Thermal Infrared Video Stabilization for Aerial Monitoring of Active Wildfires

Mario Miguel Valero, Steven Verstockt, Bret Butler, Daniel Jimenez, Oriol Rios, Christian Mata, LLoyd Queen, Elsa Pastor, and Eulàlia Planas

***Abstract*—Measuring wildland fire behavior is essential for fire science and fire management. Aerial thermal infrared (TIR) imaging provides outstanding opportunities to acquire such information remotely. Variables such as fire rate of spread (ROS), fire radiative power (FRP), and fireline intensity may be measured explicitly both in time and space, providing the necessary data to study the response of fire behavior to weather, vegetation, topography, and firefighting efforts. However, raw TIR imagery acquired by unmanned aerial vehicles (UAVs) requires stabilization and georeferencing before any other processing can be performed. Aerial video usually suffers from instabilities produced by sensor movement. This problem is especially acute near an active wildfire due to fire-generated turbulence. Furthermore, the nature of fire TIR video presents some specific challenges that hinder robust interframe registration. Therefore, this article presents a software-based video stabilization algorithm specifically designed for TIR imagery of forest fires. After a comparative analysis of existing image registration algorithms, the KAZE feature-matching method was selected and accompanied by pre- and postprocessing modules. These included foreground histogram equalization and a multireference framework designed to increase the algorithm's robustness in the presence of missing or faulty frames. The performance of the proposed algorithm was validated in a total of nine video sequences acquired during field fire experiments. The proposed algorithm yielded a registration accuracy between 10 and 1000× higher than other tested methods, returned 10× more meaningful feature matches, and proved robust in the presence of faulty video frames. The ability to automatically cancel camera movement for every frame in a video sequence solves a key limitation in data processing pipelines and opens the door to a number of systematic fire behavior experimental analyses. Moreover, a completely automated process supports the development of decision support tools that can operate in real time during an emergency.**

***Index Terms*—Fire behavior, image registration, KAZE, remote sensing, unmanned aerial systems (UAS), video stabilization, wildland fire.**

Manuscript received September 17, 2020; revised December 28, 2020 and January 21, 2021; accepted February 1, 2021. Date of publication February 12, 2021; date of current version March 11, 2021. This work was supported in part by the Spanish Ministry of Education, Culture, and Sport under Ph.D. Grant FPU13/05876, in part by the Spanish Ministry of Economy and Competitiveness under Project CTM2014- 57448-R and Project CTQ2017-85990-R, funded with FEDER funds, and in part by the Autonomous Government of Catalonia under Grant 2017-SGR-392. The work of Mario Miguel Valero was supported in part by the Erasmus+ Traineeship Program and Obra Social La Caixa Research Mobility Grants. *(Corresponding author: Eulàlia Planas.)*

Mario Miguel Valero is with the Centre for Technological Risk Studies, Universitat Politècnica de Catalunya, 08019 Barcelona, Spain and also with the Wildfire Interdisciplinary Research Center, San Jose State University, San Jose, CA 95192 USA (e-mail: mm.valero@pm.me).

Steven Verstockt is with the IDLab, Ghent University—imec, 9502 Ghent, Belgium (e-mail: steven.verstockt@ugent.be).

Bret Butler and Daniel Jimenez are with the Missoula Fire Sciences Lab, US Forest Service Rocky Mountain Research Station, Missoula, MT 59808 USA (e-mail: bret.butler@usda.gov; dan.jimenez@usda.gov).

Oriol Rios, Christian Mata, Elsa Pastor, and Eulàlia Planas are with the Centre for Technological Risk Studies, Universitat Politècnica de Catalunya, 08019 Barcelona, Spain (e-mail: oriolrios@gmail.com; christian.mata@upc.edu; elsa.pastor@upc.edu; eulalia.planas@upc.edu).

LLoyd Queen is with the National Center for Landscape Fire Analysis, University of Montana, Missoula, MT 59812 USA (e-mail: lloyd.queen@mso.umt.edu).

This article has supplementary downloadable material available at https://doi.org/10.1109/JSTARS.2021.3059054, provided by the authors.

Digital Object Identifier 10.1109/JSTARS.2021.3059054

I. Introduction

AERIAL thermal infrared (TIR) imaging is widely used to acquire detailed spatial information about active wildfires. Data such as the location of the fire perimeter, its rate of spread (ROS), fireline intensity, and fire radiative power (FRP) can be computed from georeferenced TIR footage [1]–[9]. This information has subsequently been used for additional fire behavior analysis and evaluation of suppression activities [10]–[12]. Furthermore, there is a growing trend to incorporate observed fire perimeter evolution into operational fire spread simulators in order to improve forecasts using data assimilation techniques [13]–[17].

However, there are a number of unsolved issues in the data processing pipeline that prevent full exploitation of unmanned aerial vehicle (UAV) potential for wildfire remote sensing. Limitations begin at the very first processing step with image registration. For aerial imagery to be useful, it must be georeferenced so that image information can be projected onto geographic coordinates. In theory, there are a few alternatives to achieve this goal, but none has proved to be completely successful thus far. A few modern dedicated airborne monitoring systems, such as the NASA Ikhana UAV [18], [19], are able to geocorrect acquired imagery on board, automatically and even in real time, using high-accuracy positioning systems and powerful computing units. However, completely autonomous on-board processing is only possible if very accurate information about camera position and orientation is provided by high-performance global positioning systems and inertial measurement units (IMUs). Although achieving sufficient location and orientation accuracy is technically feasible at present, the equipment capable of doing this is usually tailor-made, heavy, and expensive. For example, the autonomous modular sensor (AMS), installed aboard the NASA Ikhana UAV, weighs 109 kg [20]. Most frequently, data are acquired by less sophisticated platforms, usually consisting

of commercial off-the-shelf (COTS) small unmanned aerial systems (UAS) or hand-held cameras operated manually from a helicopter—see, for instance, [3], [5], [10], [21], [22]. In none of these studies could aerial footage be automatically georeferenced, i.e., registered onto a digital elevation model.

To date, thermal image geocorrection for wildfire research has required manual identification of ground control points (GCPs) [2], [3], [6], [7], [9], [23], [24]. This approach is easy to implement and allows achieving high-accuracy results under certain conditions, but it has very important limitations in practice.

First, identifying and annotating GCPs in every video frame is extremely time consuming. A minimum of four GCP pairs are needed to estimate the projective transformation between two planes, and higher amounts of GCPs are required when remote sensing information must be projected onto nonflat terrain. Moreover, manual identification and annotation of GCPs are prone to errors in feature detection, localization, and matching. Resulting inaccuracies are hard to quantify and minimize due to the small amounts of feature matches typically available, which prevent the application of statistical methods. While the existence of these errors encourages the annotation of additional GCPs, reaching a sufficient amount of point matches is usually unfeasible.

Furthermore, the user time required for point annotation imposes a limit to the maximum temporal resolution at which fire behavior can be studied. Modern TIR imaging sensors are capable of recording high-frequency video with high spatial resolution and low hardware requirements [25]. There are a number of fire dynamics aspects that could be studied if high-frequency infrared (IR) video could be correctly georegistered [26]. However, due to the cost of GCP annotation, the video sampling frequency is usually reduced to some value between 1 and 0.1 Hz, which is around 30–300 times lower than the technical limit.

Finally, limiting the information used for registration to a few points per image may result in further unexpected data losses. Several GCPs may stop being identifiable for various reasons, and this may prevent image registration. For example, GCPs may disappear from the field of view or they may become occluded by trees, buildings, or other aircraft. Similarly, a common practice in experimental controlled burns consists of using fire beacons as GCPs, but this approach may fail if beacons burn or blow out.

Due to these limitations, image registration has recently been identified as one of the most significant sources of uncertainty when computing fire behavior properties from aerial IR imagery [27], and several authors have reported significant challenges when registering, stabilizing, and georeferencing aerial TIR video of active fires [6], [7], [20].

A promising strategy to solve existing limitations consists of automated video stabilization. Most of the time that remote sensing equipment is deployed during an active wildfire, the fire is recorded from a quasi-static overhead position. A nominal vantage point is selected and a camera is set in that position either in a tower or on hovering rotary-wing aircraft. With this configuration, and if the camera was totally still, GCP annotation in one single reference frame would suffice to georeference the complete video sequence. However, the wildfire environment is characterized by highly turbulent winds fostered by the interaction among atmospheric wind, topography, and buoyancy originated from the fire [28]–[33]. This environment generates undesired displacements in any rotary-wing aircraft hovering near the fire [7], with small UAS being especially affected by aerodynamic turbulence. Mechanical stabilization systems can reduce camera movement, but they do not cancel it completely [25]. Video jitter has been observed even when installing IR cameras on ground boom lifts and towers [24]. On the contrary, if camera movement can be estimated and canceled using software-based solutions, resulting still video becomes significantly easier to analyze both qualitatively and through computer vision algorithms. Stable footage can be georeferenced using a single geometric transformation, and the optimum transformation can be robustly estimated by combining GCPs visible at different times.

Video stabilization is a rather mature field in other remote sensing areas, and it has successfully been used in combination with GCPs to improve the geocorrection accuracy of visible imagery [34]. However, to the best of our knowledge, this approach has not yet been extended to thermal IR imagery of fire. Stabilization of fire imagery acquired in the TIR range entails additional challenges due to limitations in image resolution and a lower level of detail. Because of the high-temperature measurement ranges required to observe fire, cold details in the image background are rarely visible in aerial TIR imagery. Furthermore, the fire itself, which is sometimes the only visible item, is highly dynamic, and this hinders the identification of persistent features in consecutive frames. Some authors have proposed the use of phase correlation in the Fourier–Mellin space for the registration of nonfire aerial TIR imagery [35]. However, we are not aware of any system capable of solving this issue in a wildfire scenario. In fact, recent review articles have concluded that in the field of forest fire monitoring, *a simple and low-cost image processing approach that can be used for image motion calculation and elimination is still in high demand* [23].

Therefore, this article presents a video stabilization algorithm specifically designed for aerial TIR imagery of wildfire. First, Section II describes the type of data to be processed. Afterward, Section III details the study design we followed, the registration methods that were considered, and the proposed video stabilization algorithm. Section IV presents and discusses the results of this study, which include data about registration accuracy, method robustness, and video stabilization performance, plus a demonstration of algorithm deployment in three real study cases. Finally, Section V summarizes the contributions of this article and future work.

II. Data

The TIR video used in this study was recorded during nine field experiments conducted between 2008 and 2018. Remote sensing data were acquired by different teams, using dissimilar setups and equipment. These nine fires were selected as a representative sample of the varied fire behavior and imaging conditions present in wildfire research.

In all scenarios, experimental fires propagated on flat terrain and fire evolution were recorded from vantage points using TIR cameras. Still, the scenarios showed significant differences in

TABLE I
CAMERA PROPERTIES AND PARAMETERS USED TO RECORD THE ANALYZED FOOTAGE

Scenario	Camera commercial name	Spectral range (wavelength, μm)	Brightness temperature range ($^{\circ}C$)	Image resolution (pixels)	Field of view ($^{\circ}$)	Recording frequency (Hz)
1	Optris PI 640	[7.5, 13]	[20, 900]	640 x 480	60 x 45	32
2	Optris PI 400	[7.5, 13]	[200, 1500]	382 x 288	60 x 45	27
3	Optris PI 400	[7.5, 13]	[200, 1500]	382 x 288	60 x 45	27
4	FLIR SC660	[7.5, 13]	[300, 1500]	640 x 480	45 x 30	1
5	FLIR SC660	[7.5, 13]	[300, 1500]	640 x 480	45 x 30	1
6	FLIR SC660	[7.5, 13]	[0, 550]	640 x 480	45 x 30	1
7	FLIR AGEMA Thermovison 570-Pro	[7.5, 13]	[30 800]	240 x 320	24 x 18	4
8	FLIR AGEMA Thermovison 570-Pro	[7.5, 13]	[30 800]	240 x 320	24 x 18	4
9	FLIR AGEMA Thermovison 570-Pro	[7.5, 13]	[30 800]	240 x 320	24 x 18	4

the experimental setup, fuels, and spatial scale. This variability affected the observed fire behavior. Appreciable differences occurred in flaming activity and ROS, both of which have an important impact on image content as well as its evolution with time. Moreover, the distance between camera and fire ranged from a few to more than a hundred meters. Different camera models, provided by various manufacturers, were used to monitor fire behavior from different perspectives. The disparity in camera resolution, together with variations in their distance to fire, contributed to inhomogeneous pixel sizes. Moreover, although all cameras worked in the same spectral range (long-wave IR), radiance measurement ranges also differed among tests. The radiance measurement range selected to record a TIR video sequence has a critical impact on the amount of image detail. Background objects with brightness temperature below the selected measurement range are invisible, whereas zones with a brightness temperature above the selected range are likely to cause sensor saturation. Both situations are highly detrimental to image processing techniques. Finally, video recording frequency varied between 1 and 32 frames per second.

The experiment in *Scenario 1* was conducted in 2018 at the Centre for Technological Risk Studies (Universitat Politècnica de Catalunya—BarcelonaTech). A homogeneous bed of straw was burned over a 1.5×3 m combustion table to reproduce fire spread on a flat horizontal surface with no wind. *Scenarios 2* and *3* were recorded at the Tall Timbers Research Station in Tallahassee, FL, USA, in April 2017. These video sequences were acquired during a set of small-scale experimental burns on mixed rough/long leaf pine fuels. *Scenarios 4*, *5,* and *6* correspond to tests S3, S4, and S5, respectively, pertaining to the Prescribed Fire Combustion and Atmospheric Dynamics Research Experiment (RxCADRE, FL, USA, 2012) [24], [36]. Burned vegetation was a mix of grass and shrubs, predominantly turkey oak. Finally, *Scenarios 7*, *8,* and *9* were recorded during another set of large-scale field experiments conducted in the Ngarkat Conservation Park, South Australia [11], [37]. These three video sequences correspond to a series of controlled burns in horizontal mallee-heath shrub plots with areas ranging from 4 to 25 ha.

Despite the significant variability among test scenarios, video sequences 1–6 have an important property in common: they were all acquired from stable vantage positions with a clear overhead view of the fire. This property was exploited here for algorithm development and validation: the application of known synthetic vibrations to the stable video facilitated the quantitative measurement of image registration performance and, therefore, the comparison of different registration methods.

Conversely, video sequences 7–9 were recorded from a hovering helicopter under circumstances similar to an actual wildfire. These datasets were used for the demonstration of the proposed video stabilization algorithm in independent footage with real jitter.

Sample frames from each video sequence are displayed in Fig. 1, while Table I summarizes the most relevant technical information about the deployed thermal cameras.

III. METHODOLOGY

Video stabilization software is primarily based on image registration. Owing to the wide range of applications of image registration and the variety of image types available, a good number of registration methods have been developed [38], [39], some of them specifically designed for remote sensing applications (e.g., [40]–[44]). However, to date, there has been no detailed analysis of image registration in wildfire monitoring scenarios. As the first step to build our video stabilization algorithm, we conducted a comparative analysis of existing image registration techniques in order to assess their performance on wildfire TIR imagery. Subsequently, we selected the best-performing method and designed additional pre- and postprocessing operations to optimize the accuracy and robustness of the algorithm. Section III-A summarizes the tested registration methods, Section III-B describes the study design, and Section III-C details the implementation of our complete video stabilization algorithm.

A. Comparative Analysis of Image Registration Methods

Despite the wide variety of existing image registration methods, the majority of the algorithms follow one of three approaches: image similarity maximization, phase correlation, or feature matching. In this study, we included some of the most widespread methods within each of these categories.

1) Image Similarity Metrics and the Optimization Problem: Assuming that two images are correctly aligned when their similarity is maximum, registration may be understood as an optimization problem with a certain similarity metric as a cost function. This constitutes the most generic approach with barely any need for *a priori* information.

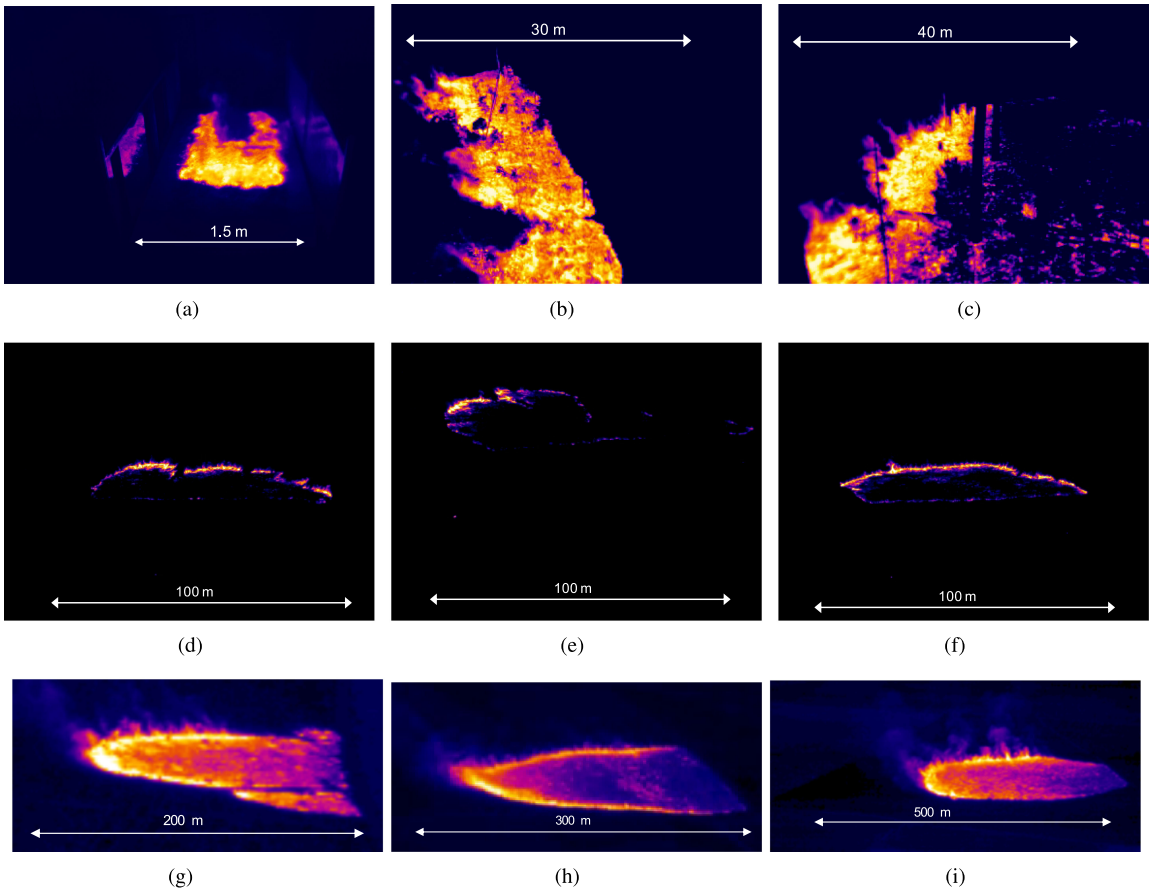

Fig. 1. Sample frames of the nine video sequences used in this study. Length values indicate approximate ground distances. (a) Scenario 1. (b) Scenario 2. (c) Scenario 3. (d) Scenario 4. (e) Scenario 5. (f) Scenario 6. (g) Scenario 7. (h) Scenario 8. (i) Scenario 9.

Image similarity can be estimated using various metrics, and how image similarity is measured has implications for registration algorithms. After a comparative analysis of various popular alternatives, our previous results encouraged the use of mutual information (MI) as a similarity metric for TIR image registration in wildfire contexts [45].

MI is a widely adopted metric to measure image similarity during registration problems in remote sensing applications [41], [43]. It has favorable properties for image similarity measurement because it does not rely on any assumption about the nature of the images or their relation. Therefore, MI provides precise and reliable information about how much information two images share even if they were acquired by different sensors or under dissimilar lighting conditions. Furthermore, MI has showed sharper peaks than cross correlation at the point of correct alignment [46] and higher robustness in front of Gaussian noise and nonlinear intensity relationships between images [43].

Nevertheless, MI itself is a similarity measure, not a registration technique. Similarity information is not accompanied by an estimation of the optimal transformation that should be applied to register two images. This limitation has traditionally made it challenging to find global optima [41]. Although some authors have addressed this issue [41], [47] and proposed mathematical and computational models that facilitate and accelerate convergence [43], poor convergence and high computation cost remain two critical limitations of MI-based registration algorithms. At present, there is active research aimed at improving the MI maximization scheme [48].

2) Phase Correlation and the Fourier–Mellin Transform: Phase correlation was first proposed in [49] as an algorithm to align two images that were shifted with respect to each other. Like similarity maximization algorithms, it uses all intensity information contained in the image. The main difference between phase correlation and other intensity-based methods resides in the mathematical space used for image comparison. By studying images in the frequency domain, phase correlation takes advantage of the Fourier shift theorem, which states that a displacement in the space domain is transformed into a change in phase in the frequency domain. Based on this property, relative translation between two identical images can be found uniquely without the need for an optimization algorithm.

However, the phase correlation method is valid only for pure translations. In order to accommodate rotations and changes in scale, Reddy and Chatterji [50] proposed a revised algorithm that combines phase correlation with a log-polar transformation. The use of polar coordinates allows handling rotations as translations of the independent variable, while expressing the signal modulus in logarithmic scale also converts scaling to a translational movement. These properties allow applying the Fourier shift theorem in the different spaces in order to sequentially retrieve

rotation, scaling, and translation. Complete details about the mathematical algorithm and its original implementation can be found in [50].

The log-polar version of the Fourier transform, also known as Fourier–Mellin transform, has been extensively used in image registration, video stabilization, and global motion estimation problems [51]–[55]. Moreover, its original formulation has been revised and coupled with additional algorithms to increase its robustness and reduce its computational requirements [53]. Therefore, registration based on phase correlation and the Fourier–Mellin transform was included in our comparative study.

3) Feature-Based Methods: In order to reduce the cost of a full 2-D intensity comparison between two images, it is a common approach to limit the comparison to a number of sampled features. Usually, small local features that are invariant to scale and rotation are sought and characterized by a combination of descriptors that allow their comparison. Because feature description is highly dependent on the type of identified features, detectors and descriptors are usually coupled together. Widely used feature descriptors include the scale-invariant feature transform [56], speed-up robust features (SURF) [57], maximally stable extremal regions (MSER) [58], [59], and KAZE [60].

Features detected in the images to be registered are then compared and matched. However, estimation of the correct geometric transformation from a set of feature matches might not be trivial. Due to the nature of feature detection methods, the amount of mismatches is usually sufficient to mislead least-squares estimators [61]. A more robust approach that has been widely employed with success relies on random sampling, with the most famous algorithm of this type being the random sample consensus (RANSAC) [62]. RANSAC is an iterative algorithm in which certain hypotheses are generated from reduced sample subsets and subsequently verified with the complete dataset. After its original publication in 1981, several revisions and computational optimizations have been proposed [61], [63], [64].

Feature matching has been successfully used in medium-altitude UAS monitoring systems. Li *et al.* [65] used edge detectors for precise image registration after a previous coarse motion estimation step. However, feature selection, description, and matching vary greatly depending on the image nature, which prevents the design of a universal feature-based algorithm. Fire TIR imagery poses important challenges for feature-based registration as points, edges, and corners are not easily identifiable in TIR fire imagery. When they are, they usually correspond to flames, which are highly dynamic. Therefore, SURF, KAZE, and MSER feature detectors and descriptors were included in this study because each of them is built upon different principles and low-level image characteristics.

B. Study Design

In order to assess the performance of different image registration approaches in a wildfire scenario, we performed a systematic comparative study in a controlled environment. Registration accuracy was analyzed through the application of synthetic movement to still TIR video. One hundred frames were randomly sampled from each of the video sequences 1–6. Afterward, random similarity transformations were applied to the selected frames. Maximum perturbation ranges were set to $\pm 20\%$ of frame width and height for horizontal and vertical translation, respectively, $\pm 25^\circ$ for rotation and $\pm 20\%$ for scaling. After perturbation, the transformed frames were registered back to their original position using each of the tested registration algorithms.

Under normal working conditions, video stabilization is to be performed by registering each video frame with a previous, i.e., different, frame. However, it is difficult to conduct a generic analysis of registration performance under such realistic operation conditions as method response to the recording frequency may vary widely among registration strategies.

In order to account for this, we compared the performance of selected registration algorithms under two different sets of conditions. First, we registered each perturbed frame back to its original position using the same original frame as a reference for registration. Registering each frame with itself instead of a previous frame allowed decoupling the algorithm response to camera movement and recording frequency. We refer to this approach as *idealized working conditions*. Afterward, we tested the same registration methods under *realistic working conditions* by sampling video frames at 1 Hz and registering each perturbed frame with a previous stable frame, sampled 1 s, before the perturbed frame.

Registration accuracy was evaluated twofold. On the one hand, by comparing retrieved translations, rotations, and scaling with the originally applied transformations. On the other, by computing similarity between the output registered frame and the original—target—frame. Image similarity was measured using 2-D correlation as recommended in [45]. Each registered frame was evaluated using as reference the stable version of the same frame, both under idealized and realistic working conditions. Therefore, the maximum achievable correlation was always exactly 1. Furthermore, computation times required by each registration algorithm were measured for every frame and averaged over the complete study sample. These times were measured on a laptop computer equipped with an Intel i7-4700MQ CPU and 8.0 GB of RAM.

In addition to registration accuracy, robustness was assessed for feature-based methods. Feature matching registration algorithms are based on the detection of relevant features in every frame of a video sequence and the pairing of these features in consecutive frames. Having a sufficient number of feature matches that conform to a single transformation is essential to obtain a reliable estimation of the relative movement between two images. Therefore, it is generally desirable to use a combination of feature detectors and descriptors that provides large amounts of point matches and a high inlier/outlier ratio.

C. Video Stabilization Algorithm

Based on the results obtained during the comparative analysis (see Sections IV-A and IV-B), the KAZE feature detection and description algorithms [60] were selected for interframe image registration. This section describes how KAZE feature matching was incorporated into a broader algorithm for video stabilization.

First, we added an image preprocessing step to facilitate feature identification. Our approach is based on histogram equalization, whose original implementation was adapted to account for the specificities of fire TIR imagery. In wildfire aerial remote sensing, it is frequent to use high-temperature measurement ranges in order to avoid sensor saturation. This requirement limits the amount of cold details available in the image. Specifically, if the lower end of the temperature measurement range falls above the brightness temperature of the image background, the intensity value of all background pixels is set to the lower limit of the camera measurement range. Consequently, any background detail is lost. This fact has serious implications not only for image registration and georeferencing but also for contrast enhancement. If the amount of foreground pixels is too small in comparison with the size of the background, histogram equalization may fail completely. Such failure occurred in scenarios 4 and 5 of our study.

We worked around this limitation by applying the histogram equalization methodology only to the bounding box of the image foreground, which usually includes a balanced combination of foreground and background pixels. This approach was implemented following three steps: first, the image foreground was identified through pixel intensity segmentation; second, histogram equalization was applied to the minimum rectangular bounding box containing the selected foreground; finally, the processed foreground was merged back into the original background in order to maintain frame size.

In addition to preprocessing, we propose the introduction of a filtering postprocessing step that increases the robustness of interframe registration. Regardless of how reliably image features can be detected and matched, sporadic registration failure is unavoidable. Extreme camera movements, vision occlusion by smoke or objects, and hardware malfunction are some examples of the events that may produce faulty frames. Invalid frames should neither be registered nor used as reference to register other frames.

In order to account for faulty frames and prevent them from affecting the overall stabilization performance, we propose the use of a sliding multireference framework (see Fig. 2). A frequent approach for video stabilization consists of registering each new frame with the latest that has already been stabilized. This approach provides high registration accuracy and is adaptable to changes in image content. However, it is not robust against the presence of faulty frames, especially if these are frequent. Instead, we suggest registering every new frame to the latest five stabilized frames, obtaining the five corresponding transformations that would map the new frame to the reference and using the median of these transformations for registration.

The mathematical representation of 2-D similarity transformations is 3×3 matrices. Therefore, the direct implementation of the proposed approach would consist of computing median values for every element in these matrices. However, this would result in more general transformation types not complying with similarity specifications. In order to keep a meaningful description of the applied registration, we suggest applying median filtering individually to the different movement components. Therefore, our algorithm retrieves the estimated translation, rotation, and scale for all five reference frames, computes their median values, and builds a new registration matrix with the resulting coefficients.

IV. Results and Discussion

A. Registration Accuracy

Figs. 3 and 4 summarize the differences between actual and estimated image misalignment in the form of pseudo-Bland–Altman plots. Bland–Altman plots are a common tool used to compare the accuracy of two methods that were designed to measure the same magnitude. This is accomplished by graphically displaying measurement differences between both methods along the complete range of measured samples. When ground truth sample values are unknown, their value is estimated as the average of measurements provided by both methods. Additionally, bias and limits of agreement are superimposed on the scatter plot. Bias is computed as the average difference, whereas limits of agreement are estimated as bias plus and minus $1.96\times$ the difference standard deviation [66]. Both bias and limits of agreement are accompanied by their respective 95% confidence intervals, which were computed here using the approximated estimations proposed in [67]. Confidence intervals are not always computed in the literature when using Bland–Altman plots, although they have been considered essential by some authors [68]. Because in our case ground truth data are known, Figs. 3 and 4 display pseudo-Bland–Altman plots where horizontal axes contain ground truth values instead of method output average.

Tables II and III summarize mean-squared errors in the estimation of individual motion components together with the global registration quality achieved by each method. These results are in agreement with Figs. 3 and 4 and show that the best accuracy was achieved by feature matching algorithms, especially when using the KAZE detector and descriptor together.

Phase correlation systematically overestimated translations (positive bias in pseudo-Bland–Altman plots, Figs. 3 and 4) and reached errors exceeding 100% of frame dimensions. MI-based registration showed better accuracy, especially for translations, although it was seriously affected by scale and errors remained high in general. Furthermore, both MI optimization and phase correlation tended to fail completely at moderately high rotation angles.

In addition to KAZE, other feature-based algorithms achieved satisfactory performance. SURF feature matching proved especially capable of estimating rotation with high accuracy under idealized working conditions ($\mathrm{MSE}_{\mathrm{rot}} = 0.0040$). Nonetheless, Fig. 3 suggests that SURF successful performance was limited to light rotations. The accuracy of SURF rotation estimation diminished considerably when rotation angles approached $\pm 20^{\circ}$. Similar error behavior was observed for MSER+SURF and MSER+KAZE combinations. Furthermore, these errors followed linear tendencies as observed in rotation plots of Figs. 3 and 4. This fact indicates that absolute errors in rotation estimation coincided with the real rotation value, which suggests that such errors were caused by complete registration failure. Although more subtle, similar failure tendencies were observed in translation and scaling estimation for SURF+SURF, MSER+SURF, and MSER+KAZE combinations. Conversely,

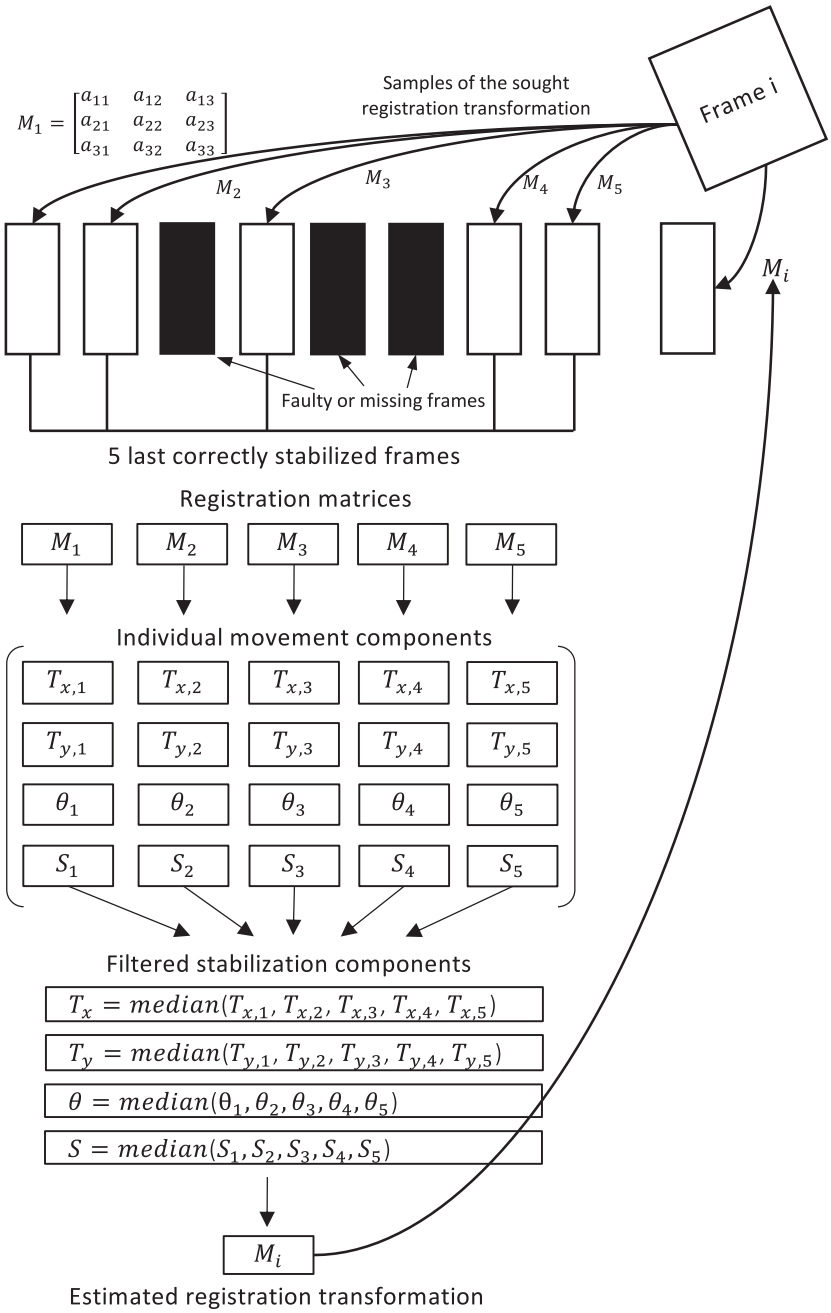

Fig. 2. Workflow of the proposed multireference stabilization algorithm.

KAZE algorithms proved robust in front of all applied perturbations. Finally, SURF suffered a larger decrease in performance under realistic operation conditions than KAZE (see Tables II and III).

Overall, feature matching approaches outperformed phase correlation and MI optimization. Not only did they achieve better registration accuracy but they also ran significantly faster. Among feature-based methods, preliminary results suggested that KAZE was the most accurate and robust. However, the slight performance differences observed using random sampling encouraged a more detailed analysis, which is presented in Section IV-B. Because registration failure was likely due to the lack of feature matches, special attention was paid to method robustness.

B. Feature Matching Robustness

There are two dominant factors that affect the amount of inlier image features effectively used by registration algorithms: image content and the relative position of compared images. We studied both by applying systematic transformations along the complete video duration in sequences 1–6.

Instead of limiting the study to a reduced set of randomly selected frames and perturbations, frames were evenly sampled at 1 Hz for the complete duration of each sequence, which allowed assessing the response of feature detection to image content. In addition, synthetic translations, rotations, and scaling were applied to each sampled frame sequentially and independently to observe the effect of each type of movement. Each

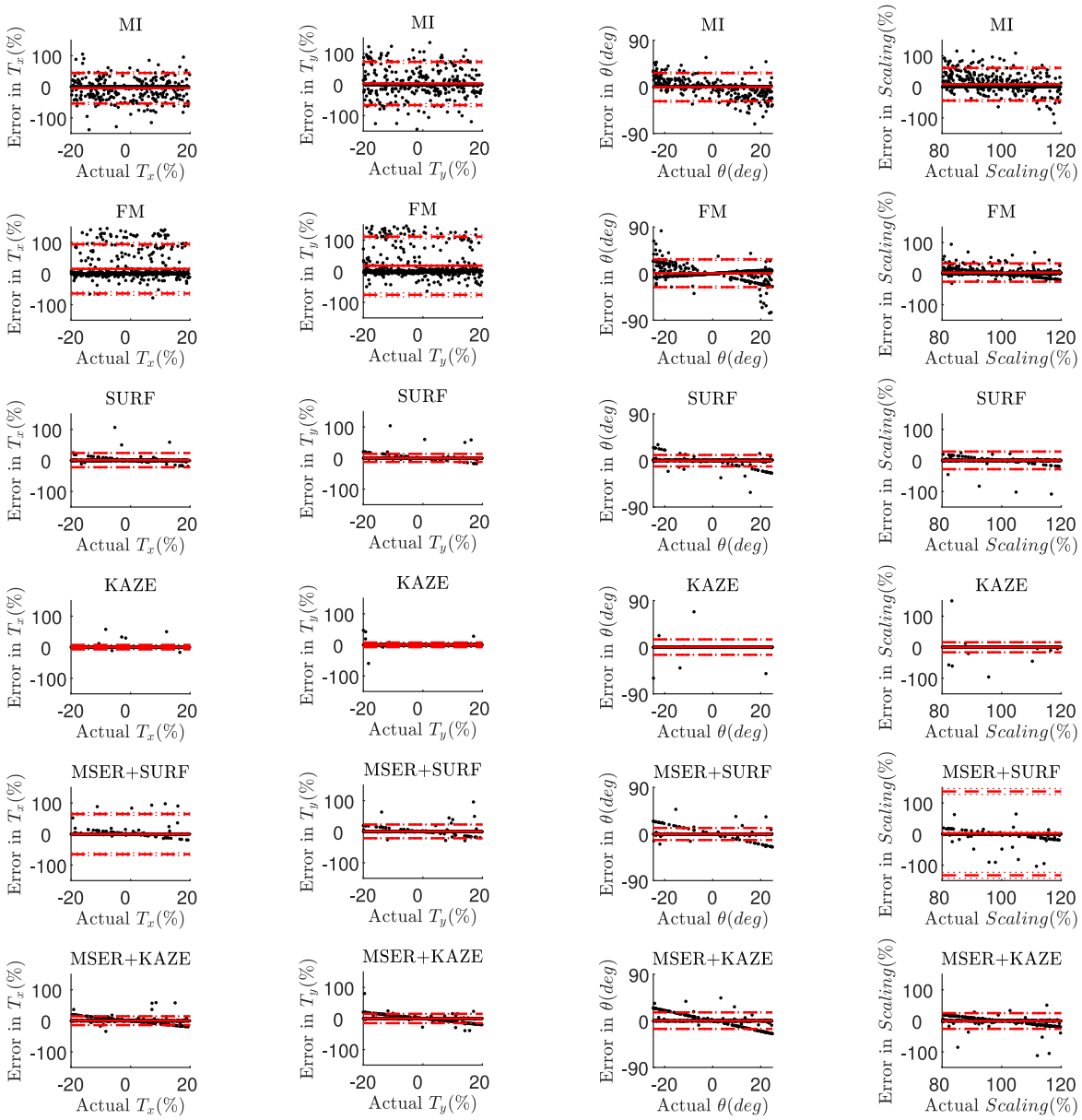

Fig. 3. Pseudo-Bland–Altman plots for motion estimation errors committed by the studied registration methods in idealized working conditions. MI = Mutual information optimization; FM = Fourier–Mellin-based phase correlation. Motion estimation errors are plotted against ground truth, i.e., actual, movement. Camera movement components are: translation in the X direction (T_x), translation in the Y direction (T_y), rotation (θ), and scaling. T_x and T_y were normalized using frame width and height, respectively. Black dots correspond to individual frames randomly sampled along all studied scenarios and randomly perturbed. Red solid lines indicate mean bias. Red dashed lines account for limits of agreement (LoA), whereas red dotted lines represent 95% confidence intervals for estimated bias and LoA.

TABLE II
REGISTRATION ACCURACY ACHIEVED BY TESTED ALGORITHMS WITH SYNTHETIC MOVEMENT UNDER IDEALIZED CONDITIONS

Registration method	MSE in X translation	MSE in Y translation	MSE in Rotation	MSE in Scaling	Mean reg. quality	Avg. computation time (s)
MI	0.0647	0.1272	0.0240	0.0790	0.8451	34.44
Phase-correlation	0.1879	0.2601	0.0231	0.0237	0.6891	2.22
SURF	0.0138	0.0044	0.0040	0.0209	0.9418	0.15
KAZE	0.0014	0.0015	0.0071	0.0068	0.9925	1.15
MSER + SURF	0.1091	0.0128	0.0044	0.4744	0.9741	0.21
MSER + KAZE	0.0053	0.0058	0.0082	0.0166	0.8631	0.28

Mean-squared errors (MSE) in horizontal and vertical translations were normalized using frame width and height, respectively, whereas normalized rotations were referred to 90°. Registration quality was defined as the similarity between the output registered frame and the ground truth target frame, measured by means of 2-D correlation.

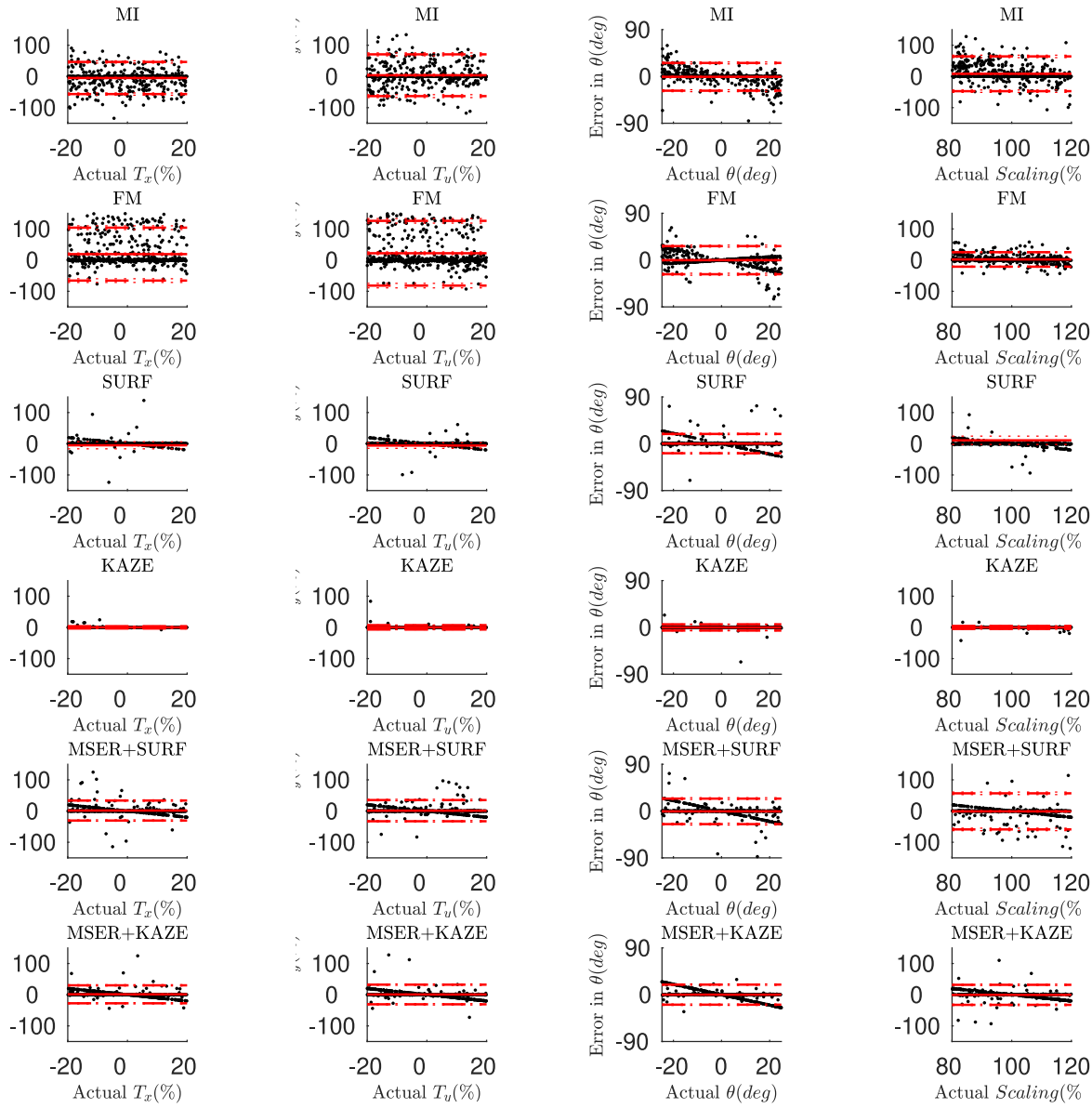

Fig. 4. Pseudo-Bland–Altman plots for motion estimation errors committed by the studied registration methods in realistic working conditions. MI = Mutual information optimization; FM = Fourier–Mellin-based phase correlation. Motion estimation errors are plotted against ground truth, i.e., actual, movement. Camera movement components are: translation in the X direction (T_x), translation in the Y direction (T_y), rotation (θ), and scaling. T_x and T_y were normalized using frame width and height, respectively. Black dots correspond to individual frames randomly sampled along all studied scenarios and randomly perturbed. Red solid lines indicate mean bias. Red dashed lines account for limits of agreement (LoA), whereas red dotted lines represent 95% confidence intervals for estimated bias and LoA.

TABLE III
REGISTRATION ACCURACY ACHIEVED BY TESTED ALGORITHMS WITH SYNTHETIC MOVEMENT UNDER REALISTIC CONDITIONS

Registration method	MSE in X translation	MSE in Y translation	MSE in Rotation	MSE in Scaling	Mean reg. quality	Avg. computation time (s)
MI	0.0714	0.1181	0.0229	0.0880	0.8480	44.38
Phase-correlation	0.2238	0.3287	0.0235	0.0145	0.6809	2.88
SURF	1.9727	1.2217	0.0111	2.7840	0.8600	0.18
KAZE	0.0004	0.0013	0.0012	0.0006	0.9749	1.32
MSER + SURF	0.0270	0.0302	0.0194	0.0872	0.8141	0.19
MSER + KAZE	0.0217	0.0259	0.0120	0.0268	0.7442	0.27

Mean-squared errors (MSE) in horizontal and vertical translations were normalized using frame width and height, respectively, whereas normalized rotations were referred to 90°. Registration quality was defined as the similarity between the output registered frame and the ground truth target frame, measured by means of 2-D correlation.

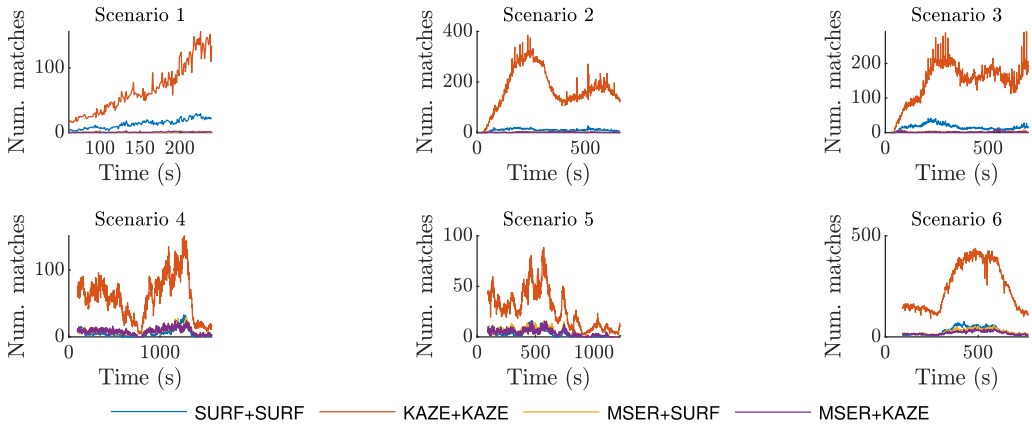

Fig. 5. Number of feature matches used by different combinations of feature detector + feature descriptor when registering each frame with the previous one. Values averaged along all perturbations applied to each frame.

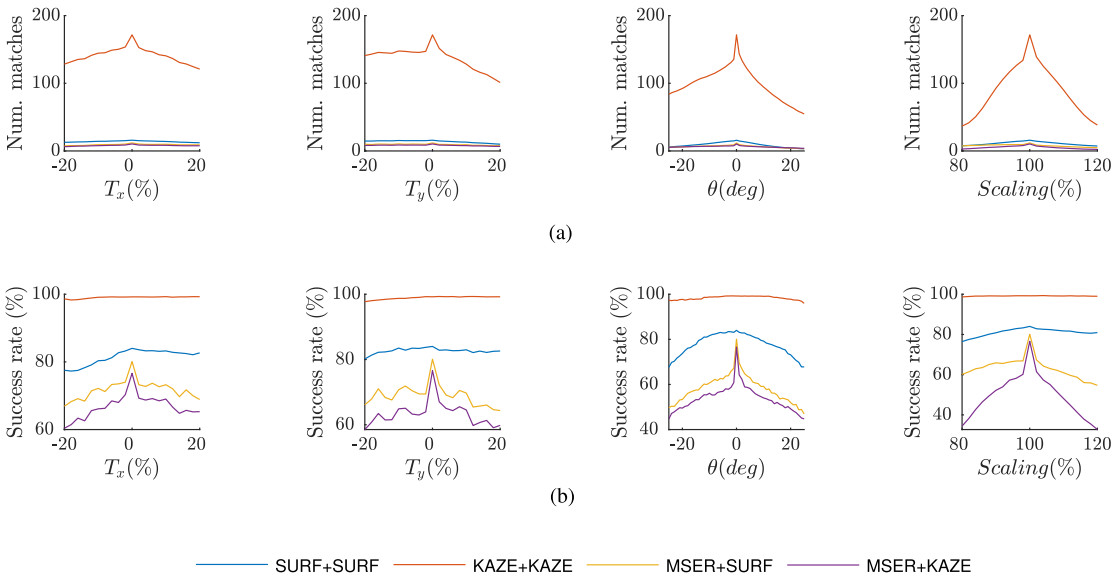

Fig. 6. Robustness of feature-based registration algorithms and its dependence on image alignment. Values averaged along video sequences 1–6 for each combination of feature detector + feature descriptor. (a) Number of feature matches used when registering each frame with the previous one, after outlier rejection. (b) Percentage of frames where registration transformation could be estimated.

distorted frame was then registered back to its original position. In this case, the previous stable frame was used as a registration reference in order to reproduce normal working conditions. Fig. 5 shows the variation in feature matches with image content, whereas Fig. 6(a) displays its dependence on image alignment. The data shown in both of these figures correspond to the number of matches effectively used to estimate registration transformations, i.e., inliers accepted after RANSAC optimization.

In addition to the number of feature matches, there is a second variable representative of registration robustness. If the selected feature detectors and descriptors are not suitable for the type of images at hand, there are occasions when not enough features can be found, they cannot be matched across images or those connections do not follow a unique transformation. In such cases, the algorithm cannot suggest any registration at all. This fact can affect the robustness of the complete video stabilization system, which may be able to handle the existence of missing frames only if they are scarce. We assessed this aspect using the percentage of frames for which each method was able to estimate a registration transformation [see Fig. 6(b)]. Note that Fig. 6(b) only evaluates the capability of providing a transformation estimation and it does not provide information about registration accuracy.

Fig. 5 shows a strong dependence of the number of matches on the portion of the image filled with fire. Footage sections with a higher amount of successful feature matches correspond to times when fire fills a larger portion of the image. Furthermore, the number of feature matches also depends on the alignment between consecutive frames, as displayed in Fig. 6(a). Despite these dependencies, results show a significantly better performance of KAZE features in all studied cases. KAZE's higher robustness is also conveyed by Fig. 6(b), KAZE being the only method capable of maintaining a success rate close to 100% at all translations, rotations, and changes in scale.

This difference can be explained by the feature detection improvements introduced by the KAZE algorithm. Most feature

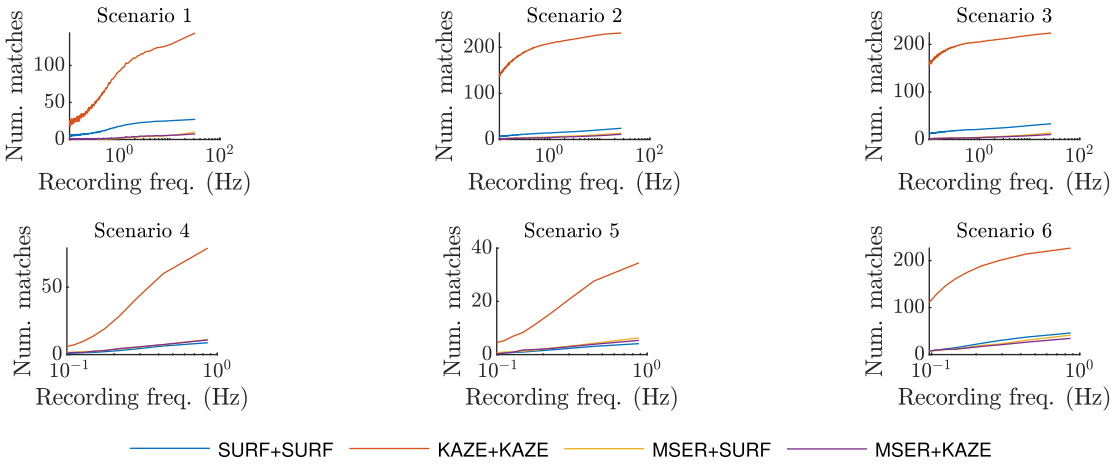

Fig. 7. Number of feature matches available for registration—after outlier rejection—and its variation with video recording frequency. Values averaged along each video sequence. Each frame was registered to the previous frame, with no synthetic perturbations applied.

detectors are based on edges and corners, which are barely visible in TIR imagery and especially in fire video. Additionally, traditional methods such as SURF include a multiscale filtering step in which the original image is convolved with a Gaussian kernel of increasing standard deviation. Although this approach allows reducing noise and detecting salient image features, it does not respect the natural boundaries of objects. In a general scenario, this usually results in localization accuracy reduction. In TIR fire imagery, consequences are worse because salient features are usually scarce and weak. Gaussian blurring can eventually filter them out along with noise, thus losing any opportunity for robust registration.

On the contrary, the KAZE feature descriptor uses a nonlinear scale space based on nonlinear diffusion filtering [60]. This approach implements locally adaptive blurring, thus allowing the reduction of noise while retaining object boundaries. Figs. 5 and 6 demonstrate the importance of this property in TIR fire imagery.

The strong relationship between the number of feature matches and fire size suggests that the majority of the identified features belong to fire. This constitutes an important aspect to consider when analyzing video since fire is highly dynamic and it may change substantially between frames. Among other issues, if video recording frequency is too low, the amount of successful feature matches may diminish dramatically. This hypothesis was examined here and it was corroborated by the results displayed in Fig. 7. However, unacceptable feature match amounts are reached at exceptionally low recording frequencies in the order of 0.1 Hz. Typical nominal recording frequencies in field experiments are in the order of 1 Hz, while popular camera models allow around 20–30 Hz. Results displayed in Fig. 5 and 6 were obtained at a nominal sampling frequency of 1 Hz and this frequency provided acceptable results in all studied cases.

C. Video Stabilization Performance

The complete video stabilization algorithm described in Section III-C was applied to video sequences 1–6 after sampling their frames at a frequency of 1 Hz and perturbing the sampled frames with random similarity transformations. A sampling frequency of 1 Hz was selected as representative of the worst case conditions expectable in wildfire behavior studies. At present, there are no technical limitations that require the use of temporal resolutions below 1 Hz. On the other hand, the algorithm performance is expected to improve monotonically at higher sampling frequencies.

Enough feature matches for registration were found in 92.7% of the total amount of processed frames. Registration accuracy was assessed through the comparison of actual and retrieved image perturbations. Mean-squared errors in normalized translations, rotations, and scaling were below 1% (see Table IV). Fig. 8 shows the registration accuracy achieved for each frame along each video sequence. In addition to measuring absolute performance, Fig. 8 compares the proposed algorithm with the baseline method that does not include image preprocessing or the multireference scheme.

Fig. 8 demonstrates how the introduction of both the multireference framework and the improved histogram equalization contribute toward higher registration accuracy and robustness. In general, image histogram equalization facilitates the detection of relevant image features, which reduces registration errors. On the other hand, registering each frame to multiple reference images avoids discontinuities in the processing workflow when faulty frames appear. The downside of the multireference approach is a slight decrease in registration accuracy, probably caused by the fact of using older reference frames (see results for scenarios 3 and 4 in Fig. 8). Nonetheless, the overall registration quality remained acceptable and the achieved increase in robustness outweighs this limitation.

D. Measurement of Video Stability in a Real Scenario

The performance of video stabilization algorithms is easy to assess in controlled experiments with synthetic camera movement. However, ground truth frame transformations are unknown in a real application scenario. To overcome this limitation, we propose the use of a purely image-based measure of video stability.

TABLE IV
AVERAGE STABILIZATION RESULTS ACHIEVED BY THE PROPOSED ALGORITHM

Sequence	Mean Squared Error			
	X translation	Y translation	Rotation	Scaling
1	4.11e-05	6.66e-05	1.65e-05	2.66e-05
2	9.79e-05	1.31e-04	1.17e-04	1.01e-03
3	2.40e-03	2.48e-03	1.35e-04	2.61e-03
4	6.01e-04	6.63e-03	5.02e-05	2.48e-03
5	3.33e-03	2.95e-03	4.50e-05	1.66e-02
6	1.98e-06	3.38e-06	1.09e-06	3.03e-06

Mean-squared errors (MSE) in horizontal and vertical translations were normalized using frame width and height, resp., whereas normalized rotations were referred to 90°.

TABLE V
VALUES OF THE PROPOSED METRIC FOR VIDEO STABILITY ESTIMATION, COMPUTED AS THE AVERAGE IMAGE SIMILARITY BETWEEN EVERY TWO CONSECUTIVE FRAMES

Scenario	Stability before processing	Stability after processing
1	0.333	0.992
2	0.359	0.991
3	0.251	0.988
4	0.029	0.946
5	0.024	0.954
6	0.266	0.988

TABLE VI
INCREASE IN VIDEO STABILITY PROVIDED BY THE PROPOSED ALGORITHM WHEN APPLIED TO UNSTABLE VIDEO SEQUENCES 7–9

Scenario	Stability before stabilization	Stability after stabilization	Stabilization success ratio
7	0.7009	0.8848	1.0000
8	0.6863	0.9099	0.9986
9	0.6640	0.9197	1.0000

Video stability was estimated as the average 2-D correlation between every two consecutive frames.

Our previous work [45] recommended the use of 2-D correlation to measure the similarity between fire TIR images. Here, we found that the 2-D correlation can also be useful to estimate overall video stability if used to compare pairs of consecutive frames. Table V summarizes video stability indicators computed as the average similarity between every two consecutive frames along each complete sequence. Stability values estimated in this fashion are in agreement with registration quality displayed in Fig. 8, with very good stability achieved in scenarios 1, 2, 3, and 6 and slightly poorer results in scenarios 4 and 5.

The main reason for the lower stability performance achieved in scenarios 4 and 5 is the higher brightness temperature range used to record those video sequences (see Table I). Because of the brightness temperature range used in scenarios 4 and 5, nothing with a brightness temperature below 300 °C is visible in the thermal video. The lower amount of detail hinders feature detection and matching, which results in less robust image registration and, ultimately, poorer video stability.

E. Algorithm Deployment Demonstration

Finally, algorithm deployment was demonstrated in three real scenarios where aerial TIR video of wildfire suffered from significant jitter. These study cases (scenarios 7–9 in Fig. 1) correspond to three field experiments recorded with a TIR camera from a helicopter hovering over the burning plot. The camera was operated manually by the onboard personnel, which added manual jitter to the helicopter movement.

Video sequences 7–9 are of great value to study a number of fire behavior aspects (see, e.g., [11], [37]). However, video instability has so far limited their analysis. In past studies, individual frames were geocorrected manually with the help of GCPs, but the manual approach restricted temporal resolution to 0.1 Hz. The capability of georeferencing every frame in this footage automatically provides an important source of new information relevant for the study of wildfire dynamics.

All three video sequences were successfully stabilized using the proposed algorithm. Additionally, the new stability metric described in Section IV-D was used to assess the result. Table VI summarizes the increase in video stability, estimated through the average interframe 2-D correlation, as well as the portion of frames successfully processed.

A qualitative visualization of the stabilization results for scenarios 7–9 is provided as supplementary materials. The first of these videos shows a virtually perfect stabilization in scenario 7, while scenario 8 demonstrates the system robustness in the presence of faulty video sections. However, scenarios 8 and 9 showcase a limitation of our algorithm: it is unable to cancel projective transformations. Projective transformations appear between consecutive frames when there is a significant change in the camera location. During the experiments in scenarios 8 and 9, the surveillance helicopter moved around the fire to seek a clear view of relevant fire dynamics. Our stabilization algorithm was affected in both cases, producing rotation artifacts in the output video. This constitutes an important limitation that should be addressed in the future.

It should be noted that scenarios 7–9 constitute real study cases of wildfire behavior research during field experiments, not real-scale wildfires. While the successful deployment of our algorithm in these tests constitutes an important contribution to the wildfire science community, additional work is needed to monitor real-scale wildfire behavior. Higher burning intensity and adverse fire weather conditions challenge the deployment of active fire monitoring systems, for example, by preventing aircraft from hovering at a stable position and increasing aerodynamic turbulence. These effects are bound to intensify undesired sensor movement, which will likely produce additional projective transformations in the video sequence. Therefore, correct handling of projective transformations is required to extend this work to real-scale wildfire monitoring.

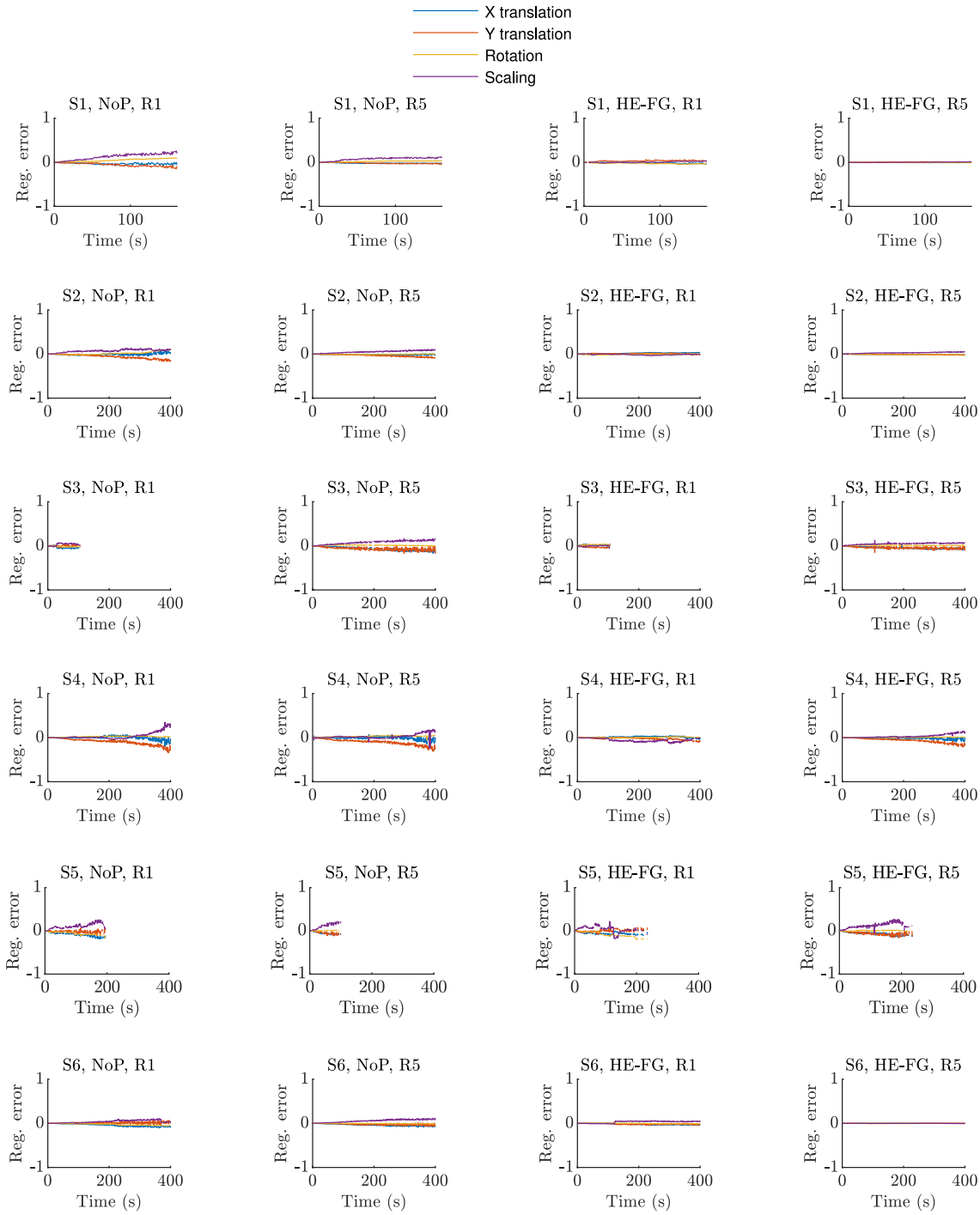

Fig. 8. Errors committed in the estimation of camera movement with the following configurations (from left to right): no preprocessing + single reference, no preprocessing + multiple reference, proposed preprocessing + single reference, proposed preprocessing + multiple reference. S: scenario; NoP: no preprocessing; HE-FG: foreground histogram equalization; R1: one single reference frame; R5: multireference approach using five reference frames. Errors were computed by comparing estimated camera movement with ground truth applied perturbations. Translations were normalized using frame size. Rotations were normalized taking 90° as reference. Each row corresponds to a different video sequence.

V. Conclusion

Aerial TIR imagery can be used to measure a number of wildland fire behavior metrics, such as ROS, FRP, and fireline intensity [3], [8], [9]. However, significant uncertainty has been reported in values estimated that way, mainly due to errors in the delineations of fire front positions and limited accuracy in image registration [27]. Until now, there existed no video stabilization algorithm able to estimate and cancel camera motion in a wildfire scenario, despite the strong need for such a system [23]. This article bridges this gap by proposing a purely software-based

robust and efficient video stabilization algorithm specifically designed for TIR imagery of forest fires.

Our first contribution is a comparative analysis of state-of-the-art image registration algorithms. TIR imagery of wildland fires has a number of specific properties that challenge interframe registration, including low spatial resolution, lack of background detail, and rapid variability of image content. In order to address this problem, we tested a set of popular registration methods pertaining to three different families: maximization of image similarity, phase correlation, and feature matching. In contrast with what one could intuitively expect, feature-based registration methods outperformed alternatives relying on phase correlation and MI maximization. Despite the dynamic nature of fire, KAZE features [60] could be robustly detected and matched, outliers being correctly discarded by a RANSAC algorithm.

Automated video stabilization was accomplished by embedding feature-based interframe registration into a broader algorithm. In particular, we introduced two additional processing modules, developed to meet the specific requirements of TIR fire imagery. First, a modified implementation of histogram equalization was recommended for frame preprocessing. Second, a sliding multireference registration framework improved the system robustness in the presence of faulty or missing frames. The performance of the proposed system was validated in six scenarios with synthetic video jitter and demonstrated in three more aerial video sequences with real camera movement.

Finally, we propose a new metric to measure IR video stability under real operation conditions, where ground truth camera movement is unknown. In such cases, video stability may be estimated through interframe image similarity, measured using 2-D correlation. The application of this new metric was demonstrated in the three unstable video sequences recorded from a helicopter (scenarios 7–9), which were stabilized using the algorithm described in this article.

The ability to register every frame in a TIR video sequence to a common reference is bound to support wildfire behavior studies. Video instability has so far restricted the analysis to a subset of frames processed manually. Previous studies provided results with a temporal resolution of 10^{-1} Hz, whereas the IR video can typically be recorded at a frame rate of 20–50 Hz. The automated processing of every frame will increase the amount of information available for fire behavior analyses.

Furthermore, our multireference registration approach facilitates uncertainty quantification in output fire behavior metrics, which has also been demanded by previous authors [27]. The capability of obtaining multiple samples of the unknown transformation allows estimation of the most probable value as well as quantification of the confidence associated with the applied transformation. Such uncertainty in the location of measured data can then be combined with other sensor uncertainties and propagated through the rest of processing algorithms. Similarly, the increased availability of georeferenced data will enable the computation of probability distributions for fire behavior descriptors, complementing averaged values with dispersion estimations and confidence intervals.

As future work, the described algorithm should be extended to more general applications, including the correct estimation of projective transformations and the analysis of TIR imagery acquired by fixed-wing aircraft.

ACKNOWLEDGMENT

The authors would like to thank V. Hoff for his help with data collection.

REFERENCES

[1] P. J. Riggan, R. G. Tissell, and J. W. Hoffman, "Application of the firemappertm thermal-imaging radiometer for wildfire suppression," in *Proc. IEEE Aerosp. Conf. Proc.*, 2003, vol. 4, pp. 1863–1872.

[2] E. Pastor, A. Àgueda, J. Andrade-Cetto, M. Muñoz, Y. Pérez, and E. Planas, "Computing the rate of spread of linear flame fronts by thermal image processing," *Fire Saf. J.*, vol. 41, no. 8, pp. 569–579, Nov. 2006.

[3] R. Paugam, M. J. Wooster, and G. Roberts, "Use of handheld thermal imager data for airborne mapping of fire radiative power and energy and flame front rate of spread," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3385–3399, Jun. 2013.

[4] F. Manzano-Agugliaro, J. Pérez-Aranda, and J. L. De LaCruz, "Methodology to obtain isochrones from large wildfires," *Int. J. Wildland Fire*, vol. 23, pp. 338–349, 2014.

[5] D. A. Stow, P. J. Riggan, E. J. Storey, and L. L. Coulter, "Measuring fire spread rates from repeat pass airborne thermal infrared imagery," *Remote Sens. Lett.*, vol. 5, no. 9, pp. 803–812, Oct. 2014.

[6] M. B. Dickinson *et al.*, "Measuring radiant emissions from entire prescribed fires with ground, airborne and satellite sensors—RxCADRE2012," *Int. J. Wildland Fire*, vol. 25, no. 1, pp. 48–61, 2016.

[7] T. J. Zajkowski *et al.*, "Evaluation and use of remotely piloted aircraft systems for operations and research—RxCADRE2012," *Int. J. Wildland Fire*, vol. 25, no. 1, pp. 114–128, 2016.

[8] J. M. Johnston, M. J. Wooster, R. Paugam, X. Wang, T. J. Lynham, and L. M. Johnston, "Direct estimation of Byram's fire intensity from infrared remote sensing imagery," *Int. J. Wildland Fire*, vol. 26, pp. 668–684, 2017.

[9] M. M. Valero, O. Rios, E. Pastor, and E. Planas, "Automated location of active fire perimeters in aerial infrared imaging using unsupervised edge detectors," *Int. J. Wildland Fire*, vol. 27, no. 4, pp. 241–256, 2018.

[10] Y. Pérez, E. Pastor, E. Planas, M. Plucinski, and J. Gould, "Computing forest fires aerial suppression effectiveness by IR monitoring," *Fire Saf. J.*, vol. 46, no. 1–2, pp. 2–8, Jan. 2011.

[11] M. Plucinski and E. Pastor, "Criteria and methodology for evaluating aerial wildfire suppression," *Int. J. Wildland Fire*, vol. 22, no. 8, pp. 1144–1154, 2013.

[12] B. Butler *et al.*, "Observations of energy transport and rate of spreads from low-intensity fires in longleaf pine habitat—RxCADRE2012," *Int. J. Wildland Fire*, vol. 25, no. 1, pp. 76–89, 2016.

[13] M. C. Rochoux, S. Ricci, D. Lucor, B. Cuenot, and A. Trouvé, "Towards predictive data-driven simulations of wildfire spread—Part I: Reduced-cost ensemble Kalman filter based on a polynomial chaos surrogate model for parameter estimation," *Natural Hazards Earth Syst. Sci.*, vol. 14, no. 11, pp. 2951–2973, 2014.

[14] M. C. Rochoux, C. Emery, S. Ricci, B. Cuenot, and A. Trouvé, "Towards predictive data-driven simulations of wildfire spread—Part II: Ensemble Kalman filter for the state estimation of a front-tracking simulator of wildfire spread," *Nat. Hazards Earth Syst. Sci.*, vol. 15, no. 11, pp. 1721–1739, 2015.

[15] O. Rios, E. Pastor, M. M. Valero, and E. Planas, "Short-term fire front spread prediction using inverse modelling and airborne infrared images," *Int. J. Wildland Fire*, vol. 25, no. 10, pp. 1033–1047, 2016.

[16] M. Valero, O. Rios, C. Mata, E. Pastor, and E. Planas, "An integrated approach for automated tactical monitoring and data-driven spread forecasting of wildfires," *Fire Saf. J.*, vol. 91, pp. 835–844, 2017.

[17] O. Rios, M. M. Valero, E. Pastor, and E. Planas, "A data-driven fire spread simulator: Validation in Vall-Llobrega's fire," *Front. Mech. Eng.*, vol. 5, pp. 1–11, 2019.

[18] V. Ambrosia, J. Myers, and E. Hildum, "NASA's autonomous modular scanner (AMS)—Wildfire sensor: Improving wildfire observations from airborne platforms," in *Proc. 34th Int. Symp. Remote Sens. Environ. - The GEOSS Era: Towards Oper. Environ. Monit.*, 2011, pp. 1–4.

[19] V. Ambrosia *et al.*, "The Ikhana unmanned airborne system (UAS) western states fire imaging missions: From concept to reality (2006–2010)," *Geocarto Int.*, vol. 26, no. 2, pp. 85–101, 2011.

[20] R. Allison, J. Johnston, G. Craig, and S. Jennings, "Airborne optical and thermal remote sensing for wildfire detection and monitoring," *Sensors*, vol. 16, no. 8, 2016, Art. no. 1310.

[21] L. Merino, F. Caballero, J. R. Martínez-de Dios, I. Maza, and A. Ollero, "An unmanned aircraft system for automatic forest fire monitoring and measurement," *J. Intell. Robot. Syst.*, vol. 65, no. 1/4, pp. 533–548, Aug. 2011.

[22] V. Ciullo, L. Rossi, T. Toulouse, and A. Pieri, "Fire geometrical characteristics estimation using a visible stereovision system carried by unmanned aerial vehicle," in *Proc. 15th Int. Conf. Control, Automat., Robot. Vis.*, 2018, pp. 1216–1221.

[23] C. Yuan, Y. Zhang, and Z. Liu, "A survey on technologies for automatic forest fire monitoring, detection, and fighting using unmanned aerial vehicles and remote sensing techniques," *Can. J. Forest Res.*, vol. 45, no. 7, pp. 783–792, 2015.

[24] J. J. O'Brien *et al.*, "High-resolution infrared thermography for capturing wildland fire behaviour: RxCADRE2012," *Int. J. Wildland Fire*, vol. 25, no. 1, pp. 62–75, 2016.

[25] M. M. Valero *et al.*, "On the use of compact thermal cameras for quantitative wildfire monitoring," *in Advances in Forest Fire Research* 2018, D. X. Viegas, Ed. Coimbra, Portugal: Univ. Coimbra Press, 2018, ch. 5, pp. 1077–1086.

[26] D. Frankman, B. W. Webb, B. W. Butler, D. Jimenez, and M. Harrington, "The effect of sampling rate on interpretation of the temporal characteristics of radiative and convective heating in wildland flames," *Int. J. Wildland Fire*, vol. 22, no. 2, pp. 168–173, 2012.

[27] D. Stow *et al.*, "Assessing uncertainty and demonstrating potential for estimating fire rate of spread at landscape scales based on time sequential airborne thermal infrared imaging," *Int. J. Remote Sens.*, vol. 40, no. 13, pp. 4876–4897, 2019.

[28] J. Trelles and P. Pagni, "Fire-induced winds in the 20 October 1991 Oakland Hills fire," *Fire Saf. Sci.*, vol. 5, pp. 911–922, 1997.

[29] A. Sullivan and I. Knight, "Estimating error in wind speed measurements for experimental fires," *Can. J. Forest Res.*, vol. 31, no. 3, pp. 401–409, 2001.

[30] M. G. Cruz and M. E. Alexander, "Uncertainty associated with model predictions of surface and crown fire rates of spread," *Environ. Model. Softw.*, vol. 47, pp. 16–28, 2013.

[31] C. B. Clements and D. Seto, "Observations of fire-atmosphere interactions and near-surface heat transport on a slope," *Boundary-Layer Meteorol.*, vol. 154, no. 3, pp. 409–426, 2015.

[32] C. B. Clements *et al.*, "Fire weather conditions and fire-atmosphere interactions observed during low-intensity prescribed fires—RxCADRE2012," *Int. J. Wildland Fire*, vol. 25, no. 1, pp. 90–101, 2016.

[33] F. Pimont, J. L. Dupuy, R. R. Linn, R. Parsons, and N. Martin-StPaul, "Representativeness of wind measurements in fire experiments: Lessons learned from large-eddy simulations in a homogeneous forest," *Agricultural Forest Meteorol.*, vol. 232, pp. 479–488, 2017.

[34] G. Zhou, "Near real-time orthorectification and mosaic of small UAV video flow for time-critical event response," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 739–747, Mar. 2009.

[35] J. R. Martínez-De-Dios and A. Ollero, *Automatic Stabilization of Infrared Images Using Frequency Domain Methods*, F. Kongoli, Ed. London, U.K.: InTech, 2012, pp. 435–452.

[36] R. D. Ottmar *et al.*, "Measurements, datasets and preliminary results from the RxCADRE project-2008, 2011 and 2012," *Int. J. Wildland Fire*, vol. 25, pp. 1–9, 2016.

[37] M. Cruz, S. Matthews, J. Gould, and P. Ellis, "Fire dynamics in malleeheath: Fuel, weather and fire behaviour prediction in south Australian semi-arid shrublands," Bushfire Cooperative Res. Centre, Melbourne, VIC, Australia, Tech. Rep. A.10.01, 2010.

[38] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, no. 4, pp. 325–376, 1992.

[39] B. Zitová and J. Flusser, "Image registration methods: A Survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, 2003.

[40] L. M. G. Fonseca and B. S. Manjunath, "Registration techniques for multisensor remotely sensed imagery," *Photogram. Eng. Remote Sens.*, vol. 62, no. 9, pp. 1049–1056, 1996.

[41] H. M. Chen, P. K. Varshney, and M. K. Arora, "Performance of mutual information similarity measure for registration of multitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 11, pp. 2445–2454, Nov. 2003.

[42] Y. Bentoutou, N. Taleb, K. Kpalma, and J. Ronsin, "An automatic image registration for applications in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 9, pp. 2127–2137, Sep. 2005.

[43] J. P. Kern and M. S. Pattichis, "Robust multispectral image registration using mutual-information models," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1494–1505, May 2007.

[44] A. Wong and D. A. Clausi, "ARRSI: Automatic registration of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1483–1492, May 2007.

[45] M. M. Valero *et al.*, "Image similarity metrics suitable for infrared video stabilization during active wildfire monitoring: A comparative analysis," *Remote Sens.*, vol. 12, no. 3, pp. 540–563, 2020.

[46] K. Johnson, A. Cole-Rhodes, I. Zavorin, and J. Le Moigne, "Mutual information as a similarity measure for remote sensing image registration," *Proc. SPIE*, *Geo-Spatial Image Data Exploitation II*, vol. 4383, pp. 51–61, 2001.

[47] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Image registration by maximization of combined mutual information and gradient information," *IEEE Trans. Med. Imag.*, vol. 19, no. 8, pp. 809–814, Aug. 2000.

[48] Y. Zhuang, K. Gao, X. Miu, L. Han, and X. Gong, "Infrared and visual image registration based on mutual information with a combined particle swarm optimization—Powell search algorithm," *Optik*, vol. 127, no. 1, pp. 188–191, 2016.

[49] C. Kuglin and D. Hines, "The phase correlation image alignment method," in *Proc. IEEE Int. Conf. Cybern. Soc.*, 1975, pp. 163–165.

[50] B. R. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. Image Process.*, vol. 5, no. 8, pp. 1266–1271, Aug. 1996.

[51] J. Martinez-de Dios and A. Ollero, "A real-time image stabilization system based on Fourier-Mellin transform," in *Proc. Image Anal. Recognit.*, Oct. 2004, pp. 376–383.

[52] R. Goecke, A. Asthana, N. Pettersson, and L. Petersson, "Visual vehicle egomotion estimation using the Fourier-Mellin transform," in *Proc. IEEE Intell. Veh. Symp.*, 2007, pp. 450–455.

[53] S. Kumar, H. Azartash, M. Biswas, and T. Nguyen, "Real-time affine global motion estimation using phase correlation and its application for digital image stabilization," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3406–3418, Dec. 2011.

[54] L. Lai and Z. Xu, "Global motion estimation based on Fourier Mellin and phase correlation," in *Proc. Int. Conf. Civil, Mater. Environ. Sci.*, 2015, pp. 636–639.

[55] G. Rabatel and S. Labbé, "Registration of visible and near infrared unmanned aerial vehicle images based on Fourier-Mellin transform," *Precis. Agriculture*, vol. 17, no. 5, pp. 564–587, 2016.

[56] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[57] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[58] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *Proc. Brit. Mach. Vis. Conf.*, 2002, pp. 384–393.

[59] D. Nistér and H. Stewénius, "Linear time maximally stable extremal regions," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 183–196.

[60] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "KAZE features," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 214–227.

[61] P. H. S. Torr and A. Zisserman, "MLESAC: A new robust estimator with application to estimating image geometry," *Comput. Vis. Image Understanding*, vol. 78, no. 1, pp. 138–156, 2000.

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

[63] H. Wang and D. Suter, "Robust adaptive-scale parametric model estimation for computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1459–1474, Nov. 2004.

[64] B. J. Tordoff and D. W. Murray, "Guided-MLESAC: Faster image transform estimation by using matching priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1523–1535, Oct. 2005.

[65] H. Li, W. Ding, X. Cao, and C. Liu, "Image registration and fusion of visible and infrared integrated camera for medium-altitude unmanned aerial vehicle remote sensing," *Remote Sens.*, vol. 9, no. 5, pp. 441–469, 2017.

[66] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet*, vol. 327, no. 8476, pp. 307–310, 1986.

[67] J. M. Bland and D. G. Altman, "Measuring agreement in method comparison studies," *Stat. Methods Med. Res.*, vol. 8, no. 2, pp. 135–160, 1999.

[68] A. Carkeet, "Exact parametric confidence intervals for Bland-Altman limits of agreement," *Optometry Vis. Sci.*, vol. 92, no. 3, pp. 71–80, 2015.

Mario Miguel Valero received the M.Eng. degree in biomedical engineering from the Universidad Politécnica de Madrid, in 2014 and the Ph.D. degree in remote sensing from the Universitat Politècnica de Catalunya Spain, in 2019.

He is an Aerospace Engineer. After two years of experience as a Data Scientist in the private industry, he is currently an Assistant Professor with San Jose State University, San Jose, CA, USA. His research interests include experimental and computational fluid dynamics, environmental modeling, remote sensing, computer vision, machine learning, and data assimilation. His work focuses on unresolved societal challenges and safety and environmental sustainability.

Steven Verstockt received the master's degree in informatics from Ghent University, Ghent, Belgium, in 2003. In 2008, he began working on the Ph.D. degree on multimodal video fire analysis with the Multimedia Lab of the Electronics and Information Systems Department, Ghent University-imec.

Since 2012, he has been a Postdoctoral Researcher with Ghent University-imec. In September 2015, he started a tenure track professorship in multimedia. His current research focuses on multimodal data processing, video analysis, computer vision, and GEO-ICT.

Bret Butler received the Ph.D. degree in mechanical engineering, in 1992 from Brigham Young University, Provo, Utah, USA, where he studied energy transport in particle-laden flames.

Since 1992, he has been a Research Mechanical Engineer with the USDA Forest Service, Washington, DC, USA. His research focuses on fundamental heat and combustion processes in wildland fire. Applications for his research include fire behavior models, links between fire behavior and effects, and firefighter safety. He is the Lead Scientist for the quantitative safety zone guidelines used throughout the world. He is also the Lead Scientist for the high-resolution surface wind models that have been developed by the U.S. Forest Service.

Daniel Jimenez received the B.S. degree in chemical engineering from the Rutgers University, in 1994 and M.S. degree in civil/environmental engineering from the Washington State University, in 2003.

He is a Chemical Engineer and has been with Missoula Fire Science Lab since 1998. During that time, he has worked primarily within the fire behavior discipline area and his research has included studies on heat transfer and fluid flow model development and validation. This research is used to predict incident heat, temperature, wind profiles, and fire effects on forest landscapes, vegetation, and soil types as well as play a critical role in firefighter safety. In collaboration with scientists at the Fire Lab and elsewhere, he has worked on the development and delivery of computer software to support fire and fuel managers and has been instrumental in developing new firefighter safety guidelines.

Oriol Rios received the M.Sc. degree in fire safety science from Gent(BE), Ghent, Belgium, Lund (SE), Lund, Sweden, and Edinburgh University, Edinburgh, U.K. in 2013, and received the Ph.D. degree from the Polytechnic University of Catalunya, Barcelona, Spain, in 2018, in fire modeling and monitoring.

In 2017, he joined the Fire Safety Engineering Team at CERN (European Organization for Nuclear Research) to work on performance-based design and exploit fire safety solutions throughout CERN underground accelerators and detectors.

Christian Mata received the Erasmus-Mundus master's degree in computer vision and robotics from Heriot-Watt University, Scotland, U.K., Université de Bourgogne, Dijon, France, and Universitat de Girona, Girona, Spain, in 2009, and the Ph.D. degree in computer vision and medical image processing from ViCOROB Research Group, Universitat de Girona, and LE2I, Université de Bourgogne, in 2015.

Since 2016, he has been a Postdoctoral Researcher with CERTEC Research Group, Universitat Politècnica de Catalunya. His current research focuses on computer vision and image processing projects.

LLoyd Queen received the Ph.D. degree in remote sensing from the University of Nebraska, Lincoln, NE, USA, in 1988.

He is a Professor of remote sensing with the Department of Forest Management, University of Montana, Missoula, MT, USA. He teaches courses, workshops, and field courses in remote sensing, GIS, and fire management. Since May 2001, he has been the Director of the National Center for Landscape Fire Analysis and is the former Chair of the Forest Management Department. He was a Presidential Fellow at the University of Nebraska and was a Postdoctoral Scientist at the University of Minnesota before coming to the University of Montana, in 1995.

Elsa Pastor received the Ph.D. degree in industrial engineering from the Universitat Politcnica de Catalunya Spain, in 2004.

She is an Associate Professor with the Chemical Engineering Department, Universitat Politècnica de Catalunya· BarcelonaTech, Barcelona, Spain, and a Research Scientist with the Center for Technological Risk Studies. She develops teaching and research activities in diverse fields related to wildfire management and technological risk analysis. Over the last 15 years, she has studied several aspects of fire behavior and dynamics by a multidisciplinary approach, combining both experimental and modeling techniques in a wide range of scenarios. She has profited from diverse fire environments (i.e., wildfires, wildfire research burning campaigns, outdoor large-scale industrial testing fields, compartment fires, laboratory setups, etc.) to observe, monitor, and analyze flames and their effect to different types of assets and ecosystems. She is currently leading the European Project WUIVIEW (wuiview.org), aimed at designing, setting-up, and operating a virtual workbench service for the analysis of fire risk in the surroundings of buildings at the wildland-urban interface.

Eulàlia Planas received the Ph.D. degree in industrial engineering from the Universitat Politècnica de Catalunya Spain, in 1996.

She is an Associate Professor with the Chemical Engineering Department, Universitat Politècnica de Catalunya, Barcelona, Spain. She is also the Head of the Centre for Technological Risk Studies. Her main research lines are the study of hydrocarbon pool-fires; the mathematical modeling of major accidents; risk analysis in the transportation of hazardous materials; and the study of wildfires. In the field of wildfire research, she has developed infrared image processing systems to quantify fire progression (rate of spread, fire intensity, and flame geometry) and aerial fire attack effectiveness. She has also worked on providing systems to deliver fire behavior forecasts for decision-making, based on data assimilation and inverse modeling. She also developed methodologies based on CFD modeling to study the effects of burning residential fuels on structures, relying on performance-based criteria to assess houses vulnerability and sheltering capacity. She also got involved extensively on experimental fire research.