Automated Classification of Gaze Direction Using Spectral Regression and Support Vector Machine

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

[1]Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL 33146
[2]Department of Electrical and Computer Engineering, University of Denver, Denver, CO 8028
[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, s.cadavid@umiami.edu, dmessinger@miami.edu, and jeffcohn@pitt.edu

Abstract

This paper presents a framework to automatically estimate the gaze direction of an infant in an infant-parent face-to-face interaction. Commercial devices are sometimes used to produce automated measurement of the subjects' gaze direction. This approach is intrusive, requiring cooperation from the participants, and cannot be employed in interactive face-to-face communication scenarios between a parent and their infant. Alternately, the infant gazes that are at and away from the parent's face may be manually coded from captured videos by a human expert. However, this approach is labor intensive. A preferred alternative would be to automatically estimate the gaze direction of participants from captured videos. The realization of a such a system will help psychological scientists to readily study and understand the early attention of infants. One of the problems in eye region image analysis is the large dimensionality of the visual data. We address this problem by employing the spectral regression technique to project high dimensionality eye region images into a low dimensional sub-space. Represented eye region images in the low dimensional sub-space are utilized to train a Support Vector Machine (SVM) classifier to predict the gaze direction (i.e., either looking at parent's face or looking away from parent's face). The analysis of more than 39,000 video frames of naturalistic gaze shifts of multiple infants demonstrates significant agreement between a human coder and our approach. These results indicate that the proposed system provides an efficient approach to automating the estimation of gaze direction of naturalistic gaze shifts.

978-1-4244-4799-2/09/$25.00 ©2009 IEEE

1. Introduction

Human face-to-face communication plays an important role in behavioral science and developmental psychology [24]. Gaze direction is an important visual channel used by humans in face-to-face communication [10, 27]. Most of the time, an eye tracker device is utilized for estimation of the gaze direction. Some of the eye trackers rely on intrusive techniques such as measuring the reflection of light, (e.g., infra red) that is shone onto the eye. Reflected light is sensed by a video camera or other optical sensors and analyzed to find the eye rotation [9, 10]. In other types of eye trackers, the corneal reflection called Purkinje image and the center of the pupil are typically used as features to track the gaze direction over time. There are also other techniques based on the electrical potentials (Electro-Oculogram, EOG) measured with contact electrodes placed near the eyes [9]. All of these techniques are intrusive and require a controlled condition to function properly. In a live face-to-face communication, utilizing intrusive techniques is not feasible. Capturing videos from participants during a dyadic interaction is more popular and comfortable. Therefore, developing non-intrusive techniques that directly measure the gaze direction from visual data becomes essential.

Eye tracking from visual data has attracted the interests of many researchers. Several computer vision approaches have been developed for tracking eyes from images [27]. These approaches can be classified into two categories: 1) model-based approaches and 2) holistic-based approaches [9, 27]. In the first category, usually, a geometrical model for the eye (i.e., the iris contour, the eyeball, and the pupil) is proposed. Then, the geometrical model is used to interpret the gaze direction in a given eye image. For example, Daugman [14, 16] developed a simple algorithm that performs a coarse-to-fine search for a circular contour in the image corresponding to the limbus, and then searches for the pupil. Similarly, the authors of [3] presented an approach called "one-circle" algorithm for measuring the eye gaze using a monocular image that zooms only on one eye. The geometric-based techniques require calibration and high resolution images of the eye such that the iris contour and pupil are visible.

In the second category, instead of modeling the geometry of the eye, the entire eye image is used to measure the gaze direction. These approaches are similar to appearance-based object detection and recognition methods. For example, in [15] the eye images are used as inputs to a neural network. The neural network is trained by requiring the user to look at a given point on a computer monitor and subsequently capturing an image of the eye as it looks at the given point. Similar methods are reported by Kar-Han et al. in [15]. They presented an appearance-based method for estimating eye gaze direction by employing an appearance manifold. Their approach is capable of estimating gaze with accuracy comparable to that obtained by commercial devices. Appearance-based approaches have the advantage of being easy to implement and do not require calibration for every subject.

Some of the gaze estimation studies assume that a very strong correlation exists between the gaze direction, defining people's visual focus of attention (VFOA), and their head pose. Motivated by this assumption, research has been conducted to use computer vision techniques to estimate VFOA from head pose in the case where gaze direction cannot be estimated directly from the eyes [1, 20, 21]. Ba et al. [1] recently proposed a model for recognizing the visual focus of attention (VFOA) of seated people from their head pose and contextual activity cues. Their model comprises the VFOA of a meeting participant as the hidden state, and his head pose as the observation. To account for the presence of moving visual targets due to the dynamic nature of the meeting, the locations of the visual targets were used as input variables to the head pose observation model. Contextual information is introduced in the VFOA dynamics through a slide activity variable and speaking or visual activity variables that relate peoples focus to the meeting activity context. They used five hours of videos to evaluate their approach. Their results show that the propose model is effective in VFOA estimation.

The focus of this paper is to develop a holistic-based framework to estimate gaze direction. The gaze direction of an infant in a face-to-face interaction is classified as either 1) looking at the parent's face or 2) looking away from the parent's face. In our approach, we track facial images in captured videos using an Active Appearance Model (AAM). The AAM consists of a shape component and an appearance component that jointly represent the shape and texture variability seen in the object. We utilize the shape component to register (warp) the facial images to a mean facial image. The resulting facial image is a shape- and pose-normalized facial image. We then segment the eye region (eye patch) from the normalized facial image. The eye patch is obtained by cropping the facial image utilizing the mesh nodes that surround the eye region as a boundary. We make use of the appearance component of the eye patch as our representation for estimating gaze direction. Although, the appearance component is a useful representation for estimating gaze direction, it possesses an extremely large dimensionality. For instance, an eye patch with a size of 30×100 pixels has a dimensionality of 3,000 in the image space. Despite the huge dimensionality of the visual data, events such as gaze shifting have low dimensions embedded in a large dimensional space. Traditional techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis have limited ability of reducing the dimensionality of complex nonlinear gaze shift data. Recently, several nonlinear data reduction techniques such as Isomap [23], Locally Linear Embedding [22], and Laplacian Eigenmap [19] have been presented for dimensionality reduction. Laplacian Eigenmap and its variants such as Laplacianface [12] and Orthogonal Locally Linear Embedding [6] have shown promising results in face recognition [12] and age estimation [11] from facial images. We are inspired by [7] and adopt the spectral regression technique to learn projection functions that map AAM representations into a sub-space termed the gaze direction subspace. Reduced feature points presented in the sub-space are employed to estimate gaze direction based on a Support Vector Machine (SVM) classifier.

The remainder of this paper is organized as follows. An eye region representation based on the Active Appearance Model is introduced in Section 2. Section 3 describes our approach to data dimensionality reduction. Section 4 reviews the Support Vector Machine classifier employed for estimating gaze direction. Section 5 shows the experimental results and conclusions. Lastly, future work is discussed in Section 6.

2. Eye Region Representation: Active Appearance Model

Determining an adequate eye image representation for effectively estimating gaze direction is a challenging problem. One of the challenges encountered is to track the eye region across each frame of a video sequence. AAM is a proven method for tracking facial features reliably over a series of video frames [8, 18]. In this section, we review the AAM and describe the AAM-based eye region representation exploited in this work.

AAM is a statistical representation of an object (e.g., face) introduced by Cootes et al. [8] and improved by others [18] over the past few years. AAM consists of a shape 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 the PCA method to 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 (mesh nodes) match the mean shape using the thin-plate spline algorithm [5]. 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 shape-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.

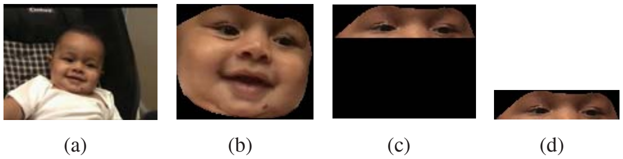

Figure 1. Eye patch appearance model: (a) A sample facial image along with the AAM component, (b) normalized-appearance face image, (c) eye region contained within the facial image, and (d) normalized-appearance eye patch

The shape-normalized appearance component contained within the eye region (eye patch), $\mathbf{g}_{eyes} \in \mathbf{g}$, in vectorized form is utilized as a representation to estimate gaze direction. The eye patch is obtained by cropping the facial image utilizing the mesh nodes that surround the eye region as a boundary. The segmentation boundary is constructed by linking the mesh nodes that lie along the eyebrows with the line formed by the nodes at the temples. Figure 1 illustrates a sample facial image decomposed into the normalized-appearance eye patch. Due to the intra-subject variations in facial appearance, we normalize each feature vector by subtracting from a neutral feature vector obtained from the subject: $X_{normalized} = X_{gaze} - X_{neutral}$. The neutral feature vector corresponds to the subject's facial image consisting of a frontal gaze direction. The $X_{normalized}$ feature vector, which is known as a delta feature, represents the displacement in facial appearance and decreases the biasing effect of intra-subject variation on gaze direction estimation. This representation is employed to estimate gaze direction. However, due to the curse of dimensionality of delta features (i.e., 3,000 dimensions, corresponding to an eye patch of 30 × 100 pixels), the classification of gaze direction is difficult. Therefore, reducing the dimensionality of the visual data becomes vital and is addressed in the following section.

3. Gaze Direction 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^N, into a low dimensional space, R^n ($n << N$) [25]. Linear techniques have limited ability to represent complex nonlinear data such as gaze shifts in a low dimensional sub-space. Recently developed nonlinear dimensionality reduction techniques such as Isomap [23], Laplacian Eigenmap [19], and Locally Linear Embedding [22] 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 [12] and face aging recognition [11]. In this paper, we employ the Laplacian Eigenmap followed by the spectral regression technique [7] to eye patches (i.e., the vectorized appearance feature) into a gaze direction sub-space by learning a projection matrix. For the remainder of this section, we review the regularized locality preserving indexing (LPI) technique [7] via spectral regression.

Recently, Cai et al. [7] presented the regularized locality preservation algorithm which has demonstrated success in representing large dimensional data in a low dimensional sub-space. Similarly to LPI, the regularized locality preserving algorithm aims 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 sub-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:

$$
\begin{aligned}
Y &= \underset{YDY^T=1}{\arg\min} \sum_{i=1}^{k}\sum_{j=1}^{k}(y_i - y_j)^2 W_{ij} \\
&= \underset{YDY^T=1}{\arg\min}\ YLY^T
\end{aligned}
\tag{3}
$$

   where similarity matrix W is constructed by finding the k nearest neighbor points using the Euclidean norm in R^N 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, $W_{ij} = \exp^{-\frac{||x_i - x_j||^2}{t}}$ [19]. 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. In this paper, the supervised method is used to calculate W. 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 the scaling factor of the solution.

   The optimization problem 3, which is also known as Laplacian Eigenmap [19], can be solved efficiently by calculating the eigenvectors of the generalized eigenproblem $LY = \lambda DY$.

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

$$
\mathbf{a} = \arg\min_{\mathbf{a}}\{\sum_{i=1}^{k}(\mathbf{a}^T x_i - y_i)^2 + \alpha||\mathbf{a}||^2\} \tag{4}
$$

   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 [4] followed by least squares regression. We use this approach to learn a projection matrix, **a**, to represent each eye patch in a low dimensional gaze direction sub-space. We have successfully used this technique in measuring the intensity of facial expressions (i.e. Action Units 6 and 12 described by the Facial Action Coding System) [17].

4. Gaze Direction Classification

After the projection of the $X_{normalized}$ features into the gaze direction sub-space, the features are utilized to classify the gaze direction as either looking at the parent's face or away from the parent's face. This is a binary (two-class) classification problem and we employ an SVM classifier to solve this problem.

SVMs have been successfully used in the field of machine learning and pattern recognition. A linear binary SVM classifier is determined by two parallel hyper-planes separating the margin between two classes. We refer our reader to [26] 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 $K(x, y) = \phi(x).\phi(y)$ 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 gaze estimation. Our experiments indicate that the RBF kernel has the highest performance in classifying gaze direction.

5. Experiments and Results

Studying the gaze shift patterns of an infant in an early face-to-face communication is a topic of interest in developmental psychology. The framework developed in this paper was applied to automatically classify the gaze direction of infants from a single camera that only captures infant's face. We classify the infant's gaze direction as either looking towards the parent's face or away from the parent's face.

This study included a subset (eight subjects) of a large infantparent dyads face-to-face study [13]. Infants participated in this study were six-month-old. The dyadic interaction was videotaped with a camera directed at the infant's face (used for automatic gaze measurement), a camera directed at the parent's face, and a camera that captured both infant and parent interacting (used as ground truth for gaze coding). Videos were recorded simultaneously at 30 fps using the three cameras. Infants are placed in an infant seat bolted to a table so that the infant is at the eye level of the mother who is seated in a position in front of the infant. The procedure is a three minute naturalistic face-to-face interaction ('play as you normally would at home') in which mothers are free to move their faces as they will, but typically move their faces toward and away, and up and down with respect to the infant's face.

The videos captured from the infant's camera were used in this paper for automatic gaze measurement. All of the captured frames of infants' videos were used in our experiment except for those where the face was occluded by the infant's hand or foot or where the eyes of the infant were not visible due to severe head pose. More than 39,000 frames were used in our experiments.

An expert coder manually coded the gaze direction (i.e., looking at the parent's face or looking away from the parent's face) occurring in each video frame acquired from the camera that captured both infant and parent interacting. The manual coding is then employed for both the training and testing of our system.

The infant's facial videos were tracked and modeled using the AAM algorithm provided by [18], and the delta features were extracted for every video frame of all eight infants. The $X_{normalized}$ features of every 15^{th} frame were used to learn the projection matrix **a** based on spectral regression for constructing the gaze direction sub-space. W in Eq. 3 was calculated using the supervised method in this paper. The dimension of the gaze sub-space is 28 which corresponds to the smallest eigenvalues construct matrix Y in Eq. 3.

Our experiments are based on Leave-One-Subject-Out cross validation to predict the gaze direction. SVM training was performed on every m^{th}, $m = 5, 7, 10, 20, 25, 30$, and 50, video frame of seven subjects 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 gaze directions, we calculate the Cohen's kappa coefficient [2], which is a statistical measure of inter-rater agreement and is defined as:

$$\kappa = \frac{Pr(o) - \Pr(e)}{1 - Pr(e)} \tag{5}$$

where $Pr(o)$ is the probability of observed agreement and $Pr(e)$ is the probability of random agreement. κ ranges between 0 and 1. Cohen's kappa is known to be more robust than simple percentage agreement since it takes into account the chance of random agreement. As a rule of thumb, κ between .4 and .6 is regarded as fair agreement, between .6 to .75 as good agreement and above as excellent agreement [2].

Table 1 illustrates both the κ coefficient and the percentage agreement between the actual and predicted gaze directions using the SVM classifiers that were calculated for eight infants under study (the results of this table is based on $m = 5$). This table demonstrates that our approach has good performance in classifying the gaze direction. The average κ coefficient and percentage agreement between our technique and a human coder are .79 and 91%, respectively. The average κ coefficient and percentage agreement between two human coders are .75 and 90%, respectively.

A human coder has also coded the gaze direction using only the videos captured from the infant's camera (the same videos that were used by our automated system). The average κ coefficient and percentage agreement between the human coder and the automated system using infants' videos in coding the gaze direction were .61 and 84.23%, respectively (this result is only based on the available codes for seven subjects excluding subject one). Obviously, this result shows that our automated system outperforms a human in coding the gaze direction from the infants' video.

Sub.	Cohen's Kappa	Percentage Agreement(%)
1	.68	82.3
2	.90	95.0
3	.63	88.0
4	.78	91.5
5	.89	95.9
6	.54	87.7
7	.68	84.0
8	.92	96.5
Average	.79	91

Table 1. Cohen's Kappa coefficient between the actual and predicted gaze direction data calculated for the eight subjects; Percentage agreement is also presented in this table. 20% of the data ($m = 5$) was used to train the SVM classifier.

To justify the use of the nonlinear data reduction technique employed in this work (spectral regression), we compare its performance to the traditional linear data reduction technique, PCA. The κ coefficient and percentage agreement ($m = 5$) achieved by substituting the data reduction component of our system with PCA were .55 and 80%, respectively. Clearly, the PCA method is outperformed by the spectral regression technique, which yielded a Kappa and percentage agreement of .79 and 91%, respectively.

Figure 2 shows the effect that the number of training frames has on the accuracy of the system when classifying the gaze direction. As the figure illustrates, even by using a small training set (2% of the frames, $m = 50$), the system demonstrates a high agreement with the human coder (the average k coefficient and percentage agreement are .77 and 90%, respectively).

6. Conclusions and Future work

In this paper, we presented a framework for estimating the gaze direction of naturalistic gaze shifts in eye patches. We utilized the concept of Regularized Locality Preservation via spectral regression to reduce the dimensionality of the eye patches modeled by the AAM. Our approach was employed to estimate the gaze direction of infants in a live face-to-face communication. The statistical agreement (i.e., the Cohen's kappa coefficient and percentage agreement) between a human coder and our system in estimating the gaze direction of naturalistic gaze shifts is significantly high.

References

[1] S. O. Ba, H. Hung, and J.-M. Odobez. Visual activity context for focus of attention estimation in dynamic meetings.

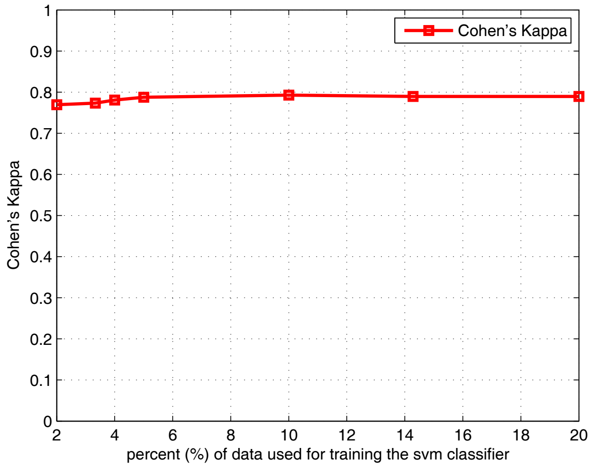

Figure 2. Cohen's Kappa coefficient versus percentage of data used in training the SVM classifiers.

Idiap-RR Idiap-RR-02-2009, Idiap, rue marconi 19, 1920, martigny switzerland, January 2009.

[2] R. Bakeman. Behavioral observations and coding. In H. T. Reis and C. K. Judd, editors, *Handbook of research methods in social psychology*, pages 138–159. New York: Cambridge University Press, 2000.

[3] S. Baluja and D. Pomerleau. Non-intrusive gaze tracking using artificial neural networks. Technical report, CMU CS Technical Report CMU-CS-94-102, 1994.

[4] 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.

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

[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.

[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.

[8] 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.

[9] J. G. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:1148–1161, 1993.

[10] A. T. Duchowski. *Eye Tracking Methodology: Theory and Practice*. Springer, 2007.

[11] 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.

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

[13] L. Ibanez, D. S. Messinger, L. Newell, M. Sheskin, and B. Lambert. Visual disengagement in the infant siblings of children with an autism spectrum disorder (asd). *Autism: International Journal of Research and Practice*, 12:523535, 2008.

[14] W. Jian-Gang and E. Sung. Study on eye gaze estimation. *IEEE Transactions on Systems, Man, and Cybernetics*, 32:332–350, 2002.

[15] T. Kar-Han, D. J. Kriegman, and N. Ahuja. Appearance-based eye gaze estimation. In *Proceedings of Workshop on Applications of Computer Vision*, 2002.

[16] D. G. Lowe. Local feature view clustering for 3d object recognition. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 2001.

[17] M. H. Mahoor, S. Cadavid, D. S. Messinger, and J. F. Cohn. A framework for automated measurement of the intensity of non-posed facial action units. In *2nd IEEE Workshop on CVPR for Human communicative Behavior analysis (CVPR4HB), Miami Beach, FLorida*, 2009.

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

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

[20] J.-M. Odobez and S. Ba. A cognitive and unsupervised map adaptation approach to the recognition of focus of attention from head pose. In *ICME*, 2007.

[21] K. Otsuka, Y. Takemae, J. Yamato, and H. Murase. A probabilistic inference of multiparty-conversation structure based on markov-switching models of gaze patterns, head directions, and utterances. In *ICMI*, 2005.

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

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

[24] 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.

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

[26] 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.

[27] L. R. Young and D. Sheena. Survey of eye movement recording methods. *Behavior research methods and instrumentation*, 7(5):397–429, 1975.