SIFT texture description for understanding breast ultrasound images

Joan Massich[1,2]*, Fabrice Meriaudeau[2], Melcior Sentís[3], Sergi Ganau[3], Elsa Pérez[4], Domenec Puig[5], Robert Martí[1], Arnau Oliver[1], and Joan Martí[1]

[1] Computer Vision and Robotics Group, University of Girona, Spain.
jmassich@atc.udg.edu
[2] Laboratoire Le2i-UMR CNRS, University of Burgundy, Le Creusot, France.
[3] Department of Breast and Gynecological Radiology, UDIAT-Diagnostic Center, Parc Taulí Corporation, Sabadell, Spain.
[4] Department of Radiology, Hospital Josep Trueta of Girona, Spain.
[5] Department of Computer Engineering and Mathematics, University Rovira i Virgili, Tarragona, Spain.

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 the low-level texture descriptors case, SIFT descriptors are extracted from a regular grid. The high-level texture descriptor is build as a Bag-of-Features (BoF) of SIFT descriptors. Experimental results are provided showing the validity of the proposed approach for describing the tissues in breast US images.

Key words: breast cancer, ultrasound, texture, SIFT

1 Introduction

Breast cancer is the second most common cancer (1.4 million cases per year, 10.9% of diagnosed cancers) after lung cancer, followed by colorectal, stomach, prostate and liver cancers. In terms of mortality, breast cancer is the fifth most common cause of cancer death. However, it places as the leading cause of cancer death among females both in western countries and in economically developing countries [3].

Medical imaging plays an important role in breast cancer mortality reduction, contributing to its early detection through screening, diagnosis, image-guided biopsy, treatment follow-up and suchlike procedures [5]. Despite Digital Mammography (DM) still remains as the image modality of reference for diagnose purposes, Ultra-Sound (US) offers useful complementary diagnose information

* This work was partially supported by the Spanish Science and Innovation grant nb. TIN2012-37171-C02-01 and TTIN2012-37171-C02-02 and the Regional Council of Burgundy.

due to its capabilities for differentiating between solid lesions that are benign or malignant [6]. It is estimated that between 65 $\sim$ 85% of the biopsies prescribed using only mammography imaging could be avoided if US information had been taken into account while issuing the diagnose [7].

In US images, texture is a major characteristic to distinguish between different breast tissues, which also allows assessing of the lesion's pathology [6]. Thus, the importance of incorporating texture data from US images into Computer Aided Diagnosis (CAD) systems. A comprehensive list of texture descriptors used for detection, segmentation or diagnose tasks applied to US breast images is given in Cheng *et al.* [1], where most of the descriptors are *ad-hoc* descriptors or based on well-known texture descriptors such as co-occurrence matrices, wavelet coefficients or Gray-Level Difference Method (GLDM).

This article explores the usage of Self-Invariant Feature Transform (SIFT) descriptors for encoding the US characteristic texture produced by the speckle noise present within the images. Its performance is evaluated using a multi-label annotated dataset.

2 Material and methods

In order to develop segmentation methodologies applied to delineate breast lesions in US data, a set of 700 US images was acquired at the *UDIAT Diagnostic Centre of Parc Taulí* in Sabadell (Catalunya), between 2010 and 2012. All the images were provided with accompanying Ground Truth (GT) delineation of the lesions present in the image. From this image database, a reduced dataset of 16 images corresponding to different patients was selected and complemented with multi-label GT in order to evaluate the texture description of the observable tissues in the breast.

Figure 1 illustrates a breast image from the dataset with its associated GT.

3 Using SIFT as a low-level texture descriptor in order to differentiate the tissues present in breast US images

Self-Invariant Feature Transform (SIFT) [4] transforms key-points into scale and rotation invariant coordinates relative to local features. The SIFT descriptor at

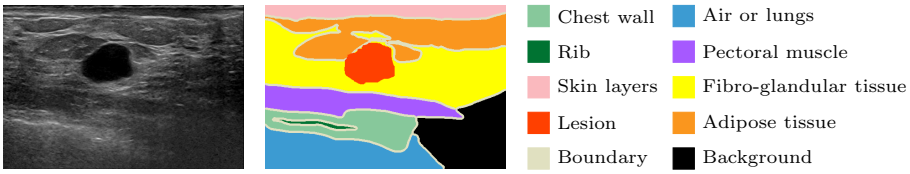
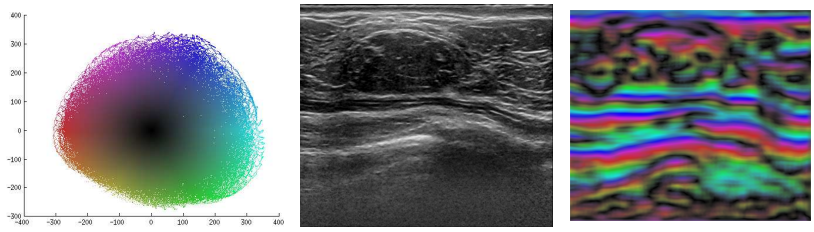


Fig. 1: Dataset sample. From left to right: image sample, accompanying multi-label GT, tissue label GT color-coding.



(a) (b) (c)

Fig. 2: Low level SIFT descriptor example. (a) Arbitrary coloring of the projected SIFT space. (b) Original image. (c) Recoding of the extracted SIFT descriptors using the color coding in (a).

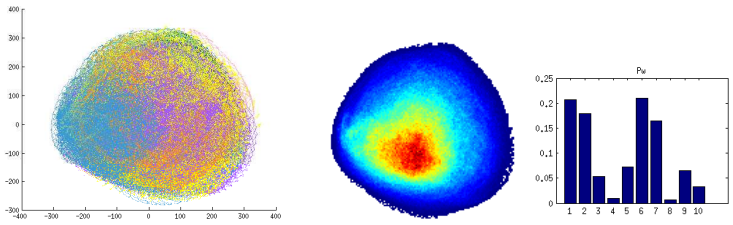
a particular key-point, samples the magnitude and orientation of the gradients surrounding this key-point to generate a 128 element feature. When setting up SIFT as a texture descriptor, the key-points are considered to be a regular grid in order to generate evenly sparse SIFT descriptors.

The usage of SIFT descriptor brings invariability to scale, rotation and minor affine transformations along with robustness to illumination changes [4], which allows to characterize the tissues despite the variability from US adquisition.

Figures 2 to 6 are used to analyze both qualitatively and quantitatively the usage of SIFT as a low-level descriptor to encode US texture. In order to study the image in terms of SIFT descriptors the images are mapped to this SIFT space by extracting a SIFT descriptor at every pixel position. In order to visually keep track of this SIFT, their 128 dimensions are projected into a two dimensional space using Principal Component Analysis (PCA). When combining features using PCA is convenient to know the ratio known as explained variation, which is in this case is given by $\frac{\lambda_1+\lambda_2}{\sum_{i=1}^{128}\lambda_i} = 21.6\%$. For the rest of the article all the calculation are carried out in the projected space, under the assumption that if any separability can be found, in a higher space with greater explained variation would have greater separability. Figure 2 offers a visual interpretation of a breast US image in terms of low-level SIFT descriptors. On it, fig. 2a shows the scatter plot resulting from projecting all the SIFT descriptors extracted from all the images within the dataset to the space described by their two principal components. The descriptors, when projected into this two dimensional space, have been arbitrary colored in such a manner that two close samples share similar color. In this manner, this arbitrary coloring can be used to color code the SIFT descriptors and visually assess the images in terms of SIFT (see fig. 2c).

In order to analyze the tissue distribution in this texture space of SIFT descriptors, the analysis of the Maximum A Posteriori (MAP) estimator has been chosen, as described in equation 1.

$$P(\omega|\bar{x}_a) = \frac{P(\bar{x}_a|\omega) \cdot P(\omega)}{P(\bar{x}_a)} \tag{1}$$



(a) (b) (c)

Fig. 3: 2D visualization of the SIFT hyper space. (a) Projected space colored according to GT tissue labeling. (b) Occurrence density.

Where $P(\omega|\bar{x}_a)$ is the probability that the sample a belongs to class $\omega \in W$ (see fig. 1b as a reminder of the GT available classes) based on $\bar{x}_a$ which is the feature vector representing the sample a, such that x_a^i is the ith feature. $P(\bar{x}_a|\omega)$ corresponds to the Maximum Likelihood (ML) of the feature distribution for a particular class ω, while $P(\omega)$ and $P(\bar{x}_a)$ are the priors for the class and feature respectively.

alin the figures in fig. 3

Figure 3 uses the entire dataset to illustrate the underlying problem and the priors extracted from the same dataset. Fig. 3a shows a scatter plot where every sample has been colored according to its GT. Fig. 3b shows an occurrence study of the samples carried out in a discretization of the space in fig. 3a that illustrates the distribution of the SIFT descriptors present within the images, which corresponds to the $P(\bar{x}_a)$ term in eq. 1. Fig. 3c illustrates the class prior corresponding to the proportion of samples present in the dataset for each class.

Figure 4 shows the feature distribution study for every class, corresponding to the $P(\bar{x}_a|\omega)$ in eq. 1. On it, tendencies of features for representing different classes can be observed illustrating the similarities and dissimilarities between classes. As example the reader is referred to fig. 1 to illustrate the fact that adipose tissue class contains Cooper's ligaments which are highly dense fibers, and fibro-glandular tissue is made of dense fibers and unstructured fat. Or, the difficulty of producing accurate GT delineations. Take for instance the label background which is meant for regions where the breast structures are no clear enough for the reader. A concept which is highly reader dependent.

Using eq. 1 a MAP estimation of the probability for each class can be calculated at every position based on the data shown in fig. 3 and fig. 4. Equation 2 illustrates produce the preferred labeling of the space which is illustrated in figure 5a. On it, the marginals $P(\omega_i|x^j)$ where $j \in \{1, 2\}$ are also represented to obtain a deeper understanding of the MAP.

$$labeling(\bar{x}) = \arg\max_i \quad P(\omega_i|\bar{x}) \qquad \text{where } i \in [1..|W|] \tag{2}$$

For comparison purposes, figure 5 is complemented by repeating the experiment carried out in fig.figure 5a, this time using intensity as single feature such that $x_a \in [0..255]$ (see fig. 5b). When comparing both labeled spaces, the SIFT

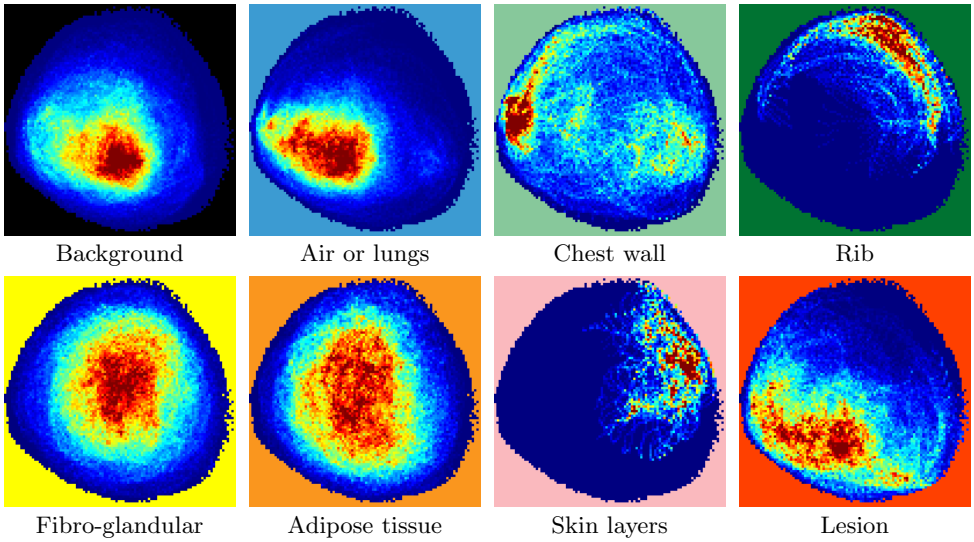
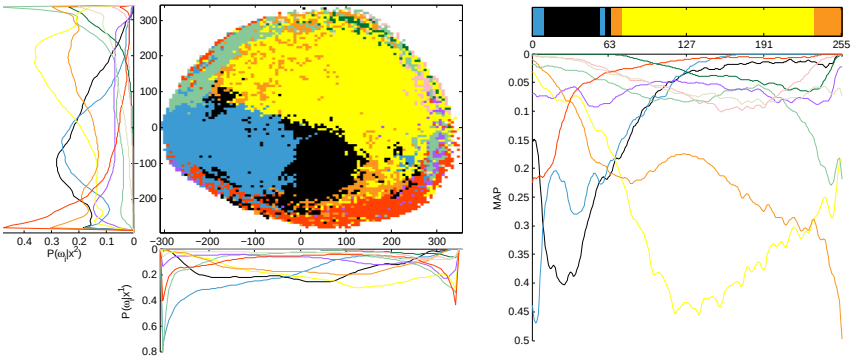


Fig. 4: Distribution of the SIFT descriptors for some classes in the GT

feature is preferred since exist no mode for some of the classes when using intensity.

In order to generate cross-validated quantitative results, it has been sampled from the dataset, 5 folds of 10.000 samples for each class per fold. At each round 4 folds have been used for training the ML term in eq. 1 ($P(\bar{x}_a|\omega)$) and the ramaining fold has been used for testing the classification. Figure 6 uses boxplots to illustrate the distribution of the confusion matrix across the folds. In the figure, the samples are grouped by the actual class of the sample and divided by the predicted classes. The top label represents the samples' actual class, whereas the predicted class is color coded at the bottom. Boxplots in blue represent the results of classifying the samples using intensity, whereas the bloxpots in red represent the results obtained when using SIFT. The lack of variability within the boxplots illustrates a repeatability of the results across the samples, which gets accentuated when using SIFT. The results show that the preferred labels which cover larger portion of the feature space achieve better results than the other classes. This is more clear for the intensity case since there are classes with no mode and therefore all the samples of this class are misclassified (see fig. 5). The sensitivity or True-Positive Ratio (TPR) allows to obtain a general sense of performance across all the labels. The 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. Notice that the large variability reported is due to missclassification of the labels with no mode, as can be observed in figure 5.



(a) SIFT (b) Intensity

Fig. 5: Qualitative evaluation of the MAP labeling of the feature space .

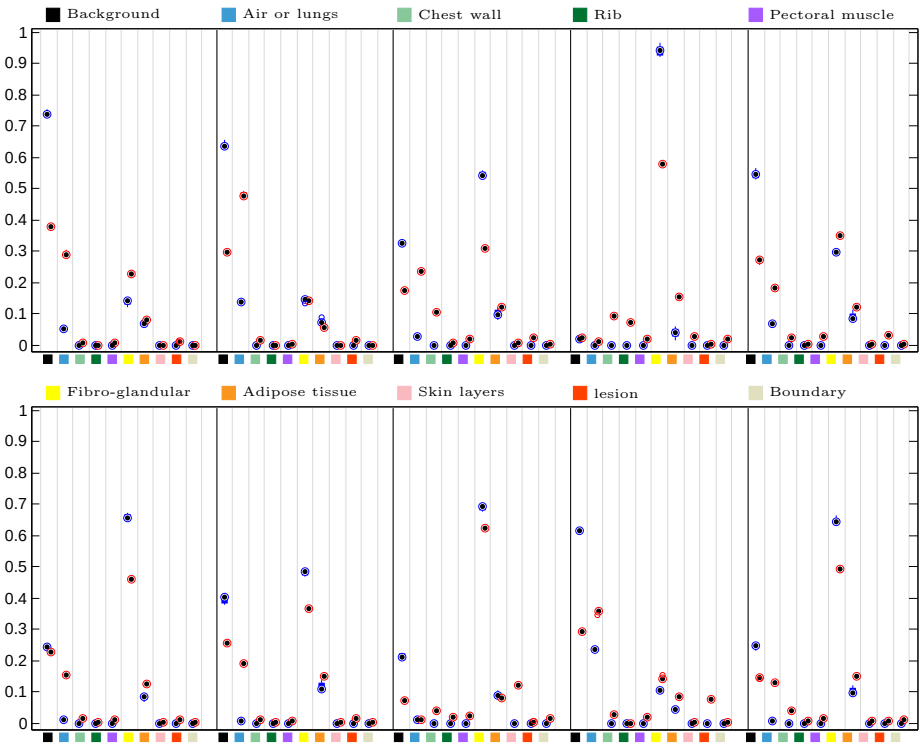


Fig. 6: Confusion matrix results distribution represented as boxplots. The results are grouped by actual class of the samples and subdivided by the predicted label.

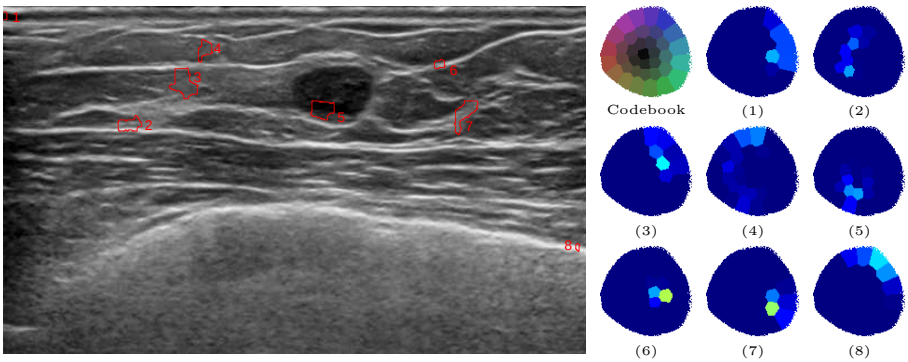


Fig. 7: SIFT-BoF descriptors qualitative analysis. (Left) image example. (Right) Dictionary representation colored using the location of the keypoint location in fig. 3a space. (1-8) Occurrence of the dictionary's key-points associated to each region highlighted in the original image.

4 High-level texture descriptor using Bag-of-Features (BoF) and SIFT descriptors

Texture is an area property related to spatial repetition of structures, similar statistical properties of the area or both. A technique to embed statistical properties of a low level descriptor is Bag-of-Features (BoF) which analyses the occurrence of a set of keywords (or key-points) within a particular region [2].

In this proposal, the words or features representing the images are SIFT descriptors. In order to determine the words forming the dictionary or codebook needed to generate the BoF descriptors, the space of SIFT descriptors is clustered in order to produce a hard quantification of this space. In this case, a k-means procedure with $k = 36$ is used to generate the codebook. To generate the BoF-SIFT feature, all the SIFT descriptors are substituted for the closest SIFT descriptor in the codebook. Finally the texture description from a particular area is expressed as the keywords' occurrence in this area. The descriptor is normalized so that the sum of all the occurrences is 1.

For this application, the super-pixels of the images have been extracted using Quick-Shift (QS)in order to extract the BoF from the obtained super-pixel regions.Figure 7 illustrates the BoF-SIFT feature. On the original image, some of the super-pixels extracted have been highlighted. Figure also shows a codebook partitioning the feature space in 36 groups along with the BoF descriptors for the 8 highlighted. For the visualization of the BoF features a jet color coding has been used to represent the occurrence of each word within the codebook.

In order to quantitatively assess the performance of SIFT embedded within in a high-level feature descriptor such as BoF, a dataset of super-pixels with its associated GT and BoF-SIFT descriptor has been build up. In order for the super-pixels to be eligible it has been set up that the superpixel should be larger than 50 pixels and be fully contained within the same GT label. This second

constrain has been relaxed for skin and rib classes allowing super-pixels with 75% label contained to be eligible. The study has been carried out only for all the main classes, thus excluding background and boundary classes. To carry out the evaluation 20 folds of 8 super-pixels (one per class) have been selected forming a set of 152 samples for training and 8 samples for testing at each round. The experiments have been repeated under the same conditions with 3 different codebooks in order to take into account the variability introduced by the codebook building. As for the case of low-level texture description the experiments have been duplicated using intensity for comparison purposes. The classification has been carried out using Support Vector Machine (SVM). The TPR results achieved are $29 \pm 3.6\%$ for the case of intensity and $33.5 \pm 2.3\%$ for the case of SIFT, showing their similar performance and the improvement from using high-level texture descriptor over the low-level texture descriptor.

5 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. During the study, SIFT information have been used both as a low-level texture descriptor and encoded within a high-level texture descriptor using BoF. When comparing the performance of using the SIFT space with the intensity to characterize tissues both spaces reach similar results.

One of the limitations of the study here presented is that in order to get a graphical representation of the SIFT space, the 128 dimensions have been reduced to two compromising its performance.

The fact that despite all this limitations, SIFT and intensity spaces, still produce similar results encourage further studies on using SIFT.

References

1. H D Cheng, Juan Shan, Wen Ju, Yanhui Guo, and Ling Zhang. Automated breast cancer detection and classification using ultrasound images: A survey. *Pattern Recognition*, 43(1):299–317, 2009.
2. Gabriella Csurka, Christopher Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray. Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV*, volume 1, page 22, 2004.
3. Ahmedin Jemal, Freddie Bray, Melissa M. Center, Jacques Ferlay, Elizabeth Ward, and David Forman. Global cancer statistics. *CA: A Cancer Journal for Clinicians*, 61(2):69–90, 2011.
4. David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
5. Robert A Smith, Debbie Saslow, Kimberly Andrews Sawyer, Wylie Burke, Mary E Costanza, WPIII Evans, Roger S Foster, Edward Hendrick, Harmon J Eyre, and Steven Sener. American cancer society guidelines for breast cancer screening: update 2003. *CA: a cancer journal for clinicians*, 53(3):141–169, 2003.
6. A Thomas Stavros. *Breast ultrasound*. Lippincott Williams & Wilkins, 2004.

7. Yading Yuan, Maryellen L Giger, Hui Li, Neha Bhooshan, and Charlene A Sennett. Multimodality computer-aided breast cancer diagnosis with ffdm and dce-mri. *Academic radiology*, 17(9):1158, 2010.