Tagged-ICP: An Iterative Closest Point Algorithm with Metadata Knowledge for Improved Matching of 3D Protein Structures

Peter Ankomah, Peter Vangorp, Ardhendu Behera, and Yonghuai Liu

Edge Hill University

Abstract

Three-dimensional shapes are important in representing physical objects in digital form. This digital representation is useful in applications in numerous fields including chemistry, biology, and engineering. The benefits in analyzing and processing such 3D digital data have given rise to vast amounts of available 3D data and related applications.

In recent times, techniques for determining the functional and structural relationships amongst proteins consider the whole 3D structure (spatial coordinates), proceeding from earlier techniques that were based on sequence information. However, techniques considering the 3D structure of all atomic positions of the protein are too demanding for fast similarity searches especially when the protein is made up of very large numbers of atoms.

Iterative Closest Point (ICP) is the standard algorithm for performing 3D shape matching tasks. ICP has several issues that can affect the process of 3D shape matching. These include the need for an initial transformation to ensure an optimal match, inability of algorithm to converge when rotations are large, and computational cost and complexity of the distance calculation.

We propose an improvement of the ICP algorithm, Tagged-ICP for matching 3D protein structures that takes into consideration known feature descriptions of the points. The search for correspondence in our algorithm matches atoms based on their meta data (atom types), making this approach more meaningful. Our algorithm also reduces the number of distance calculations by a factor depending on the partition. The neighbourhood information also increases the partitions and reduces the size of the search space even further.

Our experimental results based on the publicly accessible Protein Data Bank show that matching becomes inherently meaningful and the complexity of the distance calculation is reduced. Our results also demonstrate improvements in speed, accuracy and convergence on larger rotations over the standard ICP algorithm.

Keywords: Iterative Closest Point (ICP), 3D point cloud, 3D shape matching, 3D protein structures, Atom types

1 Introduction

Background and motivation Comparing shapes is basic to many core problems including matching MRI scans in medical imaging [Nabavi et al., 2000], developing complete objects from partial scans [Levoy et al., 2000] and studying similarity, molecular design, and protein docking in chemistry [Axenopoulos et al., 2016]. Devices and applications allowing 3D scanning have become ubiquitous [Yuan et al., 2016] and the benefits (e.g., high precision) in analysing 3D data have resulted in the emergence of new applications in different fields, processing these vast amounts of available 3D data. In fact, the number of 3D protein structures in the Protein Data Bank (PDB) [Berman et al., 2000] doubles every 18 months, requiring more advanced methods for organising the data [Holm and Sander, 1998].

In the field of molecular biology, comparing structures is used to investigate the functions and interactions between different molecules such as proteins and nucleic acids [Angaran et al., 2009]. Comparing structures of proteins is also fueled by the accepted principle that the 3D structure of a protein is linked with its function [Kinoshita and Nakamura, 2005]. Molecular biology projects have produced a vast number of 3D structures that have enabled the discovery of functions of proteins [Berman et al., 2000; Ellingson and Zhang, 2012].

Several approaches have been used to detect structural and functional relationships amongst proteins. The earliest algorithms were based on sequence information [Schmitt et al., 2002] such as comparing pairs of amino acids (molecules) in a protein structure [Needleman and Wunsch, 1970]. 3D protein comparisons require more sophisticated algorithms to capture, visualise and match the structures based on spatial coordinates. 3D Proteins can be represented as rigid objects and a transformation found to align them, however, more reliable matches would be attained if the coordinates are associated with some predefined properties [Schmitt et al., 2002] such as the atom types.

An improved shape matching algorithm can benefit this problem. A key computer vision problem is to best align two shapes by minimizing the distance between the source and target shapes in order to determine the extent of similarity or dissimilarity. This technique is used in tasks such as facial and fingerprint recognition, machine vision, assistive and automotive technologies [Burlacu et al., 2016].

The Iterative Closest Point algorithm (ICP) [Besl and McKay, 1992; Chen and Medioni, 1992] is a well-known and dominant shape matching algorithm [Attia and Slama, 2017] because of its simplicity and straight forwardness. The algorithm is very effective at registering (aligning) point clouds because of its speed and accuracy [Attia and Slama, 2017]. ICP iteratively registers a data point cloud to a model point cloud with the aim of best aligning them, whether or not they partially or fully overlap. The registration method of the algorithm finds a transformation that reduces the misalignment (distance) between the point clouds until a defined error threshold or the maximum number of iterations is met [Donoso et al., 2017a].

Several studies have used ICP to compare structures of proteins by representing them as point clouds of atoms [Weskamp et al., 2004; Shulman-Peleg et al., 2008].

Limitations of existing solutions ICP requires a good initial transformation to ensure that point clouds converge at an acceptable minimal. The algorithm may also not converge at all from a particular transformation [Besl and McKay, 1992]. In fact, larger transformations have actually been found to decrease the efficiency of the algorithm [Attia and Slama, 2017]. The presence of outliers (non-uniform points) can also affect alignment with ICP [Chen and Belaton, 2014]. Another notable problem is that ICP performs well in some data sets and context and worse in the others [Donoso et al., 2017b].

Overview This paper discusses the standard Iterative Closest Point algorithm and evaluates its performance against our variant on matching 3D protein data sets. Related works are discussed in Section 2. Section 3 details the analysis and implementation of our novel algorithm, Tagged-ICP that improves the matching of 3D proteins by considering metadata knowledge of the atoms. The section details how our algorithm also reduces the search space by partitioning the space according to atom types. Section 4 details the experimental setup and design as well as the results, analysis, and implications. Section 5 concludes this paper with the reiteration of the problems with ICP, the solution our algorithm provides, and our ongoing and future work.

2 Related Work

The standard ICP algorithm [Besl and McKay, 1992; Chen and Medioni, 1991] repeatedly computes the transformation that aligns a data point cloud and a model point cloud and applies this transformation to the data point cloud until it reaches a set error threshold or the maximum number of iterations.

Chen and Medioni [1991] and Besl and McKay [1992] independently published similar methods for creating a complete 3D model from a physical object using point clouds captured using 3D scanning from different angles of the object. The distance from each point in the data point cloud to each point in the model point cloud is computed to establish corresponding points. A set of transformations are then computed to register (align) these surfaces in an iterative way. The shapes are registered when the root mean square error (standard deviation of the distances between the two clouds) is acceptably small.

The ICP algorithm is designed to always converge. It also does not require the use of any extracted features or meta-data about the points in the cloud.

2.1 Algorithmic Improvements

Several variants of the algorithm have been developed to improve one or more aspects of the algorithm [Phillips et al., 2007]. Some of the improvements include reduction in overall computational cost, smaller mean square error, faster convergence speed and optimal selection of points for overall algorithm efficiency. The rise in the development of these variants has been fuelled by the simplicity and effectiveness of the ICP algorithm itself. Mora et al. [2016] provide detailed surveys of many variants and their improvements to the ICP algorithm.

Algorithm 1 Iterative Closest Point algorithm as formulated by Besl and McKay [1992].

function ITERATIVE CLOSEST POINT(P, X)
  Input: data and model point cloud P, X
  Initialisation: $P_0 \leftarrow P$
  for iteration $k := 0$ to $k_{\max}$ **do**
    closest points Y_k $\leftarrow$ CLOSEST POINT SEARCH(P_k, X)
    transformation M_k, MSE d_k $\leftarrow$ REGISTRATION(P_0, Y_k)
    $P_{k+1} \leftarrow$ TRANSFORM(M_k, P_0)
    if change in MSE $d_{k-1} - d_k <$ threshold **then**
      terminate
  return P_{k+1}, M_k, d_k

2.2 Use of Additional Meta-data

Some improvements to the ICP algorithm enable the use of known meta-data in the form of features and constraints (e.g., colors and labels) to make the search for correspondence between points more meaningful. Some features such as global orientation information can change when transformations are applied to them, as such, there is always the need to factor them when calculating the mean square error. On the other hand, invariant features such as colors, labels and local point positions for rigid shapes are not affected by transformations. These invariant features can also be used to improve the registration [Combes and Prima, 2009] by reducing the search space for correspondence [Sharp et al., 2002].

Schutz et al. [1998] proposed a multi-feature ICP variant that includes the color and global orientation information in the distance computation. the research showed that convergence has been considerably improved with the addition of the feature information. Similarly, Thirion [1996] demonstrated an ICP variant using geometric invariant feature points. The research showed that, for a rigid object, the relative positions of the points (local orientation) can serve as invariant feature points because their positions relative to the rigid shape do not change when transformations are applied.

The development of machine learning algorithms has seen their application to the registration problem [Aoki et al., 2019]. Wang and Solomon [2019] proposed a learning-based registration method based on the idea of deep learning. The method takes two point clouds and is able to predict the rigid transformation to align them. Having trained the machine learning model with ModelNet (a custom-built large-scale 3D CAD model dataset) [Wu et al., 2015], the algorithm was found to outperform ICP in terms of efficiency. This is due to the algorithm predicting a rigid transformation in a single pass compared to the iterative classical ICP.

2.3 Applications

ICP is appropriate for aligning protein surfaces which are often not similar [Axenopoulos et al., 2013]. This research focused on extracting features of the shape that were rotation invariant to aid in matching the protein surfaces for protein docking. ICP has also been extensively used for matching biological structures such as measuring the spatial structural similarity of biological data such as proteins. Lu et al. [2016] used ICP for alignment refinement after applying a data reduction method to reduce the missing residues when matching two proteins that are not identical. This two-stage process ensured that the matching was less sensitive to noise and ICP was able to be used effectively. Similarly, ICP was used in the comparison of binding sites on proteins for drug discovery [Bertolazzi et al., 2010]. The research overcame the problem of ICP needing a good initial transformation by formulating a continuous global optimization algorithm that iteratively updated random points in a cluster based on the worst matched point of the cluster. Using sample protein data taken from PDB [Berman et al., 2000], Bertolazzi et al. [2010] proposed a method to detect similarities in protein binding sites when such similarities actually exists. Ellingson and Zhang [2012] developed an ICP based algorithm for superimposing

and comparing protein binding-sites by representing protein atoms as point clouds, with atoms having descriptor labels of their properties to aid in the matching.

Other geometric information was also used to improve the precision of the algorithm including multiple initial local alignments derived from adapting 3D Delaunay triangulation [Cignoni et al., 1998]. Zhou et al. [2014] adapted ICP to group a 3D representation of the activities of bio-molecules. The research used ICP for computing the structural distances and another alignment method to compute the chemical distances. These distance measures were then combined and clustered.

3 The Tagged ICP Algorithm

Our algorithm is based on the basic ICP algorithm (see Algorithm 1). To establish correspondences, ICP has to compute the distances between each point of the model point cloud and each point of the data point cloud. This is an expensive and critical step that becomes a performance bottleneck as the number of points increases. Besl and McKay [1992] proposed using binary search data structures such as *k*d-tree to reduce the complexity of the distance calculation process.

In order to improve this expensive process, our algorithm uses known metadata to optimize the search for correspondences and to make it more meaningful. This optimisation reduces the search space because it only computes the distance between two points with the same metadata information which we call the *tag*. The tag consists of the point's own atom type (carbon, hydrogen, etc.) as well as the atom types of its k nearest neighbours, unordered. In the pre-processing step, the model cloud is partitioned, grouping atoms with identical tags. During the ICP algorithm, the closest point search only searches the partition with the same tag as the atom in the data shape. All results in this paper use $k = 3$. Note that this improvement can be combined with other optimisations of the inner **for** loop such as the *k*d-tree search suggested by Besl and McKay [1992].

Algorithm 2 Tagged Iterative Closest Point algorithm. The metadata of the protein is used to speed up the search for correspondence and improve the registration accuracy: The model point cloud X has been partitioned by tag and only the partition corresponding to the tag of p is searched.

```
1:  function CLOSEST POINT SEARCH(P_k, X)
2:      Input: data and model point cloud P_k, X
3:      Initialisation: closest points Y_k ← empty list
4:      for all points p in P_k do
5:          closest point distance d_min ← ∞
6:          closest point y ← null
7:          for all points x in X_p.tag do
8:              distance d ← DISTANCE(p, x)
9:              if d < d_min then
10:                 closest point y ← x
11:                 closest point distance d_min ← d
12:         append closest point y to closest points Y_k
13:     return Y_k
```

3.1 Computational Complexity

Given a data point cloud P with N_P points and a model point cloud X with N_X points, the computational complexity of a naive Closest Point Search algorithm is $O(N_P N_X)$. The Tagged Closest Point Search has the same computational complexity but it reduces the number of distance calculations by a factor depending on the partitioning. Given fractional partition sizes $f_{P,t}$ and $f_{X,t}$ for each point cloud and each tag t such that $\sum_t f_{P,t} = 1$ and $\sum_t f_{X,t} = 1$, the number of distance calculations is reduced by a factor $F = \sum_t f_{P,t} f_{X,t} \leq 1$. The factor F would be smallest for a larger number of equal-size partitions. However, the fractional partition sizes in our dataset are given by the abundances of the atoms and their neighbourhoods. If the tag consisted only of the point's own atom type and omitted the neighbourhood information, the computation time would only be reduced by a factor of $F \approx 0.45$ in practice. The neighbourhood information increases the number of partitions and reduces their size to lower F further.

4 Experiments and Metrics

Our algorithm was implemented in C#. We also made use of Unity3D [Unity Technologies, 2019] as the IDE to develop the algorithm and visualise and inspect the shape matching results (Fig. 1).

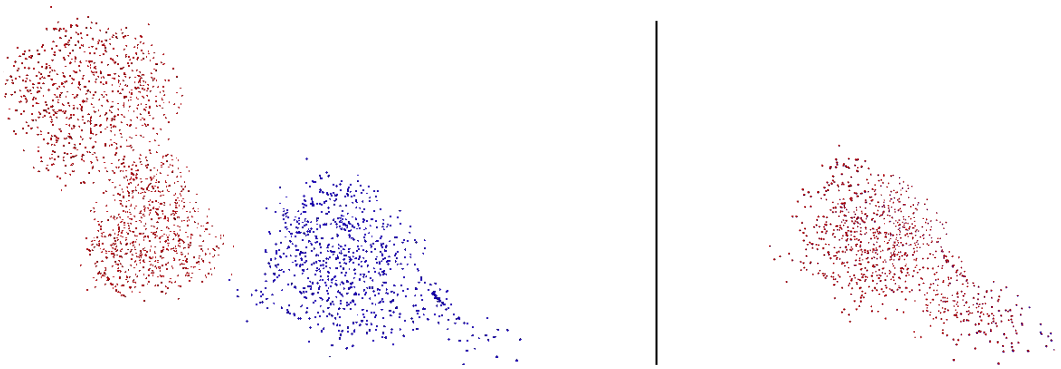

Figure 1: Same shape PDB ID: 2HAX with data shape at angle 30 degrees before match, and after Tagged-ICP match

The experiment setup allows a PDB (Protein Data Bank) [Berman et al., 2000] file to be uploaded using Cell Unity [Gehrer, 2015]. Cell Unity allows spatial information of a PDB file to be extracted and represented as a structure similar to the ball-and-stick model. For each protein dataset, the PDB ID was entered into Cell Unity and the application retrieved the PDB flat file and automatically generated a 3D rigid object representation of the protein based on the spatial coordinates of the atoms in the file. The atoms are represented by spheres and the spheres are able to show the 3D position of the atoms in the protein molecule. We then set our model and reference objects as intended for the generated shapes in our Unity application. The experiment was designed to match sets of the same and different protein structures. Three protein structures with PDB identifiers 2JKF (shapes with mutations), 2JKG (shapes with noise), and 2HAX (matching same shapes) were used in the experiment. For each test, the shapes were aligned along their centers of mass. Data shapes were then rotated at different angles (30, 60, 90, 120, 150 degrees) along the x-axis and the algorithms were run with 100 iterations. The aim of the experiment was to compare the performance in terms of the convergence and match quality of the Tagged-ICP to the original ICP algorithm over different initial rotation angles, noise levels, and when attempting to match two mutations of the same protein. Large initial rotation angles are known to cause difficulties for ICP algorithms. Noise added to the positions of the atoms simulates inaccuracies in protein measurement methods. Mutations are a common use case for matching the identical functional parts and highlighting the differences caused by the mutation. The standard ICP algorithm was chosen because we wanted to measure the effect of only our improvements. The success of this improvements means they can be incorporated into other variants for further experiments on 3D proteins. Using other variants of ICP meant that their improvements will also affect the convergence and match quality and we would not have measured our improvements accurately. We also did not place much emphasis on the translation vector because we aligned the shapes along their centers of mass, meaning they become reasonably close. In order to compare and analyse the performance of the matching of standard ICP with Tagged-ICP, we show two types of graphs.

The convergence graphs represent the root mean squared error of the alignment of the two shapes at different particular iterations. This data (iteration, mean square error) is generated in real time after each iteration. This allows us to understand which algorithm converged faster (iterated less to converge) in terms of the iteration count at convergence and not the time in seconds. The match quality graphs represent the cumulative percentage of points with their mean squared error when the shape matching is complete. This data (point, alignment error) allows us to understand the effective match quality based on the error of the cumulative sum of points. A higher cumulative percentage of points with a lower mean square error implies a better quality of match than a higher percentage of points with a higher mean square error. These graphs provide insights into the behaviours of different techniques: how fast and smoothly they evolve from one iteration to another. Each algorithm uses its own implementation of the closest point search. However, for a fair comparison of match quality, ICP finally runs Tagged-ICP's closest point search method, implying that the match quality graphs for ICP are determined by its registration process only. This is done to identify distortions between ICP's convergence and the quality of match for the different implementations. For instance, from these graphs we can understand that a bad or good convergence graph for ICP shows whether the match quality graph was improved because of Tagged-ICP's

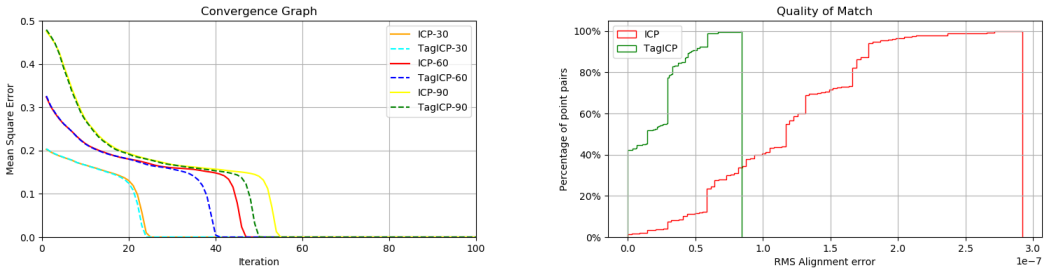

Figure 2: Same shape PDB ID: 2HAX. Convergence at angles 30, 60 and 90 degrees rotation about the x axis, and final match quality at 90 degrees

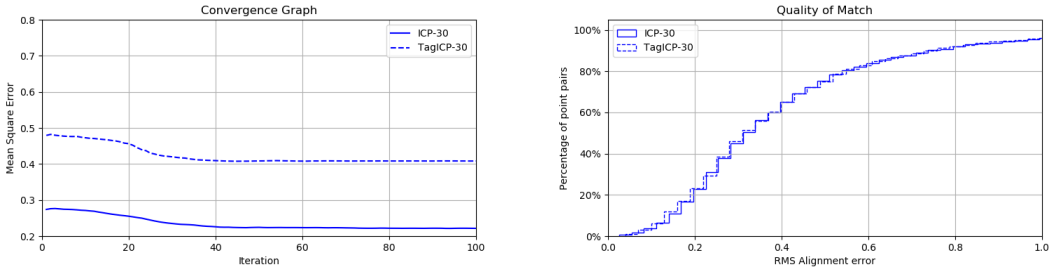

Figure 3: Mutation PDB ID: 2JKF and 2JKG. Convergence and final match quality

closest point search being used. Generating the data at these different phases allows us to study the progress of the convergence as well as the final alignment error for each correspondence.

5 Performance

Fig. 2 shows that for identical shapes, increasing angles result in a higher number of iterations required before convergence, as expected. We also see that Tagged-ICP performed better by convergence with fewer iterations than Standard ICP. Fig. 3 shows that for the mutations protein dataset (Same protein with the mutant one having single to many changes in the molecule in terms of the numbers of atoms and their positions), ICP had a better convergence quality. We can also deduce from the noise experiments (Fig. 4) that the final mean squared errors are related to the noise levels because, for each algorithm, the errors at convergence are directly proportional to the noise level so increasing noise levels shows convergence at higher mean squared error, as expected. On the other hand, the final converged errors do not correspond to the match quality graphs for ICP. That is because the convergence graphs use the algorithm's own closest points search method which gives ICP an unfair advantage . The match quality graphs show the fair comparison using the same, more meaningful closest point search method. This is because Tagged-ICP's closest points search method was used in a final iteration of the ICP which makes use of known feature descriptions to get a meaningful closest point search. Tagged-ICP's closest point search method was used in the final iteration for Standard ICP's match quality in order to compare the differences in the estimated transformations.

Similarly, the graphs show that the convergence speed is related to angle and not to the noise. We can see from the convergence graphs that the higher angles converge at much higher iteration. On the other hand, this pattern is not seen in the noise graphs. This is because the large angles of rotation make the data shape far away from the reference and thus make them differ significantly whilst the noise level is multiplied by a random distance factor that is within the diameter of the data shape.

From the graphs, we can deduce that increasing the levels of noise reduces the quality of the resulting match by increasing the final mean squared error, as expected.

6 Conclusion and Future Work

The standard ICP algorithm may not converge when rotation angles are high. The search for correspondence is also computationally expensive. We have presented an improvement to the ICP algorithm for matching 3D

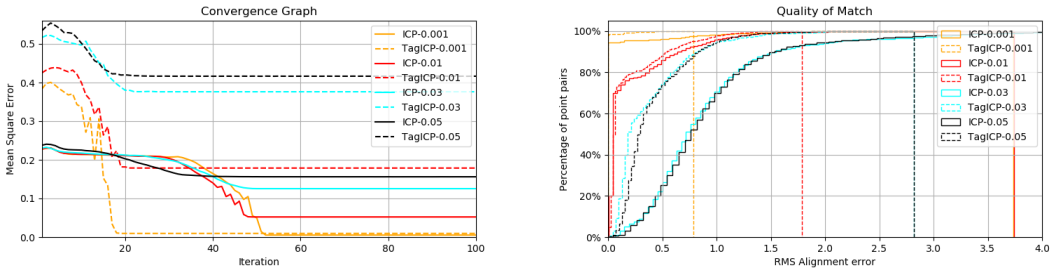

Figure 4: Noise PDB ID: 2JKF. Convergence and final match quality at varying noise levels, 0.001 to 0.05

protein structures that shows improvements over the standard ICP algorithm. Our improvement to the closest point search also demonstrates a smaller search space for searching each correspondence thus reducing the overall time and resulting complexity of the closest point search method. Some future work on Tagged-ICP include caching the partitioning of points based on tags which might further improve on speeds.

References

Angaran, S., Bock, M. E., Garutti, C., and Guerra, C. (2009). MolLoc: a web tool for the local structural alignment of molecular surfaces. *Nucleic Acids Res*, 37(suppl_2):W565–W570.

Aoki, Y., Goforth, H., Srivatsan, R. A., and Lucey, S. (2019). PointNetLK: Robust & Efficient Point Cloud Registration using PointNet. *arXiv:1903.05711 [cs]*.

Attia, M. and Slama, Y. (2017). Efficient Initial Guess Determination Based on 3D Point Cloud Projection for ICP Algorithms. In *International Conference on High Performance Computing Simulation*, pages 807–814.

Axenopoulos, A., Daras, P., Papadopoulos, G. E., and Houstis, E. N. (2013). SP-Dock: Protein-Protein Docking Using Shape and Physicochemical Complementarity. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 10(1):135–150.

Axenopoulos, A., Rafailidis, D., Papadopoulos, G., Houstis, E. N., and Daras, P. (2016). Similarity Search of Flexible 3D Molecules Combining Local and Global Shape Descriptors. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 13(5):954–970.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Res*, 28(1):235–242.

Bertolazzi, P., Guerra, C., and Liuzzi, G. (2010). A global optimization algorithm for protein surface alignment. *BMC Bioinformatics*, 11(1):488.

Besl, P. J. and McKay, N. D. (1992). A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256.

Burlacu, A., Cohal, A., Caraiman, S., and Condurache, D. (2016). Iterative closest point problem: A tensorial approach to finding the initial guess. In *2016 20th International Conference on System Theory, Control and Computing (ICSTCC)*, pages 508–513.

Chen, J. and Belaton, B. (2014). An Improved Iterative Closest Point Algorithm for Rigid Point Registration. In Wang, X., Pedrycz, W., Chan, P., and He, Q., editors, *Machine Learning and Cybernetics*, volume 481, pages 255–263. Springer Berlin Heidelberg, Berlin, Heidelberg.

Chen, Y. and Medioni, G. (1991). Object modeling by registration of multiple range images. In *1991 IEEE International Conference on Robotics and Automation Proceedings*, pages 2724–2729 vol.3.

Chen, Y. and Medioni, G. (1992). Object modelling by registration of multiple range images. *Image and Vision Computing*, 10(3):145–155.

Cignoni, P., Montani, C., and Scopigno, R. (1998). DeWall: A fast divide and conquer Delaunay triangulation algorithm in Ed. *Computer-Aided Design*, 30(5):333–341.

Combes, B. and Prima, S. (2009). Prior affinity measures on matches for ICP-like nonlinear registration of free-form surfaces. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 370–373.

Donoso, F., Austin, K., and McAree, P. (2017a). Three new Iterative Closest Point variant-methods that improve scan matching for surface mining terrain. *Robotics and Autonomous Systems*, 95:117–128.

Donoso, F. A., Austin, K. J., and McAree, P. R. (2017b). How do ICP variants perform when used for scan matching terrain point clouds? *Robotics and Autonomous Systems*, 87:147–161.

Ellingson, L. and Zhang, J. (2012). Protein Surface Matching by Combining Local and Global Geometric Information. *PLoS ONE*, 7(7):e40540.

Gehrer, D. (2015). Cellunity - an interactive tool for illustrative visualization of molecular reactions. In *19th Central European Seminar on Computer Graphics (CESCG)*, page 7.

Holm, L. and Sander, C. (1998). Touring protein fold space with Dali/FSSP. *Nucleic Acids Res*, 26(1):316–319.

Kinoshita, K. and Nakamura, H. (2005). Identification of the ligand binding sites on the molecular surface of proteins. *Protein Sci*, 14(3):711–718.

Levoy, M., Rusinkiewicz, S., Ginzton, M., Ginsberg, J., Pulli, K., Koller, D., Anderson, S., Shade, J., Curless, B., Pereira, L., Davis, J., and Fulk, D. (2000). The digital michelangelo project 3d scanning of large statues. page 14.

Lu, J., Xu, G., Zhang, S., and Lu, B. (2016). An effective sequence-alignment-free superpositioning of pairwise or multiple structures with missing data. *Algorithms Mol Biol*, 11(1):18.

Mora, H., Mora-Pascual, J. M., García-García, A., and Martínez-González, P. (2016). Computational Analysis of Distance Operators for the Iterative Closest Point Algorithm. *PLOS ONE*, 11(10):e0164694.

Nabavi, A., Mamisch, C. T., Gering, D. T., Kacher, D. F., Pergolizzi, R. S., Wells, W. M., Kikinis, R., Black, P. M., and Jolesz, F. A. (2000). Image-guided therapy and intraoperative MRI in neurosurgery. *Minimally Invasive Therapy & Allied Technologies*, 9(3-4):277–286.

Needleman, S. B. and Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3):443–453. ZSCC: 0013205.

Phillips, J. M., Liu, R., and Tomasi, C. (2007). Outlier Robust ICP for Minimizing Fractional RMSD. In *Sixth International Conference on 3-D Digital Imaging and Modeling (3DIM 2007)*, pages 427–434.

Schmitt, S., Kuhn, D., and Klebe, G. (2002). A New Method to Detect Related Function Among Proteins Independent of Sequence and Fold Homology. *Journal of Molecular Biology*, 323(2):387–406.

Schutz, C., Jost, T., and Hugli, H. (1998). Multi-feature matching algorithm for free-form 3D surface registration. In *Fourteenth International Conference on Pattern Recognition*, volume 2, pages 982–984 vol.2.

Sharp, G., Lee, S., and Wehe, D. (2002). ICP registration using invariant features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):90–102.

Shulman-Peleg, A., Shatsky, M., Nussinov, R., and Wolfson, H. J. (2008). MultiBind and MAPPIS: webservers for multiple alignment of protein 3D-binding sites and their interactions. *Nucleic Acids Res*, 36(suppl_2):W260–W264.

Thirion, J.-P. (1996). New feature points based on geometric invariants for 3D image registration. *Int J Comput Vision*, 18(2):121–137.

Unity Technologies (2019). *Unity*.

Wang, Y. and Solomon, J. M. (2019). Deep Closest Point: Learning Representations for Point Cloud Registration. *arXiv:1905.03304 [cs]*.

Weskamp, N., Kuhn, D., Hüllermeier, E., and Klebe, G. (2004). Efficient similarity search in protein structure databases by k-clique hashing. *Bioinformatics*, 20(10):1522–1526.

Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., and Xiao, J. (2015). 3d shapenets a deep representation for volumetric shapes. pages 1912–1920.

Yuan, C., Yu, X., and Luo, Z. (2016). 3D point cloud matching based on principal component analysis and iterative closest point algorithm. In *2016 International Conference on Audio, Language and Image Processing (ICALIP)*, pages 404–408.

Zhou, L., Griffith, R., and Gaeta, B. (2014). Combining spatial and chemical information for clustering pharmacophores. *BMC Bioinformatics*, 15(16):S5.