Tree Representation and Implicit Tree Matching for a Coarse to Fine Image Matching Algorithm

Julian Mattes¹ and Jacques Demongeot^{1,2}

¹ TIMC-IMAG, Faculty of Medicine, 38 700 La Tronche, France
Julian.Mattes@imag.fr

² Chaire de Biomathématiques, Institut Universitaire de France
jacques.demongeot@mesr.fr

Abstract. In medical imaging arises the problem of matching two images of the same objects but after movements or slight deformations. We present a new method for a primitive global transformation and an improvement of a recent matching strategy which makes it more robust. The strategy consists of two steps. We consider the grey level function (modulo a normalization) as a probability density function. First, we apply a density based clustering method in order to obtain a tree which classifies the points on which the grey level function is defined. Secondly, we use the identification of the hierarchical representations of the two images to guide the image matching. The general transformation invariance properties of the representations permit to extract invariant image points. But in addition, we design a new robust coarse to fine identification of the trees which applies an implicit error measure in a prediction – correction scheme using thin plate splines to interpolate the transformation function in a finer way at each step. Therefore, we will find the correspondence between invariant points even if these have locally moved. The method’s results for matching and motion analysis on real images will be discussed.

1 Introduction

The use of structural information in images for recognition tasks has been realized since the inception of machine vision ([14]). A graph structure is useful to represent information about topology and shape of and between extracted features in images and therefore to compare two images, but also, as we will show, to match images in which *local* movements appear (e.g., rigid transformations of *individual* objects in the image or images of deformable objects). For this, we have to construct the graph in such a way that its structure is kept even if deformations appear or features have moved and to use tree identification to find the correspondence between these features.

A grey level function $g : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$, $n = 2$ or $n = 3$, associates to each physical point in the image scene (represented by $\mathbb{R}^n$) a value corresponding to its physical properties ($\in \mathbb{N}$ in the technical realization) and will be considered as given by the acquisition device. g is eventually smoothed and normalized, in order

to have a continuous (or even derivable) density function. The present approach applies density based classification methods ([16], [5]) to the grey level function (considered as density function) in order to construct trees. These methods confine points in clusters at several levels if the points are in regions of (relative to the level) high density. We will call the resulting tree (section 2) a *confinement* (or *density*) *tree representation of the image*. If we choose the classification method proposed in [16] and [5] the resulting tree corresponds to the component tree defined in [4].

Matching is done now by identification of the two density trees corresponding to the two images. The idea behind this tree representation matching strategy is that the density tree will adapt itself to global and local movements in the image (Figure 1). Under reasonable conditions, we will demonstrate that the proposed representation is invariant with respect to an arbitrary *topological transformation* ϕ (for its definition see the proof in section 2). This allows us to define *characteristic* points (*candidates* for landmarks [2]) and to find corresponding point (i.e., landmark) pairs, associated to nodes in the tree, even if local movements appear. Moreover it allows us to detect if in one of the two images a cluster – intermediary node or leaf in the tree – has been changed by noise effects (section 3). All this has a series of applications as detailed in section 4.

We proposed this approach recently in [11,13], where we defined a cost function for a given association of the two tree structures and we minimized among possible associations in order to obtain the “best” association between the trees. Independently, a matching method based on the same hierarchical decomposition of the image in regions has been proposed in [7,8]. The correspondence between the regions is established using a weighted combination of similarity criteria, such as the difference of regions’ areas, contour lengths, or barycenters.

The present contribution lies in a new identification algorithm of hierarchical decompositions of images. The initialization step (to rediscover the composition of global translation, rotation, and scaling) is a new robust method based on gravity centers of image regions corresponding to the nodes of the tree. After, the method goes from coarse to fine using a prediction – correction scheme and a *local rigid* transformation is applied before the *thin plate spline interpolation* [2] of the topological transformation ϕ is refined *at each time* in order to apply an implicit error measure [1] to ever finer levels. The *prediction – correction scheme* enables us to avoid the association of not corresponding regions and to stop the matching process before it becomes noise sensitive.

We can remark that Leu and Huang [9] used a comparison of ordered trees to define a distance between objects for object recognition. However, a tree identification method, able to handle instabilities in the tree, was not proposed and the algorithm’s suitability to match images where non-affine deformations appear was not investigated. Moreover, the tree construction approach does not use statistical methods, but is of geometrical nature based, in a segmented image, on a closed contour whereas simply edge detection is not sufficient.

A recent general overview of matching algorithms can be found in [10].

Fig. 1. *Illustration of the density based tree representation.*

The remainder of the paper is organized as follows. In section 2, we detail the construction of the confinement tree and its invariance property. In section 3, we show how to identify the hierarchical representations and we present some results on real images in section 4, before we conclude in section 5.

2 Construction and Invariance Properties of the Confinement Tree

Confronted with the classification task based on a density function we were led to a mathematical entity denoted *confiner*, appearing in a natural way when we consider level lines in a landscape. Given a density function $g : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$, $n \in \mathbb{N}$, the *confiners* are defined as the maximal connected subsets (i.e., *components*, Gaal [3], p. 105) C_l of the level sets $L_l = \{x \in \mathbb{R}^n | g(x) \geq l\}$, $l \in \mathbb{R}_{\geq 0}$ (see Figure 1). In the classification domain we found these components first in a contribution by Wishart [16] and further investigations by Hartigan [5], where they are called *high density clusters*. Considering them taken on several levels l_k , $k = 1, \dots, r$ (r from resolution) including the 0 level, they define obviously a tree (by “set inclusion”) as illustrated in Figure 1. The tree is finite if g has a finite number of extrema. Let us call the level sequence $(l_k)_{k=0}^r$ with $l_0 = 0$ a *resolution* (of $\mathbb{R}_{\geq 0}$). Then, we call the corresponding tree *confinement* (or *density*) *tree* at resolution $(l_k)_{k=0}^r$ and if g is a grey level function ($n = 2, 3$) of an image I_g we use the term *confinement tree representation of I_g* . The confiners taken at level l_k are the nodes (intermediary nodes and leaves) of the tree at level k . We have chosen this classification method because of the invariance theorem and its consequences shown below. Curiously, this property has already been mentioned in the contributions of Wishart [16] and Hartigan [5] (without assuming the paradigm to hold and only in the case of an affine transformation) even if they did not address a matching problem.

In practice, we consider all grey levels of the image and we delete confiners with a *mass* (calculated as the sum of all normalized grey values associated to the pixels in the confiner) less than 1% of the mass of all confiners at this level. We choose the discrete *d4-distance* for defining the connectivity between pixels.

Time to calculate the tree is $O(n + r)$, where n is the number of pixels and r is those of levels.

The following hypothesis is the foundation of the transformation invariance property of the investigated matching strategy [11,8,7,13]. We can find similar hypotheses in Optical Flow research [6].

Medical Imaging Paradigm *Let $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^n, n = 2$ or 3 be the transformation expressing movements of physical points in an image scene (rendered in the images A and B) which appear when passing from A to B , g_A the density associated to image A , g_B that associated to image B , and $X \in \mathbb{R}^n$. Then we can assume (apart from local noise effects):*

- (i) $g_A(X) = g_B(\phi(X))$ (strong version);
- (ii) $\varphi(g_A(X)) = g_B(\phi(X))$ for some strictly monotonous function φ .

Let us remark, that: (1) for the paradigm to hold, we should consider images covering the whole space in which movements appear (thus, 3D images in general) and that the paradigm is a realistic assumption in a larger context than in medical imaging but not in general; (2) if for two images A and B the detector (or detector location) has changed, we have to assume the general case; if we simplify and assume $g_A(X) = \alpha g_B(\phi(X)) + \beta$, we can calculate α and β *a priori*.

General Invariance Theorem *Let A, B, g_A, g_B, φ , and ϕ be defined as in the paradigm and let $(l_k)_{k=0}^r$ be a resolution of the grey level space. Then: if ϕ is a topological transformation, then, apart from noise effects, the confinement tree at resolution $(l_k)_{k=0}^r$ of image A and those at resolution $(\varphi(l_k))_{k=0}^r$ of image B are identical, moreover, the confiners of A are transformed into the confiners of B and reciprocally.*

Proof (i) The proof of the strong version (i.e., φ is the identity function) is based on the fact that (a) according to its definition (Gaal [3], pp. 186, 187) a *topological transformation* is a bijective and bicontinuous (i.e., ϕ and ϕ^{-1} are both continuous) application and (b) a component stays a component under a topological transformation ([3], Lemma IV.8.1). For details see [11].

(ii) For the general case, let us first suppose that ϕ is the identity function. We have $L_k^A := \{X | g_A(X) \geq l_k\} = \{X | \varphi(g_A(X)) \geq \varphi(l_k)\} = \{X | g_B(X) \geq \varphi(l_k)\} =: L_k^B$. The components of L_k^A are the nodes at level k in the confinement tree of A , the components of L_k^B are those of the B 's tree, equally at level k . As $L_k^A = L_k^B$, this defines an isomorphism between the trees.

(iii) If ϕ is general and $h := g_B \circ \phi$, (ii) implies that there exists an isomorphism between the trees associated to h and to A , and from (i) an isomorphism between the trees associated to $h - B$. Then, trees of A and B are isomorph.

The importance of the theorem lies on the fact that the tree is invariant referred to local movements. This makes it more stable for matching than other well known hierarchical representations would be. In the presented formulation, the theorem is more general than previous versions [11,8,7,13].

3 Implicit Tree Matching

Major innovations are developed in this section. Their motivation is to make the strategy more robust with respect to previous publications [11,8,7,13]. For this aim, the new algorithm follows the *coarse to fine principle of implicit tree matching* (below) and uses a *prediction – correction scheme*. Both innovations are based on a classical implicit error measure [1] defined below, which enables, together with the application of a local rigid transform at *each iteration step k* (step (iii) below), to make fully use of proceeding coarse to fine in the two hierarchical decompositions of the images. In order to use an implicit error measure we need the correspondence between tree structure and image regions; just the two (eventually weighted) trees are not sufficient. In this way “implicit tree matching” differs from tree matching algorithms as presented in [17].

3.1 Implicit Error Measure

We call an *error-measure* between objects on two different images A, B *implicit* or *transformation based* if it depends on a transformation $\phi_k : A \rightarrow B$ (at step k) between the images A and B . A well known implicit matching algorithm is *ICP* [1]. It uses the (implicit) error measure

$$\epsilon(C_A, C_B) = \frac{1}{|C_B|} \sum_{P_B \in C_B} d_{\min}^2(P_B, C_A), d_{\min}(P_B, C_A) = \min_{P_A \in C_A} \|P_A - P_B\|, \quad (1)$$

where $|C_B|$ is the number of points in C_B , and where C_A (resp. (C_B)) is a set of points in image A (resp. image B). C_A is the reference point set.

Coarse to Fine Principle of Implicit Tree Matching. (see Figure 2). Let be given a selection order for A confiners (cf. subsection 3.2) and a transformation function $\phi_k : A \rightarrow B$ obtained at iteration step $k - 1$. Iteration step k :

Substep (i) Find the corresponding confiner $\mathcal{C}_B^k$ in image B of the currently selected confiner $\mathcal{C}_A^k$ using the error measure $relPos_k(.,.)$ (see below).

Substep (ii) Match locally rigidly confiners $\mathcal{C}_A^k$ and $\mathcal{C}_B^k$ and so define points in B corresponding to sampled contour points of locally moved confiner $\mathcal{C}_A^k$.

Substep (iii) Calculate the transformation function ϕ_{k+1} as the thin-plate spline interpolation [2] of the corresponding point pairs.

Substep (iv) Select the next A confiner $\mathcal{C}_A^{k+1}$.

Let us **remark** that substep (iii) can be omitted; in this case we have to store the rigid transformation for confiners with not updated successors but we avoid parameters due to interpolation.

Euclidian Error Measure. Given a transformation $\phi_k : A \rightarrow B$ and the current confiner $\mathcal{C}_A$ in image A , we have to find the corresponding confiner $\mathcal{C}_B$ in B based on the transformed external contour $\phi_k(\mathcal{C}_A)$ of *reference confiner* $\mathcal{C}_A$ and on the contour C_B of $\mathcal{C}_B$. We choose the confiner $\mathcal{C}_{B_{\min}}$ minimizing the following error measure $relPos_k(.,.)$ based on $\epsilon(.,.)$ of equation (1):

$$relPos_k(\mathcal{C}_A, \mathcal{C}_B) = \epsilon(\phi_k(\mathcal{C}_A), C_B) \frac{\max(A(\mathcal{C}_A), A(\mathcal{C}_B))}{\min(A(\mathcal{C}_A), A(\mathcal{C}_B))}, \quad (2)$$

Fig. 2. *Coarse to Fine Principle of Implicit Tree Matching.* On the left side we are able to find the coarse association (between A1 and B1, A2 and B2); but we are not able to find the fine association because $\text{relPos}(A13, B12) < \text{relPos}(A13, B13)$. Therefore (right hand side), first, we match locally rigidly A1 with B1, we choose corresponding point pairs on the original and locally transformed contour of A1, we interpolate the transformation function and, finally, we transform A11, A12, and A13 using this function. After we repeat the procedure for A2, B2, A21, and A22. The A-confiners on the right have been translated and rotated around 25° with respect to the left hand side.

where $A(\mathcal{C})$ denotes the area of confiner $\mathcal{C}$.

In order to associate to ϕ_k an error value, let denote by $\mathcal{G}(A)$ (resp. $\mathcal{G}(B)$) the set of gravity centers of clusters associated to the nodes of the tree corresponding to A (resp. B). We define:

$$\epsilon_{euclid}(\phi_k) = \epsilon(\phi_k(\mathcal{G}(A)), \mathcal{G}(B)) + \epsilon(\mathcal{G}(B), \phi_k(\mathcal{G}(A))). \quad (3)$$

3.2 Matching Algorithm

Initialization. The initialization is based on the point sets $\mathcal{G}(A)$ and $\mathcal{G}(B)$ defined above. In practice, we take only the nodes appearing directly after bifurcations into account, in order to reduce search space. Our aim for initialization is to find the transformation ϕ_0 representing translation, rotation, and scaling (referred below as *semi-rigid* transformation) such that the error measure $\epsilon_{euclid}(\phi_0)$ is smallest. We proceed as proposed by Thirion for *extremal points* in [15]. Roughly spoken, for each given pair of points in $\mathcal{G}(A)$, we take a pair of points in $\mathcal{G}(B)$, we calculate the semi-rigid (or rigid) transformation S defined by the association of the two point pairs, we transform all points in A according to this transformation and determine the error $\epsilon_{euclid}(S)$. For a given pair of points in $\mathcal{G}(A)$ we can limit the number of possible corresponding points in $\mathcal{G}(B)$ using a preselection based on invariant features ([15]; in our case the features are, for instance, the normalized confiner mass, its area, etc.). Finally, we take the transformation associated to the two pairs of points minimizing the error.

It remains to apply the **coarse to fine principle** detailed hereafter.

Confiner Selection Order. First, we have to define in which order we select confiners in image A for which we look for corresponding confiners in B . We consider just confiners corresponding to a node directly after a bifurcation to update the interpolation function and we *update confiners* with the same father in order of increasing predicted error (see next step). We start with the confiners after the first bifurcation. After, we choose at each step the *bifurcating node* (cf. Figure 5) having the largest mass (section 2) among the bifurcating nodes with not updated children. The selected confiners are its children.

Predicted Identification. At iteration step k we choose for a given confiner $\mathcal{C}_A$ the confiner $\mathcal{C}_B$ minimizing $relPos_k(\mathcal{C}_A, \mathcal{C}_B)$ in equation (2).

Interpolation Point Pairs. We update $\phi_k \mapsto \phi_{k+1}^{temp}$ (i.e., the *corresponding points list* on which ϕ_k is based) at each time when we associated 2 confiners $\mathcal{C}_A$ and $\mathcal{C}_B$ according to the following interpolation point pair selection scheme illustrated in Figure 3. We match locally the confiners for which we detected a correspondence in order to obtain a semi-rigid transformation ϕ_{k+1}^{loc} . First, we remove that points from the corresponding point list that are in a tubular neighborhood (with thickness η) around the transformed confiner $\phi_{k+1}^{loc}(\mathcal{C}_A)$. After, we sample points on the contour of $\mathcal{C}_A$, apply ϕ_{k+1}^{loc} on them, put the so obtained point pairs in the corresponding point list and get ϕ_{k+1}^{temp} .

Fig. 3. *Interpolation Point Pair Selection Scheme.*

Prediction – Correction Scheme. In the previous prediction step we obtained the temporary transformation ϕ_{k+1}^{temp} . This enables us to define an *intrinsic criterion* to verify if the predicted association was correct:

If ($\epsilon_{euclid}(\phi_{k+1}^{temp}) \geq \epsilon_{euclid}(\phi_k)$ or another A confiner has been associated to $\mathcal{C}_B$) then { $\phi_{k+1} := \phi_k$; /* reject identification */ } else $\phi_{k+1} := \phi_{k+1}^{temp}$;

We *stop* identification in a whole *subtree*, if we reject identification for the subtree root and all its children.

4 Applications

The first part of our new method, concerning global alignment of two 2D or 3D images, has a wide range of applications. Remember for instance, that we use such an alignment for reconstructing images from their 2D slides when their orientation has been lost during acquisition. Applications of the second part of our method include the study of the relative motion of grains in microscopic

metallurgical images or of cells in microscopic biological images. On the other hand, we can also use our method to detect abnormalities in breast or brain images (with respect to reference images coming from the same patient at a former time or from an atlas), or to detect where in an image movements occur. In the first case this serves to detect a tumor and its location. In the second, the method helps us to apply optical flow or active contour techniques at the found place. A further interesting application of the proposed method is the comparison of two brain images resulting from functional MRI. Here occur small local deformations as well as global shifts. The aim is to locate positions where the brain activation (and therefore the intensity of the functional signal) changed.

Experimental Evaluation. Figure 4 and the middle row of Figure 5 illustrate the effectiveness of our (semi-)rigid matching algorithm initializing our coarse to fine algorithm. This second part is illustrated in the bottom row of Figure 5 which proves that very few confiners in a branch are sufficient for propagating the transformation allowing the correct association of confiners at higher level. All experiments have been performed on a 450 MHz Intel Pentium III with less than 1s for calculation of the confinement tree as for the initialization step using a tree where confiners with a mass of less than 2.5% (with respect to their level) are deleted. See Figure 4 and 5 for first assessment, a more detailed evaluation can be found in [12].

Fig. 4. *On the left: initial upper slide of a 256x256 brain image. The points represent the gravity centers of confiners corresponding to the nodes after bifurcations in the confinement tree. The line corresponds to the confiner covering the whole brain and is just displayed for visualization of the transformation. On the right: lower slide. The points and the line are the transformed points and line of the left picture. The crosses are the gravity centers of the right image. The transformation represents the composition of translation, rotation and scaling as found by the initialization step of our method described in section 3.2. We can notice that the global barycenter of the whole brain is perfectly preserved at a subpixel precision (0.7 pixel).*

Fig. 5. First column (top and middle): initial 512x480 image of cells (below) with its corresponding confinement tree. Second column (top and middle): same as in the first column for the moved image. Second row: it shows the result of the initialization step described in section 3.2. The white points in the picture on the left are the gravity centers of the confiners after the bifurcating nodes as described in section 3.2. These points and the lines are transformed and displayed on the right picture after transformation. The transformed points are still displayed as white points, the gravity centers of the right image are displayed as crosses. The lines marked by arrows are the external contours of the confiners of the nodes in the tree above marked with the same number. They are displayed also (as darker lines) in the bottom row together with the corresponding confiners in the moved image. The other lines in the middle row are just shown to better visualize the initial transformation. The bottom row is another illustration of our coarse to fine principle. After matching rigidly the confiners at lower level (picture on the right), we can easily find the correspondence of the confiners at higher level.

5 Conclusion

We showed the use of a tree representation of the grey level function known in statistics (*density* or *confinement tree*) to define a hierarchical decomposition of an image in regions. This decomposition is first used for a new robust method for the alignment of two 2D (trivially to extend to 3D) image slides and after for a new coarse to fine identification algorithm which identifies the two image decompositions. With respect to previous versions of this second part, the method (a) avoids external parameters using an *intrinsic* criterion to determine when it is impossible to identify two regions and (b) is robust as it does just require that correspondence between several regions is identified. Especially, it would be possible that *no* local maximum in the first image has a corresponding local maximum in the second. Both, the new method for finding global and that for finding local transformations between two images improve matching and motion analysis of non segmented grey level images.

References

1. Besl, P.J., McKay N.D.: A method for registration of 3-D shapes. IEEE Trans. PAMI **14** (1992) 239–256 647, 650
2. Bookstein, F.: Principal wraps: Thin-plate splines and the decomposition of deformations. IEEE Trans. PAMI **11** (1989) 567–585 647, 650
3. Gaal, S.A.: Point set topology. Academic Press, New York, (1964) 648, 649
4. Hanusse, P., Guillotaud, P.: Sémantique des images par analyse dendronique. In: AFCET, 8th RFIA, Vol. 2. Lyon (1992) 577–588 647
5. Hartigan, J.A.: Statistical theory in clustering. J. of Classification **2** (1985) 63–76 647, 648
6. Horn, B., Shunck, B.G.: Determining optical flow. Art. Intel. **17** (1981) 185–203 649
7. Kok-Wiles, S.L.: Comparing mammogram pairs for the detection of mammographic lesions. PhD thesis, Brasenose College, University of Oxford, May 1998, (1998) 647, 649, 650
8. Kok-Wiles, S.L., Brady, J.M., Highnam, R.: Comparing mammogram pairs for the detection of lesions. In: Karssemeijer, N. (ed.): 4th Int. Workshop of Digital Mammography, Nijmegen, Netherlands, June 1998. Kluwer, Amsterdam, (1998) 647, 649, 650
9. Leu, J.G., Huang, I.N.: Planar shape matching based on binary tree shape representation. Pattern Recognition **21** (1988) 607–622 647
10. Maintz, J.B.A., Viergever, M.A.: A survey of medical image registration. Medical Image Analysis **2** (1998) 1–36 647
11. Mattes, J., Demongeot, J.: Dynamic confinement, classification, and imaging. In: 22nd Annual Conference of the Gesellschaft für Klassifikation e.V., Dresden, Germany, March 1998. Studies in Classification, Data Analysis, and Knowledge Organization. Springer-Verlag, Berlin Heidelberg New York (1999) 205–214 647, 649, 650
12. Mattes, J., Demongeot, J.: Statistical invariants and structures for image matching and representation. IEEE Trans. Medical Imaging (submitted) 653

13. Mattes, J., Richard, M., Demongeot, J.: Tree representation for image matching and object recognition. In: G. Bertrand, M. Couprie, L. Perroton (Eds.): DGCI'99. LNCS 1568 (1999) 298-309 [647](#), [649](#), [650](#)
14. Pavlidis, T.: Analysis of set patterns. *Pattern Recognition* **1** (1968) [646](#)
15. Thirion, J.-P.: New feature points based on geometric invariants for 3D image registration. *International Journal of Computer Vision* **18** (1996) 121–137 [651](#)
16. Wishart, D.: Mode analysis: A generalization of the nearest neighbor which reduces chaining effects. In: Cole, A. J. (Ed.): *Numerical Taxonomy*. Academic Press, London (1969) 282–319 [647](#), [648](#)
17. Zhang, K., Statman, R., Shasha, D.: On the editing distance between unordered labeled trees. *Information Processing Letters* **42** (1992) 133–139 [650](#)