A novel phase congruency based descriptor for dynamic facial expression analysis

Shojaeilangari, Seyedehsamaneh; Yau, Wei-Yun; Teoh, Eam Khwang

2014

Shojaeilangari, S., Yau, W.-Y., & Teoh, E. K. (2014). A novel phase congruency based descriptor for dynamic facial expression analysis. Pattern Recognition Letters, 49, 55-61.

https://hdl.handle.net/10356/81599

https://doi.org/10.1016/j.patrec.2014.06.009

A Novel Phase Congruency based Descriptor for Dynamic Facial Expression Analysis

Seyedehsamaneh Shojaeilangari[1*], Wei-Yun Yau[2], Eam-Khwang Teoh[1]

[1] *School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore*

[2]*Institute for Infocomm Research, A*STAR, Singapore*

[*]Corresponding author.Tel: (+65) 84534486, Fax: (+65) 67933318
E-mail: seyedehs1@e.ntu.edu.sg(SeyedehsamanehShojaeilangari).
Address: 50 Nanyang Ave, S1-B4C-14, Electrical and Electronic Engineering Department NTU, Singapore, 639798

Abstract

Representation and classification of dynamic visual events in videos have been an active field of research. This work proposes a novel spatio-temporal descriptor based on phase congruency concept and applied it to recognize facial expression from video sequences. The proposed descriptor comprises histograms of dominant phase congruency over multiple 3D orientations to describe both spatial and temporal information of a dynamic event. The advantages of our proposed approach are local and dynamic processing, high accuracy, robustness to image scale variation, and illumination changes. We validated the performance of our proposed approach using the Cohn-Kanade (CK^+) database where we achieved 95.44% accuracy in detecting six basic emotions. We also validated its robustness to illumination and scale variation using our own collected dataset.

Keywords:

Phase congruency; Spatio-temporal descriptor; Emotion recognition; Facial expression

1 Introduction

Detection of human facial expression from video sequences is a challenging problem due to real-world constrains such as background clutter, partial occlusion, viewpoint variations, scale changes, and lighting conditions. Finding a suitable feature representation is a vital step to model the facial expression and subsequently recognize it in a video sequence.

The traditional method for video representation is the extension of successful techniques used in static image analysis to support the dynamic requirement for video processing. 3D Scale Invariant Feature Transform (SIFT), Spatio-temporal Local Binary Pattern (LBP), and spatio-temporal descriptor based on 3D gradient are typical examples of video representation successfully applied in facial affect analysis or human action recognition [1, 2]. In this paper, we followed the same idea for feature representation by extending the Phase Congruency (PC) concept. We applied our proposed approach to dynamic facial emotion recognition from video sequences.

Local energy-based and phase-based models have emerged as successful tools to detect various image patterns such as step edges, corners, valleys, and lines. The phase-based feature extraction model proposes that the features of a signal are observed at locations where its Fourier components are in harmony. Such concept is also seen in the human visual system that the image features are perceived at points where the phase values of its Fourier components are maximally in congruence [3].

There are some advantages for PC-based feature extraction approaches over gradient-based techniques [4]. The gradient operators such as Prewitt, Sobel, Laplace, and Canny edge detector fail to precisely identify and localize all image features, especially in region affected by illumination changes. Unlike the gradient-based approaches which look for sharp changes of image intensity, PC is a dimensionless quantity which is robust to image contrast and illumination changes.

In Figure 1, we show the advantages of PC-based line detection over Canny and Sobel methods. This figure illustrates that PC is able to localize the sharp line similar to the gradient operators. However, for features that are not sharp (gradual intensity variation), PC is able to detect such feature better than the traditional gradient operators as shown in Figure 1. Indeed, PC captures the discontinuities even at small intensity differences which are missed by the typical image gradient-based edge descriptors. It can be helpful for facial features detection including skin folds due to aging and expression.

This paper explores the effectiveness of PC-based feature representation for video classification. The proposed descriptor, named Histogram of Dominant Phase Congruency (HDPC), comprises histograms of dominant PC over multiple 3D orientations to describe both the spatial and temporal information of a dynamic event.

To construct HDPC descriptor, the spatio-temporal PC values are calculated for multiple orientations. Therefore, each pixel of a video is characterized by multiple oriented PC values. Thus the PC values are able to encode various features at different scales and orientations for both spatial and time domains. After calculating the oriented PC values, the next step is to find the maximum PC for each pixel while preserving the dominant orientation. In other words, each pixel is represented by

a vector where its length is equal to maximum PC, and its direction is determined by the dominant orientation. Keeping the dominant PC and its orientation information will preserve the key feature contributing to a dynamic event. The final step of our novel descriptor is building a local histogram of PC directions over all pixels over a spatio-temporal patch.

The novelties of our proposed approach are:

(1) Extending the PC concept to spatio-temporal domain to extract both static and dynamic information from a video sequence by applying the 3D log-Gabor filter.

(2) Designing a bank of oriented 3D log-Gabor filters to detect the image features at various orientations.

(3) Selecting dominant PC to capture the most significant motion information while preserving its direction.

(4) Proposing histogram of dominant spatio-temporal PC to summarize the acquired information of each local 3D region.

This paper is organized as follows: Section 2 summarizes the literature review. The proposed method for feature extraction is explained in section 3, including computing the 3D PC for sequenced images, the proposed HDPC algorithm, and summary of the algorithm's properties. Section 4 and 5 describe the experimental results and conclusion respectively.

2 Related works

Automated analysis of facial expression has been the subject of many researches due to its potential applications such as human-computer interaction, automated tutoring systems, image and video retrieval, smart environments, and driver warning systems. Although considerable progress has been reported in the literature, there are still challenges for a robust and automated analysis. Most previous works are focused on facial emotion recognition via static images. The static analysis systems ignore the dynamics of facial expression due to expensive computational time involved or the complicated temporal mode [5-9]. However, it is confirmed by human visual system that the

judgement about an expression is more reliable when its temporal information is also taken into account [10].

To exploit the temporal information of facial expression, different techniques have been developed. There were several reported attempts to track the facial expression over time for emotion recognition via Hidden Markov Models (HMM). A multilevel HMM is introduced by Cohen et al. [11] to automatically segment the video and perform emotion recognition. Their experimental results indicated that the multilevel HMM have better performance than the one layered HMM. Cohen et al. [12] introduced a new architecture of HMMs for automatic segmentation and recognition of human facial expression from live videos.

Dynamic Bayesian Networks (DBN) is another successful method for sequence-based expression analysis. Kwang-Eun and Kwee-Bo [13, 14] developed a facial expression recognition system based on combining the Active Appearance Model (AAM) for feature extraction and DBN for modelling and analysing the temporal phase of an expression. They claimed that their proposed approach is able to achieve robust categorization of missing and uncertain data and temporal evolution of the image sequences.

Optical Flow (OF) is also a widely used approach for facial features tracking and dynamic expression recognition. Cohn et al. [15] developed an optical flow based approach to automatically discriminate the subtle changes in facial expression. They considered sensitivity to subtle motion when designing the OF which is crucial for spontaneous emotion detection.

Methods based on local features or interest points such as SIFT have shown to perform well for object recognition and then extended for video analysis. Camara-Chavez and Araujo [16] proposed a method for event detection in a video stream by combining Harris-SIFT with motion information in the context of human action recognition. They used the Harris corner detection for key-point extraction and the phase correlation method was used to measure the motion information.

Guoying and Pietikainen[1] presented a successful dynamic texture descriptor based on the Local Binary Pattern (LBP) operator and applied it on facial expression recognition as a specific dynamic

event. Their proposed dynamic LBP descriptors were calculated on Three Orthogonal Planes (TOP) of the video volume, resulting in LBP-TOP descriptor. Local processing, simple computation and robustness to monogenic gray-scale changes are the advantages of their method.

Dollar et al. [17] developed a general framework for dynamic behaviour detection from videos by proposing descriptors to encode the spatio-temporal cuboids surrounding the points of interest. Extracted cuboids are clustered to form a dictionary of cuboid prototypes and then the information of location and type of cuboid prototypes is kept for further processing. They argued that the proposed representation is robust to many data variations. Their experimental results on different databases including facial expression and human activity show that their method is applicable for these tasks.

Guha and Ward [18] explored the effectiveness of sparse representations in the context of facial expressions and human actions recognition. They extracted a set of spatio-temporal descriptors named Local Motion Pattern (LMP) for the key points of video sequences. A compact and rich representation was then suggested by learning the overcomplete dictionary and its corresponding sparse model. Their work presented a new local spatiotemporal feature that is distinctive, scale invariant, and fast to compute.

Our method does not require large training data which is needed for some techniques such as HMM, and also does not require key point extraction and feature tracking. The main disadvantage of gradient-based key point extraction is that any slight variation in the illumination will produce candidate key points similar to those produced by large motion. The inability to specify in advance what level of response corresponds to a significant feature is the shortcoming of many feature detectors. The PC based descriptor is able to detect a wide range of feature types and identifying their uniqueness in the image.

3 Methodology

To detect a dynamic event such as facial expression, the feature representation is a vital step that should be able to describe the visual event in both spatial and temporal domains. To obtain such a representation, we proposed the spatio-temporal PC concept since it is able to capture significant

information in both the spatial and temporal domains. Additionally, we designed a new descriptor called Histogram of Dominant Phase Congruency (HDPC) which encodes the dominant oriented PC of local 3D patches to increase its robustness to noise or outliers. More precisely, the proposed approach involves two main steps: (1) Calculating the spatio-temporal PC at various orientations, and (2) encoding the pixel's dominant oriented PC. We describe the details of each stage in the following subsections.

3.1 Spatio-temporal PC calculation

The monogenic signal framework provides the extended form of analytic signal to 3D by using a vector-valued odd filter (Riesz filter) which is represented in Fourier domain as follows [19]:

$$H_1(u,v,w) = i\frac{u}{\sqrt{u^2+v^2+w^2}}$$

$$H_2(u,v,w) = i\frac{v}{\sqrt{u^2+v^2+w^2}} \quad (1)$$

$$H_3(u,v,w) = i\frac{w}{\sqrt{u^2+v^2+w^2}}$$

where the u, v and w are the Fourier domain coordinates and i represents the imaginary part of the signal. The monogenic signal f_M is then calculated as follow:

$$f_M(x,y,z) = [f(x,y,z) * g(x,y,z), f(x,y,z) * g(x,y,z) * h_1(x,y,z), f(x,y,z) * g(x,y,z) * h_2(x,y,z), f(x,y,z) * g(x,y,z) * h_3(x,y,z)] \quad (2)$$

where f is the original signal, h_1, h_2 and h_3 are the spatial domain representations of H_1, H_2 and H_3 respectively, g is a bandpass filter, and $*$ denotes convolution operation. Indeed, the 3D image is first filtered using a bandpass filter such as log-Gabor filter. An oriented 3D log-Gabor filter is defined by the following Eq:

$$G(w,\varphi,\theta) = \exp[(-\frac{\left(\log\left(\frac{w}{w_0}\right)\right)^2}{2\left(\log\left(\frac{k}{w_0}\right)\right)^2} + \frac{(\varphi-\varphi_0)^2}{2{\sigma_\varphi}^2} + \frac{(\theta-\theta_0)^2}{2{\sigma_\theta}^2}) \quad (3)$$

where w_0 is the filter's centre frequency, parameter *k* controls the bandwidth of the filter, φ_0 and θ_0 denote the filter angles, and σ_φ and σ_θ are the respective filter spread.

The monogenic signal defined by Eq. 2 consists of 4 components:

$$f_{M,1}(x,y,z) = f(x,y,z) * g(x,y,z),$$
$$f_{M,2}(x,y,z) = f(x,y,z) * g(x,y,z) * h_1(x,y,z), \quad (4)$$
$$f_{M,3}(x,y,z) = f(x,y,z) * g(x,y,z) * h_2(x,y,z),$$
$$f_{M,4}(x,y,z) = f(x,y,z) * g(x,y,z) * h_3(x,y,z)$$

The even and odd components of a monogenic signal are represented as follows:

$$even_{MG}(x,y,z) = f_{M,1}(x,y,z) \quad (5)$$

$$odd_{MG}(x,y,z) = \sqrt{f_{M,2}(x,y,z)^2 + f_{M,3}(x,y,z)^2 + f_{M,4}(x,y,z)^2} \quad (6)$$

and therefore, similar to 1D analytic signal, the new representation of the monogenic signal is the combination of even and odd terms:

$$f_{MG} = even_{MG}(x,y,z) + i \times odd_{MG}(x,y,z) \quad (7)$$

Now, the phase congruency for multiple orientations and scales of log-Gabor filter is defined as:

$$PC_o(x,y,z) = \frac{\lfloor E_o(x,y,z) - T_o \rfloor}{\sum_{sc} \sqrt{\left(even_o^{sc}(x,y,z)\right)^2 + \left(odd_o^{sc}(x,y,z)\right)^2} + \varepsilon} \quad (8)$$

where o and *sc* represent the orientation and scale variables respectively. PC_o is the phase congruency for a specific orientation and symbol $\lfloor \ \rfloor$ denotes that the enclosed quantity is not allowed to be negative. E_o and T_o are the orientation specific energy and noise threshold respectively which can be calculated by Eq.9 and Eq.10, and ε is a small offset to avoid division by zero.

$$T_o = \sum_{sc} exp^{mean[\log\left(\sqrt{\left(even_o^{sc}(x,y,z)\right)^2 + \left(odd_o^{sc}(x,y,z)\right)^2}\right)]} \quad (9)$$

$$E_o(x,y,z) = \sqrt{(\sum_{sc} even_o^{sc}(x,y,z))^2 + (\sum_{sc} odd_o^{sc}(x,y,z))^2} \quad (10)$$

To detect the image features at various orientations, a bank of oriented log-Gabor filters are designed. However, the multi-orientation representation for the PC calculation may cause high

dimensionality and expensive computational time. Therefore, we selected the dominant orientation where the PC value is the maximum and then designed a proper feature representation.

3.2 Proposed HDPC feature

This section describes the proposed local volumetric feature. The outline of the proposed descriptor is illustrated in Figure 2. The first step for HDPC feature extraction is the partitioning of volumetric data into local regions. The local regions are determined by dividing the data into predefined number of 3D blocks as shown in Figure 2a. To preserve the geometric information of the descriptors, each block is divided into predefined number of 3D grids named cells as illustrated in Figure 2b. The block-based approach is used to combine the extracted information from pixel-level, region-level, and volume-level. The video sequence can be partitioned using overlapping or non-overlapping blocks. Then, for each pixel of a cell, multiple 3D PCs at various orientations are calculated using Eq. 8. As Figure 2c shows, each pixel is characterized by several oriented PC components.

The next step is the selection of dominant orientation where the PC component is maximum. Therefore, each pixel of the 3D data is represented by an orientation and a maximum PC as shown in Figure 2d. Since we construct our descriptor based on the dominant PC for each cell, it will be less sensitive to noise.

Finally, the sub-histogram for predefined number of orientations (bin numbers) can be constructed. Each bin will be weighted by its dominant PC. On the other hand, for each orientation which is defined for log-Gabor filter, we accumulate the dominant PCs over each cell. The block feature consists of the cell's histograms.

The final HDPC descriptor is the concatenation of all the block's features. Therefore, our proposed spatiotemporal descriptor is able to handle different sequence length, allowing the use of variable length video segments which is common in real applications.

3.3 Unique properties of HDPC

- Since the PC values are computed for multiple scales of the bandpass filter, HDPC descriptor has the ability to capture the image information at various scales, and thus it is robust to image scale.
- Since the number of discretized directions defined for log-Gabor filter can be varied, HDPC is also able to detect the image features and motion information in multiple directions.
- Using PC values instead of gradient magnitude for each pixel to construct the local histogram makes HDPC robust to illumination variation. PC is a dimensionless quantity which has been shown to be insensitive to lighting [19].

4 Experimental Results

We used the extended Cohn-Kanade dataset (CK^+) [20] for facial expression recognition task in our experiments. It contains 593 video sequences recorded from 123 university students ranging from 18 to 30 years old. In this database, the subjects expressed a series of 23 facial displays including single or combined action units. Six of the displays are labeled as prototype basic emotions (joy, surprise, anger, fear, disgust, and sadness). In this work, we used all the 309 sequences from the dataset that have been labeled with at least one of the six basic emotions.

Due to limited dataset, we adopted theLeave-One subject-Out (LOO) cross validation approach in our experiment. For the database with *N* subjects, we performed *N* experiments. For each step, the video samples of *N-1* subjects are kept for training and the remaining samples for testing. Finally, the average classification accuracy on all the test samples is calculated as the true detection rate. This evaluation method makes the result obtained to be subject independent as there is no information of the same subject in both the training and test samples.

For classification task, a Support Vector Machine (SVM) with polynomial kernel function has been used in the experiments. SVM has been originally proposed for binary categorization, and then developed for multi-class problems [21]. For our first database, we used one-against-all technique that

constructs 6 binary SVM classifiers to categorize each emotion against all the others. Classification of a new instance is done by a winner-takes-all strategy, where the classifier with the highest output function assigns the class. Regarding the parameter selection of SVM, we carried out grid-search on the parameters as suggested in [5] for LOO cross-validation. The parameters producing the best result are chosen.

We applied a pre-processing stage before feature extraction. The images are aligned such that they have a constant distance between the two eyes. Since the facial landmarks' locations are given in the dataset, the alignment is done manually and no eye detection algorithm is used. The face images are then rotated to line up the eye coordinates. Finally, the faces are cropped using a rectangle of size 100×100.

We carried out the first experiment on different settings including various numbers of blocks and cells to evaluate its effect on feature dimension and classification detection rate. The results are tabulated in Table 1. Based on the results, partitioning the volume data into 75 blocks (5×5×3) and 9 cells (3×3×1) outperforms the other settings in term of classification detection rate.

For bandpass filtering with log-Gabor filter defined by Eq. 3, proper parameter setting is important to have acceptable results. A 16-oriented filter bank (4 values for θ ranged over 0^0-180^0, and 4 values for φ ranged over 0^0-180^0) are found to be suitable for our experiment. Table 2 gives the effect of parameter $\frac{k}{w_0}$ on the classification performance. Based on this table, the value of 0.85 produced the best result. We also did an experiment to evaluate the log-Gabor wavelength on classification detection rate. As Table 3 shows, result using 3 scales of the bandpass filter with wavelengths of {8, 12, 16} is superior to the other settings in term of detection rate.

In the CK^+ database, there are a few samples with illumination and skin colour variations. However, the variation is only minor and not sufficient to test the effect of illumination variation on our proposed descriptor. We have recorded some video sequences of surprise and happy expressions under different illumination conditions (8 video sequences for surprise and 7 sequences for happy). We used the recorded samples to evaluate our classifier trained using the CK^+ database (which has

only minor illumination variation). Our proposed method is able to detect the true label for all sequences (100% accuracy). Figure 3 shows the recorded signals for surprise, and a sample face of each sequence together with the recognition results.

We also tested the ability of the proposed descriptor for emotion detection from small scale and low resolution videos. We down sampled a recorded signal of happy expression with 5 sampling rate (1/2, 1/4, 1/6, 1/8, and 1/10). We again used the down sampled sequences to evaluate our classifier trained using the original CK^+ database. The recognition results again do not degrade compared to the original. The result shows the ability of the method to be applied even for low resolution video analysis.

Table 4 summarizes a comparison between the different approaches reported in the literatures and also our method applied to the same database. Brief information of the methods including number of subjects, number of video samples, static or dynamic process, evaluation measurements, and classification accuracy is given in this table. Note that direct comparison of the results is unfair since there are some differences in the experimental setup such as pre-processing approaches, number of samples, and evaluation methods among the reported results. Nevertheless, the table gives a qualitative performance difference among the various reported approaches on the same database and serves as a reference for the readers. The result shows that our person-independent result is comparable to the other approaches, and ranked just below the spatio-temporal LBP. However, our method is more robust to illumination variation as well as large change in scale.

5 Conclusion

In this paper, we proposed a novel descriptor for dynamic visual event analysis which has several desirable properties. Histogram of Dominant Phase Congruency (HDPC) is a spatio-temporal descriptor which is able to describe the motion features in addition to appearance features by extending the Phase Congruency to 3D and incorporating histogram binning. As such, it is also able to detect the features at different orientations and scales as well as robustness to illumination variation.

We have shown that it is an effective representation method for facial expression. Experimental results on facial expression Cohn-Kanade (CK^{+}) database achieved an accuracy of 95.44%. The robustness of the proposed descriptor to illumination and low resolution conditions were evaluated using our own collected facial expression data. The high performance of the method suggests that it is applicable for dynamic video events recognition in natural situations.

Acknowledgment

This research is supported by the Agency for Science, Technology and Research (A*STAR), Singapore.

References

[1] Guoying, Z., Pietikainen, M., 2007. Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, pp. 915-928.

[2] Ping, L., Jin, W., She, M., Honghai, L., 2011. Human action recognition based on 3D SIFT and LDA model. IEEE Workshop on Robotic Intelligence In Informationally Structured Space (RiiSS). pp. 12-17.

[3] Szilagyi, T., Brady, M., 2009. Feature extraction from cancer images using local phase congruency: A reliable source of image descriptors. IEEE International Symposium on Biomedical Imaging: From Nano to Macro. ISBI '09. pp. 1219-1222.

[4] Kovesi, P., 2000. Phase congruency: A low-level image invariant. Psychological Research, vol. 64, pp. 136-148.

[5] Shan, C., Gong, S., McOwan, P. W., 2009. Facial expression recognition based on Local Binary Patterns: A comprehensive study. Image and Vision Computing, vol. 27, pp. 803-816.

[6] Kim, D. H., Jung, S. U., Chung, M. J., 2008. Extension of cascaded simple feature based face detection to facial expression recognition. Pattern Recognition Letters. vol. 29, pp. 1621-1631.

[7] Bashyal, S., Venayagamoorthy, G. K., 2008. Recognition of facial expressions using Gabor wavelets and learning vector quantization. Engineering Applications of Artificial Intelligence. vol. 21, pp. 1056-1064.

[8] Gu, W., Xiang, C., Venkatesh, Y. V., Huang, D., Lin, H., 2012. Facial expression recognition using radial encoding of local Gabor features and classifier synthesis. Pattern Recognition. vol. 45, pp. 80-91.

[9] Xie, X., Lam, K.-M., 2009. Facial expression recognition based on shape and texture. Pattern Recognition, vol. 42, pp. 1003-1011.

[10] Wehrle, T., Kaiser, S., Schmidt, S., Scherer, K. R., 2000. Studying the Dynamics of Emotional Expression Using Synthesized Facial Muscle Movements. Journal o f Personality and Social Psychology. vol. 78, pp. 105-119.

[11] Cohen, I., Garg, A., Huang, T., 2000. Emotion recognition from facial expressions using multilevel HMM. Neural Inform. Process. Syst.

[12] Cohen, I., Sebe, N., Garg, A., Chen, L. S., Huang, T. S., 2003. Facial expression recognition from video sequences: temporal and static modeling. Computer Vision and Image Understanding - Special issue on Face recognition. vol. 91, pp. 160-187.

[13] Kwang-Eun, K., Kwee-Bo, S., 2010. Development of a Facial Emotion Recognition Method Based on Combining AAM with DBN. International Conference on Cyberworlds (CW). pp. 87-91.

[14] Kwang-eun, K., Kwee-Bo, S., 2010. Facial emotion recognition using a combining AAM with DBN. International Conference on Control Automation and Systems (ICCAS). pp. 1436-1439.

[15] Cohn, J. F., Zlochower, A. J., Lien, J. J., Kanade, T., 1998. Feature-point tracking by optical flow discriminates subtle differences in facial expression. Third IEEE International Conference on Automatic Face and Gesture Recognition. pp. 396-401.

[16] Camara-Chavez, G., De Albuquerque Araujo, A., 2009. Harris-SIFT Descriptor for Video Event Detection Based on a Machine Learning Approach. 11th IEEE International Symposium on Multimedia ISM '09. pp. 153-158.

[17] Dollar, P., Rabaud, V., Cottrell, G., Belongie, S., 2005. Behavior recognition via sparse spatio-temporal features. 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. pp. 65-72.

[18] Guha, T., Ward, R. K., 2012. Learning Sparse Representations for Human Action Recognition. IEEE Transactions On Pattern Analysis And Machine Intelligence. vol. 34.

[19] Kovesi, P., 1999. Image Features from Phase Congruency. Videre: Journal of Computer Vision Research, vol. 1, pp. 1-26.

[20] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I., 2010. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 94-101.

[21] Übeyli, E. D., 2008. Multiclass support vector machines for diagnosis of erythemato-squamous diseases. Expert Systems with Applications. vol. 35, pp. 1733-1740.

[22] Caifeng, S., Shaogang, G., McOwan, P. W., 2005. Robust facial expression recognition using local binary patterns. IEEE International Conference on Image Processing, ICIP 2005. pp. II-370-3.

[23] Bartlett, M. S., Littlewort, G., Fasel, I., Movellan, J. R., 2003. Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction. Computer Vision and Pattern Recognition Workshop. CVPRW '03. pp. 53-53.

[24] Littlewort, G., Bartlett, M. S., Fasel, I., Susskind, J., Movellan, J., 2004. Dynamics of Facial Expression Extracted Automatically from Video. Computer Vision and Pattern Recognition Workshop. CVPRW '04. pp. 80-80.

[25] Yeasin, M., Bullot, B. Sharma, R., 2004. From facial expression to level of interest: a spatio-temporal approach. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2004. pp. II-922-II-927 Vol.2.

[26] Asteriadis, S., Tzouveli, P., Karpouzis, K., Kollias, S., 2009. Estimation of behavioral user state based on eye gaze and head pose—application in an e-learning environment. Multimedia Tools and Applications.vol. 41, pp. 469-493.

[27] Ying-li, T., 2004. Evaluation of Face Resolution for Expression Analysis. Computer Vision and Pattern Recognition Workshop. CVPRW '04. pp. 82-82.

Table 1: Effect of number of blocks and number of cells on detection performance of HDPC feature for CK^+ database.

No. Blocks	No. cells	No. features	Detection rate (%)
2×2×2	2×2×2	1024	84.73
2×2×2	3×3×1	1152	91.84
2×2×2	4×4×2	4096	91.92
4×4×2	2×2×2	4096	91.64
4×4×2	3×3×1	4608	92.25
4×4×2	4×4×2	16384	93.55
5×5×2	3×3×1	7200	93.99
5×5×3	3×3×1	10800	95.44
5×5×3	4×4×2	38400	95.13

Table 2: Effect of log-Gabor bandwidth on classification accuracy for CK^+ database. The results are based on 75 blocks (5×5×3), and 9 cells (3×3×1).

$\frac{k}{w_0}$	Detection Rate (%)
0.55	93.00
0.65	93.99
0.75	94.10
0.85	95.44

Table 3: Effect of log-Gabor scales on classification accuracy for CK$^+$ database. The results are based on 75 blocks (5×5×3), and 9 cells (3×3×1).

# Scales	Wavelength	Detection Rate(%)
2	{4, 8}	95.20
3	{8, 12, 16}	95.44
4	{2, 8, 12, 16}	95.20
5	{2, 8, 12, 16, 32}	91.43

Table 4: Comparison of the reported results for the CK$^+$database. The number of sequences, statice or dynamic process, evaluation measurement, and classification detection rate.

Method	#Subject	#Sequence	Dynamic	Evaluation	Recognition Rate (%)
LBP+SVM [22]	96	320	N	10-fold	88.4
Gabor+AdaSVM [23]	90	313	N	10-fold	86.9
Gabor+adaboost+SVM[24]	90	313	N	LOO	93.8
Optical flow+HMM[25]	97	-	Y	5-fold	90.9
Multistream NN [26]	90	284	Y	-	93.66
Geometricalfeatures+NN [27]	97	375	N	-	93.8
Spatio-temporal LBP [1]	97	375	Y	10-fold	96.26
Proposed method (HDPC+SVM)	118	309	Y	LOO	95.44

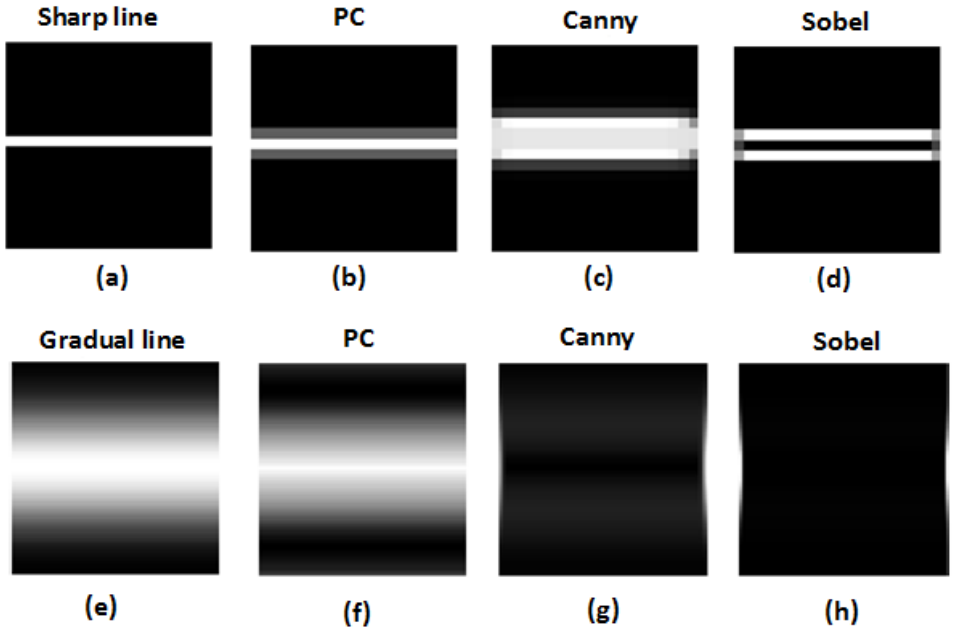

Figure 1: Comparison of methods for line detection; (a) sharp line; (b) line detection based on Phase Congruency; (c) line detection based on Canny; (d) line detection based on Sobel; (e) Gradual line with intensity range of [0 3]; (f) Line detection based on Phase Congruency; (g) Line detection based on Canny; (h) Line detection based on Sobel.

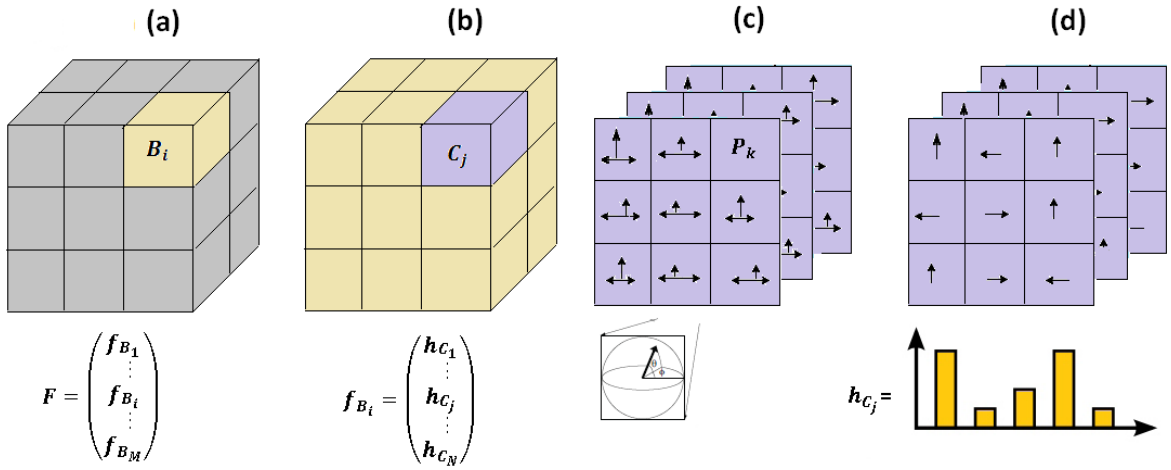

Figure 2: Descriptor computation; (a) The volume data is divided into a number of 3D grids. Each grid is denoted by a block (B_i). The final descriptor (F) consists of the block's feature. (b) Each block is divided into number of 3D cells (C_j). The block feature consists of the cell's histograms. (C) Each cell includes a number of points (P_k) which are characterized by several oriented PC components. (D) A PC component with Dominant orientation is selected for each pixel and then used to compute the histogram of a cell.

Figure 3: Variation of illumination; (a) Recorded samples of surprise expression under different illumination conditions; (b) Sample face of each sequence; (c) Classification results