

Motion detection and tracking in night-time video

Eran Stoyerman^a, Ari Krauss^a, Doron Shterman^a

^aImageSat International (ISI)

ABSTRACT

Previously, we reported on the RUNNER micro-satellite platform, describing the system and presenting results with True Color video products. By utilizing a 3CCD focal-plane-array (FPA) architecture, the RUNNER can capture Multi-Spectral (MS) images, where each spectral band is sampled independently. Using short video clips acquired by the system, we previously demonstrated real-time motion detection of moving vehicles in daylight conditions. More recently, we demonstrated the creation of informative night-time videos. In the present work, we extend this and present a framework for detecting and tracking moving objects in night-time satellite videos. Night-time illumination is typically very low and non-uniform, resulting in images with very low Signal-to-Noise Ratio (SNR). Under such conditions, object identification in individual frames is often not feasible. However, by exploiting the temporal dimension of video, moving vehicles can still be detected. When combined with geo-location information, this capability enables monitoring of traffic conditions in specific areas. Due to the limited availability of night-time data, training an AI-based model was not feasible at this stage. Instead, we employed a classical image-processing and tracking pipeline designed to operate under very low SNR conditions. The results demonstrate a high detection rate with a relatively low false-alarm rate. Notably, the proposed approach successfully detects moving vehicles that were not identifiable in individual frames due to the low SNR.

Keywords: 3CCD, True Color, Night-time imaging, Motion detection, Tracking, Micro-sat

1. INTRODUCTION

Earth-observation (EO) satellite capabilities have expanded considerably in recent years. Very-high-resolution imagery and multispectral and hyperspectral data spanning wavelengths from the ultraviolet (UV) to the long-wave infrared (LWIR) are now routinely acquired from space. However, until recently, the temporal dimension played only a limited role in high-resolution EO data. Conventional high-resolution EO systems have primarily been designed to acquire individual images or strips during an orbital pass, rather than continuous sequences of the same ground area. The emergence of satellite systems capable of staring, and lately systems with CMOS focal-plane-array (FPA) camera have introduced a new observation mode in which short videos or bursts of consecutive images can be collected over the same scene [1,2].

The availability of satellite video introduces the temporal dimension as an additional source of information and opens new possibilities for EO applications. Consecutive frames enable the detection and tracking of moving objects over large areas, including vehicles, ships, trains, and aircraft [3,4]. This capability has potential applications in traffic monitoring, urban analysis, disaster response, and security-related applications [5]. Furthermore, temporal information can provide information that is not readily available from an individual image. The redundancy and sub-pixel variations between consecutive satellite images can, for example, be exploited for multi-frame super-resolution, improving the effective spatial resolution beyond that of individual frames [6,7].

The RUNNER-1 satellite, developed by ImageSat International (ISI), was among the first EO satellites capable of acquiring very-high-resolution video imagery. In previous work [8], we demonstrated that this capability can be exploited to detect and track moving objects, primarily vehicles and vessels, over city-scale regions. RUNNER-1 also provides sufficiently high imaging sensitivity to acquire informative imagery under nighttime conditions, as demonstrated in Fig. 1. Nighttime remote sensing is by itself an established EO domain and has been used to characterize human activity, urbanization, infrastructure, disasters, and conflicts [9,10]. However, most previous nighttime EO studies have relied on individual images or temporally separated observations rather than very-high-resolution video sequences.



Figure 1. Night-time image of Tel-Aviv city captured by RUNNER-1

Combining nighttime imaging with motion detection therefore represents a natural extension of these capabilities. However, we found that algorithms developed for motion detection in daytime satellite video perform poorly when applied directly to nighttime imagery. Moving-object detection in satellite video is already challenging because targets such as vehicles occupy only a small number of pixels and often exhibit low contrast relative to the background [3,5]. These difficulties become considerably more pronounced at night. Moving objects may be barely distinguishable in individual frames, while the density of detectable spatial features is substantially lower than in daytime imagery. This is particularly relevant for vehicles, since much of the light emitted by their headlights is directed approximately horizontally and forward rather than toward a nadir-viewing satellite. Consequently, conventional image-registration and motion-estimation approaches that rely on sufficiently dense and stable spatial features become less effective.

This paper presents algorithms specifically developed to address these challenges. Combined with updated daytime motion-detection and tracking methods, the proposed approach enables the detection and tracking of moving objects in nighttime satellite video. By exploiting information accumulated across the temporal dimension, the method can detect moving objects that are only marginally visible - or, in some cases, effectively indistinguishable in individual frames. These results demonstrate the value of the temporal dimension not only for observing motion, but also as an additional source of information for extracting signals that cannot be reliably recovered from a single EO image.

2. RUNNER MICRO-SATELLITE PLATFORM

The RUNNER micro-satellite platform is a cost-effective satellite platform developed by ImageSat International (ISI). Its optical payload is an electro-optical imager capable of acquiring very-high-resolution (<1 m) true-color imagery and video products. The payload utilizes a three-CCD (3CCD) focal-plane-array (FPA) architecture, in which the spectral bands are acquired simultaneously by independent two-dimensional detectors. Consequently, each spectral band is sampled independently at the same ground sampling distance (GSD), without the spatial sampling differences inherent to systems combining a high-resolution panchromatic channel with lower-resolution multispectral channels. This architecture provides high spatial resolution, high signal-to-noise ratio (SNR), and full-resolution spectral information in each acquired frame.

Compared with other modern spaceborne multispectral imaging architectures, the RUNNER 3CCD imager does not rely on color-filter-array demosaicing for multispectral image reconstruction, nor does it require off-chip time-delay integration (TDI) or pan-sharpening to achieve the required SNR and spatial resolution. Furthermore, unlike conventional TDI-based

push broom imagers, the two-dimensional FPA enables repeated acquisition of the same ground area at short temporal intervals and therefore supports video acquisition at frame rates of several tens of frames per second. The high sensitivity and SNR of the RUNNER payload enable informative imagery to be acquired under low-light and night-time conditions. The ability to acquire sequences of consecutive frames further enables applications that cannot readily be achieved using conventional TDI-based imaging systems. These include multi-frame image enhancement and super-resolution, three-dimensional reconstruction, temporal cloud rejection through the exploitation of transient cloud gaps, and the detection and tracking of moving objects over large areas [8].

One of the challenges associated with FPA-based imaging, relative to TDI-based imaging, is the requirement to maintain sufficient line-of-sight stability throughout the exposure time. This requirement becomes particularly demanding under low-light conditions, for which longer exposure times are required. In RUNNER nighttime imaging, exposure times of up to approximately 30ms are used. The resulting stability requirements are addressed by accurate and stable attitude-control and pointing mechanisms, which ensure sufficiently low ground-image drift during each exposure. When coverage of larger areas is required, step-and-stare acquisition can be employed, sequentially directing the line of sight toward adjacent ground areas while maintaining the required stability during each acquisition period.

The first satellite of the RUNNER constellation was successfully launched into low Earth orbit aboard a Falcon 9 launch vehicle in June 2023. Table 1 summarizes the main parameters of the satellite. Subsequent RUNNER satellites are planned to provide similar or improved performance.

Table 1. Main parameters of the RUNNER satellite and electro-optical payload

Parameter	Value
Clear Aperture	0.35 [m]
f/#	9.8
Sensor pixel pitch	4.6 [μ m]
GSD	0.7 [m]
FOV	0.78 [deg]
Spectral Response	Blue: 400-495 [nm] Green: 495-580 [nm] Red: 580-670 [nm]
Swath	5.6 [km]
Frame Rate	Up to 30 [fps]
MTF	>10% @Nyquist
SNR	>100 per channel

The RUNNER payload supports conventional still-image acquisition as well as video and video-strip imaging modes. In conventional video mode, the two-dimensional FPA continuously observes the same ground area and acquires consecutive full-resolution frames at rates of up to several tens of frames per second. The exposure time is adapted to the available illumination and imaging conditions, with substantially longer exposures used for nighttime acquisitions, reaching approximately 30 ms. In video-strip mode, consecutive video frames are acquired while the observed ground footprint is continuously moving along a ground track, rather than remaining stationary. This mode extends the ground area covered within a single video acquisition while preserving spatial overlap between consecutive frames. The supported ground-track velocity is determined by the instantaneous SNR and therefore depends strongly on the illumination conditions of the observed scene. Under favorable illumination and high-SNR conditions, shorter exposure times allow higher ground-track velocities while maintaining sufficient image quality and inter-frame overlap. Under low-light conditions, longer

integration times are required to achieve adequate SNR, consequently limiting the allowable ground-track velocity to avoid excessive image smear and loss of useful temporal overlap.



Figure 2. RUNNER satellite platform during environmental testing phase

Both acquisition modes provide spatial and temporal information that can be exploited for motion detection and tracking. In nighttime sequences, individual moving targets may exhibit very low contrast and may be difficult to distinguish in a single frame, while their repeated appearance across consecutive frames provides additional temporal information. The processing approach presented in the following sections exploits this temporal redundancy to enhance the detectability of weak moving targets and estimate their trajectories, including targets that are barely distinguishable in individual frames.

3. MOTION DETECTION AND TRACKING ALGORITHMS

Motion detection in daytime satellite video is relatively straightforward due to the high signal-to-noise ratio (SNR) and the large number of detectable spatial features. Classical image-processing and feature-based methods were therefore successfully applied to RUNNER daytime imagery. However, direct application of these methods to nighttime imagery resulted in poor performance. The substantially lower target-to-background SNR and the sparse distribution of detectable features make both accurate frame registration and reliable identification of moving objects considerably more challenging.

Several AI-based approaches were initially evaluated, motivated by the demonstrated capabilities of modern learning-based methods for feature extraction, image matching, and object detection. However, available pretrained models were found to be insufficiently robust for the characteristics of the RUNNER nighttime imagery. Furthermore, the limited number of available nighttime video sequences was insufficient for effective task-specific retraining. Consequently, the approach adopted in this work combines enhanced versions of the classical algorithms previously developed for daytime motion detection with additional processing stages specifically designed for detecting and tracking objects in low-SNR nighttime imagery. The resulting processing chain enables moving objects to be detected and tracked even under conditions in which they are barely distinguishable in individual frames.

The complete processing chain is summarized in figure 3.

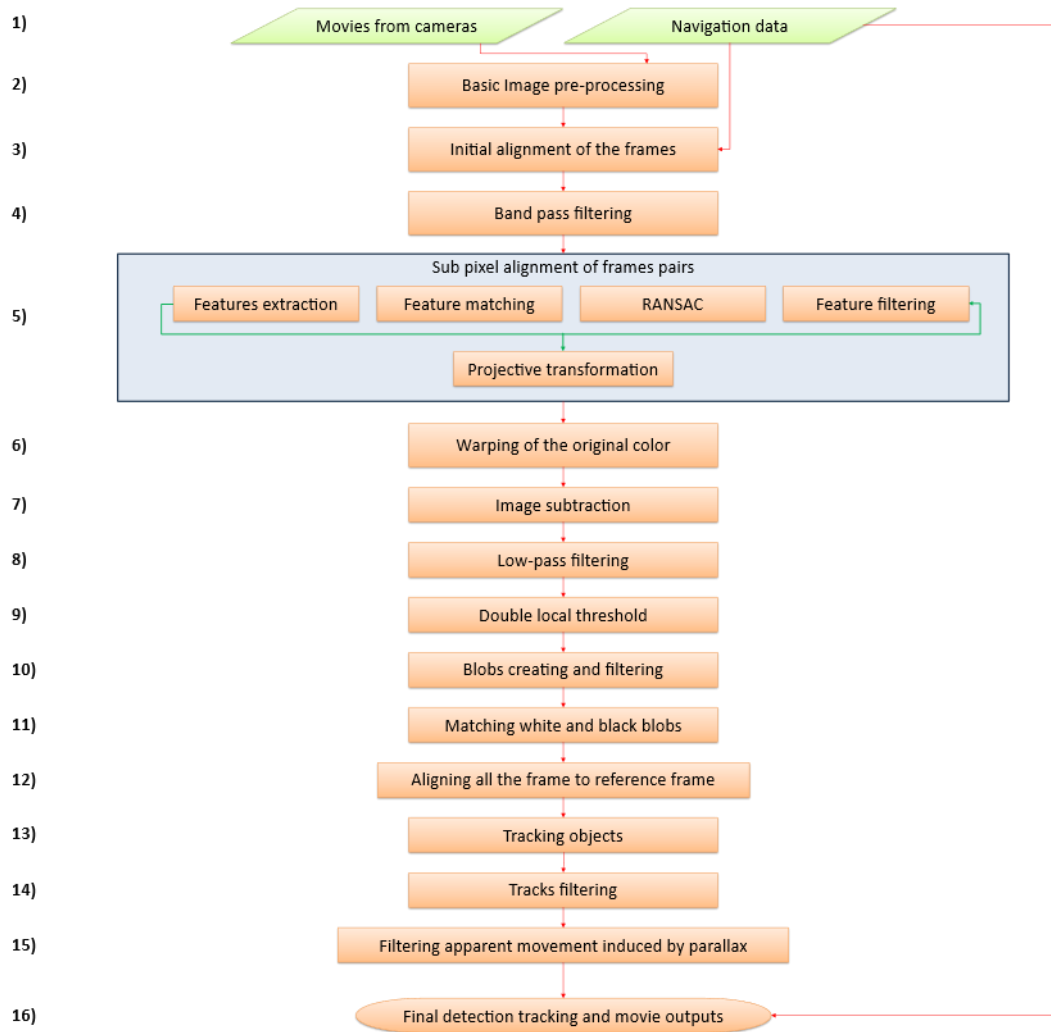


Figure 3. Proposed algorithm chain

Compared with the previously developed daytime processing chain, the nighttime implementation introduces several modifications aimed specifically at operation under low-SNR and sparse-feature conditions. These include dedicated band-pass filtering prior to feature extraction, iterative feature filtering for robust sub-pixel registration, locally adaptive double-threshold detection, positive/negative blob association, and tracking algorithms. Together, these modifications shift part of the detection process from individual-frame evidence to evidence accumulated across multiple frames. The individual processing stages are described below, with the numbering corresponding to the blocks in the flowchart (figure 3).

3.1 Input

The processing chain receives an unprocessed, geographically unanchored video sequence, typically consisting of several tens of frames. Satellite navigation and attitude data, obtained from the GPS receiver and star tracker, are also provided. These data are used for initial frame alignment and, at the final processing stage, for assigning geographical coordinates to the detected object trajectories.

3.2 Image pre-processing

Standard two-point non-uniformity correction (NUC) is first applied independently to each spectral channel, followed by correction of geometric distortions. Inter-channel registration is then performed to accurately align the three-color channels of the RUNNER-1 3CCD payload.

Point-spread-function (PSF) correction may be applied to improve the visual quality of the imagery for display purposes, however, PSF correction is not used in the motion-detection processing chain, since deconvolution may amplify image noise and consequently reduce the robustness of subsequent detection stages. Finally, bilateral filtering may be applied to suppress noise while preserving spatial features required for registration and motion detection.

3.3 Initial frame alignment using navigation data

The satellite position obtained from GPS data, together with the camera line-of-sight (LOS) information derived from the star trackers data, is used to obtain an initial alignment of all frames relative to a common reference frame, typically the middle frame of the sequence. This navigation-based alignment provides a useful initial estimate but is generally accurate only to within several pixels. Since frame subtraction is highly sensitive to registration errors, an additional image-based sub-pixel registration stage is required.

3.4 Band-pass filtering

The color imagery is converted to grayscale, and a narrow band-pass finite-impulse-response (FIR) filter is applied. The purpose of this stage is to suppress both low-frequency background variations and high-frequency noise while enhancing spatial structures useful for feature extraction. This processing significantly improves the availability and stability of interest points under low-SNR conditions and in scenes partially affected by clouds.

3.5 Sub-pixel alignment of frame pairs

Accurate motion detection by frame subtraction requires registration at the sub-pixel level. Each frame is therefore paired with another frame acquired several frames later. The temporal separation is selected according to the video frame rate and the expected velocity range of the objects to be detected, a separation of approximately four frames is typically used. A projective transformation between each frame pair is estimated using the following feature-based procedure:

3.5.1 Feature extraction

Several feature-extraction methods were evaluated, including learning-based features such as DINO [11], SIFT [12], ORB [13], and related approaches. The current implementation uses a SIFT-like feature extractor to identify interest points and calculate their corresponding descriptors. Future work will investigate task-specific learning-based feature extraction to improve robustness under very-low-SNR conditions.

3.5.2 Feature matching

Candidate correspondences between the extracted interest points are identified using a sum-of-squared-differences (SSD) distance metric. Only the few hundred highest-quality matches are retained for subsequent processing. The RoMa learning-based image-matching model [14] was also evaluated as an end-to-end matching approach, however, without task-specific training, it did not provide the registration accuracy required for the present application.

3.5.3 RANSAC estimation

Incorrect feature correspondences are rejected using a variant of the Random Sample Consensus (RANSAC) algorithm [15]. Random subsets of matched points are repeatedly evaluated to identify the projective transformation supported by the largest consistent set of correspondences. The outputs of this stage are an estimated projective transformation and a set of corresponding points between the two frames.

3.5.4 Iterative feature filtering

Registration accuracy is further improved by calculating the residual distance between each feature location in the first frame and the corresponding transformed feature location in the second frame. Matches exhibiting large residual errors are rejected as outliers, and the projective transformation is recalculated using RANSAC on the remaining correspondences. This procedure is repeated iteratively until a stable set of matches is obtained. The resulting projective transformation is used for accurate frame registration, while the retained feature correspondences are also used in the subsequent full-sequence stabilization stage.

3.6 Warping of the original color images

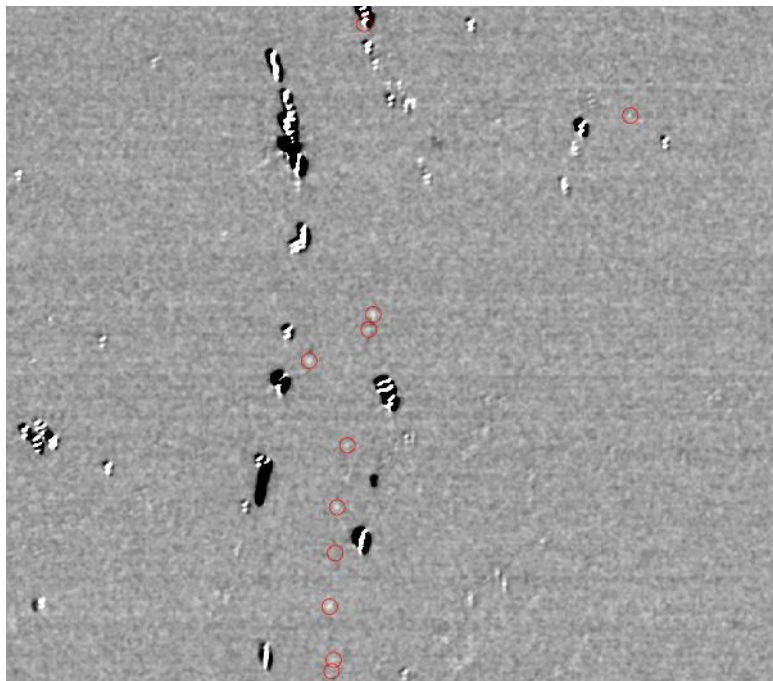
The final projective transformation obtained for each frame pair is applied to the original color imagery rather than to the band-pass-filtered images used for feature extraction. Following warping, the registered color frames are converted to grayscale using empirically selected linear coefficients optimized for the low SNR task. Each grayscale frame is subsequently normalized to compensate for global illumination variations between frames.

3.7 Image subtraction

For each accurately registered frame pair, the warped second frame is subtracted from the first frame. Ideally, stationary scene content is thereby suppressed, whereas objects that changed position between the two acquisitions produce a characteristic differential response. A moving object typically appears as a positive response at one location and a negative response at a nearby location, corresponding to its positions in the two frames.

3.8 Low-pass filtering

A small Gaussian FIR filter is applied to the difference images to suppress high-frequency noise and enhance the detectability of spatially coherent responses associated with moving objects. This stage is particularly important for nighttime imagery, where individual moving targets may have SNRs too low for reliable detection at the single-pixel level. [Video 1](#) is an example of movie after subtraction and filtering.



Video 1. Subtracted image after filtering. The red circles show detected moving objects after false detections have been filtered.

As can be seen in the video, some of the detected objects are not visible in single frames. The large white and black spots are apparent motions due to parallax from high buildings and were filtered out.

<https://youtu.be/qQi8boe1t7I>

3.9 Double local threshold

Two binary detection maps are generated from each filtered difference image: one representing significant positive changes and the other significant negative changes. Rather than applying a single global threshold, spatially varying thresholds are calculated from the local image statistics using a running-window standard deviation:

$$Threshold(x, y) = mean(running_window_{xy}) + N\sigma \cdot std(running_window_{xy})$$

Pixels exceeding the positive local threshold are assigned to the positive-change map, whereas pixels below the negative local threshold are assigned to the negative-change map. The threshold level is controlled by the parameter $N\sigma$, which is

positive for the positive-change and negative for the negative-change. This locally adaptive procedure enables weak motion signatures to be detected despite substantial spatial variations in the nighttime background and noise level.

3.10 Blob generation and filtering

Connected components are extracted independently from the positive- and negative-change binary maps using eight-neighbor connectivity [16]. For each resulting blob, several properties are calculated, including area, center of gravity, eccentricity, and mean intensity/color in the two original frames. These properties provide an initial mechanism for rejecting false detections. Blobs that are excessively small or large, or exhibit excessive eccentricity, are removed before the subsequent motion-association stage.

3.11 Matching positive and negative blobs

A moving object generally produces two nearby responses in the frame-difference image: a positive blob corresponding to its location in one frame and a negative blob corresponding to its location in the other. This characteristic signature provides an additional means of distinguishing genuine motion from random noise.

A spatial range search is therefore performed between positive and negative blobs, and candidates for which no corresponding opposite-polarity blob is found within the expected displacement range are rejected. Since the temporal separation between the paired frames is known, the distance between the matched positive and negative blobs also provides an initial estimate of the object's apparent velocity.

3.12 Full-sequence stabilization

Tracking an object over multiple frames requires the entire video sequence to be represented in a common coordinate system in which the stationary ground remains approximately fixed. For this purpose, reliable interest points identified during the pairwise registration stage (stage 3.5) are tracked across multiple frames. Only features that can be consistently tracked over a sufficient number of frames are retained.

These features are then used to estimate a projective transformation from each individual frame to a common reference frame, typically the middle frame of the sequence. Each frame is warped according to the resulting transformation, producing a stabilized video sequence suitable for temporal object tracking.

3.13 Object tracking

The positions of the filtered blobs are transformed into the stabilized reference-frame coordinate system. A Multiple Hypothesis Tracking (MHT) algorithm [17] is then used to associate detections across consecutive frames and construct candidate moving objects trajectories.

For each detected blob, candidate associations are searched for within subsequent frames based on spatial proximity and similarity in properties such as size, intensity/color, and eccentricity. Multiple candidate trajectories may initially be retained. Track-level properties are subsequently calculated, including track duration, mean velocity, fraction of missing detections, motion consistency. When several candidate trajectories originate from the same detection, the most probable track is selected according to these properties.

3.14 Track filtering

The calculated track properties are used to eliminate residual false motion detections. The most discriminative parameters were found to be the number of frames in which the object is detected, the consistency of its apparent motion, and the consistency of its blob intensity/color characteristics. Applying filtering at the track level provides considerably stronger rejection of random noise detections than can be achieved from individual frames alone.

3.15 Suppression of parallax-induced apparent motion

Changes in the camera LOS during video acquisition produce parallax-induced apparent motion of elevated structures, particularly high-rise buildings. These structures may generate motion signatures like those produced by genuinely moving objects.

Several blob and track properties, including size, shape, and apparent velocity, are used to distinguish these artifacts from actual moving objects. In addition, the expected trajectory and velocity profile of parallax-induced motion are estimated. Such trajectories are generally approximately linear, with a slowly varying apparent velocity determined by the imaging geometry. Combining this geometric information with the blob and track-level characteristics enables most parallax-induced detections to be rejected. However, the relatively short duration of the currently available video sequences limits

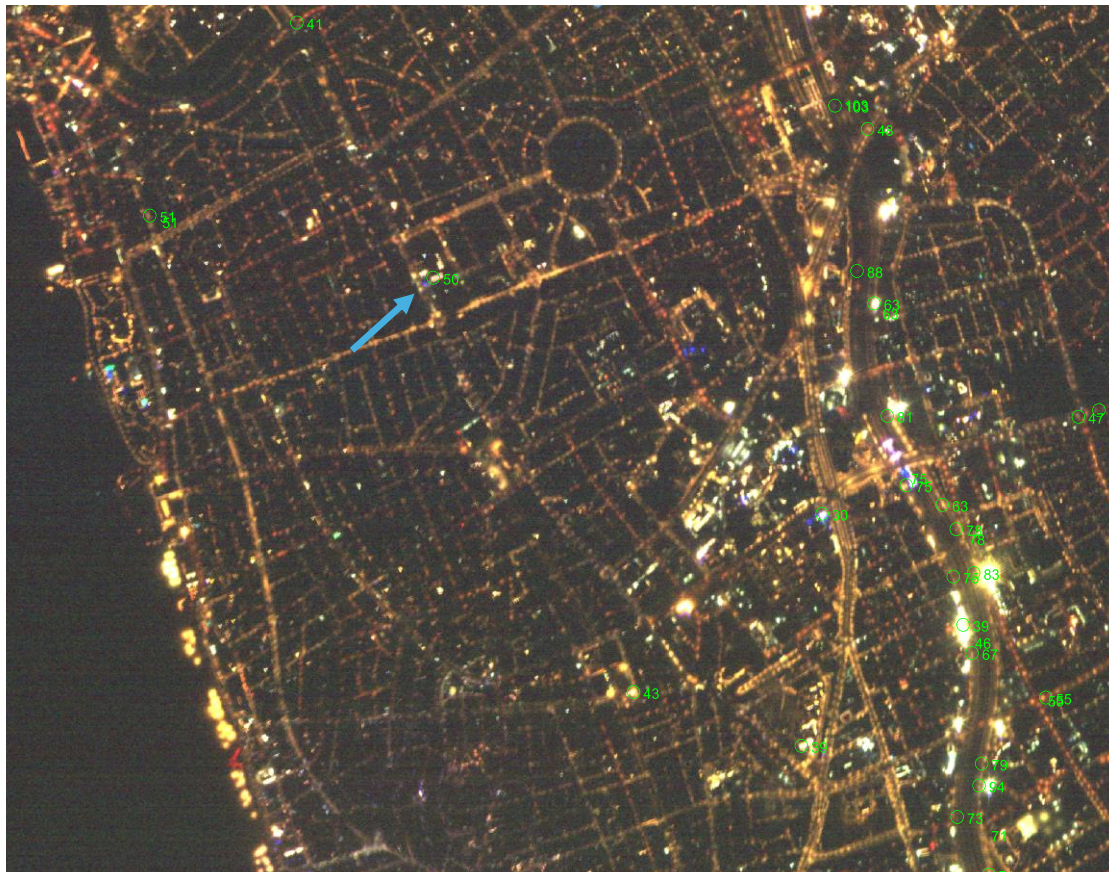
the accuracy with which the complete velocity profile can be estimated, and some residual parallax artifacts may therefore remain. An example is shown in Video 2.

3.16 Final outputs

The final processing products include:

1. The geographical trajectories of the remaining detected moving objects calculated using satellite navigation information.
2. Track information for each detected object, including its estimated velocity and size.
3. A stabilized color video containing geographical information and the detected object trajectories. For visualization, the video is enhanced using an adaptive Gaussian filter, with stronger smoothing applied to dark regions and weaker smoothing to brighter regions.
4. A stabilized image generated by weighted averaging of the registered video frames. Temporal averaging provides improved SNR relative to individual frames.

The absolute geographical accuracy of the stabilized imagery and extracted trajectories can be further improved to better than several meters by registering the imagery against an accurately georeferenced orthophoto.



Video 2. Final detection result. The green circles show the detected moving objects after filtering, the numbers indicate the calculated mean velocity of the objects. The background video is stabilized so that the ground is stationary. One parallax-induced apparent motion was not filtered out (marked with a blue arrow).

<https://youtu.be/65pV-kpJUx8>

4. SUMMARY AND FUTURE WORK

Until recently, the temporal dimension played only a limited role in very-high-resolution Earth-observation (EO) imagery, partly due to the widespread use of TDI-based imaging architectures. RUNNER-1 is among a new generation of micro-satellites capable of acquiring very-high-resolution video [8], thereby providing high-temporal-frequency information in addition to conventional spatial and spectral information. Combining this additional data dimension with algorithms capable of extracting information under both high and low-SNR conditions has the potential to significantly expand the capabilities of small-satellite EO systems, particularly when deployed as part of satellite constellations. Potential applications include defense, homeland security, transportation monitoring, and urban analysis.

In this paper, we presented algorithms for extracting spatio-temporal information from low-SNR satellite video. The proposed processing chain combines accurate sub-pixel image registration, change detection, object tracking and rejection of false motion induced by effects such as parallax. The effectiveness of the approach was demonstrated by detecting and tracking moving objects in low-SNR nighttime video acquired by the RUNNER-1 micro-satellite. Importantly, the use of temporal information enables the detection of moving objects that are barely distinguishable, or in some cases effectively indistinguishable, in individual frames. The results therefore demonstrate the value of the temporal dimension not only for observing motion, but also for extracting information that cannot be reliably obtained from a single EO image.

In the near future, ISI plans to launch additional satellites with capabilities similar to or exceeding those of RUNNER-1. The development of time-domain processing algorithms will consequently be extended from individual satellites to constellations, potentially providing substantially richer spatio-temporal information through observations acquired from different satellites, viewing geometries, and acquisition times.

Beyond motion detection and tracking, future work will investigate additional methods for exploiting the temporal information available in video sequences. These include multi-frame three-dimensional reconstruction, removing clouds through the exploitation of transient cloud gaps and multi-frames super-resolution. More generally, combining information from multiple frames may enable the recovery of spatial or temporal information that is weak or unavailable in individual observations.

Furthermore, the algorithms are currently being adapted for onboard execution. Rather than downlinking complete high-data-rate video sequences, onboard processing could enable the satellite to autonomously detect and track objects or events of interest, extract their relevant parameters, and prioritize the information to be transmitted. Compact information products, such as object coordinates, trajectories, velocities, alerts, or selected regions of interest, could then be transmitted with substantially lower bandwidth and latency than the original video data. Combined with onboard decision-making and autonomous satellite operation, this approach could provide relevant information to users in near real time, through low-bandwidth communication channels [18].

REFERENCES

- [1] J. Anger, T. Ehret, C. de Franchis, and G. Facciolo, "Fast and accurate multi-frame super-resolution of satellite images," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-1-2020, pp. 57–64, 2020. doi: 10.5194/isprs-annals-V-1-2020-57-2020.
- [2] J. Anger, T. Ehret, and G. Facciolo, "Parallax estimation for push-frame satellite imagery: Application to super-resolution and 3D surface modeling from SkySat products," *2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 2679–2682, 2021. doi: 10.1109/IGARSS47720.2021.9554340.
- [3] Q. Yin, Q. Hu, H. Liu, F. Zhang, Y. Wang, Z. Lin, W. An, and Y. Guo, "Detecting and tracking small and dense moving objects in satellite videos: A benchmark," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, Art. no. 5612518, 2022. doi: 10.1109/TGRS.2021.3130436.
- [4] J. Feng, D. Zeng, X. Jia, X. Zhang, J. Li, Y. Liang, and L. Jiao, "Cross-frame keypoint-based and spatial motion information-guided networks for moving vehicle detection and tracking in satellite videos," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 177, pp. 116–130, 2021. doi: 10.1016/j.isprsjprs.2021.05.005.

- [5] M. Shu, Y. Zhong, and P. Lv, "Small moving vehicle detection via local enhancement fusion for satellite video," *International Journal of Remote Sensing*, vol. 42, no. 19, pp. 7189–7214, 2021. doi: 10.1080/01431161.2021.1944694.
- [6] J. Anger, T. Ehret, C. de Franchis, and G. Facciolo, "Fast and accurate multi-frame super-resolution of satellite images," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-1-2020, pp. 57–64, 2020.
- [7] P. Wang and E. Sertel, "Multi-frame super-resolution of remote sensing images using attention-based GAN models," *Knowledge-Based Systems*, vol. 266, Art. 110387, 2023. doi: 10.1016/j.knosys.2023.110387.
- [8] D. Shterman, A. Krauss, and E. Shtoyerman, "True color very-high-resolution imaging with runner microsat platform," *Proc. SPIE 13699, International Conference on Space Optics—ICSO 2024*, 1369970 (2025). doi: 10.1117/12.3071582.
- [9] N. Levin et al., "Remote sensing of night lights: A review and an outlook for the future," *Remote Sensing of Environment*, vol. 237, Art. 111443, 2020. doi: 10.1016/j.rse.2019.111443.
- [10] Q. Zheng, K. C. Seto, Y. Zhou, S. You, and Q. Weng, "Nighttime light remote sensing for urban applications: Progress, challenges, and prospects," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 202, pp. 125–141, 2023. doi: 10.1016/j.isprsjprs.2023.05.028.
- [11] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin, "Emerging Properties in Self-Supervised Vision Transformers," *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9650–9660, 2021. DOI: 10.1109/ICCV48922.2021.00951.
- [12] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [13] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An Efficient Alternative to SIFT or SURF," *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2564–2571, 2011. DOI: 10.1109/ICCV.2011.6126544
- [14] J. Edstedt, Q. Sun, G. Bökman, M. Wadenbäck, and M. Felsberg, "RoMa: Robust Dense Feature Matching," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19790–19800, 2024
- [15] M. A. Fischler and R. C. 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. doi: 10.1145/358669.358692
- [16] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, vol. 1, Addison-Wesley, 1992, pp. 28–48.
- [17] I. J. Cox and S. L. Hingorani, "An Efficient Implementation of Reid's Multiple Hypothesis Tracking Algorithm and Its Evaluation for the Purpose of Visual Tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 2, pp. 138–150, Feb. 1996. doi: 10.1109/34.481539
- [18] E. Shtoyerman, R. Cohen, R. Cohen, O. Fatkiev, G. Zaidman, G. Pinchas, and D. Shterman, "Semi-Autonomous Reconnaissance Satellite: A Laboratory Demonstration of End-to-End Onboard Mission Autonomy," *Proc. 40th Annual Small Satellite Conference*, Salt Lake City, Utah, USA, 23–26 Aug. 2026.