Computers in Biology and Medicine ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Multilayer descriptors for medical image classification

Alessandra Lumini [a], Loris Nanni [b], Sheryl Brahnam [c,*]

[a] DISI, University of Bologna, Via Sacchi 3, 47521 Cesena, Italy
[b] Department of Information Engineering, Via Gradenigo 6/A, 35131 Padova, Italy
[c] Computer Information Systems, Missouri State University, 901 S. National, Springfield, MO 65804, USA

ARTICLE INFO

Article history:
Received 30 December 2014
Accepted 19 November 2015

Keywords:
Texture descriptors
Ensemble
Local binary patterns
Local phase quantization
Support vector machine
Multilayer descriptors

ABSTRACT

In this paper, we propose a new method for improving the performance of 2D descriptors by building an *n*-layer image using different preprocessing approaches from which multilayer descriptors are extracted and used as feature vectors for training a Support Vector Machine. The different preprocessing approaches are used to build different *n*-layer images ($n=3$, $n=5$, etc.). We test both color and gray-level images, two well-known texture descriptors (Local Phase Quantization and Local Binary Pattern), and three of their variants suited for *n-layer* images (Volume Local Phase Quantization, Local Phase Quantization Three-Orthogonal-Planes, and Volume Local Binary Patterns). Our results show that multilayers and texture descriptors can be combined to outperform the standard single-layer approaches. Experiments on 10 datasets demonstrate the generalizability of the proposed descriptors. Most of these datasets are medical, but in each case the images are very different. Two datasets are completely unrelated to medicine and are included to demonstrate the discriminative power of the proposed descriptors across very different image recognition tasks.

A MATLAB version of the complete system developed in this paper will be made available at https://www.dei.unipd.it/node/2357.

1. Introduction

The amount of multidimensional visual information (e.g., 2D images, videos, 3D surface models of objects, and 3D tomographic images) uploaded to the internet on a daily basis has increased enormously in the past few years. The video hosting website YouTube, for instance, received more than 100 h of new video every minute in 2014. In the field of medical imaging, individual radiology departments routinely produce an enormous amount of multidimensional information that is warehoused in private and public medical databases. Such large quantities of data are difficult to manually label for further access and reuse. There is, as a result, an urgent need to develop sophisticated, efficient, and accurate image classification algorithms that are able to provide best matches for specific classification tasks. Many applications require the ability of discriminating among classes of images, to distinguish, for instance, pedestrians from objects for advanced driver assistance, for disease recognition from medical images, for landmark recognition for touristic purposes, and many other applications. In these cases, computer vision plays a central role. Key to successful classification is the ability of representing images based on visual characteristics such as texture, color, and shape [1].

Automated visual tasks such as detection, localization, categorization, and recognition by visual features are important subjects of study in computer vision and image analysis. These tasks often prove difficult, however, due to high within-class variability that can be caused by a variety of factors: noise, distortions, illumination, scale changes, occlusion, and so on. Extracting features within visual databases is one of the most important steps in the classification process as success depends on the method adopted for extracting features from a given set of images. Particularly important are invariant image descriptors since they extract information from images which is invariant to noise, illumination, distortion, etc.

One class of invariant descriptors is color composition, which in a visual scene is robust to noise, image degradations, changes in size, resolution, and orientation. Most existing systems use various color descriptors [2,3] in order to retrieve relevant samples. Unfortunately, the classification performance of color descriptors is negatively affected by their lack of discriminative power.

Descriptors based on shape are considered more generic than color and are the most widely used descriptors in many application areas such as object detection and action recognition. There are many ways to represent shapes. Some examples include axial representation [4], primitive-based representation [5], histograms

* Corresponding author. Tel.: +1 417 8739979.
E-mail address: sbrahnam@missouristate.edu (S. Brahnam).

http://dx.doi.org/10.1016/j.compbiomed.2015.11.010
0010-4825/© 2015 Elsevier Ltd. All rights reserved.

of oriented gradients [6], contour-based representation [7], and probability density function [8]. Most shape descriptors are robust to rigid transformations, noise, occlusions, and missing data.

Another class of descriptors is texture. Texture has received considerable attention [9–11] and has proven valuable in medical imaging, image retrieval, object recognition, and industrial product identification (e.g., classifying types of marble, ceramic tiles, parquet slabs, etc.). Various methods have been introduced to analyse texture in digital images. The most well-known methods are based on statistical approaches, such as histograms of gray-level pixel values and the Gray-Level Co-occurrence (GLC) matrix [12], wavelet transforms and Gabor filters [13], and the Local Binary Pattern (LBP) operator [14,15]. The LBP operator, in particular, has proven an efficient method for describing texture in 2D. It has received much attention from the scientific community, and many powerful variants have recently been proposed [16].

Much work has also been done exploring the integration of more than one feature to improve classification accuracy [17]; however, the performance of these combinations are strictly application dependent. In some cases less efficient features can degrade more efficient features with a loss of overall system accuracy.

Most recently, there has been a move to explore color in three dimensions. In [18] the RGB space is considered as three planes in a 3D image, and a 3D version of the Local Ternary Pattern (LTP) is then proposed that extracts features directly from the 3D image, instead of from each color space (i.e., separately from R, G, and B). Moreover, they show that it is possible to generate a multilayer image starting from gray value images using a 2D circular symmetric Gaussian filter. Please note that in some works (e.g. [18]) the authors use the term “volumetric descriptors” to denote descriptors extracted from a RGB image or any kind of images represented by more than a layer. In this work we use the term multilayers to denote such a concept to avoid misunderstanding with the the notion of “volumetric descriptor” that in the context of medical image analysis refers to some measure of volume in anatomically meaningful regions on a medical image.

In this paper we start from the idea in [18] that considered the RGB image as a three-layer image and extend this representation to preprocessed images, showing that it is possible to improve the performance of single-layer descriptors by building an n-layer image ($n=3$, $n=5$, etc.) using different preprocessing approaches.

Formally speaking if a single layer image can be considered as mappings from a spatial coordinate system into a value space, i.e. a mapping from $[1,2,\ldots,m]\times[1,2,\ldots,l]\rightarrow\mathfrak{R}$ for an $m\times l$ image, a n-layer image is a mapping to a n dimensional space: $[1,2,\ldots,m]\times[1,2,\ldots,l]\rightarrow\mathfrak{R}^n$. An RGB color image is a particular case of a multilayer image where $n=3$. This definition is different from that used in the literature for existing multilayer descriptors (e.g. [18,22]) since they were proposed for dynamical textures.

The n-layer images are here described using multilayer descriptors, and these feature vectors are fed into a Support Vector Machine (SVM) [19]. We test both color and gray-level images, two well-known texture descriptors (LBP and Local Phase Quantization), and three of their variants suited for *multilayer* images (viz., Volume Local Phase Quantization, Local Phase Quantization Three-Orthogonal-Planes, Volume Local Binary Patterns). Our results show that multilayer and texture descriptors can be combined to outperform standard single-layer approaches. The proposed approach is applied to ten datasets to demonstrate the generalizability of this approach. In this work, we do not deal with the segmentation problem, since all the datasets considered in the experiments contain only segmented image. The segmentation of large images that can contain parts of different textures is out of the scope of this paper.

The remainder of this paper is organized as follows. In Section 2 we describe the base approaches used in our system, and in Section 3 we present our system. We conclude in Section 4 by summarizing the significance of our work and by highlighting some future directions of exploration.

2. Base methods

In this section we provide an overview of the methods used for building the ensemble of descriptors. We describe the multilayer descriptors in Section 2.1. A very brief description of the preprocessing methods employed to generate the *n-layer* images is available in Section 2.2.

2.1. Multilayer descriptors

In this subsection we briefly explain the methods used for building the ensemble of *multilayer* descriptors.

2.1.1. Volumetric local binary pattern (VLBP)

VLBP, introduced by [22], is an extension of LBP, with the notion of self-similarity, central to conventional image texture, extended to the spatiotemporal domain. VLBP deals with dynamic texture analysis on the 2D time series and not on the full 3D data; therefore, it only provides rotation invariance towards rotations around the z-axis.

VLBP extends standard LBP to handle dynamic texture analysis by considering the joint distribution v of the gray-levels of $3P+3$ pixels ($P>1$ is the number of local neighboring points around the center in one frame):

$$V=v(g_{t_c-L,c},g_{t_c-L,0},\ldots,g_{t_c-L,P-1},g_{t_c,c},g_{t_c,0},\ldots,g_{t_c,P-1},g_{t_c+L,c},g_{t_c}+L,0,\ldots,g_{t_c+L,P-1})$$

where $g_{t,p}$ $t\in[t_c-L,t_c,t_c+L],p\in[0,\ldots,P-1]$ corresponds to the gray value of P equally spaced pixels on a circularly symmetric neighbor of the center pixel c, in the center frame t_c and its previous/posterior frames with time interval L.

To obtain the gray-scale invariance as in LBP, the gray value of the center $g_{t_c,c}$ is subtracted from the gray values of the circularly symmetric neighborhood of each frame. To achieve invariance with respect to the scaling, the resulting code is binarized by considering only the signs of the differences. Finally, a unique $VLBP_{L,P,R}$ code is obtained by binary factorization (i.e., by weighting each pixel difference by a different binomial factor). The histogram obtained from the $VLBP_{L,P,R}$ code is normalized with respect to size variations by setting the sum of its bins to unity. Rotation invariance is obtained by rotating the neighbor set in three separate frames clockwise synchronously so that a minimal value is selected. Here we have used the following parameters $R=1$; $P=4$; $L=1$.

2.1.2. Volume local phase quantization (VLPQ)

VLPQ, introduced in [23], is based on the binary encoding of the phase information of the local Fourier transform at low frequency points and is an extension to the LPQ operator [21] used for spatial texture analysis. Since dynamic texture consists of a sequence of texture in the spatio-temporal domain, the Fourier transform estimation is performed locally using Short-Term Fourier Transform (STFT). Given a sequence f(x), STFT is computed over an M-by-M-by-N neighborhood $\mathcal{N}_\mathbf{x}$ of x (where M is the spatial size and N the size in the temporal domain):

$$F(\mathrm{u},\mathrm{x})=\mathrm{w}_\mathrm{u}^T\mathrm{f}_\mathrm{x},$$

where $\mathbf{w_u}$ is the basis of the 3D DFT at frequency $\mathbf{u}$, and $\mathbf{f_x}$ is a vector containing all pixels from the neighborhood $\mathcal{N}_\mathbf{x}$. The computation is efficient since STFT can be evaluated for each pixel using 1D convolutions for each dimension, due to the separability of the basis functions.

Please cite this article as: A. Lumini, et al., Multilayer descriptors for medical image classification, Comput. Biol. Med. (2015), http://dx.doi.org/10.1016/j.compbiomed.2015.11.010