# Recognition of Facial Expressions in the Presence of Occlusion

Fabrice Bourel
Claude C. Chibelushi
Adrian A. Low
School of Computing
Staffordshire University
Beaconside
Stafford ST18 0DG
F.Bourel@staffs.ac.uk
C.C.Chibelushi@staffs.ac.uk
A.A.Low@staffs.ac.uk

**Abstract**

We present a new approach for the recognition of facial expressions from video sequences in the presence of occlusion. Although promising results have been reported in the literature on automatic recognition of facial expressions, most techniques have been assessed using experiments performed in controlled laboratory conditions which do not reflect real-world conditions. The goal of the work presented herein, is to develop recognition techniques that will overcome some limitations of current techniques, such as their sensitivity to partial occlusion of the face. The proposed approach is based on a localised representation of facial features, and on data fusion. The experiments show that the proposed approach is robust to partial occlusion of the face.

## 1 Introduction

The automatic recognition of facial expressions has been studied with much interest in the past 10 years [9, 22, 24]. An important potential application of automatic facial expression recognition is human computer interaction [19, 30]. It may also be used as a tool in human behavioral research [22]. This paper presents a new approach for recognising facial expressions, from image sequences, in the presence of partial occlusion. The approach is based on a localised representation of facial expression features, and on fusion of classifier outputs. Data fusion or sensor fusion has been used successfully in many fields such as pattern recognition [6, 7, 25] or distributed computing [4, 5]. The ability to handle occluded facial features is important for achieving robust recognition. Despite its importance, the effect of occlusion on automatic facial expression recognition has not received much attention from the research community. As indicated by Pantic [22], no facial expression recognition techniques can currently handle partial occlusion. To the



BMVC 2001 doi:10.5244/C.15.23

knowledge of the authors, the effect of the latter on recognition accuracy has not been assessed either. Causes for occlusion may be facial hair, glasses, hands or other objects.

By the nature of their structures, facial models using local facial information extraction have an advantage over models that consider the whole face. For instance, when a part of the face is occluded, performing optical flow on the whole face [12] may cause some problems in representing the extracted data for later classification. Approaches using a deformable model fitted to the whole face [16, 15] may be adapted to handle missing features, if the structure of the model and the classifier both handle partial data in a generic way. Face models based on feature points and a local geometric representation of the face [27, 28] can be adapted easily for achieving robustness to partial occlusion. The latter will cause the loss of some model parameters but the parameters not affected by occlusion may be still used for classification. The classifier also has to be able to handle partial input data and may need to be retrained.

Although sensitivity to occlusion is still an unresolved problem, many techniques for facial expression recognition have been reported. Black and Yacoob [2] use a local parameterized model of image motion from optical flow analysis. They utilize a planar model for rigid facial motion and an affine-plus-curvature model for non rigid motion. Otsuka and Ohya [21] estimate the motion of the right eye and the mouth by optical flow. Then, they perform a 2D wavelet transform to extract low-frequency feature vectors. Essa and Pentland [12] first locate the nose, eyes and mouth. Then, from two consecutive normalised frames, a 2D spatio-temporal motion energy representation of facial motion is used as a dynamic face model. Kimura and Yachida [15] fit what they call a 'potential net' on the face. Then, applying a Gaussian filter to an edge enhanced image, they determine a 'potential field'. The net is then deformed by the elastic force of the potential field. Information encoded in the nodes of the net is used for classification. Wang *et al.* [28] use 19 facial feature points. The face is represented as a labeled graph model where facial feature points are treated as interconnected nodes. Graphs are compared to templates for recognition using a minimum-distance classification technique. Cohn *et al.* [8] use feature points that are automatically tracked using a hierarchical optical flow method [18]. The feature vectors, used for the recognition, are created by calculating the displacement of the facial points. The displacement is obtained by subtracting its normalised position in the first frame from its current normalised position. Following Cohn's approach, Lien *et al.* [17] perform three methods: dense optical flow, feature point tracking and high gradient edge extraction. Then, the vectors from each extraction steps are vector quantized and used for classification. Tian *et al.* [27] recently proposed a feature-based method using geometric facial features. The extracted facial regions (mouth, eyes, brows and cheeks) are represented with geometric parameters. The furrows are also detected using a Canny edge detector to measure orientation and quantify their intensity. The geometric parameters are then fed into a neural network for recognition.

Following the facial feature extraction is the classification step. Six basic expression classes (anger, disgust, fear, joy, sadness, surprise), defined by Ekman [10], are often used. Expression analysers can classify the encountered expression using the FACS system [11]. The FACS system defines 46 action units (AUs) for describing facial expressions. AUs are anatomically related to contraction of specific muscles. AUs can be classified either as a single or combination of AUs [2, 23, 8, 17, 27]. Most approaches for the classification of expressions from image sequences are based on neural networks [23, 27], rule-based models [2, 29], template based-methods [28, 20, 12, 15] using minimum-distance classifi-

cation, statistical models using discriminant function analysis [8], or probabilistic models such as hidden Markov Models [17, 21].

In the context of robust recognition of occluded expressions, using a modular classification and fusion approach may have several advantages over monolithic classification [25]. In the data fusion approach, failure of one or several classifiers does not necessarily degrade the classification process. New classification modules can also be appended easily when other regions of the face are to be examined. Monolithic classifiers are sensitive to incomplete input data, as would occur in case of partial occlusion of the face.

This paper presents and assesses a data fusion approach aiming to achieve robustness to partial occlusion. The paper is organised as follow: Section 2 gives an overview of the system. Section 3 and Section 4 describe respectively the spatio-temporal model used to represent the expressions, and the classification and data fusion framework. Section 5 performs experiments on the effect of occlusion and discusses the results. We conclude in Section 6 by summarising the findings and suggesting future research directions.

# 2 Overview of the approach

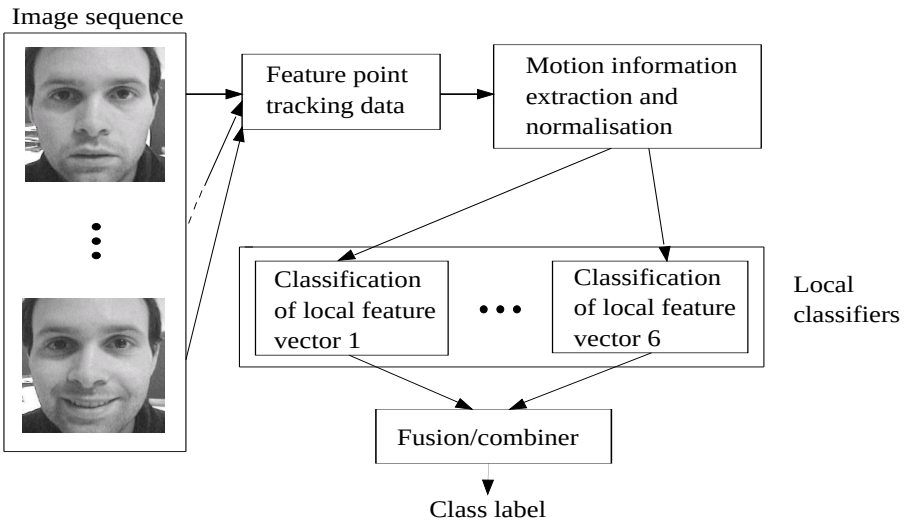


*Figure 1:* Architecture of the recognition system.

The framework for automatic recognition of facial expressions, shown in Figure 1, consists of a feature point tracker, a feature extractor, a group of local classifiers, and a fusion module. The feature extractor processes 12 facial feature points which are tracked over an entire image sequence by the robust tracker described in [3] (Figure 2). The 12 facial feature points are set manually on the first frame of the image sequence.

As described in [3], the tracker, based on the Kanade-Lucas tracker [18], is capable of following and recovering any of the 12 facial points lost due to lighting variations, rigid or non-rigid motion, or (to a certain extent) change of head orientation. Automatic

recovery, which uses the nostrils as a reference, is performed based on some heuristics exploiting the configuration and visual properties of faces. It was shown in [3] that the enhanced tracker outperforms the Kanade-Lucas tracker, in terms of recovering lost or drifting points.

When the tracking is over, independent local spatio-temporal vectors, are created from geometrical relations between the feature points (Figure 3). These vectors contain the motion information relative to the start of each sequence. Dynamic information is used in order to achieve shape independence. Shape independence is sought so that the effect of person identity on expression recognition is minimised. In particular, it has been shown in speaker recognition that lip shape conveys identity information [6].

Classification is applied to local facial regions (mouth and eyebrows) instead of global facial information, and the local regions are classified independently. This way, occlusion of a local region will not affect the classification of the other regions. Decision-level fusion [25] of classifier outputs is used. For the purpose of this paper, 4 classes of expressions are used out of the 6 defined in [10]; they are the facial expressions for 'anger', 'joy', 'sadness' and 'surprise'.

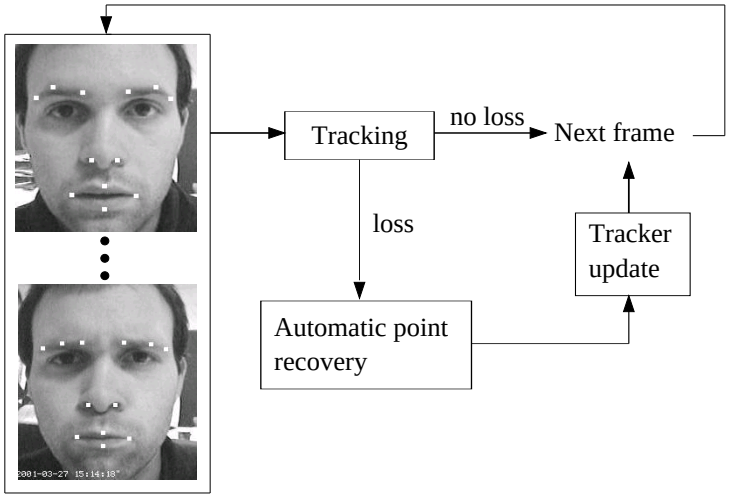


*Figure 2:* Automatic feature point tracking scheme.

# 3 Spatio-temporal representation of expressions

Facial point tracking is followed by a determination of the start and endpoints of the dynamic part of the sequence for each local region. The neutral end-segments are then discarded. Thereafter, the duration of the sequence is normalised to approximately 2.66 seconds by linear interpolation. At 30 frames per second, this corresponds to 80 video frames. The normalisation is applied independently to each local region. As a result of duration normalisation, two expressions performed in a different amount of time will become similar on the normalised temporal scale. Figure 4 illustrates the removal of the

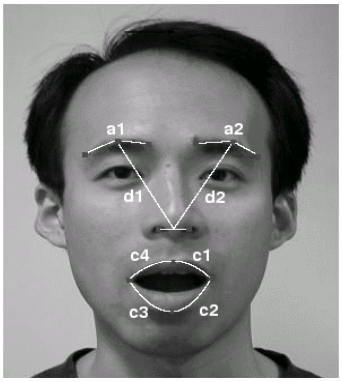


*Figure 3:* Facial model used to produce the six local feature vectors.

neutral segments at the beginning and end of a sequence.

The mouth model is defined using the four feature points shown in Figure 2. From these four points, we extract four parabolic coefficients, which model the shape of the mouth in each frame (Figure 3). The upper face model also consists of four coefficients. Two coefficients model the shape of the eyebrows by measuring their angle of deformation (Figure 3). The other two coefficients measure the distance from the eyebrows to the nostrils (Figure 3). From the eight coefficients, a set of six local feature vectors are created: $V_{(c1,c4)}, V_{(c2,c3)}, V_{a1}, V_{a2}, V_{d1}, V_{d2}$. The components of the feature vectors are created by taking the difference between the coefficients for each frame, and the average of the first set of coefficients representing the segment removed from the beginning of the sequence.

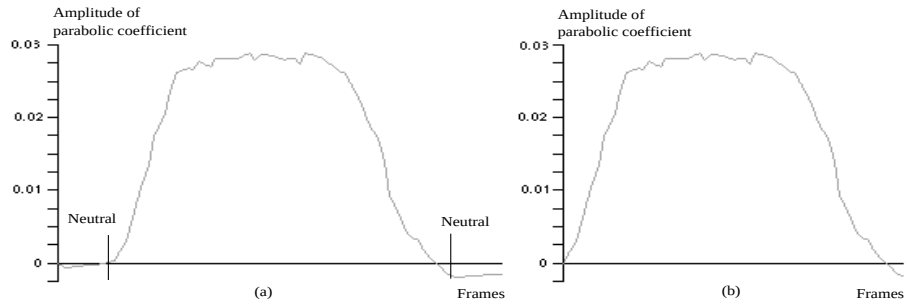


*Figure 4:* Removal of the neutral segments at the start and end of a motion sequence. (a) Neutral segments detected. (b) Neutral segments removed.

# 4 Facial expression classification

Data fusion has been used by many researchers to improve the accuracy of pattern classification [6], when several sources of information are independent or complementary. In our system, each local expert classifier outputs a recognition score for each known expression. The scores are then combined to produce a final classification result (Figure 5).

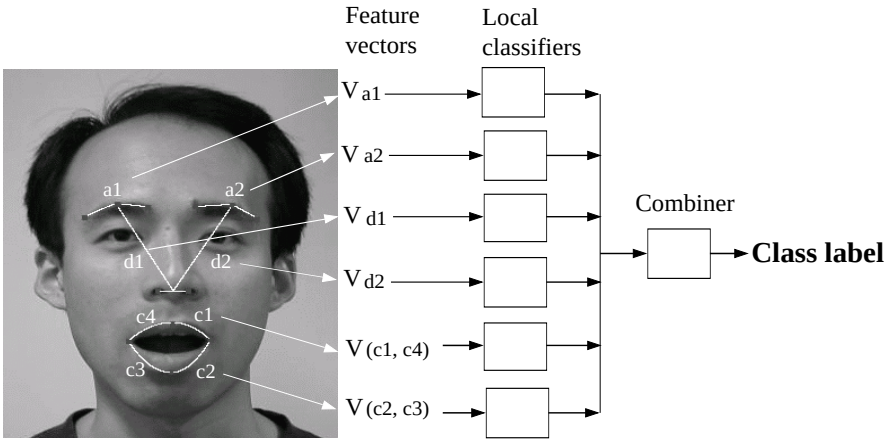


*Figure 5:* Expression classification using decision-level fusion.

## 4.1 Local classifiers

Each local classifier is a rank-weighted $k$-nearest-neighbour classifier. It produces a weighted cumulative score for each known class of expression. The weighted class score is proportional to the rank of each nearest neighbour belonging to the class. This is an adaptation to the distance-weighted $k$NN classifier [1]. Therefore, the class of the first nearest neighbour will have the highest score, then the class of the second a lower score and so on until we reach $k$ neighbours. For instance, if $k = 4$, the class of the first nearest neighbour receives a score of 4, the class of the second neighbour 3, until the fifth one and the rest which receive a score of 0.

## 4.2 Combining scheme

The combination is implemented as a *sum* scheme performed at the decision level. It is used here for its simplicity. When all local classifiers are combined, all the classes of expression get a final score by summing the class-specific score from each local classifier. The unknown pattern is assigned to the class which has the highest score. Other approaches to decision fusion such as linear combination, Bayesian inference, neural networks, fuzzy logic are the subject of our ongoing investigations [6, 25, 13].

# 5 Experimental results

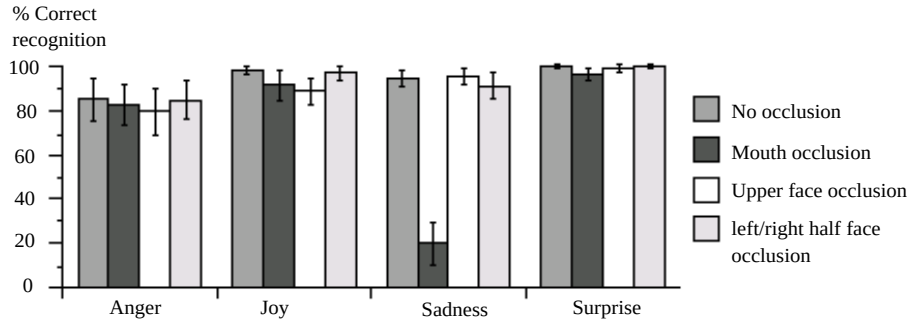


*Figure 6:* Effect of partial occlusion on facial expression recognition. Error bars show the 99% confidence intervals.

The aim of these experiments is to study the robustness, of the proposed recognition approach, to the occlusion of face regions. The experiments have been carried out using 100 video sequences, where 25 sequences are available for each expression class. Partitioning into test and training sets is based on the leave-one-out test methodology. The confidence limits are estimated under the assumption that recognition rates have of a binomial distribution [26]. The 99% confidence levels produce, in the worst case, confidence intervals of $\pm 10.3\%$ for the test sample size and values of recognition rates measured. The image sequences are from the Cohn-Kanade Action-Unit-coded Carnegie Mellon University facial expression database [14]. The 30 subjects from the database vary in ethnicities, age and skin color.

Tests with no occlusion of any facial features serve as baseline. The effect of occlusion is investigated by simulating missing facial regions. The test set covers occlusion of the mouth, upper face, or left/right half of the face. The motivation for these tests are many. For example, when somebody performs an expression such as a smile, it may happen that, he puts his hand or an object in front of his mouth. Facial hair or glasses can also occlude face regions.

The results are shown in Figure 6. Recognition of ‘anger’ and ‘surprise’ expressions is robust to the occlusion of the mouth. This is as expected because some discriminative information is also embedded in the upper part of the face. Surprisingly, ‘joy’ is accurate despite the absence of the mouth, which is important when making a smile. This unexpected result could be due to virtual lack of motion of the eyebrows during a smile. This sets ‘joy’ apart from the other expression classes under study. However, a problem may arise if a neutral expression is considered as a valid class for recognition. The feature vectors, for ‘joy’ and the neutral expression, would be similar. The discriminative features for the ‘sadness’ expression are mainly conveyed by the mouth, as illustrated by the sensitivity to mouth occlusion of the recognition of ‘sadness’.

Both left/right face sides were found to give the same results on average. They are therefore represented as one single set in Figure 6. The recognition results under occlusion are very close to the baseline results corresponding to no occlusion. This shows that reducing the amount of facial information by half still retains sufficient discriminative

information for expression recognition. However, removing the data corresponding to half of the face does not necessarily mean that we remove redundant data. Indeed, expressions can be unilateral and bilateral [11].

Overall, the experimental results show that the proposed approach is robust to partial occlusion of the face. However, analysing expressions using geometric features misses the regional information embedded in image intensity, texture, or edges. Hence, ongoing investigations are studying the use of both geometric features and texture. The effect on robustness to partial occlusion, of the granularity of the subdivision of the face into local regions, is also under study.

# 6 Conclusion

A method for recognizing facial expressions in the presence of occlusion has been presented. Facial feature points are automatically tracked over an image sequences, and used to represent a face model made of several regions of interest. Decision-level fusion combines local interpretation of the face model into a global recognition score. It has been experimentally shown that the proposed approach is robust to partial occlusion of the face. It has also been observed that there is an interaction between recognition robustness for some expressions and what region of the face is occluded.

# References


[1] T. Bailey and A. K. Jain. A Note on Distance-Weighted k-Nearest Neighbor Rules. *IEEE Transactions on Systems, Man, and Cybernetics*, 8:311–313, 1978.

[2] M. J. Black and Y. Yacoob. Recognizing Facial Expressions in Image Sequences Using Local Parameterized Models of Image Motion. *International Journal on Computer Vision*, 25(1):23–48, October 1997.

[3] F. Bourel, C. C. Chibelushi, and A. A. Low. Robust Facial Feature Tracking. *Proc. 11th British Machine Vision Conference, Bristol, England*, 1:232–241, September 2000.

[4] R. R. Brooks and S. S. Iyengar. Robust Distributed Computing and Sensing Algorithm. *IEEE Computer*, 29(6):53–60, June 1996.

[5] R. R. Brooks and S. S. Iyengar. *Multi-Sensor Fusion: Fundamentals and Applications with Software*. Prentice Hall, 1998.

[6] C. C. Chibelushi, F. Deravi, and S. D. Mason. A Review of Speech-Based Bimodal Recognition. *IEEE Transactions on Multimedia, Accepted for publication*.

[7] C. C. Chibelushi, F. Deravi, and S. D. Mason. Adaptive Classifier Integration for Robust Pattern Recognition. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, 29(6):902–907, December 1999.

[8] J. Cohn, A. Zlochower, J. Lien, and T. Kanade. Feature-point tracking by optical flow discriminates subtle differences in facial Expression. *Proc. 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pages 396–401, 1998.

[9] G. Donato, M. S. Barlett, J. C. Hager, P. Ekman, and T. J. Sejnowski. Classifying Facial Actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(10):974–989, October 1999.

[10] P. Ekman and W. Friesen. *Unmasking the Face*. Prentice Hall, New Jersey, 1975.

[11] P. Ekman and W. Friesen. The Facial Action Coding System. *Consulting Psychologists Press, San Francisco, CA*, 1978.

[12] I. A. Essa and A. P. Pentland. Coding, Analysis, Interpretation and Recognition of Facial Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):757–763, July 1997.

[13] A. K. Jain, R. P. W. Duin, and J. Mao. Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, January 2000.

[14] T. Kanade, J. Cohn, and Y. Tian. Comprehensive Database for Facial Expression Analysis. *Proc. 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'2000), France*, pages 46–53, March 2000.

[15] S. Kimura and M. Yachida. Facial Expression Recognition and its Degree Estimation. *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pages 295–300, 1997.

[16] A. Lanitis, C. J. Taylor, and T. F. Cootes. Automatic Interpretation and Coding of Face Images Using Flexible Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):743–756, July 1997.

[17] J. J. Lien, T. Kanade, J. F. Cohn, and C. Li. Detection, Tracking, and Classification of Action Units in Facial Expression. *Journal of Robotics and Autonomous System*, 31:131–146, 2000.

[18] B. D. Lucas and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. *Proc. International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.

[19] B. Myers, J. Hollan, and I. Cruz. Strategic Directions in Human-Computer Interaction. *ACM Computing surveys*, 28(4):795–809, December 1996.

[20] N. Oliver, A. P. Pentland, and F. Berard. LAFTER: Lips and Face Real Time Tracker with Facial Expression Recognition. *Proc. Computer Vision and Pattern Recognition Conference, S.Juan, Puerto Rico*, pages 123–129, June 1997.

[21] T. Otsuka and J. Ohya. Recognition of Facial Expressions Using HMM with Continuous Output Probabilities. *5th IEEE International Workshop on Robot and Human Communication*, pages 323–328, 1996.

[22] M. Pantic and Rothkrantz L. J. M. Automatic Analysis of Facial Expressions: The State of the Art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1424–1445, December 2000.

[23] M. Rosenblum, L. S. Davis, and Y. Yacoob. Human Emotion Recognition from Motion Using a Radial Basis Function Network Architecture. *IEEE Workshop on Motion of Non-Rigid and Articulated Objects, Austin, Texas*, pages 43–49, November 1994.

[24] A Samal and A. P. Iyengar. Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey. *Pattern Recognition*, 25(1):65–77, January 1992.

[25] M. Sarkar. Modular Pattern Classifiers: A Brief Survey. *Proc. IEEE Conference on System, Man, and Cybernetics*, 4:2878–2883, 2000.

[26] M. R. Spiegel. *Schaum's Outline of Theory and Problems of Statistics, 2nd Edition*. McGraw-Hill Inc., 1990.

[27] Y. Tian, T. Kanade, and J. F. Cohn. Recognizing Action Units for Facial Expression Analysis. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 23(2):97–115, February 2001.

[28] M. Wang, Y. Iwai, and M. Yachida. Expression Recognition from Time-Sequential Facial Images by Use of Expression Change Model. *Proc. 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pages 324–329, 1998.

[29] Y. Yacoob and L. S. Davis. Recognizing Human Facial Expressions from Long Image Sequences Using Optical-Flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(6):636–642, June 1996.

[30] J. Ziegler. Interactive Techniques. *ACM Computing surveys*, 28(1):185–187, March 1996.