JOURNAL OF SPACECRAFT AND ROCKETS
Vol. 56, No. 6, November–December 2019

Engineering Notes

Unsupervised Method of Infrared Spacecraft Image Foreground Extraction

Jian-Feng Shi* and Steve Ulrich†
Carleton University, Ottawa, Ontario K1S 5B6, Canada
and
Stéphane Ruel‡
Neptec Design Group Ltd., Kanata, Ontario K2K 1Y5, Canada

DOI: 10.2514/1.A34449

Nomenclature

$\mathcal{B}(\boldsymbol{x}, K_B)$	=	box filter operator
$\mathcal{BE}(\boldsymbol{x})$	=	extends the inner border pixels to the image boundary
C_b	=	weight factor for blue channel
C_g	=	weight factor for green channel
C_r	=	weight factor for red channel
$\mathcal{C}(\boldsymbol{x})$	=	closed regions in an image as filled areas
C_L	=	CLAHE clip limit
$\mathcal{CL}(\boldsymbol{x}, C_L)$	=	CLAHE function
$\mathcal{CTG}(\boldsymbol{I})$	=	convert color image into standard grayscale image
c_i	=	cumulative distribution for the ith intensity histogram cluster
$\mathcal{F}$	=	coordinate frames
$\mathcal{GTC}(\boldsymbol{S}_r, \boldsymbol{S}_g, \boldsymbol{S}_b)$	=	combine three-channel image into color image
$\mathcal{HE}(\boldsymbol{x})$	=	histogram equalization
$\boldsymbol{I}$	=	input color image
$\boldsymbol{I}_b$	=	input blue-channel image
$\boldsymbol{I}_g$	=	input green-channel image
$\boldsymbol{I}_r$	=	input red-channel image
J	=	Jaccard Index, also known as *intersect over union* (IoU)
$\bar{J}$	=	mean Jaccard index
$\boldsymbol{K}_B$	=	box filter kernel
K_B	=	size of the box filter kernel
$\boldsymbol{K}_D$	=	elliptical dilation kernel structuring element
K_D	=	size of the dilation kernel
$\boldsymbol{K}_{sx}$	=	3×3 Scharr x-axis kernel
$\boldsymbol{K}_{sy}$	=	3×3 Scharr y-axis kernel
$\mathcal{MD}(\boldsymbol{x}, K_D)$	=	morphological dilation
$\mathcal{M}(\boldsymbol{x})$	=	image intensities greater than the mean value
$\mathcal{N}(\boldsymbol{x})$	=	image nomalization
$\mathcal{O}(\boldsymbol{x})$	=	Otsu thresholding function
P_i	=	probability distribution for the ith intensity histogram cluster
$\boldsymbol{R}$	=	rotation from target body frame to camera frame
$\boldsymbol{S}_{bf}$	=	box filtered image
$\boldsymbol{S}_b$	=	FC-HSF blue-channel response map
$\boldsymbol{S}_D$	=	dilated prior mask image
$\boldsymbol{S}_e$	=	edge response map
$\boldsymbol{S}_g$	=	FC-HSF green-channel response map
S_h	=	histogram equalization adjusted pixel image
$\boldsymbol{S}_{K-1}$	=	prior image mask
$\boldsymbol{S}_r$	=	FC-HSF red-channel response map
$\boldsymbol{t}$	=	translation vector from camera to target body expressed in camera frame
X	=	x-axis direction
x	=	ground truth foreground pixels
$\boldsymbol{x}$	=	input image
Y	=	y-axis direction
y	=	predicted foreground pixels
Z	=	z-axis direction
$\tilde{\nabla}(\boldsymbol{x})$	=	image gradient approximation
μ_i	=	mean for the ith intensity histogram cluster
σ_i	=	standard deviation for the ith intensity histogram cluster

Subscripts

b	=	body frame
c	=	camera frame

Presented as Paper 2018-5281 at the 2018 AIAA SPACE and Astronautics Forum and Exposition, Orlando, FL, 17–19 September 2018; received 22 December 2018; revision received 28 June 2019; accepted for publication 8 July 2019; published online 20 August 2019.

*Ph.D. Candidate, Department of Mechanical and Aerospace Engineering, 1125 Colonel By Drive. Senior Member AIAA.

†Associate Professor, Department of Mechanical and Aerospace Engineering, 1125 Colonel By Drive. Senior Member AIAA.

‡Vice President of Innovation, 302 Legget Drive.

I. Introduction

SPACECRAFT guidance, navigation, and controls for on-orbit servicing require the chaser vehicle to have precise knowledge of the target satellite relative pose. Visual targets and sensor hardware used for relative pose measurements can be active or passive systems. Although active sensors have higher precision and robustness, they require installing fiducial markers on the target spacecraft, which is challenging to implement for most space-flight missions. Passive systems such as cameras or LIDAR do not need to modify the target vehicle. Space cameras provide an economical alternative to LIDAR systems with less mass, power, and geometric constraints. Monochromatic images captured by a monocular camera can be used to determine the relative pose between space vehicles. Some examples of using camera-based spacecraft navigation are works by Wu et al. [1], who use direct projection point matching minimization using four points defined by cooperative LED lighting; this method requires fiducial markers to be installed on the target and can be prone to errors caused by image conditions. Fourie et al. [2] use a stereo camera for depth perception and position; this method computes a point cloud map to estimate the target geometric center; however, the orientation is not estimated in the work of Fourie et al., and overall computation time is over 300 ms per frame, which requires both target and chaser vehicles to move slowly to operate in real time. Tweddle and Saenz-Otero [3] use four-point correspondence geometry of Haralick et al.'s iterative nonlinear solution [4]; it is another variation of the direct projection point matching minimization. Their [3] approach requires cooperative circular markers to be installed on the target; their work focused on the back-end pose matching and stochastic filtering but did not provide details on segmentation using adaptive threshold and marker recognition. In our experience we found that this front-end

image processing can cause significant errors to the back-end pose estimation. Sharma et al. [5,6] use Hough transform [7] for straight edge detection and use multiple confirmations to strengthen the feature recognition; they then use the efficient perspective-*n*-point [8] and refinement using the Newton-Raphson method [9]. Similarly, the front-end feature-recognition-based features can be unstable. Although the multiple confirmation processes may increase robustness, image error from environmental conditions in low-level line features cannot be avoided.

Regional-based pose estimation has exploited the color distinctiveness to perform object segmentation [10]; however, in most relative navigation systems, the captured image is typically in the grayscale due to data or hardware restrictions. Single-channel grayscale images are more difficult to process than color images because they lack the distinctiveness from color depth. Top-down machine learning methods train classifiers to use image features [11] or computationally develop convolutional neural network perception filters [12]. Top-down approaches, however, require numerous labeled images or image features, and in some scenarios, training can be difficult to stabilize. By contrast, image-driven saliency detection techniques can be faster and simpler to implement. Saliency is important for pose estimation especially in the regional-based approach, where the main target can be identified and segmented from the surrounding background without fiducial markers or other visual aid. Zhang et al. [13] provide a survey of unsupervised methods in image segmentation. The recent developments in convolutional neural networks and generative adversarial networks (GAN) may also be used to perform image transformation [14]. Other deep learning methods in pose estimation have been developed recently [15–17]. Although the deep learning techniques are powerful, some techniques such as GAN are difficult to train, the memory needs can be large, and the quality of the training images can be demanding. Three recent real-time-capable non-deep-learning saliency detection methods are regional contrast (RC) [18], minimum barrier distance (MBD) [19], and graph manifold ranking (GMR) [20]. The RC method was proposed by Cheng et al. [18] to extract saliency from local regions as the weighted sum of color contrast; RC can be computed efficiently and can produce precise foreground from color images. Zhang et al. [19] proposed a fast raster-scanning algorithm to approximate the MBD transform [21]. A central idea in the MBD is the *image boundary connectivity cue,* which assumes that background regions are connected to the image border, a variation of center surround [22]. An extended version, MB+, was proposed by adding *image boundary contrast* map using border pixels as color contrast seeds in the whitened color space. Finally, Yang et al. [20] proposed GMR saliency using document ranking; GMR optimized a graph-based ranking cost function while replacing documents with image superpixels. The aforementioned saliency detection method has better precision working with colored images; however, their performance degrades if the image is monochromatic. Grayscale image saliency includes using orientation of the image gradient [23] and image texture [24]. The fst and fst+ saliency detection [25] for grayscale images have successfully extracted the spacecraft foreground from a cluttered background when passing over the Earth; however, certain images are still too challenging, for example, the thermal images from this investigation. We develop an unsupervised image alteration to improve the regional-based pose estimation.

The regional-based pose estimation evolved from the image segmentation problem. Typical solutions to the image segmentation problem include complex top-down classification [26] or simple bottom-up image-driven techniques [27]. Whichever the method, many believe that segmentation is too high-level task for purely image-driven methods to succeed [28]. Seminal papers by Kass et al. [29], and Mumford and Shah [30] introduced concepts of energy-minimizing active contours (Snakes) and the combined minimization of regional smoothness and boundary length, respectively. Caselles et al. [31] and Chan and Vese [32] followed by adopting an implicit region boundary representation using the level-set function pioneered by Osher and Sethian [33]. The level-set function provided advantages of implicit boundaries without specific parameterization and evolving embedding functions that can elegantly describe boundary topology changes such as splitting and merging. Additional contributions by Kervrann and Heitz [34], Zhu et al. [35], and Leventon et al. [36] instilled maximum likelihood and maximum a posteriori criterion to estimate the region boundary with statistical parameterizations. Leventon et al. [36] and Rousson and Paragios [37] further stipulated the use of principal components analysis (PCA) shape template to align the regional boundary. Cremers [38] extended the use of kernel PCA [39] to the PCA template and used affine flow fields [40] in discriminating the object from its background. Brox and Weickert [41] and Bibby and Reid [42] refined the shape prior, pixelwise posteriors and proposed a framework for segmenting multiple regions. Rosenhahn et al. [43] combined 3D pose estimation with the level-set formulated region segmentation. The two approaches complemented each other to produce a framework for the region-based pose estimation [44,45].

The level-set segmentation regional-based pose estimation approach [10] uses the foreground and background probability before classifying the working image intensity histogram. Foreground and background histograms are used to make distinctions between the target and the background according to their intensity profiles. To this end, we introduce an innovative false-colored high-frequency salient feature (FC-HSF) image alteration method to enhance the foreground and background histogram distinction in a thermal image. The performance of the proposed method is evaluated in experiments with a laboratory-scale model of Radarsat spacecraft spinning about a single axis.

II. False-Color High-Frequency Saliency Feature Image

Foreground extraction is challenging for some IR images; an example is the Radarsat Satellite Model (RSM) video, where a miniature imitation of the Radarsat is spinning about the pitch axis against a laboratory curtain backdrop. A sample image from the RSM video sequence and its foreground extraction maps are provided in Fig. 1. Figure 1a is the original monocular infrared image captured under laboratory heating conditions. Figure 1b is the ground truth mask for the input image. Figure 1c is the PWP3D [10] probability posterior computed using the *prior* mask histogram template. A three-dimensional (3D) projection mask based on the initial ground truth pose is used as the *prior* mask. Figures 1d, 1e, and 1f are the saliency map computed using GMR [20], MB+ [19], and RC [18] methods, respectively. Figures 1g and 1h are the saliency map and foreground mask computed using the fst [25] and fst+ [25], respectively. Figures 1i–1k are the FC-HSF red, green, and blue channels for Scharr edge gradient, extended border, and region surround maps, respectively. Figure 1l is the final combined FC-HSF image. Figure 1m is the normalized foreground and background FC-HSF image histogram. Figure 1n is the PWP3D [10] probability posterior computed using the FC-HSF image. Figure 1o is the estimated 3D model mesh projection on the original input image. Figure 1p is the posterior mask of the 3D projection. The various saliency methods and developed models failed to generate a precise RSM foreground response as shown in Figs. 1c–1h for level-set probability posterior [25], GMR [20], MB+ [19], RC [18], fst [25], and fst+ [25] methods, respectively. Furthermore, because the foreground and background image intensities are near each other, different pose may have similar histogram profile as the initial template. An example is shown in Fig. 2, where the first column starting from row 2 shows the level-set pose estimation [25] gradient descent as 3D mesh projections on the input image for various iteration steps. The second column in Fig. 2 shows the foreground and background image histogram profiles for the respective iteration step. After four iterations the gradient descent converged to an incorrect local minimum pose, and the histogram profile of the converged pose is similar to the initial template histogram profile in Fig. 2 column 2 row 1.

To increase distinctiveness in the histogram profile and image, we add the gradient and gradient region to the input image, $\boldsymbol{I}$, as red and blue color channels. Because this method uses all three color channels to generate a false-colored image and uses the high-frequency content of the image, we call this approach the false-colored high-frequency saliency feature (FC-HSF) image. The procedure for creating the FC-

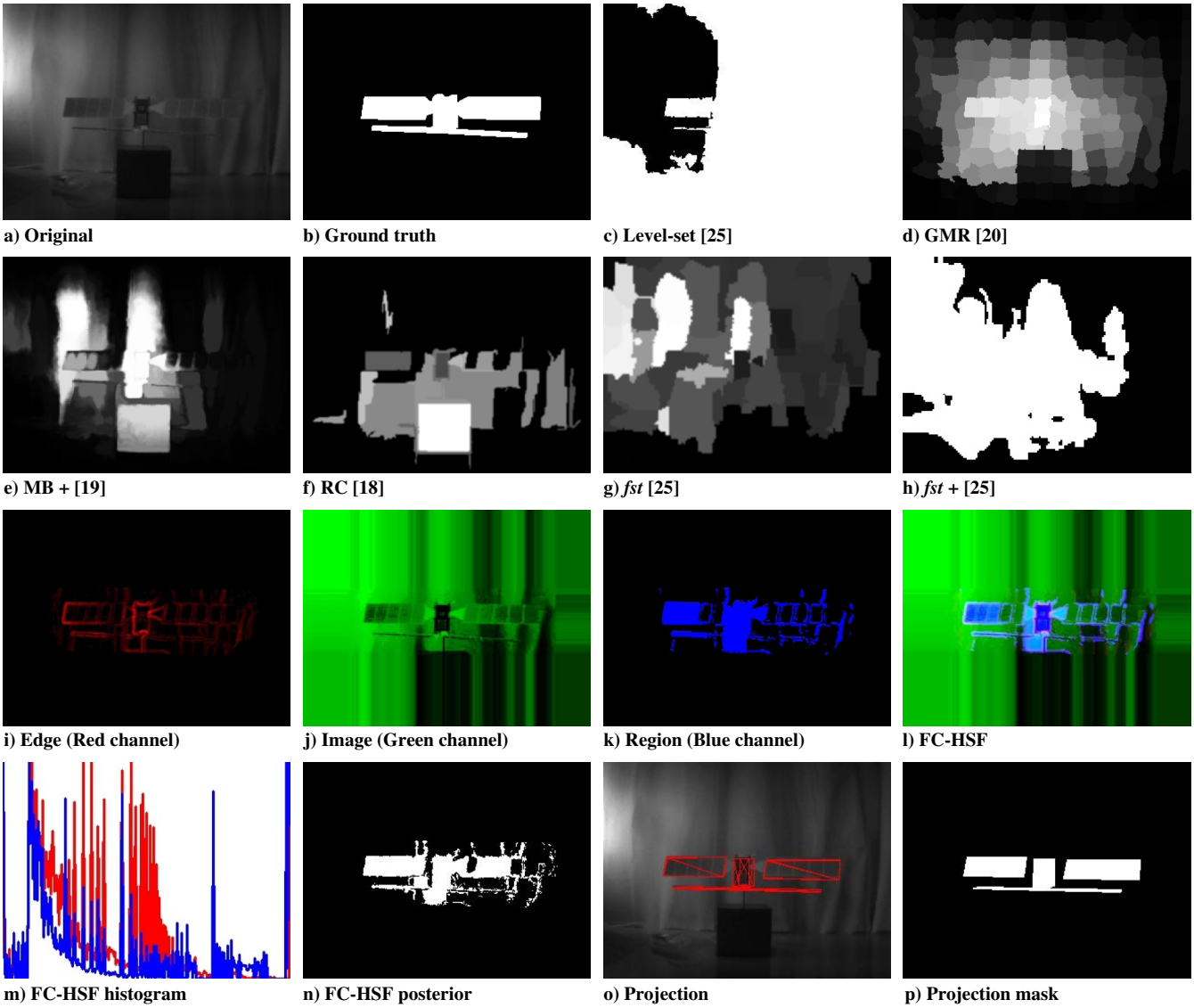


Fig. 1 RSM infrared image preprocessing and level-set-based pose estimation. Methods (*c*) to (*h*) do not produce good foreground mask due to lack of distinctiveness in the monochromatic image. Combining the image and its gradient improves separation of the foreground and background image histogram in (*m*) for better posterior foreground mask generation in (*n*).

HSF is provided as follows: let us define $\boldsymbol{S}_{K-1}$ as the prior mask; we dilate the prior mask to create a probable region of interest to clear away the clutters in the background. We cannot use the prior mask directly, because it will cause the pose estimation to converge to the prior foreground instead of the actual input. We define the morphological dilation function, $\boldsymbol{S}_D \leftarrow \mathcal{MD}(\boldsymbol{x}, K_D)$, as

$$\boldsymbol{S}_D = \boldsymbol{x} \bigoplus \boldsymbol{K}_D = \bigcup_{k \in \boldsymbol{K}_D} \boldsymbol{x}_k \tag{1}$$

where $\bigoplus$ is the dilation morphology operator, $\boldsymbol{x}$ is some arbitrary image, $\boldsymbol{x}_k$ is the translation of $\boldsymbol{x}$ by k, $\boldsymbol{K}_D$ is a $K_D \times K_D$ elliptical dilation kernel structuring element that is set to $K_D = 36$ pixels, and $\boldsymbol{S}_D$ is the resulting dilated mask image. We define $\boldsymbol{S}_g \leftarrow \mathcal{CTG}(\boldsymbol{I})$ as a function that converts a color image into grayscale image as follows:

$$\boldsymbol{S}_g = 0.299 \cdot \boldsymbol{I}_r + 0.587 \cdot \boldsymbol{I}_g + 0.114 \cdot \boldsymbol{I}_b \tag{2}$$

where $\boldsymbol{I}_i$, $i \in \{r, g, b\}$, is the red, green, blue channels of the input image, respectively. The contrast limited adaptive histogram equalization (CLAHE) [46] operation balances the image intensity in the entire local regions and enhances contrast without being influenced by extreme global intensities. Let us define $\mathcal{CL}(\boldsymbol{x}, C_L)$ as the CLAHE function, where $\boldsymbol{x}$ is some arbitrary grayscale image and C_L is the clip limit, and we set C_L to 4 in our implementation. We also perform the box filter operator as $\boldsymbol{S}_{bf} \leftarrow \mathcal{B}(\boldsymbol{x}, K_B)$ on the CLAHE filtered image to limit noise. The box filter uses a square matrix kernel to filter the input image

$$\boldsymbol{S}_{bf} = \boldsymbol{K}_B * \boldsymbol{x} \tag{3}$$

where $*$ is the convolution operator and $\boldsymbol{K}_B$ is a $K_B \times K_B$ square kernel matrix of all ones divided by K_B^2; we use $K_B = 3$ in our implementation. Faster implementation of Eq. (3) uses a four-point formulation; details can be found in Bay et al. [47]. Let us define $\boldsymbol{S}_e$ as an edge response map, computed by

$$\boldsymbol{S}_e = \tilde{\nabla}(\boldsymbol{x}) = \mathcal{N}(|\boldsymbol{K}_{sx} * \boldsymbol{x}| + |\boldsymbol{K}_{sy} * \boldsymbol{x}|) \tag{4}$$

Image normalization is computed by

$$\boldsymbol{x}_N = \mathcal{N}(\boldsymbol{x}) = \frac{255}{\arg\max_{i \in \{1,\ldots,N\}} x_i} \boldsymbol{x} \tag{5}$$

where N is the total number of pixels in $\boldsymbol{x}$, and x_i is the ith pixel in the image $\boldsymbol{x}$, assuming that image $\boldsymbol{x}$ is not all zeros. The gradients from

Eq. (5) are computed by convoluting the input image with the 3×3 Scharr kernel, $\boldsymbol{K}_{si}$, where $i \in \{x, y\}$, as follows:

$$\boldsymbol{K}_{sx} = \begin{bmatrix} -3 & 0 & 3 \\ -10 & 0 & 10 \\ -3 & 0 & 3 \end{bmatrix}, \qquad \boldsymbol{K}_{sy} = \begin{bmatrix} -3 & -10 & -3 \\ 0 & 0 & 0 \\ 3 & 10 & 3 \end{bmatrix} \tag{6}$$

Our gradient approximation is not threshold binarized, and the strength of the gradient is a useful detail in characterizing the target object and the background scene in the red-channel histogram. We extract intensities greater than the mean value from the gradient response denoted by $\tilde{\mathcal{M}}(\boldsymbol{x})$ to avoid weak edges from the background. We also compute the image region by taking the Otsu threshold [27] of the gradient image, $\mathcal{O}(\boldsymbol{x})$; this results in a binarized

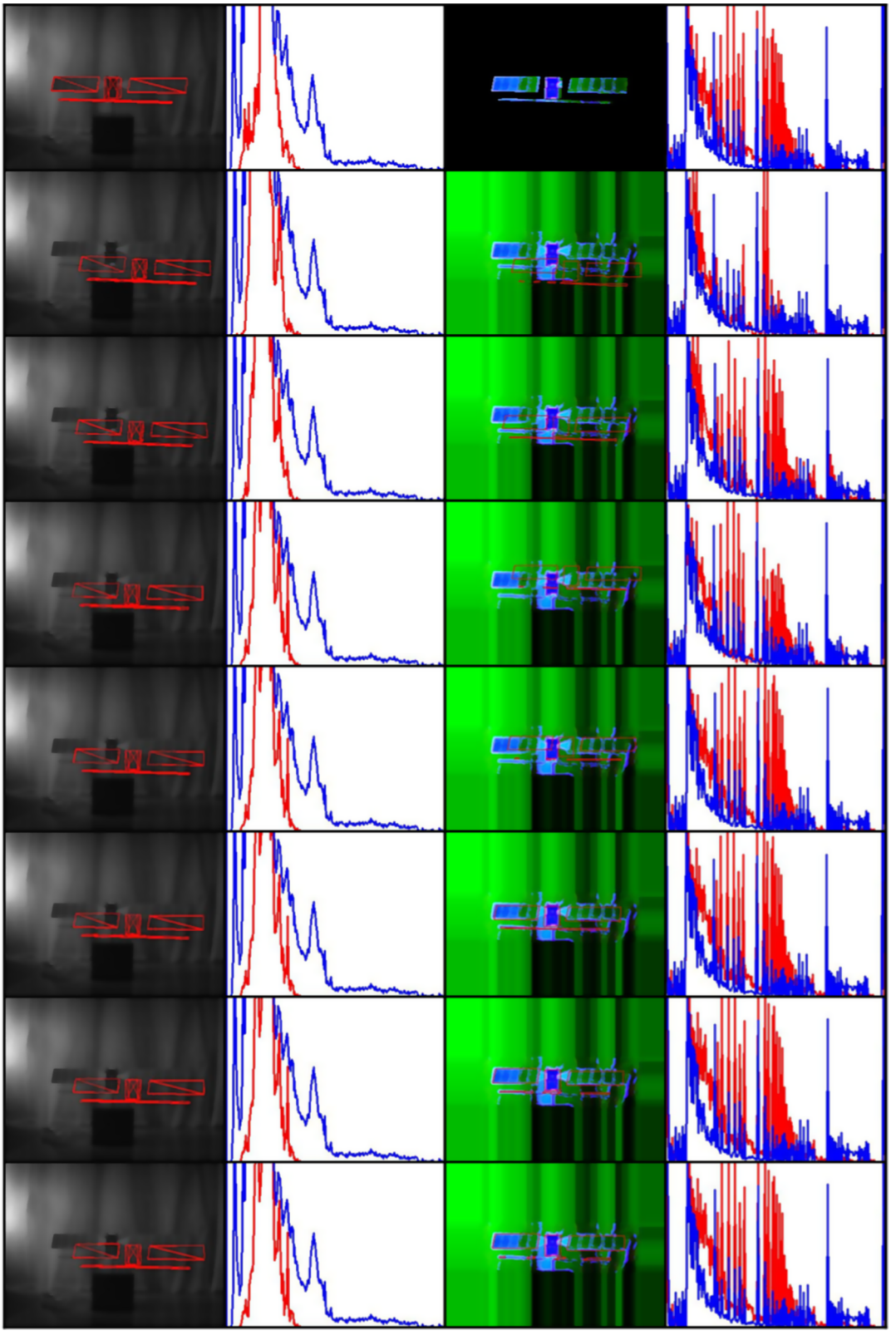


Fig. 2 RSM infrared image pose initialization. The left column is the original PWP3D pose estimation that failed to produce a correct pose estimation result. The right column is the enhanced FC-HSF + PWP3D method that produced the correct gradient descent solution. https://youtu.be/fqtMsGsnexE.

edge response map. The Otsu threshold [27] minimizes the *weighted within-class variance*, which optimally separates clustered regions in the normalized intensity histogram. The *weighted within-class variance* is defined as [27]

$$\sigma^2(\tau) = c_1(\tau)\sigma_1^2(\tau) + c_2(\tau)\sigma_2^2(\tau) \tag{7}$$

where τ is a threshold value; σ_i^2, $i \in \{1, 2\}$, is the intensity cluster variance; and c_i is the cumulative distribution function (CDF) computed as

$$c_1(\tau) = \sum_{i=0}^{\tau} P(i), \qquad c_2(\tau) = \sum_{i=\tau+1}^{255} P(i) \tag{8}$$

where P is the normalized histogram at the ith intensity, which is also the pixel intensity probability. The means of each of the clustered histogram regions are

$$\mu_1(\tau) = \sum_{i=0}^{\tau} \frac{iP(i)}{c_1}, \qquad \mu_2(\tau) = \sum_{i=\tau+1}^{255} \frac{iP(i)}{c_2} \tag{9}$$

The variance of each of the clustered histogram regions is

$$\sigma_1^2(\tau) = \sum_{i=0}^{\tau} [i - \mu_1(\tau)]^2 \frac{P(i)}{c_1}, \qquad \sigma_2^2(\tau) = \sum_{i=\tau+1}^{255} [i - \mu_2(\tau)]^2 \frac{P(i)}{c_2} \tag{10}$$

The τ intensity threshold is minimum when σ is tested to be minimum. Our tests show that the Otsu threshold [27] gives a more conservative response than using the mean threshold. Let us define $\mathcal{C}(\boldsymbol{x})$ as an operator for finding and drawing closed regions in an image as filled areas. Practically, this can be achieved using the *findContours* and *drawContours* functions in OpenCV 3.3.[§] When $\mathcal{C}(\boldsymbol{x})$ is applied on the Otsu threshold [27], edge image produces target object inner surround regions with precise boundary definitions. We intersect both the red and blue channels with the dilated prior mask image $\boldsymbol{S}_D$ to remove image clutter. If we use only the inner region based on the dilated prior, a sharp contrast at the region boundary from all the image channels will cause a strong barrier during the convergence and will give too much reliance on the dilated prior mask. To avoid the over-reliance, we smooth the border by outwardly extending the same pixel value at the dilation boundary to all the zero regions in the green-channel image. The red and blue channels are not affected because they do not have a range of pixel intensities at the border of the dilated prior. Let us define $\mathcal{BE}(\boldsymbol{x}, \boldsymbol{y})$ as an operator that extends the inner border pixels to the image boundary, where the inner border is the rectangular outer boundary of the dilated prior mask $\boldsymbol{y}$; this is achieved by applying the *copyMakeBorder* function with the *BORDER_REPLICATE* option in OpenCV 3.3. We perform histogram equalization (HE), $\mathcal{HE}(\boldsymbol{x})$, on the border extended image to gain additional contrast enhancement. The histogram-equalized response map has stronger contrast and is faster than the CLAHE response. Histogram equalization is computed by equalizing the PDF as follows:

$$S_h = 255 \cdot \sum_{i=0}^{h} P(i) \tag{11}$$

where $h \in \{0, \ldots, 255\}$ and S_h is the transformed pixel intensity. Finally, we combine all the color channels into a false-colored image using the $\mathcal{CTC}(\boldsymbol{R}, \boldsymbol{G}, \boldsymbol{B})$ operator; this false-colored image enhances the high-frequency edge response [48] and center-surround [22] in the input grayscale image. Adjustable weighting factors, $\{C_r, C_g, C_b\}$, may be applied to each of the respective color channels independently; in our example, all the adjustable weighting factors are set to one. The FC-HSF procedure is summarized in Algorithm 1. The corresponding FC-HSF pipeline is provided in Fig. 3.

Algorithm 1 The FC_HSF algorithm to generate the FC-HSF Image

procedure FC_HSF ($\boldsymbol{I}, \boldsymbol{S}_{K-1}$)
  $\boldsymbol{S}_D \leftarrow \mathcal{MD}(\boldsymbol{S}_{K-1}, K_D = 36)$
  $\boldsymbol{S}_g \leftarrow \mathcal{CTG}(\boldsymbol{I})$
  $\boldsymbol{S}_e \leftarrow \bar{\nabla}(\mathcal{B}(\mathcal{CL}(\boldsymbol{S}_g, C_L = 4), K_B = 3))$
  $\boldsymbol{S}_r \leftarrow \mathcal{M}(\boldsymbol{S}_e) \bigcap \boldsymbol{S}_D$
  $\boldsymbol{S}_b \leftarrow \mathcal{C}(\mathcal{O}(\boldsymbol{S}_e) \bigcap \boldsymbol{S}_D)$
  $\boldsymbol{S}_g \leftarrow \mathcal{HE}(\mathcal{BE}(\boldsymbol{S}_g, \boldsymbol{S}_D))$
  FC-HSF Image: $\boldsymbol{S}_I \leftarrow \mathcal{CTC}(\boldsymbol{S}_r, \boldsymbol{S}_g, \boldsymbol{S}_b)$

Figures 1i–1k show the red, green, and blue color channel images for the RSM, respectively. The combined HC-HSF and its histogram template are shown in Figs. 1l and 1m, respectively. Figure 1n shows a probability posterior foreground mask computed using the new histogram template. The new foreground mask is more precise when comparing Fig. 1n with Fig. 1c. The improvement is mainly due to the blue-channel region map enhancing foreground and background contrast. We can further improve the posterior mask by using the pose projection from Fig. 1o as the final mask shown in Fig. 1p.

The standard *Jaccard index*, also known as *intersect over union* (IoU), was used to measure foreground mask extraction performance. The IoU is defined as

$$J = \frac{|x \cap y|}{|x \cup y|} \tag{12}$$

where x is the ground truth pixels and y is the predicted pixels. The mean Jaccard index, $\bar{J}$, can be computed by averaging the individual Jaccard indexes of the test images from the video sequence.

III. Results and Discussion

Details of the level-set segmentation regional-based pose estimation method are provided in Shi et al. [25]. We use a laboratory video of the RSM spinning about its pitch axis captured by a 320 × 240 ICI-9320*P* infrared camera. The position vector, $\boldsymbol{t}$, the rotation matrix, $\boldsymbol{R}$, the projection image, $\boldsymbol{I}$, the spacecraft body coordinate, $\mathcal{F}_b$, and the camera coordinate system, $\mathcal{F}_c$, are provided in Fig. 4.

We apply FC-HSF to the initialization example shown in Fig. 2. Figure 2 columns 1 and 2 represent single image pose estimation on the original input image without any image processing. Figure 2 columns 3 and 4 are the pose estimation on FC-HSF images. Figure 2 column 1 shows the 3D projection mesh overlay on the original image over the entire gradient descent process. Figure 2 column 2 shows the normalized foreground and background histogram of the estimated pose in red and blue lines, respectively. Figure 2 column 3 shows the projection silhouette overlay on the FC-HSF image. Figure 2 column 4 shows the normalized foreground and background histogram of the FC-HSF image in red and blue lines, respectively. Figure 2 row 1 image 1 shows the prior projection on the input image. Figure 2 row 1 image 2 shows the template histogram foreground and background generated using the prior mask on the unprocessed initial input in red and blue lines, respectively. Figure 2 row 1 image 3 shows the foreground FC-HSF template image. Figure 2 row 1 image 4 shows the foreground and background template histogram on the FC-HSF image using red and blue lines, respectively. Figure 2 rows 2–8 are gradient descent steps 0, 2, 4, 6, 10, 20, and 30, respectively. In Fig. 2, the initial pose misalignment is 0.05 m in each of the X, Y, and Z spacecraft body coordinates, and 0.05 deg in each of the pitch-yaw-roll Euler rotation sequence from $\mathcal{F}_b$. The FC-HSF histogram shows a more distinctive histogram profile than the unaltered input image histogram, and the FC-HSF solution converges to the correct pose, whereas the unaltered input image does not.

Figure 5 shows a 360 deg rotation sequence for the RSM infrared image.[¶] The top four rows in Fig. 5 represent rotations from

[§]Data available online at https://opencv.org.

[¶]Dataset and video are available at http://ai-automata.ca/research/hisafe.html.

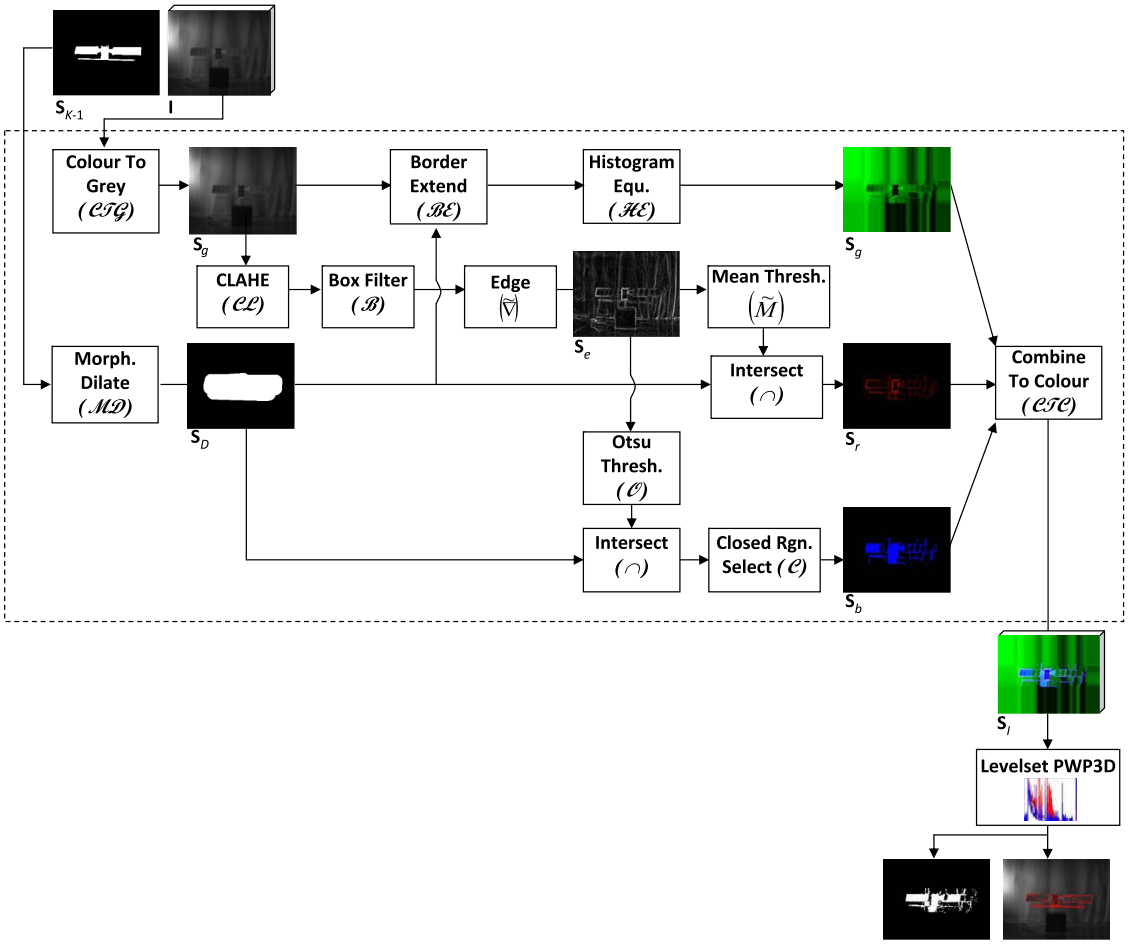


Fig. 3 The *FC_HSF* pipeline. The *FC_HSF* algorithm inside the dashed border uses the three-channel input image and the previous frame mask. The algorithm produces the *FC_HSF* image and is fed into the PWP3D function for pose estimation and posterior mask generation.

0 to 179 deg; the bottom four rows represent rotations from 180 to 359 deg. All images in Fig. 5 are selected based on equal angle separations. Figure 5 row 1 is the 3D CAD projection overlay on the original image. The red edge lines are the 3D model image projection using the estimated pose. Figure 5 row 2 is the FC-HSF image. Figure 5 row 3 is the normalized estimated pose foreground and background histogram of the FC-HSF image in red and blue lines, respectively. Figure 5 row 4 is the level-set function distance map transformation for each of the estimated pose. The 0th level-set contour is plotted in a white line, whereas distances inside and outside the zero level are plotted in red and blue shades, respectively. Brighter shade indicates a closer distance to the spacecraft boundary contour. We use the FC-HSF image and histogram profile for PWP3D regional-based pose estimation gradient descent [25]. The estimated pose tracked well against the actual model, whereas the highest observed errors occurred during 90 and 270 deg. Over these two rotation ranges, the visible area is the smallest in the entire sequence. We compared using unmodified image input and FC-HSF image input and found that the unmodified image input method diverges from the actual pose shortly after the first image. We also compared using the initial image as the histogram template and using sequential posteriors as the template. The results show that the former performed better due to accumulated errors from sequential frames that quickly diluted the following template histograms, suggesting that the level-set method is sensitive to the prior mask accuracy.

Figure 6 shows foreground mask generation comparison for six video frames. The FC-HSF+PWP3D method consistently produces the sharpest foreground extraction of the target RSM. A quantitative mean value IoU measurement of the RSM thermal image video sequence foreground pixel extraction is provided in Table 1. The FC-HSF+PWP3D produced the best foreground extraction mask compared with state-of-the-art methods. The PWP3D level-set method uses the input mask to train the foreground and background pixel sets for statistical segmentation and pose estimation. When the

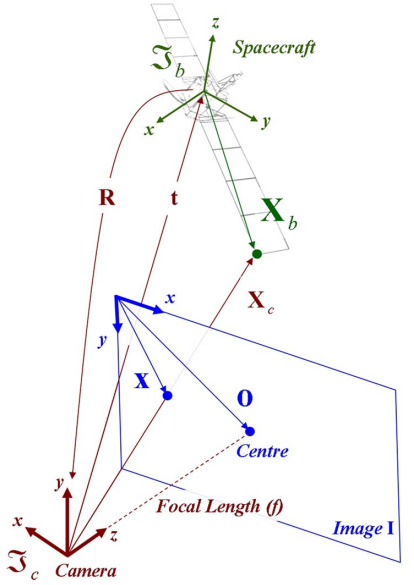


Fig. 4 Camera coordinate system definition. *R* and *t* are the pose of the target spacecraft relative to the camera frame. The point X_b measured from the spacecraft body frame is projected onto the image plane *I* as *x*.

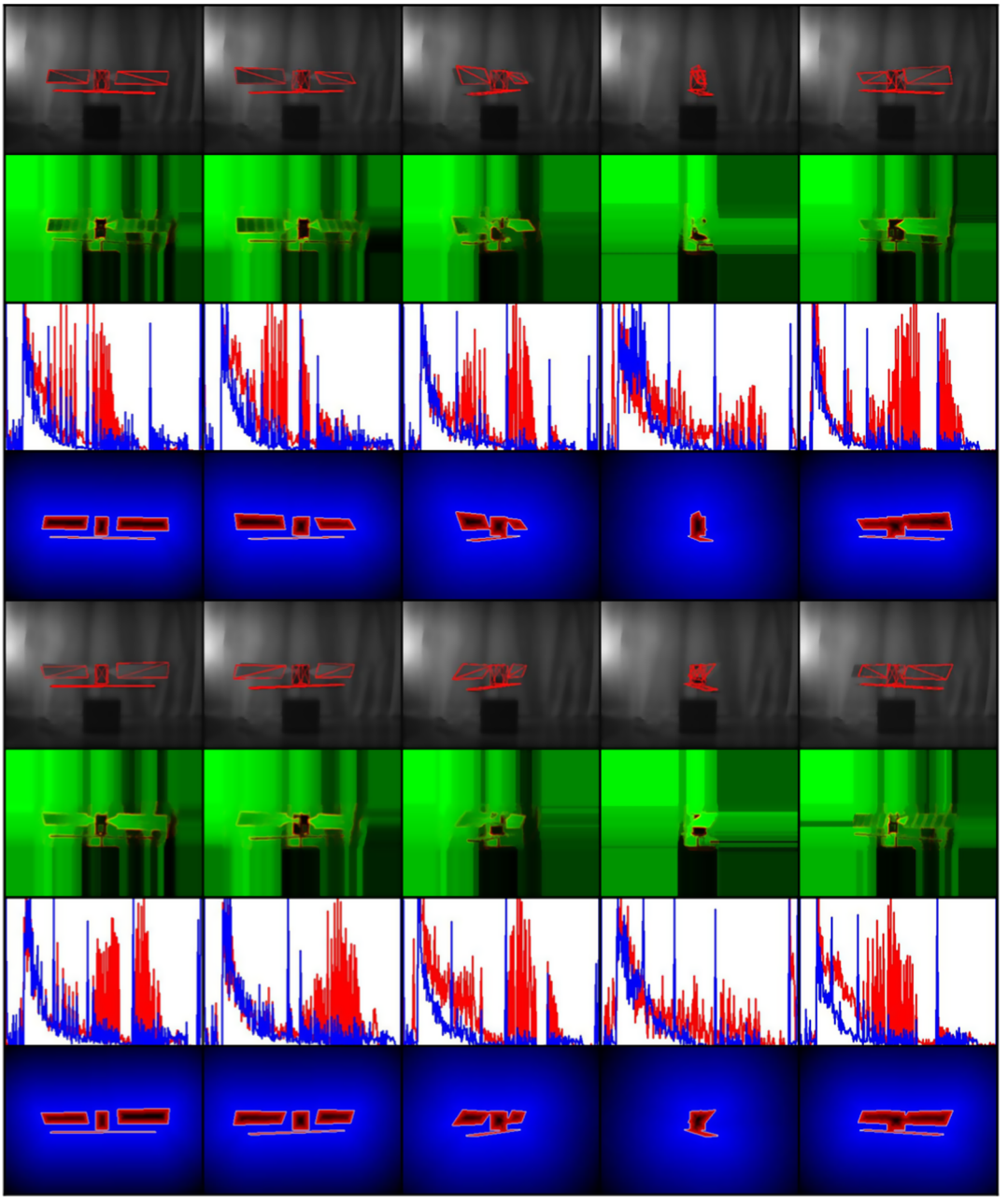
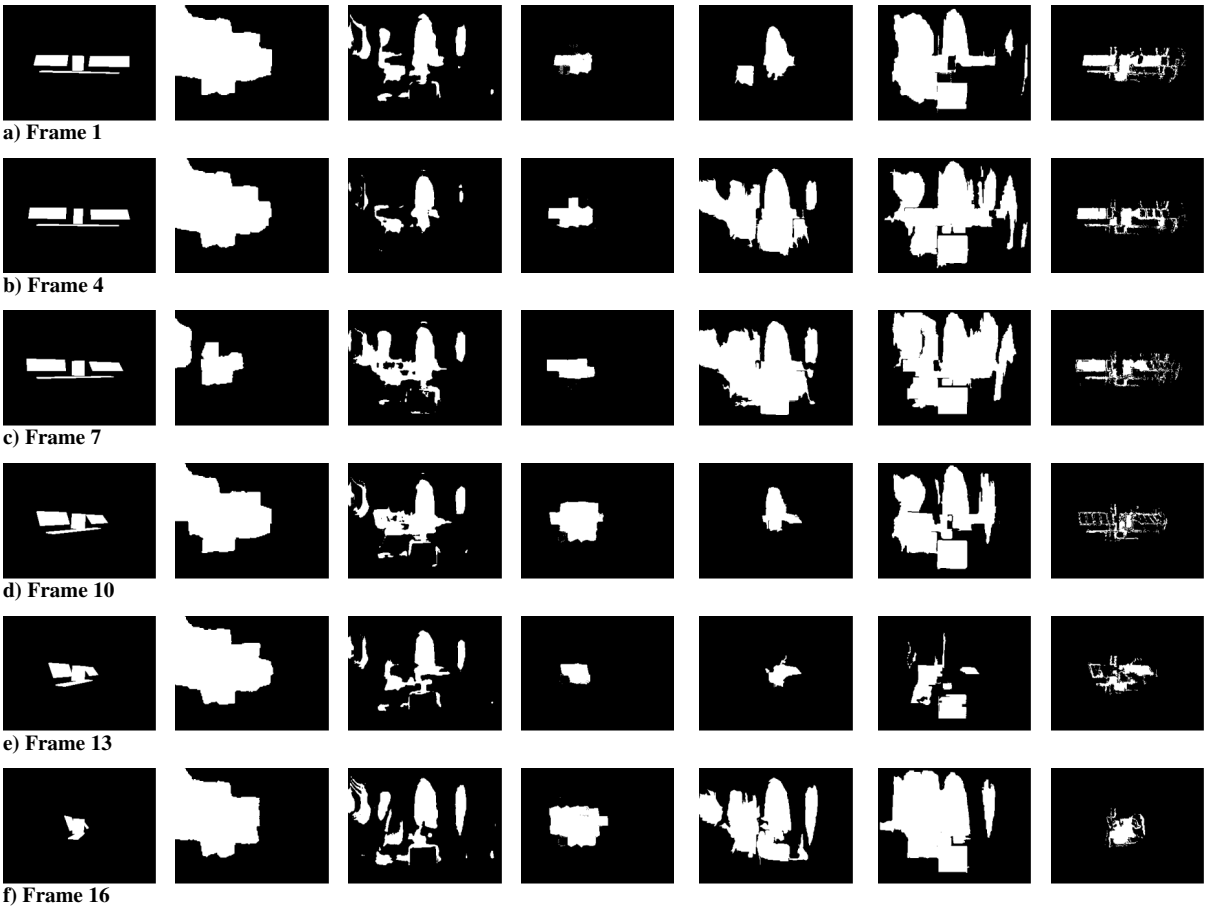


Fig. 5 Monocular infrared image pose estimation using the FC-HSF + PWP3D method. https://youtu.be/Sg4y_bAERgE.

input image is monochromatic, the pixel histogram between the foreground and the background is nondistinctive and will result in poor segmentation and pose estimation. In the case of color images, additional channels elevate the distinctiveness of the image histogram. While the monochromatic image does not have the depth of an image from multiple wavelengths of light, the novelty of our method uses image gradients to supplement the lack of color depth and enhance the distinctiveness of the image histogram between the foreground and background. The main experiment of this work is from a laboratory setting. Our technique can be transferred to the space environment, specifically when the target spacecraft is passing over the Earth shown in Fig. 7. In the case of thermal imagery, only the monochromatic image is available. The additional gradient information will increase the distinctiveness of the image the same way as the RSM. Future work will include a thorough evaluation of our method on flight IR and grayscale images.

IV. Conclusions

The state-of-the-art methods in camera-based pose estimation for spacecraft rendezvous have focused on the back-end pose estimation and stochastic filtering while provided little development in the front-end image processing. Unfortunately, the quality of the input image and the ability to extract useful features from the image can largely affect the back-end pose estimation results. Our method of using regional-based pose estimation requires distinctive foreground and background pixel probability that is difficult to obtain in monochromatic images. To this end, the false-colored high-frequency saliency feature image alteration was developed to improve histogram foreground and background distinctiveness, and the effectiveness of our approach was demonstrated using real-world infrared images. Given a challenging background including complex clutter image patterns, our false-colored image using image gradient and prior image location estimate enhanced the regional-based pose

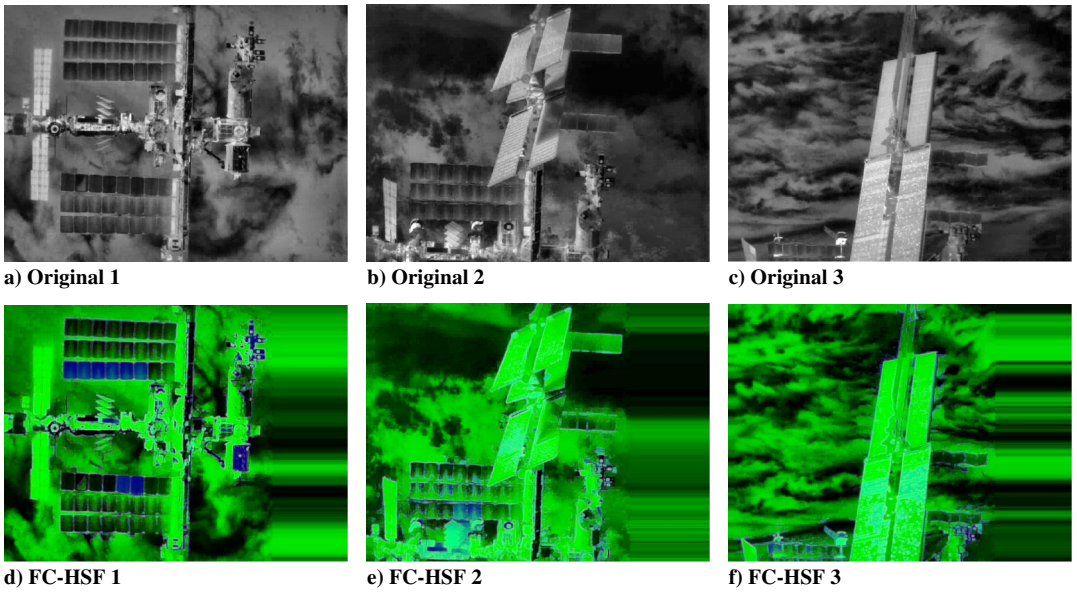


a) Frame 1
b) Frame 4
c) Frame 7
d) Frame 10
e) Frame 13
f) Frame 16

Fig. 6 Foreground mask computed using state-of-the-art methods compared with our FC-HSF+PWP3D method. From left to right: GT, fst, fst+, GMR, MB+, RC, FC-HSF+PWP3D.

Table 1 Comparison of mean IoU between the various saliency foreground extraction methods and our FC-HSF+PWP3D method

Method	fst	fst+	GMR	MB+	RC	FC-HSF + PWP3D
$\bar{J}$	0.1417	0.1678	0.3174	0.2056	0.1161	0.4079



a) Original 1
b) Original 2
c) Original 3
d) FC-HSF 1
e) FC-HSF 2
f) FC-HSF 3

Fig. 7 FC-HSF of the ISS captured by the *Neptec* TriDAR thermal camera during STS-135 mission.

estimation gradient descent process. Future work will include using temporal information from the image sequence and adding efficient top-down approach for more precise foreground extractions. Stochastic pose filtering can also be added to stabilize the pose prediction.

Acknowledgments

This research was jointly funded by the NSERC Scholarship CGSD3-453738-2014, CSA STDP, and OCE VIP II Award 24053.

References

[1] Wu, Y., Yang, G., Lin, J., Raus, R., Zhang, S., and Watt, M., "Low-Cost, High-Performance Monocular Vision System for Air Bearing Table Attitude Determination," *AIAA Journal of Spacecraft and Rockets*, Vol. 51, No. 1, Jan.–Feb. 2014, pp. 66–75. doi:10.2514/1.A32465

[2] Fourie, D., Tweddle, B., Ulrich, S., and Saenz-Otero, A., "Flight Results of Vision-Based Navigation for Autonomous Spacecraft Inspection of Unknown Objects," *AIAA Journal of Spacecraft and Rockets*, Vol. 51, No. 6, Nov.–Dec. 2014, pp. 2016–2026. doi:10.2514/1.A32813

[3] Tweddle, B., and Saenz-Otero, A., "Relative Computer Vision-Based Navigation for Small Inspection Spacecraft," *AIAA Journal of Guidance, Control, and Dynamics*, Vol. 38, No. 5, 2015, pp. 969–978. doi:10.2514/1.G000687

[4] Haralick, R., Joo, H., Lee, C., Zhuang, X., Vaidya, V., and Kim, M., "Pose Estimation from Corresponding Point Data," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 19, No. 6, 1989, pp. 1426–1446. doi:10.1109/21.44063

[5] Sharma, S., and D'Amico, S., "Comparative Assessment of Techniques for Initial Pose Estimation Using Monocular Vision," *Acta Astronautica*, Vol. 123, June–July 2016, pp. 435–445. doi:10.1016/j.actaastro.2015.12.032

[6] Sharma, S., Ventura, J., and D'Amico, S., "Robust Model-Based Monocular Pose Initialization for Noncooperative Spacecraft Rendezvous," *AIAA Journal of Spacecraft and Rockets*, Vol. 55, No. 6, Nov.–Dec. 2018, pp. 1414–1429. doi:10.2514/1.A34124

[7] Duda, R., and Hart, P., "Use of the Hough Transformation to Detect Lines and Curves in Pictures," *Communications of the ACM*, Vol. 15, No. 1, 1972, pp. 11–15. doi:10.1145/361237.361242

[8] Lepetit, V., "EPnP: An Accurate O(n) Solution to the PnP Problem," *International Journal of Computer Vision*, Vol. 81, No. 2, 2009, pp. 155–166. doi:10.1007/s11263-008-0152-6

[9] D'Amico, S., Benn, M., and Jørgensen, J. L., "Pose Estimation of an Uncooperative Spacecraft from Actual Space Imagery," *International Journal of Space Science and Engineering*, Vol. 2, No. 2, 2014, pp. 171–189. doi:10.1504/IJSPACESE.2014.060600

[10] Prisacariu, V., and Reid, I., "PWP3D: Real-Time Segmentation and Tracking of 3D Objects," *International Journal of Computer Vision*, Vol. 98, No. 3, 2012, pp. 335–354. doi:10.1007/s11263-011-0514-3

[11] Wang, J., Jiang, H., Yuan, Z., Cheng, M., Hu, X., and Zheng, N., "Salient Object Detection: A Discriminative Regional Feature Integration Approach," *International Journal of Computer Vision*, Vol. 123, No. 2, 2017, pp. 251–268. doi:10.1007/s11263-016-0977-3

[12] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet Classification with Deep Convolutional Neural Networks," *Proceedings of Conference on Advances in Neural Information Processing Systems*, The NIPS Foundation, La Jolla, CA, 2012, pp. 1097–1105.

[13] Zhang, H., Fritts, J., and Goldman, S., "Image Segmentation Evaluation: A Survey of Unsupervised Methods," *Computer Vision and Image Understanding*, Vol. 110, No. 2, 2008, pp. 260–280. doi:10.1016/j.cviu.2007.08.003

[14] Yi, Z., Zhang, H., Tan, P., and Gong, M., "DualGAN: Unsupervised Dual Learning for Image-to-Image Translation," *Proceedings of the IEEE International Conference on Computer Vision*, IEEE, New York, 2017, pp. 2868–2876. doi:10.1109/ICCV.2017.310

[15] Sharma, S., Beierle, C., and D'Amico, S., "Pose Estimation for Non-Cooperative Rendezvous Using Neural Networks," *Proceedings of the 2018 IEEE Aerospace Conference*, IEEE, New York, 2018. doi:10.1109/AERO.2018.8396425

[16] Andrychowicz, M., Baker, B., Chociej, M., Jozefowicz, R., McGrew, B., Pachocki, J., Petron, A., Plappert, M., Powell, G., Ray, A., et al., "Learning Dexterous In-Hand Manipulation," *Computing Research Repository (CoRR)*, abs/1808.00177, 2019.

[17] Li, Y., Wang, G., Ji, X., Xiang, Y., and Fox, D., "DeepIM: Deep Iterative Matching for 6D Pose Estimation," *Computer Vision—Proceedings of the European Conference on Computer Vision 2018*, Lecture Notes in Computer Science, Vol. 11210, Springer, Cham, 2018, pp. 683–698. doi:10.1007/978-3-030-01231-1_42

[18] Cheng, M., Mitra, N., Huang, X., Torr, P., and Hu, S., "Global Contrast Based Salient Region Detection," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 37, No. 3, 2015, pp. 569–582. doi:10.1109/TPAMI.2014.2345401

[19] Zhang, J., Sclaroff, S., Lin, Z., Shen, X., Price, B., and Měch, R., "Minimum Barrier Salient Object Detection at 80 FPS," *Proceedings of the 2015 IEEE International Conference on Computer Vision*, IEEE, New York, 2015, pp. 1404–1412. doi:10.1109/ICCV.2015.165

[20] Yang, C., Zhang, L., Lu, H., Ruan, X., and Yang, M., "Saliency Detection via Graph-Based Manifold Ranking," *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, New York, 2013, pp. 3166–3173. doi:10.1109/CVPR.2013.407

[21] Strand, R., Ciesielski, K., Malmberg, F., and Saha, P., "The Minimum Barrier Distance," *Computer Vision and Image Understanding*, Vol. 117, No. 4, 2013, pp. 429–437. doi:10.1016/j.cviu.2012.10.011

[22] Itti, L., Koch, C., and Niebur, E., "A Model of Saliency-Based Visual Attention, for Rapid Scene Analysis," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 11, 1998, pp. 1254–1259. doi:10.1109/34.730558

[23] Jung, C., Kim, W., Yoo, S., and Kim, C., "A Novel Monochromatic Cue for Detecting Regions of Visual Interest," *Journal of Image and Vision Computing*, Vol. 32, Nos. 6–7, 2014, pp. 405–413. doi:10.1016/j.imavis.2014.04.001

[24] Yacoob, Y., and Davis, L., "Segmentation Using Meta-Texture Saliency," *IEEE 11th International Conference on Computer Vision*, IEEE, New York, 2007. doi:10.1109/ICCV.2007.4408930

[25] Shi, J. F., Ulrich, S., and Ruel, S., "Regional Method for Monocular Infrared Image Spacecraft Pose Estimation," *Proceedings of the AIAA Space and Astronautics Forum and Exposition*, AIAA Paper 2018-5281, Sept. 2018. doi:10.2514/6.2018-5281

[26] Badrinarayanan, V., Kendall, A., and Cipolla, R., "Segnet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 12, 2017, 2481–2495. doi:10.1109/TPAMI.2016.2644615

[27] Otsu, N., "A Threshold Selection Method From Gray-Level Histograms," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 9, No. 1, 1979, pp. 62–66.

[28] Rosenhahn, B., Brox, T., and Weickert, J., "Three-Dimensional Shape Knowledge for Joint Image Segmentation and Pose Tracking," *International Journal of Computer Vision*, Vol. 73, No. 3, 2007, pp. 243–262. doi:10.1007/s11263-006-9965-3

[29] Kass, M., Witkin, A., and Terzopoulos, D., "Snakes: Active Contour Models," *International Journal of Computer Vision*, Vol. 1, No. 4, 1988, pp. 321–331. doi:10.1007/BF00133570

[30] Mumford, D., and Shah, J., "Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems," *Communication Pure Applied Math*, Vol. 42, No. 5, 1989, pp. 577–685. doi:10.1002/(ISSN)1097-0312

[31] Caselles, V., Kimmel, R., and Sapiro, G., "Geodesic Active Contours," *International Journal of Computer Vision*, Vol. 22, No. 1, 1997, pp. 61–79. doi:10.1023/A:1007979827043

[32] Chan, T., and Vese, L., "Active Contours Without Edges," *IEEE Transactions on Image Processing*, Vol. 10, No. 2, 2001, pp. 266–277. doi:10.1109/83.902291

[33] Osher, S., and Sethian, J., "Fronts Propagating with Curvature-Dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations," *Journal of Computational Physics*, Vol. 79, No. 1, 1988, pp. 12–49. doi:10.1016/0021-9991(88)90002-2

[34] Kervrann, C., and Heitz, F., "Statistical Deformable Model-Based Segmentation of Image Motion," *IEEE Transaction on Image Processing*, Vol. 8, No. 4, 1999, pp. 583–588. doi:10.1109/83.753745

[35] Zhu, S., Lee, T. S., and Yuille, A., "Region Competition: Unifying Snakes, Region Growing, Energy/Bayes/MDL for Multi-Band Image Segmentation," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 9, 1996, pp. 884–900. doi:10.1109/34.537343

[36] Leventon, M., Grimson, W., and Faugeras, O., "Statistical Shape Influence in Geodesic Active Contours," *IEEE 5th IEEE EMBS International Summer School on Biomedical Imaging*, Vol. 1, IEEE, New York, 2000, pp. 316–323. doi:10.1109/SSBI.2002.1233989

[37] Rousson, M., and Paragios, N., "Shape Priors for Level-Set Representations," *3rd European Conference on Computer Vision*, Springer, Berlin, 2002, pp. 78–92.

[38] Cremers, D., "Statistical Shape Knowledge in Variational Image Segmentation," Ph.D. Thesis, Dept. of Mathematics and Computer Science, Univ. Mannheim, Mannheim, Germany, 2002.

[39] Cremers, D., Kohlberger, T., and Schnörr, C., "Nonlinear Shape Statistics in Mumford-Shah Based Segmentation," *3rd European Conference on Computer Vision*, Springer, Berlin, 2002, pp. 93–108.

[40] Cremers, D., and Schnörr, C., "Motion Competition: Variational Integration of Motion Segmentation and Shape Regularization," *Patter Recognition*, Vol. 2449, Lecture Notes in Computer Science, Springer, Berlin, 2002, pp. 472–480. doi:10.1007/3-540-45783-6_57

[41] Brox, T., and Weickert, J., "Level-Set Based Image Segmentation with Multiple Regions," *Pattern Recognition*, Springer, Berlin, Aug. 2004, pp. 415–423.

[42] Bibby, C., and Reid, I., "Robust Real-Time Visual Tracking Using Pixel-Wise Posteriors," *9th European Conference on Computer Vision*, Springer, Berlin, 2008, pp. 831–844.

[43] Rosenhahn, B., Brox, T., Cremers, D., and Seidel, H. P., "A Comparison of Shape Matching Methods for Contour Based Pose Estimation," *Combinatorial Image Analysis*, Springer, Berlin, 2006, pp. 263–276.

[44] Brox, T., Rosenhahn, B., Gall, J., and Cremers, D., "Combined Region and Motion-Based 3D Tracking of Rigid and Articulated Objects," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 3, 2010, pp. 402–415. doi:10.1109/TPAMI.2009.32

[45] Schmaltz, C., and Rosenhahn, B., "Region-Based Pose Tracking with Occulusions Using 3D Models," *Machine Vision and Applications*, Vol. 23, No. 3, 2012, pp. 557–577. doi:10.1007/s00138-010-0317-5

[46] Zuiderveld, K., "Contrast-Limited Adaptive Histogram Equalization," *Graphics Gems IV*, Academic Press, San Diego, CA, 1994, pp. 474–485.

[47] Bay, H., Ess, A., Tuytelaars, T., and Gool, L., "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, Vol. 110, No. 3, June 2008, pp. 346–359. doi:10.1016/j.cviu.2007.09.014

[48] Hubel, D., and Wiesel, T., "Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex," *Journal of Physiology*, Vol. 160, No. 1, 1962, pp. 106–154. doi:10.1113/jphysiol.1962.sp006837

J. A. Christian
Associate Editor