# Object Motion Estimation and Detection for Satellite Imagery

Sreepathi B [1] & Vijaya lakshmi.k [2]

*Abstract*: **Enhanced multispectral satellite images with multi-angular look capability have tremendous potential applications. The object tracking algorithm includes moving object estimation, target modeling, and target matching three-step processing. Potentially moving objects are first identified on the time-series images. The target is then modeled by extracting both spectral and spatial features. In target matching procedure, Bhattacharyya distance, histogram intersection, pixel count similarity is combined in a novel regional operator design. The algorithm uses set of multi-angular sequence images acquired by the WorldView-2 satellite. The tracking performance is analyzed by the calculation of recall, precision, and *F*1 score of the test. Tracking is performed in the context of higher-level applications that require the location and shape of the object in every frame. The importance of tracking includes use of appropriate image features, selection of motion models, and detection of objects.**

Keywords: **multispectral satellite imagery contrast enhancement, , objects tracking.**

## 1. INTRODUCTION

Object tracking is an important task within the field of computer vision. The proliferation of high-powered computers, the availability of high quality and inexpensive video cameras, and the increasing need for automated video analysis has generated a great deal of interest in object tracking algorithms[1]. There are three key steps in object tracking: detection of interesting moving objects, tracking of such objects from frame to frame, and analysis of object tracks to recognize their behavior.

---------------------------------------------------------

*1. Information Science & Engineering Dept, VTU RYMEC, Cantonment Bellary*
sreepathib@gmail.com

*2. Information Science & Engineering Dept, VTU RYMEC, Cantonment Bellary*
viju483@gmail.com

Object tracking in a complex environment has long been an interesting and challenging problem. In the remote sensing context, it has often been applied to the use of aerial or satellite imagery to track ground vehicle traffic. Algorithms have been developed to demonstrate vehicle tracking using low-rate video or visible imagery sequences collected by sensors on aircraft [3]–[6]. Airborne spectral imagery, as well as spectral combined with polarimetric imaging, have also been used to demonstrate the capability of remote sensing platforms to track vehicles. In another application, satellite imagery along with other data sources has been used to track ships in the ocean. Additionally, synthetic aperture radar airborne and satellite sensors have also been shown to have surface object tracking capabilities [7].In its simplest form, tracking can be defined as the problem of estimating the trajectory of an object in the image plane as it moves around a scene. In other words, a tracker assigns consistent labels to the tracked objects in different frames of a video. Additionally, depending on the tracking domain, a tracker can also provide object-centric information, such as orientation, area, or shape of an object.

Image enhancement techniques improve the quality of an image as perceived by a human. These techniques are most useful because many satellite images when examined on a color display give inadequate information for image interpretation. Contrast generally refers to the difference in luminance or grey level values in an image and is an important characteristic.[2]-[3].It can be defined as the ratio of the maximum intensity to the minimum Intensity over an image . Contrast enhancement techniques expand the range of brightness values in an image so that the image can be efficiently displayed in a manner desired by the analyst. Current satellite platforms offer limited utility for surface object tracking primarily due to inherent tradeoffs in resolution and spatial/temporal coverage. Low resolution satellites located in low earth orbits with adequate resolution (~1 meter) to resolve surface objects offer limited spatial coverage (10's of sq. km.) and long repeat intervals (days)[6]. Satellites with short repeat intervals (minutes) are located in geosynchronous orbits and offer poor ground resolution (~1 km). The first crude image taken by the satellite Explorer- 6 shows a sunlit area of the Central Pacific Ocean and its cloud cover. Satellite

images have many applications in meteorology, agriculture, geology, forestry, landscape, biodiversity conservation, regional-planning, education, intelligence and warfare. Images can be in visible colors and in other spectral resolution together with a very agile satellite platform [8]. All of these exciting features of to further demonstrate the capability of object tracking with the help of space borne sensors.

In this paper, an object tracking algorithm is proposed for use with multi-temporal and multi-spectral satellite imagery. Multiple objects can be tracked simultaneously with user initialized starting points. At this point, the tracking algorithm only relies on the intensities of the pixels and does not include any kinematic capability to account for the smooth movement of the objects.

## 2. PROPOSED OBJECT TRACKING PROCESS

The proposed object tracking process is illustrated in Fig. 1. The algorithm consists of three steps, which are moving object detection, target modeling, and target tracking. Potentially moving objects are first identified based on the time-series images. A reference target is then modeled by extracting spectral and spatial features.

Finally a regional Feature matching operator is applied to estimate the object motion trajectory.

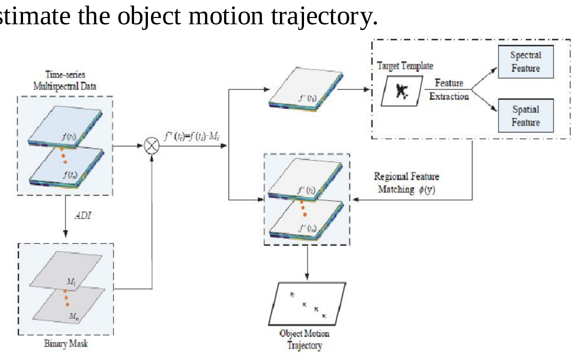


Fig. 1: Flowchart of the proposed object tracking algorithm.

### A. *Moving Object Detection*

Moving object detection is the basic step for further analysis of video. Every tracking method requires an spectra. One example is the WorldView-2 (WV2) satellite which was launched by Digital Globe in 2009 [14]. The collected panchromatic imagery has a spatial resolution of 0.5 m, and multispectral imagery has 2 m spatial resolution. WV2 overcomes some of the limitations for object tracking by satellite by providing the combination of high spatial and object detection mechanism either in every frame or when the object first appears in the video. It handles segmentation of moving objects from stationary background objects. This focuses on higher level processing. It also decreases computation time. Due to environmental conditions like illumination changes, shadow object segmentation becomes difficult and significant problem. This temporal information is usually in the form of frame differencing, which highlights regions that changes dynamically in consecutive frames [9]. Given the object regions in the image, it is then the tracker’s task to perform object correspondence from one frame to the next to generate the tracks as shown in Fig 2:

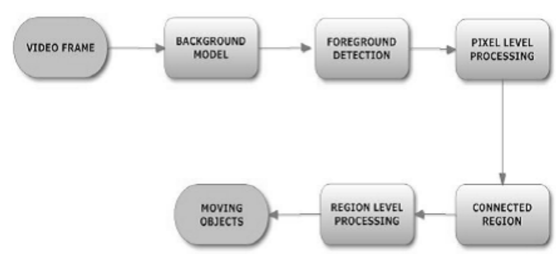


Fig 2: Frame of moving object detection system

The first step is to distinguish foreground objects from stationary background. To achieve this, we can use a combination of various techniques along with low-level image post-processing methods to create a Foreground pixel map at every frame. To extract individual object features such as bounding box, area, perimeter etc.

*Foreground Detection*: The main purpose of foreground detection is to distinguishing foreground objects from the stationary background. Almost, each of the video surveillance systems uses the first step is detecting foreground objects. Only pixels belonging to foreground objects need to be dealt with [10].

*Pixel Level Post-Processing:* The output of foreground detection contains noise. Generally, it affects by various noise factors [10]. To overcome this dilemma of noise, it requires further pixel level processing.

*Detecting Connected Regions*: After detecting foreground regions and applying post-processing operations to remove noisy regions, the filtered foreground pixels are grouped into connected regions [9]. After finding individual regions that correspond to objects, the bounding boxes of these regions are calculated.

*Region Level Post-Processing:* As pixel-level noise removed, still some artificial small regions remain just because of the bad segmentation. To remove this type of regions, regions that have smaller sizes than a

Pre-defined threshold are deleted from the foreground pixel map [8]. Once segmenting regions we can extract features of the corresponding objects from the current image.

Every tracking method requires an object detection mechanism either in every frame or when the object first appears in the video. A common approach for object detection is to use information in a single frame. However, some object detection methods make use of the temporal information computed from a sequence of frames to reduce the number of false detections [8]. Given the object regions in the image, it is then the tracker's task to perform object correspondence from one frame to the next to generate the tracks. One of the approaches for detecting moving objects is to calculate the frame difference pixel by pixel. For a sequence of images $f(\mathbf{x}, t1)$, _ _ _, $f(\mathbf{x}, tn)$ acquired from time $t1$ to $tn$, an accumulative difference image (ADI) can be formulated. The values of the ADI at location **x** are counts, calculated:

$$A_j(\mathbf{x}) = \begin{cases} A_{j-1}(\mathbf{x}) + 1 & if\ [f(\mathbf{x}, t_r) - f(\mathbf{x}, t_j)] > \eta \\ A_{j-1}(\mathbf{x}) & otherwise \end{cases}, \quad (1)$$

Where $Aj(\mathbf{x})$ is the ADI after the $j$th image $f(\mathbf{x}, tj)$ in The sequence, $f(\mathbf{x}, tr)$ is the reference image on which we want to find the moving objects, and $\eta$ is the Threshold, which is approximated to be 5% - 10% of the maximum image difference between the reference and the $j$th images in this work. For a sequence of images collected within a short time frame, the ADI can be calculated based on the accumulation of all the frames [9]. Any pixel with large ADI value is highly likely to be a moving object in the reference image. A threshold that is a count lower than the total number of frames can be further applied to the ADI to filter out the possible background pixels, and a binary mask image $Mi(\mathbf{x})$ can be found accordingly to highlight the moving objects. The filtered $i$th image in the sequence can be found as:

$$f'(\mathrm{x}, t_i) = f(\mathrm{x}, t_i) \cdot M_i(\mathrm{x}). \quad (2)$$

In this paper the moving object detection is performed on all the frames by iteratively using each frame as the reference image.

### B. *Target Model*

The target is defined as the interested object to be tracked. In most object tracking problems, we cannot be sure that the object would remain exactly the same as it was when the target model was defined. It may undergo scale changes, appearance changes and illumination changes. A threshold of 80% works well. To characterize the target, first a feature space is chosen. The reference target model is represented by its pdf q in the feature space. A target is represented by an ellipsoidal region in the image. The target model is represented by spectral and spatial features as illustrated in Fig. 1. The spectral and spatial features are defined below respectively. The spectral probability density function (spdf) is chosen as the feature space to represent an object. The similarity between the model's pdf and candidate's pdf is the criterion for finding the most probable candidate. In practice, $m - bin$ histograms are estimated for each wavelength to approximate the spdfs. In the subsequent frame, a target, candidate is defined at location y, and is characterized by the pdf p.y...The PDF of the reference target model is estimated by discrete densities and $m$-bin histograms have been shown to be sufficient for object tracking purpose [10]. The $m$-bin histograms H$b$ of the reference target model in the $b$th spectral band can be Expressed as:

$$\mathrm{H}_b = \{h_1, \cdots, h_m\}_b \quad b = 1 \cdots S, \quad (3)$$

Where $S$ is the number of spectral band S. The histogram of the $l$th bin in **H**$b$ can be computed as:

$$h_l = \sum_{i=1}^{N_t} K(|\mathbf{x}_i - \mathbf{x}^*|)\delta\,[ind(\mathbf{x}_i) - l], \quad (4)$$

Where $\delta$ is the Kronecker delta function, $ind$ ($\mathbf{x}i$) is the index of histogram bin associated with image intensity at location $\mathbf{x}i$, $K$ (\|**x**$i$ $^*$ **x***\|) is the isotropic kernel which could be monotonic decreasing to reduce the background occlusion effect, or could be Uniform to satisfy low computational cost and $Nt$ will be introduced in Eq. (5). In this paper a uniform kernel is used to accelerate the computation. The width of the kernel is determined by the size of the window. The spatial feature is defined as the geometric area of the reference target surface in the window region R, and is estimated based on its pixel count $Nt$ which can be estimated as:

$$N_t = \sum_{\mathbf{x}\in\mathbb{R}} \mathrm{sgn}\left[f'(\mathbf{x}, t_1)\right], \quad (5)$$

Where sgn is the signum function and is defined as:

$$\mathrm{sgn}(x) = \begin{cases} -1 & if\ x < 0 \\ 0 & if\ x = 0 \\ 1 & if\ x > 0 \end{cases} \quad (6)$$

Given the reference target model, To find a best matching target candidate window in other image sequences. The target candidate windows are those centered on potential moving objects identified in the ADI, and are also represented using both spectral and spatial features. The histogram of a target candidate centered at **y** can be represented as:

$$\mathbf{P}_b(\mathbf{y}) = \{p_1, \cdots, p_m\}_b \quad b = 1 \cdots S, \quad (7)$$

Where *pl* is the histogram of the *l*th bin in **P***b*(**y**). The pixel count of a target candidate centered at **y** can also be estimated in a similar way as Eq. (5), and is denoted as *N*(**y**).

C. *Target Matching*

A sliding window over other image sequences is used to indicate the possible presence of the reference target. A regional feature matching operator is applied to find the similarity between the targets model and the pixels within the window. The first step in our object tracking algorithm is matching the objects (Op's) in previous image (In−1) to the new objects (Oi's) detected in current image (In) as shown in Fig 3. Three different metrics are used to define the operator.

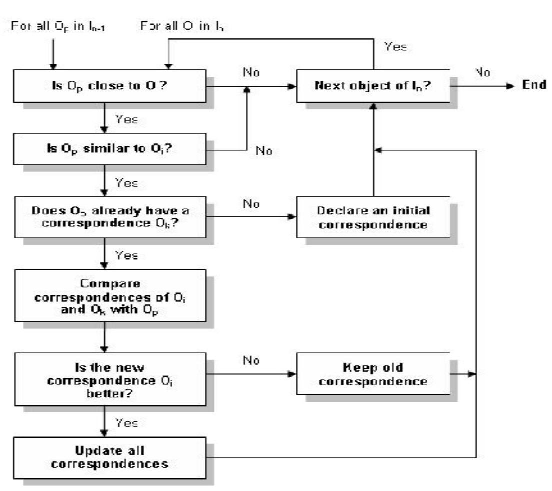


Fig 3: object matching method

For spectral feature matching, Bhattacharyya distance and histogram intersection are used. The Bhattacharyya distance [17] can be found as:

$$d_B(\mathbf{y}) = \frac{1}{S}\sum_{b=1}^{S}\sqrt{1-\rho_B\left[\mathbf{H}_b, \mathbf{P}_b(\mathbf{y})\right]}, \quad (8)$$

Where

$$\rho_B\left[\mathbf{H}_b, \mathbf{P}_b(\mathbf{y})\right] = \sum_{i=1}^{m}\sqrt{\frac{h_i \cdot p_i(\mathbf{y})}{\sum_{i=1}^{m} h_i \cdot \sum_{i=1}^{m} p_i(\mathbf{y})}}. \quad (9)$$

Bhattacharyya Coefficient is an absolute similarity measure and needs no bias correction. Many researchers used this measure to find similarity in images or sections of images. They compared Bhattacharyya coefficient with different optimization techniques. For Bhattacharyya distance [11], low values indicate good matching. The histogram intersection [18] can be calculated using:

$$d_I(\mathbf{y}) = \frac{1}{S}\sum_{b=1}^{S}\rho_I\left[\mathbf{H}_b, \mathbf{P}_b(\mathbf{y})\right], \quad (10)$$

Where

$$\rho_I\left[\mathbf{H}_b, \mathbf{P}_b(\mathbf{y})\right] = \sum_{i=1}^{m}\min\left[h_i, p_i(\mathbf{y})\right]. \quad (11)$$

For histogram intersection, high values indicate well matching. The spatial feature matching is based on the pixel count within the sliding window. As mentioned earlier, a pixel count of the reference target *NT* and a pixel count of the moving object *N*(**y**) can be estimated. It is assumed that the potential

Target candidate and reference target should have similar pixel counts, even though the target has been moved or rotated. The similarity between the reference target and the potential region based on Pixel count can be found as:

$$d_P(\mathbf{y}) = 1 - \frac{|N(\mathbf{y}) - N_t|}{N_t}. \quad (12)$$

A threshold could be applied to $d_P$ (**y**) to filter out the pixels with low similarities. Regional maxima can then be found based on the *dP* (**y**) values. Regional maxima are defined as the connected pixels with a constant $d_P$ (**y**) value, and their boundary pixels all have lower $d_P$ (**y**) values [12]. A binary image RMAX(**y**) can then be generated to identify the locations of regional maxima. The opening morphological filter is the erosion followed by the dilation. The definition of the erosion and dilation operators can be found in. A regional matching operator is finally designed based on these three metrics, and can be defined as:

$$\phi(\mathbf{y}) = \left\{ w_1 \cdot \left[1 - \hat{d}_B(\mathbf{y})\right] + w_2 \cdot \hat{d}_I(\mathbf{y}) \right\} \cdot \mathrm{RMAX}(\mathbf{y}), \quad (13)$$

Where *dB* and *dI* are normalized Bhattacharyya distance and histogram intersection whose original values are divided by their maxima, and *w*1 and *w*2 are weighting coefficients. In this paper *w*1 and *w*2 are selected equally as 0.5. A high $\phi(\mathbf{y})$ value indicates a good matching with the target model.

## 3. EXPERIMENTAL RESULTS

The imagery was collected over Rio de Janeiro (Brazil) in January 2010. The 16-bit multispectral 8-band images were used for object tracking. The images were geo-registered by Digital Globe but not corrected for parallax [13]. A subset collected over a harbor area as shown in Fig. 4 used for validation. Some intermediate results of target matching in the second frame of the harbor region are presented in Fig. 5.

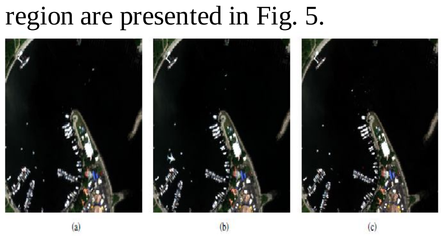

Fig (4): Subsets of five frames collected over a harbor region at different times (a) 13:09:23, (b) 13:09:54, (c) 13:10:46.

In the RGB image as shown in Fig. 5 (a), the red outline indicate the interested ship target, whose target template is extracted based on the first frame. The binary mask image *Mi*, 1-*dB*, *dI*, RMAX after applying opening morphological filter, and the combined feature matching score $\phi$, are also presented in Fig. 5. Note in Fig. 5(b) there is much false detection due to uncorrected parallax effects. Because the ship within the red outline shows both high spectral and spatial similarities to the target Template, it can be easily identified and tracked in Fig. 5(f) based on its high feature matching score $\phi$. The ground truth of the interested moving targets has been identified manually per pixel in the image sequences. The performance of the proposed algorithm is analyzed on a per pixel basis by the recall and precision, which are calculated based on the number of true positives (*tp*), false positives (*fp*), and false negatives (*fn*). The recall (*R*) and precision (*P*) can be calculated and A *F*1 score [14] which is also called *F1*measure can then be found as

$$R = \frac{tp}{tp + fn}, \quad P = \frac{tp}{tp + fp}, \quad F_1 = 2 \cdot \frac{R \cdot P}{R + P}. \quad (14)$$

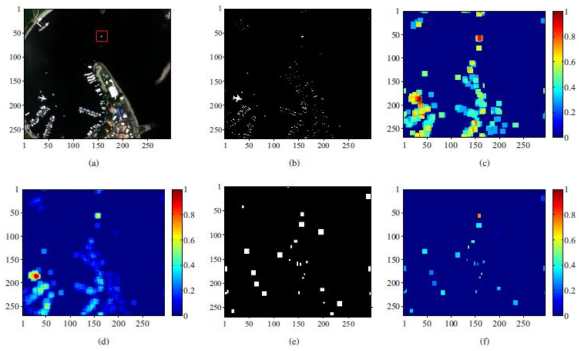

Fig. 5: Intermediate results of target matching in the second frame of harbor region. The target template is extracted based on the first frame. (a) RGB image. The red outline indicates the interested ship target. (b) Binary mask image *Mi*. (c) 1 - *dB*. (d) *dI*. (e) RMAX. (f) Combined feature matching score $\phi$.

The harbor area was first selected for algorithm testing. The RGB image of the first frame is shown in Fig. 6 (a). Our task was to track the ship shown within the red outline. The actual size of the target and the size of the searching window are listed in Table. I. The estimated trajectory of the ship movement based on the proposed object tracking algorithm is shown in Fig. 6 (b). The calculated *R*, *P*, and *F*1 values are evaluated based on 50 pixels that are identified as ship target in the last four frames, and are listed in TableII.

TABLE I: Parameters of target template.

| Target | Actual size (pixel) | Window (pixel$\times$pixel) |
|---|---|---|
| Ship | 17 | $11\times11$ |
| Aircraft A | 136 | $19\times19$ |
| Aircraft B | 104 | $19\times19$ |

TABLE II: Overall performance of the proposed algorithm.

| Region | Object | $R$ | $P$ | $F_1$ |
|---|---|---|---|---|
| Harbor | Ship | 0.690 | 0.919 | 0.782 |
| Airport | Aircraft A | 0.833 | 0.495 | 0.621 |
| | Aircraft B | 0.925 | 0.435 | 0.592 |

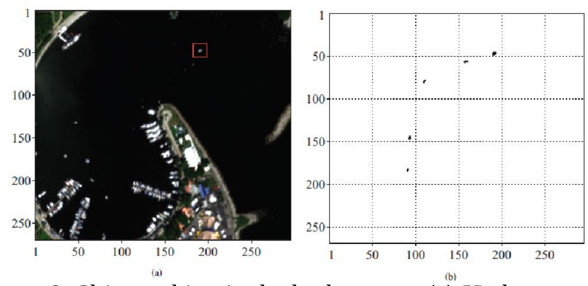

Fig. 6: Ship tracking in the harbor area. (a) Harbor area RGB image, where the ship is within red outline. (b) Ship movement trajectory.

## 4. CONCLUSION

The paper is presented with an object tracking algorithm applied to a sequence of multispectral satellite images. This unique algorithm is developed for an optical multi-angular data set, and is designed based on moving object estimation, target modeling, and target matching three-step processing. Potentially moving objects are first identified on the time-series images. A reference target is then modeled by extracting spectral and spatial features. The use of both spectral and spatial feature ensures better tracking accuracy than using each of them alone as has been observed in the intermediate results of target matching. The tracking performance is analyzed by the calculation of recall, precision, and $F1$ score of the test.

The technique introduced in this paper also has potential applications in other research areas, such as object velocity estimation, and traffic control. Based on the spatial resolution of the available data set and also the relatively long time interval between two frames, it is very difficult to visually identify identical objects with small geometric size in the image sequences. Tracking small objects, such as vehicles, in a denser scene is more challenging.

## 5. REFERENCES

[1] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," ACM Comput. Survey. vol. 38, Dec. 2006.

[2] I. Szottka and M. Butternut, "Tracking multiple vehicles in airborne image sequences of complex urban environments," in 2011 Joint Urban Remote Sensing Event (JURSE), Apr. 2011, pp. 13–16.

[3] K. Palaniappan, F. Bunyak, P. Kumar, I. Ersoy, S. Jaeger, K. Ganguli, A. Haridas, J. Fraser, R. Rao, and G. Seetharaman, "Efficient feature extraction and likelihood fusion for vehicle tracking in low frame rate airborne video," in 13th Conference on Information Fusion (FUSION), Jul. 2010, pp. 1–8.

[4] K. Palaniappan, R. Rao, and G. Seetharaman, "Wide-Area Persistent Airborne Video: Architecture and Challenges," in Distributed Video Sensor Networks, B. Bhanu et al, Ed. Springer London, 2011, pp. 349–371.

[6] C. Carrano, "Ultra-scale vehicle tracking in low spatial resolution and low frame-rate overhead video," in Proc. SPIE, vol. 7445, 744504, 2009.

[7] J. Ker ekes, M. Muldowney, K. Strackerhan, L. Smith, and B. Leahy, "Vehicle tracking with multi-temporal hyperspectral imagery," in Proc. SPIE, vol. 6233, 62330C, 2006.

[8] G. Machismo, F. Pacifica, and C. Pad wick, "On the relative predictive value of the new spectral bands in the worldwiew-2 sensor," in 2010 IEEE International Geosciences and Remote Sensing Symposium, Jul. 2010, pp. 2723–2726.

[9] N. Paragios and R. Deriche. Geodesic active contours and level sets for the detection and tracking of moving objects. IEEE Trans. Patt. Analy. Mach. Intel. 22, 3, 266–280, 2000.

[10] S. Y. Elhabian, K. M. El-Sayed, "Moving object detection in spatial domain using background removal techniques- state of the art", Recent patents on computer science, Vol 1, pp 32-54, Apr, 2008.

[11] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," IEEE Trans. Pattern Anal. Mach. Intel., vol. 25, no. 5, pp. 564–577, May 2003.

[12] P. Soille, Morphological Image Analysis: Principles and Applications, 2nd ed. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2003.

[13] http://www.grss-ieee.org/2011-ieee-digitalglobe-data-fusion-contest/

[14] C. J. Van. Rijsbergen, Information Retrieval, 2nd ed. Newton, MA, USA: Butterworth-Heinemann, 1979.

[15] Lingfei Meng, student member, IEEE, and john P. KEREKES, Senior Member, IEEE.