

SMRFI: Shape Matching via Registration of Vector-Valued Feature Images

Lisa Tang and Ghassan Hamarneh

Medical Image Analysis Lab, Computing Science, Simon Fraser University, Canada

{lisat, hamarneh}@cs.sfu.ca

Abstract

We perform shape matching by transforming the problem of establishing shape correspondences into an image registration problem. At each vertex on the shape, we calculate a shape feature and encode this feature as image intensity at appropriate positions in the image domain. Calculating multiple features at each vertex and encoding them into the image domain results in a vector-valued feature image. Establishing point correspondence between two shapes is thereafter treated as a registration problem of two vector-valued feature images. With this shape representation, various existing image registration strategies can now be easily applied. These include the use of a scale-space approach to diffuse the shape features, a coarse-to-fine registration scheme, and various deformable registration algorithms. As our validation shows, by representing shapes as vector-valued images, the overall method is robust against noise and occlusions. To this end, we have successfully established 2D point correspondences of shapes of corpora callosa, vertebrae, and brain ventricles.

1. Introduction

Establishing shape correspondences is important for statistical shape analysis and object retrieval applications. In medical imaging, it helps us to analyze shapes for understanding diseases, develop probabilistic models and atlases, and retrieve anatomical shapes in medical databases.

Most approaches frame shape matching as an optimization that maximizes a similarity measure computed between two shapes under proximity/neighborhood and spatial constraints. The similarity measure is often calculated based on shape features, *e.g.* curvature, convexity, shape context [8]. In encoding proximity constraints, Xie and Heng [23] established correspondences between segments before establishing point correspondences on the matched segments, while Myronenko et al. [15] imposed a motion coherence constraint over the deformation field from which correspondences are derived. More explicit regularization can be done with order-preserving assignment [19], or by minimiz-

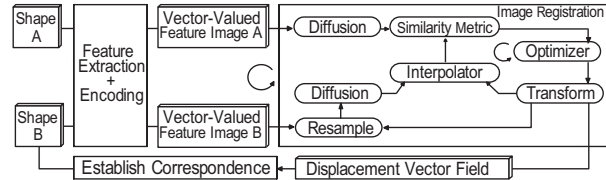


Figure 1. Pipeline of the proposed shape matching framework.

ing cost functions that either encourage the preservation of “local structure” [5] as shapes deform or penalize transformations that do not preserve distances between neighbouring points [25]. More recently, Kaick et al. [21] simultaneously enforced proximity and order-preservation constraints by formulating the problem as a quadratic assignment.

Interestingly, the seemingly complicated task of enforcing proximity constraints can be implicitly done in image registration. Given two images, the fundamental goal of registration is to find a spatial transformation T that maps pixels in one image to corresponding pixels of another image [7, 12]. The spatial transformation T is found by minimizing an energy function that measures the dissimilarity between one image and another on which T is applied. This function is often defined as a sum of an *external* (or image fidelity) term that encourages transformations that align pixels with similar intensity values and an *internal* (or regularization) term that measures the smoothness of T . The former is analogous to dissimilarity between matched shape descriptors, which are used to match similar vertices, while the latter is analogous to the proximity constraints, which are needed to ensure that neighbors map to neighbors.

Motivated by these similarities, we propose to treat shape correspondence as an image registration problem. Given shapes represented by points in 2D or 3D (*e.g.* contours or surfaces), we extract shape features and encode them at appropriate positions in the image domain. Extracting K sets of different types of features thus forms a vector-valued *feature image* of K components. Performing the registration of these feature images then allows us to derive point correspondences from the displacement vector fields generated from image registration. With appropriate choices of spatial transformations, neighborhood constraints can be effec-

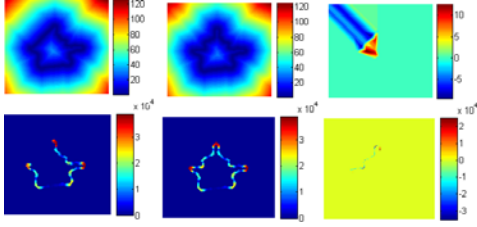


Figure 2. Occluded shape (left), original shape (middle), and the difference in their representations (right). Occlusions in shapes can significantly affect their DT representations (top row). Conversely, the impact of occlusions on the feature-based representation (bottom) is localized and not as significant.

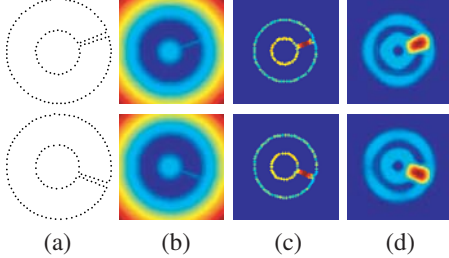


Figure 3. (a) Two shapes that differ by a rotation of 40° . (b) Their DTs; note how the distinctive region of the shape is barely visible. (c) Their feature-encoded representations, in contrast, have varying values *along* the contour, thereby giving more meaningful characterizations of the shapes. (d) The diffused feature images.

tively enforced. Fig.1 shows the pipeline of this framework.

The idea of matching shapes via image registration first appeared in [16]. In that work, shapes were embedded into images via signed Euclidean distance transforms (DT), which are then registered by minimizing their sum of squared differences. More recently in 2007, Munim and Farag [14] represented shapes with vector distance functions (or VDT) and matched shapes by registering these functions. While their methods are promising, we recognize that the use of VDTs or DTs has several problems. First, they cannot always describe a shape sufficiently well as they only represent shapes implicitly and do not encode shape features. Second, they are sensitive to noise, so the behavior of local registration of DTs is often unstable in the presence of spurious structures [5, 6]. Third, they cannot cope well with occlusions in shapes (Fig.2). Neither can they represent symmetric shapes well as these shapes have similar DTs [6], and so registration of these images are prone to failure (Fig. 3).

In contrast, the use of features offers a more descriptive and robust representation of the shapes. By constructing a multi-scale space of feature images (see Fig.4a), such a representation can be insensitive to noise and outliers. Thus, performing registration on multiple levels, each of which uses feature images that are diffused at the corresponding scale, coarse and fine levels of registration will be bet-

ter driven by global and local shape features respectively, thereby increasing the capture range and robustness of the algorithm.

The proposed method not only exploits the power of well-established image registration techniques and existing shape descriptors, but it can be extended in several ways. With the image space as the embedding space, it can be applied to match shapes of any dimensions and of any nature (*e.g.* shapes that are open or closed; consist of organized or unorganized points; have equal or unequal cardinality, etc.). As each vertex is now characterized by both its features and location, we can make use of both intrinsic features and spatial information to represent a shape. By making use of multiple complementary shape features, we can also characterize each shape more descriptively. Registration of these representations are now driven by shape descriptors, rather than only from its contour (as encoded by DT). Additionally, by maximizing a similarity metric computed between two feature images with respect to appropriate spatial transformations, the solution does not need to be constrained by the finite space of the assignment permutations. Hence, an outlier point is not required to match with any other point in another shape. Lastly, because shapes are now represented as images, this formulation allows for the borrowing of numerous image registration and regularization methods that can now be easily applied to enforce neighbourhood constraints in shapes.

We note that there exists several independent work on registration of vector-valued images and of feature images, albeit not for shape matching. In [18], multi-channel images are registered using a similarity measure based on information theory. While relevant, this work was not intended for registration or matching of shapes. Zhang and Rangarajan [24] performed affine registration of images containing projected features (*i.e.* scalar-valued), which can thus be regarded as one particular choice of features in our framework. In addition to affine transformations, however, our framework also can solve for other spatial transformations. Furthermore, rather than employing one type of feature, our method can combine different types of features that complement one another (*e.g.* geometric and medial-based features), thereby allowing for more informative descriptions of the shape.

2. Methods

2.1. Overview

We establish shape correspondence through image registration. Given two shapes, we first extract shape features for each point and encode each of these features as an image intensity value of the pixel corresponding to the physical location of the point. Thus, extracting multiple features gives rise to two vector-valued feature images. Next, we apply

isotropic diffusion to these images and perform linear registration with few degrees of freedom. Linear registration is then performed in a coarse-to-fine manner such that as we go from one level to the next finer level, more localized and less diffused features are used to solve for spatial transformations, which also increase in degrees of freedom. Nonlinear registration is used afterwards, wherein local displacement of point locations are estimated. Similarly, as registration proceeds from one level to the next, we apply a lesser amount of diffusion to images which contain more localized shape features to solve for the displacement vector field that would bring each image pixel into alignment with a corresponding pixel of another image. From the computed vector field, correspondence is established by visiting each point location in the displacement map and computing its new location. With this location, pair-wise assignment can be achieved by nearest-neighbour assignment.

2.2. Shape Features

We begin with a 2D shape that is represented with a set of points P , each of which is represented as p_i and is located at (x_i, y_i) . For each p_i , we extract a set of K features $f_{1_i}, f_{2_i}, \dots, f_{K_i}$. These can be geometric, topological, and if available, appearance-based features (*i.e.* if the shape is extracted from an underlying source image). We note here that the optimal set of features to be extracted is data-dependent and is indeed a research topic outside the scope of this work. In this work, we chose a set of K features, f_1, f_2, \dots, f_k , each belonging to a category listed in Table 1 (note the acronyms shown for the different features).

2.3. Shape representation as vector-valued feature images

Given K sets of feature values, we normalize each set to the range $[0, 1]$ to prevent bias towards features with larger ranges [1]. Specifically, if f_k is found to be normally distributed, we normalize it according to:

$$\tilde{f}_k = (f_k - \mu_{f_k})(6\sigma_{f_k})^{-1} + 1/2 \quad (1)$$

where μ_{f_k} and σ_{f_k} are the mean and variance of f_k respectively. The additional shifting and rescaling guarantee that 99% of \tilde{f}_k is in $[0, 1]$ and fewer out-of-range values will be truncated [1]. If f_k is not normally distributed, we simply apply linear rescaling [1].

After normalization, we construct a vector-valued feature image by encoding each set of features into individual feature components. In doing so, we assign \tilde{f}_k of point p_i to the pixel corresponding to the nearest physical location of the point. When dealing with sparse pointsets, we use B-spline functions to interpolate new points between existing points. Repeating this encoding process for the source shape ζ_s and target shape ζ_t results in two feature images,

Table 1. Features employed in this work.

Geometric	Convexity (CONV) [11]. Curvature (CURV) [11]. Area integral invariant (AREA) [13].
Appearance	Statistics on intensity values (RATIO): Ratio between the average intensity values in the circular region within and outside the contour. Observed transport (AOT): An adaptation of [17] to encode appearance information contextually.
Medial-based	Local feature size (LFS) [23]. Skeletal points (SKEL) [4]: Skeletal points of the shape are encoded as one component.

which we now denote as F and M (short for “fixed” and “moving” images, which are terms commonly used in image registration).

2.4. Registration of feature images

2.4.1 Linear registration

We first seek the transformation parameters $T=\{t_1, \dots, t_k\}$ that would create pixel-wise intensity correspondences between F and M . As T is optimal when it matches similar features (*i.e.* intensity values), we hence minimize the *mean of squared differences* measure S computed between F and M :

$$S(F, M(T[x])) = \frac{1}{|\Omega|} \int_{\Omega} \sqrt{\sum_{i=1}^K w_k (F_k(x) - M_k(T[x]))^2} dx \quad (2)$$

where $x \in \Omega$ is a spatial location, F_k is the k -th component of F and w_k is the weight assigned to the k -th component, with $\sum_{k=1}^K w_k = 1$. More details on w_k will be presented in Section 2.5.

Image registration (and hence shape matching), entails minimizing S with respect to T , so we investigated two optimization algorithms: a gradient descent-based algorithm and a genetic algorithm (known as one-plus-one evolutionary optimization [7]). When the former method is used, we differentiate $S=S(F, M(T[x]))$ with respect to each of the transformation parameters t_i , *i.e.*:

$$\frac{\partial S}{\partial t_i} = \frac{2}{|\Omega|} \int_{\Omega} \left(\sum_{i=1}^k w_k (F_k(T) - M_k) \nabla M_k(T) \right) \frac{\partial T}{\partial t_i} \quad (3)$$

To decrease computation, we adopt the narrow-band approach [16] such that S is computed only over pixels that are within distance R to the shape contour, thereby effectively reducing the size of Ω to minimize the penalty introduced by the additional dimension employed (*i.e.* 2D images for 1D contours, 3D images for 2D surfaces). The choice of R will be discussed in Section 2.4.3.

2.4.2 Non-linear (deformable) registration

Numerous techniques have been developed to solve for the local deformations between two images [7, 12, 20]. We illustrate how two of such methods, namely parametric and non-parametric, can be adapted for shape matching. In the parametric method, images are incrementally deformed by manipulating a regular lattice of control points according to the tensor product of a third-order B-spline [7]. Through the use of a B-spline control-point grid, a smoothness constraint is implicitly enforced. Thus, this technique can model both small and large deformations [7], support dense registration that is continuous, and is generally robust to noise [16], thereby making it suitable for modeling the deformations of the embedded shapes. In solving for the displacement parameters of each control point, we minimize S as defined previously using a variant of the LBGS optimization algorithm [7]. For this method, we employ a multi-resolution control-point grid such that at each level of deformable registration, the spacing in x and y dimensions between the control points is decreased by G_x, G_y . This multi-level scheme, which is a common and successful practice in image registration [7, 12], can avoid entrapment in local minima and allows for stable deformations.

While effective, this method can be very expensive. Accordingly, we have also investigated the feasibility of Thirion’s [7] demons algorithm, a non-parametric method that is less costly. This method computes a deformation field D that describes the local displacement of each pixel according to an optical flow equation [7]. To enforce an elastic-like behaviour on D , we adopt Thirion’s [7] regularization approach and smooth it with a Gaussian kernel between iterations with decreasing scales (i.e. the standard deviation of the kernel decreases from 4 to 1 pixel).

In short, these two methods differ in terms of how regularization and proximity constraint are enforced, which ultimately govern how shapes can be deformed. As shown in the literature, demons algorithm can be highly effective for correcting small deformations. For shapes that differ by large deformation, the parametric method will be superior. Thus, the selection of either method should be made based on the type of deformations exhibited in the shapes (*i.e.* the type of regularization desired) and also on the type of correspondences permitted (*e.g.* one-to-one correspondences, or one-to-many correspondences, which may exist in medical images or anatomical shapes).

2.4.3 Multi-level and multi-scale registration

We employ a scale-space approach to increase robustness against noise and increase capture range of the algorithm. To do so, we apply Gaussian smoothing, initially at a scale σ_o , and perform registration in a coarse-to-fine manner (in L levels). As we proceed from one level to the next, we solve

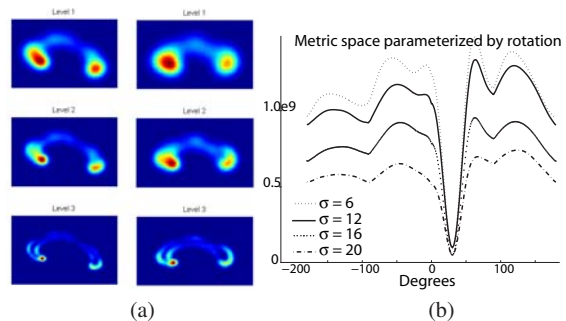


Figure 4. (a) Two sets of feature images of corpora callosa shapes with CURV. Coarse-to-fine registration with large-to-small scale smoothing (starting at level 1) can gradually improve the capture range of scale-space based optimization. (b) The metric value between the two shapes in Fig.3 as rotation varies. Note how the heights of local minima have decreased when diffusion is applied. Note too that the amount of diffusion (determined by σ) applied has not affected the location of the global minimum.

for T with more degrees of freedom, solving each within a smaller search space, and registering images that have been smoothed at lower scales $\sigma_1, \dots, \sigma_{L-1}$, which are decreased exponentially with decay γ . Fig.4a provides an example of this scheme. This process of performing finer registration and incorporation of more local features from level to level constitutes the core of our multi-feature, multi-scale, and multi-level registration framework. Since the scale of the kernel determines the level of detail at which the shape should be represented, which in turn is task-dependent, the choices of σ_o and γ were obtained empirically. We empirically found that it is optimal to set σ_o to be one third of the shape’s mean size and $\gamma \in [.5, 1.1]$, depending on the shape’s complexity, and set R (size of the narrow band) to be twice the value of σ_L .

Note that diffusion of images not only generates a smooth image gradient, but tends to convexify the objective function, thereby increasing the capture range of the optimization. This is reflected in Fig.4b wherein we see that the heights of the local minima around the global minimum are decreased when a large amount of diffusion is applied to the images, implying a smoother optimization path.

2.5. Feature weights

The weightings in the metric calculation allow registration at a particular level to be tuned to a particular subset of features. Expert or prior knowledge can be used to determine the values of these weights. If a large set of expert training correspondences is available, one can set the weight to be inversely proportional to the variance of the feature values across points known to correspond in a training set. Machine learning may also be used, for instance, as done by Ward et al. [22], to discover the discrimina-

tory features that correspond to human intuition of saliency and determine the weights accordingly. In this work, we assign weights so that medial-based features are emphasized initially, then shift focus to more localized geometric features, and finally, if available, to appearance-based features. Coarse-to-fine registration is henceforth driven by information pertaining to the global structure of the shape, to more localized geometric features, and if available, to textural and appearance-based features.

3. Results

We applied our method to establish point correspondences in both synthetic and real data and present both qualitative and quantitative validation results. In performing global alignment, we first seek the translation parameters and subsequently the affine-transform parameters. We then perform non-linear registration using the deformable registration algorithms described previously. For all experiments, we also performed B-spline-based registration using the shapes’ DTs to provide a comparison of our method against another image-based shape matching method [16].

We began our investigation on synthetic shapes (Fig.3), which occupy the image space of size 200×200 . We applied known translation and similarity transformations to these shapes to generate ground-truth correspondences. As no non-rigid deformations were introduced, performance is evaluated by calculating the mean Euclidean distances (MED) between corresponding points after registration has been performed. Table 2 summarizes the results of these tests. In addition, results using the one-plus-one evolutionary and gradient-descent based optimization were compared for solving for the optimal global transformation. From Table 3, we can see that the use of gradient-based optimization yielded higher success rates than the use of the one-plus-one algorithm. However, the latter had faster convergence and was less sensitive to the optimization parameters used. Thus, when applying our method to more complicated scenarios (*e.g.* involving 3D shapes and/or images with much larger resolution), one could employ the latter algorithm to solve for initial alignment and use the gradient-based algorithm for finer registration.

Next, we tested the overall method on anatomical shapes extracted from real medical images. These include: a) mid-sagittal contours of corpora callosa (CC) extracted from a set of 20 MR images; and b) contours of brain ventricles (BV) extracted from 14 MR images; and c) mid-axial contours of thoracic vertebrae (TV) extracted from CT volumes in 4 spinal studies. For validation based on ground-truth [9], we performed 3 to 7 repeated trials of manual landmarking on these datasets. The manually marked correspondences were later used to quantify the method’s accuracy, which

Table 2. Results of using DT and feature-based representations for synthetic shapes. Ten trials of random rotations $[-40^\circ, 40^\circ]$ and translations $[-10, 10]$ were applied to the shape shown in Fig.3. The scales of features are shown in parentheses (expressed as fraction of the shape’s size). Success rate is the percent of trials where the recovered translation parameters are within ± 5 and the rotation is within $\pm 2^\circ$.

Encoding method	MED	Success Rate
DT	3.79	75%
CURV (0.1, 0.16, 0.2)	2.87	92%
CURV, CONV, AREA (0.15)	3.11	87%

Table 3. Comparison of optimization methods. MED is the mean Euclidean distances between corresponding points after registration. $\overline{n_{itns}}$ is the average number of iterations taken using the same number of levels and similar convergence criteria.

Method	MED	Success Rate	$\overline{n_{itns}}$
One-plus-one	6.07	78%	154
Gradient-descent	4.55	83%	260

Table 4. Configuration of feature and weights used for each dataset. Values in parentheses are the feature scales (expressed as a fraction of image’s mean size). Non-linear registration starts at second last level. The notation scale (a/b) and weight (c/d) indicate that weight c was used for the feature extracted at scale a and weight d was used for the feature extracted at scale b .

Features	Weights at level l				
	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$
CC SKEL	.8	.5	0	0	0
CC AREA (0.2)	.1	.5	.5	.5	.5
CC AOT	.1	0	.5	.5	.5
BV SKEL	.8	0	0	0	NA
BV CURV (0.2)	.1	.5	.5	.5	NA
BV LFS	.1	.5	.5	.5	NA
TV SKEL	1	0	0	0	NA
TV CONV (0.20/ 0.15)	.5/ 0	.5/ 0	.25/ .25	.25/ .25	NA
TV AREA (0.25/ 0.20)	.5/ 0	.5/ 0	.25/ .25	.25/ .25	NA

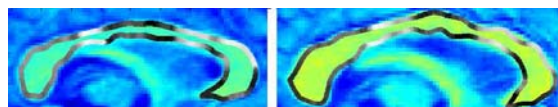


Figure 5. Two different CC contours overlaid on their source images. The appearance-based features, which are shown along the contour in grayscale, provide rich descriptions when local geometric features do not (Fig.4a).

is defined as the target error ε of each correspondence¹. A computed correspondence is then considered correct if its target error is lower than twice the corresponding variability in the repeated trials of manually marked correspondences. Robustness of the method is then calculated as the percentage of correct correspondences established.

For each dataset, we explored various schemes, each differed in terms of choice of K features, as well as values

¹For each known correspondence $p_i \mapsto q_i$, and the established correspondence, $p_i \mapsto q_j$, ε is the geodesic distance between q_i and q_j

of σ_o and γ (which together determine $L, \sigma_1, \dots, \sigma_L$) and some critical registration parameters (e.g. G_x, G_y). Specifically, we first examined the use of different geometric features. We then included appearance-based features for the registration of the CC data and found that this inclusion had increased the accuracy of the established correspondences. This is in accordance with [22], which showed how expert knowledge and appearance information can be used to improve the accuracy of point correspondence. As Fig.5 shows, the top and bottom regions of the CC shape have few meaningful geometric features, but contain descriptive appearance-based information. However, inclusion of appearance-based features for the BV data did not improve accuracy. This is likely due to the similarity in the textural properties around the contour of the ventricle shapes.

For some relatively more challenging cases (i.e. shapes that differ by large stretches and/or bending), we added medial-based features in the initial stages of global alignment so that registration is driven more by structural features. Then, in subsequent levels, geometric features were used to solve for local deformations. Fig.6 shows one configuration of such setup for registering two TV shapes.

Based on these initial experiments, we completed validation on each dataset using the configurations shown in Table 4. Fig.7 and Fig.8 show some of the established correspondences. Table 5 summarizes the performance of registration using DT and using our feature-based representation, as well as the results of applying two non-linear registration methods (Sec. 2.4.2). Using the parametric non-linear registration method yielded more accurate matching of CC shapes. The geodesic distances to known ground-truth positions of 4 out of 5 target points were reduced and the overall reduction in the average geodesic distance dropped from 3.2 to 2.6 [9]. For the BV and TV shapes, our method established more accurate correspondences on a subset of the landmark correspondences. In some cases, our method may have suffered from the low descriptive power of the features we used for describing the BV shapes, and thus yielded lower success rates. However, as we have only examined few of the many shape descriptors established in the literature, we expect improved results with a richer set of features and improved feature selection strategy.

To assess the sensitivity of the method to the weights, we also repeated the experiments on the CC and TV shapes using equal weights and results show that the use of unequal weights has generally led to faster convergence and slightly higher accuracy, thus demonstrating the effectiveness of employing complementary features.

Finally, to test the method's robustness against noise and occlusions, repeated trials of registration were performed on the CC and TV shapes of which data points are either removed or perturbed in the normal direction to the contour by a distance drawn from a Gaussian distribution. In describ-

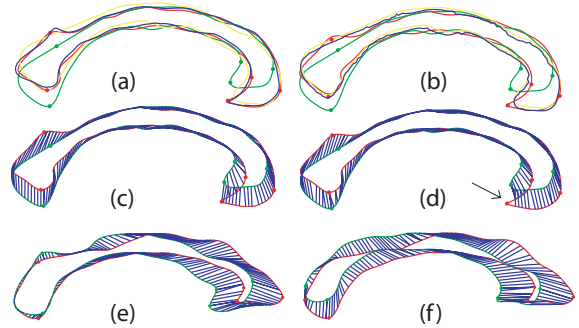


Figure 7. Landmark correspondences between different pairs of CC shapes (ζ_s in red; ζ_t before global registration in green, after global registration in yellow, and after non-linear registration in blue). Result using (a) feature-based and (b) DT-based registration. (c) and (d) show result of the same pairs with established correspondences. As DT does not encode saliency of feature points, its solution converged to local minimum, leading to inexact correspondences (highlighted by arrow). (e) and (f) result of another pair of shapes using feature-based registration.

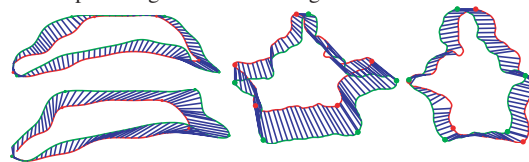


Figure 8. Point correspondences established for some BV and TV shapes.

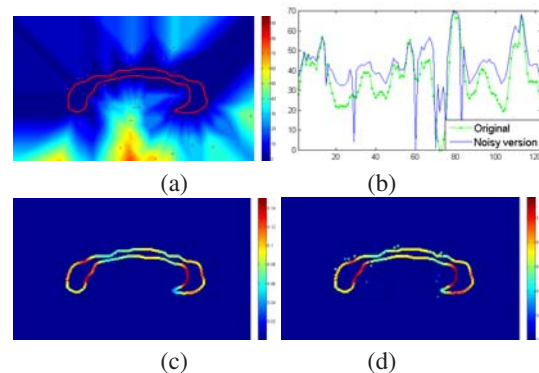


Figure 9. (a) The *difference* between the DT of the original shape and the DT of its noisy version dominated the *entire* image domain. (b) Profiles of AREA along the contour of the original shape and its noisy version. (c) Feature image describing the original shape. (d) Feature image describing its noisy version. Note the consistency in the images despite the presence of noise, and so its impact on registration is suppressed.

ing these polluted shapes, we extracted AREA at different scales. Due to its insensitivity to noise, this feature allows for a faithful description of the original shape. As Fig.9b depicts, the intensity profiles of this feature along the contour extracted before and after noise was added are fairly consistent. Thus, the impact of noise on the generated feature

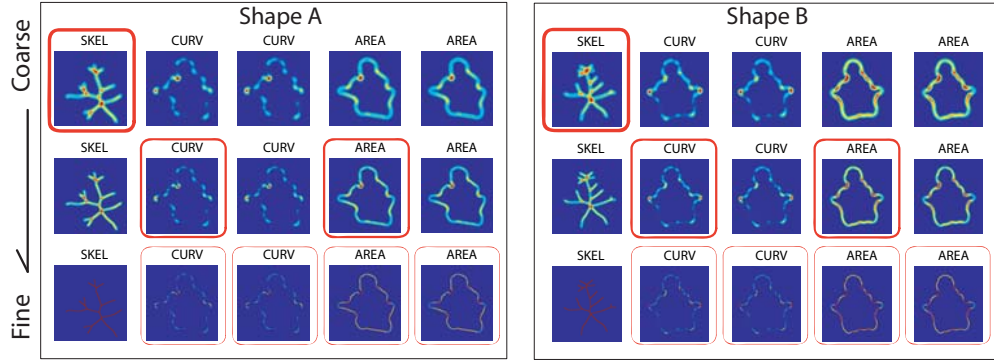


Figure 6. Scheme of linear registration for TV shapes. Each row i corresponds to the feature images used at level i . The thickness of the outlining boxes reflects the weight assigned to the k -th feature. Thus, initial stage of registration is driven by SKEL. Subsequent stages then focus on local geometric features (*i.e.* CURV and AREA) extracted at different scales.

Table 5. Quantitative evaluation of the established correspondences of landmark points p_1 to p_5 . $\bar{\tau}$ is the mean geodesic distance between the assigned point and ground-truth corresponding point. Non-linear registration was done with the parametric method (marked with B) or the demons algorithm (marked with D). The feature images (denoted as FI) consist of components shown in Table 4.

Method used	$\bar{\tau}$					Overall $\bar{\tau}$	Success rate	
	p_1	p_2	p_3	p_4	p_5			
CC	DT+B	2.4	2.5	3.0	2.9	4.2	3.0	87%
	FI+B	1.8	2.1	2.9	2.3	5.1	2.6	85%
	FI+D	2.1	3.2	3.0	3.2	4.7	3.2	91%
BV	DT+B	1.3	5.5	7.1	0.9	-	3.7	89%
	FI+B	2.6	4.8	6.7	1.8	-	4.0	74%
	FI+D	4.1	4.6	7.1	1.1	-	4.2	72%
TV	DT+B	1.5	2.0	2.9	9.1	2.3	3.6	77%
	FI+B	0.8	2.6	5.1	7.3	5.9	4.3	79%
	FI+D	2.5	2.6	2.7	8.4	8.5	4.2	68%

image is insignificant (Fig.9d). On the other hand, the impact of noise on the DT representation is apparent (Fig.9a). Consequently, the use of DT resulted in lower success rates (88% vs. 91% for CC and 97% vs. 83% for TV). When occlusions are present, the use of DT yielded 79% and 89% success rates for CC and TV shapes, respectively, while the use of feature images yielded 80% and 83% success rates for the respective shapes.

In this preliminary study, we started with the matching of anatomical shapes, which motivated this work, but have begun to match shapes from MPEG7 and other standard datasets [9]. Some qualitative results are shown in Fig.10 and Fig.11. Validation results using these databases will be reported in future work.

4. Conclusion

We have shown how point correspondence can be established via the registration of features images. To illustrate

the effectiveness of our method, we began with a simple selection of features. Future work include examining: 1) the use of more robust shape descriptors (*e.g.* (geodesic) shape context [8]); 2) possible training procedures for choosing the best set of features and weights; 3) the application of different transformation models developed in the literature, *e.g.* TPS [12], fluid [20], and poly-affine [2]; 4) the method's sensitivity to the feature set, weights, and the amount of smoothing applied; and 5) the robustness of the method when applied to match shapes from standard datasets (*e.g.* [9]). With further improvement in these areas, we anticipate improved results.

While one might argue that the increased dimensionality is costly, the additional computational expense is less than expected thanks to the adoption of the narrow-band approach. The advantages, in turn, are plenty. By encoding spatial and structural, geometric, textural and topological information as vector-valued images, we benefit from a much richer description of shapes that can be robust against noise and missing data, as we have shown in our experiments. Through the use of an embedding space and appropriate spatial transformation models, proximity or neighbourhood regularization can be easily enforced. The overall method can also be applied to match shapes of any nature and any dimension. Additionally, as we have demonstrated in the paper, we now have convenient access to rich libraries of tools on image registration (*e.g.* ITK, VTK, etc.) that can now be easily adopted for shape matching. In fact, as image registration is still an active research topic, advances in volumetric image registration will have a direct impact on shape matching as a result of the proposed formulation.

At a high level, the proposed framework introduces a new perspective to examining shape matching and leads to several possible extensions: 1) the use of articulated registration [10] and other procedures [3] to align locally rigid shapes or recover large deformations; 2) the adoption of groupwise registration [2] of feature images so that



Figure 10. Matching of shapes from a benchmarking dataset [9].

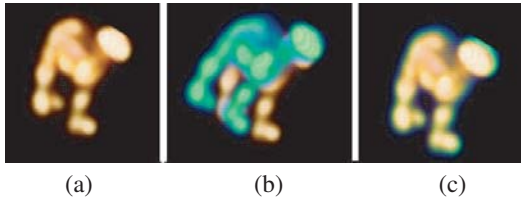


Figure 11. Preliminary results of shape matching in 3D. (a) Volume rendering of F_i of a 3D feature image (F_i embeds Gaussian curvatures and is shown with a ‘warm’ colormap). (b) Fusion of F_i and M_i before registration (M_i shown with a ‘green’ colormap). (c) Result of global registration.

our method can be used to establish dense correspondences across all shapes in a training set for statistical model construction; and 3) application to multi-modal medical image registration wherein the images involved are acquired from different imaging devices and hence, would benefit if registration can be driven by image features that are more consistent and/or meaningful than the raw image data. Ultimately, the work aims to fulfill a more ambitious goal of unifying the framework for the matching of shapes and of images.

5. Acknowledgements

We thank Aaron Ward for his assistance in proofreading and MITACS, NSERC, and Pacific Nuclear Medicine Inc. for providing partial funding for this work.

References

- [1] S. Aksoy and R. M. Haralick. Feature normalization and likelihood-based similarity measures for image retrieval. *PR Letters*, 22(5):563–582, 2001.
- [2] V. Arsigny, X. Pennec, and N. Ayache. Polyrigid and polyaffine transformations: A novel geometrical tool to deal with non-rigid deformations. *MIA*, 9(6):507–523, 2005.
- [3] G. Christensen, R. Rabbitt, and M. Miller. Deformable templates using large deformation kinematics. *Image Processing*, 5(10):1435–1447, 1996.
- [4] W.-B. Goh and K.-Y. Chan. Structural and textural skeletons for noisy shapes. In *Advances in Visual Computing*, volume 3804, pages 454–461, 2005.
- [5] B.-W. Hong, E. Prados, L. A. Vese, and S. Soatto. Shape representation based on integral kernels: Application to image matching and segmentation. In *CVPR*, volume 1, pages 833–840, 2006.
- [6] X. Huang, N. Paragios, and D. Metaxas. Shape registration in implicit spaces using information theory and free form deformations. *PAMI*, (8):1303–1318, 2006.
- [7] L. Ibanez, W. Schroeder, L. Ng, J. Cates, and the Insight Consortium. The ITK software guide, 2006.
- [8] V. Jain and R. Zhang. Robust 2d shape correspondence using geodesic shape context. In *Pacific Graphics*, pages 121–124, 2005.
- [9] J. Karlsson and A. Ericsson. Ground truth correspondence measure for benchmarking. In *ICPR*, pages 568–573, 2006.
- [10] X. Li, T. Peterson, J. Gore, and B. Dawant. Automatic registration of whole body serial micro CT images with a combination of point-based and intensity-based registration techniques. In *ISBI*, pages 454–457, 2006.
- [11] R. Loke, M. Bayer, D. Mann, and J. du Buf. Diatom recognition by convex and concave contour curvature. In *OCEANS’02*, volume 4, pages 2457–2465, 2002.
- [12] J. Maintz and M. Viergever. A survey of med image registration. *MIA*, 2(1):1–36, 1998.
- [13] S. Manay, D. Cremers, H. Woo, A. Yezzi, and S. Soatto. Integral invariants for shape matching. *PAMI*, 28(10):1602–1618, 2006.
- [14] H. Munim and A. Farag. Shape representation and registration using vector distance functions. In *CVPR*, pages 1–8, 2007.
- [15] A. Myronenko, X. Song, and M. Carreira-Perpinan. Non-rigid point set registration: Coherent point drift. In *Advances in Neural Info Processing Sys*, pages 1009–1016, 2007.
- [16] N. Paragios, M. Rousson, and V. Ramesh. Non-rigid registration using distance functions. *Compr Vision and Img Understanding*, 89(2-3):142–165, 2003.
- [17] A. Pitiot, H. Delingette, A. W. Toga, and P. M. Thompson. Learning object correspondences with the observed transport shape measure. In *IPMI*, pages 25–37, 2003.
- [18] G. Rohde, S. Pajevic, Carlo, and P. Basser. A comprehensive approach for multi-channel image registration. In *WBIR*, pages 214–223, 2003.
- [19] C. Scott and R. Nowak. Robust contour matching via the order-preserving assignment problem. *TMI*, 15(7):1831–1838, 2006.
- [20] P. Thompson and A. Toga. *Handbook of Med Imaging*, chapter “Warping strategies for intersubject registration”, pages 569–601. San Diego: Academic Press, 2000.
- [21] O. van Kaick, G. Hamarneh, H. Zhang, and P. Wighton. Contour correspondence via ant colony optimization. volume 0, pages 271–280. Computer Society, 2007.
- [22] A. Ward and G. Hamarneh. Statistical shape modeling using MDL incorporating shape, appearance, and expert knowledge. In *MICCAI (2)*, pages 278–285, 2007.
- [23] J. Xie and P. Heng. Shape modeling using automatic landmarking. In *MICCAI (2)*, pages 709–716, 2000.
- [24] J. Zhang and A. Rangarajan. A unified feature-based registration method for multimodality images. *BioMed Img*, 1:724–727, 2004.
- [25] Y. Zheng and D. Doermann. Robust point matching for non-rigid shapes: A relaxation labeling based approach. Technical report, University of Maryland, College Park, 2004.