Neurocomputing 175 (2016) 564–578

Contents lists available at ScienceDirect

Neurocomputing

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

Spontaneous facial micro-expression analysis using Spatiotemporal Completed Local Quantized Patterns

Xiaohua Huang [a,b], Guoying Zhao [a,*], Xiaopeng Hong [a], Wenming Zheng [b,c], Matti Pietikäinen [a]

[a] Center for Machine Vision and Signal Analysis, Department of Computer Science and Engineering, P.O. Box 4500, FI-90014, University of Oulu, Finland
[b] Research Center for Learning Science, Southeast University, Nanjing, JiangSu 210096, China
[c] Key Laboratory of Child Development and Learning Science (Ministry of Education), Southeast University, Nanjing, JiangSu 210096, China

ARTICLE INFO

Article history:
Received 24 April 2015
Received in revised form
26 September 2015
Accepted 27 October 2015
Communication by Su-Jing Wang
Available online 10 November 2015

Keywords:
Micro-expression
LBP-TOP
Vector quantization
Discriminative

ABSTRACT

Spontaneous facial micro-expression analysis has become an active task for recognizing suppressed and involuntary facial expressions shown on the face of humans. Recently, Local Binary Pattern from Three Orthogonal Planes (LBP-TOP) has been employed for micro-expression analysis. However, LBP-TOP suffers from two critical problems, causing a decrease in the performance of micro-expression analysis. It generally extracts appearance and motion features from the sign-based difference between two pixels but not yet considers other useful information. As well, LBP-TOP commonly uses classical pattern types which may be not optimal for local structure in some applications. This paper proposes SpatioTemporal Completed Local Quantization Patterns (STCLQP) for facial micro-expression analysis. Firstly, STCLQP extracts three interesting information containing sign, magnitude and orientation components. Secondly, an efficient vector quantization and codebook selection are developed for each component in appearance and temporal domains to learn compact and discriminative codebooks for generalizing classical pattern types. Finally, based on discriminative codebooks, spatiotemporal features of sign, magnitude and orientation components are extracted and fused. Experiments are conducted on three publicly available facial micro-expression databases. Some interesting findings about the neighboring patterns and the component analysis are concluded. Comparing with the state of the art, experimental results demonstrate that STCLQP achieves a substantial improvement for analyzing facial micro-expressions.

1. Introduction

Micro-expression is a subtle and involuntary facial expression. It usually occurs when a person is consciously trying to conceal all signs of how he is feeling. Unlike regular facial expressions, it is difficult to hide micro-expression reaction. As a result, the importance of micro-expression study is apparent in many potential applications for the security field. Psychological research shows that facial micro-expressions generally remain less than 0.2 s, as well they are very subtle [8]. Short duration and subtle change causes human difficulties in recognizing facial micro-expressions. In order to improve this ability, the micro-expression training tool developed by Ekman and his team was used to train people to better recognize micro-expressions. Even so, human can achieve just around 40% recognition accuracy [10]. Therefore, there is a great need for a high-quality automatic system to recognize facial micro-expressions.

Some earlier studies on automatic facial micro-expression analysis have primarily focused on posed or synthetic facial micro-expressions [29,30]. Recently, researchers have conducted spontaneous facial micro-expression analysis [22,27,41,42,43]. In contrast to posed facial micro-expressions, spontaneous facial micro-expressions can reveal genuine emotions that people try to conceal. It is very challenging to extract useful information from subtle changes of micro-expressions. It is noted that geometry-based or appearance-based feature extraction method has been commonly employed to analyze facial expressions. Specifically, geometric-based features represent the face geometry, such as the shapes and locations of facial landmarks, but they are sensitive to global changes, such as pose change and illumination variation. Instead, appearance-based features describe the skin texture of faces. Among these methods, Local Binary Pattern from Three Orthogonal Planes (LBP-TOP) has demonstrated its simplicity and efficiency for facial expression recognition [16,50]. As a result, LBP-TOP has been widely used in micro-expression analysis [27,42,6].

* Corresponding author.
E-mail addresses: xiaohua.huang@ee.oulu.fi (X. Huang), gyzhao@ee.oulu.fi (G. Zhao), xhong@ee.oulu.fi (X. Hong), wenming_zheng@seu.edu.cn (W. Zheng), mkp@ee.oulu.fi (M. Pietikäinen).

http://dx.doi.org/10.1016/j.neucom.2015.10.096
0925-2312/

Pfister et al. [27] proposed a spontaneous micro-expression recognition method using LBP-TOP. Yan et al. [42] applied LBP-TOP on their CASME 2 database achieving micro-expression recognition rate of 63.41%. As well, other works used LBP-TOP to investigate whether micro-facial movement sequences can be different from neutral face sequences [6].

However, it should be noted that there is still a gap to achieve a high-quality micro-expression analysis. Consequently, several works have attempted to improve the LBP-TOP. Ruiz-Hernandez and Pietikäinen [32] used the re-parameterization of second order Gaussian jet on the LBP-TOP achieving promising micro-expression recognition result on the SMIC database [27]. Wang et al. [37] extracted Tensor features from Tensor Independent Color Space (TICS) for micro-expression recognition, but their results on the CASME 2 database showed no improvement if we compare their highest achievable accuracies with the previous results. Furthermore, Wang et al. [38] used Local Spatiotemporal Directional Features (LSDF) with robust principal component analysis for micro-expressions, but yet did not obtain improvement for micro-expression analysis. In addition, recent work [39] reduced redundant information of LBP-TOP by using Six Intersection Points (LBP-SIP) to obtain better performance than the LBP-TOP. Even so, there is still much room for improvement in the recognition performance.

In our preliminary work [17], Completed Local Quantized Pattern (CLQP) was proposed by using the completed information and vector quantization to improve the performance of the original LQP proposed by Hussain and Triggs [18,19]. It achieved considerable results on texture classification and neonatal facial expression classification tasks. This paper proposes Spatio-Temporal Completed Local Quantized Pattern (STCLQP) by extending our spatial domain approach [17] for micro-expression analysis. In this work, STCLQP exploits three useful information, including sign-based, magnitude-based and orientation-based difference of pixels. Furthermore, STCLQP designs compact and discriminative codebooks for spatiotemporal domain. Different from our preliminary work, this work considers a more discriminative codebook and an application in spatiotemporal domain.

To explain the concepts of our approach, the paper is organized as follows. Section 2 discusses recent related work. Section 3 describes the completed local quantized patterns (CLQP) in spatial domain. Section 4 provides the extension of CLQP to spatio-temporal domain and explains its implementation to micro-expression analysis. Section 5 discusses parameter settings, datasets and provides experimental results with relevant discussion. Finally, Section 6 concludes the paper.

2. Related work

The video feature extraction problem has been addressed from different perspectives. Some works describe a video clip from the shape features. In [24], Liu et al. proposed a sketch-based method to organize video clip, where they used sketch annotations to enhance the narrations. The sketch-based approach achieves context-aware sketch recommendation for video shape extraction. In [1], Belongie et al. proposed the shape context using a circular local pattern and histogram to measure similarity between shapes. In [34], Shin and Chun used eighteen major feature points defined in MPEG 4, and then applied the dense optical flow method to track the feature points for sequential frames. In [20], Jain et al. used the shape feature around the eyebrows, eyes, nose, chin, inner lips and outer lips to describe the feature of each frame in video clip.

On the other hand, there are a few works to use a texture descriptor to describe the appearance and motion features of the video clip. As we know, Gabor and Local Binary Pattern (LBP) are two most representative ones for facial expression recognition. Since Gabor feature is related to the perception in human visual system, some facial expression recognition systems to date have utilized the Gabor energy filters [23,47]. Instead, LBP is simple to implement, fast to compute and has led to high accuracy in texture-based recognition tasks [26]. In recent years significant progress has been made in using LBP for facial expression recognition [9,33]. Due to its simplicity, the LBP operator was further extended to video sequences [50], named the LBP from Three Orthogonal Planes (LBP-TOP) of a space time volume. Basically, the LBP-TOP description is formed by calculating the LBP features from the planes and concatenating the histograms. In addition, LBP-TOP has efficient computation. Thus, it has become attractive for researchers in many fields.

In recent years, many extensions of LBP-TOP are developed in an application such as human action recognition [25] and lip reading [2]. Local Ternary Pattern from Three Orthogonal Planes (LTP-TOP) proposed by Nanni et al. [25] quantized intensity differences of neighboring pixels and center pixel into three levels to increase the robustness against noise. However, the LTP-TOP is sensitive to the quantization levels. In [2], for increasing the robustness against intensity noise, Local Ordinal Contrast Pattern (LOCP) used a pairwise ordinal contrast measurement of pixels from a circular neighborhood starting at the center pixel.

Recently, Completed Local Binary Pattern from Three Orthogonal Planes (CLBP-TOP) was proposed to exploit the useful information from intensity level [28]. In their work, completed local binary pattern [13] was extended into temporal domain. Magnitude and center pixel are severed as the complementary to the LBP-TOP, increasing the robustness of the LBP-TOP against noise. However, central pixel intensity information is very sensitive to noises caused by such things as illumination changes [51]. Alternatively, several researches [48,40,36,35,16] have demonstrated that orientation is useful because of its robustness to illumination changes. In [48], Zhang et al. proposed the Histogram of Gabor Phase Pattern, in which they combine the spatial histogram and the Gabor phase information coding schemes. In [40], Xie et al. proposed local Gabor XOR patterns, which encodes the Gabor phase by using local XOR pattern. In [36], Vu and Caplier presented multiple features combining patterns of oriented edge magnitude and patterns of dominant orientations. Therefore, it begs a question if orientation would be more effective for LBP-TOP.

Furthermore, LBP-TOP and CLBP-TOP inherit from the sparse sampling problem, yielding inadequate spatiotemporal descriptors. In fact, LBP histograms with small a number of bins tend to fail to provide enough discriminative information about the image appearance [31,46]. Instead, as the number of increases, more discriminative information will be provided, although this will cause the number of local patterns to increase exponentially. For example, with 8 sampling points around each pixel, there are 256 possible local patterns, while with 16 sampling points the dimensionality of histogram is 65,536. More importantly, the histogram becomes extremely sparse given a limited number of pixels. For example, it is observed that only half of the 65,536 local patterns in the case of 16-point sampling occur in spontaneous micro-expression database [22].

To address the sparseness problem, some specific codebooks were designed to reduce the number of possible codes and make the resulting histogram compact and evenly distributed. For example, in [26], Ojala et al. presented a type of codebook, namely ‘uniform pattern’, which consists of several binary patterns containing at most two bitwise transitions from 0 to 1 or vice versa when the bit pattern is traversed circularly. Although similar