Secure content based image retrieval in medical databases

Reda Bellafqira, Gouenou Coatrieux, Dalel Bouslimi, Gwénolé Quellec

Secure content based image retrieval in medical databases

Reda Bellafqira[1*], Gouenou Coatrieux[1], Dalel Bouslimi[1], Gwénolé Quellec[1]

[1] Institut Mines- Telecom; Telecom Bretagne; Unite INSERM 1101 Latim, Brest, France.
[*] Corresponding author, reda.bellafqira@telecom-bretagne.eu, Tel: (+33)610992391.

Abstract *- **In this paper, we propose an implementation in the encrypted domain of a content based image retrieval (CBIR) method. It allows a physician to retrieve the most similar images to a query image in an outsourced database while preserving data confidentiality. Image retrieval is based on image signatures we build in the hormomorphically encrypted wavelet transform domain. Experimental results show it is possible to achieve retrieval performance as good as if images were processed non-encrypted***

Index Terms *- **E-health, Image Processing, Medical Informatics.***

I. INTRODUCTION

Nowadays, medical systems produce a huge amount of images ; images which are more and more stored in an outsourced way. In order to take advantage of these data, several image content based image retrieval (CBIR) methods have been proposed so as to provide diagnosis aid, for instance [1]. In [2], Quellec et $al.$ propose a CBIR method which allows a server to retrieve the most similar images to a query image submitted by a physician. This solution is based on the comparison of image signatures ; the signature of an image being built from the distributions of its wavelet coefficients.
In this paper, we propose to implement this method in the encrypted domain that is to say carrying out a search that ensures the image content confidentiality. More clearly ; the images the server stores and receives as query are encrypted with no capability to decrypt them. The proposed solution exploits homomorphic encryption [3], the main interest of which is that it allows us to perform operations ($e.g.$," +"," ×") onto encrypted data with the guarantee the decrypted result equals the one carried out onto unencrypted data.
The rest of this paper is organized as follows. We give a brief overview of the CBIR method proposed by Quellec et $al.$ and present its implementation with the homomorphic encryption algorithm of Paillier in section I. Performance of the proposed implementation are given in section II.

II. SECURE CBIR SCHEME

II.1. CBIR based on wavelet coefficient histograms.

The CBIR method proposed in [2] compares images through the L^1-distance between wavelet-based image signatures. The signature of an image corresponds to the coefficient histograms of the four subbands LL, HL, LH and HH of the first decomposition level of its discrete wavelet transform (see Figure 1). Implementing this method

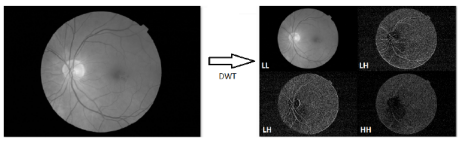

FIGURE 1 – : First DWT decomposition level of an image

in the encrypted domain imposes thus to : 1) perform the DWT onto the encrypted image, 2) compute the wavelet subbands' histograms. The solution we propose takes advantage of the homomophic encryption as we explain in what follows.

II.2. Secure CBIR based on wavelet coefficient histograms

In this work, we opted for the Paillier cryptosystem, an asymmetric homomorphic algorithm, so as to cipher the query image and the images stored in the server database. This cryptosystem has the probabilistic property. That is, the encryption of the same plaintext message yields, in general, to different ciphertexts. The Paillier cryptosystem [4] has also the additive homomorphic property : if we consider two plaintext a and b, we have :

$$E[a] \times E[b] = E(a+b) \quad (1)$$

where $E[.]$ represents the Paillier encryption function. In the sequel, we describe the implementation of DWT and the computation of the wavelet-based image signature in the homomorphic-encrypted domain.

- DWT in the encrypted domain :
  This one is performed as in [5] where the low and high pass filters of a separable DWT are as follow

$$E[A_j(k)] = \prod_{l\in\mathbb{Z}} E[A_{j-1}(l)]^{H_d(2k-l)} \quad (2)$$

$$E[D_j(k)] = \prod_{l\in\mathbb{Z}} E[A_{j-1}(l)]^{G_d(2k-l)} \quad (3)$$

  where $H_d(k)$, $G_d(k)$ and j represent the low-pass and high-pass decomposition filter coefficients, and the input signal decomposition level, respectively. $A_j(k)$ and $D_j(k)$ are the approximation and detail coefficients. $A_0(l)$ is equal to the input signal. By applying (2) and (3) on the lines and on the columns of the encrypted image, one obtains the encrypted version of the image wavelet transform. Notice that in our implementation, we used 2D Haar wavelet transform due to its simplicity.
- Wavelet-based image signature in the encrypted domain :
  As explained in section II.1, building the image signature in the encrypted domain implies being able to build

the histograms of the coefficients of the encrypted wavelet subbands. To do so, we refer to the method proposed by Hsu et *al.* [6], which allows identifying the interval to which an encrypted coefficient belongs without knowing its original value. More clearly, if $\{T_1, \ldots, T_k\}$ is a set of intervals and φ a wavelet coefficient, this method allows us to determine the interval Tu to which φ belongs from $E[\varphi]$ and $\{E[T_1], \ldots, E[T_k]\}$ as follow :

$$u = arg \min_i (E[\varphi]E[T_i]^{-1}) \quad (4)$$

Using this strategy, the histogram of each image subband can be approximately built from its encrypted version.

- Proposed secure CBIR scenario :
  Here, we consider that the images stored in database are encrypted using the Paillier cryptosystem parameterized with the physician public key. To carry out a search in this database, the physician firstly generates the set of intervals $\{T_1, \ldots, T_k\}$ dividing $\mathbb{Z}_N$ into k non overlapping intervals. $\{T_1, \ldots, T_k\}$ and the query image are then encrypted using the physician public key and then sent to the server. This latter calculates the signatures of the query image and the images stored in the database, without, however, decrypting these images. Their confidentiality is thus ensured. in order to compare images, the server applies the DWT on the encrypted image so as to obtain the encrypted versions of the image subbands LL, HL, LH and HH and next builds their respective histograms based on the encrypted intervals sent by the physician. The L^1 distance in-between signatures is then used to determine the $d - most$ similar images that will be sent back to the physician.

III. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed secure CBIR scheme has been evaluated on a medical image database for diabetic retinopathy screening. This dataset contains 400 images of patient eyes (ten photographs of 40 distinct subjects) of $2240 * 1488$ pixels associated with their diagnosis [2]. In this experiment, we consider that the intervals $\{T_1, \ldots, T_k\}$ are of same size Δ and that the server returns the five most similar images to a query image, i.e. $d = 5$. To evaluate the performance of our scheme, we measure the mean precision, which corresponds to the rate of returned images that have the same pathology as the query image. We give in Figure 2 the variation of this mean for different values of Δ. It can be seen that it decreases when Δ increases. But, at the same time, deceasing Δ implies increasing the number of intervals k, and consequently the computing complexity of our scheme. Compared to the performance of [2] with the same parameterization, obtained mean precision values are equivalent.

IV. DISCUSSION- CONCLUSION

In this paper, we propose a secure implementation of a CBIR method which allows carrying out a search into an encrypted image database. It takes advantage of homomorphic encryption that makes possible to calculate the DWT

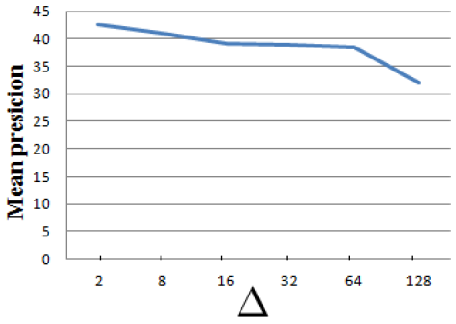

FIGURE 2 – : First DWT decomposition level of an image

of an image and build the histograms of its different subbands. Experimental results show that working on encrypted data does not influence the mean precision of the adopted CBIR method. Obviously, our implementation is slower than working with clear images but it guarantees the confidentiality and the privacy of processed images. Future works will focus on reducing the complexity of our scheme.

REFERENCES

[1] M. Rahman et *al.*, *"A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback"*, *IEEE Transactions on Information Technology in Biomedicine*, January 2007.

[2] G.Quellec et *al.*, *"Wavelet optimization for content-based image retrieval in medical databases"*, *ELSEVIER Medical image Analysis*, April 2010, Volume 14, Issue 2, pp 227–241.

[3] Z. Erkin et *al.*, *"Protection and Retrieval of Encrypted-Multimedia Content : When Cryptography Meets Signal Processing"*, *Eurasip J. Inform. Security*, 2007.

[4] P.Paillier. *"Public-key cryptosystems based on composite degree residuosity classes"*, *in Proc. Adv. Cryptology*, 1999, pp. 223–238.

[5] P.Zheng et *al.*, *"Discrete wavelet transform and data expansion reduction in homomorphic encrypted domain"*, *IEEE Trans Image Process*, 2013 Jun ;22(6) :2455-68.

[6] Hsu CY et *al.*, *"Image feature extraction in encrypted domain with privacy-preserving SIFT"*, *IEEE Trans Image Process.* , 2012 Nov ;21(11) : 4593-607.

[7] Segun Aina et *al.*, *"Spontaneous expression classification in the encrypted domain" 9th IMA International Conference on Mathematics in Signal Processing*, Fri, 14 Mar 2014.

[8] Julien Bringer et *al.*, *"Privacy-Preserving Biometric Identification Using Secure Multiparty Computation"*, *IEEE Signal Processing Magazine*, 2013.