Good Features to Track for RGBD images

Maxim Karpushin, Giuseppe Valenzise, Frederic Dufaux

GOOD FEATURES TO TRACK FOR RGBD IMAGES

Maxim Karpushin[†]*, Giuseppe Valenzise*[*]*, Frédéric Dufaux*[*]

[†] LTCI, Télécom ParisTech – Université Paris-Saclay
[*] L2S, CNRS – CentraleSupélec – Université Paris-Sud

ABSTRACT

RGBD (texture-plus-depth) image representation enriches traditional 2D content with additional geometrical information, having the potential to improve the performance of many computer vision tasks. In image matching, this has been partially studied by considering how depth maps can help render feature descriptors more distinctive. However, little has been done to design keypoint *detection* approaches able to leverage the availability of depth information. In this paper, we propose a novel and robust approach for detecting corners from RGBD images. Our method modifies a classical corner detection strategy, based on local second-order moment matrices, by computing derivatives in a coordinate system which reflects the local properties of object surfaces. Our results demonstrate a higher stability to out-of-plane rotations of the proposed RGBD corner detector both in terms of feature repeatability and in a visual odometry application.

Index Terms— RGBD, texture+depth, corner detection, keypoint extraction, local features

1. INTRODUCTION

The task of establishing local visual correspondences between two or more images consists in finding a set of local features that are visually similar in different pictures up to certain transformations, such as rotations, translations, changes of scale, etc. A number of computer vision problems can broken down into determining the number and the strength of these correspondences, including image retrieval [1], indexing [2], classification [3], object tracking [4], visual odometry [5] and simultaneous localization and mapping (SLAM) [6].

During the last decade, the problem of finding visual correspondence has been thoroughly studied for conventional images [7], leading to a variety of techniques for local features extraction [8, 9, 10], and stimulating standardization activities in MPEG [11, 12]. Nonetheless, the emergence of new and richer video formats, such as 3D meshes, plenoptic, lightfield and RGBD (texture+depth) images calls for novel image matching tools, and in particularly for improved local features able to leverage these extended representations of visual content. In this work we focus on the RGBD representation, as this kind of content has become relatively easy to capture thanks to the popularity of acquisition devices such as the Microsoft Kinect. Differently from point clouds, the RGBD format offers a simple pixel-based representation consisting of a conventional *texture* image, and of a *depth* map, which stores the distance of objects in the scene from the camera plane.

RGBD content can be matched based on its texture only, through some conventional 2D image matching algorithm. However, it has been extensively shown [13, 14, 15, 16, 17, 18, 19, 20] that involving the complementary depth information in a proper way into the feature extraction process may render the features from the texture map more robust to certain visual deformations – notably, perspective distortions, out-of-plane rotations and viewpoint position changes – to which texture-only local features are instead very sensitive. While several techniques to enhance local texture description employ depth [13, 14, 15, 21, 22], little has been done to exploit geometric information in the very first step of feature extraction, i.e., keypoint *detection*. In our previous work [19] we proposed a scale-invariant *blob* detector for RGBD images. In this paper, we further study the problem of keypoint extraction, by dealing with another class of widely used interest points: *corners*.

Specifically, we consider the popular Good Features to Track (GFTT) [23] and Harris corner detection principles, and extend them to RGBD content, making the detected keypoints more robust to 3D distortions with a moderated computational effort. More precisely, the contribution of this paper is twofold: i) we describe a generic technique to exploit depth into a local feature detection-related processing of the texture map by means of adaptive local axes; ii) using the proposed technique, we extend the GFTT and Harris corner detection approaches to texture+depth images. The proposed detectors are then tested using a standard keypoint repeatability evaluation scenario on synthetic RGBD images as well as in a visual odometry setting on real RGBD images acquired with the Microsoft Kinect sensor.

The rest of the paper is organized as follows. In Section 2 we describe the related work on keypoint detection. Section 3 introduces the notion of local adaptive axes and describes the proposed extension of GFTT and Harris detectors for RGBD content. Section 4 presents the experimental results obtained in the two testing scenarios in comparison to four other keypoint detectors. Concluding remarks and future work are given in Section 5.

2. BACKGROUND AND RELATED WORK

Visual correspondences are generally obtained through image matching algorithms. These typically consist of three stages: (1) keypoint detection, aiming at finding interesting points in an image, which have chances to be located in transformed images as well; (2) feature description, providing a compact signature of the previously detected keypoints based on their neighborhoods; and (3) descriptor matching, defining a comparison protocol for the local descriptors. In this paper we focus on the first step, i.e., detection of keypoints in RGBD images.

Keypoint detectors can be broadly classified into blobs and corners detectors, according to which kind of image characteristics they search for. Popular examples of blob detectors include the Difference-of-Gaussians detector (DoG) integrated into the Scale Invariant Feature Transform (SIFT) pipeline [8], and the Fast Hesssian detector used in the Speeded Up Robust Features (SURF) [9]. While several extensions of blob-based local feature descriptors to RGBD

exist in the literature [13, 15, 16, 20], only [19] provided an explicit formulation of a scale space construction for texture+depth data, as well as a complete blob detector pipeline based on it. Up to the authors' knowledge, this paper is the first to propose a texture+depth *corner* detector.

The first and still very popular corner detector robust under moderate content deformation was proposed by Harris and Stephens [24]. They observed that the ratio of eigenvalues of the second-order moment matrix of image gradient, $\mathcal{M}$, smoothed by a convolutional kernel W, i.e.,

$$\mathcal{M} = \begin{pmatrix} \left(\frac{\partial I}{\partial u}\right)^2 & \frac{\partial I}{\partial u}\frac{\partial I}{\partial v} \\ \frac{\partial I}{\partial u}\frac{\partial I}{\partial v} & \left(\frac{\partial I}{\partial v}\right)^2 \end{pmatrix} * W, \tag{1}$$

may serve as a clue of a repeatable corner: if the ratio is low enough at a given position within the image, a distinctive corner is discovered (here $I(u,v)$ is the image intensity at pixel (u,v)). Moreover, they showed that the explicit computation of the eigenvalues is unnecessary, and the quantity

$$R = \det\mathcal{M} - k\mathrm{tr}^2\mathcal{M} \tag{2}$$

is positive in corner regions and negative near edges (k is a constant). Local maxima of R in the spatial image variables reveal corners, which are more distinctive (and thus stable under content deformations) for larger values of R. Thus, the Harris keypoint detector employs thresholded local maxima of R as keypoints.

Shi and Tomasi [23] proposed a different detection principle, also based on the eigenvalues of (1). Studying which image locations might be reliably tracked, they observed that local maxima of the *minimum* eigenvalue $\lambda_{\min}$ of $\mathcal{M}$ also reveal distinctive image corners. Additionally, thresholding the minimum eigenvalue ensures well-conditioning of $\mathcal{M}$, which facilitates the selection of stable keypoints. This led to a corner detector known as Good Features to Track (GFTT). Numerous contributions related to the second-order matrix-based detectors have been proposed afterwards. To name a few, Sipiran and Bustos [25] extended Harris detector for meshes. Tommasini et al. [26] proposed an efficient technique of spurious feature rejection for GFTT.

In this work we develop a generic technique allowing a local keypoint selection test (specifically, a corner test) to be performed within the *surface metric* instead of the standard image plane metric. With this technique we aim at better keypoint repeatability under various 3D deformations (notably viewpoint position changes): our key observation is that the surface metric reflects intrinsic surface properties being *covariant* to camera position changes. This allows to reveal keypoints in texture image in a way intrinsically independent to the camera position. In the following section we explain how this could be done with an affordable computational cost. In the rest of this paper, we focus on second-order moment matrix based detectors, such as Harris and GFTT, whose principle has been also used as additional stability criterion in more complex detectors to filter out keypoint candidates situated near edges [8, 27]. The proposed approach can be extended to alternative corner detection strategies, such as SUSAN (the Smallest Univalue Segment Assimilating Nucleus) [28] or Accelerated Segment Test, an extended version of SUSAN initially developed in [29, 30] and further employed in scale-covariant local feature detection pipelines [10, 27].

3. THE PROPOSED APPROACH

Perspective distortions are the major effect of out-of-plane rotations and arbitrary camera position changes, and hinder the repeatability of keypoints. In order to model and compensate for these transformations, we propose to perform a change of basis, obtained by computing *local adaptive axes*. This new system of coordinates will replace pixel-wise the regular image plane grid, enabling to interpret the perspective distortions at each pixel of the texture map by means of the surface metric defined by the depth map.

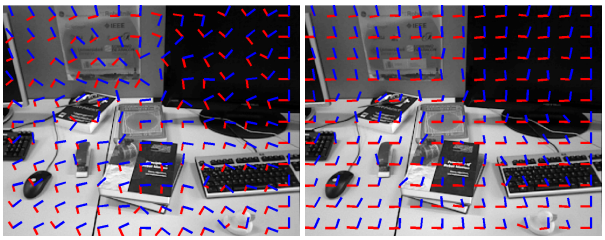

Fig. 1. Local adaptive axes computed for a RGBD image from *Freiburg long office household* sequence: raw vectors given by PCA before (left) and after regularization (right).

To this end, following our prior work [17, 20] we first consider that the texture image describes the photometric pixel intensities mapped to the manifold defined by the depth map. Hence, to detect keypoints on the textured manifold, we apply the corner test in the *tangent plane* for each point of the surface. To do so, we sample directional image derivatives $\frac{\partial I}{\partial \vec{\xi}}$, $\frac{\partial I}{\partial \vec{\eta}}$ in the local axes $(\vec{\xi}, \vec{\eta}) \in \mathbb{R}^{2\times 2}$, which are thus a basis of the tangent plane projected on the camera plane. We use these derivatives to replace those in (1), obtaining the new second-order matrix:

$$\mathcal{M}_{\text{proposed}} = \begin{pmatrix} \left(\frac{\partial I}{\partial \vec{\xi}}\right)^2 & \frac{\partial I}{\partial \vec{\xi}}\frac{\partial I}{\partial \vec{\eta}} \\ \frac{\partial I}{\partial \vec{\xi}}\frac{\partial I}{\partial \vec{\eta}} & \left(\frac{\partial I}{\partial \vec{\eta}}\right)^2 \end{pmatrix} * W. \tag{3}$$

Harris corner or GFTT minimum eigenvalue tests and the non-local maximum suppression are then carried out in the same way as before. The detection is thus performed on the manifold but not in the image plane, becoming intrinsically linked to the content and less dependent to the relative pose of the camera with respect to the surface.

In the following we explain how to compute a (regular) local axes field $(\vec{\xi}, \vec{\eta})$ to perform the proposed on-manifold corner detection approach.

3.1. Local adaptive axes field computation

The following procedure to compute local axes is carried out at each image point (u,v). To obtain the adaptive local axes $(\vec{\xi}, \vec{\eta})$, we first select a basis $(\vec{a}, \vec{b}) \in \mathbb{R}^{3\times 2}$ of the tangent plane (in the scene coordinates) and project it on the camera plane using the pinhole camera model. The basis is obtained by means of PCA-based normal estimation [31], which is a common technique to estimate surface normals in point clouds. It consists in computing the eigenvector corresponding to the smallest eigenvalue of the point cloud covariance matrix, which gives a good estimate of the surface normal vector. In applications that need only the normals, the other two eigenvectors are neglected, but we make use of them since these two vectors provide an orthonormal basis of the tangent plane. An illustration is given in Fig. 1 (left). It is worth noticing that this estimation can be implemented in a computationally efficient way using integral images technique [32].

3.2. Local adaptive axes regularization

As observed in Fig. 1 (left), local axes computed with PCA are quite irregular: when passing from a point on the surface to its neighbor, the corresponding basis vectors may change significantly, due to noise or fine variations in the depth. This might have a significant impact when computing the eigenvalues of the second-order moment matrix $\mathcal{M}$, as well as in the non-local maximum suppression of GFTT or Harris corner scores.

To cope with this, we need to regularize the local axes field, e.g., by enforcing its continuity along relatively smooth continuous surfaces. We propose to do it in the following simple way: since we are free to choose the basis $(\vec{a}, \vec{b})$ of the tangent plane, we rotate each pair $(\vec{a}, \vec{b})$ around the normal vector so that the first basis vector $\vec{a}$ always lies in XZ plane (X and Y are standard image axes, Z is oriented forward). With no loss of generality we may assume that such a basis always exists. An example of such regularized local axes vector field is given in Fig. 1 (right), where $\vec{a}$ vectors are displayed in red. Formally, we compute the two basis vectors as follows. Let $\mathbf{V} = \left(\vec{n}\ \vec{a}\ \vec{b}\right)$ be the result of the PCA decomposition, where $\vec{n}$ is the normal vector, and $\vec{a} = (a_x\ a_y\ a_z)^T$, $\vec{b} = (b_x\ b_y\ b_z)^T$. We look for a rotated pair $(\vec{a}^*, \vec{b}^*)$ so that $\vec{a}^* = (a_x^*\ 0\ a_z^*)$. The rotation around the normal $\vec{n}$ is then expressed by means of a rotation matrix $\mathbf{R}$ as follows:

$$\vec{a}^* = \mathbf{V}\mathbf{R}\mathbf{V}^{-1}\vec{a},\ \mathbf{R} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{pmatrix}. \quad (4)$$

We develop this expression using the relation between $\vec{a}$ and $\mathbf{V}$:

$$\vec{a}^* = \mathbf{V}\mathbf{R}\mathbf{V}^{-1}\vec{a} = \mathbf{V}\mathbf{R}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \vec{n} & \vec{a} & \vec{b} \end{pmatrix}\mathbf{R}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}. \quad (5)$$

This gives us the following equation to derive the rotation angle:

$$a_y^* = a_y\cos\alpha - b_y\sin\alpha = 0, \quad (6)$$

for which we can safely assume the following solution:

$$\cos\alpha = \frac{b_y}{\sqrt{a_y^2 + b_y^2}},\ \sin\alpha = \frac{a_y}{\sqrt{a_y^2 + b_y^2}}. \quad (7)$$

We plug the expression (7) into $\mathbf{R}$ in (4) with no need to derive α explicitly. Finally, according to Eq. (5), the resulting pair $(\vec{a}^*, \vec{b}^*)$ corresponds to the second and third column of the $\mathbf{VR}$ matrix respectively. For further regularity of the resulting local axes field, we may need to flip the on-screen projections $\vec{\xi}$ and $\vec{\eta}$ of $\vec{a}^*$ and/or $\vec{b}^*$ so that they form a properly oriented pair, e.g., ensuring that $\vec{\xi}$ is oriented towards left with respect to the camera, as in the example in Fig. 1 (right).

This technique allows to render a continuous vector field efficiently (with few floating point multiplications and additions and one square root computation per pixel) and stabilize the keypoint detection.

4. EXPERIMENTS AND DISCUSSION

We test the proposed approach in two scenarios: a standard detector repeatability test and a visual odometry application. In both of them we perform the keypoint detection using the original Harris corner test[1] and GFTT minimal eigenvalue test using matrix (3).

[1]We set $k = 0.04$ as it is done in OpenCV implementation of Harris corner detector

The two proposed variants are compared to other popular state-of-the-art corner detectors for conventional images: we involve GFTT, Harris and Fast AST [29] detectors implemented in OpenCV and the original implementation of BRISK detector [10], which is a scale-covariant extension of Adaptive Generic AST [30].

4.1. Repeatability

The repeatability score is a standard measure to evaluate the keypoint detector performance [33, 34]. For a given sequence of views of the same scene, it consists in detecting keypoints in all the images and then computing the fraction of repeated keypoints between a reference view and the remaining ones. A keypoint is considered as repeated if it occupies (approximately) the same physical area of the scene. We perform this test by means of overlap error: all the keypoints are considered as spheres of a unit size, and the volumetric overlap of at least 50% is required between a reference and a test keypoint to count the detected keypoint as repeated. The necessary ground truth to perform this check is provided within the data. In this test we keep default values of all the parameters for all the methods, except the score thresholds: we chose appropriate threshold values for each image sequence so that all the detectors provide approximatively 1000 keypoints per image. This ensures a fair comparison between detectors.

The experiment is performed on three different sequences of synthetic RGBD images representing viewpoint position changes and out-of-plane rotations (in total 70 images of 960×540 pixels). The resulting repeatability scores in function of reference–test viewpoint angles difference are displayed in Fig. 2. Both the proposed extensions demonstrate a moderately improved repeatability compared to all the other methods, with the only exception on the last image sequence, notably for GFTT. BRISK detector exhibits lower matching scores, since it is the only scale-covariant detector used in the test, but for a fair comparison we ignore its keypoint scales in the overlap error computation.

4.2. Visual odometry

Our second experiment is performed on two image sequences from *Freiburg* dataset [36]. This experiment is particularly suitable to test whether a given feature is good to track, since it consists in tracking distinctive visual landmarks from one frame to next one and using the matches to compute the observer trajectory (position and orientation). The goal is to estimate the trajectory as close as possible to the given ground-truth measured with GPS and/or inertial sensors.

To estimate the trajectory in this experiment we use the algorithm described in [22]. We provide it with at most 500 keypoints with the highest detector responses at each image, matched through Binary Robust Appearance and Normal descriptor (BRAND) [22]. This descriptor uses jointly the texture and the depth map, and is suitable for keypoints that are not scale-covariant, since it re-estimates the characteristic scale from the depth map. Following the same protocol as in [18], we compute *translation error*, showing how accurately the position is estimated, and *rotation error*, measuring the estimated orientation accuracy. Due to space limitations, we present the results only on two sequences from *Freiburg* in Fig. 3.

On both test sequences, the proposed variant of GFTT is able to achieve smaller or comparable translational and rotational errors to other approaches. Indeed, on *Long office household* sequence the proposed GFTT features are the only able to provide the precision within 12 cm and 4°. On the same sequence, however, the proposed variant of Harris detector turns out to be much less accurate.

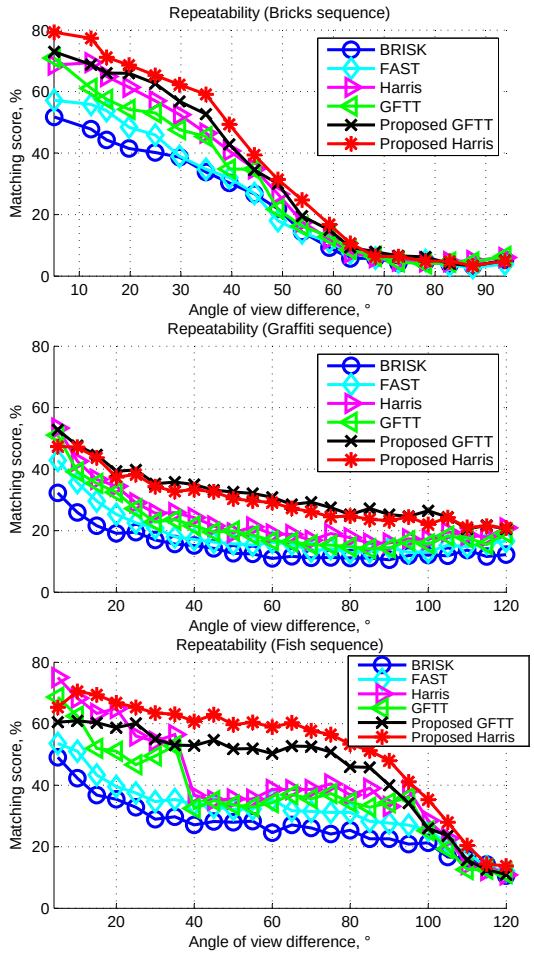

Fig. 2. Keypoints repeatability with different corner detectors on *Bricks*, *Graffiti* and *Fish* sequences used in [16, 19, 20, 35].

The possible reason is a very fast motion of the camera in this sequence, causing a significant directional blur in the texture. In these conditions, the Harris corner test may produce unstable keypoints (the original Harris detector does not perform well neither), whereas the well-conditioning of the second order matrix $\mathcal{M}$ guaranteed by GFTT is able to produce repeatable keypoints.

5. CONCLUSION

This paper describes an efficient general technique for corner detection in texture+depth images based on second-order moment matrix. The proposed approach aims at reliable detection under significant out-of-plane rotations and arbitrary viewpoint changes. These visual deformations induce perspective distortions, which are known to affect the stability of keypoint detection in traditional imaging. The proposed approach is tested with the popular Harris and Good Features to Track corner detectors. Repeatability tests and visual odometry experiments have shown overall improvement of the feature performance. Moreover, the proposed method has linear complexity in function on the image size. In future work, feature invariance under significant scale changes and varying lightning conditions will be addressed.

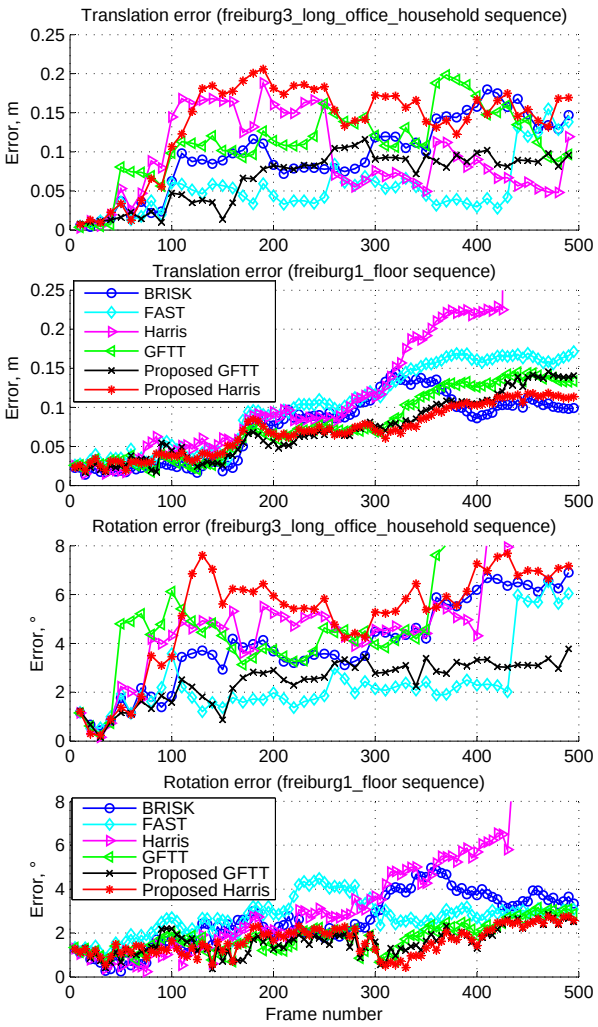

Fig. 3. Visual odometry results on Kinect images from *Freiburg* dataset: translation (top two) and rotation errors (bottom two) on the first 500 frames of *Long office household* and *Floor* sequences respectively.

6. REFERENCES

[1] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 5, pp. 530–534, 1997.

[2] Y.-G. Jiang, J. Yang, C.-W. Ngo, and A. Hauptmann, "Representations of keypoint-based semantic concept detection: A comprehensive study," *IEEE Trans. Multimedia*, vol. 1, no. 12, pp. 42–53, 2010.

[3] U. L. Altintakan and A. Yazici, "Towards effective image classification using class-specific codebooks and distinctive local features," *IEEE Trans. Multimedia*, vol. 17, no. 3, pp. 323–332, 2015.

[4] S. N. Sinha, J.-M. Frahm, M. Pollefeys, and Y. Genc, "Feature tracking and matching in video using programmable graphics hardware," *Machine Vision and Applications*, vol. 22, no. 1, pp. 207–217, 2011.

[5] B. Kitt, A. Geiger, and H. Lategahn, "Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme," in *Proceed. of IEEE Intelligent Vehicles Symposium*, San Diego, CA, USA, 2010.

[6] A. Gil, O. M. Mozos, M. Ballesta, and O. Reinoso, "A comparative evaluation of interest point detectors and local descriptors for visual SLAM," *Machine Vision and Applications*, vol. 21, no. 6, pp. 905–920, 2010.

[7] Y. Li, S. Wang, Q. Tian, and X. Ding, "A survey of recent advances in visual feature detection," *Neurocomputing*, vol. 149, pp. 736–751, 2015.

[8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Intern. J. of Comp. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[9] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comp. Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[10] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proceed. of IEEE Intern. Conf. on Comp. Vision*, Barcelona, Spain, November 2011.

[11] ISO/IEC JTC 1/SC 29/ WG 11, "ISO/IEC CD 15938-13 compact descriptors for visual search," MPEG document N14681, ISO/IEC, Sapporo, Japan, July 2014.

[12] ISO/IEC JTC 1/SC 29/ WG 11, "CDVA: Requirements," MPEG document N14509, ISO/IEC, Valencia, Spain, March 2014.

[13] C. Wu, B. Clipp, X. Li, J.-M. Frahm, and M. Pollefeys, "3D model matching with viewpoint-invariant patches (VIP)," in *Proceed. of IEEE Intern. Conf. on Comp. Vision and Pattern Rec.*, Anchorage, Alaska, USA, June 2008.

[14] K. Koser and R. Koch, "Perspectively invariant normal features," in *Proceed. of IEEE Intern. Conf. on Comp. Vision*, Rio de Janeiro, Brazil, October 2007.

[15] D. Gossow, D. Weikersdorfer, and M. Beetz, "Distinctive texture features from perspective-invariant keypoints," in *Proceed. of IEEE Intern. Conf. on Pattern Rec.*, Tsukuba, Japan, November 2012.

[16] M. Karpushin, G. Valenzise, and F. Dufaux, "Local visual features extraction from texture+depth content based on depth image analysis," in *Proceed. of IEEE Intern. Conf. Image Proc.*, Paris, France, October 2014.

[17] M. Karpushin, G. Valenzise, and F. Dufaux, "Improving distinctiveness of BRISK features using depth maps," in *Proceed. of IEEE Intern. Conf. Image Proc.*, Québec city, Canada, September 2015.

[18] M. Karpushin, G. Valenzise, and F. Dufaux, "Keypoint detection in RGBD images based on an efficient viewpoint-covariant multiscale representation," in *Proceed. of Europ. Sign. Proc. Conf.*, Budapest, Hungary, August 2016, EURASIP.

[19] M. Karpushin, G. Valenzise, and F. Dufaux, "Keypoint detection in rgbd images based on an anisotropic scale space," *IEEE Trans. Multimedia*, vol. 18, no. 9, pp. 1762 – 1771, 2016.

[20] M. Karpushin, G. Valenzise, and F. Dufaux, "A scale space for texture+depth images based on a discrete Laplacian operator," in *IEEE Intern. Conf. on Multimedia and Expo*, Torino, Italy, July 2015.

[21] F. Tombari, S. Salti, and L. Di Stefano, "A combined texture-shape descriptor for enhanced 3D feature matching," in *Proceed. of IEEE Intern. Conf. Image Proc.*, Brussels, Belgium, September 2011.

[22] E. R. do Nascimento, G. L. Oliveira, A. W. Vieira, and M. F. Campos, "On the development of a robust, fast and lightweight keypoint descriptor," *Neurocomputing*, vol. 120, pp. 141–155, 2013.

[23] J. Shi and C. Tomasi, "Good features to track," in *Proceed. of IEEE Intern. Conf. on Comp. Vision and Pattern Rec.*, Seattle WA, USA, June 1994.

[24] C. Harris and M. Stephens, "A combined corner and edge detector.," in *Alvey vision conference*, Manchester, UK, 1988, vol. 15, p. 50.

[25] I. Sipiran and B. Bustos, "Harris 3D: a robust extension of the harris operator for interest point detection on 3D meshes," *The Visual Computer*, vol. 27, no. 11, pp. 963–976, 2011.

[26] T. Tommasini, A. Fusiello, E. Trucco, and V. Roberto, "Making good features track better," in *Proceed. of IEEE Intern. Conf. on Comp. Vision and Pattern Rec.*, Santa Barbara, CA, USA, June 1998.

[27] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an efficient alternative to SIFT or SURF," in *Proceed. of IEEE Intern. Conf. on Comp. Vision*, Barcelona, Spain, November 2011.

[28] S. M. Smith and J. M. Brady, "SUSAN – a new approach to low level image processing," *Intern. J. of Comp. Vision*, vol. 23, no. 1, pp. 45–78, 1997.

[29] E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking," in *Proceed. of IEEE Intern. Conf. on Comp. Vision*, Beijing, China, October 2005.

[30] E. Mair, G. D. Hager, D. Burschka, M. Suppa, and G. Hirzinger, "Adaptive and generic corner detection based on the accelerated segment test," in *Proceed. of Europ. Conf. on Comp. Vision*, Crete, Greece, September 2010, Springer.

[31] R. B. Rusu and S. Cousins, "3D is here: Point cloud library (PCL)," in *Proceed. of IEEE Intern. Conf. on Rob. and Autom.*, Shanghai, China, May 2011.

[32] F. C. Crow, "Summed-area tables for texture mapping," *ACM SIGGRAPH computer graphics*, vol. 18, no. 3, pp. 207–212, 1984.

[33] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *Intern. J. of Comp. Vision*, vol. 65, no. 1-2, pp. 43–72, 2005.

[34] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, no. 10, pp. 1615–1630, 2005.

[35] M. Karpushin, G. Valenzise, and F. Dufaux, "An image smoothing operator for fast and accurate scale space approximation," in *Proceed. of IEEE Intern. Conf. Acoust., Speech and Sign. Proc.*, Shanghai, China, March 2016.

[36] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *Proc. of the Intern. Conf. on Intelligent Robot Systems*, Vilamoura, Algarve, Portugal, October 2012.