Quasi-Random Scale Space Approach to Robust Keypoint Extraction in High-Noise Environments

Alexander Wong, Akshaya Mishra, David A. Clausi, and Paul Fieguth
Vision and Image Processing (VIP) Research Group
Department of Systems Design Engineering
University of Waterloo, Waterloo, Canada
{a28wong,akmishra,dclausi,pfieguth}@uwaterloo.ca

Abstract

A novel multi-scale approach is presented for the purpose of robust keypoint extraction in high-noise environments. A multi-scale representation of the noisy scene is computed using quasi-random scale space theory. A gradient second-order moment analysis is employed at each quasi-random scale to identify initial keypoint candidates. Final keypoints and their characteristic scales are selected based on the local Hessian trace extrema over all quasi-random scales. The proposed keypoint extraction method is designed to reduce noise sensitivity by taking advantage of the structural localization and noise robustness gained through the use of quasi-random scale space theory. Experimental results using scenes under different high noise conditions, as well as real synthetic aperture sonar imagery, show the effectiveness of the proposed method for noise robust keypoint extraction when compared to existing keypoint extraction techniques.

1 Introduction

An important task in computer vision is the extraction of distinctive keypoints of interest from scenes, which is crucial in a wide variety of applications such as image registration [1, 2], object recognition [3, 4], and 3D reconstruction [5]. An ongoing challenge in the design of keypoint extraction algorithms is dealing with images and videos acquired under the presence of noise. For example, surveillance videos acquired in outdoor environments at night are often contaminated with a high level of shot noise due to poor lighting conditions. Another example of noise contamination in a real-world scenario is that of underwater images and videos acquired using sonar systems, which are contaminated by a high level of speckle noise due to the interference of backscatter signals. The presence of noise in the acquired scene is highly undesirable from the perspective of keypoint extraction, as noise possesses distinctive structural characteristics that makes the differentiation between keypoints and image noise a challenging task. Therefore, it is highly desirable to identify methods that are capable of reliably extracting unique keypoints of interest in high-noise environments.

Keypoint extraction methods can be generally divided into two main categories: i) single-scale methods, and ii) multi-scale methods. In single-scale methods [6, 7, 8], distinctive keypoints are extracted directly from the acquired scene. For example, the Harris keypoint extraction method [6] computes structural distinctiveness based on the second order moments extracted from the image, and identifies the local maxima of the computed metric as keypoints in the image. There are two main limitations associated with single-scale methods. First, such methods are highly sensitive to the presence of noise, resulting in the detection of numerous erroneous keypoints not associated with the actual structural characteristics of the scene. Second, such methods have no sense of scale, hence often producing poorly localized keypoints, and even missing important keypoints, in scenes characterized by structures and objects at different scales.

In multi-scale methods [9, 10, 11, 12], a scene is first decomposed into a multi-scale representation, and distinctive keypoints are then identified, along with their characteristic scales, based on a metric of structural distinctiveness across all scales. For example, in the approach proposed by Lowe [10], the scene is first decomposed into a Gaussian scale space representation, and identifies the keypoints as the local optima of the difference of Gaussians function applied to the scale space representation. There are two main benefits to existing multi-scale methods. First, such methods produce keypoints with better feature localization than single-scale methods for structures at different scales, making them more suited for scenes characterized by multi-

978-0-7695-4040-5/10 $26.00 © 2010 IEEE
DOI 10.1109/CRV.2010.11

IEEE computer society

scale characteristics. Second, such methods allow for the identification of the characteristic scale at each keypoint, which is very important for matching features captured at different scales.

While addressing the issues associated with multi-scale feature localization, existing multi-scale methods are often limited in their ability to handle high noise situations such as underwater sonar imaging. A major factor contributing to these limitations is the use of Gaussian scale space, which has been shown to produce unsatisfactory scale space representations of images characterized by high noise levels [14]. The underlying goal of the proposed keypoint extraction method is to address the issues associated with noise through the use of quasi-random scale space theory [14], which was shown to provide better structural localization across all scales compared to existing scale space formulations, especially under high noise scenarios. This paper is organized as follows. The proposed quasi-random scale space (QRS) keypoint extraction method is presented in Section 2. Experimental results are presented and discussed in Section 4. Finally, conclusions are drawn and future work is discussed in Section 5.

2 QRS keypoint extraction method

The QRS keypoint extraction method is centered around quasi-random scale space theory [14] and is comprised of the three main steps. First, a quasi-random scale space representation is computed from the scene in question. Second, structural distinctiveness is computed at each quasi-random scale using a gradient second order moment analysis to identify a initial set of keypoint candidates. Third, the final set of keypoints is determined on the local Hessian trace extrema at each keypoint candidate across all quasi-random scales. Each of the steps are described in detail in the subsequent sections.

2.1 Quasi-random scale space

In the first step of the QRS keypoint extract method, a multi-scale representation of the scene in question is computed. In existing multi-scale methods, such a multi-scale scene representation is computed based on Gaussian scale space theory. However, it has been shown that such an approach produces multi-scale representations that are sensitive to high noise conditions [14], as well as provide noticeable structural delocalization at coarser scales [13]. To address this issue, we propose to instead use quasi-random scale space theory [14] to compute the multi-scale scene representation, which can be described as follows.

Let S be set of sites in a discrete lattice $\pounds$ upon which the scene is defined and $s \in S$ be a site in $\pounds$. Further, let the acquired scene $I = \{I(s)|s \in S\}$, gradient

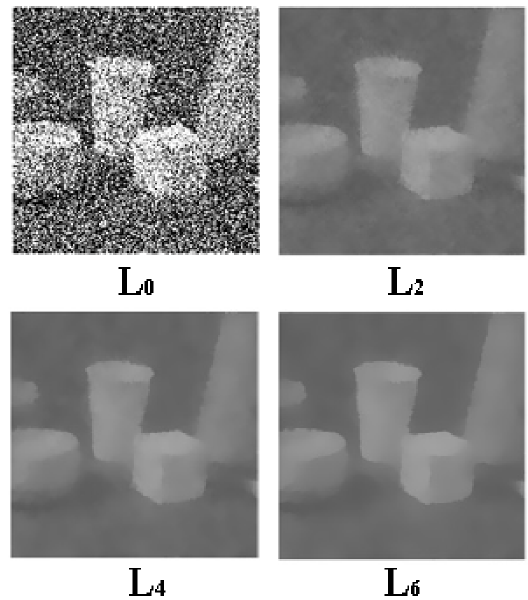

Figure 1. An example of a multi-scale representation computed using quasi-random scale space theory. The produced representation exhibits strong structural localization at all scales, as well as low noise sensitivity.

$G_i = \{G_i(s)|s \in S\}$, scale space representation $L_i = \{L_i(s)|s \in S\}$, and residual fine scale structure $C_i = \{C_i(s)|s \in S\}$ be random fields on S. Initializing with $L_0(s) = I(s)$, the scale space decomposition can be expressed s the following recursion relationship,

$$L_{i-1}(s) = L_i(s) + C_i(s), \tag{1}$$

where C is interpreted as the inter-scale residual. Formulating the computation of $L_i(s)$ as an inverse problem, where the "measurement", and "noise" are represented by $L_{i-1}(s)$ and $C_i(s)$ respectively, we can estimate the "state" $L_i(s)$ as a Bayesian least-squares estimate,

$$\hat{L}_i(s) = \arg_{\hat{L}_i} \min \left\{ E\left(\left(\hat{L}_i(s) - L_i(s) \right)^2 | L_{i-1}(s) \right) \right\}. \tag{2}$$

Given Eq. 2, the analytical solution for $L_i(s)$ can be expressed as

$$\hat{L}_i(s) = \underbrace{\int L_i(s)\, p\left(L_i(s)\, | L_{i-1}(s)\right) dL_i(s)}_{E(L_i(s)|L_{i-1}(s))}. \tag{3}$$

What Eq. 2 implies is that the optimal estimate of scale space representation $L_i(s)$ is the mean conditioned on the previous scale space representation $L_{i-1}(s)$.

To compute $L_i(s)$ based on Eq. 2, it is necessary to compute the posterior distribution $p\left(L_i(s)|L_{i-1}(s)\right)$, which is

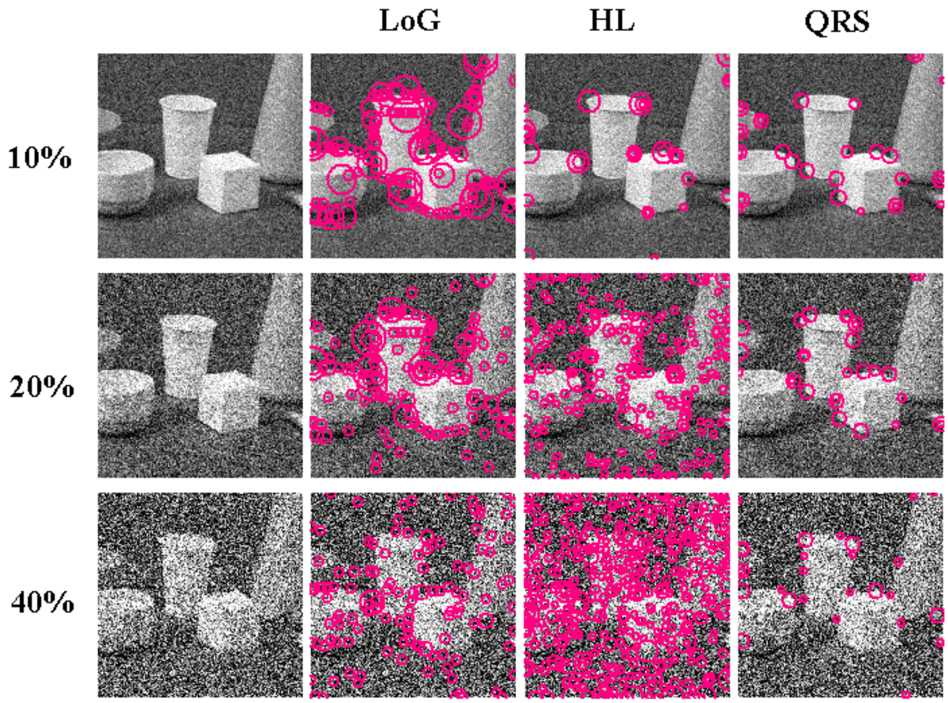

Figure 2. Keypoint extraction results for TEST1 at different noise levels.

difficult to do given the complex, nonlinear nature of images. To compute $p\left(L_i(s)|L_{i-1}(s)\right)$ in an accurate yet efficient manner, a quasi-random density estimation approach is employed. First, n samples are drawn from a Sobol quasi-random sequence [15] with respect to site s at scale i, which promotes low discrepancy samples to be drawn. To use only samples with high relevancy in the posterior estimation, a Gaussian mixture model is fitted to the distribution $p(L_i(s))$ and all samples within the Gaussian distribution to which $L_{i-1}(s)$ belongs are accepted as the realizable sample set of $p\left(L_i(s)|L_{i-1}(s)\right)$.

Given the selected sample set Ω, the posterior distribution $\hat{p}\left(L_i(s)|L_{i-1}(s)\right)$ is estimated as

$$\hat{p}\left(L_i(s)|L_{i-1}(s)\right) = \frac{p^*\left(L_i(s)|L_{i-1}(s)\right)}{\int_0^1 p^*\left(L_i(s)|L_{i-1}(s)\right) dL_i(s)}, \quad (4)$$

where p^* can be expressed as

$$p^*\left(L_i(s)|L_{i-1}(s)\right) = \frac{1}{\sqrt{2\pi}\sigma_{L_i}} \sum_{k=\Omega} f_1(k)\, f_2(k)\, f_3(k) \cdot \exp\left(-\frac{1}{2}\left(\frac{L_i - L_{i-1}(s_k)}{\sigma_{L_{i-1}}}\right)^2\right), \quad (5)$$

where $f_1(k)$, $f_2(k)$, and $f_3(k)$ are objective functions, assessing sample relevance, on the basis of intensity, gradient, and spatial offset respectively [14].

An example of a multi-scale representation computed using quasi-random scale space theory is shown in Fig. 1; observe particularly the strong structural localization and low noise sensitivity.

2.2 Keypoint candidate selection

In the second step of the QRS keypoint extract method, a set of keypoint candidates $\{p'_1, p'_2, \ldots, p'_n\}$ is selected at each quasi-random scale. A widely accepted approach to identifying suitable keypoints of interest is to assess the structural distinctiveness of local neighborhoods around sites in a scene. In QRS, a gradient second order moment analysis approach is employed and can be described as follows. For each quasi-random scale L_i, the gradient second order moment matrix at a particular site s can be defined as

$$\Phi_{L_i}(s) = \begin{bmatrix} \{\Delta_x L_i(s)\}^2 & \{\Delta_x L_i(s)\}\{\Delta_y L_i(s)\} \\ \{\Delta_x L_i(s)\}\{\Delta_y L_i(s)\} & \{\Delta_y L_i(s)\}^2 \end{bmatrix} \quad (6)$$

where Δ_x and Δ_y are the gradients along the x and y-directions, respectively. Based on $\Phi_{L_i}(s)$, the structural distinctiveness at site s is computed based on the Noble response metric [16], which provides a reliable indication of significant change in the orthogonal directions,

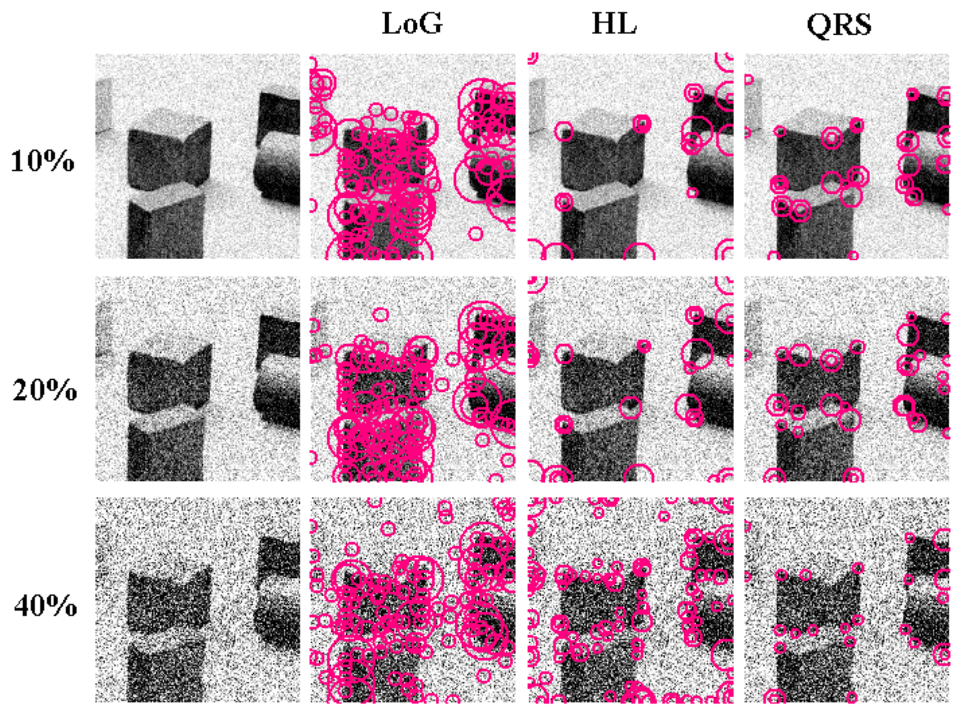

Figure 3. Keypoint extraction results for TEST2 at different noise levels.

$$\rho_{L_i}(s) = \frac{det\left(\Phi_{L_i}(s)\right)}{trace\left(\Phi_{L_i}(s)\right)}, \quad (7)$$

where det and $trace$ denote the determinant and trace of a matrix. The local maxima of $\rho_{L_i}(s)$ are selected as keypoint candidates, with L_i selected as their characteristic scale.

3 Final keypoint selection

In the third and final step of QRS, the final set of keypoints $\{p_1, p_2, \ldots, p_m\}$ is determined based on the initial set of keypoint candidates $\{p'_1, p'_2, \ldots, p'_n\}$. A widely accepted approach to selecting the final set of keypoints is to identify the keypoint candidates whose characteristic scales attain an extremum for some objective function over all scales. A comprehensive experimental study conducted by Mikolajczyk and Schmid [17] found that the Hessian trace function provided the most reliable selection of correct keypoints and their characteristic scales. Motivated by this, in QRS, the final set of keypoints $\{p_1, p_2, \ldots, p_m\}$ are selected as all keypoint candidates whose characteristic scales attain a local Hessian trace extremum over all quasi-random scales.

4 Experimental Results

To evaluate the effectiveness of the proposed QRS keypoint extraction method, two sets of experiments were performed. The first set of experiments involves investigating the effectiveness of QRS at detecting distinctive keypoints at different noise levels. To achieve this goal, three noise-free test images (one synthetic image with structures at different scales, and two real-world scenes with objects of different scales) were contaminated by Gaussian noise with standard deviations $\sigma = \{10\%, 20\%, 40\%\}$ of the image dynamic range, and keypoints were then extracted from the noise-comtaminated images.

The second set of experiments involves investigating the effectiveness of QRS at detecting distinctive keypoints in real-world high-noise environments. To achieve this goal, keypoints were extracted from four underwater scenes from EdgeTech (West Wareham, Massachusetts) acquired using side-scan synthetic aperture sonar systems. Due to the interference of backscatter signals during the sonar imaging process, the underwater scenes are contaminated by high levels of multiplicative speckle noise. Extraction of keypoints from sonar imagery is important for applications such as underwater tracking as well as 3D reconstruction of sunken objects.

For comparison purposes, the Laplacian of Gaussian (LoG) method proposed by Lindeberg [9] and the Harris-

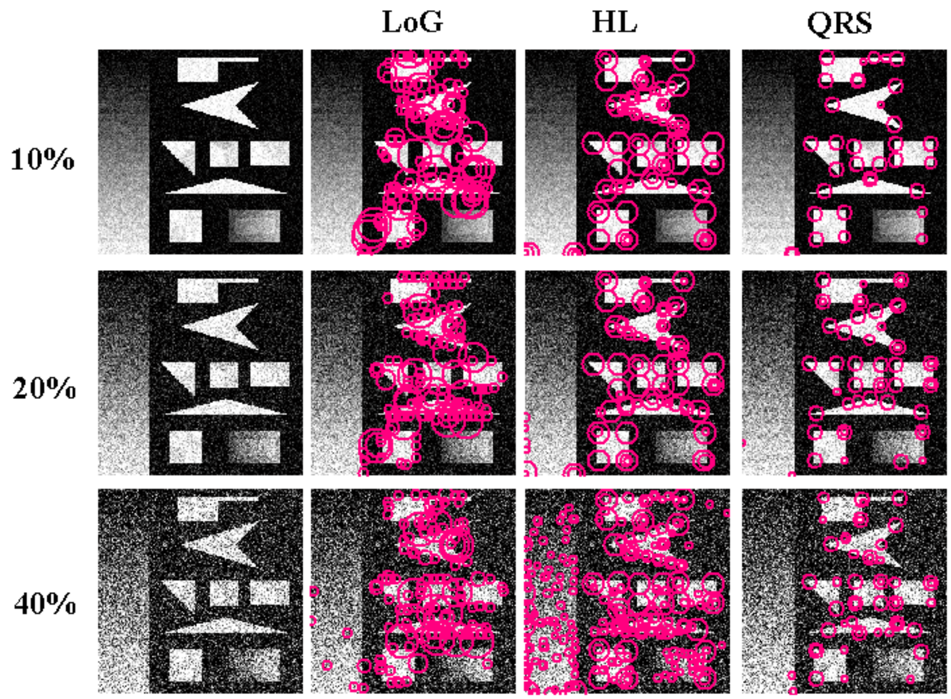

Figure 4. Keypoint extraction results for TEST3 at different noise levels.

Laplacian (HL) method proposed by Mikolajczyk and Schmid [12] were also tested. All of the tested methods are implemented using the parameters proposed in the original works. The QRS keypoint extraction method was implemented using 8 quasi-random scales.

The extracted keypoints and their corresponding characteristic scales extracted by each tested method for the three noise-contaminated test images at different noise levels are shown in Figs. 2-4. Several observations can be made from the keypoint extraction results. QRS is able to detect noticeably more distinctive keypoints of interest than HL for TEST1 and TEST2, while producing fewer erroneous keypoints than HL for TEST3 under the $\sigma = 10\%$ case. Under high noise scenarios ($\sigma = \{20\%, 40\%\}$), QRS produces noticeably fewer erroneous keypoints than HL for all test images. When compared to LoG, QRS produces noticeably fewer erroneous keypoints under all noise scenarios for all test images. These results indicate the effectiveness of QRS at extracting distinctive keypoints of interest under high-noise environments.

The extracted keypoints and their corresponding characteristic scales extracted by each tested method for the four side-scan synthetic aperture sonar images are shown in Fig. 5. QRS is able to detect distinctive keypoints while producing significantly fewer erroneous keypoints than HL and LoG for all four sonar images. These results indicate the effectiveness of QRS at extracting distinctive keypoints of interest under real-world high-noise environments.

5 Conclusions

A novel keypoint extraction method named QRS based on quasi-random scale space theory is presented. Based on the quasi-random scales space representation computed from the scene in question, a gradient second order moment analysis is employed at each scale to identify an initial set of keypoint candidates. A local Hessian trace extrema detection approach is then taken to identify the final set of keypoints and their characteristic scales. Experimental results involving scenes under different noise levels as well as real synthetic aperture sonar imagees demonstrate the noise robustness of QRS when compared to existing keypoint extraction methods. Future work involves a more thorough performance analysis of QRS, as well as investigating strategies for extending QRS for affine and illumination invariant keypoint extraction.

Acknowledgment

This research has been sponsored by the Natural Sciences and Engineering Research Council (NSERC) of Canada through individual Discovery Grants as well as GEOIDE (GEOmatics for Informed Decisions) which is a

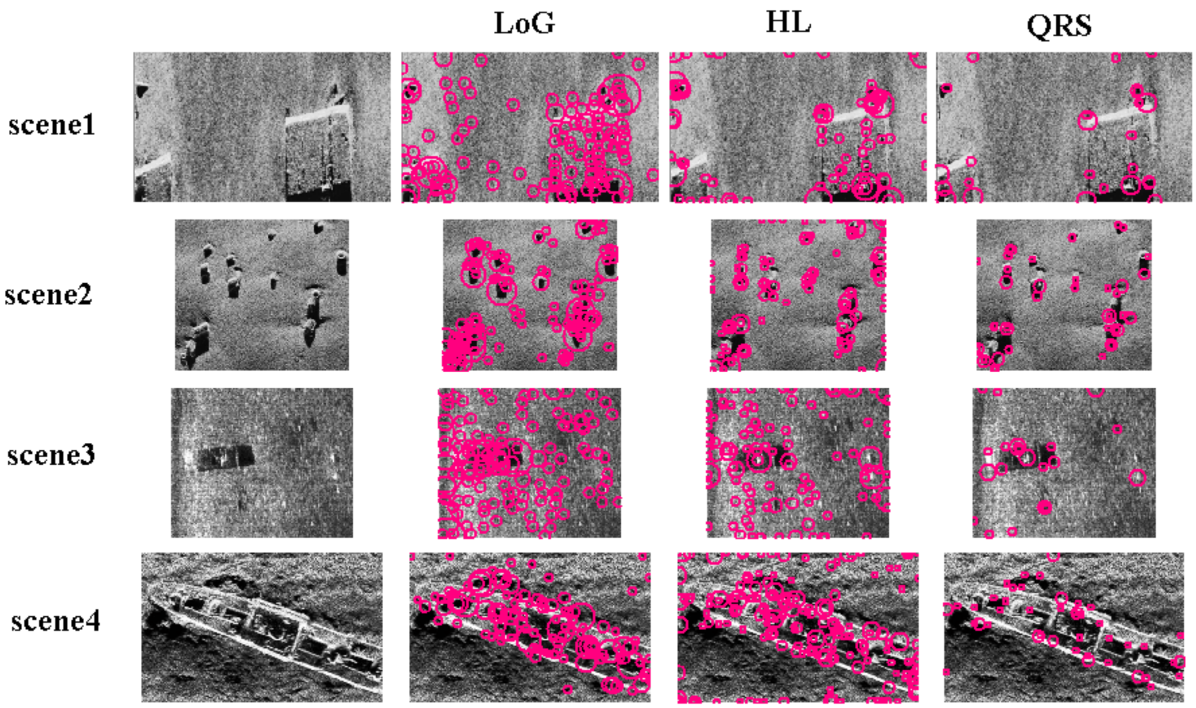

Figure 5. Keypoint extraction results for real synthetic aperture sonar imagery.

Network of Centres of Excellence under NSERC. We thank EdgeTech for providing us with the side-sonar synthetic aperture sonar images used in our testing.

References

[1] A. Wong and D.A. Clausi, "ARRSI: automatic registration of remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 45, no. 5, pp. 1483-1493, 2007.

[2] C. Tsai, C. Li, G. Yang, and K. Lin, "The Edge-Driven Dual-Bootstrap Iterative Closest Point Algorithm for Registration of Multimodal Fluorescein Angiogram Sequence," *IEEE Transactions on Medical Imaging*, 2009.

[3] D. Lowe, "Local feature view clustering for 3D object recognition," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Kauai, Hawaii, pp. 682-688, Dec 2001.

[4] G. Csurka, C. Dance, L. Fan, J. Willamowsky, and C. Bray. Visual categorization with bags of keypoints. In Int. Work. on Statistical Learning in Comp. Vis., ECCV, 2004.

[5] C. Wu, B. Clipp, X. Li, J. Frahm, and M. Pollefeys, "3D Model Matching with Viewpoint-Invariant Patches (VIP)," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, Alaska, 2008.

[6] C. Harris, "Determination of ego-motion from matched pointsk," *Proc. Alvey Vision Conf.*, Cambridge, UK, 1987.

[7] L. Kitchen and A. Rosenfeld, "Gray level corner detection," *Pattern Recognition Letters*, pp. 95-102, 1982.

[8] S. Smith and J. Brady, "SUSANA new approach to low-level image processing," *International Journal of Computer Vision*, vol. 23, no. 1, pp. 45-48, 1997.

[9] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision* vol. 30, no. 2, pp. 79-116, 1998.

[10] D. Lowe, "Object recognition from local scale-invariant features," *Proc. 7th International Conference on Computer Vision*, Kerkyra, Greece, pp. 1150-1157, 1999.

[11] Y. Dufournaud, C. Schmid, and R. Horaud, "Matching images with different resolutions," *Proc. IEEE Conference on Computer Vision and Pattern Recognition* Hilton Head Island, South Carolina, USA, pp. 612-618, 2000.

[12] K. Mikolajczyk and C. Schmid, "Scale & Affine Invariant Interest Point Detectors," *International Journal of Computer Vision,* vol. 60, no. 1, pp. 63-86, 2004.

[13] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 12, no. 7, pp. 629-639, 1990.

[14] A. Mishra, A. Wong, D.A. Clausi, and P. Fieguth, "Quasi-random nonlinear scale space," *Pattern Recognition Letters,* (accepted), 2010.

[15] I. Sobol, "Uniformly Distributed Sequences with an Additional Uniform Property," *USSR Computational Mathematics and Mathematical Physics*, vol. 16, pp. 236-242, 1977.

[16] A. Noble, "Descriptions of Image Surfaces," PhD thesis, Oxford University, 1989.

[17] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points," *Proc. 8th International Conference on Computer Vision*, Vancouver, Canada, pp. 525- 531, 2001.