Facial Expression Recognition Using Gabor Motion Energy Filters

Tingfan Wu
Dept. Computer Science Engineering
UC San Diego
tingfan@gmail.com

Marian S. Bartlett Javier R. Movellan
Institute for Neural Computation
UC San Diego
{marni,movellan}@mplab.ucsd.edu

Abstract

Spatial Gabor energy filters (GE) are one of the most successful approaches to represent facial expressions in computer vision applications, including face recognition and expression analysis. It is well known that these filters approximate the response of complex cells in primary visual cortex. However these neurons are modulated by the temporal, not just spatial, properties of the visual signal. This suggests that spatio-temporal Gabor filters may provide useful representations for applications that involve video sequences. In this paper we explore Gabor motion energy filters (GME) as a biologically inspired representation for dynamic facial expressions. Experiments on the Cohn-Kanade expression dataset show that GME outperforms GE, particularly on difficult low intensity expression discrimination.

1. Introduction

Recent years, there have been good progress in computer vision applications to facial expression recognition. In fact some of applications, like digital cameras with automatic smile detection, have found their way to daily life consumer electronics [20]. Many of the current facial expression recognizers (see [24]) focus on the problem of analyzing the expression in *static* images. Some approaches are based on parameterized shape models, i.e. models of the shape of key facial features like the eyebrows, eye, and mouth (See Chang et al.[7]).

Other approaches rely on appearance based representations that describe the visual texture of face regions, without explicitly recognizing the location and shape of facial features. Appearance based discriminative approaches have proven highly robust for face detection [19], identity recognition [16] and expression recognition [3]. Popular appearance based features include Gabor energy filters [3], Haar wavelets [21], and local binary patterns (LBP)[17]. Hybrid shape/appearance methods, particularly within the Active appearance models (AAM) literature are also popular [8, 12].

The importance of facial dynamics in recognizing facial expressions has been established in many vision [22, 5, 25] and psychological experiments [4, 2]. Facial expression experts commonly report that motion, not just static patterns, is particularly critical for analyzing subtle, low intensity, facial expressions. Psychological experiments have also shown the importance of dynamics for recognizing subtle facial expressions [2]. Static approaches incorporate temporal information by integrating the output of the frame by frame recognizers using dynamical models such as hidden Markov models [13], LBP codebooks [23] or dynamic Bayesian network (DBN) [18]. We call this approach *Late Temporal Integration* for it uses temporal information only after high level categories (e.g., facial expressions) have been already extracted.

An alternative approach, which we call *Early Temporal Integration* utilizes features that represent low level spatiotemporal patterns, prior to the abstraction of high level categories [25].

The performance of the human vision system is far superior to that of any current computer vision system. Therefore using biologically inspired methods to represent video signals may help us gain a better understanding of the computational principles that guide visual processing in the brain. Most importantly it may also help us build better computer vision systems. In this paper, we explore the use of biologically inspired spatiotemporal Gabor Motion Energy filters (GME) [1] for low level integration of spatiotemporal information in facial expression recognition tasks.

We show that GMEs (a early integration method) outperform spatial Gabor energy filters (a popular representation for late integration methods) on the Cohn-Kanade dataset of basic expressions of emotion. We show that GME filters are particularly effective for representing low intensity facial expressions, an aspect that may be important for recognition of spontaneous expressions in daily life conditions.

2. Gabor Filters

We first present a general framework for n-dimensional Gabor energy filters. We then use this framework to instantiate spatial Gabor Energy filters (GE) and spatiotemporal Gabor Motion Energy filters (GME).

The impulse response of an n-dimensional Gabor filter is the product of an n-dimensional complex sinusoid and an n-dimensional Gaussian envelope:

$$g(x;k,A,u_0,x_0,p) = ke^{j2\pi u_0^T x+p} \cdot e^{-\pi[A(x-x_0)]^T[A(x-x_0)]}$$

where $x, u_0, x_0 \in R^n$. The complex Gabor function consists of a pair of $even$ (real part) and odd (imaginary part) Gabor filters. The two filters have a relative phase difference of $\pi/2$ and an absolute phase offset p. The Fourier transform of the filter's impulse response, i.e., the filter's transfer function, is as follows

$$\hat{g}(u;k,A,u_0,x_0,p) = \frac{k}{|A|}e^{j(u-u_0)^T x_0+p-\pi(u-u_0)^T(A^TA)^{-1}(u-u_0)}.$$

Thus n-dimensional Gabor filters are essentially bandpass filters centered at spatial-(temporal) location x_0, with peak frequency u_0 and a Gaussian envelope linearly transformed by $A, (A \in R^{n\times n})$. k is the normalization scalar ensuring that $||g||_1 = 1$. It is well known that the magnitude $|g|$ of 2-d Gabor filters approximate the response of neurons in primary visual cortex (V1). In practice, we set $x_0 = 0, p = 0$ thus simplifying the equations as follows

$$g(x;A,u_0) = ke^{j2\pi u_0^T x}e^{-\pi x A^T A x} \quad (1)$$

$$|\hat{g}(u;A,u_0)| = \frac{k}{|A|}e^{-\pi(u-u_0)^T(A^TA)^{-1}(u-u_0)}. \quad (2)$$

In the following two sections, we instantiate spatial GE and spatiotemporal GME filters from the above generic framework by specifying the corresponding parameters u_0 and A. To make the notation clear, we will denote spatial parameters with the subscript s and spatiotemporal parameters with subscript st.

2.1. Spatial Gabor Energy Filter Bank (GE)

The GE filter bank proposed here is based on [11, 3]. It consists of a set of self-similar 2-D Gabor filters: The entire filter bank can be generated from a "mother" filter by changing its orientation and scale. The mother filter is shown in Fig.1(a)($\theta = 0$). It consists of a sinusoidal carrier with filter peak frequency $u_s = [\,F_0\ 0\,]^T$ and an axis aligned Gaussian envelope. We use a canonical Gaussian envelope with covariance matrix $S^{-2} = \left[\begin{smallmatrix} a^2 & 0 \\ 0 & b^2 \end{smallmatrix}\right]$, where a, b defines the bandwidths parallel and orthogonal to the carrier direction.

Converting from polar coordinates into Cartesian coordinates, we obtain

$$a = \sqrt{\frac{\pi}{\ln 2}}\left(\frac{2^{\Delta F}-1}{2^{\Delta F}+1}\right)F_0, \quad (3)$$

$$b = \sqrt{\frac{\pi}{\ln 2}}\tan\left(\frac{\Delta\theta}{2}\right)F_0. \quad (4)$$

where ΔF is the frequency bandwidth and $\Delta\theta$ is the orientation bandwidth, We set these parameters according to standard parameters found in V1 cortex [14].

Finally, a rotation matrix R_s is used to generate Gabor filters for different orientations. Combining the bandwidth scale and rotation matrix, we set

$$A_s = S_s R_s = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix},$$

$$u_s = \begin{bmatrix} F_0 \\ 0 \end{bmatrix} R_s.$$

Finally, the representations of the filter in space and frequency domains are

$$g(x;F_0,a,b,\theta) = Ke^{j2\pi\left[\begin{smallmatrix} F_0 \\ 0 \end{smallmatrix}\right]R_s x}e^{-\pi x^T R_s^T\left[\begin{smallmatrix} a^2 & 0 \\ 0 & b^2 \end{smallmatrix}\right]R_s x},$$

$$\hat{g}(u;F_0,a,b,\theta) = \frac{K}{|ab|}e^{-\pi(u-\left[\begin{smallmatrix} F_0 \\ 0 \end{smallmatrix}\right])^T R_s^T\left[\begin{smallmatrix} a^2 & 0 \\ 0 & b^2 \end{smallmatrix}\right]^{-1}R_s(u-\left[\begin{smallmatrix} F_0 \\ 0 \end{smallmatrix}\right])}.$$

Figure 1(b) and 1(c) show the GE filter bank in the frequency and spatial domains respectively.

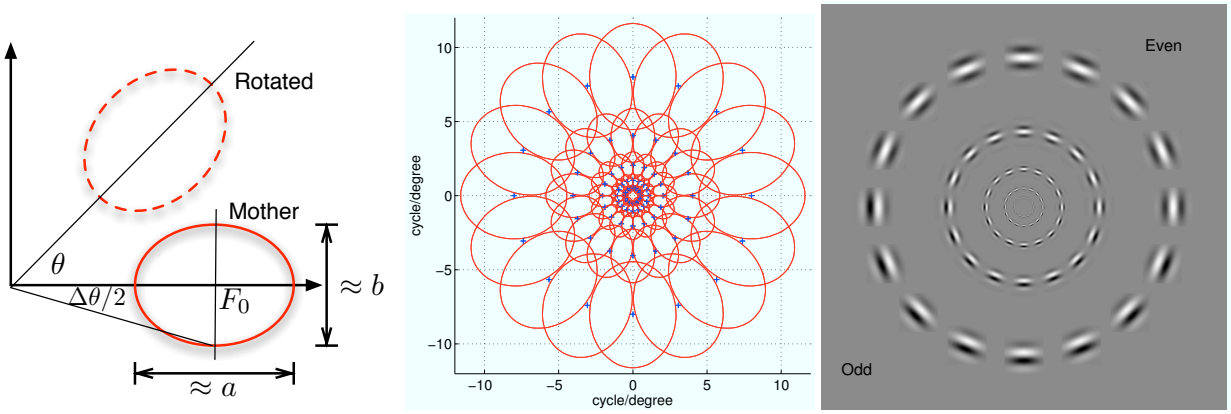
2.2. Spatiotemporal Gabor Motion Energy Filter (GME)

In the literature, there are two different types of spatiotemporal Gabor filters [15](see Fig.2): frequency-tuned and velocity-tuned. Frequency-tuned filters have a stationary Gaussian envelope while velocity-tuned filters have moving Gaussian envelopes. Velocity-tuned filters (Fig.2(b)) are non-separable, while frequency-tuned filter are spatial-temporal separable. Most neurons in primary visual cortex are frequency tuned, rather than velocity tuned. Thus here we focus on frequency-tuned filters (Fig.2(a)).

Due to the separability of frequency tuned GME filters, we can construct them by adding 1D temporal filters on top of the GE filters described in the previous section $g_s(\cdot)$: the GE filters do the frame-by-frame spatial analysis. The output of the spatial filters is then combined through time using 1D Gabor temporal filters $g_t(\cdot)$:

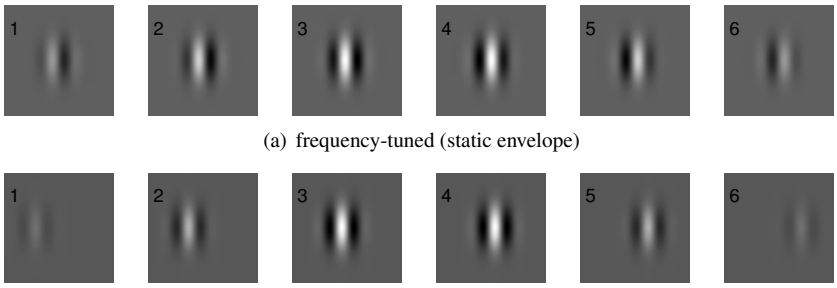
$$g_t(w,c) = e^{j2\pi wt}e^{-\pi t^2}, \quad (5)$$

where w is the preferred temporal frequency of the filter.



(a) parameters in frequency spectrum (b) half-magnitude contour in frequency spectrum (c) ensemble of filters in special domain

Figure 1. The spatial GE filter bank,



(a) frequency-tuned (static envelope)

(b) velocity-tuned (moving envelope)

Figure 2. Examples of GME filters, each example is a 6-frame long GME. (only the real part is shown.)

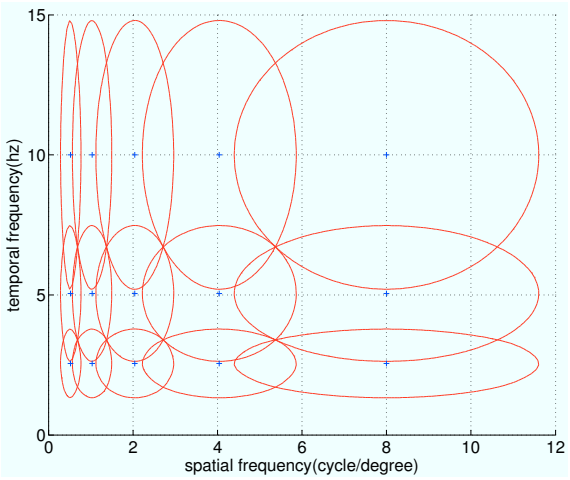


Figure 3. Half-magnitude contour of GME filter spectrum.

We design the peak frequencies of the temporal filters using similar heuristics to those used for the spatial filters. We start at the highest possible peak frequencies that keep the half-magnitude response withing the Nyquist limit. Then the next lower peak frequency is obtained by halving the previous peak frequency. The lowest possible frequency is set at 2.5Hz to prevent serious aliasing effects. The bandwidth of each filter, was designed using (3) and (4) as commonly used for spatial filters.

The combined spatiotemporal filters $g_{st} = g_t \circ g_s$ can be written in compact matrix form

$$A_{st} = S_{st} R_{st} = \begin{bmatrix} S_s & \mathbf{0} \\ \mathbf{0}^T & c \end{bmatrix} \begin{bmatrix} R_s & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix}$$

$$u_{st} = \begin{bmatrix} u_s \\ w \end{bmatrix}^T R_{st} = \begin{bmatrix} F_0 \\ 0 \\ w \end{bmatrix}^T \begin{bmatrix} R_s & \mathbf{0} \\ \mathbf{0}^T & 1 \end{bmatrix}.$$

As shown above, the rotational matrix R_{st} only rotate the filter along the temporal axis. There are never spatiotemporal-temporal rotations.

A spatiotemporal GME filter is then defined as

$$\begin{aligned} & g_{st}(x, t; F_0, w, \theta, a, b, c) \\ & = K e^{j2\pi \begin{bmatrix} F_0 \\ 0 \\ w \end{bmatrix}^T R_{st} \begin{bmatrix} x \\ t \end{bmatrix}} e^{-\pi \begin{bmatrix} x \\ t \end{bmatrix}^T R_{st}^T \begin{bmatrix} a^2 & & \\ & b^2 & \\ & & c^2 \end{bmatrix} R_{st} \begin{bmatrix} x \\ t \end{bmatrix}}. \end{aligned} \quad (6)$$

A frequency-tuned GME filter appears like a planar moving sinusoid encapsulated in a stationary spatiotemporal Gaussian envelope which fades in and out in time (see Figure 2(a)) The corresponding Fourier transform is

$$\hat{g}_{st}(u; F_0, w, \theta, a, b, c) = \frac{K}{|abc|} e^{-\pi(u-\begin{bmatrix} F_0 \\ 0 \\ w \end{bmatrix})^T R_{st}^T \begin{bmatrix} a^2 & & \\ & b^2 & \\ & & c^2 \end{bmatrix}^{-1} R_{st}(u-\begin{bmatrix} F_0 \\ 0 \\ w \end{bmatrix})} \quad (7)$$

The frequency spectrum of the filter bank is shown in Figure 3.

3. Experiment

3.1. Methods

We evaluated the performance of GE and GME filters on the Cohn-Kanade facial expression database [10], a popular and widely used database to evaluate facial expression recognition algorithms. This database consists of 100 students aged from 18 to 30 years old, of which 65% were female, 15% were African-American, and 3% were Asian or Latino. Subjects were instructed to perform a series of 23 facial displays, six of which were prototypical emotions including angry, disgust, fear, joy, sad and surprise. For our experiments, we selected 317 image sequences from 93 subjects. The selection criterion was that a sequence can be labeled as one of the six basic emotions and the video clip is longer than 13 frames. The faces were detected automatically by a variation of Viola and Jones detector [9] and then normalized to 96×96 patches based on the location of the eyes.

The facial expression dynamics of video clips in the database always start from neutral expression and end on apex, the maximum intensity of a expression. Most previous research evaluated spatial features using only the last frames which contain the most expressive expressions. However, spontaneous expressions observed in real-life are never this extreme. To make the problem harder and closer to reality, we pre-processed the training and testing data into following 2 conditions,

- **onset**– the first 6 frame – low intensity expressions
- **apex** –the last 6 frame – extreme intensity expressions

Each video clip was first convolved with all the filters in a filter bank. Then the responses from all filters were concatenated into a long feature vector. Only the magnitude (energy) of the responses was used. We aggregated the result from different time frames via statistical operators such as $\min$, $\max$, and mean. In this case, we were able to process video clips of arbitrary length.

Table 1. Average ROC on **onset** and **apex** sequences classification over 6 expressions

	onset		apex	
	GE	GME	GE	GME
Anger	0.6673	0.8294	0.9398	0.9575
Disgust	0.6696	0.6770	0.9784	0.9784
Fear	0.6104	0.6670	0.9648	0.9698
Joy	0.7896	0.8773	0.9913	0.9895
Sad	0.7095	0.7838	0.9877	0.9809
Surprise	0.7783	0.8793	0.9942	0.9927
mean	0.7041	**0.7856**	0.9760	**0.9781**

We used linear support vector machine (SVM) [6] as our classifier. Using the bootstrap method, each time we randomly selected 60 subjects as the training set, while the rest of subjects served as the testing set. To avoid over-fitting, we applied a double cross-validation method. The first layer of cross-validation was used to find the best parameters and then trained with the whole training data using the selected parameter. ROC scores were used to evaluate the performance of the system on the testing set. Each experiment was repeated 10 times and the average was reported.

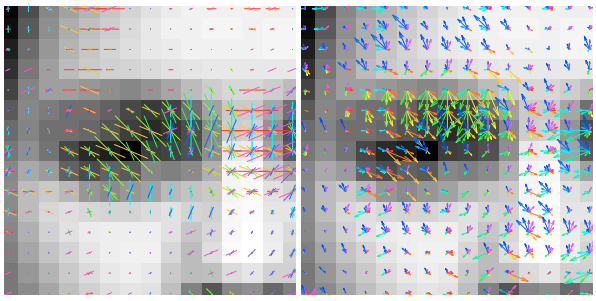
3.2. Results

Table 1 compares the cross-validation performance on GE and GME on two different conditions. The GME features outperforms GE in both conditions. In particular, GME is superior to GE by 7% on the **onset** low intensity condition. In contrast, the performance gaps between GE and GME at **apex** is small.

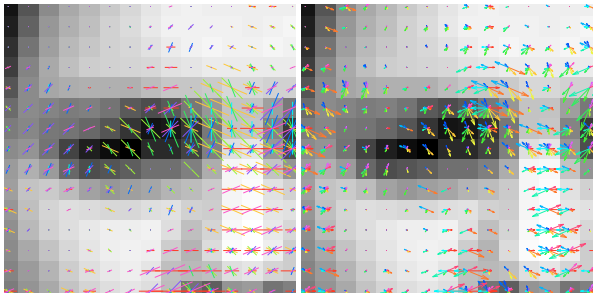
. Figure 4 shows the learned SVM weights in the eyebrow region. Since the dynamic range of features were normalized, these weights represent the relative importance of each feature. For GE features, the weights are represented by the line segments in each location on the face. The "orientation" of the line is perpendicular to the wavefront of the filters' sinusoidal carriers. The colors of the line segments also represent the "orientations", which makes it easier to identify clusters of lines pointing to the same direction. The length of the lines represent the magnitude of the weight which shows the importance of the feature. As for GME features, in addition to line segments, arrows are added to show the corresponding motion directions of the features.

3.3. The Use of Linear Classifiers

In our experiment, we used a *linear*-SVM as our classifier. Though non-linear SVM is known to be superior than linear SVM for its ability on handling non-linear problems, empirically, we found little performance difference between non-linear Gaussian kernel ($K(x, y) = e^{\gamma x^T y}$) and linear kernel. In addition, the best Gaussian kernel parameter γ is very small, which corresponds to a very smooth Gaussian



(a) GE onset (b) GME onset



(c) GE apex (d) GME apex

Figure 4. The right-eye closeup for learned SVM weights of angry expression using GME (b)(d) and GE (a)(c) features. The direction of the arrows and line-segments are color coded for better visualization.

an indication of low-linearity problem.

The linearity might result from the high dimensional feature space spanned by non-linear GE and GME features. Unlike regular linear filters, GME filters incur non-linearity by taking the magnitude of complex filter responses. The current filter bank has a collection of 15 spatiotemporal frequency combinations, 16 orientations. They are then applied to every location on the face image, which amounts to 2,211,840 features per video sequence. Such large set of features are capable of encoding facial movements of various scales and directions. In such high dimensional space, our problem may become linearly separable and thus the performance gap between non-linear and linear classifiers is minimal. In comparison to non-linear SVM, linear SVM has fewer hyper-parameters and much faster training/classification speed.

3.4. Motion Is Important for Recognizing Subtle Expressions

The importance of facial dynamics in recognizing facial expressions has been established in many vision [22, 5, 25] and psychological experiments [4, 2]. Spatiotemporal facial behavior tends to include a large number of subtle expressions. It is therefore crucial to detect subtle expressions for real applications. Psychological experiments have shown the importance of dynamics for recognizing subtle facial expressions [2]. In this paper, the experiments further identified that the major advantage of dynamic features over static features is on low intensity expressions. The result is consistent with psychophysical findings in the literature [2].

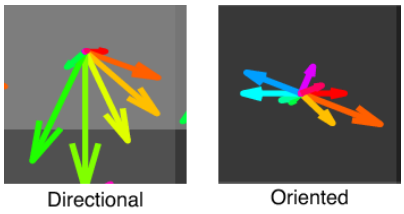


Figure 5. GME patterns

In this paper, we explore this issue by comparing the cross validation performance and the learned SVM weights in **onset** and **apex** conditions. The fact that GME outperforms GE at the onset condition (see Table 1) by a large margin suggested that dynamic motion information is more informative than the static features at low intensity (onset). At the early onset, the spatial cues may be too subtle to be detected in a noisy context. On the contrary, GME characterize expressions by facial motion patterns, such as the direction and speed of facial features movement. The motion information remains strong for low intensity expressions. On the other hand, when the expression comes to the apex, the intensity of expression is so high that spatial feature are sufficient for good discrimination. In fact, motion information close to the apex is minimal. We were actually surprised at how well GMEs performed at the apex, considering how little motion information is available at that point. After careful inspection of the GME weights in Fig.4(b)(d), we found there are actually two different types of local GME patterns in one pixel (see Fig.5). One is a "directional pattern" formed by an arrow and its adjacent directions. This pattern shows very clear directional movement. The other type is an "orientation pattern" formed by two arrows of opposite directions and their adjacent directions.

We conjecture that in practice the SVM learned to use GMEs to create "orientation patterns" that are functionally equivalent to a GE filter of corresponding orientation. This is further supported by the fact that the spatial distribution of "orientation patterns" is very similar to the distribution of the GE weights (Fig.4). A possible explanation is that the even (symmetric) part of a GME filter has non-zero DC-response which functions similar to a GE filter with the same frequency and orientation. This enables GME to pickup information that GE can. In summary, GME is capable of describing both dynamic and static textures, which make it as as good as GE when there is no motion information and better when there is little spatial texture and large motion information.

The change of SVM weights at different onset stages

also supports our hypothesis that motion information is critical for early onset and low intensity facial expressions. Figure 4 compares the weight spatial distribution and "directional" and "orientation" patterns between GME and GE. At early onset where GME significantly outperforms GE, GME captures much more information around the eyebrow and upper eyelid region. The "directional" patterns in these regions suggest the use of motion information. In contrast, GME and GE gives similar performance at the apex where the spatial distribution of GME and GE are also similar to each other. Most GME pattern at apex are "orientation", a strong indication of static texture signal.

4. Conclusion

In this work, we explored the use of Gabor Motion Energy filters (GME) as a basic representation for Early Temporal Integration in facial expression classification. We compared GMEs with Gabor Energy filters (GE) a popular representation for Late Temporal Integration methods.

We found that GMEs (Early Integration) outperformed GEs (Late Integration) spatial Gabor filters. Careful visualization of the learned SVM weights and analysis of classification performance revealed why GME were superior in their ability to capture both static and dynamic texture information.

In addition we found that low level motion information, as captured by GMEs, may be particularly critical for classifying low intensity expressions. This may be particularly important for recognition of spontaneous expressions in real-life situations. Such expressions tend to be of much lower intensity than voluntary expressions of the type available in the Cohn-Kanade database.

Acknowledgment

Support for this work was provided by NSF grants SBE-0542013 (Science of Learning Center) and CNS-0454233. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] E. Adelson and J. Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2(2):284–299, 1985.

[2] Z. Ambadar, J. Schooler, and J. Cohn. Deciphering the enigmatic face. The Importance of Facial Dynamics in Interpreting Subtle Facial Expressions. *Psychological Science*, 16:403–410, 2005.

[3] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Automatic recognition of facial actions in spontaneous expressions. *Journal of Multimedia*, 1(6):22–35, 2006.

[4] J. Bassili. Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Journal of personality and social psychology*, 37(11):2049–2058, 1979.

[5] M. Black and Y. Yacoob. Recognizing facial expressions in image sequences using local parameterized models of image motion. *International Journal of Computer Vision*, 25(1):23–48, 1997.

[6] C. Chang and C. Lin. LIBSVM: a library for support vector machines, 2001.

[7] Y. Chang, C. Hu, R. Feris, and M. Turk. Manifold based analysis of facial expression. *Image and Vision Computing*, 24(6):605–614, 2006.

[8] T. Cootes, G. Edwards, C. Taylor, et al. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[9] M. Eckhardt, I. Fasel, and J. Movellan. Towards practical facial feature detection. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(3):379–400, 2009.

[10] T. Kanade, J. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Proceedings of the fourth IEEE International conference on Automatic Face and Gesture Recognition*, page 46, 2000.

[11] C. Lee. Fingerprint feature extraction using Gabor filters. *Electronics Letters*, 35:288, 1999.

[12] 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 *Proc. International Conference on Automatic Face and Gesture Recognition*, pages 155–160, 2006.

[13] L. Ma, D. Chelberg, and M. Celenk. Spatio-temporal modeling of facial expressions using Gabor-wavelets and hierarchical hidden Markov models. In *IEEE International Conference on Image Processing, 2005. ICIP 2005*, volume 2, 2005.

[14] J. R. Movellan. Tutorial on gabor filters. Technical report, MPLab Tutorials, UCSD MPLab, 2005.

[15] N. Petkov and E. Subramanian. Motion detection, noise reduction, texture suppression, and contour enhancement by spatiotemporal gabor filters with surround inhibition. *Biol. Cybern.*, 97(5):423–439, 2007.

[16] P. Phillips. Support vector machines applied to face recognition. *Advances in Neural Information Processing Systems*, pages 803–809, 1999.

[17] C. Shan, S. Gong, and P. McOwan. Facial expression recognition based on Local Binary Patterns: A comprehensive study. *Image and Vision Computing*, 27(6):803–816, 2009.

[18] Y. Tong, J. Chen, and Q. Ji. A unified probabilistic framework for spontaneous facial action modeling and understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99(1), 5555.

[19] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[20] J. Whitehill, M. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan. Towards practical smile detection. *Pattern Analysis and Machine Intelligence*, (11):2106–2111, 2009.

[21] J. Whitehill and C. Omlin. Haar features for FACS AU recognition. In *Proc. IEEE Intl Conf. Face and Gesture Recognition*, 2006.

[22] P. Yang, Q. Liu, X. Cui, and D. Metaxas. Facial expression recognition using encoded dynamic features. In *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008*, pages 1–8, 2008.

[23] P. Yang, Q. Liu, and D. Metaxas. Boosting coded dynamic features for facial action units and facial expression recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–6, 2007.

[24] 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*, 31(1):39–58, 2009.

[25] G. Zhao and M. Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE transactions on pattern analysis and machine intelligence*, 29(6):915, 2007.