Static Facial Expression Analysis in Tough Conditions: Data, Evaluation Protocol and Benchmark

Abhinav Dhall[1] Roland Goecke[2,1] Simon Lucey[3] Tom Gedeon[1]
[1]Research School of Computer Science, Australian National University, Australia
[2]Faculty of Information Sciences and Engineering, University of Canberra, Australia
[3]Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia
abhinav.dhall@anu.edu.au, roland.goecke@ieee.org, simon.lucey@csiro.au, tom.gedeon@anu.edu.au

Abstract

Quality data recorded in varied realistic environments is vital for effective human face related research. Currently available datasets for human facial expression analysis have been generated in highly controlled lab environments. We present a new static facial expression database Static Facial Expressions in the Wild (SFEW) extracted from a temporal facial expressions database Acted Facial Expressions in the Wild (AFEW) [9], which we have extracted from movies. In the past, many robust methods have been reported in the literature. However, these methods have been experimented on different databases or using different protocols within the same databases. The lack of a standard protocol makes it difficult to compare systems and acts as a hindrance in the progress of the field. Therefore, we propose a person independent training and testing protocol for expression recognition as part of the BEFIT workshop. Further, we compare our dataset with the JAFFE and Multi-PIE datasets and provide baseline results.

1. Introduction

Realistic face data plays a vital role in the research advancement of facial expression analysis systems. Human facial expression databases till now have been captured in controlled 'lab' environments. We present a static facial expressions database. The database covers unconstrained facial expressions, varied head poses, large age range, different face resolutions, occlusions, varied focus and close to real world illumination. The static database has been extracted from the temporal dataset *Acted Facial Expressions in the Wild (AFEW)* [9]. Therefore, we name it the *Static Facial Expressions in the Wild (SFEW)* database.

With the advances in computer vision in the past few years, the analysis of human facial expressions has been made possible. Facial expressions are the facial changes in response to a person's internal affective state, intentions, or social communications. Facial expression analysis includes both the measurement of facial motion and the recognition of facial expressions, which are generated by the change in a person's facial muscles, which convey the affect of the individuals to the observers. It finds its use in human computer interaction (HCI), affective computing, human behavior analysis, ambient environment and smart homes, pain monitoring in patients, stress, anxiety and depression analysis, lie detection and medical conditions such as autism. Facial expression analysis is an active field of research for over a decade now and methods work well but in lab controlled environments.

On the basis of the descriptor type, facial expression analysis methods can be divided into three categories: geometric based [2, 27, 7], appearance based [8, 26, 25] and a combination of both [14]. Furthermore, facial expression analysis methods can also be classified into image based [18, 13] and video based [23, 28]. Human facial expressions are dynamic in nature and, therefore, video based methods are more robust since they encode the facial dynamics, which are not available in static, image-based methods. Studies [1] have also proven the effectiveness of video based methods over the static ones. However, there are scenarios where temporal data is not available and image based facial expression analysis methods come into picture. Typical applications of image based facial expression analysis classifying expressions in consumer level photographs, smile detection [25], expression based album creation [7] etc.

The *JAFFE database* [15] is one of the earliest static facial expressions dataset. It contains 219 images of 10 Japanese females. The subjects posed for six expressions (*angry*, *disgust*, *fear*, *happy*, *sad* and *surprise*) and the neutral expression. It has been extensively used in expression research. However, it has a limited number of samples, subjects and has been created in a lab controlled environment.

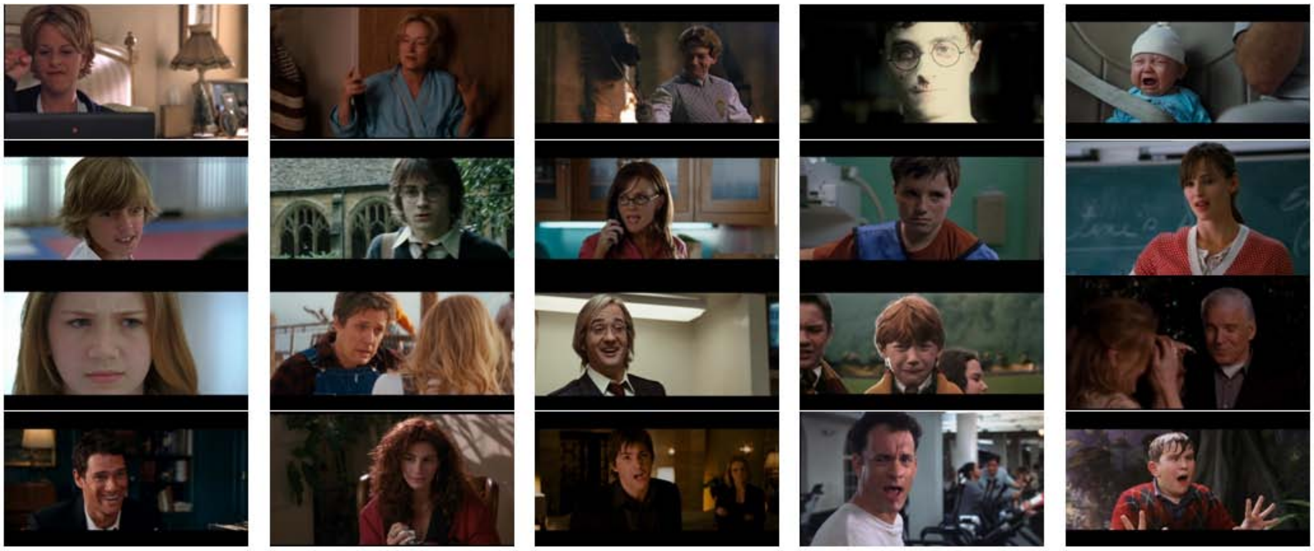


Figure 1. Sample images from the SFEW database.

The *CMU Pose Illumination and Expression (PIE)* [21] database is another popular and widely used database. It contains facial expression images posed by 68 subjects. Similar to PIE is the *Multi-PIE* [10] database, which contains 337 subjects. The main limitation of these databases is that they have been recorded in lab-controlled environments. Figure 2 contains some sample frames from the JAFFE and Multi-PIE datasets. It is evident from the figure that these samples do not represent the real world conditions.

The *MMI database* [19] is a facial expressions database, which contains both images and videos of 75 subjects shot in a lab-controlled environment. Similar to MMI is the *AR facial expressions database*, which contains 4000 images of 126 subjects. However, both these databases do not capture the conditions found in real world situations well.

The *Labeled Faces in the Wild database (LFW)* [11] is a static face recognition database created from face images found on the internet. It contains natural head movements, varied illumination, age, gender and occlusion. LFW has a strict defined training and testing protocol, which helps researchers in comparing the performance of their methods to that of others. Similar to LFW is the *Pubfig database* [12], which contains 58797 images of 200 people collected from the internet. Both these databases have been created for face recognition research. SFEW is similar in spirit to the LFW and Pubfig databases.

While movies are often shot in somewhat controlled environments, they provide close to real world environments that are much more realistic than current datasets that were recorded in lab environments. Though actors also pose in movies, clearly, (good) actors attempt mimicking real world human behaviour in movies. Our dataset in particular addresses the issue of static facial expressions in *difficult conditions* that are approximating real world conditions, which provides for a much more difficult test set than currently available datasets. Figure 1 displays some sample images from the database, which are very similar to real world scenarios.

Recently, the Facial Expression Recognition and Analysis Challenge (FERA) 2011 [22] competition was organised for comparing the state of the art temporal facial expression analysis methods. A subset of a new dataset *GEMEP* was used and both person independent and person dependent protocols were defined. The GEMEP dataset [3] consists of actors speaking dialogues and expressing emotions. Though are work is similar in spirit but has noticable differences: the FERA GEMEP subset consists of just 10 subjects and has been captured in lab conditions and is temporal. On the other hand, SFEW is a static dataset, which captures facial expressions in tough conditions.

2. Database Details

2.1. AFEW

AFEW [9] is a dynamic temporal facial expressions data corpus consisting of close to real world environment extracted from movies. It was collected on the basis of *Subtitles for Deaf and Hearing impaired (SDH)* and *Closed Caption (CC)* for the purpose of searching expression related content and extracting time stamps corresponding to video clips which represent some meaningful facial motion. The database contains a large age range of subjects from 1-70 years. The information about the clips has been stored in an extensible XML schema and the subjects in the clips have been annotated with attributes like *Name*, *Age of Actor*, *Age of Character*, *Pose*, *Gender*, *Expression of Person* and the overall *Clip Expression*.

A semi-automatic approach was followed during the creation of the database. The database contains clips from 37 movies. The movies have been chosen keeping in mind the need for different realistic scenarios and large age range of subjects to be captured (Table 1). The whole system of the database creation was divided into two steps. The first step consists of subtitle parsing. The subtitles are searched for a list of expression keywords such as ‘smiles’, ‘cries’, ‘sobs’, ‘scared’, ‘shouts’, 'laughs', 'shocked' etc. Video clips associated with the subtitles, which match the search criteria are played based on their time stamps information (extracted from the subtitle). Then in the second step, a human observer annotates the clips with the information about the actors and the expressions, which is stored in the XML schema. There are a total of 957 video clips in the database labeled with six basic expressions *angry*, *disgust*, *fear*, *happy*, *sad*, *surprise* and the *neutral* class. The audio corresponding to the extracted video clips is also stored for the scope of multimodal experiments. The database also contains video sequences, which have multiple actors (see, for example, Figure 3). In these situations, all actors have been annotated for their individual expression and the sequence has an overall video/scene expression.

2.2. Database construction Process

SFEW has been developed by selecting frames from AFEW. The database covers unconstrained facial expressions, varied head poses, large age range, occlusions, varied focus, different resolution of face and close to real world illumination. Frames were extracted from AFEW sequences and labelled based on the label of the sequence. In total, SFEW contains 700 images and that have been labelled for six basic expressions *angry*, *disgust*, *fear*, *happy*, *sad*, *surprise* and the *neutral* class and was labelled by two independent labellers. The database can be downloaded at:

http://cs.anu.edu.au/few

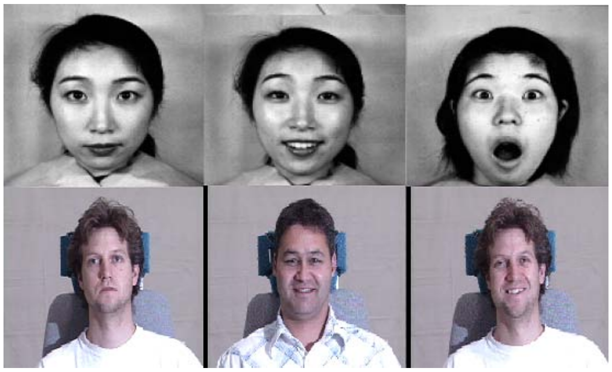
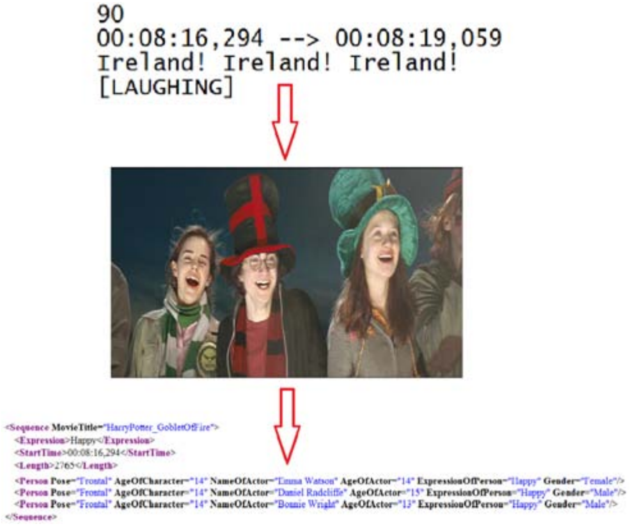


Figure 2. Sample images from the JAFFE and Multi-PIE databases. Note the controlled environment, in which they were recorded.



```
90
00:08:16,294 --> 00:08:19,059
Ireland! Ireland! Ireland!
[LAUGHING]
```

```
<Sequence MovieTitle="HarryPotter_GobletOfFire">
  <Expression>Happy</Expression>
  <StartTime>00:08:16,294</StartTime>
  <Length>2765</Length>
  <Person Pose="Frontal" AgeOfCharacter="14" NameOfActor="Emma Watson" AgeOfActor="14" ExpressionOfPerson="Happy" Gender="Female"/>
  <Person Pose="Frontal" AgeOfCharacter="14" NameOfActor="Daniel Radcliffe" AgeOfActor="15" ExpressionOfPerson="Happy" Gender="Male"/>
  <Person Pose="Frontal" AgeOfCharacter="14" NameOfActor="Bonnie Wright" AgeOfActor="13" ExpressionOfPerson="Happy" Gender="Male"/>
</Sequence>
```

Figure 3. The screenshot described the process of database formation. For example in the screenshot, when the subtitle contains the keyword ‘laughing’, the corresponding clip is played by the tool. The human labeller then annotates the subjects in the scene using the GUI tool. The resultant annotation is stored in the XML schema shown in the bottom part of the snapshot. Note the structure of the information about a sequence containing multiple subjects. The image in the screenshot is from the movie ‘Harry Potter and The Goblet Of Fire’.

3. Experiments

In this section, we perform experiments on SFEW, and present a simple baseline based on the *SPI* protocol. We compute state of the art descriptors on the data and also compare their performance on SFEW with JAFFE and Multi-PIE. The results demonstrate that the current state of the art methods, which perform very well on the existing datasets are not robust when it comes to being applied in more real world like conditions.

3.1. Person Independent Experiment Protocol

The *Strictly Person Independent (SPI) Protocol* for the SFEW database is divided into two sets. Each set has seven subfolders corresponding to the seven expression categories. The images are divided on the basis of their expression labels in their respective expression folder. The sets are created in a strict person independent manner. There are 346 images in Set 1 and 354 images in Set 2. There

Movie sources
21
About a Boy
American History X
Aviator
Black Swan
Did You Hear About The Morgans?
Dumb and Dumber
When Harry met Sally
Four weddings and a funeral
Frost/Nixon
Harry Potter and The Philosopher Stone
Harry Potter and The Chamber of Secrets
Harry Potter and The Goblet of Fire
Harry Potter and The Half Blood Prince
Harry Potter and The Order of Phoenix
Harry Potter and The Prisoners of Azkaban
Informant
It's Complicated
I Think I Love My Wife
Kings Speech
Little Manhattan
Notting Hill
One Flew Over Cuckoo's Nest
Pretty In Pink
Pretty Woman
Remember Me
Run Away Bride
Saw 3D
Serendipity
Social Network
Terminal
Term of Endearment
The Hangover
The Devil Wears Prada
Town
Valentine Day
Unstoppable
You've got mail

Table 1. Movie sources for the SFEW and AFEW databases.

are a total of 95 subjects in the database. For the purpose of consistent evaluation of different algorithms, the experiment will be twofold: first, train on set 1 and test on set 2 and then train on set 2 and test on set 1. The evaluation metrics for measuring the performance of FER systems are *accuracy*, *precision*, *recall* and *specificity*. The training and testing protocol is part of the BEFIT workshop in the form of a *Static Facial Expressions in the Wild Challenge*[1].

This challenge forms part of a broader plan of facial expressions in the wild (Table 2). Experiments for SFEW and AFEW are divided into three categories:

Prot.	**SFEW**	**AFEW**
	Train-Test sets have the:	
SPS	same single subject	same single subject
PPI	seen & unseen subjects	seen & unseen subjects
SPI	unseen subjects (Part of BEFIT workshop)	unseen subjects

Table 2. Different training and testing protocol scenarios for SFEW and AFEW. *SPS* - Strictly Person Specific, *PPI* - Partial Person Independent, *SPI* - Strictly Person Independent.

1. *SPS* - Strictly Person Specific,
2. *PPI* - Partial Person Independent, and
3. *SPI* - Strictly Person Independent.

Table 2 defines the scope of these protocols. The BEFIT workshop challenge falls under SPI for SFEW. Data, labels and other protocols will be made available on the database website.

3.2. Local Binary Patterns

In recent times, the local binary pattern (LBP) [16, 17] of descriptors has been extensively used for face analysis experiment. Our method for facial expression recognition is based on the local binary pattern (LBP) class of descriptors. The LBP descriptor assigns binary labels to pixels by thresholding the neighbourhood pixels with the central value. Therefore, for a centre pixel p of an image I and its neighbouring pixels N_i, a decimal value is assigned to it.

$$d = \sum_{i=1}^{k} 2^{i-1} I(p, N_i) \tag{1}$$

$$where \;\; I(p, N_i) = \begin{cases} 1 & \text{if } c < N_i \\ 0 & \text{otherwise} \end{cases}$$

An extension of LBP, local phase quantisation (LPQ) has been shown to perform better [17] than LBP and to be invariant to blur and illumination to some extent. LPQ is based on computing the short-term Fourier transform (STFT) on a local image window. At each pixel, the local Fourier coefficients are computed for four frequency points. Then, the signs of the real and the imaginary part of the each coefficient is quantised using a binary scalar quantiser, for calculating the phase information. The resultant 8-bit binary coefficients are then represented as integers using binary coding. LPQ [2] descriptor is calculated on grids and then concatenated for an image.

[1] website http://fipa.cs.kit.edu/befit/workshop2011

[2] We used the implementation available at http://www.cse.oulu.fi/MVG/Downloads

3.3. HOG and PHOG

We also experimented with the histogram of oriented gradients (HOG) descriptor [6], HOG counts occurrences of gradient orientation in localised portions of an image and has been used extensively in computer vision. Its extension, the pyramid of histogram of oriented gradients (PHOG) descriptor [4] has shown good performance in object recognition [4][3].

3.4. Comparison with JAFFE and Multi-PIE

First, we compared SFEW with JAFFE and Multi-PIE. Two experiments were conducted: (1) a comparison of SFEW, JAFFE and Multi-PIE on the bases of four common expression classes (*disgust*, *neutral*, *happy* and *surprise*) and (2) a comparison of SFEW and JAFFE on seven expression classes. For Multi-PIE, images for 50 subjects were extracted both frontal and non-frontal with 15 degree. In total, there were 400 images. JAFFE contains a total of 213 images. For the four common expression classes, there are a total of 120 images and all 213 images constitute the seven class experiment.

The faces are localised using the Viola-Jones [24] face detector, which gives the rough location of the face. We compute the descriptors on the cropped faces from both the databases. The cropped faces were divided into 3×3 blocks for LPQ and the neighbourhood size was set to 8. For PHOG, bin length $= 8$, pyramid levels $L = 3$ and angle range $= [0, 360]$. For classification, we used a support vector machine [5]. The type of kernel was C-SVC, with a radial basis function (RBF) kernel

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) = \exp(-\gamma ||x_i - x_j||^2), \ \ \gamma > 0. \tag{2}$$

We used a five-fold cross validation script [5], which creates five subsets of the dataset.

For the ***four expression class*** experiment, the classification accuracy on the Multi-PIE subset is 86.25% and 88.25% for LPQ and PHOG, respectively. For JAFFE, it is 83.33% for LPQ and 90.83% for PHOG. For SFEW, it is 53.07% for LPQ and 57.18% PHOG. Please note that these subsets are not person independent since the script randomly creates these subsets.

For the ***seven expression class*** experiment, the classification accuracy for JAFFE is 69.01% for LPQ and 86.38% for PHOG. For SFEW, it is 43.71% for LQ and 46.28% for PHOG. Figure 4 shows the performance accuracy comparison of the three databases. It is evident that LPQ and PHOG have high performance accuracy on JAFFE and Multi-PIE but significantly lower accuracy for SFEW. This is due to the close to real world conditions in the SFEW database. SFEW contains both high and very low resolution faces, which adds to the complexity of the problem. Furthermore, current methods have typically been developed and experimented on lab-controlled data. Expression analysis in (close to) real world situations is a non-trivial task and requires more sophisticated methods at all stages of the approach, such as robust face localisation/tracking, illumination and pose invariance.

3.5. SPI Baseline

Furthermore, as part of the *SPI* BEFIT challenge, *accuracy*, *precision*, *recall* and *specificity* should be calculated as follows:

$$\text{Overall Accuracy} = \frac{tp + tn}{tp + fp + fn + tn} \tag{3}$$

$$\text{Class Wise Precision} = \frac{tp}{tp + fp} \tag{4}$$

$$\text{Class Wise Recall} = \frac{tp}{tp + fn} \tag{5}$$

$$\text{Class Wise Specificity} = \frac{tn}{tn + fp} \tag{6}$$

Here, *tp* = true positive, *fp* = false positive, *fn* = false negative, and *tn* = true negative. Researchers should report the average score of these attributes over the two sets.

Based on the SPI protocol we compute the baseline scores (Table 3). In the strict person independent setup, we combined the two descriptors for better performance. Empirically the parameters are chosen for the descriptors, For PHOG, pyramid level L = 4, bin length = 16, angle range = [0-360] and for LPQ block size is 4×4. To reduce the complexity, principal component analysis is applied and 98% of the variance is kept. Further classification is performed with a non-linear SVM. The **baseline classification accuracy** calculated by averaging the accuracy for the two sets is 19.0%. Again, this low accuracy is attributed to the complex nature of conditions in the database. Clearly, the

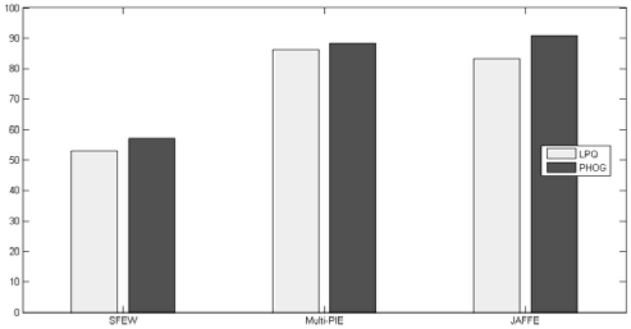


Figure 4. Four expression class accuracy comparison of SFEW, JAFFE and Multi-PIE based on LPQ and PHOG descriptors, and SVM classification.

[3] We used the PHOG implementation available at http://www.robots.ox.ac.uk/ vgg/research/caltech/phog.html

Emotion	**Angry**	**Disgust**	**Fear**	**Happy**	**Neutral**	**Sad**	**Surprise**
Precision	0.17	0.15	0.20	0.28	0.22	0.16	0.15
Recall	0.21	0.13	0.18	0.29	0.21	0.16	0.12
Specificity	0.48	0.66	0.64	0.51	0.61	0.60	0.66

Table 3. Average expression classwise Precision, Recall and Specifity results on the SFEW database based on the *SPI* protocol

current techniques are not robust enough for uncontrolled environment experiments.

4. Future Work and Conclusions

As part of a planned extension, we will make available the location (x, y) of the faces in the image, which can be helpful for accurate initialisation for feature extractors. We will compute a face aligned version of the dataset similar to the aligned LFW dataset. We will also provide dense landmark annotation of the faces using person-dependent AAMs [20]. Furthermore, as discussed in Section 3.1 and Table 2, strictly person dependent and partial person independent training and testing protocols will be posted on the database website.

We have presented a facial expression analysis database that contains face images in close to real world conditions extracted from movies. As part of the BEFIT workshop, we have presented a strict person-independent training and testing protocol, which should be used by researchers in future for evaluating their methods on the database and reporting their results. We have also compared the performance of state of the art descriptors such as LPQ and PHOG on SFEW with that of the widely used JAFFE and Multi-PIE databases. Empirically, it is proven that these methods are clearly not suitable for facial expression analysis in uncontrolled environment. Moreover, we also provide baseline results, which can be referred to by researchers. We believe that this dataset will be a useful resource for facial expression analysis research.

References

[1] Z. Ambadar, J. Schooler, and J. Cohn. Deciphering the enigmatic face: The importance of facial dynamics to interpreting subtle facial expressions. *Psychological Science*, 16(5):403–410, 2005. 1

[2] A. Asthana, J. Saragih, M. Wagner, and R. Goecke. Evaluating AAM Fitting Methods for Facial Expression Recognition. In *Proceedings of the IEEE International Conference on Affective Computing and Intelligent Interaction*, ACII'09, pages 598–605, 2009. 1

[3] T. Bänziger and K. Scherer. Introducing the Geneva Multimodal Emotion Portrayal (GEMEP) Corpus. In K. Scherer, T. Bänziger, and E. Roesch, editors, *Blueprint for affective computing: A sourcebook*. Oxford, England: Oxford University Press, 2010. 2

[4] A. Bosch, A. Zisserman, and X. Munoz. Representing Shape with a Spatial Pyramid Kernel. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, CIVR '07, pages 401–408, 2007. 5

[5] C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines, 2001. `http://www.csie.ntu.edu.tw/ cjlin/libsvm`. 5

[6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, CVPR'05, pages 886–893, 2005. 5

[7] A. Dhall, A. Asthana, and R. Goecke. Facial expression based automatic album creation. In *Proceedings of the 17th international conference on Neural information processing: models and applications - Volume Part II*, ICONIP'10, pages 485–492, 2010. 1

[8] A. Dhall, A. Asthana, R. Goecke, and T. Gedeon. Emotion recognition using PHOG and LPQ features. In *Proceedings of the Ninth IEEE International Conference on Automatic Face Gesture Recognition and Workshops (FG'2011), Facial Expression Recognition and Analysis Challenge Workshop (FERA)*, pages 878–883, 2011. 1

[9] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon. Acted Facial Expressions in the Wild Database. In *Technical Report*, 2011. 1, 2

[10] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. In *Proceedings of the Eighth IEEE International Conference on Automatic Face and Gesture Recognition*, FG'2008, pages 1–8, 2008. 2

[11] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, 2007. 2

[12] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and Simile Classifiers for Face Verification. In *Proceedings of the IEEE International Conference of Computer Vision*, ICCV'09, 2009. 2

[13] Z. Li, J.-i. Imai, and M. Kaneko. Facial-component-based bag of words and PHOG descriptor for facial expression recognition. In *Proceedings of the 2009 IEEE international conference on Systems, Man and Cybernetics*, SMC'09, 2009. 1

[14] S. Lucey, I. Matthews, C. Hu, Z. Ambadar, F. de la Torre, and J. Cohn. AAM Derived Face Representations for Robust Facial Action Recognition. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, FG'2006, pages 155–162, 2006. 1

[15] M. J. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expressions with gabor wavelets. In *Proceedings of the IEEE International Conference on Automatic Face Gesture Recognition and Workshops*, FG'98, 1998. 1

[16] T. Ojala, M. Pietikinen, and T. Menp. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, pages 971–987, 2002. 4

[17] V. Ojansivu and J. Heikkil. Blur Insensitive Texture Classification Using Local Phase Quantization. In *Proceedings of the 3rd International Conference on Image and Signal Processing*, ICISP'08, pages 236–243, 2008. 4

[18] M. Pantic and L. Rothkrantz. Facial action recognition for facial expression analysis from static face images. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, pages 1449–1461, 2004. 1

[19] M. Pantic, M. F. Valstar, R. Rademaker, and L. Maat. Web-based database for facial expression analysis. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, ICME'05, pages 317–321, 2005. 2

[20] J. Saragih and R. Goecke. Learning AAM fitting through simulation. *Pattern Recognition*, 42(11):2628–2636, 2009. 6

[21] T. Sim, S. Baker, and M. Bsat. The CMU Pose, Illumination, and Expression Database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, Dec. 2003. 2

[22] M. Valstar, B. Jiang, M. Mehu, M. Pantic, and S. Klaus. The first facial expression recognition and analysis challenge. In *Proceedings of the Ninth IEEE International Conference on Automatic Face Gesture Recognition and Workshops*, FG'11, pages 314–321, 2011. 2

[23] M. Valstar and M. Pantic. Fully automatic facial action unit detection and temporal analysis. In *Proceeding of IEEE International Conference on Computer Vision and Pattern Recognition Workshop*, CVPR-W'06, pages 149–149, 2006. 1

[24] P. A. Viola and M. J. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, CVPR'01, pages 511–518, 2001. 5

[25] J. Whitehill, G. Littlewort, I. R. Fasel, M. S. Bartlett, and J. R. Movellan. Toward practical smile detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 2106–2111, 2009. 1

[26] L. Xu and P. Mordohai. Automatic facial expression recognition using bags of motion words. In *Proceedings of the British Machine Vision Conference*, BMVC'10, pages 1–13, 2010. 1

[27] Z. Zeng, M. Pantic, G. Roisman, and T. Huang. A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 39–58, 2009. 1

[28] F. Zhou, F. De la Torre, and J. Cohn. Unsupervised discovery of facial events. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, CVPR'10, pages 2574–2581, 2010. 1