Range Image Registration Based on Photometry

Diego Thomas

DOCTOR OF PHILOSOPHY

Department of Informatics
School of Multidisciplinary Sciences
The Graduate University for Advanced Studies (SOKENDAI)

March 2012

Abstract

3D modeling of a real scene stands for constructing a virtual representation of the scene, generally simplified that can be used or modified at our will. Constructing such a 3D model by hand is a laborious and time consuming task, and automating the whole process has attracted growing interest in the computer vision field. In particular, the task of registering (i.e. aligning) different parts of the scene (called range images) acquired from different viewpoints is of crucial importance when constructing 3D models. During the last decades, researchers have concentrated their efforts on this problem and proposed several methodologies to automatically register range images. Thereby, key-point detectors and descriptors have been utilized to match points across different range images using geometric features or textural features. Several similarity metrics have also been proposed to identify the overlapping regions. In spite of the advantages of the current methods, several limitation cases have been reported. In particular, when the scene lacks in discriminative geometric features, the difficulty of accounting for the changes in appearance of the scene observed in different poses, or from different viewpoints, significantly degrades the performance of the current methods. We address this issue by investigating the use of photometry (i.e. the relationship between geometry, reflectance properties and illumination) for range image registration. First, we propose a robust descriptor using albedo that is permissive to errors in the illumination estimation. Second, we propose an albedo extraction technique for specular surfaces that enlarges the range of materials we can deal with. Third, we propose a photometric metric under unknown lighting that allows registration of range images without any assumptions on the illumination. With these proposed methods, we significantly enlarge the practicability and range of applications of range image registration.

Acknowledgements

I owe my deepest gratitude to my supervisor, Akihiro Sugimoto, for welcoming me in his laboratory and providing me the opportunity to work in a unique research environment during the last three years. I will always be grateful for the numerous and constructive advice he taught me to succeed in my research, and, the precious time he spent with me for discussions or the multiple times he read and corrected my articles to make them acceptable for submission. Professor Sugimoto also introduced me to respectful researchers with whom I had the opportunity to discuss and cooperate on successful projects.

I thank the members of my PhD committee, Professor Yoichi Sato, Associate professor Imari Sato, Professor Shin'ichi Satoh and Associate professor Helmut Prendinger, for their constructive comments and invaluable advice. I am indebted to them for investing time and effort discussing range image registration and suggesting directions to guide me through my PhD.

I am particularly grateful to Dr. Andre Gagalowicz, Dr. Roby T.Tan and Dr. Yasuyuki Matsushita. Conversations with them have been fruitful and greatly contributed to my PhD.

I thank all my lab mates for sharing their experiences and reporting their feedback on my work; including people from the Czech Technical University of Prague Jan Heller and Tomas Vojir, Dr. Yohan Thibault from the university of Paris-Est. I owe special thanks to Yohan, he has been particularly helpful during the early years of my stay in Japan.

Certainly none of this work would exist without the financial and human

support of the National Institute of Informatics (NII), I thank the NII for funding my PhD and the NII secretaries for their patience in helping me to decipher Japanese paperwork. My stay in Japan was also really enjoyable thanks to the nice people I have met in NII and that became invaluable friends with whom I spent wonderful time.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 3D Vision

Among all our five senses, vision plays the most important role in our daily life. Our visual perception helps us understanding, interacting and moving into our environment. The objective of 3D vision is to simulate the human visual perception (i.e. make the computer "see"). Though it is a natural and easy task for a human to recognize a friend, to grasp an object or to avoid a wall, it is difficult for a computer, or a robot. Why is it so? First of all we need to agree on what is the visual system we want to simulate.

A visual system, in broad terms, is a collection of devices that transform measurements of light into information about spatial and material properties of a scene. It contains visual sensors such as eyes for the human or digital cameras for the computer, and computational units such as the brain for the human or the CPU for the computer. While the sensors record intensity of the light that hits the photosensitive cells (pixels in the case of digital cameras), the computational unit "interprets" the recorded values to tell what are the characteristics of the scene we are looking at.

To get a grasp on how difficult 3D vision is, we can simply notice the time it takes for a human to fully control its sense of vision (which requires half of our brain capacities at all time). And even so it happens to mistake

what we are seeing. Obviously, to make the computer "see" like a human is tremendously hard. The objective of 3D vision is to successfully perform simple tasks (for the human) such as recognition for example, which could relieve the human from several boring tasks (verification or surveillance for example).

We need models and assumptions to interpret the captured images and to tell meaningful information from them. These models are a virtual representation of our world, generally simplified that can be used or modified at our will. For example, deformation models and gravity models can explain motion of the same object on different planets. Image formation models can explain the appearance of an object surface. And most importantly 3D models give a geometric understanding of the scene.

Constructing 3D models of an observed scene is of outmost importance for computer vision. And even though many researches have been done on this topic during the last decades, it remains a challenging problem, yet to be fully solved. An introduction to 3D vision is given in [51]. An overview of practical applications for computer vision is given in [99]. A more advanced presentation of computer vision is given in [25].

1.2 3D modeling

To understand our environment, it is essential to know its 3D geometry. 2D images alone are insufficient. This is supported by the fact that there are many famous examples of paintings "trompe-l'oeil". If we look at an object from different viewpoints (from above and from behind), in different illuminations (at the light of a candle or under a neon light) or with different paintings, its appearance may change completely. It is thus difficult from the visual observations alone to recognize an object, its position or its characteristics for example. On the contrary, even with our eyes closed, by simply touching an object we can infer its 3D shape and recognize it. This is because, from the 3D shape of an object we can naturally imagine its 3D model (i.e. its 3D representation). Then, by combining both the visual

observations and the 3D models of objects, the human can easily interpret its surroundings. For a visual system to be efficient, it is thus essential to have 3D models of the environment.

From 2D images to 3D models, the process can be divided into two main steps: (1) 3D reconstruction. Namely, the objective of 3D reconstruction is to extract 3D information from the 2D observations. Such 3D information can be 3D coordinates of points or surface normals for example. (2) 3D modeling. The objective of 3D modeling is to *fuse* all 3D information extracted in step (1) into a compact 3D representation (or model). Such model can be a mesh, a cloud of points, a 3D grid of voxels or implicit surfaces for example.

Famous applications areas for 3D models are, for example,

- **Surveillance and security.** With augmented virtual environment, 3D models help observers comprehend multiple streams of temporal data and imagery from arbitrary views of the scene. 3D models are particularly useful for moving object detection, tracking, and 3D display for effective dynamic event comprehension and situational awareness ([85]).

- **Industrial inspection and quality control.** Defects experienced during construction are costly and preventable. 3D models are being utilized for renovation, retrofit and expansion projects in industrial, commercial and heavy-civil sectors of construction, and active quality control purposes ([1]).

- **Reverse engineering.** From simple distance measurement up to the control of micrometric deformation, symmetry, etc., all the operations that allow the measurement and analysis of geometric forms are called reverse engineering of shapes. 3D models are used to understand the structure and mechanisms of objects ([55]).

- **Face and gesture recognition.** 3D models are being utilized to improve the quality and robustness of face and gesture recognition

when dealing with complex background or changes in appearance.

- **Autonomous vehicles.** Access to 3D models increases the operator situational awareness and allows better mission planning and execution, as the models can be visualized from different viewpoints and used for relative measurements ([83]).

- **Space and applications.** Servicing satellites in space requires accurate and reliable 3D information. Such information can be used to create virtual models of space structures for inspection (geometry, surface flaws, and deployment of appendages), estimation of the relative position and orientation of a target spacecraft during autonomous docking or satellite capture, replacement of serviceable modules, detection of unexpected objects and collisions ([84]).

- **Medical image analysis.** 3D models offer many possibilities to image analysis that combines geometry, physics, and approximation theory. They have proven to be effective in segmenting, matching, and tracking anatomic structures. Deformable models are capable of accommodating the significant variability of biological structures over time and across different individuals. Furthermore, they support highly intuitive interaction mechanisms that, when necessary, allow medical scientists and practitioners to bring their expertise to bear on the model-based image interpretation task ([54]).

- **Analysis and simulation.** Testing the resistance and behavior of a product under various conditions is a crucial step in designing a product. In car industry for example, crash tests are performed for security check. While these experiments require heavy and expensive experimental set-up, 3D models allow cheap and easy simulation.

Accurately modeling a scene by hand, however, requires expertise and time. Therefore, automating the 3D modeling process in controlled (laboratory set-up for example) or uncontrolled (outdoor for example) environ-

ments is of major importance for the computer vision community and has attracting growing interest in the past decades.

Obtaining the 3D model of an object is certainly not as simple as moving a camera around it. First, we do not have such an efficient visual system as the human eye. A camera records 2D projections of the observed scene, and inferring the 3D geometry from a single 2D image is impossible. To extract the 3D surface from 2D images, multiple observations from different viewpoints are required.

By observing the same point from different viewing positions, it becomes possible to retrieve its 3D coordinates. From a set of observations, it is therefore possible to reconstruct the 3D surface corresponding to the overlap region between all 2D images. This process is called stereo vision in the case of two viewing positions. Currently, using more viewing positions is popular and inferring 3D coordinates is well studied under the name of multiple view geometry ([22, 32]).

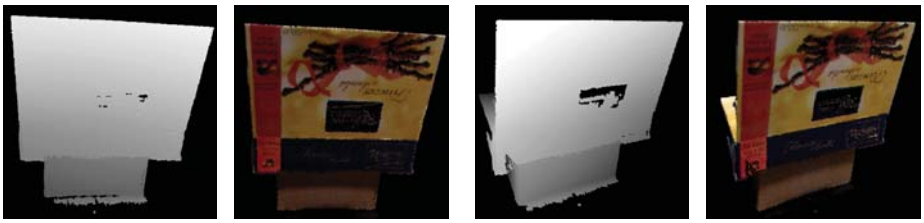
On the other hand, the recent advances in 3D acquisition devices such as modern laser range scanners or structured light sensors (for example, the new Microsoft *Kinect* or *Primesense* sensors) open new possibilities for 3D modeling. In particular, recent acquisition devices can retrieve both the 3D shape and a color image of an object from a fixed viewpoint. The acquired 3D image in this case is called a range image (Fig.1.1). The first step of the 3D modeling process (i.e. the 3D reconstruction step) can thus be omitted.

However, from one viewpoint some parts of the scene are occluded. As a consequence, from one viewpoint it is only possible to acquire a part of the scene. Therefore, multiple range images, acquired from various positions are required to acquire all parts of the scene.

All range images of the scene are obtained in the local coordinate systems defined by the different viewpoints. The transformations that relate all range images together are in general unknown and it is thus necessary to align all overlapping range images together. This process is called range image registration [79]. The difficulty of range image registration is that

(c) *can00* (f) *can20*

Data *can*.



(i) *box00* (l) *box20*

Data *box*.

(o) *hand00* (r) *hand20*

Data *hand*.

Figure 1.1: Examples of range images with color images used in this thesis

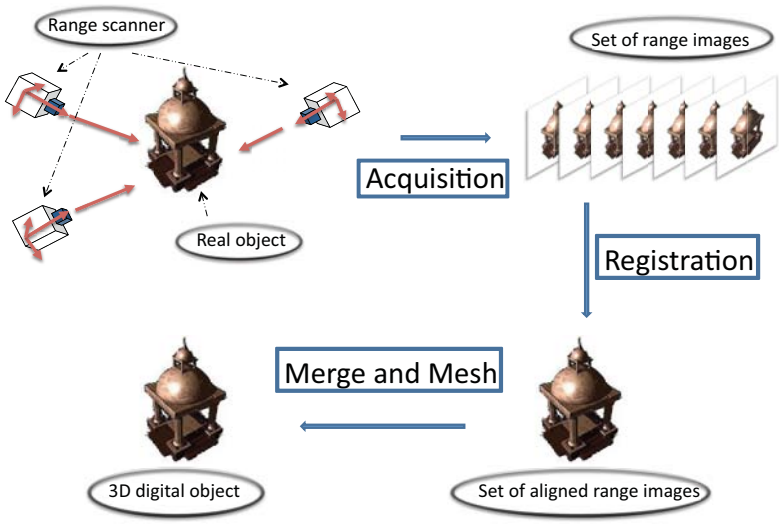


Figure 1.2: The 3D modeling process

we do not know which areas of the different range images are overlapping. This problem can be solved either by identifying correspondences across adjacent scans or by minimizing cost functions that model the goodness of the alignment.

All aligned range images are finally merged and integrated into a single 3D model (such as a mesh) that can be used in various vision systems. The 3D modeling process is summarized and illustrated in Fig. 1.2.

The large amount of work on integration methods gave us satisfying tools to merge multiple range images together and generate 3D models. Therefore, the most critical step for automated 3D modeling is the range image registration process. It has attracted growing interest in the past decades, but though impressive techniques have been proposed, there are still several situations where aligning multiple range images together remains an open challenge.

1.3 3D registration

Many studies have been proposed that allow accurate 3D registration ([5, 7, 39, 50]). However, they rely highly on geometry and/or texture of the object to be modeled and inevitably fail in some situations. In particular, registering range images of an object using photometry is still an unresolved problem.

1.3.1 Registration using geometric features

Aligning overlapping range images using geometry is the most popular approach to 3D registration and has been extensively studied over the past decades. When using geometry to guide the registration process, in general, discriminative feature descriptors are used to identify key-point correspondences that allow estimation of the transformation. Popular descriptors are for example, the position of the point, the normal at the surface or the curvature.

Another approach to 3D registration is to minimize a cost function, which models the quality of the alignment. Cost functions based on geometry have been investigated. For example, useful geometrical cost functions have been proposed that represent the distance between the two aligned range images or the geometrical distribution of points.

When geometric features of points at the surface are sufficiently discriminative, the large toolbox of 3D registration methods allows for accurate alignment. However, many man-made objects (such as a cup or a ball for example) present symmetries in their shape or repetitive patterns. In such a case, the captured range images are devoid of salient geometric features and using geometry alone for registration inevitably fails.

1.3.2 Registration using textural information

In addition to geometry, textural information is also available. By textural information, we denote the information derived from the appearance of the

object's textured surface. This can be the color reflected by the object surface towards the scanning viewpoint, the chromaticity or the intensity for example. By using the color images in addition to geometry it becomes possible to some extent to overcome the problem of registering range images devoid of salient geometric features.

Several methods have been proposed that use textural information such as color or chromaticity in addition to geometry to guide the registration process. For example, the popular SIFT method uses differences of Gaussians in the intensity image to identify key-points and define scale-invariant descriptor. Several 2D texture-based cost functions such as the cross-correlation are also available for aligning textured images.

However, unlike geometry, the textural information is not an intrinsic attribute of the object surface (i.e. it does not depend on the object surface material only). It is actually the result of the complex interaction between geometry, reflectance and light. As a consequence, the direct use of textural information is, in general, unreliable. In particular, when the changes in the object pose, viewpoint or illumination induce drastic changes in the object surface appearance; texture base registration methods fail accordingly.

1.3.3 Registration using photometric features

The recent advances and breakthroughs in understanding and modeling image formation ([4, 64, 75]) bring new possibilities for 3D registration using photometry. By photometry, we denote the relationship between geometry, reflectance properties and incident illumination. This phenomenon is illustrated in Fig. 1.3.

By understanding and modeling the complex process behind image formation, it becomes possible to deal with changes in the object surface appearance, and even use it (rather than ignoring it) for the task of 3D registration.

By contrast with the textural features, we denote by photometric features the surface intrinsic features that define its reflectance properties. For

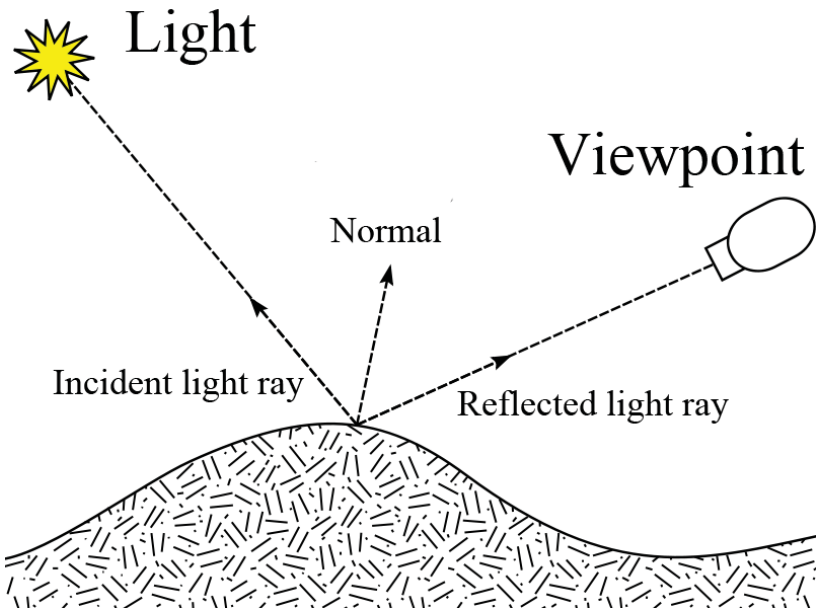


Figure 1.3: Photometry

instance, by contrast with the captured color, the albedo is a well-known reflectance attribute that depends solely on the object material (for the diffuse reflection, the albedo of a point represents how much light is reflected by it when illuminated by a point light source facing its normal). It is a photometric feature well fitted for matching and comparing similarity of points at the surface of an object.

A few works have been reported that investigate the use of albedo for range image registration. However, they rely on a precise estimation of albedo from the captured color and geometry, and thus on a precise estimation of the surrounding illumination. In particular, no robust descriptor exists to palliate the inevitable inaccuracies in the illumination estimation. The exact surrounding illumination is in general unknown and therefore existing techniques that use albedo for registration purpose are not applicable in real situations.

On the other hand, compact and accurate representations of image for-

mation have been identified ([4, 60, 63, 70, 75, 76]). Thereby, by using the spherical harmonics, the complex mechanism behind image formation can be accurately expressed in a linear subspace. Though this recent advances have been widely applied in computer graphics ([47]) and inverse rendering ([9, 52, 80]), no work has been reported that use this new model for range image registration.

In particular, when using photometry to align overlapping range images, no specific cost functions are available. In the following, we investigate the use of photometry (i.e. the relationship between geometry, reflectance properties and illumination) for range image registration.

1.4 Contribution

In this dissertation, we address the problem of accurately aligning overlapping pairs of range images of an object devoid of salient geometric features (a bottle for example). We assume the rigid surface and known reflectance property (such as Lambertian, specular or Lambertian plus specular for example). We also assume no inter-reflections, neither cast-shadows. We consider the scenario where the viewpoint and illumination are fixed while the pose of the object is changing.

We propose to use photometry to tackle this challenging problem and we enforce to leverage as much as possible the assumptions on the surrounding illumination.

The contributions of this dissertation are three-fold. First, we define a robust 3D descriptor that allows registration of Lambertian range images using albedo even under a rough estimation of "simple" illumination (i.e. an ambient and a directional light source). Second, by extracting albedo in parts of specular objects illuminated by a few unknown point light sources, we extend the range of application of our previously proposed registration method. Third, by using the new linear representation of image formation, we define a useful photometric metric for range image registration of Lambertian objects under unknown and general illumination.

1.4.1 Robust 3D descriptor

We propose a robust descriptor for the registration of overlapping Lambertian range images that makes use of local distribution of albedo. We use adaptive regions to model the local distribution of reflectance, which enables us to stably extract reliable attributes of each point against illumination estimation. We use the popular level set method to grow robust and adaptive regions to define these attributes. A similarity metric between two attributes is also defined to match points in the overlapping area, and remaining mismatches are efficiently removed using the rigidity constraint of surfaces. Our experiments using synthetic and real data demonstrate the robustness and effectiveness of the proposed method.

The proposed descriptor allows for accurate alignment of Lambertian range images even under rough approximation of the surrounding illumination. Our method is the first one that allows application of registration using albedo for real data, when the scene is illuminated by a simple light, such as a single directional light source. It thus becomes possible to account for significant changes in appearance caused by illumination and therefore extend the range of applications for range image registration. A part of this work appeared in [91, 95].

1.4.2 Specularity removal

When the object to be modeled present specularities, the simple Lambertian reflection model is insufficient to extract albedo from the captured color and geometry. Therefore, the previously proposed registration method cannot be directly applied in such a case.

We propose a method for estimating albedo in parts of specular surfaces for the purpose of range image registration of specular objects devoid of salient geometric properties under a few unknown point light sources. Our method uses illumination consistency on two range images to detect specular highlights. Directions and intensities of the surrounding light sources are estimated from the detected specular highlights, which allows identification

of regions where the diffuse refection components can be extracted. This technique can handle various kind of illumination situations and can be applied to a wide range of materials. Our experiments using synthetic data and real data demonstrate the usefulness of our proposed method for range image registration.

By extracting albedo in parts of specular range images we allow for the accurate registration of the range images even under a few unknown point light sources. While specular reflection used to be simply ignored in precedent works (i.e. the object surface is assumed to be mostly Lambertian), we use them to reliably estimate reflectance attributes that are used to guide the registration. In particular, it becomes possible to align range images of an object that presents mostly specular regions. A part of this work appeared in [92, 93, 94].

1.4.3 Photometric metric under unknown lighting

The image formation process is complex, and under general unknown illumination it becomes impossible to extract albedo from range images without knowing the exact point correspondences. As a consequence, feature based matching registration approach is inevitably limited to known or simple illumination.

On the other hand, matching features is not the only way we can take for registering overlapping range images. We propose a photometric cost function for evaluating the goodness of a rigid transformation aligning two overlapping range images under unknown lighting, and under the assumption of Lambertian surface. Our metric is based on photometric re-projection error but not on feature detection and matching. By using the spherical harmonics representation of image formation, we estimate illumination and albedo from the correspondences induced by the input transformation. We then synthesize the color of one image using albedo of the other image to compute the photometric re-projection error. This way allows us to derive a photometric metric under unknown lighting for range image alignment.

We use a hypothesize-and-test method to search for the transformation that minimizes our photometric cost function. Employing the spherical representation of each range image efficiently generates transformation candidates. Experimental results using synthetic and real data show the usefulness of the proposed metric.

Our proposed cost function is the first one that makes use of photometry for evaluating the goodness of the alignment. Its generality and independence against illumination allows accurate registration of Lambertian range images under unknown general illumination. The proposed method thus further enlarges the range of applications for range image registration. In particular, the ideal laboratory conditions are no more required for aligning overlapping range images. A part of this work appeared in [96].

1.5 Dissertation outline

The remaining of this dissertation consists of five chapters. In Chapter 2 we summarize the state of the art in 3D registration and identify the open challenges. Chapter 3 presents our robust registration method using local distribution of albedo. In this chapter, we assume the Lambertian reflectance and a rough estimation of incident illumination, and propose a robust descriptor for matching points across two overlapping range images. In Chapter 4, we present our albedo extraction technique for specular objects illuminated by a few light sources. We assume that the illumination is unknown but composed of a few point light sources. We detect the point light sources direction from the specular highlights, which allows us to separate the reflection components in some parts of the range images. In Chapter 5, we present our photometric metric under unknown lighting. We assume the Lambertian reflectance but allow for general, unknown illumination. We use the photometric consistency to evaluate the current alignment and employ spherical representation for searching the best transformation. Last but not least, Chapter 6 concludes the dissertation and gives the promising directions for improving 3D registration using photometry.

Chapter 2

Literature Review

Registration has been the topic of much recent work. This interest comes from the importance of image registration in various applications, as well as its complicated nature.

At first, because only 2D images were available, research focused on registering 2D images. The historically first survey paper [26] covers mainly the methods based on image correlation. Probably the most exhaustive reviews of general-purpose image registration methods are in [12] and [111]. Registration techniques applied particularly in medical imaging are summarized in [35, 45, 53, 100]. In [2] the surface based registration methods in medical imaging are reviewed. Volume-based registration is reviewed in [19]. The registration methods applied mainly in remote sensing are described and evaluated in [24, 30, 56]. Different registration methods were evaluated in [106].

With the birth of 3D sensors (such as range finder or structured light sensors) and the improvements in stereo vision, registering 3D images (also called range images) became a major topic in the computer vision literature. When using range images, the 3D geometry becomes directly available, which opened new possibilities as well as new challenges for range image registration. A review of range image registration methods is given in [78].

We classify the state of the art in range image registration into three classes: (1) the methods that use geometry only; (2) the methods that combine geometry and textural features; and (3) the methods that use photometric features.

While the first class is the most popular and most studied case, the second one has attracted many research recently and the third one is still largely unexplored.

2.1 Registration based on geometric features

The most famous approach to registration is the ICP (Iterative Closest Point) [7, 109]. This feature-based method iterates two steps: matching of each point of the first image with its closest point on the other image; estimation of the transformation between the two images using the matched point correspondences. ICP converges with a local estimate and therefore requires a rough alignment. To achieve robust and accurate alignment, many variants have been proposed ([6, 8, 16, 41, 61, 73, 77, 107, 108].

Like all feature-based methods, the ICP-like methods rely on the invariance and distinctiveness of the detected features. Much effort has been done on extracting such geometric features. For example, Khoualed *et al.* [43] proposed the semantic shape context to combine local descriptors and global context information using the Bag-of-Word paradigm. Nguyen *et al.* [62] proposed the Robust-Closest-Patch algorithm (RCP), which uses low curvature patches to guide the registration.

Different approaches have been proposed to match the detected features. For example Liu *et al.* [48] proposed an approach based on graduated assignment called ICPMC that uses entropy maximization to define a probabilistic model for point correspondences. In other work, Herman *et al.* [34] proposed the EM-ICP (Expectation-Maximization ICP) using mixture of gaussians. Zhang *et al.* [108] proposed a variant of ICP using probability fields. The major drawback of ICP-like methods is that they may get trapped in local minima and thus depend highly on the initial alignment.

On the other hand, other approaches exist that do not rely on the detected features but put emphasis on rather minimizing the alignment error ([10, 31, 39, 68]). These methods, sometimes called correlation-like methods or template matching [24], do not explicitly use the point correspondences. The feature detection step is then merged with the matching part. In [31], the similarity measure is defined as the cross correlation of the spherical representations of surfaces, and customized according to the surface-intrinsic attributes while the spherical harmonics speeds up the optimization process. Torsello *et al.* [98] proposed an algorithm that uses projection of pairwise alignments onto a reference frame and diffusion along a graph of adjacent. Nishino *et al.* [65] proposed to minimize a global error by using conjugate gradient search with M-estimator.

In other work, Pulli [71] proposed to first align the scans pair-wisely with each other using the standard ICP and use the pairwise alignment as constraints that the multiview step enforces while evenly diffusing the pairwise registration errors. Eggert *et al.* [20] iterate between pairing each point of each scan with exactly one other point and minimizing the total distance between paired points. This avoids many thresholding issues in point pairing, but can prevent the algorithm from converging to a correct solution. Jin *et al.*'s approach [40] was to incrementally build a surface model, against which new views can be registered and already registered views can be reregistered.

Acceptable results have been obtained in registering 3D shapes with complex geometry. However, registering simple textured shapes (like a cylinder), which are devoid of salient geometric features, is still an open problem.

2.2 Registration based on textural features

To overcome the problem of objects devoid of salient geometric features, many interest point detectors and feature descriptors have been proposed in the last decade to establish point correspondences using color-like features.

From the extensive research on 2D image processing, many distinctive

2D textural features are available ([11, 17, 23, 28, 29, 37, 57, 104]). The most popular one is Lowe *et al.*'s [50] scale-invariant feature descriptor called the scale invariant feature transform (SIFT). They proposed to use the difference of Gaussian function to define the features for key point. Some variations of SIFT have been proposed to speed up the algorithm while maintaining the same accuracy. Bay *et al.* [5] proposed a descriptor called speeded-up robust features (SURF) that makes use of an integral image to speed up computation and comparison. Tola *et al.* [97] also proposed a local descriptor that can be quickly computed and even be used in poor quality images.

Several similarity measures for optimization methods, defined for 2D images, are also available. Methods using gradient descent techniques [103], Gauss-Newton [87] or Levenberg-Marquardt [82] optimization frameworks have been discussed. Such approaches are sometimes referred to as *energy minimization methods* when the so-called regularization or penalty term is added to the dissimilarity measure ([102]).

However, as those techniques did not directly use available 3D information but used 2D information projected from 3D, they were sensitive to texture deformation caused by projection. In fact, the same texture captured from different viewpoints produces differently deformed 2D textures, which makes the use of these techniques problematic.

On the other hand, methods combining both 3D geometry and textural information are also available. In early work, Johnson and Kang [42] proposed to deal with textured 3D shapes by using color. Okatani *et al.* [66] proposed to use chromaticity for registration. However, color and chromaticity of a point are not distinctive by themselves and good quality of correspondences cannot be achieved.

Godin *et al.* [27] proposed to use dense attributes of range image elements as a matching constraint. Weik [105] proposed to use texture intensity gradient and intensity difference. Brusco *et al.* [13] proposed to incorporate texture information in the concept of spin-images. Pulli *et al.* [72] proposed a new mismatch error to improve registration summarizing discrepancies

in color, range and silhouette between pairs of images. Ozuysal *et al.* [67] proposed a fast key-point feature detection using binary features and class posterior probabilities. Candocia's method [14] was to jointly register image pairs in their spatial coordinates and pixel values respectively.

Heo *et al.* [33] proposed a combination of mutual information and SIFT descriptor for stereo matching, applied to simultaneous depth and consistent color estimation. Seo *et al.* [86] extended the SIFT feature for range image alignment by accounting for the projective distortion. However, SIFT-like methods do not account for changes in pose-illumination relationship, but rely on the color normalization assumption (i.e. intensity changes uniformly with changes in illumination, and normalization of features with respect to the overall intensity is sufficient).

Because color or chromaticity or intensity of a Lambertian surface depends on the object pose and illumination conditions, the performance of these methods is degraded when the pose change has significant effects on the object appearance.

2.3 Registration based on photometry

In optics, photometry is the science of measurement of visible light in terms of its perceived brightness to human vision. In this dissertation, we will denote photometry as the study of image formation.

The research in photometry leads to various image formation models, drawn from empirical observations and theoretical insights. The light that is reflected by an object surface towards the camera sensors is the result of a complex relationship between incident illumination, surface reflectance properties and geometry. The bidirectional reflectance distribution function (BRDF) is the function that defines how light is reflected at an opaque surface.

The BRDF was first defined by Fred Nicodemus around 1965 [63] and widely studied since then. The most popular and simplest BRDF is the Lambertian model, which is defined along the Lambert's law. Many fa-

mous models that account for more complex material (such as the Phong reflectance [70], Cook-Toorence [76] or Orren-Nayar [60] models for example) are also available.

The complex mechanisms behind reflectance results in complex BRDF models. The image formation models then become even more complicated when combining BRDF models with general illumination. While BRDF models have been widely used in computer graphics [69] where the illumination and geometry are perfectly controlled (and often simplified), it is much more challenging to apply them to computer vision where either geometry or illumination (sometimes even both) are unknown or approximated.

The recent breakthroughs in BRDF analysis ([4, 75]) allow for more compact models of image formation, which have been successfully applied in inverse rendering ([80]) or object recognition. However, using photometry for the purpose of range image registration is still largely unexplored.

Albedo (or reflection coefficient), for example, is a reflectance property that represents the diffuse reflectivity or reflecting power of a surface. It is defined as the ratio of reflected radiation from the surface to incident radiation upon it and depends on solely the object material properties. It is a photometric property invariant to the pose of the object, the illumination conditions and the viewpoint, and is thus useful for the purpose of matching. Recently, Cerman *et al.* [15] proposed to use albedo for registering pairs of range images. The method was to use the difference of albedo to match points across the range images. Though this is the first work that proposed to use reflectance attributes for range image registration, this point-based approach is sensitive to data noise and requires precise knowledge on illumination conditions. Therefore it is not practically applicable to real data.

Chapter 3

Local descriptor using albedo distribution

3.1 Introduction

The most common approach to registering range images is to find correspondences in points between two overlapping range images and then accordingly estimate the transformation in aligning the two range images. Several methods for registering range images can be found in the literature that use geometric features for computing correspondences in points. However, we assume that the range images to be registered have simple textured shapes (like cylinders) and are thus devoid of salient geometric features. Consequently, photometric features in addition to geometric features are required to compute correspondences in points.

Reflectance properties as a photometric feature are promising because of their independence of the pose of the object relative to the sensor. Retrieving these properties has provided a major research area in physics-based vision called reflectance from brightness (with a known shape and illumination). Cerman *et al.* [15] recently proposed a method, which we call ICP using albedo (ICPA), using the reflectance properties (which is the albedo for Lambertian objects) of the object surface in the standard iterative clos-

est point (ICP) process. The illumination conditions have to be precisely known a priori so that the reflectance of the surface of an object can be accurately retrieved from its shape and brightness. Consequently, the direct use of albedo values as a matching constraint, as achieved by ICPA, requires global illumination to be accurately estimated, which is difficult to attain in practice under real illumination conditions.

We introduce a region-based approach to using reflectance attributes, namely the albedo, for robust fine registration of Lambertian objects under rough estimates of illumination. Because retrieving the albedo on the surface of a Lambertian object is sensitive to estimates of illumination, the albedo of a point cannot be directly used under rough estimates of illumination. We thus employ the local distribution of albedo for registration. Our proposed method uses adaptive regions to model the local distribution of albedo on the object surface, which leads to robust extraction of attributes against illumination estimates. These regions are grown using a level-set method, allowing us to exclude outliers and then to define more reliable attributes. We define a robust metric, using the principal component analysis (PCA) of each region to find correspondences in points. This is a stable and powerful metric to maximize the number of correct matches, even under rough estimates of illumination. Moreover, we reject remaining mismatches by enforcing the rigidity constraint on surfaces and then estimate transformation using the weighted least squares (WLS) method. Our method has advantages with rough estimates of illumination and with large amounts of noise. These advantages allow us to use simple models of illumination to register range images. Our experiments using synthetic and real data demonstrate that our method is robust. We assume in this work that the surfaces' textures present sufficient saliency to constrain the matching of two overlapping range images. We do not consider uniform or 'salt and pepper' textures, or repetitive patterns. We also assume that the objects do not present self-occlusions, shadows nor inter-reflections. Note that a part of this work appeared in [91, 95]

3.2 Related work

During the past few decades, many approaches to registering range images have been discussed. The most well-known approach to fine registration is the iterative closest point (ICP) [7, 109]. This method iterates two steps: it matches each point of the first range image with its closest point on the other range image, and estimates the transformation between the two range images using the correspondences in the matched points. The ICP converges monotonically to a local minimum and therefore requires rough alignment. Many discriminative geometric features have been proposed ([6, 39, 41, 62, 107]) to achieve robust and accurate alignment. Other approaches based on graduated assignment such as the ICP Markov Chain (ICPMC) [48] use maximization of entropy to define a probabilistic model for correspondences between points. However, even though such methods can deal with complex 3D shapes, they do not work well for simple textured shapes (like cylinders), which are devoid of salient geometric features.

To overcome the problem with shapes devoid of salient geometric features, many approaches using photometric features have also been discussed. Godin *et al.* [27] proposed using the dense attributes of range image elements as matching constraints. Weik [105] proposed using texture intensity gradients and intensity differences. Johnson and Kang [42] proposed dealing with textured 3D shapes by using color. Okatani *et al.* [66] proposed using chromaticity for registration. Brusco *et al.* [13] proposed incorporating texture information in the concept of spin-images. Pulli *et al.* [72] proposed new mismatch error to improve registration using both color and geometric information. However, because the color or chromaticity of a Lambertian surface depends on the pose of the object and illumination conditions, these methods perform poorly when the change in pose significantly affects the appearance of the object.

However, the albedo is a photometric property that is unaffected by the pose of the object, the illumination conditions, or the viewpoint, and is thus useful for matching. Cerman *et al.* [15] proposed using the albedo

difference to match points to register range images. However, this point-based approach is sensitive to data noise and requires detailed knowledge on illumination conditions. Therefore it cannot be applied in practice to real data.

In other approaches, Lowe *et al.* [50] proposed a scale-invariant feature descriptor called the scale invariant feature transform (SIFT) that makes use of differences in Gaussian functions to define the features for key point. Some variations of SIFT have been proposed to speed up the algorithm while maintaining the same accuracy. For example, Bay *et al.* [5] proposed a descriptor called speeded-up robust features (SURF) that makes use of an integral image to speed up computation and comparison. Tola *et al.* [97] also proposed a local descriptor that can be quickly computed and even be used in poor quality images. However, as these techniques did not directly use available 3D information but used 2D information projected from 3D, they were sensitive to texture deformation caused by projection. In fact, the same texture captured from different viewpoints produces differently deformed 2D textures, which makes the use of these techniques problematic. Moreover, these approaches focused more on computational efficiency than on accuracy.

In contrast to previous work, the method we propose can handle changes in photometric appearance even under rough estimates of illumination. It is robust to data noise and can thus be easily applied to real data. Moreover, it makes use of the albedo distribution as well as normals and 3D point positions, which leads to accurate and robust results.

3.3 Proposed method

Our proposed method uses the local distribution of albedo on the surface to define discriminative attributes that are robust to data noise. We define a similarity metric to efficiently match points in the overlapping part of two range images and use the rigidity constraint of surfaces to refine matching. The transformation aligning two range images is then computed using the WLS approach.

3.3.1 Overview of proposed method

The registration process is iteratively carried out by successively estimating rigid transformation, until a convergence criterion is satisfied or a maximum of iterations is completed. Matches are obtained by evaluating the similarities between attributes of points. These attributes are defined by adaptive regions representing the local distribution of albedo on the surfaces of objects. That is, each region represents the shape of the texture patterns around a point. The region grows inside homogeneous areas and stops at the border of pattern boundaries. The transformation is then estimated by minimizing the distances between matching points. Fig.3.1 has a flowchart of our proposed approach.

Generation of adaptive region Because point-based matching methods are sensitive to data noise, we propose using information from a local distribution of albedo. We define attributes, viz., a region, for each point of two range images that represent the local distribution of albedo around the point. The region is adaptively defined using a level-set approach and transformed into their local coordinate system defined by the principal axis.

Evaluation of similarities using the albedo We define a similarity metric to match points based on their attributes to estimate correspondences in points between two range images. This similarity metric represents the photometric similarities of attributes weighted by their geometric similarity. We define a list of possible matches for each point based on similarity scores and then dynamically create a list of matches.

Elimination of incorrect matches We eliminate possibly incorrect matches by using the rigidity constraint of the surfaces. That is, the distance between two points on the surface does not depend on the pose of the object. Consequently, the distance between two points in one range image and their two corresponding points in the other range image should be the same. If not, the rigidity constraint is violated. We test the rigidity constraint for

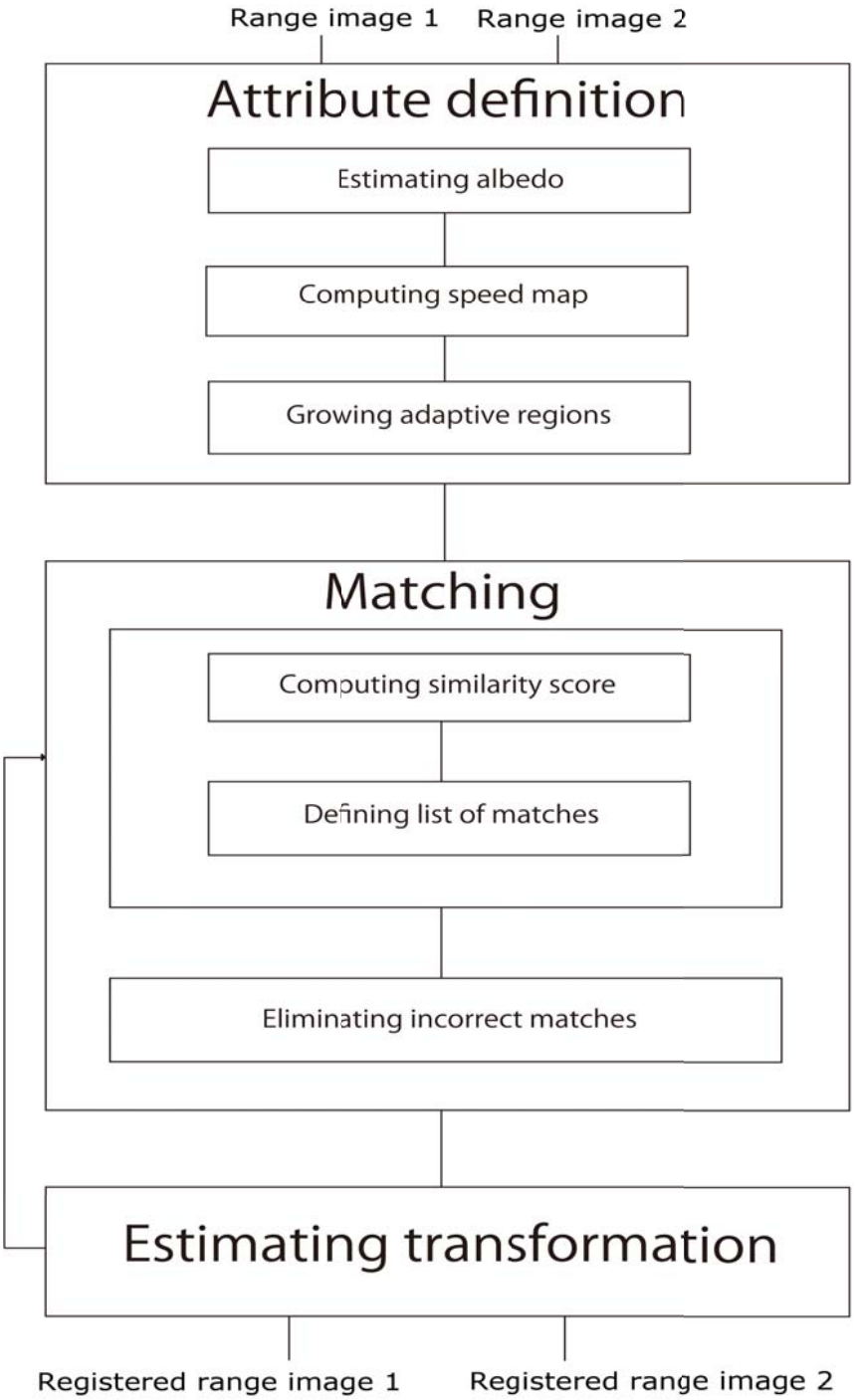


Figure 3.1: The flowchart of our proposed method.

a match with all other matches and if the number of violations exceeds a tolerance threshold, the match is eliminated.

Estimation of rigid transformation The matching list that is obtained is used as the input of a WLS algorithm to compute the current rigid transformation aligning two range images.

This framework allows simple textured range images to be robustly and accurately registered even with large amounts of data noise and rough estimates of illumination. We explain our method in more detail in what follows.

3.3.2 Generation of adaptive region

We define a region for each point of the two range images to obtain reliable attributes for each to find correspondences between points. The main idea here is to obtain a reliable representation of the local distribution of albedo. Therefore, these regions should be adaptively defined depending on the local distribution of albedo around the point of interest. Level-set methods, which are widely used for segmentation, appear to effectively model complex shapes in textured images and are robust to data noise. Therefore, we adaptively grow regions using a level-set method.

Level-set method

A region is defined by a contour that we define with a level-set method (fast marching algorithm [38]). A contour is defined as the zero level-set of a higher dimensional function called the level-set function, $\psi(X,t)$ (see Fig. 3.2). The level-set function is then evolved under the control of a differential equation. The evolving contour can be obtained at any time by extracting the zero level-set $\Gamma(t) = \{X \mid \psi(X,t) = 0\}$.

We use a simple form of the level-set equation:

$$\frac{d}{dt}\psi = -P(x)\|\nabla\psi\|, \tag{3.1}$$

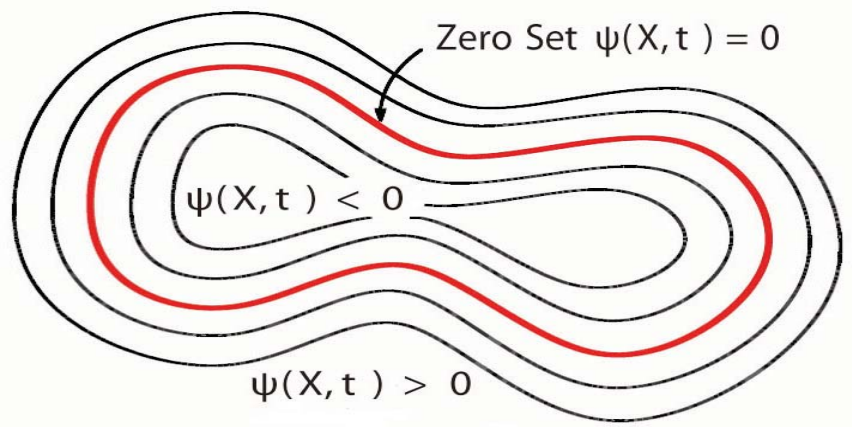


Figure 3.2: Concept of zero set in a level-set.

where P is a propagation (expansion) term. This propagation term of the level-set function is next defined in terms of a speed image. In our approach, the propagation of the contour is defined using the gradient of the albedo such that the propagation term is high in uniform areas and low close to pattern boundaries. We define the zero level-set as the contour propagated at a certain time T (for example, $T = 0.2$ seconds).

Speed image A speed image represents how fast a contour can propagate for every point of the range image. This speed should depend on the homogeneity of the albedo for every point compared with their local neighborhood. The speed image in the proposed method is computed from the gradient magnitude image of albedo that is obtained by applying the Gradient Magnitude filter to the albedo image[1]. Mapping should be done in such a way that the propagation speed of the front is very low with high albedo

[1]We remark that in the case of general shape models, points may not be aligned in a grid manner and the Gradient Magnitude Filter cannot be directly applied. However, it is still possible to compute the gradient magnitude image as far as the neighborhood relationship between points is provided. In fact, we can compute the gradient magnitude of a point as the sum of the differences in albedo between the point and its adjacent points. The obtained gradient magnitude image represents the local homogeneity of albedo and can thus be effectively mapped to the speed image.

gradients while it speeds up in low gradient areas. This arrangement makes the contour propagate until it reaches the edges of albedo patterns and it then slows down in front of these edges.

We employ the Sigmoid function, S, to compute the speed image since it provides numerous control parameters that can be customized to shape an accurate speed image. In fact, it has a mechanism for focusing attention on a particular set of values and progressively attenuating the values outside that range.

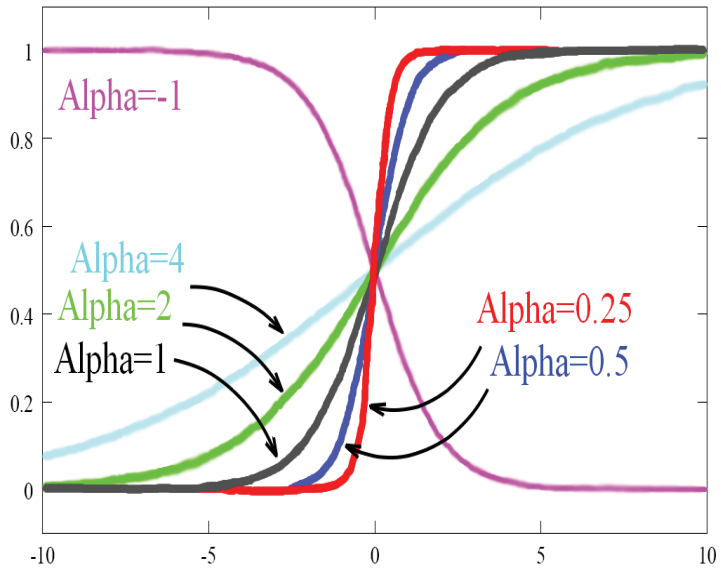
$$S(I) = \frac{1}{1 + e^{-(\frac{I-\beta}{\alpha})}}, \tag{3.2}$$

where I is the gradient magnitude image of albedo, α defines the width of the gradient intensity range, and β defines the gradient intensity around which the range is centered (see Fig. 3.3). As suggested by Ibanez *et al.* [38], α and β are found as follows. For the gradient magnitude image, let us denote the minimum value along the contour of the albedo patterns as $K1$. We denote the average value of the gradient magnitude in the homogeneous regions of the albedo image as $K2$. Then, β is $\frac{K1+K2}{2}$ while α is $\frac{K2-K1}{6}$. Figure 3.4 shows a concrete example of the albedo, gradient and speed images obtained using a synthetic data.

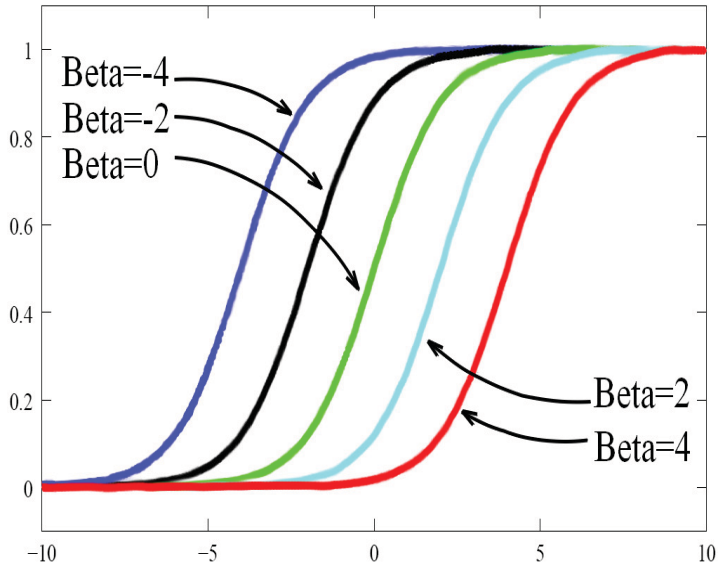
Generation of regions

For each point p, we define a time image T_p. For each pixel x of T_p, $T_p(x)$ represents the time required for the level-set function to propagate from p to x. Starting from point p, a region is grown by using the four-neighborhood and by adding points into the region, such that T_p on these points is less than a threshold (e.g., 0.2 sec) (see Fig. 3.5). The maximum size of the region is enforced, which allows us to discriminate points in excessively homogeneous areas.

This region grows in homogeneous areas and stops in front of the contour of albedo patterns. Consequently, while the size of the region increases, the homogeneity of the region is preserved. Moreover, the growth of the region is adapted to the distribution of albedo and to data noise in the neighborhood



(a) α defines width of window.

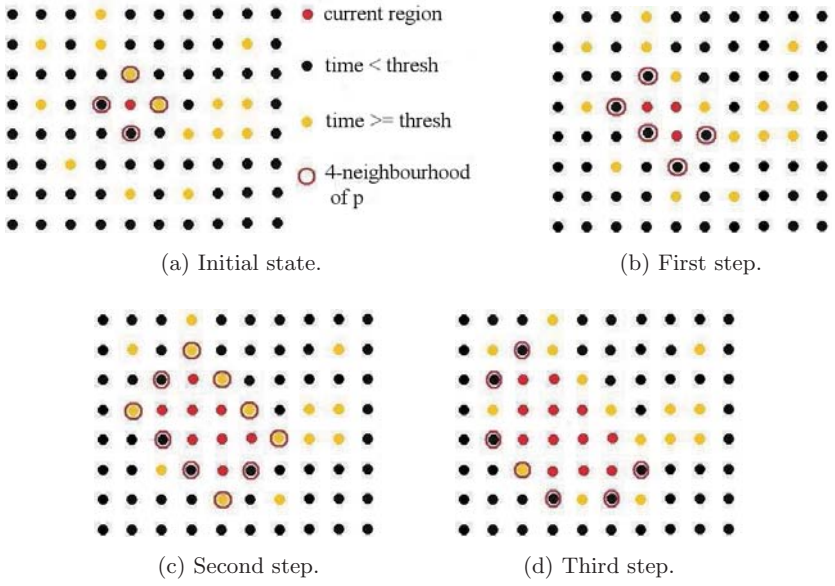


(b) β defines center of window.

Figure 3.3: Effects of various parameters in Sigmoid function.

(a) Albedo image. (b) Gradient image. (c) Speed image.

Figure 3.4: Example of albedo, gradient and speed images obtained using synthetic data.



(a) Initial state. (b) First step.

(c) Second step. (d) Third step.

Figure 3.5: Adaptively defined region using four-neighbourhood.

of each point.

We then transform each region into their local coordinate system so that the comparison between two regions becomes independent of the pose of the object. That is, we transform a region into the coordinate system defined by the normalized principal axis computed for this region. Given point x and its region $R(x)$, the PCA is carried out to obtain the eigenvalues and eigenvectors of the covariance matrix of $R(x)$. The three eigenvectors are then normalized to define a new basis $B = (e_1, e_2, e_3)$ where e_1 corresponds to the normalized eigenvector with the greatest eigenvalue and e_3 corresponds to the eigenvector with the smallest eigenvalue. $R(x)$ is then transformed to the new basis B to obtain the final attribute of x. We remark that the transformation is done independently of the color of points inside the region.

As a result, a reliable region is adaptively generated depending on each point. The local distribution of albedo of 3D points inside this region defines an attribute for each point. The attributes obtained in this way enhance robustness in evaluating similarity to find correspondences. Figures 3.6, 3.7 and 3.8 show a concrete example of the different steps in the region growing process. Figure 3.6 shows the initial state. Two corresponding points in two overlapping range images are regarded as seed points for the growth of the regions, and a maximum size of the regions is enforced. Figures 3.7 and 3.8 shows the regions at different time. Namely after the first, the third, the sixth and the tenth iteration of the region growing process. As expected, we observed that the regions of two corresponding points grow in a similar way.

3.3.3 Restrictions on search area

Poor estimates of the albedo are obtained with rough estimates of illumination. These will particularly be far less useful around the border of the range images. Consequently, we do not take into consideration points near the borders of the range images to reduce their influence. We denote the restricted area of the range image, P, as $C(P)$.

(a) Range image 1.

(b) Range image 2.

Initial seed points in two range images (in red).

(c) Range image 1.

(d) Range image 2.

Maximum size of the regions (in vivid green).

Figure 3.6: Initial state for the growth of the regions of two matched points in two range images.

(a) Range image 1.

(b) Range image 2.

The regions at the first iteration (in red).

(c) Range image 1.

(d) Range image 2.

The regions at the third iteration (in red).

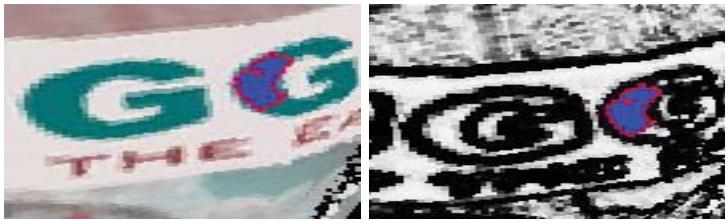
Figure 3.7: The growth of the regions at different time.

(a) Range image 1.

(b) Range image 2.

The regions at the sixth iteration (in red).

(c) Range image 1.



(d) Range image 2.

The regions at the tenth iteration (in red).

Figure 3.8: The growth of the regions at different time.

Moreover, we dynamically control a search area (area where a possible match for p will be searched), $\Omega(p)$, for each point $p \in P$ in the other range image Q, whose center is the projection of p on Q (i.e., the point of Q closest to p in the sense of the Euclidean distance). $\Omega(p)$ is defined such that the closer to convergence registration becomes, the smaller $\Omega(p)$ becomes. This control enhances the stability and accuracy of registration. $\Omega(p)$ is defined independently of $C(Q)$.

Points in large homogeneous areas are not discriminative enough to be used in registration. Such points are detected using the size of their regions. Indeed, the regions of such points are close to the maximum given beforehand. Therefore we do not take into account points whose regions are larger than 95% of the maximum size of regions.

3.3.4 Evaluation of similarities using the albedo

We define a similarity metric between two points using their attribute to find correspondences across two range images.

Letting p be a point in $C(P)$ and q be a point in $\Omega(p)$, we denote the regions corresponding to p and q by $R(p)$ and $R(q)$, respectively. For each point $m \in R(p)$, we define its corresponding point $n(m)_q \in R(q)$ (Fig. 3.9). The corresponding point $n(m)_q$ is defined by

$$\arg \min_{x \in R(q)} (\|T(\overrightarrow{pm}) - \overrightarrow{qx}\|_2). \tag{3.3}$$

For each pair $(m, n(m)_q)$, we define a weight $\omega_{(m,q)}$ such as

$$\begin{aligned} \omega_{(m,q)} &= 0 \ \text{ if } \|T(\overrightarrow{pm}) - \overrightarrow{qn(m)_q}\|_2 > thresh, \\ \omega_{(m,q)} &= 1 \ \text{ if } \|T(\overrightarrow{pm}) - \overrightarrow{qn(m)_q}\|_2 \leq thresh, \end{aligned} \tag{3.4}$$

where $thresh$ is a distance threshold (e.g., 0.4 mm if the resolution of range images is 0.5 mm). We can similarly define the corresponding point and weight for each point in $R(q)$.

The similarity function between two points p and q is then defined as

the weighted sum of the differences of the albedo of corresponding pairs:

$$\begin{aligned} L(p,q) = & \frac{size(R(p)) + size(R(q))}{(\sum_{m\in R(p)} \omega_{(m,q)} + \sum_{m\in R(q)} \omega_{(m,p)})^2} \times \\ & \Big\{ \sum_{m\in R(p)} \omega_{(m,q)} \|\overrightarrow{alb(m)} - \overrightarrow{alb(n(m)_q)}\|_2^2 + \\ & \sum_{m\in R(q)} \omega_{(m,p)} \|\overrightarrow{alb(m)} - \overrightarrow{alb(n(m)_p)}\|_2^2 \Big\}, \end{aligned} \tag{3.5}$$

where $size(R(\cdot))$ is the number of points in $R(\cdot)$ and $\overrightarrow{alb(m)}$ is the albedo vector of point m, computed using the Lambertian model of reflectance for each color channel:

$$\overrightarrow{alb(m)} = \frac{\overrightarrow{c(m)}}{\overrightarrow{norm(m)}^\top M \overrightarrow{norm(m)}}, \tag{3.6}$$

where $\overrightarrow{norm(m)}$ is the normal of the surface at point m, M is the illumination matrix and $\overrightarrow{c(m)}$ is the RGB vector of point m.

If p and q are matches and two regions $R(p)$ and $R(q)$ represent the same part of the object viewed from different viewpoints, then $m \in R(p)$ and $n(m)_q \in R(q)$ are two sampled points with small distance (the distance between the two points is smaller than the resolution of the range images) viewed from different viewpoints. Thus, their albedo is likely to be similar. Therefore, the function, L, becomes small for points p and q. In contrast, L increases for points with different regions. As we can see, support by corresponding points inside the region defines the similarity between two points of interest. This leads to similarity being robustly and stably evaluated.

Note that if $\sum_{m\in R(p)} \omega_{(m,q)}$ or $\sum_{m\in R(q)} \omega_{(m,p)}$ is less than $0.6\times size(R(p))$, *resp.* $0.6\times size(R(q))$, then the pair, (p,q), is not considered to be a possible match. This is because if (p,q) is a correct match, we can expect that there will be a sufficient number of matches between $R(p)$ and $R(q)$. Moreover, considering computational time, if the bounding boxes of $R(p)$ and $R(q)$ differ (up to the resolution of the range images), then the pair, (p,q), is not considered to be a possible match.

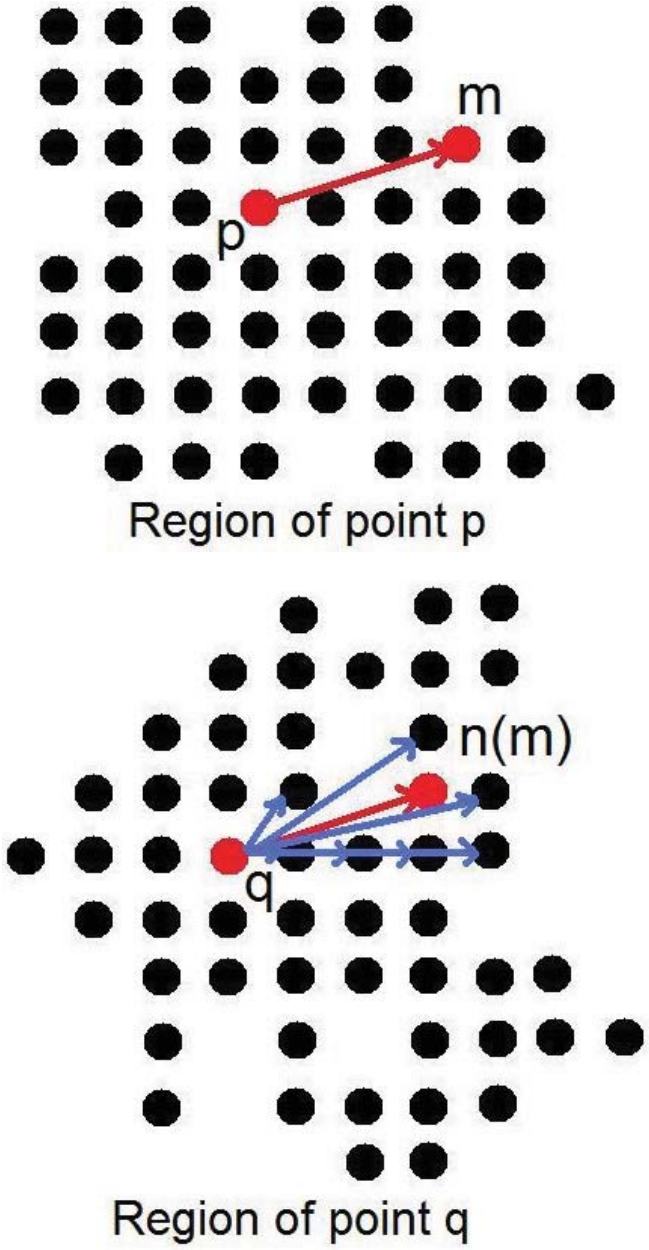


Figure 3.9: Searching for corresponding point of m.

3.3.5 Matching

We dynamically create a matching list based on similarity scores computed as explained above. We search for a set of matches such that each point has at most one corresponding point and that the sum of the scores between all matches is minimized.

We compute a list of possible matches for each point sorted in the ascending order of similarity scores. Taking into consideration computational time, we enforce a maximum tolerance threshold for possible matches. The matching list is then iteratively and dynamically constructed. The match with the lowest similarity score at each iteration is chosen and added to the matching list. The two matching points are then removed from all the lists of possible matches and these lists are updated accordingly (resorted). We iterate this process until no more possible matches remain to obtain the final matching list.

At the end of this step, we have a reliable and consistent list of matches that does not contain any isolated points. Indeed, the region grown from an isolated point is empty and this point will not be a candidate for any match.

3.3.6 Elimination of incorrect matches

The list of matches that is obtained cannot be always directly used as input in the step to estimate transformation. This is because large amounts of noise or repetitive patterns in the albedo distribution may cause incorrect matches. We therefore remove such incorrect matches to enhance the robustness of estimating transformation further. To evaluate the accuracy of matches, we use the rigidity constraint of surfaces. This is because the rigidity constraint does not depend on the intensity or normals and it is therefore robust against data noise.

For two corresponding pairs, (p, q) and (p', q'), in the range images P and Q, we consider point pairs (p, p') and (q, q'), which represent the same points viewed from different viewpoints. Assuming that surfaces are rigid, we can see that distances $\|\overrightarrow{pp'}\|_2$ and $\|\overrightarrow{qq'}\|_2$ should not differ too much. That

is, we define d by representing the difference between $\|\overrightarrow{pp'}\|_2$ and $\|\overrightarrow{qq'}\|_2$:

$$d = |\|\overrightarrow{pp'}\|_2 - \|\overrightarrow{qq'}\|_2|. \tag{3.7}$$

If (p, q) and (p', q') are correct matches, then d should be smaller than a threshold, $Tdist$ (e.g., 1.0mm, for a resolution of 0.55mm). This gives us the rigidity constraint (see Fig. 3.10).

Each pair in the list of matches is evaluated with all the other pairs in the list. If the number of pairs that violates the rigidity constraint exceeds a certain percentage, $Perc$ (e.g., 50%), of the pairs, then the current pair is considered to be an incorrect match and is removed from the list.

Assuming that the majority of the obtained pairs are correct matches, this method efficiently removes incorrect matches from the list of pairs obtained in Section 3.3.4. To handle cases where the majority of matches are incorrect, we dynamically update the parameters, $Tdist$ and $Perc$, in such a way that the elimination step is tolerant to mismatches at the beginning of registration, and stringent against mismatches at the end of registration. This is because, at the beginning of registration, we may have a large number of mismatches and only a rough estimate of the transformation is sufficient. In contrast, as registration becomes more accurate, there are less mismatches and we aim at eliminating as many mismatches as possible to refine the estimation of transformation.

Our proposed approach is simpler than RANSAC in implementation. In fact, the RANdom SAmple Consensus (RANSAC) algorithm iterates the following three steps until the probability of finding the correct transformation becomes sufficiently large. (1) Three matches are randomly chosen among all matches in the list. (2) The rigid transformation aligning the three matches is estimated. (3) All the other matched points in one range image are transformed to the other range image using the estimated transformation and the consistency of matched points with the list is evaluated. The transformation that is the most consistent with the list of corresponding points is identified as the solution, and matches inconsistent with the solution are identified as outliers. On the other hand, our proposed technique to

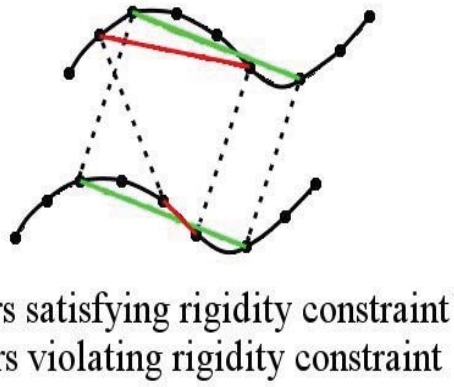


Figure 3.10: Principle of rigidity constraint.

eliminate incorrect matches requires only to verify the validity of the rigidity constraint for each pair of matches. We remark that our technique needs $O(n^2)$ operations to remove incorrect matches.

3.3.7 Estimation of rigid transformation

We use the WLS method [49] to estimate transformation as accurately as possible. It weights each pair with the Euclidean distance between two corresponding points during the least squares minimization. These weights represent the feasibility of the correspondence of paired points. Minimization is iterated while updating the weight of each pair. The resulting transformation obtained with this method is more accurate than that with the standard least squares method.

3.4 Computational complexity analysis

Our proposed algorithm has its input of two range images with N points and outputs the rigid transformation aligning the two range images. Here we briefly give analysis on the computational complexity to our proposed algorithm. We refer to Fig. 3.1 for the different steps of our method and give the computational complexity for each of these steps.

The first part of our proposed method (Attribute definition) takes $O(N)$ operations. We first estimate the albedo of each point of the two range images using eq. 3.6, which takes $O(N)$ operations. In order to compute the speed images, gradient images and sigmoid images are computed, which is also done in $O(N)$ operations. Let Cst denote a given maximum number of points for a region. Growing the region of a point takes $O(Cst)$ operations (because we do not compute the propagation time outside the maximum region). Therefore, computing the region for each point of the two range images takes $O(NCst)$ operations. Because Cst is fixed during the process, the computational complexity of the Attribute definition is finally $O(N)$.

The second part of our method (Matching) takes $O(N^3 \log(N))$ operations. For a point p, the corresponding point is searched in $\Omega(p)$, where $\Omega(\cdot)$ is defined in Section 3.3.3. Let ω be the size of $\Omega(\cdot)$. For one possible match, the similarity score is computed in $O(Cst^2)$ operations. Thus computing all similarity scores takes $O(N\omega Cst^2) = O(N\omega)$ operations (Cst is constant). The following loop is then executed until no possible match exists: (1) The list of possible matches is sorted ($O(N\omega \log(N\omega))$ operations). (2) The best match is taken and the two corresponding points are removed from the list ($O(N\omega)$ operations). (3) Go to (1) with $O((N-2)\omega)$ possible matches. The computational complexity of this loop is $O(\sum_{i=0}^{i=N/2} i\omega \log(i\omega))$. We observe that

$$\sum_{i=0}^{i=N/2} i\omega \log(i\omega) \leq N\omega(\log(\prod_{i=0}^{i=N/2} i) + N\log(\omega)) \leq N\omega(\log(N^N) + N\log(\omega)) \leq N^2\omega \log(N\omega). \tag{3.8}$$

As a consequence, the loop is executed in $O(N^2\omega \log(N\omega))$ operations. Even though in practice ω is monotonically decreasing at each iteration, ω remains constant dominated by N in the worst case. Thus $\omega = O(N)$ and the computational complexity of the loop becomes $O(N^3 \log(N^2)) = O(N^3 \log(N))$. We then eliminate incorrect matches with $O(N^2)$ operations. Thus the computational complexity of the "Matching part" is finally $O(N^3 \log(N))$.

The computational complexity of the Weighted Least Square algorithm is $O(N^2)$ and thus the total computational complexity of our proposed algorithm is: $O(N) + O(N^3 \log(N)) + O(N^2) = O(N^3 \log(N))$.

In this work, we focus on accuracy of the transformation estimation and leave reducing the computational cost for future work. However, some ideas are available to reduce the computational cost of our algorithm in implementation. The computation of each point's attribute or similarity score is done independently and thus our method can be highly parallelized. Theoretically, with an infinite number of processors, the computational complexity of the "Attribute definition" part could go from $O(N)$ to $O(1)$ if parallelized. With modern graphic cards, a large number of processors available would drastically improve the computational time of our technique. It would also be interesting to use a random algorithm to select possible matches, as proposed in [27] for example. This would decrease the number of possible matches and thus speed up the algorithm.

3.5 Experiments

We evaluated our method using synthetic and real data and compared it with ICPA and ICP using both chromaticity and geometric features (which we call ICP-CG). This comparative study is thus useful for determining the effectiveness of different methods of registering overlapping range images of Lambertian surfaces devoid of salient geometric features. We selected these two methods for two main reasons:

- ICPA is, to the best of our knowledge, the most recent method that uses the albedo for registering overlapping range images. Because the albedo is invariant to changes in object poses under fixed illumination and fixed viewpoints, it is useful for computing correspondences in points between overlapping range images.
- ICP-CG[2] is the standard algorithm for registering overlapping range

[2] The ICP-CG algorithm is an ICP algorithm that uses Euclidean distance and chro-

images. Because chromaticity is tolerant to some extent against changes in illumination, using chromaticity in addition to geometric features improves the accuracy of registration.

Because the objects used in our experiments were simple in shape (like cylinders), registration using only geometric characteristics did not work well. Consequently, standard root mean square (RMS) point-to-point Euclidean error was not relevant for evaluating the registration results in our case. This paper discusses our evaluation of the registration results by comparing the transformation we obtained with the ground truth transformation.

We use an angular measure of errors for rotation like Barron *et al.* [3] and the Euclidean error for translation. Let (R_g, T_g) be the ground truth transformation and (R_e, T_e) be the estimated transformation with R_g, R_e rotations and T_g, T_e translations. A rotation, $R = \cos(\frac{\alpha}{2}) + \overrightarrow{u}\sin(\frac{\alpha}{2})$, is represented using quaternions, where α is the angle of rotation and $\overrightarrow{u}$ is the unit vector representing rotation axis. Let res be the resolution and d the depth of range images, we define err, which is the error of the obtained transformation as:

$$err = \frac{\Theta d + ||T_g - T_e||}{res}, \tag{3.10}$$

where Θ is the angle between the normalized ground truth rotation $\frac{R_g}{||R_g||}$, and the normalized estimated rotation, $\frac{R_e}{||R_e||}$. The err represents the error

maticity similarity to find corresponding points. At iteration k, the current transformation $Trans$ aligning the two range images is estimated. For each point p of the range image P, the projection q of point $Trans(p)$ on the range image Q is found and then the closest point to p in the sense of chromaticity distance is found in the region $\Omega(p)$ as the corresponding point of p, where $\Omega(p)$ is defined in section 3.3.3. If $\sigma_R(x)$, $\sigma_G(x)$ and $\sigma_B(x)$ are the red, green and blue channels for the chromaticity of a point x, then the chromaticity distance $dchrom(p, q)$, between two points p and q is:

$$dchrom = \sqrt{(\sigma_R(p) - \sigma_R(q))^2 + (\sigma_G(p) - \sigma_G(q))^2 + (\sigma_B(p) - \sigma_B(q))^2}. \tag{3.9}$$

The obtained list of corresponding pairs is then sorted and only a certain percentage (70% in the experiments) of the pairs with the best score is kept.

in terms of neighboring points. It is thus an objective and informative criterion to evaluate the accuracy of the different methods of registration. All results from these experiments have been presented with the estimated albedo image. We compute the resolution res of a range image as the average of the 2D distances between adjacent points projected onto the 2D grid induced by the laser scanner (the depth is not considered). Note that the resolution is not computed based on neighboring relationship in the 3D space.

3.5.1 Definition of parameters

ICPA and ICP-CG require three parameters: max_it (the maximum iterations for registration), $conv\_thresh$ (the threshold for convergence), and $percentage$ (the percentage of matches to be eliminated).

Our proposed method requires five parameters: max_it, $conv\_thresh$, and $thresh$ (the threshold to grow the regions), max_size (the maximum size of a region), and tol_thresh (the threshold for tolerance to similarity). The two thresholds, $Tdist$ and $Perc$, are dynamically defined for the elimination step depending on the current state of registration. At the beginning of registration, $Tdist = 8 \times$"resolution of the image" and $Perc = 70\%$, and at the end of registration, $Tdist = 2 \times$"resolution of the image" and $Perc = 30\%$. The tolerance tol_thresh was set to 0.01 for synthetic data and to 0.5 for real data. We remark that the restricted search area in Section 3.3.3 was equally applied to our method, ICPA and ICP-CG.

3.5.2 Evaluation with synthetic data

We conducted experiments with synthetic data to test the robustness of the proposed method against data noise and changes in illumination. The synthetic data were obtained with 3D modeler software (3D Studio Max) (see Table 3.1). The exact albedo image and the exact illumination, modeled with a direction and a color vector, were known. The illumination represents a single distant light-source point. Assuming the object to be Lambertian,

Table 3.1: Description of synthetic data.

Nb_Points	Resolution	Expected_rot (angle; axis)	Expected_trans
30650	0.01 mm	$(18.00; 0.01, 0.99, -0.03)$	$(-0.02, 0.00, -0.01)$

(a) First image. (b) Second image.

Figure 3.11: Input synthetic data.

we rendered the brightness at the surface using the Lambertian reflection model. That is, given the albedo ρ and normal $\overrightarrow{norm}$ at a surface point, and light-source direction $\overrightarrow{l}$ and light-source brightness L, the brightness, I, at this point is computed as: $I = \rho L \overrightarrow{norm} \cdot \overrightarrow{l}$. We estimated the albedo using an approximation of the exact illumination to test the robustness of our proposed method (see Figs. 3.11 and 3.12).

We manually established a rough pre-alignment of the two range images before applying our method, which allowed us to simulate the case where the input data were captured from two viewpoints rotationally differentiated by 18.00 degrees around the axis $(0.01, 0.99, -0.03)$. We used the same sets of parameters for all synthetic experiments: $max\_it = 10$; $conv\_thresh = 0.02$ radian; for ICPA, $percentage = 30\%$; for the proposed method, $thresh = 0.05$ seconds, $max\_size = 0.1$ mm and $tol\_thres = 0.1$.

(a) Superimposed input images.

(b) Registration results with albedo image.

Figure 3.12: Results obtained with our proposed method in ideal case.

To verify the effects against estimated illumination, we randomly rotated the direction vector of exact illumination. That is, let the angle between the perturbed direction vector and the ground truth direction vector of illumination be (θ, ϕ), in which ϕ is a uniform random number from 0 degree to 360 degrees. We evaluated our method with different values of θ, and therefore different estimates of illumination. The value of θ was changed from 0 to 13 degrees by 0.7 degrees. For all values of θ, we applied our method 30 times under the same initial conditions.

Figure 3.13 plots the quantitative evaluation of registration results in terms of averages and variances in error in results obtained with different estimates of illumination. The results obtained with ICPA have also been shown for comparison. Since ICP-CG failed in registration because of geometrical symmetries in the shape of the object, we did not carry out comparative experiments with ICP-CG. As expected, we find that our method is in average more accurate and more stable than ICPA in estimating transformation on average. The results obtained with ICPA became inaccurate and unstable as soon as the estimated illumination differed slightly from

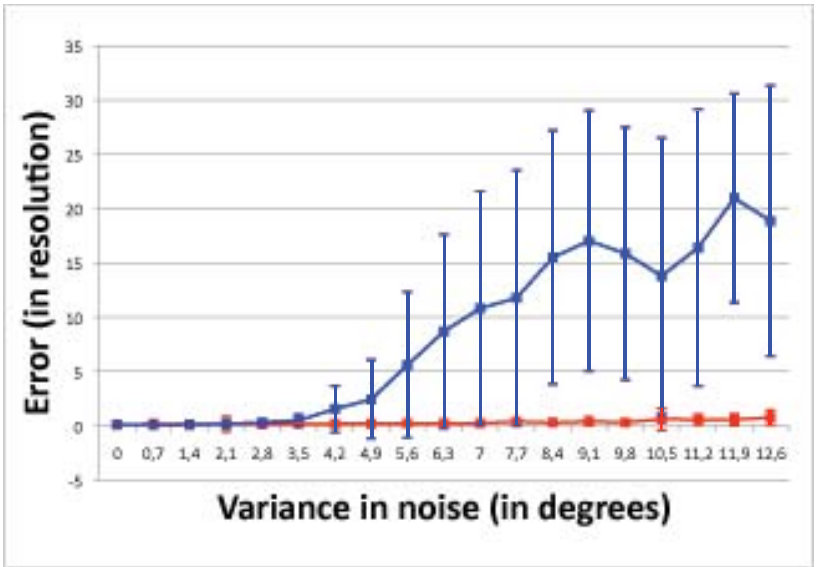


Figure 3.13: Results for various different illuminations. (red: our method, blue: ICPA)

being exact (for θ greater than 4.2 degrees). The results with ICPA were largely unsatisfactory when the estimates of illumination were not close to being exact. In contrast, our proposed approach gained satisfactory results for θ up to 12 degrees. Note that, in addition to small errors, the variance was also small, which proves the robustness and reliability of our proposed approach against noise in illumination. Therefore, we can conclude that it is more robust to changes in illumination than ICPA. Moreover, for exact estimates of illumination, our proposed method achieves registration that is as accurate as that with ICPA.

To verify the robustness of our descriptor against changes in pose, we evaluated our method under various initial relative poses of the two input synthetic range images. Namely, we rotated the second range image around the vertical axis, with different rotation angles, ranging from 1.0 degree to 53.0 degrees. We notice that when the rotation angle increases, the number of maximum iterations required for convergence increases (starting from

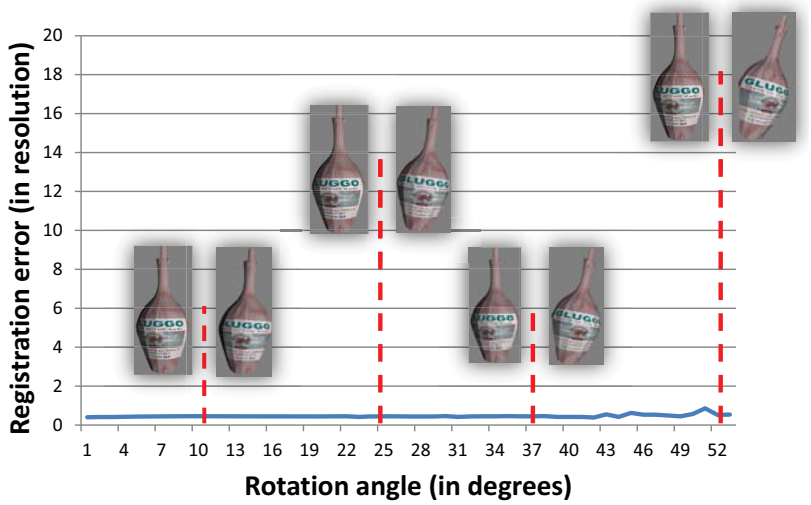


Figure 3.14: Results for different initial relative poses when the second range image is rotated in the reverse clock-wise direction.

$max\_it = 10$ for a rotation angle of 1.0 degree up to $max\_it = 25$ for a rotation angle of 53.0 degrees). Figure 3.14 and Figure 3.15 show the results obtained with our method when the second range image is rotated in the reverse clock-wise direction and in the clock-wise direction (respectively). As expected, we can see that our proposed descriptor is almost insensitive to changes in pose. In Figure 3.15, the failure that arises for a rotation angle greater than 52.0 degrees can be explained by the fact that many points in the second range image become invisible from the incident light. As a consequence, most of points do not have their descriptors, which prevents us from identifying the point correspondences.

Figure 3.16 plots quantitative results of registration obtained with our proposed method under various noise in intensities where the ground truth illumination was used. We applied Gaussian noise with a variance σ of several percent to the average of the image intensities. The value σ was changed from 0% to 10% by 0.5%. For each different noise intensity, we applied our method 30 times under the same initial conditions. We observe

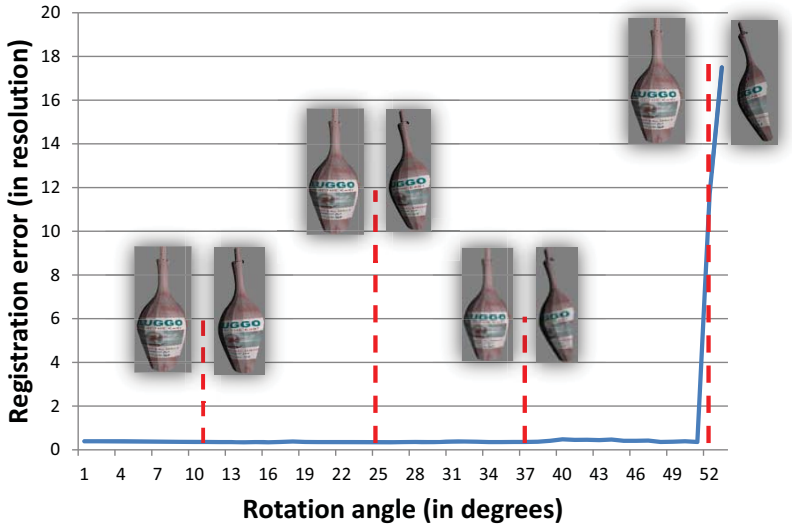


Figure 3.15: Results for different initial relative poses when the second range image is rotated in the clock-wise direction.

that even with noise with a variance of 9.5%, the largest errors are under the resolution of the range sensor.

We also conducted intensive experiments under noise added to normals (Fig. 3.17). We randomly perturbed each normal vectors. That is, let the angle between the direction of a perturbed normal and a ground truth normal be (α, ϕ), in which ϕ is a uniform random number from 0 degree to 360 degrees. We evaluated our method with different values of α, and therefore different perturbations in the normals. The value α was changed from 0 to 10 degrees by 0.6 degrees. For each different intensity of noise, we applied our method 30 times under the same initial conditions. We observe that even with noise of angle 10 degrees, the largest errors are of the same order as the resolution of the range sensor. We find that our method is stable against both geometric and photometric noise from these results.

Figure 3.18 has some particular examples of results obtained under different types of noise with our proposed method and with ICPA. The first example corresponds to noise of 17 degrees added to illumination, where

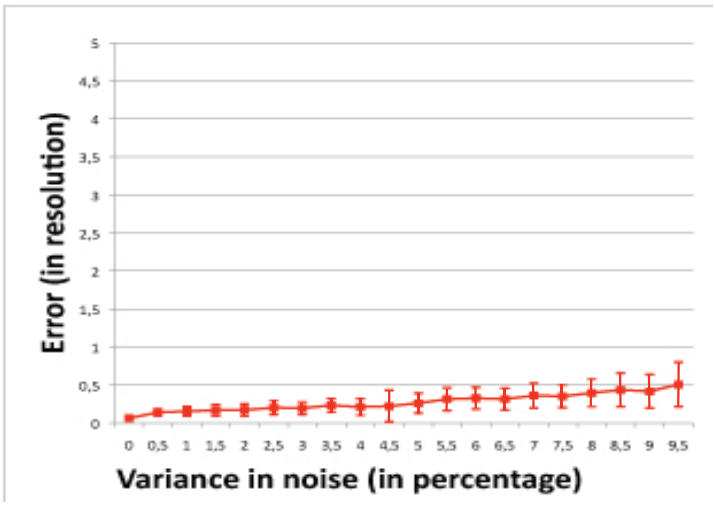


Figure 3.16: Results for noise in intensities with our method.

we obtained error of 0.49 with our method and error of 24.39 with ICPA. The second example corresponds to noise of 10% added to intensity, where we obtained error of 0.62 with our method and error of 2.43 with ICPA. The last example corresponds to noise of 6 degrees added to normals, and we obtained error of 0.62 with our method and error of 5.30 with ICPA. The quantitative results confirm the improvements in accuracy with our proposed method than with ICPA.

3.5.3 Evaluation with real data

We conducted experiments using real objects to test the effectiveness of the proposed method. In addition to ICPA and ICP-CG, we also compared the proposed method with SURF. SURF is a fast and robust well-known detector and descriptor that is often used for matching points between overlapping images. SURF uses only 2D gray images and is most effective for images with large amount of distinctive patterns. Using approximations of Hessian matrices allows fast detection of distinctive points at multiple scales. Note that we used the source code for SURF provided at

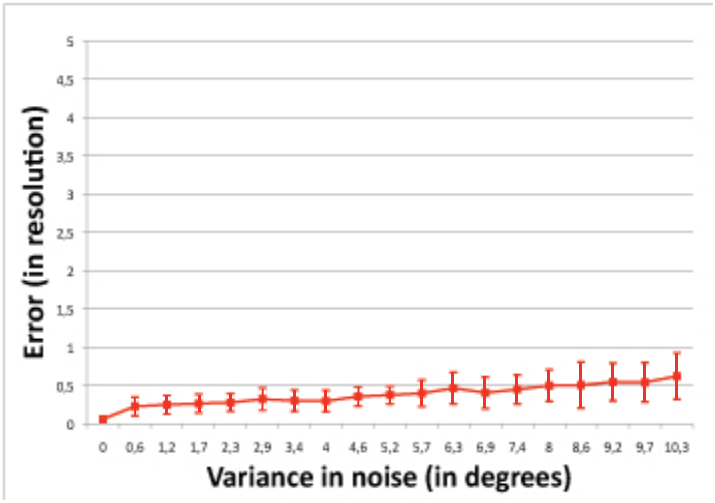


Figure 3.17: Results for noise in normals with our method.

http://www.vision.ee.ethz.ch/∼surf/ without modifications.

We employed a Konica Minolta Vivid 910 range scanner, which captures the 3D shapes and textures of objects. A mechanical system was used to rotate objects (see Fig. 3.19). The data *can* obtained using this equipment are presented in Fig. 3.20. Because the position and orientation of the range scanner are unknown, it is difficult to obtain the ground truth from the experimental setup. Accordingly, we manually computed the ground truth transformation for all data to evaluate the registration results. That is, we chose about 10 corresponding points in two range images and computed the transformation that minimized the distance between all corresponding points. We employed the ground truth obtained in this way to evaluate errors using equation (3.10).

The proposed method starts with an estimate of the geometric transformation and with a rough estimate of global illumination conditions to estimate the albedo. The initial estimate of registration is obtained by just superimposing the two captured range images. Global illumination is manually estimated roughly by rotating the direction vector of illumination and changing its brightness, which corresponds to spot light. We used the

(a) Intensity image. (b) Albedo image. (c) Our method. (d) ICPA.

(e) Intensity image. (f) Albedo image. (g) Our method. (h) ICPA.

(i) Intensity image. (j) Albedo image. (k) Our method. (l) ICPA.

Figure 3.18: Examples of registration results obtained with different types of noise. From top to bottom: results with 17 degrees of noise added to illumination, those with 10% of noise added to intensities, and results obtained with 6 degrees of noise added to normals.

Figure 3.19: Experimental set up to capture range images.

(a) First image. (b) Second image. (c) Superimposed.

Figure 3.20: Range images captured from different viewpoints.

Table 3.2: Description of data *can* used for experiment.

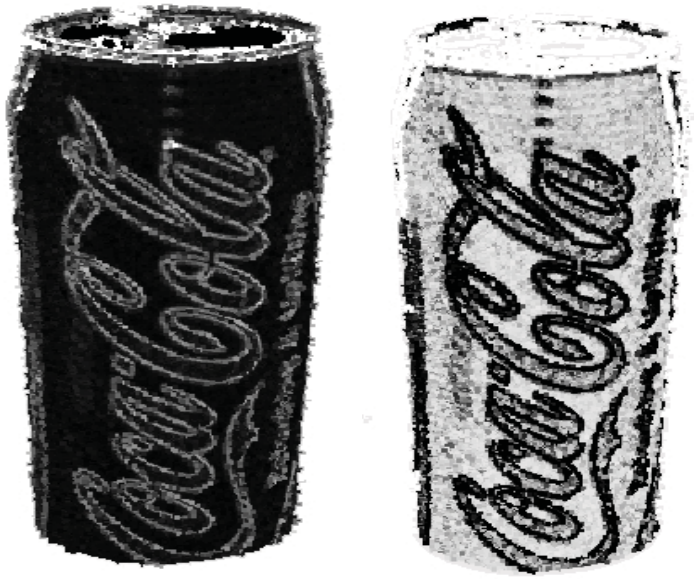
Nb_Points	Res	Expected_rot (angle; axis)	Expected_trans
35000	0.55 mm	$(17.83; 0.00, -0.94, -0.34)$	$(-8.05, 0.41, -1.47)$

same sets of parameters for all experiments: $max\_it = 10$; $conv\_thresh =$ 0.002radian; for ICP-CG and ICPA, $percentage = 30\%$; for the proposed method, $thresh = 0.2$sec, $max_size = 5.5$mm and $tol\_thres = 0.5$. This means that we do not need extra tuning for the parameters depending on objects.

We obtained two range images of a rotationally symmetric can that has a height of about 10cm and a diameter of about 5cm (Fig.3.20). Fig. 3.21 has the gradient image and the speed image computed from the estimate of albedo. We compared the results obtained with ICPA, ICP-CG, and SURF (Figs. 3.22 and 3.23) to demonstrate how effective our method is. Note that the same initial estimate was used for the four methods. The details on data are listed in Table 3.2 and the quantitative results are listed in Tables 3.3 and 3.4. Nb_Points is the number of points in the range image in these tables, Res is the resolution of the range image, $NbMPts$ represents the number of matched points at the end of the registration, $Final\_rot$ is the estimated rotation after registration, $Final\_trans$ is the estimated translation after registration, $Error$ is the error between the estimated transformation and the ground truth transformation, and $Time$ is the processing time of the registration. Note that for our method, ICPA and ICP-CG, we had 10 iterations to converge, and we observe that the improvement in the estimations of transformations became small after 3 iterations.

Because of the lack of geometric features, ICP-CG failed in registering the two range images. In addition, the registration result obtained with ICPA is also not satisfactory. The result obtained with SURF is more accurate than those with ICPA and ICP-CG because of the robustness and discriminative power of the SURF descriptor. However, we observed an error of about 4 times the resolution of the range images. This comes from

(a) Albedo first image. (b) Albedo second image.



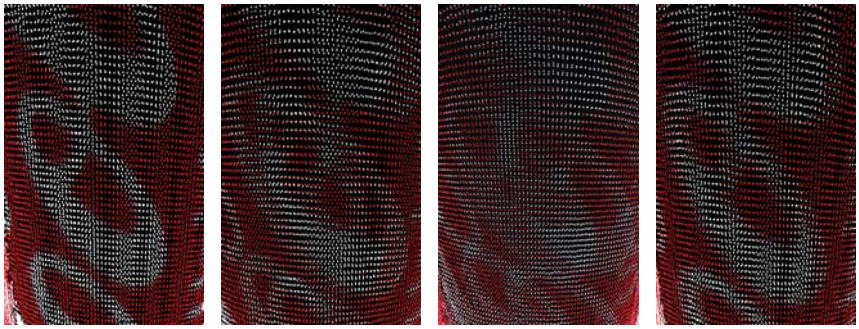
(c) Gradient image. (d) Speed image.

Figure 3.21: Estimate of albedo, gradient, and speed images.

(a) Ground Truth.

(b) Our method. (c) ICPA. (d) ICP-CG. (e) SURF.

Figure 3.22: Results obtained with data *can* using different methods.



(a) Our method. (b) ICPA. (c) ICP-CG. (d) SURF.

Figure 3.23: Zoom in on the square part in Fig. 3.22.

Table 3.3: Registration results using data *can*.

	NbMPts	Final_rot (angle; axis) Final_trans
Proposed Method	2400	$(17.83; -0.01, -0.94, -0.34)$ $(-8.00, 0.38, -1.52)$
ICPA	25000	$(8.47; -0.02, -0.96, -0.28)$ $(-5.11, -1.12, -0.95)$
ICP-CG	24000	$(4.01; -0.06, -0.98, -0.21)$ $(-2.92, -1.01, -0.57)$
SURF	222	$(15.33; 0.02, -0.94, -0.35)$ $(-6.65, 0.16, -1.04)$

Table 3.4: Quantitative evaluation of registration, using data *can*.

	Proposed Method	ICPA	ICP-CG	SURF
Error	0.22	10.66	16.53	3.91
Time	3.0 mn	5.2 mn	5.4 mn	0.0 mn

Table 3.5: Description of data *can2*.

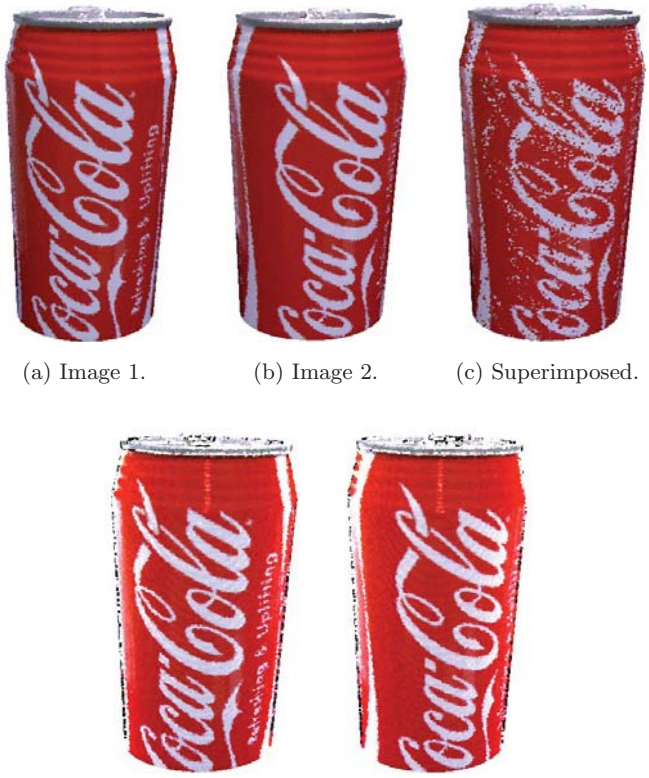
Nb_Points	Res	Expected_rot (angle; axis)	Expected_trans
94000	0.3 mm	$(17.52; 0.00, -0.94, -0.34)$	$(-9.16, 0.55, -1.38)$

Table 3.6: Registration results using data *can2*.

	NbMPts	Final_rot (angle; axis) Final_trans
Proposed Method	838	$(17.46; -0.02, -0.94, -0.34)$ $(-9.06, -0.50, -1.44)$
ICPA	44000	$(1.60; 0.35, 0.46, -0.82)$ $(-1.60, -1.36, -0.30)$
ICP-CG	44000	$(6.69; -0.01, 0.96, 0.29)$ $(1.50, -0.15, -0.09)$
SURF	330	$(12.09; 0.05, -0.98, -0.19)$ $(-5.71, -0.53, -1.41)$

the distorted texture patterns due to 2D projection. In contrast, we can see significant improvements in the registration obtained with the proposed method. It is remarkable that 15% of matched pairs were eliminated as incorrect matches in our method and the accuracy obtained was about 0.2 times the resolution of the range images. The error obtained is below the accuracy of the range sensor; it is not possible to obtain better accuracy to register these data. When the search area (see Section 3.3.3) was not dynamically restricted by $\Omega(\cdot)$, the results obtained were less accurate for all ICP variants. Indeed, the errors here were 6.74, 12.83, and 19.96 times the resolution of the range images, with the proposed method, ICPA and ICP-CG respectively.

We captured two range images of the same object under the same situation using a lens with a different focal length to check the effect of the resolution of range images in our proposed method. The new lens zoomed



(a) Image 1. (b) Image 2. (c) Superimposed.

(d) Albedo 1. (e) Albedo 2.

Figure 3.24: Range images captured from different viewpoints with zoom in, and albedo estimations.

Table 3.7: Quantitative evaluation of registration, using data *can2*.

	Proposed Method	ICPA	ICP-CG	SURF
Error	0.42	40.60	55.28	35.6
Time	16.13 mn	31.0 mn	29.6 mn	0.0 mn

(a) Ground Truth.

(b) Our method.

(c) ICPA.

(d) ICP-CG.

(e) SURF.

Figure 3.25: Results obtained with data *can2* using different methods.

in on the object surface, which increased the resolution of the range images. Description of the new data *can*2 is listed in Table 3.5 and Fig. 3.5. Figure 3.25 has qualitative evaluation of registration and Tables 3.6 and 3.7 list quantitative evaluation of registration. We observe that the results obtained with our proposed method are almost the same even for range images with different resolutions. This can be understood by the fact that we use neighborhood relationship to grow regions, which is independent to the resolution. Remark that for these data, we changed the parameter max_size to 3.0 mm only to reduce the computational time.

Figures 3.29, 3.30, and 3.31 show the results obtained with the different objects called *hand*, *box*, and *candy*. Table 3.8 lists the description of data before registration. Note that *hand* has a height of about 20cm and a width of about 10cm, *box* has a height of about 10cm and a width of about 20cm, and *candy* has a height of about 10cm and a width of about 15cm. Figs. 3.27 and 3.28 show initial estimates of registration and global illumination. Figs. 3.29, 3.30, and 3.31 compare the results obtained with our method, ICPA, ICP-CG, and SURF. The quantitative results for these experiments are listed in Tables 3.9 and 3.10. Identified incorrect matches were 35% for *hand* data, 7% for *box* data and 10% for *candy* data. Note that for the data *box*, we slightly changed the tolerance for outliers. That is, we set $Tdist$ to be $4 \times$ the "resolution of the image" at the end of registration.

Table 3.11 has the average of computational time required for the different steps of registration for our proposed method, ICPA and ICP-CG. We used an Intel Core 2 Duo CPU 3GHz and all real data for the experiments. We do not show the time required for the registration using SURF. This is because when using SURF the registration is almost instantaneous compared with the other methods. Remark that the computational time were long compared to standards for registration methods. This is because we aimed at estimating the transformation aligning two range images as accurately as possible without considering time consumption. In this sense, we enforced ten iterations for the registration while in general, with only

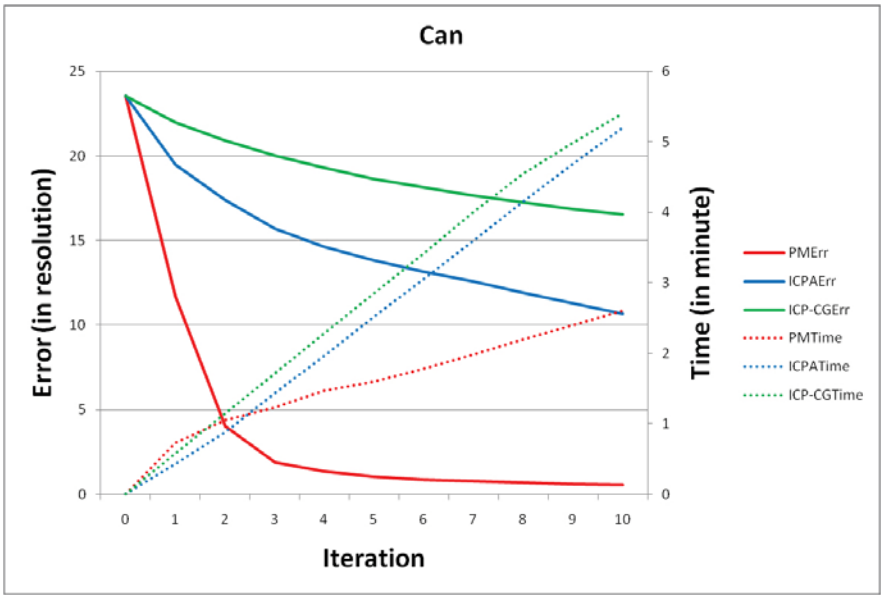


Figure 3.26: Error versus computational time for each iteration during registration using data *can* (PM stands for Proposed Method).

three iterations the methods almost converged. If a compromise between accuracy and processing time is searched, relaxing the threshold value for convergence will reduce the processing time while the registration accuracy decreases (we expect that an acceptable compromise will be around three iterations). In addition, we remark that the processing time of our proposed method was shorter than that of ICPA and ICP-CG. This is because in our proposed method, points without discriminative albedo (points in a large uniform area for example) are not considered in the matching step. As a consequence, the number of corresponding points obtained after matching was less with our proposed method and thus the outlier elimination step required less time.

We also show in Fig. 3.26 the computational time as well as error attained at each iteration during registration for our proposed method, ICPA, and ICP-CG when using data *can* (we had similar results for the other data). We observe that the error drastically decreases after the first few iterations

Table 3.8: Description of data *hand*, *box*, and *candy*.

	Nb_Points	Res	Expected_rot (angle; axis) Expected_trans
hand	50000	0.55 mm	$(17.66; -0.02, -0.94, -0.34)$ $(-7.32, 6.64, -0, 80)$
box	100000	0.55 mm	$(17.62; -0.02, -0.94, -0.34)$ $(-4.96, -0.72, -2.06)$
candy	50000	0.55 mm	$(13.99; -0.02, -0.93, -0.36)$ $(-11.70, 1.58, 2.82)$

with our proposed method. A small error means that the estimated transformation is close to the exact one. The initial estimate of transformation is far from the exact one and therefore the estimated transformation drastically changes during the first few iterations of registration with our proposed method. Consequently, the size of the search area (Ω), which decreases proportionally to the change in the estimated transformation, also drastically decreases during the first few iterations. Therefore, for each point of the two range images, the search for the corresponding point is restricted to a small area after the first few iterations, which decreases the computational time during the subsequent iterations. In contrast, the error decreases slowly with ICPA and ICP-CG during the iterations of registration and therefore the size of Ω also decreases slowly. Consequently, the computational time for each iteration of registration with ICPA and ICP-CG remains approximately constant. This explains the reason why the computational time for our proposed method decreases drastically with the first few iteration, while in contrast, that for ICPA and ICP-CG remains approximately constant. Finally, after ten iterations, the computational time of registration with our proposed method is far less than that with ICPA and ICP-CG.

Methods using only geometric features are not sufficient because of the lack of discriminative geometric features. Moreover, color information like chromaticity cannot be directly used because of changes in color due to il-

lumination. In fact, neither ICPA nor ICP-CG attained satisfactory results due to data noise and inaccurate estimates of illumination. The SURF could not accurately register the range images, either. This is because of distortions due to 2D projection (e.g., as seen with data *candy*). In contrast, our method succeeded in registering range images with an accuracy of around the resolution of the range images for all data. Through all these experiments, our proposed method was the most accurate and stable in terms of the accuracy of estimated transformation.

3.6 Conclusion

We introduced region-based registration of range images using reflectance attributes obtained under rough estimates of illumination conditions. Our method stably extracts reliable attributes that capture the local distribution of albedo on the object surface. These attributes are defined by adaptively growing regions that are generated using a level-set method. Such attributes are used to evaluate the similarity of points to robustly obtain correspondences in points even under rough estimates of illumination conditions. Our method also efficiently removes mismatches by using the rigidity constraint of surfaces, which enhances the robustness of the registration process. Our experiments using synthetic and real data demonstrated improvements in the robustness and the accuracy of registration results under rough estimates of illumination.

In this chapter, we assumed the Lambertian reflectance. However, real objects do present both Lambertian and specular reflection components. When the specular component has few impact on the object's appearance, using the Lambertian model is justified. However, when the specular component plays a significant role in the object's appearance, the estimated albedo using the Lambertian model becomes unreliable for matching points across pairs of range images. In particular, the specular highlights (which are not a surface intrinsic attribute) are propagated into the albedo estimate. In the next chapter, we will investigate how to recover albedo from two specular

(a) First image. (b) Second image. (c) Superimposed.

(d) First image. (e) Second image. (f) Superimposed.

(g) First image. (h) Second image. (i) Superimposed.

Figure 3.27: Initial state for data *hand* (top), *box* (middle), and *candy* (bottom).

(a) Albedo first image.

(b) Albedo second image.

(c) Albedo first image.

(d) Albedo second image.

(e) Albedo first image.

(f) Albedo second image.

Figure 3.28: Estimated albedo for data hand, box, and candy.

(a) Ground truth.

(b) Our method. (c) ICPA. (d) ICP-CG. (e) SURF.

Figure 3.29: Registration results obtained with data *hand*.

(a) Ground truth.

(b) Our method. (c) ICPA. (d) ICP-CG. (e) SURF.

Figure 3.30: Registration results obtained with data *box*.

Table 3.9: Registrations results, using data *hand*, *box*, and *candy*.

	NbMPts	Final_rot (angle; axis) Final_trans
hand		
Proposed Method	2500	$(17,84;-0.01,-0.93,-0.36)$ $(-7.05,6.66,-0.66)$
ICPA	36000	$(17.68;-0.01,-0.93,-0.36)$ $(-6.90,6.61,0.60)$
ICP-CG	37000	$(17.46;-0.01,-0.94,-0.34)$ $(-6.80,6.64,-0.55)$
SURF	189	$(7.21;0.26,-0.68,-0.68)$ $(-4.53,6.55,2.63)$
box		
Proposed Method	12000	$(17.75;-0.02,-0.94,-0.33)$ $(-4.56,-1.11,-2.21)$
ICPA	57000	$(17.64;-0.03,-0.92,-0.35)$ $(-0.14,-1.62,-2.15)$
ICP-CG	57000	$(17.67;-0.02,-0.94,-0.35)$ $(-1.51,-1.17,-1.98)$
SURF	210	$(17.39;-0.00,-0.95,-0.30)$ $(-4.25,-1.63,-2.70)$
candy		
Proposed Method	2000	$(13.72;-0.01,-0.94,-0.35)$ $(-11.46,1.43,2.62)$
ICPA	29000	$(9.94;-0.03,-0.92,-0.40)$ $(-10.52,1.30,2.54)$
ICP-CG	30000	$(11.23;-0.01,-0.94,-0.34)$ $(-10.72,1.18,2.33)$
SURF	210	$(6.83;-0.28,0.59,0.75)$ $(-5.84,2.45,1.78)$

(a) Ground truth.

(b) Our method.

(c) ICPA.

(d) ICP-CG.

(e) SURF.

Figure 3.31: Registration results obtained with data *candy*.

Table 3.10: Quantitative evaluation of registrations, using data *hand*, *box*, and *candy*.

hand				
	Proposed Method	ICPA	ICP-CG	SURF
Error	0.82	1.14	1.20	16.02
Time	4.5 mn	9.8 mn	9.7 mn	0.0 mn
box				
	Proposed Method	ICPA	ICP-CG	SURF
Error	1.27	9.13	6.56	2.96
Time	12.0 mn	44.5 mn	41.1 mn	0.0 mn
candy				
	Proposed Method	ICPA	ICP-CG	SURF
Error	0.81	5.14	4.03	23.25
Time	4.5 mn	7.7 mn	7.4 mn	0.0 mn

Table 3.11: Average of time consumption of each step of our proposed method, ICPA, and ICP-CG, with the percentage of time used for each step.

	Attribute definition	Matching	Outlier elimination
Proposed method	0.03 mn / 0.2%	7.47 mn / 94.8%	0.40 mn / 5.0%
ICPA	0.00 mn / 0.0%	14.12 mn / 72.0%	5.49 mn / 28.0%
ICP-CG	0.00 mn / 0.0%	12.46 mn / 72.0%	4.85 mn / 28.0%

range images illuminated by a few unknown point light sources.

Chapter 4

Estimating Albedo of Specular Objects

4.1 Introduction

In Chapter 3, we proposed to use local distribution of albedo (i.e. the ratio of the diffuse reflected light over the irradiance) to enhance robustness for range image registration, and demonstrated that albedo is a powerful photometric feature for range image registration of textured objects devoid of salient geometric features. This is because, contrary to color or chromaticity that depend on the pose of the object, albedo depends on only the object material. Though this method achieves robust registration under a rough estimation of illumination, it is limited to Lambertian objects illuminated by a single distant light source.

Many registration methods are based on the ICP (Iterative Closest Point) framework that consists of iteratively refining the transformation until convergence using the current point correspondences between the different images. Estimating accurate point correspondences is thus a key issue for this approach and can be decompose into two parts: (1) extracting reliable and discriminative features and (2) defining a similarity measure to match points.

Albedo at a point can be directly computed when both the diffuse reflection and the incident illumination at this point are known. However, under unknown illumination or if the surface exhibits specular reflections (like shiny surfaces for example), computing albedo becomes a demanding problem. As a consequence, the method we proposed in the chapter 3, as well as previous work that makes use of albedo ([15]), assume the Lambertian surface (diffuse reflection only) and known incident illumination. In these approaches, specular reflections at the surface of an object are not considered.

In this chapter, our objective is to estimate albedo from the captured range images in the challenging situation of specular objects illuminated by a few unknown point light sources. The proposed method will be used as a pre-process of registration. For the registration process to work successfully, sufficient number of feature points with discriminative descriptors need to be detected. We consider our albedo extraction method to be successful when the registration achieves accurate result. We assume that two range images together with color images in different poses are captured from a fixed viewpoint under fixed unknown illumination conditions (that consists of a few distant point light sources). We also assume that neither shadows nor inter-reflections exist and that each light source illuminating the object produces specular reflections.

We propose a method for extracting albedo from two range images of a specular object under a few fixed and unknown point light sources. To compute albedo at the surface, incident illumination and diffuse reflection components are required. For each range image, we generate candidates of light source directions, using normals at the surface and local peak of intensity. Illumination consistency on two range images allows us to identify light source directions among the candidates. The detected light source directions then enable us to define regions where the reflection components are accurately separated. We compute albedo in these regions and extrapolate it by using neighboring similarities. In this way, we obtain albedo over the

range images. The estimated albedo is used as an input of an existing ICP-like registration algorithm to show the usefulness of our proposed method. Fig. 4.1 illustrates the flowchart of our proposed method. Our intensive experiments show the effectiveness of the proposed method. The contributions of this work are (1) an efficient technique to discriminate between specular highlights and high intensity texture regions; (2) identification of regions where the reflection components can be separated even under unknown illumination and (3) an extrapolation technique to maximize the amount of points with estimated albedo. These contributions result in robust and accurate extraction of albedo for range image registration in practical situations, which enlarges the range of possible situations for registration. We note that a part of this work appeared in [92, 93, 94].

4.2 Related work

To deal with specular reflections at the surface of an object illuminated by a few point light sources, recent work on reflectance analysis can be used. Several methods to separate or decompose reflection components of textured surfaces can be found in the literature ([46, 58, 81]). For example, Lin *et al.* [46] proposed to separate reflection components from a sequence of images by computing the median intensity of corresponding pixels in the image sequence. However, this method requires a large number of images as well as pixel correspondences over all images. It is thus inappropriate for range image registration.

Tan *et al.* [89] proposed a method to separate reflection components of textured surfaces from a single image. By assuming the dichromatic reflection and a single distant light source, a specular free image is generated by locally and non-linearly shifting each pixel's intensity and maximum chromaticity. This specular free image has exactly the same geometrical profile as the diffuse components. Though this method achieves accurate separation of reflection components, it cannot handle multiple light sources and high intensity textures.

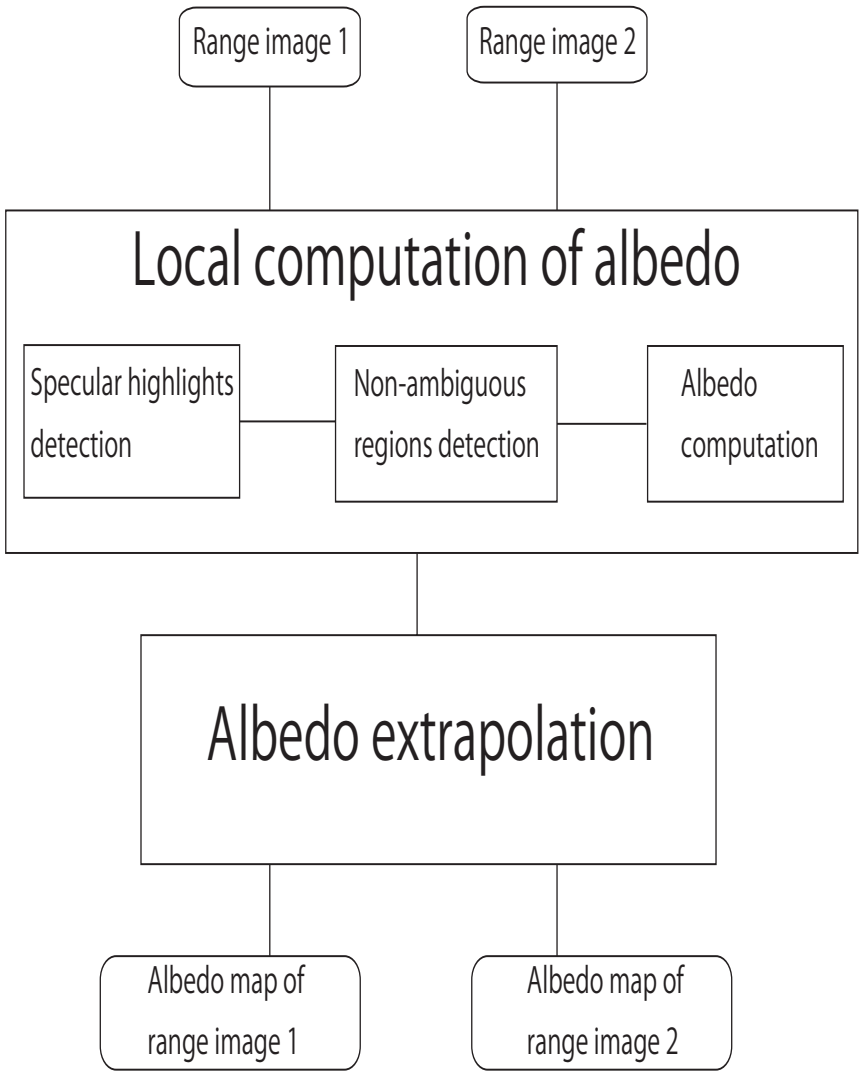


Figure 4.1: Basic flow of the proposed method.

In contrast to previous work, our proposed method can handle changes in photometric appearance, non-Lambertian surfaces and a few unknown point light sources even in the presence of high intensity textures.

4.3 Local computation of albedo

Computing albedo at the surface requires diffuse reflection components and light source directions. When an object is illuminated by a single distant light source and we are given the corresponding illumination chromaticity, a method exists that separates the reflection components of the textured surface [89]. On the other hand, in our case, the illumination environment is not restricted to a single light source and, thus, such a separation technique cannot be applied to the whole surface. However, even in the case of multiple light sources, there exist some regions where the incident illumination can be approximated by a single light source. We thus divide the whole image into regions so that we have a region that is approximated by a single light source illumination. We call such regions non-ambiguous. We can then separate the reflection components of non-ambiguous regions to locally compute albedo. We note that a region is called ambiguous if the region is illuminated by multiple light sources.

4.3.1 Detection of specular highlights

For a smooth surface without high intensity texture, a specular highlight is centered on the mirror-like reflection direction, which is useful to estimate incident illumination direction. If the surface exhibits regions with high intensity texture, however, it becomes difficult to distinguish between specular highlights and regions with high intensity texture. Therefore, we first detect all highlights at the surface that can be either a specular highlight or a high intensity texture region. We then employ illumination consistency between two range images to discriminate specular highlights from high intensity texture regions.

Highlight detection If we consider a region with homogeneous texture, then a specular highlight will exhibit a local peak of intensity. This is because the specular reflection component increases as the viewing direction comes closer to the mirror-like reflection direction. We thus detect local peaks of intensities at the surface.

Points with lowest intensities in the image are first removed to focus on only significant specular highlights (with sufficient intensity). Then we obtain several connected regions. For each connected region, the average avg and standard deviation std of the intensities are computed, and each pixel $\mathbf{x}$ such that $\mathbf{I}(\mathbf{x}) > avg + std$ is selected, where $\mathbf{I}(\mathbf{x})$ is the intensity at $\mathbf{x}$. Then, if the current connected region is separated into several connected parts, the same process is applied to each connected part. The detection stops when the number of connected regions becomes stable. Each connected region represents one possible specular highlight.

Specular highlights Some of the detected highlights may be high intensity texture regions, which may cause inaccurate estimation of incident illumination directions. We first compute light source direction of each highlight by assuming that all the highlights are specular highlights, and then employ illumination consistency to discriminate between specular highlights and high intensity texture regions.

The illumination condition is assumed to be fixed when two range images are captured. This means that the light source directions producing corresponding specular highlights are the same in two range images. We will call this illumination consistency.

Normals at the surface are available for two range images. We can thus estimate the incident illumination direction that can produce such highlight. To be more specific, we first compute the average of the incident light vectors in the highlight region, where an incident light vector at a point $\mathbf{x}$ is computed by rotating the viewing directions at a point $\mathbf{x}$ around the normal at point $\mathbf{x}$ with an angle of π. This is because for smooth surfaces, the view-

ing direction in this region is roughly centered on the mirror-like specular reflection direction.

The highlight regions are then clustered into groups that are produced by similar light sources. Namely, consider the sets $(H_{1,j})_{j\in[0,n_1]}$ and $(H_{2,j})_{j\in[0,n2]}$ of the highlight regions of two range images, with n_1 and n_2 the number of highlight regions. We combine highlight regions using the criterion below:

$$\forall i \in [1,2], \forall (j,j') \in [0,n_i], \text{ if } \operatorname{acos}(\mathbf{l}_{i,j} \cdot \mathbf{l}_{i,j'}) < Th_l,$$
$$\text{then the corresponding regions are combined,}$$

where $\mathbf{l}_{i,j}$ is the estimated normalized light direction for the highlight region $H_{i,j}$, $(\mathbf{l}\cdot\mathbf{l}')$ is the scalar product of two vectors $\mathbf{l}$ and $\mathbf{l}'$, and Th_l is a threshold (for example 10 degrees). When two regions $H_{i,j}$ and $H_{i,j'}$ are combined into a group, $H_{i,j'}$ is added to $H_{i.j}$, $l_{i,j} = \frac{l_{i,j}+l_{i,j'}}{2}$ and $H_{i,j'}$ is removed from the list of highlight regions.

We then eliminate high intensity texture regions using the illumination consistency constraint. For each range image, each light source illuminating the object produces specular reflections. This means that the distribution of normals at the surface should be wide enough, so that for each light source there exists at least one region at the surface for which the viewing direction is close to the mirror-like reflection direction.

Assume a region as a specular highlight in a range image and consider its corresponding light source direction. If no specular highlights can be found in the other range image with its similar corresponding light source direction, then the same light source does not illuminate the object in the other range image, which contradicts to the assumption of fixed illumination. Accordingly, we use the criterion below:

$$\forall i \in [1,2], \forall j \in [0,n_i],$$
$$\text{if for } i' \in [1,2], i' \neq i, \forall j' \in [0,n_{i'}] \ \operatorname{acos}(\mathbf{l}_{i,j} \cdot \mathbf{l}_{i',j'}) > Th_l,$$
$$\text{then the region } H_{i,j} \text{is eliminated.}$$

Fig. 4.2 illustrates the illumination consistency constraint under a fixed viewpoint and fixed illumination condition. We notice that we focus on situations

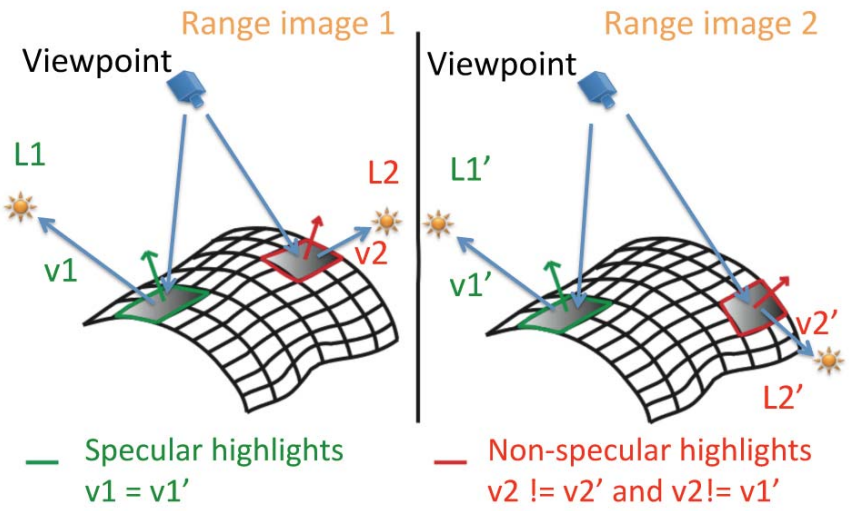


Figure 4.2: Illumination consistency constraint.

when the object presents symmetries in its shape. As a consequence, the geometry of the surface viewed in different pose may be exactly the same, while its texture changes. That is what is illustrated in Fig.4.2. The color part highlighted in red corresponds to the texture of the object and thus moves accordingly to the object pose. By contrast, the color highlighted in green represents a specular highlight and moves according to the relationship between the object shape and the light source. In this example, both the geometry and illumination does not change in the two different poses. That is why the specular highlight (highlighted in green) remains fixed.

We finally obtain consistent specular highlights on two range images with their estimated incident light direction. These specular highlights are then used to compute the illumination chromaticity of each light source. The estimated light sources directions are used to detect non-ambiguous regions each of which is mostly illuminated by a single dominant light source. Details of these procedures are given in the next sections.

4.3.2 Detection of non-ambiguous regions

For each specular highlight, we have estimated its mostly dominant light source direction. If the incident illumination of a region is a single distant light source, we can use the method [89]. We can not, however, directly apply the method [89] to the whole surface, because the illumination environment can be composed of multiple light sources. In fact, the method [89] requires a normalized image that simulates pure white illumination. However, we cannot obtain a normalized image if the scene is illuminated by unknown multiple light sources with different colors. This is because the normalization process is not additive, not even linear.

Since each detected light source is distant from the surface, the incident light rays coming from one light source is the same for all points at the surface. By using the detected incident light directions, we compute a shadow map for each detected light source. Namely, for a light L with its directional vector $\mathbf{l} = (l_x, l_y, l_z)$, we define the shadow map S induced by L proportional to the energy received from L by each point at the surface. More precisely, for a point $\mathbf{x}$ on the surface with normal $\mathbf{n}$ and with angle Θ between $\mathbf{l}$ and $\mathbf{n}$, we define

$$S(\mathbf{x}, L) = \cos\Theta.$$

We consider a set of light sources $\{L_i\}_{i\in[1:n]}$. To detect non-ambiguous regions, we use the criterion below:

$\mathbf{x}$ is in a non-ambiguous region if and only if $\exists i \in [1:n]$, s.t $\forall \mathrm{j} \in [1:\mathrm{n}], \mathrm{j} \neq \mathrm{i}, \mathrm{S}(\mathbf{x}, \mathrm{L_i}) \geq \mathrm{S}(\mathbf{x}, \mathrm{L_j})$ and $\frac{\mathrm{S}(\mathbf{x},\mathrm{L_i})-\mathrm{S}(\mathbf{x},\mathrm{L_j})}{\mathrm{S}(\mathbf{x},\mathrm{L_j})} \geq \beta$, with $\beta = 1$ for example.

For each non-ambiguous region, we attach the light source that emits the most energy inside this region and regroup regions with the same corresponding light sources. We remark that it is preferable to over-detect ambiguous regions rather than non-ambiguous regions. This is because errors in the albedo estimation may propagate during the subsequent extrapolation process.

As a consequence, we obtain non-ambiguous regions in two range images in which we can reliably and adaptively separate reflection components

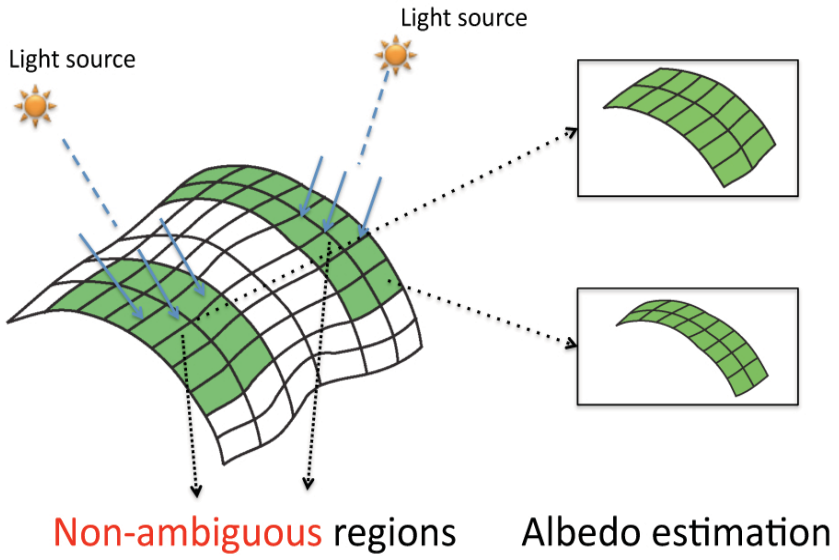


Figure 4.3: Definition of non-ambiguous regions.

using a single distant light source. Fig. 4.3 illustrates the detection of non-ambiguous regions at the surface in the presence of two point light sources.

4.3.3 Estimating albedo of non-ambiguous regions

For each non-ambiguous region, the incident illumination can be approximated by a distant single light source whose illumination chromaticity can be estimated. We can thus independently apply the method proposed in [89] to each non-ambiguous region for separating reflection components of these parts of the surface. We briefly recall the method proposed in [89].

The dichromatic reflection model at a pixel $\mathbf{x}$ can be expressed as:

$$\mathbf{I}(\mathbf{x}) = \omega_d(\mathbf{x})\mathbf{B}(\mathbf{x}) + \omega_s(\mathbf{x})\mathbf{G}(\mathbf{x}), \tag{4.1}$$

where $\mathbf{I} = (I_r, I_g, I_b)$ is the color vector of image intensity, $\mathbf{x} = (x, y)$ is the image coordinates, $\omega_d(\mathbf{x})$ and $\omega_s(\mathbf{x})$ are the weighting factors for diffuse and specular reflections, $\mathbf{B}(\mathbf{x})$ represents the color vector of the diffuse reflection and $\mathbf{G}(\mathbf{x})$ represents the color vector of the specular reflection. Note that we assume that the specular reflection intensity is equal to the illumination

intensity, without any inter-reflections. The first part of the right-hand side in (4.1) represents the diffuse reflection component and the second part represents the specular reflection component. The basic idea for separating reflection components is to iteratively compare the intensity logarithmic differentiation of an input image and its specular-free image. We remark that a specular-free image is an image that has exactly the same profile as the diffuse image.

The input image should be a normalized image that simulates a pure white illumination. Accordingly, the input image is normalized by the illumination chromaticity. To compute illumination chromaticity, several methods based on color constancy can be found in the literature. In particular, the method [90] achieves robustness as well as accurate estimation of the illumination chromaticity by using specular reflection intensity. The specular-free image is generated by shifting each pixel's intensity and maximum chromaticity nonlinearly. Given a normalized and a specular-free image, the reflection components are then iteratively separated until the normalized image has only diffuse pixels.

As a result, a diffuse normalized image is obtained. This estimated diffuse image is then used, together with the estimated light source direction corresponding to the non-ambiguous region and the diffuse reflection model, to estimate albedo in this region.

4.4 Extrapolating albedo into ambiguous regions

Up to here, we have computed albedo in non-ambiguous regions. However, in ambiguous regions, albedo is still unknown and matching points in these regions is not yet possible. We remark that albedo has been computed in several parts in the surface and we expect that several points in the ambiguous regions have albedo similar to points in non-ambiguous regions. We thus estimate albedo in the ambiguous region by extrapolating albedo computed in non-ambiguous regions.

We consider a small region at the surface without specular highlights.

The energy reflected at points inside this region is then mostly diffuse. As a consequence, the chromaticity of points inside this region with the same surface color is similar to each other. Therefore, by comparing chromaticity of points inside the regions, we can identify points having similar albedo.

For a point $\mathbf{x}$ at the surface, the chromaticity $\boldsymbol{\sigma}(\mathbf{x})$ of the point $\mathbf{x}$ is defined as follows:

$$\boldsymbol{\sigma}(\mathbf{x}) = \frac{\mathbf{I}(\mathbf{x})}{I_r(\mathbf{x}) + I_g(\mathbf{x}) + I_b(\mathbf{x})},$$

where, $\boldsymbol{\sigma} = (\sigma_r, \sigma_g, \sigma_b)$.

Starting from the diffuse points in the ambiguous region that have a neighbor in a non-ambiguous region, albedo values are iteratively and locally extrapolated until the size of the ambiguous region converges to a constant value. At each iteration, considering a point $\mathbf{x}$ at the border of the ambiguous region, we extract the point $\mathbf{y}$ in the neighborhood of $\mathbf{x}$ such that $\epsilon = ||\boldsymbol{\sigma}(\mathbf{x}) - \boldsymbol{\sigma}(\mathbf{y})||$ is minimal and albedo of $\mathbf{y}$ is known. If ϵ is smaller than a threshold Th_ϵ (for example $Th_\epsilon = 0.1$), then we set the albedo value of $\mathbf{x}$ to the one of $\mathbf{y}$ and remove $\mathbf{x}$ from the ambiguous region. Namely, we process as follows:

$$\begin{gathered}\mathbf{y} = \text{argmin}_{\mathbf{p}\in V(\mathbf{x})}(||\boldsymbol{\sigma}(\mathbf{x}) - \boldsymbol{\sigma}(\mathbf{p})||),\\ \text{if } ||\boldsymbol{\sigma}(\mathbf{x}) - \boldsymbol{\sigma}(\mathbf{y})|| < Th_\epsilon\\ \text{then } alb(\mathbf{x}) = alb(\mathbf{y}),\\ \text{and we remove } \mathbf{x} \text{ from the ambiguous region,}\end{gathered}$$

where $alb(\mathbf{x})$ is the albedo of point $\mathbf{x}$ and $V(\mathbf{x})$ is a neighborhood of $\mathbf{x}$ such that $\forall \mathbf{p} \in V(\mathbf{x}), \|\mathbf{x} - \mathbf{p}\|_2 < Th_V$ and $\mathbf{p}$ is in a non-ambiguous region, with Th_V a threshold (for example $Th_V = 0.06$ mm if the resolution of range image is 0.01 mm). Fig. 4.4 illustrates different stages of the extrapolation procedure.

As a result, we extrapolate albedo to the rest of points on the surface that are not inside a specular highlight. We then obtain albedo over the surface. The estimated albedo thus becomes useful for registering range

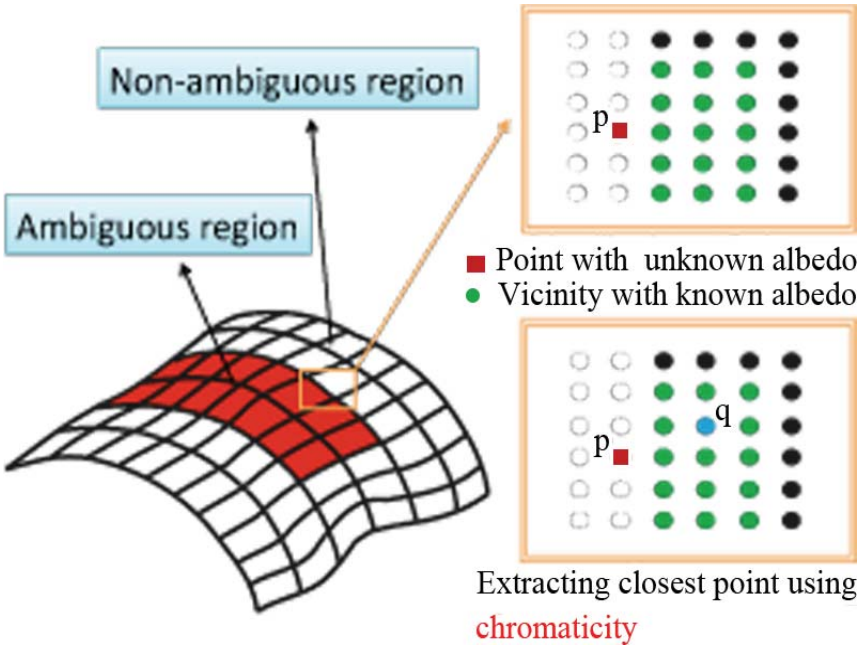


Figure 4.4: Albedo extrapolation.

images. The obtained range image where each point has its corresponding albedo is called the albedo map.

4.5 Computational complexity analysis

Our proposed algorithm has its input of two range images with N points and outputs the estimated albedo values for each point of the two range images. Here we briefly give analysis on the computational complexity to our proposed algorithm. We refer to Fig. 4.1 for the different steps of our method and give the computational complexity for each of these steps.

The first part of our proposed method (Local computation of albedo) takes $O(N^2)$ operations. We first detect specular highlights using the illumination consistency between two range images, which takes $O(N^2)$ computations. In order to detect the non-ambiguous regions, shadow maps induced by each incident light source are computed, which is done in $O(N)$

operations (assuming the number of incident light sources is always below a constant value). Therefore, identifying the non-ambiguous regions takes $O(N)$ computations. Estimating albedo values in each non-ambiguous region takes $O(N)$ computations, and the computational complexity for the local computation of albedo is finally $O(N^2)$.

The second part of our method (Albedo extrapolation) takes $O(N)$ operations. We use a constant size Th_V for the neighborhood $V(\cdot)$ of a point (see section 4.4). Therefore, for each diffuse point of the ambiguous regions of the range images, identifying its corresponding point in its vicinity is done in constant time ($O(1)$). The computational complexity of the albedo extrapolation step is thus $O(N)$.

4.6 Experiments

In order to show the usefulness of our method, we use our estimated albedo image as an input of the range image registration method we proposed in Chapter 3. We recall that the method we proposed in Chapter 3 uses adaptive regions defined from the local distribution of albedo. A similarity metric between two points of interest is then defined based on the albedo similarity of corresponding points inside the regions weighted by the geometric similarity of the regions.

Because objects used in our experiments have a shape devoid of salient geometric properties, the registration using only geometric characteristics does not work well. As a consequence, the standard RMS using point-to-point Euclidean distance errors is not relevant to evaluate the registration results in our case. In this work, we evaluate registration results by comparing the obtained transformation with the ground truth transformation.

As proposed in Chapter 3, we use the combination of the angular measure of errors for the rotation and the standard Euclidean error for the translation, weighted by the resolution of the range images (3.10). In these experiments, all results are shown with estimated albedo images.

4.6.1 Evaluation with synthetic data

We conducted experiments with synthetic data to evaluate the robustness of our proposed method against changes in illumination conditions and noise in both normals and intensities. The synthetic data were obtained with a 3D modeler software (3D Studio Max), each range image had about 30000 points with a resolution of 0.01 mm. The exact albedo image is known. We simulated intensity at the surface with a known specular reflection component and synthetic light sources using the Torrance and Sparrow reflection model [59].

Before applying our method, we manually established a rough pre-alignment of two range images. This alignment allowed us to simulate the case where the input data were captured from two different viewpoints rotationally differentiated by 18.00 degrees around the axis $(0.006, 0.999, -0.026)$ and with a translation of $(-0.02, 0.00, -0.01)$.

In order to see the effects against data noise, we randomly transformed the normals and intensity of the two range images. More precisely, let the latitude and longitude angles between the direction of the perturbed normal and the ground truth normal be (α, ϕ), in which ϕ is a number uniformly generated from 0 degree to 360 degrees. The normals were perturbed with different values of α. On the other hand, the surface intensity was perturbed with Gaussian noise with 0 mean and λ variance, where λ is a percentage of the average over the ground truth intensity of the surface.

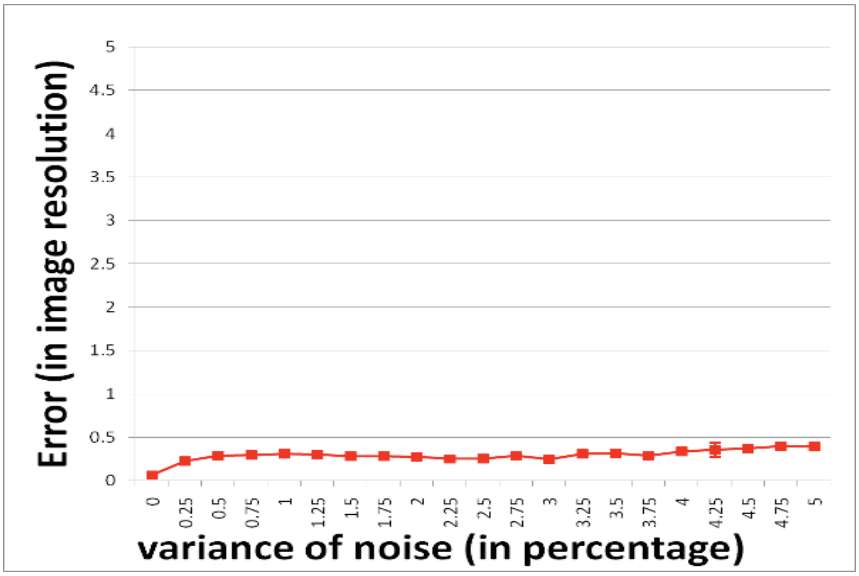
We evaluated our method with different values of α and λ. The value α was changed from 0 to 7 degrees by 0.6 degrees. The value λ was changed from 0 to 5 percents by 0.25 percents. For each values of α, λ, we applied our method 50 times under the same initial conditions.

Figure 4.5 shows quantitative evaluation of registration results under various different level of noise in both normals and in intensity. Our method achieves robustness for both noise in normals and intensity. We observe that even with a noise in intensity of variance 5%, the largest error remains under 0.5 times the resolution of the image. For noise in normals, we observe

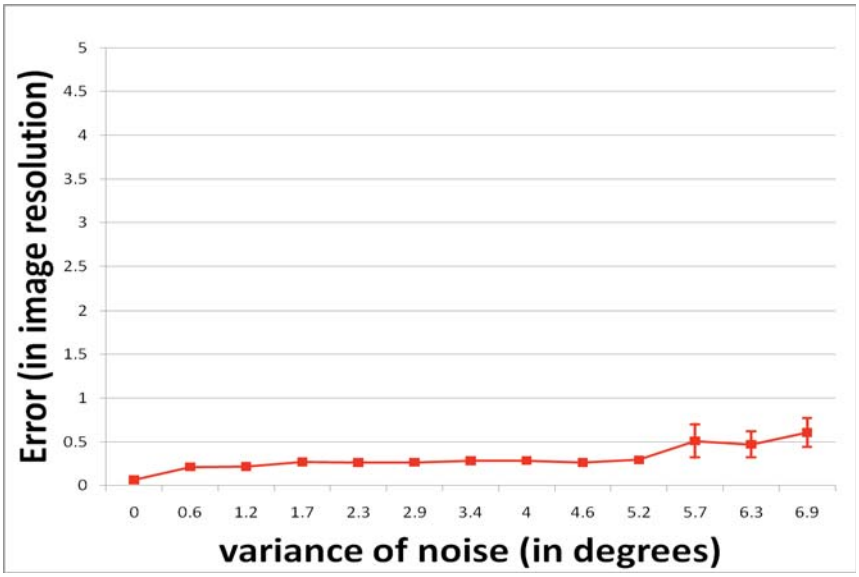
that even with a noise of variance 7 degrees, the largest error remains under 1.0 times the resolution of the image. Fig. 4.6 shows an example of the input range images and the estimated albedo. In this example, we obtained and error of 0.2 times the resolution of the image. We observe that the registration achieves accuracy of the same precision of the acquisition device accuracy. We also observe that as expected, the specular effects are correctly removed and that the features are globally invariant to the viewpoint, the pose of the object and the illumination. Moreover, the obtained albedo is consistent for the two range images.

In order to see the effects against illumination conditions, we rendered two images with various kinds of illumination. The light source direction is computed using the normal at a point $\mathbf{x}$ and the viewpoint, and the light source position is defined at an arbitrary distance on the light direction. This is because we need specular highlights at the surface and we prefer to choose a random point $\mathbf{x}$ at the surface that represents the perfect specular reflection from the viewpoint, rather than choosing the position of the light source randomly.

We changed the position between the specular highlights that define the light source directions. One light was fixed and considered as a reference light. We then evaluated our method with three different values of d, where d is the distance of two different specular highlights: 1.2, 1.0 and 0.8. For each value of d, our method was applied 50 times with a random light direction. Table 4.1 shows the results obtained with our method. The value *Ratio* is the ratio of ambiguous points over the total number of points in the two range images. We observe that the largest error remains under 1.0 time the resolution of the image. Figs. 4.7 illustrates the estimated albedo obtained with our method when using two light sources with $d = 1.0$. For comparison, the result obtained with the method proposed in chapter 3 is also shown. The method proposed in chapter 3 had an error of 24.8 times the resolution of the image. In contrast, the method proposed in this chapter obtained accurate result, with an error of 0.244 times the resolution of the image. We

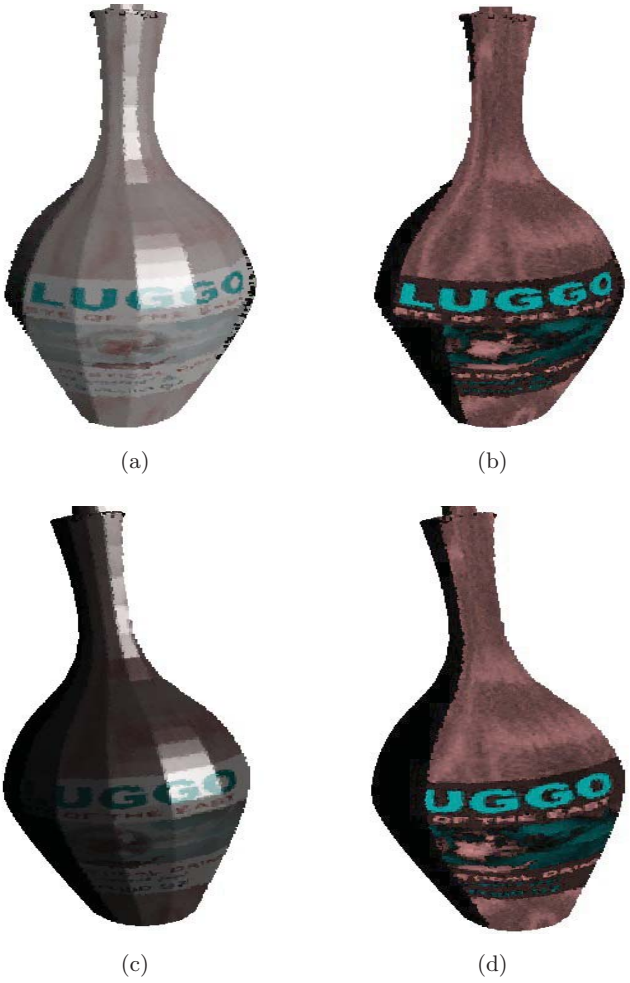


(a) Noise in intensities.



(b) Noise in normals.

Figure 4.5: Results under various noise.



(a) (b)

(c) (d)

Figure 4.6: The input synthetic data and estimated albedo images. (a) input image 1 and (b) its estimated albedo image. (c) input image 2 and (d) its estimated albedo image.

Table 4.1: Results obtained with two light sources.

d	Error	Variance of Error	Ratio	Variance of ratio
1.2	0.886	0.194	0.445	0.041
1.0	0.315	0.022	0.406	0.066
0.8	0.467	0.088	0.354	0.048

observe that the ratio of ambiguous points was of 0.393 in this experiment.

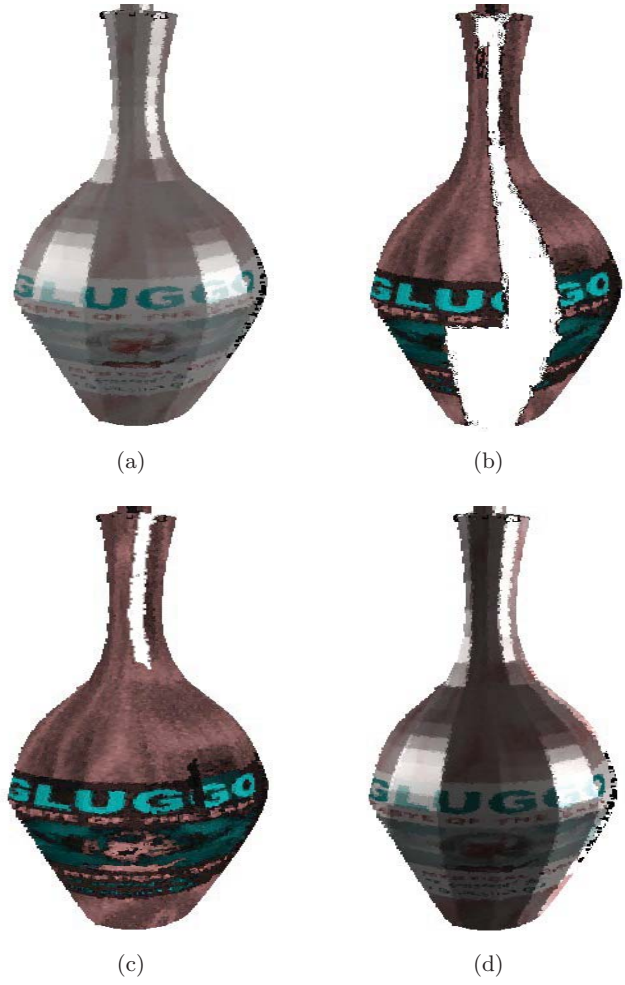
4.6.2 Evaluation with real data

We also conducted experiments using real data. We evaluated our method by comparing with registration results obtained using the albedo image computed with the diffuse reflectance method. We also compared with registration results obtained using chromaticity. We selected these two comparison methods for the reasons below.

- To our best knowledge, existing methods that make use of albedo for registering overlapping range images (like proposed in [15] and in chapter 3 for example) approximate the reflection model with the diffuse reflection model. This approximation is relevant for many types of objects when the viewpoint is far from the mirror-like reflection direction. It is thus often useful to compute albedo for registering overlapping range images.

- Chromaticity is tolerant to some extent against changes in illumination. Using chromaticity for registering overlapping range images is, therefore, more reliable than using brightness of the object.

We employed a Konica Minolta Vivid 910 range scanner, which captures the 3D shape and the texture of an object. A mechanic system was used for object rotations. Because the position and orientation of the range scanner are unknown, it is difficult to obtain the ground truth from the experimental setup. In order to obtain the ground truth, we manually chose about 10



(a) (b)

(c) (d)

Figure 4.7: Simulation with two light sources. (a) input image; (b) our albedo image in non-ambiguous regions; (c) our obtained albedo image; (d) albedo image using the diffuse reflection model.

Table 4.2: The ground truth transformation for data *globe*.

Exp_Rotation	Expected_Translation
$(22.49, 0.02, 0.94, 0.33)$	$(9.32, 0.12, -1.54)$

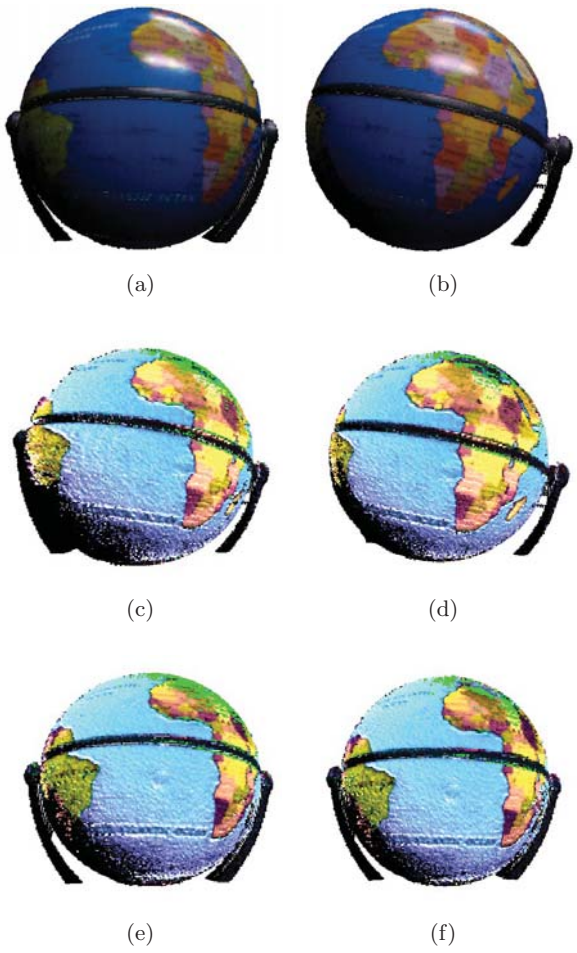
Table 4.3: Results obtained for the data *globe*.

Error	Rotation	Translation
Proposed method		
0.54	$(22.32, 0.03, 0.94, 0.33)$	$(9.11, 0.21, -1.49)$
Diffuse reflection model		
0.90	$(22.38, 0.02, 0.95, 0.32)$	$(9.00, 0.22, -1.39)$
Chromaticity		
1.61	$(22.10, 0.04, 0.92, 0.30)$	$(8.90, -0.09, -1.46)$

corresponding points in two range images and computed the transformation that minimizes the distance between each corresponding points. We used this ground truth to evaluate errors using equation (3.10). We note that for this experiments we took care not to have saturated pixels and to remove the gamma correction of the camera.

We obtained to range images of a sphere with specular reflection components under fixed and uncontrolled illumination (Fig. 4.8). Fig. 4.8 shows estimated albedo images. Each range image had about 31000 points with a resolution of 0.53 mm. The ground truth transformation is shown in Table 4.2. The results obtained with three methods are shown in Fig. 4.9. The quantitative results of the registration by the three methods are shown in Table 4.3.

As we can see in Table 4.3 and in Fig. 4.9 (c), using chromaticity to establish matches between two range images of a specular object with different poses does not work. Indeed, the specular highlights are not removed, which tends to degrade the accuracy of matching. Similarly, the diffuse approximation performed worse than our proposed method. The specular reflections



(a) (b)

(c) (d)

(e) (f)

Figure 4.8: The data *globe*. Input image 1 (a) and 2 (b). Albedo of image 1 in non-ambiguous regions (c) and after extrapolation (d). Albedo of image 2 in non-ambiguous regions (e) and after extrapolation (f). Ambiguous regions are displayed in vivid green.

(a) (b) (c)

Figure 4.9: Results obtained with different methods. (a) our method; (b) using diffuse reflection model; (c) using chromaticity.

at the surface are ignored in the diffuse reflection model. In fact, we can observe in Fig. 4.9(b) that the specular highlights are present. These remaining specular highlights degraded the accuracy of matching. On the contrary, our proposed method successfully removed the specular highlights and, at the same time, estimated the albedo for almost all surface points. The obtained albedo image (see Fig. 4.9(a)) is thus reliable and accurate enough for matching points. The quantitative results also confirm the effectiveness of our proposed method.

4.7 Conclusion

We presented a method for estimating albedo of specular surfaces for the purpose of range image registration of specular objects devoid of salient geometric properties under unknown multiple point light sources. By using highlights at the surface and illumination consistency on two range images, we estimated the incident illumination. We then used the illumination information and the dichromatic model of reflection in order to locally estimate albedo. Locally estimated albedo was then extrapolated into the whole surface to obtain reliable albedo map. A range image registration technique was then used to estimate the transformation aligning two range images. Experiments using synthetic data and real data confirmed the efficiency of our proposed method.

In this chapter as well as in Chapter 3, we assumed the surrounding illumination to be composed of only a few point light sources. Such kind of illumination is rather limited since the surrounding illumination generally presents extended light sources. Under general illumination, however, the mechanisms behind image formation are complex (even for the Lambertian reflection). Estimating photometric features from unregistered range images under unknown general illumination is not possible and so is matching points across range images using photometric features under unknown general illumination. However, matching features is not the only way we can take for registering overlapping range images. In particular, a similarity metric between two overlapping range images that makes use of photometry is yet to be defined. In the next chapter, we will investigate the derivation of a global photometric metric, independent to illumination, which evaluates the quality of a given rigid transformation aligning a pair of overlapping range images.

Chapter 5

Photometric metric under unknown lighting for range image registration

5.1 Introduction

The mechanisms behind image formation are complex and estimating albedo from a single range image under unknown illumination is impossible. As a consequence, using photometric features to match points across range images as proposed in the previous chapters has limited range of applications. However, matching points across the range images is not the only way we can take for registration.

In this chapter, we exploit the use of photometry to successfully register two overlapping range images with color of an object's surface devoid of salient geometric features under general and distant, unknown lighting. Differently from the methods proposed in the previous chapters, we do not explicitly search for the point correspondences across the range images but rather derive a photometric evaluation function and search for the transformation that minimizes the evaluation function. This expands applications such as on-site modeling of archaeological objects (whose surfaces often ex-

hibit symmetries). We consider the situation where the object pose changes during acquisition, while the viewpoint and illumination stay fixed. We assume a Lambertian reflection with no inter-reflections, nor any cast shadows.

When an object's surface lacks discriminative geometric features, photometric features are used to guide the registration ([5, 66]). In particular, illumination-invariant photometric features are preferable. Without known correspondences between two range images or without known illumination, however, computing such features is not possible. Therefore, the ICP-like approach where photometric features are used to find correspondences does not work under unknown lighting.

Another approach to registration can be found where alignment is achieved by minimizing a cost function ([10, 68]) without using explicit point correspondences. In general, the cost function is based on re-projection error. Defining such functions for 2D images has been widely studied and famous methods such as the cross-correlation method are available. However, defining a photometric metric for 3D images remains an open challenge.

We propose a novel photometric metric for registering two range images of an object lacking in geometric features under unknown lighting and show its usefulness with a practical registration method. The spherical harmonics give a compact yet accurate representation of image formation. For a given transformation and its induced point correspondences, this representation allows the estimation of illumination and albedo, both of which are used with one range image to synthesize the color image of the other range image. The difference between the synthesized color images and the captured color images (we will refer to it as the photometric re-projection error) defines our photometric metric without using any a-priori information on the incident lighting. This function evaluates the consistency of the relationship between geometry, texture and illumination.

We use a hypothesis-and-test registration method to demonstrate the usefulness of our proposed photometric metric. Our method carries out registration not by estimating transformations from point correspondences but

by generating transformations and evaluating them to find the best one. A transformation is defined by a triplet of correspondences, and with about 10^4 points in each range image, we have potentially around 10^{12} possibilities. Directly evaluating each possibility is thus computationally unrealistic, even using RANSAC-like methods. To efficiently generate transformations, we use the spherical representation of each range image. The sphere is homeomorphic to a close-zero genus surface, and a natural and efficient representation for a 3D surface. Over the sphere, we generate rigid transformations and evaluate them to reach the best one for the final result.

Though for a closed surface[1] its spherical representation is pose-invariant, it is not the case for range images (which represent non-closed surfaces). Therefore, the rigidity of a transformation defined in the spherical domain is not inherited in the original domain. We thus iteratively refine the spherical representation throughout our registration process not to miss any rigid transformation.

The contribution of our work is two-fold: (1) we derive a photometric function to evaluate given transformations under general and distant, unknown lighting; and (2) we demonstrate the effectiveness of our photometric metric by implementing a practical registration method using the hypothesis-and-test search strategy. To the best of our knowledge, this is the first work that registers range images devoid of salient geometric features under general and distant, unknown lighting using photometry. We note that a part of this work appeared in [96].

5.2 Related work

In general, pair-wise registration methods can be divided into two categories: (1) the one that search for the best correspondences between points of the two range images to find the best transformation; (2) the one that search for the best transformation as the minimum of a cost function. While the

[1] A closed surface is defined as a surface that is compact and without boundary; a non-closed surface is a compact surface with boundary.

methods that fall into the first category are in general efficient, methods falling into the second category are in general more robust and accurate.

5.2.1 Registration using point correspondences

In general, methods falling into the first category consist of feature detection and description, followed by feature matching (the transformation aligning two range images is then estimated using the obtained correspondences). The iterative closest point (ICP) [7] and its variants are popular methods where points are matched to their closest point. The list of correspondences is iteratively updated until the estimated transformation converges to a stable solution. The major drawback of ICP is that it may get trapped in local minima and thus depends highly on the initial alignment.

Many interest point detectors and feature descriptors have been proposed in the last decade to establish point correspondences using color-like features. In early work, Johnson and Kang [42] or Okatani *et al.* [66] proposed to use color or chromaticity to match points. However, color and chromaticity of a point are not distinctive by themselves and good quality of correspondences cannot be achieved.

The SIFT and SURF descriptors ([5, 50]) are the most popular textural features, which are extensively used for aligning 2D images. They are scale-invariant feature descriptors that make use of difference of Gaussian (for SIFT) and integral images (for SURF) to define features for each keypoint. However, they are 2D features and thus suffer from projective distortion when used to align range images. Seo *et al.* [86] extended the SIFT feature for range image alignment by accounting for the projective distortion. However, SIFT-like methods do not account for changes in pose-illumination relationship, but rely on the color normalization assumption (i.e. intensity changes uniformly with changes in illumination, and normalization of features with respect to the overall intensity is sufficient). The above mentioned methods thus all suffer from changes in pose-illumination relationship.

In the previous chapters, we explored the use of albedo for range im-

age registration. We proposed a robust feature descriptor based on adaptive regions using albedo and a technique to estimate albedo for specular objects illuminated by a few unknown point light sources. However, this approach assumes known illumination or a few unknown point light sources. In practical situations, however, illumination is more complex and difficult to estimate.

In a broader sense, photometry states for the relationship between geometry, reflectance properties and incident illumination. As a consequence, from a single range image it is not possible to estimate one of the three without knowing the others. Therefore, the feature-based approach, where features are computed for two range images independently, fails in some situations when illumination is not known.

5.2.2 Registration by minimizing a cost function

Matching features is not the only way we can take to align range images. Other transformation search methods can be found in the literature where a cost function is minimized over a parameter space. Some use optimization strategy such as Gauss Newton ([10, 31, 39, 68]), and some use the hypothesis-and-test strategy such as RANSAC or brute-force search ([21]). In [31], for example, the similarity measure is defined as the cross correlation of the spherical representations of surfaces, and it is customized according to the surface-intrinsic attributes while the spherical harmonics speed up the optimization process. The former is efficient but sensitive to the initial alignment while the latter does not depend on initialization even though it may be less efficient.

Several geometric cost functions, as well as 2D textural cost functions have already been explored. However, less work has been done on defining a 3D photometric metric for aligning pairs of range images, and as far as we know no photometric metric under unknown lighting has been reported.

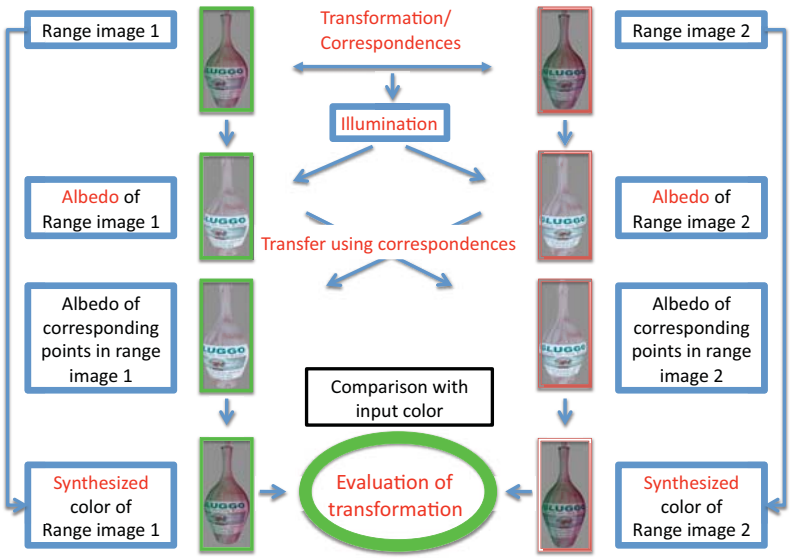


Figure 5.1: Procedural evaluation of a given transformation.

5.3 The photometric metric

When geometry alone is not sufficient to discriminate between good or bad relative orientations of range images, photometry should be used. We introduce our photometric metric under unknown lighting that does not compare features but compute re-projection error. By doing so, we simultaneously take into account geometry, reflectance properties and illumination to derive a metric that makes full use of photometry. Figure 5.1 illustrates the derivation of our proposed metric. A brief review on image formation modeling using spherical harmonics is given in Section 5.3.1. Then, we introduce our photometric metric in Section 5.3.2 and discuss the point correspondence identification in Section 5.3.3.

5.3.1 Review of spherical harmonics representation

The spherical harmonics have been shown to be a powerful tool to model image formation [74], and, for the Lambertian reflectance, up to the second-order spherical harmonic expansion is known to be sufficient to approximate the image formation with more than 98% accuracy [4]. We briefly recall the principles of the spherical harmonics representation and refer to [4] for more details.

We consider a convex, Lambertian object illuminated by distant isotropic lights. The intensity of the reflected light is known to be a function of the normal and albedo. Namely, according to the Lambert's law and for an incident light ray $\vec{l}$ of intensity l, the intensity of the reflected ray at a point $\mathbf{x}$, is $l\max(\cos(\theta), 0)$, where θ is the angle between the incident light ray and the normal at the surface at point $\mathbf{x}$. Then, the irradiance $E(\mathbf{x})$ at a point $\mathbf{x}$ for a distant global illumination L and the diffuse reflection kernel[2] R is given by an integral over the sphere.

$$E(\mathbf{x}) = \int_0^{2\pi} \int_0^{\pi} L(\theta, \phi) R(G_{\mathbf{x}}(\theta, \phi)) \sin(\theta) \mathrm{d}\theta \mathrm{d}\phi, \tag{5.1}$$

where (θ, ϕ) are the incident angles in the global coordinate system and $G_{\mathbf{x}}$ is the transformation from the global coordinate system to the local coordinate system centered around the normal of the point $\mathbf{x}$.

The irradiance at a point $\mathbf{x}$ is then scaled by the albedo $\rho(\mathbf{x})$ to have the color of the point: $I(\mathbf{x}) = \rho(\mathbf{x})E(\mathbf{x})$. We remark that the irradiance can be viewed as the convolution over the sphere of the incident illumination L and the reflection kernel R. Equivalently to the Fourier series for the circle, the *spherical harmonics* are a convenient tool to examine the convolution over the sphere. They allow a compact representation of the image formation. Namely, the color $I(\mathbf{x})$ of a point $\mathbf{x}$ on the Lambertian surface is approximated as

$$I(\mathbf{x}) \approx \rho(\mathbf{x}) \sum_{l=0}^{2} \sum_{m=-l}^{l} L_{l,m} R_{l,m}(\mathbf{x}), \tag{5.2}$$

[2] $R(\theta) = \max(\cos(\theta), 0)$.

where $L_{l,m}$ and $R_{l,m}(\mathbf{x})$ represent the spherical harmonic coefficients of L and $R \circ G_{\mathbf{x}}$, respectively ($\circ$ stands for the function composition operator). We notice that the spherical harmonic coefficients $R_{l,m}(\mathbf{x})$ of the Lambertian reflection kernel $R \circ G_{\mathbf{x}}$ are known as functions of the normal at point $\mathbf{x}$ [4].

5.3.2 Evaluation metric for a transformation

We use the spherical harmonics representation of image formation to derive our photometric evaluation metric for a given rigid transformation. We remark that our photometric metric does not suffer from scale ambiguity that arises when estimating photometric features, neither requires any a-priori information on the incident illumination.

A given transformation between two range images (range images 1 and 2) induces point correspondences across the two images. By contrast with [88] where the popular Lambertian reflection model is used, we use the spherical harmonics representation of image formation to derive a linear system from the point correspondences with the illumination as unknown. It is thus straightforward to obtain an estimate of the illumination with respect to the given transformation. The estimated illumination allows us to compute albedo at points of the range images, which is transferred to their corresponding points. The transferred albedo are used together with the estimated illumination and geometry to synthesize colors of the two range images. The synthesized colors are then compared with the captured colors of the two range images to evaluate the photometric consistency (i.e. photometric re-projection error) of the alignment induced by the given rigid transformation.

Let T, $(\mathbf{x}_i, \Gamma(T(\mathbf{x}_i)))_{i\in[0,n-1]}$ and $(\Gamma(T^{-1}(\mathbf{y}_i)), \mathbf{y}_i)_{i\in[0,m-1]}$ respectively denote a given transformation and the induced point correspondences, where $\mathbf{x}_i \in \mathbf{R}^3$ belongs to range image 1 (denoted as I^1), $\mathbf{y}_i \in \mathbf{R}^3$ belongs to range image 2 (denoted as I^2) and Γ denotes the point correspondences identification operator (see Section 5.3.3). If T accurately aligns the two range images, then two corresponding points represent the same point of

the surface viewed in different poses, and their albedo is the same $\rho(\mathbf{x}_i) = \rho(\Gamma(T(\mathbf{x}_i)))$ (similarly $\rho(\mathbf{y}_i) = \rho(\Gamma(T^{-1}(\mathbf{y}_i)))$).

R is known and depends on only the surface normals. Therefore, using (5.2), we can derive a linear system (5.3) with L as unknowns, where L is a row vector in 9D and M is a $9 \times (n+m)$ matrix, where n and m are the number of corresponding points from range images 1 and 2 respectively.

$$LM = 0, \tag{5.3}$$

where $L = [L_{0,0}, L_{1.-1}, L_{1,0}, L_{1,1}, L_{2.-2}, L_{2,-1}, L_{2,0}, L_{2,1}, L_{2,2}]$ and $M = [M_i]_{i\in[0,n+m-1]}$ with

$$\begin{cases} M_i = [I^1(\mathbf{x}_i)R_{0,0}(\Gamma(T(\mathbf{x}_i))) - I^2(\Gamma(T(\mathbf{x}_i)))R_{0,0}(\mathbf{x}_i), \cdots, \\ \qquad I^1(\mathbf{x}_i)R_{2,2}(\Gamma(T(\mathbf{x}_i))) - I^2(\Gamma(T(\mathbf{x}_i)))R_{2,2}(\mathbf{x}_i)]^\top \text{ if } i < n \\ M_i = [I^2(\mathbf{y}_i)R_{0,0}(\Gamma(T^{-1}(\mathbf{y}_i))) - I^1(\Gamma(T^{-1}(\mathbf{y}_i)))R_{0,0}(\mathbf{y}_i), \cdots, \\ \qquad I^2(\mathbf{y}_i)R_{2,2}(\Gamma(T^{-1}(\mathbf{y}_i))) - I^1(\Gamma(T^{-1}(\mathbf{y}_i)))R_{2,2}(\mathbf{y}_i)]^\top \text{ if } n \le i < n+m. \end{cases}$$

The matrix M is known, and we can estimate the illumination $\tilde{L}(T)$ with respect to the given transformation T using the SVD, up to an unknown scaling factor λ ($\lambda \neq 0$). We can then estimate albedo of each point.

$$\rho(\mathbf{x}) = \frac{1}{\lambda}\left(\frac{I(\mathbf{x})}{\sum_{l=0}^{2}\sum_{m=-l}^{l}\tilde{L}_{l,m}(T)R_{l,m}(\mathbf{x})}\right). \tag{5.4}$$

A given transformation is not likely correct in most cases. Therefore, contrary to the case where correspondences are assumed to be perfect ([88]), we need to carefully choose an attribute for our evaluation. For example, comparing the estimated albedo of corresponding points is not effective. This is because the photometric solution for a given transformation has scale ambiguity, and regardless of the relationship between geometry, illumination, and albedo, a solution with a small scaled albedo always gives better results. Namely, the re-projection error $\|\tilde{L}(T)M\|$ or the residual error in albedo $\|\rho(\mathbf{x}_i) - \rho(\Gamma(T(\mathbf{x}_i)))\|$ is different for $\tilde{L}$ and $\lambda\tilde{L}$, with $\lambda \neq 1$ while $\tilde{L}$ and $\lambda\tilde{L}$ correspond to the equivalent photometric solution. We thus use the

captured color images as the ground truth to evaluate the transformation T. This is justified by the fact that the estimated photometric properties should be coherent with the correspondences and the captured images.

Corresponding points $(\mathbf{x}_i, \Gamma(T(\mathbf{x}_i)))$ should have the same albedo if T is accurate. We thus synthesize the color of $\mathbf{x}_i$ in range image 1 as follows:

$$\tilde{I}_T^1(\mathbf{x}_i) = \rho(\Gamma(T(\mathbf{x}_i))) \sum_{l=0}^{2} \sum_{m=-l}^{l} \tilde{L}_{l,m}(T) R_{l,m}(\mathbf{x}_i). \tag{5.5}$$

Similarly, we synthesize the colors $\tilde{I}_T^2(\mathbf{y}_i)$ of points $\mathbf{y}_i$ in range image 2.

We now define our photometric re-projection error of T.

$$Eval(T) = \frac{\sum_{i=0}^{n-1} \|I^1(\mathbf{x}_i) - \tilde{I}_T^1(\mathbf{x}_i)\| + \sum_{i=0}^{m-1} \|I^2(\mathbf{x}_i) - \tilde{I}_T^2(\mathbf{x}_i)\|}{n+m}. \tag{5.6}$$

We remark that the unknown scaling factor λ that arises when estimating albedo in (5.4) is no longer present in (5.6).

We notice that the shape of $Eval$ depends on T. Namely, $\tilde{I}_T^1$ and $\tilde{I}_T^2$ change even for the same point depending on T. Thus, the derivation of $Eval$ is procedural and we do not have an analytical formula for the function. In addition, $\tilde{I}_T^1$ and $\tilde{I}_T^2$ are only piece-wise continuous with sufficiently similar transformations. This is because the distribution of albedo over the surface is only piece-wise continuous. As a consequence, the values of the entries of the matrix M in (5.3) vary piece-wise continuously with sufficiently similar transformations and so does the estimated photometric properties as well as the synthesized images $\tilde{I}_T^1$ and $\tilde{I}_T^2$.

5.3.3 Point correspondences identification

Every time our photometric evaluation function is computed, we have to identify the point correspondences from the given transformation T. Equivalently, we have to define the operator Γ. Though it is a simple task, it is the most time consuming for evaluating a given transformation. It is thus of major importance to perform it as fast as possible. We propose to use the projection operator of 3D points to a virtual image plane to realize fast point correspondences estimation.

We consider a virtual camera and denote by K and S, the intrinsic and extrinsic parameters (respectively) of the virtual camera. We first create a look-up table for each range image with respect to the virtual camera. Namely, for each range image, each entry of the look-up table takes the list of points that project to the corresponding pixel in the image plane of the virtual camera. To be more precise, let us consider a point $\mathbf{y}$ in range image 2 (represented using the quaternions) and the corresponding look-up table L_T_2. Then the coordinates (i, j) of the corresponding pixel in the virtual image plane are obtained as $(i, j, 1) = KS\mathbf{y}$, and the point $\mathbf{y}$ is appended to the list $L\_T_2(i, j)$. Thereafter for a given transformation, we search for the closest points from range image 1 to range image 2 as follows: (1) a point $\mathbf{x}$ of range image 1 is transformed by the given transformation ($\mathbf{y} = T\mathbf{x}$); (2) the point $\mathbf{y}$ is projected into the virtual camera plane ($(i, j, 1) = KS\mathbf{y}$); (3) $\Gamma(\mathbf{x})$, the closest point of $\mathbf{x}$ in range image 2, is searched inside the list of the corresponding entry of the look-up table of range image 2 ($L\_T_2(i, j)$). Searching for the closest points from range image 2 to range image 1 is carried out similarly.

By using a virtual camera such that one entry in the look-up tables does not contain more that an arbitrary fixed number of points, we can ensure that the point correspondences identification between the two range images is done in $O(n+m)$ operations, with n and m the number of points in range images 1 and 2, respectively. This can be done by increasing the resolution of the virtual camera. A too high resolution of the virtual camera, however, may result in holes in the look-up tables, which may prevent us from finding the corresponding points. We thus set the intrinsic K and extrinsic S depending on the intrinsic parameters of the scanning device. Because of possible errors in the estimated scanning device parameters, we choose a resolution for the virtual camera equal to half the one of the scanning device, for better stability in the point correspondences identification. Namely,

$$K = \begin{pmatrix} f \cdot \frac{m_x}{2} & \gamma & \frac{u_0}{2} \\ 0 & f \cdot \frac{m_y}{2} & \frac{v_0}{2} \\ 0 & 0 & 1 \end{pmatrix} \tag{5.7}$$

$$S = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{5.8}$$

where f is the focal length, m_x and m_y are the scale factors relating pixels to distance, γ represents the skew coefficient between the x and y axis (which is often 0), u_0 and v_0 represent the principal point (which would be ideally in the centre of the image) of the scanning device.

Stable points identification Since range images are acquired by changing the poses of an object, even points in the overlapping area may not correspond exactly. This is due to different digitization of the overlapping area depending on the pose. As a consequence, even for the best transformation aligning the range images, there may be some point correspondences $(\mathbf{x}, \Gamma(T(\mathbf{x})))$ that do not satisfy the statement $\rho(\mathbf{x}) = \rho(\Gamma(T(\mathbf{x})))$. This is because the distribution of albedo at an object's surface is not continuous. In such a case, the quality of the estimated photometric properties (illumination and albedo) may be significantly degraded, which would reduce the reliability of our photometric metric for evaluating the quality of a transformation. To overcome this problem, we first identify a stable point, i.e., a point whose photometric features are sufficiently similar to those of its corresponding point even though the correspondence may not be exact. We then use only stable points to evaluate our photometric function.

We extract stable points independently for two range images as a preprocessing step. Namely, our objective here is to identify points, each of which has the property of having albedo values and surface normals inside its small vicinity that are sufficiently similar to each other. Using only

stable points, we ensure that even if a corresponding pair does not exactly represent the same point viewed in different poses, albedo and the surface normals of the corresponding points remain sufficiently similar with each other to achieve accurate estimation of illumination.

It is well known that in a small vicinity and for diffuse reflection, the difference in chromaticity approximates well the difference in albedo. Accordingly, we define a stable point using both difference of chromaticity and difference of normals in a small vicinity. Namely, a point $\mathbf{x}$ is identified to be stable if

$$\begin{aligned} \forall \mathbf{y} \text{ such that } \|\mathbf{y}-\mathbf{x}\| < \epsilon_s, \\ \|c(\mathbf{x})-c(\mathbf{y})\| < \epsilon_c \text{ and } \|\mathrm{n}(\mathbf{x})-\mathrm{n}(\mathbf{y})\| < \epsilon_n, \end{aligned} \tag{5.9}$$

where $\mathbf{y}$ is a point in the range image concerned, c is chromaticity, n represents the surface normals, and ϵ_s, ϵ_c, and ϵ_n are three thresholds.

5.4 Analysis of the photometric metric

We analyze the behavior of our photometric metric under various parameters. We consider the two range images shown in figure 5.2, of resolution about 0.01 mm. Starting from the ground truth transformation that perfectly aligns the two range images we generate several transformations by randomly perturbing the parameters of the ground truth transformation and plot the photometric re-projection error in function of the registration error (5.10) for each generated transformation. We notice that when the function is not defined (insufficient number of corresponding points), the photometric metric returns $+\infty$. We then clamp our metric to [0: 255] for better visualization (i.e. all metric values greater than 255 are set to 255). The random perturbation was obtained by perturbing the rotation angles inside the range $[-0.3; 0.3]$ radians and the translation values inside the range $[-0.7; 0.7]$ mm. First a uniform noise was applied with range $[-0.3; 0.3]$ radians for perturbations in the angles and $[-0.7; 0.7]$ mm for the perturbations in the translation. In order to increase the density of transformations

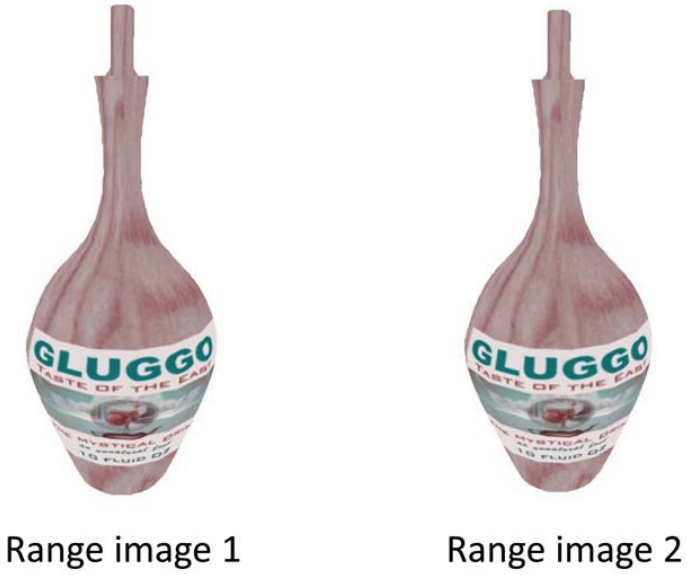


Figure 5.2: The two input range images.

generated close to the ground truth, we then successively applied a uniform noise with ranges $[-0.15; 0.15]$, $[-0.07; 0.07]$ and $[-0.03; 0.03]$ radians for perturbations in the angles and $[-0.3; 0.3]$, $[-0.1; 0.1]$ and $[-0.01; 0.01]$ mm for the perturbations in the translation.

We performed this procedure for different illumination conditions and different values of ϵ_s. We chose to test our photometric metric against different values of ϵ_s because it is the parameter that reflects the size of the neighborhood used in identifying the stable points.

5.4.1 Our metric against different illumination conditions

Figures 5.5 and 5.6 illustrate our photometric metric with respect to the different illumination situations illustrated in figures 5.3 and 5.4, for the two range images shown in figure 5.2. We remark that many points are accumulated on the line $error = 255$, and that the more the registration error increases, the more points are accumulated on this line. This is due to clamping our metric to $[0 : 255]$. In all situations, we used the same range

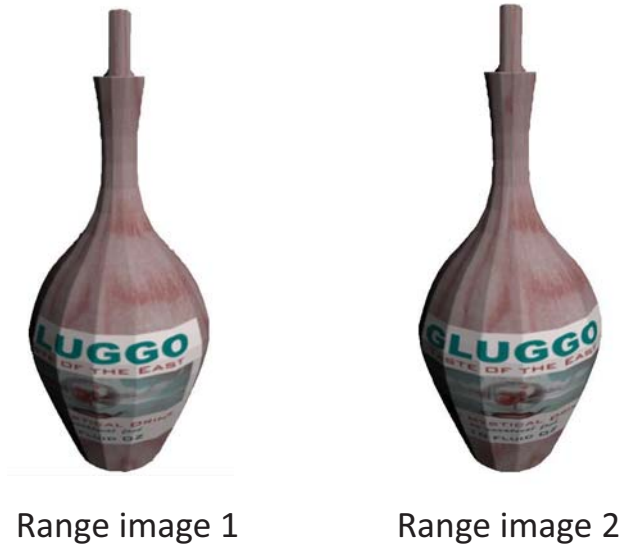
images, with the same initial positions. The situation 1 (figure 5.3 (a)) is obtained by illuminating the two range images with a light probe using the spherical harmonics representation of image formation. The situations 2, 3 and 4 (figures 5.3 (b) and 5.4) are obtained by illuminating the range images with an ambient light and a point light source in different positions and with different colors. In situations 2, 3 and 4 the classic Lambertian reflection model was used to simulate image formation. For all situations the stable points were identified using the parameters $\epsilon_s = 0.03$, $\epsilon_c = 0.01$, and $\epsilon_n = 0.2$.

First, we observe that while the variation in the photometric re-projection error is large for transformations with large registration error (most of the metric values are then above 255), it becomes small for transformations with small registration errors (most of the metric values are then below 255). This is naturally explained by the fact that the transformations with larger registration errors are much more different from each other than those with smaller registration errors (and so for the point correspondences).

Second, we observe that the behavior of our photometric metric is similar in all situations, even though the illumination conditions are completely different (extended light sources in situation 1, point light source with different positions and different colors in the other three situations). This is because no a-priori information on the illumination is used in deriving the photometric cost function.

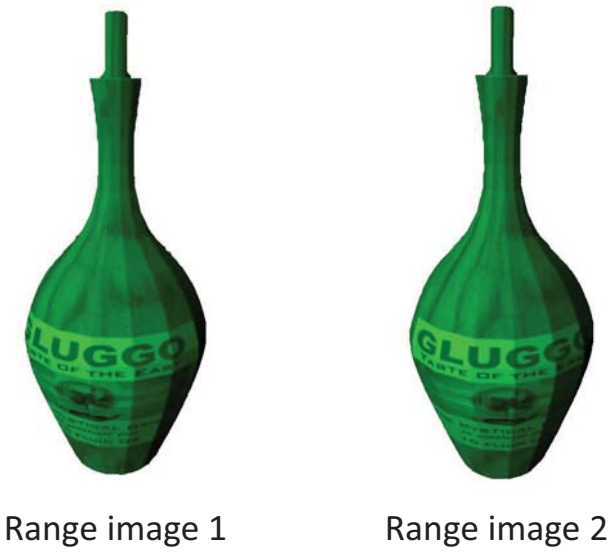
Third, we observe that for the same registration error, our photometric metric can have different values. Two factors explain this phenomena. (1) While the registration error is the same, the transformations are different. For example, from the ground truth position, sliding one range image to the left or to the right with the same translation magnitude gives the same registration error. However, the point correspondences are different, which gives different photometric re-projection errors (Fig. 5.7). (2) The texture at the surface of an object is not continuous. As a consequence, while in some directions, the normals and albedo values vary smoothly, in some other

(a) Situation 1

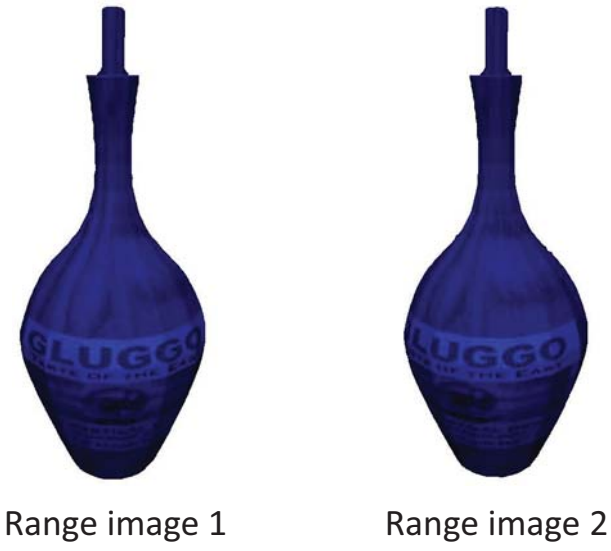


(b) Situation 2

Figure 5.3: The situations 1 and 2.

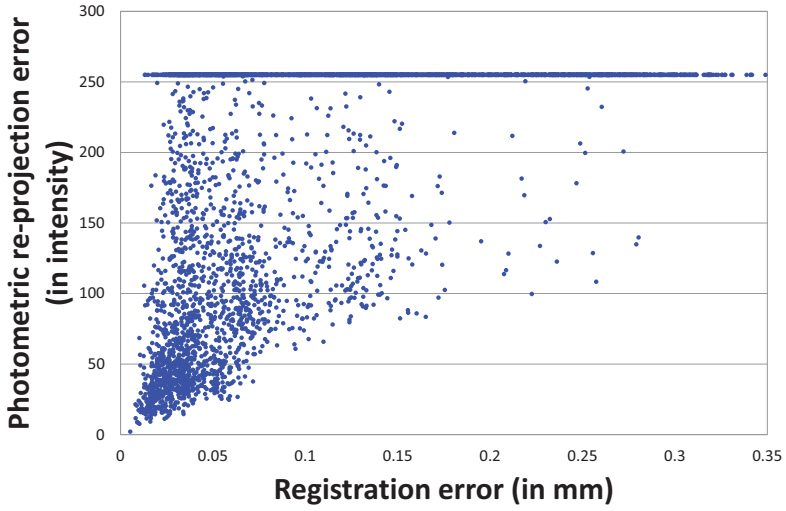


(a) Situation 3

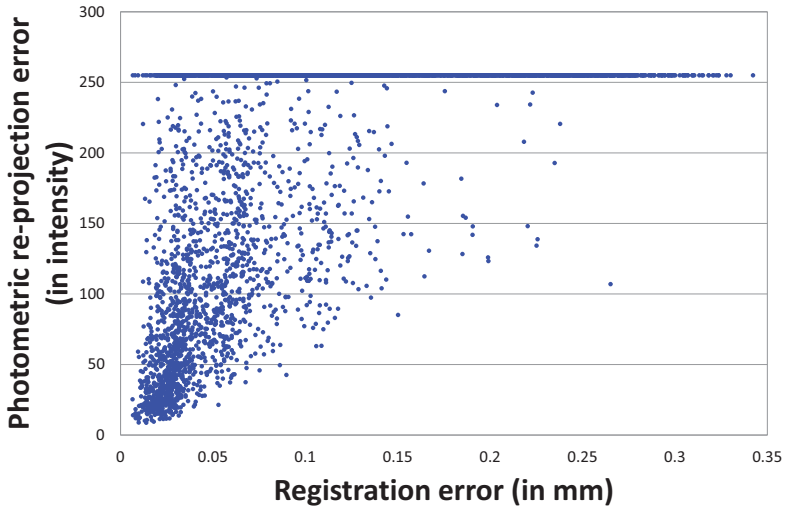


(b) Situation 4

Figure 5.4: The situations 3 and 4.

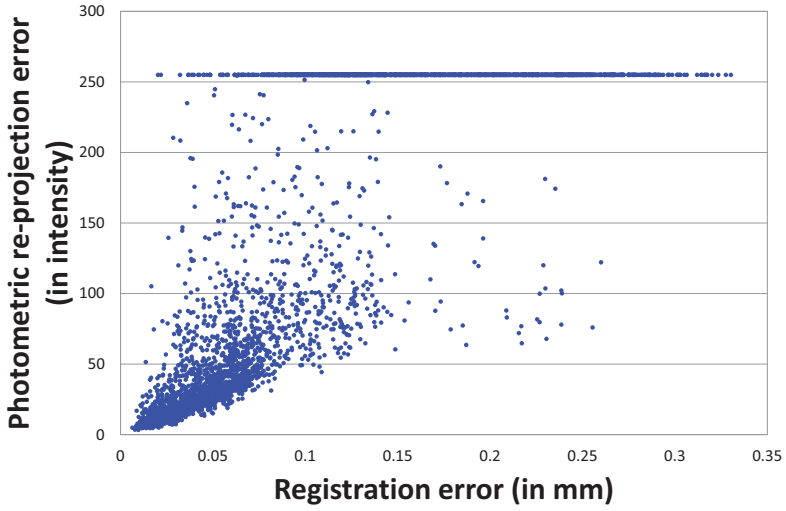


(a) Situation 1

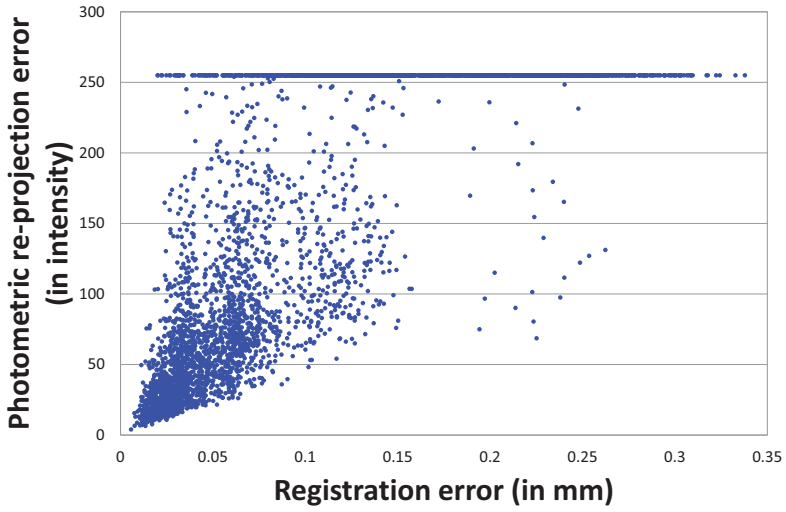


(b) Situation 2

Figure 5.5: Photometric re-projection error in function of the registration error for situations 1 and 2.



(a) Situation 3



(b) Situation 4

Figure 5.6: Photometric re-projection error in function of the registration error for situations 3 and 4.

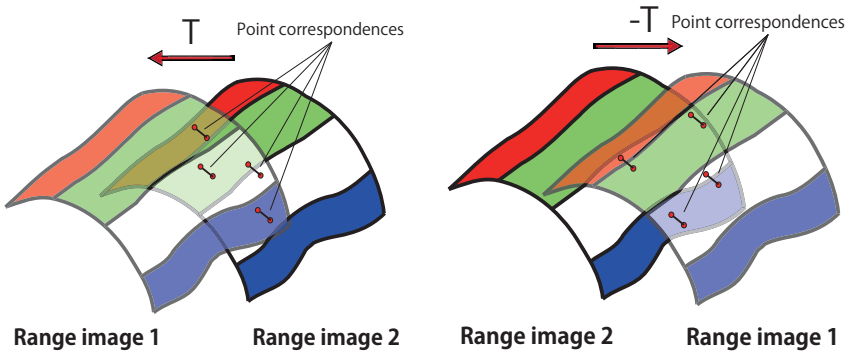


Figure 5.7: Two different transformations, having the same registration errors give different photometric re-projection errors.

directions the albedo values may change brutally (Fig. 5.8). This explains why in some cases even for a small change in the registration error, the photometric re-projection error changes drastically and why in some cases even for significant change in the registration error it changes slowly.

Finally, we observe that, as expected the transformations with the minimum photometric re-projection errors are also the ones with the minimum registration errors. We notice, however, that the extreme case where two different transformations give the minimum photometric re-projection error exists. Namely, this situation occurs if albedo and normals at the surface present the same symmetries. For example, if we consider a white ball illuminated by a spot light and rotated with a rotation R, then the minimum photometric re-projection error is given by both the solution $T = Identity$, with albedo equal to the appearance and illumination equal to an ambient light and the solution $T = R$, with uniform white albedo and illumination equal to the spot light. This extreme case rarely occurs in real situations and we may ignore such case from the practical point of view.

From these observations, we can conclude that (1) our proposed photometric metric will be difficult to minimize using traditional optimization methods, (2) a transformation with a small photometric re-projection error

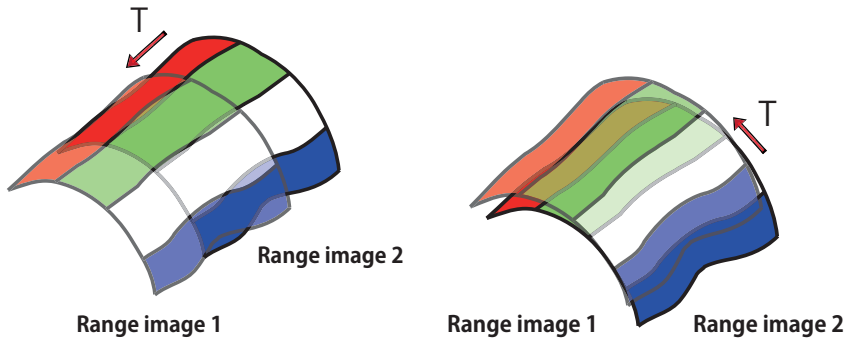


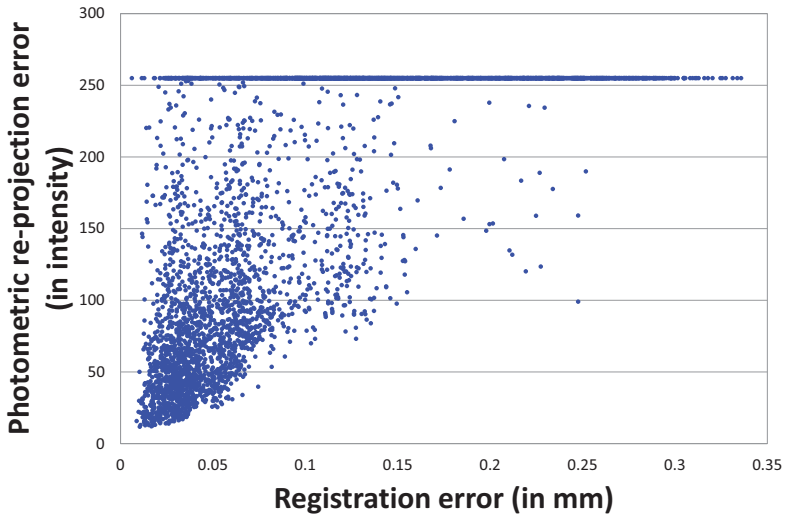
Figure 5.8: The photometric re-projection error varies differently for different directions of the transformation.

is close to the ground truth transformation error aligning the range images and (3) the minimal solution is almost insensitive to changes in illumination.

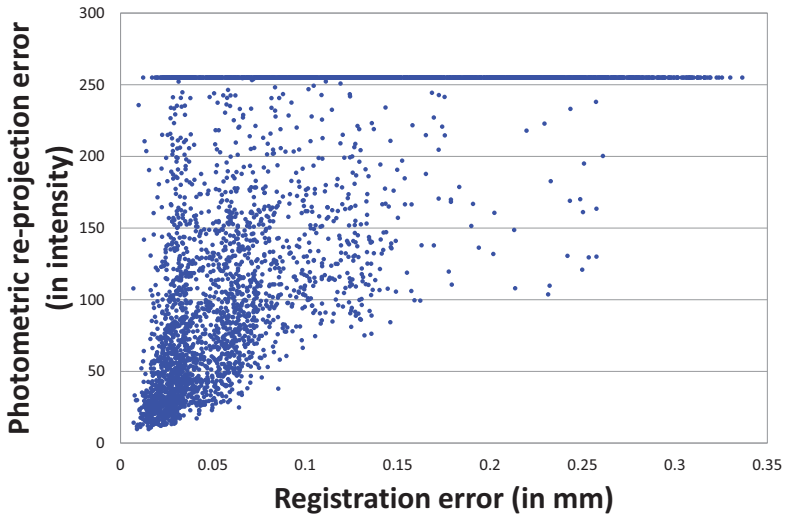
5.4.2 Our metric against different values of ϵ_s

Figures 5.9 and 5.10 illustrate our photometric metric with respect to different values for ϵ_s, for the two range images shown in figure 5.2. On one hand, we observe that when ϵ_s increases, the values taken by our metric become more concentrated for small registration errors and more sparse for large registration errors. The first effect can be explained by the fact that for a large epsilon, the chromaticity and normals around the stable points are homogeneous inside a large area. Therefore, for slightly different transformations, the attributes of the corresponding points change slightly and then the matrix M in equation (5.3) also. Thus the photometric re-projection error varies more slowly with respect to the transformation when ϵ_s increases. The second effect can be explained by the fact that when ϵ_s increases, the number of stable points decreases, and then the number of situations where the number of correspondences is insufficient for estimating *Eval* increases.

On the other hand, we observe that when ϵ_s increases, our photometric metric becomes less discriminative for evaluating the quality of the registration. This is because the attributes of the stable points become less discrim-

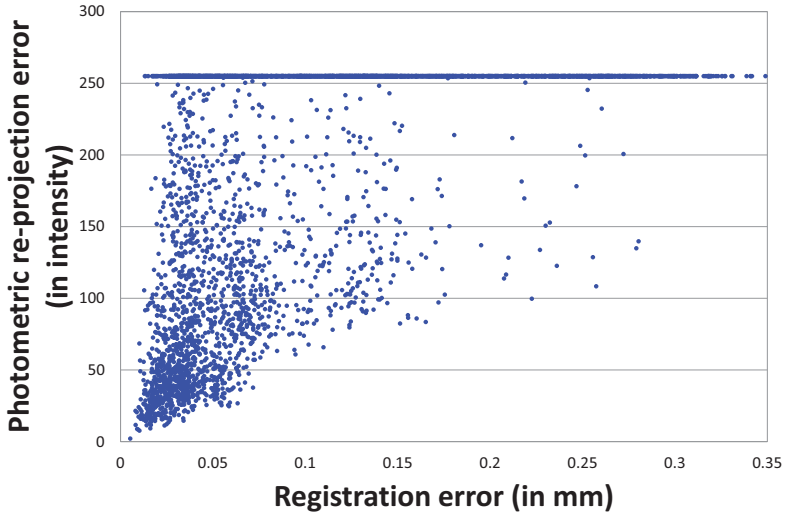


(a) $\epsilon_s = 0.015$

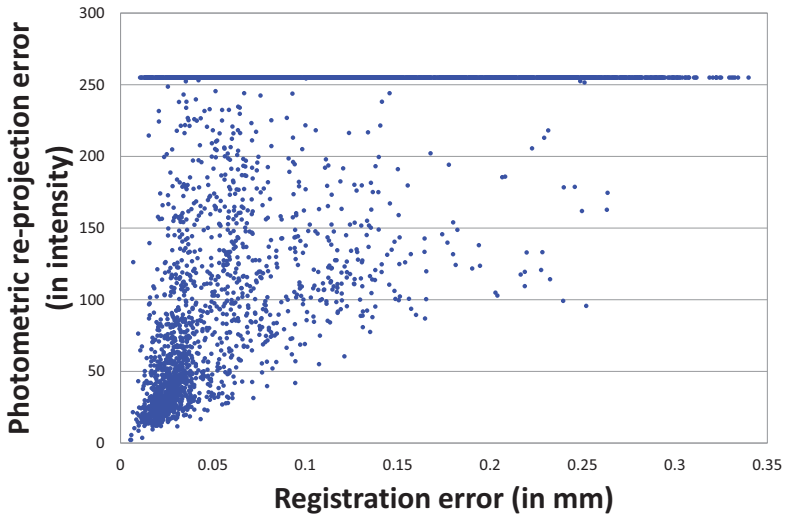


(b) $\epsilon_s = 0.02$

Figure 5.9: Photometric re-projection error in function of the registration error for $\epsilon_s = 0.015$ and 0.02.



(a) $\epsilon_s = 0.03$



(b) $\epsilon_s = 0.04$

Figure 5.10: Photometric re-projection error in function of the registration error for $\epsilon_s = 0.03$ and 0.04.

inative (similar chromaticity and normals for correspondences obtained for slightly different transformations).

In conclusion, a large value for ϵ_s brings more robustness to our metric, while a small ϵ_s brings more discriminative power to our metric (and thus increase its accuracy).

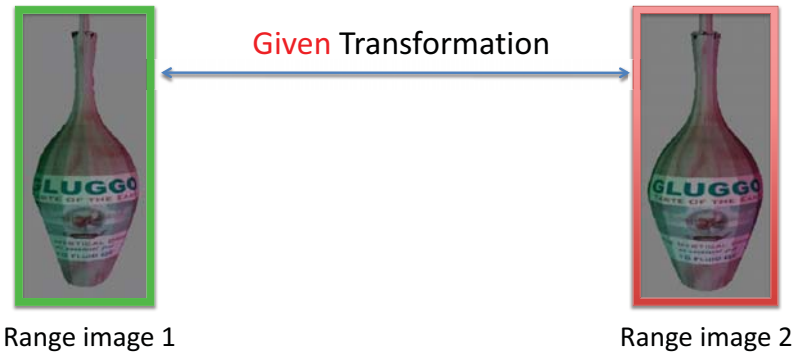
5.4.3 Time complexity analysis

Our proposed algorithm has its input of two range images with n and m points respectively and a given transformation T; it outputs the quantitative evaluation of the correctness of the transformation T with respect to the two input range images. Here we briefly give analysis on the computational complexity to our proposed algorithm. We refer to figures 5.11, 5.12, 5.13 and 5.13 for the different steps of our method and give the computational complexity for each of these steps.

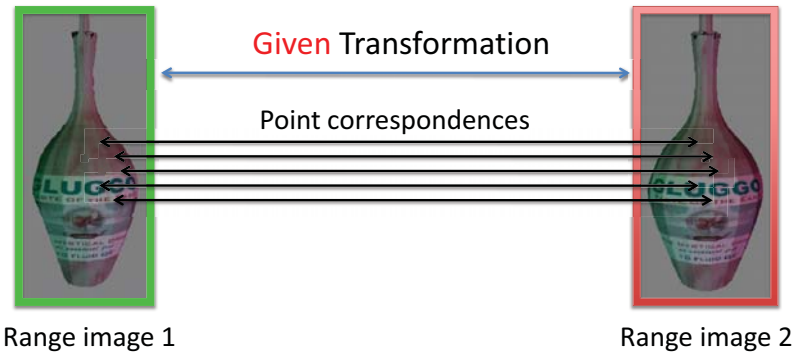
The point correspondences identification is perform in $O(n+m)$ operations. To build the linear system, we need to compute each element of the matrix M, whose size is $9 \times (n+m)$. It thus takes $O(n+m)$ operations (it also takes $O(n+m)$ operations to solve a linear system of size $n+m$ with 9 unknowns). It takes $O(n+m)$ operations to compute albedo at each point with respect to the estimated illumination and to transfer it to the corresponding point. To synthesize the colors of both range images and to compute the photometric re-projection error also takes $O(n+m)$ operations. Finally, our procedural photometric evaluation function requires $O(n+m)$ operations. We notice that, besides the computation of $Eval$, in each step all operations are done independently for all points. It is thus well fitted for parallel computing.

5.5 Registration

Given two overlapping range images, we seek a rigid transformation that minimizes the photometric re-projection error (5.6). When minimizing the

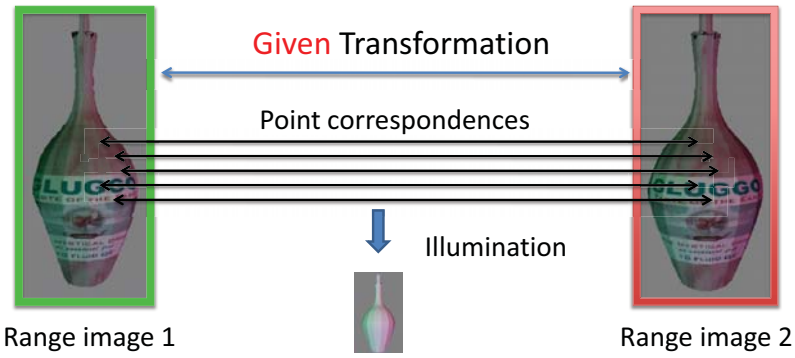


(a) Input

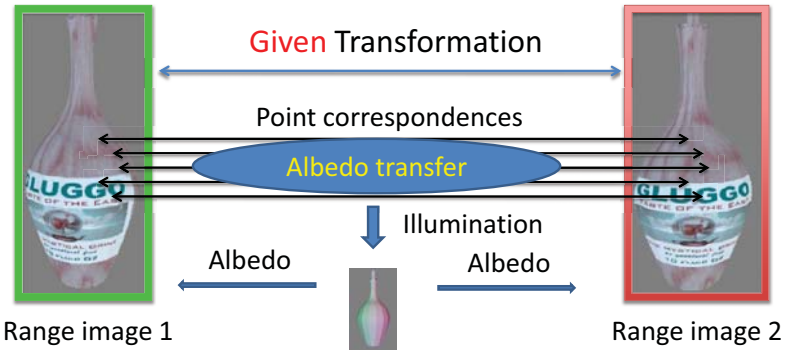


(b) Point correspondences identification

Figure 5.11: The procedural evaluation of our photometric evaluation function (part 1).



(a) Illumination estimation



(b) Albedo computation and transfer

Figure 5.12: The procedural evaluation of our photometric evaluation function (part 2).

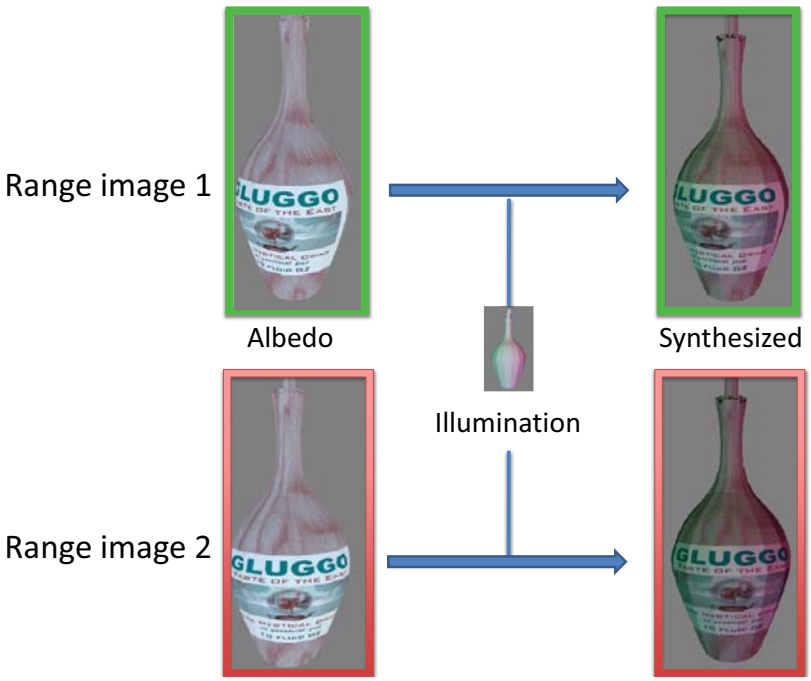


Figure 5.13: The synthesized colors.

photometric re-projection error, we have to decide the strategy we use. As we discussed in Section 5.4, our proposed photometric metric is hard to minimize using traditional optimization methods. Therefore, we choose the hypothesis-and-test search. Fig. 5.15 illustrates the flowchart of our proposed registration method.

5.5.1 Transformation search

The hypothesis-and-test search is performed by testing the quality of the registration for a set of rigid transformation candidates. The search ends when a rigid transformation accomplishing accurate registration is found or when all candidates are tested (the rigid transformation minimizing the cost function is then selected).

The most famous hypothesis-and-test search is the RANSAC method where candidates are generated from random triplets of correspondences.

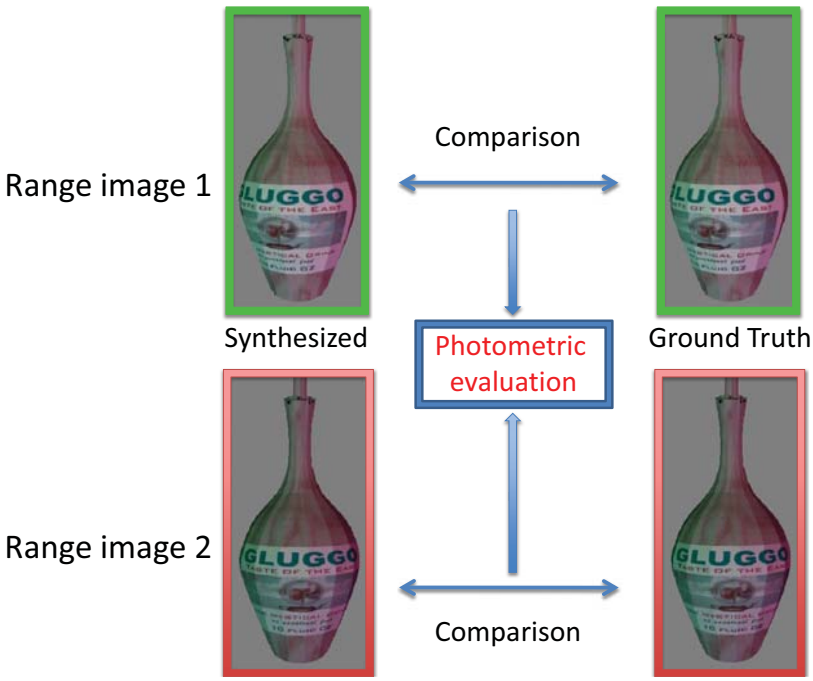


Figure 5.14: The photometric re-projection error.

Straightforwardly using RANSAC is computationally unrealistic. This is because we have potentially around 10^{12} possibilities for range images with 10^4 points. Therefore, how to efficiently search the best rigid transformation aligning range images becomes a critical issue.

On one hand, rigid transformations aligning two range images can be equivalently represented by sets of rigid point correspondences induced by the transformations. Therefore, if we represent the range image in another domain while keeping the rigidity of point correspondences, we can discuss the problem of searching the best transformation aligning the range images in this new domain.

On the other hand, the unit sphere is a convenient representation of a close-zero genus 3D surface. For closed surfaces, the spherical representation is pose invariant [110]. Therefore, the local structure in the spherical domain does not change and the rigidity of point correspondences is kept.

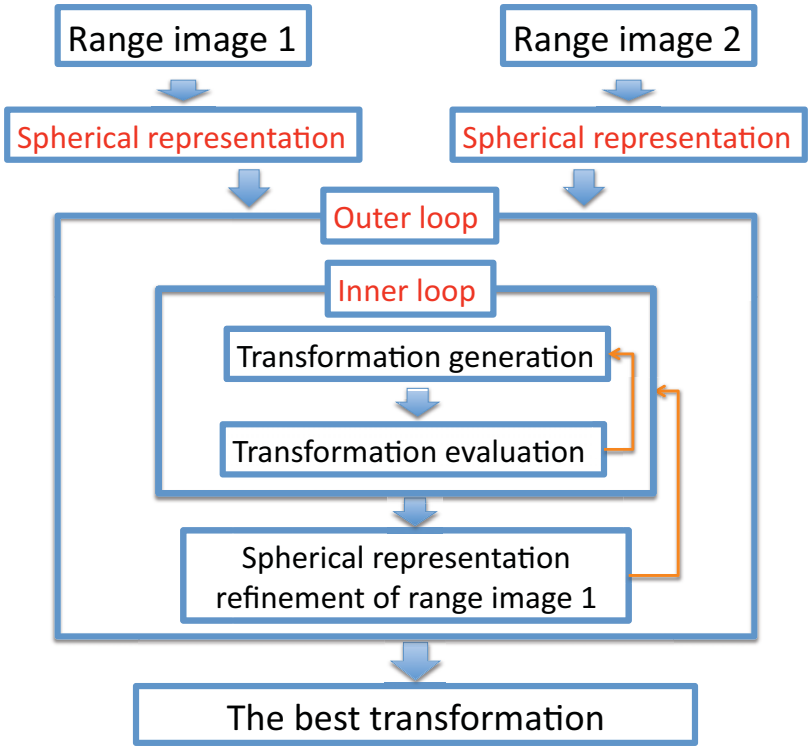


Figure 5.15: Flowchart of transformation search.

In addition, the rigid transformations aligning two spheres belong to $SO(3)$. We thus employ the spherical representation for range images. This representation reduces the transformation parameter space from $SO(3) \times \mathbf{R}^3$ to $SO(3)$.

The spherical representation of range images is, unfortunately, not pose invariant because surfaces in a range image are not closed. As a consequence, the local structure in the spherical domain may change in the original domain. This means that the rigidity of point correspondences in the spherical domain may not be kept in the original domain. To tackle this problem, we introduce refinement of the spherical representations throughout the registration process to reduce changes of the local structure in the spherical

domain as much as possible. Due to the possibility of violating rigidity of point correspondences in the original domain, we also have to generate the rigid transformations in the original domain from the point correspondences obtained in the spherical domain using the method proposed by Horn [36] as follows. A 3D rotation in $SO(3)$ gives us point correspondences in the spherical domain. In the original domain, we use the same point correspondences as the input of [36] to obtain the corresponding rigid transformation in $SO(3) \times \mathbf{R}^3$.

We remark that though we can use RANSAC method to generate transformation candidates from $SO(3)$ we prefer to use an exhaustive search to ensure convergence to the optimal solution. To reduce the computational time, we reduce $[0 : 360]^3$ to $[0 : \frac{20}{step}]^3$, where $step$ increases during the iteration. In the experiments, we set $step = 1$ at the first iteration and then $step = step^i$ for the iteration i.

5.5.2 Spherical representation and refinement

Our spherical representation method is inspired by the method proposed in [110] for closed surfaces. The input is an unorganized point set represented in the global coordinate system, and the output is a structured mesh with corresponding coordinates on the unit sphere that preserves the local structure. We note that the spherical representation of each range image is computed independently.

Spherical representation. We first orthogonally project all the 3D points of a range image along the z axis (viewing direction) to a plane. We then compute the convex hull of the projected points and identify the vertices of the convex hull. The vertices are used to generate Delaunay triangulations. The set of 3D points in the range image corresponding to the vertices of the convex hull is then projected to the unit sphere by normalizing the coordinates of each point. Next, we select a point (in the range image) that is not included in the vertices of the convex hull and carry out the

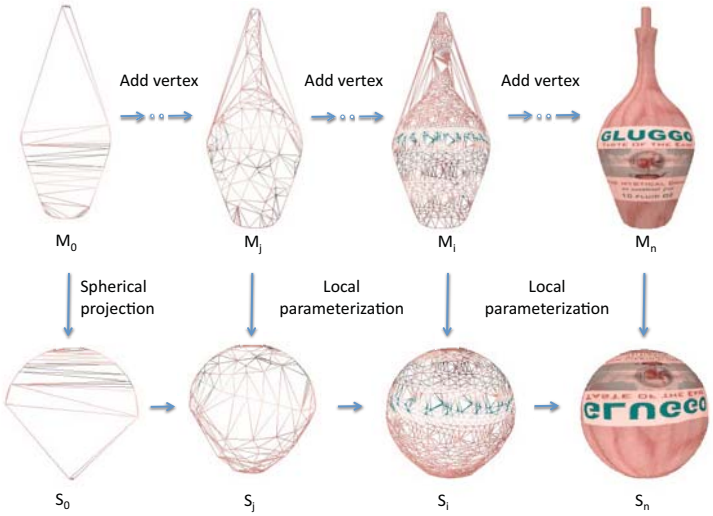


Figure 5.16: Illustration of the progressive spherical representation construction.

following process: we progressively construct triangulations by adding the point and compute the local position of the point with respect to the new triangulation. The local position is computed in the flattened vicinity of the point, obtained using conformal mapping [44], to accurately represent the local structure. The point is then positioned on the sphere using this local position. This series of processes is carried out until all the points in the range image are involved. The concrete procedure is described in Algorithm 1, and the progressive spherical representation is illustrated in figures 5.16 and 5.17.

Spherical representation refinement. We consider two range images: range image 1 and range image 2. Without loss of generality, we consider the problem of aligning range image 1 to range image 2. After each iteration, we refine the spherical representation of range image 1 with respect to range image 2.

Algorithm 1 Progressive spherical representation

Require: range image RI

Ensure: spherical representation S of RI that preserves the local structure of RI

$Mesh_0 \leftarrow$ triangulated 2D convex hull of RI, centered around its center of mass

$PtMesh \leftarrow$ set of vertices belonging to $Mesh_0$

$S_0 \leftarrow$ projection of $Mesh_0$ to the sphere

$List \leftarrow$ list of points in RI but not in $PtMesh$

$nb \leftarrow$ size of $List$

for $i = 0$ to $nb - 1$ **do**

$\mathbf{\Psi} \leftarrow$ a point of $List$

$(\mathbf{a}, \mathbf{b}, \mathbf{c}) \leftarrow$ three points of the enclosing triangle of $\mathbf{\Psi}$ in $Mesh_i$

$Mesh_{i+1} \leftarrow$ Delaunay triangulation of $Mesh_i$ where $\mathbf{\Psi}$ has been added

$P \leftarrow$ polygon in $Mesh_{i+1}$ containing the points $(\mathbf{\Psi}, \mathbf{a}, \mathbf{b}, \mathbf{c})$

$U \leftarrow$ flatten polygon obtained by conformal mapping of P [44]

$(\alpha, \beta, \gamma) \leftarrow$ barycentric coordinates of $\mathbf{\Psi}$ in U for $(\mathbf{a}, \mathbf{b}, \mathbf{c})$

$\mathbf{\Psi}' \leftarrow \alpha\mathbf{a}' + \beta\mathbf{b}' + \gamma\mathbf{c}'$

$S_{i+1} \leftarrow S_i + \{\mathbf{\Psi}'\}$, with the same connectivity as in $Mesh_{i+1}$

$List \leftarrow List - \{\mathbf{\Psi}\}$

end for

return S_n

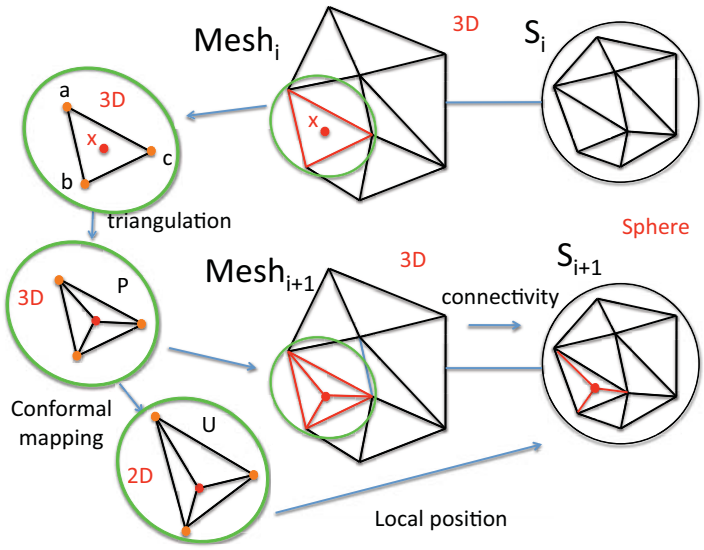


Figure 5.17: Illustration of the loop "for" of Algorithm 1.

The overlapping areas between the two range images from the current best transformation are first identified. The bijection B between points of the two overlapping areas is then computed. Namely, for a point $\mathbf{x}$ in the overlapping area O_1 of range image 1, $B(\mathbf{x}) = closest(\mathbf{x})$ if $\mathbf{x} = closest(closest(\mathbf{x}))$, $B(\mathbf{x})$ is undefined otherwise. Here, *closest* stands for the closest point (in the sense of the Euclidean distance) in the overlapping area of the other range image. Then, for each point of O_1, its coordinates on the sphere are set to those of its closest point. The remaining points of O_1 that do not have an image for B are placed on the sphere using the local positions as we did in the spherical representation above.

We notice that we identify the closest points in the spherical domain in the same way as explained in section 5.3.3.

5.5.3 Time complexity analysis

Our proposed algorithm has its input of two range images with n and m points respectively and outputs the transformation T that best aligns the

two input range images. Here we briefly give analysis on the computational complexity to our proposed algorithm. We refer to figure 5.15 for the different steps of our method and give the computational complexity for each of these steps.

At each inner iteration, our method performs two step: (1) 3D rigid transformation candidate generation; (2) evaluation of the candidate transformation. During step (1), we identify point correspondences in the spherical domain then estimate the rigid transformation for the correspondences in the original domain. The closest point identification in the spherical domain is done in $O(n)$ computations. Estimating the rigid transformation for the point correspondences is done in $O(n)$ computations and evaluating the transformation takes $O(n+m)$ computations. Therefore for l inner iterations, the inner loop takes $O(l \times (n+m))$ computations.

Identifying the current overlapping areas takes $O(n+m)$ computations. Therefore refining the spherical representation of range image 1 is done in $O(n+m)$ computations. Finally, for l inner iterations and h outer iterations, our registration method takes $O(h \times l \times (n+m))$ computations. We notice that all inner iterations are completely independent. Therefore using parallel implementation would drastically reduce the computational time.

5.6 Experiments

To demonstrate the usefulness of our proposed method, we evaluate our algorithm in several challenging situations using synthetic and real data. For the comparison, we used three methods: the proposed method using chromaticity instead of our evaluation function (Method 1); the method proposed in Chapter 3 using albedo with a given directional light source (Method 2), and the method proposed in Chapter 3 using chromaticity instead of albedo (Method 3). Whereas comparing our method with Method 1 shows the advantage of using photometry, comparing it with Methods 2 and 3 shows the advantage of our search strategy. We also compared our method with the alignment obtained by matching using SIFT (we used the available

code provided by Andrea Vedaldi [101]). We notice that all data are devoid of salient geometric features, and thus, using geometric feature-based registration methods does not work in these cases.

We consider the problem of aligning range image 1 to range image 2 and we assume we are given the ground truth (obtained manually for real data). We evaluate the registration result using the distance between the estimated position of points of range image 1 after registration and their ground truth position. Namely, given T_g and T_e the ground truth transformation and the estimated transformation respectively, the registration error $err(T_e)$ is computed as follows:

$$err(T_e) = \frac{\sum_{i=1}^{n} \|T_g(\mathbf{x}_i) - T_e(\mathbf{x}_i)\|_2}{n}, \tag{5.10}$$

where n is the number of points in range image 1 and $\{\mathbf{x}_i\}_{i\in[1;n]}$ are the points of range image 1.

ϵ_s, ϵ_c, and ϵ_n were set to $3 * res$, 0.02, and 0.1 respectively for all experiments with synthetic data and to $2 * res$, 0.05, and 0.2 for all experiments with real data.

5.6.1 Synthetic data

(a) The data *vase*

The synthetic data, called *vase* were obtained using a 3D modeler software (3D Studio Max) (see Table 5.1). The exact albedo is known and we simulated lighting under different illuminations. This data set is challenging for registration in that the shape is rotationally symmetric, the texture of the objects presents several repetitive patterns, and no exact correspondences exist between the two range images.

The first set-up is illustrated in figure 5.18. The range images were illuminated by the light probe *galileo* from the Debevec database [18], and the color was synthesized using spherical harmonics with a Lambertian reflection kernel. Figure 5.19 shows the results obtained with the five methods

Table 5.1: Description of the data *vase*.

Nb_Points	Resolution	Expected_rot (angle; axis)
30650	0.01 mm	$(-20.0; 0.01, 0.93, 0.34)$
Expected_translation		$(0.19, -0.01, 0.03)$

mentioned above. In this situation, the illumination is dominated by ambient light, and the changes in appearance between the two range images are small. This situation is thus well adapted to using chromaticity. Nevertheless, our evaluation function worked well compared with the chromaticity cross correlation. Moreover, because of repetitive textures and distant initial positions, Methods 2 and 3, i.e., methods using an ICP-like framework, fell into a local minimum and failed pitifully in aligning the range images, even when using chromaticity. Our proposed search strategy, on the contrary, efficiently found the global minimum. The estimated transformation T_e by our proposed method was a rotation of $(-19.9; 0.00, 0.94, 0.35)$ and a translation of $(0.19, -0.01, 0.03)$, and $Eval(T_e)$ was[3] 1.2. It was a rotation of $(-19.9; 0.00, 0.94, 0.35)$ and a translation of $(0.19, -0.01, 0.03)$ when using chromaticity instead of our evaluation function.

The second set-up is illustrated in figure 5.20. The illumination was composed of three directional light sources of different intensities as well as an ambient light source. The images were rendered using the standard Lambertian model. Figure 5.21 shows the results obtained with the five methods. In this situation, the illumination induces significant changes in the object appearance (e.g. the color of several points changed from reddish to white), and the use of chromaticity is no longer effective for both our search strategy and ICP-like methods. Method 2 failed in aligning the range images because it assumes that the object is illuminated by only a single light source. Our

[3]The color was coded in RGB with values between 0 and 255. Since the color of a point is approximated with 98% accuracy in Eq. (5.2), the re-projection error of around 1.2 means that the optimal solution is found provided that the distribution of RGB values is uniform over the range.

Two input range images

Light probe

Figure 5.18: The first set-up.

Proposed method | Method 1 | Method 2 | Method 3 | SIFT

Method used	Proposed method	Method 1	Method 2	Method 3	SIFT
Error (in mm)	0.005	0.006	0.415	0.415	0.156

Figure 5.19: Results obtained with the five methods for the data *vase*.

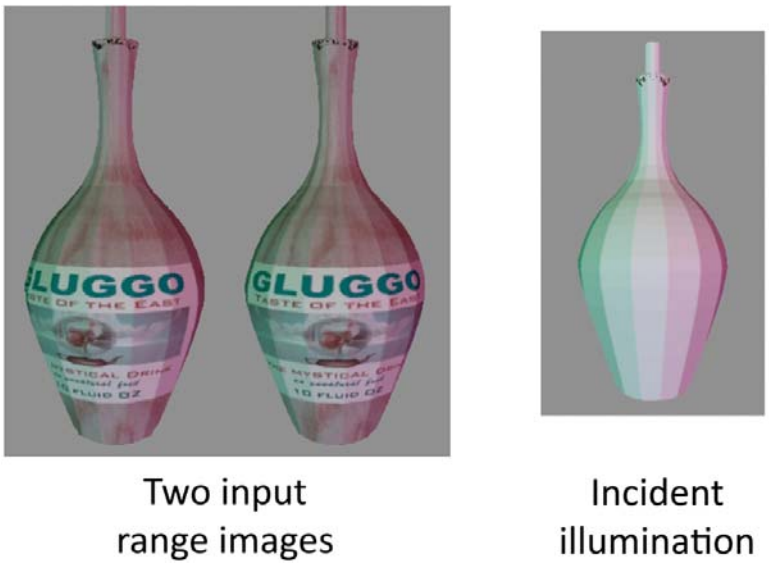


Figure 5.20: The second set-up.

method is the only one that achieved accurate alignment (similar to the one obtained in the previous set-up). The estimated transformation T_e by our proposed method was a rotation of $(-20.1; 0.00, 0.94, 0.35)$ and a translation of $(0.19, -0.01, 0.03)$, and $Eval(T_e)$ was 3.4.

Results obtained using SIFT key-point detector and descriptor for estimating point correspondences are also illustrated in both figure 5.19 and figure 5.21. Due to the projective deformations, changes in intensity and repetitive patterns, the SIFT-based method did not work in our situation.

The results obtained with our method for the various situations presented in Section 5.4.1 are showed in figure 5.22, along with the graphs of the photometric re-projection error evaluated during the registration process. We can see that during the registration, the lower bound of our photometric metric is always evaluated, which allowed us to converge and validated our search strategy. However, we remark that when ϵ_s is too small, the registration result is degraded. This is because the quality of the point correspondences becomes poor.

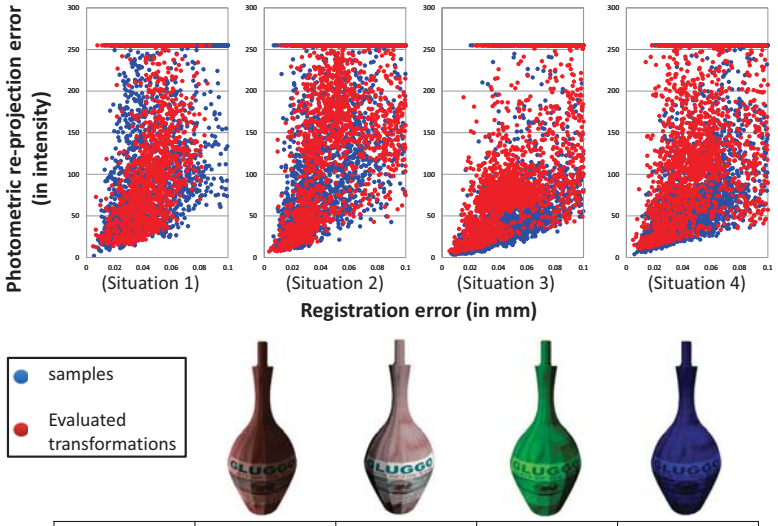
Method used	Proposed method	Method 1	Method 2	Method 3	SIFT
Error (in mm)	0.005	0.410	0.415	0.415	0.120

Figure 5.21: Results obtained with the five methods for the data *vase*.

Figure 5.23 shows the results of intensive experiments when adding noise into the color. A random noise was added to the rendered colors with different level independently in each RGB channel. For each level of noise, our method was run 40 times. Figure 5.23 (a) shows the errors after registration and figure 5.23 (b) shows the number of runs for which our method failed in registering the two range images (i.e. the error was greater than 5 times the resolution of the range image). From these results, we can see that our method can achieve accurate results even in the presence of noise in the captured colors. However, when the noise becomes larger, our method happens to fail from time to time. This can be explained by the discretization of the searching space. To reduce the number of failure, it is possible to refine the discretization of the searching space (that will increase the computational time).

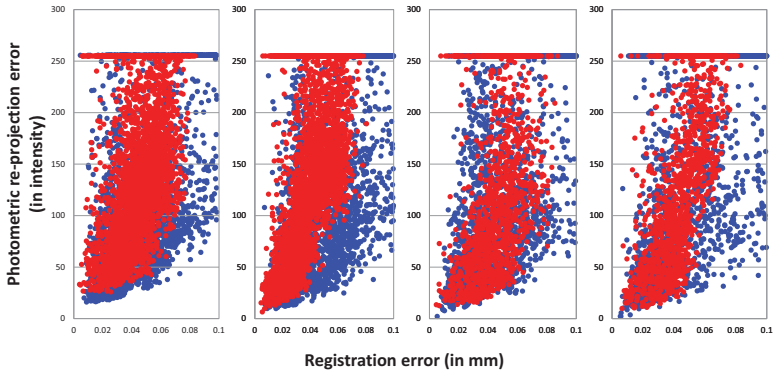
(b) The data *Al*

Results obtained with our method for another synthetic data called *Al* in two different scenarios are shown in figures 5.25 and 5.27. This synthetic



	(Situation 1)	(Situation 2)	(Situation 3)	(Situation 4)
Error	0.007	0.007	0.011	0.011

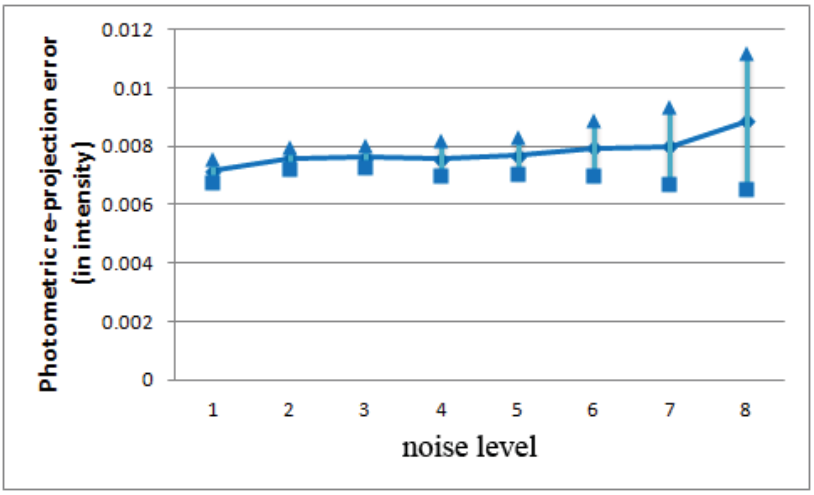
(a) Different illumination conditions



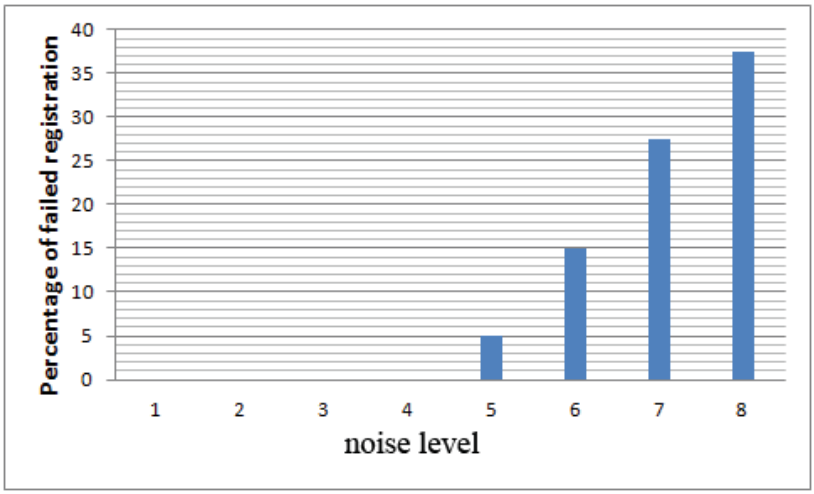
	(a)	(b)	(c)	(d)
Error	0.028	0.005	0.007	0.007

(b) Different ϵ_c

Figure 5.22: Registration results obtained with our method for the various situations presented in Section 5.4. The transformations evaluated during the registration process are plotted in red and superimposed onto the graphics shown in Section 5.4



(a) Errors after registration



(b) Number of registration failure

Figure 5.23: Experiments with different level of noise added to the color.

data was of resolution about 0.01 mm.

For the scenarios 1 (figure 5.24), the two range images were rotated by 40.0 degrees around the vertical axis and illuminated by a fixed directional light source. The observed colors were synthesized using the classic Lambertian reflection model. This situation is challenging in that the change in pose between the two range images induces significant change in the intensity. Moreover, there are a large number of occluded points and few number of points with salient photometric features, which make the matching problem more difficult to solve.

For the scenario 2 (figure 5.26), the two range images were rotated by 20.0 degrees around the vertical axis. The texture of the object was removed and the observed colors were synthesized using the light probe showed in figure 5.18 and using the spherical harmonics reflection model. This situation is challenging in that the object is textureless. In this case, using textural information alone such as color, chromaticity or even albedo does not help registering the two range images.

Our method successfully registered the two range images in all scenarios. In particular, it is interesting to notice that our method successfully registered the two range images even for the case of textureless surface. This because the relationship between normals and lighting direction is sufficiently discriminative in this situation. We note that when not using texture, our photometric metric derivation problem reduces to the shape-from-shading problem.

To verify the robustness of our method against the initial relative pose between the input range images, we evaluated our method under various rotations. We restricted our experiment to the situation were the rotation axis is fixed, while the rotation angle varies. Namely, we captured several range images of the synthetic data *Al*, from the same viewpoint but under different poses. Each pose was obtained by rotating the 3D model around the vertical axis, with different rotation angles (ranging from 1.0 degree up to 50.0 degrees) and in the clock-wise direction. Figure 5.28 shows the

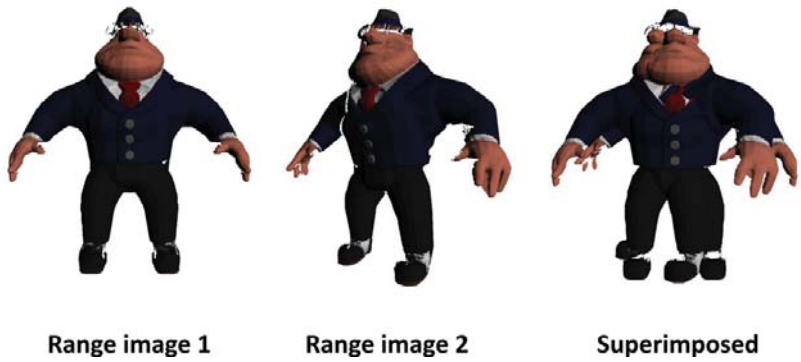


Figure 5.24: The input range images with initial estimate of registration for the scenario 1.

Method used	Proposed method	Method 1	Method 2	Method 3	SIFT
Error (in mm)	0.007	0.008	0.191	0.222	0.543

Figure 5.25: Registration results obtained with our method for the data *Al* for the scenario 1.

Figure 5.26: The input range images with initial estimate of registration for the scenario 2.

Method used	Proposed method	Method 1	Method 2	Method 3	SIFT
Error (in mm)	0.008	0.024	0.122	0.171	0.403

Figure 5.27: Registration results obtained with our method for the data *Al* for the scenario 2.

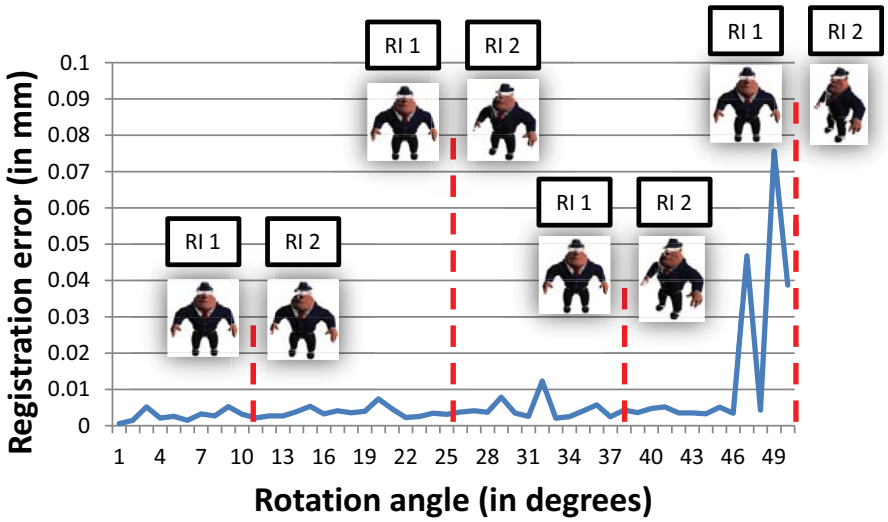
results obtained with our method. We can see that for a rotation angle up to 46.0 degrees, we could always obtain accurate registration results (for a rotation angle of 46.0 degrees, the percentage of overlapping area between the two range images was about 80.0 percent). However, for a rotation angle greater than 46.0 degrees, we observe failures in the registration. This can be explained by the fact that the initial spherical representations of the two input range images (computed independently) become too much different. As a consequence, as we can see in figure 5.28 (b) it prevents us from generating a searching space fine enough for convergence. In particular, the transformation that correctly aligns the two range images (which is also the minimum of our photometric evaluation function) is not generated.

5.6.2 Real data

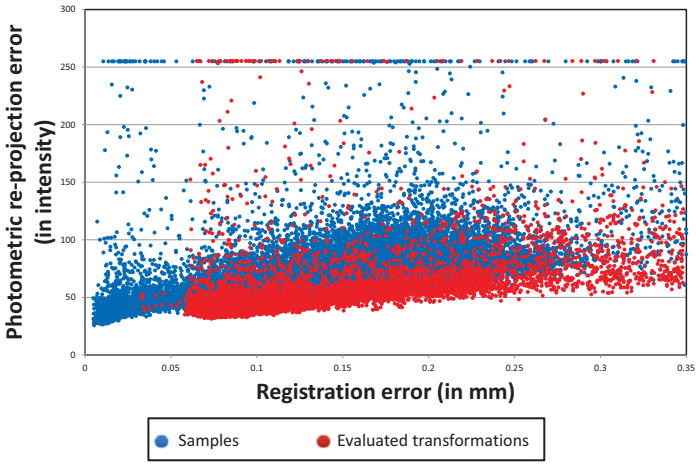
We employed a Konica Minolta Vivid 910 range scanner, which captures the 3D shape and the texture of an object. The ground truth transformation was obtained manually. We note that there is a gamma correction factor in the obtained color images which should preferably be cancelled. In our experiments, however, we did not know this factor, and thus the gamma correction was not cancelled.

(a) The data *can*

We obtained two range images of a rotationally symmetric can that is approximately 10.0 cm high and has a diameter of about 5.0 cm (figure 5.29). Details on the data called *can* are given in Table 5.2, and the results are shown in figure 5.30. These data are challenging in that the quality of the image is low, and there is an unknown gamma correction factor. Moreover, this data exhibits several repetitive patterns such as similar letters while the texture is either red or white with large uniform areas. Nevertheless, our proposed method accurately registered the two range images. The obtained accuracy was under the resolution of the range sensor, and our method worked extremely well compared to the other methods. The



(a) Registration results obtained with our method for the data *Al* for the scenario 1 under various different initial poses ("RI 1" and "RI 2" stand for "range image 1" and "range image 2", respectively).



(b) Plot of the registration process for a rotation with an angle of 50 degrees.

Figure 5.28: Experiments against various initial relative pose.

Table 5.2: Description of the data *can*.

Nb_Points	Resolution	Expected_rot (angle; axis)
28000	0.55 mm	$(20.0; 0.01, 0.93, 0.34)$
Expected_translation		$(9.00, 0.10, -1.00)$

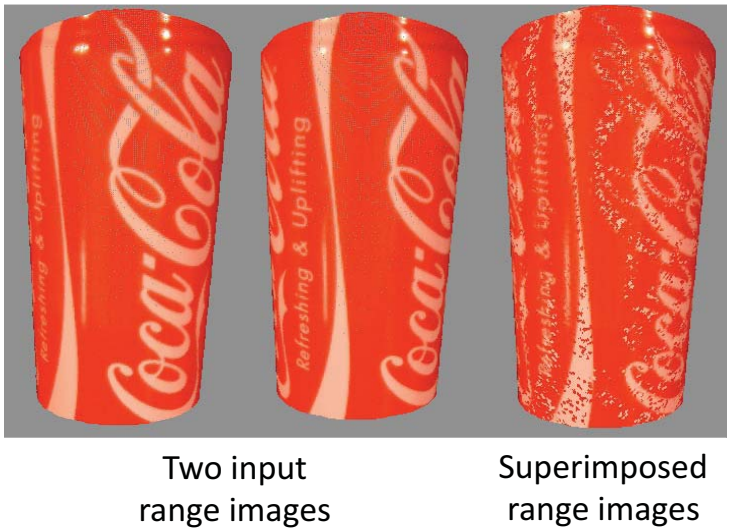
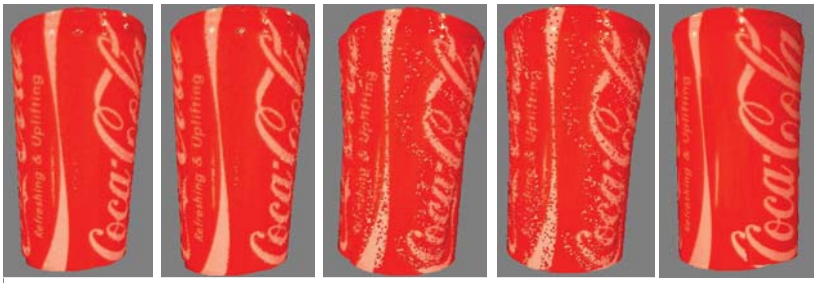


Figure 5.29: Input range images and initial positions for the data *can*.

estimated transformation T_e obtained with our proposed method was a rotation of $(19.7; 0.01, 0.94, 0.34)$ and a translation of $(8.93, 0.37, -1.19)$, and $Eval(T_e)$ was 2.56. We obtained a rotation of $(19, 2; 0.03, 0.95, 0.30)$ and a translation of $(9.06, -0.18, -1.11)$ by Method 1.

(b) The data *hand*

Another data item called *hand* is presented in figure 5.31 and in Table.5.3. Registration results are shown in figure 5.32. For this data, the intensity of a point in two range images changed drastically (e.g. points at the middle of the images). Therefore, the use of chromaticity to evaluate the goodness of transformations becomes unreliable. Our method is the only one that



Proposed method | Method 1 | Method 2 | Method 3 | SIFT

Method used	Proposed method	Method 1	Method 2	Method 3	SIFT
Error (in mm)	0.517	0.869	17.6	16.5	8.80

Figure 5.30: Results obtained with the five methods for the data *can*.

Table 5.3: Description of the data *hand*.

Nb_Points	Resolution	Expected_rot (angle; axis)
50000	0.55 mm	$(20.0; 0.01, 0.93, 0.34)$
Expected_translation		$(6.60, -2.3, -0.40)$

achieved accurate registration of the two range images. The gap in accuracy between our proposed method and Method 1 became larger than that for the data *can*. This is because drastic changes in intensity degrade reliability of chromaticity while our method uses a photometric metric. Method 2 and Method 3 still failed pitifully.

Results obtained using SIFT key-point detector and descriptor for estimating point correspondences are also illustrated in both figure 5.30 and figure 5.32. Due to the repetitive patterns for the data *can* and significant changes in intensity for the data *hand*, the SIFT-based method did not work well in our situation.

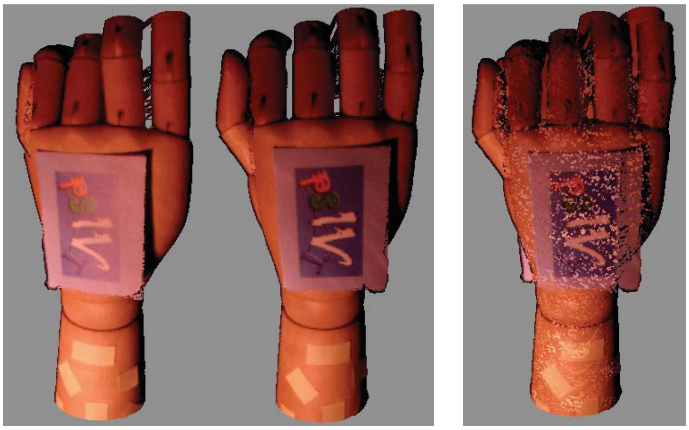


Figure 5.31: Input range images and initial positions for data *hand*.

Method used	Proposed method	Method 1	Method 2	Method 3	SIFT
Error (in mm)	0.551	3.06	16.8	21.3	6.29

Figure 5.32: Results obtained with the five methods for data *hand*.

Table 5.4: Description of data *Hand 2*, *Base* and *Cylinder*.

	Nb_Points	Res	Expected_rot (angle; axis) Expected_trans
Hand 2	70000	0.55 mm	$(31.0; -0.02, -0.97, -0.21)$ $(-11.0, 0.76, -1.24)$
Base	210000	0.55 mm	$(20.2; 0.05, 0.91, 0.42)$ $(309, 4.57, -48.13)$
Cylinder	80000	0.55 mm	$(32.0; 0.03, 0.96, 0.26)$ $(10.0, -0.50, -4.00)$

(c) Other data

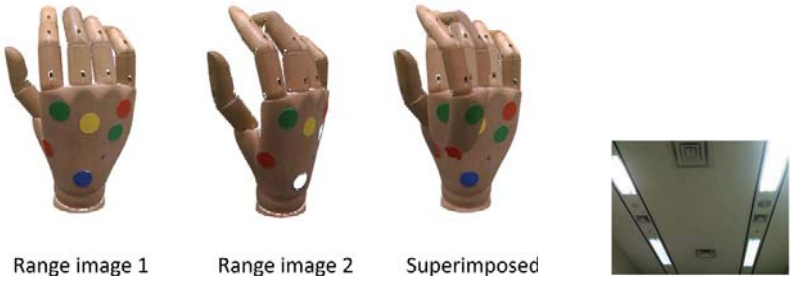
Figures 5.34, 5.36, 5.38, 5.40 and 5.42 show the results obtained with the different objects called *Hand 2*, *Base*, *Cylinder 1*, *Cylinder 2* and *Cylinder 3*. Table 5.4 lists the description of the data before registration. Figures 5.38, 5.40 and 5.42 show the result obtained with the data *Cylinder* for the same changes in pose but under different illuminations. Note that data *Base* had a height of about 15 cm and a width of about 15 cm, and data *Hand 2* and *Cylinder* had a height of about 15 cm and a width of about 5 cm. Figures 5.33, 5.35, 5.37, 5.39 and 5.41 show initial estimates of registration.

The data *Hand 2* is challenging in that there are only few points with discriminative photometric features. Because in Method 2 and Method 3 we do not use key-point detector, the number of outliers in matching becomes larger than the number of inliers. This results in failed registration. The SIFT approach uses key-point detector and key-point descriptor identified in the intensity images for matching. In the intensity images, however, the texture patterns (the round stickers) are repetitive, which lead to some mismatches. Because there are only a few key-points detected (we detected about 35 key-points) this leads to failure in the registration. On the other hand, our proposed method uses a global error metric and a hypothesis-and-test search, which allowed us to successfully register the two range images.

We notice that in this case the ambient illumination allows for accurate registration even when using chromaticity.

The data *Base* is challenging in that its shape is rotationally symmetric while its texture does not exhibit clear key-points with distinctive features that could be used for matching. Because in Method 2 and Method 3 we do not use key-point detector, the number of outliers in matching becomes larger than that of inliers. This results in failed registration. The SIFT approach uses key-point detector and key-point descriptor identified in the intensity images for matching. In the intensity images, however, the texture patterns are not distinctive enough, which leads to some mismatches and results in failed registration. On the other hand, our proposed method uses a global error metric and a hypothesis-and-test search, which allowed us to successfully register the two range images. We notice that using chromaticity in this case did not work. This is because the difference of chromaticity between points of the two range images is not discriminative enough to find the best transformation.

The data *Cylinder* is challenging in that due to its reflective properties there are several missing points and large noise in the depth values. This is because the accuracy of the laser scanner decreases when the texture at the surface becomes black (the laser beam is then not properly reflected), which is the case for all letters. This effect has dramatic impact on the Methods 2 and 3 proposed in Chapter 3 for growing the regions. The missing points may prevent the region from growing in one range image while it will continue growing in the other range image. In addition, the noise in the depth values amplify the distortion between the descriptors of the same point in the two range images. This results in failed registration. The SIFT method did not work in this case neither. This can be explained by the repetitive patterns of the letters and deformations due to perspective projection. On the other hand, with our proposed method, we could obtain accurate registration results for all situations.



(a) Data *Hand 2*. (b) Lighting.

Figure 5.33: Initial estimates of registration for the data *Hand 2*.

Method used	Proposed method	Method 1	Method 2	Method 3	SIFT
Error (in mm)	0.181	1.64	11.0	11.6	8.78

Figure 5.34: Registration results for the data *Hand 2*.

(a) Data *Base*. (b) Lighting.

Figure 5.35: Initial estimates of registration for the data *Base*.

Method used	Proposed method	Method 1	Method 2	Method 3	SIFT
Error (in mm)	0.823	19.7	27.4	27.7	34.95

Figure 5.36: Registration results for the data *Base*.

(a) Data *Cylinder 1*. (b) Lighting.

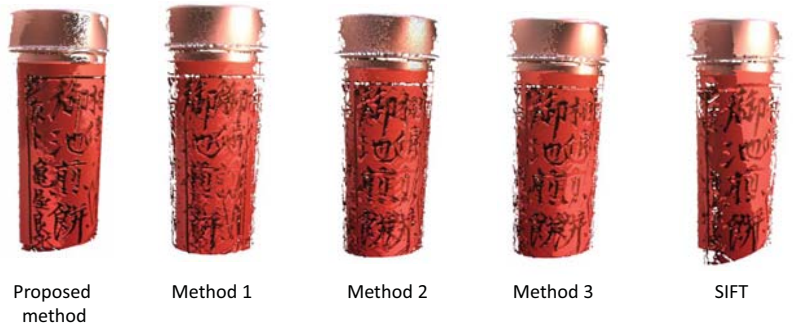
Figure 5.37: Initial estimates of registration for the data *Cylinder 1*.

Method used	Proposed method	Method 1	Method 2	Method 3	SIFT
Error (in mm)	0.372	1.41	17.0	13.0	6.49

Figure 5.38: Registration results for the data *Cylinder 1*.

(a) Data *Cylinder 2*. (b) Lighting.

Figure 5.39: Initial estimates of registration for the data *Cylinder 2*.



Method used	Proposed method	Method 1	Method 2	Method 3	SIFT
Error (in mm)	1.13	19.4	14.3	14.6	5.65

Figure 5.40: Registration results for the data *Cylinder 2*.

(a) Data *Cylinder 3*. (b) Lighting.

Figure 5.41: Initial estimates of registration for the data *Cylinder 3*.

Method used	Proposed method	Method 1	Method 2	Method 3	SIFT
Error (in mm)	0.883	17.3	15.5	16.2	8.34

Figure 5.42: Registration results for the data *Cylinder 3*.

5.7 Conclusion

We introduced a new photometric metric for registering range images of Lambertian surfaces under general and distant, unknown lighting. Our metric evaluates photometric re-projection error by taking into account the relationship between reflectance properties, geometry and illumination. We use captured color of range images as the ground truth to eliminate scale ambiguity that arises when estimating photometric features. We also demonstrated the effectiveness of our metric by using the hypothesis-and-test strategy for the registration where a range image is represented over the sphere and its representation is refined throughout the registration process. We notice that different approaches to using our photometric metric for registration exist. For example, we can use a coarse registration method as an initialization of a brute-force search in the vicinity of the initial estimate of the transformation. We chose to use the spherical representation of range images to reduce the searching space because it allows us not to depend on the quality of an initial coarse registration, making the overall registration method more stable against the initial conditions.

Chapter 6

Conclusion

6.1 Summary

Accurate 3D models of real objects are of crucial importance. Thereby, many automated 3D modeling methods have been proposed in the last decades. From a set of observations, the approach is to first reconstruct partial 3D surfaces of the object, align them together and then fuse them together to obtain a full compact representation of the object. With the development of 3D sensors, the acquisition of partial 3D surfaces (also called range images) is more and more accessible and aligning the acquired range images (known as range image registration) becomes the most limited step of the 3D modeling process.

Though remarkable advances have been done in the last decade for registering range images there are still several limitation cases. In particular aligning range images devoid of salient geometric features when the appearance of the object's surface changes significantly during the acquisition process remains an open problem.

On the other hand, the recent advances in image formation understanding brings new possibilities for registering range images using photometry. In this dissertation we addressed the use of photometry for accurately registering pairs of range images devoid of salient geometric features. First, we

designed, implemented and evaluated a robust local descriptor that overcomes the drawbacks of current methods using albedo for Lambertian objects under simple illumination. Second, we proposed an albedo estimation strategy for the case of specular objects illuminated by a few unknown point light sources that enlarges the range of applications of our previously proposed registration method. Third, we proposed a photometric metric for registering Lambertian range images under unknown general illumination and proved its usefulness through a practical registration method.

In this work, we made significant advances in using photometry for registering pairs of overlapping range images. In a broad sense, we could enlarge the practicability and range of applications of range image registration.

Robust 3D descriptor. Our proposed robust 3D descriptor allowed us to accurately register pairs of Lambertian range images devoid of salient geometric features even under a rough estimate of the incident illumination. While current methods used restrictive assumptions on the acquisition environment, such as color constancy or known illumination, we significantly improved the practicability of the registration process by allowing deviations from the color constancy assumption, even under rough estimations of illumination.

Specularity removal By estimating albedo in parts of specular surfaces illuminated by a few unknown point light sources, we could obtained accurate registration of pairs of range images even in the presence of large specular highlights on the surface, which was not possible with current range image registration methods. We thus enlarged the range of applications of range image registration.

Photometric metric under unknown lighting The image formation is a complex process and, for a user who is not expert on this domain, extracting the reflectance properties of a surface under general illumination from its range images is a difficult task, even when the illumination is known.

The task becomes even harder, if not impossible, when the illumination is unknown. Our proposed photometric metric allows us to register pairs of Lambertian range images acquired under unknown general illumination, without any required knowledge on the image formation process. The few parameters required (for the stable points identification and searching step) are easy to understand, which makes our registration method easy to use regardless of the surrounding illumination. This approach thus enlarged the practicability and range of applications of range image registration.

6.2 Future perspectives

In a broad sense, photometry accounts for the relationship between geometry, surface reflectance and illumination. The natural combination of geometry and surface reflectance behind image formation is of great interest for evaluating the similarity between different range images. While photometry has received a lot of attention in the field of inverse rendering, the attempts of using it for range image registration are rare. Therefore, using photometry for registering range images deserves more attention in the future. In particular, four directions deserve to be investigated.

Speed-up Though in this work we have not paid close attention to the computational cost of our registration techniques, speeding-up our proposed methods will be an interesting direction for future work. In particular, with the recent development of real time depth sensors (such as the PrimeSense or Kinect structured light sensors), much interest arises in registering range images in real time. It becomes clear that in the near future, real time execution will be mandatory for registration techniques. In our proposed methods, most of the computations are done independently, which is well suited for parallel implementation, and we believe that there is great potential for speed-up.

Variable illumination Investigating the use of photometry for registering range images acquired under variable illumination is a challenging topic that would drastically increase the range of applicability for range image registration. Once the exact correspondences are known, it is possible to retrieve both the different illuminations and albedo, in the same manner as we did in Chapter 5. However, the estimation of the photometric attributes is likely to become unstable because of the increased number of unknowns. The objective would then be to define a robust and stable photometric metric, suitable for optimization.

Non Lambertian objects To deal with non Lambertian surfaces is of major importance since in general real objects present both diffuse and specular reflection components. When accounting for both reflection components, the image formation models become much more complex than for the case of diffuse reflection only. How to integrate such model to reliably evaluate the similarity between different range images would be a challenging objective that, if reached, would open new possibilities for automatically registering range images.

Simultaneous registration of multiple scans In this dissertation, we focused on aligning pairs of overlapping range images. However, with the recent development of real-time 3D sensor, registering multiple range images simultaneously becomes of great interest. The method proposed in Chapter 5 can be extended to more than two scans so that pair-wisely given transformations are simultaneously evaluated. The extension of our method in this direction will be interesting.

Bibliography

[1] B. Akinci, F. Boukamp, C. Gordon, D. Huber, C. Lyons, and K. Park. A formalism for utilization of sensor systems and integrated project models for active construction quality control. *In Automation in Construction*, 15:124–138, 2006. (Cited on page 3.)

[2] M. Audette, F. Ferrie, and T. Peters. An algorithmic overview of surface registration techniques for medical imaging. *In Medical image analysis*, 4:201–217, 2000. (Cited on page 15.)

[3] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *International Journal of Visual Computing*, 12(1):43–77, 1992. (Cited on page 44.)

[4] R. Basri and D. Jacobs. Lambertian reflectance and linear subspace. *IEEE Trans. on PAMI*, 25(2):218–233, 2003. (Cited on pages 9, 11, 20, 102 and 103.)

[5] H. Bay, T. Tuytelaars, and L. V. Gool. Surf: Speeded up robust features. *In Proc. of ECCV'06*, pages 404–417, 2006. (Cited on pages 8, 18, 24, 97 and 99.)

[6] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. on PAMI*, 24(4):509–522, 2002. (Cited on pages 16 and 23.)

[7] P. J. Besl and N. D. McKay. A method for registration of 3-D shapes.

IEEE Trans. on PAMI, 14(2):239–256, 1992. (Cited on pages 8, 16, 23 and 99.)

[8] G. Blais and M. Levine. Registering multiview range data to create 3d computer objects. *In IEEE transaction on PAMI*, 17(8):820–824, 1995. (Cited on page 16.)

[9] S. Boivin and A. Gagalowicz. Inverse rendering from a single image. *In Proc. of IS&T CGIV*, 2002. (Cited on page 11.)

[10] D. Breitenreicher and C. Shnörr. Intrinsic second-order geometric optimization for robust point set registration without correspondence. *In Proc. of EMMCVPR2009*, pages 274–287, 2009. (Cited on pages 17, 97 and 100.)

[11] P. Brivio, A. Ventura, A. Rampini, and R. Schettini. Automatic selection of control points from shadow structures. *In International Journal of Remote Sensing*, 13:1853–1860, 1992. (Cited on page 18.)

[12] L. Brown. A survey of image registration techniques. *In ACM Computer surveys*, 24:326–376, 1992. (Cited on page 15.)

[13] N. Brusco, M. Andreetto, A. Giorgi, and G. M. Cortelazzo. 3D registration by textured spin-images. *In Proc. of 3DIM'05*, pages 262–269, 2005. (Cited on pages 18 and 23.)

[14] F. Candocia. Jointly registering images in domain and range by piecewise linear comparametric analysis. *In IEEE Trans. on Image Processing*, 12(4):409–418, 2003. (Cited on page 19.)

[15] L. Cerman, A. Sugimoto, and I. Shimizu. 3D shape registration with estimating illumination and photometric properties of a convex object. *In Proc. of CVWW'07*, pages 76–81, 2007. (Cited on pages 20, 21, 23, 73 and 90.)

[16] Y. Chen and G. Medioni. Object modeling by registration of multiple range images. *In IEEE International Conference on Robotics and Automation*, pages 2724–2729, 1991. (Cited on page 16.)

[17] X. Dai and S. Khorram. Development of a feature-based approach to automated image registration for multitemporal and multisensor remotely sensed imagery. *In Proc. of IGARSS'97*, pages 243–245, 1997. (Cited on page 18.)

[18] P. Debevec. *Light Probe Image Gallery*, 2004. `http://ict.debevec.org/~debevec/Probes/`. (Cited on page 130.)

[19] L. Ding, A. Goshtaby, and M. Satter. Volume image registration by template matching. *In Image and Vision Computing*, 19:821–832, 2001. (Cited on page 15.)

[20] D. Eggert, A. Fitzgibbon, and R. Fisher. Simultaneous registration of multiple range views for use in reverse engineering, and location of 3d objects. *Technical report 804, Dept. of Artificial Intelligence, University of Edinburgh*, 1996. (Cited on page 17.)

[21] O. Enqvist, F. Jiang, and F. Kahl. A brute-force algorithm for reconstructing a scene from two projections. *In Proc. of CVPR'11*, 2011. (Cited on page 100.)

[22] O. Faugeras and Q.-T. Long. The geometry of multiple images. *In The MIT Press*, 2001. (Cited on page 5.)

[23] J. Flusser and T. Suk. A moment-based approach to registration of images with affine geometric distortion. *In Trans. on Geoscience and Remote Sensing*, 32:382–387, 1994. (Cited on page 18.)

[24] L. Foncesca and B. Manjunath. Registration techniques for multisensor remotely sensed imagery. *In Photogrammetric Engineering and Remote Sensing*, 62:1049–1056, 1996. (Cited on pages 15 and 17.)

[25] D. Forsyth and J. Ponce. Computer vision. a modern approach. 2003. (Cited on page 2.)

[26] B. Ghaffary and A. Sawchuk. A survey of new techniques for image registration and mapping. *In Proc. of SPIE: Applications of Digital Image Processing*, 432:222–239, 1983. (Cited on page 15.)

[27] G. Godin, D. Laurendeau, and R. Bergevin. A method for the registration of attributed range images. *In Proc. of 3DIM'01*, pages 179–186, 2001. (Cited on pages 18, 23 and 43.)

[28] A. Goshtasby and G. Stockman. Point pattern matching using convex hull edges. *In Trans. on System, Man and Cybernetics*, 15:631–637, 1985. (Cited on page 18.)

[29] A. Goshtasby, G. Stockman, and C. Page. A region-based approach to digital image registration with subpixel accuracy. *In Trans. on Geoscience and Remote Sensing*, 24:390–399, 1986. (Cited on page 18.)

[30] E. Gulch. Results of test on image matching of isprs wg. *In ISPRS Journal of Photogrammetry and Remote Sensing*, 46:1–18, 1991. (Cited on page 15.)

[31] B. Gutman, Y. Wang, T. Chan, P. M. Thompson, and A. W. Toga. Shape registration with spherical cross correlation. *In 2nd MICCAI Workshop on Mathematical Foundations of Computational Anatomy*, pages 56–67, 2008. (Cited on pages 17 and 100.)

[32] R. Hartley and A. Zisserman. Multiple view geometry in computer vision. second edition. *In Cambridge University Press*, 2003. (Cited on page 5.)

[33] Y. Heo, K. Lee, and S. Lee. Simultaneous color consistency and depth map estimation for radiometrically varying stereo images. *In Proc. of ICCV'09*, pages 1771–1778, 2009. (Cited on page 19.)

[34] J. Herman, D. Smeets, D. Vandermeulen, and P. Suetens. Robust point set registration using em-icp with information-theoretically optimal outlier handling. *In Proc. of CVPR'11*, pages 2465–2472, 2011. (Cited on page 16.)

[35] D. Hill, P. Batchelor, M. Holden, and D. Hawkes. Medical image registration. *In Physics in Medicine and Biology*, 46:R1–R45, 2001. (Cited on page 15.)

[36] B. Horn. Closed-form solution of absolute orientation using orthonormal matrices. *J. Opt. Soc. Amer. A*, 5(7):1127–1135, 1987. (Cited on page 125.)

[37] Y. Hsieh, D. McKeown, and F. Perlan. Performance evaluation of scene registration and stereo matching for cartographic feature extraction. *In Trans. on PAMI*, 14:214–237, 1992. (Cited on page 18.)

[38] L. Ibanez, W. Schroeder, L. Ng, j. Cates, and the Insight Software Consortium. The ITK software guide second edition. *The ITK Software Guide Second Edition*, 2005. (Cited on pages 27 and 29.)

[39] B. Jian and B. C. Vemuri. A robust algorithm for point set registration using mixture of gaussian. *In Proc. of ICCV'05*, 2:1246–1251, 2005. (Cited on pages 8, 17, 23 and 100.)

[40] H. Jin, T. Duchamp, H. Hoppe, J. McDonald, K. Pulli, and W. Stuetzle. Surface reconstruction from misregistered data. *In Proc. of SPIE*, 2573(7):324–328, 1995. (Cited on page 17.)

[41] A. E. Johnson and M. Hebert. Surface registration by matching oriented points. *In Proc. of 3DIM'97*, pages 121–128, 1997. (Cited on pages 16 and 23.)

[42] A. E. Johnson and S. B. Kang. Registration and integration of textured 3D data. *Image and vision computing*, 17(2):135–147, 1999. (Cited on pages 18, 23 and 99.)

[43] S. Khoualed, U. Castellani, and A. Bartoli. Semantic shape context for the registration of multiple partial 3d views. *In Proc. of BMVC'09*, 2009. (Cited on page 16.)

[44] A. Lee, W. Sweldens, P. Shroder, L. Cowsar, and D. Dobkin. Maps: multiresolution adaptive parameterization of surfaces. *In Proc. of SIGGRAPH'98*, pages 343–352, 1998. (Cited on pages 126 and 127.)

[45] H. Lester and S. Arridge. A survey of hierarchical non-linear medical image registration. *In Pattern Reconition*, 32:129–149, 2001. (Cited on page 15.)

[46] S. Lin, Y. Li, S. Kang, X. Tong, and H. Shum. Diffuse-specular separation and depth recovery from image sequences. *In Proc. of ECCV'02*, pages 210–224, 2002. (Cited on page 74.)

[47] I. Lisle and S. Huang. Algorithms for spherical harmonic lighting. *In Proc. of GRAPHITE'07*, pages 235–238, 2007. (Cited on page 11.)

[48] Y. Liu. Automatic range image registration in the markov chain. *IEEE Trans. on PAMI*, 32(1):12–29, 2010. (Cited on pages 16 and 23.)

[49] Y. Liu, h. Zhou, X. Su, M. Ni, and R. J.Lloyd. Transforming least squares to weighted least squares for accurate range image registration. *In Proc. of 3DPVT'06*, pages 232–239, 2006. (Cited on page 41.)

[50] D. G. Lowe. Object recognition from local scale-invariant features. *In Proc. of ICCV'99*, 2:1150–1157, 1999. (Cited on pages 8, 18, 24 and 99.)

[51] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry. An invitation to 3d vision. from images to geometric models. *In Imaging, Vision, and Graphics. Interdisciplinary applied Mathematics*, 26, 2004. (Cited on page 2.)

[52] T. Machida, H. Takemura, and N. Yokoya. Dense estimation of surface reflectance properties based on inverse global illumination rendering. *In Proc. of ICPR'04*, 2:895–898, 2004. (Cited on page 11.)

[53] J. Maintz and M. Viergever. A survey of medical image registration. *In Medical Image Analysis*, 2:1–36, 1998. (Cited on page 15.)

[54] T. McInerney and D. Terzepoulos. Deformable models in medical image analysis: a survey. *In Medical Image Analysis*, 1(2):91–108, 1996. (Cited on page 4.)

[55] F. Menna and S. Troisi. Low cost reverse engineering techniques for 3d modeling of propellers. *In Proc. of ISPRS'10*, pages 452–456, 2010. (Cited on page 3.)

[56] J. L. Moigne. First evaluation of automatic image registration methods. *In Proc. of IGARSS'98*, pages 315–317, 1998. (Cited on page 15.)

[57] S. Moss and E. Hancock. Multiple line-template matching with em algorithm. *In Pattern Recognition Letters*, 18:1283–1292, 1997. (Cited on page 18.)

[58] S. Nayar, X. Fang, and T. Boult. Separation of reflection components using color and polarization. *Int'l J. Computer Vision*, 21(3), 1996. (Cited on page 74.)

[59] S. Nayar, K. Ikeuchi, and T. Kanade. Surface reflection: Physical and geometrical perspectives. *Trans on PAMI'91*, 13(7):611–634, 1991. (Cited on page 86.)

[60] S. Nayar and M. Oren. Generalization of the lambertian model and implications for machine vision. *In International Journal on Computer Vision*, 14(3):227–251, 1995. (Cited on pages 11 and 20.)

[61] P. Neugebauer. Geometrical cloning of 3d objects via simultaneous registration of multiple range images. *In International Conference on Shape Modeling and Application*, pages 130–139, 1997. (Cited on page 16.)

[62] V.-D. Nguyen, V. Nzomigni, and C. V. Stewart. Fast and robust registration of 3D surfaces using low curvature patches. *In Proc. of 3DIM'99*, pages 201–208, 1999. (Cited on pages 16 and 23.)

[63] F. Nicodemus. Directional reflectance and emissivity of an opaque surface. *In Applied Optics*, 4(7):767–775, 1965. (Cited on pages 11 and 19.)

[64] D. Nishino. Directional statistics brdf model. *In Proc. of ICCV'09*, pages 476–483, 2009. (Cited on page 9.)

[65] K. Nishino and K. Ikeuchi. Robust simultaneous registration of multiple range images. *In Proc. of ACCV'02*, pages 454–461, 2002. (Cited on page 17.)

[66] Okatani, I.S., and A. Sugimoto. Registration of range images that preserves local surface structures and color. *In Proc. of 3DPVT'04*, pages 786–796, 2004. (Cited on pages 18, 23, 97 and 99.)

[67] M. Ozuysal, M. Calonder, V. Lepetit, and P. Fua. Fast keypoint recognition using random ferns. *In IEEE Trans. on PAMI*, 32(3):448–461, 2010. (Cited on page 19.)

[68] C. Papazov and D. Burschka. Stochastic optimization for rgid point set registration. *In Proc. of ISVC'09*, pages 1043–1054, 2009. (Cited on pages 17, 97 and 100.)

[69] M. Pharr and G. Humphreys. Physically based rendering, second edition: From theory to implementation. *Morgan Kauffmann*, 2010. (Cited on page 20.)

[70] B. Phong. Illumination for computer generated pictures. *In Communications of ACM*, 18(6):311–317, 1975. (Cited on pages 11 and 20.)

[71] K. Pulli. Multiview registration for large data set. *In Proc. of 3DIM'99*, pages 160–168, 1999. (Cited on page 17.)

[72] K. Pulli, S. Piiroinen, T. Duchamp, and W. Stuetzle. Projective surface matching of colored 3D scans. *In Proc. of 3DIM'05*, pages 531–538, 2005. (Cited on pages 18 and 23.)

[73] P. Rademacher. Range image registration via consistency of empty space. *http://www.cs.unc.edu/ibr/projects/emptyspace/*, 1999. (Cited on page 16.)

[74] R. Ramamoorthi. Modeling illumination variation with spherical harmonics. *Face Processing: Advanced Modeling Methods*, pages 385–424, 2006. (Cited on page 102.)

[75] R. Ramamoorthi and P. Hanrahan. An efficient representation for irradiance environment maps. *In Proc. of SIGGRAPH'01*, pages 497–500, 2001. (Cited on pages 9, 11 and 20.)

[76] R.Cook and K. Torrence. A reflectance model for computer graphics. *In Proc. of SIGGRAPH'81*, 15(3):301–316, 1981. (Cited on pages 11 and 20.)

[77] M. Rutishauser, M. Stricke, and M. Trobina. Merging range images of arbitrarily shaped objects. *In Proc. CVPR*, pages 573–580, 1994. (Cited on page 16.)

[78] J. Salvi, C. Matabosch, D. Fofi, and J. Forest. A review of recent image registration methods with accuracy evaluation. *In Image and Vision Computing*, 25(5):578–596, 2007. (Cited on page 15.)

[79] J. Salvi, C. Matabosch, D. Fofi, and J. Forest. A review of recent range image registratio methods withaccuracy evaluation. *Image and Vision Computing*, 25(5):578–596, 2007. (Cited on page 5.)

[80] I. Sato. Appearance sampling of real objects for variable illumination. *In International Journal of Computer Vision*, 75(1):29–48, 2007. (Cited on pages 11 and 20.)

[81] Y. Sato and K. Ikeuchi. Temporal-color space analysis of reflection. *J. Optics Soc. Am. A*, 11, 1994. (Cited on page 74.)

[82] H. Sawhney and R. Kumar. True multi-image alignment and its applications to mosaicing and lens distortion correction. *In IEEE Trans. on PAMI*, 21:235–243, 1999. (Cited on page 18.)

[83] S. Se and P. Jasiobedzki. Stereo-vision based 3d modeling for unmanned ground vehicles. *In Proc. of SPIE'07*, 2007. (Cited on page 4.)

[84] S. Se, P. Jasiobedzki, and R. Wildes. Stereo-vision based 3d modeling of space structures. *In Proc. of SPIE'07*, 2007. (Cited on page 4.)

[85] I. Sebe, J. Hu, S. You, and U. Neumann. 3d video surveillance with augmented virtual environments. *In Proc. of IWVS'03*, 2003. (Cited on page 3.)

[86] J. Seo, G. Sharp, and S. Lee. Range data registration using photometric features. *In Proc. of CVPR'05*, 2:1140–1145, 2005. (Cited on pages 19 and 99.)

[87] R. Sharma and M. Pavel. Multisensor image registration. *In Proc. of The Society for Information Display*, pages 951–954, 1997. (Cited on page 18.)

[88] T. Tachikawa, S. Hiura, and K. Sato. Robust estimation of light directions and diffuse reflectance of known shape object. *In Proc. of Vision, Modeling and Visualization Workshop*, pages 37–44, 2009. (Cited on pages 103 and 104.)

[89] R. Tan and K. Ikeuchi. Separating reflection components of textured surfaces using a single image. *IEEE Trans. on PAMI'05*, 27(2):178–193, 2005. (Cited on pages 74, 76, 80 and 81.)

[90] R. Tan, K. Nishino, and K. Ikeuchi. Color constancy through inverse intensity chromaticity space. *J. Optics Soc. Am. A*, 21(3):321–334, 2004. (Cited on page 82.)

[91] D. Thomas and A. Sugimoto. Robust range image registration using local distribution of albedo. *In Proc. of 3DIM'09*, pages 1654–1661, 2009. (Cited on pages 12 and 22.)

[92] D. Thomas and A. Sugimoto. Estimating albedo of specular objects. *In Proc. of MIRU'10*, 2010. (Cited on pages 13 and 74.)

[93] D. Thomas and A. Sugimoto. Range image registration of specular objects. *In Proc. of CVWW'10*, 2010. (Cited on pages 13 and 74.)

[94] D. Thomas and A. Sugimoto. Range image registration of specular objects under complex illumination. *In Proc. of 3DPVT'10*, 2010. (Cited on pages 13 and 74.)

[95] D. Thomas and A. Sugimoto. Robustly registering range images using local distribution of albedo. *In Computer Vision and Image Understanding*, 115:649–667, 2011. (Cited on pages 12 and 22.)

[96] D. Thomas and A. Sugimoto. Illumination-free photometric metric for range image registration. *In Proc. of WACV'12*, 2012. (Cited on pages 14 and 98.)

[97] E. Tola, V. Lepetit, and P. Fua. Daisy: an efficient dense descriptor applied to wide baseline stereo. *IEEE Trans. on PAMI*, 2009. (Cited on pages 18 and 24.)

[98] A. Torsello, E. Rodola, and A. Albarelli. Multiview registration via graph diffusion of dual quaternions. *In Proc. of CVPR'11*, pages 2441–2448, 2011. (Cited on page 17.)

[99] E. Trucco and A. Verri. Introductory techniques for 3d computer vision. 1998. (Cited on page 2.)

[100] P. van den Elsen, E.-J. Pol, and M. Viergever. Medical image matching-a review with classification. *In IEEE Engineering in Medicine and biology*, 12:26–39, 1993. (Cited on page 15.)

[101] A. Vedaldi. *SIFT code for Matlab*, 2006. `http://www.vlfeat.org/~vedaldi/code/sift.html`. (Cited on page 130.)

[102] V.Hajnal, D. Hill, and D. hawkes. Medical image registration. *In CRC Press, Baton Rouge*, 2001. (Cited on page 18.)

[103] P. Viola and W. Wells. Alignment by maximization of mutual information. *In International of Computer Vision*, 24:137–154, 1997. (Cited on page 18.)

[104] N. Vujovic and D. Brzakovic. Establishing the correspondence between control points in pairs of mammographic images. *In Trans. on Image Processing*, 6:1388–1399, 1997. (Cited on page 18.)

[105] S. Weik. Registration of 3-D partial surface models using luminance and depth information. *In Proc. of 3DIM'97*, pages 93–101, 1997. (Cited on pages 18 and 23.)

[106] J. West, J. M. Fitzpatrick, M. Wang, B. D. Jr, C. Maurer, R. Kessler, and R. Maciunas. Retrospective intermodality registration techniques for images of head: surface-based versus volume-based. *In IEEE Transaction on Medical Imaging*, 18:144–150, 1999. (Cited on page 15.)

[107] Z. Xie, S. Xu, and X. Li. A high-accuracy method for fine registration of overlapping point of clouds. *Image and Vision Computing*, 28(4):563–570, 2010. (Cited on pages 16 and 23.)

[108] H. Zhang, O.Hall-Holt, and A. Kaufman. Range image registration via probability fiels. *In Proc. International conference on Computer Graphics*, pages 546–552, 2004. (Cited on page 16.)

[109] Z. Zhang. Iterative point matching for registration of free form curves and surfaces. *Int'l J. Computer Vision*, 13:119–152, 1994. (Cited on pages 16 and 23.)

[110] K. Zhou, H. Bao, and J. Shi. 3D surface filtering using spherical harmonics. *Computer Aided Design*, 36:363–375, 2004. (Cited on pages 123 and 125.)

[111] B. Zitova and J. Flusser. Image registration methods: a survey. *In Image and Vision Computing*, 21:977–1000, 2003. (Cited on page 15.)