Real Time Facial Expression Recognition in Video using Support Vector Machines

Philipp Michel
Computer Laboratory
University of Cambridge
Cambridge CB3 0FD, United Kingdom
pmichel@cantab.net

Rana El Kaliouby
Computer Laboratory
University of Cambridge
Cambridge CB3 0FD, United Kingdom
rana.el-kaliouby@cl.cam.ac.uk

ABSTRACT

Enabling computer systems to recognize facial expressions and infer emotions from them in real time presents a challenging research topic. In this paper, we present a real time approach to emotion recognition through facial expression in live video. We employ an automatic facial feature tracker to perform face localization and feature extraction. The facial feature displacements in the video stream are used as input to a Support Vector Machine classifier. We evaluate our method in terms of recognition accuracy for a variety of interaction and classification scenarios. Our person-dependent and person-independent experiments demonstrate the effectiveness of a support vector machine and feature tracking approach to fully automatic, unobtrusive expression recognition in live video. We conclude by discussing the relevance of our work to affective and intelligent man-machine interfaces and exploring further improvements.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces; H.1.2 [**Models and Principles**]: User/Machine Systems—*Human factors*; I.5.4 [**Pattern Recognition**]: Applications; I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis—*Tracking.*

General Terms

Design, Experimentation, Performance, Human Factors.

Keywords

Facial expression analysis, support vector machines, feature tracking, emotion classification, affective user interfaces.

1. INTRODUCTION

The human face possesses superior expressive ability [8] and provides one of the most powerful, versatile and natural

means of communicating motivational and affective state [9]. We use facial expressions not only to express our emotions, but also to provide important communicative cues during social interaction, such as our level of interest, our desire to take a speaking turn and continuous feedback signalling understanding of the information conveyed. Facial expression constitutes 55 percent of the effect of a communicated message [18] and is hence a major modality in human communication.

Reeves & Nass [21] posit that human beings are biased to attempt to evoke their highly evolved social skills when confronting a new technology capable of social interaction. The possibility of enabling systems to recognize and make use of the information conferred by facial expressions has hence gained significant research interest over the last few years. This has given rise to a number of automatic methods to recognize facial expressions in images or video [20].

In this paper, we propose a method for automatically inferring emotions by recognizing facial expressions in live video. We base our method on the machine learning system of Support Vector Machines (SVMs). A face feature tracker gathers a set of displacements from feature motion in the video stream. These are subsequently used to train an SVM classifier to recognize previously unseen expressions. The novelty of our approach consists of its performance for real time classification, its unobtrusiveness and its lack of preprocessing. It is particularly suitable for ad-hoc, incrementally trained and person-independent expression recognition.

The remainder of this paper is organized as follows. Section 2 gives an overview of some prior work in the area of facial expression analysis. In Section 3 we outline the overall design of our approach to expression recognition. Section 4 describes how face localization, feature extraction and tracking are accomplished, while Section 5 gives an overview of support vector machines and details how they are used in our approach. We evaluate our system in Section 6 and conclude in Section 7.

2. RELATED WORK

Pantic & Rothkrantz [20] identify three basic problems a facial expression analysis approach needs to deal with: face detection in a facial image or image sequence, facial expression data extraction and facial expression classification.

Most previous systems assume presence of a full frontal face view in the image or the image sequence being analyzed, yielding some knowledge of the global face location. To give the exact location of the face, Viola & Jones [27] use the Ad-

aboost algorithm to exhaustively pass a search sub-window over the image at multiple scales for rapid face detection. Essa & Pentland [13] perform spatial and temporal filtering together with thresholding to extract motion blobs from image sequences. To detect presence of a face, these blobs are then evaluated using the eigenfaces method [24] via principal component analysis (PCA) to calculate the distance of the observed region from a face space of 128 sample images. The PersonSpotter system [22] tracks the bounding box of the head in video using spatio-temporal filtering and stereo disparity in pixels due to motion, thus selecting image regions of interest. It then employs skin color and convex region detectors to check for face presence in these regions.

To perform data extraction, Littlewort et al. [17] use a bank of 40 Gabor wavelet filters at different scales and orientations to perform convolution. They thus extract a "jet" of magnitudes of complex valued responses at different locations in a lattice imposed on an image, as proposed in [16]. Essa & Pentland [13] extend their face detection approach to extract the positions of prominent facial features using eigenfeatures and PCA by calculating the distance of an image from a feature space given a set of sample images via FFT and a local energy computation. Cohn et al. [4] first manually localize feature points in the first frame of an image sequence and then use hierarchical optical flow to track the motion of small windows surrounding these points across frames. The displacement vectors for each landmark between the initial and the peak frame represent the extracted expression information.

In the final step of expression analysis, expressions are classified according to some scheme. The most prevalent approaches are based on the existence of six basic emotions (anger, disgust, fear, joy, sorrow and surprise) as argued by Ekman [10] and the Facial Action Coding System (FACS), developed by Ekman and Friesen [12], which codes expressions as a combination of 44 facial movements called Action Units. While much progress has been made in automatically classifying according to FACS [23], a fully automated FACS-based approach for video has yet to be developed. Dailey et al. [7] use a six unit, single layer neural network to classify into the six basic emotion categories given Gabor jets extracted from static images. Essa & Pentland [13] calculate ideal motion energy templates for each expression category and take the euclidean norm of the difference between the observed motion energy in a sequence of images and each motion energy template as a similarity metric. Littlewort et al. [17] preprocess image sequences image-by-image to train two stages of support vector machines from Gabor filter jets. Cohn et al. [4] apply separate discriminant functions and variance-covariance matrices to different facial regions and use feature displacements as predictors for classification.

3. IMPLEMENTATION OVERVIEW

We use a real time facial feature tracker to deal with the problems of face localization and feature extraction in spontaneous expressions. The tracker extracts the position of 22 facial features from the video stream. We calculate displacements for each feature between a neutral and a representative frame of an expression. These are used together with the label of the expression as input to the training stage of an SVM classifier. The trained SVM model is subsequently used to classify unseen feature displacements in real time, either upon user request or continuously, for every frame in
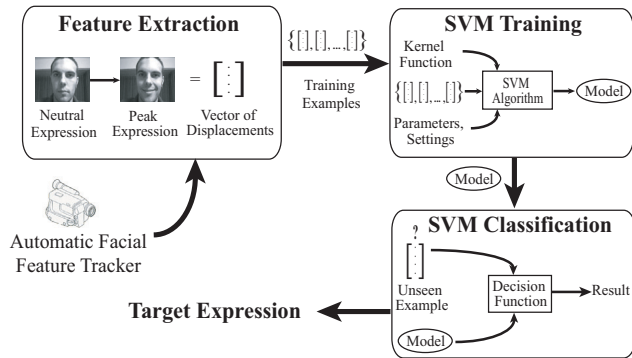

Figure 1: The stages of SVM-based automated expression recognition.

the video stream. Figure 1 illustrates the implementation diagrammatically.

Commodity hardware is used, such as a commercial digital camcorder connected to a standard PC, allowing classification at 30 frames per second. We have found our approach to perform well even at significantly lower frame rates (and hence more intermittent motion) and under a variety of lighting conditions. We assume a full frontal view of the face, but take into account the user-dependent variations in head pose and framing inherent in video-based interaction. Our approach works for arbitrary user-defined emotion classes. For illustration purposes, we use the universal basic emotions throughout the remainder of this paper.

4. FEATURE EXTRACTION & TRACKING

We chose the feature displacement approach due to its suitability for a real time video system, in which motion is inherent and which places a strict upper bound on the computational complexity of methods used in order to meet time constraints. Cohn et al. [5] posit that approaches based on feature point tracking show 92% average agreement with manual FACS coding by professionals and are hence highly applicable to expression analysis. Our earlier experiments using an SVM classifier on displacements of manually defined facial landmarks in image sequences from the Cohn-Kanade facial expression database [15] yielded high classification accuracy. Furthermore, they exhibited short training and classification delays which prompted us to investigate the application of a combined feature displacement / SVM approach for real time video.

The tracker we employ uses a face template to initially locate the position of the 22 facial features of our face model in the video stream and uses a filter to track their position over subsequent frames as shown in Figure 2. For each expression, a vector of feature displacements is calculated by taking the euclidean distance between feature locations in a neutral and a "peak" frame representative of the expression, as illustrated in Figure 3. This allows characteristic feature motion patterns to be established for each expression, as given by Figure 4. Feature locations are automatically captured when the amount of motion is at a minimum, corresponding to either the initial neutral phase or the final phase of a spontaneous expression, when motion has settled around its peak frame.
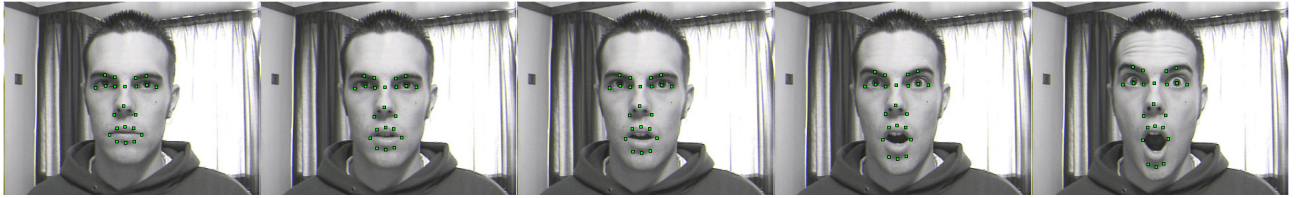
Figure 2: Facial features localized and tracked over a sequence of video frames (taken at intervals in a 30fps stream).
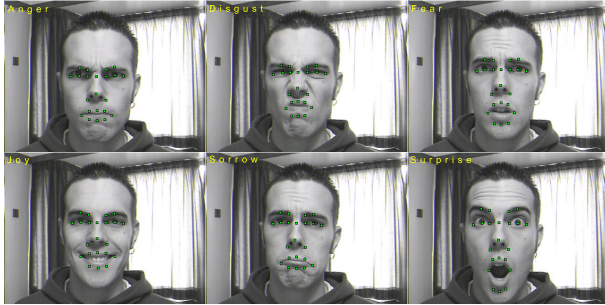

Figure 3: Peak frames for each of the six basic emotions, with features localized.

5. CLASSIFICATION

5.1 Support Vector Machine Overview

Machine learning algorithms receive input data during a training phase, build a model of the input and output a hypothesis function that can be used to predict future data. Given a set of labelled training examples

$$S = ((\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_l, y_l)), y_i \in \{-1, 1\}$$

learning systems typically try to find a decision function of the form

$$h(\mathbf{x}) = sgn(\langle \mathbf{w} \cdot \mathbf{x} \rangle + b)$$

that yields a label $\in \{-1, 1\}$ (for the basic case of binary classification) for a previously unseen example $\mathbf{x}$.

Support Vector Machines [1, 6] are based on results from statistical learning theory, pioneered by Vapnik [25, 26], instead of heuristics or analogies with natural learning systems. These results establish that the generalization performance of a learned function on future unseen data depends on the complexity of the class of functions it is chosen from rather than the complexity of the function itself. By bounding this class complexity, theoretical guarantees about the generalization performance can be made.

SVMs perform an implicit embedding of data into a high dimensional feature space, where linear algebra and geometry may be used to separate data that is only separable with nonlinear rules in input space. To do so, the learning algorithm is formulated to make use of kernel functions, allowing efficient computation of inner products directly in feature space, without need for explicit embedding. Given a nonlinear mapping $\boldsymbol{\Phi}$ that embeds input vectors into feature space, kernels have the form

$$K(\mathbf{x}, \mathbf{z}) = \langle \boldsymbol{\Phi}(\mathbf{x}) \cdot \boldsymbol{\Phi}(\mathbf{z}) \rangle$$
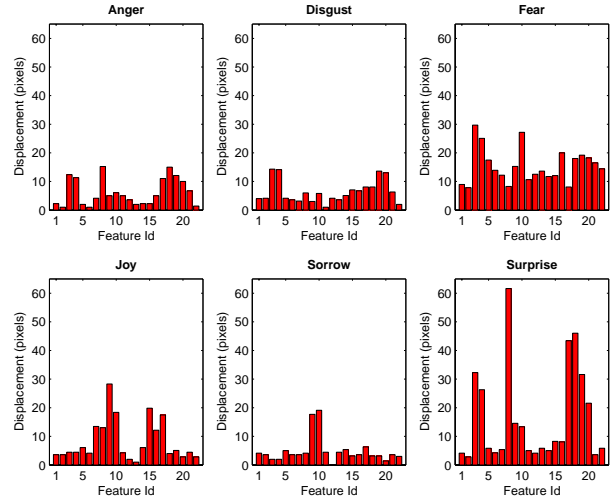

Figure 4: Feature motion patterns for the six basic emotions.

SVM algorithms separate the training data in feature space by a hyperplane defined by the type of kernel function used. They find the hyperplane of maximal margin, defined as the sum of the distances of the hyperplane from the nearest data point of each of the two classes. The size of the margin bounds the complexity of the hyperplane function and hence determines its generalization performance on unseen data. The SVM methodology learns nonlinear functions of the form:

$$f(\mathbf{x}) = sgn(\sum_{i=1}^{l} \alpha_i y_i K(\mathbf{x}_i \cdot \mathbf{x}) + b)$$

where the α_i are Lagrange multipliers of a dual optimization problem. It is possible to show that only some of the α_i are non-zero in the optimal solution, namely those arising from training points nearest to the hyperplane, called support vectors. These induce sparseness in the solution and give rise to efficient approaches to optimization. Once a decision function is obtained, classification of an unseen example $\mathbf{x}$ amounts to checking on what side of the hyperplane the example lies.

The SVM approach is highly modular, allowing domain-specific selection of the kernel function used. Table 1 gives the kernels used during our evaluation of SVM-based expression classification. In contrast to previous "black box" learning approaches, SVMs allow for some intuition and human understanding. They deal with noisy data and overfitting (where the learned function perfectly explains the training

Kernel	Formula
Linear	$\mathbf{x} \cdot \mathbf{z}$
Polynomial	$(\gamma \mathbf{x} \cdot \mathbf{z} + c)^{degree}$
Radial Basis Function (RBF)	$\exp(-\gamma|\mathbf{x} - \mathbf{z}|^2)$
Sigmoid	$\tanh(\gamma \mathbf{x} \cdot \mathbf{z} + c)$

Table 1: Typically used kernel functions. $c, \gamma, degree$ are parameters used to define each particular kernel from the family given.

set but generalizes poorly) by allowing for some misclassifications on the training set. This handles data that is linearly inseparable even in higher space. Multi-class classification is accomplished by a cascade of binary classifiers together with a voting scheme. SVMs have been successfully employed for a number of classification tasks such as text categorization [14], genetic analysis [28] and face detection [19]. They currently outperform artificial neural networks in a variety of applications. Their high classification accuracy for small training sets and their generalization performance on data that is highly variable and difficult to separate make SVMs particularly suitable to a real time approach to expression recognition in video. They perform well on data that is noisy due to pose variation, lighting, etc. and where often minute differences distinguish expressions corresponding to entirely different emotions.

5.2 SVM Classification for Real Time Video

Our implementation uses `libsvm` [2] as the underlying SVM classifier. We encapsulate its stateless functionality in an object-oriented manner to work in an incrementally trained interactive environment. This avoids having to supply the set of training examples in its entirety before any classification can proceed and allows the user to augment the training data on-the-fly. It also enables us to export the entire state of a trained SVM classifier for later use. Hence, data gathered across several training sessions is preserved and can be re-used for classification. In addition, it allows for convenient combination of training data from multiple individuals to accomplish person-independent classification.

The user requests for training examples to be gathered at discrete time intervals and provides a label for each. This is combined with the displacements output by the feature extraction phase and added as a new example to the training set. The SVM is then retrained. Unseen expressions to be classified pass the same feature extraction process and are subsequently assigned the label of the target expression that most closely matches their displacement pattern by the SVM classifier.

Most computational overhead resides in the training phase. However, due to the fact that the training set is interactively created by the user and hence limited in magnitude and that the individual training examples are of constant and small size, overhead is low for typical training runs. This is also aided by the sparseness of the SVM solution, manifested by the fact that the number of support vectors which define the decision surface only increases sub-linearly as more examples are added to the training data. This is illustrated by Figure 5. Because evaluation of an SVM decision function on unseen input essentially amounts to checking which of the two subspaces defined by a separating hyperplane a point lies in, classification overhead is negligible. This allows our
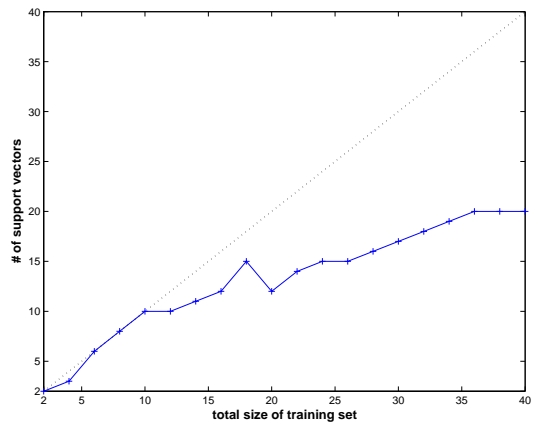
Figure 5: Number of support vectors defining the decision surface as training set size is increased.

approach to perform classification both directly upon user request and continuously in real time for every frame in the video stream, with the current result being constantly reported back to the user.

6. EVALUATION

We evaluate our system by considering classification performance for the six basic emotions. Our approach makes no assumptions about the specific emotions used for training or classification and works for arbitrary, user-defined emotion categories.

To establish an upper bound on the recognition accuracy achievable for a combined feature displacement / SVM approach to expression recognition, we initially evaluated our method on still images from the Cohn-Kanade facial expression database. Features were manually defined for each image and displacements were subsequently extracted from pairs of images consisting of a neutral and a representative frame for each expression. A set of ten examples for each basic emotion was used for training, followed by classification of 15 unseen examples per emotion. We used the standard SVM classification algorithm together with a linear kernel. Table 2 gives the percentage of correctly classified examples per basic emotion and the overall recognition accuracy. Figure 6 shows the mean and standard deviation of the feature displacements for each basic emotion as extracted from the training data. It can be seen from these results that each emotion, expressed in terms of feature motion, varies widely across subjects. Particular emotions (eg. fear) or indeed particular combinations (eg. disgust vs. fear) are inherently harder to distinguish than others. However, the results also expose certain motion properties of expressions which seem to be universal across subjects (eg. raised mouth corners for 'joy', corresponding to a smile).

Kernel choice is among the most important customizations that can be made when adjusting an SVM classifier to a particular application domain. We experimented with a range of polynomial, gaussian radial basis function (RBF) and sigmoid kernels and found RBF kernels to significantly outperform the others, boosting overall recognition accuracy on the still image data to 87.9%. The human 'ceiling' in correctly classifying facial expressions into the six ba-

Emotion	Percent correct
Anger	82.2%
Disgust	84.6%
Fear	71.7%
Joy	93.0%
Sorrow	85.4%
Surprise	99.3%
Average	**86.0%**

Table 2: Recognition accuracy of SVM classification on displacements extracted from still images.
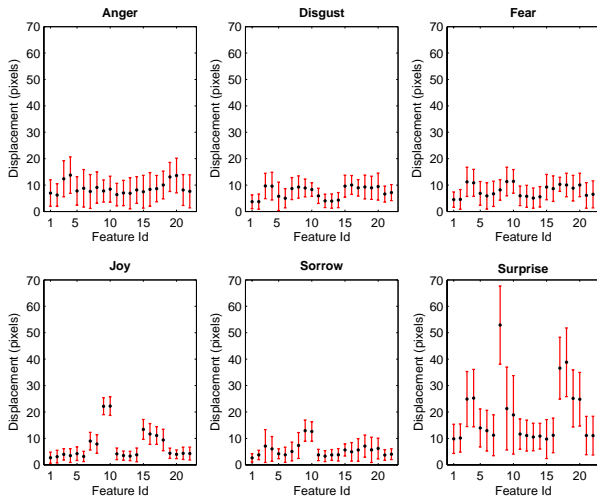

Figure 6: The mean and standard deviation for the characteristic motion of the 22 facial features for each basic emotion.

sic emotions has been established at 91.7% by Ekman & Friesen [11].

We next established the optimum recognition accuracy for video-based spontaneous expressions by having an expert user familiar with our approach and aware of how the basic emotions are typically expressed provide ten examples per basic emotion. A set of 72 test examples (12 per basic emotion) were then SVM classified using an RBF kernel. Table 3 gives the confusion matrix for these trials. It can be seen that for an expert user enthusiastically expressing emotions relatively little penalty in accuracy is incurred when moving from images to video. We are able to customize our tracker to a particular person by calibrating the face template used to initially locate facial features. This further increases recognition performance slightly and allows a tradeoff between generality and precision to be made.

We then evaluated our approach for the most challenging but ultimately desirable scenario of "ad-hoc" interaction, whereby wholly inexperienced users are asked to express emotions naturally in an unconstrained setup in terms of lighting conditions, distance from the camera, pose and so forth. Six individuals were asked to provide one training example per basic emotion and subsequently supply a number of unseen examples for classification. The resulting confusion matrix is shown in Table 4. The significantly lower accuracy achieved in these trials can mainly be attributed to the often significant amount of head motion by

Emotion	Percent correct
Anger	66.7%
Disgust	64.3%
Fear	66.7%
Joy	91.7%
Sorrow	62.5%
Surprise	83.3%
Average	**71.8%**

Table 5: Recognition accuracy for person-independent classification from video.

the subjects during feature extraction, resulting in inaccurate displacement measurements. In addition, since subjects were not given any instructions on how to express each basic emotion, cross-subject agreement on what constituted, for instance, an expression of fear was considerably lower, resulting in much more variable and difficult to separate data.

Finally, we tested the ability to recognize expressions of an expert user given only training data supplied by a different expert user, essentially person-independent classification. One example per basic emotion was supplied for training and 12 unseen examples were classified. This trial was repeated for various suppliers of training and testing data. Table 5 gives the recognition accuracy. Recognition was best for emotions having more uniform expression across subjects (eg. joy or surprise). Trials established that the more expressive the person supplying the training data, the higher the accuracy that can be achieved. Furthermore, we found it to be crucial that training and classification be performed under similar conditions, particularly with regard to the distance from the camera.

7. CONCLUSION AND DISCUSSION

In this paper we presented an approach to expression recognition in live video. Our results indicate that the properties of a Support Vector Machine learning system correlate well with the constraints placed on recognition accuracy and speed by a real time environment. We evaluated our system in terms of accuracy for a variety of interaction scenarios and found the results for controlled experiments to compare favorably to previous approaches to expression recognition. Furthermore, usable results are achieved even for very unconstrained ad-hoc interaction, which we consider a crucial prerequisite for the application of expression recognition in novel multimodal interfaces.

We currently deal with the prevalent problem of inaccuracies introduced by head motion by normalizing all feature displacements with respect to a single fiducial feature (the nose root) which provides an approximation to head motion in the video stream. This does not take into account rotational motion such as nodding or head tilting. We expect an additional increase in recognition accuracy if feature displacements can be properly adjusted in the presence of such motion. Moreover, using automatic SVM model selection [3] to determine optimal parameters of the classifier for displacement-based facial expression data should also increase classification accuracy further.

Incorporating emotive information in computer-human interfaces will allow for much more natural and efficient interaction paradigms to be established. It is our belief that methods for emotion recognition through facial expression

Emotion	Anger	Disgust	Fear	Joy	Sorrow	Surprise	*Overall*
Anger	**10**	0	0	0	0	0	83.3%
Disgust	0	**12**	0	0	0	0	100.0%
Fear	2	0	**10**	0	0	0	83.3%
Joy	0	0	0	**9**	0	3	75.0%
Sorrow	0	0	2	0	**10**	0	83.3%
Surprise	0	0	0	0	0	**12**	100.0%

Total accuracy: 87.5%

Table 3: Person-dependent confusion matrix for training and test data supplied by an expert user.

Emotion	Anger	Disgust	Fear	Joy	Sorrow	Surprise	*Overall*
Anger	**19**	4	3	1	3	2	59.4%
Disgust	2	**18**	1	2	5	3	58.1%
Fear	7	3	**15**	0	3	1	51.7%
Joy	1	5	1	**21**	4	1	63.6%
Sorrow	2	3	7	3	**18**	0	54.4%
Surprise	2	3	1	0	2	**25**	75.8%

Total accuracy: 60.7%

Table 4: Person-dependent confusion matrix for training and test data supplied by six users during ad-hoc interaction.

that work on real time video without preprocessing while remaining cheap in terms of equipment and unobtrusive for the user will play an increasing role in building such affective and intelligent multimodal interfaces.

8. REFERENCES

[1] B. Boser, I. Guyon, and V. Vapnik. A training algorithm for optimal margin classifiers. In D. Haussler, editor, *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pages 144–152, 1992.

[2] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[3] O. Chapelle and V. Vapnik. Model selection for support vector machines. In S. Solla, T. Leen, and K.-R. Müller, editors, *Advances in Neural Information Processing Systems*, volume 12, pages 230–236. MIT Press, Cambridge, MA, USA, 2000.

[4] J. F. Cohn, A. J. Zlochower, J. J. Lien, and T. Kanade. Feature-point tracking by optical flow discriminates subtle differences in facial expression. In *Proceedings International Conference on Automatic Face and Gesture Recognition*, pages 396–401, 1998.

[5] J. F. Cohn, A. J. Zlochower, J. J. Lien, and T. Kanade. Automated face analysis by feature point tracking has high concurrent validity with manual faces coding. *Psychophysiology*, 36:35–43, 1999.

[6] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and other Kernel-based Learning Methods.* Cambridge University Press, Cambridge, UK, 2000.

[7] M. N. Dailey, G. W. Cottrell, and R. Adolphs. A six-unit network is all you need to discover happiness. In *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society*, Mahwah, NJ, USA, 2000. Erlbaum.

[8] C. Darwin. *The Expression of the Emotions in Man and Animals.* John Murray, London, UK, 1872.

[9] P. Ekman. *Emotion in the Human Face.* Cambridge University Press, Cambridge, UK, 1982.

[10] P. Ekman. Basic emotions. In T. Dalgleish and T. Power, editors, *The Handbook of Cognition and Emotion.* John Wiley & Sons, Ltd., 1999.

[11] P. Ekman and W. Friesen. *Pictures of Facial Affect.* Consulting Psychologists Press, Palo Alto, CA, USA, 1976.

[12] P. Ekman and W. Friesen. *Facial Action Coding System (FACS): Manual.* Consulting Psychologists Press, Palo Alto, CA, USA, 1978.

[13] I. Essa and A. Pentland. Coding, analysis, interpretation and recognition of facial expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):757–763, 1997.

[14] T. Joachims. Text categorization with support vector machines: Learning with many relevant features. In *Proceedings of ECML-98, 10th European Conference on Machine Learning*, pages 137–142, Heidelberg, DE, 1998. Springer Verlag.

[15] T. Kanade, J. Cohn, and Y. 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, 2000.

[16] M. Lades, J. C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Würtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42:300–311, 1993.

[17] G. Littlewort, I. Fasel, M. Stewart Bartlett, and J. R. Movellan. Fully automatic coding of basic expressions from video. Technical Report 2002.03, UCSD INC MPLab, 2002.

[18] A. Mehrabian. Communication without words. *Psychology Today*, 2(4):53–56, 1968.

[19] E. Osuna, R. Freund, and F. Girosi. Training support vector machines: An application to face detection. In

Proceedings of Computer Vision and Pattern Recognition '97, pages 130–36, 1997.

[20] M. Pantic and L. J. M. Rothkrantz. Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, 2000.

[21] B. Reeves and C. Nass. *The Media Equation: How People treat Computers, Television, and New Media like Real People and Places.* Cambridge University Press, New York, NY, USA, 1996.

[22] J. Steffens, E. Elagin, and H. Neven. Personspotter—fast and robust system for human detection, tracking, and recognition. In *Proceedings International Conference on Automatic Face and Gesture Recognition*, pages 516–521, 1998.

[23] Y.-L. Tian, T. Kanade, and J. F. Cohn. Recognizing action units for facial expression analysis.

IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(2):97–115, 2001.

[24] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[25] V. N. Vapnik. *The Nature of Statistical Learning Theory.* Springer-Verlag New York, Inc., New York, NY, USA, 1995.

[26] V. N. Vapnik. *Statistical Learning Theory.* Wiley, New York, NY, USA, 1998.

[27] P. Viola and M. Jones. Robust real-time object detection. Technical Report 2001/01, Compaq Cambridge Research Lab, 2001.

[28] A. Zien, G. Rätsch, S. Mika, B. Schölkopf, T. Lengauer, and K. Müller. Engineering support vector machine kernels that recognize translation initiation sites. *Bioinformatics*, 16(9):799–807, 2000.