A Framework for Automated Measurement of the Intensity of Non-Posed Facial Action Units

Mohammad H. Mahoor[1], Steven Cadavid[2], Daniel S. Messinger[3], and Jeffrey F. Cohn[4]

[1]Department of Electrical and Computer Engineering, University of Denver, Denver, CO 8028
[2]Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL 33146
[3]Department of Psychology, University of Miami, Coral Gables, FL 33146
[4]Department of Psychology, University of Pittsburgh, Pittsburgh, PA 15260
Emails:
mmahoor@du.edu, scadavid@umsis.miami.edu, dmessinger@miami.edu, and jeffcohn@pitt.edu

Abstract

This paper presents a framework to automatically measure the intensity of naturally occurring facial actions. Naturalistic expressions are non-posed spontaneous actions. The Facial Action Coding System (FACS) is the gold standard technique for describing facial expressions, which are parsed as comprehensive, nonoverlapping Action Units (Aus). AUs have intensities ranging from absent to maximal on a six-point metric (i.e., 0 to 5). Despite the efforts in recognizing the presence of non-posed action units, measuring their intensity has not been studied comprehensively. In this paper, we develop a framework to measure the intensity of AU12 (Lip Corner Puller) and AU6 (Cheek Raising) in videos captured from infant-mother live face-to-face communications. The AU12 and AU6 are the most challenging case of infant's expressions (e.g., low facial texture in infant's face). One of the problems in facial image analysis is the large dimensionality of the visual data. Our approach for solving this problem is to utilize the spectral regression technique to project high dimensionality facial images into a low dimensionality space. Represented facial images in the low dimensional space are utilized to train Support Vector Machine classifiers to predict the intensity of action units. Analysis of 18 minutes of captured video of non-posed facial expressions of several infants and mothers shows significant agreement between a human FACS coder and our approach, which makes it an efficient approach for automated measurement of the intensity of non-posed facial action units.

1. Introduction

Human face-to-face communication plays an important role in behavioral science and developmental psychology [27]. Facial expressions are the most important visual channels used by humans in face-to-face communication. Efficient measurement of facial expression is necessary to understand the functionality of face-to-face communication. The most comprehensive measurement approach to quantify facial expressions of partners (e.g., mother and infant) in a face-to-face communication, is manual coding of Action Units based on the Facial Action Coding System (FACS) [11, 22]. FACS provides a description of all possible and visually detectable facial variations in terms of 44 Action Units (AUs). Usually, a trained human FACS coder identifies the occurrence of an action unit and codes its intensity in a given facial image. Although the FACS coding is a precise tool for studying facial expressions, it is labor intensive. Therefore, automating the FACS coding and measuring the intensity of AUs would make it easier and widely accessible as a research tool in behavioral science.

Automated recognition of facial expressions and action units has caught the attention of many researcher in the field of computer vision since the early 1980s. Most of the studies on automated expression analysis focus on classification of basic expressions (i.e., joy, fear, sadness, disgust, surprise, and anger) [27]. The utilized image analysis techniques include holistic analysis, spatio-temporal analysis and analytic spatial analysis. We refer our reader to [27] for more details on these techniques. More recent systems [17, 2, 9, 18] have achieved some success in the more difficult task of recognizing facial action units.

In a real face-to-face communication, we deal with non-posed facial expressions. Posed facial expressions and ac-

978-1-4244-3993-5/09/$25.00 ©2009 IEEE

tion units are those that are created by asking subjects to deliberately make specific facial actions or expressions. On the other hand, non-posed facial expressions and action units are representative of facial expressions in daily life. They typically occur in uncontrolled conditions and are combined with head pose variation, head movement and often more complex facial action units. Non-posed AUs have intensities measurable from absent to maximal appearance using a six-point intensity metric (i.e., 0 to 5). Most of the developed systems for facial expression and action unit classification are evaluated using posed expressions data (e.g., Cohn-Kanade face expression database [16]).

Recently, Bartlett et al. [2] attempted to measure the intensity of action units in posed and non-posed facial expressions. Their system is based on Gabor wavelet and support vector machines. Gabor wavelets are 2-D sine waves modulated by a Gaussian envelope in different spatial scales and orientation and are applied to features describing local appearance of the face. Suites of extracted Gabor features are utilized to train support vector machines (SVMs) classifying the action units. They have reported average correlation values of .3 and .63 between a human coder and the predicted intensity of action units of non-posed and posed expressions, respectively. The correlation value is moderate and statistically significant for posed expressions, but low and insignificant for non-posed expressions. This means that measuring the intensity of non-posed expressions is more challenging than measuring the intensity of posed expressions.

Most recently, Lucy et al. [18] studied the effective representation of face for detection of non-posed action units. They have investigated the employment of the Active Appearance Model (AAM) [10, 19] to derive an effective facial representation. They evaluated the effectiveness of this representation using the RU-FACS spontaneous expression database [1]. They attempted to detect the existence of action units, however, a measure of intensity was not performed. Presently, there has been no significant work in measuring the intensity of non-posed action units.

In [23] Reilly et al. presented their effort in capturing the dynamic of facial expression by describing its intensity and timing of formation. Their system uses the Local Linear Embedding (LLE) technique to reduce the dimensionality of the facial images and employs SVMs to classify combination of action units and facial expressions. Compared to our work, they manually extract 24 facial landmarks on the mouth and use only shape information for capturing the dynamics of facial expressions. A study by Lucy et al. [18] reveals that shape information is not sufficient for modeling the dynamics of facial expressions and the combination of appearance and shape is the most successful representation for action unit recognition. Also, posed facial expressions data (Cohn-Kanade database) was used by Reilly et al. to evaluate their system.

The focus of this paper is to develop a framework to measure the intensity of action units from absent to maximal (i.e., 0-5 metric) in non-posed facial expressions. In our approach, we track and represent facial images in captured videos using AAM. AAM consists of a shape component and an appearance component that jointly represent the shape and texture variability seen in the object. Although, the appearance component in conjunction with the shape component is a useful representation for facial expression analysis, it has extremely large dimensionality. For instance, a facial image with a size of 128×128 pixels has a dimensionality of 16,384 in the image space. Despite the huge dimensionality of the visual data, activities such as facial expressions have low dimensions embedded in a large dimensional space [12]. Traditional techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis have limited ability of reducing the dimensionality of complex nonlinear facial expression data. Recently, several nonlinear data reduction techniques such as Isomap [26], Locally Linear Embedding [24], and Laplacian Eigenmap [21] have been presented for dimensionality reduction. Laplacian Eigenmap and its variants such as Laplacianface [14] and Orthogonal Locally Linear Embedding [6] have shown promising results in face recognition [14] and age estimation [13] from facial images. We are inspired by [7] and adopt the spectral regression technique to learn projection functions that map AAM representations into separate sub-spaces called action unit sub-spaces. Reduced feature points presented in separate sub-spaces are employed to measure the intensity of an individual action unit based on Support Vector Machine (SVM) classifiers.

The remainder of this paper is organized as follows. A face representation based on the Active Appearance Model is introduced in Section 2. Section 3 develops our approach to data dimensionality reduction. Section 4 reviews the SVM classifiers employed for predicting the intensity of action units. Section 5 shows the experimental results and conclusions. Future work is discussed in Section 6.

2. Face Representation: Active Appearance Model

Determining an adequate facial image representation for effectively measuring the intensity of facial expressions and action units is a challenging problem. A most recent study by Lucy et al. [18] reveals that AAM [10, 19] is a successful facial representation technique for action unit recognition. In this section, we review the AAM and describe the AAM-based facial representation exploited in this work.

AAM is a statistical representation of an object (e.g., face) introduced by Cootes et al. [10] and improved by others [19] over the past few years. AAM consists of a shape

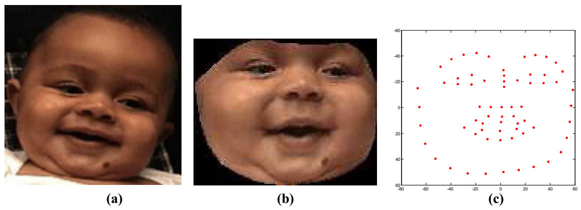

Figure 1. (a) A sample facial image along with the AAM components (b) normalized-appearance face image and (c) normalized shape data.

component, $\mathbf{s}$, and an appearance component, $\mathbf{g}$, that jointly represent the shape and texture variability seen in the object. The shape component represents a target structure by a parameterized statistical shape model obtained from training. The shape model is defined by a linear model:

$$\mathbf{s} = \mathbf{s}_0 + \sum_{i=1}^{m} p_i^{(s)}\, \mathbf{s}_i \tag{1}$$

where $\mathbf{s}_0$ is the mean shape vector, $\mathbf{s}_i$ is a set of orthogonal modes (i.e., eigenvectors) of shape variation calculated by applying PCA on the covariance matrix of the training shape data, and $\mathbf{p}^{(s)} = [p_1^{(s)}, ..., p_m^{(s)}]^T$ is a vector of shape parameters. The appearance statistical model is built by warping each image instance so that its control points match the mean shape using the thin-plate spline algorithm [4]. Then, the intensity variation is sampled from the shape-normalized image over the region covered by the mean shape. Similarly, by applying PCA to the appearance data a linear model is defined:

$$\mathbf{g} = \mathbf{g}_0 + \sum_{i=1}^{m} p_i^{(g)}\, \mathbf{g}_i \tag{2}$$

where $\mathbf{g}_0$ is the mean normalized grey-level vector, $\mathbf{g}_i$ is a set of orthogonal modes (i.g., eigenvectors) of intensity variation and $\mathbf{p}^{(g)} = [p_1^{(g)}, ..., p_m^{(g)}]^T$ is a set of grey-level parameters. This generates shape data on facial landmarks and appearance data on the gray-level intensity of each pixel in the face model. Figure 1 illustrates a sample facial image decomposed into the shape-normalized appearance component and the shape component where consists of 66 facial landmark points.

The concatenation of the shape-normalized appearance and the shape component, $[\mathbf{g}, \mathbf{s}]$, in the form of a vector is shown to be an effective representation for facial analysis [18]. Due to the variations in subject facial appearance, we normalize each concatenated feature vector by subtracting from a neutral feature vector obtained from the subject: $X_{normalized} = X_{expression} - X_{neutral}$. The $X_{normalized}$ feature vector, which is known as delta feature, decreases the biasing effect of subject variation on action unit measurement. This representation is employed to measure the intensity of facial action units. However, due to the curse of dimensionality of delta features (i.e., 10,132 dimensions, corresponding to an 100×100 image pixels and 132 coordinate values of 66 facial landmark points), the classification of action unit intensity is difficult. Therefore, reducing the dimensionality of the visual data becomes vital and is addressed in the following section.

3. Action Unit Sub-Space Learning

The problem of dimensionality reduction arises in the areas of computer vision, artificial intelligence and data mining. Traditionally, linear techniques such as PCA and LDA are utilized to project a feature vector from a high dimensional space, R^D, into a low dimensional space, R^d ($d << D$) [28]. Linear techniques have limited ability to represent complex nonlinear data such as facial expressions in a low dimensional sub-space. Recently, developed nonlinear dimensionality reduction techniques such as Isomap [26], Laplacian Eigenmap [21], and Locally Linear Embedding [24] have shown success in reducing the dimensionality of complex data. These techniques are also known as manifold learning methods since they assume that the original feature data lies on a low dimensional manifold embedded in a high dimensional space. These techniques are computationally efficient and have locality-preserving properties.

Laplacian Eigenmap and its variants (e.g., Orthogonal Locally Linear Embedding [6]) have been successfully used in face identification [14] and face aging recognition [13]. In this paper, we employ the Laplacian Eigenmap followed by the spectral regression technique [7] to project facial images (i.e., the appearance and shape features) into action unit sub-spaces by learning separate projection matrices. In the following, we review the locality preserving indexing (LPI) technique [5] and then present the regularized LPI via spectral regression.

3.1. Locality Preserving Indexing

Given a set of k points $x_1, ..., x_k$ in R^D, we want to find a set of points $y_1, ..., y_k$ in R^d ($d << D$) such that y_i corresponds to x_i. In other words, we look for a mapping function, $\mathbf{a}$ such that: $y_i = \mathbf{a}^T\, x_i$. Given a similarity matrix W describing the closeness of the points in space R^D, the mapping is obtained by solving the following optimization problem:

$$\begin{aligned}\mathbf{a} &= \underset{\mathbf{a}^T X D X^T \mathbf{a}=1}{\arg\min} \sum_{i=1}^{k}\sum_{j=1}^{k}(\mathbf{a}^T x_i - \mathbf{a}^T x_j)^2 W_{ij} \\ &= \underset{\mathbf{a}^T X D X^T \mathbf{a}=1}{\arg\min}\ \mathbf{a}^T X L X^T \mathbf{a}\end{aligned} \tag{3}$$

where D is a diagonal matrix whose elements are column sums of W ($D_{ii} = \sum_j W_{ij}$) and $L = D - W$ is the Laplacian graph. The constraint $\mathbf{a}^T XDX^T\mathbf{a} = 1$ effectively fixes a scaling factor of the solution.

The similarity matrix W is constructed by finding the k nearest neighbor points using the Euclidean norm in R^D and weights are assigned as follows: $W_{ij} = 1$ if two points x_i and x_j are neighbors and $W_{ij} = 0$, otherwise. Alternatively, weights can be assigned by using the heat kernel, $W_{ij} = \exp^{-\frac{||x_i - x_j||^2}{t}}$ [21]. The weight assignment can also be supervised if we know the class that these points belong to. Therefore, if two points belong to different classes, the assigned weight is zero, otherwise the weight is calculated as described above. The supervised method is used to calculate W in this paper.

The objective function 3 incurs a heavy penalty if neighboring points x_i and x_j are mapped far apart in the low dimensional space. This ensures that if two points are close in the R^D space, then they are also close in the R^d space. By using the Lagrangian multiplier technique, the optimization problem 3 can be solved by computing eigenvalues, λ, and eigenvectors for the generalized eigen-problem:

$$XLX^T\mathbf{a} = \lambda\, XDX^T\,\mathbf{a}. \tag{4}$$

The eigenvectors associated with the smallest eigenvalues construct vector $\mathbf{a}$.

To obtain a stable solution of the eigen-problem 4, the matrix XDX^T needs to be non-singular. However, the number of features (i.e., $X_{normalized}$) are much larger than the number of sample facial images under study and this makes XDX^T singular. One potential solution is to use Singular Value Decomposition to solve this problem [5], but it is computationally very expensive. An alternative approach for finding the projection matrix $\mathbf{a}$, is to use the following two-step method called regularized LPI via spectral regression [7]

3.2. Regularized Locality Preserving Indexing: Spectral Regression

The problem with LPI technique is its high computational complexity and therefore cannot be applied to large dimensional visual data. Recently, Cai et al. [7] presented the regularized locality preservation algorithm which has shown success in representing large dimensional data in a low dimensional space. Similar to LPI, the regularized locality preserving algorithm wants to find the mapping function $\mathbf{a}$ that maps a set of points $X = [x_1, x_2, ..., x_k]$ represented in a high dimensional space, to points $Y = [y_1, y_2, ..., y_k]$ in a low dimensional space ($y_i = \mathbf{a}^T x_i$). This approach is computationally efficient and is described as follows:

1. Find Y by solving the following optimization problem:

$$\begin{aligned} Y &= \underset{YDY^T=1}{\arg\min} \sum_{i=1}^{k}\sum_{j=1}^{k}(y_i - y_j)^2 W_{ij} \\ &= \underset{YDY^T=1}{\arg\min}\; YLY^T \end{aligned} \tag{5}$$

   where W is the weight matrix described in the previous section. This optimization problem which is also known as Laplacian Eigenmap [21] can be solved efficiently by calculating the eigenvectors of the generalized eigen-problem $LY = \lambda DY$. Matrix YLY^T is not a singular matrix [3].

2. Find $\mathbf{a}$ such that $Y = \mathbf{a}^T X$ by solving a regularized least squares problem:

$$\mathbf{a} = \arg\min_{\mathbf{a}}\{\sum_{i=1}^{k}(\mathbf{a}^T x_i - y_i)^2 + \alpha||\mathbf{a}||^2\} \tag{6}$$

   The regularization term guarantees that the least squares problem is well-posed and has a unique solution.

This technique is called spectral regression since it performs spectral analysis on the Laplacian graph [3] followed by least square regression. We use this approach to learn separate projection matrices $\mathbf{a}$, to represent each facial image in low dimensional action unit sub-spaces. Since the supervised method is used to calculate the similarity matrix W, we solve separate optimization problems for obtaining $\mathbf{a}_{AU12}$ and $\mathbf{a}_{AU6}$

4. Measuring the Intensity of Facial Action Units

After the projection of $X_{normalized}$ features into action unit sub-spaces, the problem is to measure the intensity of action units described in six levels from these features. This is a multi-label classification problem and we employ SVM classifiers to solve this problem. The SVMs have been used in the field of machine learning and pattern recognition, and has shown success in recognition of facial expressions and action units [18, 17]. We refer our reader to [29] for technical details on SVM classification.

Kernel functions are usually employed to efficiently map input data which may not be linearly separable to a feature space where linear methods can then be applied. Based on the kernel mapping approach, every inner product is replaced by a nonlinear kernel function kernel $K(x, y) = \phi(x).\phi(y)$ where x and y are two input data sets. There are different types of kernel mappings such as the polynomial kernel and the Radial Basis Function (RBF) kernel.

SVMs using kernel functions demonstrate good classification accuracy even when only a modest amount of training data is available, making them particularly suitable for a dynamic, interactive approach to expression recognition. Our experiments show that the RBF kernel has the highest performance in classifying the intensity of action units. Since SVM is a binary classifier, the "one-against-one" technique [15] is used to extend the binary SVM classifier into a multi-label classifier. The LibSVM library is employed in this work [8].

5. Experiments and Results

Parent-infant communication is a topic of considerable interest in developmental psychology. Much of this communication occurs during the exchange of expressions of positive affect. Positive affect can be parsimoniously indexed by the intensity of infant smiling (AU12, Lip Corner Puller) and co-occurring Cheek Raising and Lid Compression (AU6) [20]. They are the most challenging cases of infant's expression (e.g., low facial texture in infant's face). Our ability to accurately measure intensity variations in these actions provides a framework for measuring early human communication dynamics. This framework can be easily extended to other FACS AUs in infants and adults.

In our study, videos were recorded (at 30 fps) of facial expressions of four 6-month infants and two mothers during a face-to-face interaction. The infants were engaged by their mothers in order to incite emotional expression from the infants. Video of the facial expressions was captured for a three minute interval. There was no condition placed on subject's head movement. All of the captured frames were used in our experiment except for those where the face was occluded by the subject's hand or foot or where the eyes of the subjects were not visible due to severe head pose.Overall, 24,680 frames were used in our experiments.

A certified FACS/BabyFACS coder manually reliably coded the presence and intensity of AU12 and AU6 (i.e., 0 represents absence and 1-5 represent intensity) occurring in all video frames; these codes were employed for training and testing of our system. The facial videos were tracked and modeled using the AAM algorithm provided by [19] and the delta features were extracted for every video frame of all four infants. The $X_{normalized}$ features of every 15^{th} frame were used to learn the projection matrix **a** based on spectral regression for the AU12 sub-space and the AU6 sub-space.

Our experiments are based on Leave-One-Subject-Out (LOSO) cross validation to predict the intensity of action units. SVM training was performed on every $m^{th}, m = 2, 5, 7, 10, 12, 15, 20, 25, 30, 50$ frame of video excluding one subject. Testing was performed on the left out subject. This scenario was repeated for all other subjects.

In order to compare the predicted and manually coded intensities of action units, we calculate Intra-Class Correlation (ICC). ICC ranges from 0 to 1 and is a measure of correlation or conformity for a data set when it has multiple targets [25]. In other words, ICC measures the reliability studies in which n targets of data are rated by k judges (i.e., in this paper $k = 2$ and $n = 6$). ICC is similar to Pearson correlation and is preferred when computing consistency between judges or measurement devices.

The ICC in is defined as:

$$ICC = \frac{(BMS - EMS)}{(BMS + (k-1) * EMS)} \tag{7}$$

where BMS is the between-targets mean squares and EMS is the residual mean squares defined by Analysis Of Variance (ANOVA). That is, the ICC indicates the proportion of total variance due to differences between targets. See [25] for additional details.

Table 1 shows the ICC coefficient[1] between the actual and predicted intensity of AU6 and AU12 using the SVM classifiers calculated for four infants and two mothers under study (the results of this table is based on $m = 5$). As the table demonstrates, our approach has high performance on measuring the intensity of AU6 and AU12 (ICC = .81 and .84, respectively). The number of frames with AU events is presented in Table 2.

Figure 2 shows the effect that the number of training frames has on the accuracy of the system when classifying the intensity of action units. As the figure illustrates, even by using a small training set (2% of the frames, $m = 50$), the system has a high agreement with the human coder (the overall ICC = .79). In another experiment, we used the PCA technique (i.e., Eigenface) instead of spectral regression for data dimensionality reduction. We utilized 100 Eigenfaces (capturing 95% of the total variance) obtained from the same dataset used for training the spectral regression. The overall ICC in measuring the intensity of AU6 and AU12 based on PCA and SVM is .61 (5% of the frames used for training the SVMs). Compared to the results obtained using the spectral regression for data dimensionality reduction (i.e., .81 and .84), the nonlinear spectral regression technique shows better results than the linear PCA technique.

6. Conclusions and Future work

In this paper, we presented a framework for measuring the intensity of non-posed facial action units in facial images. We utilized the idea of Regularized Locality Preservation via spectral regression to reduce the dimensionality of facial images modeled by AAM. Our approach was employed to measure the intensity of AU12 and AU6 in facial

[1] For comparison, the overall Pearson correlation coefficients are .82 and .85 for AU6 and AU12, respectively.

	Intra-Class Correlation		Number of Frames
Sub.	AU6	AU12	
Infant 1	.83	.90	4968
Infant 2	.87	.93	4385
Infant 3	.92	.80	3706
Infant 4	.76	.85	2268
Mother 1	.78	.83	4968
Mother 2	.7	.65	4385
Overall	.81	.84	24,680

Table 1. Intra-Class Correlation coefficient between the actual and predicted intensity of AU6 and AU12 calculated for the four infants and two mothers; 20% of data ($m = 5$) used to train the SVM classifier.

Intensity	AU6	AU12
0	4405	5262
1	5988	5281
2	5378	4536
3	4338	4528
4	3485	3932
5	1086	1141
Overall	24,680	24,680

Table 2. Number of frames with AU events processed in this study.

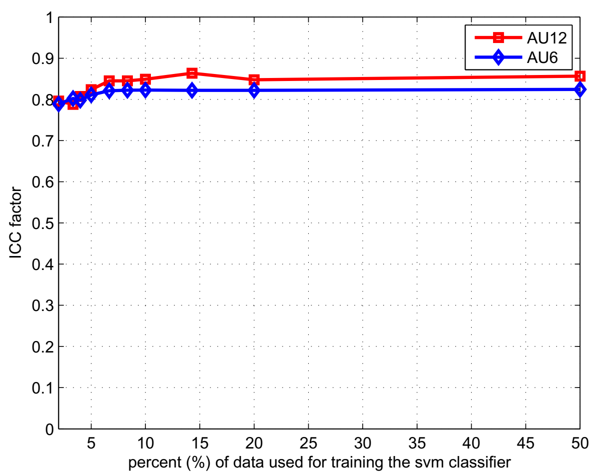

Figure 2. ICC coefficient versus percentage of data used in training the SVM classifiers.

expressions of infants in a live face-to-face communication. The statistical agreement (i.e., the ICC coefficient) between a human FACS coder and our system in quantifying the intensity of non-posed action unit are significantly high.

We will utilize our framework to measure the intensity of other FACS action units related to negative facial expressions. A major problem in measuring the intensity of non-posed action units is head movement. We plan to decompose the facial variations into separate factors such as head movement and facial expression and only use the facial expression factor in quantifying the facial action units.

References

[1] Ru-facs expression database. http://mplab.ucsd.edu/databases/databases.html. 2

[2] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Fully automatic facial action recognition in spontaneous behavior. pages 223–230, 2006. 1, 2

[3] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems 14*, pages 585–591. MIT Press, 2002. 4

[4] F. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on PAMI*, 11(6):567–585, June 1989. 3

[5] D. Cai, X. He, and J. Han. Document clustering using locality preserving indexing. *IEEE Trans. Knowl. Data Eng*, 17(12):1624–1637, 2005. 3, 4

[6] D. Cai, X. He, J. Han, and H. Zhang. Orthogonal laplacianfaces for face recognition. *IEEE Transactions on Image Processing*, 15(11):3608–3614, November 2006. 2, 3

[7] D. Cai, X. He, W. V. Zhang, and J. Han. Regularized locality preserving indexing via spectral regression. In *CIKM '07: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 741–750, New York, NY, USA, 2007. ACM. 2, 3, 4

[8] C.-C. Chang and C.-J. Lin. Libsvm-a library for support vector machines. http://www.csie.ntu.edu.tw/ cjlin/libsvm/. 5

[9] J. Cohn and T. Kanade. Use of automated facial image analysis for measurement of emotion expression. In J. A. C. . J. B. Allen, editor, *The handbook of emotion elicitation and assessment. Oxford University Press Series in Affective Science*. 2006. 1

[10] T. Cootes, D. Cooper, C. Taylor, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, Jan. 1995. 2

[11] P. Ekman and W. Friesen. The facial action coding system: A technique for the measurement of facial movement. In *Consulting Psychologists*, 1978. 1

[12] A. Elgammal and C. Lee. Nonlinear manifold learning for dynamic shape and dynamic appearance. 106(1):31–46, April 2007. 2

[13] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang. Image-based human age estimation by manifold learning and

locally adjusted robust regression. *IEEE Transactions on Image Processing*, 17(7):1178–1188, 2008. 2, 3

[14] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. Face recognition using laplacianfaces. 27(3):328–340, March 2005. 2, 3

[15] C.-W. Hsu, C.-C. Chang, and C.-J. Lin. A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University, 2003. 5

[16] T. Kanade, J. Cohn, and Y.-L. Tian. Comprehensive database for facial expression analysis. In *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00)*, pages 46 – 53, March 2000. 2

[17] G. Littlewort-ford, M. S. Bartlett, and J. R. Movellan. Are your eyes smiling? detecting genuine smiles with support vector machines and gabor wavelets. In *In Proceedings of the 8th Joint Symposium on Neural Computation*, 2001. 1, 4

[18] S. Lucey, A. B. Ashraf, and J. Cohn. Investigating spontaneous facial action recognition through aam representations of the face. In K. Kurihara, editor, *Face Recognition Book*. Pro Literatur Verlag, Mammendorf, Germany, April 2007. 1, 2, 3, 4

[19] I. Matthews and S. Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, November 2004. 2, 5

[20] D. Messinger, M. H. Mahoor, S. Chow, and J. F. Cohn. Continuously automated measurement of facial expression in infant-mother interaction: A pilot study. *Infancy*, 14, 2009. 5

[21] P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15:1373–1396, 2003. 2, 3, 4

[22] H. Oster. Baby facs: Facial action coding system for infants and young children. Technical report, Unpublished monograph and coding manual. New York University, 2006. 1

[23] J. Reilly, J. Ghent, and J. McDonald. Investigating the dynamics of facial expression. In *Proceedings of the International Symposium on Visual Computing*, 2006. 2

[24] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. 290(5500):2323–2326, December 2000. 2, 3

[25] P. Shrout and J. Fleiss. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(2):420–428, 1979. 5

[26] J. Tenenbaum, V. de Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. 290(5500):2319–2323, December 2000. 2, 3

[27] Y.-L. Tian, T. Kanade, and J. Cohn. Facial expression analysis. In S. L. . A. Jain, editor, *Handbook of face recognition*, pages 247–276. Springer, New York, New York, 2005. 1

[28] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1), 1991. 3

[29] V. Vapnik, S. E. Golowich, and A. J. Smola. Support vector method for function approximation, regression estimation and signal processing. In M. Mozer, M. I. Jordan, and T. Petsche, editors, *Proceedings of the 1996 Neural Information Processing Systems Conference, December 2-5, 1996, Dever, CO, USA*, pages 281–287, 1997. 4