Expert Systems with Applications 42 (2015) 5658–5667

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Visual orientation inhomogeneity based scale-invariant feature transform

Sheng-hua Zhong [a,b], Yan Liu [c,*], Qing-cai Chen [d]

[a] College of Computer Science & Software Engineering, Shen Zhen University, Shen Zhen, Guang Dong, PR China
[b] Department of Psychological & Brain Science, The Johns Hopkins University, Baltimore, MD 21218-2686, USA
[c] Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China
[d] Shenzhen Graduate School, Harbin Institute of Technology, Shen Zhen, Guang Dong, PR China

A R T I C L E I N F O

Article history:
Available online 16 January 2015

Keywords:
Orientation inhomogeneity
Real-world distribution
Scale-invariant feature transform
Least discriminability

A B S T R A C T

Scale-invariant feature transform (SIFT) is an algorithm to detect and describe local features in images. In the last fifteen years, SIFT plays a very important role in multimedia content analysis, such as image classification and retrieval, because of its attractive character on invariance. This paper intends to explore a new path for SIFT research by making use of the findings from neuroscience. We propose a more efficient and compact scale-invariant feature detector and descriptor by simulating visual orientation inhomogeneity in human system. We validate that visual orientation inhomogeneity SIFT (V-SIFT) can achieve better or at least comparable performance with less computation resource and time cost in various computer vision tasks under real world conditions, such as image matching and object recognition. This work also illuminates a wider range of opportunities for integrating the inhomogeneity of visual orientation with other local position-dependent detectors and descriptors.

1. Introduction

Inspired by the highly discriminatory property of local position-dependent gradient orientation histograms, researchers have proposed a variety of means to detect and describe local features in images, such as scale-invariant feature transform (SIFT) (Lowe, 1999, 2004), histogram of oriented gradients (HOG) (Dalal & Triggs, 2005), gradient location and orientation histogram (GLOH) (Mikolajczyk & Schmid, 2005), and speeded up robust feature (SURF) (Bay, Ess, Tuytelaars, & Gool, 2008). As we known, the dimension of the image feature descriptor has an impact on the running time. The lower dimensions indicate faster interest point matching. However, lower dimensional feature vectors tend to be less distinctive in general. So our goal is to develop both a detector and descriptor that, in comparison to the state-of-the-art, is fast to compute without sacrificing much performance (Bay et al., 2008).

From the research in neuroscience (Girshick, Landy, & Simoncelli, 2011), we know the orientation perception of human is inhomogeneous. Neuroscientists measured the performance in several orientation-estimation tasks and found that orientation discriminability in human observation is worst at oblique angles and best at cardinals (horizontal and vertical). They pursued the physiological instantiation of this phenomenon and found that the non-uniformities in the representation of orientation in the V1 population contribute to non-uniformities in perceptual discriminability. Specifically, a variety of measurements have shown that cardinal orientation is represented by a disproportionately large fraction of V1 neurons, and that those neurons also tend to have narrower tuning curves (Li, Peterson, & Freeman, 2003).

Although we know the property and the physiological evidence of human's orientation perception, we do not know whether this property is useful and helpful to human's visual tasks or it is only a limitation of human's perception. In this paper, we will investigate the real-world orientation distribution in different semantically organized categories. Then, we will provide a human-like feature detector and descriptor by drawing lessons from the orientation inhomogeneity of human visual perception. Unlike the existing standard SIFT algorithm or other detectors and descriptors, the proposed V-SIFT detects, preserves and processes the non-uniformly information from different visual orientation in each stage. The information from cardinals (horizontal and vertical) is retained, but the information from the least discriminatory orientation (oblique orientation) is ignored in our proposed V-SIFT.

The remainder of this paper is organized as follows. Section 2 reviews the related work of the SIFT algorithm. Section 3 details three stages in the proposed visual orientation inhomogeneity SIFT

* Corresponding author.
E-mail addresses: csshzhong@szu.edu.cn (S.-h. Zhong), csyliu@comp.polyu.edu.hk (Y. Liu), qingcai.chen@hitsz.edu.cn (Q.-c. Chen).

http://dx.doi.org/10.1016/j.eswa.2015.01.012
0957-4174/

Fig. 1. Sample images from the Urban and Natural Scene dataset.

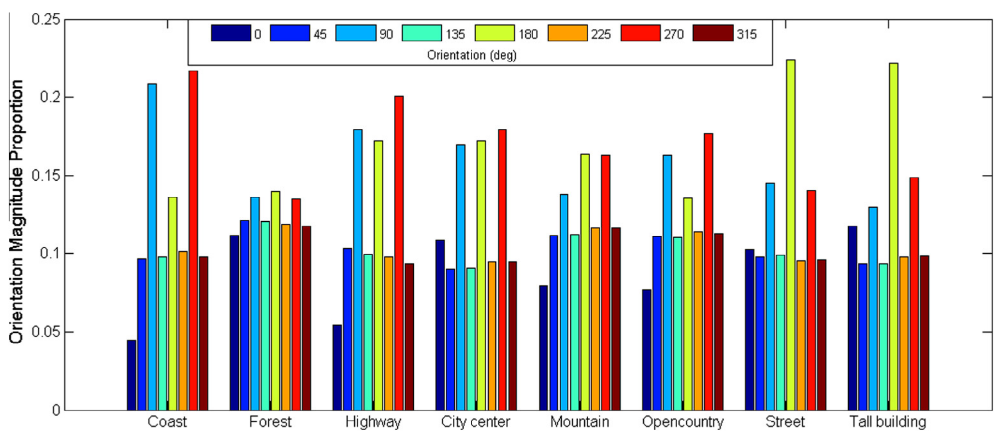

Fig. 2. Average distribution of oriented edges in eight categories. The cardinal orientation especially 90°, 180°, 270° predominate in all categories.

Table 1
Magnitude proportion of cardinal vs. oblique orientation. A paired *t*-test is used to determine the significance of a difference between the cardinal orientation angles vs. oblique orientation angles.

Category	Orientation	Mean ± SEM	*P* Value	Category	Orientation	Mean ± SEM	*P* Value
Coast	Cardinal	0.6999 ± 0.00406	<0.0001	Mountain	Cardinal	0.5436 ± 0.00228	<0.0001
	Oblique	0.3001 ± 0.00406			Oblique	0.4564 ± 0.00228	
Forest	Cardinal	0.5399 ± 0.00407	<0.0001	Open country	Cardinal	0.5718 ± 0.00280	<0.0001
	Oblique	0.4601 ± 0.00407			Oblique	0.4282 ± 0.00280	
Highway	Cardinal	0.6564 ± 0.00532	<0.0001	Street	Cardinal	0.6255 ± 0.00335	<0.0001
	Oblique	0.3436 ± 0.00532			Oblique	0.3745 ± 0.00335	
City center	Cardinal	0.7539 ± 0.00487	<0.0001	Tall building	Cardinal	0.7027 ± 0.00505	<0.0001
	Oblique	0.2461 ± 0.00487			Oblique	0.2973 ± 0.00505	

(V-SIFT) algorithm. Section 4 provides the experimental results from a comparison between V-SIFT and standard SIFT on feature detection and matching experiments. In addition, we demonstrate the performance for object classification task based on these features detectors and describers. Finally, Section 5 concludes this paper and outlines the future work.

2. Related work on SIFT

Scale-invariant feature transform is an algorithm to detect and describe local features in images developed by Lowe (1999, 2004). The SIFT descriptor is invariant to translations, rotations and scaling transformations in the image domain, and it is robust to moderate perspective transformations and illumination variations.

The standard SIFT algorithm firstly detects interest points by searching for the scale-space extrema of differences-of-Gaussians (DoG) within a difference-of-Gaussians pyramid. Then the position-dependent histograms of local gradient directions around the interest points are statistically accumulated as the SIFT descriptor. In the end, the SIFT descriptor is utilized to match the corresponding interest points between different images. Experimentally, the SIFT algorithm has been proven to be very useful in practice for image matching and object recognition under real-world conditions, including image copy detection (Ling, Yan, Zou, Liu, & Feng, 2013), multi-object recognition (Kim, Rho, & Hwang, 2012), image stitching (Brown & Lowe, 2007), neurosurgery (Qian, Hui, & Gao, 2013), human action recognition (Liu, Shao, & Rockett, 2013), video tracking (Saeedi, Lawrence, & Lowe, 2006), and so on.