Information Sciences 424 (2018) 235–249

Contents lists available at ScienceDirect

Information Sciences

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

Perceptual uniform descriptor and ranking on manifold for image retrieval

Shenglan Liu[a,b,c,*], Jun Wu[a], Lin Feng[a], Hong Qiao[d,e], Yang Liu[b], Wenbo Luo[f], Wei Wang[b]

[a] School of Innovation and Entrepreneurship, Dalian University of Technology, Dalian, Liaoning, 116024, China
[b] Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, Liaoning, 116024, China
[c] State Key Laboratory of Software Architecture(Neusoft Corporation), Shen Yang, Liaoning, 110179, China
[d] the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation,Chinese Academy of Sciences, Beijing 100190 China
[e] CAS Centre for Excellence in Brain Science and Intelligence Technology (CEBSIT), Shanghai 200031, China
[f] School of Computer Science, Liaoning Normal University, Dalian, Liaoning, 116081, China

ARTICLE INFO

Article history:
Received 23 September 2016
Revised 6 August 2017
Accepted 4 October 2017
Available online 5 October 2017

Keywords:
Manifold
Gestalt psychology
Perceptual uniform descriptor
Ranking
Image retrieval

ABSTRACT

Incompatibility of image descriptor and ranking has been often neglected in image retrieval. In this paper, Manifold Learning and Gestalt Psychology Theory are involved to solve the problem of incompatibility. A new holistic descriptor called Perceptual Uniform Descriptor (PUD) based on Gestalt psychology is proposed, which combines color and gradient direction to imitate human visual uniformity. PUD features in the same class images distributes on one manifold in most cases, as PUD improves the visual uniformity of the traditional descriptors. Thus, we use manifold ranking and PUD to realize image retrieval. Experiments were carried out on four benchmark data sets, and the proposed method is shown to greatly improve the accuracy of image retrieval. Our experimental results in Ukbench and Corel-1K datasets demonstrate that N-S score reached 3.58 (HSV 3.4) and mAP at 81.77% (ODBTC 77.9%) respectively by utilizing PUD which has only 280 dimensions. The results are higher than other holistic image descriptors including local ones as well as state-of-the-arts retrieval methods.

1. Introduction

Feature extraction and ranking are two important topics in Content Based Image Retrieval (CBIR).It is well known that image representation plays an important part in CBIR systems [5,35], and thus the performance of these systems depends mainly on the discrimination and effectiveness of features. The process of feature extracted can be divided into three steps: 1) image preprocessing; 2) detection of discriminative image regions; 3) feature statistical strategy in these regions. Most of the works concentrated on one or more steps to improve their descriptors.

First, in order to describe certain properties of natural images which may contain various types of image noise, image preprocessing is an indispensable step. Many image denoising and image sharpening algorithms have been presented to reduce the effect of noise on image content and strengthen discriminative information in some regions.

* Corresponding author at: School of Innovation and Entrepreneurship, Dalian University of Technology, Dalian, Liaoning, 116024, China.
E-mail addresses: liusl@mail.dlut.edu.cn (S. Liu), fenglin@dlut.edu.cn (L. Feng).

https://doi.org/10.1016/j.ins.2017.10.010
0020-0255/

Second, discriminative image regions are detected. Based on that, the descriptors can be classified into global-based and local-based. Color Histogram (CH) [23], Local Binary Patterns (LBP) [25,26] and Histogram of Gradient (HOG) [4], which describe the color, texture and edge features respectively, are provided based on the global image regions. Motivated by the visual perception mechanisms for image retrieval, Liu et al. provided micro-structure descriptor (MSD) [21] which defined the micro-structures through the similarity of edge orientation and the underlying colors, and introduced structure element correlation statistics to characterize the spatial correlation among them (Color difference histogram (CDH) [22] is another version.). On the contrary, local-based descriptors focus on describing local regions which contain certain information. Lowe et al. [15] introduced a local descriptor called scale-invariant feature transform (SIFT), which aimed at detecting and describing some local neighborhoods around key points in scale space. HMAX model [32] based on the hierarchical visual processing in the primary visual cortex (V1) utilized Gabor filters in different scales and orientations in S1 unit [29]. More details about performance comparisons among other local descriptors are presented in [24].

Finally, corresponding feature statistics methods in these regions are provided. As one of the most common methods, the histogram-based strategy has been applied in many descriptors, such as CH, LBP and HOG. Moreover, color moment [36], color correlogram [13] and color coherence vector [28] were proposed to emphasize the spatial relationship of feature elements.

Besides image feature extraction methods, image Ranking methods have also been rapidly developed in CBIR. Lots of researches have been devoted to improving the ranking results, such as $L1$-norm [1], Euclidean distance [18], Hamming distance [7] etc. Previous research has showed that ranking by $L1$-norm is simple and can obtain a better result than that by Euclidean distance [27,37]. In addition, the graph based ranking methods, such as PageRank [31] and manifold ranking [10], are also widely used for image retrieval. The manifold ranking is proposed based on manifold learning and relates to perception.

In most image retrieval schemes, image feature extraction and ranking are two independent processes. This likely accounts for the incompatibility between descriptor and ranking method (for example, an image representation which is compatible with $L1$-norm ranking, may not obtain expectant results while using manifold ranking methods, see Section 7).

In computer vision, we hope the computer to imitate human's perception for learning image and other visual data [16]. In the process of human cognition, visual uniformity is beneficial to learn image, and has been used for the extraction of the image features [14]. Visual uniformity is consistent with human perception of the image. Thus, we point out that the image features extraction by visual uniformity is more likely to distribute on the manifold. In 2000, three types of research related to manifold learning were published in "Science" [11,33,38], in which Lee [33] points out that "human perception is in the way of manifold" (This phenomenon is illustrated in Section 2). In this paper, we construct the image feature and ranking model based on the manifold, which aims to realize the uniformity in CBIR.

In this paper, according to visual organization principle and the theory in "The manifold way of perception", we use human's visual perception to construct the image visual feature, and retrieve images via manifold ranking. The main contributions of this paper are stated as follows:

(1) Perceptual Uniform Descriptor (PUD) is proposed by using the visual principle of Gestalt psychology, so that it can better distribute on a manifold.

(2) The incompatible problem between image descriptors and ranking methods is analyzed. The concept of a manifold is involved as a bridge for descriptors and ranking methods in CBIR.

The rest of the paper is organized as follows: Section 2 states the motivation of our proposed image retrieval scheme. Principles of Gestalt psychology are introduced in Section 3. Sections 4 and 5 present our image descriptor. Sections 6 and 7 refer to manifold ranking for image retrieval. In Section 8, experimental results and analysis are reported. Section 9 concludes the paper.

2. Motivation

The human visual system can pinpoint and analyze objects in complex images in a very short time. The main aims of many studies related to human brain visual mechanism and cognitive psychology are to simulate vision systems that have the equal performance to humans in object recognition. According to the analysis that the image variability can actually be considered as a manifold embedded in the image space, Seung and Lee [34] introduced the idea that human visual perception can be expressed by manifolds. The brain must encode the visual information by some ways. For image analysis, the descriptors that are in accordance with the distribution of manifolds have more discriminative information.

Due to the connection with low-level visual features, human visual attention system related to the perception and understanding of visual images facilitates the construction of image feature representation. Some psychophysical and neurobiological studies demonstrate that human visual system is sensitive to the low-level visual features, such as color and edge information. However, the holistic images usually contain some redundant regions where less discriminative information is useful for image analysis. And the image representation in these regions may not only impair the performance of descriptor but also consume too much time. To detect the special regions that human eyes perceive predominantly, the studies in the cognitive psychology give some inspiration about perceiving the objects. The Gestalt Laws of perception introduces some principles that help people group similar pixels or patches in the image. Among these principles, proximity, similarity and good continuation are fundamental to define the perceptually uniform regions. The closer the image pixels are or the