

INFORMATION POINT SET REGISTRATION FOR SHAPE RECOGNITION

Zheng Cao¹, Jose C. Principe¹, Bing Ouyang²

¹Department of Electrical and Computer Engineering, University of Florida

²Harbor Branch Oceanographic Institute, Florida Atlantic University

ABSTRACT

This paper proposes a way of enhancing shape recognition through point set registration. Firstly, a modified version of shape context (SC) is developed, which is invariant to rigid transformation and flipping. With the point correspondence obtained by the modified SC, an affine transformation based on the maximum correntropy criterion (MCC) is performed on the query shape. This point set registration could be further refined by non-rigid morphing with the minimization of Cauchy-Schwarz divergence (D_{CS}). Not only does this information theoretical learning (ITL) approach renders excellent registration result, but a new shape similarity measure can also be derived from the registration.

Index Terms— Point set registration, flipping invariant SC, MCC, shape similarity

1. INTRODUCTION

Shapes of objects are often represented by points on contour, surface or point clouds. Point set registration and shape recognition are not the same problem, yet they are related to each other. Literature that deals with point set registration alone assumes that the shapes to be registered are of the same kind (e.g. fish and distorted fish). As such, for most registration algorithms, the cost functions to be optimized do not stress the similarity between two shapes. The iterative closest point (ICP) [1] assigns binary correspondence to point sets according to the nearest distance criterion, then the average distance is reduced by least squares. Robust point matching (RPM) [2] uses a similar point correspondence except that the assignment is soft. Other works model shapes with PDF. Gaussian mixture model (GMM) is used to model the points, then the two GMMs are aligned by likelihood maximization[3] or statistical discrepancy minimization [4].

The shape context (SC) [5] is a famous descriptor in shape matching (find matching/correspondent points on two shapes). The fundamental difference between SC and point correspondence in aforementioned approaches [1][2] is that SC is solely based on the property of the shape itself, regardless of a shape's interaction with the other. Thus a similarity

measure for two shapes can be derived directly from SC [5]. The inner distance shape context (IDSC) [6] is a descriptor similar to SC in principles, but it emphasizes on articulation invariance. Both SC and IDSC are among the most popular descriptors in shape matching and subsequent shape recognition, but they are not studied as much in registration. In fact, point set registration based on such point correspondence is usually faster than PDF-based registration. It is also more likely to facilitate rigid transformation that keeps a shape's morphological properties. However, correspondences established by SC or other descriptors are not perfect. They are subject to noise, outliers and occlusion, and sensitive to local difference in shapes in varying degrees. By applying the information theoretical learning (ITL) [7] methodologies, this paper aims at mitigating such imperfection at the registration stage. The other goal of this paper is to study whether registration can help improving shape recognition, which has been usually done with shape matching only. The intuition comes from the fact that many shape similarity measures, such as L_p distance, Hausdorff distance and Frechet distance, are available only when the shapes are aligned [8].

The rest of this paper is organized as follows. Section 2 presents a modified SC which is flipping invariant. Section 3 demonstrates first the ITL-based point set registration, then discusses the criterion for evaluating shape similarity. Section 4 shows experimental results. Section 5 gives conclusions.

2. FLIPPING INVARIANT SHAPE CONTEXT

Major properties of SC include having rich representation ability, being rotational invariant and global [5]. The principle of SC is to describe any point by its relationship with all N points on the same point set. This relationship includes the distance r between two points, and the angle θ formed by the tangent line at the point and the line connecting the two points. Suppose there are n_r bins for distance and n_θ bins for angles. The cost of matching a point x_i on the query shape X and a point y_j on the template shape Y is

$$C(x_i, y_j) = \sum_{k=1}^{n_d * n_\theta} \frac{[h_i(k) - h_j(k)]^2}{h_i(k) + h_j(k)} \quad (1)$$

where h_i, h_j are the histograms of x_i and y_j . Point correspondences can then be found with the $N * N$ cost matrix using the

This work was supported by U.S. Department of Energy Grant DE-EE0006787 and HBOI foundation Special License Place Aquaculture Fund.

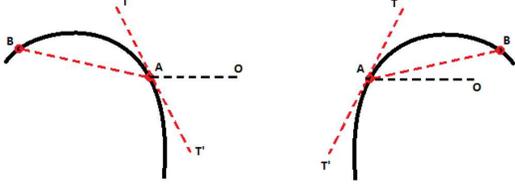


Fig. 1: Angle BAT could have different values due to flipping and tangent line direction ambiguity.

Hungarian method.

Two practical problems emerge from calculating the angle $\angle BAT$ (see Figure 1), which is done by subtracting $\angle TAO$ from $\angle BAO$. Firstly, the angle is not invariant to flipping, which is a very common affine variation of any shape. Second, there is ambiguity in the direction of tangent line (T and T'). This makes a total number of 4 possible conditions. Previous attempts to solve flipping include the SYM-FISH [9], which reorders the histogram by putting the largest values first and placing the rest with the found order (shorter path from 1st to 2nd largest value). Much information has been lost this way and the new histogram is less discernible from histograms of other points. For the same reason, applying DFT on the histogram and taking the absolute value [10] is not a good idea. There have also been efforts to eliminate the tangent line direction ambiguity. The most common approach is to force the points to be ordered in a clockwise (or counter-clockwise) manner. However, experiments suggest that techniques that find the point set's orientation (such as the one used by authors of IDSC [6]) are not always reliable.

Assume that in the left figure, $\angle BAO = \beta$ and $\angle TAO = \theta$, such that $\angle T'AO = \theta + \pi$. It follows that in the right figure, $\angle BAO = \pi - \beta$, $\angle TAO = \pi - \theta$ and $\angle T'AO = -\theta$. Therefore, the desired angle can be one of the following:

$$\begin{aligned} (i) \quad \angle BAT &= \beta - \theta & (ii) \quad \angle BAT' &= \beta - \theta - \pi \\ (iii) \quad \angle BAT &= \theta - \beta & (iv) \quad \angle BAT' &= \theta - \beta + \pi \end{aligned} \quad (2)$$

Condition (i) in (2) corresponds to the original histogram h_j . Respective histograms of conditions (ii) — inverted tangent line, (iii) — flipping, and (iv) — both of (ii) and (iii), can easily be obtained from h_j . Consider any of the n_d bins in h_j that is formed by different angles and the same distance. For (ii), all bins are circularly shifted by $n_\theta/2$. For (iii), the bins are flipped, i.e. value in any bin $(2\pi k/n_\theta, 2\pi(k+1)/n_\theta)$ is switched with the value in bin $(2\pi(n_\theta - k - 1)/n_\theta, 2\pi(n_\theta - k)/n_\theta)$. For (iv), both operations are carried out.

To determine point correspondence, one needs first to choose the best condition from (i) to (iv). (There is no necessarily the "correct" condition because the two shapes may belong to different specie/category.) With h_i and the 4 varieties of h_j , 4 cost values can be computed using (1). The cost values form 4 cost matrices for i, j ranging from 1 to N . Instead of using the costly Hungarian method, a greedy search that finds the smallest matching cost for each point of

the query shape is applied. The condition that generates least sum of the first $N/2$ values is considered the best one.

Shapes need to be preprocessed before computing SC. After being extracted from 2-D intensity image, the contour is smoothed by projecting points onto the local regression line. Tangent at any point is computed using a number (e.g. 10) of neighborhood points instead of using just its two neighbors to ensure smoothness [11]. Finally, the points are downsampled.

3. REGISTRATION AND SIMILARITY MEASURE

3.1. Affine and Non-rigid Transformation

In this paper, registration is mainly based on information shape matching [12], but with major differences. The affine and non-rigid transformation are performed in consecutive but separate steps, with each step employing a different cost function. By this practice, the transformation spaces are kept separated and calculation of transformation is simplified. Moreover, the PDF-based non-rigid transformation can bring variation to SC correspondence based affine transformation.

With SC point correspondence available, affine registration becomes a well-defined optimization problem. Any transformation performed on query shape X , now known as a set of ordered points $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$, can be denoted as $f(\mathbf{X}) = \mathbf{X}\mathbf{A}$. For 2-D data, \mathbf{X} is $N \times 3$ and the affine transformation matrix \mathbf{A} is 3×3 because of the homogeneous coordinates being used. A common practice to find a reasonable \mathbf{A} is to minimize the mean squared error (MSE) between $\mathbf{X}\mathbf{A}$ and the template point set $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^N$, but the presence of outliers (erroneous correspondence) may make MSE sub-optimal. Instead, the maximum correntropy criterion (MCC) is picked, which is more robust to outliers [13]:

$$\mathbf{A} = \operatorname{argmax} \sum_{i=1}^N G_\sigma(\mathbf{x}_i\mathbf{A}, \mathbf{y}_i) \quad (3)$$

where $G_\sigma(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{\|x-y\|^2}{2\sigma^2})$. When kernel size σ approaches infinity, effect of (3) resembles MSE. More properties and applications of MCC can be found in [13][14][15]. By taking derivative of \mathbf{A} and solving the equation one gets

$$\begin{aligned} \mathbf{D} &= \operatorname{diag}(G_\sigma(f(\mathbf{x}_1), \mathbf{y}_1), \dots, G_\sigma(f(\mathbf{x}_N), \mathbf{y}_N)) \\ \mathbf{A}_{new} &= (f(\mathbf{X})^T \mathbf{D} f(\mathbf{X}))^{-1} (f(\mathbf{X}) \mathbf{D} \mathbf{Y}) \\ \mathbf{A}_j &= \mathbf{A}_{j-1} \mathbf{A}_{new}, \quad f(\mathbf{X}) = \mathbf{X}\mathbf{A}_j \end{aligned} \quad (4)$$

Solution (4) is a fixed point update solution rather than an analytic one, but convergence can be guaranteed [16]. \mathbf{A} is set to identity for initialization. The stopping criterion is $\|\mathbf{X}\mathbf{A}_j - \mathbf{X}\mathbf{A}_{j-1}\| < 10^{-4} \|\mathbf{Y}\|$ or a maximum of 50 iterations. Kernel size σ is initialized as the mean distance of all points in the template, which is obtained during computation of SC. With changed $f(\mathbf{X})$ in every iteration, annealing is necessary to ensure fast convergence. Annealing rate is set as 0.9 for the first 20 iterations and 1 for the rest.

The non-rigid transformation can be formulated as $f(\mathbf{T}) = \mathbf{T} + \mathbf{K}\mathbf{W}$, where $\mathbf{T} = \mathbf{X}\mathbf{A}$ is from the affine step, \mathbf{K} is the $N \times N$ TPS matrix [2] computed from \mathbf{T} and \mathbf{W} is the $N \times 3$ transformation matrix to be determined. The D_{CS} is an ITL measure that describes the similarity of two PDFs. A regularized D_{CS} -based cost function can be written as

$$J = -2\log \sum_{i=1}^N \sum_{j=1}^N G_{\sigma}(\mathbf{y}_i, \mathbf{t}_j + \mathbf{K}\mathbf{W}_j) + \log \sum_{i=1}^N \sum_{j=1}^N G_{\sigma}(\mathbf{t}_i + \mathbf{K}\mathbf{W}_i, \mathbf{t}_j + \mathbf{K}\mathbf{W}_j) + \lambda * \text{tr}(\mathbf{W}^T \mathbf{K}\mathbf{W}) \quad (5)$$

The fixed point solution for \mathbf{W} that minimizes J is shown in (6).

$$\begin{aligned} \mathbf{G}_{T0}(i, j) &= G_{\sigma}(\mathbf{t}_i + \mathbf{K}\mathbf{W}_i, \mathbf{t}_j + \mathbf{K}\mathbf{W}_j), \text{ for all } i, j \\ \mathbf{G}_{Y0}(i, j) &= G_{\sigma}(\mathbf{y}_i, \mathbf{t}_j + \mathbf{K}\mathbf{W}_j), \text{ for all } i, j \\ \mathbf{G}_T &= \frac{\mathbf{G}_{T0}}{\mathbf{1}^T \mathbf{G}_{T0} \mathbf{1}}, \quad \mathbf{G}_Y = \frac{\mathbf{G}_{Y0}}{\mathbf{1}^T \mathbf{G}_{Y0} \mathbf{1}} \\ \mathbf{D}_{G.T} &= \text{diag}(\mathbf{1}^T \mathbf{G}_T), \quad \mathbf{D}_{G.Y} = \text{diag}(\mathbf{1}^T \mathbf{G}_Y) \\ \mathbf{W} &= [(\mathbf{D}_{G.Y} - \mathbf{D}_{G.T} + \mathbf{G}_T)\mathbf{K} + \lambda \mathbf{I}]^{-1} \\ &\quad * [\mathbf{G}_Y^T \mathbf{Y} - (\mathbf{D}_{G.Y} - \mathbf{D}_{G.T} + \mathbf{G}_T)\mathbf{T}] \end{aligned} \quad (6)$$

Non-rigid transformation is useful only when it is meaningful, local and small in amount. The "meaningfulness" is ensured by the good initial position brought by affine registration. Without such initialization, the transformation will fall to local minima. This applies, and is supposed to apply only to similar-looking shapes. Locality is ensured by small kernel size (0.2 times the mean distance of template, with a relatively small kernel annealing rate $\alpha=0.8$), while the transformation amount can be controlled by regularization parameter λ . Times of iteration can be set as small as 5. On average, for a pair of shapes, the affine transformation takes 10ms, whereas non-rigid transformation takes 28ms in MATLAB on a 2.4GHz dual processor Intel CPU.

3.2. Shape Similarity Criterion

When SC is applied as shape descriptor, it automatically becomes a shape similarity criterion [5], regardless of whether point set registration takes place or not. With the alignment of two shapes, it is possible that a similarity measure can benefit from such interaction. This paper uses the following similarity criterion, where larger cost indicates greater similarity:

$$\text{corr_cost}(\mathbf{X}, \mathbf{Y}) = \sum_{i=1}^N G_{\sigma}(\mathbf{y}_i, f_{\text{nonrigid}}(f_{\text{affine}}(\mathbf{x}_i))) \quad (7)$$

Correntropy is applied again not only because of its good property of being local, but also because it naturally derives from (6) (or (4) when non-rigid transformation, which is less important than affine transformation, is not performed) so no

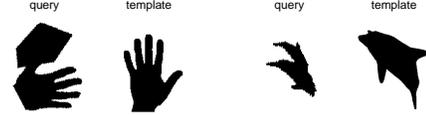


Fig. 2: Hand and dolphin - two query point sets to be registered.

additional computation is required. Advantage of this correntropy cost over SC cost is that while the majority of correspondences are correct, some others are not (e.g. ear matches with tail), and the SC cost associated with such matches are not necessarily high, as suggested by the K-cardinality assignment [17] experiment. The correntropy cost, on the other hand, is able to suppress these bad matches such that their effects are nearly negligible. A valid choice of kernel size in the cost can be the σ after annealing in affine transformation.

A similarity cost can mix with others by simple linear combination [5]. Likewise, the correntropy cost and SC cost can be heuristically combined into a new cost:

$$\text{new_cost}(\mathbf{X}, \mathbf{Y}) = \frac{\text{corr_cost}(\mathbf{X}, \mathbf{Y})}{\text{SC_cost}(\mathbf{X}, \mathbf{Y})} \quad (8)$$

Since the pair correspondence in (7) is based on SC, the correntropy cost (or the combined new cost) still retains some limitations of SC. When two points should match but are not actually matched by SC, then neither will the correntropy cost return correct cost values for them.

4. EXPERIMENTAL RESULTS

4.1. Point Set Registration

Two pairs of point sets are tested. The first pair includes a hand and another hand with occlusion. The second pair has one and two dolphins respectively, where the two dolphins are overlapped and differ in size (see Figure 2). For the proposed method, the SC parameters are set as $N=130$, $n_r=7$ and $n_{\theta}=12$, and parameters for affine transformation are set as indicated in the last section. These settings are kept the same for all other experiments. As only affine point set registration is to be compared, no non-rigid transformation is applied here.

Figure 3 shows the results. None of the other five registration approaches are as good as the proposed method. Copap [18] emphasizes on order preservation so it has trouble dealing with occlusion. SYM-FISH omits useful information, which leads to worse correspondence. Unlike MCC, MSE as a cost function is incapable of handling imperfect correspondence. Meanwhile, registration suffers from inadequate correspondence found by surprise [12] even with MCC. Rigid CPD is heavily dependent on initial position which is not desired. In both hand and dolphin cases, these five approaches will still render bad result even if the query image is manually flipped. Note that SC is actually more susceptible to occlusion than local descriptors such as the turning angle [19]. Yet the combination of SC and MCC has made up for this weakness.

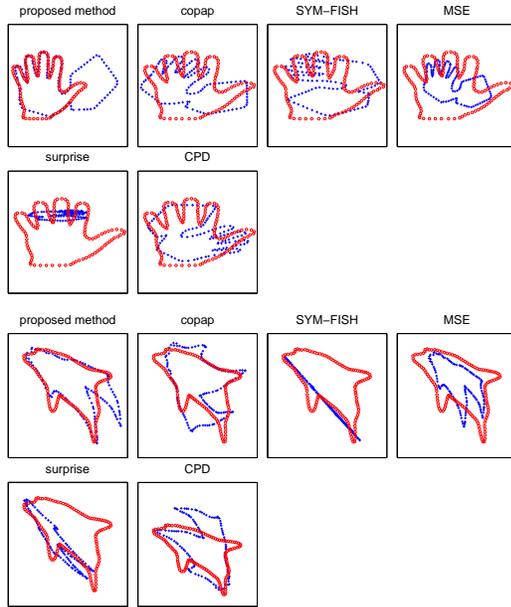


Fig. 3: For either hand or dolphin, all approaches in the first row adopt SC or modified SC as shape descriptor. The "surprise" (5th plot) is a different point correspondence criterion, while the last plot shows CPD, a popular approach that is not based on correspondence.

Table 1: Shape retrieval results. Last 3 rows are literature results.

| Method | 1st | 2nd | 3rd | 4th | 5th | 6th | 7th | 8th | 9th | 10th |
|---------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| SC cost | 99 | 97 | 97 | 97 | 96 | 94 | 94 | 88 | 84 | 81 |
| Corr. (affine only) | 99 | 98 | 98 | 98 | 98 | 98 | 97 | 96 | 91 | 75 |
| SC+Corr. | 99 | 99 | 98 | 98 | 98 | 98 | 98 | 95 | 94 | 81 |
| SC cost [20] | 97 | 91 | 88 | 85 | 84 | 77 | 75 | 66 | 56 | 37 |
| Shock Edit [20] | 99 | 99 | 99 | 98 | 98 | 97 | 96 | 95 | 93 | 82 |
| IDSC+DP | 99 | 99 | 99 | 98 | 98 | 97 | 97 | 98 | 94 | 79 |

4.2. Shape Retrieval: Kimia 99 dataset

The Kimia 99 dataset [20] consists of 9 different types of objects, where each type has 11 instances. When any one of the 99 instances serves as the template shape, the closest 10 matches are found. Non-rigid transformation is applied here, for which λ is set to 1. Table 1 summarizes the results as the number of top 1 to top 10 correct matches. The baseline correntropy cost (2nd row) without non-rigid transformation already produces better retrieval results than the original SC cost (1st row). Results can be further improved by introducing non-rigid transformation and using the combined cost (8). This result is slightly better than the result of Shock Edit, and comparable to that of IDSC+DP. Also notice that IDSC+DP uses many more contour points (300). The SC result in literature (4th row) is poor due to insufficient preprocessing.

4.3. Marine Animal Classification

Classification of marine animals is crucial in undersea applications such as hydrokinetic site monitoring and fishery stock

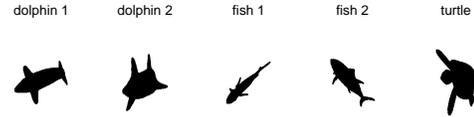


Fig. 4: Template shapes of 3 animal species.

Table 2: Marine animal classification results.

| Method | Accuracy(%) |
|---------------------|--------------|
| SC cost | 94.37 |
| Corr. (affine only) | 97.00 |
| SC+Corr. | 97.43 |
| IDSC+DP | 95.77 |

assessment. Lidar cameras mainly catch the shape information of an animal. The simulation dataset comprises of three species of animals: dolphin, fish and turtle. Objects in each specie is generated by projecting the same 3-D model onto different 2-D planes. For each specie, there are 1000 "query objects" whose labels are to be found, and another 50 "candidate template objects" with known labels. For any query object, feature is collected by pairwise comparison with every template, then 1-nearest neighbor classification is carried out. To speed up comparison, a few highly representative "mode" templates are chosen from the candidates by K-means clustering. A cluster is split if it has more than 1 specie. When $K = 3$, five templates are chosen: 2 for dolphin, 2 for fish and 1 for turtle, since more variation exists in dolphin and fish than in turtle. Chosen templates are shown in Figure 4. For non-rigid transformation, $\lambda=100$. IDSC+DP is implemented using the MATLAB code provided by its author. Its parameters are set as $N=130$, $n_r=7$ and $n_\theta=12$ (same as the proposed method), and $k=4$, $\tau=0.3$ as recommended by the author.

Classification results are shown in Table 2. Like in the retrieval experiment, the baseline correntropy cost with only affine transformation has significantly improved upon the SC cost result, and performs better than IDSC+DP. Result is even better with non-rigid transformation and combined cost (8). In addition, if part of the query objects are intentionally flipped, then performance difference between the proposed method and IDSC+DP will widen, because IDSC+DP does not have flipping invariance mechanism.

5. CONCLUSION

Initialized by a flipping invariant shape context, the optimization of the MCC function provides outstanding point set registration result. This result can be further utilized by a correntropy-based similarity measure. Shape recognition performance exceeds SC itself and matches up state-of-the-art approaches, while extra computational cost on registration is small. Future goals include developing a similarity measure for aligned shapes that is less affected by descriptor imperfection, and incorporating the registration-recognition framework with other descriptors such as IDSC.

6. REFERENCES

- [1] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, 13(2):119–154, 1994.
- [2] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 89(2):114–141, 2003.
- [3] A. Myronenko and X. Song. Point set registration: Coherent point drift. *Computer Vision and Image Understanding*, 32(12):2262–2275, 2010.
- [4] B. Jian and B. C. Vemuri. Robust point set registration using gaussian mixture models. *PAMI*, 33(8):1633–1645, 2011.
- [5] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *PAMI*, 24(4):509–522, 2002.
- [6] H. Ling and D. W. Jacobs. Shape classification using the innerdistance. *PAMI*, 29(2):286–299, 2007.
- [7] J. C. Principe. *Information Theoretic Learning, Renyis Entropy and Kernel Perspectives*. New York, NY: Springer, 2010.
- [8] R. C. Veltkamp. Shape matching: Similarity measures and algorithms. In *Shape Modeling International*, 2001.
- [9] X. Cao, H. Zhang, S. Liu, X. Guo, and M. Lin. Symfish: A symmetry-aware flip invariant sketch histogram shape descriptor. In *Proc. ICCV*, 2013.
- [10] S. Yang and Y. Wang. Rotation invariant shape contexts based on feature-space fourier transformation. In *Fourth International Conference on Image and Graphics*, pages 575–579, 2007.
- [11] J. Feldman and M. Singh. Information along contours and object boundaries. *Psychological Review*, 112(1):243–252, 2005.
- [12] E. Hasanbelliu, L. S. Giraldo, and J. C. Principe. Information theoretic shape matching. *PAMI*, 36(12), 2014.
- [13] W. Liu, P. P. Pokharel, and J. C. Principe. Correntropy: properties and applications in non-gaussian signal processing. *IEEE Transactions on Signal Processing*, 55(11):5286–5298, 2007.
- [14] B. Chen and J. C. Principe. Maximum correntropy estimation is a smoothed map estimation. *Signal Processing Letters, IEEE*, 19(8):491–494, 2012.
- [15] S. Zhao, B. Chen, and J. C. Principe. Kernel adaptive filtering with maximum correntropy criterion. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pages 2012–2017. IEEE, 2011.
- [16] B. Chen, J. Wang, H. Zhao, N. Zheng, and J. C. Principe. Convergence of a fixed-point algorithm under maximum correntropy criterion. *IEEE Signal Processing Letters*, 22(10):1723–1727, 2015.
- [17] M. Dell’Amico and S. Martello. The k-cardinality assignment problem. *Discrete Applied Mathematics*, 76(1):103–121, 1997.
- [18] C. Scott and R. D. Nowak. Robust contour matching via the order preserving assignment problem. *IEEE Transactions on Image Processing*, 15(7), July 2006.
- [19] L. Chen, R. Feris, and M. Turk. Efficient partial shape matching using smith-waterman algorithm. In *CVPRW*, 2008.
- [20] T. B. Sebastian, P. N. Klein, and B. B. Kimia. Recognition of shapes by editing their shock graphs. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(5):550–571, 2004.