Pattern Recognition Letters 000 (2017) 1–9

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Learning effective binary descriptors for micro-expression recognition transferred by macro-information

Xianye Ben [a,*], Xitong Jia [a], Rui Yan [b], Xin Zhang [a], Weixiao Meng [c]

[a] School of Information Science and Engineering, Shandong University, Jinan 250100, PR China
[b] Computer Science Department, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
[c] School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150080, PR China

ARTICLE INFO

Article history:
Available online xxx

MSC:
41A05
41A10
65D05
65D17

Keywords:
Micro-expression recognition
Binary feature descriptor
Coupled metric learning
Smooth support vector machine

ABSTRACT

In this paper, we propose three effective binary face descriptor learning methods, namely dual-cross patterns from three orthogonal planes (DCP-TOP), hot wheel patterns (HWP) and HWP-TOP for macro/micro-expression representation. We use feature selection to make the binary descriptors compact. Because of the limited labeled micro-expression samples, we leverage abundant labeled macro-expression and speech samples to train a more accurate classifier. Coupled metric learning algorithm is employed to model the shared features between micro-expression samples and macro-information. Smooth SVM (SSVM) is selected as a classifier to evaluate the performance of micro-expression recognition. Extensive experimental results show that our proposed methods yield the state-of-the-art classification accuracies on the CASMEII database.

1. Introduction

Micro-expression, which distinguishes from macro-expression because of the brief duration, subtle motion range and fewer action areas, is a facial expression controlled by human psychological activities. Currently, the tremendous potential of micro-expressions in national security, judicial trail and lie test has been eye-catching, thus attracted great attentions [8]. Micro-expression, a quick expression with only 1/25–1/5 s duration, reveals the emotions that are intentionally concealed. Discovered by Haggard and Isaacs [18], the micro-expression is thought to be related to the self-defense system, expressing the suppressed emotions. As the face recognition technologies [6–10,26,27,36] evolves, the recognition of spontaneous micro-expression has drawn more and more attention and a lot of progress has been made. Like a conventional face recognition system, the micro-expression system contains two very important stages, namely feature extraction and classification [42]. For the feature extraction, the objective is to explore the discriminant features so as to distinguish the facial motions. For the classification, the goal is to design a functionally efficient classifier in order to recognize various micro-expressions.

For micro-expression recognition, effective features are usually sought by descriptors and machine learning algorithms. At present, the main popular features in micro-expression recognition includes local binary pattern from three orthogonal planes (LBP-TOP) features [43], and optical flow features [2], etc. LBP-TOP features, as the extension of LBP features calculated in three planes (XY, YT, XT), are mainly used to extract the texture features of the micro-expression, and the consideration of time factors makes it more representative. Li et al. [21] used the LBP features to describe the temporal and spatial local texture features of the cropped face sequences and the support vector machine (SVM) was used to perform leave-one-out cross-validation test on the SMIC database, which reached a recognition rate of 48.78% for three types of emotions (Positive, negative, surprise). LBP-TOP is an effective feature extraction operator, but the redundant information in the orthogonal overlapping plane increases the computational complexity, which leads to poor discriminant performance. To address this problem, Wang et al. [34] proposed two methods of feature extraction: LBP-Six Intersection Points (SIP) and LBP-Three Mean Orthogonal Planes (MOP). Moreover, by taking the color information into account, Wang et al. [31] proposed a new color space model, tensor independent color space (TICS). The micro-expression recognition performed in TICS using LBP-TOP features achieved a higher recognition rate than RGB and grayscale color space, because the redundant information in LBP-TOP was removed. Guo et al. [16] combined the LBP-TOP feature with a near-

* Corresponding author.
E-mail address: benxianye@gmail.com (X. Ben).

http://dx.doi.org/10.1016/j.patrec.2017.07.010

est neighbor classifier, but it was highly influenced by the sample dimension, which greatly increased the computational complexity.

The optical flow features, mainly used to study the pixels and expression changes in a short period of time, are widely applied in micro-expression recognition. Liong et al. [22] proposed a new optical strain weighted feature extraction scheme with the optical flow information as the weight function of the LBP-TOP feature, which to some extent improved the recognition accuracy. However, this method had a feeble robustness as it was easily affected by noise. Liu et al. [23] came up with the main directional mean optical-flow (MDMO) feature, in which both local statistical motion information and the spatial position were included, for micro-expression recognition, and the computational efficiency was ensured by its fewer dimensions. Shreve et al. [29,30] obtained a 100% recognition rate of seven expressions on the UFS database with optical flow features, but the non-spontaneous samples made this work unpractical.

Besides LBP-TOP and optical flow features, other micro-expression features are extracted for identification. Polikovsky et al. [25] divided the human face into specific areas, used 3D-Gradient orientation histogram descriptor to extract facial motion information. Wu et al. [37] proposed an automatic micro-expression recognition system, using the Gabor filter for feature extraction and Gentle SVM for classification. As for the short duration, low intensity and usually local movements of micro-expressions, Wang et al. [32] argued that micro-expression contained sparse features extracted by Robust PCA. Ben et al. [4] proposed the maximum margin projection with tensor representation (MMPTR), which projected the micro-expression samples into a new tensor space to ensure the largest inter-class Laplacian scatter and smallest intra-class Laplacian scatter. Ruiz-Hernandez et al. [28] suggested that with a re-parameterization of the second local order Gaussian jet on Local Binary Patterns (LBP), a more robust and reliable histogram could be generated for micro-expression recognition. In order to learn a compact and discriminative codebook and gather valid information for micro-expression analysis, Huang et al. [19] proposed a spatiotemporal completed local quantized patterns (STCLQP) for the recognition. Although these feature extraction methods for micro-expression recognition have achieved encouraging recognition performance on the micro-expression datasets including SMIC [21] and CASME I [39] and CASMEII [38], their performance can still be improved by more effective binary face descriptors. In addition, the lack of a micro-expression database with sufficient training samples makes the micro-expression recognition performance low. Therefore, how to extract effective binary face descriptors for micro-expression recognition and learn the shared information between micro-expression and macro-expression/speech to enlarge the inter-emotion margins and reduce the intra-emotion variations is a challenging issue in the micro-expression recognition.

Second-order binary statistics descriptors [6] contain more informative features within a facial image than LBP, and binary descriptor extension to three orthogonal planes (TOP) (XY, YT, XT) [43], which have achieved encouraging performance. In this paper, we propose three effective binary face descriptor learning methods, including dual-cross patterns from three orthogonal planes (DCP-TOP), hot wheel patterns (HWP) and HWP-TOP for macro/micro-expression representation. Then, we use feature selection to make the binary descriptors compact. In addition, the labeled micro-expression samples are limited, so that we can leverage abundant labeled macro-expression and speech samples to train a more accurate classifier. Coupled metric learning can simultaneously project cross-domain samples into the common subspace [3]. Therefore, we employed coupled metric learning algorithm to model the shared features between micro-expression samples and macro-information. In the end, Smooth SVM (SSVM) is selected as a classifier to evaluate the performance of micro-expression recognition. Extensive experiments show that our proposed methods yield the state-of-the-art classification accuracies on the CASMEII database. The contributions of this paper are summarized as follows:

(1) We propose a binary descriptor, namely hot wheel patterns (HWP) to encode the discriminative features of a macro-expression image. With this descriptor, the connection with micro-expression image sequences can be obtained.
(2) We propose two binary descriptors, namely DCP-TOP and HWP-TOP to encode the discriminative features of a micro-expression image sequence, of which the goal is to enlarge the inter-emotion differences and reduce the intra-emotion variations.
(3) We use the coupled metric learning algorithm to model the shared features of both micro-expression samples and macro-information.

The remainder of this paper is organized as follows. Section 2 presents effective binary descriptors and their dimensions. Section 3 provides the micro-expression recognition model transferred by macro-information. Section 4 introduces the experimental results and analysis. Finally, Section 5 concludes this paper with discussion and future developments.

2. Effective binary descriptors

As one of basic but important research problems at the area of pattern recognition, local feature descriptors play a vital role in search of corresponding points of an image and the description of object features. The core requirement of image feature descriptors should be robust and distinguished. The output of feature descriptors is statistical histograms extracted from those images. Generally, binary descriptors consist of three major processing procedures, local sampling, image filtering, and pattern encoding. This section analyzes the local sampling and pattern encoding for the Dual-cross patterns (DCP) and the proposed hot wheel patterns (HWP), DCP-TOP and HWP-TOP.

2.1. Dual-cross patterns (DCP)

DCP [6] is a kind of local binary descriptors committed to perform local sampling and pattern encoding in three directions with $\pi/4$ angle, as the most information gather in the axis of abscissas, the axis of ordinates, and the diagonal. For instance, after normalization of a face image, parts of the expression information like eyes, nose, mouth, and eyebrows extend horizontally or vertically, while they converge in diagonal directions. Thus these four directions constitute the main information of the facial expression.

Based on the above analysis, Fig. 1 shows its local sampling.

As shown in Fig. 1 (a), for each pixel in an image, there are 16 points in the local neighborhood, which are in 8 directions, such as 0, $\pi/4$, $\pi/2$, $3\pi/4$, π, $5\pi/4$, $3\pi/2$, and $7\pi/4$ based on the horizontal direction. In addition, there are two circles around the center pixel, A_0, ...,A_7 are spaced on an inner circle of radius R_1, while B_1, ...,B_7 are located on an outer circle of radius R_2. Points between the two circles form the resulting sampled point pairs as $\{A_0, B_0; \ldots; A_7, B_7\}$.To quantify the textural information, a unique number has been designed in each sampling direction as follows

$$DCP_i = S(I_{A_i} - I_O) \times 2 + S(I_{B_i} - I_{A_i}), 0 \le i \le 7 \quad (1)$$

where $S(t) = \begin{cases} 1, & t \ge 0 \\ 0, & t < 0 \end{cases}$, and I_O, I_{A_i}, I_{B_i} represent the gray value of points O, A_i, B_i respectively.

Then DCP_i is further divided into two cross encoders: $DCP-1$ and $DCP-2$, as shown in Fig. 1(b) and (c). And the codes at each