# Stereopsis Based Registration with Photometric Features

Daesik Kim, Jaewoong Kim, Jeonghyun Seo and Sukhan Lee

Intelligent Systems Research Center, Sungkwankwan University,
Suwon, Korea
{daesik80, lsh}@ece.skku.ac.kr

**Abstract.** In this paper, we present a robust registration method based on photometric features. The correct estimation of the transformation between the range data measured at distinct viewpoints is the key issue on registration. There are two main approaches: one is the range data oriented approach and the other is the image data oriented approach. In the range data oriented approach, the computational time is normally high, and we need relatively accurate range data in order to compute the correct transform. In case of the image based approach, only the interest points (photometric features) are generally used to estimate the transform so it is more efficient. In practice, the extracted features can be erroneous, and, of course, the range data may have errors. In this paper, we present the method to select the correct photometric features and range data, and also propose the technique to merge the points while re-estimating transformation matrix. The experiment shows the proposed approach is robust to the erroneous data.

**Keywords:** registration, photometric features, stereopsis

## 1 Introduction

Most registration methods try to optimize the distances between the two range data measured at distinct viewpoints, and some methods employ 3D geometric features derived from the range data or photometric features extracted from the 2D photometric images.

The iterative closest point (ICP) algorithm and its variants are a popular approach for range data registration [1][2]. It is an iterative procedure that minimizes the distances between the points in the two views. However, the ICP method has the disadvantage that a good initial alignment and the accurate range data are required to achieve a good registration. Photometric features based approach is the alternative to the ICP related methods. It is relatively efficient and robust.

In this paper, we introduce an approach that uses photometric features and its range data to registration. Even if the photometric features are distinctive, that can not be perfectly identified. In order to correctly match the features located in the different images, we applied the epipolar geometric constraints with RANSAC.

After selecting the proper features that satisfy the epipolar constraints, the estimation of the transformation is required. If the range data is not fully reliable, we can not directly estimate the transform since the range data are so erroneous. In order to detect the correct transformation matrix, we first randomly select four corresponding points, and the check the orthogonality of the computed rotation matrix and calculate the translation error. We performed this procedure iteratively and estimated the transformation matrix with the reliable data.

The rest of this paper is organized as follows. Section 2 addresses the problems of the photometric features based registration. Section 3 reviews the multiple view geometry, and Section 4 introduces the features we used for experiments. Section 5 and 6 presents the methods to estimate the fundamental and the transformation matrix, respectively. Section 7 shows the technique to merge and refine 3D range data and re-estimate the transformation matrix. Experimental results are shown in Section 8, and we conclude the paper in Section 9.

## 2 Problem Description

In case of the photometric feature based registration, there are two main problems. One is from mismatched features, and the other is from wrong estimated 3D data.

Features based matching is known as efficient, but these features, of course, contain some amount of errors. For example, if the same pattern is appears repeatedly in a scene, it is not easy to distinguish where one point in a scene matches the point in another scene. If we trust mismatched features, the corresponding 3D data are naturally incorrect. Thus, these mismatched points generate the wrong transformation matrix. The method to deal with this problem is described in Section 5.

The transformation matrix can be estimated with four corresponding 3D points. The $4 \times 4$ transformation matrix includes $3 \times 3$ rotation matrix and $3 \times 1$ translation matrix. An ideal $3 \times 3$ rotation matrix has three vectors which are orthogonal each other. However, in practice, most 3D points from stereopsis have some amount of errors. Therefore, the direct estimation of the transformation may be wrong. The method to deal with this problem is described in Section 6.

## 3 Multiple View Epipolar Geometry

The fundamental matrix represents the epipolar geometry algebraically and contains most of the information about the relative position and orientation between the two views. Suppose we have two images acquired by cameras with distinct view points, then the fundamental matrix $\mathbf{F}$ is the unique $3 \times 3$ rank 2 homogeneous matrix which satisfies

$$\mathbf{x}_2^T \mathbf{F} \mathbf{x}_1 = 0 \tag{1}$$

for all corresponding points $\mathbf{x}_1$ and $\mathbf{x}_2$ in two image planes.

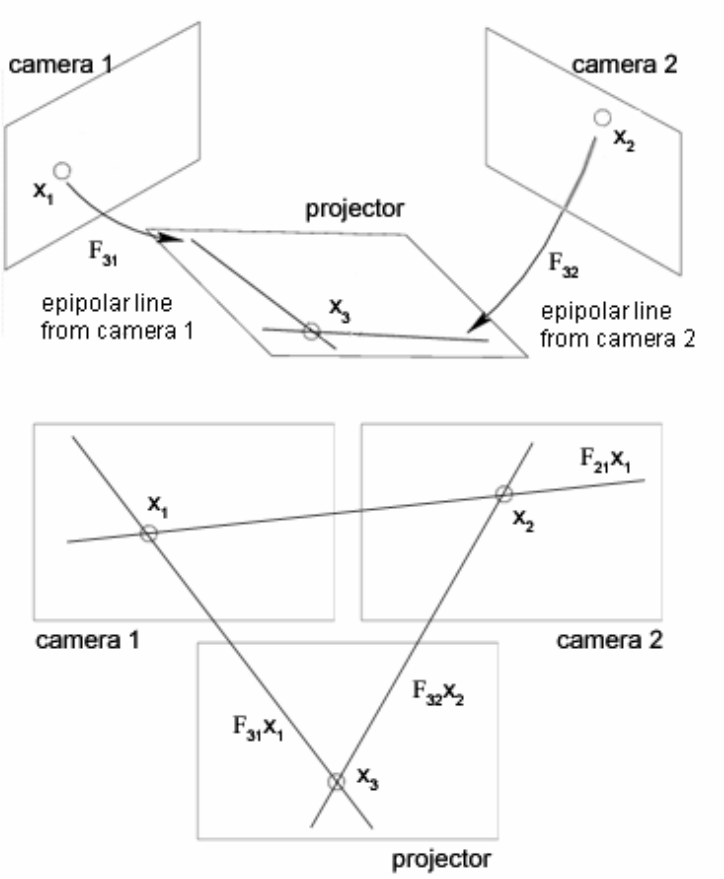


**Figure 1.** If the image point $\mathbf{x}_1$ of the camera 1 corresponds to the image point $\mathbf{x}_2$ in the camera 2, the point $\mathbf{x}_3$ of the camera 3 corresponding to these two points is the intersecting point of the epipolar lines $\mathbf{F}_{31}\mathbf{x}_1$ and $\mathbf{F}_{32}\mathbf{x}_1$. Similarily, the point $\mathbf{x}_2$ must lie on the intersecting point of the two epipolar lines corresponding to $\mathbf{x}_1$ and $\mathbf{x}_3$.

As shown in Figure 1, Suppose the image points $\mathbf{x}_1$ and $\mathbf{x}_2$ are a matched pair, then the corresponding point $\mathbf{x}_3$ lies on the epipolar line corresponding to $\mathbf{x}_1$ and the epipolar line corresponding to $\mathbf{x}_2$, and consequently lies on the point intersecting two epipolar lines $\mathbf{F}_{31}\mathbf{x}_1$ and $\mathbf{F}_{32}\mathbf{x}_2$. Namely,

$$\mathbf{x}_3 = (\mathbf{F}_{31}\mathbf{x}_1)\times(\mathbf{F}_{32}\mathbf{x}_2) \tag{2}$$

However, if the three camera centers are collinear and points are located on the plane defined by the centers, then three view epipolar constraints doesn't work. In this case, we can compute the geometrical constraints with trifocal tensor. For more details, see [3].

# 4 Feature Extraction

## 2.1 SIFT

"Scale Invariant Feature Transform (SIFT)" algorithm, proposed by Lowe [4] is one of the most effective, fast and reliable feature descriptors. It has several important properties: (1) Rotation invariance, (2) Robustness to illumination changes, (3) Affine compensation. In our experiment, we perform 2D local feature matching using SIFT.

## 2.2 Corners

Corners are also distinctive points. Different from SIFT, corners do not have the ID. That means there is no descriptor to identify itself. In order to identify each corner, we use the $5 \times 5$ or $7 \times 7$ neighbor intensity of the corners; then compared the correlation among the corners at different view point.

# 5 Estimation of Fundamental Matrix and Trifocal Tensor

RANSAC is known as one of the good methods to estimate the fundamental matrix. However, when the corresponding features are too erroneous, the matrix may be wrong estimated with only distinct two viewpoints.

## 5.1 Stereo Camera

As we mentioned above, the features can be defined whether these are correctly matched with three different view points. A stereo camera provides two images (left and right images), so we can obtain four images when the camera moves to another position. When the stereo camera is calibrated, the fundamental matrix between these left and right images is clearly given and we can trust this fundamental matrix without doubt. When the camera moves, we can compute the fundamental matrix $\mathbf{F}_{31}$ related to the image1 and image 3, and the fundamental matrix $\mathbf{F}_{32}$ related to the image 2 and image 3.

The reliability of the fundamental matrix $\mathbf{F}_{31}$ or $\mathbf{F}_{32}$ can be computed with equation (2). Thus, given more than three images, we can eliminate the mismatched features and refine the fundamental matrix with RANSAC. Of course, we can define the trifocal tensor with similar strategy.

### 5.2 Structured Light

In case of structured light system which uses a projector and a camera, we can obtain one image. Thus, we can extract the features from only this image. However, we can easily deduce the positions on the projector image which correspond the features of the camera image because the projector can illuminate the light. Because we can obtain the reliable fundamental matrix between the projector and the camera, the remaining procedure to compute the fundamental matrix $\mathbf{F}_{31}$ and $\mathbf{F}_{32}$ is same as the one that we did with stereo camera in precious section.

## 6 Estimation of Transformation Matrix

For a rigid 3D transformation, at least of three non-collinear point correspondences are required for a unique solution. With more correspondences, the accuracy of the transformation can be increased, however that does not be always guaranteed due to the erroneous data. Thus we estimate the transformation matrix with RANSAC.

Given four non-collinear 3D point correspondences, then we can form the following equation

$$\mathbf{m}_1 = \mathbf{T}\mathbf{m}_2, \tag{3}$$

then, $\mathbf{T}$ can be computed simply by

$$\mathbf{T} = \mathbf{m}_1\mathbf{m}_2^{-1} \tag{4}$$

where $\mathbf{T}$ is the $4\times4$ transformation matrix including $3\times3$ rotation matrix $\mathbf{R}$ and $3\times1$ translation matrix $\mathbf{t}$:

$$\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix}. \tag{5}$$

Since the matrix $\mathbf{R}$ is the orthogonal matrix, the matrix must be satisfy the following condition

$$\left(\mathbf{r}_1 \times \mathbf{r}_3\right) \cdot \left(\mathbf{r}_2 \times \mathbf{r}_3\right) = 0, \tag{6}$$

where

$$\mathbf{R} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \end{bmatrix}. \tag{7}$$

That means the vectors $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$ are orthogonal each other.

If randomly selected four corresponding 3D points are correct, the estimated transformation $\mathbf{T}$ must be the homogeneous matrix whose last row must be the vector $\mathbf{t}_4 = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}$, and the rotation matrix $\mathbf{R}$ must be the orthogonal matrix.

In practice, the matrix $\mathbf{R}$ may not be orthogonal due to the erroneous data, so we gather the 3D corresponding points that suffice the following equation,

$$\left(\overline{\mathbf{r}}_1 \times \overline{\mathbf{r}}_3\right) \cdot \left(\overline{\mathbf{r}}_2 \times \overline{\mathbf{r}}_3\right) < \lambda, \tag{8}$$

or

$$\left(\overline{\mathbf{r}}_1 \cdot \overline{\mathbf{r}}_2\right) < \lambda \text{ and } \left(\overline{\mathbf{r}}_1 \cdot \overline{\mathbf{r}}_3\right) < \lambda \text{ and } \left(\overline{\mathbf{r}}_2 \cdot \overline{\mathbf{r}}_3\right) < \lambda, \tag{9}$$

where $\lambda$ is the threshold which is the value close to zero and $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$ are the normalized vectors of the rotation matrix $\mathbf{R}$. Normally, we can compute the orthogonallity more accurately with Equation (9) than (8).

Therefore, with RANSAC, we can select the correct 3D points that satisfy the equation (9) and eliminate the miscalculated 3D points, and then we can compute the matrix $\mathbf{R}$ again with only correct correspondences.

As mentioned above, the estimated rotation matrix $\mathbf{R}$ is not really orthogonal, so we need to enforce orthogonality on $\mathbf{R}$. Assume the SVD of $\mathbf{R}$ is $\mathbf{R} = \mathbf{U}\mathbf{D}\mathbf{V}^{\mathsf{T}}$. Since the three singular values of a $3 \times 3$ orthogonal matrix are all 1, we can simply replace $\mathbf{D}$ with the $3 \times 3$ identity matrix $\mathbf{I}$ so that the resulting matrix $\mathbf{U}\mathbf{I}\mathbf{V}^{\mathsf{T}}$ is exactly orthogonal.

After estimating the matrix $\mathbf{R}$, we need to check the correctness of the translation $\mathbf{t}$. We know two centroid of the 3D corresponding points and the rotation matrix $\mathbf{R}$, we can compute the translation $\hat{\mathbf{t}}$. If the computed translation $\hat{\mathbf{t}}$ is very close the matrix $\mathbf{t}$ estimated by equation (5), we can trust the transformation $\mathbf{T}$ is correct.

## 7 3D Points Merging and Transformation Matrix Refinement

When $\mathbf{m}_1$ is 3D point in one scene, and $\mathbf{m}_2$ is the corresponding 3D point in the other scene, we seek a rigid transformation $\mathbf{T}$ that minimizes the mean squared difference between two scenes:

$$E = \sum \left\| \mathbf{m}_1 - \mathbf{T}(\mathbf{m}_2) \right\|^2, \tag{10}$$

where $\mathbf{m}_1, \mathbf{m}_2$ is the corresponding points, and $\mathbf{T}$ is the transformation matrix. In practice, there is no matrix $\mathbf{T}$ which makes the error $E$ zero due to the erroneous data.

Thus, the estimation of the transformation matrix with erroneous 3D data does not guarantee the correct estimation, and the errors are accumulated when the number of registration increases. Therefore, we need to *tune* these 3D points in order to minimize the error.

Suppose we are given the camera parameters, then we can compute the following error:

$$E_1 = \left\| \mathbf{A} - \mathbf{P}_1 \mathbf{T}(\mathbf{m}_2) \right\|, \tag{11}$$

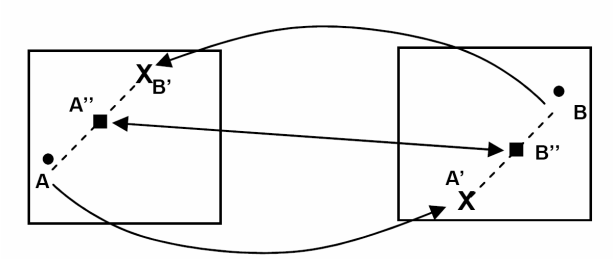


**Figure2**. An example of 3D points merging in a pixel space. $\mathbf{B}$ is transformed to $\mathbf{B}'$ and $\mathbf{A}$ is transformed to $\mathbf{A}'$. If the distance between $\mathbf{A}$ and $\mathbf{B}'$, or the distance between $\mathbf{B}$ and **A'** is bigger than the pixel size, average the 3D point corresponding $\mathbf{A}$, $\mathbf{B}'$, $\mathbf{B}$, $\mathbf{A}'$.

where $\mathbf{A}$ is the 2D point in the camera 1, the $\mathbf{P}_1$ is the perspective projection matrix of camera 1, $\mathbf{T}$ is the transformation matrix, and $\mathbf{m}_2$ is the 3D points corresponding point **B** in Figure 2. Here, $\mathbf{B}' = \mathbf{P}_1\mathbf{T}(\mathbf{m}_2)$.

Similarly, we can compute the following error:

$$E_2 = \left\| \mathbf{B} - \mathbf{P}_2\mathbf{T}^{-1}(\mathbf{m}_1) \right\|, \tag{12}$$

where $\mathbf{B}$ is the 2D point in the camera 2, the $\mathbf{P}_2$ is the perspective projection matrix of camera 2, $\mathbf{T}^{-1}$ is the inverse of the transformation matrix, and $\mathbf{m}_1$ is the 3D points corresponding point **A** in Figure 2. Here, $\mathbf{A}' = \mathbf{P}_2\mathbf{T}(\mathbf{m}_1)$.

The 3D point merging and transformation matrix refining process is as follows:

(1) If the distance $|\mathbf{A} - \mathbf{B}'|$, or the distance $|\mathbf{B} - \mathbf{A}'|$ is bigger than the actual camera pixel size,
(2) average the 3D points corresponding $\mathbf{A}$, $\mathbf{B}'$, $\mathbf{B}$, $\mathbf{A}'$,
(3) then, move the $\mathbf{A}$ and $\mathbf{B}$ to $\mathbf{A}''$ to $\mathbf{B}''$ respectively.
(4) Do this process throughout the image.
(5) Re-estimate the transform matrix,
(6) and repeat this process until the changed value is less than the threshold.

## 8 Experimental Results

In our experiment, Bumblebee stereo camera was used. We captured the multiple images at distinct view point, and the methods proposed in this paper are performed. Experimental results are shown in Figure 3.

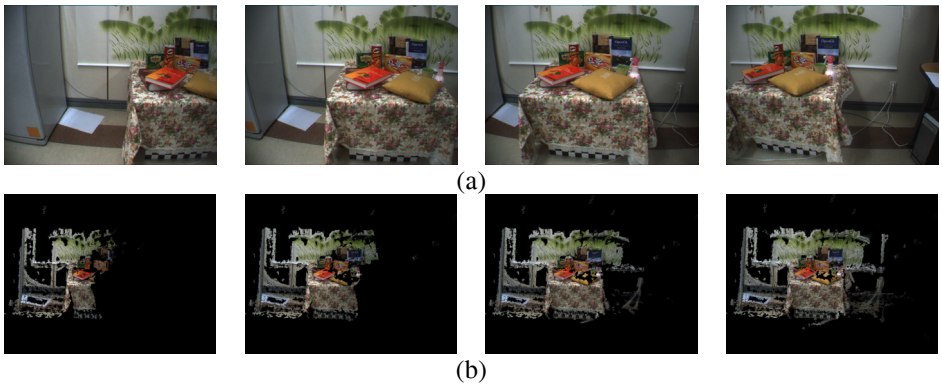

**Figure 3**. Multiple View Range Data Registration. (a) 2D images taken at distinct view point. (b) 3D point cloud image. The points are merged after the new data are measured at different view points

## 9 Conclusion

In this paper, we present the refining these 2D and 3D data method and propose the technique to merge the points while re-estimating transformation matrix. The experiment shows the proposed approach is robust to the erroneous data.

## Acknowledgments

This paper was performed for the Intelligent Robotics Development Program, one of the 21st Century Frontier R&D Programs funded by the Ministry of Commerce, Industry and Energy of Korea. This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MOST) (No. R01-2006-000-11297-0). This work is partly supported by the Science and Technology Program of Gyeonggi province.

## References

1. P. J. Besl and N. D. McKay, "A Method for Registration of 3-D Shapes," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 14(2):239-256, 1992.
2. G. C. Sharp, S. W. Lee and D. K. Wehe, "ICP Registration using invariant features", IEEE Transaction on Pattern Analysis and Machine Intelligence, 24(1):90-102, 2002.
3. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Second Edition, Cambridge University Press, 2003.
4. D. G. Lowe. "Object recognition from local scale invariant features," *Proc. International Conference on Computer Vision*, pp. 1150-1157, 1999.

5. D. G. Lowe. “Distinctive Image Features from scale Invariant Keypoints”, International Journal of Computer Vision 60(2):91-110, 2004.