TESTING SALIENCY BASED TECHNIQUES FOR PLANETARY SURFACE SCENE ANALYSIS

Brian Yeomans[1], **Affan Shaukat**[1], **and Yang Gao**[1]

[1]*Surrey Space Centre (STAR Lab), University of Surrey, GU2 7XH Guildford, United Kingdom. E-mail: yang.gao@surrey.ac.uk*

ABSTRACT

Saliency based techniques inspired by biological vision systems offer significant potential advantages over the approaches typically adopted for analysing images gathered from the planetary surface. Using saliency, the algorithms deployed are significantly more computationally efficient as only a subset of the image is processed in depth, and thus they offer the potential for near real time performance, facilitating time critical applications such as hazard detection. Saliency based analysis techniques have however in the main been developed with a focus on terrestrial objects and environments, and so require substantial modification for the relatively monochromatic scenes found on the planetary surface. This paper describes the development and testing of a scene analysis model utilising saliency based techniques which has been optimised for the planetary surface environment. The model has been widely tested using datasets from multiple sources and has been implemented as a hazard detection algorithm as part of the recently completed FASTER FP7 project. The model is shown to achieve rates of computational analysis which are more than adequate for navigation purposes in the context of wheeled planetary exploration; it incorporates equalisation techniques designed to expand the limited colour space seen in the planetary environment context; and uses intensity, colour opponency, shape and orientation-based visual stimuli in a weighted matrix to generate a saliency map; stereo disparity information is then used to locate the salient objects in 3D space. Of note is that despite the limited hue variation in the planetary environment, colour is found still to be an important contributor to the success of the model. The model weighting approach is tuneable, enabling optimisation for a number of specific object types and classes to be achieved, which means that the model can be utilised for many more applications than purely hazard detection, such as improving the computational speed of a feature detection / tracking model using saliency to reduce the scale of the image data to which point based feature detection / tracking, as used for SLAM, is applied. Other applications include data mining of planetary imagery, identifying images of interest appropriate for further detailed study, based on parameters tuned to select the desired feature set of interest.

Key words: Saliency, biological vision, FASTER, ExoMars, SLAM.

1. INTRODUCTION

This paper describes a novel approach to planetary rover hazard detection, using saliency based techniques to identify objects of interest in the rover view which potentially represent hazards to navigation. The model has previously been extensively tested using PANGU generated simulation data [12] and real world data from the Ral Space "Seeker" dataset [15], as described in [14]. This paper describes the evolution of the developed algortihm for use in a hazard detection application, deployed to identify and localise rock hazards on the primary rover "Bridget", as part of the recently completed FASTER FP7 Framework Project [10].

The objective of the FASTER (Forward Acquisition of Soil and Terrain data by Exploration Rover) project was to facilitate higher average travel velocities whilst reducing risks arising during the traverse, from both geometric hazards and non geometric hazards such as loose sand. It comprised two Exploration Rovers working cooperatively. At the start of each traverse leg, the primary rover generated an initial map using images stitched together from a mast mounted stereo camera pair with pan / tilt capability. A high mobility, lightweight hybrid wheel / leg vehicle went ahead of the primary rover, under its control, to scout the planned path, gathering LIDAR data on geometric hazards, and sinkage data from the wheel / legs to assess terrain conditions and trafficability. LIDAR data and data on terrain conditions was shared between the rovers, and a revised global map constructed. In the case of elevated risk due to the identification of poor trafficability, path re-planning was conducted to avoid the dangerous zones, thus enhancing overall mission safety. As the primary rover moved, a second forward mounted stereo camera dedicated to hazard detection provided a continuous update, identifying potential hazards using the system described in this paper, and triggering further path re - planning cycles as required.

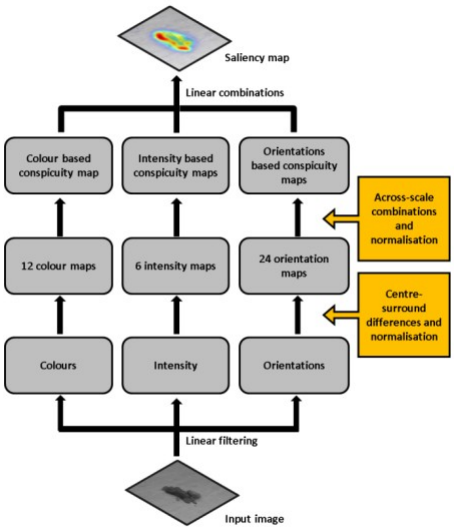

Figure 1. Itti Koch Niebur Saliency Model [8]

2. BACKGROUND TO SALIENCY BASED ANALYSIS

Animals, and primates in particular, have a highly developed ability to interpret complex information in the visual scene in real time, despite the limited speed of the "computational hardware" available for the purpose, when compared with that of a modern processor [8]. This process appears to function by means of selection of a subset of the available sensory information, which is then subjected to futher processing. This selection appears to take the form of the identification of a sub region or regions of the scene. This "focus of attention" is identified though a rapid, bottom up initial process highlighting saliency regions for further detailed study.

In the context of machine vision, saliency based analysis therefore seeks to mimic this process through an initial, rapid scan of the scene to identify and select zones of interest for futher study. Regions are identified by decomposing the visual input into a series of topographic feature maps, each of which is assembled through analysis based on some particular aspect contributing to the conspicuity of the region, such as intensity, colour balance, shape, orientation, and across scale / frequency dependent features [8]. Within each feature map, different spatial locations compete for attention so that only regions which are locally distinctive can persist in the feature map. These individual maps then are combined, typically on a weighted basis, into a master binary "saliency map", which is used to identify locally conspicuous features across the whole scene to be selected for futher processing.

3. HARDWARE EQUIPMENT AND SOFTWARE PLATFORM

Images utilised in testing the developed algorithms were sourced from Mars Curiosity Rover images [3], RAL Space data sets [15], PANGU simulated data [12], and recordings of live image streams sourced from a Point Grey Bumblebee 2 1394A interface stereo camera [2] mounted on the ExoMars propotype rover "Bridget" [9]. The saliency analysis uses solely monocular images, which in the case of the Bumblebee 2 were sourced from the left hand camera. Disparity information from the Bumblebee 2 stereo pair was used to generate the position of detected objects in 3D space as described in Section 6. The Bumblebee 2 is mounted so the camera axis points forward and angled down from the horizontal by approximately 22° so the terrain approximately 5 m in front of the rover is in the centre of the field of view of the camera. This distance would not normally be exceeded in a single leg of the rover traverse and therefore this orientation of the camera provides good data on the region about to be traversed.

The software used was developed to integrate with that used for the FASTER project, and utilises the well known Robot Operating System (ROS) middleware platform [5]. Vision processing and data handling routines were written and implemented within a C^{++} ROS framework making extensive use of the open source vision processing libraries of OpenCV [4].

4. CHOICE OF ALGORITHM

Much of the reseach work in saliency models is based on terrestial environments, which typically exhibit a wider range of colour components and a higher incidence of regular geometric patterns, particularly in an indoor or outdoor built environment, when compared with the relatively low featured, unstructured, somewhat monochromatic visual environments typical of the planetary surface. To address this question, a number of alternative algorithms were selected and evaluated using typical planetary surface imagery of the Martian surface, or images from the Airbus Mars Yard [1] collected during the FASTER project field trials.

4.1. Itti Koch Niebur Algorithm

This algorithm was originally proposed in 1998 and so is perhaps the most well established. The structure of the algorithm is shown in Figure 1.

The approach is to decompose visual input into a series of topographic maps. Different spatial locations within each feature map then compete for saliency, so that only locations which stand out from their surroundings can persist [8]. All feature maps feed into a master saliency map

which then encodes for local conspicuity over the entire scene. The method generates forty two feature maps in all, which are subsequently combined into one master saliency map, and so the computational load is necessarily fairly high. A fundamental aspect of the "Itti" approach is the construction of subsets of six "across scale" feature map sets derived using Gaussian pyramid scaling. Within each subset, salient features are identified by "centre" (fine scale) / "surround" (coarse scale) differences. This approach exploits theories of primate visual processing which consider that visual neurons are most sensitive in a small, centre region whereas stimuli presented in a broader region concentric with the centre tends to inhibit the response [8]. The processing which is applied develops across - scale differences using several scales to achieve multi - scale extraction of features.

This across scale difference generation is applied firstly to an Intensity image I obtained from the red, green and blue image channels as $I = (r + g + b)/3$. Then, colour difference channels are computed based on the principle that within their receptive field, neurons are excited by one colour, such as red, and inhibited by their "opponent" colour (green), whereas the converse is true in the surrounding zone. Therefore colour difference channels are computed which adopt this colour opponency principle and compute red / green plus green / red, and blue / yellow plus yellow / blue double opponency channels respectively. The across scale analysis described above is then also applied to the colour difference channels to generate twelve new feature maps .

Finally, sets of orientation information generated using a Gabor filter are produced, based on orientations of 0°, 45°, 90° and 135°. Each of these are also analysed using the same across scale processing to generate an additional twenty four feature maps in total. [8]

Having generated all of these forty two (6 + 12 + 24) individual feature maps, the next challenge is to combine them into a master saliency map in a manner which preserves the desired information whilst minimising false positives. The approach adopted is to combine sub maps using a weighting matrix whose coefficients are determined based in the incidence of local maxima in the individual feature maps. Local maxima in each feature map are identified, the number of peaks and average peak value computed, and the weighting coefficients calculated based on $(1 - m)^2$ where m = average peak value. This approach to weighting has the effect of promoting maps with a small number of strong peaks whilst suppressing maps with a large number of similar value peaks.

Finally, the "grey scale" result of combining the individual feature maps is thresholded to a binary saliency map using Otsu's thresholding method [11]. This method assumes that the image contains two sets or classes of pixels, the foreground class and the background class. It then computes the threshold value which minimises the combined spread (the intra - class variance) of the background and foreground classes. Computationally, this is equivalent to maximising the inter - class variance σ_B^2, which is much easier to calculate:

$$\sigma_B^2 = W_b W_f (\mu_b - \mu_f)^2 \quad (1)$$

where W_b, W_f are the background and foreground weights and μ_b, μ_f the background and foreground means.

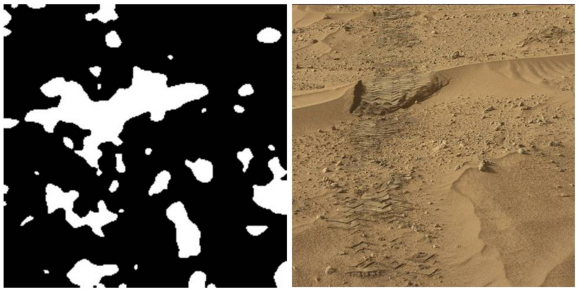

Figure 2. Itti saliency map (original image courtesy NASA)

Figure 2 shows an example binary saliency map, shown side by side with the original image. It can be seen that results are mixed. Whilst the gap feature in the sand ridge is clearly identified, it can also be seen that a very substantial number of false positives are generated, particularly from textures in trafficked sand.

4.2. Hou Algorithm

The second algorithm considered was that of Hou [7]. In contrast to Itti's algorithm, this approach is computationally simpler and therefore faster and so more suited to a real time application. It operates on grey scale images, and adopts a different approach from Itti's across scale difference methodology to identify salient features.

The principle adopted in the Hou method is to mimic the behaviour of visual systems which suppress the response to frequently occurring features whilst remaining sensitive to features that deviate from the norm [7]. This is achieved using a method which approximates the salient or "innovation" component of an image by removing the statistically redundant components.

This is based on an analysis of the log spectrum of the image. From the log spectrum $L(f)$ is subtracted an averaged spectrum $A(f)$ to derive the "spectral residual" $R(f)$ and so:

$$R(f) = L(f) - A(f) \quad (2)$$

The spectral residual, which can be thought of as the non trivial or unexpected portion of the image, is converted back to the spatial domain to generate a saliency map

Figure 3. Hou saliency map

Figure 4. Rudinac saliency map

using an Inverse Fourier Transform [7]. There is no requirement for a weighting exercise here as no merging of individual sub maps is required. The grey scale saliency map generated is thresholded to a binary representation using Otsu's method, in the same manner as described above [11].

Figure 3 shows results from applying the Hou algorithm, this time using Mars Yard imagery. It can be seen that the algorithm is successful in identifying the rocks in the field of view, however the sand texturing in the foreground continues to cause problems through generation of small, false positive detections. These false positives in the main are sized below the level which would present a significant obstacle to the rover, and therefore conceivably a filtering process with an upper limit of detection sizes below that which would present a real obstacle to the rover might still result in a workable hazard detection system. However it was decided to investigate whether extending the Hou approach through the inclusion of colour information might improve false positive rejection, despite the relatively monochromatic nature of the planetary environment.

4.3. Rudinac Algorithm

This algorithm, described in [13], can be considered as following essentially the same computationally efficient process as the Hou algorithm, with the addition of analysis of colour opponency information. In the case of this approach, spectral residual information is computed not simply from intensity data but also from red - green and blue - yellow colour difference data, giving three channels in total. These channels are defined as follows:

$$M_I(Intensity) = \frac{r+g+b}{3} \tag{3}$$

$$M_{R-G}(Red-Green) = \frac{r-g}{max(r,g,b)} \tag{4}$$

$$M_{B-Y}(Blue-Yellow) = \frac{b-min(r,g)}{max(r,g,b)} \tag{5}$$

The three sub saliency maps generated are then combined into a single global map.Weighting of the relative contribution of the components can also be implemented, as with the Itti algorithm.The grey scale map generated is again thresholded using Otsu's method [11].

Figure 4 shows an example of the output from this algorithm. In comparison with the Hou algorithm, this approach is much more successful in suppressing false positives from the sand textures which retaining detection of actual obstacles. As the primary difference between the algorithms is the incorporation, in the Rudinac algorithm, of colour information, it would appear that small differences between the colour of the rocks and the sand background are sufficient to push those differences above the binarisation threshold and to push the sand texture effects below the threshold.

5. ALGORITHM ENHANCEMENTS

Given the relative performance of the three routes tested, an approach based on the Rudinac algorithm was selected as the basis for further study into how the techniques adopted could be further improved.

5.1. Histogram Equalisation

One way of maximising the impact of the limited colourspace information available is to expand the range of intensity data to highlight otherwise small intensity colour differences in each of the colour channels. This can be achieved by a process of contrast improvement using histogram equalisation, enabling the intensities of each of the Red, Green and Blue image channels to be more broadly distributed on the histogram of intensity values. Histogram equalisation generates an equalised image using a normalised histogram with a bin for each possible intensity value. Therefore the normalised histogram p_n of an image f with intensity values which range between 0 and L for $n = 0, 1...L-1$ is given by:

$$p_n = \frac{\text{number of pixels with intensity } n}{\text{total number of pixels}} \tag{6}$$

The equalised image g is given by:

Figure 5. Effect of Histogram Equalisation

$$g_{i,j} = floor((L-1)\sum_{n=0}^{f_{i,j}} p_n \qquad (7)$$

The result of incorporating this pre - processing step to each of the three colour channels, prior to further processing, can be seen in Figure 5.

The left image shows the original scene from the Airbus Mars Yard. The centre image shows the saliency map without histogram equalisation, and the right hand image the saliency map generated once each of the three colour channels have been histogram equalised. Spurious objects, at the top and the far lower left of the image, can also be seen, which derive from the Mars Yard wallpaper background and the rover's deployed soil sensing equipment respectively. These are in a constant position by reference to the image frame and are removed by masking the image before further processing.

It can be seen that the histogram equalisation pre - processing step is quite successful in improving rock detection in the middle distance. Incorporation of colour based information in the map generation can also be seen to maintain rejection of false positives from sand texturing despite significant lighting variations.

6. HAZARD LOCALISATION

For hazard detection as implemented in the FASTER project, the regions identified from the saliency map were treated as potential navigation hazards, localised in 3D space and incorporated in the mission mapping process.

To localise the identified potential hazards in 3D space, the following steps were undertaken:

1. Rectangular bounding boxes were generated around each saliency map "blob".
2. The centre of each rectangle was computed.
3. Using the disparity map generated from the Bumblebee 2 stereo image pair, the locations of the bounding box centres were identified in 3D space, relative to the camera coordinate frame, and taken as the centroids of the "proto - rocks " identified.

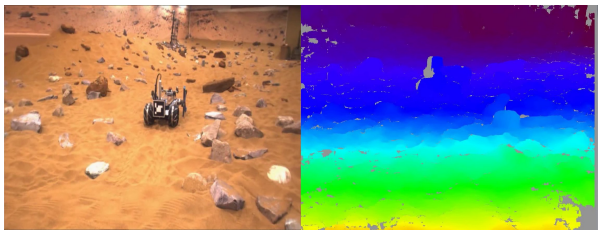

Figure 6. Disparity Map Generation

Figure 6 shows a side by side comparison of the camera image and the generated disparity map, here showing the scout rover clearly in the field of view. It can be seen that here and there there are a few grey areas which represent gaps in the disparity information. This was found to be inevitable to some degree with a moving scene, as on occasion the disparity algorithm cannot generate an accurate disparity value.The potential consequence is that if a blank zone falls precisely on a saliency blob centroid, an invalid depth measure will be generated. The fact that the value is invalid can be clearly identified however, and the erroneous value filtered out, and so it was found that in practice these missing data are transitory and do not materially affect the reliability of the output.

Once the centroids of the anticipated hazards are identified, the dimensions of the rock hazards were estimated as follows:

1. The diameter of the rocks were taken as the width of the bounding rectangles
2. The height of the rocks were estimated using a statistical model of rock diameter / height relationships developed by Golombek [6].

The rock hazard data was then further processed as follows:

1. The stream of rock detection data generated was filtered to remove both very small detections, and those both large and beyond the range of the rover during the current traverse of some 4 - 5 m. This is because detections of small rocks would not represent real hazards and may well be false positives. Detections beyond the rover range in the current traverse of around 4 to 5 m were also removed as again, they do not represent current hazards.
2. Finally, the hazard location and dimension data were output to a data fusion process which transforms the coordinates of the rock centroids from the camera frame to the world frame for inclusion in the global map.

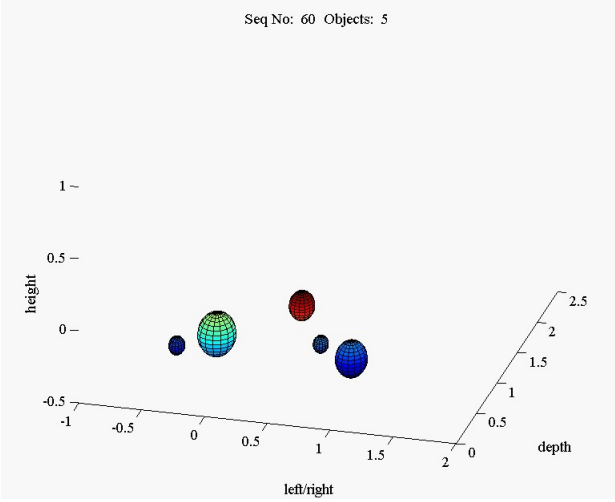

Figure 7. Tracking Detected Objects

A qualitative view of the validity of detected data was obtained by visualising the detected rocks as spheres and examining their movement frame by frame. Figure 7 shows one such frame from a video sequence designed to validate the results. What is expected is that detected rocks appear in the camera frame and move smoothly through it, exiting once the obstacle is no longer in the field of view. In contrast, poor performance would be indicated by detections which appear and then disappear at high frequency or jump discontinuously from one location to another. In fact, the video sequences generated showed a pattern of smooth motion, confirming that the algorithm and computation routines appeared to be operating effectively.

7. CONCLUSIONS AND FUTURE WORK

The primary purpose of investigating saliency based approaches to hazard detection in a planetary surface environment was to improve the computation speed of the image analysis algorithm. Generation of a saliency map as the first stage of processing enables only those parts of the image which contain relevant information to be focussed on. The experience from these tests has shown that the algorithm was easily able to keep up with the forward progress of the rover during its traverse and thus provided near to real time performance in the context of the rover's forward velocity of $3 - 4cms^{-1}$.

Incorporation of colour information was shown to be effective in retaining detection of rock hazards whilst suppressing texture originated false positive detection.The value of the inclusion of colour information was shown to be further enhanced through the use of histogram equalisation to enhance contrast in each of the three R, G , B colour channels. The overall result was found to be quite effective at acurately identifying all rock hazards although it was not surprisingly seen to be somewhat less sensitive in cases where the hue and textural components of the rocks were very similar to those of the background sand.

With respect to future work, the benefit of futher tuning of the algorithm, by adjusting the weighting of the feature sub maps which contribute to the final saliency map, should be investigated. Whilst in the tests described in this paper, a general purpose weighting was utilised, it would be possible to modify the weightings to emphasise or de - emphasise components, allowing a more targetted pre - analysis of the scene. This would enable a more general application of the algorithm, for example to identify regions of interest targetted for further analysis using point based feature identification and tracking, as in SLAM, or for pre - analysis of image collections to identify and select images or image sections relevant for more detailed study.

ACKNOWLEDGMENTS

The FASTER project received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 284419. The authors would like to thank the entire FASTER team for their assistance and support in establishing and implementing the mission scenarios which have enabled these algorithms to be tested under realistic mission conditions.

REFERENCES

[1] Airbus Mars Yard Ready for Red Planet Rover. `http://www.esa.int/Our_Activities/Space_Science/Mars_yard_ready_for_Red_Planet_rover`. Accessed on 19th March 2015.

[2] Bumblebee 2 Stereo Camera. `http://www.ptgrey.com/bumblebee2-firewire-stereo-vision-camera-systems`. Accessed on 19th March 2015.

[3] http://mars.nasa.gov/msl/multimedia/images/. `http://www.ros.org/`. Accessed on 14th April 2015.

[4] Open Computer Vision. `http://opencv.org/`. Accessed on 19th March 2015.

[5] Robot Operating System. `http://www.ros.org/`. Accessed on 19th March 2015.

[6] MP Golombek, RA Cook, T Economou, WM Folkner, AFC Haldemann, PH Kallemeyn, Jens Martin Knudsen, RM Manning, HJ Moore, TJ Parker, et al. Overview of the mars pathfinder mission and assessment of landing site predictions. *Science*, 278(5344):1743–1748, 1997.

[7] Xiaodi Hou and Liqing Zhang. Saliency detection: A spectral residual approach. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[8] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11):1254–1259, 1998.

[9] CGY Lee, J Dalcolmo, S Klinkner, L Richter, G Terrien, Ambroise Krebs, Roland Yves Siegwart, L Waugh, C Draper, Roland Yves Siegwart, et al. *Design and Manufacture of a full size Breadboard ExoMars Rover Chassis*. Citeseer, 2006.

[10] Yashodhan Nevatia, F Bulens, J Gancet, Y Gao, S Al-Mili, RU Sonsalla, TP Kaupisch, M Fritsche, E Allouis, T Jorden, et al. Safe long-range travel for planetary rovers through forward sensing. In *Proceedings of the 12th Symposium on Advanced Space Technologies in Robotics and Automation (ASTRA 2013), Noordwijk, the Netherlands*, 2013.

[11] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

[12] Stephen Parkes, Iain Martin, Martin Dunstan, and D Matthews. Planet surface simulation with pangu. In *Eighth International Conference on Space Operations*, pages 1–10, 2004.

[13] Maja Rudinac and Pieter P Jonker. Saliency detection and object localization in indoor environments. In *ICPR*, pages 404–407, 2010.

[14] Affan Shaukat, Conrad Spiteri, Yang Gao, Said Al-Milli, and Abhinav Bajpai. Quasi-thematic feature detection and tracking for future rover long-distance autonomous navigation. In *Proc. ESA Conference on Advanced Space Technologies in Robotics and Automation, ASTRA, Noordwijk, Netherlands*, 2013.

[15] Mark Woods, Andrew Shaw, Estelle Tidey, Bach Van Pham, Lacroix Simon, Raja Mukherji, Brian Maddison, Gary Cross, Aron Kisdi, Wayne Tubby, et al. Seekerautonomous long-range rover navigation for remote exploration. *Journal of Field Robotics*, 31(6):940–968, 2014.