Non-stationary VFD Evaluation Kit: Dataset and Metrics to Fuel Video-Based Fire Detection Development

Cristiano Rafael Steffens(✉), Ricardo Nagel Rodrigues, and Silvia Silva da Costa Botelho

Center of Computer Science, Federal University of Rio Grande, Avenida Itália – Km. 8, Rio Grande, Brazil
{cristianosteffens,ricardonagel,silviacb}@furg.br
http://c3.furg.br

Abstract. Datasets play a major role in the advance of computer vision techniques nowadays. Open, complete and challenging ground truth data, combined with standardized metrics are essential to push the development and allow the proper evaluation of computer vision algorithms. Even though a significant amount of work on VFD (video-based fire detection) systems has been developed, compare different algorithms is a laborious task due to the lack of common evaluation schemes and evaluation datasets. We address both of these issues by presenting a dataset of fire videos along with frame by frame annotations to be used for non-stationary fire detection algorithms training and validation. By the time, this is the largest dataset released on this subject matter. Standard video file formats and open markup languages where used to allow compatibility and convenient integration with the most popular computer vision libraries. The dataset includes hand-held, robot attached and drone attached footages and aims to boost the development of fully autonomous firefighter robots. The presented ground truth and metrics adapt to the majority of the state-of-the-art techniques and provides a reliable and unbiased solution to compare them. The dataset, example source-code and documentation are publicly available under the Creative Commons 3.0 license on GitHub.

Keywords: Dataset · Database · Evaluation · Validation · Fire

1 Introduction

The video-based fire detection has been a research topic longer than two decades now. The first research in this area dates back to 1993, when Healey *et al.* [24] presented a real-time system for automatic fire detection using color video input from stationary cameras. The physical properties of fire are considered for the

C.R. Steffens—Acknowledges the financial support of CNPq-Brazil, Finep and CAPES.

F. Santos Osório and R. Sales Gonçalves (Eds.): LARS 2015/SBR 2015, CCIS 619, pp. 135–151, 2016.
DOI: 10.1007/978-3-319-47247-8_9

development of algorithms that explore the spectral, spatial, and temporal properties of fire events. Latter, in 1996 Foo [16] proposed a knowledge based system based on heuristics of statistical measures such as mean, median, standard deviation, and first-order moment derived from the histogram and image subtraction analyses of successive image frames. Also in 1996, Plumb and Richards [36] present a prototype VFD system capable of determining the location and heat release rate of the fire employing transient temperatures using temperature-sensitive, color-changing sensors.

As research kept improving the results of VFD systems and video acquisition hardware became ubiquitous, some techniques have made their way to become commercial products. Only to name a few, we can list SIGNIFIRE video flame, smoke and intrusion detection system [15], AlarmEye AE3000 PC Based Video Fire Detection System [25], and FireVu Video Smoke Detection [1], which promise early and effective detection. Stipaničev *et al.* [40] lists some commercial VFDs for forest fire detection, showing their strengths and weaknesses, and presents IPNAS, a complete terrestrial tower based structure featuring cameras, wireless communication and software for forest fire surveillance. Despite the maturity that these solutions have reached, while all vendors advocate towards the capabilities of their VFD systems, it remains difficult to compare them.

On academic research, ideally, algorithms should be published with sufficient details to replicate the claimed results or, at least, with an executable binary and datasets. In this sense, the present work aims to offer a public test database and common evaluation metrics, allowing researchers to test and establish clear comparisons and consequently better benchmarks. Proprietary ground truth data is a barrier to independent evaluation of metrics and algorithms. The interested parties should be able to duplicate the metrics produced by various types of algorithms, validating them against the ground truth data and so comparing the results.

Current publications in the Video-based fire detection field can be grouped in two main categories according to their input data: stationary – where the camera is fixed to a tower or building – or non-stationary – where the camera is carried by a person or any moving equipment such as robots, cars or drones. The majority of the current research focuses on stationary systems, which are, usually, aimed to forest and outdoor surveillance. Usually, the proposed solutions are tested using a set of videos provided by the Bilkent VisiFire[1] Sample Video Clips as is the case for Celik *et al.* [4], Toreyin *et al.* [42–45], Habiboglu *et al.* [20] or by KMU Fire & Smoke[2] database, as is the case for Park *et al.* [35], Kwak *et al.* [28] and Shidik *et al.* [38]. More recently, Gunay *et al.* [19], Jin *et al.* [26] and Kong *et al.* [27] used the ICV database[3]. Other recent research such as Labati *et al.* [29] does not mention nor provide further access to the dataset. All the aforementioned datasets are incomplete, in the sense that they do not provide annotations for a standardized evaluation. Once the mentioned datasets

[1] Available at http://signal.ee.bilkent.edu.tr/VisiFire/Demo/SampleClips.html.
[2] Available at http://cvpr.kmu.ac.kr/Dataset/Dataset.htm.
[3] Available at http://vision.inha.ac.kr. Password protected.

lack essential information, such as the amount of negative and positive samples, it is impossible to compute the accuracy, specificity, fall-out and false negative rate among other relevant statistics.

Non-stationary VFDs are more recent and therefore there are only a few published researches targeting this problem. As far as we know, there is no publicly available fire detection dataset for tests on non-stationary videos. For instance, the dataset used in Borges *et al.* [2] is not available under the provided link. In personal contact the authors claimed the project was executed with private partners and neither code nor datasets where release. On the other hand, Chenebert *et al.* [9] do not provide any further information about the test data that has been used.

In fact, the lack of a standardization for evaluating the output of detection algorithms results in a situation where the reported performance results are strongly dependant on the definition of what is a correct detection. Since the evaluation process is not standardized, a comparison between two different algorithms remains difficult. Back in 2015, Tolouse *et al.* [46] presented a first attempt to bring some common metrics proposing a dataset and framework for benchmarking wildland fire segmentation algorithms. The dataset, which is composed of 100 RGB images acquired from the internet and specialized researchers, contains images of wildland (outdoor vegetation) fire in different contexts, such as fuel, background, luminosity, and smoke. All images of the dataset are characterised according to the principal colour of the fire, the luminosity, and the presence of smoke in the fire area. With this characterisation, the authors claim it to be possible to determine on which kind of images each algorithm is efficient. Although the authors do not provide a downloadable file, the dataset can be used via Octave/Matlab code trough their site[4].

The first contribution of this work, which extends a paper previously published by Steffens *et al.* [39], is the creation of a new non-stationary dataset composed by 20 annotated videos, featuring a wide range of difficulties including occlusion, different scales, camera vibration and a variety of bright and contrast conditions. As a second contribution we address the evaluation problem by presenting a new evaluation scheme composed by the following elements:

– An algorithm to find correspondences between a fire detector output and the annotated fire regions;
– Two separated rigorous and precise methods for evaluating any algorithm's performance on the proposed dataset. These two methods are intended for two different applications: fire location and frame-by-frame classification;
– C++ source code that implements these procedures.

The publicly available dataset and evaluation scheme proposed provide an straightforward manner to compare the performance of different algorithms. With an easy to use software library, it allows to develop machine learning based approaches once the training data can be easily generated. In this sense, researchers can focus solely on their methods, which will further prompt

[4] Available at http://firetest.cs.wits.ac.za/benchmark.

researchers to work on more difficult versions of the mobile robot based fire detection problem.

2 Dataset

In order to propose a new dataset and evaluation metrics the first step is to study how researchers tested their prior work. The non-stationary video based fire detection was first introduced by Borges and Izquierdo [2], which presents a method that analyzes the frame-to-frame changes of specific low-level features describing potential fire regions in order to classify newscast videos as containing or not hydrocarbon flames. The algorithm is structured in two main steps: color based classification and texture based classification. A statistical based equation is proposed for the color classification. In a second step color, area size, surface coarseness, boundary roughness and skewness of the R channel within the estimated fire regions are used as descriptors. The final classification is given by a Bayes classifier. Borges and Izquierdo [2] evaluate their detector by checking the classification of an entire video, disregarding the detection time and the location on the frame.

In 2011, Chenebert et al. [9] proposed a non-temporal texture driven approach for fire detection, combining a color-threshold equation previously proposed in [8]. The main idea of this work was to use supervised learning classifiers such as neural networks and regression trees. As texture descriptors they used ten bins histograms on the hue and saturation channels from the HSV color space. The authors also propose the use of the Gray-level Covariance Matrix [23], energy, entropy, contrast, homogeneity and correlation to find these parameters. Based on the descriptors, the authors claim that they were able to process 12 frames per second with an average precision close to 87.83 % using classification trees. The results are given considering exclusively if each frame was classified as fire or non-fire.

Given the wide range of possible applications for the proposed dataset, we release it as an unique wide pack that developers can use in different ways. In machine learning based fire detection techniques researchers may break it in training, test and evaluation sets while other approaches may possibly use the whole dataset only for evaluation. Therefore, we provide functions that allow software developers to access a specific frame and its associated annotations, making it easy to implement sample splitting techniques such as cross-validation or bootstrapping. Another important property of the state of the art detectors that leads us to release the dataset as an unique package is that some techniques, such as the ones proposed in [7,8,33,34], are dependant on a video sequence, using the flickering frequency and optical flow in order to determine the position and location of the fire region.

2.1 Video Properties

The dataset was created using 28022 frames distributed in 24 videos and published on the internet under the Creative Commons 3.0 license. It features a total of 14397 fire frames, which stand for 51.37 %, while the amount of 13625 non-fire frames represents the remaining 48.62 %. Considering that some fire frames present more than one annotated fire region, the dataset is composed by 17917 annotations. In order to provide a complete and challenging non-stationary video set, enabling a true evaluation of the detection algorithms, the files present the following properties:

– Variety of fire sources: different liquid and solid fuels produce different flames. The fuel and oxidizer may have a negligible influence on the flickering according to Hamins et al. [22] but still affect the color and shape.
– Uneven illumination: the videos were recorded in different and uncontrolled light conditions. It is worth to mention that fire is, by itself, a source of light that affects the surrounding objects. The dataset also features a large contrast and brightness range.
– Camera movement: the videos that compose the dataset where recorded using either hand-held or robot-attached cameras presenting forward and back, up and down, left and right movement and roll, pitch and yaw rotation.
– Different color accuracy settings: Different cameras were used resulting in a heterogeneous set of videos.
– Clutter: the fire flames can be obscured by surrounding objects, affecting them.
– Partial Occlusion: occlusion has been reported to be one of the biggest challenges to prior approaches specially when using flame contour and optical flow.
– Motion blur: camera shaking is very common when using a robot-attached camera due to the relative motion between the camera and the scene while the shutter is open.
– Scale and projection: fire does not have a specific size, scale or view point. The distances vary from less than 1 m to nearly 15 m meters.
– Reflection: as fire is a light source it may affect the objects that surround it, which will produce reflection and/or any other optical phenomenon.

2.2 Annotations

Fire can often assume random shapes, colors and transparency characteristics, which may have a direct impact on the detected area. For some image regions, deciding whether or not it represents a fire region can be a challenge. Besides the intrinsic flame properties, several factors, such as the image low resolution, scale and occlusion may turn this determination ambiguous. Due the lack of an objective criterion for including (or excluding) a fire region we resort to human judgment for this decision.

Sometimes the fire flickering process results in small fire flame regions that are completely separated from the main fire source. As fire flickers between 2 Hz and 10 Hz it is almost impossible to annotate the exact contour of each single fire flame. Therefore, the annotations are given by a rectangle that embraces the whole fire region. In cases where the fire flames are distant enough to be separated in rectangles without intersection our approach is to annotate them

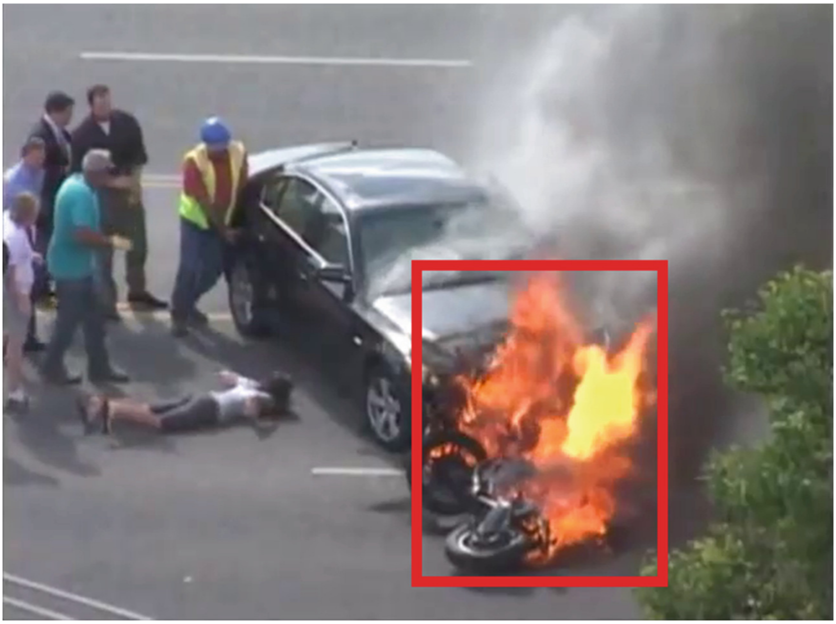

Fig. 1. Approach for the validation process.

as distinct regions, even if they are actually related to the same source. The main reason to separate them is to keep the data clean and allow it to be used with machine learning approaches. Small fire sparkles are left out. For every frame that presents visible fire one or more annotations are made. A example is shown in the Fig. 1.

The annotations are released as XML files. The XML file format was chosen once it is a W3C standard, endorsed by software industry market leaders and easy to read and understand. Every video is provided with its own annotation file, making it simple to extend the dataset or separate it in smaller parts for training, testing and evaluation steps.

The heat map presented in Fig. 2 gives an idea of the fire region placement throughout the dataset. Regions with cold colors present less fire occurrences. Hot colors represent a higher number of occurrences. It can be noted that the majority of the ground truth annotations occur at the center of the image, endorsing some assumptions that have been made in [2]. The average area of a annotation is 61512 pixels (aprox. 250 $\times$ 250 px square). In proportion, fire regions size stands for 8.92 % of the frame size. Figure 3 shows a few excerpts of the videos that compose the dataset.

2.3 Software Tools

To use the ground truth data available in XML format along with the corresponding video files, we also released relevant software artifacts. An OpenCV [3] based

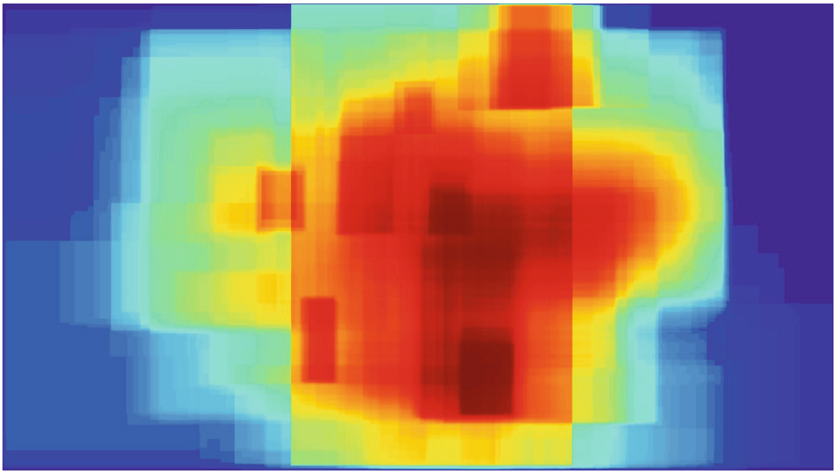

Fig. 2. Heat map showing the density of the annotated regions throughout the dataset.

file manipulation library is provided so that developers can easily access the dataset content calling standard methods. We also provide the software implementation and documentation to enable researchers to compare their algorithms against the ground truth data using the methodology presented in Sect. 3. The interface of the annotation tool is shown in Fig. 4. Quick commands, usage tips and current frame information is presented on the bottom of the frame.

A ground truth visualization tool is provided, allowing the user to load a video file and the corresponding XML file into an application, displaying the annotations overlaid on the image. It implements some features that allow the user to get pertinent data such as image statistics and histograms.

There will be occasions where there is new ground truth data to add or existing data to modify. Therefore a ground truth editor is provided that allows the users to do it graphically. As the annotations are delivered in a XML format the user can also edit the file directly using any text editor.

3 Evaluation Criteria

One challenge in comparing fire detection systems is the lack of agreement on the desired output. The reported performance results are highly dependant on the definition of what is a correct detection result. Therefore we propose three different evaluation approaches: frame based, where it becomes a classification problem, location based, where it becomes a location problem and time based, where the detection delay is considered. We do not consider the approach presented by [2] where they only take in account the classification as a whole video containing or not fire.

On the frame based evaluation, each frame is an instance as in a binary classifier. This approach is the most commonly used to evaluate fire detection

Fig. 3. Sample frames for some of the videos in the dataset. The two top rows show some videos that present fire flames, while the bottom rows show some negative samples.

systems. Even though the authors did not fully explore the potential of the approach, it can be assumed that it has been used before in [5,6,9], where the recall has been labeled as Detection Rate and the precision has been labeled as False Alarm Rate. Considering the whole frame as fire or non-fire makes it

Fig. 4. Ground truth annotation tool interface.

possible to compute the true positive rate TPR (a.k.a. Recall or Hit Rate), true negative rate TNR (a.k.a. Sensitivity), positive predictive value PPV (a.k.a. precision), negative prediction value NPV, false positive rate FPR (a.k.a Fall-Out), false discovery rate FDR, and false negative rate FNR. The corresponding equations are presented in Eqs. 1, 2, 3, 4, 5 and 6 where P and N represent respectively the positive and negative samples count in the dataset and T_P and T_N are the number of true positive and true negative detections.

$$TPR = \frac{T_P}{P} \tag{1}$$

$$SPC = \frac{T_N}{N} \tag{2}$$

$$PPV = \frac{T_P}{T_P + F_P} \tag{3}$$

$$NPV = \frac{T_N}{T_N + F_N} \tag{4}$$

$$FPR = \frac{F_P}{N} \tag{5}$$

$$FDR = \frac{F_P}{F_P + T_P} \tag{6}$$

Considering the fire detection problem as a frame-by-frame binary classification task also makes it possible to compute the accuracy ACC (Eq. 7), F_1 score (Eq. 8) and Matthews correlation coefficient MCC (Eq. 9). In so far as we know, the first time MCC and F Score were used used to compare fire detectors points back to Collumeau *et al.* [11]. Latter, they have also been used in

[39,46]. Although these metrics are not usually considered, they can provide important insights by considering both precision and recall, resulting in a better measurement to evaluate the quality of a detector.

$$ACC = \frac{T_P + T_N}{P + N} \tag{7}$$

$$F_1 = \frac{2 \times PPV \times TPR}{PPV + TPR} \tag{8}$$

$$MCC = \frac{T_P T_N - F_P F_N}{\sqrt{(T_P + F_P)(T_P + F_N)(T_N + F_P)(T_N + F_N)}} \tag{9}$$

The F-Measure was first introduced by Chinchor [10] as a measure that combines both precision and recall in one single metric through the harmonic mean. When the recall and precision values are considered to have the same weight that measure is named as F1-Score. Yet the Matthews correlation coefficient, was introduced in [31] as another balanced metric for binary classifiers that considers true and false positives and negatives. It can be used even if the classes are of different sizes. The MCC scale goes from -1 up to $+1$, where -1 indicates a total disagreement between prediction and observation, 0 indicates a random prediction and $+1$ indicates that the classifier output perfectly matches the ground truth labels.

While the frame based approach may be appropriated for most cases, it may also be interesting to evaluate the location of the detection. Many researchers have yet proposed benchmarking metrics for face and object detection. Most of them, however, assume that there will be only one detection that matches the ground truth annotation which is not appropriated for the fire detection problem. As fire does not have a fixed shape and color and can separate into many flame sparks it is better to use a many to one approach.

Detections are considered true or false positives based on the area overlap with the ground truth rectangles. The similarity function S is given by the Eq. 10 where d_i is the detector output and gt_i is the annotation. A detection is considered as correct when $S > 0.5$.

$$S = \frac{d_i \cap gt_i}{d_i} \tag{10}$$

Once the location overlap based approach does not have the negative sample count we can only compute metrics that do not rely on the negative data. Using the positive data and considering that by definition the precision is the fraction of retrieved instances that are relevant, we are able to compute it using the Eq. 11. The recall, which is defined as the fraction of relevant instances that are retrieved, is given by Eq. 12. D represents the number of detections while GT represents the number of annotations in the evaluation dataset.

$$PPV = \frac{T_P}{D} \tag{11}$$

$$TPR = \frac{GT - F_N}{GT} \tag{12}$$

Another fundamental information when evaluating fire detection systems is the time gap between the moment the fire starts to be detected and the first time it appears in the ground truth. The latency Lt is given by the Eq. 13 where d_i is the frame in which the first detection occurs and gt_i is the first time fire appears in the ground truth annotations. Usually the fire detectors present a intrinsic latency because they use the flame flickering as input.

$$Lt = min(d_i) - min(gt_j) \tag{13}$$

One last important metric when evaluating the quality of a non-stationary video-based fire detection system is the accuracy of the detection location. The average similarity $\overline{S}$ may be an useful information to direct the development of algorithms for active vision and mobile robot systems.

4 Experimental Setup

The dataset has been used to compare three different state-of-the-art techniques. The source code for each solution was implemented in C++ and the default parameters were used. The methods proposed by Celik [4] and Zhou [47], are both based on temporal evaluation. Chenebert's [9] approach is texture based and non-temporal. The results for this experimental setup are given in Tables 1 and 2, which respectively show the performance considering the frame by frame classification and the location based metrics.

Table 1. Frame by frame results.

Metrics	Better	Celik [4]	Zhou [47]	Chenebert [9]
TPR	↑	0.739	0.987	0.990
SPC	↑	0.317	0.022	0.724
PPV	↑	0.654	0.638	0.857
NPV	↑	0.410	0.501	0.979
FPR	↓	0.682	0.977	0.275
FDR	↓	0.345	0.361	0.142
FNR	↓	0.260	0.012	0.009
ACC	↑	0.585	0.635	0.890
F_1 Score	↑	0.694	0.775	0.919
MCC	↑	0.060	0.036	0.773

While the temporal based methods show a high recall and could make us think that the results are satisfactory, the balanced metrics show that, in fact,

Table 2. Location based results

Metrics	Better	Celik [4]	Zhou [47]	Chenebert [9]
L_{PPV}	↑	0.251	0.019	0.832
L_{TPR}	↑	0.732	0.440	0.979
F_1 Score	↑	0.384	0.037	0.902
$\overline{S}$	↑	0.250	0.020	0.801

the predictions are almost random, having a small correlation with the expected outputs. On the same hand, when the location based metrics are considered, Table 2 proves that the Chenebert's method outperforms both Celik's and Zhou's methods, being the only one to report a high L_{PPV} and L_{TPR}. The mean similarity $\overline{S}$ also shows that the stationary camera based systems do not present a good detection/ground truth area intersection. The Borges *et al.* method was not tested once the code is not publicly available and it requires some threshold values that are not presented on the original paper.

5 Additional Considerations

While the proposed evaluation kit presents a complete set of videos, annotations and software artifacts, we also have to acknowledge some of its limitations. First and most important, as has been mentioned in the Sect. 2.2 the annotations do not provide the exact contours of the fire regions, which in turn does not enable developers to compute the Hafiane's criterion [21]. The Hafiane's criterion (a.k.a. Hafiane quality index) is a supervised evaluation criterion for region based segmentation methods. It considers the position, shape, and size of the segmented regions and has been used before for the evaluation of VFD systems in [11,46]. While we think the Hafiane's criterion could be an useful metric, we have to consider that, in turn, the aforementioned publications used datasets with only a few hundred images or less.

Also regarding the annotations, we reinforce they are based on OpenCV XML format and have not been tested with other computer vision and deep learning libraries. However, we believe this wont pose a major problem, once the format is a standard in the software development industry. In comparison to the framework for benchmarking of wildland fire segmentation algorithms from [46] the proposed evaluation kit lacks relevant information from the context, such as climate conditions, burning material or camera pose. That could be an important improvement for future updates in the dataset.

The videos are in the MPEG-4 file format, which requires specific codecs. On the other hand, mp4 codecs are available for all mainstream operational systems. By default the videos are stored in the YUV color format, with 8 bits bit depth. As they were obtained from different and converted to this format latter for the sake of standardization, this might result in an overall non-significant quality loss. When using OpenCv library it will, by default, convert the frames to BGR (blue, green, red) color space to simplify manipulation.

Aside from being the first dataset aimed to evaluate and validate non-stationary fire detection systems, the presented dataset is also larger than others which have been previously proposed. For instance, [11] used 76 pictures, [46] used 100 images. When it comes to datasets used in other publications, but not publicly available or annotated in details, it is important to notice that the presented kit has almost balanced fire and non-fire samples. Labati *et al.* [29] used 72852 frames in total, but the data was very unbalanced, with only 1328 smoke samples. Similarly Toreyin *et al.* [43] used a dataset with 83745 frames in 61 sequences, but only 19 of them contained fire regions. That is an important remark, given that many machine learning and data mining algorithms do not work properly on imbalanced datasets.

Although, it is aimed to be used in non-stationary systems, it can also come in hand to evaluate stationary VFD systems. Stationary approaches can be divided in two large groups: (I) systems based on single images and (II) systems based on multiple frames. The latter ones usually depend on fire flickering features extracted from videos and therefore they can not be trained or evaluated with a shaking camera. The first ones, on the other hand, depend only on color, boundaries or texture clues extracted from a single frame, and therefore they can be trained using the presented dataset.

The proposed dataset has been integrated to CvWorks – Computer Vision Framework[5] and all parts of the kit are available on Github[6]. The dataset is downloadable and can be easily extended due its annotation model. Public datasets play an important role for the development and evaluation of computer vision solutions. Datasets such as Image-net [12], Pascal VOC [13] and Calltech [14,18] had a great impact on the development of object recognition. So had Kitty [17] and TUM-RGB-D [41] to the development of autonomous vehicles, and AM-FED [32] and CK+ [30] for the development of human action and emotion recognition. In the same way, our dataset can be used for the development of machine learning approaches such as deep neural networks, which are a current trend in this field.

For methods that need to compute a threshold value, we recommend to follow the same procedure as Rudz *et al.* [37]. In this approach, thresholds are found using a third of the images of the dataset taken randomly. The value of the threshold that maximises the F_1 Score (Eq. 8) for these images is estimated with a direct pattern search algorithm. Once the threshold value is found it should be used for all testing steps. Otherwise, adjusting the threshold and parameters during the process would result in over-fitting, invalidating the whole results.

[5] Available at http://www.cvworks.c3.furg.br.

[6] Available at https://github.com/steffensbola/furg-fire-dataset.

6 Conclusion

The validation kit presented, featuring a dataset of videos annotated at the frame level, evaluation metrics, along with the software artifacts, enable researchers to accurately evaluate their video-based fire detection methods. The dataset is publicly available under a non-restrictive copyright license which allows it to be freely shared and redistributed. Open file formats are used to provide large compatibility and easy use. Example implementations with source code are provided using the OpenCv computer vision library, which currently is a *de facto* standard for computer vision research. For researches who intend to use the provided kit with another programming language we provide evaluation metrics that are unambiguous and straightforward to implement.

In this work, as a demonstration, three state-of-the-art VFD algorithms were implemented and their performances analysed. The obtained results justify the effort to build the presented dataset and evaluation scheme, once they show there is still room for further improvements in non-stationary VFD systems. The dataset is a convenient resource to support the development of machine learning approaches and benchmark tools, enabling the researchers to focus on new algorithms and improvements rather than implementing their own metrics and creating their own datasets from scratch. Aside from being a standardized validation for non-stationary VFD systems, the proposed dataset can boost research and push the development of autonomous fire hazard combat alternatives such as early stage fire alarm, automatic fire extinguishers and firefighting robots.

Our evaluation kit addresses a computer vision application that did not have a publicly available evaluation scheme yet. Therefore, the importance of the presented dataset for fire detection algorithms can be compared to what ground truth data represented in fields like face, object detection and tracking decades ago, fostering their development to reach a level where they have become reliable enough to be used in real life applications.

References

1. AD Group: Firevu video smoke detection (2016). http://www.firevu.com/
2. Borges, P.V.K., Izquierdo, E.: A probabilistic approach for vision-based fire detection in videos. IEEE Trans. Circ. Syst. Video Technol. **20**(5), 721–731 (2010)
3. Bradski, G.: The OpenCV library. Dr. Dobb's J. Softw. Tools **25**(11), 120–126 (2000)
4. Celik, T.: Fast and efficient method for fire detection using image processing. ETRI J. **32**(6), 881–890 (2010)
5. Celik, T., Demirel, H.: Fire detection in video sequences using a generic color model. Fire Saf. J. **44**(2), 147–158 (2009)
6. Celik, T., Ozkaramanli, H., Demirel, H.: Fire and smoke detection without sensors: image processing-based approach. In: 15th European Signal Processing Conference, EUSIPCO, pp. 147–158 (2007)
7. Chen, J., He, Y., Wang, J.: Multi-feature fusion based fast video flame detection. Build. Environ. **45**(5), 1113–1122 (2010)

8. Chen, T.H., Wu, P.H., Chiou, Y.C.: An early fire-detection method based on image processing. In: 2004 International Conference on Image Processing, ICIP 2004, vol. 3, pp. 1707–1710. IEEE (2004)
9. Chenebert, A., Breckon, T.P., Gaszczak, A.: A non-temporal texture driven approach to real-time fire detection. In: 2011 18th IEEE International Conference on Image Processing (ICIP), pp. 1741–1744. IEEE (2011)
10. Chinchor, N.: MUC-4 evaluation metrics. In: Proceedings of the 4th Conference on Message Understanding, MUC4 1992, pp. 22–29. Association for Computational Linguistics, Stroudsburg, PA, USA (1992). http://dx.doi.org/10.3115/1072064.1072067
11. Collumeau, J.F., Laurent, H., Hafiane, A., Chetehouna, K.: Fire scene segmentations for forest fire characterization: a comparative study. In: 2011 18th IEEE International Conference on Image Processing (ICIP), pp. 2973–2976. IEEE (2011)
12. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 248–255. IEEE (2009)
13. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL visual object classes (VOC) challenge. Int. J. Comput. Vis. **88**(2), 303–338 (2010)
14. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. Comput. Vis. Image Underst. **106**(1), 59–70 (2007)
15. Fike Corporation: Signifire video flame, smoke and intrusion detection system, January 2016. http://www.fike.com/products/signifire-video-flame-smoke-intrusion-detection-system/
16. Foo, S.Y.: A rule-based machine vision system for fire detection in aircraft dry bays and engine compartments. Knowl.-Based Syst. **9**(8), 531–540 (1996)
17. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: the kitti dataset. Int. J. Robot. Res. **32**(11), 1231–1237 (2013). 0278364913491297
18. Griffin, G., Holub, A., Perona, P.: Caltech-256 object category dataset (7694) (2007). http://authors.library.caltech.edu/7694
19. Gunay, O., Cetin, A.E.: Real-time dynamic texture recognition using random sampling and dimension reduction. In: 2015 IEEE International Conference on Image Processing (ICIP), pp. 3087–3091. IEEE (2015)
20. Habiboglu, Y.H., Gunay, O., Cetin, A.E.: Real-time wildfire detection using correlation descriptors. In: 19th European Signal Processing Conference (EUSIPCO 2011), Special Session on Signal Processing for Disaster Management and Prevention, pp. 894–898 (2011)
21. Hafiane, A., Chabrier, S., Rosenberger, C., Laurent, H.: A new supervised evaluation criterion for region based segmentation methods. In: Blanc-Talon, J., Philips, W., Popescu, D., Scheunders, P. (eds.) ACIVS 2007. LNCS, vol. 4678, pp. 439–448. Springer, Heidelberg (2007). doi:10.1007/978-3-540-74607-2_40
22. Hamins, A., Yang, J., Kashiwagi, T.: An experimental investigation of the pulsation frequency of flames. Symp. (Int.) Combust. **24**(1), 1695–1702 (1992). Elsevier
23. Haralick, R.M.: Ridges and valleys on digital images. Comput. Vis. Graph. Image Process. **22**(1), 28–38 (1983)
24. Healey, G., Slater, D., Lin, T., Drda, B., et al.: A system for real-time fire detection. In: 1993 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Proceedings, CVPR 1993, pp. 605–606. IEEE (1993)
25. InnoSys Industries Inc: Alarmeye ae3000 pc based video fire detection system (2016). http://www.innosys-ind.com/

26. Jin, D., Li, S., Kim, H.: Robust fire detection using logistic regression and randomness testing for real-time video surveillance. In: 2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA), pp. 608–613. IEEE (2015)
27. Kong, S.G., Jin, D., Li, S., Kim, H.: Fast fire flame detection in surveillance video using logistic regression and temporal smoothing. Fire Saf. J. **79**, 37–43 (2016)
28. Kwak, J., Ko, B.C., Nam, J.Y.: Forest smoke detection using CCD camera and spatial-temporal variation of smoke visual patterns. In: 2011 Eighth International Conference on Computer Graphics, Imaging and Visualization (CGIV), pp. 141–144. IEEE (2011)
29. Labati, D., Genovese, A., Piuri, V., Scotti, F.: Wildfire smoke detection using computational intelligence techniques enhanced with synthetic smoke plume generation. IEEE Trans. Syst. Man Cybern. Syst. **43**(4), 1003–1012 (2013)
30. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 94–101. IEEE (2010)
31. Matthews, B.W.: Comparison of the predicted and observed secondary structure of t4 phage lysozyme. Biochim. Biophys. Acta (BBA)-Protein Struct. **405**(2), 442–451 (1975)
32. McDuff, D., El Kaliouby, R., Senechal, T., Amr, M., Cohn, J.F., Picard, R.: Affectiva-MIT facial expression dataset (AM-FED): naturalistic and spontaneous facial expressions collected in-the-wild. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 881–888. IEEE (2013)
33. Mueller, M., Karasev, P., Kolesov, I., Tannenbaum, A.: Optical flow estimation for flame detection in videos. IEEE Trans. Image Process. **22**(7), 2786–2797 (2013)
34. Nguyen-Ti, T., Nguyen-Phuc, T., Do-Hong, T.: Fire detection based on video processing method. In: 2013 International Conference on Advanced Technologies for Communications (ATC), pp. 106–110. IEEE (2013)
35. Park, J., Ko, B., Nam, J.Y., Kwak, S.: Wildfire smoke detection using spatiotemporal bag-of-features of smoke. In: 2013 IEEE Workshop on Applications of Computer Vision (WACV), pp. 200–205. IEEE (2013)
36. Plumb, O., Richards, R.: Development of an economical video based fire detection and location system. US Department of Commerce, Technology Administration (1996)
37. Rudz, S., Chetehouna, K., Hafiane, A., Laurent, H., et al.: Investigation of a novel image segmentation method dedicated to forest fire applications. Meas. Sci. Technol. **24**(7), 075403 (2013)
38. Shidik, G.F., Adnan, F.N., Supriyanto, C., Pramunendar, R.A., Andono, P.N.: Multi color feature, background subtraction and time frame selection for fire detection. In: Robotics, Biomimetics, and Intelligent Computational Systems (ROBIONETICS), 2013 IEEE International Conference on, pp. 115–120 (2013). doi:10.1109/ROBIONETICS.2013.6743589
39. Steffens, C.R., Rodrigues, R.N., da Costa Botelho, S.S.: An unconstrained dataset for non-stationary video based fire detection. In: IEEE LARS/SBR 2015. Uberlândia, October 2015
40. Stipaničev, D., Vuko, T., Krstinić, D., Štula, M., Bodrožić, L.: Forest fire protection by advanced video detection system-croatian experiences. In: Third TIEMS Workshop-Improvement of Disaster Management System, pp. 26–27. Citeseer (2006)

41. Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D.: A benchmark for the evaluation of RGB-D SLAM systems. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 573–580. IEEE (2012)
42. Toreyin, B.U., Cetin, A.E.: Wildfire detection using LMS based active learning. In: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2009, pp. 1461–1464. IEEE (2009)
43. Töreyin, B.U., Dedeoğlu, Y., Güdükbay, U., Cetin, A.E.: Computer vision based method for real-time fire and flame detection. Pattern Recogn. Lett. **27**(1), 49–58 (2006)
44. Toreyin, B., Dedeoglu, Y., Cetin, A.: Wavelet based real-time smoke detection in video. In: 13th European Signal Processing Conference, pp. 1–4, September 2005
45. Toreyin, B., Dedeoglu, Y., Cetin, A.: Contour based smoke detection in video using wavelets. In: 14th European Signal Processing Conference, pp. 1–5, September 2006
46. Toulouse, T., Rossi, L., Akhloufi, M., Celik, T., Maldague, X.: Benchmarking of wildland fire colour segmentation algorithms. IET Image Process. **9**(12), 1064–1072 (2015)
47. Zhou, X.L., Yu, F.X., Wen, Y.C., Lu, Z.M., Song, G.H.: Early fire detection based on flame contours in video. Inf. Technol. J. **9**(5), 899–908 (2010)