MOLECULAR IMAGE REGISTRATION USING MUTUAL INFORMATION AND DIFFERENTIAL EVOLUTION OPTIMIZATION

Bartosz Telenczuk[‡], María J Ledesma-Carbayo[§], Javier A Velazquez-Muriel[♯], Carlos O S Sorzano [♯ℵ], Jose-Maria Carazo[♯] and Andrés Santos[§]*

[‡] Wroclaw University of Technology, Poland
[§]ETSI Telecomunicación, Universidad Politécnica de Madrid, Spain
[♯] Centro Nacional de Biotecnologia-CSIC, Madrid, Spain
[ℵ] Escuela Politécnica Superior, Univ. San Pablo - CEU, Madrid, Spain

ABSTRACT

In this work we propose a novel rigid image registration approach to determine the position of high-resolution molecular structures in medium-resolution macromolecular complexes. Mutual information similarity measure is used as an alternative to the cross-correlation coefficient commonly applied in this context. The optimum of the objective function is sought by means of differential evolution algorithm. This global optimization technique yields robust registration, exhibits fast convergence and is easy to use. In order to additionally improve its accuracy we combine it with a local gradient search strategy. The registration framework is tested both on simulated and experimental data sets forcing large rotations and translations. Results in terms of success rate and execution time, indicate the suitability of the proposed approach.

1. INTRODUCTION

Three-dimensional electron microscopy is a powerful technique that allows imaging macromolecular structures nearly at their native state [1]. The micrographs obtained by the microscope are X-ray projections of the specimen under study and they are reconstructed into a volume which is compatible with the experimental data acquired. The resolution of these reconstructed volumes range between 6 and 25 Å.

This medium-resolution information can be complemented with high-resolution data coming from X-ray diffraction experiments, Nuclear Magnetic Resonance or molecular modelling. These techniques produce high-resolution data of some of the domains (pieces) of the macromolecular complex under study. The combination of both kinds of information (medium and high resolution data) provides a powerful tool to understand how each piece (known at high resolution) interacts within the whole complex (known at a medium resolution) to perform a given biological function [1].

In order to combine these two data sources the spatial location of the domain within the complex is needed. This is a rigid registration problem in which the domain has six degrees of freedom (three translations and three rotations) to fit within the bigger complex. In principle, there is no clue about its location and the full space must be sought. Traditionally, it is done by maximizing the correlation of the two volumes. Recently, the correlation has been computed locally (only within the region occupied by the domain) and some information about the molecule surface has been added [2]. However, the problem remains open since not always the algorithm converges to the global maximum.

In this article, we study the possibility of replacing the cross-correlation similarity measure by mutual information. It is experimentally shown that the goal function in this latter case has fewer local maxima than that of cross correlation. Furthermore, we employ a hybrid optimization technique that combines global and local strategies to correctly and efficiently identify the position of the high-resolution domain within the medium resolution particle. In order to assess the usefulness of the proposed registration algorithm, simulated as well as experimental data were used.

2. METHODS

Main steps of the registration technique proposed to solve our problem is shown in Figure 1. The key points we have worked on to get a successful result are the similarity measure and the optimization approach. The registration algorithm has been implemented in C++ within the framework provided by Insight Segmentation and Registration Toolkit (ITK 2.1) [3].

*Corresponding author: mledesma@die.upm.es. Partial support is acknowledged to: the European Union (NoE EMIL LSHC-CT-2004-503569 and Marie Curie HPMT-CT-2001-420); the Spanish Health Ministry (research project IM3 PI052204) and Comunidad de Madrid (grant GR/SAL/0234).

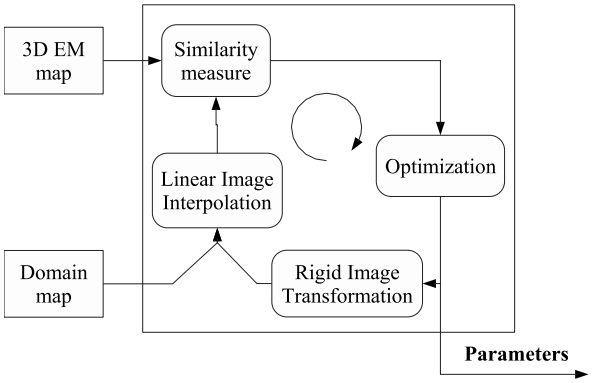

Fig. 1. Diagram of image registration steps

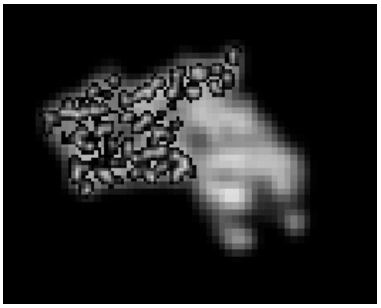

Fig. 2. Example of the combination of the high-resolution domain PDB-1a8d02 into the medium resolution fragment C of the *Tetanus* toxin PDB-1A8d, as it would be reconstructed from electron micrographs. Notice that both data sources are three-dimensional.

2.1. Similarity measure

The similarity measure establishes correspondence between images by comparing pixel intensities or other image features. In the literature many measures were proposed which proved to be suitable for various applications [4]. In the context of our problem the similarity measure more widely used is the cross-correlation coefficient (CCC) [2]. In this work we propose to use mutual information (MI) [5] as an alternative, as it is a good measure to detect nonlinear correlations between voxels' intensities adequate for images from different modalities or different resolutions. We have actually used the implementation of MI proposed by [6] and included in ITK 2.1.

A smoothing preprocessing step was performed on the images to balance their resolutions. Moreover in order to reduce the influence of the background noise, MI was computed only for those voxels whose intensity was greater than a given threshold. Such masking is similar to calculation of local similarity measures described in other works [7].

2.2. Optimization

The nature of our problem results in a similarity measure function with many local minima apart from the global minimum. The main reason for this local minima is the unconstrained location of the high-resolution domain in the whole complex and the possible symmetries found in the molecular structures under study (see figures 3 and 4).

In such case finding the optimal solution poses a difficult task. Extensive search although frequently successful requires much computation time which grows exponentially with increasing number of parameters. On the other hand gradient search based methods are easily attracted by local minima and fail to find the best transformation parameters. However computational cost and robustness are well balanced by so-called global optimization algorithms.

Recently many new schemes for finding global optima of nonlinear functions have been proposed. Some of them include: Nelder-Mead simplex algorithm, simulated annealing, genetic algorithms and stochastic equations. Differential Evolution (DE) is an evolutionary strategy introduced by Storn and Price in 1996 [8] which proved to be very efficient in many complex optimization tasks. The main advantages of DE are its convergence properties, robustness and simplicity of usage and implementation.

In the method a population of random parameter vectors is generated and in every generation a set of new vectors is constructed from already existing elements by the simple rule:

$$\underline{v}_{i,G+1} = \underline{x}_{r_1,G} + F \cdot \left(\underline{x}_{r_2,G} - \underline{x}_{r_3,G}\right), \quad i = 1, 2, .., NP, \quad (1)$$

where $\underline{v}_{i,G+1}$ is a new perturbed vector, $\underline{x}_{i,G}$ is a population vector, r_1, r_2, r_3 are random numbers, G is the generation number, F is a weighting factor and NP is the population size.

The resulting vector $\underline{v}_{i,G+1}$ is then compared to another vector randomly drawn from the population. The one that yields the best value of the objective function is retained for the next generation. Additionally, in order to increase diversity of the population, crossing-over with probability CR can be introduced. The algorithm is repeated a selected number of generations G (there is no stopping condition).

In most cases DE optimization converges to a solution near to the global minimum, however it often fails to find precisely the optimal transformation. In order to improve the accuracy we introduced a hybrid algorithm (DE+G). First, a rough solution is found by means of DE and then it is refined by Regular Step Gradient Descent (RSGD) method.

3. EXPERIMENTS

In order to assess the usefulness of the proposed registration algorithm, simulated as well as experimental data were used. The full complex of the simulated data was generated from the C fragment of the *Tetanus* toxin (entry 1a8d in the Protein

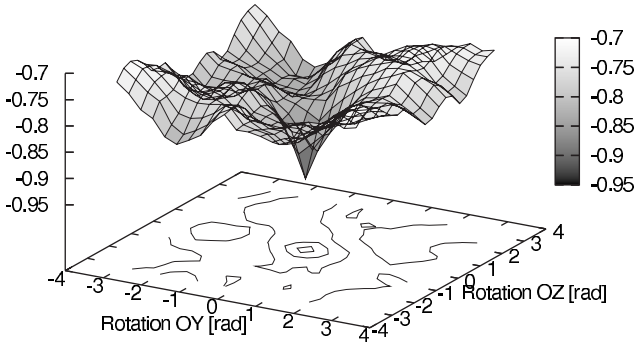

Fig. 3. Local correlation coefficient calculated at different rotation angles.

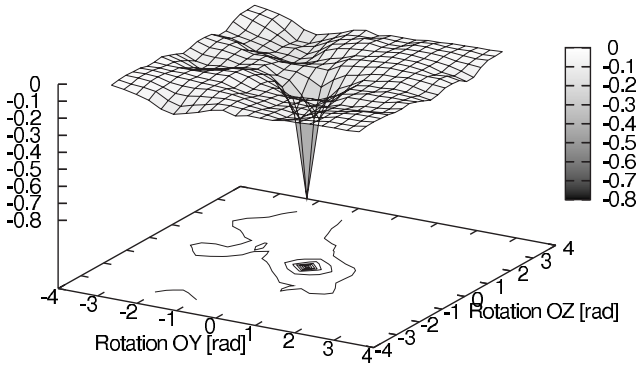

Fig. 4. Mutual information calculated at different rotation angles.

Data Bank (PDB, http://www.rcsb.org/*pdb*)) and its resolution was computationally lowered to 8Å. Its second domain (1a8d02, from aminoacid 247 to aminoacid 452), an all β structure, was lowered to 3Å and the goal was to find the location of this domain within the full complex. The sampling rate for the full complex was 2Å/pixel while for the domain was 1Å/pixel (Figure 2).

The chaperonin GroEL of *E. Coli* was used as the first experimental case (entry 1081 in the Macromolecular Structure Database, MSD (http://www.ebi.ac.uk/msd/index.html)). This entry is an experimental reconstruction at 6 Å resolution of the protein. The resolution of domains 1, 2 and 3 of chain B of GroEL (entry 1oel in PDB) was lowered to 3Å. The sampling rate was 1 Å/pixel for the three domains and the protein.

A second set of experimental data was used: the structure of N-ethyl maleimide sensitive factor at 11Å resolution as experimentally reconstructed (MSD entry: 1059) was used as the full complex to find its D2 domain (PDB entry: 1d2n). The resolution of the domain was lowered to 3Å. The sampling rate was 1 Å/pixel for both volumes.

The previously described data were originally pre-aligned, and therefore the location of the domains in the full proteins were known. This data setting allowed us to perform multiple experiments altering the location of the subdomains with different ranges of translations and rotations.

The following experiments were then conducted on the Tetanus toxin data set. First, the MI and CCC similarity measure functions were compared for different ranges of rotations. Secondly, the hybrid optimization (DE+G) method proposed was tested in terms of efficiency and success rate in comparison to DE and RSGD. A registration is considered successful if the difference between the attained MI value and the known optimum one is smaller than 10% of the latter.

Finally, different experiments were conducted taking into account all the test data sets to tune the algorithm parameters in order to get a good compromise between success rate and efficiency.

4. RESULTS

The comparison of MI vs CCC similarity measures can be observed in figures 3 and 4 calculated for different angles of rotation around two perpendicular axes. Figure 3 shows the landscape of the CCC objective function. There is a global minimum but surrounded by many local minima. Figure 4 shows the MI similarity function with a single well pronounced global minimum corresponding to the best alignment of the two images.

Optimization	Correct	Mean Error [rad]			Time [s]
		α	β	γ	
Gradient	15/31 (48%)	-0.14 $\pm$0.51	0.11 $\pm$0.60	-0.10 $\pm$1.49	5.6
DE	31/31 (100%)	-0.03 $\pm$0.16	0.02 $\pm$0.15	0.02 $\pm$0.16	41.9
Hybrid	31/31 (100%)	-0.002 $\pm$0.008	-0.012 $\pm$0.038	0.000 $\pm$0.019	45.4

Table 1. Efficiency and accuracy of optimization algorithms after 31 experiments with angles in range of $-\pi \div \pi$ rad. Successful rate, mean error of final angles and approximate evaluation time of one registration are given.

Table 1 compares evaluation time and accuracy of the three different optimization methods. RSGD is the fastest one, but fails for more than 50 % of initial conditions. Although DE optimization needs much more metric evaluations (> 1000) and consequently is much slower, it proves to be robust and accurate. The accuracy of the method can be further improved with little computational overhead by usage of hybrid approach which gives ten times better estimation being only 10 % slower.

In order to evaluate the described registration technique in real world problems we performed a series of tests in which correct alignment of the molecules was estimated for 50 ran-

dom initial orientations of the domain. In spite of the fact that initial rotation angles range from $-\pi$ to π radians most of the solutions come very close to global optimum of MI while only few registrations fail. Unfortunately if translations are also introduced number of correct solutions decreases. In Table 2 accuracy of the hybrid optimization ($NP = 30$, $G = 60$, $F = 0.4$, $CR = 0.9$) is evaluated as a function of a range of initial displacements along one of the axes (OZ). When initial translation is increased the number of successful registrations slightly decreases.

Initial	Mean Error [rad] / [pixels]				Correct
ΔZ	α	β	γ	Z	
± 1	-0.11	-0.04	0.02	0.05	45/50
	± 0.75	± 0.39	± 0.56	± 0.48	(90%)
± 5	0.22	0.03	0.12	0.18	40/50
	± 0.62	± 0.87	± 0.90	± 0.65	(80%)
± 9	-0.01	0.18	0.07	0.22	39/50
	± 0.76	± 0.82	± 0.72	± 0.58	(78%)
± 15	-0.07	0.16	-0.08	0.17	38/50
	± 0.89	± 0.81	± 0.92	± 2.59	(76%)

Table 2. Accuracy of registration as a function of the range of initial translation parameters (ΔZ). Translations X and Y were fixed while angles varied in range of $-\pi \div \pi$ radians.

Finally the same type of experiment was conducted taking into account all data sets and all possible translations and rotations. It should be remarked that the range of initial translations is large (± 20 pixels). Different values for the algorithm parameters were tested to find a good compromise between success rate and efficiency. Table 3 shows the success rates for parameter values $NP = 60$, $G = 200$, $F = 0.1$, $CR = 0.5$. With this setting the registration time was 75 seconds in a standard PC (AMD Athlon XP 3200+). Success rate can be improved by increasing the parameters NP and G, at the expense of longer the computation times.

Images		
Domain	Molecule	Correct
PBD-1a8d02	PBD-1a8d	65%
PBD-1oelB1	MSD-1081	95%
PBD-1oelB2	MSD-1081	76%
PBD-1oelB3	MSD-1081	95%
PBD-1d2n	MSD-1059	86%

Table 3. Successful registration rate for the simulated (PBD-1a8d) and experimental (MSD-1059, MSD-1081) data sets. Registration was run for 20 random initial parameters - all translation parameters (X,Y and Z) varied in the range of $-20 \div 20$ and all rotations in the range of $-\pi \div \pi$.

5. CONCLUSIONS

In this paper, we have examined the problem of fitting atomic models into three dimensional electron microscopy maps (3D EM). We proposed using mutual information as a similarity measure which outperforms more popular correlation coefficient. Its accuracy was further improved by applying mask to input images which reduces contribution of background noise.

Moreover we have described a new optimization algorithm which is the combination of differential evolution and gradient search strategies. Our experiments prove that it is robust and flexible and performs very well even applied to complex, non-linear objective functions. The algorithm compares favorably to local search methods which usually depend strongly on initial conditions.

In our study registration was successful in more than 65 % of cases (for the worst case) in a trial setting of large translations and rotations. This percentage could be improved at the expense of longer computation times. Results seem promising and should be confirmed on more experimental data.

6. REFERENCES

[1] M. van Heel, B. Gowen, and R. Matadeen, "Single-Particle electron cryo-microscopy: Towards atomic resolution." *Quarterly Review of Biophysics*, vol. 33, pp. 307–369, 2000.

[2] P. Chacon and W. Wriggers, "Multi-resolution contour-based fitting of macromolecular structures." *J Mol Biol*, vol. 317, no. 3, pp. 375–384, Mar 2002.

[3] L. Ibanez, W. Schroeder, L. Ng, and J. Cates, *The ITK Software Guide*. Insight Software Consortium, 2003.

[4] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image Vision and Computing*, vol. 21, pp. 977–1000, 2003.

[5] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," *International Journal of Computer Vision*, vol. 24, no. 2, pp. 137–154, Sept. 1997.

[6] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellyn, and W. Eubank, "Nonrigid multimodality image registration," in *Medical Imaging 2001: Image Processing*, July 2001, pp. 1609–1620.

[7] M. G. Rossmann, M. C. Morais, P. G. Leiman, and W. Zhang, "Combining X-ray crystallography and electron microscopy." *Structure (Camb)*, vol. 13, no. 3, pp. 355–362, Mar 2005.

[8] R. Storn and K. Price, "Minimizing the real functions of the ICEC'96 contest by differential evolution," in *IEEE Conference on Evolutionary Computation*, 1996.