SIFT texture description for understanding breast ultrasound images

Joan Massich, Fabrice Meriaudeau, Melcior Sentís, Sergi Ganau, Elsa Pérez, Domenec Puig, Robert Martí, Arnau Oliver and Joan Martí

Abstract

Texture is a powerful cue for describing structures that show a high degree of similarity in their image intensity patterns. This paper describes the use of Self-Invariant Feature Transform (SIFT), both as low-level and high-level descriptors, applied to differentiate the tissues present in breast US images.
For such a task, a subset of 16 images has been randomly selected from a larger dataset of 700 Ultra-Sound (US) images acquired at the *UDIAT Diagnostic Centre of Parc Taulí* in Sabadell (Catalunya), between 2010 and 2012. This subset has been complemented with multi-label Ground Truth (GT), as illustrated in figure 1.

Fig. 1: Dataset sample. From left to right: image sample, accompanying multi-label GT, tissue label GT color-coding.

SIFT as a low-level descriptor, tested using Maximum A Posteriori (MAP)

In this experiment, it has been analyzed how separable are the tissue classes present in breast ultrasound images, when using low-level descriptors based on SIFT to encode US texture. Here a Bayesian framework has been assumed to perform the tissue discrimination and its results are presented both qualitatively (see fig. 2-4) and quantitatively (see fig. 5).
All the pixel positions of all the images are used as a key-point for extracting a SIFT descriptor and mapped in the $128D$ feature space of SIFT. This SIFT space is then projected into a $2D$ space to visually assess how the tissue classes are distributed in such space. From this projected space, models (see fig. 2) and priors (see fig. 3) are extracted to infer the MAP probability. Figure 4 shows this MAP probability illustrating how the tissue classes are separated based on the observed data.
In order to generate cross-validated quantitative results, the descriptors have been randomly sampled as follows: (10.000 samples $\times$ 10 classes) $\times$ 5 folds. At each round 4 folds have been used for training the models and the remaining fold has been used for testing. Figure 5 shows the corss-validated confusion matrix resulting from classifying breast tissues based on a Bayesian framework using either low-level SIFT texture descriptors or intensity. The trance of the confusion matrix corresponds to the sensitivity which provides a general sense of performance across all the labels. The True-Positive Ratio (TPR) value obtained for the intensity case is $16.6 \pm 27.5\%$, whereas for the SIFT case is $18.8 \pm 17.2\%$ which show that both feature spaces produce similar results.

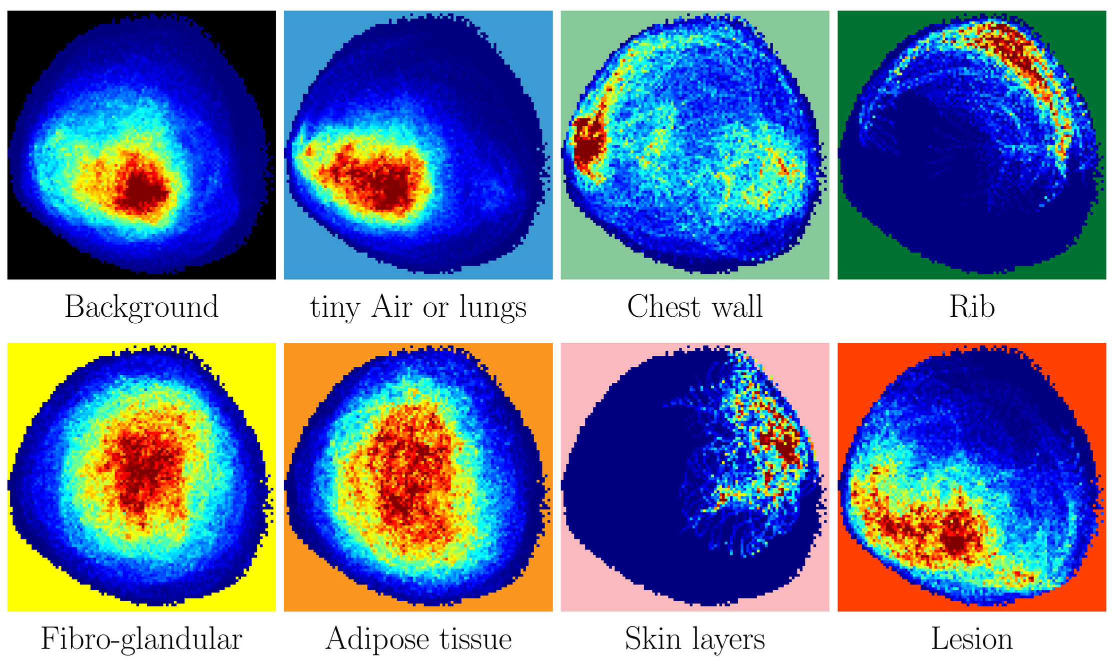

Fig. 2: Distribution of the SIFT descriptors for some classes in the GT.

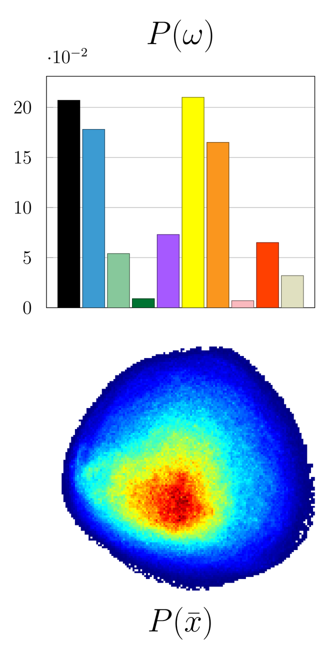

Fig. 3: Data prior knowlage.

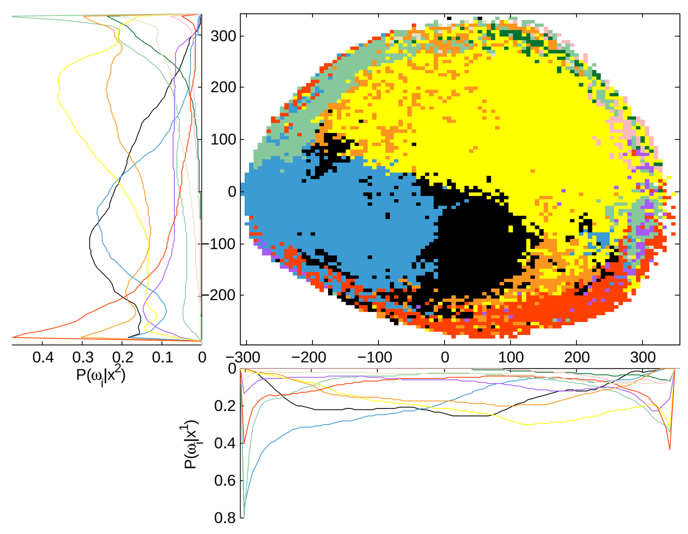

Fig. 4: Qualitative evaluation of the MAP labeling of the feature space.

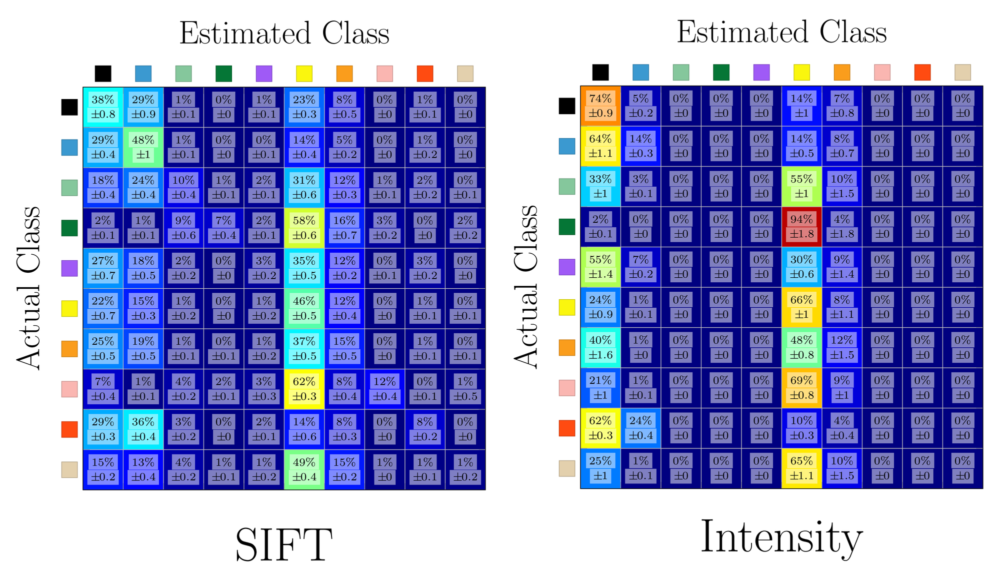

Fig. 5: Confusion matrix showing quantitative results obtained from (10.000 samples $\times$ 10 classes) $\times$ 5 folds cross-validation.

SIFT as a high-level descriptor using Bag-of-Features (BoF), tested using Radial Basis Function (RBF)-Support Vector Machine (SVM) classifier

Texture is an area property related to spatial repetition of structures, statistical similarities, or both. In this experiment superpixels are extracted using Quick-Shift (QS) methodology to generate this areas. Some of this superpixels are illustrated in figure 6. For each superpixel, a high-level texture descriptor is build as a BoF of SIFT descriptors. Initially the SIFT space is clustered to generate a codebook. Here k-means procedure with $k = 36$ is used to generate these codebooks. Figure 6 shows an arbitrary coloring of each cluster to illustrate how the codebook produce a hard quantification of the SIFT space. Finally the BoF descriptor for a superpixel corresponds to the occurrence study of the codebook using the SIFT descriptors belonging to this particular superpixel (see 1-8 in fig 6).
In order to produce quantitative results, dataset of superpixels is generated based on the images and the multi-label GT available. For a superpixel to be eligible, an area larger than 75% need to belong to a single GT label. The resulting dataset contains 20 folds of 8 superpixels (one per class).
The experiment has been repeated for different coodebooks to take into account the variability of the coodebook within the results. Figure 7 compares the results of using BoF based on SIFT or intensity for encoding US texture. The sensitivity achieved is $29 \pm 3.6\%$ for the intenisty and $33.5 \pm 2.3\%$ for SIFT.

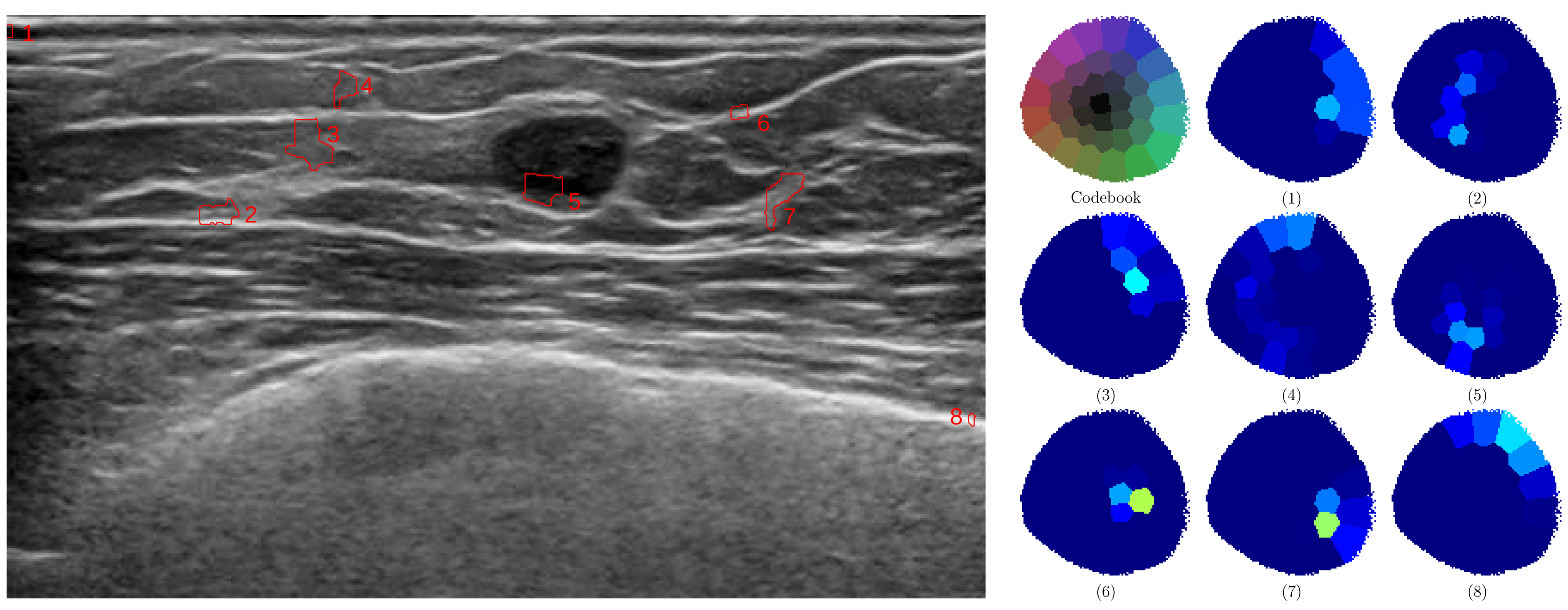

Fig. 6: SIFT-BoF descriptors qualitative analysis. (Left) image example. (Right) Dictionary representation colored using the location of the keypoint location in fig. ??a space. (1-8) Occurrence of the dictionary's key-points associated to each region highlighted in the original image.

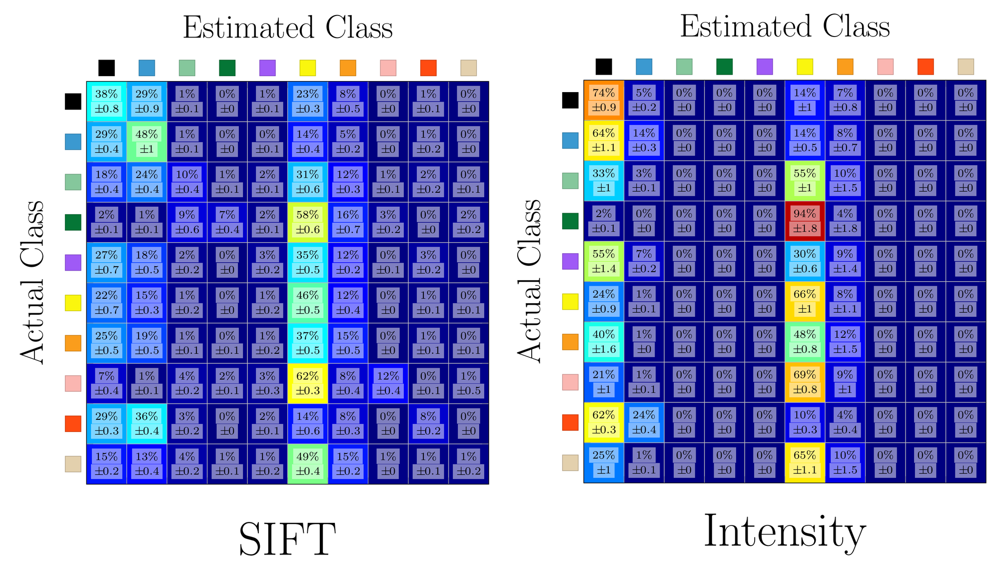

Fig. 7: some caption

Conclusion

The present study was designed to explore the usage of SIFT feature space as a texture for characterizing the different tissues present in a breast US image. The usage of SIFT either as a low-level or high-level texture descriptor has been evaluated in comparison to intensity features, which are the features most commonly used. The fact that SIFT and intensity descriptors