Superpixel Finite Element Segmentation for RGB-D Images

Superpixel Finite Element Segmentation for RGB-D Images

D. Kerr, S. A. Coleman
School of Computing and Intelligent Systems,
University of Ulster, Magee College,
Londonderry, N. Ireland, U.K.
e-mail: {d.kerr, sa.coleman}@ulster.ac.uk

B. W. Scotney
School of Computing and Information Engineering, University of Ulster, Coleraine,
N. Ireland, UK
e-mail: bw.scotney@ulster.ac.uk

This publication appears in:

2017 2nd Asia-Pacific Conference on Intelligent Robot Systems (ACIRS 2007)

2017 3rd International Conference on Robotics and Vision (ICRV 2017)

June 16-18, 2017

Wuhan, China.

Superpixel Finite Element Segmentation for RGB-D Images

D. Kerr, S. A. Coleman
School of Computing and Intelligent Systems,
University of Ulster, Magee College,
Londonderry, N. Ireland, U.K.
e-mail: {d.kerr, sa.coleman}@ulster.ac.uk

B. W. Scotney
School of Computing and Information Engineering,
University of Ulster, Coleraine,
N. Ireland, UK
e-mail: bw.scotney@ulster.ac.uk

***Abstract*—Computer vision research has advanced from focusing solely on intensity images to the use of depth images, or combinations of RGB, intensity and depth images, mainly due to the recent development of low cost depth cameras. These images can be efficiently represented as a space-variant image by segmenting the images using a superpixel representation. Whilst superpixel representations offer advantages in terms of reduced processing requirements they present challenges in further processing as many existing image processing techniques require regularly distributed image data. We overcome this issue by making use of the Finite element framework for processing these images and demonstrate the application of the technique for detecting access holes in disaster management situations.**

Keywords-RGB-D imaging; image segmentation; SLIC; finite element framework; feature detection

I. INTRODUCTION

The recent popularity of low cost depth sensing technology such as the Microsoft Kinect or ASUS Wavi-Xtion has been driven by the integration of this imaging technology into consumer gaming technology. Low cost and widespread availability of these imaging devices has seen their use go beyond the original gaming application areas to general imaging research devices for computer vision and robotic applications. The ability of the devices to capture not only depth but also colour (RGB) and intensity images makes them a popular platform for many computer vision application areas. These imaging devices are commonly referred to as RGB-D to signify they are capable of **D**epth sensing and **R**ed, **G**reen, **B**lue colour sensing.

Depth sensing technologies have been used previously in areas as diverse as robot navigation, surveillance, object detection and recognition and human interaction [1, 2, 3, 4]. These domains regard depth imaging as a particularly important feature as it can be effectively used to obtain a reliable 3-D description of a scene, an important benefit when considering the diverse application areas. Technology for obtaining 3-D descriptions of a scene is not a recent innovation, as the idea of creating a depth image has been experimented with for a number of years, and many different technologies have been investigated in order to obtain a depth description of a scene or object. However, the introduction of consumer level depth sensing technology has provided researchers with a low-cost depth sensing technology which they are now exploiting in many unimagined ways.

Prior to the introduction of RGB-D devices depth sensing technology was known as range imaging, referring to the fact that the image contained the distance (or range) from the camera (or imaging device) to the imaged point in the scene. The literature contains many references to depth sensing and range imaging; however, we will use the term depth imaging in line with the use of the modern depth image capture technologies.

A depth image is regarded as a 2-D image with each pixel location containing a distance measurement rather that a pixel intensity. Distances are estimated from the imaging device to the surface points or objects within the imaged scene [5]. When using a device such as a Kinect to capture RGB-D images, the main advantage offered over a traditional camera is that it provides additional information in the form of depth measurements, thereby providing more information within the scene to be recovered [6]. It is important to remember that when discussing depth sensing technology such as this, that any single depth image only contains information about the surfaces of the scene visible from the imaging device. Therefore, as a single depth image is not capable of representing a 3-dimensional scene the information in these images is often referred to as 2½-D information [5].

Image segmentation is a common image processing technique which partitions a digital image into multiple segments. The purpose of image segmentation is to subdivide an image into its constituent regions or objects thereby assigning a label to every pixel in an image. The subdivision is based on the principle that pixels with the same label should share certain visual characteristics, for example distinct regions to correlate strongly with objects or features of interest. The level to which the subdivision is performed is application specific. Many techniques have been proposed over the years to segment an image including thresholding, region growing, morphological methods, clustering, graph-based methods, shape-based methods and machine learning. More recently, superpixel-based segmentation techniques have become popular due to the numerous benefits they offer in terms of efficient image representations. Superpixels may be used as an efficient primitive from which local image features can be computed [7]. The basis of the idea behind superpixels is that the rectilinear pixel-grid commonly used to represent a digital image is not a natural representation of a visual scene. The pixel grid is considered an "artifact" of the digital imaging process and superpixels are seen as a way of providing a more natural and efficient image representation and processing framework on which subsequent processing operations can be performed [7].

In this paper, we present a novel framework for efficient processing of RGB-D images, using superpixel segmentation and a finite element-based image processing framework to extract domain specific image features in a perceptually meaningful and computationally efficient manner. We make use of the access hole RGB-D dataset [8] to demonstrate the operation of our technique. This paper is organised as follows: Section II contains an overview of superpixel segmentation and details of the specific superpixel image representations used in the work presented here. Section III discusses the finite-element based image processing framework used to extract image features and Section IV presents experimental results for the proposed technique. Finally, in section V conclusions are presented and future work is discussed.

II. Superpixel Image Representation

An alternative to pixels as image primitives are superpixels which provide an efficient and semantically meaningful image primitive from which local image features may be computed. Superpixels have an advantage over pixels in that they reduce the complexity of subsequent image processing methods by capturing the redundancy found in digital images. They have demonstrated their applicability in a number of areas including segmentation, depth estimation, and localisation. The requirement for superpixels to be used in subsequent image processing operations means that the superpixel segmentation technique must be fast, reliable and produce robust segmentations. The theory behind superpixel image representations is that when considering a digital image Figure 1 (a), individual pixels, Figure 1(b) provide little descriptive or semantic information when considered in isolation. Thus, superpixels seek to provide a more efficient image representation on which to base further image processing operations.

Figure 1. (a) Digital image (b)Pixel When considering a digital image (a) individual pixels (b) provide little descriptive or semantic information when considered in isolation.

The majority of superpixel methods [9, 10, 11, 12, 13] suffer from a number of deficiencies including high computational cost, poor quality segmentation, inconsistent size and shape, and multiple parameters. However, the recently developed simple linear iterative clustering (SLIC) technique [7] adapts a k-means clustering approach to efficiently perform a local clustering of pixels using a 5-D space that includes colour and pixel coordinates. The inclusion of the pixel coordinates enforces compactness and regularity in the superpixel shape.

With SLIC the image is first divided into a grid approximated with the desired amount of superpixels. The centre of each grid is then used to initialise a corresponding k-means, and the k-means centres and clusters are then refined by using the Lloyd algorithm [14], which find evenly spaced sets of points in subsets of Euclidean spaces and thus partition the image into well-shaped and uniformly sized convex cells. An example of such a SLIC superpixel segmentation is illustrated in Figure 2.

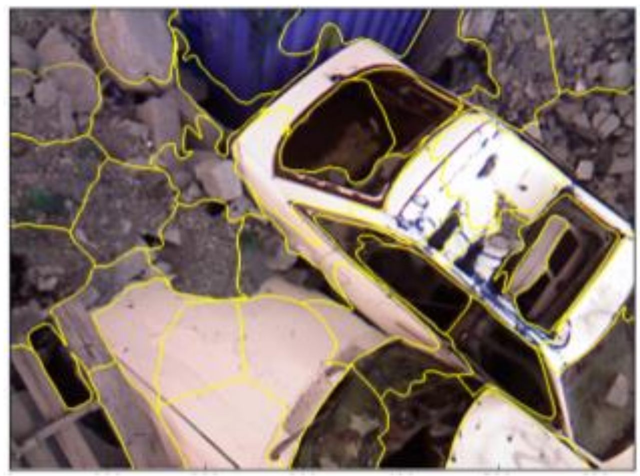

Figure 2. SLIC superpixel segmentation of RGB image using an initial number of 50 superpixels.

Figure 3. SLIC superpixel segmentation of RGB image with average RGB values within each of the 50 superpixel regions computed.

The approach developed in this paper makes use of the SLIC superpixel segmentation approach to segment RGB-D images into semantically meaningful regions for subsequent processing. We segment both the RGB image (as illustrated in Figure 2) and the depth image using this approach for subsequent processing. Once the image has been segmented the average pixel value within the region may be computed and used for subsequent processing operations as illustrated in Figure 3. Similar segmentations using the depth image are illustrated in Figure 4, and Figure 5 presents the averaging process for depth images. The superpixel segmentation approach also has the advantage in that locations with missing data in the depth image may be approximated using the average depth value from within the superpixel region resulting in a hole filling process.

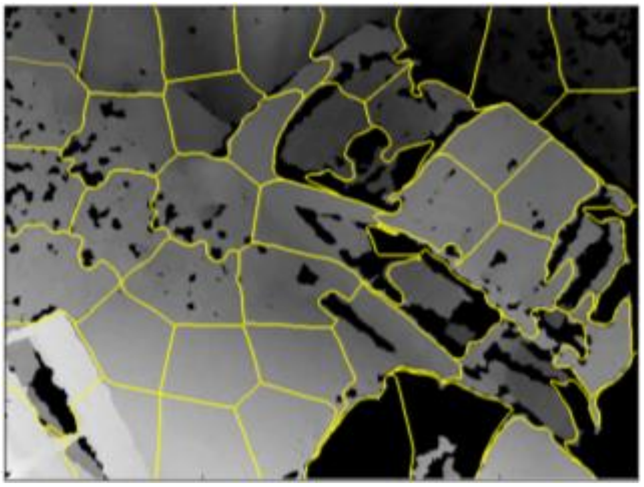

Figure 4. SLIC superpixel segmentation of depth image using an initial number of 50 superpixels.

Figure 5. SLIC superpixel segmentation of depth image with average RGB values within each of the 50 superpixel regions computed.

In addition to using superpixel segmentations of approximately 50 regions we also use segmentations of 100, 500, and 1000 regions respectively.

As previously mentioned superpixels have the advantage of reducing image dimensionality and processing requirements. If we consider the original RGB-D images to have a resolution of 640 $\times$ 480 pixels in each RGB and depth image then each image is composed of a total of 307200 individual pixels. In the case of using an image segmented into 50 superpixel regions the image can now be effectively represented using only 50 superpixels, or 1000 superpixels in the case of a segmentation of 1000 regions. Whilst beneficial in terms of reducing processing requirements, superpixel representations present challenges in that many conventional image processing techniques are not suited to this image representation.

III. AUTONOMOUS FEATURE EXTRACTION OPERATORS

In order to process further such superpixel image representation, consideration has to be given to the irregular distribution of superpixel regions. Here we present a novel framework where a superpixel segmented image is considered as a space-variant image in which the unweighted spatial centre of each superpixel region corresponds to a node in the space variant image. With each node we associate a single pixel value using the average pixel value, calculated from all the pixel values within the corresponding original superpixel region.

Next, we define as nodes those space-variant pixels so as the image plane may be triangulated using these nodes as the triangle vertices in an (irregular) mesh. In practice, we use Delaunay triangulation to obtain this triangulation [15]. This is illustrated in Figure 6: a SLIC superpixel segmentation has the space-variant nodes indicated in blue (corresponding to superpixel region centres). With each of these nodes the average pixel value from within the region is associated. We then perform the Delaunay triangulation of the image plane using the nodes as the triangle vertices in an irregular mesh. In Figure 6 the Delaunay triangulation has the edges indicated in white.

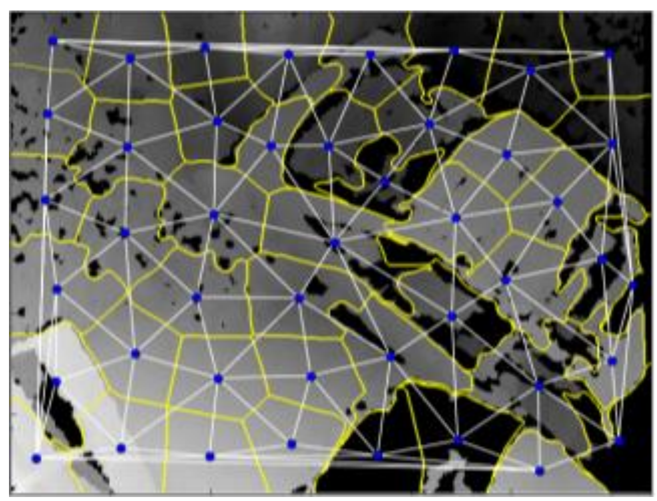

Figure 6. SLIC superpixel segmentation with nodes (indicated in blue) corresponding to superpixel region centres and Delaunay triangulation of the image plane using the nodes as the triangle vertices in an irregular mesh.

We then use a family of autonomous finite element based image processing operators [16, 17] presented in [17] for use on non-uniformly sampled intensity images and in [16] for use on range images. Here, the term autonomous indicates that these operators were developed in such a way that they can change size and shape across the image plane in accordance with the space-variant pixel distribution as illustrated in Figure 7. Our image processing operators are then constructed over polygonal neighbourhoods Ω_i, centred on each node *i* and comprising those triangles that share node *i* as a vertex; neighbourhoods are thus adaptive in size and shape, as illustrated in Figure 7.

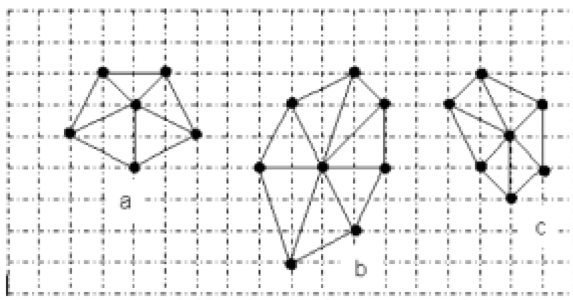

Figure 7. Neighbourhoods of space variant operators.

Construction of space variant operators on such neighbourhoods differs from that of image processing operators on a regular rectangular grid in that it is no longer appropriate to build explicitly an entire operator; each operator throughout the irregular mesh may be different with respect to the operator neighbourhood size, shape, and the number of nodal points in the operator. When using the polygonal neighbourhoods, we work on an element-by-element basis, taking advantage of the flexibility offered by the finite element method as a means of adaptively changing the irregular operator size and shape to encompass the data available in any local neighbourhood. A finite element based approach is used to define and construct the detectors: in each neighbourhood, the detectors are built on weak forms of operators that are based on first order derivative approximations on each triangular element. On each element e_m within neighbourhood Ω_i, the functional is defined as

$$E_i^{e_m}(U) = \int_{e_m} \underline{b}_i \cdot \underline{\nabla} U \psi_i^{\sigma_m} \, d\Omega \tag{1}$$

where *U* is the representation of the image data, and $\underline{b}_i$ is an image-dependent unit vector. The function $\psi_i^{\sigma_m}$ is chosen to be a Gaussian function restricted to the element e_m, and defined by

$$\psi_i^{\sigma_m}(x,y) = \frac{1}{2\pi\sigma_m^{\;2}} e^{-\left(\frac{(x-x_i)^2+(y-y_i)^2}{2\sigma_m^{\;2}}\right)} \tag{2}$$

By defining the longest edge of element e_m from the node *i* as l_{e_m} the value of σ_m is chosen to on any element e_m is given as $l_{e_m}/1.96$. For each element e_m within neighbourhood Ω_i, the spatial relationship between the node *i* on which the Gaussian function $\psi_i^{\sigma_m}$ is centred and each of the nodes in the element is readily available from nodal numbering and locational information routinely stored in the finite element method. A complete first derivative operator over the neighbourhood Ω_i is thus achieved by the process of finite element assembly of the operators $E_i^{e_m}$ for all of the elements e_m contained in Ω_i. The integral required to compute each $E_i^{e_m}$ is efficiently evaluated by the use of isoparametric mappings that relate each element to a "standard" right-angled triangle on which numerical integration can be performed accurately using a Gaussian quadrature rule that just requires function evaluations of $\psi_i^{\sigma_m}$ and of the basis functions used to represent the image *U* within e_m (locally indexed for the three nodes of the element). If the image representation on each element is linear, then the gradient $\underline{\nabla} U$ is locally constant, and each element integral may be accurately approximated by just four function evaluations of $\psi_i^{\sigma_m}$ (i.e., using a four-point Gauss rule). Thus, through this finite element based approach, the operator is able to automatically alter its shape and size as required, as illustrated by the three different operators in Figure 7. Operator a has a central node with 5 adjoining nodes, operator b illustrates 7 adjoining nodes and operator c illustrates 6 adjoining nodes.

Local first order derivative operators *X* and *Y* along the *x*- and *y*-coordinate directions respectively are generated by appropriate choices of the unit direction vector in equation (1). These operators are then combined to provide a gradient magnitude measure.

IV. APPLICATION SCENARIO

Here we demonstrate how this novel image processing framework can be used to efficiently process RGB-D images. We make use of a publicly available RGB-D dataset [8] collected by an unmanned aerial vehicle. The dataset explores the possibility that in the case of collapsed buildings there should be a way of automatically identifying potential access holes to guide rescuers to trapped people. Once we detect regions with depth discontinuities, indicated by large gradient magnitude values in the depth image we check the corresponding region in the RGB image to determine if the region consists of a dark coloured area (indicating an absence of light). Such regions are then highlighted as possible locations where rescuers could be directed with the aim of finding trapped individuals.

Figure 8. Possible access hole locations highlighted using image segmented with 100 regions

In Figure 8 we illustrate the original image with areas of possible interest for access holes highlighted in yellow. These regions have been identified using a superpixel segmentation of 100 regions. This image demonstrates the superpixel regions that our framework has identified as regions with both large depth discontinuities using the depth image and consisting of dark coloured regions using the RGB image. Similar results are presented in Figure 9 where the image is segmented using 1000 superpixels.

Figure 9. Possible access hole locations highlighted using image segmented with 1000 regions

V. Conclusion

In this paper we have outlined a novel framework for efficient processing of digital images which have been converted to a computationally efficient superpixel representation. We have outlined how the Finite-element image processing operators can be applied to the superpixel image when it is considered as a space-variant image and demonstrated the application of this technique for the purpose of detecting access holes in the case of collapsed buildings for emergency management situations.

Acknowledgment

The research leading to these results has received funding from the European community's Seventh Framework Programme under grant agreement No. 607691 (SLANDAIL).

References

[1] L. Bo, X. Ren and D. Fox, Unsupervised Feature Learning for RGB-D Based Object Recognition, In International Symposium on Experimental Robotics, (ISER), June 2012.

[2] C. J. Taylor and A. Cowley, Segmentation and Analysis of RGB-D data, RSS 2011 Workshop on RGB-D Cameras, June 2011.

[3] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox., RGB-D Mapping: Using Depth Cameras for Dense 3D Modeling of Indoor Environments, Proc. of International Symposium on Experimental Robotics (ISER), 2010.

[4] D. Holz, S. Holzer, R.B. Rusu and S. Behnke, Real-Time Plane Segmentation using RGB-D Cameras, RoboCup Symposium 2011.

[5] P.J. Besl, Active, optical range imaging sensors, Machine Vision and Apps, Vol.1, pp127-152, 1988.

[6] O. Bellon, and L. Silva, New Improvements on Range Image Segmentation by Edge Detection Techniques, Proceedings of the workshop on Artificial Intelligence and Computer Vision, Nov. 2000.

[7] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2012). SLIC superpixels compared to state-of-the-art superpixel methods. IEEE transactions on pattern analysis and machine intelligence, 34(11), 2274-2282.

[8] C. Kong, A. Ferworn, J. Tran, S. Herman, E. Coleshill and K. Derpanis, “Toward the Automatic Detection of Access Holes in Disaster Rubble,” in IEEE International Workshop on Safety, Security & Rescue Robotics (SSRR-2013), Oct 24-26 2013, Linköping, Sweden, 2013

[9] P. Felzenszwalb and D. Huttenlocher, “Efficient Graph-Based Image Segmentation,” Int’l J. Computer Vision, vol. 59, no. 2, pp. 167-181, Sept. 2004.

[10] J. Shi and J. Malik, “Normalized Cuts and Image Segmentation,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 888-905, Aug. 2000.

[11] O. Veksler, Y. Boykov, and P. Mehrani, “Superpixels and Supervoxels in an Energy Optimization Framework,” Proc. European Conf. Computer Vision, 2010.

[12] A. Vedaldi and S. Soatto, “Quick Shift and Kernel Methods for Mode Seeking,” Proc. European Conf. Computer Vision, 2008.

[13] A. Levinshtein, A. Stere, K. Kutulakos, D. Fleet, S. Dickinson, and K. Siddiqi, “Turbopixels: Fast Superpixels Using Geometric Flows,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, no. 12, pp. 2290-2297, Dec. 2009.

[14] Lloyd, Stuart P. (1982), "Least squares quantization in PCM", IEEE Transactions on Information Theory, 28 (2): 129–137, doi:10.1109/TIT.1982.1056489

[15] Shewchuk, J., “Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator”, Applied Computational Geometry', Springer-Verlag, Vol. 1148, pp. 203-222, 1996.

[16] Coleman, Sonya A., Shanmugalingam Suganthan, and Bryan W. Scotney. "Gradient operators for feature extraction and characterisation in range images." Pattern Recognition Letters 31.9 (2010): 1028-1040.

[17] Kerr, Dermot, Sonya Coleman, and Bryan Scotney. "A Space Variant Gradient Based Corner Detector for Sparse Omnidirectional Images." Journal of Mathematical Imaging and Vision 38.2 (2010): 119-131.