

# **AUTOMATIC SALIENT OBJECT DETECTION IN UAV IMAGERY**

*Jan Sokalski, Toby P. Breckon*  
School of Engineering  
Cranfield University, UK

*Ian Cowling*  
Blue Bear Systems Research  
Bedford, UK

## **ABSTRACT**

Due to the increased use of Unmanned Aerial Vehicle (UAV) platforms in land-sea search and surveillance operations a suitable general technique for the automatic extraction of visually significant information is needed in order to augment current human-performed manual analysis of received video imagery. This paper presents a novel image processing based approach that builds on existing salient object detection work within related domains. Our proposed approach uses an image contrast map derived from the combination of seminal work in this area, multi-scale mean-shift segmentation with additional histogram enhancement and additional multi-channel edge information. This is used to construct a robust saliency map from a given UAV aerial image in the presence of environmental, transmission and motion noise affecting image quality. The approach is generally targeted towards the detection of salient objects in the rural, uncluttered and relatively uniform environments. A range of results are presented over such representative environments.

## 1. Introduction

Modern search and surveillance operations can be complex and expensive involving numerous personnel and multiple airborne search platforms. These are often employed to search wide areas of open expanse such as mountainous, oceanic or desert areas. Two recent examples of such operations are the disappearance of Steve Fosset who disappeared in his small plane in the vast Rockies mountain range (2007) and additionally the search-and-rescue operation in the South Pacific for the debris of Air France Flight 447 (2009). In the former example of Fosset's crash the crash scene was only discovered by chance a year after the event and the search and rescue involved among other resources, around 5000 online volunteers providing the analysis of thousands of digital satellite images covering hundreds of square kilometres. As for Flight 447, here it took a prolonged search period of around 5 days before the first debris was found in the Pacific Ocean via search operation involving tens of ships and aircrafts. This is a classic scenario for the automatic image analysis of remotely gathered satellite and airborne imagery.

In both of these scenarios essentially we are looking for something unique in a fairly simple and uniform environment (for example floating wreckage on the ocean in the case of Flight 447). Although the course of events for a given search and rescue mission may vary depending on the scenario, environment and many additional factors some common themes can be found. The most prominent of which is that very often the object being searched for may differ significantly from the environment in which the search is taking place. This makes this scenario particularly prominent as an application for computer vision based saliency detection.

One of the many aims of computer vision is to automatically extract visually important parts of the scene image. Therefore computer vision based saliency detection has been developed as an entity within itself ((7), (1), (3), (10)). Saliency detection is essentially the detection of visually unique objects within a given image and this is well suited to the problem of automatic image analysis for the search and rescue imagery cases discussed. However, although image saliency has already been used within more complex computer vision systems (4), (12) with different goals it has surprisingly not yet been applied

to any system addressing automatic search in Unmanned Aerial Vehicle (UAV) imagery. Here we aim to create such system capable of the automatic detection of salient objects in UAV imagery for search and rescue (or surveillance) operations.

This paper introduces an overview of such a system based on (10), (13). We combine the seminal work in this area, multi-scale mean-shift segmentation with novel histogram enhancement and additionally multi-channel edge information to construct a robust saliency map from a given UAV aerial image. This is then used to highlight salient objects of interest within a stream of UAV video imagery and hence reduce operator workload for manual image analysis. A range of related work has been carried out in computer vision on saliency detection which we will adapt towards our UAV video search purposes (10), (13).

## 2. Related work

Before we detail related work in the area of computer vision saliency detection it is important to make one point regarding the type of images considered by the authors of most work in this area. The majority of computer vision image saliency detection work considers high quality and high resolution colour images with minimal noise where the salient object is the main subject of the image (e.g. Figure 1). By contrast, here we deal with UAV surveillance imagery (Figure 2a/2b) where they may contain multiple salient objects within a lower quality, low resolution image suffering from a range of motion turbulence, transmission and compression noise in addition to environmental noise such as weather and atmospheric conditions.

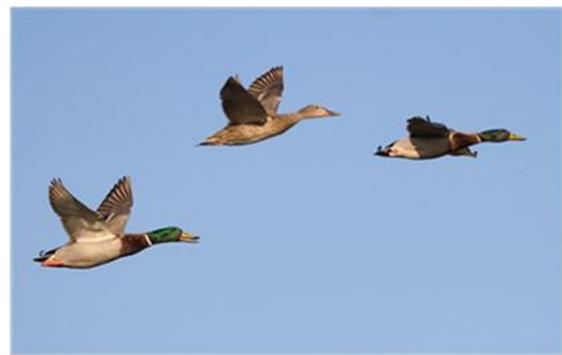
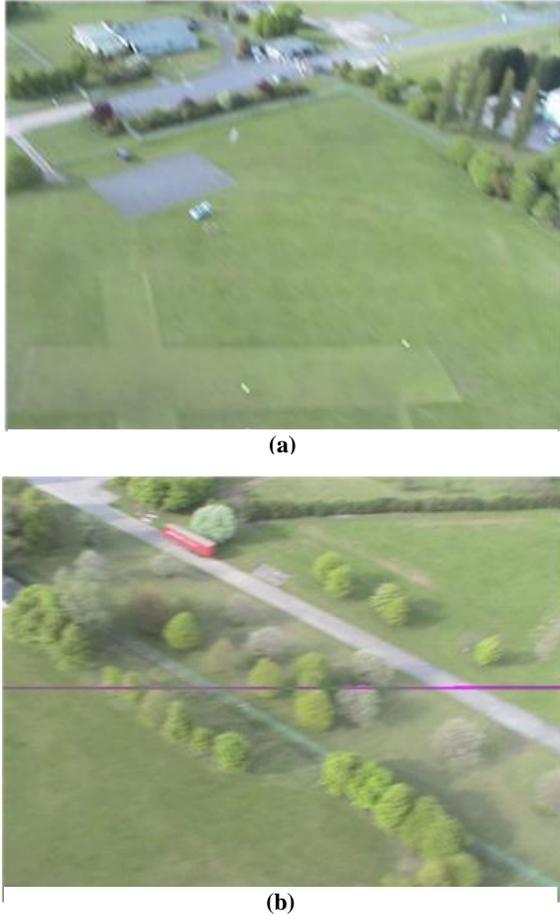


Figure 1 Typical image considered.



**Figure 2** Images captured by UAV operated by Blue Bear System Research

The images presented in Figure 2 expose all aspects of imagery collected by a UAV platform. For example varying light conditions, colour saturation, blur and noise distortion due to variations in weather, platform speed, altitude and stability and additionally transmission related noise. Finally it is important to note the difference in the type of visual saliency we wish to detect in such UAV surveillance images where the salient objects are by their nature much less conspicuous than the traditional “*subject of the image*” type examples used by traditional authors in this domain (7), (9), (13).

The first comprehensive methodology for saliency detection in computer vision was introduced by Itti and Koch (10) who based their work on equivalency to the mechanisms of the human visual system which has become a point of reference in the field of saliency detection. Several of the techniques introduced in their methodology are still the basis for more current research in this area (9). Itti and Koch

(10) created saliency map as a linear combination of three conspicuity maps received from calculating three low-level features: colours, intensity, and orientation. An image pyramid (6) with the fixed number of levels is then used during the process to give multi-scale evaluation of saliency. This is built up on low level features of Itti and Koch’s algorithm which itself does not incorporate any explicit region information (10).

Further work in this area improved upon (10), such as that of Hu et al (8) which used a subspace estimation algorithm based on GPCA (Generalized Principle Component Analysis) to directly measure the global visual contrast using region level information. Furthermore Liu and Gleicher (13) used mean-shift segmentation (5) to supply information on saliency of individual regions. As a basic low level feature and measure of saliency Liu and Gleicher used pixel contrast while Hu et al (8) resigned from using this in favour of a polar transformation of features. Image contrast is now the most widely used low-level saliency feature and in addition to the work of Itti and Koch (10), Liu and Gleicher (13) was also incorporated into later works of Kadir and Brady (11), Ma and Zhang (14). Overall contrast is seen as the most powerful and widely used low level image feature for the task (10).

However some other authors have additionally employed alternative features such as measures of orientation saliency (10), (3). Itti and Koch (10) and Chalmond et al (3) both used oriented Gabor filters to obtain such orientation information as an additional saliency measure. In addition to such low level features a number of other approaches have been employed in saliency detection such as the work of Hou and Zhang (7) which used a Fourier image to create a *log spectrum representation*. This was an incorporated with basic information theory work to obtain *spectral residual* which contains conspicuous elements beyond the average Fourier spectrum of a set of natural images. Further work expanded upon this approach (Wang and Li (15)) by considering the application of the spectral residual process to each channel of a perceptual HSV colour space image (6). Other work advanced the use of information theory in this field (1) based on an architecture utilizing information maximization.

The work that we investigate here, with application to saliency detection in UAV images, is

primarily based on the work of Liu and Gleicher (13) and Itti and Koch (10) as we look to take image saliency from its traditional form of main “subject of the image” saliency towards salient object of interest detection in the wide area search scenario.

### 3. Intended saliency detection scenario

The area on which search and rescue missions are performed can often be narrowed down to a set of fairly uniform environments (e.g. in case of a plane crash on the ocean we would be dealing with a uniform water environment upon which we would be searching for a range of salient debris objects). This more exact description of the type of environment for such UAV search and rescue or surveillance missions enables us to determine more precisely the type of images upon which saliency detection will be performed. This knowledge that the environment will be generally uniform defines the background of the images as being largely invariant over a given search mission (e.g. Figure 4a). As for the salient objects of interest (the foreground), we will be interested in finding items such as survivors, debris or crash sites (e.g. Figure 4a, Figure 11, Figure 12). In general therefore, the proposed architecture should thus be able to extract multiple salient objects at varying scales within the scene. Additionally due to the varying altitudes of the UAV platform the architecture for saliency detection should additionally be capable of incorporating some scale adjustment based on varying UAV altitude.

From this overview of the intended scope of saliency detection for UAV search and rescue (or surveillance) missions we move on to detail in full the algorithmic architecture of the proposed approach.

### 4. Algorithm Description

We propose a multi-stage salient detection system which combines low-level contrast features, mean shift segmentation with additional histogram information and multichannel edge features gathered over several feature maps. Figure 3 presents a block diagram of proposed saliency detection architecture in which we see the various steps of feature detection

and amalgamation. Each of these steps in the overall architecture are described in Sections 4.1-4.5 below.

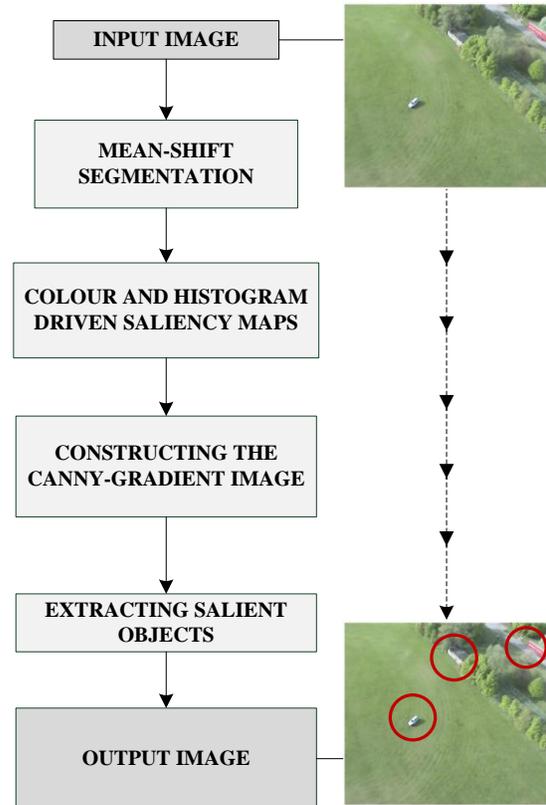


Figure 3 Block diagram of proposed methodology.

#### 4.1. Mean-shift Segmentation

Using mean-shift segmentation (5) with low-level contrast features has proven to add significant information about the saliency of image regions (Liu and Gleicher (13) Here we follow this approach as using mean-shift segmentation (5) with appropriate radii in both the spatial and colour domains has the ability to preserve small entities in image whilst at the same time merging larger, uniform image regions. Figure 4b illustrates the use of mean-shift segmentation on example UAV image (input image is Figure 4a). It is important to note that the mean-shift has left all of the possible objects of interest as feature regions within the scene, including the person wearing a bright life jacket type garment (marked with red circle in Figure 4a and 4b for clarity).

## 4.2. Colour and Histogram Driven Saliency Maps

The mean-shift segmented image (Section 4.1) is first transformed into the LUV colour space (i.e.  $L^* u^* v^*$  (6)) and a two level Gaussian image pyramid (6) is created out of which the final contrast map is derived. In prior work Liu and Gleicher (13) assigned weights to each pixel giving higher values to those closer to the centre of the image. In our case this is not needed as salient objects of interest can appear in any part of the UAV image (whereas in the Liu and Gleicher original work they were thus concentrating on salient objects appearing in a focal point). Thus here, we alternatively calculate contrast as a Euclidean distance which is a characteristic feature of the LUV colour space, as shown in Equation 1:

$$C_{i,j,l} = \sum_{q \in \Theta} d(p_{i,j,l}, p_q) \quad (1)$$

Where in Equation 1  $C_{i,j,l}$  is accumulated contrast value for every pixel,  $p_q$ , in  $\Theta$ -sized neighborhood of pixel  $(i,j)$ ,  $p_{(i,j,l)}$ , at level  $l$  within the pyramid (13). The contrast maps from each level within the pyramid are then rescaled to the original image size and additively combined into one image. At this point of recombination we also add additional information about colour distribution from the HSV (Hue, Saturation and Variance colour space) histogram of the mean-shift image (6). Each contrast pixel value is effectively multiplied by an inversely weighted probability of its occurrence in a normalized two-dimensional histogram of hue and saturation (H and S channel histogram of HSV image). Using this information each contrast pixel value in the final resized contrast map is suppressed if its frequency is high within the histogram distribution whilst those which appear with limited frequency are enhanced. Thus contrast pixel values with an associated high probability of hue and saturation occurrence within the image are effectively suppressed (by the assumption that they are the majority background information) whilst the less probable contrast pixel values within the image are enhanced (with the assumption they are the limited, isolated salient objects within the image). In this way we combine both of the traditional contrast map based approaches for saliency detection (13) with the additional

enhancement of likelihood of occurrence probability of a given colour within the image. This specifically targets our saliency approach presented here towards the detection of isolated, small scale salient objects relative to the image size in contrast to the approach of Liu and Gleicher (13) which weighted their saliency in relation to the object being in the centre or the most likely focal point of the image.



(a)

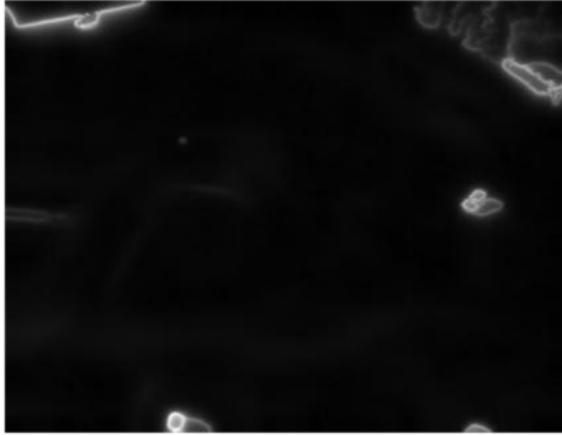


(b)

**Figure 4** Mean-shift segmentation. (a): original image, (b) segmentation result

In Figure 5 we see these steps illustrated on the original input image from taken from Figure 4a. Firstly, in Figure 5a we see the saliency map extracted from two level Gaussian pyramid performed on the LUV colour space and secondly we see this enhanced with the additional histogram-based weighting which gives the results we see in Figure 5b. As we can see adding histogram information suppresses the background noise, and preserves only possible objects of interest (i.e. one

red and two silver cars, the building near the right top corner and also the person wearing life-jacket-with ref. to Figure 4a).



(a)



(b)

**Figure 5** Resized contrast map of image pyramid of mean-shift images (a) and contrast map with added histogram information (b)

### 4.3. Constructing the Canny-Gradient Image

In this next stage we gather multiple sources of edge information into two separate images, one for Canny edges (2) and one for gradient edges, where the latter are defined with the use of morphological operation of erosion and dilation (6) as follows:

$$gradient(i) = dilate(i) - erode(i) \quad (2)$$

Where in equation (2)  $i$  is the input image over a set of nine feature images. Overall edge information is

gathered on a set of nine feature images as follows: the mean-shift image (Section 4.1.), the contrast map with added histogram information (Section 4.2) and seven feature images obtained following Itti and Koch's original seminal method (10). These seven feature images are three original color channels (R,G,B) normalized by intensity (equation (3)) and four *broadly tuned color channels*. Figure 6 below illustrates the R,G,B channels normalized by intensity (Figure 6b, 6c, 6d) for a given input image (Figure 6a), where intensity is defined for the input image as follows:

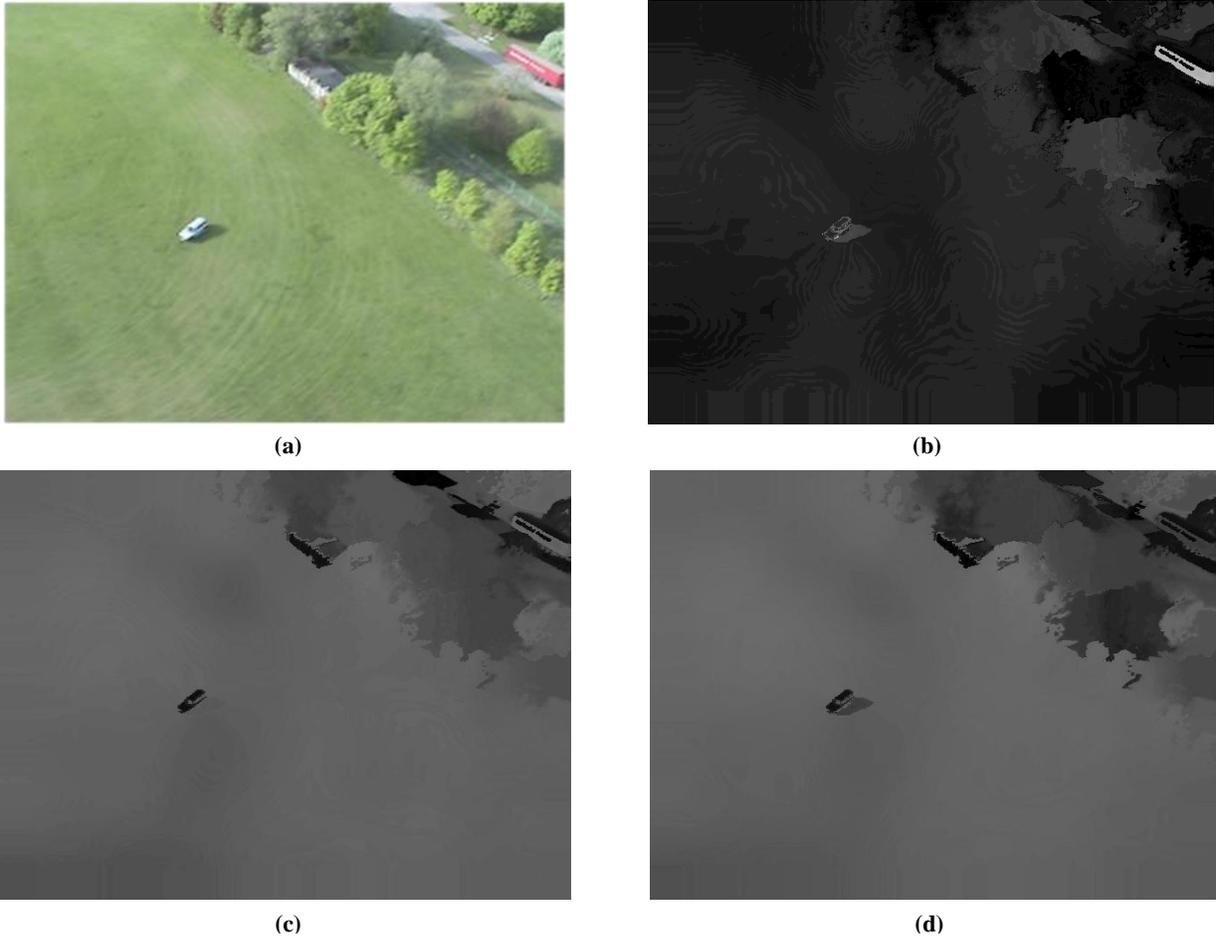
$$intensity = \frac{(r + g + b)}{3} \quad (3)$$

Where  $r,g,b$  are the pixel values from each of the red, blue and green channels from the image respectively (6). From Figure 6 and Figure 7 we can see that those regions which are of interest are easily distinguishable across different image channels. For example, red trailer is the brightest region on R-normalized (Figure 6b) image and bright blue car is a conspicuous dark region on B-normalized (Figure 6c) and G-normalized (Figure 6d) images. The reason behind this is that a salient region within the image contains the least probable value in either of R,G,B channels at any given time based on its size within the image and this is highlighted by the inverse weighting that originates from the HSV histogram weights used earlier in Section 4.2.

In addition the broadly tuned channels, as defined by Itti and Koch (10) can be expressed as follows:

$$\begin{aligned} R &= r - \frac{(g + b)}{2} \\ G &= g - \frac{(r+b)}{2} \\ B &= b - \frac{(g + r)}{2} \\ Y &= \frac{(r + g) - |r - g|}{2} \end{aligned} \quad (4)$$

for red (R), green (G) and blue (B) channels respectively. Y channel is a novelty introduced by Itti and Koch (10). Similarly as for the normalized channels, if we look at the images of four broadly tuned channels R,G,B and Y (for yellow) presented in Figure 7a and Figure 7b we see a similar phenomenon for a given input image.



**Figure 6** Red (b), green (c), blue (d) channels of original image (a) normalized by intensity image. (Logarithmic transform was applied on images (b)/(c)/(d) for visibility purposes due to small scale of reproductions).



**Figure 7a** Broadly tuned channels: red (i), green (ii) obtained for input image presented in Figure 6a



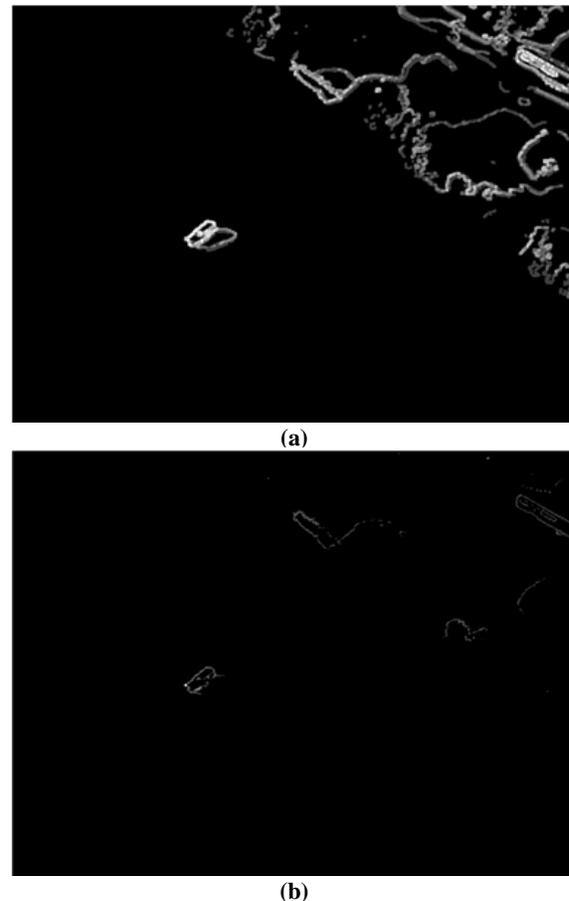
**Figure 7b** Broadly tuned channels: blue (i) and yellow (ii) obtained for input image presented in Figure 6a. (Brightness of the image presented in Figure 7b (i) was slightly adjusted for printing/reading purposes).

Once the set of nine feature images is composed in this manner we additively gather edge information over the entire set into two resulting images, by using both the Canny edge detection operator (2) and gradient edge information (6). Each of the resulting images is obtained as the sum of responses over each of these operators (Canny and gradient) respectively.

In the final step these two resulting images are multiplicatively combined leaving only the most conspicuous edges (effectively a gradient dependent weighted logical AND operation). Figure 8 presents the accumulated Canny edges (Figure 8a) and gradient edges (Figure 8b) obtained in this manner for the original input image presented in Figure 6a. From Figure 8 we can see that the salient objects present in the original Figure 6a example UAV image are prominent in the resulting salient edge response obtained by the accumulation using both of these methods. Having now obtained these two images of accumulated edges we multiplicatively combine them to obtain our final saliency map image. This operation will preserve only the most salient parts of the image by both independent measures of gradient in the combined nine feature saliency maps. An example of the final image obtained for the image presented in Figure 6a is presented below in Figure 9. This shows the final combined edge map information over range of 9 inputs combining both our own work in developing a novel histogram driven saliency map with the seminal work of Itti and Koch (10). From Figure 9 we can again see the prominence of salient objects originating from the input example Figure 6a remaining in this Canny-gradient image representing

the combination of multiple saliency detection contrast maps.

We move now to the extraction of coherent salient objects within this robust saliency map.



**Figure 8** Accumulated Canny edges (7a) and accumulated gradient edges (7b).



**Figure 9** Canny-gradient image.

#### 4.4. Extracting Salient Objects

The constructed Canny-gradient image (Section 4.3.) is now processed using dilation and connected components to extract coherent salient objects within the scene. Using prior morphological dilation will merge separated objects within the Canny-gradient image prior to further connectivity processing. As can be seen in Figure 10 the subsequent application of connected components then identifies coherent objects (uniquely coloured within Figure 10) within the image. Here (in Figure 10) we see these items additionally bounded with a bounding red rectangle as an indicator of convex shape/area which will be used in later constraint processing.



**Figure 10** Connected components analysis of canny gradient image from Figure 9. Red rectangles bound particular identified salient objects.

It is at this later stage any criteria for size constraint on the objects to be detected within the

original scene can be easily applied. As previously discussed this can be derived based on the type of object being looked for (e.g. people, vehicles, debris) and in addition the altitude of the UAV platform. If we are interested in finding only objects of a particular scale – we can discard those below certain size threshold or only identify objects within a certain bounds of size and shape.

## 5. Results

The presented methodology (Section 4) was evaluated over a set of images gathered from a UAV platform developed by Blue Bear Systems Research. Results for saliency detection on a few sample images are presented in Figure 11. As we can see it was tested over varying environmental and daylight conditions as well as varying UAV altitude. Figure 11a and Fig 11b represent images collected over a highly uniform environment with weather and lighting conditions which resulted in relatively low contrast and colour saturation within the resulting image. However despite this we were able to successfully detect all of the objects of interest including the people wearing mock life jackets (within this test scenario). Furthermore three images in Figure 11 also present results obtained for more complex environment. In the image in Figure 11c all objects of interest are detected, together with people wearing bright mock life jackets. Near the right-bottom corner of the image one false positive occurs as a result of noise within the image. However, if we want to be able to detect people at such scale within the presence of noise such false positives must be taken into consideration. In Figure 11d and Figure 11e we can see more cluttered environment with some buildings. As it can be seen in this image – those parts of the images where buildings occur are marked as salient due to uniqueness within the overall scene. This is to be expected given our definition of saliency and also our definition of the intended area of operation for application to UAV search and rescue and surveillance missions.

Automatic Salient Object Detection In UAV Imagery



(a)



(b)



(c)



(d)



(e)

**Figure 11** Results obtained for images collected by UAV operated by BBSR.



(a)



(b)



(c)

**Figure 12** Results obtained for publicly available aerial images from crash sites both on land and on sea.

Due to the limitation in collecting data from specific search and rescue missions a small relevant data set of publically available images from this genre was used in order to more realistically validate

the algorithm. Three example results on such images are presented in Figure 12. Figure 12a and Figure 12b present scenes from a plane crash. No scale constraints were imposed in order to see the outcome of the approach. However, knowledge of the UAV altitude could easily be used to discard smaller objects within Figure 12a and Figure 12b. Lastly, Figure 12c presents an oceanic environment which is entirely different from those presented thus far. As it can be seen every object of interest within this scene is successfully marked as salient. Again, no scale constraints were imposed for this example (Figure 12c). In all examples (Figure 12a/b/c) we see the detected salient objects encircled with a red bounding rectangle.

Although our methodology was originally developed to operate on images obtained over a very uniform environment Figure 13 also shows performance on a highly cluttered environment. Further development of such system might result in enhancement to this methodology to work on images of such complex environments.



**Figure 13** Result obtained for very complex scene.

## 6. Conclusion

A novel methodology of extracting salient objects from images obtained by UAV platforms has been proposed. A saliency map has been created as a Canny-gradient image which is comprised of multiple edge information gathered from several salient feature map images (the mean-shift image, contrast saliency map with histogram enhancement and multiple channel colour feature images (Itti,

Koch and Niebur 1998)). An assumption was made that the methodology would be operating mainly on images of uniform environments within the common UAV search and rescue (or surveillance) mission.

The methodology has been tested over a range of images gathered under different environmental conditions and with varying UAV altitude. The obtained results appear promising as we have been able to detect every possible salient object of interest within these test examples. Currently the only drawback is occasional false positive detection due to the noise in the environment. However as one of the guiding assumptions was to be able to detect people as salient objects from certain higher UAV altitudes these false positives (due to noise) are to be expected within this saliency detection scenario at the present time.

Finally, future work in this area could include the integration of a terrain database to further act as a prior for saliency detection within certain environmental background areas. Additionally aspects of saliency detection within a certain *a priori* search environment or the combination of optical/thermal saliency detection could also be considered.

## 7. References

1. Bruce, N., and J. Tsotsos. "Saliency Based on Information Maximization." *Advances in Neural Information Processing Systems*, Vol. 18, p.155-162, 2006.
2. Canny, J. "A computational approach to edge detection." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, Issue 6, p.679-698, November 1986.
3. Chalmond, B., B. Francesconi, and S. Herbin. "Using Hidden Scale for Salient Object Detection." *IEEE Transactions on Image Processing*, Vol. 15, Issue 9, p.2644-2656, September 2004.
4. Chen, L., X. Xie, X. Fan, W. Ma, H. J. Zhang, and H. Zhou. "A visual attention model for adapting images on small displays." *ACM Multimedia Systems Journal*, Vol. 9, p.353-364, October 2003.
5. Comaniciu, D., and P. Meer. "Mean Shift Analysis and Applications." *Proceedings of the Seventh IEEE International Conference on Computer Vision*, Vol.2, p.1197-1203, September 1999.
6. Gonzalez, R., and R. Woods. *Digital Image Processing*. Prentice Hall 2002.
7. Hou, X., and L. Zhang. "Saliency Detection: A Spectral Residual Approach." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR07)*. 2007.
8. Hu, Y., D. Rajan, and L. Chia. "Robust Subspace Analysis for Detecting Visual Attention Regions in Images." *Proc. Of the 13th ACM international conference on Multimedia*, p.716-724, 2005.
9. Hu, Y., X. Xie, W. Ma, L. Chia, and D. Rajan. "Salient Region Detection using Weighted Feature Maps based on the Human Visual Attention Model." *Advances in Multimedia Information Processing – PCM 2004* p.993-1000, 2004.
10. Itti, L., C. Koch, and E. Niebur. "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.20, p.1254-1259, November 1998.
11. Kadir, T., and M. Brady. "Saliency, Scale and Image Description." *International Journal Of Computer Vision*, Vol.45, p.83-105, 2001.
12. Liu, F., and M. Gleicher. "Automatic image retargeting with fisheye-view warping." *Proceedings of the 18th annual ACM symposium on User interface software and technology*, p.153-162, 2005.
13. Liu, F., and M. Gleicher. "Region Enhanced Scale-invariant Saliency Detection." *IEEE International Conference on Multimedia and Expo*. p.1477-1480, 2006.
14. Ma, Y.F., and H.J. Zhang. "Contrast-based Image Attention Analysis by Using Fuzzy Growing." *Proceedings of the eleventh ACM international conference on Multimedia*. Vol. 1, p.374-381, 2003.
15. Wang, Z., and B. Li. "A Two Stage Approach To Saliency Detection in Images." *IEEE International Conference on Acoustics, Speech and Signal Processing*, p.965-968, 2008.