Vol.2012-CVIM-182 No.34
2012/5/24

光学解析に基づく距離画像の位置合わせ

トマ　ディエゴ[1,a)]　杉本晃宏[1,b)]

概要：奥行きと色を同時に撮影できる距離画像センサを使って異なる位置から取得した２枚の距離画像の位置合わせを行う手法を提案する．撮影物体の幾何特徴を利用して位置合わせを行う従来手法に対して，光学特徴を利用した位置合わせ手法を提案する．物体表面の色の見え方は，照明条件，視点，物体表面の向き，物体表面の反射特性によって決まるが，これらの関係に基づいて光学特徴を抽出し，それを用いた類似度を定義することによって位置合わせのための変換を評価し，頑健で正確な位置合わせを実現する手法を提案する．画像撮影の照明条件における仮定を，大まかに分かっている単一点光源の場合，未知の少数の点光源の場合，未知の複雑照明の場合，と段階的に緩和し，より一般的に適用できるように手法を順次，補強・拡張する．光学解析を位置合わせ問題に取り入れた提案手法によって，これまでは実現できなかった未知複雑照明下での位置合わせが可能となった．

キーワード：距離画像，位置合わせ，光学解析，アルベド

Range Image Registration Based on Photometry

Diego Thomas[1,a)]　Akihiro Sugimoto[1,b)]

Abstract: In this paper we address the use of photometry for accurately registering pairs of range images devoid of salient geometric features. Our contribution is three fold: first, we propose a robust local descriptor that overcomes the drawbacks of current methods using albedo for Lambertian objects under simple illumination; second, we propose 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 propose a photometric metric for registering Lambertian range images under unknown general illumination and prove its usefulness through a practical registration method. With these methods, we make advances in using photometry for registering pairs of overlapping range images and significantly enlarge the practicability and range of applications of range image registration.

Keywords: Range images, registration, photometry, albedo.

1. Introduction

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. The 3D modeling process is summarized and illustrated in Fig. 1.

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 ([3], [5], [14], [20]).

Aligning overlapping range images using geometric features is the most popular approach to 3D registration and has been extensively studied over the past decades. Popular geometric features are for example, the position of the point, the normal at the surface or the curvature. In spite of the advantages of using geometric features, several limitation cases have been reported. In particular, using geometric features alone for registration inevitably fails if

[1] 国立情報学研究所
National Institute of Informatics, Tokyo 101–8430, Japan
[a)] diego_thomas@nii.ac.jp
[b)] sugimoto@nii.ac.jp

the captured range images are devoid of salient geometric features.

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. However, the appearance of the object changes depending on its pose, viewpoint or illumination. As a consequence, the direct use of textural information is, in general, unreliable.

The recent advances and breakthroughs in understanding and modeling image formation ([2], [22], [26]) bring new possibilities for 3D registration using photometry. By photometry, we denote the relationship between geometry, reflectance properties and incident illumination. By contrast with the textural features, we denote by photometric features the surface intrinsic features that define its reflectance properties (such as the albedo for example). 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. Existing methods have thus a limited range of applications.

In this work, we further investigate 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.

2. Local descriptor using albedo distribution

The most common approach to registering range images is (1) to find correspondences in points between two overlapping range images and then (2) accordingly estimate the transformation in aligning the two range images. The most well-known approach to fine registration is the iterative closest point (ICP) [5], [38]., which iterates the two above mentioned steps until convergence is reached.

Several methods for registering range images can be found in the literature that use geometric features for computing correspondences in points ([4], [14], [15], [18], [21], [37]). 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, textural features, such as color ([7], [10], [16]), chromaticity ([23]) or intensity ([3], [20], [36]), or photometric features such as the albedo, in addition to geometric features are required to compute correspondences in points.

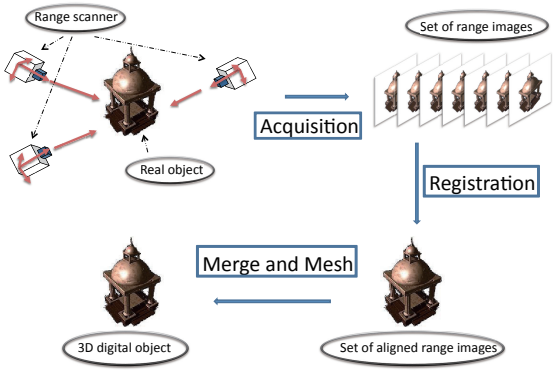


図 1: The 3D modeling process

In particular 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.* [8] 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.

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 [29], [33]

2.1 Proposed method

Our proposed method carries out the registration process 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, which are defined by adaptive regions representing the local distribution of albedo on the surfaces of objects. Incorrect matches are then eliminated using the rigidity constraint of surfaces. The transformation is then estimated by minimizing the distances between matched points. Fig.2 has a flowchart of our proposed approach.

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.

2.1.1 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. 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.

2.1.1.1 Level-set method

A region is defined by a contour that we define with a level-set method (fast marching algorithm [13]). A contour is defined as the zero level-set of a higher dimensional function called the level-set function, $\psi(X,t)$. 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:

$$\frac{d}{dt}\psi = -P(x)\|\nabla\psi\|, \tag{1}$$

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. Namely, we define a (propagation) speed image by applying a sigmoid filter to the gradient magnitude image. We define the zero level-set for a given point as the contour propagated at a certain time T (for example, $T = 0.2$ seconds) from the point. As a result, a reliable region is adaptively generated depending on 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. The local distribution of albedo of 3D points inside this region then defines an attribute for each point. The attributes obtained in this way enhance robustness in evaluating similarity to find correspondences.

2.1.2 Evaluation of similarities

We define a similarity metric between two points using their attribute to find correspondences across two range images (*Range Image 1* and *Range Image 2*).

Letting p be a point in *Range Image 1* and q be a point in *Range Image 2*, 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). The corresponding point $n(m)_q$ is defined by

$$\arg\min_{x \in R(q)} (\|T(\overrightarrow{pm}) - \overrightarrow{qx}\|_2). \tag{2}$$

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

$$\begin{aligned} \omega_{(m,q)} &= 0 \ \text{ if } \|T(\overrightarrow{pm}) - \overrightarrow{qn(m)_q}\|_2 > thresh, \\ \omega_{(m,q)} &= 1 \ \text{ if } \|T(\overrightarrow{pm}) - \overrightarrow{qn(m)_q}\|_2 \le thresh, \end{aligned} \tag{3}$$

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:

$$\begin{aligned} L(p,q) = \frac{size(R(p)) + size(R(q))}{(\sum_{m\in R(p)} \omega_{(m,q)} + \sum_{m\in R(q)} \omega_{(m,p)})^2} \times \\ \Big\{ \sum_{m\in R(p)} \omega_{(m,q)} \|\overrightarrow{alb(m)} - \overrightarrow{alb(n(m)_q)}\|_2^2 + \\ \sum_{m\in R(q)} \omega_{(m,p)} \|\overrightarrow{alb(m)} - \overrightarrow{alb(n(m)_p)}\|_2^2 \Big\}, \end{aligned} \tag{4}$$

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 chan-

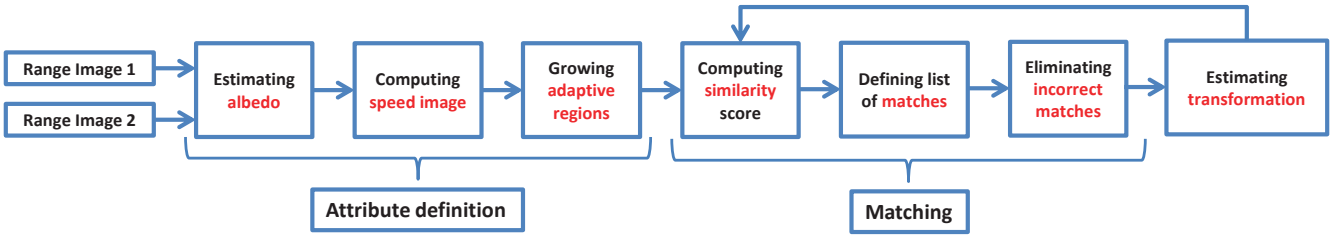


図 2: The flowchart of our proposed method.

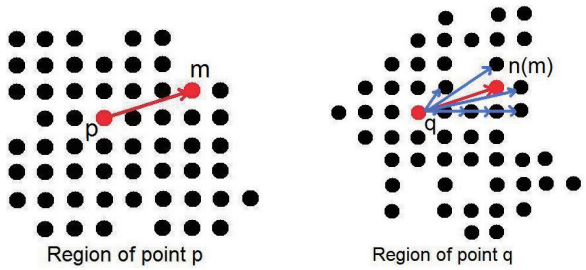


図 3: Searching for corresponding point of m.

nel:

$$\overrightarrow{alb(m)} = \frac{\overrightarrow{c(m)}}{\overrightarrow{norm(m)}^{\top} M \overrightarrow{norm(m)}}, \qquad (5)$$

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.

2.1.3 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.

2.1.3.1 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 $\|\overrightarrow{qq'}\|_2$:

$$d = |\|\overrightarrow{pp'}\|_2 - \|\overrightarrow{qq'}\|_2|. \qquad (6)$$

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. 4).

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.

2.1.3.2 Estimation of rigid transformation

We use the WLS method [19] 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

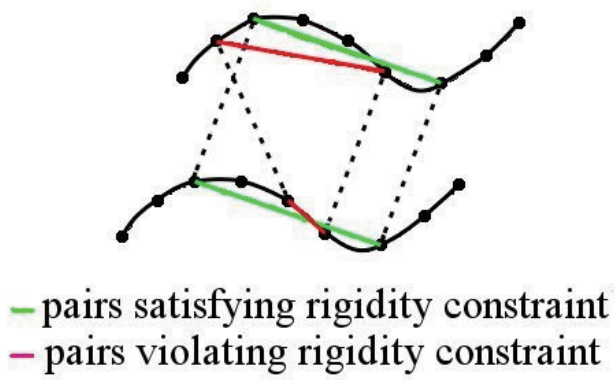


図 4: Principle of rigidity constraint.

of each pair. The resulting transformation obtained with this method is more accurate than that with the standard least squares method.

2.2 Experiments

We evaluated our method using synthetic and real data and compared it with ICPA (which is the most recent registration method that uses albedo).

To evaluate the registration results, we use an angular measure of errors for rotation like Barron *et al.* [1] 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:

$$err = \frac{\Theta d + ||T_g - T_e||}{res}, \qquad (7)$$

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 in terms of neighboring points. It is thus an objective and informative criterion to evaluate the accuracy of the different methods of registration.

2.2.1 Evaluation with synthetic data

We conducted experiments with synthetic data to test the robustness of the proposed method against data noise, changes in illumination and changes in initial relative poses. The synthetic data were obtained with a 3D modeler software (3D Studio Max) (see Fig. 5). The exact albedo image and the exact illumination (which represents a single distant light-source point), modeled with a direction and a color vector, were known. We estimated the albedo using an approximation of the exact illumination to test the robustness of our proposed method.

Figure 6 plots the quantitative evaluation of registration results in terms of averages and variances in error in results obtained with different estimates of illumination. In this experiment, we randomly rotated the direction vector of exact illumination to estimate albedo values. For all values of the noise variance, we applied our method 30 times under the same initial conditions. The results obtained with ICPA have also been shown for comparison. As expected, we find that our method is in average more accurate and more stable than ICPA in estimating the correct transformation. Note that, in addition to small errors, the variance of errors obtained with our method was also small, which proves the robustness and reliability of our proposed approach against noise in illumination. Moreover, for exact estimates of illumination, our proposed method achieves registration that is as accurate as that with ICPA.

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

図 5: Input synthetic data.

Figure 7 plots quantitative results of registration obtained with our proposed method under various noise in intensities and normals where the ground truth illumination was used. We applied Gaussian noise with a variance of several percent to the average of the image intensities to generate noise in the intensity, and randomly perturbed each normal vectors using a uniform distribution to generate noise in normals. For each different noise in intensity and normals, we applied our method 30 times under the same initial conditions. We observe that even with noise in intensity with a variance of 9.5%, and noise in normals of angle 10 degrees the largest errors are under the resolution of the range sensor. We find that our method is stable against both geometric and photometric noise from these results. Note that in these experiments, the standard ICP using both geometry and chromaticity did not work.

2.2.2 Evaluation with real data

We conducted experiments using real objects to test the effectiveness of the proposed method. In addition to ICPA, we also compared the proposed method with ICP using both chromaticity and geometric features (which we call ICP-CG) and SURF [3]. Note that we used the source code for SURF provided at http://www.vision.ee.ethz.ch/∼surf/ without modifications.

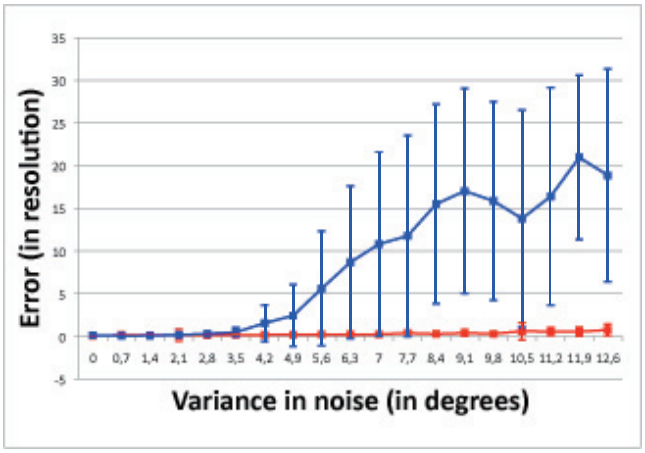
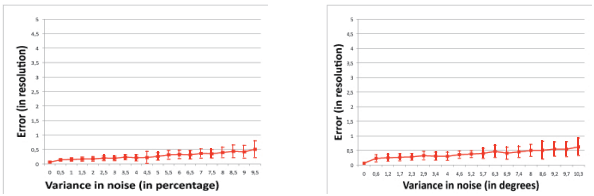


図 6: Results for various different illuminations. (red: our method, blue: ICPA)



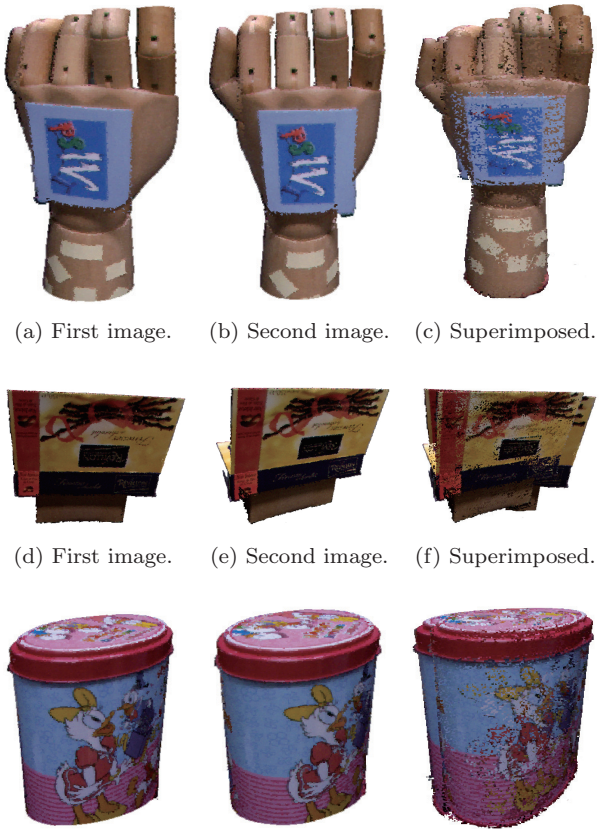
(a) (b)

図 7: Results for various noise in (a) intensities and (b) normals with our method.

We employed a Konica Minolta Vivid 910 range scanner, which captures the 3D shapes and textures of objects. 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 (7).

Figure 8 shows the three sets of range images that we used in these experiments. The quantitative results for these experiments are listed in Table 1. As we can see, our proposed method was the only one to always achieve accurate registration of the two input range images.

In this section, 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 section, we will investigate how to recover albedo from two specular range images illuminated by a few unknown point light sources.



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

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

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

図 8: Initial state for data *hand* (top), *box* (middle), and *candy* (bottom).

表 1: Quantitative evaluation of registrations, using data *hand*, *box*, and *candy*.

hand				
	Our method	ICPA	ICP-CG	SURF
Error	0.82	1.14	1.20	16.02
box				
	Our method	ICPA	ICP-CG	SURF
Error	1.27	9.13	6.56	2.96
candy				
	Our method	ICPA	ICP-CG	SURF
Error	0.81	5.14	4.03	23.25

3. Estimating Albedo of Specular Objects

In the method we proposed in the section 2, as well as in previous work that makes use of albedo ([8]), the specular reflections at the surface of an object are not considered. In this section, we propose a method for extracting albedo from two range images of a specular object under a few

fixed and unknown point light sources. 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 our previously proposed registration algorithm to show the usefulness of the proposed method. Fig. 9 illustrates the flowchart of our proposed method. Our intensive experiments show the effectiveness of the proposed method. We note that a part of this work appeared in [30], [31], [32].

3.1 Local computation of albedo

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 [28]. However, 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. Nevertheless, 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.

3.1.1 Detection of specular highlights

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 first identify possibles specular highlights as the local peaks of intensities at the surface. However, some of the detected highlights may just be high intensity texture regions, which would cause inaccurate estimation of incident illumination directions. We employ illumination consistency (i.e. the fact that the light source directions producing corresponding specular highlights are the same in two range images) to discriminate between specular highlights and high intensity texture regions.

Normals at the surface are available for two range images. We can thus estimate the incident illumination direction that can produce such highlight by back-tracing the viewing direction to the center of the specular highlight region in the mirror-like reflection direction.

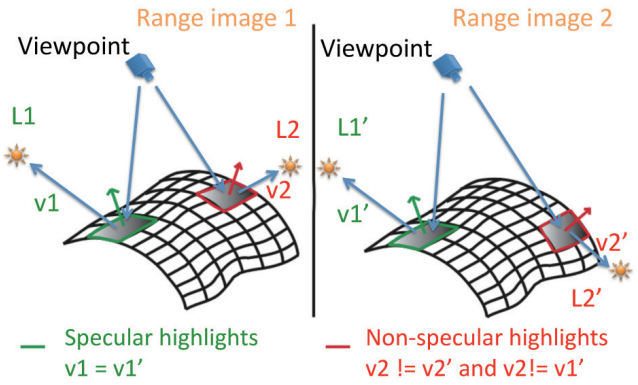


図 10: Illumination consistency constraint.

The highlight regions are then clustered into groups that are produced by similar light sources (i.e. light sources with similar incident direction) and the high intensity texture regions are identified and eliminated using the illumination consistency constraint. 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.

Fig. 10 illustrates the illumination consistency constraint under a fixed viewpoint and fixed illumination condition. We notice that we focus on situations when each light source illuminating the object produces specular highlights at the surface and 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.10. 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.

3.1.2 Detection of non-ambiguous regions

For each specular highlight, we have estimated its mostly dominant light source direction. If the incident

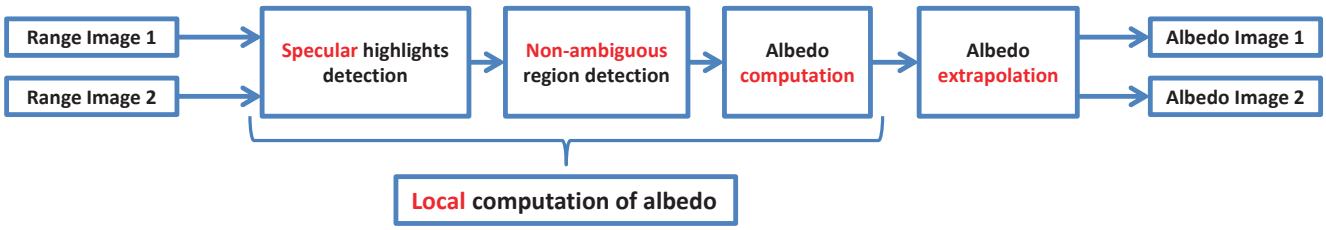


図 9: Basic flow of the proposed method.

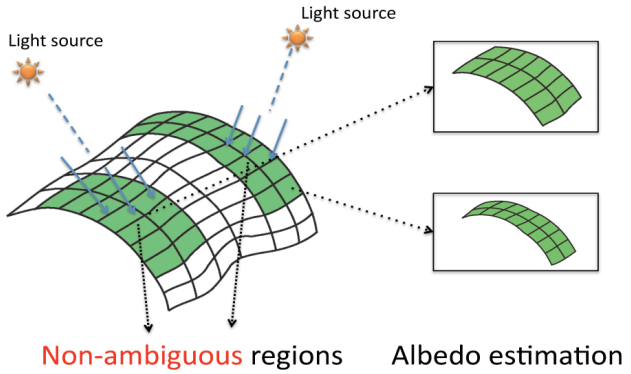


図 11: Definition of non-ambiguous regions.

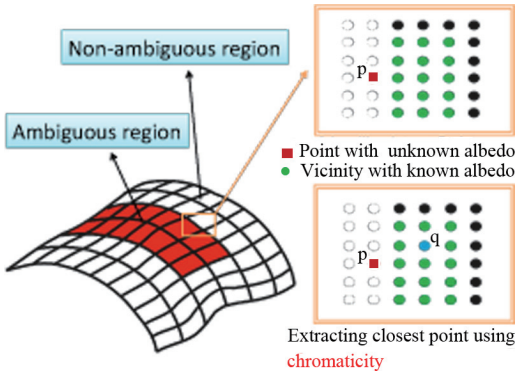


図 12: Albedo extrapolation.

illumination of a region is a single distant light source, we can use the method [28]. We can not, however, directly apply the method [28] to the whole surface, because the illumination environment can be composed of multiple light sources. In fact, the method [28] 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.

As a consequence, we obtain non-ambiguous regions in two range images in which we can reliably and adaptively separate reflection components using a single distant light source, as proposed in [28]. Fig. 11 illustrates the detection of non-ambiguous regions at the surface in the presence of two point light sources.

3.2 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. Fig. 12 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 images.

3.3 Experiments

In order to show the usefulness of our method, we use

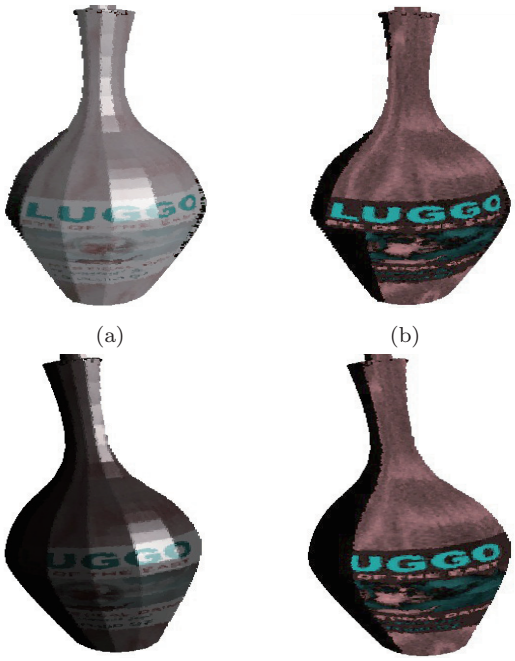


図 13: The input synthetic data and estimated albedo images. (a) input image 1 and (b) its estimated albedo image. (c) input image 2 and (d) its estimated albedo image.

our estimated albedo images as an input of the range image registration method we proposed in Section 2.

Figure 13 shows results obtained with the synthetic range images presented in Fig. 5 on which specular reflection is simulated. In this example, we obtained and error[*1] of 0.2 times the resolution of the image after registration. 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, which allowed us to obtain registration accuracy of the same precision of the acquisition device accuracy.

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 obtained to range images of a sphere with specular reflection components under fixed and uncontrolled illumination (Fig. 14). Fig. 14 shows estimated albedo images and the quantitative results of the registration by the three methods are shown in Table 2.

As we can see in Table 2, 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

[*1] We used equation 7 to compute the error.

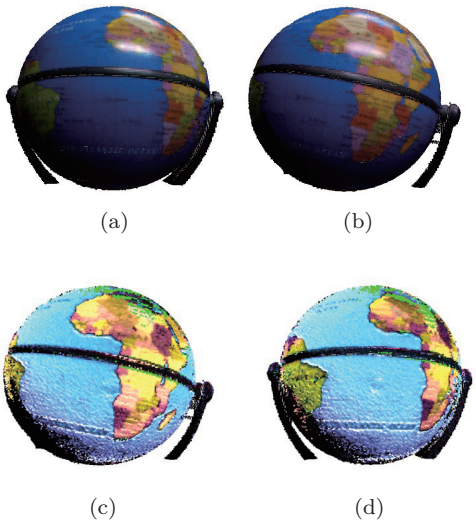


図 14: The data *globe*. Input image 1 (a) and 2 (b). Albedo of image 1 (c) and 2 (d). Points with unknown albedo are displayed in vivid green.

表 2: Results obtained for the data *globe*.

Error	Rotation	Translation
Ground truth		
0.00	$(22.5, 0.02, 0.94, 0.33)$	$(9.0, 0.1, -1.5)$
Proposed method		
0.54	$(22.3, 0.03, 0.94, 0.33)$	$(9.0, 0.2, -1.5)$
Diffuse reflection model		
0.90	$(22.4, 0.02, 0.95, 0.32)$	$(9.0, 0.2, -1.4)$
Chromaticity		
1.61	$(22.1, 0.04, 0.92, 0.30)$	$(8.9, -0.1, -1.5)$

accuracy of matching. Similarly, the diffuse approximation performed worse than our proposed method. This is because the specular reflections at the surface are ignored in the diffuse reflection model. On the contrary, our proposed method allowed us to obtain reliable albedo values for matching points.

In this section as well as in Section 2, 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. In the next section, we will investigate the derivation of a photometric metric, independent to illumination, which evaluates the quality of a given rigid transformation aligning a pair of overlapping range images.

4. Photometric metric under unknown lighting

The mechanisms behind image formation are complex and estimating albedo from a single range image under unknown and general illumination is impossible. As a

consequence, using photometric features to match points across range images as proposed in the previous sections has limited range of applications. However, matching points across the range images is not the only way we can take for registration. 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 ([6], [11], [14], [24]), and some use the hypothesis-and-test strategy such as RANSAC or brute-force search ([9]). 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.

In this section, we propose a novel photometric metric for evaluating the correctness of a given transformation and demonstrate its usefulness with a practical registration method. 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. We note that when the object's pose changes while the illumination stays fixed, its appearance in both shape and color changes.

We define our photometric metric using the spherical harmonics representation of image formation, without using any a-priori information on the incident lighting. This function evaluates the consistency of the relationship between geometry, texture and illumination. We then 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. Over the sphere, we generate rigid transformations and evaluate them to reach the best one for the final result. We note that a part of this work appeared in [34].

4.1 The photometric metric

We introduce our photometric metric under unknown lighting that does not compare features but compute photometric 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 15 illustrates the derivation of our proposed metric.

4.1.1 Review of spherical harmonics representation

The spherical harmonics have been shown to be a powerful tool to model image formation [25], 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 [2]. We briefly recall the principles of the spherical harmonics representation and refer to [2] for more details.

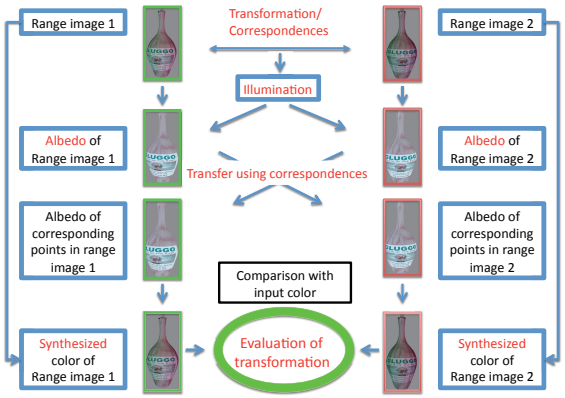


図 15: Procedural evaluation of a given transformation.

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.

$$E(\mathbf{x}) = \int_0^{2\pi} \int_0^{\pi} L(\theta, \phi) R(G_{\mathbf{x}}(\theta, \phi)) \sin(\theta) \mathrm{d}\theta \mathrm{d}\phi, \quad (8)$$

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

$$I(\mathbf{x}) \approx \rho(\mathbf{x}) \sum_{l=0}^{2} \sum_{m=-l}^{l} L_{l,m} R_{l,m}(\mathbf{x}), \quad (9)$$

where $L_{l,m}$ and $R_{l,m}(\mathbf{x})$ represent the spherical harmonic

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

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 $\mathbf{x}$ [2].

4.1.2 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. We use the spherical harmonics representation of image formation to derive a linear system from the point correspondences with the illumination as unknown. The estimated illumination then allows us to compute albedo values at points of the range images, which are transferred to their corresponding points. The transferred albedo values 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 denote a given transformation and $(\boldsymbol{x}_i, \Gamma(T(\boldsymbol{x}_i)))_{i\in[0,n-1]}$ and $(\Gamma(T^{-1}(\boldsymbol{y}_i)), \boldsymbol{y}_i)_{i\in[0,m-1]}$ denote the induced point correspondences, where $\boldsymbol{x}_i \in \boldsymbol{R}^3$ belongs to range image 1 (denoted as I^1), $\boldsymbol{y}_i \in \boldsymbol{R}^3$ belongs to range image 2 (denoted as I^2) and Γ denotes the point correspondences identification operator (see Section 4.1.3). 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 $\rho(\boldsymbol{x}_i) = \rho(\Gamma(T(\boldsymbol{x}_i)))$ (similarly $\rho(\boldsymbol{y}_i) = \rho(\Gamma(T^{-1}(\boldsymbol{y}_i)))$).

R is known and depends on only the surface normals. Therefore, using (9), we can derive a linear system $LM = 0$ with L as unknown, where $L = [L_{0,0}, L_{1.-1}, L_{1,0}, L_{1,1}, L_{2.-2}, L_{2,-1}, L_{2,0}, \ L_{2,1}, L_{2,2}]$ is a row vector in 9D and $M = [M_i]_{i\in[0,n+m-1]}$ is a $9\times(n+m)$ matrix, where n and m are the number of corresponding points from range images 1 and 2 respectively, and $M_i = [I^1(\boldsymbol{x}_i)R_{0,0}(\Gamma(T(\boldsymbol{x}_i))) - I^2(\Gamma(T(\boldsymbol{x}_i)))R_{0,0}(\boldsymbol{x}_i), \cdots, I^1(\boldsymbol{x}_i)R_{2,2}(\Gamma(T(\boldsymbol{x}_i))) - I^2(\Gamma(T(\boldsymbol{x}_i)))R_{2,2}(\boldsymbol{x}_i)]^\top$ if $i < n$, $M_i = [I^2(\boldsymbol{y}_i)R_{0,0}(\Gamma(T^{-1}(\boldsymbol{y}_i))) - I^1(\Gamma(T^{-1}(\boldsymbol{y}_i)))R_{0,0}(\boldsymbol{y}_i), \cdots, I^2(\boldsymbol{y}_i)R_{2,2}(\Gamma(T^{-1}(\boldsymbol{y}_i))) - I^1(\Gamma(T^{-1}(\boldsymbol{y}_i)))R_{2,2}(\boldsymbol{y}_i)]^\top$ if $n \le i < n+m$.

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.

$$\rho(\boldsymbol{x}) = \frac{1}{\lambda}\left(\frac{I(\boldsymbol{x})}{\sum_{l=0}^{2}\sum_{m=-l}^{l}\tilde{L}_{l,m}(T)R_{l,m}(\boldsymbol{x})}\right). \quad (10)$$

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 $\|\rho(\boldsymbol{x}_i) - \rho(\Gamma(T(\boldsymbol{x}_i)))\|$ 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 if T is accurate.

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

$$\tilde{I}^1_T(\boldsymbol{x}_i) = \rho(\Gamma(T(\boldsymbol{x}_i)))\sum_{l=0}^{2}\sum_{m=-l}^{l}\tilde{L}_{l,m}(T)R_{l,m}(\boldsymbol{x}_i). \quad (11)$$

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

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

$$Eval(T) = \frac{\sum_{i=0}^{n-1}\|I^1(\boldsymbol{x}_i) - \tilde{I}^1_T(\boldsymbol{x}_i)\| + \sum_{i=0}^{m-1}\|I^2(\boldsymbol{x}_i) - \tilde{I}^2_T(\boldsymbol{x}_i)\|}{n+m}. \quad (12)$$

We remark that the unknown scaling factor λ that arises when estimating albedo in (10) is no longer present in (12).

We notice that the shape of $Eval$ depends on T. Namely, $\tilde{I}^1_T$ and $\tilde{I}^2_T$ 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, $\tilde{I}^1_T$ and $\tilde{I}^2_T$ 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 the linear system defined above vary piece-wise continuously with sufficiently similar transformations and so does the estimated photometric properties as well as the synthesized images.

4.1.3 Point correspondences identification

Though it is a simple task, identifying the point correspondences from the given transformation T is the most time consuming one for our evaluation metric. It is thus of major importance to perform it as fast as possible. We use *projective data association* [27] to realize fast point correspondences estimation.

For two range images I^1 and I^2, their corresponding depth maps D_1 and D_2 with the associated intrinsic matrix K, and the given transformation T aligning I^1 to I^2,

the corresponding point $\Gamma(\boldsymbol{x}) \in I^2$ of a point $\boldsymbol{x} \in I^1$ is identified as follows: (1) $\boldsymbol{x}$ is transformed into the camera coordinate system of I^2 ($\boldsymbol{y} = T\boldsymbol{x}$); (2) the point $\boldsymbol{y}$ is perspective projected into image coordinates ($(i, j, 1) = K\boldsymbol{y}$); (3) $\Gamma(\boldsymbol{x})$, the closest point in I^2 of $\boldsymbol{x}$, is then identified as the point associated to the pixel (i, j) in D_2. Searching for the closest points from I^2 to I^1 is carried out similarly.

4.1.4 Stable points identification

Points in the overlapping area do not always correspond exactly. This is due to different digitization of the overlapping area depending on the object pose. As a consequence, even for the best transformation aligning the range images, there may be some point correspondences $(\boldsymbol{x}, \Gamma(T(\boldsymbol{x})))$ that do not satisfy the statement $\rho(\boldsymbol{x}) = \rho(\Gamma(T(\boldsymbol{x})))$. 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. To overcome this problem, we first identify a stable point, i.e., a point whose albedo and normal values 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 from two range images as a pre-processing step.

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 $\boldsymbol{x}$ is identified to be stable if

$$\begin{aligned} &\forall \boldsymbol{y} \text{ such that } \|\boldsymbol{y} - \boldsymbol{x}\| < \epsilon_s, \\ &\|c(\boldsymbol{x}) - c(\boldsymbol{y})\| < \epsilon_c \text{ and } \|\boldsymbol{n}(\boldsymbol{x}) - \boldsymbol{n}(\boldsymbol{y})\| < \epsilon_n, \end{aligned} \tag{13}$$

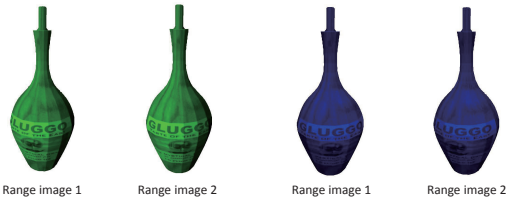
where $\boldsymbol{y}$ is a point in the range image concerned, c is chromaticity, $\boldsymbol{n}$ represents the surface normals, and ϵ_s, ϵ_c, and ϵ_n are three thresholds.

4.1.5 Analysis of the photometric metric

We analyze the behavior of our photometric metric under various parameters. Starting from the ground truth transformation that perfectly aligns two range images we generate several transformations by randomly perturbing the parameters of the ground truth transformation and plot the photometric re-projection error as the function of the registration error (equation (14)) for each generated transformation. 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.

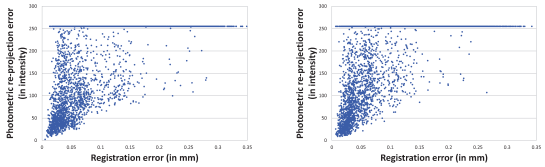
Figure 17 illustrates our photometric metric with respect to the different illumination conditions illustrated in Figure 16. For all situations the stable points were identified using the parameters $\epsilon_s = 0.03$, $\epsilon_c = 0.01$, and

(a) Situation 1 (b) Situation 2

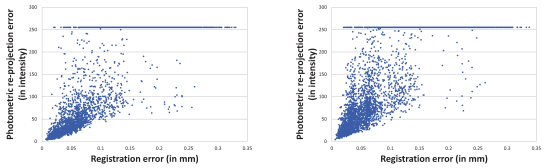


(c) Situation 3 (d) Situation 4

図 16: The four different illumination conditions.



(a) Situation 1 (b) Situation 2



(c) Situation 3 (d) Situation 4

図 17: Photometric re-projection error in function of the registration error for the different situations.

$\epsilon_n = 0.2$. From these different plots, we can see that (1) a transformation with a small photometric re-projection error is close to the ground truth transformation aligning the range images, (2) the minimal solution is almost insensitive to changes in illumination and (3) our proposed photometric metric will be difficult to minimize using traditional optimization methods.

4.2 Registration

Given two overlapping range images, we seek a rigid transformation that minimizes the photometric re-projection error (12). When minimizing the photometric re-projection error, we have to decide the strategy we use. As we discussed in Section 4.1.5, our proposed photometric metric is hard to minimize using traditional optimization methods. Therefore, we choose the hypothesis-and-test search. Fig. 18 illustrates the flowchart of our proposed registration method.

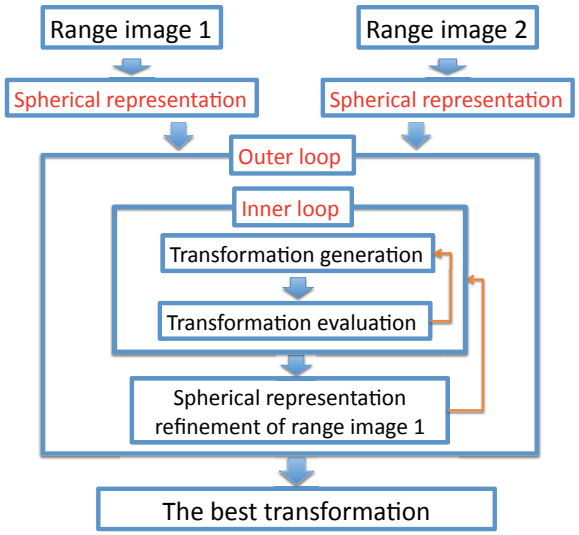


図 18: Flowchart of transformation search.

4.2.1 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. 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 [39]. Therefore, the local structure in the spherical domain does not change and the rigidity of point correspondences is kept. 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 [12] 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 [12] 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.

4.2.2 Spherical representation and refinement

Our spherical representation method is inspired by the method proposed in [39] 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.

4.2.2.1 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 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 [17], 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.

4.2.2.2 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.

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 4.1.3.

4.3 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 Section 2 using albedo with a given directional light source (Method 2), and the method proposed in Section 2 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 [35]). 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:

$$err(T_e) = \frac{\sum_{i=1}^{n} \|T_g(\mathbf{x}_i) - T_e(\mathbf{x}_i)\|_2}{n}, \qquad (14)$$

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.

4.3.1 Synthetic data

We used the synthetic data presented in Figure 5 where the exact albedo is known and we simulated lighting under Lambertian reflection. 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.

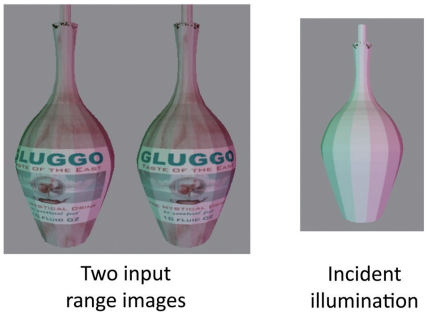


図 19: Input data.

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

図 20: Results obtained with the five methods.

The set-up is illustrated in Figure 19. 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 20 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. Results obtained using SIFT key-point detector and descriptor for estimating point correspondences are also illustrated in figure 20. Due to the projective deformations, changes in intensity and repetitive patterns, the SIFT-based method did not work in our situation. Our method is the only one that achieved accurate alignment.

4.3.2 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.

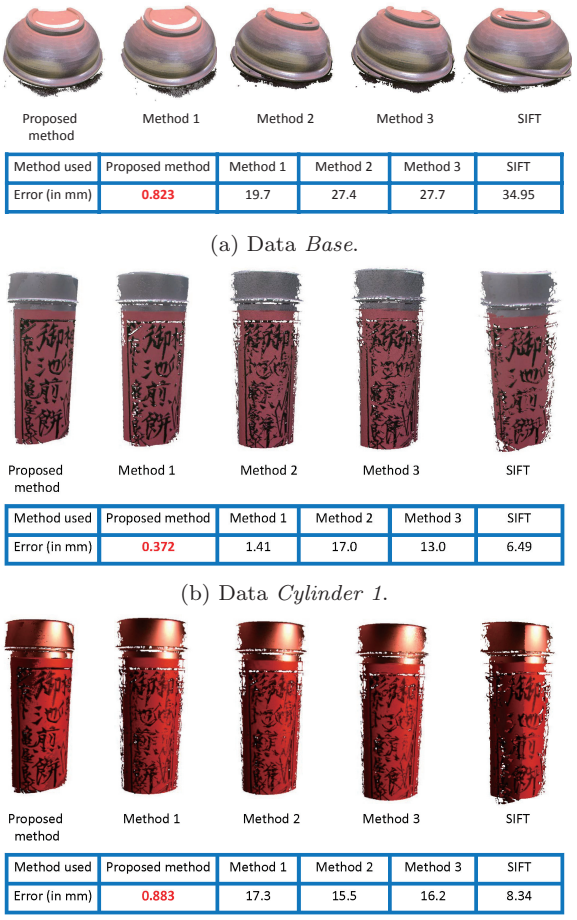
Figure 22 shows the results obtained with the different objects called *Base*, *Cylinder 1* and *Cylinder 2* (shown in figure 21). Figures 22 (b) and (c) show the result obtained with the data *Cylinder* for the same changes in pose but

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

(c) Data *Cylinder 1*. (d) Lighting.

(e) Data *Cylinder 2*. (f) Lighting.

図 21: Initial state for data *Base*, *Cylinder 1* and *Cylinder 2*.



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

(a) Data *Base*.

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

(b) Data *Cylinder 1*.

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

(c) Data *Cylinder 2*.

図 22: Registration results for data *Base*, *Cylinder 1* and *Cylinder 2*.

under different illumination conditions.

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 1* and *Cylinder 2* are challenging in that 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 because 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 either. This can be due to the repetitive patterns of the letters and deformations caused by perspective projection. On the other hand, our proposed method could obtain accurate registration results for all situations.

5. Conclusion

In this paper 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.

参考文献

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

[2] R. Basri and D. Jacobs. Lambertian reflectance and linear subspace. *IEEE Trans. on PAMI*, 25(2):218–233, 2003.

[3] H. Bay, T. Tuytelaars, and L. V. Gool. Surf: Speeded up robust features. *In Proc. of ECCV'06*, pages 404–417, 2006.

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

[5] P. J. Besl and N. D. McKay. A method for registration of 3-D shapes. *IEEE Trans. on PAMI*, 14(2):239–256, 1992.

[6] 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.

[7] 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.

[8] 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.

[9] O. Enqvist, F. Jiang, and F. Kahl. A brute-force algorithm for reconstructing a scene from two projections. *In Proc. of CVPR'11*, 2011.

[10] 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.

[11] 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.

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

[13] 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.

[14] 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.

[15] A. E. Johnson and M. Hebert. Surface registration by matching oriented points. *In Proc. of 3DIM'97*, pages 121–128, 1997.

[16] A. E. Johnson and S. B. Kang. Registration and integration of textured 3D data. *Image and vision computing*, 17(2):135–147, 1999.

[17] 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.

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

[19] 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.

[20] D. G. Lowe. Object recognition from local scale-invariant features. *In Proc. of ICCV'99*, 2:1150–1157, 1999.

[21] 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.

[22] D. Nishino. Directional statistics brdf model. *In Proc. of ICCV'09*, pages 476–483, 2009.

[23] 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.

[24] C. Papazov and D. Burschka. Stochastic optimization for rgid point set registration. *In Proc. of ISVC'09*, pages 1043–1054, 2009.

[25] R. Ramamoorthi. Modeling illumination variation with spherical harmonics. *Face Processing: Advanced Modeling Methods*, pages 385–424, 2006.

[26] R. Ramamoorthi and P. Hanrahan. An efficient representation for irradiance environment maps. *In Proc. of SIGGRAPH'01*, pages 497–500, 2001.

[27] S. Rusinkiewitcz and M. Leroy. Efficient variants of the icp algorithm. *Proc. 3D Digital Imaging and Modeling (3DIM '01)*, pages 145–152, 2001.

[28] 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.

[29] D. Thomas and A. Sugimoto. Robust range image registration using local distribution of albedo. *In Proc. of 3DIM'09*, pages 1654–1661, 2009.

[30] D. Thomas and A. Sugimoto. Estimating albedo of specular objects. *In Proc. of MIRU'10*, 2010.

[31] D. Thomas and A. Sugimoto. Range image registration of specular objects. *In Proc. of CVWW'10*, 2010.

[32] D. Thomas and A. Sugimoto. Range image registration of specular objects under complex illumination. *In Proc. of 3DPVT'10*, 2010.

[33] D. Thomas and A. Sugimoto. Robustly registering range images using local distribution of albedo. *In Computer Vision and Image Understanding*, 115:649–667, 2011.

[34] D. Thomas and A. Sugimoto. Illumination-free photometric metric for range image registration. *In Proc. of WACV'12*, 2012.

[35] A. Vedaldi. *SIFT code for Matlab*, 2006. `http://www.vlfeat.org/~vedaldi/code/sift.html`.

[36] S. Weik. Registration of 3-D partial surface models using luminance and depth information. *In Proc. of 3DIM'97*, pages 93–101, 1997.

[37] 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.

[38] Z. Zhang. Iterative point matching for registration of free form curves and surfaces. *Int'l J. Computer Vision*, 13:119–152, 1994.

[39] K. Zhou, H. Bao, and J. Shi. 3D surface filtering using spherical harmonics. *Computer Aided Design*, 36:363–375, 2004.