Neurocomputing 184 (2016) 99–106

Contents lists available at ScienceDirect

Neurocomputing

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

Capturing correlations of local features for image representation[☆]

Xiaopeng Hong [a,*], Guoying Zhao [a], Stefanos Zafeiriou [b], Maja Pantic [b], Matti Pietikäinen [a]

[a] Department of Computer Science and Engineering, University of Oulu, Finland
[b] Department of Computing, Imperial College London, UK

ARTICLE INFO

Article history:
Received 22 January 2015
Received in revised form
22 May 2015
Accepted 13 July 2015
Available online 29 December 2015

Keywords:
Local feature
Feature pooling
Covariance matrix

ABSTRACT

Local descriptors are popular ways to characterize the *local* properties of images in various computer vision based tasks. To form the *global* descriptors for image regions, the first-order feature pooling is widely used. However, as the first-order pooling technique treats each dimension of local features separately, the pairwise correlations of local features are usually ignored.

Encouraged by the success of recently developed second-order pooling techniques, in this paper we formulate a general second-order pooling framework and explore several analogues of the second-order average and max operations. We comprehensively investigate a variety of moments which are in the central positions to the second-order pooling technique. As a result, the superiority of the second-order standardized moment average pooling (*2Standmap*) is suggested. We successfully apply 2Standmap to four challenging tasks namely texture classification, medical image analysis, pain expression recognition, and micro-expression recognition. It illustrates the effectiveness of 2Standmap to capture multiple cues and the generalization to both static images and spatial-temporal sequences.

1. Introduction

How to extract informative cues from images plays a central role in various computer vision based tasks, such as texture classification, facial analysis, and medical image processing. Image representations are expected to have the discriminative information for distinguishing instances from different classes and also can handle the possibly large intra-class divergence in appearance, such as illumination, colour, rotation, and scale. As a result, local features, which characterize the properties of small image patches, are of great interest [1,2]. The local features are proven informative and robust against noise and background clutters [1,3–7]. Feature pooling then, summarizes the local features within an image area and form a global description [8].

Most approaches rely on the first-order pooling, typically implemented by an average or a max operation [8]. The former one usually works together with the popular bag-of-words framework [1,3–7] to approximate the distribution of visual words, while the latter is normally in conjunction with the sparse coding framework [9–11]. These approaches achieve promising results in a number of applications, such as texture classification, object classification, and scene classification. However, as the first-order pooling treats each dimension of local features separately, the correlations between local features, which are important for a non-periodic and higher-level appearance analysis, are inevitably ignored.

More recently, Carreira et al. propose the second-order pooling technique to take into account the feature correlations [12]. The investigation in the second-order average and max operations indicates that the second-order pooling technique with appropriate non linear mappings[1] substantially outperforms the commonly used first-order one.

Encouraged by the success of second-order pooling techniques, in this paper we formulate a more general second-order pooling framework which enables to explore several analogues of the second-order average and max operations. Through comprehensive investigation of a variety of moments which are in the central positions to the second-order pooling technique, the superiority of the second-order standardized moment average pooling (*2Standmap*) is suggested. We then successfully apply the 2Standmap to four challenging applications namely texture classification, medical image analysis, pain expression recognition, and micro-expression recognition. Promising results on these four applications illustrates the effectiveness of 2Standmap to capture multiple

☆ This work is supported by the Academy of Finland and Infotech Oulu

* Corresponding author.
E-mail addresses: xiaopeng.hong@ee.oulu.fi (X. Hong), gyzhao@ee.oulu.fi (G. Zhao), s.zafeiriou@imperial.ac.uk (S. Zafeiriou), m.pantic@imperial.ac.uk (M. Pantic), mkp@ee.oulu.fi (M. Pietikäinen).

[1] Usually, the second-order average pooling is incorporated with the Log-Euclidean mapping, and the second-order max pooling is followed by the power normalization.

http://dx.doi.org/10.1016/j.neucom.2015.07.134
0925-2312/

cues and the generalization to both static images and spatial-temporal sequences.

The remainder of this paper is organized as follows: Section 2 introduces the related work. The second order pooling framework is summarized and analysed in detail in Section 3. Section 4 provides experimental results of comprehensive studies and the applications in four tasks. Finally, the conclusion is given in Section 5.

2. Related work

The most relevant works to ours are the coder-free second-order pooling approaches. In [12], the second-order average and max operations (hereafter termed *2AveP* and *2MaxP* respectively) are investigated to characterize super pixels for semantic segmentation. The second-order feature pooling takes into account the pairwise correlation of local features. Considering its superiority in performance compared with the first-order pooling method, the second-order pooling is applied to a few hot topics, such as object recognition, human pose estimation [13], touch saliency prediction [14], and face representation [15]. Moreover, the covariance matrix (CovM) [16–20] also belongs to the same umbrella. Distinct from the 2*AveP* which focuses on the raw moment (moments about zero) in [12], the CovM uses the second order central-moment (namely the moments about the mean) to characterize the pairwise correlation of low level local features.

To the best knowledge of the authors, there is no explicit comparison at the variety of second order moments. In this paper, we aim to fill this gap by carrying out a comprehensive study on this issue in the context of texture classification and medical image analysis. For fair comparisons, besides those simple raw features, such as the intensity (or the intensity of the three colour channels), the first and the second-order partial derivatives with respect to x and y as recommended in [20], we also evaluate three groups of typical local features, which include the histogram based descriptors (as used in [12]), the orthogonal transform based descriptors, and the statistics based descriptors. According to the experimental results, the second-order standardized central moment average pooling (2Standmap) is suggested. We further apply the 2Standmap to four challenging applications, and show that this simple coder-free pooling technique performs well even with the commonly used classifiers, such as the nearest neighbour classier and the linear support vector machine.

3. The second order pooling framework

Fig. 1 illustrates the overview of the investigated framework combining low-level raw features and mid-level local descriptors. Given an image region $\mathcal{I}$ with a size of $W \times H$, for each pixel $\mathbf{x} = (x, y)$, a vector $\mathbf{f}(\mathbf{x})$ concatenating all local features is extracted. We then perform the second order pooling to obtain the global descriptor $\mathbf{g}(\mathcal{I})$.

3.1. The second order pooling framework

2nd order average and max operation: Given the d-dimensional $\mathbf{f}(\mathbf{x})$ at pixel $\mathbf{x}$, we apply some pre-defined mapping θ and obtain the pre-processed feature vectors $\mathbf{p}(\mathbf{x}) = \theta(\mathbf{f}(\mathbf{x}))$. We then employ the *second order average pooling* [12] to summarize the pair wise connection of features as a matrix:

$$\mathbf{G}_{2AvgP}(\mathcal{I}) = c\sum\nolimits_{\mathbf{x} \in \mathcal{I}} \mathbf{p}(\mathbf{x})\mathbf{p}(\mathbf{x})^T, \tag{1}$$

and *second order max pooling* [12], as the following matrix

$$\mathbf{G}_{2MaxP}(\mathcal{I}) = \max\nolimits_{\mathbf{x} \in \mathcal{I}} \mathbf{p}(\mathbf{x})\mathbf{p}(\mathbf{x})^T, \tag{2}$$

where c is a normalization factor and the max operation is performed to the outer products of $\mathbf{p}$ in a per-entry manner.

A variety of moments: According to different points the moments are about, there are two types of second moments [21] commonly used. The central moments are the moments about the mean. And the raw moments refer to the moments about zero. More specifically, for central moments, the pre-processing function θ is expressed as

$$\theta_{cen}(\mathbf{f}(\mathbf{x})) = \mathbf{f}(\mathbf{x}) - \mu, \tag{3}$$

where μ is the mean of $\{\mathbf{f}(\mathbf{x})\}_{\mathbf{x} \in \mathcal{I}}$. In the case of raw moments, it becomes:

$$\theta_{raw}(\mathbf{f}(\mathbf{x})) = \mathbf{f}(\mathbf{x}) - \mathbf{0} = \mathbf{f}(\mathbf{x}). \tag{4}$$

Moreover, according to whether the features are normalized, the moments can be categorized to two types: the normalized and the unnormalized moments [21]. The former ones are the moments divided by the corresponding variances, while the latter preserved the scale of the original feature vectors.

As a result, there are four combinations of moments, namely the *normalized central moment* (a.k.a. *standardized moment*), the *normalized raw moment*, the *unnormalized central moment*, and the *unnormalized raw moment*.

It is notable that the second order average pooling used in [12] is the *unnormalized raw moment* here and the correlation matrix descriptor [20] is the *standardized moment*. This paper thus

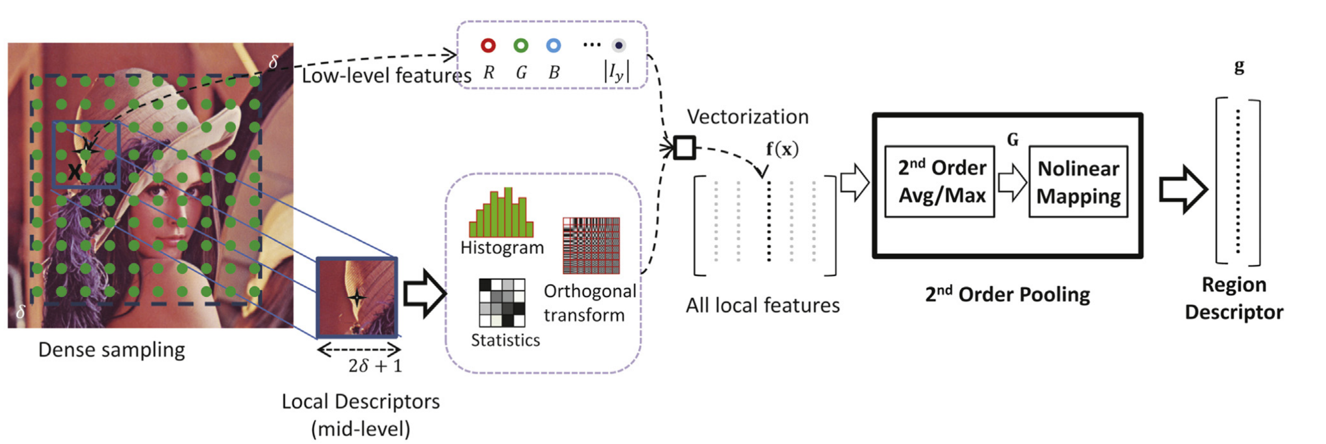

Fig. 1. Overview of the second pooling framework.