Medical image information representation: Gabor Filter solution for the Big Data

N Bourkache, Sahbi Sidhom, M Laghrouche

Medical image information representation: Gabor Filter solution for the Big Data.

N. Bourkache, Sahbi Sidhom., and M. Laghrouche.

Abstract—In the health field, several thousand images are generated every day in medical imaging establishments. The volume of information involved is still far from being fully controlled. On the other hand, the development of machine learning tools today opens the way to a new generation of image analysis in this context of "BigData". Moreover, our approach is part of this research dynamic. In order to test the robustness of our algorithm and its degree of adaptation to BigData, we tested, in a first phase of analysis, our algorithm on an image-database containing 320 mammograms. The precision obtained is estimated at 75% for a recall of 33%. In a second analysis phase, we performed the test on an image data-base containing 1000 medical images. The precision obtained is estimated at nearly 70% for a recall of 33%. Although the precision obtained in this first step is far from perfect, our processing algorithm remains promising and shows a good adaptation to the management of "Digdata

Key words— image representation; image retrieval; Gabor Filter; image databases; Big Data.

1 INTRODUCTION

CURRENTLY, the number of medical imaging exams is estimated to be close to 4 billion per year worldwide. If we take the example of cancer, we see that imaging is closely linked, even essential, in all phases: Diagnosis, treatment and follow-up after treatment. In this area, the interest of imagery is paramount. Because, early detection (by imaging) may well reveal anomalies that predict imminent cancer! in such cases, and in order for appropriate measures to be taken early, it is desirable that the conclusion of the radiologist is as precise as possible. In this context, Computer Aided Diagnosis (CAD) systems are essential and then help in decision-making. In this article, we will present our medical imagery analysis tool based on Gabor wavelets.

2 RELATED WORKS

Since the 1980s, image analysis tools have been constantly evolving. Particularly, the CBIRS (Content-Based Image Retrieval System) take an important place and involves the interest of the scientific community since the 90s. In this type of systems, the image analysis is based on the extraction of the morphological features of the image namely: Color, shape, or texture. In the medical field (especially in cancer screening), recent work is mainly oriented towards diagnostic assistance and decision-making. We find in [1] [2] [3] Classification algorithms of Lung Nodules into Benign or Malignant represented in a CBIR System. in CT *scans* (Computed Tomography) several methods in images classification and analysis are proposed. For example, [4] propose a classification and analysis of pulmonary nodules in CT images using random forest. In prostate cancer diagnosis, [5] offers a cancer classification based in genetic algorithm. In breast cancer, [6] present a new method based on the expert annotation and automatic selection of cell types by their transcriptome profiles. [7] offers a similarity measure method for mammogram retrieval. For large data volumes of the order of BigData, several approaches are proposed [8] [9] [10].

3 RESEARCH WORK

in this paper we represent our image analysis algorithm based on the extraction of morphological features (digital signature) from medical images. the objective of this work is to provide a learning tool and diagnostic aid in breast cancer.

3.1 Gabor Filter Algorithm in image processing: Texture representation

In this approach, we had chosen to study the texture parameter to be able to construct the digital component of images. Parameters will be collected in the form of a vector, called "texture vector" or "digital signature" This one should be not sensitive to image transformations: particularly for translation and rotation of the image. For this important reason, we have chosen in our study the coding by Gabor wavelets.

As illustration cf. Fig. 1, the main algorithm for image indexing applying the Gabor filter Model.

This architecture represents the principal steps of features image representation.

- *N.B. is with the University of Mouloud Mammeri (Tizi-Ouzou) & LAMPA Lab., Algeria. E-mail: nourbourkache@yahoo.fr*
- *S.S. is with the University of Lorraine (Nancy) & LORIA Lab., France. E-mail: sahbi.sidhom@loria.fr*
- *M.L. is with the University of Mouloud Mammeri (Tizi-Ouzou) & LAMPA Lab., Algeria. E-mail: larouche_67@yahoo.fr*

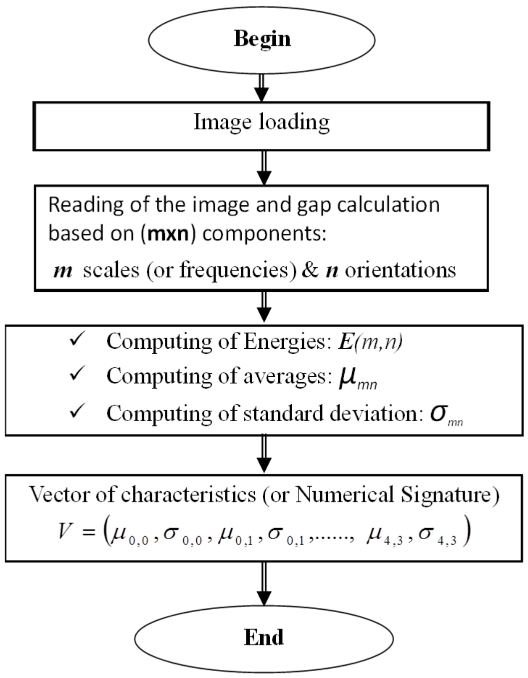

Fig. 1. Flowchart of the main algorithm for image indexing applying the Gabor (filter) Model.

Practically, the Gabor wavelet proves to be an interesting tool for texture analysis applied to image and it is largely adopted in performance measures.

3.2 Gabor Filter analysis

For an image I(x, y) having dimensions MxN, its conversion into discrete Gabor Wavelet is given by the following convolution formula:

$$G_{mn}(x,y) = \sum_s \sum_t I(x-s, y-t)\psi^*_{mn}(s,t) \qquad (1)$$

Ψ^* is the combined of $\Psi(x,y)$ such as the formula:

$$\psi(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + j2\pi \cdot f \cdot x\right] \qquad (2)$$

After applying the Gabor filter to the image with various orientations and levels of filtering, we get two computed formula of σmn (average) and μmn (standard deviation):

$$\mu_{m,n} = \frac{E(m,n)}{MxN} \qquad (3)$$

with: $$E(m,n) = \sum_x \sum_y \left|G_{m,n}(x,y)\right| \qquad (4)$$

$$\sigma_{m,n} = \frac{\sqrt{\sum_x \sum_y \left(\left|G_{m,n}(x,y)\right| - \mu_{m,n}\right)^2}}{MxN} \qquad (5)$$

The values of σmn and μmn represent the components of the characteristics vector (V). Thus, for four orientations and five scales, this vector V had the following formula:

$$V = (\mu_{0,0}, \sigma_{0,0}, \mu_{0,1}, \sigma_{0,1}, \ldots\ldots, \mu_{4,3}, \sigma_{4,3}) \qquad (6)$$

At this level, images are represented by the characteristic vectors in the space of numeric attributes.

3.3 Searching step

The research phase, the similarity measure between images is defined by a set of distances in the same defined space.

The similarities are computed with the image-query Q and the image-targets T (Stored in the image-database) using for each vector values the distance *D(Q,T)* using the formula:

$$D(Q,T) = \sum_m \sum_n d_{mn}(Q,T) \qquad (7)$$

where:

$$d_{mn} = \sqrt{(\mu^Q_{mn} - \mu^T_{mn})^2 + (\sigma^Q_{mn} - \sigma^T_{mn})^2} \qquad (8)$$

The fig. 2, illustrates the various steps of the search process. The extraction of image characteristics is always carried out by the Gabor Filter Model applied to the query image.

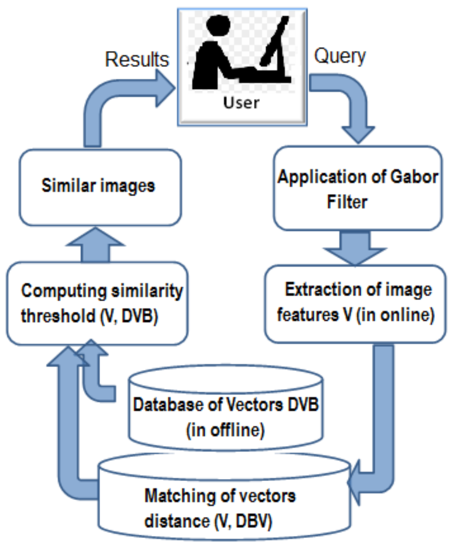

Fig. 2. Principal steps of the Information Retreival (IR) process by image content: application of the Gabor Filter Model.

4 CORPUS D'ANALYSE ET RÉSULTATS.

Dans notre première phase de test, nous avons réalisé une analyse sur 320 Mammographies. L'étude de performances basée sur l'estimation de la précision en fonction du rappel a donné les résultats obtenus sont illustrés dans la table 1 et la figure.3

4.1 Analysis Corpus and results.

In our first test phase, we performed an analysis on 320 mammograms. The performance study based on the estimation of precisions according to the recall. the results obtained are illustrated in table 1 and fig. 3.

TABLE 1
AVERAGE REFERENCE VALUES OBTAINED IN THE FIRST TEST PHASE

Main precision and recall values obtained	
Recall	Precision
33 %	75 %
50 %	64 %
66 %	53 %

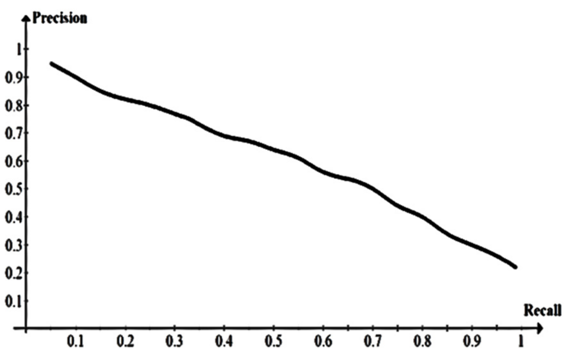

Fig. 3. Representation of the average values of the precisions according to the recall of the 1st test phase

In the second test phase, we performed an analysis on a corpus of 1000 medical images. The results obtained are illustrated in Table. 2 and Fig 4.

From results we note that the transition from our 1st corpus of images (320 mammograms) to the 2nd corpus (1000 medical images) has relatively affected the average values of the precision obtained. If we take for example the three reference values given in table 2, we see that the values of the average precisions have dropped a little compared to those obtained in table 1. However, we also find that tripling the corpus of images (going from 320 images to 1000 images) has not weakened the performance of our system too much. This encourages us to process a larger corpus or even the transition to processing Big Data.

TABLE 2
AVERAGE REFERENCE VALUES OBTAINED IN THE SECOND TEST PHASE

Main precision and recall values obtained	
Recall	Precision
33 %	70 %
50 %	59 %
66 %	42 %

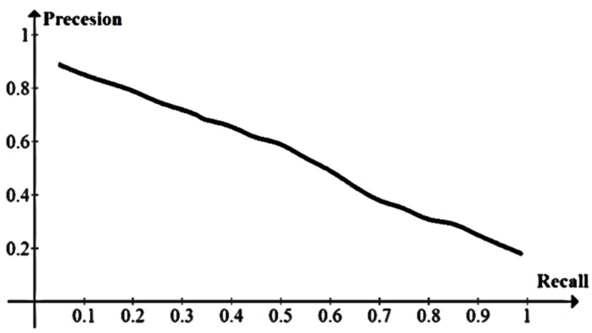

Fig. 4. Representation of the average values of the precisions according to the recall of the 2nd test phase

5 CONCLUSION AND PERSPECTIVES

In this work, we were able to apply the Gabor filter to different corpora of images (on the number side as well as the homogeneity side) the performances obtained are not perfect at this level. On the other hand, the observations made allow us to take a correct path in the treatment of Bigdata. Moreover, since the low-level analysis of an image by the Gabor filter is carried out pixel by pixel, the processing time is relatively considerable for large image corpora. Even if major processing is done offline, the transition to BigData can, a priori, impose a more demanding computing time! to overcome this requirement, we are thinking of carrying out the processing on supercomputers (clusters) where each node of the cluster is responsible for processing, in parallel with the other nodes, a well-defined part of the global corpus. We then envisage carrying out this task in other work to come.

REFERENCES

[1] Bhavanishankar K., Sudhamani M.V. (2020) Classification of Lung Nodules into Benign or Malignant and Development of a CBIR System for Lung CT Scans. In: Smys S., Tavares J., Balas V., Iliyasu A. (eds) Computational Vision and Bio-Inspired Computing. ICCVBIC 2019. Advances in Intelligent Systems and Computing, vol 1108. Springer, Cham. https://doi.org/10.1007/978-3-030-37218-7_64. Janv. 2020

[2] Xie, Y., Xia, Y., Zhang, J., Song, Y., Feng, D., Fulham, M., Cai, W.: Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT. IEEE Trans. Med. Imaging **38**, 991–1004 (2018)

[3] Pham, T. D., Watanabe, Y., Higuchi, M., &Suzuki, H. (2017). Texture Analysis and Synthesis of Malignant and Benign Mediastinal Lymph Nodes in Patients with Lung Cancer on Computed Tomography. Scientific Reports, 7, 1-10

[4] Baboo, S.S., Iyyapparaj E.: A classification and analysis of pulmonary nodules in CT images using random forest. In: 2018 2nd International Conference on Inventive Systems and Control (ICISC), pp. 1226–1232. IEEE (2018)

[5] Lu, X., & Chen, D. (2017). Cancer Classification through Filtering Progressive Transductive Support Vector Machine Based on Gene Expression Data. AIP Conference Proceedings, 1864 (1), 1-5.

[6] Identification and transfer of spatial transcriptomics signatures for cancer diagnosis (2020) 22: 6. https://doi.org/10.1186/s13058-019-1242-9. Janv. 2020.

[7] A similarity measure method fusing deep feature for mammogram retrieval (2019). Journal of X-Ray Science and Technology, vol. Pre-press, no. Pre-press, pp. 1-17, 2019. DOI: 10.3233/XST-190575

[8] Bertrand Thirion, «Introduction aux approches BigData pour l'imagerie médicale», Collège de France, 2 mai 2018: https://www.college-de-france.fr/site/gerard-berry/symposium-2018-05-02-09h35.htm

[9] Gérard Berry, «Imagerie médicale et apprentissage automatique: vers une intelligence artificielle?»: https://www.college-de-france.fr/site/gerard-berry/symposium-2017-2018.htm

[10] Isabelle Thomassin-Naggara et Olivier Clatz, «Technologies innovantes au chevet du dépistage du cancer du sein : quel futur pour les femmes ?», Collège de France, 2 mai 2018: https://www.college-de-france.fr/site/gerard-berry/symposium-2018-05-02-15h15.htm

First N. BOURKACHE is a Teaching researcher at the Mouloud Mammeri Tizi-Ouzou University & LAMPA Lab., Algeria.

Second S. Sidhom Assistant Professor Lorraine University & LORIA Lab. (Kiwi Team)
https://members.loria.fr/SSidhom/
https://www.researchgate.net/profile/Sahbi_Sidhom

Third M. LAGHROUCHE Profesor at the Mouloud Mammeri Tizi-Ouzou University & LAMPA Lab., Algeria.