



HAL
open science

A groupwise mutual information metric for cost efficient selection of a suitable reference in cardiac computational atlas construction

Corné Hoogendoorn, Tristan Whitmarsh, Nicolas Duchateau, Federico M Sukno, Mathieu de Craene, Alejandro Frangi

► **To cite this version:**

Corné Hoogendoorn, Tristan Whitmarsh, Nicolas Duchateau, Federico M Sukno, Mathieu de Craene, et al.. A groupwise mutual information metric for cost efficient selection of a suitable reference in cardiac computational atlas construction. SPIE Conference on Medical Imaging, 2010, San Diego, United States. pp.76231R, 10.1117/12.844428 . hal-02320752

HAL Id: hal-02320752

<https://hal.science/hal-02320752v1>

Submitted on 19 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A groupwise mutual information metric for cost efficient selection of a suitable reference in cardiac computational atlas construction

Corné Hoogendoorn^{a,b}, Tristan Whitmarsh^{a,b}, Nicolas Duchateau^{a,b}, Federico M. Sukno^{b,a},
Mathieu De Craene^{a,b}, and Alejandro F. Frangi^{a,b,c}

^a Center for Computational Imaging & Simulation Technologies in Biomedicine (CISTIB),
Universitat Pompeu Fabra, Barcelona, Spain

^b Networking Biomedical Research Center on Bioengineering, Biomaterials and Nanomedicine
(CIBER-BBN), Barcelona, Spain

^c Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

ABSTRACT

Computational atlases based on nonrigid registration have found much use in the medical imaging community. To avoid bias to any single element of the training set, there are two main approaches: using a (random) subject to serve as an initial reference and posteriorly removing bias, and a true groupwise registration with a constraint of zero average transformation for direct computation of the atlas. Major drawbacks are the possible selection of an outlier on one side, and an initialization with an invalid instance on the other. In both cases there is great potential for affecting registration performance, and producing a final average image in which the structure of interest deviates from the central anatomy of the population under study.

We propose an inexpensive means of reference selection based on a groupwise correspondence measure, which avoids the selection of an outlier and is independent from the atlas construction approach that follows. Thus, it improves tractability of reference selection and robustness of automated atlas construction. We illustrate the method using a set of 20 cardiac multislice computed tomography volumes.

Keywords: Cardiac atlas construction, bias reduction, reference selection, groupwise mutual information

1. INTRODUCTION

Computational atlases have found much use in the medical imaging community as of late. The robustness of their automated construction depends entirely on the success of the underlying image registration procedure. Although the elements of the registration process—similarity metric, transformation model and optimizer—in large part determine the performance, other major influencing factors are the images to be registered and the parameters employed in the registration. Usually each pair of images is registered using its own set of (near) optimal parameters, which represents a considerable amount of interaction prior to the actual atlas construction. In automated atlas construction based on nonrigid registration, one would therefore preferably carry out all registrations with as few sets of parameters as possible, meaning that these will be suboptimal for most pairs of images. Combined with possibly major differences between images, it deserves preference to find a reference as central to the population as possible to initialize the atlas construction procedure with.

The quest for centrality comes directly from the purpose of atlas construction itself: the creation of a population average that is constructed by minimizing the average deformation required to deform any member of said population into this reference. Additionally, the success of registration, given fixed parameterization and registration components, depends on the similarity between images upon initialization. This means we should maximize the similarity between the initial estimate of the population average and the members of the population.

Further author information (send correspondence to *A.F.F.*):
{corne.hoogendoorn, tristan.whitmarsh, nicolas.duchateau, federico.sukno, mathieu.decaene, *alejandfro.fragi*}@upf.edu

The option of selection based on expert opinion as proposed by Park *et al.*¹ must be discarded for scalability reasons; although the first ‘typical’ volume may be used, some definition (by the expert) of what constitutes ‘typical’ would have to be based on the entire set of volumes, or at least on a sizeable subset. The same group later proposed to select the reference based on the bending energy of the deformations, in this case represented by a thin plate spline.² From these, a matrix of pairwise distances is filled, after which multi-dimensional scaling is employed to find the initial reference. While claiming to demonstrate feasibility, their experiments encompass 50 two-dimensional images of 256×256 pixels. The amount of data we use in this work is two orders of magnitude larger, and is still at least an order of magnitude smaller than the set we consider for atlas construction. At this rate, the feasibility of the energy based approach rapidly diminishes.

Another family of approaches assumes no reference volume as a member of the set, but one generated as the voxel-wise arithmetic mean of the images under zero deformation. The central tendency is established by imposing limits on the average transformation³⁻⁵ while maximizing the similarity between the deformed images and the voxel-wise arithmetic mean of the deformed images. As such, there is no reference to introduce bias, but the stability of the method is jeopardized by the initialization: there is no guarantee that the voxel-wise mean image under zero deformation represents a valid instance of the population under study.

We present here a heuristic called groupwise mutual information (GWMI), derived from mutual information (MI), for the selection of a reference image. We show that it leads to the selection of a most typical image in a set. In addition to the centrality property, a low noise level aids in registration performance. We also demonstrate that our GWMI measure favors lower noise levels in equal anatomies. Finally, we show that employing our heuristic indeed results in a smaller mean local deformation in the mean deformation compared to candidates avoided based on our heuristic.

2. METHODS

The question of reference selection has been addressed previously using *a posteriori* selection based on the bending energy of a thin plate spline based nonrigid transformation.² However, this requires exhaustive nonrigid cross-registration, which is extremely expensive, while the majority of these registrations is ultimately discarded. Additionally it assumes that all registrations have converged to a globally optimal solution. We would limit this to only an affine cross-registration to make sure pose and scaling variations do not influence our results. After this registration, GWMI is applied to select the reference volume. This metric is cheap to compute in both time and memory.

We can extend the MI metric between two images \mathbf{I}_1 and \mathbf{I}_2 to a metric between an image \mathbf{I} and a set of images \mathcal{J} in the same fashion as in Bhatia *et al.*⁶ using

$$\text{GWMI}(\mathbf{I}, \mathcal{J}) = H(\mathbf{I}) + H(\mathcal{J}) - H(\mathbf{I}, \mathcal{J}). \quad (1)$$

With the probability of an intensity j in the set defined as

$$p(j) = \sum_{\mathbf{J} \in \mathcal{J}} p(j|\mathbf{J})p(\mathbf{J}), \quad (2)$$

and with an *a priori* uniform probability distribution over the candidate images, we are effectively summing the histograms of the images in the set \mathcal{J} defined on the same grid of bins:

$$\text{Histogram}(\mathcal{J}) = \sum_{\mathbf{J} \in \mathcal{J}} \text{Histogram}(\mathbf{J}). \quad (3)$$

The joint probabilities of an intensity i in the candidate reference with an intensity j in the set \mathcal{J} are defined analogously as

$$p(i, j) = \sum_{\mathbf{J} \in \mathcal{J}} p(i, j|\mathbf{J})p(\mathbf{J}), \quad (4)$$

also coming down to a summation of the joint histograms:

$$\text{JointHistogram}(\mathbf{I}, \mathcal{J}) = \sum_{\mathbf{J} \in \mathcal{J}} \text{JointHistogram}(\mathbf{I}, \mathbf{J}). \quad (5)$$

In the case where \mathcal{J} consists of only one image, this reduces to the standard MI.

In the same fashion we can compute a groupwise normalized MI (GWNMI):

$$\text{GWNMI}(\mathbf{I}, \mathcal{J}) = \frac{H(\mathbf{I}) + H(\mathcal{J})}{H(\mathbf{I}, \mathcal{J})}. \quad (6)$$

In the remainder of this work, any mention of GWMI can be replaced with GWNMI without loss of generality. In the Results section it will also become apparent that there is little reason to consider using one over the other as far as the outcome is concerned.

Using a leave-one-out approach where $\mathbf{I} \in \mathcal{J}$, we obtain a reference candidate through

$$\mathbf{I}_{\text{ref}} = \underset{\mathbf{I}}{\text{argmax}} \text{GWMI}(\mathbf{I}, \mathcal{J} \setminus \mathbf{I}). \quad (7)$$

That is, we attempt to find the image \mathbf{I} which carries most information about the population.

From the construction of the histograms over the sets it follows that the histogram of $\mathcal{J} \setminus \mathbf{I}$ and the joint histogram of \mathbf{I} and $\mathcal{J} \setminus \mathbf{I}$ will be dominated by a subset of \mathcal{J} (and of $\mathcal{J} \setminus \mathbf{I}$) that represents images that are alike, assuming that there is only one such subset. In such a case, the image with the highest GWMI is likely to be a member of that subset, while other members would also be viable candidates. The subset itself could be analyzed using the pairwise application of traditional MI to establish whether it consists of a single cluster of images. If not, this could be a reason to opt for atlas stratification.⁷

2.1 Atlas Construction

For a qualitative evaluation, we constructed atlases using all our subjects. Our nonrigid transformation is modeled as a composition of B-splines, each of which is constrained to be diffeomorphic,⁸ and at increasing level of detail.⁹ The total transformation is given by $T_{\text{total}} = T_8 \circ T_{16} \circ T_{32}$, with the subscripts indicating, in millimeters, the B-spline control point grid spacings. As each of the three deformations in the composition is diffeomorphic, T_{total} is also guaranteed to be diffeomorphic. For the computation of the mean deformation, the log-Euclidean framework is employed.¹⁰ Our optimizer is a quasi-Newton optimizer.¹¹

3. RESULTS

3.1 Data

Our test data consists of 20 multi-slice computed tomography (MSCT) volumes of the human heart, of 512×512 pixels in-plane and around 60 slices each at a spacing of $0.4 \times 0.4 \times 2.0$ mm. All measurements were done within a thresholding mask, including only locations where the voxel intensity value fell within the $(-1000 \dots 1500)$ HU range in all images. Because computation of the score involves addition of histograms, it is generally not possible to apply our method to other modalities without a prior intensity standardization step.

3.2 Computational Aspects

GWMI is computationally inexpensive compared to registration. However, implementation details find a tradeoff between memory consumption and time. For very large-population atlases, memory consumption is the most pressing issue. Therefore our implementation is memory efficient, having no more than two images in memory at any given time, but requiring many disk operations. Figure 1 presents timings for the algorithm with varying numbers of subjects, picked randomly from our set of 20, after affine registration to the same subject*. This latter condition allows comparison between images of equal size.

The complexity of computing the GWMI is $O(n^2)$ for n subjects; this is not different from exhaustive cross-registration. However, the implicit constant in this complexity is defined by the iteration through an image rather than performing a registration.

*The machine used was equipped with an Intel Core i7 920 CPU, running at 2.67 GHz, with 6 GB of RAM.

3.3 Noise

The property that the score should favor ‘clean’ images is illustrated in Fig. 2a. A central slice of one image in our data set was taken, and 50 increasingly noisy copies were generated to form a set. This was done using both additive and multiplicative noise. In Computed Tomography (CT), the noise is typically additive due to the photon energy distribution as well as system noise. The figure shows a decreasing score for images with increasing noise levels, reflecting a faster increase of the joint entropy compared to the individual candidate image entropy, and a decrease in the entropy of the set, the second term in Eq. (1).

3.4 Global Alignment

The GWMI score clearly increases as the images are affinely registered, as is shown in Fig. 2b. While it is not immediately apparent that the ranking of candidates changes, this may be coincidence. It should be noted, however, that the general location of the heart within real-world coordinates is quite stable, and that therefore most of the variation removed by the affine registration is due to rotation and size, weaker contributors to similarity than translation.

The affine registration that precedes the computation of the GWMI measure also requires a reference to be chosen. However, our results show that the chosen reference does not influence the GWMI-based selection much (see Fig. 3); regardless of the subject used as the reference for the affine registration, dramatic changes in ranking for subjects are only observed in the middle segment (not plotted in Fig. 3). At the top and bottom of the ranking, changes are observed only rarely, which is very desirable. With pose variations removed from the set, it is the shapes embedded in the images that are evaluated for their representativity of the other shapes. Thus, while it still deserves preference to continue the atlas construction from the images that were affinely registered to the reference chosen for the nonrigid registration (that is, chosen based on GWMI), it may not be necessary to carry out an exhaustive cross-registration with affine transformations before the selection procedure. This makes the complexity of this preprocessing step linear with respect to the number of subjects.

3.5 Typical Elements

By adding more copies of one of \mathcal{J} 's elements (\mathbf{I}) to \mathcal{J} , we show that the metric does favor the most typical element. However, the number of copies required to elevate a volume's GWMI to the level that makes it the reference of choice, clearly increases as \mathbf{I} is more of an outlier (lower GWMI score). This is illustrated in Fig. 4. Once this necessary number of copies has been added, the weaker the subject was initially, the lower its final ‘winning’ score was. In fact, this held for all subjects that required 5 or more copies, illustrating how this set of copies represents a cluster largely disjoint from the remainder of the images; such a scenario would call for further examination of the data set.

3.6 Quantification of Deformation

A comparison of GWMI scores with quantifications of the transformations is not presented; such a comparison would imply having reached the global optimum in the registration step, which is not guaranteed by the optimization strategy¹¹ but depends on the initialization. Unless the optimum found is global, the transformation required to achieve the obtained similarity metric value is not unique, and neither will any quantification of the transformation.

We do, however, perform statistics on the local mean deformations represented as vector fields derived from the B-splines. A comparison between means and percentiles of vector magnitudes is presented in Fig. 5. It shows that up to the 75th percentile the vector magnitudes in the strongest candidate are clearly smaller than those for the other candidates, but more importantly that the vector magnitudes for the weakest candidate are very clearly larger than the rest. This means that this candidate is indeed further away from the population mean than the other candidates.

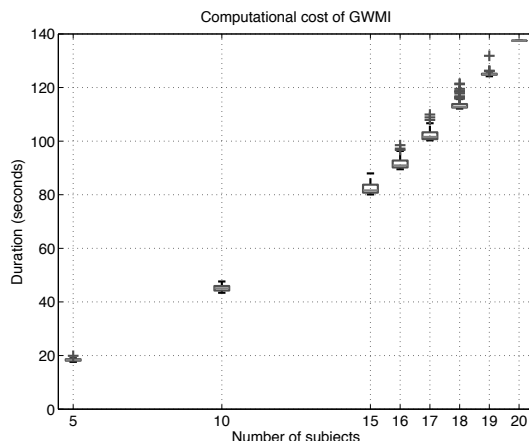


Figure 1. The computational cost in time of the GWMI, for sets of increasing sizes. The sets containing fewer than 19 subjects were obtained by random permutation of the indexes of our 20 subjects and using the first 5, 10, . . . subjects. This was repeated 100 times. The data for the sets of 19 subjects was obtained using leave-one-out.

3.7 Qualitative Inspection

Qualitative results are presented in Fig. 6. It shows the two candidates deemed most and least fit to serve as the initial reference (respectively corresponding to the solid and dotted lines in Fig. 3a). While our focus is on the registration of the heart, the intensity window we use covers everything from the air in the lungs to the bone of the ribs, sternum and spine. The only parts that are consistently eliminated from both the GWMI computation and the registration are the electrocardiogram leads and the area outside the field of view (this corresponds to the top and bottom, anterior and posterior parts of the vector fields depicted in Fig. 6). Additionally, some sternal cortical bone tends to be removed, while for one subject the leads of an implantable assist device and starvation artefacts related to the device fall outside the intensity window.

4. CONCLUSIONS

We have developed a computationally inexpensive means for the selection of an initial reference for automated atlas construction based on nonrigid registration, which requires only a number of prior affine registrations that is linear in the number of subjects. The method requires only three parameters: the resolution of the histogram and the two bounds of a possible image intensity window.

We have shown that the reference selected before the necessary affine transformation preceding the nonrigid registration does not significantly influence the choice of reference afterwards. By controlling the composition of the set we show that the metric indeed moves towards the most typical candidate, and that the metric favors clean image in the case of equal anatomy. The local mean deformation of the candidate chosen is smaller, albeit not significantly so.

A limitation of the method is its usefulness in imaging modalities in which image intensity does not relate straightforward to physical properties of the tissue imaged. This includes Magnetic Resonance, which alongside CT represents the most extensively used modality for the construction of high-resolution computational atlases.

The features combined, taking into account the limitations, should improve the robustness of automated atlas construction from CT imagery, as compared to the selection of a random reference, or initialization using a voxel-wise mean of the population under zero deformation.

ACKNOWLEDGMENTS

This work was funded through the Integrated Project euHeart (FP7/ICT-2007-224495) in the European Commission's 7th Framework Programme, through project TIN2009-14536-C02-01 of the Spanish Ministry of Science

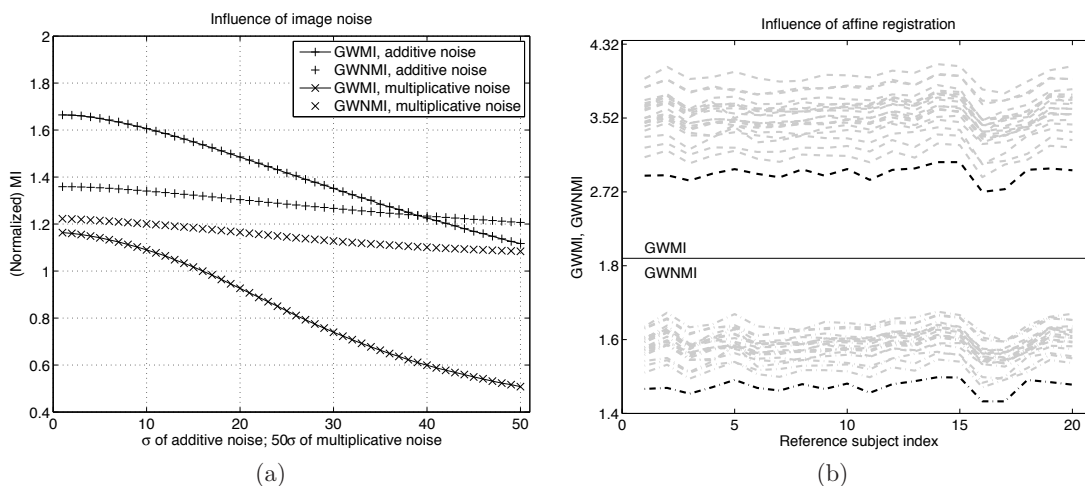


Figure 2. (a) The effects of image noise on the (normalized) scores. The (+) symbols indicate additive noise (typical for CT), the (x) indicate multiplicative noise. The normalized mutual information scores are indicated with markers only, as opposed to a combination of markers and curves for the non-normalized scores. It is clear that 'cleaner' images will obtain higher scores. (b) The effects of affine registration on the scores. The dark lines indicate the GWMI (above, dashed) and GWNMI (below, dash-dot) of the non-aligned images; the light lines indicate these scores for the affinely registered images. Each of the lines represents a different reference used for the affine registration.

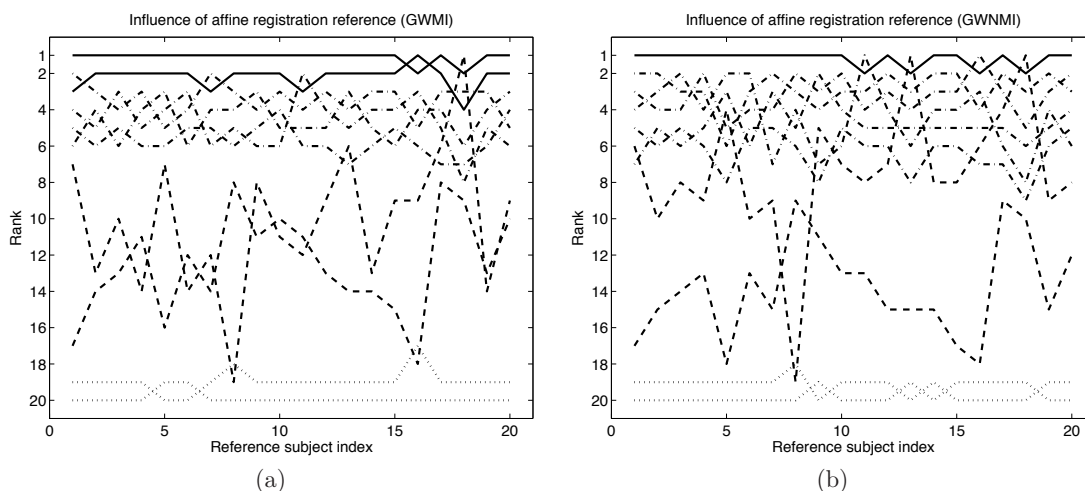


Figure 3. The influence of the subject used as the reference in the affine registration step. Each line represents the development of the ranking of a subject based on its (a) GWMI or (b) GWNMI score. Plots for 10 subjects in the lower middle segment have been removed for clarity, with both (a) and (b) showing the same set of subjects. The line style reflects a (subjective) classification of suitability regarding both ranking and consistency therein. Solid line: strong candidate; dash-dot: second-tier candidate; dashed line: very inconsistent candidate; dotted line: very weak candidate, which should be avoided at all cost. As it happens, the small changes among the two solid line subjects in (a) occur when the two dotted line subjects are used as affine registration reference.

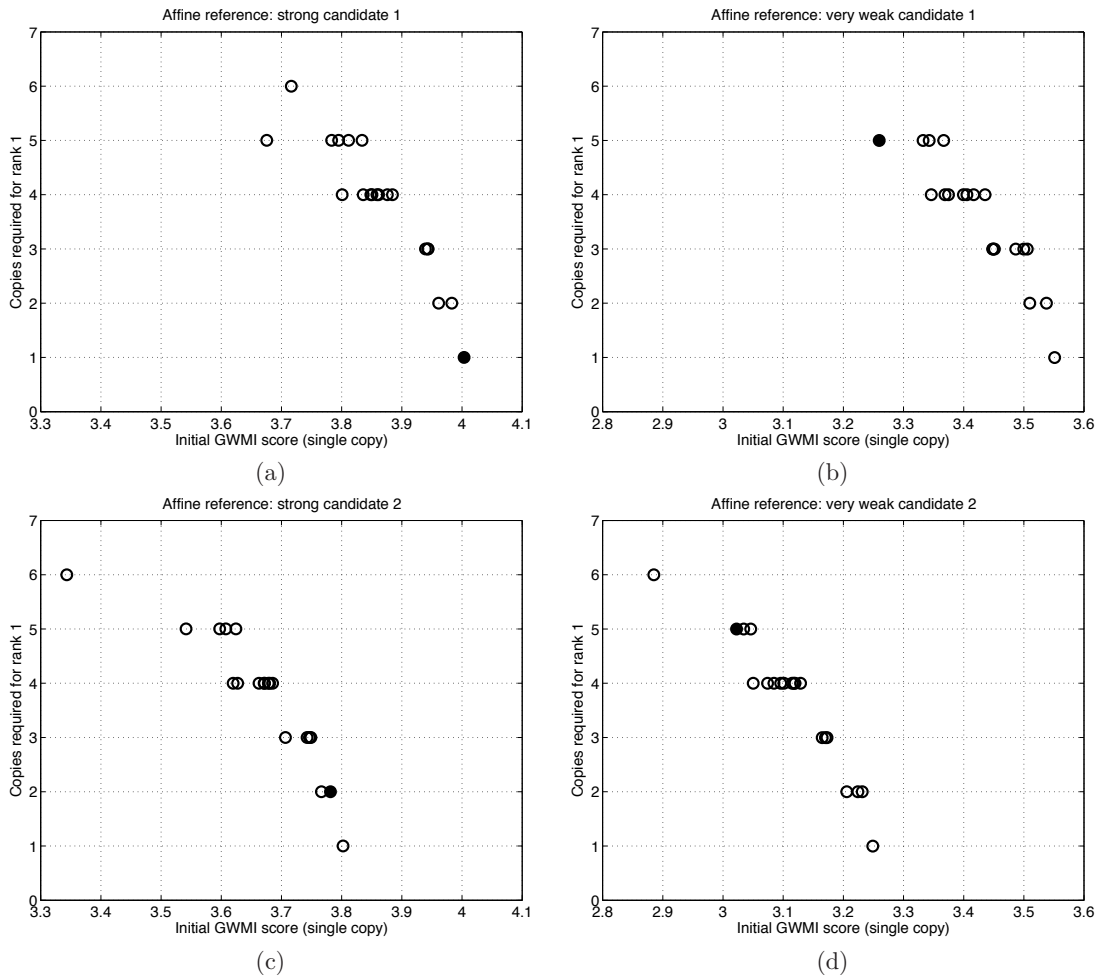


Figure 4. The GWMI versus the number of copies required to elevate the subject to the rank of highest GWMI. The references used for the affine registration are (a) and (c) the strong candidates and (b) and (d) the weak candidates as indicated in Fig. 3a (strong/weak candidate 1 is the generally top/bottom curve in Fig. 3a respectively). The solid markers indicate these 'affine reference' subjects; the open markers indicate the subjects registered to these references.

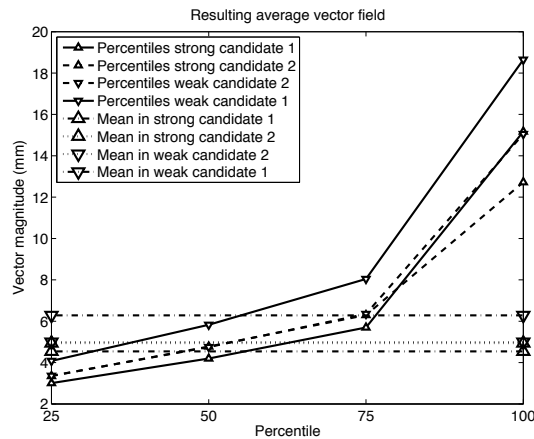


Figure 5. Percentiles and means of the vector magnitudes in the mean vector field. The (Δ) markers indicate strong candidates, (∇) markers indicate weak candidates. The solid and dashed lines with the smaller markers indicate percentiles, the dash-dot and dotted horizontal lines with the larger markers indicate the mean magnitude.

and Innovation, and through the CDTEAM project under the CENIT program of the ministry's Industrial and Technological Development Center (CDTI). T. Whitmarsh is supported by the Instituto de Salud Carlos III through PFIS predoctoral grant FI09/00757.

REFERENCES

1. Park, H., Bland, P. H., and Meyer, C. R., "Construction of an abdominal probabilistic atlas and its application in segmentation," *IEEE Trans. Med. Imag.* 22, 483–492, Apr. 2003.
2. Park, H., Bland, P. H., Hero III, A. O., and Meyer, C. R., "Least biased target selection in probabilistic atlas construction," in *Proc. MICCAI, LNCS 3750*, 419–426, 2005.
3. Joshi, S., Davis, B., Jomier, M., and Gerig, G., "Unbiased diffeomorphic atlas construction for computational anatomy," *NeuroImage* 23(S1), S151–S160, 2004.
4. Studholme, C. and Cardenas, V., "A template free approach to volumetric spatial normalization of brain anatomy," *Pattern Recogn. Lett.* 25(10), 1191–1202, 2004.
5. Zöllei, L., Learned-Miller, E., Grimson, E., and Wells, W., "Efficient population registration of 3D data," in *Proc. CVBIA, LNCS 3765*, 291–301, 2005.
6. Bhatia, K. K., Hajnal, J. V., Puri, B. K., Edwards, A. D., and Rueckert, D., "Consistent groupwise non-rigid registration for atlas construction," in *Proc. IEEE ISBI*, 908–911, 2004.
7. Blezek, D. J. and Miller, J. V., "Atlas stratification," *Med. Image Anal.* 11(5), 443–457, 2007.
8. Rueckert, D., Aljabar, P., Heckemann, R. A., Hajnal, J. V., and Hammers, A., "Diffeomorphic registration using B-splines," in *Proc. MICCAI, LNCS 4191*, 702–709, 2006.
9. Duchateau, N., De Craene, M., Silva, E., Sitges, M., Bijnens, B. H., and Frangi, A. F., "Septal flash assessment on CRT candidates based on statistical atlases of motion," in *Proc. MICCAI, LNCS 5762*, 759–766, 2009.
10. Arsigny, V., Commowick, O., Pennec, X., and Ayache, N., "A log-euclidean framework for statistics on diffeomorphisms," in *Proc. MICCAI, LNCS 4190*, 924–931, 2006.
11. Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C., "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Comput.* 16(5), 1190–1208, 1995.

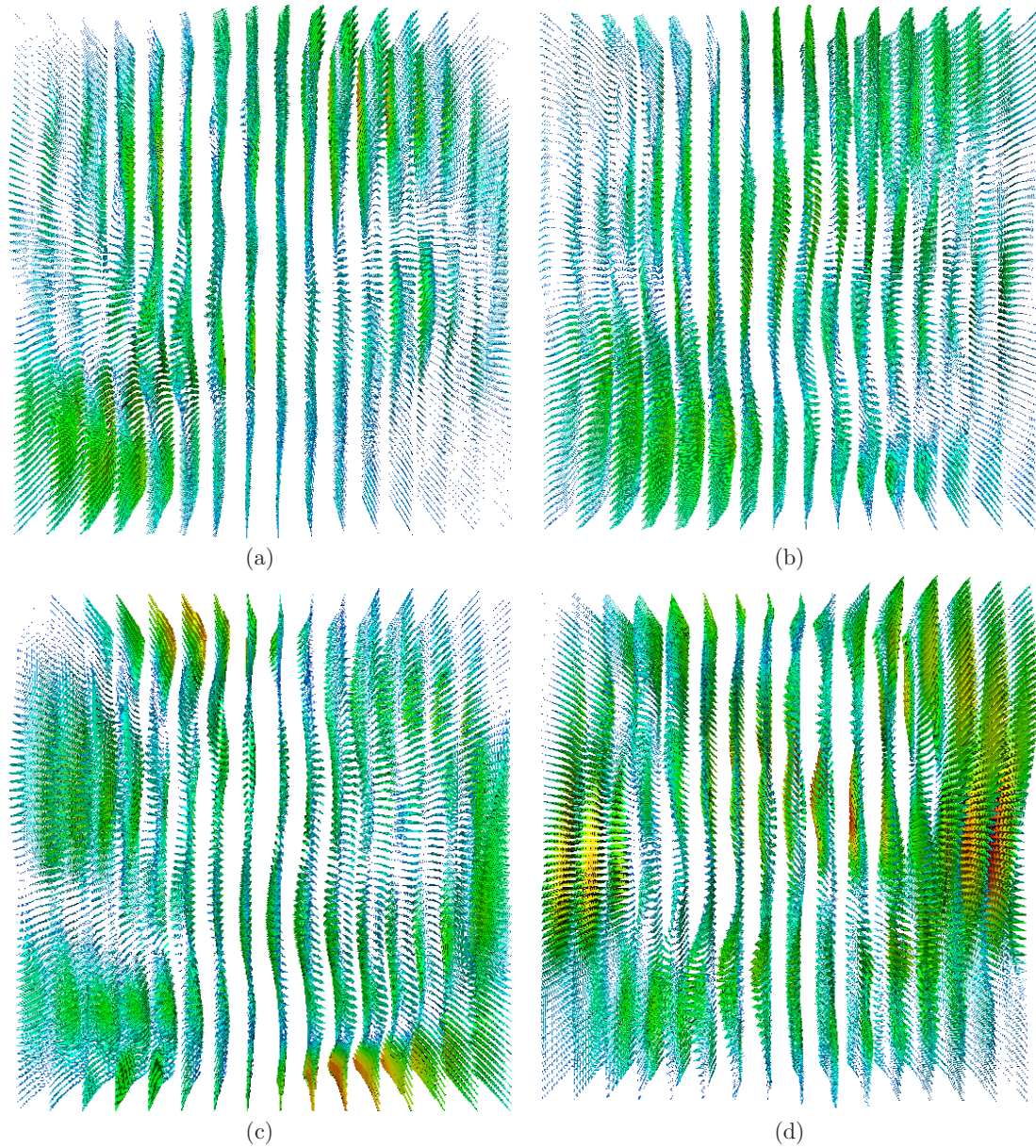


Figure 6. Vector fields of (a,b) the two strong candidates and (c,d) the two weak candidates. Through the red/dark and dense areas in both (c) and (d), it shows that in general a greater amount of deformation was required to deform the subjects' images to match those of these weak candidates, as compared to the stronger candidates shown in (a) and (b).