EARLY WILDFIRE SMOKE DETECTION BASED ON MOTION-BASED GEOMETRIC IMAGE TRANSFORMATION AND DEEP CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORKS

Süleyman Aslan[⋆] *Uğur Güdükbay*[⋆] *B. Uğur Töreyin*[‡] *A. Enis Çetin* [†§]

[⋆] Dept. of Computer Eng., Bilkent University, Ankara, Turkey
[†] Dept. of Electrical and Electronics Eng., Bilkent University, Ankara, Turkey
[§]Dept. of Electrical and Computer Eng., University of Illinois at Chicago, Chicago, IL, USA
[‡] Informatics Institute, Istanbul Technical University, Istanbul, Turkey

ABSTRACT

Early detection of wildfire smoke in real-time is essentially important in forest surveillance and monitoring systems. We propose a vision-based method to detect smoke using Deep Convolutional Generative Adversarial Neural Networks (DCGANs). Many existing supervised learning approaches using convolutional neural networks require substantial amount of labeled data. In order to have a robust representation of sequences with and without smoke, we propose a two-stage training of a DCGAN. Our training framework includes, the regular training of a DCGAN with real images and noise vectors, and training the discriminator separately using the smoke images without the generator. Before training the networks, the temporal evolution of smoke is also integrated with a motion-based transformation of images as a pre-processing step. Experimental results show that the proposed method effectively detects the smoke images with negligible false positive rates in real-time.

Index Terms— Wildfires, smoke detection, Deep Convolutional Generative Adversarial Networks (DCGAN)

1. INTRODUCTION

Wildfires are one of the most harmful hazards in rural areas. They may spread fast and cause substantial damages to flora, properties and human life. Hence, immediate and accurate wildfire detection plays instrumental role in fighting wildfires.

Among different approaches, the use of visible-range video captured by surveillance cameras are particularly convenient for wildfire detection, as they can be deployed and operated in a cost-effective manner [1]. One of the main challenges is to provide a robust vision based detection system with negligible false positive rates, while securing rapid response. If the flames are visible, this may be achieved by analyzing the motion and color clues of a video in wavelet domain [2], [3]. Similarly, wavelet based contour analysis [4] can be used for detection of possible smoke regions. Modeling various spatio-temporal features such as color and flickering, and dynamic texture analysis [5] have been shown to be able to detect fire, as well. We developed smoke and flame detection algorithms using wavelets, support vector machines, Markov models, region covariance, and co-difference matrices in the past [6]. An important feature of the wildfire detection algorithms that we developed in the past is that, they not only use spatial information, but also the temporal information [6], [7]. We focus on wildfire smoke detection, rather than flame detection. This is mainly due to the fact that smoke rises above the crowns of trees, and it has a higher chance of falling into the viewing range of cameras monitoring the forest.

Deep convolutional neural networks (DCNN) achieve superb recognition results on a wide range of computer vision problems [8], [9]. Deep neural network based wildfire detection algorithms using regular cameras have been developed by many researchers including us in recent years but none of these algorithms can handle false alarms due to cloud shadows and fog [10], [11]. Radford et al. [12] demonstrate that a class of convolutional neural networks, namely, Deep Convolutional Generative Adversarial Networks (DCGANs), can learn general image representations on various image datasets.

We propose a two-stage training approach for a DCGAN in such a way that the discriminator is utilized to distinguish ordinary image sequences without smoke from wildfire smoke. Our first contribution is the development of a discriminator network classifying regular wilderness images from wildfire images. We employ the discriminator network of the DCGAN as a classifier.

One important aspect of wildfire smoke that we also exploit is its evolution in time. We integrate the temporal progress of smoke by using a motion-based image transfor-

A.E. Çetin is on leave from Bilkent University and his work is partially funded by NSF with grant number 1739396 and NVIDIA Corporation. B.U. Töreyin's work is in part funded by TÜBİTAK 114E426 and İTÜ BAP MGA-2017-40964.

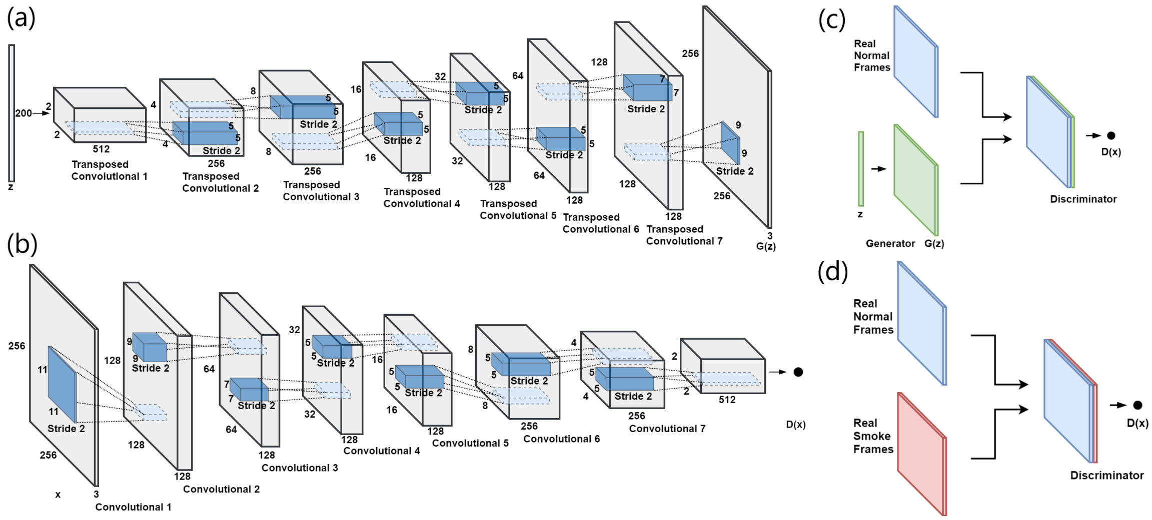

Fig. 1. The architecture of DCGAN: (a) generator network, (b) discriminator network, (c) the first stage of training, and (d) the second stage of training.

mation before training the networks. This constitutes our second contribution.

The remainder of the paper is organized as follows. In section 2, the proposed wildfire smoke detection method is described. Experimental results are presented in Section 3. The paper is concluded in the last section.

2. METHOD

The proposed wildfire smoke detection method is presented in this section. The method is based on a DCGAN structure accepting images with size 256×256 px. We use seven transposed convolutional layers for the generator, and seven convolutional layers for the discriminator with filters of varying sizes and channels. The architecture of DCGAN and the training framework are given in Figure 1.

We first train the DCGAN using images without smoke and noise distribution z. The discriminator part of the DCGAN learns a representation for ordinary wilderness video scenes and distinguishes smoke, because images containing smoke are not in the training set. Then, we refine and retrain the discriminator without generator network, where regular video images obtained from the surveillance cameras constitute the "real" training data and actual smoke images correspond to generated data. Training the DCGAN using both the regular data and noise vector z makes the recognition system more robust compared to a generic CNN structure. Moreover, the second stage of training increases the recognition accuracy.

In our model, for the training of the networks, we use instance normalization [13] before each layer in the discriminator network, and batch normalization [14] before each layer in the generator network. To initialize the layers we apply "MSRA" initialization [15]. Dropout layers [16] are added, as well, to address overfitting. Finally, we use the Adam optimizer for stochastic optimization [17]. The representations of algorithms are supported by TensorFlow system [18].

2.1. Motion-based Geometrical Image Transformation

As a pre-processing step, we apply transformations to the frames captured by the cameras. In a wildfire, smoke can usually be distinguished by its characteristic evolution compared to other moving objects. In order to exploit this temporal behavior, we first compute the estimated motion using Farnebäcks algorithm [19], then we apply a geometrical transformation as follows (see Figure 3)

$$T(k,l) = S(k - f_k(k,l), l - f_l(k,l)), \qquad (1)$$

where $T(k,l)$ ($S(k,l)$) is the pixel at position (k,l) in the resulting transformed (source) image, $f_k(k,l)$ ($f_l(k,l)$) is the estimated motion along horizontal-k (vertical-l) axis at position (k,l).

Issues, such as, extrapolation of non-existing pixels and interpolation of pixel values are handled by implementations in the OpenCV library [20]. As for the motion estimation, we ignore abrupt motions, such as, fast movements or rotations of the camera. Examples of transformed smoke frames are shown in Figure 2.

2.2. Proposed GAN-type Discriminator Network

Wildfire smoke has no particular shape or specific feature as human faces, cars, and so on. Therefore, it is more suitable to treat smoke as an unusual event or an anomaly in the observed scene.

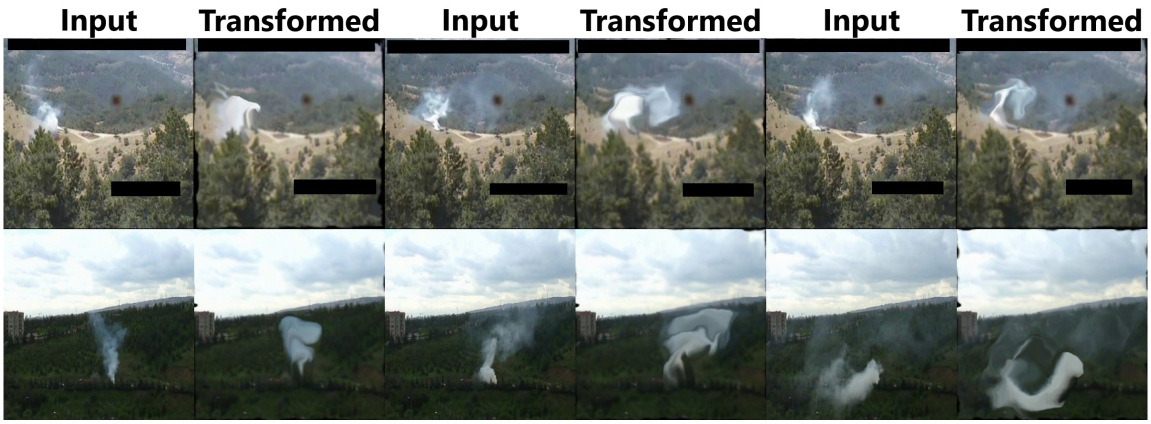

Fig. 2. Examples of transformed frames.

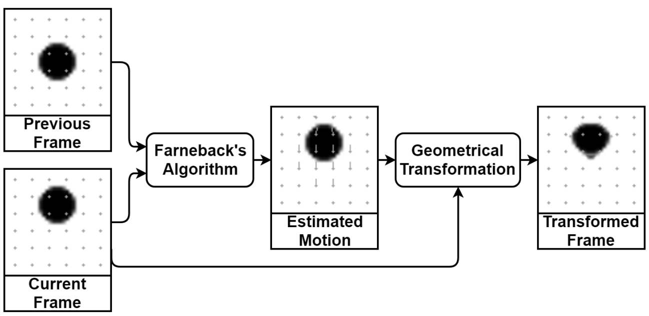

Fig. 3. The illustration of the motion-based geometrical image transformation.

The DCGAN structure is utilized to distinguish regular camera views from wildfire smoke. The discriminator part of the GAN produce probability values above 0.5 for normal wilderness video scenes and below 0.5 for images containing smoke, because smoke images are not in the training set. In the second stage of training, we refine and retrain the GAN using the gradient given in (3).

In standard GAN training, the discriminator D that outputs a probability value is updated using the stochastic gradient

$$SG_1 = \nabla_{\theta_d} \frac{1}{M} \sum_{i=1}^{M} (\log D(x_i) + \log(1 - D(G(z_i)))), \quad (2)$$

where x_i and z_i are the i-th regular image data and noise vector, respectively, and G represents the generator that generates a "fake" image according to the input noise vector z_i; the vector θ_d contains the parameters of the discriminator. After this stage, the generator network G is "adversarially" trained, as in [8]. During the first round of training we do not include any smoke videos. This GAN is able to detect smoke, because smoke images are not in the training set. To increase the recognition accuracy, we perform a second round of training by fine-tuning the discriminator using the stochastic gradient

$$SG_2 = \nabla_{\theta_d} \frac{1}{L} \sum_{i=1}^{L} (\log D(x_i) + \log(1 - D(y_i))), \quad (3)$$

where y_i represents the i-th image containing wildfire smoke. The number of smoke image samples, L, is much smaller than the size of the initial training set, M, containing regular forest and wilderness images, because wildfires are rare events. In the refinement stage characterized by (3), we do not update the parameters of the generator network of GAN, because we do not need to generate any artificial images in this stage of training.

3. EXPERIMENTAL RESULTS

In our experiments, we use 40 video clips containing no smoke frames with a duration of 4 hours 52 minutes, and 29 video clips containing only smoke frames with a duration of 3 hours and 46 minutes. For each smoke video, there is a corresponding normal video for generator network to learn, however, not all normal videos do have a corresponding smoke video.

Throughout the experiments, we first apply motion-based geometrical image transformation. For that purpose, at every second, we sample 10 previous frames at equal intervals, then we calculate the estimated motion and obtain the transformed frame. In effect, we acquire one frame per second to be input to the network. Each one of these frames contains an integrated history of the ten most recent frames. Since the video clips in our dataset differ greatly in length (from 20 seconds to 40 minutes), we normalize the number of frames by randomly discarding frames from longer videos and duplicating frames of shorter ones. That way, the dataset, composed of forty-thousand-frames in total, becomes one containing similar-length clips.

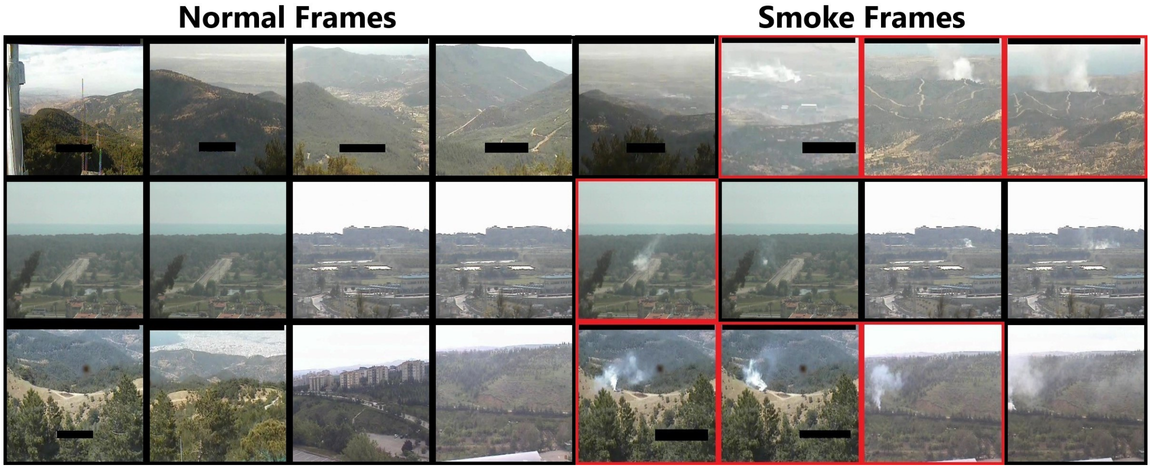

Fig. 4. Examples of frame-based classification results. Red border indicates that smoke is detected in that frame.

After this procedure, we split the data into training, validation, and test sets with a ratio of 3:1:1. We pick the parameters and stop training the network based on its performance on the validation set, then report the final results obtained on the test set.

We first evaluate the proposed method in terms of frame-based results. We compare our model by excluding the contributions one by one and training the network again with the same parameters. A few frame-based classification examples are presented in Figure 4. Our approach targets at reducing the false positive rate, while keeping the hit-rate as high as possible. Results indicate that, our approach achieves best results on the test set (cf. Table 1). Without the refinement stage, smoke detection rates are smaller however it can still be useful when there are no labeled smoke frames. For the motion-based transformation, the difference is mainly in hit-rates, and if a DCNN is used without adversarial training, the model will be more susceptible to false positives.

Table 1. Obtained true negative rate (TNR) and true positive rate (TPR) values on test set for frame-based evaluation.

Method	**TNR (%)**	**TPR (%)**
Our method	99.45	86.23
Transformation excluded	98.70	83.33
Refinement excluded	95.10	62.56
Transformation and refinement excluded	93.94	60.16
Adversarial training excluded	98.07	84.10
Adversarial training and transformation excluded	97.39	81.43

We also evaluate the approach in terms of video-based results. For the video-based evaluation, we classify a video as a smoke video, if, at least, one frame is detected as smoke. We train different versions of the network, by up(down)-weighting the cost of a false positive relative to a false negative, to trade-off specificity and sensitivity. The results indicate that a false-positive rate of 2.5% is achieved corresponding to a 6.9% miss rate (cf. Table 2). On the other hand, the proposed method has a hit-rate of 89.67% without issuing any false alarms (cf. Table 2).

Table 2. Video-based results for our method.

	TNR (%)	**TPR (%)**
Up-weighted false positives	100.00	89.67
Unweighted	97.50	93.10
Down-weighted false positives	87.50	100.00

4. CONCLUSION

We propose a wildfire smoke detection method using motion-based geometrical image transformation and DCGANs. By treating smoke as an unusual event, we develop a two-stage DCGAN training approach. Spatio-temporal dynamics of smoke event are acquired using motion-based geometric image transformation and represented within a single image accounting for ten consecutive frames.

Results suggest that the proposed method achieves low false alarm rates while keeping the detection rate high. The proposed approach may be utilized to detect other anomalous events in forests, such as, flames or people in restricted zones.

5. REFERENCES

[1] A. E. Çetin, K. Dimitropoulos, B. Gouverneur, N. Grammalidis, O. Günay, Y. H. Habibolu, B. U. Töreyin, and S. Verstockt, "Video fire detection–review," *Digital Signal Processing*, vol. 23, no. 6, pp. 1827–1843, 2013.

[2] Y. Dedeoğlu, B. U. Toreyin, U. Güdükbay, and A. E. Cetin, "Real-time fire and flame detection in video," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 2005, vol. 2 of *ICASSO'05*, pp. ii–669.

[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*, vol. 27, no. 1, pp. 49–58, 2006.

[4] B. U. Toreyin, Y. Dedeoğlu, and A. E. Cetin, "Contour based smoke detection in video using wavelets," in *Proceedings of the European Signal Processing Conference*, 2006, EUSIPCO 2006.

[5] 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*, vol. 25, no. 2, pp. 339–351, 2015.

[6] A. E. Çetin, B. Merci, O. Günay, B. U. Töreyin, and S. Verstockt, Eds., *Methods and Techniques for Fire Detection*, Academic Press, Oxford, 2016.

[7] Y. H. Habiboğlu, O. Günay, and A. E. Çetin, "Covariance matrix-based fire and flame detection method in video," *Machine Vision and Applications*, vol. 23, no. 6, pp. 1103–1113, Nov 2012.

[8] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 1, pp. 436444, May 2015.

[10] O. Günay, B. U. Töreyin, K. Köse, and A. E. Çetin, "Entropy-functional-based online adaptive decision fusion framework with application to wildfire detection in video," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2853–2865, May 2012.

[11] Y. Zhao, J. Ma, X. Li, and J. Zhang, "Saliency detection and deep learning-based wildfire identification in UAV imagery," *Sensors*, vol. 18, no. 3, pp. Article No. 712, 19 pages, 2012.

[12] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *CoRR*, vol. abs/1511.06434, 2015.

[13] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *CoRR*, vol. abs/1607.08022, 2016.

[14] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR*, vol. abs/1502.03167, 2015.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, ICCV'15, pp. 1026–1034.

[16] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[17] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.

[18] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al., "Tensorflow: a system for large-scale machine learning," in *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, 2016, OSDI'16, pp. 265–283.

[19] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Proceedings of the 13th Scandinavian Conference on Image Analysis*, Berlin, Heidelberg, 2003, SCIA'03, pp. 363–370, Springer-Verlag.

[20] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools,* *http://www.drdobbs.com/open-source/the-opencv-library/184404319*, 2000.