Contents lists available at ScienceDirect

Neurocomputing

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

Tensor decomposition and application in image classification with histogram of oriented gradients

Tan Vo, Dat Tran *, Wanli Ma

Faculty of Education, Science, Technology and Mathematics, University of Canberra, ACT 2601, Australia

ARTICLE INFO

Article history:
Received 28 February 2014
Received in revised form
23 May 2014
Accepted 8 June 2014

Keywords:
Image classification
HOG
Tensor
CP decomposition
ANN

ABSTRACT

In the field of visual data mining, Histogram of Oriented Gradients (HOG) and its variants have been widely used. The speed and ability to extract image features that are robust against many types of distortions such as scaling, orientation, affine and illumination that HOG offers have made it a popular choice for the task of detecting images in scenes for classification. However, the high dimensionality nature of HOG descriptors (features), usually in the order of multiple thousands of them per image, would require careful consideration in place to achieve accurate and timely categorization of objects within images. This work explores the possibility of processing HOG features as tensors, or multi-dimensional arrays. A direct result of that is tensor decomposition techniques such as canonical polyadic (CP) decomposition performed on the high-order HOG tensors as the mean for dimensionality reduction by filtering. This work focuses on the impact of this approach on both accuracy and efficiency, comparing it against the standard practice of processing HOG features. Validating with the Caltech-101 dataset, the results achieved with artificial neural network (ANN) classification indicate that the proposed method not only improves the overall system performance, it also achieves the edge in accuracy by a notable margin.

1. Introduction

In recent years, the role that image classification plays in visual data mining is getting more and more important. This is even more specific in the case of large scale visual recognition, where substantial amounts of images are required to be processed and categorized by machine in a both effective and accurate manner. The performance of image classification relies heavily on the process that extracting image features. Among the options in the category, Histogram of Oriented Gradients (HOG) [1,2] has been considered an effective algorithm for producing these features [3], which consists of local characteristic across images. Object recognition with this algorithm is considered more robust against affine distortions such as rotation, scale and position as well as lighting distortions of objects in images.

In Dalal and Triggs's original work [1], HOG descriptors are used as feature vectors for a linear Support Vector Machine (SVM), which performs like a sliding window human detector within scenes. Since the work of Felzenszwalb et al. [4], HOG has been known in providing robust and effective features to be used for object detection. In many cases, HOG descriptors can be processed and utilized in similar fashion to scale-invariant feature transform descriptors [2] (SIFT). In particular, this work considers using HOG as a feature extracting mechanism for the image classification task. Likewise to SIFT, HOG descriptors allow an image classifier to produce matches on a given image to the images that were used to train the image classifier by constructing a visual Bag-of-Words (BoW) model of the HOG descriptor and using it in tandem with a machine learning algorithm such as Artificial Neural Network (ANN).

Due to the nature they are created, HOG descriptors are generally represented in three-dimensional space and, depending on the resolution of the image, also tends to be high in volume. Fig. 1 visualizes the HOG descriptors created on the image of a butterfly by two variants of HOG: (i) UoC/TTI, Felzenszwalb et al. [4] (upper row) and (ii) Dalal–Triggs [1] (lower row). Fig. 1 shows that there exist subtle differences between the two variants. A value of O, or number of orientations, can be configured for HOG operation and its influence on the HOG features is quite significant and consistent for both variants. This flexibility allows HOG to be used for a variety of image detection problems.

Research activities have formed a new trend in recent years, using techniques such as principal component analysis (PCA) and singular value decomposition SVD on these types of descriptors. This trend aims for the two goals is either to lessen the complexity by dimensionality reduction of the original descriptors or to create improved versions of themselves. The HOG-PCA work in pedestrian detection [5–7] represents how the first goal is achieved while the PCA-SIFT work of Yan et al. [8] demonstrates the second goal. It effectively generates more robustness and performance in the image recognition task.

* Corresponding author.
E-mail addresses: Tan.Vo@canberra.edu.au (T. Vo), Dat.Tran@canberra.edu.au (D. Tran), Wanli.Ma@canberra.edu.au (W. Ma).

http://dx.doi.org/10.1016/j.neucom.2014.06.093
0925-2312/

The potential in the work of compacting and reducing these image descriptors of the trend has motivated us to propose an approach, in which a technique called canonical polyadic (CP) decomposition [9,10] is applied on HOG descriptors. This technique is amongst a range of tensor decomposition techniques, which has application in many fields, including signal processing, computer vision and data mining [11]. CP decomposition is essentially a generalization of matrix SVD to tensors, or multidimensional array. The biometric work by Liu et al. [12] has demonstrated the possibility of performing CP decomposition on a image descriptor type similar to HOG. That, followed by the image classification work from Vo et al. [13] recently, has showed that the outcome of the CP process on image descriptors helps identifying the key subset of descriptors representing a given image.

There are two strategies being proposed and investigated for usage scenario. The first strategy is (i) category free, where the process is performed regardless of the image classes while with the other strategy, (ii) category bounded, the descriptors are grouped and processed within their classes.

2. Proposed method

2.1. Histogram of oriented gradients

Fig. 2 summarizes the steps in extracting the HOG descriptors of an image. This process begins with dividing a given image into cells of equal size as in Fig. 2(a). Within each cell, a histogram of gradient directions, or edge orientations, is accumulated over the pixels as in Fig. 2(b). The orientation $\theta(x,y)$ and the magnitude $r(x,y)$ of a pixel (x,y) are calculated with a 1-D discrete derivations mask $[-1,0,1]$ and its transpose $[-1,0,1]^\top$. The magnitude $r(x,y)$ is calculated with the color channel with the largest gradient magnitude. Let $O=9$ be the number of orientations, there will be $2\times 9=18$ directed orientation bins allocated, or one bin for every 20° in the range 0–360°: 2 orientations ($\pm$) for each of the 9 undirected gradient directions [1].

The next step in HOG is block normalization, in which blocks are generated by grouping four adjacent cells together (sliding of each cell) as visualized in Fig. 2(c). Let vector v be the stacking of the positive direction histogram in a block, the norm (l^2-norm) of a block is defined as

$$v=\frac{v}{\sqrt{\|v\|_2^2+\epsilon^2}} \quad (1)$$

where $\|v\|_2$ represents the two-norm of v and ϵ is a very small number. It is worth notice that the value of ϵ is insignificant to the outcome.

The final step Fig. 2(d) produces the actual descriptors. For each cell, four normalization factors can be obtained as the inverse of the norm of the four blocks that contain it. Four copies of the cell's undirected 9-dimensioned histogram will then be normalized with each normalization factor, separately. The results are stacked and clipped at 0.2. The process will produce a vector of $4\times 9=36$ in length. This is used as the HOG descriptor representing the cell.

Another HOG variant considered in this work is UoC/TTI [4]. This type of descriptors is created with a slightly different process, in which (i) the normalization is performed over both directed (18 bins) and undirected histograms (9 bins), i.e., produce a vector of $4\times(2+1)\times 9$ in length; (ii) the dimensionality of the result is reduced with a PCA variation to the length of $(2+1)\times 9$; and (iii) the l^1norm of the four normalized undirected histograms is computed and stored as additional four dimensions. With that, the UoC/TTI process produces 31-dimensioned descriptors ($(2+1)\times 9+4$), rather than 36 in the case of Dalal–Triggs variant with the UoC/TTI variant.

In Fig. 2(d), as blocks are visited from left to right and top to bottom, they form the final descriptor of the image. The HOG descriptors are normally structured in the form of a 3-dimensional

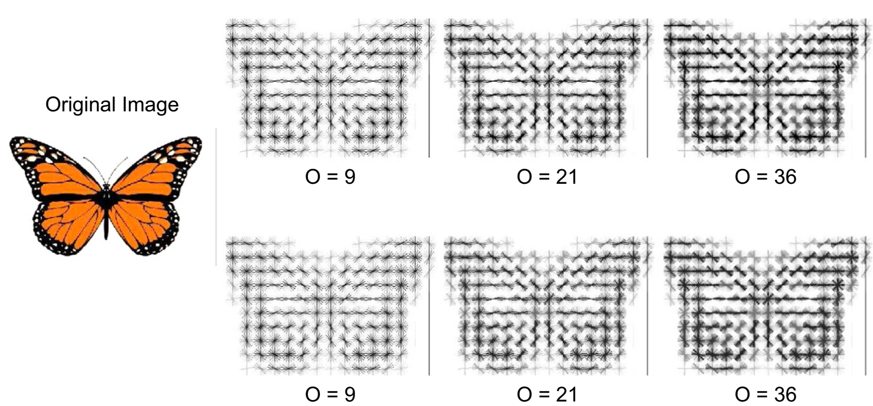

Fig. 1. HOG with different numbers of orientations using UoC/TTI (upper row) and Dalal–Triggs (lower row).

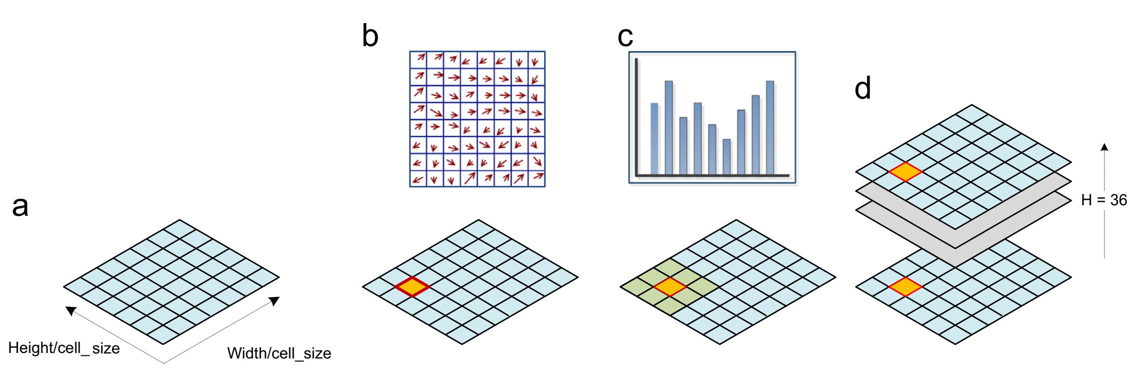

Fig. 2. Process of creating HOG descriptors.

Please cite this article as: T. Vo, et al., Tensor decomposition and application in image classification with histogram of oriented gradients, Neurocomputing (2015), http://dx.doi.org/10.1016/j.neucom.2014.06.093