Multiple-Frame Subpixel Wildfire Tracking

David R. Thompson, William Johnson, Robert Kremens

***Abstract*—We present a method to improve subpixel signal detection in airborne or orbital image sequences. The proposed technique recognizes stable *interest point* features in multiple overlapping frames. It estimates motion between consecutive frames and tracks candidate detections over time. The final detection decision combines signal strengths from multiple views to improve sensitivity. The algorithm is computationally tractable for real-time use on autonomous robotic platforms and spacecraft. Additionally, the interest points enable image-relative localization, obviating the need to transmit the entire image and reducing transmission bandwidth requirements by one or more orders of magnitude. This permits higher acquisition rates and potentially improved coverage for remote monitoring. Ground-based systems can reconstruct absolute positions from landmarks without measurements of sensor pose. We demonstrate the algorithm using airborne 4μm imagery from multiple overpasses of a controlled wildfire.**

***Index Terms*—Wildfire detection, Thermal Imagery, Smart Cameras, Computer Vision, Pattern Recognition**

I. Introduction

Subpixel target detection uses intensity or spectral information to find features that may subtend less than a full image pixel. A wide range of science, surveillance and monitoring applications involve subpixel detection from moving platforms. Here we focus on early wildfire detection systems using unpiloted spacecraft or aircraft to observe a large area with thermal infrared cameras. Such systems can automatically detect weak emissivity signals from small wildfires, and automatically alert authorities for rapid containment. This can mitigate damages, save lives, and preclude costly evacuations. The sensitivity of automated detection is a key performance bottleneck for such systems, and algorithmic improvements can significantly benefit their overall performance.

Currently there are a wide range of detection methods based on one or more 4μm or 11μm channels [1]–[3]. The 4μm atmospheric window provides the best ratio of emissive to reflective signal, and tests against longer wavelengths can exclude false positives from sun glint or other artifacts. These detection rules may also incorporate spatial information in the form of adjacent pixels. Detection products typically analyze individual image frames or single passes of linear pushbroom cameras [4]–[7]. The BIRD system is a small-satellite fire detection with multiple infrared bands [8]. The GOES ABBA product offers wildfire detection, though its low spatial resolution is generally too small for early warning [9]. Some recent research has used thermal images from different dates to detect temporal anomalies [10], [11]. Otherwise, detection rules generally operate on each frame independently.

Our detection strategy augments these rules by exploiting overlapping frames of image sequences. We hypothesize that multiple frames could improve detection by effectively increasing the detector integration time, and by providing several opportunities to detect a flickering or intermittent source near threshold. This work demonstrates a multiple-frame technique for subpixel detection that is computationally tractable for real time use onboard mobile platforms. As a side benefit, the algorithm identifies landmark features that can localize detections without needing to transmit the entire image or know the precise sensor pose. This image-relative localization can reduce transmission bandwidth requirements by one or more orders of magnitude, enabling higher acquisition rates and better coverage. Section II describes the multiple-frame detection method. Sections III and IV show an experiment using 4μm aircraft overpass images of a controlled wildfire. We conclude with remarks on applicability to airborne and orbital detection systems.

II. Method

Our approach tracks scene content to estimate motion and find projective geometric relationships between multiple images. We take as a constraint that the method should be automated. It should be feasible for field networks with limited communications that cannot download a video stream for pixel-wise coregistration on the ground. We withhold precise knowledge of sensor position since such information may not be available in the field. We assume that the scene is approximately planar which is appropriate for a high-altitude airborne or orbital view.

The procedure identifies stable, high contrast features, tracks them across consecutive frames, and uses these correspondences to estimates a *homography* that describes the geometric transformation between each adjacent image pair. It then applies a single- or multiple-channel detection rule independently to each image with a very lenient threshold. The algorithm matches consecutive detections across potentially large displacements, and associates them to form *tracks*, i.e. unique physical events with a precise geographic location, that may appear in multiple frames. Finally, the system considers the entire sequence history in the final detection decision.

A. SIFT feature detection

Consider detection in a sequence of n partially-overlapping frames. The first step identifies stable, high contrast image locations known as *interest points* that can be recognized easily across multiple views [12]. Here use Scale-Invariant Feature Transform (SIFT) keypoint approach which is among

David R. Thompson, and William Johnson are with the Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena CA, 91109 USA email: *david.r.thompson@jpl.nasa.gov*.Robert Kremens is with Rochester Institute of Technology.Manuscript received May 2011; published XXXX

the most common and well-studied [13], [14]. We find these locations with a Difference of Gaussians (DoG) operator at multiple scales. Following the notation of [15], the DoG operator applied to an image I at pixel coordinates (x, y) is a difference of circular Gaussian filters G_σ and $G_{c\sigma}$ based on a scale-dependent standard deviation σ. The widths of the two Gaussian filters differ by a constant ratio.

$$D_\sigma(x, y) = (G_{c\sigma}(x, y) - G_\sigma(x, y)) * I(x, y) \qquad (1)$$

We apply the filter several times, increasing filter widths by a factor of two with each iteration. Peaks in this scale space indicate interest points. We associate each with a 128 dimensional SIFT descriptor representing local statistics around each interest point [13]. The descriptor is a histogram of local image gradients; it is invariant to rotation and somewhat robust against affine distortion. A nearest-neighbor matching procedure finds candidate matches between frames [13]. The end result of this first step is a list of candidate interest points and descriptors in each frame. Figure 1 shows an example of the best matches in two frames from our airborne dataset. We ignore matches in bright burning areas whose appearance may change independently of camera motion. In order to capture the extra degree of freedom in three-dimensional location, we express each point in homogeneous coordinates [16] as the vector $\mathbf{x} = \lambda[x, y, 1]^T$ defined up to an arbitrary scale factor λ.

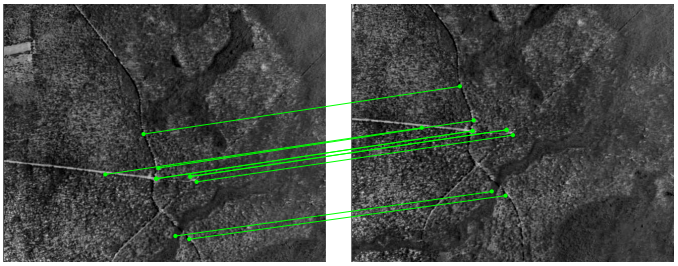

Fig. 1. SIFT features. The algorithm identifies informative locations such as road crossings and recognizes them across changes in translation or orientation. Over 500 such SIFT features were matched; for clarity we show only the highest-scoring pairs.

B. Homography estimation

We use pointwise matches to estimate linear homographies that map between pairs of frames in the image sequence [16]. Assuming a linear camera model and a planar scene, there is a homography H relating an image point $\mathbf{x}$ to its position in the following frame $\mathbf{x}'$:

$$\mathbf{x}'\lambda = H\mathbf{x}, \qquad \mathbf{x}, \mathbf{x}' \in \mathbf{R}^3 \qquad (2)$$

Again this transformation is only unique up to a scalar factor λ. We fit the homography using matched interest points in adjacent images. We define a column vector $\mathbf{h} = [\mathbf{h}_1\mathbf{h}_2\mathbf{h}_3]^T$ from the rows $\mathbf{h}_j$ of $\mathbf{H}$. For each interest point match i, equation 2 provides the following constraints on the new pixel coordinates x' and y':

$$\begin{bmatrix} \lambda^{-1}\mathbf{x}_i^T & 0^T & -x_i'\mathbf{x}_i^T \\ 0^T & -\lambda^{-1}\mathbf{x}_i^T & y_i'\mathbf{x}_i^T \end{bmatrix} \mathbf{h} = 0 \qquad (3)$$

Successive matches provide additional relationships that can be written together in matrix form as the homogeneous linear system $\mathbf{Ah} = 0$. In practice, the projection is noisy and an exact solution does not exist. Instead we find $\mathbf{h}$ to minimize $\|\mathbf{Ah}\|$, using the constraint $\|\mathbf{h}\| = 1$ to avoid the trivial result $\mathbf{h} = 0$. The solution of this problem is the unit eigenvector corresponding to the smallest eigenvalue of $\mathbf{A}^T\mathbf{A}$ [16].

Many of the initial matches are spurious so we use the Random Sample Consensus (RANSAC) algorithm to identify outliers [17]. RANSAC is an iterative approach that attempts model fittings using many random subsets of the data. Each trial selects a random subset of 4 interest point matches, and fits a homography using just this subset. We score the resulting projection using least-squares reconstruction error between projected and actual interest points. Specifically we count the number of inlier matches whose projection agrees to observation within a small spatial tolerance. We refine the best-scoring homography through nonlinear optimization with the Nelder-Mead simplex algorithm [18]. The result is a set of pairwise homographies relating neighboring frames.

C. Fire detection and tracking

Next, we apply any subpixel detection rule to each independent frame to return a list of candidate detection locations. We aim to combine these single-frame detections into a globally-consistent list of *tracks* representing unique physical fires that appear in multiple frames. Note that an obvious alternative would simply co-register all frames into an image mosaic, as in [15], prior to any detection. However, automated mosaics may be more challenging with respect to onboard memory and time. Moreover, performing a detection before the association means we need only associate sparse clouds of subpixel detections. This can be more error-tolerant than a pixel-wise mosaic where consecutive frames must match exactly.

Solving the association requires determining the correspondence between observed detections in consecutive frames. Relevant algorithms include stable-marriage matching, methods based on quadratic programming [19], and spectral analysis of affinity matrices [20]. Here we use a simple greedy approach that is efficient for scenes involving hundreds of fire pixels. We build a list of detection tracks in chronological order beginning with the first frame. We initialize a separate track for each detection $\mathbf{a}_k$ and project each into the following frame:

$$\mathbf{a'}_k\lambda = \mathbf{Ha}_k \qquad (4)$$

The apostrophe denotes that $\mathbf{a'}_k$ is a virtual or projected event rather than an observed hot pixel. We then match these projected locations to each real hot pixel in the next frame, denoted $\mathbf{b}_j$. We use a one-to-many association where $\mathbf{a'}_k$ could pair with multiple instances $\mathbf{b}_j$, provided that the matched detections all lie within an error radius r. We find the best matches $\mathbf{a'}_\star$ based on Euclidean distance:

$$\mathbf{a'}_\star = \mathrm{argmin}_k\|\mathbf{a'}_k - \mathbf{b}_j\|, \quad \|\mathbf{a'}_k - \mathbf{b}_j\| < r \qquad (5)$$

A new track is formed including the entire signal/location history of $\mathbf{a}_\star$ along with the new instance $\mathbf{b}_j$. Any old detection that projects to at least one observation in the subsequent

frame is considered to have been successfully merged with other detection events; its track is removed from the list. Unmatched detections maintain a list entry, and we propagate these projections to the next round of forward projection and association. Therefore fires that fail to match in just one frame are not lost, and tracks can skip one or more frames. However homography estimation is difficult for separated frames having less spatial overlap, and we enforce a maximum skip interval: any fire not matched after three successive frames cannot be matched again. New fires enter near the edge of the advancing field of view, and generate new tracks in the list. The end goal of this process is a list of unique tracks, each associated with a specific geographic location and appearing as a separate pixel in at least one frame. We also record the signal levels in all frames and note wherever a detection was expected but not found (Figure 2).

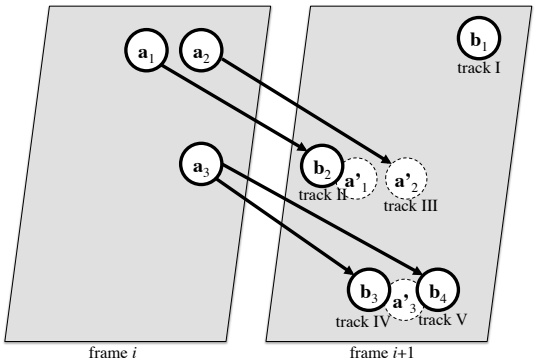

Fig. 2. Association of detections between two frames of the sequence. We project detections $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ in frame i to new locations $\mathbf{a'}_1, \mathbf{a'}_2, \mathbf{a'}_3$ in frame $i+1$. We associate each observed fire in frame $i+1$ with the closest matching projection, and any unmatched projections are recorded as absent. If a new detection has no neighboring projection within error tolerance it becomes a new track (e.g. detection b_1).

Finally, we classify each track using properties of its magnitude vector $[q_i^1, q_i^2 \ldots q_i^k]$, where one or more of the elements contains either a signal strength or a flag denoting no detection. The initial detection rule use a lenient threshold so that even very weak candidate fires have a chance of appearing in the list. We incorporate multi-frame information using a *hysteresis threshold* similar to the Canny edge detection algorithm [21]. Designers specify two thresholds τ_1 and τ_2. Any detection strength above τ_1 is accepted, provided it appears above the stricter threshold τ_2 in a neighboring frame. Thus, any pixel with an intensity above τ_2 immediately triggers a detection, while weak signals above τ_2 can still be recovered provided they have sufficient support in other sequence frames.

III. Experimental Approach

A series of experiments evaluate multi-frame subpixel detection for small wildfire detection. The tests consider an experimental controlled wildfire in the wooded area surrounding Eglin Air Force Base, Florida. The main body of this fire was over a kilometer in length. The fire evolved over several hours, growing in intensity as it moved laterally across the investigation area. In its wake, many hundreds of small fires survived and continued burning well after the main fire had passed. We use aerial overflight data collected by the WASP instrument platform [22]. The instrument observed the entire scene with multiple consecutive linear overflights over the burning region. A cooled 4μm Medium-Wave Infrared (MWIR) camera mounted inside the aircraft collected 10-20 overlapping 640×512 pixel frames at a rate of approximately one frame per 4 seconds. The camera ground-sampling distance (GSD) averaged $2.5 - 3$m per pixel, and the scene progressed by approximately 50 pixels across the short axis in each subsequent frame (Figure 1).

We perform some initial filtering operations prior to subpixel detection. We manually identify the first n frames at the beginning of each overpass where fires do not appear; these images do not show the burning area and we use them to characterize background noise. First, we eliminate individual bad pixels by averaging fire-free frames and comparing each pixel's average value to the mean. Any outliers are replaced with the scene's mean value. Second, we divide each pixel by the average value of its column to remove column-wise noise effects. Finally, we correct vignetting by fitting a radially-symmetric low-degree polynomial to the image frames and dividing by this flat-field image.

Our subpixel detection tests subsampling the cleaned high-resolution images (Figure 3). We reduce image resolution by a factor of 4 with bicubic interpolation. This proportional averaging over geographic regions approximates observations from a higher-flying platform, characterizing performance as Ground Sampling Distance increases. We select six low-resolution overpasses from the experiment for comprehensive analysis. No special criterion was used to select the sequences, apart from a general desire that they fall late in the experiment when many small pixel-scale fires are present in the burn scar (Figures 3 and 6).

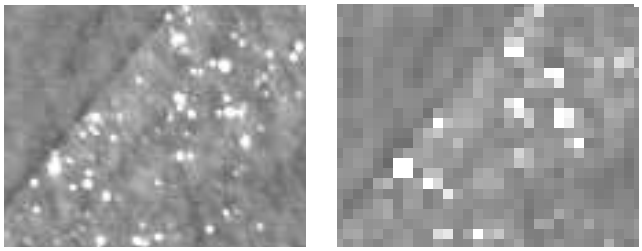

Fig. 3. Detail from original and subsampled images.

We evaluate performance for each sequence using the frame that shows the most subpixel fires. The pixel intensity of the *high-resolution* raw image clearly shows a bimodal distribution corresponding to *burning* and *non-burning* pixels. Visual inspection provides an obvious threshold that can be used to identify burning areas in these high-resolution images, and by extension, the locations of active subpixel fires in the subsampled sequence. We label a pixel in the low-resolution image as a burning pixel if it subtends any burning area. We then perform single- and multiple-frame analysis with a lenient setting of τ_1 at 2.5 standard deviations above the mean. Varying the τ_2 threshold reveals different achievable false positive and false negative rates for Receiver Operating Characteristic (ROC) analysis.

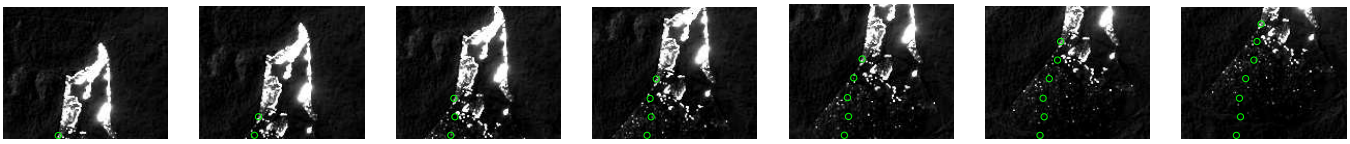

Fig. 4. A portion of test sequence 4 demonstrating the track of a single detection across frames. All image processing uses downsampled images (shown).

IV. Results

For all images analyzed the algorithm successfully identifies a homography and tracks detections to within visual accuracy (Figure 4). Figure 5 shows ROC curves illustrating detection performance for the six test sequences. Multiple frame methods offer significant improvements in each case, with better sensitivity for comparable false positive rates. Points along each ROC curve represent different values of the τ_2 threshold. The curves to the lower right of each plot (e.g. the "x" symbols) disregard multiple-frame information.

Figure 6 shows an example from sequence 6. Here we display the high-resolution image for clarity, but all tracking and detection uses the subsampled version. Small bright dots correspond to the subpixel remnant fires. We set detection thresholds so that the single- and multiple frame methods achieve similar true positive rates. We use a threshold of 2.7σ in the single-frame case ("x" markers), or $\tau_1 = 2.0\sigma, \quad \tau_2 = 3.3\sigma$ in the multiple frame case ("o" markers), with thresholds expressed as standard deviations above the mean. Red symbols are false positives, which occasionally occur on bright regions far from the burning area. For comparable false positive rates the multiple-frame approach detects many more subpixel fires: 72% of total fire pixels with 22 false positives, compared with 54% of total fire pixels with 36 false positives for the single-frame approach. This result represents just one of many possible settings that operators can choose depending on their error tolerance. However, the ROC curves favor the multiple frame method for all regimes.

V. Discussion

These experiments support our hypothesis that multiple-frame wildfire tracking can improve detection sensitivity. However, the SIFT features also provide a valuable side benefit: they could enable absolute georeferencing based on image content if there is a suitable database of previous images from the area. System designers could use an initial characterization phase to acquire high-contrast SIFT descriptors along with images of the (fire-free) surface. The ground system, with possible human assistance, would determine these descriptors' geographic locations. During regular operations the system can query this database to find the geographic locations of new scenes. Such a strategy could significantly reduce downlink volumes, since the transmission need contain just the positions and descriptors of the highest-contrast SIFT features along with the image locations of the hot pixels in that frame. This would provide sufficient information to geolocate a hot pixel. All camera pose computations, cross-frame associations and final detection decisions could take place on the ground, though not necessarily requiring manual intervention after the initial characterization phase. With such a system in place, the total bandwidth required for the frames in this study would be approximately $30KB$ per image. This represents an order-of-magnitude reduction in downlink and could significantly improve image acquisition rates for bandwidth-constrained systems. It would also permit more flexible detection decisions to be made with contextual or geographic information. For example, the system could ignore specific regions that produce sun glint, reflectance or other recurring false positive effects.

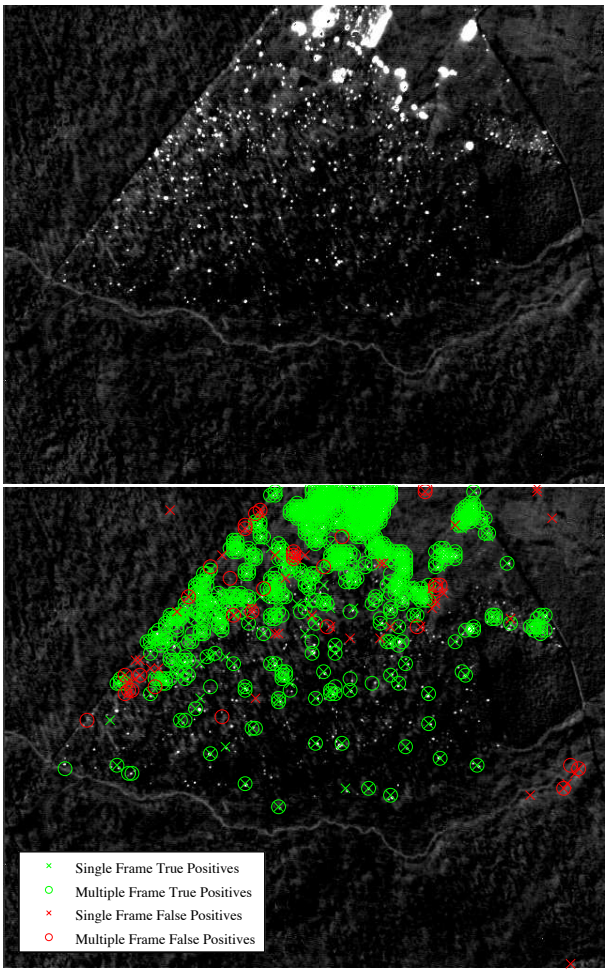

Fig. 6. Detections in frame 15 of the final sequence. The high-resolution image is shown, but the actual detection results are based on the subsampled version. Small bright white dots are burning remnants of the large fire near the top of the frame.

Several potential enhancements to this basic algorithm could improve computational requirements and perhaps performance. A natural alternative would treat detections as contiguous connected regions, and match these regions across image frames instead of individual pixels. This would dramatically reduce the size of the correspondence problem and permit

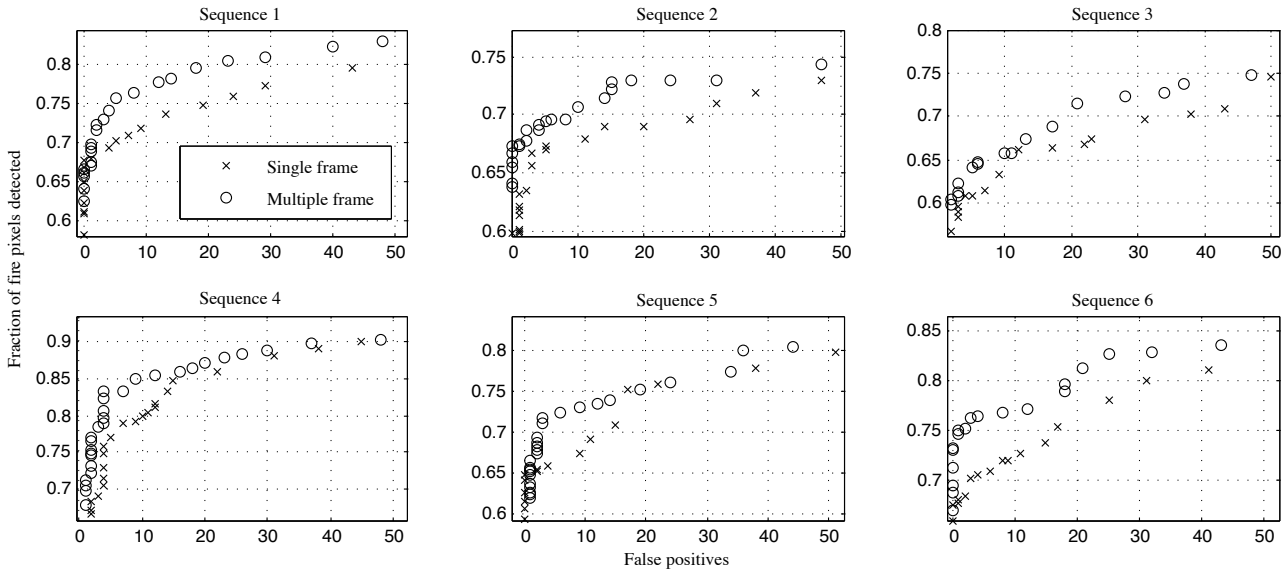

Fig. 5. Test sequence performance. Points along the curve represent different values of the τ_2 threshold. The curves to the lower right (exhibiting worst performance) ignore multiple-frame information.

application of more sophisticated matching techniques such as spectral approaches [20]. Finally, more sophisticated methods based on likelihood ratios could be used to merge the multiple detections.

ACKNOWLEDGMENT

The fire dataset was obtained through Joint Fire Science Program grant 08-1-6-01 to Roger Ottmar, Matthew Dickinson, and Elizabeth Reinhardt (Validation of Fuel Consumption Models for Smoke Management Planning in the Eastern Regions of the United States) and support from the USFS Northern Research Station and the National Fire Plan. This work utilized the VLFeat toolbox [14]. We thank Robert Staehle and Michael Burl for their ideas and counsel. We also thank the RIT WASP team for their support. Finally, we acknowledge and appreciate the support of Moustafa Chahine. The work was funded by a NASA Jet Propulsion Laboratory Innovative Spontaneous Concept grant.

REFERENCES

[1] L. Giglio, I. Csiszar, A. Restás, J. T. Morisette, W. Schroeder, D. Morton, and C. O. Justice, "Active fire detection and characterization with the advanced spaceborne thermal emission and reflection radiometer (aster)," *Remote Sensing of Environment*, vol. 112, pp. 3055–3063, 2008.

[2] J. T. Morisette, L. Giglio, I. Csiszar, and C. O. Justice, "Validation of the modis active fire product over southern africa with aster data," *International Journal of Remote Sensing*, vol. 26, no. 19, pp. 4239–4264.

[3] W. Schroeder, E. Prins, L. Giglio, I. Csiszar, C. Schmidt, J. T. Morisette, and D. Morton, "Validation of goes and modis active fire detection products using aster and etm+ data," *Remote Sensing of Environment*, vol. 112, pp. 2711–2726, 2008.

[4] S. Boles and D. Verbyla, "Comparison of three AVHRR-based fire detection algorithms for interior Alaska," *Remote Sensing of Environment*, vol. 72, no. 1, pp. 1–16, 2000.

[5] Y. Li, A. Vodacek, R. Kremens, and A. Ononye, "A new algorithm for global forest fire detection using multispectral images," *Proceedings of SPIE*, vol. 5075, p. 367, 2003.

[6] L. Giglio, J. Descloitres, C. O. Justice, and Y. J. Kaufman, "An enhanced contextual fire detection algorithm for modis," *Remote Sensing of Environment*, vol. 87, no. 2-3, pp. 273–282, 2003.

[7] Z. Li, R. Fraser, J. Jin, A. Abuelgasim, I. Csiszar, P. Gong, R. Pu, and W. Hao, "Evaluation of algorithms for fire detection and mapping across North America from satellite," *Journal of Geophysical Research*, vol. 108, no. D2, p. 4076, 2003.

[8] B. Zhukov, E. Lorenz, D. Oertel, M. Wooster, and G. Roberts, "Space-borne detection and characterization of fires during the bi-spectral infrared detection (bird) experimental small satellite mission (2001-2004)," *Remote Sensing of Environment*, vol. 100, no. 1, pp. 29–51, 2006.

[9] W. Xu, M. Wooster, G. Roberts, and P. Freeborn, "New goes imager algorithms for cloud and active fire detection and fire radiative power assessment across north, south and central america," *Remote Sensing of Environment*, vol. 114, no. 9, pp. 1876 – 1895, 2010.

[10] A. Koltunov and S. Ustin, "Early fire detection using non-linear multitemporal prediction of thermal imagery," *Remote Sensing of Environment*, vol. 110, no. 1, pp. 18 – 28, 2007.

[11] K. Kushida, "Detection of active wildland fires using multitemporal modis images," *Geoscience and Remote Sensing Letters, IEEE*, vol. 7, no. 2, pp. 301–305, April 2010.

[12] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE transactions on pattern analysis and machine intelligence*, pp. 1615–1630, 2005.

[13] D. G. Lowe, "Object recognition from local scale-invariant features," *Computer Vision, IEEE International Conference on*, vol. 2, p. 1150, 1999.

[14] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," http://www.vlfeat.org/, 2008.

[15] Y. Wang, J. Ng, M. J. Garay, and M. C. Burl, "Onboard image registration from invariant features," *International Symposium on Artificial Intelligence, Robotics and Automation in Space*, 2008.

[16] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge Univ Pr, 2003.

[17] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[18] D. Olsson and L. Nelson, "The Nelder-Mead simplex procedure for function minimization," *Technometrics*, vol. 17, no. 1, pp. 45–51, 1975.

[19] A. Berg, T. Berg, and J. Malik, "Shape matching and object recognition using low distortion correspondences," *TR CSD-04-1366*, Dec. 2004.

[20] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," *The International Conference on Computer Vision*, 2005.

[21] J. Canny, "A computational approach to edge detection," *Readings in computer vision: issues, problems, principles, and paradigms*, vol. 184, pp. 87–116, 1987.

[22] D. McKeown, J. Cocburn, J. Faulring, R. Kremens, D. Morse, H. Rhody *et al.*, "Wildfire airborne sensor program (wasp): A new wildland fire detection and mapping system," *Remote sensing for field users. Proc. 10th Biennial USDA Forest Service Remote Sensing Applications Conference*, 2005.