then the author has been quoted by name in the paper. References with [1] a number represents the important resources used.

Bibliography

1. Bulhemeur et el. (1997). Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection.

2. Duda et el. (2000). *Pattern Classification* (Second Edition ed.). New York, NY, USA: Wiley-Interscience.

3. Hastie. et el. (2009). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction* (Second Edition ed.). New York, NY, USA: Springer.

4. Tan et el. (2006). *Introduction to Data Mining.* Boston, MA , USA: Pearson Education.

5. Martinez, W. L., & Martinez, A. R. (2005). *Exploratory Data Analisys with MATLAB.* boca raton, FL, USA: Chapman Hall/CRC.

6. Khan, M., Ward, R., & Ingleby, M. (2009). Classifying pretended and evoked facial expressions of positive and negative affective states using infrared measurement of skin temperature. *Transactions on Applied Perception*