FIRE DETECTION FROM IMAGES USING FASTER R-CNN AND MULTIDIMENSIONAL TEXTURE ANALYSIS

Panagiotis Barmpoutis[1], *Kosmas Dimitropoulos*[2], *Kyriaki Kaza*[2], *and Nikos Grammalidis*[2]

[1]Department of Electrical and Electronic Engineering, Faculty of Engineering, Imperial College London, United Kingdom

[2] Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, 57001, Greece

ABSTRACT

In this paper, we propose a novel image-based fire detection approach, which combines the power of modern deep learning networks with multidimensional texture analysis based on higher-order linear dynamical systems. The candidate fire regions are identified by a Faster R-CNN network trained for the task of fire detection using a set of annotated images containing actual fire as well as selected negatives. The candidate fire regions are projected to a Grassmannian space and each image is represented as a cloud of points on the manifold. Finally, a vector representation approach is applied aiming to aggregate the Grassmannian points based on a locality criterion on the manifold. For evaluating the performance of the proposed methodology, we performed experiments with annotated images of two different databases containing fire and fire-coloured objects. Experimental results demonstrate the potential of the proposed methodology compared to other state of the art approaches.

Index Terms— *Image-based fire detection, Convolutional Neural Networks, Spatial texture analysis, Linear Dynamical Systems.*

1. INTRODUCTION

Forest fires are one of the most harmful natural hazards, having numerous and significant consequences on the environment, local communities, economy and heritage. Furthermore, climate change has increased the number of droughts, heat waves and dry spells, leading to a marked increase of fire potential across Europe and worldwide [1]. This results to an increase the length and severity of the fire season, the area at risk and the probability of large fires [2].

Thus, computer-based early fire warning systems have attracted particular attention during the last decade. Detection techniques that are based on various colour spaces [3], [4], spectral [5], spatial [6], texture characteristics [7] and deep learning methods [8], [9], have been widely investigated. More recently, due to the increasing number of social media users and the significant increase of data and specifically of images that are published on the internet, the researchers exploit the potential for the early detection of fires using public information. To this end, many approaches have been developed for the understanding of the syntax and context of messages [10], geo-tags [11] and images [12], [13], for the early detection of wildfire events.

More specifically, for fire detection from images Bedo et al., [12] applied colour and texture extractors by designing and implementing a database-driven architecture. Furthermore, Sharma et al., [14] combined two pre-trained deep Convolutional Networks (CNNs), namely VGG16 and Resnet50, to develop a fire detection system. The dataset included images that were unbalanced by including significantly more non-fire images than fire images. More recently, Zhang et al., [9] used Faster R-CNN to detect wildland forest fire smoke to avoid the complex manually feature extraction process. In addition, Toulouse et al., [13] identified the difficulty to find an annotated dataset and presented a publicly evolving wildfire annotated image database with ground truth data and examples of use.

However, false alarm rates are often high due to natural objects, which have the same characteristics with flame, large variations of flame appearance and environmental changes that complicate fire detection including clouds, sun and light reflections. Hence, the challenge in fire detection from digital images lies in the modelling and detection of the chaotic and complex nature of the fire phenomenon. In this paper, we propose an efficient approach for early fire detection from images by combining a powerful deep learning technique with multidimensional texture analysis using Linear Dynamical Systems (LDS). Specifically, we first extract candidate fire regions of each image using a Faster R-CNN network (experiments were performed applying three different architectures, namely AlexNet, VGG16 and Resnet101), then represent the candidate fire regions of an image as a cloud of points on the Grassmann manifold and finally extract a VLAD descriptor for each image. To evaluate the efficiency of the proposed methodology, we used images from two different databases containing a large number of images of wildfire, including images with dominant fire-like colours and/or fire-coloured objects. Specifically, we used annotated wildfire images from the

978-1-5386-4658-8/18/$31.00 ©2019 IEEE

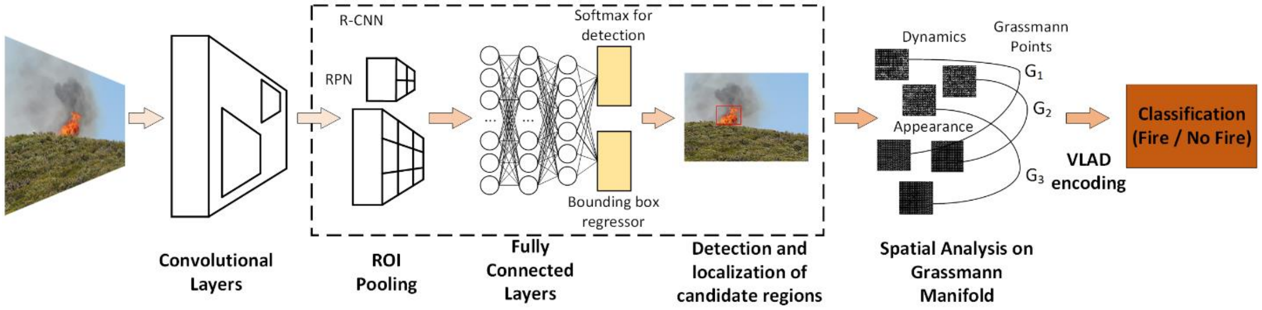

Figure 1. The proposed methodology

Corsican Fire Database (CFDB) [13] as well as images of various objects and classes from the PASCAL Visual Object Classes (VOC) dataset [15].

The rest of this paper is organized as follows: First, details of the proposed methodology are presented, followed by experimental results using the specific dataset. Finally, some conclusions are drawn and future extensions are discussed.

2. METHODOLOGY

The proposed methodology consists of two main steps, as shown in Fig. 1. More specifically, in the first step, each image is fed as input into a Faster R-CNN network for the detection and localization of candidate fire regions. In the second step, the extracted candidate fire regions are divided into rectangular patches (blocks), which are modelled using linear dynamical systems and projected to a Grassmann manifold. As a result, each region is represented as a cloud of points on the manifold. For the classification of the images, a vector representation approach, namely VLAD encoding, is applied, aiming to aggregate the Grassmannian points based on a locality criterion on the manifold.

2.1 Detection of candidate fire regions using Faster R-CNN

Region-CNN (R-CNN) [16] is one of the state-of-the-art CNN-based deep learning approaches for object detection. It first uses a selective search approach to generate a number of region proposals, i.e. bounding boxes for image classification. Then, for each bounding box, image classification is performed through a CNN, while each bounding box is finally refined using regression. Faster R-CNN [17] is an improved variant addressing important drawbacks of R-CNN. Specifically, the slow selective search procedure is replaced by another CNN, namely Region Proposal Network (RPN), which is integrated with the detection network. An end-to-end training procedure is used to simultaneously learn the class of object as well as the associated bounding box position and size. Thus, the overall computational complexity is significantly reduced, while performance is improved.

In this paper, for the detection of the candidate fire regions, we tested Faster R-CNN with three different base networks, AlexNet, VGG16 and Resnet101. We trained the Faster R-CNN for fire detection (two classes: fire and background), using transfer learning. For training, we split the set of 500 RGB annotated images from CFDB (410 are used for training and the rest for testing). In addition, non-fire images from the PASCAL VOC dataset, including many fire-coloured objects, were also used (200 in training and 350 in testing).

2.2. Spatial texture analysis using Linear Dynamical Systems and VLAD encoding

In general, LDS models are used for spatiotemporal modelling and attempt to associate the output of the system, i.e., the observation, with a linear function of a state variable, while in each time instant, the state variable depends linearly on the state of the previous time instant. Both state and output noise are zero-mean normally distributed random variables and apart from the output of the system, all other variables (state and noise variables) are hidden. The adopted system is described by the following equations:

$$x(t+1) = Ax(t) + Bv(t) \qquad (1)$$

$$y(t) = \bar{y} + Cx(t) + w(t) \qquad (2)$$

where $x \in R^n$ is the hidden state process, $y \in \mathbb{R}^d$ is the observed data, $A \in \mathbb{R}^{n \times n}$ is the transition matrix of the hidden state and $C \in \mathbb{R}^{d \times n}$ is the mapping matrix of the hidden state to the output of the system. The quantities $w(t)$ and $Bv(t)$ are the measurement and process noise respectively, with $w(t) \sim N(O, R)$ and $Bv(t) \sim N(0, Q)$, while $\bar{y} \in \mathbb{R}^d$ is the mean value of the observation data. The extracted tuple LDS descriptor, $M = (A, C)$, models both the appearance and dynamics of the observation data, represented by C and A, respectively [18].

In the proposed image-based approach, fire can be considered as a spatially-varying visual pattern. Hence, spatial texture analysis through linear dynamical systems can be applied in order to increase the reliability of the algorithm by modelling the spatially-evolving multidimensional fire regions. To exploit this information, in this paper we attempt to extract the appearance information and the dynamics of each candidate fire region using linear dynamical systems

(LDSs). Initially we divide each region into a number of image blocks with size $n \times n$ (we set $n = 16$ based on our previous research [7] for spatio-temporal analysis of fire). We then represent the region with a third-order tensor Y and apply a higher-order Singular Value Decomposition (hoSVD) to decompose the tensor and estimate the mapping and transition matrix C and A respectively [19]. In addition, in order to improve the stability of the dynamical system, we estimated the stabilized transition matrix A, adopting the methodology used in [20].

Having modelled each candidate fire region using a higher-order linear dynamical system approach, we estimate the finite observability matrix of each dynamical system, $O_m^T(M) = [C^T, (CA)^T, (CA^2)^T, \dots, (CA^{m-1})^T]$ and then, we apply a Gram-Schmidt othonormalization [21] procedure, i.e., $O_m^T = GR$, in order to represent each descriptor M as a Grassmannian point, $G \in \mathbb{R}^{m \times \mathrm{T} \times 3}$[20].

Finally, for the modelling of each fire candidate region, we apply VLAD encoding, which is considered as a simplified coding scheme of the earlier Fisher Vector (FV) representation and was shown to outperform histogram representations in bag of features approaches [22], [23]. More specifically, we consider a codebook, $\{m_i\}_{i=1}^r = \{m_1, m_2, \dots, m_r\}$, with r visual words and local descriptors v, where each descriptor is associated to its nearest codeword $m_i = NN(v_j)$. The VLAD descriptor, $\bar{V}$, is created by concatenating the r local difference vectors $\{u_i\}_{i=1}^r$ corresponding to differences $v_j - m_i$, with $m_i = NN(v_j)$, where v_j are the descriptors associated with codeword i, with $i = 1, \dots, r$.

$$\bar{V} = \{u_i\}_{i=1}^r = \{u_1, \dots, u_r\} \tag{3}$$

or

$$\bar{V} = \left\{ \sum_{\substack{v_j \, such \, that \\ m_1 = NN(v_j)}} (v_j - m_1), \dots, \sum_{\substack{v_j \, such \, that \\ m_r = NN(v_j)}} (v_j - m_r) \right\}. \tag{4}$$

The final VLAD representation is determined by the L2-normalization of vector $\bar{V}$:

$$\bar{V}_{Euclidean} = \bar{V} / \|\bar{V}\|_2 \tag{5}$$

Finally, for the classification of fire candidate regions, the VLAD representations on Grassmann manifold are fed into an SVM classifier.

3. EXPERIMENTAL RESULTS

To evaluate the efficiency of the proposed methodology, we used a dataset, consisting of 500 images that contain fire events of CFDB and 550 images that contain fire coloured objects. It is worth mentioning that CFDB is the largest dataset released in this research field [13]. The goal of this experimental evaluation is two-fold: a) Initially, we aim to show that the proposed methodology improves the detection of fires using images and b) secondly we want to demonstrate the superiority of the proposed approach against other state of the art approaches.

To achieve the first goal, we compare the recognition rates of the proposed methodology with those achieved by the Faster R-CNN and spatial texture analysis (Grassmannian VLAD encoding) when applied separately for fire detection (Fig. 2). Specifically, we present the percentages of images where fire is correctly detected (true positives) or not (false negatives) out of all fire images, together with the percentages of images correctly identified as non-fire (true negatives) and false alarms (false positives) out of all non-fire images.

Concerning the true positive rates, AlexNet Faster R-CNN is slightly better than spatial texture analysis, achieving 1.1% higher recognition rate. However, both VGG16 and Resnet101 Faster R-CNN networks significantly improve true positives achieving 100% recognition rates. Furthermore, the proposed spatial texture analysis achieves the best true negatives results against AlexNet, VGG16 and Resnet101 networks. By combining Faster R-CNN network and spatial texture analysis, the proposed algorithm retains the high detection rate offered by the deep network and

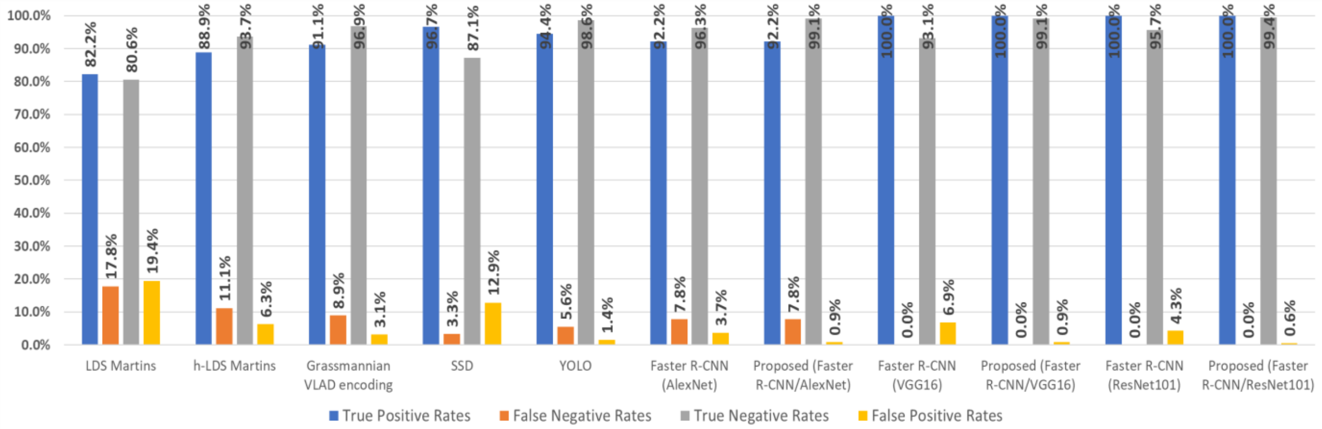

Figure 2. Detection rates for the proposed and state-of-the-art approaches

drastically minimizes the probability of false positives, achieving a true negative rate of 99.4% using the Resnet101 architecture. Thus, it is shown that the addition of the spatial texture analysis performed by linear dynamical systems on top of the deep network proposals offers a significant performance improvement. It is also worth mentioning that experiments for spatial texture analysis were performed using whole images, while, in the proposed approach, spatial texture analysis was performed only in the candidate regions. In this way, the associated computational cost of the spatial texture analysis is also significantly reduced. Fig. 3(a-b) shows two images that are correctly identified as actual fire by the proposed system, Fig. 3c shows an image that correctly detected as a non-fire object and Fig. 3d shows an image that is erroneously identified as actual fire.

Table I. Comparison results of various fire detection approaches

Method	F-Score
LDS (Martin distance)	81.5%
h-LDS (Martin distance)	91.1%
Grassmannian VLAD encoding	93.8%
SSD	92.3%
YOLO v3	96.4%
Faster R-CNN (AlexNet)	94.1%
Proposed (Faster R-CNN/AlexNet)	**95.5%**
Faster R-CNN (VGG16)	96.7%
Proposed (Faster R-CNN/VGG16)	**99.6%**
Faster R-CNN (Resnet101)	97.9%
Proposed (Faster R-CNN/ResNet101)	**99.7%**

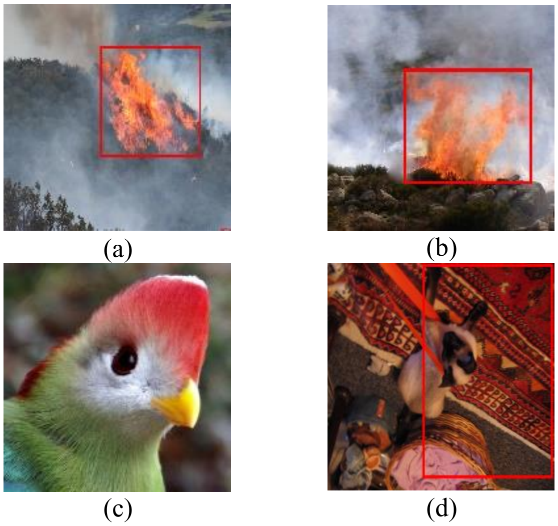
(a)
(b)
(c)
(d)

Fig. 3. Dataset images containing actual fires and fire-coloured (non-fire) objects: a-b) True detection of actual fire, c) True negative of fire-coloured object, d) False detection of fire-coloured object as fire.

Regarding the second goal, we applied the proposed methodology in three Faster R-CNN architectures. As shown in Table I, we applied the proposed methodology, based on Faster R-CNN with AlexNet, VGG16 and Resnet101 architectures, achieving F-score rates 95.5%, 99.6% and 99.7%, respectively. Furthermore, we compared the performance of the proposed approaches, estimating the fire recognition rates (Fig. 2) and the F-scores (Table I), with the performance of other approaches. As depicted in Table I, our method offers an improved F-score rate compared to the second higher rate of 97.9% yield by the Faster R-CNN Resnet101 network alone. Furthermore, the proposed methodology improves state-of-the h-LDS using Martin distance as a similarity metric [7] rates by 8.6% and Grassmannian VLAD encoding approach by 5.9%, while Faster R-CNN with VGG16 and Resnet101 as base networks achieve higher F-score rates against YOLO v3 [24] and SSD [25] approaches (trained using the same training set).

4. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented an approach for fire detection from images combining the merits of deep learning and spatial texture analysis. The two main steps of the proposed approach are a) candidate region detection using a Faster R-CNN network trained for fire detection and b) validation of detected fire regions using analysis of spatial characteristics through Linear Dynamical Systems (LDS). In order to discriminate fire-coloured objects and actual fire we used VLAD encoding that improves performance and significantly decreases detection errors. We applied our approach to both fire images and fire coloured objects. Experimental results show that the proposed approach retains high true positive rates, while simultaneously significantly reducing false positives due to fire-coloured objects. In the future, we aim to a) extend our dataset using images from social media in order to further assess the effectiveness of the proposed methodology and b) extend the proposed approach for fire detection in video sequences using dynamic textures.

5. ACKNOWLEDGEMENT

The research leading to these results has received funding from INTERREG V-A COOPERATION PROGRAMME Greece-Bulgaria 2014-2020 project ‘‘e-OUTLAND: Protecting biodiversity at NATURA 2000 sites and other protected areas from natural hazards through a certified framework for cross-border education, training and support of civil protection volunteers based on innovation and new technologies’’.

5. REFERENCES

[1] European Environment Agency, Forest Fires, online access 10/08/2018: https://www.eea.europa.eu/data-and-maps/indicators/forest-fire-danger-2/assessment

[2] M. Rodríguez and J. Manuel, "Forest fires under climate, social and economic changes in Europe, the Mediterranean and other fire-affected areas of the world", 2014.

[3] B. U. Töreyin, Y. Dedeoğlu, U. Güdükbay and A. E. Cetin, "Computer vision based method for real-time fire and flame detection", *Pattern Recognition Letters*, 27(1), pp. 49-58, 2006.

[4] K. Dimitropoulos, F. Tsalakanidou, and N. Grammalidis, "Flame detection for video-based early fire warning systems and 3D visualization of fire propagation", In 13th IASTED International Conference on Computer Graphics and Imaging, Crete, Greece, June 2012.

[5] N. Grammalidis, E. Cetin, K. Dimitropoulos, F. Tsalakanidou, K. Kose, O. Gunay, B. Gouverneur, D. Torri, E. Kuruoglu, S. Tozzi, A. Benazza, F. Chaabane, B. Kosucu, C. Ersoy, "A Multi-Sensor Network for the Protection of Cultural Heritage", *19th European Signal Processing Conference*, Barcelona, Spain, 2011.

[6] P. Barmpoutis, K. Dimitropoulos and N. Grammalidis, "Real time video fire detection using spatio-temporal consistency energy", in *Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on*, pp. 365-370, IEEE, Aug. 2013.

[7] K. Dimitropoulos, P. Barmpoutis, and N. Grammalidis. "Spatio-temporal flame modeling and dynamic texture analysis for automatic video-based fire detection." *IEEE Transactions on Circuits And Systems for Video Technology* 25.2 (2015): 339-351

[8] D. Shen, X. Chen, M. Nguyen and W. Q. Yan, "Flame detection using deep learning." *2018 4th International Conference on Control, Automation and Robotics (ICCAR)*. IEEE, 2018, pp. 416-420, April 2018.

[9] Q. X. Zhang, G. H. Lin, Y. M. Zhang, G. Xu and J. J. Wang, "Wildland Forest Fire Smoke Detection Based on Faster R-CNN using Synthetic Smoke Images", *Procedia engineering*, 211, pp. 441-446, 2018.

[10] X. Shi, B. Xue, M. H. Tsou, X. Ye, B. Spitzberg, J. M. Gawron and R. Jin, "Detecting events from the social media through exemplar-enhanced supervised learning", *International Journal of Digital Earth*, pp. 1-15, 2018

[11] M. Kibanov, G. Stumme, I. Amin and J. Lee, "Mining social media to inform peatland fire and haze disaster management." *Social Network Analysis and Mining* 7.1 (2017): 30, 2017.

[12] Bedo, M. V. N., de Oliveira, W. D., Cazzolato, M. T., Costa, A. F., Blanco, G., Rodrigues, J. F., ... & Traina, C., "Fire detection from social media images by means of instance-based learning", *International Conference on Enterprise Information Systems*, Springer, Cham, pp. 23-44, April 2015.

[13] T. Toulouse, L. Rossi, A. Campana, T. Celik and M. A. Akhloufi, "Computer vision for wildfire research: An evolving image dataset for processing and analysis", *Fire Safety Journal* 92, 188-194, 2017.

[14] J. Sharma, O. C. Granmo, M. Goodwin and J. T. Fidje, "Deep convolutional neural networks for fire detection in images", *International Conference on Engineering Applications of Neural Networks*. Springer, Cham, 2017, pp. 183-193. Springer, Cham, August 2017.

[15] M. Everingham, L. Van Gool, C. K. Williams, J. Winn and A. Zisserman, "The PASCAL Visual Object Classes (VOC) challenge.", *International journal of Computer Vision* 88.2, pp. 303-338, 2010.

[16] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation" In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580-587, 2018.

[17] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks", *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (6), 1137-1149, 2017.

[18] G. Doretto, A. Chiuso, Y. N. Wu and S. Soatto, "Dynamic textures", *International Journal of Computer Vision*, 51(2), 91-109, 2003.

[19] K. Dimitropoulos, P. Barmpoutis, N. Grammalidis, "Higher Order Linear Dynamical Systems for Smoke Detection in Video Surveillance Applications", *IEEE Transactions on Circuits and Systems for Video Technology*, 27(5), 1143-1154, 2017

[20] K. Dimitropoulos, P. Barmpoutis, A. Kitsikidis, N. Grammalidis, "Classification of Multidimensional Time-Evolving Data using Histograms of Grassmannian Points", *IEEE Transactions on Circuits and Systems for Video Technology*, 28(4), 892-905, 2018.

[21] G. Arfken, "Gram-schmidt orthogonalization", *Mathematical methods for physicists*, 3, pp. 516-520, 1985.

[22] H. Jégou, M. Douze, C. Schmid, C. and P. Pérez, "Aggregating local descriptors into a compact image representation.", *Computer Vision and Pattern Recognition (CVPR)*, 2010 IEEE Conference on. IEEE, 2010, pp. 3304-3311

[23] V. Kantorov and I Laptev, "Efficient feature extraction, encoding and classification for action recognition", In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2593-2600, 2014.

[24] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).

[25] Liu, Wei, et al. " SSD: Single Shot MultiBox Detector." European conference on computer vision. Springer, Cham, 2016.