International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:3, No:7, 2009

Extraction of Semantic Digital Signatures from MRI Photos for Image-Identification Purposes

Marios Poulos, and George Bokos

***Abstract*—**This paper makes an attempt to solve the problem of searching and retrieving of similar MRI photos via Internet services using morphological features which are sourced via the original image. This study is aiming to be considered as an additional tool of searching and retrieve methods. Until now the main way of the searching mechanism is based on the syntactic way using keywords. The technique it proposes aims to serve the new requirements of libraries. One of these is the development of computational tools for the control and preservation of the intellectual property of digital objects, and especially of digital images. For this purpose, this paper proposes the use of a serial number extracted by using a previously tested semantic properties method. This method, with its center being the multi-layers of a set of arithmetic points, assures the following two properties: the uniqueness of the final extracted number and the semantic dependence of this number on the image used as the method's input. The major advantage of this method is that it can control the authentication of a published image or its partial modification to a reliable degree. Also, it acquires the better of the known Hash functions that the digital signature schemes use and produces alphanumeric strings for cases of authentication checking, and the degree of similarity between an unknown image and an original image.

***Keywords*—**Computational Geometry, MRI photos, Image processing, pattern Recognition.

I. Introduction

THIS paper makes an attempt to solve the problem of searching and retrieving of similar MRI photos via Internet services using morphological features which are sourced via the original image. Until now the main way of the searching mechanism is based on the syntactic way using keywords. This mechanism creates many problems in the researchers or the doctors because many medical cases of photos while have the same diagnosis however, the morphological depiction of them are not similar necessary. Taking this into account an additional searching tool in this study is proposed in which this problem could be satisfied. In the implementation of this aim an attempt is being made to propose a mechanism for computing the degree of similarity between images when other mechanisms such as the hash function of the digital signature schemes fail to do so. This mechanism is based on algorithm which has been applied for identification purposes in general photos. Thus, in this way a suitable transformation of this algorithm was attempted in order to this to be applied in the medical photos. Thus, the propose method is based on a scenario of management for the control of authentication or for the detection of a degree of violation of medical photos which could be used as an additional searching tool for a doctor or a researcher in order to retrieve medical images using the morphological image features. For this implementation, a number of MRI medical photos of different thematic subjects were used and these were compared with three different thematic MRI photos in order to the ability of the proposed algorithm to be investigated.

M. Poulos is with the Laboratory of Information Technology, Ionian University, Corfu, Greece (phone:+302661087429; fax:+302661087433; e-mail: mpoulos@ionio.gr).
G. Bokos is with the Laboratory of Information Technology, Ionian University, Corfu, Greece (e-mail: gbokos@ionio.gr).

II. Methodology

This method is divided into four stages: the Pre-processing stage-1, the Pre-processing stage-2, the Processing stage, and the Meta-processing stage. Pre-processing stage-1 [1]. The input image is made suitable for further processing by image enhancement techniques using Matlab 6.5, and is transformed into a matrix of pixels. Specifically, in this stage any image which is available from any of the known image formats (tiff, bmp, jpg, etc), is transformed into a matrix (a two-dimensional array) of pixels. Consider, for example, the matrix of pixel values of the aforementioned array. The brightness of each point is proportional to the value of its pixel. This gives the synthesized image of a bright square on a dark background. This value is often derived from the output of an A/D converter [2]. The matrix of pixels, i.e. the image, will be described as N x L m-bit pixels, where m controls the number of brightness values. Using m bits gives a range of 2m values, ranging from 0 to 2m –. 1. Thus, the digital image may be denoted as the following compact matrix form:

$$f(x,y)=\begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,L-1) \\ f(1,0) & f(1,1) & \dots & f(1,L-1) \\ \vdots & \vdots & \vdots & \vdots \\ f(N-1,0) & f(N-1,1) & \dots & f(N-1,L-1) \end{bmatrix} \quad (1)$$

The coordinate vector of the above matrix is:

$$\mathbf{P}=\left[\, f(x,y)\right] \quad (2)$$

Thus, a vector $\mathbf{P}$ of 1 x N x L dimension is constructed, which is then used in the next stage.

Pre-processing stage-2. This matrix is submitted on Fast Fourier Transform. The data which comes from this step is submitted to specific segmentation (data sets) using computational geometric algorithms implemented via Matlab. Thus, onion layers (convex polygons) are created from these data sets (see Fig. 1).

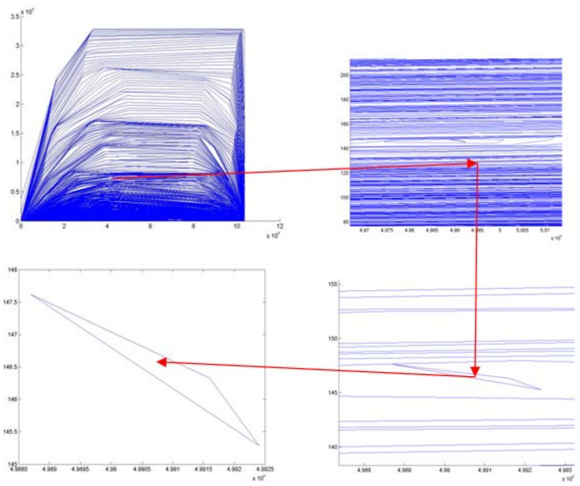

Fig. 1 Instances of Onion layers of a real image

We considered that the number of the elements of the vector $\mathbf{P}$ is K or K=NxL then the spectral density can now be calculated from the coefficients of vector $\mathbf{P}$ $\{P_k, k = 0, 1,\ldots, K-1\}$ via the Fourier Transform:

$$R_f = \sum_{k=0}^{K-1} S_k e^{-j\frac{2\pi f k}{K}} \quad f = 0, 1,\ldots, K-1 \qquad (3).$$

Finally, the magnitude or spectrum of the Fourier transform is calculated. Then, the vector S0 (equation 4) constitutes the initial point set.

$$S0 = |R_f| \qquad (4)$$

Processing stage of CGA (Computational Geometry Algorithm)

Proposition

We considered that the set of brightness values for each image contains a convex subset which has a specific position in relation to the original set. This position may be determined by using a combination of computational geometric algorithms, known as the Onion Peeling Algorithm, with the overall complexity of O(d*n log n) times.

Implementation

We consider the set of brightness values of an image to be the vector $\mathbf{S}$ (eq.4). The algorithm starts with a finite set of points $\mathbf{S} = \mathbf{S_0}$ in the plane, and the following iterative process is considered. Let $\mathbf{S}_1$ be the set $S_0 - \partial H \quad (S_0) : S$ minus all the points on the boundary of the hull of $\mathbf{S}$. Similarly, define $S_{i+1} = S_i - \partial H \quad (S_i)$. The process continues until the set is ≥ 3. The hulls $H_i = \partial H \quad (S_i)$ are called the layers of the set, and the process of peeling away the layers is called onion peeling, for obvious reasons [3]. Any point on H_i is said to have onion depth, or just depth i. Thus, the points on the hull of the original set have a depth of 0.

Extraction of the points of the convex polygon with semantic properties

The latest research has proven statistically that the smallest convex layer carries specific information, because this position gives a geometrical interpretation of the average of the image's brightness. In other words, the smallest convex polygon (layer) depicts a particular geometrical area in which this average ranges.

This feature may be characterized as unique [5, 6] to each image because the following two conditions are met:

- The selected area layer does not intersect with another layer.
- The particular depth of the smallest layer is variable in each case.

Thus, two variables are extracted from the proposed image-processing method: the area of the smallest onion layer $\mathbf{S_x}$ and the depth of this layer, which is a subset of the original image set S values, where: x is the number of amplitudes of the smallest layer. Taking into account the specific features of the aforementioned variables, it is easy to ascertain that these may be used for accurate image identification.

III. Identification Procedure

The known algorithm, Hausdorff Distance [7], was adopted for the procedure of comparing the convex polygons' points. The creation of this algorithm was done exclusively for the computation of such problems as the computation of the mean Euclidean distance of two convex polygons in Euclidean Space. In the present work the convex pairs of values (x, y) represent both the smallest onion layer that came from the original image and the smallest onion layer that came from a second image. From the computation of the distance between these two sets of points it emerges if, and to what degree, the second image is identical to the original one.

Specifically, the illustration of this algorithm is analyzed below:

Given two sets of points A={a1,a2,…,am} and B={b1,b2,…,bn} the Hausdorff Distance is defined as $\mathrm{H}(\mathrm{A},\mathrm{B}) = \max(\mathrm{h}(\mathrm{A},\mathrm{B}), \mathrm{h}(\mathrm{B},\mathrm{A}))$, where:

$h(A,B) = \max_{a \in A} \min_{b \in B} \|a-b\|$ and $\|a-b\|$ is any metric between the points a(x1, y1) and b(x2, y2), such as the Euclidian distance.

In this case, if $h(A,B) = 0$ and the number of pixels and all the points of the smallest polygon (Fig. 2) of the two images that are compared are the same, then the two images are

International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:3, No:7, 2009

identical. Otherwise, the degree of similarity between the two images is decided by the value of Hausdorff distance. A typical indicator of high similarity is decided empirically to be any value equal to or less than 0.1, $h(A,B) \leq 0.1$ and in all other cases, in which the Hausdorff distance is bigger than 0.1, the images are considered different.

IV. Empirical Observations

In this research, we selected 14 MRI different or similar medical images which were obtained from the Pictures of National Library of Medicine [8]. Specifically, three of these consisted of the basis of the comparison. The thematic areas of these were: lung Nodule, brain-transversal and brain-sagittal. The others eleven thematic MRI images were:

abdomen_mri
brain_mri_inversion_recovery
feet_mri
head_mri
lung Nodule with squamous cancer
mri_orbita
mri_orbita 2
mri_orbita_t2_fatsat
pelvis_mri
thighs_mri
thorax_mri

Specifically, there were three starting images. Sample 1.jpg, which is a Lung Nodule_mri image, Sample 2.jpg, which is a Brain-transversal, and Sample 3.jpg, which is a lBrain-sagital. For each of these three images, the researcher compare with 11 similar or different mri images. Finally, there were 14 images and their serial numbers.

Then, 11 x 3 = 33 pairs of images were created in order for the Hausdorff distance to be computed (see Fig. 2). From the results of these computations, which are presented in the table 1, it is clear that the Hausdorff distance is insignificant or zero for the images that came by modification from the original ones, and is about one or greater when the two images are different.

Having in mind the conditions of identification and similarity which were presented in the previous section, it emerges that the images that are absolutely identical are the ones that were compared with themselves and the ones that resulted from changing the file format. In each of these cases the Hausdorff distance is equal to zero. In the cases in which the comparison is about two images with minimal changes, the Hausdorff distance is much smaller than 0.1. Finally, in the cases in which the compared images are entirely different, then the Hausdorff distance is significantly larger than 0.1.

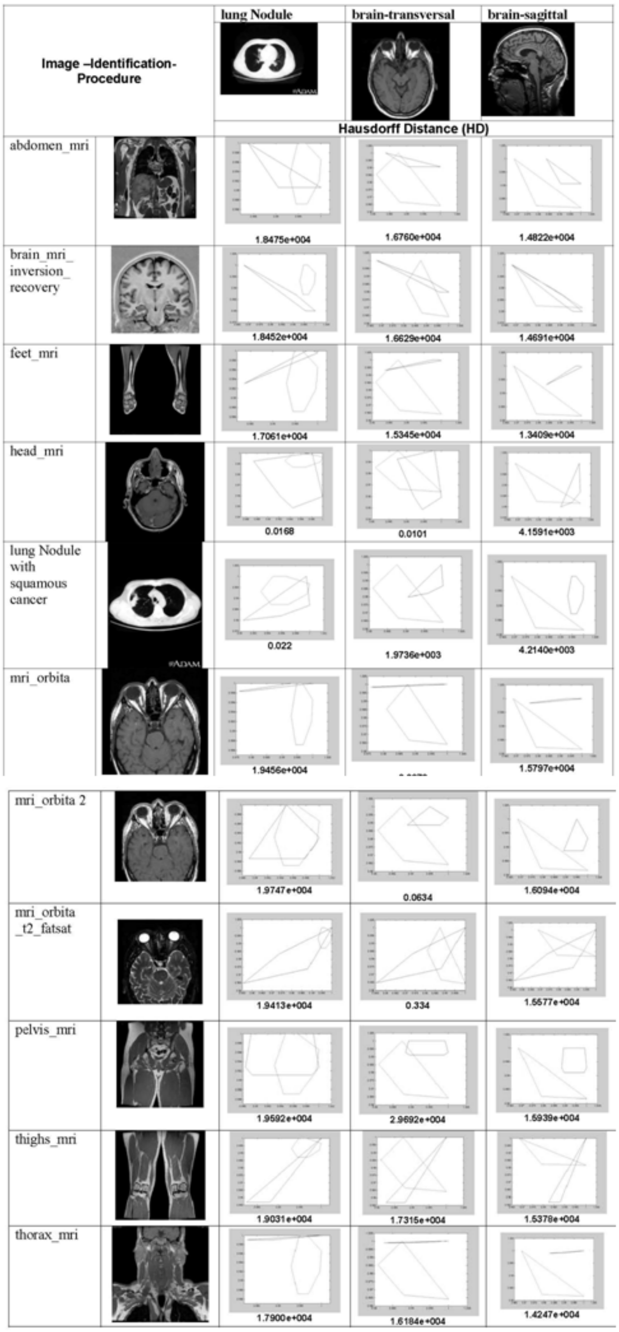

Fig. 2 The identification procedure between (14) images

V. Concluding Remarks

This study's major objective was to try to construct a serial number which, under certain circumstances, could be inserted into the information-management strategies of information organizations and big, co-operative medical libraries as a mechanism for the identification and control of the intellectual property of electronically published material. The existing techniques, and especially the digital signature schemes, could fulfill only the first, the identification, part of the objective.

This work presents the adjustment of a computational geometric algorithm for the semantic representation of the information of an image's pixel in a well-formed serial number. The idea for this construction came from the test of the onion-peeling algorithm in other areas of image-processing, such as the identification of humans by fingerprints. The aim of this application is to construct a serial number that could identify a copyright-protected published image even if its file format has changed from one type to another. Furthermore, it aims to provide a satisfactory amount of correlation similarity with other images created from the original by a small pixel alteration, and to detect and automatically reject images that are not related to the original.

For a realistic implementation of the proposed serial number, the researcher created a small database with three images transformed from jpeg format to tiff and bmp. Also, the researcher changed some pixels of the original image by the addition of small objects covering a tiny area. The changes were done to address a hypothetical scenario of image violation. In total, 14 images were created

This study adopted the Hausdorff distance, which has been applied in solving problems involving the identification of convexes in Euclidean space, for its preliminary identification decisions. The present work's experimental results show that in cases of absolute identification the Hausdorff distance was zero. The images of slight alteration of pixels, which would be suspected of violating a copyright, yielded Hausdorff distances of less than 0.1. Entirely different images produced Hausdorff distances much greater than 0.1.

Furthermore, this method may be characterized as accurate method which produces a very good degree of complexity regarding to other method. Furthermore, may be characterized as novel and helpfully because the using algorithms and the sophistication schema in which based on this technique ιt come in agreement with recent studies [9, 10].

Finally, as futures plan this method could be used as retrieved tool for the searching medical mri images for similar characteristics.

REFERENCES

[1] K. Dalal, Counting the onion. Random Struct. Algorithms 24(2): 155-165 (2004).

[2] M. Poulos, E. Magkos, V. Chrissikopoulos and N. Alexandris. "Secure Fingerprint Verification Based on Image Processing Segmentation using Computational Geometry Algorithms", Proceedings of the IASTED International Conference on Signal Processing, Pattern Recognition, and Applications, Rhodes Island, Greece, June 30- July 2, 2003, ACTA Press, 308-312.

[3] Lee, J.-W.; Khargonekar, P.P. " A convex optimization-based nonlinear filtering algorithm with applications to real-time sensing for patterned wafers" Automatic Control, IEEE Transactions on 48 (2), 224 – 235 (2003).

[4] http://doi.org/handbook_2000/metadata.html#4.3.3

[5] Shah B. and Raghavan V. and Dhatric P. and Zhao X. (2006). A Cluster-Based Approach for Efficient Content-Based Image Retrieval Using A Similarity-Preserving Space Transformation Method. Journal of the American Society for Information Science and Technology, 57, 1694--1707.

[6] Zachary J. and Iyengar SS and Barhen J. (2001). Content Based Image Retrieval and Information Theory: A General Approach. Journal of the American Society for Information Science and Technology, 52, 840--852.

[7] M. J. Atallah (1983). "A linear time algorithm for the Hausdorff distance between convex polygons". Information Processing Letters, v. 17, pp. 207-209.

[8] http://www.nlm.nih.gov/

[9] Marios Poulos, George Bokos, Fotios Vaioulis: Towards the semantic extraction of digital signatures for librarian image-identification purposes. JASIST 59(5): 708-718 (2008).

[10] Poulos M., Papavlasopoulos S. and Chrissikopoulos V. (2004). A text categorization technique based on a numerical conversion of a symbolic expression and an onion layers algorithm.Journal of Digital Information, Vol 6, Is. 1.

Marios Poulos is an assistant professor of informatics in the Department of Archives and Library Science at the Ionian University. He received his BSc from the University of Athens, Greece (1986) and his PhD from the University of Piraeus (2003). In the period from 2003 to 2008 he was a member of the staff (adjunct lecturer and assistant professor) in the Department of Archives and Library Science at the Ionian University. His research interests include medical informatics, semantic web and digital libraries. He is a member of several technical committees and working groups on subjects related to the medical informatics and information science areas.

George Bokos is a professor of library sciences in the Department of Archives and Library Science at the Ionian University. He was head of the Cataloguing & Bibliographic Services Department at the National Library of Greece from 1988 to 1995, the director of the National Library of Greece from 1996 to 1997 and a member of several technical committees and working groups, both at the national and international levels and on subjects related mainly to the automated handling and management of bibliographical information. Finally, he has been consulting with libraries and other information services for many years on library automation and the automated management of information.