Point Cloud Registration Using
MSSIR: Maximally Stable Shape Index Regions

MohammadMahdi Manafzade*, Ahad Harati**
*Ferdowsi University of Mashhad, Mashhad, Iran, mahdi.manafzadetabriz@stu-mail.um.ac.ir
** Ferdowsi University of Mashhad, Mashhad, Iran, a.harati@um.ac.ir

Abstract: *Range image registration is one of the fundamental tasks in 3D computer vision and robotics which is gaining more attention with availability of affordable range cameras. Existing recent research has considered application or extension of well known point features like SIFT to the range data; examples include Shape index SIFT and 2.5D SIFT. Compared to RGB image, the quality of range measurement is much worse in sensors like Kinect. This is expected and inherent due to the exploited structured light technique. Therefore, point features may easily mismatched as a result of higher noise level.*
In this paper we show how using region based features may overcome this challenge. MSER features are extracted from shape index image obtained from the input range image. A SIFT-like descriptor is then proposed to encode major smooth regions of the scene as stable features invariant to scale, rotation and affine transformations.
Experimental results are obtained using range image databases of Ohio State University and Stuttgart University which show improvement on the percentage of correct matched features and stability of detected features.

Keywords: range image, point cloud registration, maximally stable shape index regions, MSER.

1.   Introduction

In computer vision and computer graphics, point cloud registration is one of the most important and fundamental tasks that is widely used in various applications such as medical imaging, remote sensing and object recognition. Point cloud registration is the process of aligning two point clouds obtained from the same object or scene with slightly different viewpoints in order to reconstruct a more complete 3D model of the object or the scene. This process involves obtaining a proper transformation which consistently maps one observation to the other.

Creating virtual models for 3D object or scene recognition using data acquired from indoor and outdoor environments have been ingratiated by experts and researchers. In recent years, some methods and approaches have been proposed for registration. One of these approaches is the point-based approach. The most well-known method in this approach is Iterative Closest Point (ICP) [21, 22]. ICP is one way to register 3D point clouds with rigid transformations and employed to minimize the difference between two point clouds. Nevertheless it has some drawbacks and disadvantages. Its first drawback is time consuming specifically when the number of points is large or point clouds have large difference between their views and the second one is that ICP sometimes converges to the local minima. The algorithm is conceptually simple and iterates two steps. In first step, it matches each point of the first point cloud with its closest point on the other one and in second step it estimates the rotation and the translation matrix between the two point clouds using the matched point correspondences and mean square cost function. It iterates the steps until some stopping criteria are satisfied. The criteria can be the maximum iteration or the amount of similarity between two MSEs at the last and the penultimate iterations. In order to achieve robust and accurate alignment, many variants of this algorithm have been proposed [23]. Another approach for registration is the feature-based approach. This approach consists of two steps. First, interest and salient features are extracted and in the next step, the transformation matrix is estimated by the matched features. This approach has differences to point-based approach. The feature-based methods are usually fast. Also, they can solve the problem of sticking in local minima but in this approach, the result of the registration often suffers from errors. So in order to reduce such errors, after performing a specific feature-based method, we need to use the point-based approach like ICP.

So far, various feature-based methods have been introduced and developed by researchers that extract and describe features from range images. Some such methods are extended techniques that have already been applied in RGB and gray images like SIFT [3, 4] and SURF [5, 6]. 2.5D SIFT [19] and shape index SIFT [20] are examples of such methods. In shape index SIFT, first the shape index value is computed for all pixels of a range image. Then SIFT is applied on the shape index image and interest features are extracted and described by SIFT

descriptor. Shape index value is somewhere between -1 to 1. It describes the type of surfaces locally. An important property about shape index is that it is 3D geometric invariant. We also use shape index in our novel method because of its suitable properties. In 2.5D SIFT, it uses the original SIFT keypoint detector [3, 4]. This method has been extended for range images. However, in 2.5D SIFT; an extra preprocessing step is required. This algorithm extracts features in range images. After this step, a descriptor is obtained by using the histogram of the shape index values and the range gradient orientations around these keypoints. Another keypoint extraction method is the normal aligned radial feature (NARF) [12] presented by B. Steder et al. that operates on range images. Its detector looks for stable areas with significant change in vicinity that can be identified from different viewpoints and descriptor characterizes the area around the keypoint by calculating a normal aligned range value patch and finding the dominant orientation of the neighboring pixels. There are a lot of descriptors for characterizing feature points extracted in range images. A well-known descriptor for point clouds and range images is spin-image [17, 18]. Spin-image was proposed by A. E. Johnson. A spin-map is defined so as the function that projects 3-D points to the 2-D coordinates of a particular basis (α, β) corresponding to oriented point O. For a given point cloud P, α is the distance from projected neighbor points on the tangent plane to P and β is the distance from neighbor points to the tangent plane. Spin image is composed of $m \times n$ squares. So its x-axis and y-axis show α, β values that are split up into m and n squares respectively. For all neighbours of a certain point, α and β values are computed and then added to respective square. Spin-image is robust to clutters and occlusions and invariant to rigid transform. Some texture descriptors like LBP [9, 10, 24], 3D LBP [25] and LSP [11] have also been used for describing point clouds.

In this paper we propose a new method that extracts stable regions called maximally stable shape index regions (MSSIR) instead of interest points. The algorithm extends MSER detector for range images and can be used in registration and object recognition. MSER detector finds distinct and stable regions in gray images. P.-E. Forss´en introduced the extended MSER for color images called MSCR [2]. We utilize MSER detector in this proposed method by reason of having high repeatability. K. Mikolajczyk et al. [7, 13] compared different affine region detectors and concluded that MSER has high repeatability and often leads to better results than other region detector methods. The paper is structured as follows: In Section 2, we explain a method called Shape index. Then the MSER method is described in Section 3. In Section 4, our proposed method called MSSIR is presented and finally we compare our method to several state-of-the-art methods about registration and matching with well known datasets in section 5.

2. Shape Index

Shape index was first introduced by J. Koenderink and A. J. van Doorn [16] to characterize the type of surfaces and represent the geometry of objects locally. As shown in Fig. 1, the Shape index ranges from -1 to 1. It has considerable and useful properties such as scale, pose and 3D rotation invariance.
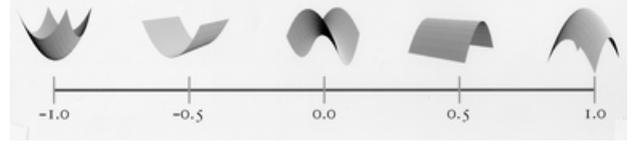

Fig. 1. Shape index value ranges from -1 to 1.

In order to obtain the shape index value for each point, a local neighborhood around it is considered and a quadratic polynomial is fit to the corresponding points. So the coefficients of the quadratic polynomial are first found by means of least square error. Equation (1) and Equation (2) indicate a quadratic polynomial form and shape index formula respectively.

$$G(x, y) = ax^2 + by^2 + cxy + dx + ey + f \qquad (1)$$

$$s = \frac{2}{\pi} \tan^{-1}(\frac{k_2 + k_1}{k_2 - k_1}) \qquad (2)$$

Where $k_1 > k_2$ and k_1 , k_2 are principle curvatures shown in Equation (3) and Equation (4).

$$k_1 = H + \sqrt{H^2 - K} \qquad (3)$$

$$k_2 = H - \sqrt{H^2 - K} \qquad (4)$$

H and K are called mean and Gaussian curvatures respectively. You can see how they are computed in Equation (5) and Equation (6).

$$H(i, j) =$$
$$\frac{(1 + f_y^2(i, j)) f_{xx}(i, j) + (1 + f_x^2(i, j)) f_{yy}(i, j) - 2 f_x(i, j) f_y(i, j) f_{xy}(i, j)}{2(\sqrt{1 + f_x^2(i, j) + f_y^2(i, j)})^3} \qquad (5)$$

$$K(i, j) = \frac{f_{xx}(i, j) f_{yy}(i, j) - f_{xy}^2(i, j)}{(1 + f_x^2(i, j) + f_y^2(i, j))^2} \qquad (6)$$

Where (i, j) is the ith and j[th] pixel of the range image. f_x, f_y, f_{xx}, f_{yy} and f_{xy} denote the first and second derivatives at (i, j) position.

3. MSER

MSER (Maximally stable extremal regions) was introduced by Matas et al. [1] for the first time. The method finds stable and distinct regions into an image. The MSERs detection uses a watershedding process that can be described in the following way:

It binarizes the image with different thresholds t and the number of levels. So the image is divided into two groups black (B) and white (W) as shown in Equation (7).

$$B := \{x \in \Omega^2 : I(x) < t\}, W := \Omega^2 \setminus B \qquad (7)$$

The gray-scale image is represented by function $I: \Omega \rightarrow [0 \dots 255]$ where $\Omega = [1 \dots W] \times [1 \dots H]$ is the set of all image coordinates and t can be selected between $[0 \dots 255]$. When the threshold t is zero, the set W contains all image coordinates but by increasing t, the cardinality of the set W reduces and the cardinality of the set B increases until the whole image becomes black. While threshold t increases, connected components are created and merged together. The extracted Areas or connected components in an image are called extremal regions and maximally stable extremal regions are those that have changed in size only a little across at least several intensity threshold levels. The MSER algorithm uses Equation (8) to find stable areas.

$$\Psi(R_i^t) = \frac{|R_j^{t-\Delta}| - |R_k^{t+\Delta}|}{|R_i^t|} \qquad (8)$$

Where $|.|$ denotes the cardinality, R_i^t is a region obtained by thresholding at a gray value t and Δ is a stability range parameter. $R_j^{t-\Delta}$ and $R_k^{t+\Delta}$ are the extremal regions obtained by moving upwards and downwards in the component tree from the region R_i^t until a region with gray value $t - \Delta$ and $t + \Delta$ are found. MSERs correspond to those nodes of the tree that have a stability value Ψ, which is a local minimum along the path. Note that whatever the value Δ increases, more stable regions are found but their number will decrease. The value is selected by using trial and error and according to experimental images. In order to get better results, we can perform mentioned procedures on the image and its negative to get both dark areas surrounded by bright background and bright areas surrounded by dark background. In Fig. 2, you can see an image that MSERs are extracted from it and watershedding process.
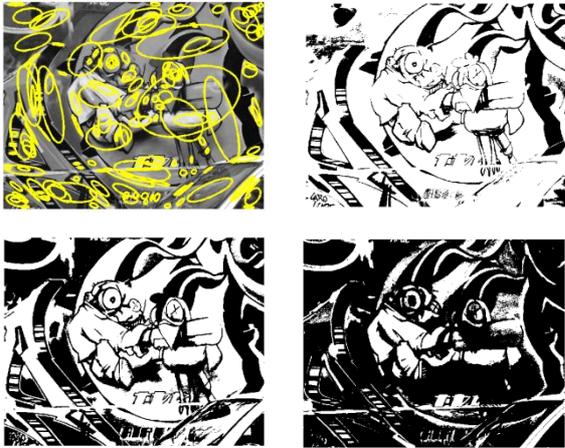

Fig. 2.a) Extracted MSERs, b, c, d) Watershedding process(Thresholding with different threshold values)

In order to describe MSERs, we have used SIFT descriptor [8]. For description, first the affine normalization is performed on regions. It is done by computing covariance matrix and estimating an ellipse using eigenvalues and eigenvectors of the covariance matrix and converting it to circle. The square patch is split up into 4*4 sub-squares and at every square, the direction and the magnitude of the gradients are computed and then discretized to eight bins. Thus the length of the descriptor will be 128 (4*4*8=128). We have utilized vlfeat library to detect MSER regions [15].

4. MSSIR

In this paper, we introduce a novel method called maximally stable shape index regions (MSSIR) to match range images and reconstruct 3d models. The proposed method can be utilized in the registration and 3D object recognition fields. A lot of methods exist which extract interest features from range images and then match them together by their descriptors such as 2.5d SIFT [19] and shape index SIFT [20]. We intend to extract stable and distinct regions called MSSIR instead of feature points within range images. Therefore our algorithm consists of the following steps:

1) Compute the shape index value for each pixel within the range image by using Equation (2).

As demonstrated before, the shape index is 3D rotation invariant. We need to transform raw range images to shape index images because a raw range image is not invariant to rotation and translation.

2) Extract MSSIRs within it.

We extract MSSIR regions by binarizing the shape index image with different threshold values. For this reason, we use Equation (8) as stability measure for regions and select them that have minimum area size variation at certain threshold interval.

3) Describe by means of SIFT descriptor [8].

We describe extracted MSSIRs by creating a histogram. The histogram contains 128 bins like SIFT descriptor. So we split up the patch (circular region) around the center of region to 4*4 sub regions. In every sub regions, we compute the orientation and the magnitude of gradients and discretize them to 8 bins with 45-degree intervals. Finally our descriptor vector will have 128 elements.

4) Apply RANSAC algorithm [14] to remove outliers and prevent wrong and incorrect matches so that we get the best rotation matrix and translation vector that transform one point cloud to the other one. Fig. 5 indicates the steps of the proposed algorithm to match two test range images. As shown in Fig. 5, the proposed detector has found areas as MSSIRs which have had similar surface type and topology. Results show that the proposed method has high matching accuracy to match extracted features and areas. Our method is also robust to partial occlusions and clutters and invariant to rotation, scale and affine transformations.

5.   Experimental Results

We have used Stuttgart range image [26] and Ohio State University range database [27]. Stuttgart range image database consists of 42 classes of different objects. Each class contains 66 range images from different views and the size of each range image is 400*400. Range images in this database contain 3D point cloud (x, y and z) and the mask of the object. Ohio State University range database contains different images and objects. Some of them have both range and colour images and the others have only range images. The size of the Selective images is 200*200. Fig. 3 and Fig. 3 show some such images of these databases.
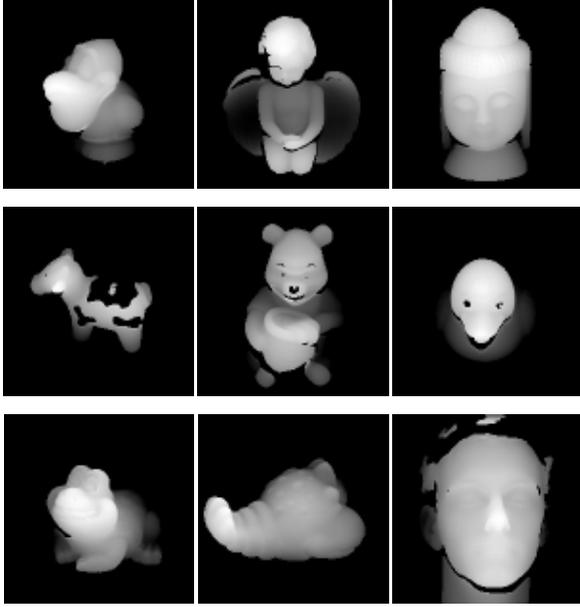

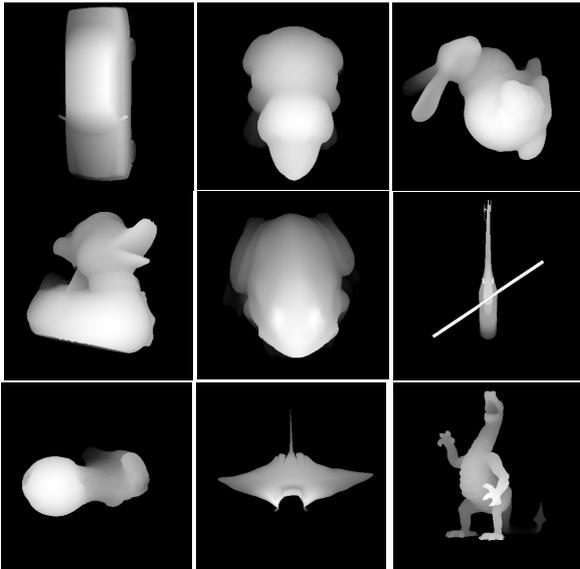Fig. 3. Example images from Ohio range image database [27]

Fig. 4. Sample images from Stuttgart range image database [26]

In TABLE I, MSER parameters have been shown used in our experiments. We have tried to select and get the most appropriate parameters but they are not optimal and depend on images.

TABLE I: MSER parameters

MinDiversity	MaxVariation	Delta	MinArea	MaxArea
0.7	0.3	7	0.0005	0.1

Fig. 6 indicates two point clouds from different views and the integrated point cloud.

Fig. 7 shows the registration result for two point clouds of the bunny object with and without Applying ICP algorithm. In Fig. 8, some test images with the matching procedure have been observed.
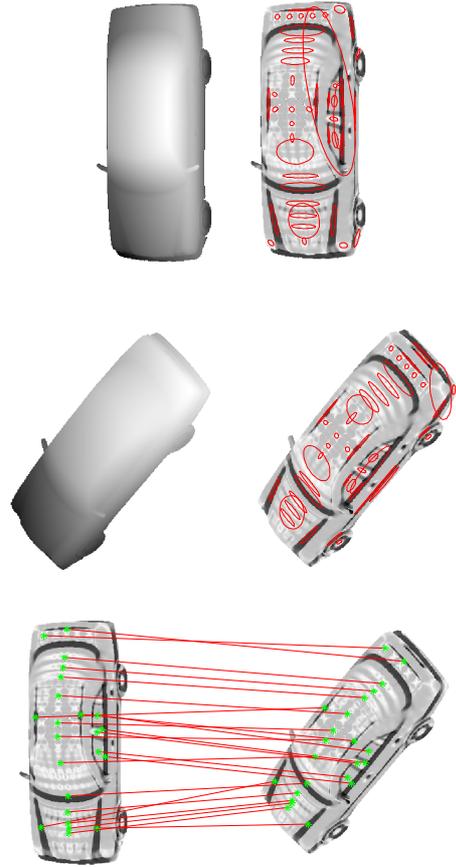

Fig. 5. Up and middle images: two car range images and their shape index images with extracted regions by MSSIR, bottom image: the process of matching

As be seen in TABLE II, The Mean squared error before applying ICP is small for the bunny point cloud in

Fig. 7 that it means that the registration using MSSIR leads to good and appropriate result alone and without performing ICP but to improve point clouds registration, we need to perform ICP. The MSE value after applying ICP has changed from 0.5364 to 0.2968.

TABLE II: The mean squared error before and after applying ICP

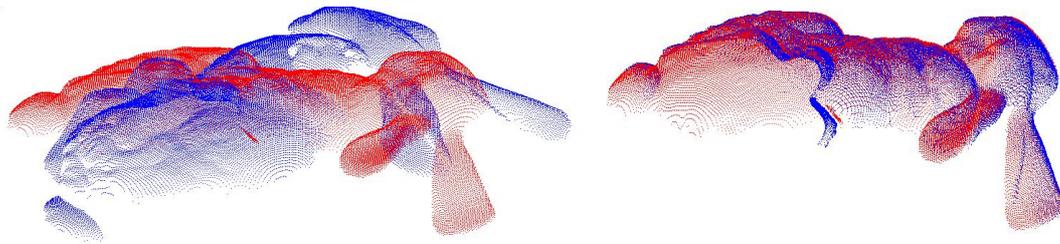| MSE before applying ICP	MSE after applying ICP
0.5364	0.2968

Fig. 6. Two overlapped point clouds of Bunny from different views before and after registration
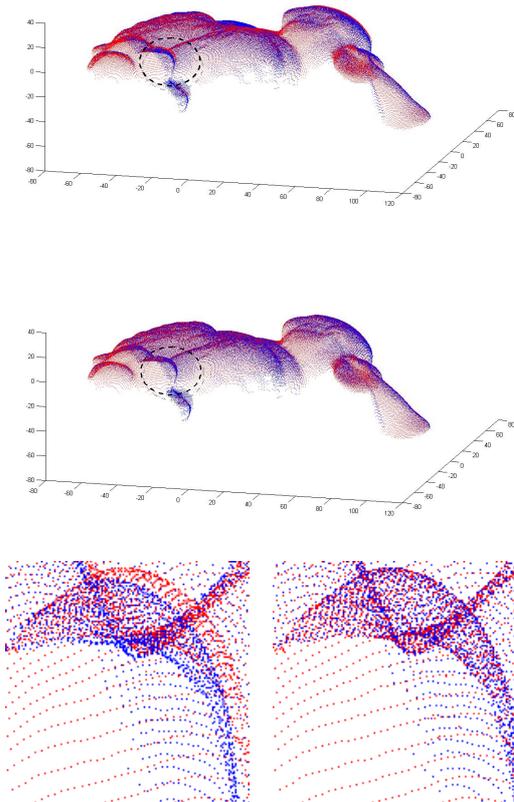

Fig. 7. Top image: Bunny point cloud registration without performing ICP, Middle image: Bunny point cloud registration with performing ICP, Bottom images and left to right: a part of bunny point cloud in zoom mode to display the registration result without and with applying ICP.

One problem in matching is symmetry in objects. This problem leads to mismatch between regions. For example, a car or the human body is symmetric with respect to a vertical axis through the middle. So for these images which have symmetrical components, solving matching problem is really difficult and complex. We can just remove these regions by setting the ratio between first to second best matching. The ratio has been set to 0.8.

In order to discern false from true matches, we have applied the RANSAC algorithm. This algorithm can remove outliers effectively. TABLE III shows the number of all matches and the number and the percentage of correct matches for some test images. We can see that the proposed method has high accuracy but note that the number of detected regions and correct matches depends on many factors such as MSER parameters especially delta parameter, the complexity of objects and shapes and other factors. In TABLE III, we have compared our method to the other state-of-the-art methods (Shape index SIFT and 2.5D SIFT) for some images in Ohio range image database. In order to evaluate the methods, we have selected similar images for all of them. As be seen, our method outperforms other methods. In other words, the percentage of correct matches in our method is more than 2.5D SIFT and Shape index SIFT methods.

TABLE III: The number and the percentage of correct MSSIR matches for different objects

	Auto	Bunny	Idea	Mole
Number of matches	23	26	71	31
Number of correct matches	23	22	67	28
percentage of correct matching	100	84.62	94.37	90.32

TABLE IV: The comparison between three methods with Ohio state university range image database. T, C and F mean the number of total, correct and false matches respectively.

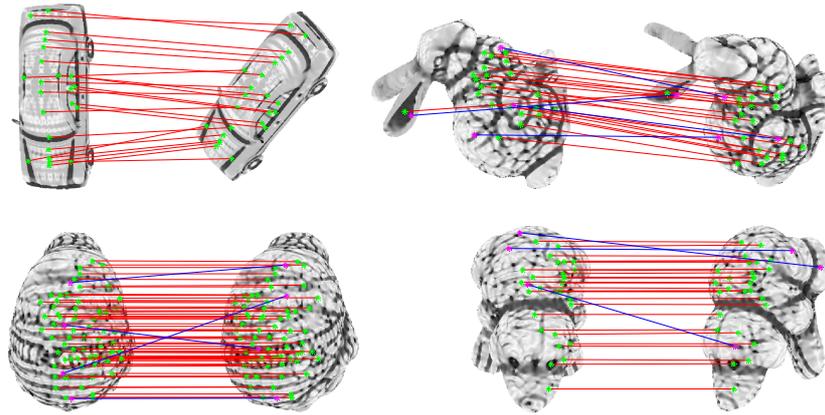| Name	2.5D SIFT			Shape index SIFT			Proposed method		
	T	C	F	T	C	F	T	C	F
Lobster2	8	8	0	32	20	12	19	15	4
Bird	9	8	1	37	30	7	22	19	3
Duck	6	4	2	25	19	6	10	10	0
Pooh	9	8	1	37	27	10	18	16	2

Fig. 8. The matching procedure for test images. Red and blue lines show correct and incorrect matches respectively.

6. Conclusion

We proposed a new method to register point clouds and range images by region-based methods. Our method is robust to occlusions, clutters and noise. In future works, we intend to change and improve our descriptor and fuse colour and depth information together. We would also like to extract and find salient regions by normal vector images instead of shape index. We will also change our region-based methods (MSSIR) to other methods like IBR, EBR and salient regions and compare their results to each other. Finally we will experiment our method with different databases and conditions.

References

[1] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," In 13th BMVC, pp. 384–393, September 2002.

[2] P.-E. Forss´en, "Maximally stable colour regions for recognition and matching," Proc. IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, USA, IEEE Computer Society, June 2007.

[3] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," IJCV, 60(2):91–110, 2004.

[4] D. G. Lowe, "Object recognition from local scale-invariant features," International Conference on Computer Vision, Corfu, Greece, pp. 1150-1157, September 1999.

[5] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features," In ECCV, pp. 404–417, 2006.

[6] H. Bay, A. Ess, T. Tuytelaars and L. V. Gool, "Speeded up robust features (SURF)," Computer Vision and Image Understanding, 110(3), pp. 346-359, 2008.

[7] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. van Gool, "A comparison of affine region detectors," IJCV, 65(1/2), pp. 43-72, 2005.

[8] P.-E. Forssén and D. G. Lowe, "Shape descriptors for maximally stable extremal regions," In International Conference on Computer Vision (ICCV), October 2007.

[9] T. Ojala, M. Pietikäinen, and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," Proc. 12th IAPR International Conference on Pattern Recognition (ICPR), vol. 1, pp. 582-585, 1994.

[10] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," Pattern Recognition, vol. 29, pp. 51-59, 1996.

[11] H. R. Pourreza, M. Masoudifar, M. ManafZade, "LSP: Local similarity pattern, a new approach for rotation invariant noisy texture analysis," Proc. 18th IEEE International Conference on Image Processing (ICIP), pp. 837-840, 2011.

[12] B. Steder, R. B. Rusu, K. Konolige, and W. Burgard, "NARF: 3D range image features for object recognition," in Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), workshop on Defining and Solving Realistic Perception Problems in Personal Robotics, Taipei, Taiwan, 2010.

[13] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," In IJC V 60(1), pp. 63-86, 2004.

[14] M. A. Fischler, R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," vol. 24, Issue 6, pp. 381-395, June 1981.

[15] Web-site. http://www.vlfeat.org/

[16] J. J. Koenderink and A. J. van Doorn, "Surface shape and curvature scales," Image and Vision Computing, 10(8), pp. 557-565, 1992.

[17] A. E. Johnson, "Spin-Images: A representation for 3-D surface matching," PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, August 1997.

[18] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(5):433–449, 1999.

[19] T.W.R. Lo, J.P. Siebert, "Local feature extraction and matching on range images: 2.5D SIFT," Computer Vision and Image Understanding 113 (2009) 1235–1250.

[20] N. Bayramo˜glu, A.A. Alatan, "Shape index SIFT: Range image recognition using local features," 20th International Conference on Pattern Recognition (ICPR), pp. 352–355, 2010.

[21] P. Besl and N. McKay, "A method of registration of 3-D shapes," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, no. 2, pp. 239-256, February 1992.

[22] Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," International Journal of Computer Vision, vol. 13, no. 2, pp. 119-152, 1994.

[23] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," 3th International Conference on 3D Digital Imaging and Modeling, pp. 145-152, 2001.

[24] T. Ojala and M. Pietikaeinen, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Trans. Pattern Analysis and Machine Intelligence, 24 (2005), 971–987.

[25] Y. Huang, Y. Wang, and T. N. Tan, "Combining Statistics of Geometrical and Correlative Features for 3D Face Recognition," ICPR04, 3 (2006), I: 330–333.

[26] http://range.informatik.uni-stuttgart.de/htdocs/html/

[27] http://sampl.ece.ohiostate.edu/data/3DDB/RID/minolta/