ICSP2012 Proceedings

Effective Geometric Features for Human Emotion Recognition

Anwar Saeed, Ayoub Al-Hamadi, Robert Niese , and Moftah Elzobi
Institute for Electronics, Signal Processing and Communications (IESK)
Otto-von-Guericke-University Magdeburg
D-39016 Magdeburg, P.O. Box 4210 Germany
{Anwar.Saeed, Ayoub.Al-Hamadi}@ovgu.de

***Abstract*—Human face carries variety of useful information. For example, person's emotion, behavior, and pain can be perceived from his facial expressions. In this paper, we make full use of eight fiducial facial points to extract geometric features used after that to infer the universal human emotions (happy, surprise, anger, disgust, fear, and sadness). We compared our results with results obtained by two different algorithms, representing the state of the art, on two separated databases. We show using features from eight facial points, our approach performs as well as an algorithm that utilizes features extracted from 68 fiducial facial points and as well as another algorithm that uses hundreds of texture features.**

***Index Terms*—Facial Expressions, Fiducial Facial Points, Support Vector Machine.**

I. Introduction

Human face is acting as an indication display for human mental states. Earlier studies have found that each human emotion has different impacts on the face expressions. Consequently, the recognition of these emotions from their impacts on the face is an active research topic in the computer vision over the last two decades. Inferring the human emotion will lead to improve the Human-Computer Interaction HCI to be as good as the Human-Human Interaction. Therefore, building reliable emotionally aware machines is achievable. These machines would be helpful in various fields as recognizing pain and depression, health support appliances, monitoring fatigue, and providing feedback on the machine tutoring system. The interpretation of the human emotions and their mechanism was done by models from the psychology of emotion (e.g., basic emotions according to Ekman, and Circumplex models [1], [2]. Ekman et al. broke the facial expressions down into smaller action units (AU). Each AU codes small visible changes in the facial muscles. Then, the coexistence of different AUs is considered as an evidence of the occurrence of human emotions. For example, in the happy emotion, the AU12 (*lip Corner Puller*) should be occurred; in anger emotion, AU23 (*Lip Tighner)* and AU24 (*Lip Pressor*) should be presented; in disgust emotion, the AU9 (*Nose Wrinkler*) or AU10 (*Upper Lip Raiser*) must be presented; in fear emotion, a combination of AU1 (*Inner Brow Raiser*), AU2 (*Outer Brow Raiser*) and AU4 (*Brow Lowerer*) must be existed; in surprise emotion, either AU1+2 or lower intensity of AU5 (*Upper Lip Raiser)* must be present, and similarly, in sadness emotion, a combination of either AU1, AU4 and AU15 (*Lip Corner Depressor*) or AU11 (*Nasolabial Deepener*) must be present. Additionally, some AUs should appear with varied intensities in different emotions [3]. Aside from mapping the AUs to the emotion recognition, several approaches recognize human emotions with the help of features extracted directly from the face region [3], [4].

Many algorithms were developed to recognize human emotions based on facial expressions. These algorithms can be divided into two distinct categories. The first category requires prior knowledge of the neutral emotion state. For example, in this category we detect and track fiducial facial points over an image sequence and utilize the changes in the points location or the behavior of their movements to recognize the human emotions [3], [4]. Several proposals were suggested to identify the neutral state:

- having a general face model which supposed to best fit the majority of human faces,
- each person has his own neutral state, which picked up in an initial registration step,
- the features at neutral state is an average of the facial features for a long-period based on the assumption that emotional expressions spread just over few frames.

The second category does not depend on the neutral emotion state. It is kind of holistic (spanning the whole face) approach; in this category we pass the detected human face image through several filters (e.g. Gabor filters, Local Binary Patterns (LBP)) and use the filters output as features passed to learning algorithms for the emotion recognition [5], [6]. The algorithm developed by Niese et al. [4] is an example of the first category. They used two types of features: geometric features extracted from 3-D points; dynamic features came from optical flow values that are extracted from specific regions associated to 3-D model of each subject. As an example of the second category, Littlewort et al. [6] convolved registered detected face with a filter bank of 72 Gabor filters with eight orientations and nine spatial frequencies. All those features are the input to individual Support Vector Machine (SVM) classifiers for each AU. Finally, they built Multivariate Logistic Regression classifier (MLR) on top of the output of AU classifiers to recognize

978-1-4673-2197-6/12/$31.00 ©2012 IEEE

Fig. 1: The used geometric features. (a) Human face in neutral state overlaid with 68 fiducial facial points [3]. (b) The selected eight facial points. (c) Human face in the surprise state overlaid with geometric features extracted from the selected points. (d) The six geometric features, f_1 and f_2 are the average of two mirrored values on the left and right sides of the face. (Images from Cohn-Kanade Dataset (CK+) database, ©Jeffrey Cohn.)

human emotions. Littlewort et al. [6] published a demo called CERT, which gives probability values for eight emotions. In this paper, we use geometric features to recognize human emotions. These features represent the changes in location of eight fiducial facial points. We do not rely on 3-D user specific information. Our geometric features are extracted from 2-D posed face images. Our results compared to those of CERT and Lucey et al. [3] approaches are presented in Sec.III.

The remainder of the paper is structured as follows. In section II, we describe our proposed algorithm. Experimental results, which are conducted on two different databases, are discussed in section III. Additionally, our results compared to results from two other algorithms will be presented in this section. Finally, the conclusion and future perspectives are given in section IV.

II. The Proposed Algorithm

Our approach is largely inspired by recent face analysis of Martinez et al. [7]. They argued that robust computer vision algorithms for face analysis and recognition should be based on configural and shape features. These features are defined as the distance between facial components (mouth, eye, eyebrow, nose, and jaw line). In our work, we rely on features representing the distances among (mouth, eye, and eyebrow). These components are sampled with a set of eight fiducial points (P_s). We use two points for the eyebrows (p_{reb},p_{leb}); two points for the eye's corner (p_{rec},p_{lec}); four points for the mouth ($p_{rm}, p_{lm}, p_{upm}, p_{lom}$). This facial point set P_s is 2D points,

$$P_s = \{p_{reb}, p_{leb}, p_{rec}, p_{lec}, p_{rm}, p_{lm}, p_{upm}, p_{lom}\}, \quad p_i \in \mathbb{R}^2.$$

We extracted six geometric features from the eight fiducial facial points. These features represent the changes in the face expressions during emotion occurrence. Then, we passed the features to SVM classifiers for emotion recognition.

A. Feature Extraction

In contrast to Lucey et al. [3] who extracted features from 68 fiducial facial points, we extracted our features just from eight points, as shown in Fig. 1. We have found that the different emotional facial expressions can be modeled distinctively by the change in the location of these eight points. Moreover, detecting and tracking eight points over image sequence is less time-consuming. Additionally, these points represent corner points that can be accurately detected [8], [9]. The feature vector **f** Eq. (1) includes the ratio of distances between fiducial points measured in action apex to that at the neutral emotion state (see Fig. 1.d).

$$\mathbf{f} = (f_1, f_2, f_3, f_4, f_5, f_6). \tag{1}$$

f_1 and f_2 are the average of two mirrored values on the left and right sides of the face. These features implicitly include some AUs defined by Ekman et al. [1]. For example, f_1 carries information about the eyebrow AU1 (*Inner Brow Raiser*) and AU4 (*Brow Lowerer*); f_2 indicates the movements of the upper lip AU10 (*Upper Lip Raiser*). The other four features cover some of the AUs related to the mouth. To our knowledge, f_5 and f_6 are used for the first time here. They are more useful for the recognition of sadness emotion. However, they (f_4, f_5) look like redundant of f_1, they are not. The effect of (f_4, f_5) on overall recognition rate is discussed in Sec.III-C. To remove the dominant effect of the large valued feature, before passing the features into a machine learning algorithm, the feature vector **f** (1) is normalized to be $\tilde{\mathbf{f}} = (\tilde{f_1}, \tilde{f_2}, \tilde{f_3}, \tilde{f_4}, \tilde{f_5}, \tilde{f_6})$ as follows.

$$\tilde{f_i} = \frac{\frac{f_i - \mu_i}{2\sigma_i} + 1}{2}, \qquad i = 1, ..., 6. \tag{2}$$

Where μ_i and σ_i are mean and standard deviation of the ith feature across the training data, respectively. If we assume f_i is normally distributed, Eq. (2) guarantees 95% of $\tilde{f_i}$ to be in the [0,1] range. Then, we truncate the out-of-range components to either 0 or 1.

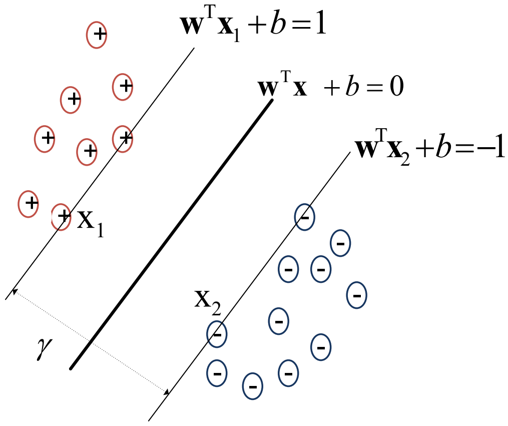


Fig. 2: The separating hyperplane ($\mathbf{w}^T\mathbf{x} + b = 0$) splits $\mathbf{x}_i$ vectors into two classes, vectors are labeled with $\mathbf{y} = 1$ in one side and $\mathbf{y} = -1$ on the other side. $\mathbf{x}_1$ and $\mathbf{x}_2$ are samples of support vectors of opposite sign. The canonical hyperplanes pass through the support vectors. The region between them is the margin band γ.

B. The machine learning algorithm

Our emotion recognition is formulated as a multi-class learning process. We assigned one class to each emotion. Among several supervised learning algorithms, we used the support vector machine (SVM) due to its well-known generalization capability. In the case of a binary classification task with training data $\mathbf{x}_i$ ($i = 1, ..., N$), having corresponding classes $y_i = \pm 1$, the decision function could be formulated as:

$$f(\mathbf{x}) = \mathrm{sign}(\mathbf{w}^T\mathbf{x} + b), \qquad (3)$$

where $\mathbf{w}^T\mathbf{x} + b = 0$ denotes a separating hyperplane, b is the bias or offset of the hyperplane from the origin in input space, and $\mathbf{w}$ is a weight vector normal to the separating hyperplane. There are two hyperplanes, called canonical hyperplanes, passing through support vectors ($\mathbf{x}_1, \mathbf{x}_2$) and satisfying $\mathbf{w}^T\mathbf{x}_1 + b = 1$ and $\mathbf{w}^T\mathbf{x}_2 + b = -1$, respectively as shown in Fig. 2. The region between the canonical hyperplanes is the margin band and is given as

$$\gamma = \frac{2}{\|\mathbf{w}\|},$$

where $\|\mathbf{w}\|$ is the square root of $\mathbf{w}^T\mathbf{w}$. Finally, choosing the optimal values $(\mathbf{w}, b)$ is formulated as a constrained optimization problem. We maximize the margin band (4) subject to the following constraints:

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \qquad \forall i. \qquad (4)$$

Several one-vs-all SVM classifiers are incorporated to handle multi-class emotion recognition. For this purpose, we employed LIBSVM [10].

III. Experimental results

To assess the reliability of our method, we compared our results with those obtained by two different algorithms. The first algorithm relies on features extracted from 68 fiducial points [3]. The second algorithm employs hundreds of texture features. These features are extracted by passing detected face through a filter bank of 72 Gabor filters with eight orientations and nine spatial frequencies [6]. The experiments were carried out on two different human emotion databases.

A. The Extended Cohn-Kanade Dataset (CK+) [3]

This database contains 593 sequences across 123 subjects. Each image sequence starts from onset (neutral state) and ends with a peak expression (last frame). The offered peak expression is fully coded by the Facial Action Coding System using FACS investigator guide. Additionally, the emotion labels were validated through perceptual judgment. As results of the aforementioned validation, only 327 of the sequences were labeled for the human emotions: 45 for anger; 18 for contempt; 59 for disgust; 25 for fear; 69 for happy; 28 for sadness; 83 for surprise. Keyframes within each image sequence were manually labeled with 68 points, and after that a gradient descent Active Appearance Model (AAM) is used to fit these points in the remaining frames. Lucey et. al. extracted two types of features from 68 facial points: similarity-normalized shape (SPTS) and canonical appearance (CAPP) features. In this work, we use 8 points out of the offered 68 points to extract geometric features (see Sec. II-A). Due to the lack of training samples, we both employed a leave-one-out cross-validation strategy. As the name suggests, all the samples were used in the training except the one used as a test sample. Table I shows our results compared to Lucey et. al. published results for the six basic emotions. We both achieved very good recognition rates for the (happy, surprise, and disgust emotions) due to their strong facial expressions. On the other hand, our approach achieved lower recognition rates for (anger and fear emotions) due to their subtle facial expressions. In contrast with Lucey et. al [3] approach, we attained very good recognition rate for sadness emotion. This improvement is achieved with the help of the features (f_5 and f_6), as will be cleared in Sec.III-C. The anger emotion was confused with sadness as well with disgust. This confusion is due to its weak face deformation. A Combination with texture features would be suggested here. As shown in Table I, we achieved an average recognition rate of 87.48% compared to 83.15% achieved by Lucey et. al [3], taken into consideration that removing contempt emotion from their classification algorithm can lead to an improve in their result. On summary, features from eight fiducial points provide good results as well as that taken from 68 points.

B. Binghamton University 3D Facial Expression Database (BU-4DFE) [11]

From this 3-D database, 2-D frontal face image sequences were extracted. After that, the eight fiducial points were detected at the first frame (neutral emotion state) with help

	Anger	Disgust	Fear	Happy	Sadness	Surprise	Contempt
Anger	75.6	13.3	0.00	0.00	11.1	0.00	-
	75.0	7.50	5.00	0.00	5.00	2.50	5.00
Disgust	8.40	88.2	1.69	0.00	1.69	0.00	-
	5.30	94.7	0.00	0.00	0.00	0.00	0.00
Fear	0.00	0.00	76.00	16.00	8.00	0.00	-
	4.40	0.00	65.2	8.70	0.00	13.0	8.70
Happy	0.00	0.00	2.90	97.1	0.00	0.00	-
	0.00	0.00	0.00	100	0.00	0.00	0.00
Sadness	3.57	3.57	3.57	0.00	89.3	0.00	-
	12.0	4.00	4.00	0.00	68.0	4.00	8.00
Surprise	0.00	0.00	1.30	0.00	0.00	98.7	-
	0.00	0.00	0.00	0.00	4.00	96.0	0.00

TABLE I: Confusion matrix of human emotion recognition for the CK+ database. The first row in each emotion represents our results. The other row shows the results of Lucy et al., as reported in [3].

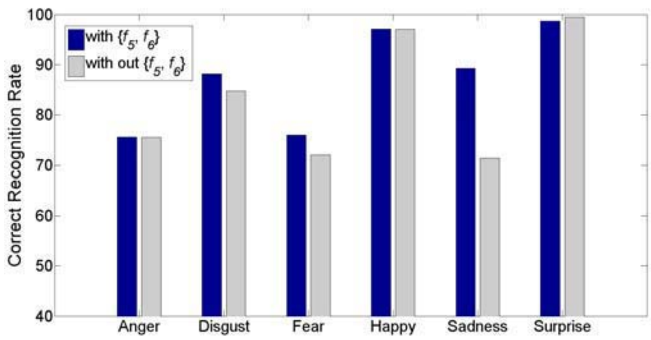
	Anger	Disgust	Fear	Happy	Sadness	Surprise
Anger	85.2	5.65	0.00	0.00	9.15	0.00
	57.0	31.7	0.00	0.00	9.86	1.44
Disgust	7.19	80.3	4.49	3.57	2.67	1.78
	12.5	73.2	2.69	1.79	9.82	0.00
Fear	0.87	2.54	64.4	16.94	9.32	5.93
	5.93	11.0	57.6	5.97	19.5	0.00
Happy	0.60	1.33	3.65	92.6	1.82	0.00
	0.00	1.82	1.82	92.7	1.22	2.44
Sadness	11.9	0.00	3.20	0.00	84.9	0.00
	7.94	1.66	0.00	0.00	90.4	0.00
Surprise	0.00	0.00	2.96	1.14	0.00	95.9
	2.83	4.02	20.7	1.15	12.1	59.2

TABLE II: Confusion matrix of human emotion recognition for the BU-4DFE database. The first row in each emotion represents our results. The other row shows the results obtained by CERT [6].

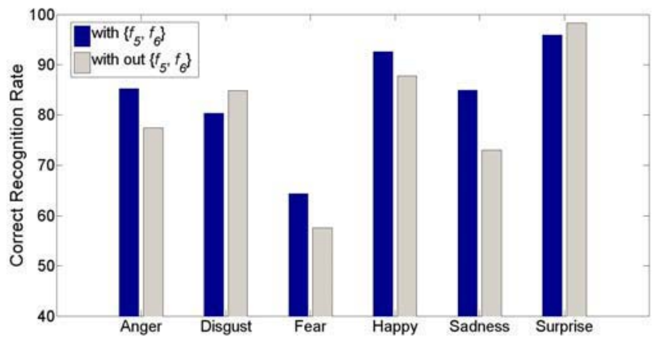
of Valstar et al. approach [9], and then tracked using a dense optical flow tracking algorithm [12] in the rest sequence. As in Sec.III-A, we used here also a leave-one-out cross-validation strategy. We compared our results with results obtained from CERT demo [6] (see Sec.I). For each facial image, CERT provides probability values for eight human emotions (happy, surprise, anger, disgust, fear, sadness, contempt, and neutral). The neutral and contempt emotions are not included in our algorithm. Hence for meaningful comparison, we decided on the emotion from the six basic emotions with higher probability to be the recognized one, even if the probability value of contempt or/and neutral emotion is higher. The CERT was trained with different databases, similarly our algorithm never used a test sample in the training, as a leave-one-out cross-validation strategy suggests. Final remarks concerning the difference between CERT and our approach, CERT has an advantage over our algorithm that it does not rely on information from neutral state. However, this leads to big confusion between anger and neural state as well as between sadness and neutral, as reported by Littlewort et al. [6]. We did not utilize any information about the head rotation (yaw, pitch, and roll) as they did. A performance comparison between both approaches can not be built, since they are trained on different databases and our dependence on prior knowledge of the neutral state. We list the results to show the state-of-art of using texture features for facial emotion recognition. Our results compared to their results on BU-4DFE database are presented in Table II. We achieved a high recognition rate for happy, surprise, anger, and disgust emotions. In contrast with our evaluation in CK+ database, we achieved lower correct recognition rate for fear emotion. The CERT shows very good results for happy, sadness, and disgust emotions recognition. On the other hand, CERT achieved lower recognition rate for surprise and anger emotions. The highly correct recognition rate of sadness by CERT illustrates the importance of texture features. However, a small confusion between anger and sadness as well as between fear and sadness is also unavoidable even we use texture features. The use of geometric approaches results in a confusion between (happy and fear) emotions. This confusion is less with the help of texture features as the results of CERT show. We achieved an average correct recognition rate of 83.88% compared to 71.68% achieved by CERT, which indicates that geometric features extracted from 8 fiducial points are performing as well as hundreds features extracted from Gabor filters distributed among the face image. We can improve our approach performance by incorporating texture features.

C. *Feature Analysis*

However, the distance between the upper and lower mouth facial points (p_{upm}, p_{lom}) is divided into two distances used afterwards to generate features (f_5, f_6), the newly generated features behave differently with each human emotion expression. For example, pulling down of lip corners can be easily detected with help of (f_5, f_6). This action is crucial for sadness emotion recognition. To evaluate the usefulness of (f_5, f_6), we recalculated the confusion matrices (shown in Tables I and II) using just features (f_1, f_2, f_3, f_4). We have found that the use of (f_5, f_6) improves the total correct recognition rate from 83.46% to 87.48% for CK+ database; and from 79.83% to 83.88% for BU-4DFE database. The detailed improvement in the correct recognition rates for each emotion are depicted in Fig.3. The use of f_5 and f_6 caused a significant increase in the correct recognition rate of sadness emotion (17.9% for CK+ and 11.9% for BU-4DFE) at cost of small reduction in the correct rate of other emotions (0.8% in surprise emotion for CK+; and 4.5% in disgust, 2.3% in surprise for BU-4DFE). To illustrate our proposed feature in an intuitive way, we depict the first two principal components of feature vectors in Fig.4. As follows from Fig.4, the features discriminatively represent (surprise, happy) emotions. On the other hand, a confusion is expected between the other emotions, mainly between anger and sadness.



(a) Correct recognition rate for CK+ database.



(b) Correct recognition rate for BU-4DFE database.

Fig. 3: A comparison of the emotion correct recognition rates showing the improvements due to the use of f_5 and f_6.

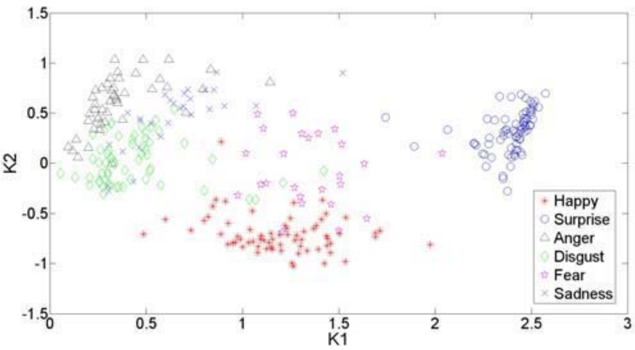


Fig. 4: The first two principal components of our proposed feature vectors for CK+.

IV. Conclusions and future work

Human emotions can be estimated by detecting and tracking just eight fiducial facial points. We have shown that our results are in good agreement with those obtained from algorithm that used 68 facial points, and other approach that used hundreds of features out of Gabor filters. Sadness emotion can be well discriminated with the help of texture features, as obviously seen from CERT results. Hence, our approach performance would be enhanced by adding texture features. Detecting and tracking eight points is less time-consuming which will open space for creating more effective applications. Furthermore, the relatively low dimensional features space helps the classifier to perform efficiently. For example, accurate heavily process algorithm for fiducial points detection [8] can be optimized with only 8 points to produce an efficient online system for human emotion recognition. The next step in our system is to correct the extracted geometric features by using an estimation of the head rotation angles (yaw, pitch, and roll).

Acknowledgment

This work is supported by Transregional Collaborative Research Centre SFB/TRR 62 ”Companion-Technology for Cognitive Technical Systems” funded by DFG and OvG-University Magdeburg.

References

[1] P. Ekman and W. Friesen, “Facial action coding system: A technique for the measurments of facial movements,” *Consulting Psychologists Press*, 1978.

[2] J. A. Russell, “A circumplex model of affect.” *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, Dec. 1980. [Online]. Available: http://dx.doi.org/10.1037/h0077714

[3] P. Lucey, J. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression,” in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, june 2010, pp. 94 –101.

[4] R. Niese, A. Al-Hamadi, B. Michaelis, and H. Neumann, “Integration of geometric and dynamic features for facial expression recognition in color image sequences,” in *Soft Computing and Pattern Recognition (SoCPaR), 2010 International Conference of*, dec. 2010, pp. 237 –240.

[5] C. Shan, S. Gong, and P. W. McOwan, “Facial expression recognition based on local binary patterns: A comprehensive study,” *Image Vision Comput.*, vol. 27, no. 6, pp. 803–816, May 2009. [Online]. Available: http://dx.doi.org/10.1016/j.imavis.2008.08.005

[6] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, and M. Bartlett, “The computer expression recognition toolbox (cert),” in *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, march 2011, pp. 298 –305.

[7] A. Martinez, “Deciphering the face,” in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, june 2011, pp. 7 –12.

[8] P. Belhumeur, D. Jacobs, D. Kriegman, and N. Kumar, “Localizing parts of faces using a consensus of exemplars,” in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, june 2011, pp. 545 –552.

[9] M. Valstar, B. Martinez, X. Binefa, and M. Pantic, “Facial point detection using boosted regression and graph models,” in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, june 2010, pp. 2729 –2736.

[10] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[11] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale, “A high-resolution 3d dynamic facial expression database,” in *Automatic Face Gesture Recognition, 2008. FG ’08. 8th IEEE International Conference on*, sept. 2008, pp. 1 –6.

[12] B. D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” 1981, pp. 674–679.