



Registration and retrieval of highly elastic bodies using contextual information [☆]

J. Amores ^{*}, P. Radeva

Computer Vision Center, Dept. Informàtica, UAB, Bellaterra, Spain

Received 18 December 2004; received in revised form 18 December 2004

Available online 14 April 2005

Communicated by E. Backer

Abstract

In medical imaging, comparing and retrieving objects is non-trivial because of the high variability in shape and appearance. Such variety leads to poor performance of retrieval algorithms only based on local or global descriptors (shape, color, texture). In this article, we propose a context-based framework for medical image retrieval on the grounds of a global object context based on the mutual positions of local descriptors. This characterization is incorporated into a fast non-rigid registration process to provide invariance against elastic transformations. We apply our method to a complex domain of images—retrieval of intravascular ultrasound images according to vessel morphology. Final results are very encouraging.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Retrieval; Contextual information; Registration; Elastic matching; Medical imaging; IVUS

1. Introduction

In the last decade, the imaging technology has achieved a strong progress in developing novel

and powerful systems for image acquisition, processing and storage. Users are exploiting the opportunity to access and retrieve images in a large and varied collection. Medical imaging represents a real content-based image retrieval (CBIR) application of large image repositories where image retrieval can be of high importance in order to help image diagnosis and therapy. Images constitute a big part of the clinical data; case-based reasoning (recovering similar pathological cases) is one of the usual diagnostic procedures; moreover,

[☆] Work supported by Ministerio de Ciencia y Tecnología of Spain, grant TIC2000-1635-C04-04.

^{*} Corresponding author. Fax: +34 935811670.

E-mail address: jaume@cvc.uab.es (J. Amores).

constructing digital atlases based on automatic retrieval of images with the same clinical interpretation is very useful as a didactic program.

Medical image retrieval is still an emerging field. Due to the fact that human organs can vary in their position and appearance, CBIR systems based on global descriptors have poor performance. Some authors suggest using more elaborated approaches recovering similar patient cases based on contextual information (Shyu et al., 1999; Hou et al., 1992; Petrakis and Faloutsos, 1997; Tagare et al., 1995). The most common descriptor is the attributed relational graph (ARG) that deals with objects and their spatial relations (Petrakis and Faloutsos, 1997). The common characteristic of the reported approaches is that they rely on very well segmented regions (often segmented by hand) (Shyu et al., 1999; Hou et al., 1992; Petrakis and Faloutsos, 1997; Tagare et al., 1995). Looking for context-based CBIR that avoids the need of precise object segmentation, we can mention the general CBIR approach presented by Huang et al. (1997) that uses a contextual descriptor called color correlogram. This correlogram takes into account the spatial relations of local properties such as color of pixels. However, it is not conceived for considering relations between structures (such as pathological regions) of medical images. Another example of context-based general CBIR is presented by Belongie et al. (2002) that use correlograms for shape matching but on binary images.

Our goal is to develop a CBIR on medical images where our objects are very elastic bodies, i.e. can significantly vary in position, shape and appearance. In particular, our system analyzes intravascular ultrasound (IVUS) images that represent cross-sectional views of the artery showing normal and diseased tissues (plaques) on the wall of the vase (Europe, 1998). In order to represent the IVUS morphology, local descriptors are necessary to deal with small pathological regions. Information about the spatial arrangement of pathological regions should also be considered due to its importance in diagnosis. This can be achieved by constructing contextual object (vessel) descriptors. Our CBIR integrates all these characteristics: local, global and contextual information,

and the image comparison is invariant against elastic transformations. These characteristics are also important in general image retrieval (Smeulders et al., 2000). Regarding transformation invariance in medical imaging, only works focused on non-rigid registration (without considering retrieval) deal with smooth and elastic transformations where traditional elastic matching methods are computationally very expensive (Bajcsy and Kovacic, 1989; Gee, 1999; Christensen et al., 1996). Due to this fact, works in medical image retrieval only use fast registration methods that do not force smooth or elastic alignments (Dahmen et al., 2000; Robinson et al., 1996; Liu et al., 2001).

In this paper we present a novel CBIR system for highly elastic bodies that has two main contributions. First, the object description is constructed in an optimal feature space that represents contextual information needing just a weak object segmentation. In contrast to other context-based approaches (Huang et al., 1997; Belongie et al., 2002), we generalize the correlogram descriptor to deal with mutual positions of different structures or parts of the object, hence (1) it incorporates all the types of information mentioned above (global, local and contextual); (2) it does not need accurate segmentations/classifications of the structures inside the image; and (3) it is flexible: it incorporates specific descriptors of the application domain, which is mandatory in medical images where general descriptors perform poorly. The second contribution consists of the fact that our approach provides invariance to highly elastic transformations without allowing changes in topology of the object and using a computationally efficient registration. Our fast registration approach combines the use of contextual information, thin-plate splines (TPS) applied on a sparse set of landmarks, and a feedback scheme that achieves an elastic and regular, smooth transformation. Using the generalized correlogram into the registration enforces matchings between parts with the same context and removes ambiguities in the possible matchings, leading to faster convergence.

The paper is organized as follows: Section 2 describes the feature space, in Section 3 the registration algorithm is analyzed in detail, in Section 4 we explain the final distance used in image

comparison, in Section 5 results are provided over the different components of the system and in Section 6 we conclude and discuss future lines.

2. Feature space

In order to achieve that the feature space takes into account all types of information relevant to retrieve the image, we include local, global and contextual information by using generalized correlograms.

2.1. Local information

By using local information we aim at describing the different types of structures inside the image. In the IVUS case, the discriminating structures are placed around the wall of the vessel (Europe, 1998). A snake is placed at the center of the image after applying an anisotropic diffusion, and it is attracted to this wall. The set of landmarks is then obtained by sampling this snake (see Fig. 1). Associated with each landmark, a local feature vector is computed that describes the type of structure where the landmark lies. In this way we include specific information about our domain, as these local feature vectors are chosen for characterizing the biological structures we deal with (Dy et al., 2003). For discriminating plaques and normal tis-

sue a good descriptor is the gray-level profile along the normal to the wall at the landmark (Nair et al., 2002) (see Fig. 1). Empirically we chose a profile of 32 pixels, i.e. a feature vector of 32 dimensions. These feature vectors are classified and labels are assigned to each landmark, giving more compact local information. For doing so, non-parametric discriminant analysis (NDA) (Bressan and Vitria, 2003) and then K -nearest neighbors is applied. NDA reduces the dimension from 32 to 10, a number chosen empirically. A set of labelled descriptors is necessary for K -NN computation. For generating them, a group of physicians segmented and labelled each biological tissue in the images. Based on this, we took landmarks located at these structures, for each landmark extracted a feature vector and applied the corresponding label.

2.2. Global and contextual information

We incorporate the local information into a generalization of correlograms that allows to provide this information along with contextual and global information about the image. Correlograms are histograms which not only measure statistics about the features of the image, but also take into account the spatial distribution of these features. We show here that using a generalization of correlograms we can provide spatial relations between

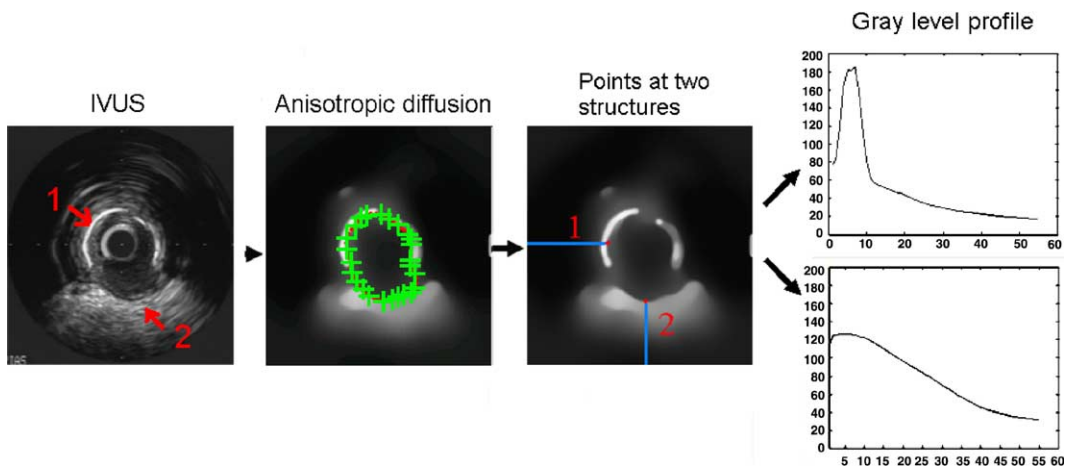


Fig. 1. From left to right: IVUS, its anisotropic diffusion where landmarks are extracted, and gray-level profiles.

structures. Let $C = \{p_i\}_{i=1}^n$, $p_i \in \mathbb{R}^2$ be a set of n landmarks. Let l_i be the label of the type of structure where the landmark p_i lies. Associated with p_i we compute a generalized correlogram h_i as follows. For every other landmark p_j we consider its label l_j and the spatial relation between p_i and p_j expressed in polar coordinates: $(p_i - p_j) = (\alpha_{ij}, r_{ij})$. We gather the label and spatial relation into one triplet $(\alpha_{ij}, r_{ij}, l_j)$. Based on the triplets of the $n - 1$ landmarks p_j , $j \neq i$, the correlogram h_i measures the joint distribution of local and spatial properties.

This distribution is calculated by a histogram that is based on a quantization of the resulting space with dimensions: angle, radius and label. Each dimension is partitioned separately. Let A_u , $u = 1, \dots, n_u$ be the bins for the angles and let R_v , $v = 1, \dots, n_v$ be the bins for the radius. The third dimension (the label) already represents a partition into n_c clusters obtained by the classifier described previously for local descriptors. The correlogram h_i is a histogram of dimension $n_u \times n_v \times n_c$ expressed as:

$$\begin{aligned}
 h_i(u, v, c) &= \#\{(\alpha_{i,j}, r_{ij}, l_j) : \alpha_{ij} \in A_u, r_{ij} \in R_v, l_j = c, j \neq i\}, \\
 u &= 1, \dots, n_u; \quad v = 1, \dots, n_v; \quad c = 1, \dots, n_c
 \end{aligned}
 \tag{1}$$

In this work we use the same log-polar spatial quantization for our correlogram as [Belongie et al. \(2002\)](#) for their shape context, which makes the correlogram more sensitive to local context. [Fig. 2\(a\)](#) shows the spatial bins used in the correlogram associated with a landmark of an IVUS image. The landmarks in this image have been classified into two types of structures: those

belonging to calcium plaque (squares), and those belonging to adventitia (circles). This correlogram has 12 intervals of angles and five intervals of radius. [Fig. 2\(b\)](#) shows a log-polar representation of the correlogram for each of the values in the third dimension: type of structure $c = 1$ and $c = 2$. In this plot, bins with a high density of landmarks from a particular type of structure are represented by a high gray level. This figure illustrates how the generalized correlogram measures the density over relative positions of the different types of structures, being a contextual descriptor.

This correlogram has scale and orientation invariance by normalizing the radius r_{ij} by the size of our object and orientating the correlogram along the tangent of the contour ([Belongie et al., 2002](#)). The main disadvantage of the spatial quantization is that the resulting correlograms are not robust against large shape changes of the object. This low robustness is accused before registering the images, however it can be avoided by using an appropriate feedback scheme in the registration, such as the one we will explain later. Correlograms are special types of histograms so that an appropriate distance used in this work is the χ^2 ([Duda et al., 2001](#)).

3. Registration: obtaining invariance against elastic transformations

We obtain invariance against elastic deformations by registering the images before their comparison. The scheme followed in the registration is the so-called point-mapping. First, a set of landmarks is extracted from each image. The landmarks are described in some feature space (in our case by correlograms), and a set of correspondences is computed which globally minimize the distance between matching landmarks in this feature space. Finally, a transformation is obtained based on the correspondences. Our registration also includes a search strategy of the final transformation.

Let I_1 and I_2 be two images so that I_1 is to be matched against I_2 . Let $X = \{p_i\}_{i=1}^n$ be the set of landmarks from I_1 , $Y = \{q_i\}_{i=1}^n$ the set from I_2 . Let $d^F(p_i, q_j)$ be the distance in the feature space

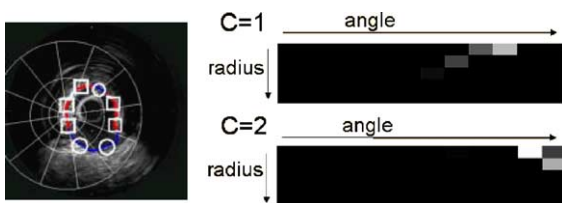


Fig. 2. Correlogram: spatial bins (a) and log-polar representation (b).

between landmarks p_i and q_j . We seek the correspondence function $\phi: \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ that minimizes $\sum_{i=1}^n d^F(p_i, q_{\phi(i)})$. This can be obtained by using an assignment optimization algorithm such as the Hungarian's method (Papadimitriou and Stieglitz, 1982). This method regards the assignment problem as a bipartite graph matching problem: we have nodes in the first set to be linked with nodes in the second, each link has an associated weight expressing the cost of matching the two nodes, and we want to obtain the set of links (matches) that minimizes the total cost and that results in a bijective mapping from the first set to the second one. In our problem, nodes are represented by landmarks and weights are represented by distances $d^F(p_i, q_j)$ between landmarks. If we have n landmarks in each image, the cost of the algorithm has order $O(n^3)$. This cost is manageable for a small set of landmarks representing the image.

The set of correspondences ϕ is not necessarily regular because geometric restrictions are not considered in the Hungarian method (in the results this fact is illustrated clearly). A regular spatial transformation T_λ is derived based on ϕ . Given $X = \{p_i\}$ and the corresponding set $Y_\phi = \{q_{\phi(i)}\}$, the transformation $T_\lambda: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ maps X close to Y_ϕ and varies smoothly in the rest of the plane \mathbb{R}^2 . We use the thin-plate spline (TPS) as an efficient elastic transformation T_λ that involves just inverting a matrix of $n \times n$, where n is the number of landmarks. The TPS transformation T_λ is obtained by minimizing $\sum_i \|T_\lambda(p_i) - q_{\phi(i)}\| + J_\lambda$, where the first term forces approximation to $q_{\phi(i)}$, the second term forces smoothness and λ represents a tradeoff between both terms. A high λ provides a smooth mapping but a more coarse approximation. J_λ is based on the second derivatives of T_λ (Bookstein, 1989) and represents the energy of the transformation.

The regular transformation T_λ is used for obtaining a more regular next set of correspondences ϕ using an iterative method. Initially, ϕ is obtained based on the distances $d^F(p_i, q_j)$. Because ϕ is not necessarily regular, we apply T_λ with high λ (regularity term). T_λ does not approach ϕ accurately (i.e. T_λ maps p_i far from $q_{\phi(i)}$). We use T_λ only to incorporate regularity into a new ϕ . This is done by recomputing the distance between landmarks

$d(p_i, q_j)$ and deriving a new ϕ by the Hungarian method. We use the following updating formula for the distance: $d_{k+1}(p_i, q_j) = d^F(p_i, q_j) + \alpha \|T_\lambda^k(p_i) - q_j\|$, where $d_{k+1}(p_i, q_j)$ is the distance in iteration $k+1$ and $T_\lambda^k(p_i)$ is a regular mapping for p_i obtained in iteration k . The term $\|T_\lambda^k(p_i) - q_j\|$ is the spatial distance from the potential correspondence q_j to the regular mapping $T_\lambda^k(p_i)$, and thus represents the amount of irregularity introduced by matching q_j with p_i . The parameter α represents the tradeoff between the two terms: similarity in the feature space and regularity. With this updating formula, the next ϕ is more regular and the regularity parameter λ can be decreased in the next iteration (so that a subsequent T_λ is more accurate). As T_λ becomes more accurate, we force the next ϕ to be close to T_λ by increasing α . In this way, λ is decreased and α is increased through the iterations, and we use an exponential ratio of change typical of annealing schemes. By this feedback, the neighbors of p_i influence in its correspondence through the term $T_\lambda^k(p_i)$ because the TPS considers the spatial dependencies between p_i and the neighbors. Therefore we are performing a (fast) relaxation-like cooperative search.

4. Similarity measure in the final comparison between images

The registration produces a transformation T which is regular and maps the characteristic points p_i from I_1 close to their corresponding ones in I_2 . However, the mapped points are not exactly the characteristic points q_i of I_2 . In order to obtain a regular final set of correspondences ϕ from $\{p_i\}_{i=1}^n$ to $\{q_i\}_{i=1}^n$, we simply take the Euclidean distances of mapped points and destination points: $d(p_i, q_j) = \|T(p_i) - q_j\|$ and compute the correspondences using the Hungarian's algorithm over this matrix of distances.

Our similarity measure is based on the sum of three factors: the distance in the feature space, the amount of deformation necessary to align both objects by TPS (Bookstein, 1989), and a local appearance difference between the aligned image and the destination image.

The distance in the feature space is in our case the distance of the correlograms. We recompute these correlograms orienting them now along the x axis of the image. This is done to reduce the little robustness that these correlograms have on the computation of the tangents, now that the alignment permits to avoid the invariance to rotation. Let $d^F(p_i, q_j)$ be the χ^2 distance between correlograms of p_i and q_j . Let us denote by bold \mathbf{d} the global distance between images, we take as distance $\mathbf{d}^F(I_1, I_2)$ the bidirectional Chamfer distance in the feature space:

$$\mathbf{d}^F(I_1, I_2) = \frac{1}{n} \sum_{i=1}^n \min_j d^F(p_i, q_j) + \frac{1}{n} \sum_{j=1}^n \min_i d^F(p_i, q_j)$$

For computing the local appearance difference between both images, let I_W be the image I_1 warped according to the obtained transformation from the registration. We take local windows around the mapped points $T(p_i)$ in I_W and matching points $q_{\phi(i)}$ in I_2 . The local appearance difference is expressed:

$$\mathbf{d}^A(I_1, I_2) = \sum_{i=1}^n \sum_{x=-w}^w \sum_{y=-w}^w G(\|(x, y)\|) \times [I_W(T(p_i) + (x, y)) - I_2(q_{\phi(i)} + (x, y))]^2$$

where $G(r)$ is a gaussian-like function of the radius r , more sensitive to close positions. The warped image I_W does not respect the original pattern of the textures, so it is better to remove them in the comparison. Therefore, we take as images I_1 and I_2 the anisotropic diffusion of the original images.

Finally, the total distance between both images is computed as a combination of the distance components defined above: $\mathbf{d}(I_1, I_2) = \alpha^F \mathbf{d}^F(I_1, I_2) + \alpha^A \mathbf{d}^A(I_1, I_2) + \alpha^E \mathbf{E}$. The weights α^F , α^A , α^E are computed as the ones minimizing the classification error on the IVUS database, following a leave-one-out procedure. Values obtained for these weights can be found in Section 5.5.

5. Results

In this section we see the results of applying each of the components of the registration

algorithm: feature space, feedback scheme, final registration results, and final retrieval results. All the experiments have been conducted on a database of 100 IVUS images, all of them presenting calcium plaque structures. Studying the registration of images presenting calcium plaque is very interesting for the following reasons: first, there is a great difficulty in the differentiation between plaques and adventitia tissue, second, there is a high variability in the shapes of both the entire vessel and the calcium plaque structures, and third it has been clinically seen that the relative spatial position of the calcium is important in diagnosis of heart diseases.

5.1. Performance of the correlograms

Our correlograms work with labels of classified landmarks (see Section 2.1). For landmark classification we use a K -nearest neighbor with $K = 7$, a parameter obtained experimentally. To exploit the complete dataset of 100 IVUS, a procedure similar to leave-one-out is used: descriptors from a new image are classified based on the descriptors from the other 99 images. As we have 100 landmarks per image, each image has 100 descriptors (one per landmark). Therefore the set of 100 descriptors from a new image is classified using 9900 descriptors from the rest of images. The classification hit rate for landmarks is 90.1%.

Fig. 3(a) and (b) shows a couple of IVUS images to be registered, Fig. 3(a) displays the image I_1 to be aligned and Fig. 3(b) the destination image I_2 . Fig. 3(c) and (d) show the anisotropic diffusion of (a) and (b) respectively. The thick curve represents the contour of the vessel from which the landmarks are extracted in each image. The image I_1 has two calcium plaques on both sides (indicated in Fig. 3(a)), and the image I_2 has three calcium plaques: two on both sides and one at the bottom (indicated in Fig. 3(b)). Taking into account global characteristics the plaques on both sides should be matched in both images, leaving alone the small plaque at the bottom of the image I_2 . We show that the global description is included into our correlograms by comparing the result of an initial coarse transformation using contextual information (correlograms) and then

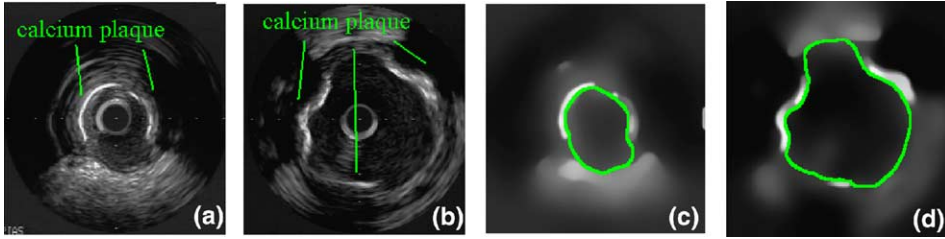


Fig. 3. Couple of images to register (a,b). Their anisotropic diffusions (c,d).

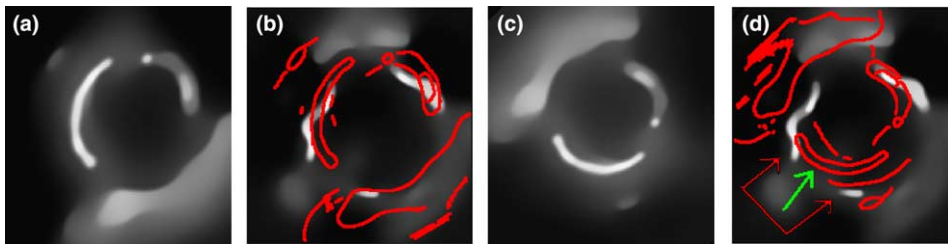


Fig. 4. Intermediate alignment using contextual information (a,b), and using only local information (c,d).

using only local information (our local feature vectors). We show transformation results on the anisotropic diffusion of the images because it is visually more clear. Fig. 4(a) shows the warped I_1 when using correlograms. Fig. 4(b) shows the edges of the warped I_1 superposed with thick lines onto the image I_2 . Note that both calcium plaques of I_1 are mapped close to the big plaques of I_2 . Fig. 4(c) and (d) show the warping result when using only local feature vectors. One of the mapped calcium plaques (indicated by a thick arrow in Fig. 4(d)) lies at an intermediate position between a big plaque and a small one (thin arrows). Using correlograms this matching is avoided as the size characteristic is included.

5.2. Evaluation of the feedback scheme

The result of applying the feedback scheme is that the transformation becomes more and more accurate and at the same time the set of correspondences becomes more and more regular. The iterative feedback algorithm includes two parameters: α , and λ , that change with exponential rate (see Section 3). Experimentally we chose the following

values: α changes from 0 to 0.25 and λ changes from 1500 to 0.1. The number of iterations we used in the experiments is 8.

Fig. 5(a) shows the initial set of correspondences in the registration of the pair of images of Figs. 3 and 5(b) shows the final set of correspondences. The initial set is very irregular, but holds information about the correct global matching. The final set is completely regular. Fig. 5(c) shows the final transformation: the contours of the aligned image are superposed with thick lines on the destination image. Note that plaques (indicated by arrows) are completely aligned.

Now we provide a quantitative evaluation of the feedback scheme. Given one query image and one target image that belongs to the same category as the query, we want the query to be closely aligned to the target. Here we quantify how this alignment evolves in the iterations of the feedback scheme. We randomly take 100 pairs so that in every pair both images are from the same category. Let (I_1, I_2) be one such a pair, I_1 represents a possible query and I_2 an image from the same category as that of the query. Given a transformation T that aligns I_1 to I_2 , we measure the goodness

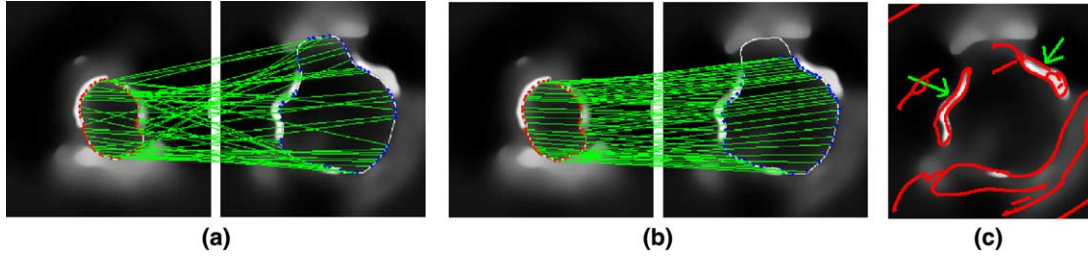


Fig. 5. Evolution of the feedback scheme.

of this transformation by the spatial distance between the structures of the aligned $T(I_1)$ and the homologous structures of I_2 . The concept homologous structure is based on whether two structures are from the same type of structure (e.g. both are plaque structures). The resulting measure represents the amount of misalignment of T . We also take another measure that considers the amount of irregularity of T which we explain below. Both measures are based on the transformation T obtained after each iteration: let T_i be this transformation after iteration i . We can then observe how the goodness of the alignment is increased throughout the iterations of the algorithm.

Let us express the amount of misalignment more concretely. Let C_1 be the set of landmarks of I_1 , and C_2 the set of I_2 . Given the transformation T_i , the misalignment for the type of structure s is expressed:

$$E^s(I_1, I_2, T_i) = \max(E_1^s, E_2^s)$$

$$E_1^s = \frac{1}{n_s^1} \sum_{p \in C_1, l(p)=s} \min_{q \in C_2, l(q)=s} (\|T_i(p) - q\|) \quad (2)$$

$$E_2^s = \frac{1}{n_s^2} \sum_{q \in C_2, l(q)=s} \min_{p \in C_1, l(p)=s} (\|T_i(p) - q\|)$$

where n_s^1 is the number of landmarks in I_1 located at the type of structure s , n_s^2 is the respective thing for landmarks of I_2 , and we express as $l(p) = s$ the fact that landmark p has label s .

Given Eq. (2), the total misalignment of I_1 to I_2 given T_i is the average misalignment considering all the types of structures in both images:

$$E(I_1, I_2, T_i) = \frac{1}{n_c} \sum_{s=1}^{n_c} E^s(I_1, I_2, T_i) \quad (3)$$

where n_c is the number of types of structure present in both I_1 and I_2 . For evaluating Eq. (2) we use a priori knowledge of what is the correct type of structure of every landmark in both images.

The amount of irregularity obtained for the alignment T_i is computed by the deformation energy of the thin-plate spline (see Section 3).

For each couple (I_1, I_2) the algorithm iterates eight times for this experiment, resulting in eight alignments T_i , $i = 1, \dots, 8$. For each iteration the median is computed across results on the 100 couples, Fig. 6 shows a graphic of the median evolution of the amount of misalignment (Fig. 6(a)) and amount of irregularity (Fig. 6(b)). The horizontal axis represent the number of the iteration $i = 1, \dots, 8$, and the vertical axis represent the misalignment in (a) or irregularity in (b). As can be seen the misalignment and the irregularity are decreased throughout the iterations, due to the simultaneous maximization in our feedback scheme of both the accuracy and regularity (see Section 3).

Finally, we see that the shape-context registration algorithm in (Belongie et al., 2002) without any cooperative feedback scheme leads to poor results when dealing with the types of images we have. We take the same couple displayed in Fig. 3, and only use the landmarks from one type of structure (calcium plaque in this case) for registration. We do so because the shape context in (Belongie et al., 2002) is not suitable for different types of structures. Fig. 7(a) and (b) show the disposition of the mentioned landmarks in both images, note the shape differences. Fig. 7(c) shows the final set of correspondences. The landmarks are not mapped close to their destination, and

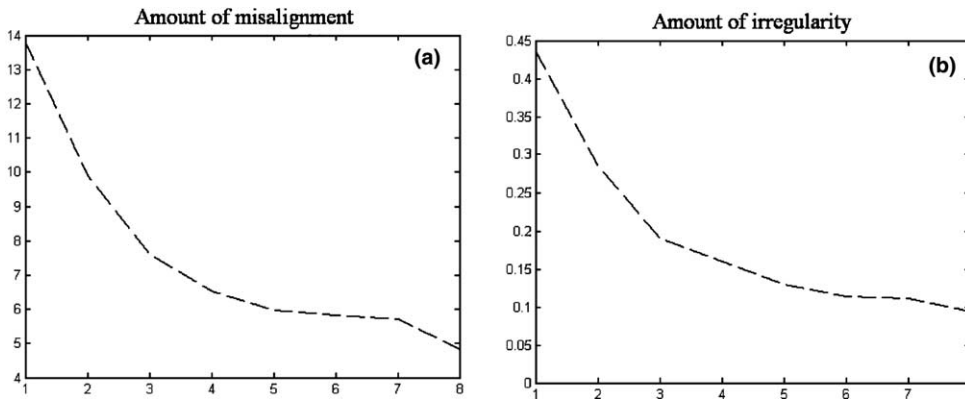


Fig. 6. Evolution of the misalignment (a) and irregularity (b) of the transformation.

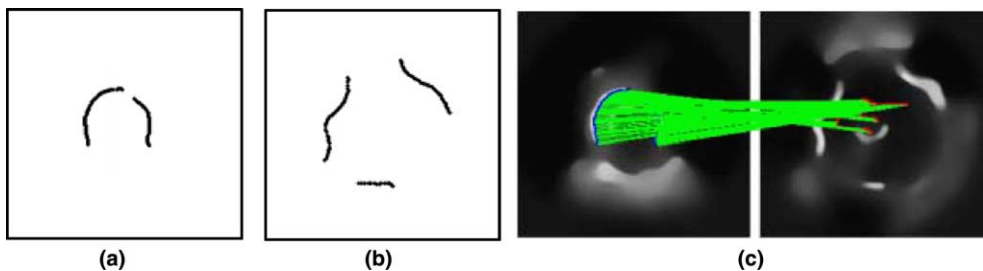


Fig. 7. (a,b) Points from the plaque structures of Fig. 3(a) and (b). (c) Final set of correspondences obtained with the shape-context registration algorithm (Belongie et al., 2002).

the final correspondences are quite irregular. The main cause is that, with the high shape difference in both objects we need to enforce step by step some regularity in the correspondences or else the resulting transformation will map the points without preserving the spatial coherence. Correlograms continue to be valid for modelling contextual and global information (as shown with our result in Fig. 5), but only if we strengthen the spatial coherence of the mapping by some feedback algorithm such as the one explained above.

5.3. Registration results

We provide here the final registration results (i.e. results after the last iteration). From our database of 100 IVUS images, there are 100×100 pairs of images, from which 1646 pairs are formed by homologous images (i.e. both images in the pair belong to the same category) not including pairs that contain the same image, e.g. (I_i, I_i) . As homol-

ogous images should be aligned completely by the registration step, we take these 1646 couples for computing the statistics of the registration accuracy. We obtained a mean amount of misalignment of 4.6 pixels, a median misalignment of 2.04 pixels and a standard deviation of 7.6 pixels. The mean distance between two neighbor landmarks is of 3.1 pixels. Thus, the misalignment is just 1.5 times the distance between landmarks that are neighbors. Experimentally we have seen that a misalignment below 6 pixels is acceptable by the physicians. 75% of the registrations have a misalignment below 4.23 pixels.

5.4. Computational cost: scaling the system

The system has been implemented in Matlab[®] code. In a Pentium IV at 2.4 GHz, the average time for all the steps prior to registration is 14 s. The bottleneck of the system is the registration done for every image in the database. The average

registration time per image is 12 s, so that the total spent time per 100 images in the database adds up to 20 min in average. This is not a high cost compared to the time spent with a physics-based registration that uses iterative PDEs that lasts several hours. Scaling the system to large number of images requires choosing prototypes representing all the images in the database. In this way, each category in the database is represented by a set of prototypes, so that each prototype represents the appearance of a set of images in the category and is thus a subcategory representant. Given a distance like the one we propose, an automatic algorithm for prototype-selection is given in (Belongie et al., 2002).

5.5. Retrieval results

In order to assess the retrieval efficiency, a database of 100 IVUS images has been used. For this database, a group of physicians have grouped the images into categories attending to clinical properties (see Amores et al., 2003). We extract each image from the database, present it as query, and the system computes the distance from this query to every other image in the database. The distance is computed as explained in Section 4 on the 100 images using a leave-one-out procedure. We obtained weights α^F , α^A , α^E that respectively have values 0.4%, 0.25%, 0.35% in average, therefore stress is put on the distance of correlograms.

Based on a particular query, the system orders the rest of the images from the database in order of similarity to this query. From this ordered list, we take only the first K images (i.e. the K most similar images to our query). Two measures of retrieval efficiency are used. The first one is the estimated number of images we need to retrieve in order to include an image from the same category as the query. We obtained an average of 2.33 images necessary for including one of the same category. For $K=2$ retrieved images the mean number of times in which an image from the same category is included is 89.7%.

The second measure is the recall vs scope (Huang et al., 1997): if query Q has N images from the same category, we compute for

$$Q : E^Q(K) = \frac{\#\{I : \text{rank}(I) \leq K, \text{category}(I) = \text{category}(Q)\}}{N}$$

and average over all the queries presented: $E(K) = \frac{1}{N_Q} \sum_Q E^Q(K)$, where N_Q is the number of query images presented to the database. Our system is compared with other approaches using contextual information. Retrieval with shape contexts in (Belongie et al., 2002) is only suitable for binary images. Huang et al. developed a correlogram that considers for every pair of colors the co-occurrence of pixels with a given distance between them. They do not use the angle for the spatial relations, do not extract landmarks but use all the pixels, and use as local information pixel-level properties such as color (thus we cannot use the type of structure as local information). They report a better performance in color retrieval than other contextual descriptors such as color coherent vectors (Huang et al., 1997). In Table 1 the $E(K)$ efficiency is shown for Huang's correlogram (computed on gray level for IVUS), for Huang's auto-correlogram, and for our context-based retrieval. Our method outperforms both types of contextual descriptors. Huang's correlogram is included for completion, although this descriptor has proved to be slightly worse than the auto-correlogram (indeed, Huang et al. do not report results for the former descriptor). Although Huang's approach showed a good performance in a non-medical color database, it is poor when dealing with IVUS images, and a more sophisticated method such as the one presented is necessary for such a domain. For $K=10$, we obtained 22% recall (see table) and 36.5% of precision, typical values for other non-medical CBIR applications. For $K=30$ we obtained as recall-precision 50–27.7% and for $K=50$ we obtained 73–24%. Fig. 8(a) shows three

Table 1
Recall vs scope measure

K	Huang's correlograms	Huang's auto-correlograms	Present method
10	0.15	0.17	0.22
30	0.36	0.38	0.50
50	0.54	0.57	0.73

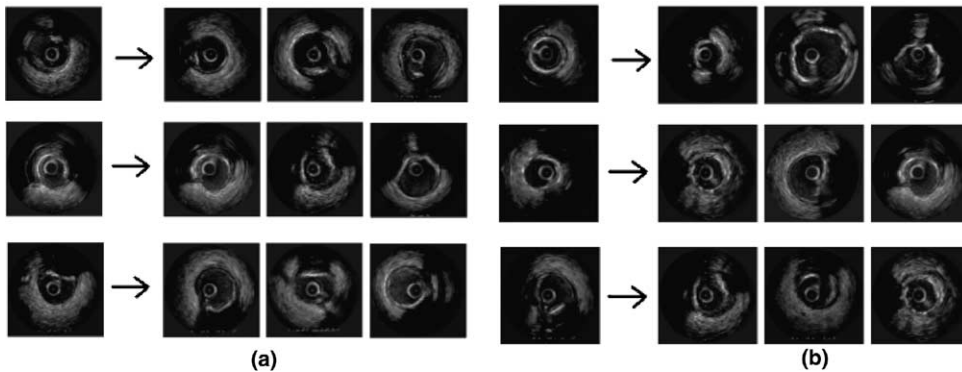


Fig. 8. Example of optimal retrieval result (a) and worst result (b) for $K = 3$.

queries with high performance. In these queries, the first $K = 3$ retrieved images are all of the same category. Fig. 8(b) shows three queries with low performance. Although in this example none of the retrieved images strictly belong to the same category as the query, the similarity of the retrieved images and the queries is quite high. For six of the nine retrieved images, the degree of embracement of the plaque in both the query and the retrieved image is the same. Most times the failure of the retrieval system is due to the interpretation of big false structures (in the local landmark classification). The correlogram fails if a big region of the image is entirely missclassified, as there is not valid information in that case. Still, the correlogram is robust when part (but not all) of the landmarks are missclassified. For example, if all the landmarks from a big region of the image are wrongly classified as adventitia structure, the correlogram fails because it interprets that there is adventitia structure in that region of the image. However, if part of the landmarks are wrongly classified as adventitia and a bigger part are correctly classified as plaque structure then the correlogram is robust because measures the density of landmarks belonging to plaque in that region of the image.

6. Conclusions

We have introduced a content-based retrieval system that deals with complex medical images

where contextual information is mandatory. We showed that using generalized correlograms we can incorporate this contextual information so that we do not need hand-based segmentation, which differs from previous work (Shyu et al., 1999; Hou et al., 1992; Petrakis and Faloutsos, 1997; Tagare et al., 1995). We reported quantitative results on every component of the system and showed that our system outperforms other contextual-based approaches such as the one by Huang et al. (1997). The important contributions are as follows.

First, we introduced a new definition of the correlogram that extends the descriptor to deal with the context of structures. Using this type of correlogram we avoid very accurate segmentations of the structures. Our correlogram can incorporate easily specific information about the medical domain, which is fundamental in medical image retrieval.

Second, we achieved invariance against elastic transformations. We used an efficient registration that achieves elastic and accurate alignments and at the same time is smooth, which differs from other medical retrieval systems (Dahmen et al., 2000; Robinson et al., 1996; Liu et al., 2001). Our method combines the use of thin-plate splines, efficient compared to solving a Navier–Stokes PDE (Bajcsy and Kovacic, 1989; Gee, 1999; Christensen et al., 1996), and a feedback scheme that enforces smoothness and accuracy.

As future work, faster indexing methods for correlograms are under investigation. An analysis

of the impact of choosing a small set of prototypes for large databases would be also of high interest.

References

- Amores, J., Radeva, P., Elastic matching retrieval in medical images using contextual information. Technical report, CVC, September 2003.
- Bajcsy, R., Kovacic, S., 1989. Multiresolution elastic matching. *Comput. Vision, Graphics Image Process.* 46 (1), 1–21.
- Belongie, S., Malik, J., Puzicha, J., 2002. Shape matching and object recognition using shape contexts. *IEEE Trans. PAMI* 24 (24), 509–522.
- Bookstein, F.L., 1989. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE TPAMI* 11 (6).
- Bressan, M., Vitria, J., 2003. Nonparametric discriminant analysis and nearest neighbor classification. *Pattern Recognition Lett.* 24 (15), 2743–2749.
- Christensen, G.E., Rabbitt, R.D., Miller, M.J., 1996. Deformable templates using large deformation kinematics. *IEEE Trans. Image Process.* 5 (10).
- Dahmen, J., Theiner, T., Keysers, D., Ney, H., Lehmann, T., Wein, B., Classification of radiographs in the image retrieval in medical applications system (irma), In: *Proceedings of 6th International RIAO Conference on Content-Based Multimedia Information Access*, Paris, 2000, pp. 551–566.
- Duda, R.O., Hart, P.E., Stock, D.G., 2001. *Pattern Classification*. John Wiley & Sons.
- Dy, J.G., Brodley, C.E., Kak, A., Broderick, L.S., Aisen, A.M., 2003. Unsupervised feature selection applied to content-based retrieval of lung images. *IEEE TPAMI* 25 (3).
- Europe, B.S. (Ed.), *Beyond Angiography. Intravascular Ultrasound: State of the Art*, vol. 1. XX Congress of the ESC, 1998.
- Gee, J.C., 1999. On matching brain volumes. *Pattern Recognition* 32, 9–111.
- Hou, T.-Y., Liu, P., Hsu, A., Chiu., M.-Y., 1992. Medical image retrieval by spatial features. *IEEE Internat. Conf. Syst. Man Cybernet.* 2, 1364–1369.
- Huang, J., Kumar, S., Mitra, M., Zhu, W., Zabih, R., Image indexing using color correlograms. In: *Proc. CVPR.*, 1997, pp. 762–768.
- Liu, Y., Dellaert, F., Rothfus, W.E., Moore, A., Schneider, J., Kanade, T., Classification-driven pathological neuroimage retrieval using statistical asymmetry measures. In: *Proc. MICCAI '01, Utrecht, The Netherlands*, 2001, pp. 655–665.
- Nair, A., Barry, D.K., Tuzcu, E.M., Schoenhagen, P., Nissen, S.E., Vince, D.G., 2002. Coronary plaque classification with intravascular ultrasound radiofrequency analysis. *Circulation* 22, 2200–2206.
- Papadimitriou, C., Stieglitz, K., 1982. *Combinatorial Optimization: Algorithms and Complexity*. Prentice Hall.
- Petrakis, E.G.M., Faloutsos, C., 1997. Similarity searching in medical image databases. *IEEE Trans. Knowledge and Data Engineering* 9 (3).
- Robinson, G.P., Tagare, H.D., Duncan, J.S., Jaffe, C.C., 1996. Medical image collection indexing: shape-based retrieval using kd-trees. *Comput. Med. Imaging Graphics* 20 (4), 209–217.
- Shyu, C.R., Brodley, C.E., Kak, A.C., Kosaka, A., Aisen, A.M., Broderick, L.S., 1999. Assert—a physician-in-the-loop content-based retrieval system for hrct image databases. *Comput. Vision Image Understand.* 75 (1–2), 111–132.
- Smeulders, A.W., Worring, M., Santini, S., Gupta, A., Jain, R., 2000. Content-based image retrieval at the end of the early years. *IEEE TPAMI* 22 (12), 1349–1380.
- Tagare, H.D., Vos, F.M., Jaffe, C.C., Duncan, J.S., 1995. Arrangement, a spatial relation between parts for evaluating similarity of tomographic section. *IEEE Trans. PAMI* 17 (9), 880–893.