BSIF: Binarized Statistical Image Features

Juho Kannala and Esa Rahtu,
Center for Machine Vision Research,
University of Oulu, Finland

Overview

• A local image descriptor is constructed by binarizing the responses to linear filters

• In contrast to previous binary descriptors, the filters are learnt from natural images using independent component analysis (ICA)

• The proposed descriptor performs well in texture classification and face recognition tasks

Background

• Our descriptor is inspired by other methods that also describe image patches using binary code strings, such as local binary patterns (LBP), local phase quantization (LPQ) and binary robust independent elementary features (BRIEF)

• All these previous approaches can be seen to consist of two stages:
  – Filtering the image with a set of linear filters
  – Quantization of filter responses

• The main difference between our method and the previous approaches is that we learn the filters from natural images instead of using hand-crafted or random filters

Method

Given an image patch X of size $l \times l$ pixels and a linear filter W_i of the same size, the filter response s_i is obtained by

$$s_i = \sum_{u,v} W_i(u,v)X(u,v) = \mathbf{w}_i^\top \mathbf{x}, \quad (1)$$

where vector notation is introduced in the latter stage. If we have n linear filters W_i, we may stack them to a matrix $\mathbf{W}$ and compute all responses at once:

$$\mathbf{s} = \mathbf{W}\mathbf{x}. \quad (2)$$

Given a random sample of natural image patches, we determine the filters W_i so that the elements s_i of $\mathbf{s}$ are as independent as possible when considered as random variables (see [1]). The binary code string $\mathbf{b}$, which corresponds to image patch $\mathbf{x}$, is obtained by binarizing each element s_i of $\mathbf{s}$ as follows:

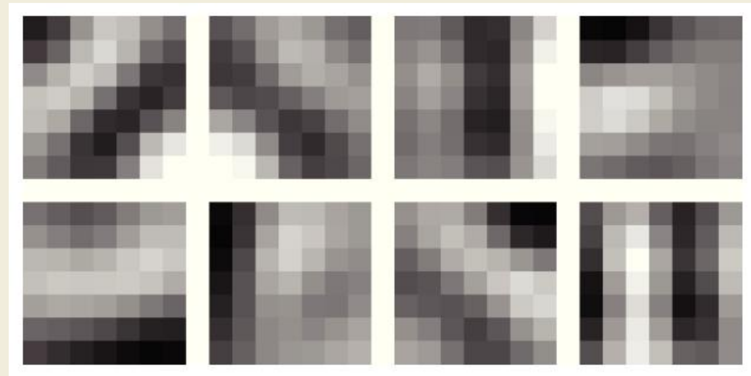
$$b_i = \begin{cases} 1 & \text{if } s_i > 0 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

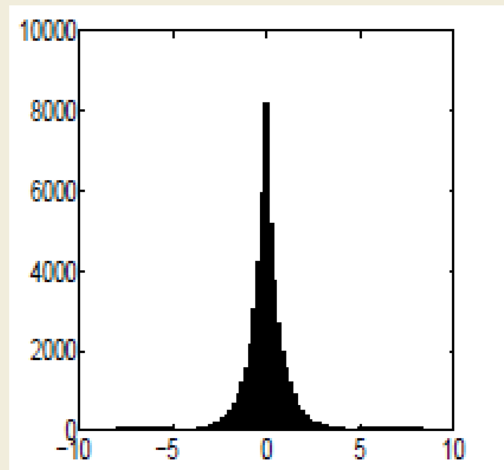
where b_i is the i^{th} element of $\mathbf{b}$.

In this manner one may compute a n-bit binary code string $\mathbf{b}$ for each pixel and thereafter image regions can be represented by histograms of pixels' binary codes.

References

[1] A. Hyvärinen et al. *Natural Image Statistics*. Springer, 2009.





Left: Examples of eight learned filters of size 7x7 pixels
Right: A typical histogram of filter response values (heavy-tailed)

Results

• We experimented our approach in two tasks: texture classification (Outex and CUReT databases) and face recognition (FRGC database)

• In both cases the classification was performed using nearest neighbor classifier and chi-squared distance metric for descriptor histograms

Texture classification

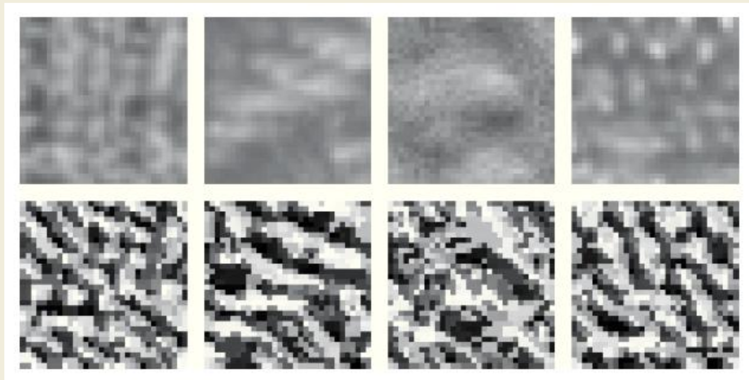


Figure 2. Samples from Outex database (top) and corresponding BSIF codes.

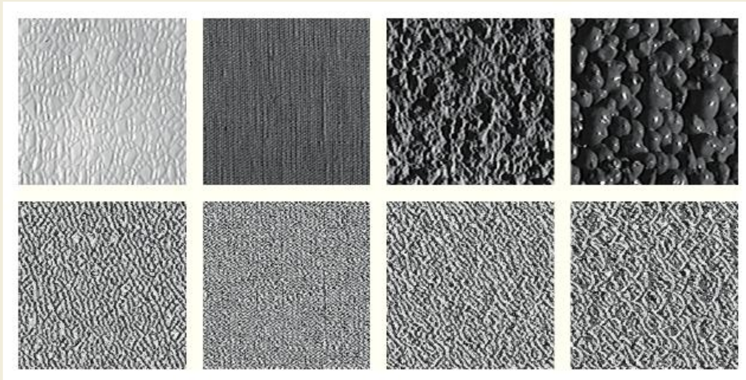
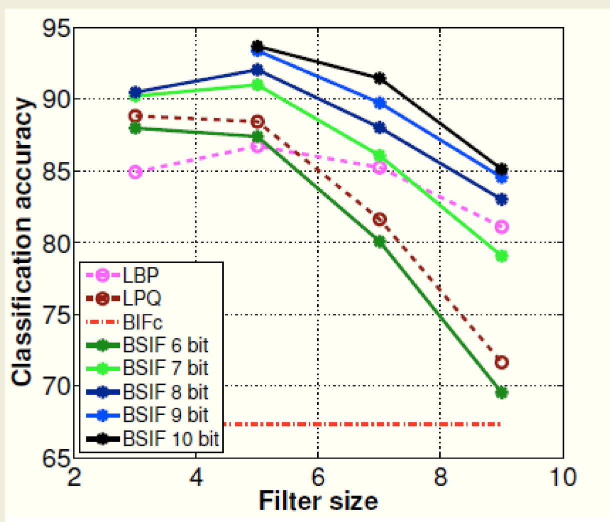
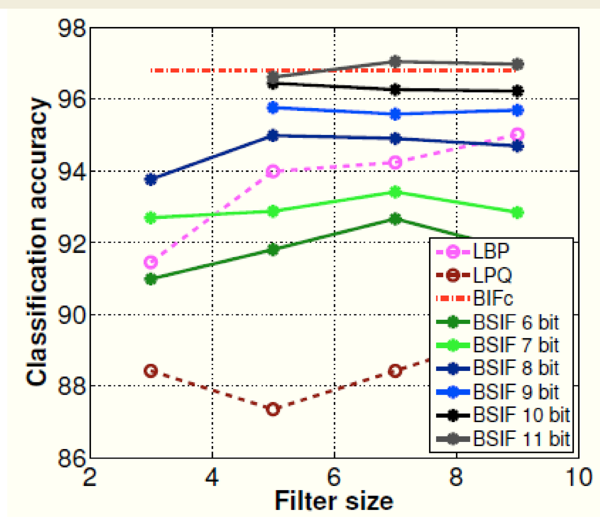


Figure 3. Samples from CUReT database (top) and corresponding BSIF codes.



(a) Outex



(b) CUReT

Face recognition

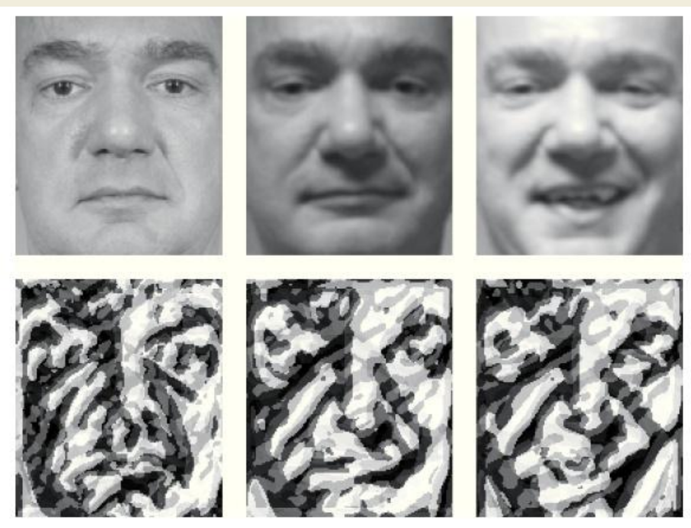
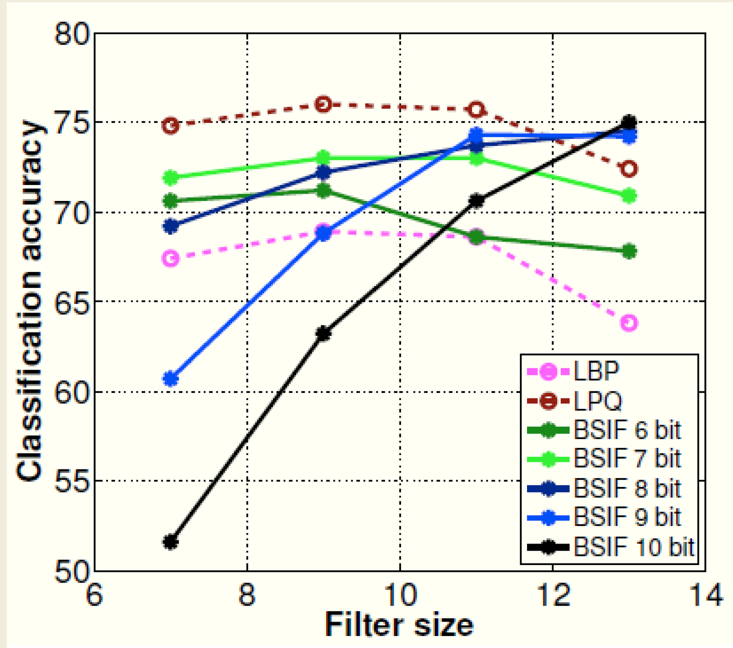


Figure 4. A sample gallery image and two probe images images from FRGC (top), and corresponding BSIF codes. The blur in the probe images is clearly observable.



(c) FRGC