A Correlation-Based Approach to Robust Point Set Registration

Yanghai Tsin[1] and Takeo Kanade[2]

[1] Siemens Corporate Research,
755 College Road East, Princeton, NJ 08540, USA
[2] Carnegie Mellon University,
5000 Forbes Ave., Pittsburgh, PA 15213, USA

Abstract. Correlation is a very effective way to align intensity images. We extend the correlation technique to point set registration using a method we call kernel correlation. Kernel correlation is an affinity measure, and it is also a function of the point set entropy. We define the point set registration problem as finding the maximum kernel correlation configuration of the the two point sets to be registered. The new registration method has intuitive interpretations, simple to implement algorithm and easy to prove convergence property. Our method shows favorable performance when compared with the iterative closest point (ICP) and EM-ICP methods.

1 Introduction

Point set registration is among the most fundamental problems in vision research. It is widely used in areas such as range data fusion, medical image alignment, object localization, tracking, object recognition, just to name a few.

One of the most effective methods in registration is correlation. In vision problems, correlation between two image patches has long been used for measuring the similarities between them. When studying discrete point sets, such as those returned by range sensors or feature detectors, however, we are given just the coordinates of a set of points. The definition of correlation is no longer directly applicable since we are given a set of geometric entities without any appearance information to compare.

Nevertheless, the presence or absence of feature points themselves tell a lot more than the coordinates of individual points. They also present the structure implied by the point sets. The simplest way of capturing such structure is to treat the feature points as binary intensity images which have only values 0 (absence) and 1 (presence). However, when noise presents, or when we have different sampling strategies in obtaining the two point sets, the binary images usually do not match.

In the following we present a technique we call kernel correlation that extends the concept of correlation to point sets. We begin by introducing kernel correlation.

2 Kernel Correlation

2.1 Definitions

Kernel correlation (KC) is defined on three levels. First, it is defined on two points. Given two points x_i and x_j, their kernel correlation (KC) is defined as

$$KC(x_i, x_j) = \int K(x, x_i) \cdot K(x, x_j) dx. \tag{1}$$

Here $K(x, x_i)$ is a kernel function centered at the data point x_i. We limit ourselves to the symmetric, non-negative kernel functions that are usually used in the Parzen window density estimation [11], instead of the broader "kernel" definition in the machine learning community. Such kernels include the Gaussian kernel, Epanechnikov kernel, tri-cubic kernel, *et. al.* [7]. In the following we use the Gaussian kernel as an example for its simplicity. The Gaussian kernel has the form:

$$K_G(x, x_i) = (\pi\sigma^2)^{-D/2} \exp\left(-\|x - x_i\|^2/\sigma^2\right). \tag{2}$$

Here $\|x - y\|$ is the Euclidean distance between two vectors x and y, and D is the dimension of the vectors.

Because the kernel functions we adopt are symmetric, it's not surprising to see that the KC defined in (1) is a function of distance between the two points. For example, the KC corresponding to the Gaussian kernel is,

$$KC_G(x_i, x_j) = (2\pi\sigma^2)^{-D/2} \exp\left\{-\|x_i - x_j\|^2/2\sigma^2\right\}. \tag{3}$$

KC's for other kernels can be shown to be functions of distance $\|x_i - x_j\|$ as well. For clarity of the presentation we will not list them here. But we will discuss their shared properties with the Gaussian kernel whenever necessary. Right side of (3) is known in the vision community as "affinity" or "proximity"[17]: a closeness measure. In this paper we show its utility in registration problems.

Next we define the KC between a point and the whole set of points $\mathcal{X}$, the *Leave-one-out Kernel Correlation* (LOO-KC),

$$KC(x_i, \mathcal{X}) = \sum_{x_j \neq x_i} KC(x_i, x_j). \tag{4}$$

According to (3) and (4), for the Gaussian case we have

$$KC_G(x_i, \mathcal{X}) = (2\pi)^{-D/2} \sum_{x_j \neq x_i} \exp\{-\|x_j - x_i\|^2/2\sigma^2\} \tag{5}$$

Leave-one-out KC defines the total affinity from a point to a point set.

Finally, we extend the KC definition to a point set: the total sum of the LOO-KC of all the points x_k in the set,

$$KC(\mathcal{X}) = \sum_i KC(x_i, \mathcal{X}) = 2 \sum_{i \neq j} KC(x_i, x_j). \tag{6}$$

If the points in the set are close to each other, the KC value is large. In this sense KC of a point set is a compactness measure of the point set.

2.2 Entropy Equivalence

If we define the density of the point set $\mathcal{X}$ as the kernel density estimate:

$$P(x) = \sum_{i=1}^{N} K(x, x_i)/N, \tag{7}$$

and adopt the Renyi's Quadratic Entropy (RQE) [15] as,

$$H_{rqe} = -log \int_x P(x)^2 dx, \tag{8}$$

KC of the point set has a simple relationship with the entropy measure,

$$KC(\mathcal{X}) \propto C + \exp\{-H_{rqe}\}. \tag{9}$$

The above observation follows directly by expanding the $\int_x P(x)^2 dx$ term in the entropy definition. In fact,

$$\begin{aligned} N^2 \cdot \int P(x)^2 dx &= \left(\sum_i \int_x K(x, x_i)^2 dx + 2 \sum_{i \neq j} \int_x K(x, x_i) K(x, x_j) dx \right) \\ &= C' + KC(\mathcal{X}). \end{aligned}$$

Here we use the fact that $\int_x K(x, x_i)^2 dx$ is a constant and the definition of $KC(\mathcal{X})$ (6). Note that the relationship does not assume any specific form of kernel function, as long as the integrals are defined.

Thus the compactness measure of KC is linked to the compactness measure of entropy. A minimum entropy system is the one with the maximum affinity (minimum distance) between all pairs of points. The information theoretic compactness measure indeed has a geometric interpretation.

We were brought to the attention of the independent work by Principe and Xu [13]. They expanded the RQE definition in the Gaussian case and defined the integral of the cross product terms as "information potential". Their purpose for such decomposition is efficient evaluation of entropy and entropy gradients in the context of information theoretic learning. In contrast, our goal is instead to configure a dynamic point set.

2.3 KC as an M-Estimator

If there are just two points involved, maximum KC corresponds to minimum distance between them. However, when we are dealing with multiple points, it's not immediately obvious what is being optimized. For instance, in the Gaussian case we have (5). What does it mean to maximize KC? It turns out that in this case we are still minimizing the distance, but in the sense of M-estimators.

In an M-estimator, instead of minimizing the usual sum of quadratic distances, $E_q = \sum_j (x_i - x_j)^2$, we are minimizing a robust version of the distance

function $E_r = \sum_j g\left((x_i - x_j)^2\right)$, where g is a robust function [8]. The advantage of changing from the quadratic distance function to the robust function is that local configuration of x_i is insensitive to remote points. To see this we compare the gradients of the above two functions.

$$\partial E_q/\partial x_i \propto \sum\nolimits_j (x_i - x_j) \tag{10}$$

$$\partial KC_G(x_i, \mathcal{X})/\partial x_i \propto \sum\nolimits_j exp(-\|x_i - x_j\|^2/2\sigma^2)(x_j - x_i). \tag{11}$$

The gradient term (10) is very sensitive to outliers in that any outlier point x_j can have arbitrarily large contribution to the gradient. Remember that the gradient is the direction (and magnitude in the quadratic function case) to update x_i. To minimize E_q, estimation of x_i will be severely biased toward the outlier points. In the KC_G case, however, there is a second term $exp(-\|x_i - x_j\|^2/2\sigma^2)$ that decays exponentially as a function of distance. Consequently, remote outliers will have no influence to local E_r minimization.

When we use kernels other than the Gaussian kernel, we can still have the M-estimator equivalence when maximizing KC. For example, by using the Epanechnikov kernel, we implicitly embedded a line process [5] into the correlation process: Points beyond a certain distance don't contribute.

Chen and Meer [3] also observed the equivalence of mode finding in a kernel density estimate and the M-estimators. The difference is that they are fitting parametric models to a set of static data (the projection pursuit example), or clustering the static point set. The introduction of KC is to robustly configure dynamic point sets.

3 Kernel Correlation for Registration

Given two finite size point sets, the *model* set $\mathcal{M}$ and the *scene* set $\mathcal{S}$, our registration method is defined as finding the parameter θ of a transformation T to minimize the following cost function,

$$\mathcal{COST}(\mathcal{S}, \mathcal{M}, \theta) = -\sum_{s \in \mathcal{S}} \sum_{m \in \mathcal{M}} KC(s, T(m, \theta)). \tag{12}$$

Notice that in the above equation each transformed model point m is interacting with all the scene points. We call (12) a *multiply-linked* registration cost function. This is in contrast to the ICP algorithm, where each model point is connected to its nearest scene point only. It can be shown that,

$$KC(\mathcal{S} \cup T(\mathcal{M}, \theta)) = KC(\mathcal{S}) + KC(T(\mathcal{M}, \theta)) - 2\mathcal{COST}(\mathcal{S}, \mathcal{M}, \theta). \tag{13}$$

$KC(\mathcal{S})$ is independent of θ. Under rigid transformation, $KC(T(\mathcal{M}, \theta))$ is also constant. This is the case because KC is a function of Euclidean distances between pairs of points (e.g. (3)). Rigid transformation reconfigures the point set as a whole and preserves the Euclidean distances between all pairs of points. Thus $KC(T(\mathcal{M}, \theta))$ is invariant. As a result $KC(\mathcal{S} \cup T(\mathcal{M}, \theta)) = C - 2\mathcal{COST}(\mathcal{S}, \mathcal{M}, \theta)$.

Due to the equivalence of KC and entropy (Section 2.2), our registration method implies finding the minimum entropy configuration of the joint point set $\mathcal{S} \cup T(\mathcal{M}, \theta)$ in the RQE sense.

By denoting the kernel density estimates (KDE) as

$$P_{\mathcal{M}}(x, \theta) = \sum_{m \in \mathcal{M}} K(x, T(m, \theta))/N, \qquad P_{\mathcal{S}}(x) = \sum_{s \in \mathcal{S}} K(x, s)/N,$$

we can show that the cost function is also proportional to the correlation of the two KDE's,

$$\mathcal{COST}(\mathcal{S}, \mathcal{M}, \theta) = -N^2 \int_x P_{\mathcal{M}} \cdot P_{\mathcal{S}} \; dx. \tag{14}$$

3.1 Convergence of a KC Registration Algorithm

It's easy to show that the cost function (12) is bounded from below. If we use gradient descent based method to minimize the cost function such that the cost function is decreasing at each step, the convergence of the cost function to a fixed point is guaranteed. Convergence properties for other registration methods, such as ICP or EM-ICP are usually difficult to study because their cost functions, defined on nearest neighbors, change from iteration to iteration as the point configuration evolves. In contrast, the KC registration function is defined globally and each step of minimization decreases the *same* cost function.

3.2 Accuracy of KC Registration

We will empirically study the accuracy of our registration algorithm in Section 5. Here we will discuss one of the simplest cases to theoretically characterize the KC registration algorithm.

Given a point set $\mathcal{M}$ and it's transformed version $\mathcal{S} = T(\mathcal{M}, \theta^*)$, a registration method should satisfy what we call the *minimum requirement for a registration algorithm*. That is, θ^* should correspond to one of the global minima of the cost function. Although this requirement seems to be trivial, we will show in our experiments that it is not met by other multiply-linked registration algorithms. Here we first give a proof that our registration algorithm meets the minimum requirement under rigid transformation. The extension to no-rigid motion is followed. We observe that

$$N^2 \int_x \left(P_{\mathcal{M}} - P_{\mathcal{S}}\right)^2 dx = N^2 \left(\int_x P_{\mathcal{M}}^2 dx + \int_x P_{\mathcal{S}}^2 dx - 2 \int_x P_{\mathcal{M}} \cdot P_{\mathcal{S}} dx \right) \tag{15}$$

$$= C + KC(T(\mathcal{M}, \theta)) + KC(\mathcal{S}) + 2 \cdot \mathcal{COST}(\mathcal{S}, \mathcal{M}, \theta). \tag{16}$$

Here C is a constant due to KC values of a point with itself. $KC(\mathcal{S})$ is independent of θ. As we discussed in the beginning of this section, $KC(T(\mathcal{M}, \theta))$ is also a constant under rigid transformation. Thus minimizing the left side of (15) is equivalent to minimizing our registration cost function. When $\theta = \theta^*$, $P_{\mathcal{M}}$ and

$P_{\mathcal{S}}$ are exactly the same and the left side of (15) is zero, the global minimum. That is, θ^* corresponds to one of the global minima of the KC registration cost function. Note that this statement is independent of the kernel functions being chosen and the kernel scale, as long as the integrals in the proof are all defined.

The KC registration framework can be extended to non-rigid transformations if we minimize a normalized KC cost function. By denoting the normalization term as $I_{\mathcal{M}} = (\int_x P_{\mathcal{M}}^2 dx)^{1/2}$ the normalized cost function is

$$\mathcal{COST}_n = -\sum_{s\in\mathcal{S}, m\in\mathcal{M}} KC(s, T(m,\theta))/I_{\mathcal{M}}. \tag{17}$$

Similar to (16), we can show that

$$N^2 \int_x (P_{\mathcal{M}}/I_{\mathcal{M}} - P_{\mathcal{S}}/I_{\mathcal{S}})^2 dx = 2N^2 + 2\cdot \mathcal{COST}_n/I_{\mathcal{S}}, \tag{18}$$

where $I_{\mathcal{S}} = (\int_x P_{\mathcal{S}}^2 dx)^{1/2}$ is independent of θ. Given that $\mathcal{S} = T(\mathcal{M}, \theta^*)$, θ^* will again be one of the global minima of the registration cost function (17), even under non-rigid transformations.

3.3 Discrete Approximation of the Registration Cost Function

In practice we don't need to enumerate each pair of model and scene points in order to evaluate the cost function (12) or (17). We can use the discrete version of (14) to approximate the registration cost. That is, we compute two discrete KDE's, $P_{\mathcal{M}}(x,\theta)$ and $P_{\mathcal{S}}(x)$ at grid points x, and use $-\sum_x P_{\mathcal{M}}(x,\theta)\cdot P_{\mathcal{S}}(x)$ to approximate the scaled cost function. Compared to the ICP or EM-ICP methods, there is no nearest neighbor finding step involved in the KC registration, which can result in significant simplification in algorithm implementation. $P_{\mathcal{S}}(x)$ plays the role of an affinity map in our algorithm. The affinity of a model point m to the scene points can be computed by correlating $K(x,m)$ with $P_{\mathcal{S}}(x)$.

4 Related Work

We store the affinity information in a density estimate. This technique bears much resemblance to the registration methods based on distance transform (DT) [2]. However, there are some important differences. First, DT is known to be extremely sensitive to outliers and noise because a single point can have influence to a large area. The influence of each point in the KC case is local. Thus KC based registration can be robust to outliers. Second, our affinity map is usually sparse for usual point sets such as an edge map or a laser range scan, with most grid points having zero values. The affinity map can be efficiently stored in data structures such as an octree. In contrast, high resolution DT in 3D is very costly. This prompted Lavallée and Szeliski to approximate 3D DT using octree spline [10].

One elegant registration method based on DT is the partial Hausdorff distance registration [9]. By minimizing partial Hausdorff distance, a registration algorithm can have up to 50 percent breakdown point. The underlying robustness mechanism is the same as the *least medium of squares* (LMedS) algorithm [16] in robust regression. However, the registration depends on a single critical point in the data set and most information provided by other points are ignored. Compared to other registration methods such as ICP and our proposed method, it is very sensitive to noise.

Scott and Longuet-Higgins [17] explored the possibility of finding correspondence by singular value decomposition (SVD) analysis of an affinity matrix, whose elements are proportional to the Gaussian KC values. Their algorithm is known to be vulnerable to perturbations such as large rotations, outliers and noise. In addition, forming a large affinity matrix for a large point set is costly.

One of the most successful point set registration algorithms is the iterative closest point (ICP) algorithm [1, 19]. A naive implementation of ICP is not robust because the cost function is a quadratic function of distance. To be robust, line-process like outlier detection or M-estimator like cost functions have been suggested [19, 4]. KC registration can be considered as multiply-linked and robust ICP. The benefits of establishing multiple-links will become clear when we compare the two algorithms in Section 5.

KC registration is mathematically most related to the EM-ICP algorithm [6] and the SoftAssignment algorithm [14], which are also multiply-linked ICP. For example, at each step, EM-ICP minimizes the following function:

$$\sum_{m\in\mathcal{M}}\sum_{s\in\mathcal{S}} exp(-\|T(m,\theta)-s\|^2/\sigma^2)\|T(m,\theta)-s\|^2/N(m,\theta). \tag{19}$$

where $N(m,\theta)=\sum_s exp(-\|T(m,\theta)-s\|^2/\sigma^2)$ is a normalization term. In fact, the KC_G cost function has the same gradient as EM-ICP, except the normalization term. Due to these mathematical similarity, KC registration and EM-ICP performs very similarly, except that EM-ICP does not meet the minimum requirement of a registration algorithm: The exactly aligned point sets does not correspond to the global minimum of the EM-ICP cost function. Depending on the point sets being registered and the kernel scale, the EM-ICP (as well as SoftAssignment) algorithms can give biased registration even for clean data. This point will be demonstrated in our experiments. For in-depth discussion on this topic, the reader is referred to our technical report [18] (pp. 56-59). In addition, the KC provides a framework for using different kernel functions for registration and its convergence proof does not rely on statistical methods such as EM.

5 Performance Evaluation

We compare the KC registration algorithm with the ICP and EM-ICP algorithm in this section. We implemented two versions of the KC algorithm. The first one is a simple 2D Matlab version that uses the Matlab "*fminsearch*" function (Nelder-Mead simplex method) for optimization. In this case the gradients are not explicitly computed. The second implementation is a C++ 3D version that computes

gradients and uses the variable metric method [12] for optimization. The 2D Matlab code is available at webpage: *http://www.cs.cmu.edu/∼ytsin/KCReg/*. The 3D ICP registration algorithm used for comparison is developed independently by the CMU 3D vision group (*http://www.cs.cmu.edu/∼3dvision/*). Both the 3D ICP and our 2D ICP code in Matlab implemented the outlier thresholding mechanism [19].

For evaluation purpose, we use a 2D range data (the *road-data*, 277 points) acquired by a SICK LMS 221 laser scanner and a 3D bunny model (the "*bunny1*"-data, 699 points) acquired by a Minolta Vivid 700 scanner. The models are shown in Figure 1. Extensive registration experiments on thousands of 2D scans and some other 3D models can be found at our website.

Fig. 1. The data set for performance evaluation.

5.1 Convergence Region

We first test the convergence properties of ICP, EM-ICP and KC in 2D. Two copies of the *road data* are generated by adding different random noise. One of the copies is then rotated on its center of mass for a certain angle. We study the convergence performance of the three registration methods by registering rotated point sets at different angles. The results are shown in Figure 2. The leftmost plot shows the registration costs as functions of the rotation angle. Note that we allow full 2D Euclidean motions and the cost is a 3D function. We plot a 1D slice of the cost function for clarity. With a kernel scale of $\sigma = 15$, both EM-ICP and KC have very smooth cost functions in the whole test range. In the ICP cost function we see a lot of local minima, which correspond to the much smaller convergence region in the center plot of Figure 2. The plot shows average registration error between corresponding points after registration. The EM-ICP has a little wider convergence region than the KC registration in this data set. However, we observed constantly larger registration error in the EM-ICP case. Here we experimentally demonstrate that EM-ICP does not meet the minimum

requirement for registration. The right plot shows the average registration error as a function of the kernel scale in the noiseless case. KC registration has zero error regardless of the kernel scale.

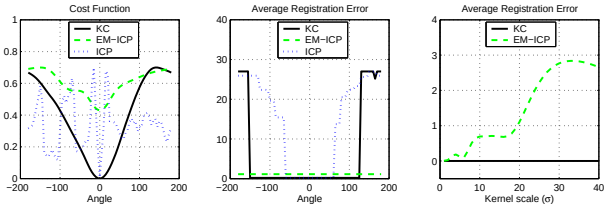

Fig. 2. 2D convergence study. The widest ICP convergence region (with varying outlier detection threshold) is shown here.

We conduct two different convergence test on the 3D data. We draw 100 random θ (6D parameter space) samples from a uniform distribution. We transform the *bunny1* data set using the random parameters and form 100 pairs for registration. The Venn diagram of the successful registrations is shown in Figure 3(a). The KC method has larger convergence region(79 versus 60, or 24 versus 5 when excluding the "easy" cases for both).

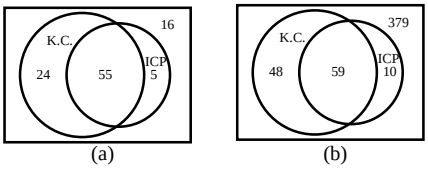

Fig. 3. Venn diagram of the sets of successfully registered model-pairs. Numbers are sizes of the regions. (a) Random transformation. (b) Pairwise.

Next, we study pairwise registration of 32 scans of the bunny model acquired from different views by the laser scanner. There are in total 496 pairs of point sets to be registered. We visually examine each of the registration results. The Venn diagram for this experiment is shown in Figure 3(b). Again, the KC method has larger success rate than ICP (107 versus 69, or 48 versus 10 excluding the "easy" cases).

Our experiments show that the KC registration method has larger convergence region. This is due to a smoothed cost function which enables an optimization algorithm to find a good registration more easily. The smoothness is provided by weighting the contributions of the multiple-links between a point and its neighbors.

5.2 Sensitivity to Noise

For both the 2D and and 3D data, we use the same method to test how sensitive the registration methods are in the presence of noise perturbation. We generate slightly rotated versions of the same point set, and add zero mean random noise to both the reference model and the rotated model. At every noise level, we register 30 pairs of noise corrupted point sets.

After registration we compute the average shift between corresponding points in the two registered point sets. If the point sets are registered well, the average shift should be close to zero because the added noise has zero mean. We can thus use the standard deviation of the average shifts over the 30 pairs as a measure of sensitivity to noise. We plot the standard deviation and average shift as a function of the noise level in Figure 4. The kernel scale is 5 for the 2D tests and 20 for the 3D tests. In both (2D and 3D) cases we observe that KC registration has smaller variance than ICP. At the same time, the registration error is small.

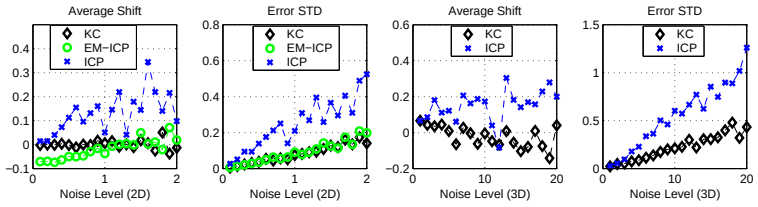

Fig. 4. Sensitivity to noise tests.

The superior capability of the KC technique can be explained by its extended interaction with neighbors. KC considered weighted effects of points in a large neighborhood, instead of just its immediate nearest neighbor, thus achieving better ability to resist noise.

5.3 Robustness

To test robustness, we register outlier corrupted point sets. We generate the reference model and the rotated model the same way as the previous section. Instead of corrupting the models with noise, we add 20% of outliers. The outlier points are randomly drawn from a uniform distribution. The corrupted point sets is illustrated in Figure 5. In both the 2D and 3D cases, we use the ICP and KC methods to register 100 pairs of outlier corrupted data.

Examples of 2D registration final results are presented in the left two plots of Figure 5. For ICP we tried three outlier-detection thresholds, 20, 5 and a concatenation of 20 and 5. The best of the three, by concatenating two ICP registrations with thresholds 20 and 5, correctly registered 43 out of 100 pairs. In contrast, KC registration robustly registered all 100 pairs.

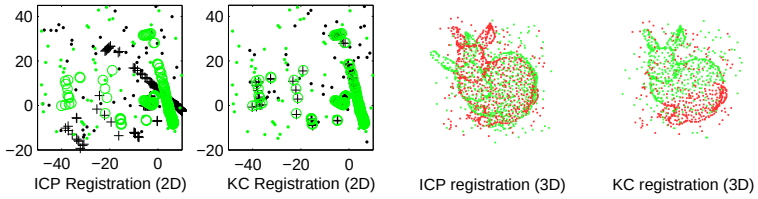

Fig. 5. Robustness tests. Note the added outlier points in all cases.

Exemplar 3D registration final results are demonstrated in the right two plots of Figure 5. The performance of the ICP algorithm is beyond our expectation. It failed only in 8 pairs of the outlier corrupted data sets. Still, KC registration can achieve better robustness. Again, KC registered all 100 pairs without mistake, by using a large range of different kernel scales.

In our experiments with KC registration in 3D, we do observe failed cases when the scale is either too small (easily fell victim to outlier distractions) or too large (containing too many outliers in the neighborhood). Thus in the presence of outliers, proper scale selection is an important and open issue in our technique.

In our experiments there are two kinds of outliers. First, the points that have a large distance to all the model points. These points are taken care of by both the M-estimator mechanism of KC, and the distance thresholding of ICP. Second, the points that fall in the neighborhood of a model point. These points can be very distracting to singly-linked methods such as ICP. For KC, each point is connected to multiple points. As long as the percentage of outliers in the local neighborhood is small, their influence can be averaged out by contributions from other inlier points. Consequently, KC is capable of registering despite of these local distractions.

6 Conclusions

In this paper we introduced a registration method by dynamically configuring point sets, whose fitness is measured by KC. KC is shown to be an M-estimator. KC is also equivalent to an entropy measure.

KC based registration can be considered as a robust, multiply-linked ICP. It has a built-in smoothing mechanism that makes it very important in dealing with noise and outlier corrupted data sets. We experimentally demonstrated that it outperforms ICP in terms of convergence region, robustness and resistance to noise, and it outperforms EM-ICP in terms of registration accuracy.

Kernel function selection is an interesting direction. The choice of kernels determines the underlying robust function to be used. We leave it to our future research.

Acknowledgment

We thank Daniel Huber and Bob Chieh-Chih Wang from Carnegie Mellon for providing the 3D and 2D test data used in this work.

References

1. P. J. Besl and N. D. McKay. A method for registration of 3-D shapes. *IEEE TPAMI*, 14(2):239–256, February 1992.
2. G. Borgefors. Hierarchical chamfer matching: A parametric edge matching algorithm. *IEEE TPAMI*, 10(6):849–865, November 1988.
3. H. Chen and P. Meer. Robust computer vision through kernel density estimation. In *(ECCV'02)*, pages Part I, 236–250. Springer-Verlag, May 2002.
4. A. Fitsgibbon. Robust registration of 2D and 3D point sets. In *BMVC'01*, 2001.
5. S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE TPAMI*, 6:1721–741, 1984.
6. S. Granger and X. Pennec. Multi-scale EM-ICP: A fast and robust approach for surface registration. In *(ECCV'02)*, June 2002.
7. T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning, Data Mining, Inference and Prediction*. Springer, 2001.
8. P. J. Huber. *Robust Statistics*. John Wiley & Sons, New York, New York, 1981.
9. D. P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge. Comparing images using the hausdorff distance. *IEEE TPAMI*, 15(9):850–863, September 1993.
10. S. Lavallée and R. Szeliski. Recovering the position and orientation of free-form objects from image contours using 3-D distance maps. *IEEE TPAMI*, 17(4):378–390, April 1995.
11. E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.
12. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, England, second edition, 1992.
13. J. Principe and D. Xu. Information-theoretic learning using Renyi's quadratic entropy. In *First International Workshop on Independent Component Analysis (ICA '99)*, pages 407–412, 1999.
14. A. Rangarajan, H. Chui, and F.L. Bookstein. The softassign procrustes matching algorithm. *Information Processing in Medical Imaging*, pages 29–42, 1997.
15. A. Rényi. On measures of entropy and information. In *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, volume 1, pages 547–561. University of California Press, 1961.
16. P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection*. Wiley-Interscience, New York, New York, 1987.
17. G.L. Scott and H.C. Longuet-Higgins. An algorithm for associating the features of two images. *Proceedings: Biological Sciences*, 244(1309):21–26, April 1991.
18. Y. Tsin. Kernel correlation as an affinity measure in point-sampled vision problems. *Techical Report, CMU-RI-03-36*, 2003.
19. Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *IJCV*, 13(2), 1994.