Intelligent multi-spectral IR image segmentation

Thomas Lu[a], Andrew Luong[b], Stephen Heim[c], Maharshi Patel[b], Kang (Frank) Chen[d], Tien-Hsin Chao[a], Edward Chow[a], Gilbert Torres[e]

[a]NASA/Jet Propulsion Lab/California Institute of Technology, Pasadena, CA, USA; [b]Univ. of Calif. Irvine, CA, USA; [c]Occidental College, CA, USA; [d]Univ. of Calif. Los Angeles, CA, USA; [e]Naval Air Warfare Center, Point Mugu, CA, USA

Abstract

We present a neural network based multi-spectral image segmentation method. A neural network is trained on the selected features of both the objects and background in the longwave (LW) Infrared (IR) images. Multiple iterations of training are performed until the accuracy of the segmentation reaches satisfactory level. The segmentation boundary of the LW image is used to segment the midwave (MW) and shortwave (SW) IR images. A second neural network detects the local discontinuities and refines the accuracy of the local boundaries. The neural net based segmentation method is compared with Wavelet-threshold and Grab-Cut methods. Test results have shown increased accuracy and robustness of this segmentation scheme for multi-spectral IR images.

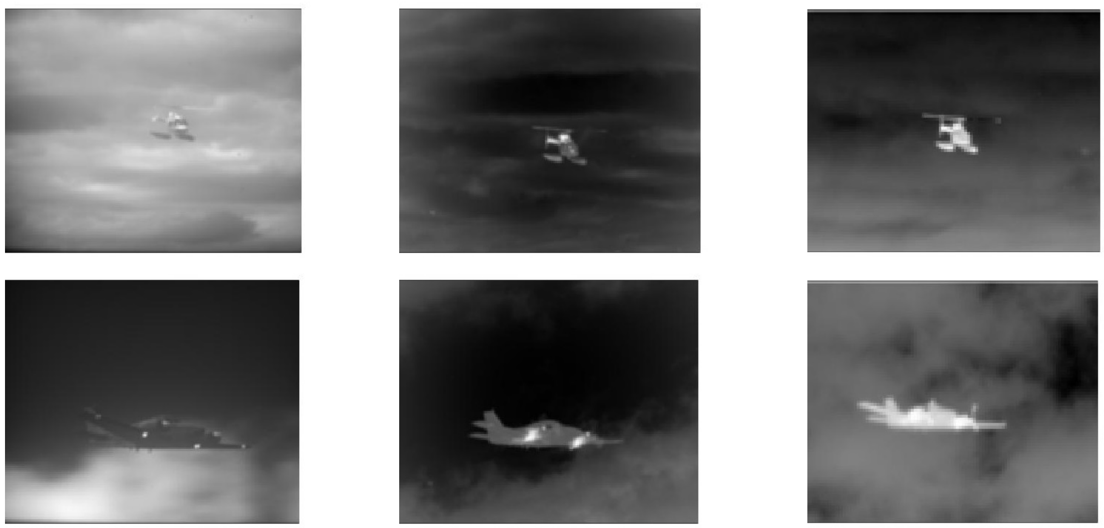
Keywords: Multi-band, Infrared, IR, segmentation, neural network learning, object recognition, classification, Background removal.

1. INTRODUCTION

Computers have been playing an important role in many aspects of our life. However, it is still rather difficult for a computer to recognize objects like human eyes. Computer vision has been an active research area for the past 30 years [1]. The advancement of computer central processing unit (CPU) and graphical processing unit (GPU) technologies have made possible massive parallel processing of artificial neural networks with millions of neurons. Intelligent algorithms such as deep learning are closing the gap between computer and human vision.

One of the key research topics in computer vision is object segmentation. Image segmentation attempts to separate an object from its background. It is quite challenging to separate an object from high background clutter. There are many ways to segment an image. They can be classified into several general approaches. The intensity-based segmentation is one of the simplest approaches to segment an image [2]. It relies on the global and local threshold techniques for separating the main object from the background. If the background were not uniform in intensity, then the threshold would not work. The similarity-based segmentation is also called Region Growing method [3]. It segments the image by grouping neighboring similar pixels into a larger region. The discontinuity-based method performs the segmentation based on detecting major differences in pixel intensity between neighboring pixels [4]. It detects the edge or boundary by finding the gradient and derivative operators. The clustering-based method segments the images by grouping the similar intensity and spatial order [5]. The pixels in the image are allocated to a region based on their distance to its center and intensity. The center points are iteratively changed and updated until coordinates no longer change. It is similar to the similarity-based method. The results of the segmentation process may vary due to user initializations. The graphing-based method represents the image as a graph [6]. It uses an edge detection method to create disjoint and to divide the image into sets or regions using edges as connections between pixels or regions.

In this paper, we present a neural network-based segmentation method that segments targets in multiple band IR images. Figure 1 shows examples of multi-band IR images. We can see the background has different cloud features in each band. The algorithm starts by using the threshold-based method. Since the background is not always uniformly low, the method adds the similarity measures by calculating the mean, standard deviation, max and min values of the sub-regions in the IR images. A neural network is trained to take all of the above parameters as inputs and give a proper classification between the object and the background. Through correlation of unique features in all the bands, the objects in each of the three wavelengths are aligned. The segmentation outline in the LW is then applied to the MW and SW. Another neural network is trained to look for object edges in each wavelength. The discontinuity-based method is used along with the neural network to detect edges accurately. In Section 2, we present the different segmentation approaches such as a Wavelet-filter/intensity threshold method, a Grab-Cut method, and the neural network segmentation methods. The results are compared and analyzed in Section 3.



(a) (b) (c)

Figure 1: Examples of multi-band IR images of a helicopter and a fixed wing aircraft in (a) SW, (b) MW, and (C) LW bands.

2. SEGMENTATION ALGORITHMS

In this section, we discuss the principles of the segmentation algorithms used in the comparison and analysis, namely, the Wavelet filter/intensity threshold, the Grab-Cut, and the neural network-based segmentation methods.

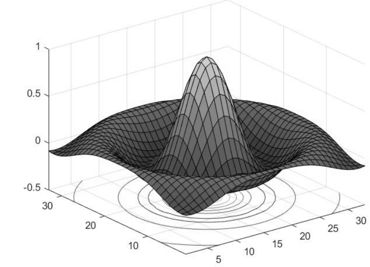


Figure 2: A two-dimensional Mexican Hat wavelet function.

2.1 Segmentation with Wavelet Filter and Intensity Thresholding

When performing image segmentation, intensity thresholding is the first method to try if there is a large distinction between target and background. The intensities of certain target and background regions within raw images are typically quite similar, however, and as such some form of pre-processing is often a necessity. When processing IR images a filtering method is needed to increase general target intensity and suppress background features. This is especially significant for our purposes as, in the case of most SW and some MW images, targets exhibit unwanted details and background features are quite prominent. Additionally, due to variance in target size, the filtering method must have the ability to scale appropriately. A wavelet based filter, shown in Figure 2, seemed fitting for this task, as these offer a modicum of control over which features to enhance based on qualities such as size and intensity. [7] The filtering method implemented in this particular segmentation process utilizes a continuous wavelet transform with the Mexican Hat wavelet. This wavelet is non-directional and as a result can detect potential target features regardless of orientation. The Mexican Hat continuous wavelet transform used to filter images prior to thresholding is given by the equation

$$\varphi(\omega_x, \omega_y) = -2\pi(\omega_x^2 + \omega_y^2)^{\frac{p}{2}} e^{-\frac{(\sigma_x\omega_x)^2 + (\sigma_y\omega_y)^2}{2}} \qquad \sigma_x, \sigma_y \in R, p > 0 \tag{1}$$

in which σ_x and σ_y can be adjusted to manipulate the scale of the wavelet used in the transform. Greater scaling values stretch the wavelet and allow for enhancement of larger features with less detail. Smaller scaling values compress the wavelet and result in the enhancement of smaller and more detailed features.

After enhancing the desired features a simple threshold function is applied to the filtered image. An example of the conversion from wavelet-filtered images to binary images in all three bands is shown in Figure 3 below. We can see apparent noise and background features misclassified as the target in all three bands.

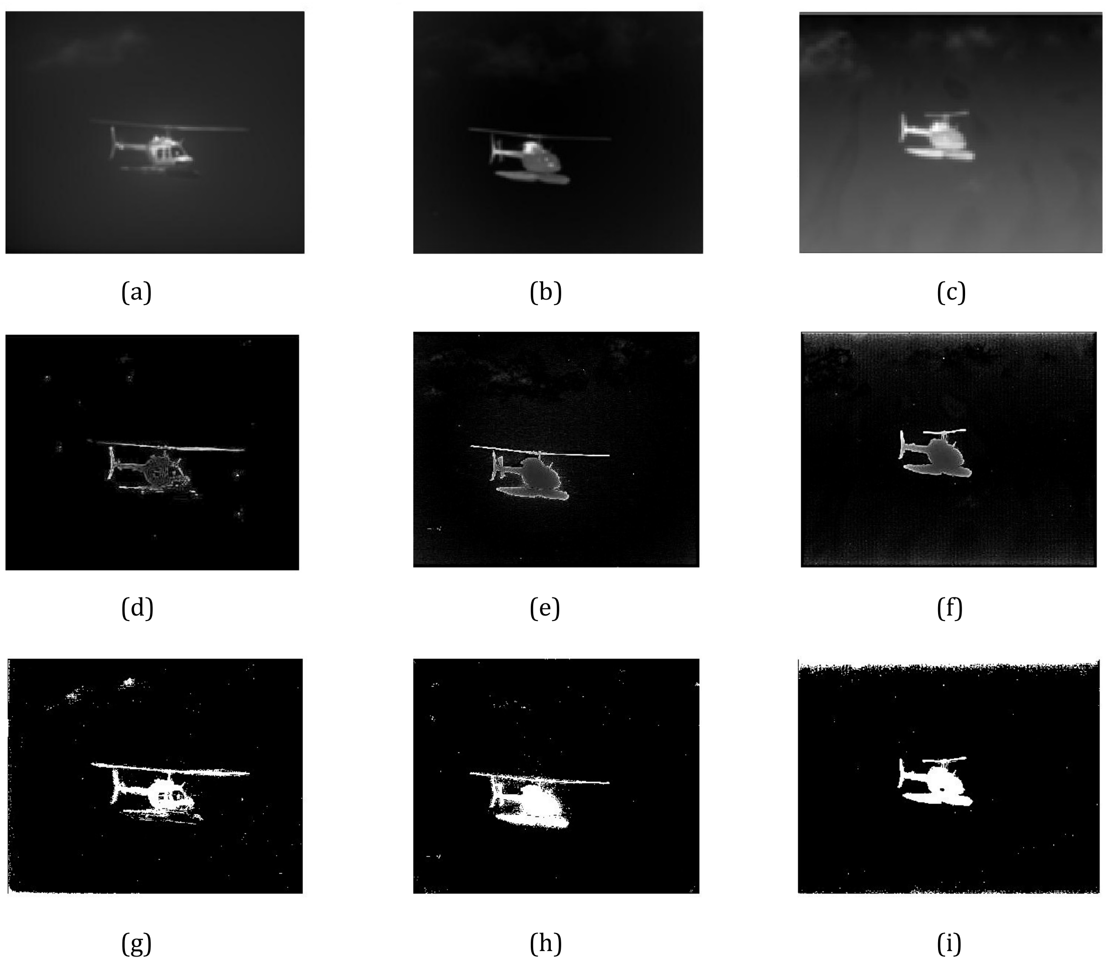
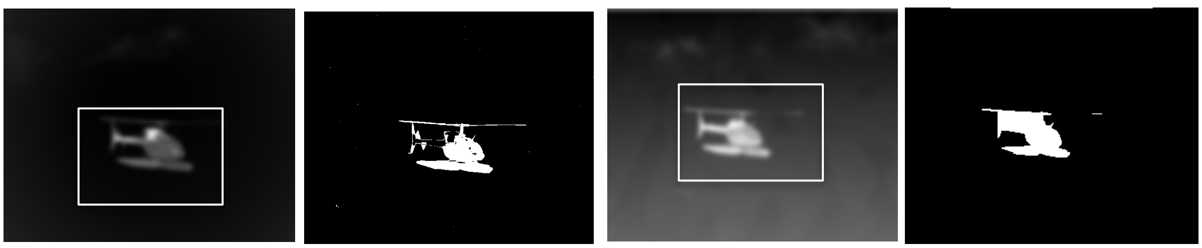


Figure 3: Original multi-band images in all three bands (a) SW, (b) MW and (c) LW; (d), (e) and (f) Wavelet filtered images in three corresponding bands; (g), (h) and (i) their corresponding binary images after thresholding.

2.2 Grab-Cut Segmentation

In image analysis, efficient extraction of foreground and background is of great practical importance. Grab-Cut [6] is an interactive foreground extraction algorithm using an iterated graphing method. It is designed to solve the "Min Cut/Max Flow" problem. An energy cost function is defined by creating a specific graph model. The energy function has the following inputs: input image, and a bounding box drawn by a person, which is the label that defines whether a pixel belongs to the foreground or background. Grab-Cut encourages neighboring pixels to have the same label and certain color distribution. [8, 9]

Initially the user draws a rectangle around the foreground region, inside which the target must be completely contained. The outer part of the rectangle will be defined as the definite background, while the inner of the rectangle contains the unknown combination of foreground (target) and background. Then the Grab-Cut algorithm segments the image iteratively to get the best result. As each frame is processed by Grab-Cut, the resulting bounding box is used in the next frame, automating the testing process.



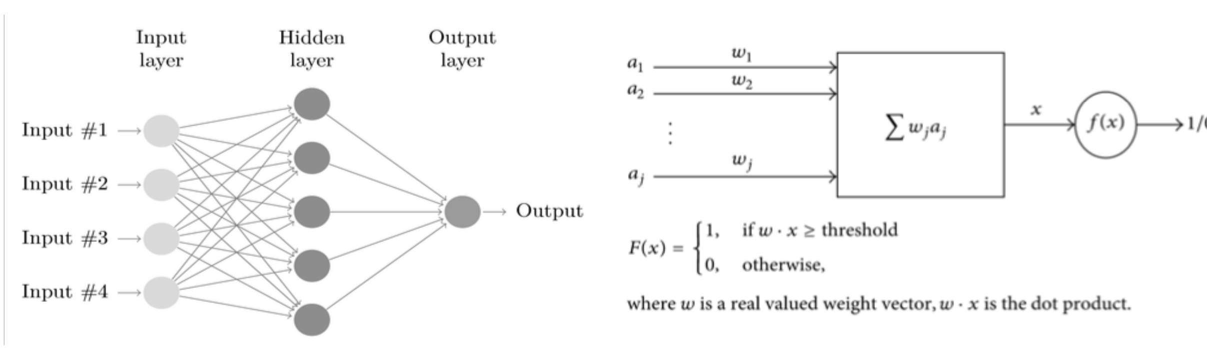
(a) (b) (c) (d)

Figure 4: Examples of Grab-Cut: (a) A MW and (c) a LW image with bounding box; (b) & (d) Segmentation results by Grab-Cut.

From the results in Figure 4, we can see that the bounding box needs to be drawn by a person, which is not convenient. In addition, the Grab-Cut segmentation method still cannot make a clear-cut of the object from the background.

2.3 Neural Network-Based Segmentation

Inspired by human brain operations, the artificial neural network is capable of non-linear classification through a training process [10]. The advantage of a neural network is its ability to adapt to new environments through training with new data. We have used a multi-layer feedforward neural network to perform segmentation for the target from the background in the LW images. Figure 5(a) illustrates a typical three layer neural network. The first is the input layer. The features are input into the neural network. The hidden layer finds the associations of the input features. The output layer gives the confidence of the decision on either “Target” or “Background”.



(a) (b)

Figure 5: A three layer feed-forward neural network: (a) illustration of the neural network architecture; (b) mathematical operation of a neural network.

The mathematical formula of the neural network can be expressed in Figure 5(b). The input of a neuron in the next layer is a weighted sum of all neurons in the previous layer. The weights are determined by the "Learning" algorithms of the neural network. A backpropagation learning method is used to train the neural network. The input features are the "Mean", "Max", "Min", "STD", "Distance to the Center", and the pixel intensity value.

As can be seen in the IR images in all three bands, it is difficult to segment the target of interest only based on intensity or a set of simple local pixel features, as clouds reflect long waves in similar amount to the target. So we use the context around the pixel to increase the accuracy of the classification. As such, we integrated texture information of a pixel's surrounding area as a feature by considering each pixel (x, y) with a window of 3x3 with (x, y) being in center. We computed minimum and maximum intensity of the pixels in the 3x3 window, along with average and standard deviation of intensity and used them as features. We utilize the fact that the farther away a pixel is from the center of the mass of the target, the less likely it is part of the target by adding an additional feature that represents distance of the pixel from the center of mass of the target.

We used tiling to reduce the execution time for the neural network-based segmentation. With tiling, we divide the image into a 4×4 square and compute standard deviation of the pixels in that tile. If standard deviation is below a threshold we pick a random point in the tile, perform classification, and use the result of the classification for all the other points in that tile. Using tiling increased the efficiency of segmentation significantly.

After producing a binary mask for the LW image using neural network segmentation, the mask is transformed and aligned to the corresponding SW and MW frames using common correlation points. These points are user selected in the starting frame of a given video, and then automatically tracked using a correlation algorithm in subsequent frames [11]. The SW and MW images are then segmented using this altered LW mask.

Using the neural network to train on the features of the "Target" and "Background", we have improved segmentation accuracy and reliability. In Figure 6, we show examples of the segmentation of the LW/MW/SW images using the neural network. Despite the cloud in the background and fuzzy edges is in the images, the neural network accurately recognized the outlines of the object in all three bands. Only parts of the rotor blades were missing due to the low reflectance of the rotating blades in those sections.

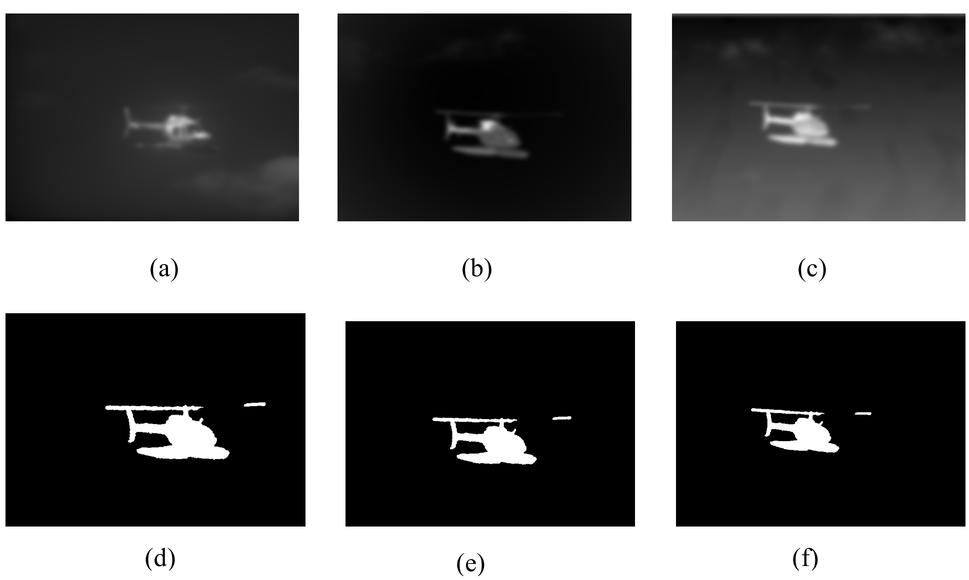


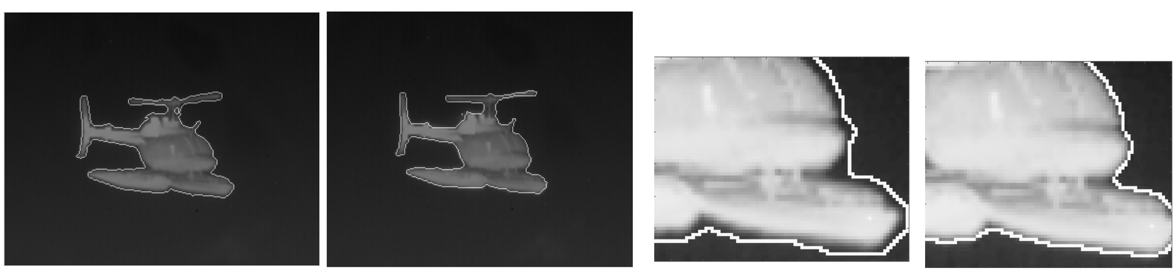
Figure 6: Neural network-based segmentation: Original IR images: (a) SW, (b) MW and (c) LW; Neural network-based segmentation masks in all three bands (d), (e) and (f).

2.4 Local Edge Search using Neural Network

While the neural network-based segmentation method produces fairly accurate segmentation masks for LW images, the segmentation results in MW and SW may not be accurate enough due to imperfect alignment among the bands. To get a more accurate segmentation in MW and SW images, we apply a local edge search algorithm to the MW and SW images with the guidance of the LW mask. The local edge search algorithm takes into account the LW segmentation mask and a few other features to find more accurate edges in the MW and SW images. The algorithm basically creates a tighter segmentation that is closer to the true edge, and picks up small details that the LW mask otherwise would not pick up in the MW and SW images.

The search function of the algorithm performs the calculations to choose the pixel that most likely is the true edge in the original image. The search begins by finding the edge point in a local region that is perpendicular to the slope of the pixel being searched. For each point in the local region, the function calculates features that are passed to a neural network that outputs a value from 0 to 1 that represents the probability of that point being the true edge.

The first feature is the intensity change at the local region point. To find the intensity change, we apply an edge filter to the local region point in the perpendicular direction. The second feature is the difference between the intensities of the current local region point and the previous true edge point. The third feature is the distance between the current local region point and the previous true edge point. The fourth feature is the change in angle among the previous three true edge points. The fifth feature is the difference in angle between the previous LW edge points and previous true edge points. We pass these five features to a neural network. This process is done on each point in the local region and we choose the edge point with the maximum class value. Figure 6 shows an example of using the local edge search to improve the accuracy of an outline. Comparing the zoom-in outlines in Figures 6(c) and 6(d), we can see that the local edge search is more accurate than using the neural network-based segmentation alone. Section 3 shows the statistical results of the different segmentation methods.



(a) (b) (c) (d)

Figure 6: Comparison of local edge search segmentation result: (a) Neural network segmentation outline; (b) Outline after local edge search; (c) Zoomed in part of outline before local edge search; (d) Zoomed in outline after local region. Tighter outline is achieved by local edge search.

3. EXPERIMENTAL RESULTS AND ANALYSIS

We have tested and compared the results of each segmentation method using our database of hand-segmented images as a ground truth. The database contains 7 images of 640x512 resolution from 10 videos. Each video includes SW, MW, and LW bands for a total of 210 test images, 70 in each band. Error results are given in percent image difference, calculated by the equation

$$Error\ (\%\ Image\ Difference) = \left(\frac{False\ Positives + False\ Negatives}{Total\ Image\ Pixels}\right) \times 100\% \qquad (2)$$

in which the sum of false positives and false negatives calculates the number of pixels which are dissimilar between the ground truth image and the image generated from a specific segmentation method. The segmentation methods to be compared are as follows:

1. Wavelet filter and intensity thresholding
2. Grab-Cut - Iterated graph cuts
3. LW neural network-based segmentation and alignment with MW & SW images
4. LW neural network segmentation and alignment with local edge search

All segmentation methods were tested on a common testing set of SW, MW, and LW images for which we have a corresponding image in our ground truth database. These images contain target objects with areas ranging from small (3200 pixels) to large (41000 pixels) in both cloudy and clear environments.

Figure 7 shows the testing results and the comparison chart of the four segmentation methods: (1) Wavelet+Threshold, (2) Grab-Cut, (3) Neural network-based segmentation + alignment, and (4) Neural network-based segmentation + alignment + local edge search. We analyze the testing results in the following sub-sections.

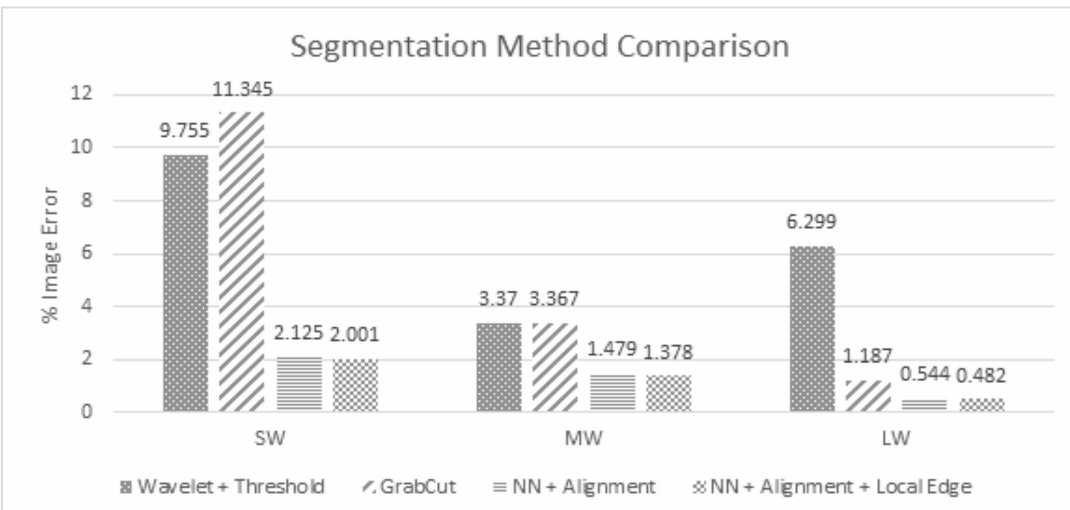


Figure 7: Comparison chart of percent image error of all tested segmentation methods: (1) Wavelet+Threshold, (2) Grab-Cut, (3) Neural network-based segmentation + alignment, and (4) Neural network-based segmentation + alignment + local edge search. The neural net segmentation and cross band alignment method with local edge search are consistently the most accurate across all bands.

3.1 Wavelet Filter and Intensity Thresholding

When performing filtering and thresholding on each set of test images, the scaling values σ_x and σ_y as well as the luminance threshold were manually optimized. Separate optimizations were required for each video as well as each band. Though the wavelet filter was able to enhance the target based on the user defined scaling values, background features were often enhanced as well. As a result, these regions of background fell above the luminance threshold and were included in the corresponding binary images. When compared to the manual segmentations of the IR images, all three bands had a relatively high average error due to frames in which background regions exhibit similar qualities to the desired target as shown in Figure 8. The average error rate for all bands is around 6.47%.

(a) (b) (c)

Figure 8: Images illustrating limitations of the Wavelet filter-threshold method: (a) Original image; (b) Wavelet filtered image; (c) Binary image after threshold. Notice the misclassified background region due to similar features between target and cloud.

3.2 Grab-Cut Segmentation

Grab-Cut generally performs better with LW images than with MW and SW images, due to their distinct pixel difference between foreground and background. However, cases where certain features of the target, such as the tail and landing gear, often confused Grab-Cut into classifying the inner area as a target. In addition, MW and SW images posed significant challenges due to foreground and background similarities. Furthermore, automating Grab-Cut to run on multiple frames of the image was especially difficult in SW images due to its interactive foreground/background correction. Inaccurate boundary tracking often made the image segmentation obsolete. The average error for the Grab-Cut segmentation method is around 5.30%, slightly better than the threshold method.

3.3 Neural Net Segmentation and Cross-Band Alignment

The segmentation on the testing set of 70 LW images was performed using a neural net trained with sample points from a training set of 40 LW images. The training set was selected from the videos containing the testing set, but the frame sets are completely disjoint. As the initial segmentation was performed directly on the 70 LW test frames, the average error (0.54%) for this band is the lowest of the three. Differences in target orientation and level of detail present an increased percentage of error when segmenting the corresponding SW and MW images with transformed LW masks.

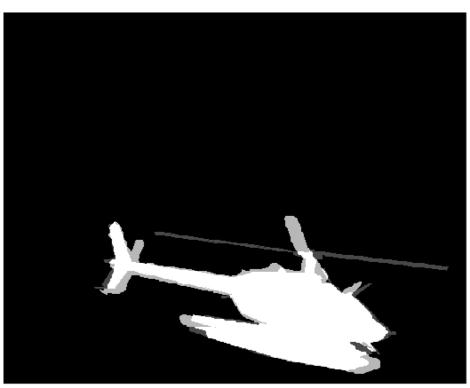


Figure 9: Image overlay of neural net segmentation + alignment and the ground truth segmentation representing the misaligned rotors in MW. Gray areas are dissimilar regions while white areas are similar regions.

When aligning the LW segmentation with MW and SW, inconsistencies between the bands due to asynchronous video contribute sizably to the total error. The largest sources of error are typically rotor blades, as it is not possible to align these features using our current methods. An example of this misalignment is shown in Figure 9. The misaligned rotor blades contribute up to 1.16 percent error. The overall average error for the neural network based segmentation method is 1.38%, over 70% better than the two previous methods.

3.4 Neural Net Segmentation with Local Edge Search

The Local Edge Search algorithm decreases the percent error for all three wavelengths. The most apparent improvement is with the LW images because there isn't the problem of having to transform the mask before performing local edge search. An issue with the algorithm is with MW and SW images. Because it uses the LW transformed masks as the guideline, the performance really depends on how well the LW outline and the MW and SW images are aligned. If there are large discrepancies between the transformed LW mask and the MW and SW images, the algorithm will likely be unable to find the true edge for those parts of the image. The overall performance of this method is the best, achieving a low average error rate of 1.28%.

Upon analyzing and comparing the results of all tested segmentation methods, it is clear that the neural network segmentation in conjunction with cross-band alignment produces an accurate segmentation across all bands, more so than either wavelet filtering and thresholding or Grab-Cut segmentation. Additionally, the local edge search algorithm further improves the accuracy of this method. Perhaps the most notable strength of the neural net segmentation and cross band alignment method is its ability to consistently ignore background features, which are captured by the other methods of segmentation. As the neural net is trained to classify such features as background in LW frames, the corresponding MW and SW segmentations performed with transformed and aligned LW masks are also devoid of similar background misclassifications. Additional work needs to be done in training the neural network for identifying the edges of the rotor blades in the MW and SW images.

4. CONCLUSIONS

From the testing we can see that in multi-spectral IR images, the shorter the wavelength, the more difficult to segment the object from the background due to variations of reflectance from the objects and the background. We have presented a novel image segmentation method using a neural network for training the texture of the LW images and then guide the segmentation of the MW and SW images, while a second neural network is used to find the local edge information in the MW and SW images. The neural network is a powerful non-linear classifier. It has the ability to be trained by image samples directly, which makes it robust and adaptable to various environments. It is also convenient to re-train a neural network when presented with new objects or new background clutters. We have tested the neural network-based segmentation method in comparison to the Wavelet-threshold and Grab-Cut methods. Both neural networks have showed better performance in locating fuzzy edges in the multi-band IR images than the other two methods. Test results have shown increased overall accuracy and robustness of the neural network based segmentation scheme for multi-spectral IR images.

ACKNOWLEDGEMENTS

The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology and was supported in part by the U.S. Department of Defense, Test Resource Management Center, Test & Evaluation/Science and Technology (T&E/S&T) Program under NASA prime contract NAS7-03001, Task Plan Number 81-12346. We thank Kevin De Jesus, Alexander Huyen, Jessi, Bustos, Kevin Payumo, Sebastian Stone and Jorge Serrano for constructive suggestions and their help in testing of the segmentation methods.

REFERENCES

[1] Lu, T., and Mintzer, D., "Hybrid neural networks for nonlinear pattern recognition," Optical Pattern Recognition, ed. by F. T. S. Yu & S. Jutamulia, Cambridge University Press, (1998).
[2] S. Nagabhushana, "Computer Vision and Image Processing", New Age International Publishers, 2005.
[3] Bryan S. Morse, Lecture 18: Segmentation (Region Based), Brigham Young University, 1998.
[4] Ravi S and A M Khan, "Operators Used in Edge Detection: A Case Study", International Journal of Applied Engineering Research, ISSN 0973-4562 vol. 7 No 11, 2012.
[5] K. Duraiswamy and V. Valli Mayil "Similarity Matrix Based Session Clustering by Sequence Alignment Using Dynamic Programming" , Computer and Information Science journal, Vol. 1, No.3, 2008.

[6] C Rother, V Kolmogorov, A Blake, “Grabcut: Interactive foreground extraction using iterated graph cuts”, ACM transactions on graphics (TOG), 2004
[7] Tan, L., Ma, J., Wang, Q., and Ran, Q., “Filtering Theory and Application of Wavelet Optics at the Spatial-Frequency Domain,” Appl. Opt. 40, 257-260, (2001)
[8] Boykov, Y., and Jolly, M.-P. 2001. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In Proc. IEEE Int. Conf. on Computer Vision.
[9] "OpenCV: Interactive Foreground Extraction Using GrabCut Algorithm." OpenCV: Interactive Foreground Extraction Using GrabCut Algorithm. N.p., n.d. Web. 26 July 2016.
[10] Ms. Sonali. B. Maind, Ms. Priyanka Wankar, “Research Paper on Basic of Artificial Neural Network”, International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), ISSN: 2321-8169, 96-100, (2014)
[11] T. Lu, T. Chao, K. Chen, A. Leung, M. Dewees, X. Yan, E. Chow, G. Torres, "Cross--correlation and image alignment for multi--band IR sensors", Proc. SPIE, Vol 9845 (2016).