

A Novel Framework for Early Fire Detection Using Terrestrial and Aerial 360-Degree Images

Panagiotis Barmpoutis^(⊠) and Tania Stathaki

Department of Electrical and Electronic Engineering, Faculty of Engineering, Imperial College London, London, UK {p.barmpoutis,t.stathaki}@imperial.ac.uk

Abstract. In this paper, in order to contribute to the protection of the value and potential of forest ecosystems and global forest future we propose a novel fire detection framework, which combines recently introduced 360-degree remote sensing technology, multidimensional texture analysis and deep convolutional neural networks. Once 360-degree data are obtained, we convert the distorted 360-degree equirectangular projection format images to cubemap images. Subsequently, we divide the extracted cubemap images into blocks using two different sizes. This allows us to apply h-LDS multidimensional spatial texture analysis to larger size blocks and then, depending on the probability of fire existence, to smaller size blocks. Thus, we aim to accurately identify the candidate fire regions and simultaneously to reduce the computational time. Finally, the candidate fire regions are fed into a CNN network in order to distinguish between fire-coloured objects and fire. For evaluating the performance of the proposed framework, a dataset, namely "360-FIRE", consisting of 100 images with unlimited field of view that contain synthetic fire, was created. Experimental results demonstrate the potential of the proposed framework.

Keywords: Fire detection \cdot Natural disasters \cdot 360-degree images \cdot Terrestrial \cdot Aerial \cdot CNN \cdot Multidimensional analysis

1 Introduction

The environmental challenges the world faces nowadays have never been greater or more complex. Global areas that are covered by forests and urban woodlands, which comprise key parts of the global carbon cycle, are threatened by the impacts of climate change and the natural disasters that are intensified and accelerated by it. To this end, to address these impacts on people and nature, it is necessary to efficiently protect the forest ecosystems, by the occurrence of natural disasters maximizing the role of nature in absorbing and avoiding greenhouse gas emissions. Forest fires are one of the most harmful natural disasters affecting life around the world. It is worth mentioning that climate change and drier conditions have led to a marked increase of fire potential across Europe [1].

Thus, computer-based early fire warning systems that incorporate remote sensing technologies have attracted particular attention in the last years. These detection systems consist of visual cameras or multispectral/hyperspectral sensors while the main fire

© Springer Nature Switzerland AG 2020 J. Blanc-Talon et al. (Eds.): ACIVS 2020, LNCS 12002, pp. 63–74, 2020. https://doi.org/10.1007/978-3-030-40605-9_6 detection challenge lies in the modelling and detection of the chaotic and complex nature of the fire phenomenon and the large variations of flame and smoke appearance in their representations. Detection techniques are based on various color spaces [2–4], spectral [4], spatial [5] and texture characteristics [6]. More recently, deep learning methods using a variety of algorithms such as YOLO [7], Faster R-CNN networks [8] and combination of fire representations in Grassmannian space and Faster R-CNN networks [9] have been designed, implemented and widely investigated. Among the most recent existing hazard-events detection approaches [9–11] the most successfully and commonly used base networks are the AlexNet [12], VGG16 [13], GoogLeNet [14] and ResNet101 [15] networks.

However, all the previously computer-based surveillance and monitoring systems for early detection of fire suffer from some limitations. Most of the frameworks taken to date use ground fixed, PTZ or human-controlled cameras with limited field of view. Other approaches require expensive and specialized aerial hardware with complex standard protocols for data collection and complex analysis methods, limiting their potential eventual widespread use by local authorities, forest agencies and experts [16]. Furthermore, the high levels of power, the long operation times and the high computational cost for the analysis that are required for the surveillance of wide areas do not allow the free of operation intervention (i.e. need to change UAVs' - unmanned aerial vehicles' - batteries) and under bad weather conditions (i.e. windy weather) appliance of these methods. Nevertheless, nowadays, 360-degree digital camera sensors become more and more popular and they can be installed in any place even in UAVs, hence proving to be a useful tool for the surveillance of wide areas [17].

In this paper, given the urgent priority around protecting the value and potential of forest ecosystems and global forest future, use of recently introduced 360-degree sensors and a novel computer-based approach for forest health surveillance and a better-coordinated global approach is proposed. More specifically, this paper makes the following contributions:

- We propose a new framework using terrestrial and aerial 360-degree digital camera sensors in an operationally and time efficient manner, aiming to overcome the limited field of view of state-of-the-art systems and human-controlled specified data capturing.
- A novel method is proposed for fire detection combining multidimensional texture analysis using Linear Dynamical Systems (LDS) and CNN ResNet101 network. Specifically, we first identify candidate fire regions of each image dividing the extracted cubemap images into two different size blocks (rectangular patches) and modelling them using LDS. Then we feed the candidate regions into the CNN network.
- To evaluate the efficiency of the proposed methodology, we created a dataset, namely "360-FIRE", consisting of 100 images of forest and urban areas that contain synthetic fire.

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

2 Methodology

The framework of the proposed methodology is shown in Fig. 1. In this, recently introduced terrestrial and aerial 360-degree remote sensing systems are used in order to capture images with unlimited field of view. Once equirectangular images are acquired and due to the existence of distortions, they are converted to cubemap projection format images. Then, the extracted cubemap images are divided into blocks using two different sizes. Larger blocks are used in order to identify a high probability of existence of fire regions' and at the same time to reduce the computational time, whilst smaller blocks are used in order to accurately identify the candidate fire regions applying h-LDS multidimensional spatial texture analysis. Finally, the candidate fire regions are fed into a CNN *ResNet101* network for the classification to fire and non-fire regions.

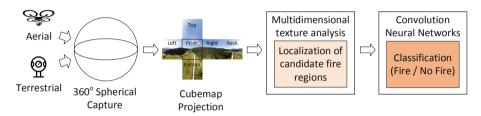


Fig. 1. The proposed methodology.

2.1 Introduction of Innovative Surveillance Schemes

To overcome the limited field of view data capturing and to achieve early and accurate detection of fire, two different formations, namely terrestrial and aerial 360-degree remote sensing systems are proposed. Terrestrial 360-degree digital cameras are ideal for areas with panoramic view while aerial 360-degree cameras are able to capture sphere images in areas where the installation of terrestrial cameras is not possible. It is worth mentioning that the required time to capture these types of data is estimated to be under a second using terrestrial cameras and under 30 s using a commercial UAV equipped with a digital camera.

Both of the proposed systems extract images in equirectangular projection (ERP) format. This native projection format is converted into cubemap projection (CMP) format in order to avoid false alarms due to the existence of distortions in the equirectangular images [18]. The CMP images consist of front, back, left, right, top and bottom images. This format is obtained by radially projecting points on the sphere to the six square faces of a cube enclosing the sphere (as illustrated in Fig. 2), and then unfolding the six faces. To this end, the spherical coordinates p(theta, phi) are calculated using the normalized coordinates u and v of the equirectangular image:

$$theta = u * 2pi \tag{1}$$

$$phi = v * pi \tag{2}$$

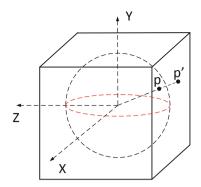


Fig. 2. Cubemap projection.

Then, the equivalent pixel coordinates p'(x', y') for each square face of CMP are estimated as follows:

$$x' = \left(\frac{w}{2}\right) \left(\frac{theta}{pi} + 1\right) \tag{3}$$

$$y' = \left(\frac{h}{2}\right) \left(\frac{phi}{pi/2} + 1\right) \tag{4}$$

where w is the width, h is the height and *theta* and *phi* are the polar coordinates of the equirectangular image. Among all projection formats, CMP is widely used in the computer graphics community. Finally, as the top image in the proposed framework represents the sky in all cases, it was not taken into account for further processing.

2.2 Localization of Candidate Fire Regions Using Multidimensional Texture Analysis

For the localization of candidate fire regions, due to the need to achieve early detection of fire events, variant sizes of fires and different distances of fires from cameras, we propose a new modeling method through the division of 360-degree images into blocks using two different sizes (Fig. 3). Initially, we divide each region into larger size blocks with size $n \times n$ (we set n = 30) which are used in order to identify areas of higher probability of fire existence. Then, depending on the probability of fire existence these blocks are divided into four smaller size blocks (n = 15) in order to accurately detect the candidate fire regions. The goal of this, is two-fold: (a) Initially, we aim to reduce the computational time of the proposed framework and (b) to increase the reliability of the candidate fire regions localization procedure. Then, we apply multidimensional texture analysis in both block sizes through the higher order linear dynamical systems. Specifically, in the proposed approach, fire can be considered as a spatially-varying visual pattern dividing each individual block in 3×3 sub-patches and considering them as a multidimensional signal evolving in the spatial domain and modelling it through the following dynamical systems:

$$x(t+1) = Ax(t) + Bv(t)$$
 (5)

$$y(t) = \overline{y} + Cx(t) + w(t) \tag{6}$$

where $x \in \mathbb{R}^n$ 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, while $\overline{y} \in \mathbb{R}^d$ is the mean value of the observation data [19, 20]. Thus, this dynamical system models both the appearance and dynamics of the observation data, represented by *C* and *A*, respectively (Fig. 4). We then represent each sub-patch with a third-order tensor Y and apply a higher order Singular Value Decomposition to decompose the tensor:

$$Y = S \times_1 U_{(1)} \times_2 U_{(2)} \times_3 U_{(3)}$$
(7)

where $S \in \mathbb{R}^{n \times n \times c}$ is the core tensor, while $U_{(1)} \in \mathbb{R}^{n \times n}$, $U_{(2)} \in \mathbb{R}^{n \times n}$ and $U_{(3)} \in \mathbb{R}^{c \times c}$ are orthogonal matrices containing the orthonormal vectors spanning the column space of the matrix and \times_j denotes the *j*-mode product between a tensor and a matrix. Since the columns of the mapping matrix *C* of the stochastic process need to be orthonormal, we can consider $C = U_{(3)}$ and

$$X = S \times_1 U_{(1)} \times_2 U_{(2)}$$
(8)

The transition matrix A can be estimated using least squares as follows:

$$A = X_2 X_1^T \left(X_1 X_1^T \right)^{-1}$$
(9)

where $X_1 = [x(2), x(3), \dots, x(t)]$ and $X_2 = [x(1), x(2), \dots, x(t-1)]$.

Assuming that the tuple M = (A, C) describe each sub-patch, we estimate the finite observability matrix of each dynamical system, $O_m^T(M) = \left[C^T, (CA)^T, (CA^2)^T, \dots, (CA^{m-1})^T\right]$ and we apply a Gram-Schmidt orthonormalization procedure [21], i.e., $O_m^T = GR$, in order to represent each descriptor as a Grassmannian point, $G \in \mathbb{R}^{m \times T \times 3}$ [5].

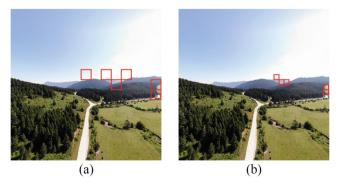


Fig. 3. Identification of candidate fire regions using: (a) larger size blocks and (b) smaller size blocks.

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, $\{m_i\}_{i=1}^r = \{m_1, m_2, ..., m_r\}$, 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, *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, ..., r.

$$\overline{V} = \{u_i\}_{i=1}^r = \{u_1, \dots, u_r\}$$
(10)

or

$$\overline{V} = \begin{cases} \sum_{v_j \text{ such that}} \left(v_j - m_1 \right), \dots, \sum_{v_j \text{ such that}} \left(v_j - m_r \right) \\ m_1 = NN(v_j) & m_r = NN(v_j) \end{cases}$$
(11)

while the final VLAD representation is determined by the L2-normalization of vector \overline{V} :

$$\overline{V}_{Euclidean} = \overline{V} / \left\| \overline{V} \right\|_2 \tag{12}$$

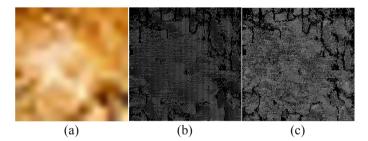


Fig. 4. Visualization of a: (a) candidate fire block, (b) transition matrix A and (c) mapping matrix C.

Finally, by ranking the similarities across all blocks, the majority rule of the *s* labels with the minimum distances is adopted in order to classify the examined block into candidate and non-candidate fire regions.

2.3 Fire Detection Using Convolution Neural Networks

After candidate fire regions are extracted, they are fed into a Convolution Neural Network (CNN). CNNs are one of the state-of-the-art deep learning approaches for hazard events detection. Although CNN architectures require large amount of training data, they are able to automatically learn very strong features. Thus, inspired by previous hazard events detection systems [9, 10] we chose to deploy ResNet101 feature extractor. With the

advent of this model, researchers have developed deepened network structures that do not increase computational complexity. In the proposed methodology we use similar architecture to the original, aiming to train the parameters using fire focused images in order to solve the fire detection problem more effectively. Furthermore, we modify the number of neurons to two in the final layer of our architecture, enabling classification into fire and non-fire.

3 Experimental Results

Through the experimental evaluation we want to demonstrate the superiority of the proposed framework against other state of the art approaches and we aim to show that the proposed methodology improves the detection of fires.

To evaluate the efficiency of the proposed methodology, we created a dataset, namely "360-FIRE", consisting of 100 images of forest and urban areas that contain synthetic fire (90 of them consist of fire events and 10 of them are fireless 360-degree images). To the best of our knowledge and as 360-degree digital camera sensors are newly introduced type of cameras, there is not any dataset consisting of 360-degree images that contain fire. Thus, in order to create the dataset, we captured 360-degree images in different environments, and we used higher order SVD analysis (as shown in Eq. 9) in order to synthesize artificial flames [24]. Specifically, we represented synthetic video frames solving linear system equations and estimating the tensor generated at time k when the system is driven by random noise V [19]. Then the synthesized data was adapted to the 360-degree images and the size of fires was suitably adjusted with regards to the distance and the assumed start time of fire. For each captured image, we created 5 different fire events corresponding to different size of fires or fire locations. To evaluate the performance of the proposed methodology true positive, false negative, true negative, false positive rates and F-Score were used.



Fig. 5. CFDB and PASCAL VOC 2007 dataset training images containing actual fires and firecoloured (non-fire) objects. (color figure online)

For the training of the proposed method for the localization of candidate fire regions using multidimensional texture analysis, we used the Corsican Fire Database (CFDB) that contain fire events and the PASCAL Visual Object Classes (VOC) 2007 dataset that contain fire-coloured objects (Fig. 5). Additionally, the CNN network was trained using the same datasets. Dataset sample images are shown in the Fig. 6. The implementation code of the proposed structure was written in Matlab and all calculations were performed

on a 4 GB GPU and on a 6-core processor Intel Core i7-8750H, CPU 2.2 GHz. It is worth mentioning that in to have a fair comparison in our experiments we used the same training and testing set. In each case an image is labeled as a fire 360-degree image if it contains at least one fire image region.

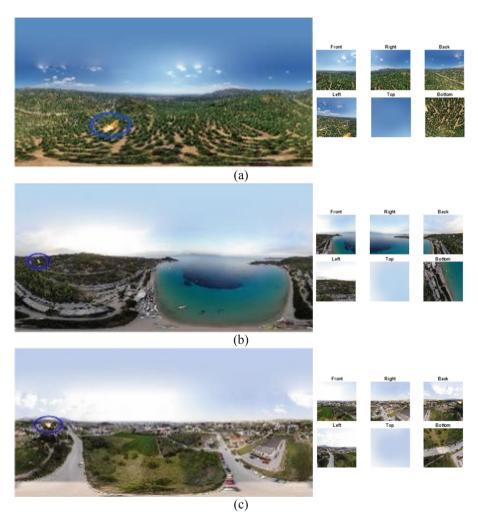


Fig. 6. "360-FIRE" dataset 360-degree images (left: equirectangular images, right: cubemap images) containing fires in different environments: (a) forest ecosystem, (b) forest ecosystem by the sea and (c) semi-urban environment.

Regarding the detection power of the proposed algorithm, we compared the proposed methodology to a spatial texture analysis method [20, 25] on its own as well as combined with a color localization method [5] and two Faster R-CNN [13, 15] architectures. As shown in Table 1, we compared the performance of the proposed approach firstly with a predefined color distribution for the localization of candidate blocks in combination

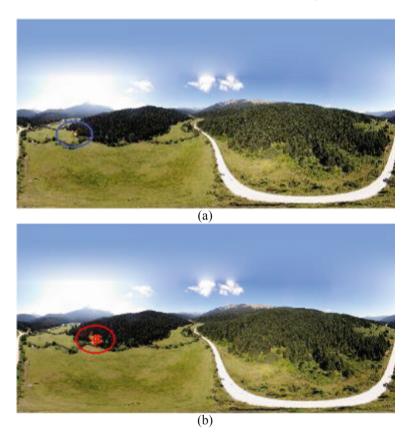


Fig. 7. Equirectangular images: (a) False negative of a long-distance fire, (b) True detection of a long-distance fire.

with the h-LDS approach for fire detection, secondly with the h-LDS approach playing the localization and detection role and finally with two Faster R-CNNs with VGG16 and Resnet101 base networks, achieving true positive rates 88.9%, 92.2%, 94.4% and 95.6% and F-score rates 77.7%, 83%, 88.1% and 88.7%, respectively. As depicted in Table 1, our method offers improved true positive rate (96.7%) and F-score (93.5%) compared to the second higher rates of 95.6% and 88.7% yield by the Faster R-CNN Resnet101 network. Furthermore, the proposed methodology increases true negative rates and reduces false negative and false positive rates.

Experimental results show that the proposed approach retains high true positive rates, while simultaneously significantly reducing false positives. This can be explained by the fact that the localization procedure eliminates many regions that are misclassified as fire or non-fire by the Faster R-CNN networks. Furthermore, it is obvious that the proposed methodology which includes dividing images into two different block sizes and extracting h-LDSs in order to estimate candidate fire regions is more efficient than the color analysis approaches, while CNN ResNet101 outperforms the discrimination ability of h-LDSs.

Method	True positive rates	False negative rates	True negative rates	False positive rates	F-score
Color & Spatial Texture Analysis (h-LDS)	88.9%	11.1%	60.0%	40.0%	77.7%
Spatial Texture Analysis (h-LDS)	92.2%	7.8%	70.0%	30.0%	83.0%
Faster R-CNN (VGG16)	94.4%	5.6%	80.0%	20.0%	88.1%
Faster R-CNN (ResNet101)	95.6%	4.4%	80.0%	20.0%	88.7%
Proposed	96.7%	3.3%	90.0%	10.0%	93.5%

Table 1. Comparison results of various fire detection approaches using cubemap projection.

Table 2. Comparison results of the proposed fire detection methodology using different projections.

Method	F-score
Equirectangular projection (ERP)	81.8%
Cubemap projection (CMP)	93.5%

However, in the proposed methodology some false negatives are noticed in small fires in long distance (Fig. 7a). In contrast, larger fires in shorter or same distance were accurately detected (Fig. 7b).

In Table 2, we present experimental results of the proposed methodology against the use of that equirectangular projection formats. More specifically, the use of cubemap projection formats achieves 11.7% higher detection rates.

Finally, we performed a computational speed test that is a crucial factor for early hazard events detection applications. Specifically, we estimated the time that is required for the processing of equirectangular images and cubemap images. Results show that the multi-thread processing of cubemaps requires 60% less computational time.

4 Conclusion

In this paper, we presented a novel framework that combines the newly introduced 360-degree digital camera sensors and modern signal and image processing techniques. This enables the near real-time environmental data acquisition, assessment, processing and analysis for the ultimate goals of ecosystem protection and forest and urban areas

monitoring. The proposed framework will allow experts and scientists to achieve a bettercoordinated global approach that will contribute to the limitation of negative impacts of climate change on forest ecosystems, air, and timber supplies.

In the future, a system for autonomous operation of UAV's which require one charge per day will be developed, in order to perform periodic flights every 30 min for the surveillance of wide areas. Additionally, in order to assess the effectiveness of the proposed we aim to extend our dataset using more data from a variety of urban, rural and forest areas. Furthermore, indoor images from buildings of cultural heritage will be captured and used. Finally, our goal is to install cameras at critical observation points for long periods in order to apply the proposed method in real hazard events.

References

- European Environment Agency: Forest Fires (2019). https://www.eea.europa.eu/data-andmaps/. Accessed 12 June 2019
- 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)
- Dimitropoulos, K., Tsalakanidou, F., Grammalidis, N.: 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 (2012)
- Grammalidis, N., et al.: A multi-sensor network for the protection of cultural heritage. In: 19th European Signal Processing Conference, pp. 889–893 (2011)
- Barmpoutis, P., Dimitropoulos, K., Grammalidis, N.: Real time video fire detection using spatio-temporal consistency energy. In: 10th IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 365–370 (2013)
- Dimitropoulos, K., Barmpoutis, P., Grammalidis, N.: Spatio-temporal flame modeling and dynamic texture analysis for automatic video-based fire detection. IEEE Trans. Circuits Syst. Video Technol. 25(2), 339–351 (2014)
- Shen, D., Chen, X., Nguyen, M., Yan, W.Q.: Flame detection using deep learning. In: 2018 4th International Conference on Control, Automation and Robotics, pp. 416–420 (2018)
- Zhang, Q.X., Lin, G.H., Zhang, Y.M., Xu, G., Wang, J.J.: Wildland forest fire smoke detection based on faster R-CNN using synthetic smoke images. Procedia Eng. 211, 441–446 (2018)
- Barmpoutis, P., Dimitropoulos, K., Kaza, K., Grammalidis, N.: Fire detection from images using faster R-CNN and multidimensional texture analysis. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 8301–8305 (2019)
- Giannakeris, P., Avgerinakis, K., Karakostas, A., Vrochidis, S., Kompatsiaris, I.: People and vehicles in danger-a fire and flood detection system in social media. In: 2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop, pp. 1–5 (2018)
- Yang, L., Cervone, G.: Analysis of remote sensing imagery for disaster assessment using deep learning: a case study of flooding event. Soft Comput. 23, 13393–13408 (2019)
- Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
- Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
- Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2014)

- He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
- Chowdary, V., Gupta, M.K.: Automatic forest fire detection and monitoring techniques: a survey. In: Singh, R., Choudhury, S., Gehlot, A. (eds.) Intelligent Communication, Control and Devices. Advances in Intelligent Systems and Computing, vol. 624, pp. 1111–1117. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-5903-2_116
- Zia, O., Kim, J.H., Han, K., Lee, J.W.: 360° panorama generation using drone mounted fisheye cameras. In: Proceedings of the IEEE International Conference on Consumer Electronics, pp. 1–3, January 2019
- 18. Kim, J.H., et al.: U.S. Patent Application No. 15/433,505 (2018)
- Doretto, G., Chiuso, A., Wu, Y.N., Soatto, S.: Dynamic textures. Int. J. Comput. Vision 51(2), 91–109 (2003)
- Dimitropoulos, K., Barmpoutis, P., Kitsikidis, A., Grammalidis, N.: Classification of multidimensional time-evolving data using histograms of Grassmannian points. IEEE Trans. Circuits Syst. Video Technol. 28(4), 892–905 (2016)
- 21. Arfken, G.: Gram-schmidt orthogonalization. Math. Methods Phys. 3, 516–520 (1985)
- Jégou, H., Douze, M., Schmid, C., Pérez, P.: Aggregating local descriptors into a compact image representation. In: CVPR 2010-23rd IEEE Conference on Computer Vision & Pattern Recognition, pp. 3304–3311 (2010)
- Kantorov, V., Laptev, I.: 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. Costantini, R., Sbaiz, L., Susstrunk, S.: Higher order SVD analysis for dynamic texture synthesis. IEEE Trans. Image Process. **17**(1), 42–52 (2007)
- Barmpoutis, P., Dimitropoulos, K., Barboutis, I., Grammalidis, N., Lefakis, P.: Wood species recognition through multidimensional texture analysis. Comput. Electron. Agric. 144, 241– 248 (2018)