Background suppressing Gabor energy filtering

ARTICLE *in* PATTERN RECOGNITION LETTERS · JANUARY 2015

Impact Factor: 1.55 · DOI: 10.1016/j.patrec.2014.10.001

3 AUTHORS:

Albert Cruz
California State University, Bakersfield
13 PUBLICATIONS **19** CITATIONS
SEE PROFILE

Bir Bhanu
University of California, Riverside
449 PUBLICATIONS **5,641** CITATIONS
SEE PROFILE

Ninad Shashikant Thakoor
University of California, Riverside
44 PUBLICATIONS **124** CITATIONS
SEE PROFILE

Pattern Recognition Letters 52 (2015) 40–47

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Background suppressing Gabor energy filtering ☆

Albert C. Cruz*, Bir Bhanu, Ninad S. Thakoor

Center for Research in Intelligent Systems, Winston Chung Hall 216, University of California, Riverside, Riverside, CA 92521, USA

ARTICLE INFO

Article history:
Received 14 May 2014
Available online 13 October 2014

Keywords:
Gabor filter
Background texture
Facial emotion recognition
Bioimaging

ABSTRACT

In the field of facial emotion recognition, early research advanced with the use of Gabor filters. However, these filters lack generalization and result in undesirably large feature vector size. In recent work, more attention has been given to other local appearance features. Two desired characteristics in a facial appearance feature are generalization capability, and the compactness of representation. In this paper, we propose a novel texture feature inspired by Gabor energy filters, called background suppressing Gabor energy filtering. The feature has a generalization component that removes background texture. It has a reduced feature vector size due to maximal representation and soft orientation histograms, and it is a white box representation. We demonstrate improved performance on the non-trivial Audio/Visual Emotion Challenge 2012 grand-challenge dataset by a factor of 7.17 over the Gabor filter on the development set. We also demonstrate applicability of our approach beyond facial emotion recognition which yields improved classification rate over the Gabor filter for four bioimaging datasets by an average of 8.22%.

1. Introduction

Features extracted from images are at the core of computer vision and pattern recognition and their applicability can vary depending on the type of images. In this paper, we focus on local appearance features because they are easily adaptable to different application areas. Specifically, we are inspired by the Gabor filter which was originally introduced in 1946 [1,2]. Since then, it has seen extensive use in many fields of pattern recognition. Gabor filtering is the process of representing an image in terms of gratings that approximate the low-level behavior of the human visual system. However, current methods prefer to use features other than the Gabor energy filter. In this paper, we explore Gabor-based features' two areas where other features are more frequently used than the Gabor filter, namely facial emotion recognition and bioimaging.

State-of-the-art features have two properties: (1) the ability to generalize to external and intrinsic factors, such as registration errors, illumination variations, blur or noise. For example, local binary pattern (LBP) features can be rotation invariant and are robust to monotonic grayscale transformation from shadows [3]. The scale invariant feature transform (SIFT) is scale invariant [4]. (2) Compactness of feature representation. LBP uses histograms to reduce feature vector size. The original formulation [5] of the Gabor energy filter does not have either of these properties.

We propose background suppressing Gabor energy filtering which removes background texture with a generalization step, and reduces feature vector size with a computational efficiency step. We improve performance over other frontal face feature representations used for facial emotion recognition on the Audio/Visual Emotion Challenge (AVEC) 2012 grand-challenge dataset [6] and the Cohn–Kanade+ (CK+) dataset [7]. We also provide results on four bioimaging datasets.

2. Related Work, motivation and contributions

The focus of this work is local appearance features, the most commonly used of which are LBP [3]. Though the features are often referred to as LBP features, they are actually histograms of a LBP coded image. LBP quantifies textures at a pixel level by encoding the microtexture of a pixel and its neighborhood with an eight-bit code. Ref. [26] conducted a survey of LBP features for use with bio-imaging data and investigated: elongated quinary patterns (EQP), local ternary patterns, improved local binary patterns and center-symmetric local binary patterns. It was found that EQP was the best performer. Ref. [27] detected mTBI from MRI images with LBP in a context based system. In facial emotion recognition, current methods often divide the frontal face into sub-regions and compute the histogram of LBP codes for each sub-region. For example, in Ref. [6], the face was divided evenly into 10×10 sub-regions, or grids, and the outer regions were discarded because they corresponded to the regions of a face where there were no facial expressions. Uniform LBP features have been used as the baseline for recent facial emotion recognition grand challenges

☆ This paper has been recommended for acceptance by S. Sarkar.

* Corresponding author. Tel.: +1 951 827 3954.
E-mail address: acruz@ee.ucr.edu (A.C. Cruz).

http://dx.doi.org/10.1016/j.patrec.2014.10.001
0167-8655/

[6]. There have been many improvements to the original LBP feature. Ref. [11] proposed three-patch and four-patch local binary patterns (TPLBP, FPLBP). Whereas LBP encodes a microtexture of a single pixel, TPLBP and FPLBP encode larger patterns and homogeneity of a region by comparing a pixel's microtexture to the microtextures of neighboring pixels. Ref. [13] proposed extending LBP to a spatiotemporal feature with the use of three orthogonal planes (LBP-TOP). Ref. [28] extended LBP to the spatiotemporal domain with monogenic signals analysis and phase-quadrant encoding and a local XOR operator in three orthogonal planes (STLMMBP).

Not all facial emotion recognition and bio-imaging methods use LBP as a local appearance feature. Ref. [29] detected myopia from retinal fundus images with a bag-of-features including SIFT. Ref. [30] clustered pigmented skin lesions from a dermoscope with LPQ and other features. Ref. [31] classified states of hESC from phase contrast images with Gabor statistics. The top approach for the facial emotion recognition and analysis challenge for discrete emotions used local phase quantization (LPQ) [16]. In LPQ, the phase of a per-pixel discrete Fourier transform (DFT) quantifies the texture. It was found that the phase of DFT of a local neighborhood is invariant to centrally symmetric blur. Sub-region histograms give LPQ a compact representation. Ref. [21] used a difference image to quantify facial motion, and a discrete cosine transform (DCT) to compress the feature vector size. Ref. [4] proposed the scale-invariant feature transform (SIFT), which quantifies local features with the maxima and minima of a difference-of-Gaussians. Recently, it was used by Ref. [23], where the SIFT features were computed at 83 fiducial feature points. A summary of related work is given in Table 1.

2.1. Motivation

We focus our work on improving the Gabor energy filter because it has been, and still is an important feature for computer vision [32], though it has no generalization or computational efficiency steps. There are approaches that have improved upon the original Gabor energy filter [5]. Ref. [33] used the imaginary part of the Gabor filter for cerebrovascular images. Ref. [34] applied a spatiotemporal Gabor filter to emotion recognition on the Cohn–Kanade dataset. Ref. [20] represented the output of Gabor energy filters with sub-region histograms. These two methods improve the Gabor energy filter, but do not address both generalization and compactness of the representation. Out of the top six approaches for AVEC 2011, only one approach used a Gabor energy filter [17]. Approaches preferred LPQ, LBP or active appearance models. We assert that the Gabor filter can still be effectively applied to facial emotion if the following technical challenges are addressed:

(1) *Generalization*: Gabor energy filters do not generalize well in unconstrained settings because a Gabor energy filter captures edge magnitudes at almost all orientations, including edges from noise due to background texture. Current local appearance features have additional steps in an effort to be more generalizable and robust. Ref. [34] addressed this by extending the Gabor filter to temporal domain with Gabor motion energy features. However, the feature vector size was increased by the number of temporal scales over the original Gabor energy filter, which already has a large feature vector size. For example, the feature vector was increased by a factor of 3.72 between Refs. [5] and [34]. We address this technical challenge with background suppressing Gabor energy filtering, which removes the edges due to background noise but retains the significant edges that correspond to the edges of the objects in a scene. We also compute texture at a pixel, microtexture level, so the method is invariant to monotonic grayscale transformations. We prefer Gabor filters over LBP because the background suppression pipeline emulates the human visual system (HVS). It requires tuned filters which can be approximated by the Gabor filter. LBP does not approximate the HVS in this way.

Table 1
Summary of related work. Size: feature vector size.

Method	Feature	Generalization	Computational efficiency	Size	Recent usage
[3]	Local binary patterns (LBP)	Rotation invariance, robust to monotonic grayscale transformations	Uniform patterns reduce number of codes, sub-region histograms	5900	Baseline features for facial emotion recognition grand-challenges [6,8]; dynamic sampling approach [9]; survey with varying classifiers [10]
[11]	Three-patch and four-patch local binary patterns (TPLBP, FPLBP)	Robust to monotonic grayscale transformations	Sub-region histograms	Not stated	Used with prototype hyperplane learning on labeled faces in the wild dataset [12]
[13]	Local binary patterns from three orthogonal planes (LBP-TOP)	Robust to monotonic grayscale transformations, temporal information	Sub-region histograms	12	Action unit detection on the man machine interaction dataset [14]
[15]	Local phase quantization (LPQ)	Robust to blur	Sub-region histograms	25 600	Top approach in facial emotion recognition and analysis sub-challenge for discrete facial emotions [16]
[5]	Gabor energy filter	–	–	595 353[a]	Entry to AVEC grand-challenge [17]
[18]	Local Gabor binary pattern histogram sequence (LGBPHS)	Illumination invariance	Sub-region histograms	151 040	Survey with other LBP features [19]
[20]	Gabor energy filter histograms	–	Sub-region histograms	2400	Entry to AVEC grand-challenge [20]
[21]	Discrete cosine transform (DCT)	Difference image, accounts for motion only	DCT to compress difference image	$<$10 000[a, b]	Applied to AVEC grand-challenge [22]
[4]	Scale invariant feature transform (SIFT)	Scale, rotation and affine invariance	Histograms	10 624	Used with regional covariance matrix for multi-view face representation on BU-3DFE [23]
Proposed method	Background suppressing Gabor energy filtering	Removes background noise, robust to monotonic grayscale transformations	Maximal response determines significant edges, sub-region histograms	6400	Detection of microtubules in bioimaging data [24,25]

[a] Assuming the image was a square image of 100 pixels width.
[b] Varies based on how much energy is retained by the DCT.

(2) *Computational efficiency*: Gabor energy filters produce a response for each filter in its bank. The feature dimensionality of a Gabor feature vector is a product of the size of the image by the number of scales and the number of orientations. For example, a Gabor energy filter bank at six orientations, three scales, and a square image of 150 × 150 results in a dimensionality of 405 000. The dimensionality of LBP is 5900 in Ref. [16] regardless of the image size. Ref. [20] addressed this with sub-region histograms, similar to LBP. However, their approach lacked a generalization step. Ref. [35] reduced Gabor feature vector size by computing features at 34 landmark points. While this greatly reduced the feature vector size, the landmark points were manually extracted. Ref. [18] reduced Gabor feature vector size by combining Gabor filters with LBP. We propose to combine maximal edge response and soft orientation histograms to create a compact representation for emotion recognition.

(3) *White box feature*: Some state-of-the-art features, such as LBP, are visually incomprehensible to a human, thus they are like a black box where it is hoped that the classifier will capture a pattern human eyes cannot see. The proposed method is a white box because it produces contours that are visually comprehensible by humans.

2.2. Contribution

We contribute a novel method that improves the Gabor energy filter. It generalizes well because of its ability to suppress background texture. It has a low feature vector dimensionality because of soft orientation histograms. We demonstrate its efficacy on the non-trivial AVEC 2012 grand-challenge dataset [6]. We thoroughly examine the impact on performance of each part of the algorithm on the CK+ dataset. The feature produces contours understandable by humans and quality of these contours are investigated with bio-imaging data.

3. Technical approach

The proposed system overview for extracting local appearance is described in Fig. 1: In the generalization step, (1) the input image is filtered by a bank of Gabor filters, all fixed in scale at the pixel-level and varying for N different orientations. (2) Background texture of the input image is estimated on a per-pixel basis and removed from the result of each filtered image. In the computational efficiency step, (3) the bank of responses is condensed into a *maximal response*, a representation that retains the most intense edges and their orientations across all of the filters in the bank. (4) The image is divided into $M \times M$ subregions to account for face morphology, and *soft orientation histograms*, where bin counts are weighted by the magnitudes of their edges, are computed for each region. The histograms from each sub-region are concatenated to form the feature vector for the input image.

3.1. Gabor energy filter

A Gabor filter is a band-pass filter that can detect edges of a specific orientation and scale. Conventionally, an image is filtered by many Gabor filters with different parameters, and the collection of filters is called a *bank*. Each filter in the bank is tuned to a different orientation and scale. Under specific conditions, the Gabor filter can approximate the behavior of the human visual system [36,37]. The first component of the human visual cortex that processes visual information is the V1 area, located in the occipital lobe. Parts of the V1 area form *cells*, and each cell responds to edges of a specific magnitude and orientation, called a *grating*. This is referred to as the classical receptive field [38]. The Gabor energy filter emulates this process by creating a bank of filters where each filter responds to a specific grating. Let f be an input image. A Gabor energy filter for a specific magnitude and orientation is:

$$g(x, y; \gamma, \theta, \lambda, \sigma, \phi) = e^{\left(\frac{\tilde{x}^2+\gamma^2\tilde{y}^2}{2\sigma^2}\right)} \cos\left(2\pi \frac{\tilde{x}}{\lambda} + \phi\right) \quad (1)$$

where x and y are the pixel location. γ is the spatial aspect ratio that is a constant, taken to be 0.5. It effects the eccentricity of the filter. θ is the angle parameter that tunes the filter to specific orientations. γ is the wavelength parameter that tunes the filter to specific spatial frequencies, or magnitudes. In pattern recognition, this is also referred to as scale. σ^2 is the variance. It determines the size of the filter. ϕ is the phase offset taken to be 0 and π. $\tilde{x}$ and $\tilde{y}$ are defined as:

$$\tilde{x} = x\cos\theta + y\sin\theta \quad (2)$$

$$\tilde{y} = -x\sin\theta + y\cos\theta \quad (3)$$

Conventionally, the scales and orientations in the bank are selected such that the half-magnitude of each filter overlaps with others [39]. The Gabor filter can be used as local appearance filter by tuning the filter to a local neighborhood while still varying the orientation: $\sigma/\lambda = 0.56$, and varying θ. Scale is an important factor for the Gabor filter, which we fix to compute edges in a local neighborhood in the same way that LBP computes a microtexture. For the rest of the paper, $g(x, y; \theta, \phi)$ is shorthand for the following: $g(x, y; 0.5, \theta, 7.14, 3, \phi)$. $f(x, y)$ is filtered by $g(x, y; \theta, 0)$ and $g(x, y; \theta, \pi)$ and the magnitude of both is taken to be the result. This is called the Gabor energy filter:

$$E(x, y; \theta) = \sqrt{((f * g)(x, y; \theta, 0))^2 + ((f * g)(x, y; \theta, \pi))^2} \quad (4)$$

where $(f * g)(x, y; \theta, \phi)$ is the convolution of $f(x, y)$ and $g(x, y; \theta, \phi)$.

3.2. Generalization step

Eq. (4) captures the edge information. It responds to edges in the same way a simple cell in the human visual system responds to a grating. However, the human visual system is able to detect edges in the presence of background texture. This is called the pop-out effect [38], and an example is given in Fig. 2. The complex cells in the human visual cortex estimate background texture to focus on edges that are

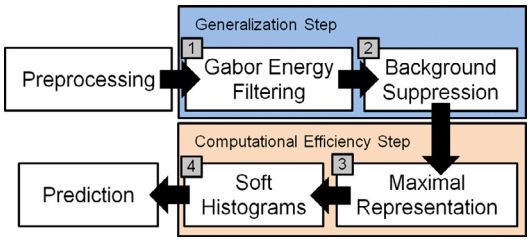

Fig. 1. System overview of the proposed texture descriptor.

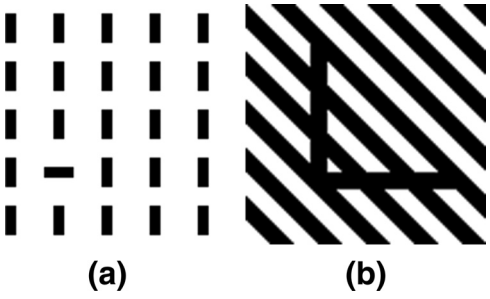

Fig. 2. Two examples of the pop-out effect. (a) In this image, the eye is drawn to the horizontal line because the repeated vertical lines form a background texture that is suppressed by the human visual system. (b) In this image, a triangle is presented along with a diagonal pattern. The removal of background texture suppresses one side of a triangle to give the illusion of an 'L'.

not consistent with the background texture. If the Gabor energy filter from Eq. (4) were applied to the images in Fig. 2, it would detect a high energy in the direction of the background texture, and a low energy for the orientations associated with the perpendicular line, or the 'L'. The background texture is referred to as the *Non-Classical Receptive Field*. In the conditions presented in Fig. 2, the human visual system suppresses the Non-Classical Receptive Field to better represent the edge information. The proposed feature should emulate this effect. The Non-Classical Receptive Field t is estimated as a weighted Gabor filter:

$$t(x, y; \theta) = (E * w)(x, y) \tag{5}$$

where the weight function w is:

$$w(x, y) = \frac{1}{\|\mathrm{DoG}(x, y)\|_1} g(\mathrm{DoG}(x, y)) \tag{6}$$

where $g(z) = H(z) * z$, where $H(z)$ is the Heaviside step function. $\mathrm{DoG}(x, y)$ is a difference of Gaussians:

$$\mathrm{DoG}(x, y; K, \sigma) = \frac{1}{2\pi K^2 \sigma^2} e^{-\frac{x^2+y^2}{2K^2\sigma^2}} - \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{7}$$

where K is a weight. σ^2 is the variance, the same as in Eq. (4). This ensures that the filter is bounded within the original Gabor filter. w resembles the ridges of a Mexican hat filter. When applied as the weight, Eq. (5) captures the edge information surrounding the current pixel. This allows background texture to be estimated on a per-pixel basis. The background suppressing Gabor energy filtered result $\tilde{b}$ is:

$$\tilde{b}(x, y; \theta) = g(E(x, y; \theta) - \alpha t(x, y; \theta)) \tag{8}$$

where α is a parameter that effects how much of the background texture is removed. When $\alpha = 0$, there is no background texture suppression, and the result is a Gabor energy filter. We constrain $\alpha = [0, 1]$. t is a weighted version of E, so α should not exceed 1.

3.3. Computational efficiency step

Eq. (8) retrieves the significant gratings of f less background texture. It is computed for N different orientations. Conventionally, the responses from the N orientations would be concatenated and taken to be the feature vector. A method is needed to reduce the feature size. A representation of $\tilde{b}$ is created that retains edges with maximum magnitude, for each pixel:

$$b(x, y) = \max\{\tilde{b}(x, y; \theta) | \theta = \theta_1, \ldots, \theta_N\} \tag{9}$$

Eq. (9) is called the maximal response. Separately, an orientation map $\Theta(x, y)$ is constructed that contains the orientation of the dominant edge in the maximal response, for each pixel:

$$\Theta(x, y) = \mathrm{argmax}_\theta\{\tilde{b}(x, y; \theta) | \theta = \theta_1, \ldots, \theta_N\} \tag{10}$$

Eqs. (9) and (10) retain the information of the most dominant edge. b retains the value of the maximum edge intensity, across all orientations, and Θ stores the specific orientation of the maximal edge. The image f is divided into M^2, equally sized, non-overlapping subregions. LBP and LPQ features use a *hard histogram*. That is, a histogram is computed that counts the number of microtextures. We use a *soft orientation histogram* to represent each sub-region. Instead of equally counting the presence of each microtexture, the votes are weighted by their magnitude from the maximal representation:

$$h(\theta_i) = \sum_{\forall(x,y)|\Theta(x,y)=\theta_i} b(x, y) \tag{11}$$

where $h(\theta)$ is an N bin histogram. A histogram is computed in each grid. The $M \times M$ grids are concatenated to form the feature vector for f.

From the parameters defined in Section 4.2, the feature size of the Gabor filter is 1 048 576 (64 orientations, 1 scale, and an image size of 128 × 128 pixels), whereas LBP feature size is 5900 (a product of a 59-bin histogram and gridding where $M = 10$). The feature size of the proposed method is a product of N and M^2. For $N = 64$ and $M = 10$, the size is 6400. Compared to the Gabor filter, the feature size is reduced by a factor of 163.84.

4. Experiments

4.1. Facial emotion recognition pipeline

Face regions-of-interest are detected with a cascade of Haar-like features [40]. The faces are registered with Avatar Image Registration, which is run for three iterations, based on tests in Ref. [16]. The features in Table 1 are compared to the proposed method. For regression, an ϵ-SVR detects the emotion label intensity and, for classification, a linear SVM detects classes [41].

4.2. Experimental parameters

Parameters are the same as in related work. For the background suppressing Gabor energy filtering: $\sigma/\lambda = 0.56$, values of θ were selected such that $\theta_{N+1} = \pi, \sigma = 4, \alpha = 1$, which are the same as in Ref. [22]. For Eq. (7), $K = 4$ and is chosen from previous work [38]. N in the computational efficiency step is taken to be 64. All local histograms are calculated in neighborhoods of 10 × 10. LBP parameters are the same as in Ref. [6]. TPLBP and FPLBP parameters are the same as in Ref. [11]. Images are resized to 128 × 128 before processing. For the Gabor energy filter, there are four scales at {1.0, 2.6, 6.8, 17.9} and eight orientations selected evenly from the range $[0, \pi]$, with all of the responses concatenated to make the feature vector. For LBP-TOP, the radii parameters for x and y are 1. For LGBPHS, the parameters are the same parameters used for the Gabor filter and LBP. For EQP, we use the parameters in Ref. [26].

4.3. Datasets

The first dataset is the AVEC 2012 grand-challenge dataset [6]. This dataset is needed to demonstrate the performance of the proposed method versus other texture descriptors. AVEC 2012 quantizes discrete emotional states with the Fontaine emotion model [42]. Emotion is described in terms of: *valence, arousal, power and expectancy*. For a more detailed explanation the reader is referred to Fontaine et al. [42]. Because AVEC 2012 has such a high number of frames, it is computationally undesirable to load all the frames into memory. We reduce this cost by downsampling the frame rate evenly by a factor sufficient to load all of the feature vectors into memory. We limit the cost to 8 GB, which is the most common system RAM size for PCs according to a recent hardware survey [43].

The second dataset is CK+ [7]. It consists of 593 videos of 123 different individuals. This dataset is used to compare the impact of performance for different parts of the proposed algorithm. A person faces a video camera and acts out an expression. Expressions are quantized in terms of facial action units (AU) [44]. AUs are the minimal set of muscle movements used in facial expressions. In CK+, they are tagged for each video. An algorithm must detect which AUs are present in each the video.

There are four bioimaging datasets: (1) From Ref. [22], 20 images of the pavement cells in *Arabidopsis thaliana* with GFP-tagged microtubules (Tub-A). The method must detect the microtubules. (2) Forty-four images of pavement cells in *A. thaliana* captured with transient light. The method must detect the cell walls. (1) and (2) are both captured using the Leica SP2. (3) Seventy-seven images of cell compartments in *Neurospora crassa* with the chitin tagged with calcofluor-white. The images were captured with a DAPI filter. The task is to detect the cell walls. (4) Five hundred and ninety-nine video frames of pollen tubes of *A. thaliana* with a GFP-tagged membrane

Table 2
Results on AVEC 2012 development set frame-level sub-challenge. For correlation, higher is better. Bold: best performer. Underline: second best performer. Size: the size of the feature vector, smaller is better. Ar: arousal. Exp.: expectancy. Val.: valence. Pow.: power. Prop: proposed.

Feature	Ar.	Exp.	Val.	Pow.	Avg.	Fact.[a]	Time[b] (s)	Size
DCT	0.034	0.078	0.076	0.063	0.063	1.1	**0**.01	8192
FPLBP	**0.425**	0.108	0.291	0.093	0.229	**1.0**	5.04	200
Gabor	0.059	0.019	0.063	0.012	0.036	70.3	3.02	5.2e5
Gabor histogram	0.171	0.080	0.082	0.067	0.100	**1**.0	4.11	2048
LBP	0.434	0.072	0.257	0.088	0.213	**1.0**	0.26	5900
LBP-TOP	0.389	0.092	0.177	0.084	0.186	**1.0**	0.29	**177**
LGB-PHS	0.131	0.091	0.143	0.066	0.107	25.65	0.39	1.9e5
LPQ	0.032	0.085	0.072	0.076	0.066	3.5	0.25	2.6e4
SIFT	0.037	0.038	0.073	0.048	0.049	3.5	0.04	2.6e4
TPLBP	0.024	0.047	0.086	0.039	0.049	283.7	3.40	2.1e6
Prop.	0.417	**0.143**	**0.347**	**0.124**	**0.258**	**1.0**	0.42	6400

[a] The downsampling factor applied to the frame rate to fit all of the feature vectors into memory; smaller is better and 1.0 indicates that all the feature vectors fit into memory without requiring downsampling.
[b] The average time to process a single face image with the parameters in Section 4.2 using face images from CK+. Computer: Windows 7, Intel Core 2 Duo E8500, Intel Q45 chipset, and 4 GB DDR at 533 MHz using MATLAB.

Table 3
Breakdown of performance of the different parts of the proposed method for different facial action units on CK+. TP: true positive. FP: false positive. FN: false negative. TN: true negative. PR: precision. RE: recall.

AU	TP	FP	FN	TN	PR	RE	F_1
(a) Generalization step only							
1	0.69	0.31	0.04	0.96	0.69	0.94	0.79
2	0.67	0.33	0.03	0.97	0.67	0.96	0.79
4	0.51	0.49	0.02	0.98	0.51	0.96	0.67
5	0.57	0.43	0.06	0.94	0.57	0.91	0.70
6	0.76	0.24	0.05	0.95	0.76	0.94	0.84
7	0.80	0.20	0.08	0.92	0.80	0.91	0.85
12	0.23	0.77	0.03	0.97	0.23	0.90	0.37
17	0.90	0.10	0.21	0.79	0.90	0.82	0.86
25	0.76	0.24	0.69	0.31	0.76	0.53	0.62
(b) Computational efficiency step only							
1	0.73	0.27	0.05	0.95	0.73	0.94	0.82
2	0.74	0.26	0.03	0.97	0.74	0.96	0.83
4	0.60	0.40	0.03	0.97	0.60	0.96	0.74
5	0.64	0.36	0.06	0.94	0.64	0.91	0.75
6	0.82	0.18	0.05	0.95	0.82	0.94	0.88
7	0.84	0.16	0.09	0.91	0.84	0.91	0.87
12	0.33	0.67	0.03	0.97	0.33	0.91	0.49
17	0.93	0.07	0.22	0.78	0.93	0.81	0.86
25	0.86	0.14	0.02	0.98	0.86	0.97	0.91
(c) Proposed method							
1	0.77	0.23	0.06	0.94	0.77	0.93	0.84
2	0.79	0.21	0.04	0.96	0.79	0.95	0.86
4	0.67	0.33	0.03	0.97	0.67	0.95	0.78
5	0.69	0.31	0.07	0.93	0.69	0.91	0.78
6	0.76	0.24	0.05	0.95	0.76	0.94	0.84
7	0.80	0.20	0.08	0.92	0.80	0.91	0.85
12	0.41	0.59	0.04	0.96	0.41	0.91	0.56
17	0.95	0.05	0.24	0.76	0.95	0.80	0.87
25	0.92	0.08	0.03	0.97	0.92	0.97	0.94

bound protein. The task is to detect the cell membrane. These four datasets are used to demonstrate the quality of edges detected with the proposed method.

4.4. Performance metrics

In the AVEC 2012 scoring system, each emotion's value is given a real number. The task is a regression problem. The algorithm detects the real valued emotion on a per-frame basis. While there are many metrics that could be used to quantify performance, the official metric for AVEC 2012 is Pearson product-moment correlation coefficient with the ground-truth [6]. Results for CK+ are given in terms of true positive rate, false positive rate, false negative rate, true negative rate, precision, recall and F_1-score. Results on bio-imaging data are given in terms of ROC-plots and the area under the curve (AUC).

Table 4
Results on the AVEC 2012 development set in terms of correlation with the ground truth for varying values of α.

α	0.000	0.250	0.500	0.750	1.000
Correlation	0.062	0.099	0.185	0.213	0.258

4.5. Emotion recognition results

The parameter α was selected empirically and results are given in Table 4 where it is found that $\alpha = 1$ gives the best performance.

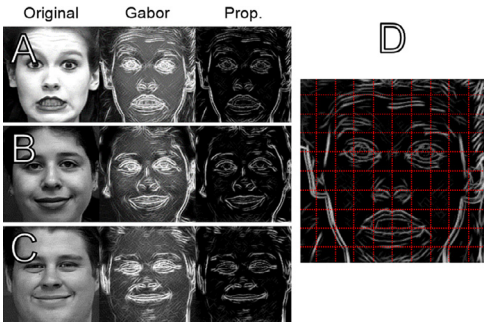

Fig. 3. Maximal response of the generalization step applied to faces from CK+. (A–C) Example faces that have been filtered by the Gabor filter and the proposed generalization step. (A) Note that the visible teeth form a pattern that is removed by the proposed generalization step. (D) Boundaries of sub-regions for soft orientation histograms.

Results on AVEC 2012 are given in Table 2. Results are given on the development set, and they are generated with a three-fold cross validation. We use the same folds from previous work [22]. In Table 2, average indicates the average correlation among the four labels of arousal, expectancy, valance and power. Size indicates the feature vector size. There is a clear dichotomy in the performance. There are three categories of performance. The proposed method, FPLBP, and LBP and LBP-TOP are the best performers. Gabor histograms and LGB-PHS have mid-grade performance. DCT, Gabor, LPQ, SIFT, and TPLBP are the worst performers. The proposed method does better than the other methods in the categories of expectancy, valence and power. FPLBP performs better for arousal, but its variance is higher. Note that the pairing of LPQ and Avatar Image Registration was the best performer in the facial emotion recognition and analysis, discrete emotional states sub-challenge [16]. LBP and FPLBP are comparable in performance to the proposed method. However, LBP and FPLBP rely on the existence of coded microtextures. An LBP image of 8 neighbors and a radius of 1 is challenging to understand with the human eye. The proposed method produces a visually understandable contour map to humans. An example of background texture suppression is given in Fig. 3. DCT and SIFT are the fastest methods, but the comparison may not be fair. A DCT is built into MATLAB. SIFT is programmed in C++ and interfaced to MATLAB with MEX.

4.6. Impact of generalization and computational efficiency

In this section, we explore the impact on performance from the generalization step and computational efficiency step. The three methods are compared: (1) a background suppressing Gabor energy feature bank is used as a feature. The response of each filter is concatenated to form the feature vector. This represents the generalization step without the computational efficiency step. (2) The second method is the computational efficiency step applied to a Gabor energy filter, without the background suppression. This method represents the computational efficiency step without the generalization step. (3) The third method is the proposed method. We use the CK+ dataset. For class probability rates please refer to Ref. [9]. The true positive rate, false positive rate, false negative rate, true negative rate, precision, recall and F_1-score are given in Table 3. The negative samples greatly outnumber the positive samples, so the true negative rate is very high for all the AUs, except for AU17, which has a positive rate of 0.76. For this reason, more attention should be paid to the true positive, false negative and F_1-score. The generalization step by itself without the computational efficiency step is the worst performer of the three in all metrics. This is due to the large feature dimensionality. Because there are 64 filters in the bank, the feature vector size is 1 048 576.

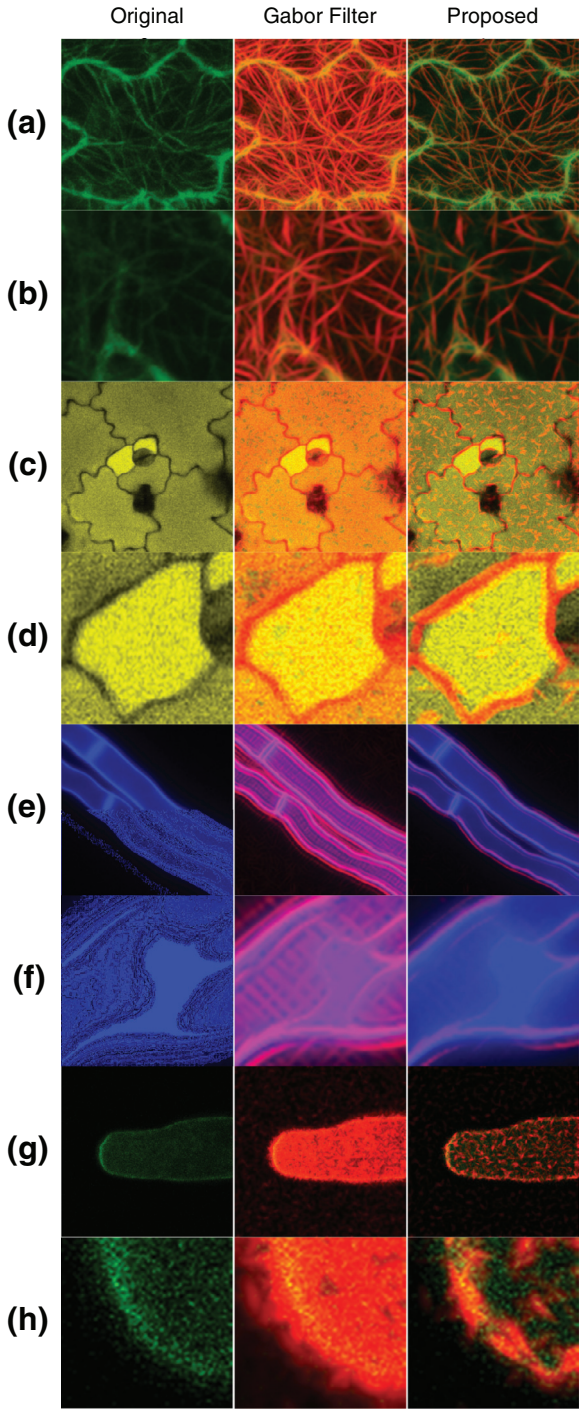

Fig. 4. Edge detection results: (a and b) Microtubules of *A. thaliana* pavement cells. Green: microtubule. (c and d) Transient light image of *A. thaliana* pavement cells. Yellow: cell. (e and f) Cell membrane of *N. crassa* hyphal stems. Blue: cell membrane. (g and h) Membrane bound protein of *A. thaliana* pollen tubes. Green: protein. For all, red/pink: detected edges. (b, d, f, and h) Close ups demonstrating false alarms with the Gabor filter. We recommend viewing this figure in color. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article)

Table 5
Summary of experiments from Table 3 in terms of average F_1-score across all AUs. Higher is better.

Method	Average F_1-score
Generalization step only	0.72
Computational efficiency step only	0.80
Proposed method (both steps)	0.81

Table 6
Results on the bioimaging dataset in terms of classification rate. Bold: best performer. Underline: second best performer.

Dataset	(1) *A. thaliana* microtubules	(2) *A. thaliana* cell membrane	(3) *N. crassa* cell membrane	(4) *A. thaliana* pollen tube
		Proposed method		
$\alpha = 0.0$	0.706 ± 0.043	0.741 ± 0.040	0.771 ± 0.015	0.709 ± 0.032
$\alpha = 0.5$	0.723 ± 0.053	0.766 ± 0.034	0.854 ± 0.021	0.721 ± 0.025
$\alpha = 1.0$	$\mathbf{0.858 \pm 0.080}$	$\mathbf{0.793 \pm 0.019}$	$\mathbf{0.910 \pm 0.041}$	$\underline{0.726 \pm 0.013}$
		Comparison to related work		
Imaginary Gabor [33]	0.712 ± 0.062	0.731 ± 0.070	0.799 ± 0.041	0.711 ± 0.054
EQP [26]	$\underline{0.725 \pm 0.080}$	$\underline{0.763 \pm 0.031}$	$\underline{0.869 \pm 0.053}$	$\mathbf{0.738 \pm 0.022}$

The computational efficiency step by itself and the proposed method has a feature vector size of 6400. Also, because each pixel is taken to be a feature, there is an extreme sensitivity to alignment. Histograms in local regions allow for some tolerance of registration errors, which is why histograms were adopted for use in LBP and LPQ features. The pairing of generalization and computational efficiency is always the best performer. A summary comparing the average F_1-score values is given in Table 5, where it can be concluded that the proposed method has the best average F_1-score across all AUs.

The true negative rate for AU25 is low when using the generalization step only. AU25 is lips part, and when a person's lips part, they open their mouth and reveal their teeth. The teeth form a pattern of perpendicular lines which are removed by the generalization step (see Fig. 3(D)). This important information is lost, explaining the poor performance. Note that the true negative rate is the highest when using only the computational efficiency step, which does not remove the teeth pattern. However, the contours of the lips fall in different sub-regions when the mouth is open. This information is captured by the sub-region histograms (see where the grid lines fall in Fig. 3(D)), which explains why the generalization step paired with the computational efficiency step has a better performance.

4.7. Bioimaging results

In this section, we present results comparing the quality of edges retained by the proposed method, versus the Gabor filter. The technical approach is as follows: a frame of bio-imaging data is filtered, and the contours, from Eq. (9), are used as features for a linear SVM. We use three-fold cross-validation. Bioimaging results are given in Table 6 for varying values of α. For the proposed method, α is the level of background suppression. $\alpha = 0$ is equivalent to a Gabor filter. We also compare our results to the imaginary part of Gabor [33] and elongated quinary patterns (EQP) [26]. The proposed method is the best performer, for all but one dataset. The quinary aspect of EQP allowed it to capture the large difference between the background pixel intensity and the edge of the *A. thaliana* pollen tubes. We posit that this is why EQP could better detect the cell membrane of the *A. thaliana* pollen tubes than the proposed method. However, the ability of the proposed method to remove background texture greatly increased the performance for all other datasets.

Examples of edge detection are given in Fig. 4. In Fig. 4(b) and (d) it can be seen that the Gabor filter detects many more edges than the proposed method because background noise is not removed in the filtering process. In (f) the Gabor filter detected a cross-hatching pattern that does not exist in the original image. In (h), the Gabor filter detects edges everywhere, whereas the objective of the task was to detect the edges of the pollen tube.

5. Conclusions

In this paper, we proposed a novel procedure that extended the Gabor filter to be robust against background noise and reduced the feature vector size by a factor of 126.56, when comparing the proposed method to a Gabor energy filter [39]. The proposed texture descriptor was found to have competitive performance on the AVEC 2012 dataset. It was demonstrated that the generalization step and the computational efficiency step improved classification accuracy, and that even more performance is improved by combining the two parts of the proposed algorithm. It was also shown that the edges detected by the proposed method are more meaningful than a Gabor filter on bio-imaging data.

Acknowledgements

Support for this work was provided for in part by NSF grant numbers 0727129, 0905671 and NSF IGERT: Video Bioinformatics grant number DGE 0903667. The contents and information do not reflect the position or policy of the U.S. Government.

References

[1] D. Gabor, Theory of communication, J. Inst. Elec. Eng. 93 (1946) 429–459.
[2] X. Wu, B. Bhanu, Gabor wavelets for 3-D object recognition, IEEE Trans. Image Process. 6(1) (1997) 47–64.
[3] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. PAMI 24(7) (2002) 971–987.
[4] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60(2) (2004) 91–110.
[5] M. Lyons, S. Akamatsu, Coding facial expressions with Gabor wavelets, in: IEEE Conference on Automatic Face and Gesture Recognition Workshops, 1998.
[6] B. Schuller, M.F. Valstar, F. Eyben, R. Cowie, M. Pantic, AVEC 2012 the continuous audio/visual emotion challenge, in: ACM International Conference on Multimodal Interaction, 2012.
[7] P. Lucey, J. Cohn, T. Kanade, J. Saragih, Z. Ambadar, The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression, in: IEEE Conference on CVPR Workshops, 2010.
[8] M. Valstar, M. Mehu, J. Bihan, M. Pantic, K. Scherer, The first facial expression recognition and analysis challenge, IEEE Trans. SMC B 42(4) (2011) 966–979.
[9] A. Cruz, B. Bhanu, N.S. Thakoor, Vision and attention theory based sampling for continuous facial emotion recognition, IEEE Trans. Affect. Comput. (2014).
[10] C. Shan, S. Gong, P.W. McOwan, Facial expression recognition based on local binary patterns: a comprehensive study, Image Vis. Comput. 27 (2009) 803–816.
[11] L. Wolf, T. Hassner, Y. Taigman, Effective unconstrained face recognition by combining multiple descriptors and learned background statistics, IEEE Trans. PAMI 33(10) (2011) 1978–1990.
[12] M. Kan, D. Xu, S. Shan, W. Li, X. Chen, Learning prototype hyperplanes for face verification in the wild, IEEE Trans. Image Process. 22(8) (2013) 3310–3316.
[13] G. Zhao, M. Pietikainen, Dynamic texture recognition using local binary patterns with an application to facial expressions, IEEE Trans. PAMI 29(6) (2007) 915–928.
[14] B. Jiang, M. Valstar, B. Martinez, M. Pantic, A dynamic appearance descriptor approach to facial actions temporal modeling, IEEE Trans. SMC B 44(2) (2013) 161–174.
[15] J. Heikkila, V. Ojansivu, Blur insensitive texture classification using local phase quantization, in: Image and Signal Processing, Springer, 2008.
[16] S. Yang, B. Bhanu, Understanding discrete facial expressions in video using an emotion avatar image, IEEE Trans. SMC B 42(4) (2012) 920–992.
[17] M. Glodek, S. Tschenchne, G. Layher, M. Schels, T. Brosch, S. Scherer, M. Kachele, M. Schmidt, H. Neumann, G. Palm, F. Schwenker, Multiple classifier systems for the classification of audio-visual emotional states, in: Affective Computing and Intelligent Interaction Workshops, Lecture Notes in Computer Science, Springer, 2011.
[18] W. Zhang, S. Shan, W. Gao, X. Chen, H. Zhang, Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition, in: IEEE International Conference on Computer Vision, 2005.
[19] S. Moore, R. Bowden, Local binary patterns for multi-view facial expression recognition, Comput. Vis. Image Understand. 115 (2011) 541–558.
[20] M. Dahmane, J. Meunier, Continuous emotion recognition using Gabor energy filters, in: Affective Computing and Intelligent Interaction Workshops, Lecture Notes in Computer Science, Springer, Berlin, 2011.
[21] L. Ma, K. Khorasani, Facial expression recognition using constructive feedforward neural networks, IEEE Trans. SMC B 34(3) (2004) 1588–1595.

[22] A. Cruz, B. Bhanu, N. Thakoor, Facial emotion recognition with anisotropic inhibited Gabor energy histograms, in: IEEE International Conference on Image Processing, 2013.
[23] W. Zheng, H. Tang, Z. Lin, T. Huang, Emotion recognition from arbitrary view facial images, in: European Conference on Computer Vision, 2010.
[24] G. Harlow, A. Cruz, L. Shuo, N. Thakoor, A. Bianchi, J. Chen, B. Bhanu, Z. Yang, Automated spatial analysis of ARK2: a key microtubule and cell polarity link, in: IEEE International Symposium on Biomedical Imaging, 2013.
[25] G. Harlow, A. Cruz, B. Bhanu, Z. Yang, S. Li, N. Thakoor, A.C. Bianchi, J. Chen, Pillars of plant cell polarity, 2013. http://posterhall.org/igert2013/posters/370.
[26] L. Nanni, A. Luminni, S. Brahnam, Local binary patterns variants as texture descriptors for medical image analysis, Artif. Intell. Med. 49(2) (2010) 117–125.
[27] A. Bianchi, B. Bhanu, Dynamic low-level context for the detection of mild traumatic brain injury, IEEE Trans. Biomed. Eng. (2014).
[28] X. Huang, G. Zho, W. Zheng, M. Pietikainen, Spatiotemporal local monogenic binary patterns for facial expression recognition, IEEE Signal Process. Lett. 19(5) (2012) 243–246.
[29] Y. Xu, J. Liu, Z. Zhang, N.M. Tan, D.W.K. Wong, S.M. Saw, T.Y. Wong, Learn to recognize pathological myopia in fundus images using bag-of-feature and sparse learning approach, in: IEEE International Symposium on Biomedical Imaging, 2013.
[30] Y. Qazaefi, S. Paris, J. Lefevre, C. Gaudy, J.J. Grob, B. Fertil, Learning from examples to automatically cluster pigmented skin lesions, in: IEEE International Symposium on Biomedical Imaging, 2013.
[31] B.X. Guan, B. Bhanu, P. Talbot, S. Lin, N. Weng, Comparison of texture features for human embryonic stem cells with bio-inspired multi-class support vector machine, in: IEEE International Conference on Image Processing, 2014.
[32] J. Yu, B. Bhanu, Evolutionary feature synthesis for facial expression recognition, Pattern Recognit. Lett. 27(11) (2006) 1289–1298.
[33] Y. Lei, M. Wang, T. Sun, G. Chen, Y. Liu, Z. Liu, The study of edge detection of cerebrovascular image based on gabor filter, in: IEEE Conference on Engineering in Medicine and Biology, 2005.
[34] T.F. Wu, C.J. lin, R.C. Wend, Probability estimates for multi-class classification by pairwise coupling, J. Mach. Learn. Res. 5 (2004) 975–1005.
[35] W. Zheng, X. Zhou, C. Zou, L. Zhou, Facial expression recognition using kernel canonical correlation analysis, IEEE Trans. Neural Netw. Lett. 171 (2006) 233–238.
[36] J.G. Daugman, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, J. Opt. Soc. Am. 2(7) (1985) 1160–1169.
[37] S. Marcelja, Mathematical description of the responses of simple cortical cells, J. Opt. Soc. Am. 70(11) (1980) 1297–1300.
[38] C. Grigorescue, N. Petkov, M.A. Westenberg, Contour detection based on nonclassical receptive field inhibition, IEEE Trans. Image Process. 12(11) (2003) 729–739.
[39] J.R. Movellan, Tutorial on Gabor filters, Technical Report, MPLab, 2008.
[40] P. Viola, M. Jones, Robust real-time face detection, Int. J. Comput. Vis. 57(2) (2004) 137–154.
[41] C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 27 (2011) 1–27.
[42] J.R.J. Fontaine, K.R. Scherer, E.B. Roesch, P.C. Ellsworth, The world of emotions is not two dimensional, Psychol. Sci. 18(12) (2007) 1050–1057.
[43] Valve, Steam hardware & software survey, September 2013, http://store.steampowered.com/hwsurvey.
[44] P. Ekman, W. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement, Consulting Psychologists Press, 1978.