Local Binary Patterns Calculated Over Gaussian Derivative Images

Varun Jain, James L. Crowley, Augustin Lux

▶ To cite this version:

Local Binary Patterns Calculated Over Gaussian Derivative Images

Varun Jain   James L. Crowley   Augustin Lux

INRIA

Univ. Grenoble Alpes, LIG, F-38000 Grenoble, France

CNRS, LIG, F-38000 Grenoble, France

firstname.lastname@inria.fr

Abstract—**In this paper we present a new static descriptor for facial image analysis. We combine Gaussian derivatives with Local Binary Patterns to provide a robust and powerful descriptor especially suited to extracting texture from facial images. Gaussian features in the form of image derivatives form the input to the Linear Binary Pattern(LBP) operator instead of the original image. The proposed descriptor is tested for face recognition and smile detection. For face recognition we use the CMU-PIE and the YaleB+extended YaleB database. Smile detection is performed on the benchmark GENKI 4k database. With minimal machine learning our descriptor outperforms the state of the art at smile detection and compares favourably with the state of the art at face recognition.**

I. INTRODUCTION

The face is one part of the body which is usually visible without much occlusions. It is the most readily available feature to distinguish people. Faces are also mirrors to our emotional state and we tend to express our emotions through the face both consciously and subconsciously. Imparting to computers the ability to analyse faces is one of the big challenges faced by the computer vision community. Facial Image Processing has traditionally been performed using Gabor filters and LBP features for static images [1]. In the present work we combine two powerful descriptors: Gaussian derivatives and LBP to obtain a new descriptor for facial image analysis. Section 2 gives a brief description of Gaussian derivatives and a few equations associated with their calculation. Gaussian derivatives have been used in a variety of image processing applications such as face detection [2], age estimation [3], head pose estimation [4] and facial expression analysis. Although they provide a compact description of the image texture and are quick to compute due to the availability of linear time complexity algorithms they are not invariant to illumination changes. LBP features discussed in Section 3 on the other hand are quite robust to illumination changes. LBP and its modifications have been exploited for all kinds of diverse tasks such as biometrics including eye localization, iris recognition, fingerprint recognition, palm-print recognition, gait recognition and facial age classification [5]. Section 5 describes our approach where we combine the descriptors described in the last 2 sections. To exhibit the versatility of our descriptor, we test the technique on two applications: Face recognition which

is discussed in section 5 and smile detection discussed in section 6.

Experimental results in section 8 show that our descriptor with the minimal use of machine learning techniques outperforms the state of the art in smile detection and is at par with other techniques in face recognition.

II. GAUSSIAN DERIVATIVES

Gaussian derivatives efficiently describe the neighbourhood appearance of an image [4]. They provide a scale and orientation invariant description which can be used for a variety of applications such as detection, tracking, indexing and reconstruction. The Gaussian support function is described by the equation:

$$G(x, y; \sigma) = e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

Here σ is the scale factor or variance and defines the spatial support. The first order derivatives are of the form:

$$G_x(x, y; \sigma) = \frac{\partial G(x, y; \sigma)}{\partial x} = -\frac{x}{\sigma^2} G(x, y; \sigma) \quad (2)$$

$$G_y(x, y; \sigma) = \frac{\partial G(x, y; \sigma)}{\partial y} = -\frac{y}{\sigma^2} G(x, y; \sigma) \quad (3)$$

First order derivatives give information about the gradient (intensity and direction). The second order derivatives are given by:

$$G_{xx}(x, y; \sigma) = \frac{\partial^2 G(x, y; \sigma)}{\partial x^2} = (\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2})G(x, y; \sigma) \quad (4)$$

$$G_{yy}(x, y; \sigma) = \frac{\partial^2 G(x, y; \sigma)}{\partial y^2} = (\frac{y^2}{\sigma^4} - \frac{1}{\sigma^2})G(x, y; \sigma) \quad (5)$$

$$G_{xy}(x, y; \sigma) = \frac{\partial^2 G(x, y; \sigma)}{\partial x \partial y} = \frac{xy}{\sigma^4}G(x, y; \sigma) \quad (6)$$

First and second order derivatives are excellent descriptors for features such as bars, blobs and corners in images. Higher order features can describe more complicated structures but are difficult to exploit because of their sensitivity to noise.

IEEE computer society

III. Local Binary Patterns

Local binary patterns is a simple yet powerful descriptor for texture analysis [6]. It requires limited computational power and is ideal for real-time applications. Its robustness to monotonic gray-scale changes make it suitable for applications such as facial image analysis where variations in illumination can have major effects on appearance.
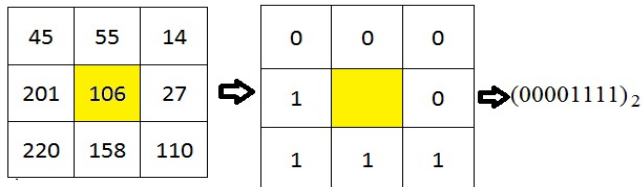

Fig. 1: Computing LBP response from a pixels local neighbourhood.

Many variations over the original LBP operator have been proposed. One extension allows LBP operator to use neighborhoods of different sizes [7]. Another modification introduced in [7] is the Uniform LBP which which can be used to reduce the length of the feature vector by using a smaller number of bins and is also invariant to rotation.

IV. Our method

In [8] Zhang proposed the calculation of LBP features over the results of filtering a facial image with 40 Gabor filters of different scales and orientations. Gaussian features take much less time to compute than Gabor features and have been shown to perform better in a variety of applications. Ruiz-Hernandez *et al.* in [9] create 3-d tensors after calculating censor transforms on Gaussian feature maps which are subsequently reduced in dimensionality using MPCA and then a kernel technique is used for face recognition. Recently a dual layered architecture was proposed by Wu *et al.* [10] which employed LBP and two layers of gabor features.

We propose to combine Gaussian derivatives with LBP and test the resulting description technique at Face recognition and Smile detection. In our method gaussian derivative images are produced from the normalized input images by using the Half octave gaussian pyramid [11] which allows for the fast calculation of gaussian derivatives.

First and second order derivative images of the following order are used: $I_x, I_y, I_{xx}, I_{yy}, I_{xy}$ from the base of the pyramid($\sigma = 1$). Next these derivative images are divided into grids of 4 X 4 local regions with 43.75% overlapping areas from which uniform LBP features were calculated. The local histograms are concatenated to obtain the final feature vector. Since we had 5 derivative images and 16 grids with each grid producing a uniform LBP histogram, we have a feature vector of : 5*16*59=4720 dimensions. Unlike the method introduced by the authors of [9], we do not use tensor mathematics and the number of gaussian features used by us is much lower than the number used in [9]. We followed the hill-climbing algorithm and started adding gaussian derivatives starting from the base

of the pyramid and stopped there because adding derivatives from the level above did not lead to an improvement in accuracy. On the other hand in [9] the authors use 6 levels of the pyramid.

The grid size of 4X4 and 43.75% of overlap area is chosen by means of cross-validation. Images are normalized to 66 X 66 pixels, this size is also chosen through cross-validation.
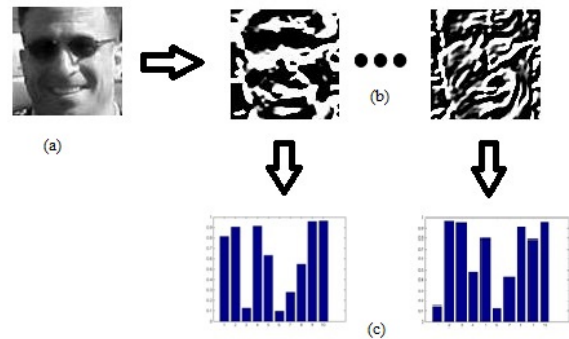

Fig. 2: Creating the features: a) original image, b) Gaussian derivative images, and c) concatenation of resulting histograms after applying LBP.

V. Face recognition

A. Related Work

Face recognition involves the identification of individuals from an image or video frame. A major challenge in face recognition is to make the system invariant to illumination since the appearance of the face can change dramatically with changes in lighting conditions. Other problems include aging, occlusions, pose and facial expressions.

People have experimented with feature based techniques for face recognition using methods such as elastic graphs in [12] where the authors generate a graph using fiducial points labeled with Gabor filter responses and in [13] where Gabor filters are replaced by HOG features. Holistic approaches are more popular involving descriptor calculation over the entire image rather than on local features of the face.

Li and Yin use the wavelet transform in conjunction with neural networks for face recognition[14]. In [15] the authors use the versatile descriptor LBP for face recognition over the FERET dataset. Ruiz-Hernandez *et al.* [9] combine LBP with Gaussian features maps and then generate a tensor which is then reduced in dimensions using Multilinear Principal Component Analysis and finally recognition is done with Kernel Discriminative Common Vector.

In [16] the authors deal with the problem of illumination by dealing with the effects of illumination on large scale and small scale features explicitly, they achieve the best results by combining their illumination normalization technique with quotient images[17]. Meanwhile in [18] an illumination invariant descriptor is presented for face recognition

which also claims to solve the bottleneck associated with heterogeneous lighting.

B. Experiments with Face Recognition

Our method is tested on the CMU-PIE and the extended YaleB databases. The performance our method is compared with the state of the art methods such as SQI [19], LTV [20], WF[21], Gradient Face [22], LGH [18] for both the databases and with the method used by Ruiz-Hernandez *et al.* in [9] on the YaleB database. Additionally we compare our performance with Gabor filters, LBP, Gaussian derivatives and LBP calculated over Gabor images for the CMU-PIE dataset. We use a subset of the original CMU-PIE database, 1428 frontal images from 68 people under 21 variations of illumination conditions are selected. No feature alignment method was used.

The results are presented in figure 3. Only one image per individual is used as the reference image. All the 21 images taken under different lighting conditions are chosen as the reference images one at a time. We use the L1 distance as the similarity measure. The reference image closest to the test image decides the identity of the person in the test image.
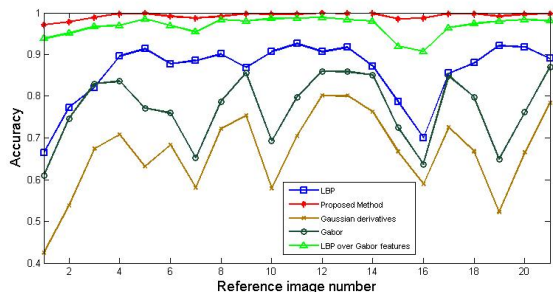

Fig. 3: Recognition rate with different reference images

Our approach achieves the highest accuracy of 100% with the image with frontal lighting. Only the performance of LBP calculated over Gabor images comes close to our method. Gaussian derivatives and LBP alone do not achieve very high recognition rates making the case for our method which combines the two.

The maximum and average recognition rates for the different descriptors are given in table 1.

	SQI	LTV	WF	GF	LGH	Ours
Max. Acc%	98.82	95.81	99.71	99.93	**100**	**100**
Avg. Acc%	89.77	80.78	89.52	96.93.93	98.19	**99.26**

TABLE I: Maximum and average accuracy attained by different methods

The extended YaleB dataset contains images from 28 individuals captured under 64 different lighting conditions with 9 pose views. We only use the images with frontal views in our experiments. Researchers have divided the database into 5 subsets in increasing order of complexity of lighting

conditions. We use the image number A+000E+00 with the simplest lighting scenario as the reference image. We use the Minkowski distance metric as the similarity measure. The p-Minkowski metric between two points a=(x1,y1) and b=(x2,y2) can be given as:

$$d^p(a,b) = [|x1 - x2|^p + |y1 - y2|^p]^{\frac{1}{p}} \qquad (7)$$

The optimum value of p varies from 0.75 to 1.25 on the YaleB database. The reference image closest to the test image decides the identity of the person in the test image. We compare the results of our technique with the state of the art in the following table.

	Set 1	Set 2	Set 3	Set 4
SQI	88.60	**100**	85.75	87.97
LTV	87.28	99.78	66.67	45.49
WF	79.39	99.78	75.88	77.07
GF	94.74	**100**	83.33	75.94
LGH	94.74	**100**	92.54	**96.43**
Ruiz-Hernandez[9]	**100**	**100**	94.7	60.1
Ours	**100**	**100**	**97.22**	79.10

TABLE II: Accuracy(%) over the 4 subsets using different methods

Our technique achieves the highest accuracy for the first 3 subsets, on the 4th subset it is beaten by SQI and LGH, which are both techniques that handle the problem of illumination explicitly. It is interesting to see how the two components of our approach namely Gaussian derivatives and LBP alone match up against their proposed combination.

	Set 1	Set 2	Set 3	Set 4
Gaussian Derivatives	97.53	93.52	62.65	25.13
LBP	99.38	99.38	55.86	33.07
Ours	**100**	**100**	**97.22**	**79.10**

TABLE III: Accuracy(%) over the 4 subsets using Gaussian derivatives, LBP and their proposed combination

VI. SMILE DETECTION

A. Related Work

A lot of people are working on facial expression analysis but very few focus on specifically on smile detection. In [23] McDuff *et al.* use smile intensity to predict how much a viewer likes a particular video. Smile detection is an integral part of emotional state estimation in humans. It also has a variety of applications in consumer surveys, gaming and user interfaces. A major issue in facial expression analysis is that most of the research is validated on posed databases. In [24] authors have argued that spontaneous expressions are different from posed expressions in both appearance and timing therefore systems developed for recognizing posed expressions might not work well on real world expressions. Spontaneous expressions are much more subtle and complex than posed expressions.

Most smile detection systems in the past have been trained on these posed databases. Deniz *et al.* in [25] present a smile detection system based on finding keypoints on the face and test their method on the DaFex and JAFFE datasets which

both contain posed smiles. Others in [26], [27], [28] have all experimented on posed databases.

The GENKI-4K database presented by Whitehill and others in [29] contains 4000 images with a wide range of subjects, facial appearance, illumination, geographical locations, imaging conditions and camera models. The images are annotated for smile content(1=smile, 0=non-smile). The difference between this dataset and other facial expression datasets is that this dataset was compiled from images on the internet rather than being captured in a controlled environment.

the location of the nose. A ten fold cross validation procedure was adopted to obtain the final results. Apart from measuring the accuracy we also measured the Balanced Error Rate(BER).

		Prediction	
		Class-1	Class+1
Truth	Class-1	a	b
	Class+1	c	d

TABLE IV: Confusion matrix for 2 class classification

The balanced error rate is the average of the errors on each class: $BER = 0.5 * (b/(a+b) + c/(c+d))$. Where a, b, c, d stand for: true negatives, false positives, false negatives and true positives respectively.
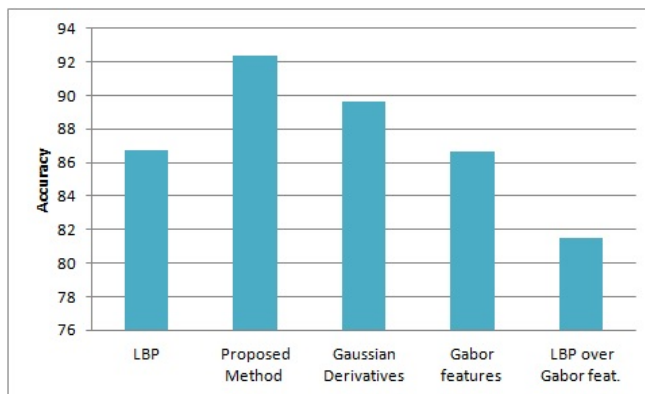
Fig. 5: Accuracy(%) of different descriptors over the GENKI-4K database

The proposed technique achieves the highest accuracy of 92.3602% with the lowest BER of 0.0702.

Fig. 6: BER of different descriptors over the GENKI-4K database

It is surprising to see that LBP calculated over Gabor features actually perform worse than both Gabor features and LBP alone. This could be because of the curse of dimensionality since we are using Support Vector Machines with a radial basis kernel and the feature vector of LBP calculated over
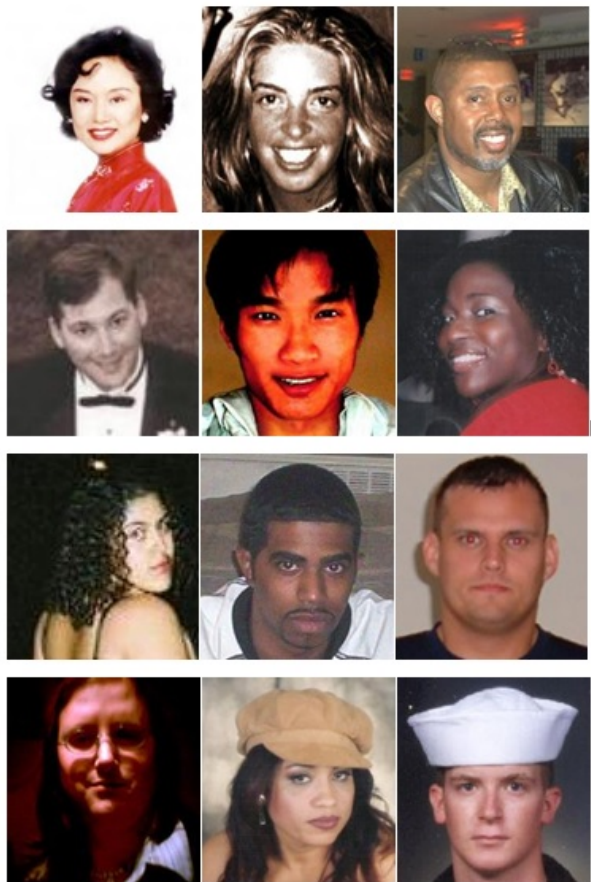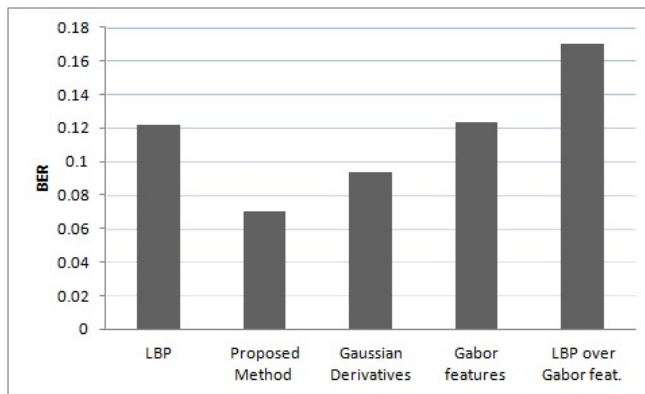

Fig. 4: Examples of (top two rows) real-life smile faces and (bottom two rows) nonsmile faces, from the GENKI4K database.

Shan in [30] presents a comprehensive study on smile detection and proposes his own method which is faster than the state of the art but not more accurate than Gabor filters combined with SVM's.

B. Experiments with Smile Detection

We perform our experiments on the GENKI-4K dataset. We use Support Vector Machines with a Radial Basis Kernel to compare the accuracies obtained by different descriptors. The images were not aligned using facial features such as eyes and

Gabor features has a dimensionality of nearly 17000 whereas the number of training instances is less than 3600.
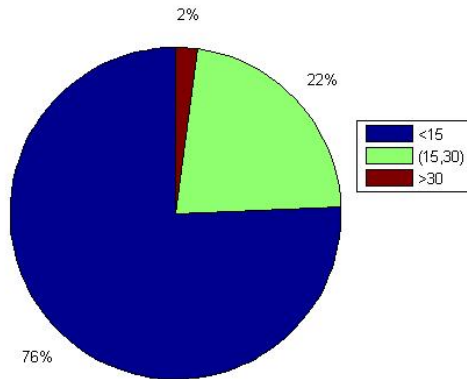

Fig. 7: Proportion of images with different pose

We divide the GENKI-4K database into 3 subsets according to the head pose(only yaw). The proportion of images that fall into the three sets is shown in figure 7.
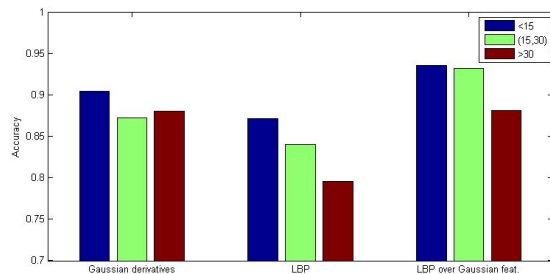

Fig. 8: Accuracy with different poses

As expected all the three techniques produce the best results for head pose close to frontal i.e. < 15 degrees, slightly lower for 15 to 30 degrees and lowest for yaw above 30 degrees except for Gaussian derivatives which perform slightly better at yaw above 30 degrees than they perform at yaw between 15 and 30 degrees. The most interesting aspect of the results is that Gaussian derivatives and LBP features calculated over Gaussian features are less susceptible to pose variation than LBP alone explaining why our method performs better than traditional LBP.

VII. CONCLUSION

This paper presents a versatile descriptor which performs well on two facial image processing tasks. It is simple to compute, robust to illumination changes and performs better than more computationally expensive methods. It is interesting to see that a general purpose descriptor like ours works better in most cases than the specialized descriptors such as LGH[18]

and SQI[19] for face recognition.

Its robustness to head pose variations helps to avoid expensive image alignment calculations. The success of the technique at smile detection suggests that it could be utilized for other facial expression problems.

REFERENCES

[1] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, 2006.

[2] J. A. Ruiz-Hernandez, A. Lux, and J. L. Crowley, "Face detection by cascade of gaussian derivates classifiers calculated with a half-octave pyramid," in *FG*, 2008, pp. 1–6.

[3] J. A. Ruiz-Hernandez, J. L. Crowley, and A. Lux, ""how old are you?" : Age estimation with tensors of binary gaussian receptive maps," in *BMVC*, 2010, pp. 1–11.

[4] V. Jain and J. L. Crowley, "Head Pose Estimation Using Multi-scale Gaussian Derivatives," in *18th Scandinavian Conference on Image Analysis*, Espoo, Finland, June 2013.

[5] J. Ylioinas, A. Hadid, and M. Pietikäinen, "Age classification in unconstrained conditions using lbp variants," in *ICPR*, 2012, pp. 1257–1260.

[6] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[7] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.

[8] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (lgbphs):a novel non-statistical model for face representation and recognition," in *ICCV*, 2005, pp. 786–791.

[9] J. A. Ruiz-Hernandez, J. L. Crowley, A. Méler, and A. Lux, "Face recognition using tensors of census transform histograms from gaussian features maps," in *BMVC*, 2009, pp. 1–11.

[10] T. Wu, N. J. Butko, P. Ruvolo, J. Whitehill, M. S. Bartlett, and J. R. Movellan, "Multilayer architectures for facial action unit recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 42, no. 4, pp. 1027–1038, 2012.

[11] J. L. Crowley, O. Riff, and J. H. Piater, "Fast computation of characteristic scale using a half-octave pyramid," in *In: Scale Space 03: 4th International Conference on Scale-Space theories in Computer Vision, Isle of Skye*, 2002.

[12] L. Wiskott, J.-M. Fellous, N. Kruger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," in *Intelligent Biometric Techniques in Fingerprint and Face Recognition*, L. C. Jain, U. Halici, I. Hayashi, and S. B. Lee, Eds. CRC Press, 1999, ch. 11, pp. 355–396.

[13] A. Albiol, D. Monzo, A. Martin, J. Sastre, and A. Albiol, "Face recognition using hog-ebgm," *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1537–1543, 2008.

[14] B. Li and H. Yin, "Face recognition using rbf neural networks and wavelet transform," in *ISNN (2)*, 2005, pp. 105–111.

[15] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *ECCV (1)*, 2004, pp. 469–481.

[16] X. Xie, W.-S. Zheng, J.-H. Lai, P. C. Yuen, and C. Y. Suen, "Normalization of face illumination based on large-and small-scale features," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 1807–1821, 2011.

[17] H. Wang, S. Z. Li, Y. Wang, and J. Zhang, "Self quotient image for face recognition," in *ICIP*, 2004, pp. 1397–1400.

[18] J.-Y. Zhu, W.-S. Zheng, and J.-H. Lai, "Logarithm gradient histogram: A general illumination invariant descriptor for face recognition," in *FG*, 2013, pp. 1–8.

[19] H. Wang, S. Z. Li, and Y. Wang, "Generalized quotient image," in *CVPR (2)*, 2004, pp. 498–505.

[20] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. S. Huang, "Total variation models for variable lighting face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1519–1524, 2006.

[21] B. Wang, W. Li, W. Yang, and Q. Liao, "Illumination normalization based on weber's law with application to face recognition," *IEEE Signal Process. Lett.*, vol. 18, no. 8, pp. 462–465, 2011.

[22] T. Zhang, Y. Y. Tang, B. Fang, Z. Shang, and X. Liu, "Face recognition under varying illumination using gradientfaces," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2599–2606, 2009.

[23] D. McDuff, R. E. Kaliouby, D. Demirdjian, and R. W. Picard, "Predicting online media effectiveness based on smile responses gathered over the internet," in *FG*, 2013, pp. 1–7.

[24] M. F. Valstar, M. Pantic, Z. Ambadar, and J. F. Cohn, "Spontaneous vs. posed facial behavior: Automatic analysis of brow actions," in *Proceedings of the 8th International Conference on Multimodal Interfaces*, ser. ICMI '06.   New York, NY, USA: ACM, 2006, pp. 162–170.

[25] O. Déniz, M. C. Santana, J. Lorenzo-Navarro, L. Antón-Canalís, and G. Bueno, "Smile detection for user interfaces," in *ISVC (2)*, 2008, pp. 602–611.

[26] Y. Shinohara and N. Otsu, "Facial expression recognition using fisher weight maps," in *FGR*, 2004, pp. 499–504.

[27] A. Ito, X. Wang, M. Suzuki, and S. Makino, "Smile and laughter recognition using speech processing and face recognition from conversation video," in *CW*, 2005, pp. 437–444.

[28] U. Kowalik, T. Aoki, and H. Yasuda, "Broaference - a next generation multimedia terminal providing direct feedback on audience's satisfaction level," in *INTERACT*, 2005, pp. 974–977.

[29] J. Whitehill, G. Littlewort, I. R. Fasel, M. S. Bartlett, and J. R. Movellan, "Toward practical smile detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 2106–2111, 2009.

[30] C. Shan, "Smile detection by boosting pixel differences," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 431–436, 2012.