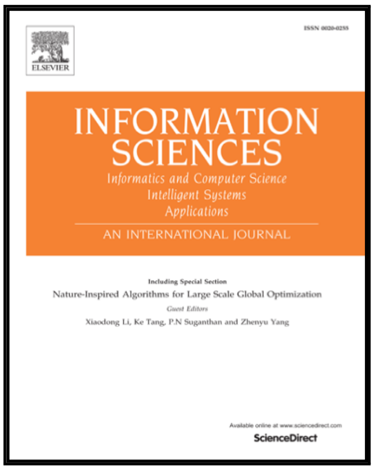
Accepted Manuscript



Rotative Maximal Pattern: A Local Coloring Descriptor for Object Classification and Recognition

Junbiao Pang, Jing Huang, Lei Qin, Weigang Zhang, Laiyun Qing, Qingming Huang, Baocai Yin

PII: S0020-0255(17)30652-7
DOI: 10.1016/j.ins.2017.04.011
Reference: INS 12832

To appear in: *Information Sciences*

Received date: 29 June 2016
Revised date: 5 April 2017
Accepted date: 7 April 2017

Please cite this article as: Junbiao Pang, Jing Huang, Lei Qin, Weigang Zhang, Laiyun Qing, Qingming Huang, Baocai Yin, Rotative Maximal Pattern: A Local Coloring Descriptor for Object Classification and Recognition, *Information Sciences* (2017), doi: 10.1016/j.ins.2017.04.011

Rotative Maximal Pattern: A Local Coloring Descriptor for Object Classification and Recognition☆

Junbiao Pang[a], Jing Huang[a], Lei Qin[b], Weigang Zhang[d], Laiyun Qing[b], Qingming Huang[b,c], Baocai Yin[a]

[a]*Beijing Key Laboratory of Multimedia and Intelligent Software Technology, College of Metropolitan Transportation, Beijing University of Technology, No.100 Pingleyuan Road, Chaoyang District 100124, China*
[b]*School of Computer and Control Engineering, University of Chinese Academy of Sciences, No.19 Yuquan Road, Shijingshan District, Beijing 100049, China*
[c]*Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, No.6 Kexueyuan South Road, Haidian District, Beijing 100190, China*
[d]*School of Computer Science and Technology, Harbin Institute of Technology at Weihai, No. 2 West Wenhua Road, Weihai 26209, China*

Abstract

Inspired by the photometric invariance of color space, this paper proposes a simple yet powerful descriptor for object detection and recognition, called Rotative Maximal Pattern (RMP). The effectiveness of RMP comes from the two components: Rotatable Couple Templates (RCTs) with max pooling, and Normalized Histogram Intersection (NHI) with the theoretical guarantee. More concretely, RCTs are the combination of two templates to code the possible rotations. NHI serves as the similarity between two color histograms. We have conducted extensive experiments on INRIA pedestrian and Pascal VOC2007 data sets for object detection tasks; we also show that our approach leads to a promising performance on Caltech 101, Scene 15, UIUCsport and Stanford 40 action data sets.

Keywords:
Object detection, Object recognition, Max pooling, Translation invariance, Self similarity, Photometric invariance

1. Introduction

Recently there has been much interest to use color information in visual object recognition and object detection, *e.g.*, local invariant features [28], Color Name (CN) [15]. There is also several work to evaluate the invariance and the discrimination of color descriptors. Essentially, the success of the invariant photometric of color heavily depends on the appearance of an object itself – whether color is the important cue to describe the object or not. If features from color barely capture the dominant pattern, the color-based descriptors tend to achieve inferior performances. Therefore, a successful approach combines the color cues with other complementary ones, in order to obtain more discriminative ability [37].

There are two ways to combine the complementary cues with color descriptors:

- *At the feature level:* The typical approach [1] combines color cues with Scale-Invariant Feature Transform (SIFT) [37].
- *At the classifier level:* The other cues and color descriptors are directly fused by classifiers, *e.g.*, Multiple Kernel Learning (MKL) [32] or boosting [8].

Technically, concatenating descriptors by the first approach results in a high-dimensional feature vector, naturally increasing both the storage requirement and the computational cost; while the second heavily depends on the generalization ability of classifiers.

Therefore, it is natural to ask whether color cues can be combined with other complementary cues in a more compact scheme without depending on the generalization of classifiers. There are many potential benefits of this scheme: reduced storage requirements, reduced computational complexity, and improved classification performances. A more compact descriptor would simplify both the pattern representations and the subsequent classifiers.

Inspired by the promising results from the fusion of both the shape-based features and the color-based ones [15], we seek an *invariant* descriptor with a *good* generalization ability, based on two motivations. First, although an enormous volume of literature has been devoted to feature fusion, there is little attention about how to obtain compacted descriptors. Second, we want to avoid the disadvantages of the traditional approaches [32]: the computational burden of the enlarged feature vector, and the requirement of the generalization of classifiers. In summary, we desire a computationally simple method to fuse multiple cues into a compacted descriptor.

*Corresponding author
Email addresses: junbiao_pang@bjut.edu.cn (Junbiao Pang), jing.huang@vipl.ict.ac.cn (Jing Huang), lqin@jdl.ac.cn (Lei Qin), wgzhang@hit.edu.cn (Weigang Zhang), lyqing@ucas.ac.cn (Laiyun Qing), qmhuang@jdl.ac.cn (Qingming Huang), ybc@bjut.edu.cn (Baocai Yin)