Int J Comput Vis
DOI 10.1007/s11263-013-0676-2

A Klein-Bottle-Based Dictionary for Texture Representation

Jose A. Perea · Gunnar Carlsson

Received: 19 December 2012 / Accepted: 19 November 2013

Abstract A natural object of study in texture representation and material classification is the probability density function, in pixel-value space, underlying the set of small patches from the given image. Inspired by the fact that small $n \times n$ high-contrast patches from natural images in gray-scale accumulate with high density around a surface $\mathscr{K} \subset \mathbb{R}^{n^2}$ with the topology of a Klein bottle (Carlsson et al. International Journal of Computer Vision 76(1):1–12, 2008), we present in this paper a novel framework for the estimation and representation of distributions around $\mathscr{K}$, of patches from texture images. More specifically, we show that most $n \times n$ patches from a given image can be projected onto $\mathscr{K}$ yielding a finite sample $S \subset \mathscr{K}$, whose underlying probability density function can be represented in terms of Fourier-like coefficients, which in turn, can be estimated from S. We show that image rotation acts as a linear transformation at the level of the estimated coefficients, and use this to define a multi-scale rotation-invariant descriptor. We test it by classifying the materials in three popular data sets: The CUReT, UIUC-Tex and KTH-TIPS texture databases.

Keywords Texture representation · Texture classification · Klein bottle · Fourier coefficients · Patch distribution · Density estimation

J. A. Perea (✉)
Department of Mathematics, Duke University, Durham, NC 27708-0320, USA
e-mail: joperea@math.duke.edu; joperea@gmail.com

G. Carlsson
Department of Mathematics, Stanford University, Stanford, CA 94305, USA
e-mail: gunnar@math.stanford.edu

1 Introduction

One representation for texture images which has proven to be highly effective in multi-class classification tasks, is the histogram of texton occurrences (Crosier and Griffin 2010; Hayman et al. 2004; Jurie and Triggs 2005; Lazebnik et al. 2005; Leung and Malik 2001; Varma and Zisserman 2005, 2009; Zhang et al. 2007). In short, this representation summarizes the number of appearances in an image of either, patches from a fixed set of pixel patterns, or the types of local responses to a bank of filters. Each one of these pixel patterns (or filter responses, if that is the case) is referred to as a texton and the set of textons as a dictionary. Images are then compared via their associated histograms using measures of statistical similarity such as the Earth Mover's distance (Rubner et al. 2000), the Bhattacharya metric (Aherne et al. 1998), or the chi square similarity test as introduced by Leung and Malik (2001).

For images in gray-scale and dictionaries with finitely many elements, the coding or labeling of $n \times n$ pixel patches, represented as column vectors of dimension n^2, can be seen as fixing a partition $\mathbb{R}^{n^2} = \mathscr{C}_1 \cup \cdots \cup \mathscr{C}_d$ of $\mathbb{R}^{n^2}$ into d distinct classes (each associated to a texton) and letting a patch contribute to the count of the i-th bin in the histogram if and only if it belongs to $\mathscr{C}_i$. For instance, if a patch is labeled according to the dictionary element to which it is closest with respect to a given norm, then the classes $\mathscr{C}_i$ are exactly the Voronoi regions associated to the textons and the norm. If labeling is by maximum response to a (normalized) filter bank, which amounts to selecting the filter with which the patch has largest inner product, then the classes $\mathscr{C}_i$ are the Voronoi regions associated to the filters with distances measured with the norm induced by the inner product used in the filtering stage. More generally, any partition of filter response (or feature) space induces a partition of patch space,