Multi-channel Information Processing for Fire Detection

Elizabeth Ren, Nour Zalmai, Hans-Andrea Loeliger
Signal and Information Processing Laboratory, ETH Zurich, Switzerland

Mathias Stäger
Siemens Schweiz AG, Building Technologies Division, Zug, Switzerland

Abstract

This paper provides a novel approach for fire detection and classification based on multi-channel signals of optical smoke detectors. The algorithm is flexible in number of optical channels as well as fire and nuisance types. Furthermore, a method that selects the smallest number of channels according to a classification goal is developed.

From a simplified physical model of light scattering, a linear increase of the signals for a short period of time after the onset of a fire or nuisance event is expected. Moreover, the relation of the linear increase of signals from different channels is directly characterized by the type of the fire or nuisance. This signal pattern can be described by a two-sided linear state space model that is efficiently fit to the measured signals. By evaluating the goodness of such a model fit at every time step with a likelihood ratio, we can robustly detect the start of a fire or nuisance event. When an event is detected, the estimated signal slopes are extracted as features for classification. The simplified physical model predicts that all extracted slopes from a given type should lie on a line. Thus, each fire and nuisance type is characterized by a line which can be learned from a labeled data set of recordings. Classification then reduces to assigning features to the nearest line. In addition, with these representative lines, channel selection can be performed by analyzing the class spread and the allowed error margin.

The method showed promising classification results on several data sets and allowed selection of channels for a classification task.

Keywords: smoke detection algorithm, fire and nuisance classification, optical smoke detectors with multiple channels, channel selection

Introduction

Optical smoke detectors rely on the physical phenomenon of light scattering by smoke particles to detect fires. However, nuisances also produce particles that scatter light. With only one optical channel,

nuisances may be barely distinguishable from fires. This can be circumvented by incorporating multiple optical channels with variations in scattering angle, wavelength and polarization of light in the detectors. For design of a well-functioning detector, the channels need to be carefully selected in order to robustly detect and classify different fire and nuisance categories specified by the application.

A rudimentary detection and classification approach consists in extracting a feature from each signal sample separately and classifying the sample accordingly. This approach is however highly sensitive to noise and may lack robustness due to neglect of temporal information, such as changing physical properties of a fire over time. Moreover, it cannot be directly adapted to a different choice of channels and classification goal.

This paper provides an algorithm that, while being flexible in number of channels and classification purpose, fully exploits data from multiple optical channels to detect and classify different types of fires and nuisances. The proposed method focuses on the starting time period of a fire or nuisance event, for which a simplified physical model can be derived. Based on this algorithm, a method to pre-select the smallest subset of channels needed for a given classification goal is described.

Physical Model for Optical Smoke Detectors

For a short time period after the start of a fire or nuisance event, the physical model of light scattering for one optical channel can be simplified with the following assumptions:

- A fire or nuisance event type T (e.g., EN 54-7 TF2 or a type of dust) produces a characteristic type and size distribution of particles.
- The signal y is proportional to the concentration of scattering particles with a coefficient R_T that is constant for this short time period. R_T is determined by the particle distribution, the wavelength and polarization of light and the scattering angle of the channel.
- The particle concentration grows approximately linearly in time with a coefficient α which depends on several factors, such as the environment or the amount of material involved.

For M optical channels ($M \in \mathbb{N}$), the recorded signals of all channels at time t are stacked into a vector $\boldsymbol{y}(t) \in \mathbb{R}^M$. For a short period of time at the beginning of an event, the signals rise linearly: $\boldsymbol{y}(t) \cong \alpha \boldsymbol{R}_T t$ with a slope per channel that is determined by α and the corresponding coordinate of $\boldsymbol{R}_T \in \mathbb{R}^M$ that is characterized by the type T of the event.

Multi-channel Signal Model

Following the physical model, each discrete-time signal from an optical channel observed around the start of an event can be described by a two-sided line: a flat left-sided line approximating the time before an

event followed by a right-sided line with a positive slope for a short time during the event (see Fig. 1). The two-sided lines (one per channel) centered around a time step k can be parameterized with two second order autonomous state space models [1], [2]:

$$\tilde{\boldsymbol{y}}_i(\boldsymbol{C}_\mathrm{p}, \boldsymbol{C}_\mathrm{f}) = \begin{cases} \boldsymbol{C}_\mathrm{p}\boldsymbol{A}_\mathrm{p}^{k-i}\boldsymbol{s}_\mathrm{p}, & \text{if } i \leq k \\ \boldsymbol{C}_\mathrm{f}\boldsymbol{A}_\mathrm{f}^{i-k}\boldsymbol{s}_\mathrm{f}, & \text{if } i > k, \end{cases} \quad \text{(Eq. 1)}$$

where the vector $\tilde{\boldsymbol{y}}_i \in \mathbb{R}^M$ holds the values of the two-sided lines for all channels at time step i. The fixed state transition matrices $\boldsymbol{A}_\mathrm{p}$, $\boldsymbol{A}_\mathrm{f}$ and state vectors $\boldsymbol{s}_\mathrm{p}$, $\boldsymbol{s}_\mathrm{f}$ are:

$$\boldsymbol{A}_\mathrm{p}^{-1} = \boldsymbol{A}_\mathrm{f} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{(Eq. 2)}$$

$$\boldsymbol{s}_\mathrm{p} = \boldsymbol{s}_\mathrm{f} = [1 \;\; 0]^\mathrm{T}. \quad \text{(Eq. 3)}$$

The observation matrix $\boldsymbol{C}_\mathrm{p} = [\boldsymbol{a}_\mathrm{p} \; \boldsymbol{b}_\mathrm{p}] \in \mathbb{R}^{M\times 2}$ contains the left-sided slopes $\boldsymbol{a}_\mathrm{p} \in \mathbb{R}^M$ and offsets $\boldsymbol{b}_\mathrm{p} \in \mathbb{R}^M$ of the channels while the observation matrix $\boldsymbol{C}_\mathrm{f} = [\boldsymbol{a}_\mathrm{f} \; \boldsymbol{b}_\mathrm{f}] \in \mathbb{R}^{M\times 2}$ contains the right-sided slopes $\boldsymbol{a}_\mathrm{f}$ and offsets $\boldsymbol{b}_\mathrm{f}$.

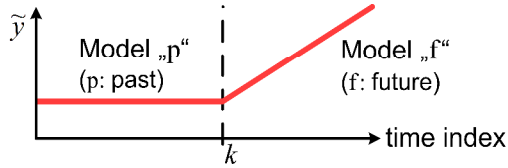

Fig. 1. Two-sided line model for one channel (here: $\boldsymbol{b}_\mathrm{p} = \boldsymbol{b}_\mathrm{f}$).

At a time step k we locally fit the model $\tilde{\boldsymbol{y}}$ with the recorded signal $\boldsymbol{y}$. This amounts to optimizing the parameters of the two-sided lines (i.e. $\boldsymbol{C}_\mathrm{p}$ and $\boldsymbol{C}_\mathrm{f}$) by minimizing a cost function which is defined as an exponentially weighted squared error sum:

$$J_k(\boldsymbol{C}) = \sum_{i=1}^{k} \gamma_\mathrm{p}^{k-i} \|\boldsymbol{y}_i - \tilde{\boldsymbol{y}}_i(\boldsymbol{C})\|^2 + \sum_{i=k+1}^{k+\Delta} \gamma_\mathrm{f}^{i-k} \|\boldsymbol{y}_i - \tilde{\boldsymbol{y}}_i(\boldsymbol{C})\|^2, \quad \text{(Eq. 4)}$$

where $\boldsymbol{C} = [\boldsymbol{C}_\mathrm{p} \; \boldsymbol{C}_\mathrm{f}]$ holds the parameters of the two-sided lines. The fit is made localized in time with the two-sided exponential window of decay $0 < \gamma_\mathrm{p} < 1$ and $0 < \gamma_\mathrm{f} < 1$ that gradually reduces the influence of samples further away from time step k. Due to real-time constraints, the right-sided lines are fit with a finite number of future samples, i.e. a time delay Δ, while the left-sided lines are fit with all past samples. The cost function in Eq. 4 can be parameterized as:

$$J_k(\boldsymbol{C}) = \kappa_k - 2\mathrm{tr}(\boldsymbol{C}\boldsymbol{\xi}_k) + \mathrm{tr}(\boldsymbol{C}\boldsymbol{W}_k\boldsymbol{C}^\mathrm{T}), \quad \text{(Eq. 5)}$$

with the quantities $\kappa_k \in \mathbb{R}$, $\boldsymbol{\xi}_k \in \mathbb{R}^{4\times M}$ and $\boldsymbol{W}_k \in \mathbb{R}^{4\times 4}$. This parameterization allows an efficient computation of the cost function since the quantities $\kappa_k, \boldsymbol{\xi}_k, \boldsymbol{W}_k$ can be computed with forward and backward recursions in time (see [2]).

Event Detection

The goal is to detect a fire or nuisance soon after the beginning of the event. For that purpose, at every time step k, we evaluate the likelihood of such an event start by comparing the cost of two model fits. The first model accounts for the expected signal form when an event starts:

H_1 – *start of fire or nuisance event:* signal slopes change from zero ($\boldsymbol{a}_\mathrm{p} = \boldsymbol{0}$) to rising ($\boldsymbol{a}_\mathrm{f} > \boldsymbol{0}$) while the offsets are identical ($\boldsymbol{b}_\mathrm{p} = \boldsymbol{b}_\mathrm{f}$).

The second model accounts for the expected signal form when there is no starting event:

H_0 – *no start of fire or nuisance event:* no change from left-sided to right-sided line ($\boldsymbol{a}_\mathrm{p} = \boldsymbol{a}_\mathrm{f}$, $\boldsymbol{b}_\mathrm{p} = \boldsymbol{b}_\mathrm{f}$).

These two models are compared by computing a log-likelihood ratio (LLR) that, as derived in [2], simplifies to a closed-form expression:

$$\mathrm{LLR}_k = -\frac{1}{2}\ln\left(\frac{\min\limits_{\boldsymbol{C}\in H_1} J_k(\boldsymbol{C})}{\min\limits_{\boldsymbol{C}\in H_0} J_k(\boldsymbol{C})}\right) = -\frac{1}{2}\ln\left(\frac{\kappa_k - \mathrm{tr}(\widehat{\boldsymbol{C}}_1\boldsymbol{\xi}_k)}{\kappa_k - \mathrm{tr}(\widehat{\boldsymbol{C}}_0\boldsymbol{\xi}_k)}\right), \qquad \text{(Eq. 6)}$$

where the optimal parameters for both model fits $\widehat{\boldsymbol{C}}_1$ and $\widehat{\boldsymbol{C}}_0$ can be computed from the quantities $\boldsymbol{\xi}_k$ and $\boldsymbol{W}_k$. At a local maximum of the LLR, an event is detected. With a suitable choice of delay Δ in the cost function in Eq. 4, short time disturbances in the signals are automatically disregarded.

Fig. 2 shows typical raw signals recorded during two test fires and the computed LLR. Vertical lines indicate the event detection times at the maximum of the LLR. The two-sided lines (i.e. model fit) of this time are displayed together with the raw signals. They capture the signal evolution well for the starting time period of the events.

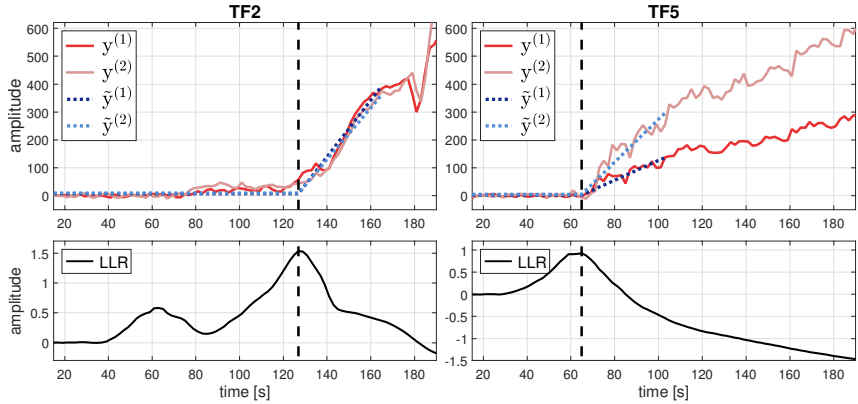

Fig. 2. Top: Examples of 2-channel signals ($y^{(1)}$,$y^{(2)}$) recorded during two EN 54-7 test fires. Vertical lines indicate the event detection time. Estimated signals from the model fit at this time are ($\tilde{y}^{(1)}$,$\tilde{y}^{(2)}$). Bottom: Computed LLR at each time step.

Feature Extraction and Classification

Given that an event is detected, the estimated right-sided slopes $\widehat{\boldsymbol{a}}_{\mathrm{f}} \in \mathbb{R}^M$ from the model fit are used as features for classification. The reason for this feature choice is that from the physical model, $\widehat{\boldsymbol{a}}_{\mathrm{f}}$ is expected to be proportional to $\boldsymbol{R}_T$, which is characteristic for the type T of the event. In other words, all $\widehat{\boldsymbol{a}}_{\mathrm{f}}$ of a type T approximately lie on a line that goes through the origin and is spanned by $\boldsymbol{R}_T$ in the M-dimensional positive space. The line direction $\boldsymbol{v}_T \propto \boldsymbol{R}_T \in \mathbb{R}^M$ can be learned from a labeled data set of recordings with principal component analysis (PCA).

Fig. 3 shows a plot of extracted features and representative lines for a data set with 4 channels. Note that only the first 3 channels are plotted. Hence, the separation of the lines would be even more apparent in $\mathbb{R}^4$.

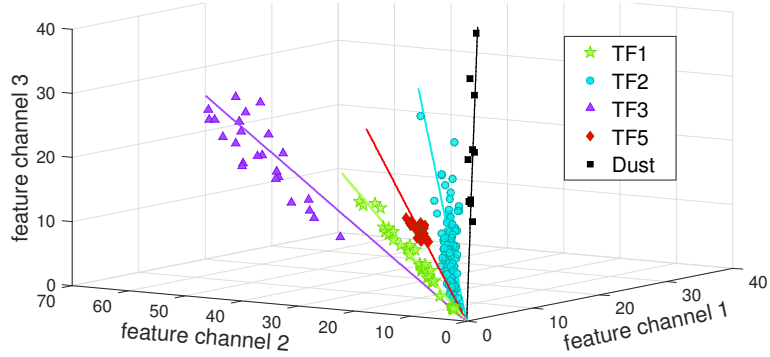

Fig. 3. Classifier illustration: Each point is a feature $\widehat{\boldsymbol{a}}_{\mathrm{f}}$ from a recording of an event. Each line was learned with $\widehat{\boldsymbol{a}}_{\mathrm{f}}$ of its event type.

We classify new events by assigning feature values to the event type of the nearest line. The proposed classifier is linear and a hyperplane defines the decision boundary between any pair of event types T and T'. All points $\boldsymbol{p} \in \mathbb{R}^M$ that lie on the separating hyperplane fulfill the hyperplane equation that is given by the line directions $\boldsymbol{v}_T$ and $\boldsymbol{v}_{T'}$:

$$(\boldsymbol{v}_T - \boldsymbol{v}_{T'})^{\mathrm{T}} \boldsymbol{p} = 0. \qquad \text{(Eq. 7)}$$

Depending on the application, different event types are grouped into classes. In this case, the classifier can be trivially extended by classifying according to the associated class of the predicted event type.

Channel Selection

Given data from M channels, we want to select a subset of M' channels for a classification task determined by the application. The brute-force method is to compare the classification performance for all possible channel sets. This approach is however computationally expensive, especially for a large number of channels. Moreover, a reliable channel selection suitable for real conditions would require a data set of recordings with plenty of variability.

Instead, we propose to directly utilize the line representations of each event type for channel selection. The line directions $\boldsymbol{v}_T$ only need to be learned once with all M channels, while the line directions of a channel subset are obtained directly from the corresponding coordinates of $\boldsymbol{v}_T$. With the help of these lines, we consider two criteria for channel selection: class separation and allowable error margin of the classifier (accounting for e.g., training with a biased data set, physical degradation of the sensors or calibration tolerances).

A measure of separation for a pair of event types T and T' is the angle between their representative lines $\measuredangle(\boldsymbol{v}_T, \boldsymbol{v}_{T'})$. The smallest class separation for a set of channels is then given by the smallest angle between any pair of lines belonging to different classes. A channel set with a large minimal angle thus has well separated classes.

We define the allowable error margin as the allowed percent of signal change in any channel before correct classification is no longer guaranteed. To compute this value, we use the fact that if the signal from any channel can differ by maximally $\pm x\%$ for some $x > 0$, then the collection of all possible new line directions of an event line is bounded by a M'-dimensional hyperrectangle with corners given by:

$$\boldsymbol{v}_T^c = \boldsymbol{v}_T + x\,\mathrm{diag}(\boldsymbol{D}_c)\boldsymbol{v}_T, \qquad \text{(Eq. 8)}$$

where the vector $\boldsymbol{D}_c \in \mathbb{R}^{M'}$ is any of the $2^{M'}$ combinations of ± 1 (e.g. in Fig. 4 the corner is specified by $\boldsymbol{D}_c = [-1,\ 1]$). The margin is thus given by the smallest x which causes a corner of one of the hyperrectangles to hit a separating hyperplane. This is found by inserting Eq. 8 into the hyperplane equation Eq. 7 and solving for x.

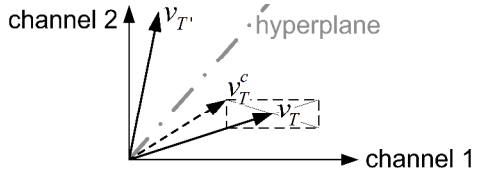

Fig. 4. Depiction of error margin evaluation for two channels.

Example based on Real World Data

We demonstrate the capability of the proposed method by applying it to a data set recorded from a measurement instrument with 4 optical channels. The data set contains recordings of various fire and nuisance event types, with 14 to 100 recordings per type. The fire event types considered are EN 54 test fires (TF1 to TF5, TF8) while the nuisance types considered are two dust types (D1: ISO 12103-1 A1 Ultrafine, D2: DMT Dolomite 10) and haze (HZ) generated by an ultrasonic haze machine. The nuisance tests were performed in such a way that the particle concentration increased at various time rates.

With the algorithm, the event start of each recording was accurately detected. Features were extracted at these times, with which the representative lines of the event types were learned. The lines for TF1 to TF3, TF5, and the dust type D1 can be seen in Fig. 3, which shows that the feature spread around the lines is quite small. First, we classify by assigning each feature to the event type of the nearest line. The results are displayed in the form of a confusion matrix in Table 1.1. Most events are assigned correctly. A majority of misclassifications are in the range of open test fires which produce rather similar particle distributions. Table 1.1 further shows that there is no confusion between fires and nuisances. Just as good results are achieved with 3-fold cross-validation which shows that the model is robust. The classifier with the lines learned from this data set was further tested on a data set with more variations, containing recordings from non-standard experiments performed with the same material as used for TF2, TF3, and TF8 but with varying amount of material and under different procedural and environmental conditions. The results can be seen in Table 1.2.

Actual

Predict	TF1	TF2	TF3	TF4	TF5	TF8	D1	D2	HZ
TF1	1	0	0	0	0	0	0	0	0
TF2	0	0.94	0	0	0	0	0	0	0
TF3	0	0.05	1	0	0	0	0	0	0
TF4	0	0.01	0	0.87	0	0.03	0	0	0
TF5	0	0	0	0.09	0.7	0.37	0	0	0
TF8	0	0	0	0.04	0.3	0.60	0	0	0
D1	0	0	0	0	0	0	1	0	0
D2	0	0	0	0	0	0	0	1	0
HZ	0	0	0	0	0	0	0	0	1

Table 1.1

Actual

Predict	'TF2'	'TF3'	'TF8'
TF1	0	0.17	0.29
TF2	0.98	0	0
TF3	0	0.83	0
TF4	0.02	0	0.19
TF5	0	0	0.14
TF8	0	0	0.38
D1	0	0	0
D2	0	0	0
HZ	0	0	0

Table 1.2

Table 1. Confusion matrix for 4-channel data sets with EN 54 test fires (Table 1.1) and with non-standard fires (Table 1.2).

For the goal of separating fire from nuisance events with the smallest number of required channels, we evaluated the classification performance (average recall: correct classification rate per class) and the two introduced criteria for channel selection (smallest angle, allowable error margin) for all channel combinations (see Fig. 5). As expected, the evaluated measures tend to increase with the number of included channels. Although often high margin, large angle and recall value do not correspond. Selecting channels therefore involves a trade-off between the different measures. For example, one of the 2-channel sets (circle symbol) achieves a high recall compared to all other channel sets, while the other two measures are comparatively low. The discrepancy between these measures may shrink for a data set of recordings with more variability. In this example, the channel set with all 4 channels seems to be preferable as it has the largest angle and

recall. Though one 3-channel set does have a higher margin (~14%), it does not achieve 100% classification performance.

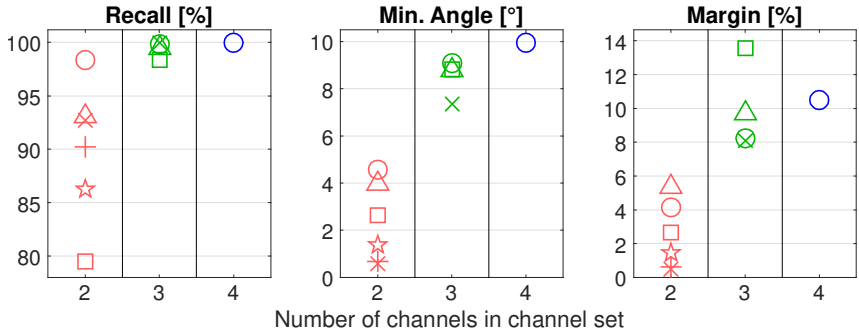

Fig. 5. Channel selection criteria for classification of fires vs. nuisances. For each number of channels, different symbols indicate different channel sets.

The method was also successfully tested on a data set with more channels, obtained from a light scattering measurement instrument with 14 optical channels described in [3]. Here, the method was helpful to make a pre-selection from all 2^{14} possible channel sets.

Conclusion and Outlook

In summary, we have presented an algorithm that can detect and classify fire and nuisance events. The algorithm is flexible in number of channels and event types. In addition, a straightforward real-time implementation is possible. Based on the proposed classifier, a method to select a set of channels for a defined set of event types and classification scenario was introduced. The algorithm was tested on several data sets and the classifier showed good results while the line representations provided useful insights for channel selection. The current channel selection method does not consider the distribution of the feature points around the lines. In the future, this could be included.

References

[1] L. Bruderer, H.-A. Loeliger, N. Zalmai, *"Local statistical models from deterministic state space models, likelihood filtering, and local typicality"*, Proc. of the IEEE International Symposium on Information Theory (ISIT), pp. 1106-1110, July 2014.

[2] N. Zalmai, R. Wildhaber, H.-A. Loeliger, *"Detection and estimation in multi-channel signals with autonomous linear state space models"*, (submitted to) IEEE Trans. on Signal Processing, 2017.

[3] M. Loepfe et al., *"Application of the Rayleigh-Debye-Gans Theory on the Scattering of Light by Standard Test Fire Aerosols"*, Proc. of the 15th International Conference on Automatic Fire Detection AUBE, vol. 1, pp. 45-52, Duisburg, Germany, Oct. 2014.