

# 國立交通大學

資訊科學與工程研究所

## 碩士論文

使用核關聯之非剛性形體對齊與對應

Non-Rigid Shape Registration Using Kernel Correlation

研究生：林育右

指導教授：莊榮宏 教授

黃世強 教授

中華民國 101 年 12 月

使用核關聯之非剛性形體對齊與對應  
Non-Rigid Shape Registration Using Kernel Correlation

研究生：林育右

Student：Yu-Yu Lin

指導教授：莊榮宏

Advisor：Jung-Hong Chuang

黃世強

Sai-Keung Wong

國立交通大學  
資訊科學與工程研究所  
碩士論文



A Thesis

Submitted to Institute of Computer Science and Engineering

College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

December 2012

Hsinchu, Taiwan, Republic of China

中華民國 101 年 12 月

# 使用核關聯之非剛性形體對齊與對應

研究生：林育右

指導教授：莊榮宏 博士

黃世強 博士



## 摘要

我們提出一個非剛性形體間之對齊與對應的方法。許多基於形變的方法改進迭代最近點法，並且將形體之對齊與對應的問題轉化為最佳化問題。然而，基於迭代最近點法的非線性能量系統必須在每次迭代時改變對應的最近點，並且移除當中被認為是不好的對應。此項行為造成最佳化的過程中改變了能量系統，因此最佳化處理時無法直截了當地求解。相反地，我們使用基於核關聯的方法來表示非線性能量系統，此法間接地給了動點一個方向，並且保證最佳化過程中擁有固定的數學表示式。我們的演算法不將每個形變物表上的取樣點對應到目標物表上的另一個取樣點，而是將其對應到目標物表上的一個合理位置，因此獲得更合意的對齊與對應結果。此外，我們的演算法因擁有固定的數學表示式而能更有效率地求出最佳解。

# Non-Rigid Shape Registration Using Kernel Correlation

Student: Yu-Yu Lin

Advisor: Dr. Jung-Hong Chuang

Dr. Sai-Keung Wong

Institute of Computer Science and Engineering

College of Computer Science

National Chiao Tung University



We present an algorithm for shape registration of non-rigid partial scans. Many deformation-based methods adapt their algorithm based on iterative closest point (ICP) and formulate the registration as an optimization problem. However, non-linear energy systems based on ICP should change the set of closest points iteratively and remove some of them for filtering out bad correspondence. This behavior changes the formulation during energy minimization and the optimization process can not be solved straightforwardly. On the contrary, we formulate the energy system using kernel correlation (KC), which implicitly gives a direction for a moving point and guarantees a fixed formulation during optimization. Our algorithm gains a more preferable result since each point on the source surface is not fitted to a point but a reasonable 3D position on the target surface, and can be more efficient due to the fixed formulation of energy equations.

## Acknowledgments

I would like to thank my parents and my dear friend Yi-Ting Lin.



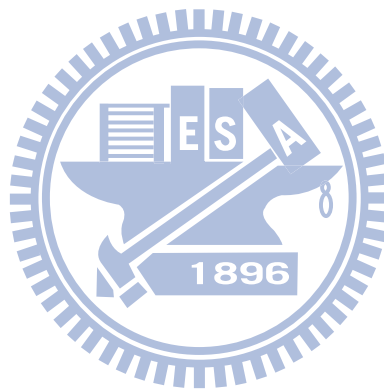
---

# Contents

---

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Related Work</b>	<b>4</b>
2.1	Rigid Registration . . . . .	4
2.2	Non-Rigid Registration for Range Images . . . . .	5
2.2.1	Template-Based Registration . . . . .	6
2.2.2	Registration Without Using A Template . . . . .	7
2.2.3	Non-Rigid Registration by Deformation . . . . .	8
2.2.4	Non-Rigid ICP-Based Energy System . . . . .	11
2.3	A Correlation-Based Approach for Point Set Registration . . . . .	12
<b>3</b>	<b>Non-Rigid Registration Using Kernel Correlation</b>	<b>13</b>
3.1	Overview . . . . .	13
3.2	Deformation Model . . . . .	18
3.3	Kernel Correlation as A Fitting Function . . . . .	20
3.4	Solving The Optimization Process . . . . .	26
3.5	The Trusted Set of Fitting Function . . . . .	29
<b>4</b>	<b>Experimental Results</b>	<b>33</b>
4.1	Implementations and Parameters . . . . .	33
4.2	Registration Errors and Time Statistics . . . . .	41

<b>5</b>	<b>Conclusions</b>	<b>45</b>
5.1	Summary . . . . .	45
5.2	Limitations . . . . .	46
5.3	Future Works . . . . .	47
	<b>Bibliography</b>	<b>49</b>



---

## List of Figures

---

2.1	(a) An ICP displacement field may not form a smoothed vector field from the source (blue) to the target (green). (b) By applying smoothing algorithm on displacement field one cloud get a better mapping result. . . . .	9
3.1	A deformation process defines a deformation model and a fitting function that transforms the source (blue) to the target surface (green). . . . .	14
3.2	A point on the source (blue) should find the correct matching position on the target (green) in a small region (red circle) according to the assumption of high frame coherence. The extracted small region for each source point is fixed during the whole optimization process. . . . .	15
3.3	(a) An ICP-based method directly selects its closest point on the target as the reference point. (b) A kernel-correlation-based approach considers all targets within its kernel. The blue and green points stand for source and target surface, respectively. . . . .	16



3.4	The pipeline of our non-rigid registration process. Given the input range images and the corresponding surface template, our algorithm transforms the template into the space of range image at each time step. . . . .	17
3.5	The kernel correlation between a point and a point set sums up the values of each point sample on the kernel function centered at $s$ . . . . .	21
3.6	The moving point $s$ (blue) will reaches an extreme position where the gradient of the $KC(s, T)$ is zero and the $KC(s, T)$ is maximized. . . . .	22
3.7	When the distance is larger then 5 times the deviation $\sigma$ , the weight is almost zero. . . . .	22
3.8	Different kernel sizes and initial positions can result in different extreme positions. . . . .	23
3.9	Point set registration using different kernel sizes. . . . .	25
3.10	(a) Point samples in 1D space, each 3 adjacent points form a neighborhood. (b) Point samples in 2D space, each 3 by 3 adjacent points form a neighborhood. (c) Point samples in 3D space, each 3 by 3 by 3 adjacent points form a neighborhood. . . . .	26
3.11	The maximal value of $KC(s, T)$ tends to appear in the densest region of point samples as shown in (a), compared to the region shown in (b). . . . .	28
3.12	An approximated bound value of kernel correlation $U_t = \overline{KC}(t, T)$ is calculated at each $t \in T$ . The red line represents the function $\overline{KC}(t, T)$ . The approximated bounding value $U$ is selected from the maximal $U_t$ . . . . .	28

3.13	(a) The movement of the point $s$ at each time step. (b) The consecutive movements of $s$ forms a trajectory. (c) The trusted set $T_s$ is the union of point sets inside the regions with kernel size of the moving point $s$ along the trajectory. . . . .	30
3.14	(a) The region $M_s$ covered by red ellipse is the set that we assume the moving point $s$ will find a matched point on $T$ according to the assumption of high frame coherence. (b) Based on $M_s$ and the effective size of the kernel function, we can find a trusted set $T_s \subseteq T$ for the point $s$ such that all points in $T_s$ will affect $s$ effectively during the entire optimization procedure. . . . .	31
3.15	(a) The moving point $s \in S$ is assumed to move within a small distance $r_{move}$ to find a matched point on the target $T$ within $M_s$ . (b) Based on $M_s$ , the trusted set $T_s$ for the moving point $s$ is bounded approximately by the sphere with radius $r_{trust} = r_{move} + r_{kernel}$ . . . . .	32
4.1	(a) The selected range image represented in mesh form. (b) A clean surface edited from the selected range image. (c) A hole filled surface. (d) A smoothed surface. (e) A simplified surface with opened mouth. . . . .	34
4.2	(a) Generate graph nodes by uniform sampling. The yellow lines partition the xy-plane into uniform grids and the blue points stand for graph nodes. (b) The deformation graph for the template. Blue lines show the connectivities of the graph. . . . .	35

4.3	Registration with $w_{fit} = 1$ , $w_{rotate} = 100$ , and $w_{regular} = 10$ . The mouth region of the template is obviously scaled and sheered. In fact, the whole region of the template is scaled and sheered. . . . .	36
4.4	Registration with $w_{fit} = 1$ , $w_{rotate} = 10^6$ , and $w_{regular} = 10$ . The template is not scaled and sheered, but is stretched when the registration process goes farther. . . . .	37
4.5	A too large kernel size leads the template to shrink. . . . .	38
4.6	A too small kernel size leads the template to bend and distort. . . . .	39
4.7	(a) A general solver that takes a set of initial affine transformations $A_0$ as input and produces $A^*$ as output, which contains the desired transformation of the template. (b) An adaptive kernel size strategy that contains an extra step of guessing the initial point for the non-linear system. <sup>8</sup> . . . . .	40
4.8	Non-rigid registration results of the face data set. . . . .	41
4.9	Non-rigid registration results of the hand data set. . . . .	41
4.10	Registration errors of the face data set. . . . .	43
4.11	Registration errors of the hand data set. . . . .	43
5.1	Image (a) and (b) are adjacent frames. (c) is the registration result of the latter frame. . . . .	46
5.2	(a) The surface template before doing registration. (b) A set of graph nodes are selected and each node is given a motion vector as an initial guess for the non-linear system. (c) The adjusted registration result. . . . .	47

- 5.3 (a) Kernel correlation with isotropic kernel function may lead a source point to a wrong 3D position. Anisotropic kernel can be designed according to either the source surface (b), or the target surface (c). . . . . 48
- 5.4 If we have a guess of matched position on the target surface for the source point as shown in (a), then the anisotropic kernel should be designed according to the target surface (b). . . . . 49



# CHAPTER 1

---

## Introduction

---

Shape registration is a fundamental task for geometry and motion reconstruction. Recent acquiring systems provide partial scanned surface data with high spatial and temporal coherence, but tracking of the surfaces is remaining a challenge due to the noise and partial overlap of the input range images. We consider the problem of tracking a sequence of input range scans under the assumption of high frame coherence.

A number of deformation-driven approaches have been proposed for the registration problem. Such approaches define a deformation model as well as a shape fitting function that provides a mapping between two scanned surfaces and solve the problem by energy minimization. Most proposed methods adapt iterative closest point (ICP) as the fitting function to explicitly map a point on the deforming source to the closest point on the target surface. In real case of registration problem, the scanned range images are usually composed of point samples, and mapping a point on the source to a point on the target is not reasonable since two sets of point samples are not always matched. Instead, two surfaces represented as point sets should be matched in the basis of point-to-position mapping. Another problem of ICP is that

---

it must filter out some closest points when they are recognized as bad correspondence, resulting in the formulation changes during energy minimization. As a result, optimization process may not be solved straightforward.

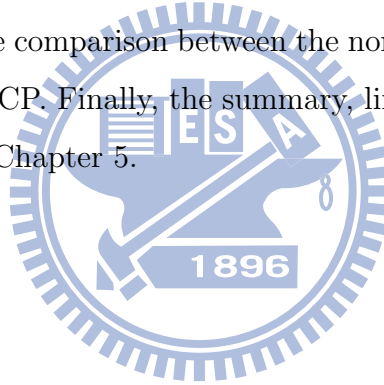
We introduce a different method that defines the shape fitting function using kernel correlation (KC) to implicitly give a moving direction for each point on the deforming source. A point on the source surface will move according to a kernel function and finally reach an extreme position on the target surface. The kernel function considers all points covered by its kernel, opposing to ICP which only considers the closest point on the target surface. The major advantage of the fitting function defined by using kernel correlation is that points in the kernel region on the target surface for each source point are unchanged through the entire optimization process. Such a formulation matches points on the deforming surface to more reasonable 3D positions on the target surface and leads to fast computation due to the fixed formulation of fitting function. It is opposite to the function defined based on ICP which explicitly matches a point on the deforming source to its closest point on the target surface and changes the set of closest points at each iterative step.

The contributions of the thesis are as follows:

- Proposes a kernel-correlation-based approach to formulate the shape fitting function. It implicitly gives a moving direction for a point on the source surface at each iteration step. The direction will change gradually over steps, which is opposite to ICP in which the direction given to a moving point may change greatly according to the change of the closest point at each iteration step. The moving point will finally be fitted to a 3D position instead of a point on the target surface, which is more reasonable for the mapping between two point samples.

- 
- The optimization process for non-rigid shape registration benefits from the fixed formulation of the shape fitting function, leading to fast computation for energy minimization procedure since the non-linear system can be solved straightforward. This is opposite to ICP-based methods since the set of the closest points is changed at each iteration, which result in different energy systems at each iteration.

The rest of the thesis is organized as follows. Chapter 2 gives a background knowledge and reviews the related works of shape registration. Chapter 3 introduces the proposed framework and the details of the registration algorithm. Chapter 4 shows the experimental results with different parameters and also depicts the comparison between the non-rigid registration using kernel correlation and ICP. Finally, the summary, limitations and the future works are discussed in Chapter 5.



## CHAPTER 2

---

### Related Work

---

In this chapter we review some methods for shape registration, including rigid registration and non-rigid registration. We will focus on methods which are driven by surface deformation.

#### 2.1 Rigid Registration

Early registration researches focus on reconstructing a complete digital model from a real world object. Such a registration process captures a sequence of partial scans of the object from different views and aligns these partial scans to a common space for reconstruction. A number of approaches have been proposed to accomplish the process [AMCO08, PB09, ART10] and we only mention about the approach using iterative closest point (ICP) [BM92].

The registration process using ICP aims to find a transformation between two partial scans  $S$  and  $T$ , and seeks to minimize an energy system by alternating between a matching step and a transformation step. The energy



system is written as

$$E(S, T, A) = \sum_{s \in S} dist(A(s), t(s)), \quad (2.1)$$

where  $A$  is an affine transformation and  $A(s)$  is the new position of  $s$  transformed by  $A$ ,  $t(s)$  is the closest point of  $s$  in  $T$ , and  $dist(s, t)$  calculates the distance between two points  $s$  and  $t$ . In each iteration, the matching step maps each point  $s \in S$  to its closest point  $t(s) \in T$  and the transformation step finds an optimal affine transformation  $A$  that minimizes the energy system  $E(S, T, A)$ . The whole energy system is optimized when some stopping criteria are satisfied.

The idea of ICP is simple and works well on registration problem. Plenty of researches have been proposed based on ICP. One main drawback of ICP is that the initial positions of two shapes must be close to get correct alignment or the registration process will reach a local minimum of the energy system.

## 2.2 Non-Rigid Registration for Range Images

Modern 3D geometry acquisition equipments such as structured light scanner and Microsoft Kinect can capture object's depth information in high frame rate and produce a sequence of range images with high spatial and temporal coherence, but with no point correspondences. The main task of non-rigid registration is to collect and align frames so that point samples between each frame become in correspondence. This is the basic step for shape reconstruction or applications that require surface tracking information. One challenge in non-rigid registration process is that the range data from 3D scanners are usually noisy and can have large regions of holes due to unreliable settings of acquisition or data occlusion. How to handle this

problem carefully and build fine correspondence relation between frames is no longer straightforward. There are several approaches for non-rigid shape registration. One class uses a template as the knowledge of underlying surface and transforms it into the space of point samples, while the others do their registration process without using template.

### 2.2.1 Template-Based Registration

Non-rigid registration using a template reduces the complexity of the process. It does not consider the problem of shape reconstruction since the template depicts the topology and geometry of the shape. In the situation of scanning the real world object, the template helps us to know the potential motion of the input range data and to capture fine registration results from the data sequence [LAGP09, LLV<sup>+</sup>12]. In addition, a template surface can naturally handle the noise and holes of the input range data, leading to a more robust registration when the noise and holes become larger, compared to the methods without using a template.

Typically, non-rigid registration using a template surface focuses on how it fits the input range data by deformation [ACP03, SSP07, LSP08]. Methods that help to find such deformation and transform the template to fit the input range data will be described in Section 2.2.3. One main drawback of this approach is that it is sometimes hard to have a proper template for the input range data and usually requires filtering to remove geometry details. Another disadvantage is that the template must be designed in case-by-case. Nevertheless, the use of template leads to a more straightforward registration since we know what the object is.

### 2.2.2 Registration Without Using A Template

Non-rigid registration without using a template means that we do not know what the input object is. How to recognize the correct shape and trace the correspondence along with shape motion is more complicated compared to the template-based registration. It must handle input data sequence with noise and holes carefully and merges the shapes from frames to form a complete shape. It should also consider the topology of the merged shape, which can not be detected instantly until the whole input data sequence is processed.

Many methods have been proposed for building such framework; for example, Wand et al. [WJH<sup>+</sup>07, WAO<sup>+</sup>09, TBW<sup>+</sup>12] used a template-like approach to build their framework by alternating between a registration step and a reconstruction step. In the registration step, they registered two adjacent frames by surface deformation. In the reconstruction step, they merged the registered frames to form a more complete surface. Popa et al. [PSDB<sup>+</sup>10] first constructed model of each frame independently and find frame-to-frame correspondence by applying cross-parameterization on successive models. Each pair of two parameterized models are merged into a new common model and the overall reconstruction is done in a bottom-up fashion. These kind of methods not only focus on registration between frames but also deal with surface reconstruction problem. The main advantage is that it provides a general framework for non-rigid shape reconstruction and registration. However, shape reconstruction step will break down when the noise and holes of the partial scans become serious. Hence, the whole process is less robust compared to the template-based methods.

### 2.2.3 Non-Rigid Registration by Deformation

An important property of the range image sequence is that it usually has high spacial and temporal coherence between frames. Many proposed methods are developed under this assumption, but have different ways finding the frame-to-frame correspondence.

One class of non-rigid registration methods use deformation-driven approach that transforms one frame of the input range data into another. Usually, this kind of approach defines a deformable model and designs a fitting function that can be found via an energy minimization process. The most popular way for defining such surface fitting is ICP. It fits two surfaces in an iterative fashion, each iteration the deformed one moves towards the target by a small step until two surfaces are close enough. Another common approach for surface fitting employs feature points, similar to the problems in correspondence finding. It usually starts with a small set of feature matchings, and then propagates the matching to the whole surface.

**ICP-Based Surface Fitting.** ICP is widely used in rigid and non-rigid registration problems. The operation of non-rigid ICP is similar to the way described in Section 2.1 except that points on the source surface  $S$  is now transformed by a set of affine transformations  $A = \{A_1, A_2, A_3, \dots\}$ , i.e., each point has its own transformation, and the new position  $A(s)$  is updated by the non-rigid transformation according to a deformation model. Various methods have been proposed for non-rigid registration problem. Li et al. [LSP08, LAGP09, LLV<sup>+</sup>12] used a template and formulate their non-rigid ICP process as an energy minimization problem. The template captures the large scale motion while fine scale geometry details is appended by synthesizing texture maps obtained from another energy minimization.

Although ICP-based algorithm handles registration problems well, it relies on some assumptions and has several drawbacks. First, the initial positions of two models to be registered should be roughly aligned first when doing ICP; otherwise, the process may be trapped at local minimum of the energy system and hence result in wrong registration result. Second, since the input range images may have noise and hole regions, the displacement mapping of two adjacent surfaces may not form a smoothing field. Papazov et al. [PB11] indicated that such cases could be refined by smoothing the displacement field. They first built an initial fitting from the closest points, and then applied smoothing algorithm to these fittings by calculating an energy minimization; see Figure 2.1.

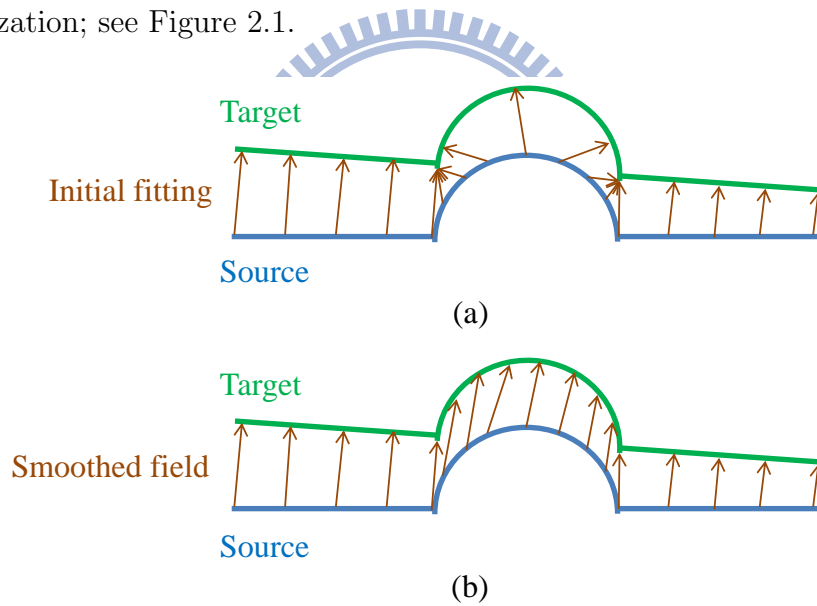


Figure 2.1: (a) An ICP displacement field may not form a smoothed vector field from the source (blue) to the target (green). (b) By applying smoothing algorithm on displacement field one cloud get a better mapping result.

A potential problem may rise in solving the ICP process. Since in each iteration the non-linear system must recompute the set of closest points, the

formulation of the non-linear system may change from iteration to iteration. Moreover, by deformation, some of the computed closest points may need to be removed since the mapping from the source to the target are recognized as bad correspondences. As a result, the whole minimization process can not be solved straightforwardly since the energy system between iterations are almost different. A more detailed description will be discussed in Section 2.2.4.

**Feature-Based Surface Fitting.** Features are a set of points on the surface. If we have matched features between frames then we can explicitly use these matching information in the registration process. The main advantage of feature-based method is that the mapping between surfaces is more preferable and robust compared to the methods depend only on local surface mapping. Typically, the fitting algorithm starts with a sparse set of matched features between frames and propagates them to the whole surface fitting. However, the process may become difficult when it tries to find feature matching between non-rigid surfaces. In addition, the computation cost for propagating the matching from sparse to dense may become expensive when the number of required matchings is large.

Features can be extracted based on pure geometry and represented as surface descriptors on points [GMGP05]. A point on the source surface is matched to the point with the most similar descriptor on the target surface. In addition to pure geometry approach, features can also be traced by resorting to image tracking technique if the related images of input range data are given. One popular image tracking method is optical flow [BHB<sup>+</sup>11, PSDB<sup>+</sup>10], in which each pixel is traced from frame to frame in 2D image space.

A hybrid method that combines feature matching and surface deformation

is often used for a robust non-rigid registration [GMGP05, TBW<sup>+</sup>12]. It first matches a sparse set of features between two frames and then registers the rest region via surface deformation.

### 2.2.4 Non-Rigid ICP-Based Energy System

By using deformation, the non-rigid registration process is usually formulated as an energy minimization problem. Non-rigid ICP is a way to define the surface fitting function which is used to map the source to the target. Li et al. [LAGP09] used non-rigid ICP for their fitting function and the equation is written as

$$E(S, T, A) = \sum_{(s, t(s)) \in C} \alpha_{point} \|A(s) - t(s)\|^2 + \alpha_{plane} |n_{t(s)}^T (A(s) - t(s))|^2, \quad (2.2)$$

where  $S$  and  $T$  stand for source and target shape, and  $A$  is a set of affine transformations  $A = \{A_1, A_2, A_3, \dots\}$  define in the deformation model,  $A(s)$  stands for the new position of the source point  $s$  according to the set  $A$ , and  $t(s) \in T$  is the closest point of  $s$  on target surface with corresponding normal  $n_{t(s)}$ , and  $C$  is a correspondence set that matches a point  $s$  to its closest point  $t(s)$ . There are two terms in the fitting function with two corresponding weighting values  $\alpha_{point}$  and  $\alpha_{plane}$ . The first one means that the new position  $A(s)$  of the source point should get closer to the closest point  $t(s) \in T$  on target, which is a typical term of ICP. The second term means that  $A(s)$  should lie on the tangent plane of  $t(s)$ , in other words,  $A(s)$  should lie on the surface of the target. The equation sums over the source points if  $A(s)$  and  $t(s)$  are in correspondence.

For surface deformation, removing points in bad correspondence is necessary since they will lead to a bad deformation. The correspondence set  $C$  may change frequently according to the computed closest points in each iteration

of the minimization process. A potential problem may arise in minimizing the energy system based on ICP that the minimization process can not be solved straightforwardly since the equations between iterations are almost different. Methods for speeding up will be useless since they are interrupted by recomputing the resources such as the linear system of each iteration.

## 2.3 A Correlation-Based Approach for Point Set Registration

A correlation-based approach was proposed in Tsin et al. [Tsi03, TK04] for rigid registration of point sets. The method is able to obtain correct results for the poses with larger deviation. The idea of correlation-based approach for shape registration is to minimize the distance of two rigid entities by maximizing a kernel function which produces a smoothed vector field over 3D space. One rigid entity on the source is moved according to the kernel function and finally matched to the target.

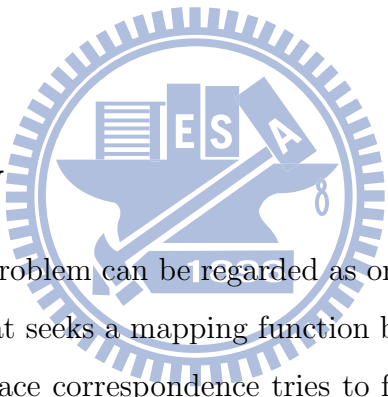
The concept between ICP and correlation-based approach is quite different. ICP directly selects one target as the reference point to indicate where a source should move to, whereas the correlation-based approach considers all targets points that have influence on the movement of the source point. The main advantage of correlation-based approach is that the objective function is smooth since the formulation of energy system is fixed over the entire minimization process; on the contrary in ICP the closest points must be changed over iterations to guarantee the convergence.



# Non-Rigid Registration Using Kernel Correlation

---

### 3.1 Overview



Non-rigid registration problem can be regarded as one kind of surface correspondence problems that seeks a mapping function between two surfaces. A deformation-driven surface correspondence tries to find a way of fitting the source surface to the target surface; as show in Figure 3.1. Such a registration process defines a deformation model and a fitting function, and is usually formulated as an energy minimization problem. The deformation model defines how a point on the source surface is moved and the fitting function drives a mapping that maps the point on the source to one on the target.

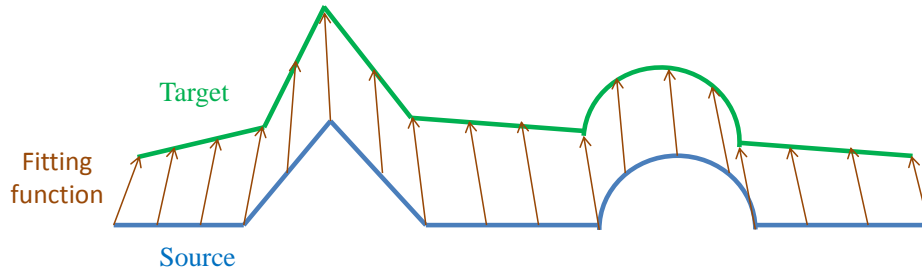


Figure 3.1: A deformation process defines a deformation model and a fitting function that transforms the source (blue) to the target surface (green).

For surface deformation, most recent methods adopt ICP-based algorithm for the fitting function [LAGP09, HCTW11]. ICP-based approach always finds the closest point in the target shape as a reference to which one point on the deforming surface moves. Such a fitting approach will change the closest points frequently in each iteration of the non-linear energy system, and the whole optimization process can not be solved straightforward. For the input range images with high spatial and temporal coherence between adjacent frames, the matched point on the target of one point on the source should be found in a small region, as shown in Figure 3.2. We wish to collect all points in the small region for that one point on source. Each point in the small region will have an affection and lead the point on source to move to the surface of target. Since the small region for each point on source is assumed to be fixed, we can write our fitting function as a fixed formulation for the non-linear energy system, opposing to ICP-based method which must recompute the closest points in each iteration.

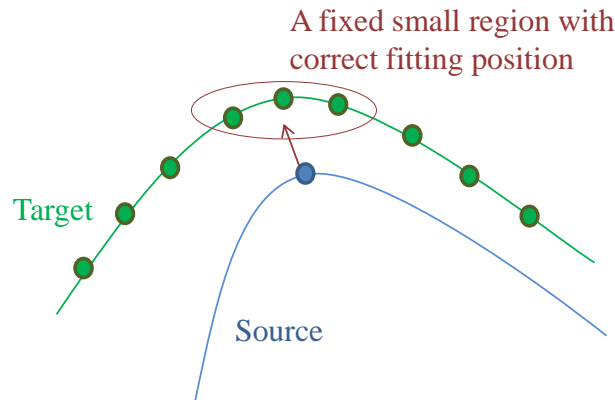


Figure 3.2: A point on the source (blue) should find the correct matching position on the target (green) in a small region (red circle) according to the assumption of high frame coherence. The extracted small region for each source point is fixed during the whole optimization process.

We use a different method, called kernel-correlation-based approach [Tsi03, TK04], as our fitting function in which each point of the deforming surface is not directly linked to its closest point, but multiply-linked to all points in a kernel as illustrated in Figure 3.3. Each point of the deforming source is associated with a kernel function and gain a moving direction at each time step according to all points in the kernel. This moving direction is changed smoothly in the movement of the source point, which is opposite to ICP-based method that one point may change its moving direction greatly according its closest point.

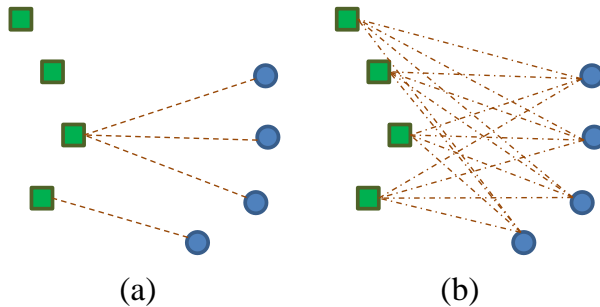


Figure 3.3: (a) An ICP-based method directly selects its closest point on the target as the reference point. (b) A kernel-correlation-based approach considers all targets within its kernel. The blue and green points stand for source and target surface, respectively.

Our non-rigid registration process by surface deformation is similar to that of Li et al. [LAGP09] and formulated as an energy minimization problem. To this end, we define our surface deformation model using deformation graph [SSP07] and the fitting function using kernel correlation. The deformation graph can be seen as a control mesh that embeds the object to deform as it deforms. The kernel-correlation fitting can be seen as a point equipped with a kernel function moving around the point samples and finally reaching an extreme position on the target surface. The formulated optimization problem is as follows: Given two point sampled sets,  $S$  and  $T$ , and a deformation graph embedded in a template, we define a set of affine transformations  $A = \{A_1, A_2, A_3, \dots, A_k\}$ , each of which represents the transformation for a node in the deformation graph. Find the set  $A$  which minimizes the energy system

$$E_{total}(S, T, A) = E_{fit}(S, T, A) + E_{constraint}(S, A), \quad (3.1)$$

where  $E_{fit}$  means the cost of fitting function that fits the deforming surface  $S$  to the target shape  $T$  and  $E_{constraint}$  is the cost that constrains the defor-

mation over the deforming surface. When the system is minimized and we get the desired non-rigid transformation from one frame to the next frame.

We consider the problem of registering a template surface and transforming it into the space of the target range images. The template depicts the shape of the input range images and is selected from one of them. Since the input range image may have noise and hole regions, we design the template by first applying hole filling and smoothing algorithm and then simplification to obtain a simplified template in order to reduce the computation cost. The main steps of the proposed non-rigid registration process is shown in Figure 3.4.

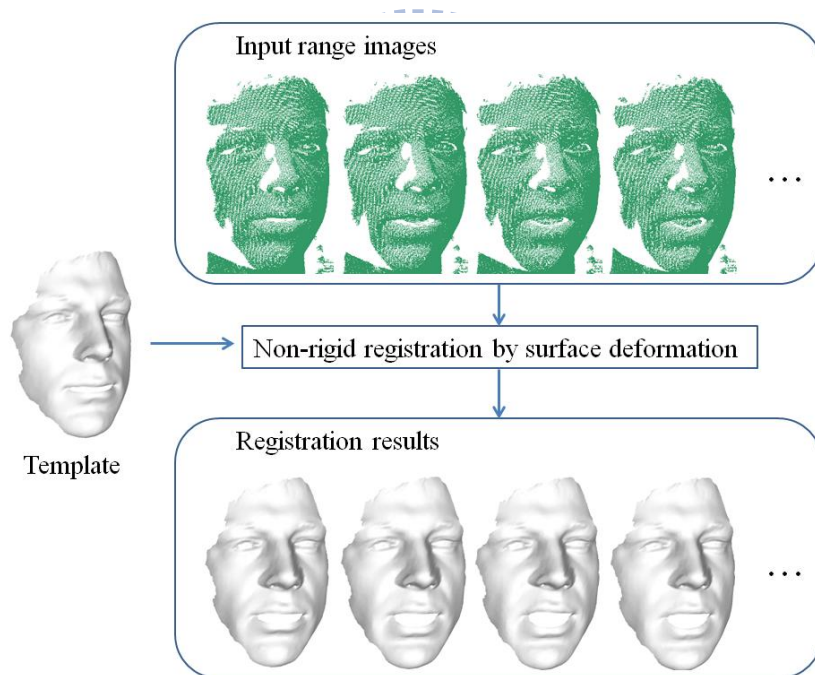


Figure 3.4: The pipeline of our non-rigid registration process. Given the input range images and the corresponding surface template, our algorithm transforms the template into the space of range image at each time step.

## 3.2 Deformation Model

In conventional surface deformation, each one point is treated as a moving entity and transform with an affine transformation [ACP03]. The main idea of deformation graph proposed in [SSP07] is that a set of points in a local region is considered as one entity and transformed by an affine transformation. That is, deformation graph can be seen as a control mesh that embeds the object to deform as it deforms. Each graph node represents a moving entity and is equipped by a 3 by 3 rotation matrix  $R$  and a 3 by 1 translation vector  $t$ . To embed deformation graph on the underlying surface, each node is also associated with a radius and affects all points on the underlying surface that are covered by the region of the radius in Euclidean space. The connectivity of deformation graph is defined by the graph edges. An edge links two nodes if the regions associated with two nodes overlap. In other words, each edge links two nodes that have influence on some points on the underlying surface.

Our deformation model adapt the deformation graph proposed in Sumner et al. [SSP07]. Each node on the deformation graph holds four variables  $\{R, t, g, r\}$  where  $R$  is the rotation matrix,  $t$  is the translation vector,  $g$  is the node position, and  $r$  is the radius. The point  $s$  on the underlying surface is transformed to  $\tilde{s}$  by some nodes in a linear combination way like radial basis functions as follows:

$$\tilde{s} = \sum_j w_j(s)[R_j(s - g_j) + g_j + t_j], \quad (3.2)$$

where  $j$  means the  $j$ 'th node and  $w_j(s)$  is a normalized weighting value that decreases when the distance between  $s$  and  $g_j$  increases. Here we use  $w_j(s) = \max(0, (1 - \|s - g_j\|^2/r_j^2)^3)$ , where  $r_j$  is the distance from the  $j$ 'th node to its  $k$ -nearest neighbors [SSP07, LAGP09] (We use  $k = 4$ ).

The use of deformation graph was proposed for shape manipulation. The surface deformation process was formulated as an optimization problem and the energy system was written as

$$E_{total} = E_{fit} + E_{rotate} + E_{regular}. \quad (3.3)$$

**User specified surface fitting.** Users can specify a set of mappings  $F = \{\{s_{k_1}, p_{k_1}\}, \{s_{k_2}, p_{k_2}\}, \dots\}$  that fits each point  $s_k$  on the underlying surface to the desired target 3D position  $p_k$ . The cost of fitting function was formulated as

$$E_{fit} = \sum_{(s,p) \in F} \|\tilde{s} - p\|^2, \quad (3.4)$$

where  $\tilde{s}$  is the new position of  $s$ . The cost function minimize the square distance between  $\tilde{s}$  and  $p$  in order to satisfy shape manipulation.

**Rotation matrix constraint.** This constraint ensures that the 3 by 3 matrix  $R$  of each node truly presents a rotation instead of other operations such as scaling or sheering.

$$E_{rotate} = \sum_j Rot(R_j), \quad (3.5)$$

where

$$\begin{aligned} Rot(R) = & (c_1 \cdot c_2)^2 + (c_2 \cdot c_3)^2 + (c_2 \cdot c_3)^2 \\ & + (1 - c_1 \cdot c_1)^2 + (1 - c_2 \cdot c_2)^2 + (1 - c_3 \cdot c_3)^2, \end{aligned} \quad (3.6)$$

and  $c_1$ ,  $c_2$  and  $c_3$  are the three columns of  $R$ . The cost function ensures that each column  $c_1$ ,  $c_2$  and  $c_3$  of a matrix  $R$  must be orthonormal. Equation 3.5 sums the square error over all nodes in deformation graph.

**Surface regularization constraint.** This constraint preserves surface regularity that the transformations of a local surface should be as rigid as possible.

$$E_{regular} = \sum_j \sum_{k \in K_j} \|R_j(g_k - g_j) + g_j + t_j - (g_k + t_k)\|^2, \quad (3.7)$$

where  $K_j$  is the neighbor set of node  $j$ . The cost function sums the square error of surface regularization over all nodes in deformation graph.

For the given set of mappings  $F$ , minimizing the energy system in Equation 3.3 results in a set of  $R_j$  and  $t_j$  for the desired deformation of the underlying surface. Our surface deformation model employs deformation graph and the cost of constraint energy in Equation 3.1 becomes as

$$E_{constraint} = E_{rotate} + E_{regular}. \quad (3.8)$$

We overwrite the surface fitting energy  $E_{fit}$  for non-rigid shape registration using kernel correlation, which will be described in the following section.

### 3.3 Kernel Correlation as A Fitting Function

Kernel correlation was proposed in Tsin [Tsi03] and Tsin et al. [TK04] for rigid registration. It measures the affinity between two shapes and gains a higher affinity if the poses of the shapes are similar. Considering the affinity measurement between two points  $s$  and  $t$ , the definition of kernel correlation using Gaussian kernel is

$$KC(s, t) = e^{-\frac{\|t-s\|^2}{2\sigma^2}}, \quad (3.9)$$

which has a higher value when the distance between two points become smaller. This implies that two points are getting closer when the affinity value is maximized. For the affinity measurement between a point  $s$  and a point set  $T$ , we define kernel correlation as

$$KC(s, T) = \sum_{t \in T} e^{-\frac{\|t-s\|^2}{2\sigma^2}}. \quad (3.10)$$

This formulation shows that a point  $s$  in 3D space is relevant to all points in the set  $T$  and the Gaussian kernel is centered at  $s$  with deviation  $\sigma$ . Notice



that we don't care about the normalized term of Gaussian function because we directly sum up the values at points instead of calculating the area under Gaussian function, as shown in Figure 3.5.

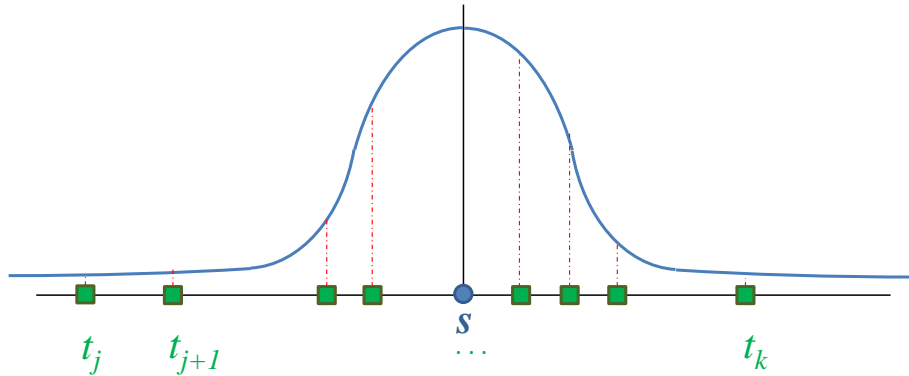


Figure 3.5: The kernel correlation between a point and a point set sums up the values of each point sample on the kernel function centered at  $s$ .

To explain the interaction between  $s$  and  $T$ , we take the first differentiation upon  $s$  and see the gradient

$$\frac{\partial KC(s, T)}{\partial s} \propto \sum_{t \in T} e^{-\frac{\|t-s\|^2}{2\sigma^2}} (t - s), \quad (3.11)$$

which is proportional to the sum of weighted directions between  $s$  and points  $t$  in  $T$ . Each point  $t$  in  $T$  implies a direction outgoing from  $s$  with a weighting value  $e^{-\frac{\|t-s\|^2}{2\sigma^2}}$ , which means that with the larger distance deviated from  $s$ ,  $t$  has the smaller weighting. Maximizing the affinity value of kernel correlation defined on a point  $s$  with respect to a point set  $T$  means moving the point  $s$  through the gradient field and finally reaching an extreme position, as shown in Figure 3.6.

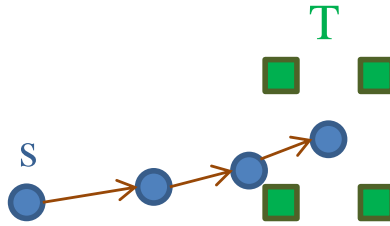


Figure 3.6: The moving point  $s$  (blue) will reach an extreme position where the gradient of the  $KC(s, T)$  is zero and the  $KC(s, T)$  is maximized.

This behavior can also be seen as a point  $s$  equipped with a kernel function moving around the point samples  $t \in T$ . At each time step,  $t$  affects  $s$  according to the kernel function and becomes ineffective to  $s$  when the distance between  $t$  and  $s$  is too large. In Figure 3.7, we see that the weighting value is below  $3.73 \times 10^{-6}$  when the distance between  $t$  and  $s$  is large then 5 times the deviation  $\sigma$ . The weighting value could be clamped to zero since it almost has no effect on the gradient. We say that each source point has an effective kernel region for the point samples in  $T$ . For Gaussian kernel, the kernel size is related to the value  $\sigma$ .

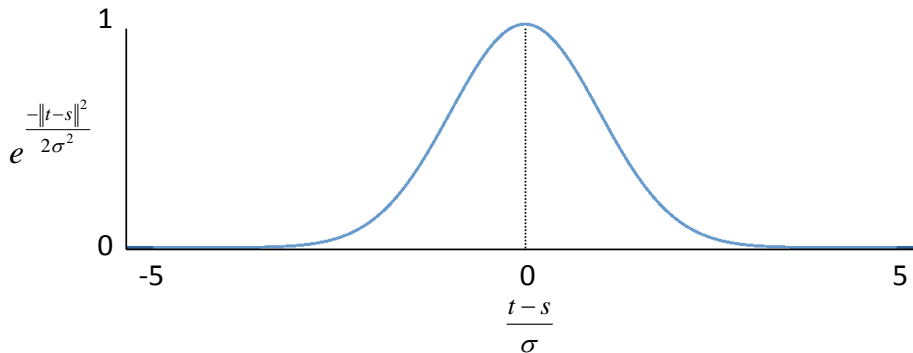


Figure 3.7: When the distance is larger than 5 times the deviation  $\sigma$ , the weight is almost zero.

Different values of  $\sigma$  can lead a moving point  $s$  to different extreme po-

### 3.3. KERNEL CORRELATION AS A FITTING FUNCTION

sitions. In Figure 3.8, point  $s$  reaches the extreme position at  $t_1$  in Figure 3.8(a), while it reaches an extreme position in the middle between  $t_1$  and  $t_2$  in Figure 3.8(b). Different initial positions of  $s$  can also result in different extreme positions. For example, the point  $s$  in Figure 3.8(c) has the same kernel size as in Figure 3.8(a) but reaches the extreme position at  $t_2$ .

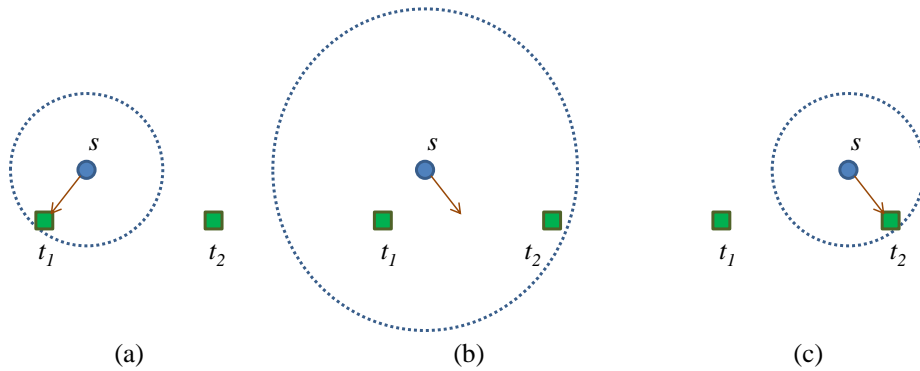


Figure 3.8: Different kernel sizes and initial positions can result in different extreme positions.

The aforementioned reveals that it is possible to have a point closer to a point set via maximizing kernel correlation when the kernel size is appropriately set up. For non-rigid registration process using kernel correlation, we equip each point  $s$  on the deforming source  $S$  with a kernel function. Each point  $s$  considers all points in its kernel and gains a moving direction according to the gradient field of the kernel function and finally reaches the extreme position on the target surface  $T$  when the kernel correlation is maximized. To define the fitting function of non-rigid registration between the deforming source  $S$  and the target  $T$ , we sum up the kernel correlation of each point in  $S$  and maximize it, or we sum up the negative kernel correlation of each

### 3.3. KERNEL CORRELATION AS A FITTING FUNCTION

---

point in  $S$  and minimize the following function:

$$E_{fit}(S, T, A) = - \sum_{s \in S} KC(s, T) = - \sum_{s \in S} \sum_{t \in T} e^{-\frac{\|t - \tilde{s}\|^2}{2\sigma^2}}, \quad (3.12)$$

where the new position  $\tilde{s}$  is transformed according to the deformation graph. The fitting function will lead each point  $s$  to its own extreme position as close as possible.

As mentioned before, different kernel sizes can result in different extreme positions for each moving point  $s$ . If kernel size is large enough to cover all points  $t \in T$ , points in  $S$  will move to extreme positions which might be close to each other, leading to a stretched surface; as illustrated in Figure 3.9(a). By appropriately setting the kernel size, we can get a better result as show in Figure 3.9(b). Conceptually, each point  $s \in S$  can have its own kernel size, but we use the same size for all  $s \in S$  for simplicity. How to set the size appropriately is currently decided by user. An important observation is that the size should be able to cover a small set of point samples for effective motion, otherwise the moving point may get stagnated due to the noneffective gradient.

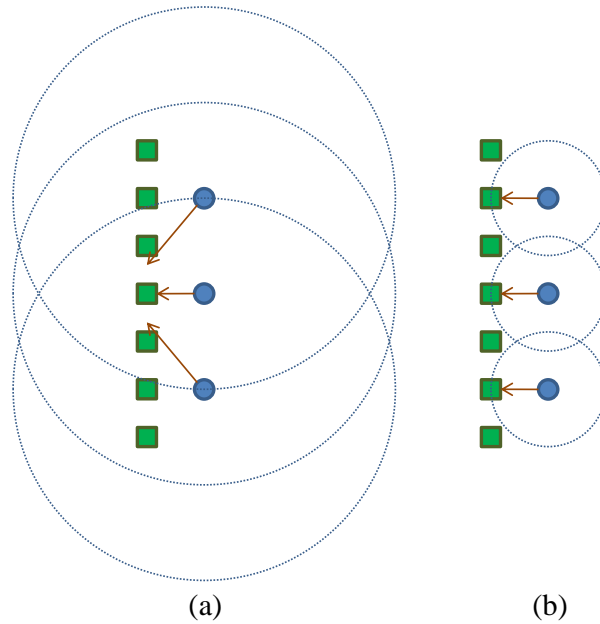


Figure 3.9: Point set registration using different kernel sizes.

A heuristic method can be used for setting the kernel size. For simplicity here we define the kernel size  $r_{kernel} = 5\sigma$ . We need a way to set appropriate  $\sigma$ . Given an arbitrary value of Gaussian deviation  $\sigma$ , one can observe how many point samples will be covered by the region with  $r_{kernel}$ . From the observation that  $r_{kernel}$  is suggested to cover a small local neighborhood on the point samples. For example, Figure 3.10 illustrates the neighborhood of point samples in 1D, 2D and 3D space. A set of points in a 3 by 3 window forms a neighborhood for range images. We tune the Gaussian deviation  $\sigma$  for different data sets such that the region with  $r_{kernel}$  covers 9 point samples. Since the exact calculation of the number of points covered by the region with  $r_{kernel}$  over Euclidean space is too computationally expensive, we approximate the value by calculating the number of points covered by the region with  $r_{kernel}$  at each point sample and extract the median number from them. Finally,  $\sigma$  is decided when the median number is 9, which means that

the region with  $r_{kernel}$  covers 9 point samples.

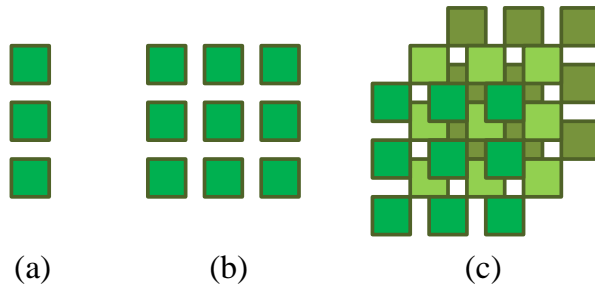


Figure 3.10: (a) Point samples in 1D space, each 3 adjacent points form a neighborhood. (b) Point samples in 2D space, each 3 by 3 adjacent points form a neighborhood. (c) Point samples in 3D space, each 3 by 3 by 3 adjacent points form a neighborhood.

### 3.4 Solving The Optimization Process

Problem statement: *Given two point sampled sets,  $S = \{s_1, s_2, \dots, s_{|S|}\}$  and  $T = \{t_1, t_2, \dots, t_{|T|}\}$ , we set up a deformation graph  $G = \{N, E\}$  representing the deformation of  $S$  and associated with a set of affine transformations  $A = \{R_1, t_1, R_2, t_2, \dots, R_{|N|}, t_{|N|}\}$ . Find the set  $A^*$  that minimizes the energy system*

$$E_{total} = w_{fit}E_{fit} + w_{rotate}E_{rotate} + w_{regular}E_{regular}. \quad (3.13)$$

This is a non-linear system which can not be trivially solved by a simple gradient descent method if we combine the fitting function with constraints. Following Sumner et al. [SSP07], the non-linear system is solved using non-linear least square problem. To do so, we have to modify our fitting function in Equation 3.12 since the non-linear least square problem minimizes an equation from a positive value to zero, but the range of Equation 3.12 is

from zero to a negative value. The modified fitting function is

$$E_{fit} = \sum_{s \in S} (U - KC(s, T))^2 = \sum_{s \in S} (U - \sum_{t \in T} e^{-\frac{\|t-\bar{s}\|^2}{2\sigma^2}})^2. \quad (3.14)$$

We add an upper bound  $U$  and take square to each term of summation over  $S$ . This bound ensures that the maximum correlation value of a moving point  $s$  will be under this bound so that  $U - \sum_{t \in T} e^{-\frac{\|t-\bar{s}\|^2}{2\sigma^2}}$  is larger than zero when minimizing the system using least square solvers. The value of bound is important and should be sufficiently tight since it affects the gradient of the non-linear system. A too large  $U$  may lead to poor convergence in optimization process, while a too small  $U$ , on the other hand, will lead the term of summation to a negative value and the optimization process may never converge to the right result.

How to evaluate the bounding value  $U$  is related to the size of the kernel function. Considering a point  $s \in S$  moving around the point samples  $t \in T$ , the maximal value of  $KC(s, T)$  may appear when  $s$  moves to the densest region of point samples, as shown in Figure 3.11. Since exact calculation of the bounding value over Euclidean space costs too much, we approximate the bound by calculating an approximated value of kernel correlation  $\overline{KC}(t, T)$  at each point sample  $t \in T$ , defined by

$$\overline{KC}(s, T) = \sum_{t \in T} e^{-\frac{1}{2}(\lfloor \frac{\|t-s\|}{\sigma} \rfloor)^2}. \quad (3.15)$$

As described above, kernel correlation has an effective region on the target  $T$ . We collect all points  $t \in T$  in the region with kernel size  $5\sigma$  and compute  $U_t = \overline{KC}(t, T)$  for each  $t \in T$ , as shown in Figure 3.12. The bounding value  $U$  used for Equation 3.14 is selected from the maximal value of  $U_t$ .

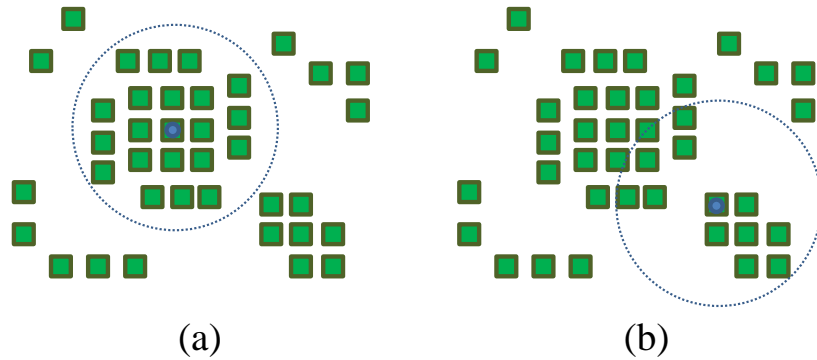


Figure 3.11: The maximal value of  $KC(s, T)$  tends to appear in the densest region of point samples as shown in (a), compared to the region shown in (b).

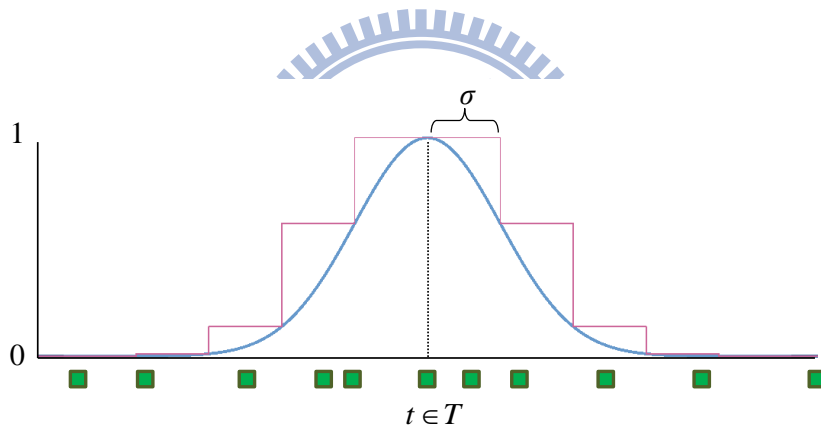


Figure 3.12: An approximated bound value of kernel correlation  $U_t = \overline{KC}(t, T)$  is calculated at each  $t \in T$ . The red line represents the function  $\overline{KC}(t, T)$ . The approximated bounding value  $U$  is selected from the maximal  $U_t$ .

Finally, we solve the non-linear least square system in Equation 3.13 by Levenberg-Marquardt algorithm [MNT04], and the number of unknown variables in the energy system is  $12n$  where  $n$  is the number of nodes in deformation graph. First, find a vector function  $f(x)$  such that  $f(x)^T f(x) =$



$w_{fit}E_{fit} + w_{rotate}E_{rotate} + w_{regular}E_{regular}$ . This vector function  $f(x)$  can be easily designed since we formulate all the terms in least square form. In our experience, using supernodal sparse Cholesky factorization for solving the linear system at each iteration of Levenberg-Marquardt is the fastest way for the entire non-linear system. The inputs to the objective function, say point samples  $T$  for each moving point  $s$  do not need to be changed through the entire optimization process, which is opposite to ICP in which the set of closest points always changes from iteration to iteration. Our non-linear system is solved straightforward and reaches fast computation compared to ICP since the system is not interrupted by recomputing the resources between iterations.

### 3.5 The Trusted Set of Fitting Function

Recall that each point  $s$  in Equation 3.14 can be seen as a point equipped with a kernel function moving around the point samples on  $T$ . The moving point  $s$  has an effective kernel region on  $T$  according to the kernel function at each time step. It is not necessary to collect all points on  $T$  for each moving point  $s$  when calculating the summation of  $KC(s, T)$  since there are many point samples on  $T$  whose distance to  $s$  is larger than  $5\sigma$  and have weights approaching to zero. These point samples with distance larger than  $5\sigma$  can be filtered out for reducing computation time.

We introduce a smaller set of point samples  $T_s \subseteq T$  for each point  $s$  in Equation 3.14 and call such set  $T_s$  a trusted set of  $s$ . For the moving point  $s$ ,  $T_s$  covers a set of point samples in  $T$  within its kernel for all time steps. The consecutive movements of  $s$  forms a trajectory; as illustrated in Figure 3.13. The trusted set  $T_s$  is the union of point sets inside the regions with kernel size

of moving point along the trajectory. However, collecting the point samples to form  $T_s$  must rely on the prediction of the movement, which can not be easily achieved.

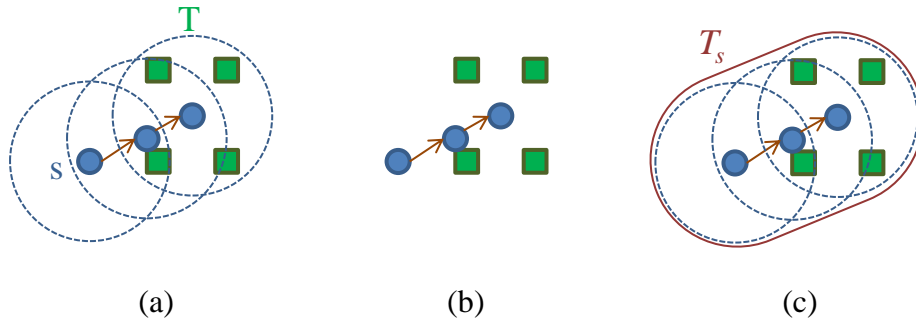


Figure 3.13: (a) The movement of the point  $s$  at each time step. (b) The consecutive movements of  $s$  forms a trajectory. (c) The trusted set  $T_s$  is the union of point sets inside the regions with kernel size of the moving point  $s$  along the trajectory.

We use an approximated method to calculate  $T_s$ . As mentioned in Section 3.1, the matched point of  $s$  should be found in a small region under the assumption of high frame coherence, which means that the consecutive movements of  $s$  can be constrained in the small region  $M_s$ . If we have such region  $M_s$  for each  $s$ , then the trusted set  $T_s$  is a bigger set expanding from  $M_s$  according to kernel correlation, as illustrated in Figure 3.14.

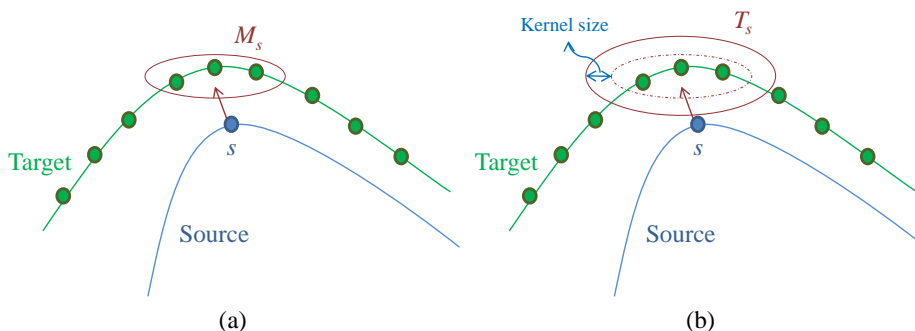


Figure 3.14: (a) The region  $M_s$  covered by red ellipse is the set that we assume the moving point  $s$  will find a matched point on  $T$  according to the assumption of high frame coherence. (b) Based on  $M_s$  and the effective size of the kernel function, we can find a trusted set  $T_s \subseteq T$  for the point  $s$  such that all points in  $T_s$  will affect  $s$  effectively during the entire optimization procedure.

Associating each  $s$  with such a trusted set posts several benefits. It reduces computation cost when summing  $KC(s, T)$  in Equation 3.14 since  $T$  is reduced to  $T_s$ , and more importantly  $T_s$  can hook the result of extreme position of the point  $s$ . We define the trusted set as follows: since the input range images have high spatial and temporal coherence between adjacent frames, the moving distance of each point  $s$  is assumed to be small. We assume that each point  $s \in S$  will move within a distance  $r_{move}$  from  $s$ . The small set  $M_s$ , which is expected to include a matched point, is defined by a spherical region centered at  $s$  with radius  $r_{move}$ , as illustrated in Figure 3.15(a). Based on  $M_s$ , the trusted set  $T_s$  is calculated by collecting the point samples bounded by the spherical region centered at  $s$  with radius  $r_{trust} = r_{move} + r_{kernel}$ , where  $r_{kernel}$  is the effective size of the kernel function, as shown in Figure 3.15(b). We set  $r_{move} = 5\sigma$  and  $r_{kernel} = 5\sigma$ .

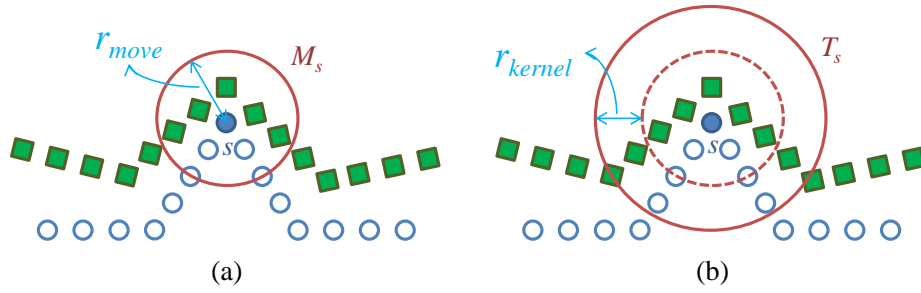
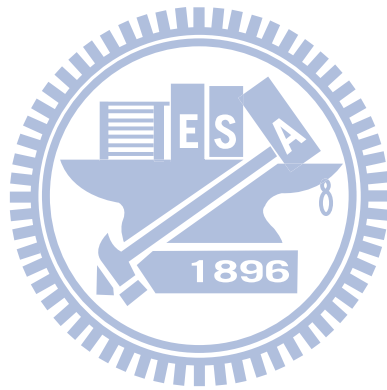


Figure 3.15: (a) The moving point  $s \in S$  is assumed to move within a small distance  $r_{move}$  to find a matched point on the target  $T$  within  $M_s$ . (b) Based on  $M_s$ , the trusted set  $T_s$  for the moving point  $s$  is bounded approximately by the sphere with radius  $r_{trust} = r_{move} + r_{kernel}$ .



# Experimental Results

---

In this chapter we first describe our implementation details and discuss the parameter settings for various experiments, and then we compare our method to ICP-based methods in terms of the registration errors and time statistics.

## 4.1 Implementations and Parameters

As illustrated in Figure 3.4, our non-rigid registration process takes a surface template and a sequence of range images as input and produces a sequence of deformed templates as output. To this end, the process needs to design a template that depicts the shape of the input range images and to generate a deformation graph that embeds the template for deformation. For energy minimization, the registration process also needs to tune some parameters for various different registration results. We use two data sets to demonstrate the implementation details and to reveal results for different parameter settings.

The surface template is selected from one of the input range images. First we edit the selected range image and keep the region which can capture the motion of the input range images. Then we apply hole filling and surface

smoothing algorithms to fill up hole regions and smooth the noisy surface. We additionally edit the mouth region to make it open and finally simplify the template for reducing the computation cost. The acquisition steps of the template is illustrated in Figure 4.1.

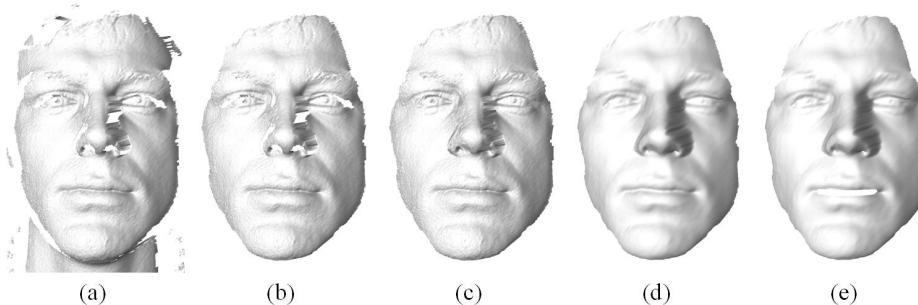


Figure 4.1: (a) The selected range image represented in mesh form. (b) A clean surface edited from the selected range image. (c) A hole filled surface. (d) A smoothed surface. (e) A simplified surface with opened mouth.

The deformation graph for the template is built as follows. Graph nodes are distributed by uniform sampling on the template. Since the template is selected from one of the input range images, we project the template on  $xy$ -plane and partition the plane into uniform grids. Each grid holds a graph node if it covers the template. A graph node is placed at the grid center with the original  $z$ -values of the template vertices in 3D space, as shown in Figure 4.2(a). We also associate each node with a radius  $r$  which is used in Equation 3.2. The value of  $r$  is given by calculating the distance from the node to its  $k$ -nearest neighbors. We use  $k = 4$ . A graph edge links two nodes that can affect the same point on the template. In other words, an edge links two nodes if the distance between them is less than  $r_1 + r_2$ , where  $r_1$  and  $r_2$  are the radii of the nodes. Figure 4.2(b) shows the deformation graph for the template. The connectivity in the nose region is complex since the radius of

each node is large due to the variations in z-values.

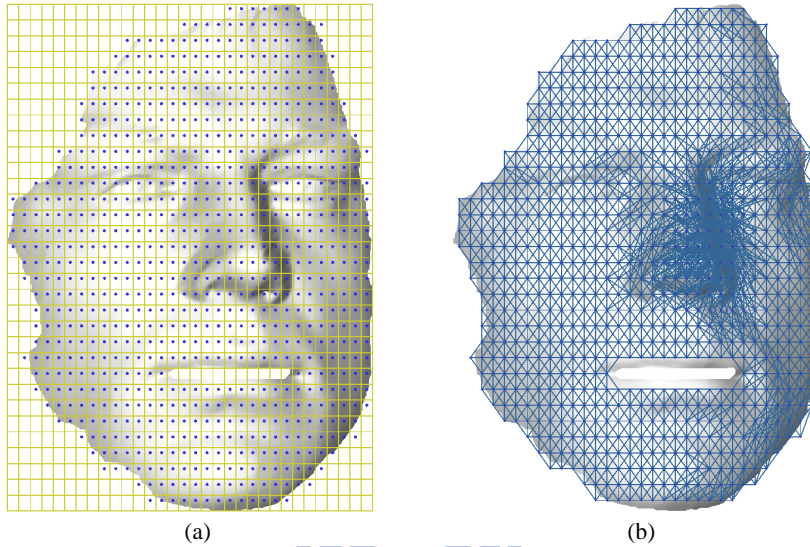


Figure 4.2: (a) Generate graph nodes by uniform sampling. The yellow lines partition the xy-plane into uniform grids and the blue points stand for graph nodes. (b) The deformation graph for the template. Blue lines show the connectivities of the graph.

For energy minimization, different parameters may lead to different registration results. We describe the affection of each parameter to our non-rigid registration process by examples.

**The weights of the energy system.** The weights  $w_{fit}$ ,  $w_{rotate}$ , and  $w_{regular}$  in Equation 3.13 should be set with care. During the optimization, the non-linear system regards the energy with largest scale as the most important term and minimizes it firstly since the gradient in the non-linear system descends most from the energy with largest scale. The non-linear system then minimizes the energy with middle scale, and so on. In our experience,  $E_{rotate}$  is the most important energy and the weight  $w_{rotate}$  should be the largest. The weights  $w_{fit}$  and  $w_{regular}$  can be set with small and middle-

scale values, respectively. We start the system with  $w_{fit} = 1$ ,  $w_{rotate} = 100$ , and  $w_{regular} = 10$ . The registration results shown in Figure 4.3 are bad since the rotation matrix constraint energy  $E_{rotate}$  can not preserve the property of a rotation matrix and the template is scaled and sheered.

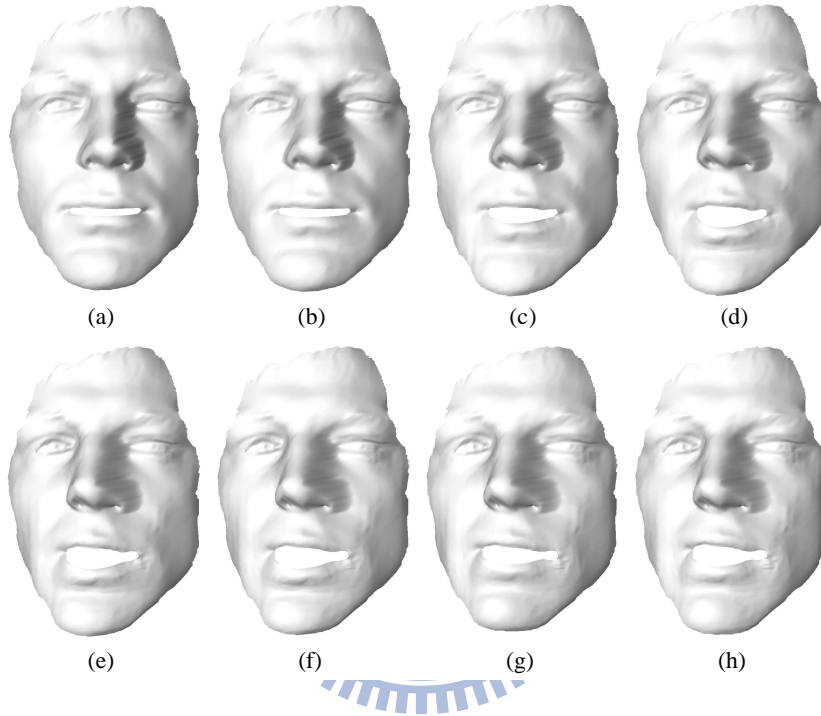


Figure 4.3: Registration with  $w_{fit} = 1$ ,  $w_{rotate} = 100$ , and  $w_{regular} = 10$ . The mouth region of the template is obviously scaled and sheered. In fact, the whole region of the template is scaled and sheered.

With the weights  $w_{fit} = 1$ ,  $w_{rotate} = 10^6$ , and  $w_{regular} = 10$ , it seems that the rotation matrix constraint energy  $E_{rotate}$  has been constrained, but the template is stretched when the registration process goes farther; as illustrated in Figure 4.4.



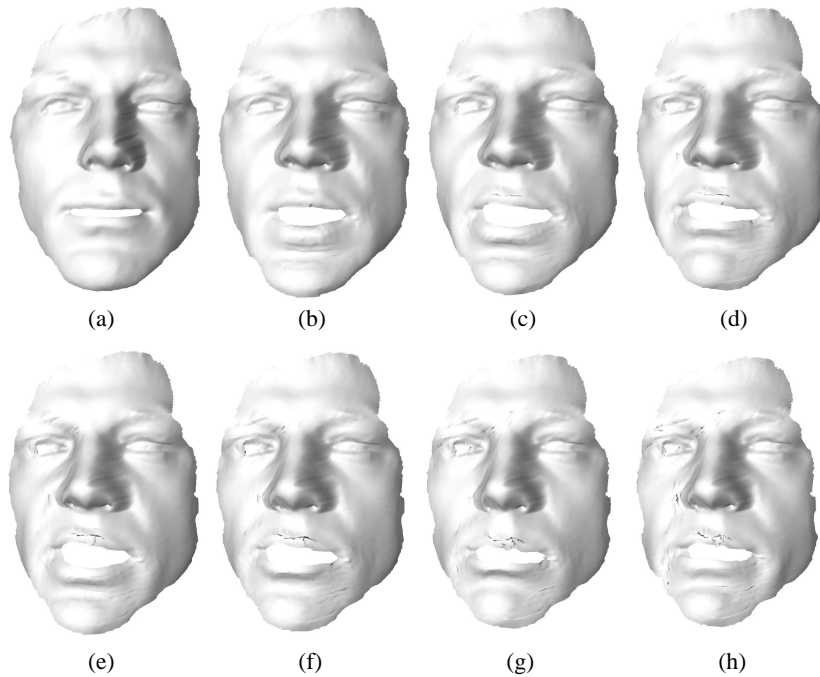


Figure 4.4: Registration with  $w_{fit} = 1$ ,  $w_{rotate} = 10^6$ , and  $w_{regular} = 10$ . The template is not scaled and sheered, but is stretched when the registration process goes farther.

With the weighting values  $w_{fit} = 1$ ,  $w_{rotate} = 10^6$ , and  $w_{regular} = 100$ , the registration result looks reasonable. Notice that this is not the only one setting for the weights. One can adjust the weights for different data sets as long as the scales of the weights are appropriately designed.

**The size of the kernel function.** This value affects the searching of extreme position for a moving point and is defined in terms of the deviation  $\sigma$  when using Gaussian kernel. Typically, an effective kernel size is  $5\sigma$  but we use  $4\sigma$  in experiments since any point sample with distance larger than  $4\sigma$  is already small enough in computation. Basically, a too large  $\sigma$  leads the template to shrink since many points on the template move to the same extreme position; as shown in Figure 4.5. On the other hand, a too small  $\sigma$

leads the template to bend and distort since many points on the template are stagnated due to noneffective motion, as illustrated in Figure 4.6. The appropriate size of  $\sigma$  is currently chosen by user and a suggestion for setting the size is described in Section 3.3.

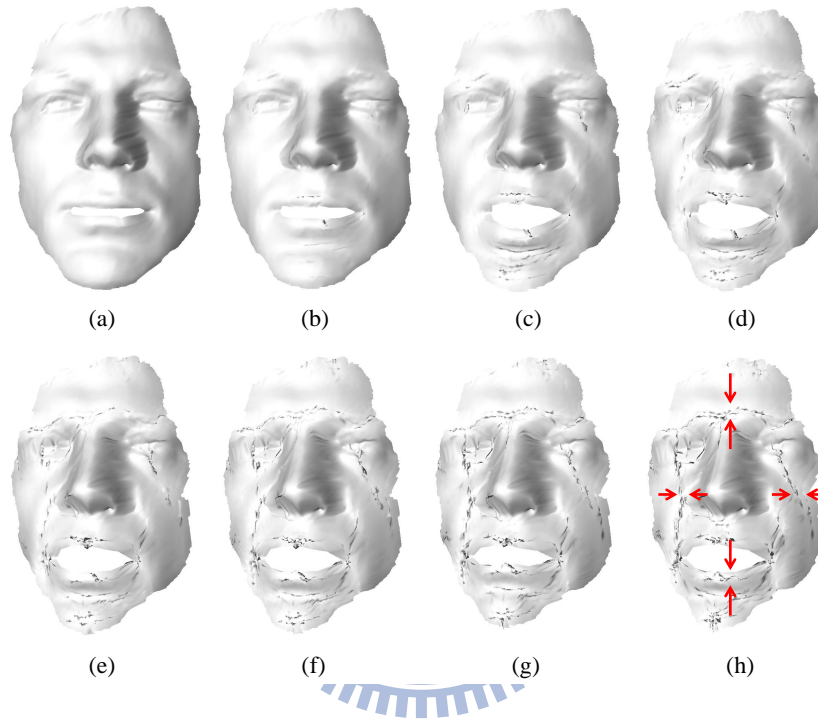


Figure 4.5: A too large kernel size leads the template to shrink.

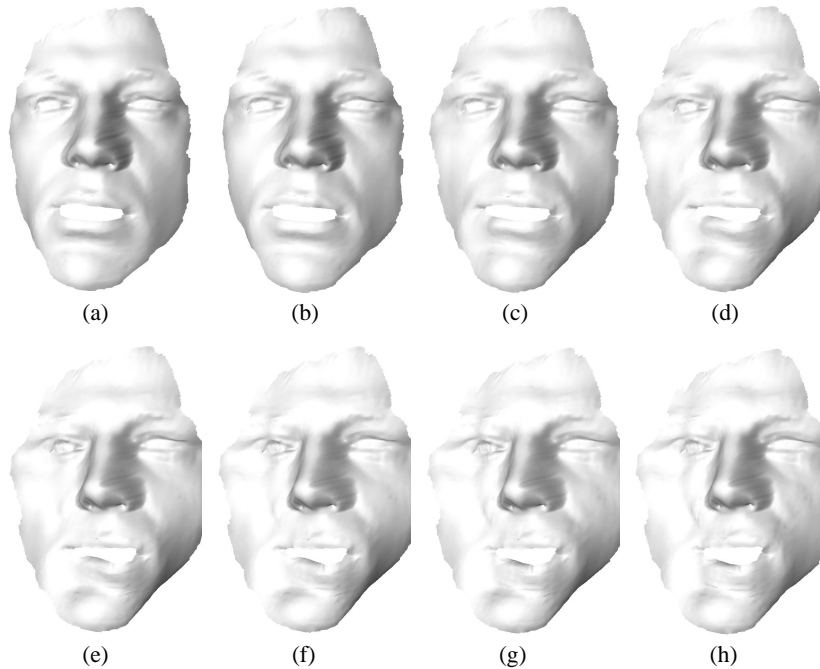


Figure 4.6: A too small kernel size leads the template to bend and distort.

**The initial point of the non-linear system.** Since we assume the input range images have high spatial and temporal coherence between adjacent frames, the initial point of the non-linear system is generally assigned with  $R = I$  and  $t = \vec{0}$  for each graph node. The kernel size  $\sigma$ , on the other hand, is expected to be small for correct registration. If the coherence between two adjacent frames becomes lower, which means the motion is large, some moving points on the template are stagnated as mentioned above. To handle this problem, we need a good initial point for the non-linear system and employ an adaptive kernel size strategy for optimization.

The adaptive kernel size strategy is illustrated in Figure 4.7. We divide the entire minimization process into several subsystems and each subsystem is solved using different parameters. In our implementation, we use two-pass optimization. The first pass is considered as an initial guess of the

initial point for the non-linear system. We solve the first pass with a larger kernel size with 10 maximum iterations. The second pass is the main step that the resultant affine transformations  $A^*$  will be used to transform the template. We solve the second pass with a smaller kernel size and with initial point that is the results of the first pass. The termination condition is  $\|A_k - A_{k+1}\| < 10^{-6}$ , where  $A_k$  is the set of affine transformations in the  $k$ -th iteration of the non-linear optimization. Notice that the adaptive kernel size strategy is not the only way to obtain a good initial point for the non-linear system. It is better to use feature matchings between adjacent frames if we have such matched features since it is more robust for shape registration.

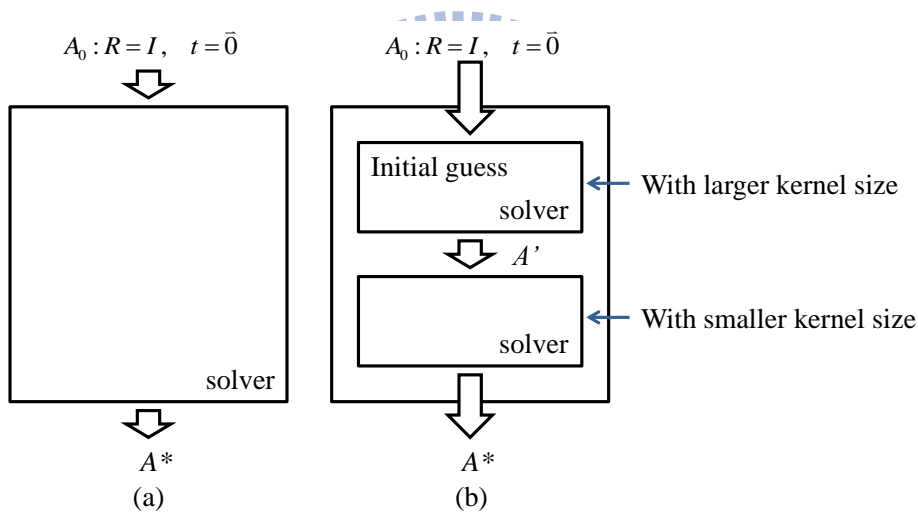


Figure 4.7: (a) A general solver that takes a set of initial affine transformations  $A_0$  as input and produces  $A^*$  as output, which contains the desired transformation of the template. (b) An adaptive kernel size strategy that contains an extra step of guessing the initial point for the non-linear system.

Finally, our energy minimization process is solved with aforementioned parameters and the registration results are shown in Figure 4.8 and 4.9.

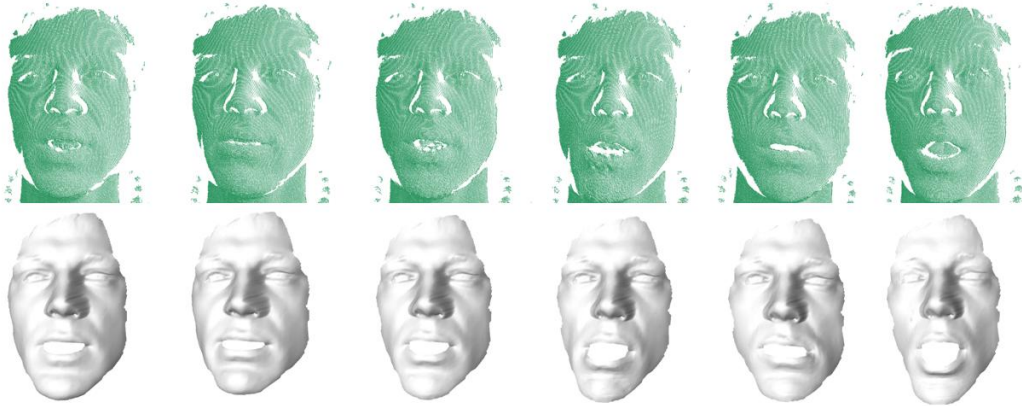


Figure 4.8: Non-rigid registration results of the face data set.

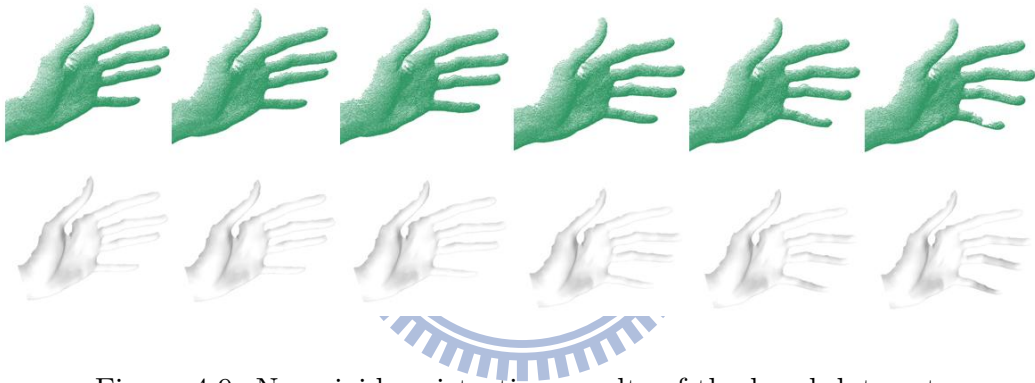


Figure 4.9: Non-rigid registration results of the hand data set.

## 4.2 Registration Errors and Time Statistics

In this section we compare our non-rigid registration method to ICP-based methods in terms of registration errors and time statistics. We implement the non-rigid ICP-based method proposed in Li et al. [LAGP09]. The ICP-based formulation for energy system is as follows [LAGP09]:

$$E_{total} = w_{fit}E_{fit} + w_{rotate}E_{rotate} + w_{regular}E_{regular}, \quad (4.1)$$

where  $E_{rotate}$  and  $E_{regular}$  are the same as Equation 3.5 and 3.7, and the surface fitting function is

$$E_{fit} = \sum_{(s,t(s)) \in C} \alpha_{point} \|A(s) - t(s)\|^2 + \alpha_{plane} |n_{t(s)}^T (A(s) - t(s))|^2, \quad (4.2)$$

where  $s$  is a point on the source surface and  $A(s)$  is the transformed position of  $s$ , and  $t(s)$  is the closest point of  $s$  on the target surface with the corresponding surface normal  $n_{t(s)}$ , and  $C$  is the set of mappings that maps one source point to its closest point and in which some of them may have been removed to filter out bad correspondences. The weights used in the test are  $w_{fit} = 1$ ,  $w_{rotate} = 10^6$ ,  $w_{regular} = 10$ ,  $\alpha_{point} = 0.1$ , and  $\alpha_{plane} = 1$ .

For comparison, the registration error of two surfaces is measured by calculating the distance between the shapes. The Hausdorff distance can be used to calculate the distance between two shapes. But calculating the maximum of minimum distances is bad for measuring the error between two shapes if there exists a large minimum distance between the shapes. We use an error measurement that averages the minimum distances between two shapes as follows. Given the registered surface template  $S$  and the corresponding range image  $T$ , we associate each  $s \in S$  with a distance  $d_s = \|s - t(s)\|$  where  $t(s)$  is the closest point of  $s$  in  $T$ . The registration error is measured by calculating the root-mean-square error (RMSE) on the template,

$$RMSE(S) = \sqrt{\frac{\sum_{s \in S} d_s^2}{|S|}}, \quad (4.3)$$

where  $|S|$  is the number of points in  $S$ .

Figure 4.10 shows the registration errors of the face data set. Colors on the template represent the distance  $d_s$  of each point  $s$  and a red color means a larger distance whereas a blue color means a smaller distance. The colors of the nose regions on the template in Figure 4.10 are always near red since

hole regions in the range image there result in large distances  $d_s$ . Another similar result is shown in Figure 4.11 for the hand data set.

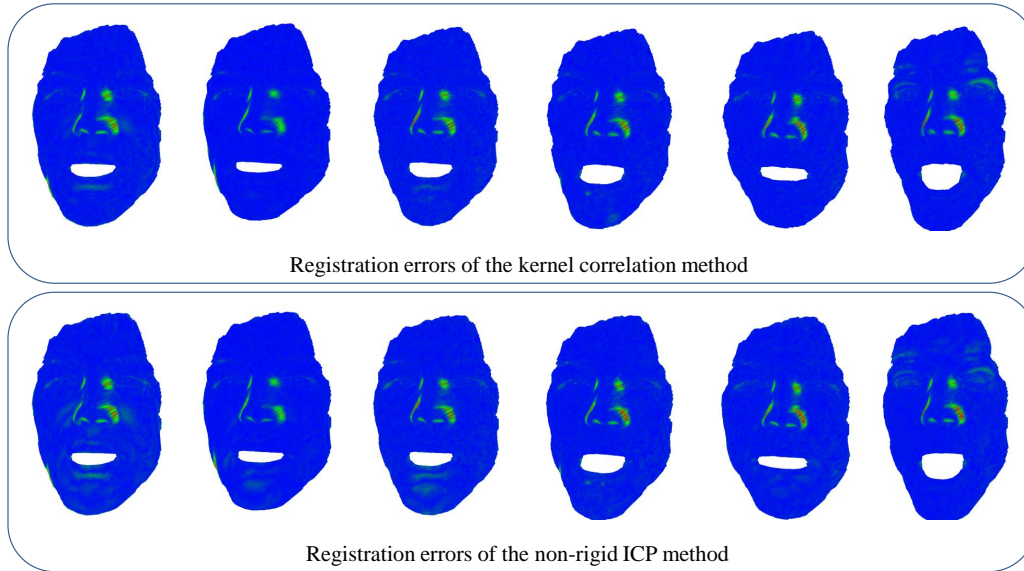


Figure 4.10: Registration errors of the face data set.

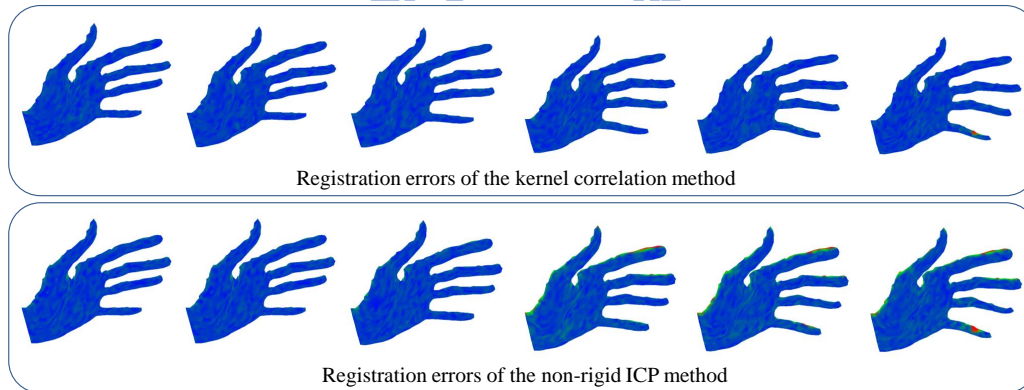


Figure 4.11: Registration errors of the hand data set.

The corresponding error statistics of the two data sets are shown in Table 4.1. The numbers in Table 4.1 show the average RMSE of 90 frames in the face data set and 100 frames in the hand data set using two methods. In

## 4.2. REGISTRATION ERRORS AND TIME STATISTICS

---

summary, the average RMSE between the proposed method and non-rigid ICP are quite close, but ours are smaller. It makes sense that our RMSE is smaller since our surface fitting function fits a point on the source surface to a reasonable 3D position rather than another point on the target surface.

	Face	Hand
Avg. RMSE per Frame (KC)	0.78597	0.383834
Avg. RMSE per Frame (N-ICP)	0.823858	0.480465

Table 4.1: The average RMSE of the face and hand data set.

The time statistics for the two data sets are listed in Table 4.2. The computations are performed on a 2.3GHz Intel<sup>®</sup> Core<sup>™</sup> i5-2410M processor with 4GB RAM. In general, our registration algorithm is more efficient compared to the non-rigid ICP since the minimization process does not need to recompute the resources from iteration to iteration in the non-linear system and is solved straightforwardly. The gap of computation time for the two methods are smaller when the number of graph nodes is small, and becomes larger when the number of graph nodes grows; as listed in Table 4.2.

	Face	Hand
# of Frames	90	100
Avg. # of Points per Frame	~56k	~37k
# of Template Vertices	8032	1441
# of Graph Nodes	802	142
Avg. Registration Time per Frame (KC)	9m:23s	7s
Avg. Registration Time per Frame (N-ICP)	1h:0m:4s	20s

Table 4.2: The time statistics for the face and hand data sets.



## CHAPTER 5

---

# Conclusions

---

We give a summary for the proposed non-rigid shape registration algorithm in this chapter and also discuss the limitations and some future works.

### 5.1 Summary

For non-rigid registration problem, many methods based on non-rigid ICP have been proposed to formulate the surface fitting function. The non-rigid ICP needs to find the set of closest points at each iteration of the non-linear energy optimization. Instead, in the proposed kernel-correlation based surface fitting function is fixed during the entire optimization process. The energy optimization can be solved straightforwardly and more efficiently. The non-rigid registration algorithm we have presented relies only on the knowledge of point samples in 3D space and does not use any color images for additional information between adjacent frames. In summary, we have presented a different way of formulating the surface fitting function for non-rigid shape registration, with better result and better computational efficiency, compared to an ICP-based method [LAGP09].

## 5.2 Limitations

The input range images are assumed to have high spatial and temporal coherence between adjacent frames. If there exists two adjacent frames with large motion gap, then the process will result in incorrect registration since it can be trapped into the local minimum of the non-linear energy system. Figure 5.1 shows an example in which the motion gap between two adjacent frames is large, especially in the mouse region. The registration result for the second frame is bad.

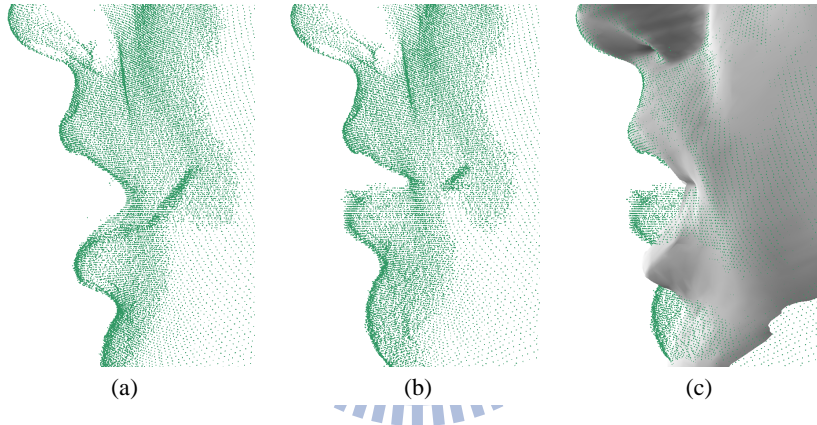


Figure 5.1: Image (a) and (b) are adjacent frames. (c) is the registration result of the latter frame.

To handle this problem, we provide a good guess for the initial point of the non-linear system. We manually select a set of graph nodes and adjust initial translation vector  $t$  for each selected graph node. Figure 5.2(a) shows the adjusted result, which is better than before.

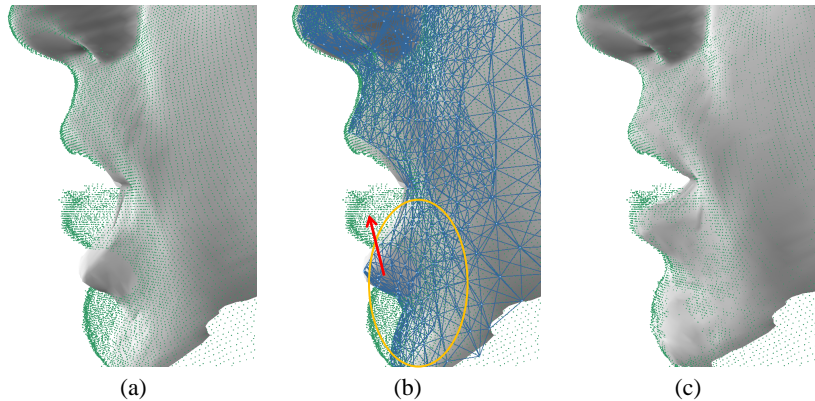


Figure 5.2: (a) The surface template before doing registration. (b) A set of graph nodes are selected and each node is given a motion vector as an initial guess for the non-linear system. (c) The adjusted registration result.

### 5.3 Future Works

There are several issues for enhancing the robustness of our non-rigid registration process. First, we set the same kernel size  $\sigma$  to each source point  $s$  using Gaussian kernel and solve the optimization process with the fixed  $\sigma$  over the entire minimization process. Setting the same  $\sigma$  for each  $s$  is fine. However, the size of  $\sigma$  is currently chosen by user and fixed in the optimization process, which may lead a bad registration result if the motion gap between two adjacent frames is large. As described in Section 4.1, we use adaptive kernel size strategy that employs two different sizes to handle the problem with large motion gap. It is recommended to study the adaptive kernel size strategy to see if the size of kernel  $\sigma$  can be automatically computed. The Gaussian kernel, on the other hand, is an isotropic kernel. In some cases the surface fitting function using kernel correlation always fits a point to a wrong 3D position, as shown in Figure 5.3(a). It is desired

to design an anisotropic kernel for kernel correlation. Figure 5.3(b) and (c) show that an anisotropic kernel can be designed according either the source or the target surface, depending on the problem we have. For non-rigid shape registration problem, the input range images have high spatial and temporal coherence between adjacent frames. The anisotropic kernel can be designed according to the source surface since the deformation between the source and the target is small. It is also suggested to employ feature matchings between shapes if we have such matched features. Figure 5.4 shows that if we have a guess of the matched position on the target surface, the anisotropic kernel for the source point  $s$  is designed according to the target surface. In fact, the anisotropic kernel should always be designed according the target surface since we are transforming the source to the target.

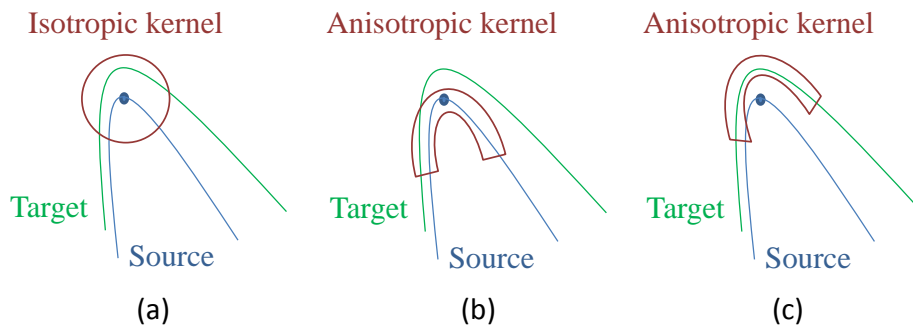


Figure 5.3: (a) Kernel correlation with isotropic kernel function may lead a source point to a wrong 3D position. Anisotropic kernel can be designed according to either the source surface (b), or the target surface (c).

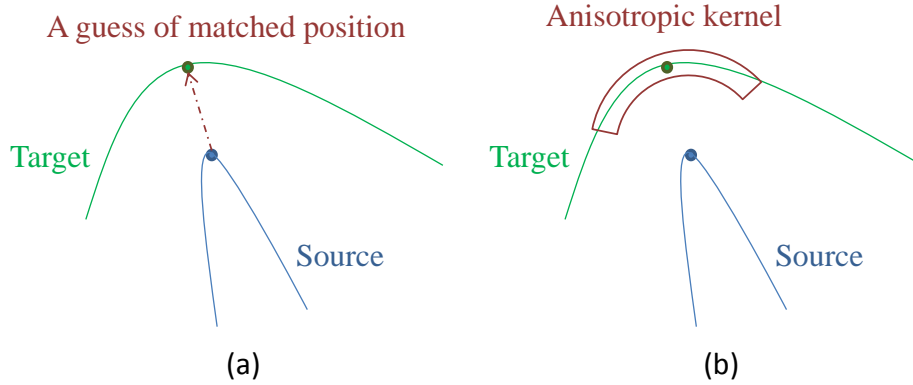


Figure 5.4: If we have a guess of matched position on the target surface for the source point as shown in (a), then the anisotropic kernel should be designed according to the target surface (b).

For reducing the computation time of our non-rigid registration algorithm, it is worth to reduce the computation cost of the energy minimization process. The number of unknown variables in the non-linear energy system is  $12n$  where  $n$  is the number of nodes in deformation graph. The computation cost can be reduced if we reduce the number of graph nodes. In deformation graph, a node presents a transformation entity. It is worth to distribute graph nodes by importance sampling on the underlying surface rather than uniform sampling. Furthermore, the number of unknown variables in the non-linear energy system can be reduced from  $12n$  to  $7n$  if we replace the 3 by 3 rotation matrix  $R$  for each graph node with a quaternion. Using a quaternion to present a rotation only needs 4 variables compared to a rotation matrix which needs 9 variables.

---

## Bibliography

---

- [ACP03] B. Allen, B. Curless, and Z. Popović. The space of human body shapes: reconstruction and parameterization from range scans. *ACM Transactions on Graphics - Proceedings of ACM SIGGRAPH*, 22(3), 2003.
- [AMCO08] D. Aiger, N. J. Mitra, and D. Cohen-Or. 4-points congruent sets for robust pairwise surface registration. *ACM Transactions on Graphics - Proceedings of ACM SIGGRAPH*, 27(3), 2008.
- [ART10] A. Albarelli, E. Rodolà, and A. Torsello. A game-theoretic approach to fine surface registration without initial motion estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.
- [BHB<sup>+</sup>11] T. Beeler, F. Hahn, D. Bradley, B. Bickel, P. Beardsley, C. Gotsman, R. W. Sumner, and M. Gross. High-quality passive facial performance capture using anchor frames. *ACM Transactions on Graphics - Proceedings of ACM SIGGRAPH*, 30(4), 2011.
- [BM92] P. J. Besl and N. D. McKay. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992.

- [GMGP05] N. Gelfand, N. J. Mitra, L. J. Guibas, and H. Pottmann. Robust global registration. In *Proceedings of the third Eurographics symposium on Geometry processing*, 2005.
- [HCTW11] H. Huang, J. Chai, X. Tong, and H.-T. Wu. Leveraging motion capture and 3d scanning for high-fidelity facial performance acquisition. *ACM Transactions on Graphics - Proceedings of ACM SIGGRAPH*, 30(4), 2011.
- [LAGP09] H. Li, B. Adams, L. J. Guibas, and M. Pauly. Robust single-view geometry and motion reconstruction. *ACM Transactions on Graphics - Proceedings of ACM SIGGRAPH Asia*, 28(5), 2009.
- [LLV<sup>+</sup>12] H. Li, L. Luo, D. Vlastic, P. Peers, J. Popović and M. Pauly, and S. Rusinkiewicz. Temporally coherent completion of dynamic shapes. *ACM Transactions on Graphics - Proceedings of ACM SIGGRAPH*, 31(1), 2012.
- [LSP08] H. Li, R. W. Sumner, and M. Pauly. Global correspondence optimization for non-rigid registration of depth scans. In *Proceedings of the Symposium on Geometry Processing*, 2008.
- [MNT04] K. Madsen, H. B. Nielsen, and O. Tingleff. Methods for non-linear least squares problems. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, 2004.
- [PB09] C. Papazov and D. Burschka. Stochastic optimization for rigid point set registration. In *Proceedings of the 5th International Symposium on Advances in Visual Computing*, 2009.

- [PB11] C. Papazov and D. Burschka. Deformable 3d shape registration based on local similarity transforms. In *Proceedings of Eurographics Symposium on Geometry Processing*, 2011.
- [PSDB<sup>+</sup>10] T. Popa, I. South-Dickinson, D. Bradley, A. Sheffer, and W. Heidrich. Globally consistent space-time reconstruction. In *Proceedings of Eurographics Symposium on Geometry Processing*, 2010.
- [SSP07] R. W. Sumner, J. Schmid, and M. Pauly. Embedded deformation for shape manipulation. *ACM Transactions on Graphics - Proceedings of ACM SIGGRAPH*, 26(3), 2007.
- [TBW<sup>+</sup>12] A. Tevs, A. Berner, M. Wand, I. Ihrke, M. Bokeloh, J. Kerber, and H.-P. Seidel. Animation cartography - intrinsic reconstruction of shape and motion. *ACM Transactions on Graphics - Proceedings of ACM SIGGRAPH*, 31(2), 2012.
- [TK04] Y. Tsin and T. Kanade. A correlation-based approach to robust point set registration. In *European Conference on Computer Vision*, 2004.
- [Tsi03] Y. Tsin. *Kernel Correlation as an Affinity Measure in Point-Sampled Vision Problems*. PhD thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, 2003.
- [WAO<sup>+</sup>09] M. Wand, B. Adams, M. Ovsjanikov, A. Berner, M. Bokeloh, P. Jenke, L. Guibas, H.-P. Seidel, and A. Schilling. Efficient reconstruction of nonrigid shape and motion from real-time 3d scanner data. *ACM Transactions on Graphics - Proceedings of ACM SIGGRAPH*, 28(2), 2009.



- [WJH<sup>+</sup>07] M. Wand, P. Jenke, Q. Huang, M. Bokeloh, L. Guibas, and A. Schilling. Reconstruction of deforming geometry from time-varying point clouds. In *Proceedings of the fifth Eurographics symposium on Geometry processing*, 2007.

