FIRE AND SMOKE DETECTION IN VIDEO WITH OPTIMAL MASS TRANSPORT BASED OPTICAL FLOW AND NEURAL NETWORKS

I.Kolesov, P.Karasev, A.Tannenbaum

E.Haber

Georgia Institute of Technology
Schools of Electrical and Computer
and Biomedical Engineering
Atlanta,GA

The University of British Columbia
Dept.of Mathematics and Earth
and Ocean Science
Vancouver, Canada

ABSTRACT

Detection of fire and smoke in video is of practical and theoretical interest. In this paper, we propose the use of *optimal mass transport* (OMT) optical flow as a low-dimensional descriptor of these complex processes. The detection process is posed as a supervised Bayesian classification problem with spatio-temporal neighborhoods of pixels;feature vectors are composed of OMT velocities and R,G,B color channels. The classifier is implemented as a single-hidden-layer neural network. Sample results show probability of pixels belonging to fire or smoke. In particular, the classifier successfully distinguishes between smoke and similarly colored white wall, as well as fire from a similarly colored background.

Index Terms— Optimal Mass Transport, Neural Network, Vision, Video, Detection, Supervised Classification.

1. INTRODUCTION

The detection of fire and smoke in video is particularly applicable in industrial monitoring and surveillance. Currently, detection is performed by point detectors that use ionization and light scattering [1].They are not effective in large, open spaces and have an inherent delay because of the time it takes for combustion particles to reach the sensor.

The method in [2] uses only color information at a given pixel. Other approaches use Fourier Descriptors to characterize fire boundaries [3]. In [4], a wavelet analysis is used to resolve the problem of selecting a window size to perform a FFT. This approach relies on deciding what is a sufficient decrease in wavelet energy, from what point in time one should start looking for this decrease, and is susceptible to noise.

The authors in [5] propose a system that models fire pixels as a hidden markov model with (temporal) wavelet coefficients of spatially-fixed pixels being the variables that change between three states. Additionally, they use boundary smoothness of regions as a variable for classification. The two properties are combined as *weak classifiers*. In [6] non-smoke regions are filtered out using background estimation and color information. Then, Lucas-Kanade Optical Flow is computed, and a neural network is trained using statistics of the flow.

(a) Training Pipeline

(b) Classification Pipeline
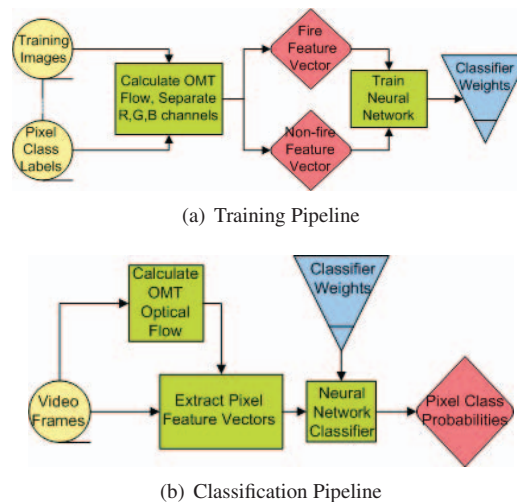
Fig. 1. TOP: Training data is created by manually marking example image sequences. Examples contain spatio-temporal pixel neighborhoods that are marked as to whether they contain fire, smoke, or neither. OMT optical flow is computed via sparse-matrix fininte-difference solver. Feature vectors are formed containing R,G,B color channels and optical flow velocities are compiled and processed by a backpropagation neural network classifier. BOTTOM: the classifier is tested by creating feature vectors for each pixel neighborhood in a new video frame using the trained classifier weights. The final output contains probability of class membership (*Smoke, Fire, None*) for each pixel.

The authors do not attempt to classify individual pixels, as done in this paper, instead they have a binary output: smoke is present in the frame or it is not. And, we propose using the optimal mass transport (OMT) Optical Flow, which is better suited for this application, as explained in Section 2.2 .

The remainder of this paper is as follows: in Section 2, the classification pipeline is described. A review of OMT and its application to image analysis is given in Section 2.1. Section 3 describes the architecture of the Neural Network classifier. We conclude with some example results in Section 4

and final remarks in

2. CLASSIFIER FEATURE SELECTION

The majority of existing methods are based on a heuristic model that roughly characterizes fire or smoke but tends to be sub-optimal. A natural approach is to learn the model from training data that describes smoke or fire by training a classifier such as a neural network. The training and testing pipelines are described graphically in figure Fig. 1 .

Computing optical flow for an image sequence rather than simple frame differencing allows one to take into account expected properties of the process being imaged; for reasons discussed in the next sections, optical flow based on optimal mass transport(OMT) is calculated for fire and the Horn and Schunck optical flow for smoke region classification.

2.1. Optimal Mass Transport

The mass transport problem was first formulated by Gaspar Monge in 1781, and concerned finding the optimal way, in the sense of minimal transportation cost, of moving a pile of soil from one site to another. This problem was given a modern formulation in the work of Kantorovich, and so is now known as the *Monge–Kantorovich problem.*

We now give a modern formulation of the Monge–Kantorovich problem. Let Ω_0 and Ω_1 be two subdomains of $\mathbb{R}^d$, with smooth boundaries, each with a positive density function, μ_0 and μ_1, respectively. We assume

$$\int_{\Omega_0} \mu_0 = \int_{\Omega_1} \mu_1$$

so that the same total mass is associated with Ω_0 and Ω_1. We consider diffeomorphisms $\tilde{u}$ from (Ω_0, μ_0) to (Ω_1, μ_1) which map one density to the other in the sense that

$$\mu_0 = |D\tilde{u}|\, \mu_1 \circ \tilde{u}, \tag{1}$$

which we will call the *mass preservation* (MP) property, and write $\tilde{u} \in MP$. Equation (1) is called the *Jacobian equation.* Here $|D\tilde{u}|$ denotes the determinant of the Jacobian map $D\tilde{u}$. In particular, Equation (1) implies, for example, that if a small region in Ω_0 is mapped to a larger region in Ω_1, there must be a corresponding decrease in density for the mass to be preserved. A mapping $\tilde{u}$ that satisfies this property may thus be thought of as defining a redistribution of a mass of material from one distribution μ_0 to another distribution μ_1.

There may be many such mappings, and we want to pick out an optimal one in some sense. Accordingly, we define the L^p Kantorovich–Wasserstein metric as follows:

$$d_p(\mu_0, \mu_1)^p := \inf_{\tilde{u} \in MP} \int \|\tilde{u}(x) - x\|^p \mu_0(x)\, dx. \tag{2}$$

An *optimal MP map*, when it exists, is one which minimizes this integral. This functional is seen to place a penalty on the distance the map $\tilde{u}$ moves each bit of material, weighted by the material's mass.

A fundamental theoretical result [7, 8], is that there is a unique optimal $\tilde{u} \in MP$ transporting μ_0 to μ_1, and that this $\tilde{u}$ is characterized as the gradient of a convex function w, i.e., $\tilde{u} = \nabla w$. Note that from Equation (1), we have that w satisfies the *Monge–Ampère* equation

$$|Hw|\, \mu_1 \circ (\nabla w) = \mu_0,$$

where $|Hw|$ denotes the determinant of the Hessian of w.

Hence, the Kantorovich–Wasserstein metric defines the distance between two mass densities, by computing the cheapest way to transport the mass from one domain to the other with respect to the functional given in (2), the optimal transport map in the $p = 2$ case being the gradient of a certain function. The novelty of this result is that like the Riemann mapping theorem in the plane, the procedure singles out a particular map with preferred geometry.

2.2. Optical Flow

Optical flow is a computational procedure to compute the motion between a set of images taken within a short time difference. The main idea is that the gray values of each image do not change between two images. This leads to the *optical flow constraint*

$$I_t + \vec{\mathbf{u}} \cdot \nabla I = 0. \tag{3}$$

where I is the image and $\vec{\mathbf{u}} = [u, v]$ is the flow field. Given two images taken in a short time interval, it is possible to solve for the optical flow field $\vec{\mathbf{u}}$ by solving the following optimization problem

$$\min_{\vec{\mathbf{u}}} \quad \frac{1}{2}\|I_t + \vec{\mathbf{u}} \cdot \nabla I\|^2 + \alpha R(\vec{\mathbf{u}}) \tag{4}$$

where $R(\vec{\mathbf{u}})$ is a regularization operator, typically chosen to be the gradient of $\vec{\mathbf{u}}$ and α is a regularization parameter.

Notice that the underlying assumption for Equation (4) is one of intensity constancy. Under this assumption an objects brightness is constant from frame to frame. This assumption holds for rigid objects with a Lambertian surface, but fails for fluid and gaseous materials. In computer vision, these are modeled by so-called *dynamic textures.* The dynamic textures typical of smoke and fire possess intrinsic dynamics and so cannot be reliably captured by the standard optical flow method. Also, the fire/smoke region tends to flow much faster than the area around it which again may cause the model given by Equation (4) to produce erroneous results.

Our goal in this paper is to obtain better optical flow field models for fire and smoke. One way to do so is to base the optical flow on the physical attributes of these processes. One simple attribute is that fire and smoke tends to approximately conserve intensity taken as a generalized mass and move the

mass in an optimal way [9, 10]. Thus, an appropriate mathematical optical constraint is not intensity preserving but rather *mass conservation* or *brightness conservation*. This model can be written as

$$I_t + \nabla \cdot \vec{\mathbf{u}} I = 0. \tag{5}$$

The justification is as follows:

$$\int_\Omega I_t dA = -\int_{\partial\Omega} I\vec{\mathbf{u}} \cdot \vec{\mathbf{N}} dL.$$

This means that total rate change in intensity can only occur through a flow (entering or exiting on the boundary). This is a conservation law. But by the **divergence theorem**

$$\int_{\partial\Omega} I\vec{\mathbf{u}} \cdot \vec{\mathbf{N}} dL = \int_\Omega \nabla \cdot (I\vec{\mathbf{u}}) dA \qquad \text{so}$$

$$\int_\Omega I_t + \nabla \cdot (I\vec{\mathbf{u}}) dA = 0 \qquad \text{and}$$

$$I_t + \nabla \cdot (I\vec{\mathbf{u}}) = 0.$$

This is precisely the condition of infinitesimal brightness (mass) conservation.

Following the reasoning in the previous section, we propose the following for optical flow for dynamic texture segmentation:

$$\min \ \alpha \int_\Omega \int_0^T I\,|\vec{\mathbf{u}}|^2\ dx\ dt\ + \frac{1}{2}\|I_t + \nabla \cdot \vec{\mathbf{u}} I\|^2. \tag{6}$$

The first term in this optimization problem represents the total energy of moving the images while the second is the mass preserving optical flow equation.

3. CLASSIFICATION WITH NEURAL NETWORKS

The problem of smoke detection is posed as a two class system that must decide if a given pixel belongs to class *smoke* or class *non-smoke*. Neural Networks compute a least-squares model fit to Bayes discriminant functions [11]; the output is a probability of a pixel belonging to a particular class, thus a user's choice of threshold level for classifying a pixel as smoke/non-smoke directly corresponds to his desired confidence level. From Bayes theorem, we know that for a multi-class problem the posterior probability can be written as

$$p(C_k|\mathbf{x}) = \frac{p(\mathbf{x}|C_k)p(C_k)}{\Sigma_i p(\mathbf{x}|C_i)p(C_i)}\ , \tag{7}$$

where $\mathbf{x}$ is assigned to the class C_k that corresponds to the discriminant function, $y_k(\mathbf{x}, \mathbf{w})$, having the largest value. Here, target values $t_k(\mathbf{x}) = \mathbf{1}$ if $\mathbf{x}$ belongs to class C_k and zero otherwise. The error committed by the neural network per output is defined to be

$$E(\mathbf{w}) = \sum_{n=1}^N (y_k(\mathbf{x_n}, \mathbf{w}) - t_n)^2\ . \tag{8}$$

In the limit that N, the number of samples approaches infinity, it can be shown [11] that back-propogation approach to minimizing the error committed by a neural network minimizes the following sum

$$\sum_{k=1}^n \int (y_k(\mathbf{x}, w) - P(C_k|\mathbf{x}))^2 p(\mathbf{x})d\mathbf{x}\ , \tag{9}$$

where n is the number of classes. This shows that as the number of data point goes to infinity the output of the discriminant function is equal to the posterior probability $y_k(\mathbf{x}, w) \approx P(C_k|\mathbf{x})$. Hence, assigning $\mathbf{x}$ to the class C_k, which maps to the discriminant function with the largest value, is equivalent to assigning $\mathbf{x}$ to the class with the highest posterior probability.

Bayes theorem is used again to determine the form of the discriminant functions. The posterior probabilities can be written as

$$p(C_k|\mathbf{x}) = \frac{p(\mathbf{x}|C_k)p(C_k)}{\Sigma_i p(\mathbf{x}|C_i)p(C_i)} = \frac{exp(a_k)}{\Sigma_i exp(a_i)}\ , \tag{10}$$

also known as the softmax function, with the substitution $a_k = ln(p(\mathbf{x}|C_k)p(C_k))$ [12] . This is precisely the activation function a neural network uses.

Assume that the class-conditional probability densities $p(\mathbf{x}|C_k)$ belong to the restricted exponential family of distributions, which take the form

$$p(\mathbf{x}|\mathbf{w}_k) = h(\mathbf{x})g(\mathbf{w}_k) \exp\left(\mathbf{w}_k{}^T\mathbf{x}\right)\ . \tag{11}$$

Substituting this density into (10), the resulting equation for $a_k(\mathbf{x})$ is linear in x:

$$a_k(\mathbf{x}) = \mathbf{w}_k{}^T\mathbf{x} + ln\ g(\mathbf{w}_k) + ln\ p(C_k)\ . \tag{12}$$

Thus, the discriminant function takes the form of an activation function acting on a linear combination of non-linear basis functions $\phi(\mathbf{x})$:

$$y(\mathbf{x}, \mathbf{w}) = f(\sum_{i=1}^M w_j\phi_j(\mathbf{x}) + w_0)\ , \tag{13}$$

where $f(\cdot)$ is the activation function.

In a neural network the non-linear basis functions, implemented as hidden units, are adaptively selected and are themselves functions of a linear combinations of inputs

$$\phi_j(\mathbf{x}) = h(\sum_{i=1}^L l_i x_i + l_0) \tag{14}$$

where $h(\cdot)$ is also a softmax function. The Neural Network used herein is fully connected and made up of a single hidden level containing twenty hidden units with a softmax nonlinearity in the hidden layer and in the output.

4. RESULTS

To obtain the following results, just six frames of a video were used in training a neural network classifier. Certain fire/non-fire and smoke/non-smoke regions were manually delineated. The number of samples used is small and it comes from the same video. The classifier can be extended to work with other videos by providing significantly more samples especially from varying video sources of positive and negative data samples.

The output of a neural network classifier is a posterior probability $p(C_k|\mathbf{x})$ for each pixel, where C_k is the class, fire/smoke or non-fire/non-smoke, and $\mathbf{x}$ is the feature vector for a given pixel. Figure 2(b) shows a sample output of the classifier for all pixels in a video frame. A threshold can be applied to make a binary decision about a pixel's class; this is shown in figure 2(c) for smoke and figure 3(b) for fire.

(a) Input Image   (b) Probability of class *Smoke*   (c) Binary decision by probability threshold
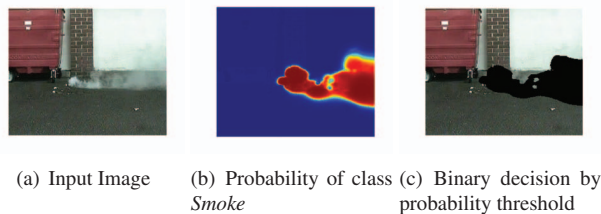
Fig. 2. Feature vectors containing image and OMT Optical Flow velocity values in a spatio-temporal pixel neigbhorhood are extracted and fed into a neural network classifier. The output assigns a probability of each pixel belonging to class *Smoke*, shown in Fig. 2(b). Binary decisions are formed by thresholding the probabilities. Notice that the white smoke is distinguished from the white wall.

Similarly the binary classification of pixels in a video frame containing fire can be seen in figure 3.

(a) Original Image   (b) Image after threshold is applied
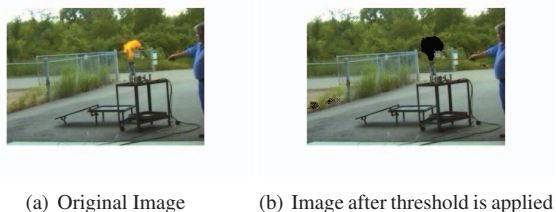
Fig. 3. An image sequence is fed into a neural network classifier that has been trained with examples of *Fire* and *Not Fire* pixel neighborhoods. The fire in the image is marked by a threshold of probabilities given by the classifier.

5. CONCLUSION

Video detection of flame/smoke is the preferred detection method because it allows a single camera to cover a large area and integrates easily into existing closed circuit surveillance systems. In this work, we presented an OMT based optical flow calculation that respects physical properties of the process being measured and an approach to training a classifier rather than creating one from heuristics.

6. REFERENCES

[1] T.Cleary and W.Grosshandler, "Survey of Fire Detection Technologies and System Evaluation/Certification Methodologies and Their Suitability for Aircraft Cargo Compartments," NISTIR 6356, National Institute of Standards and Technology, July 1999.

[2] B.Drda G.Healey D.Slater T.Lin and A.D. Goedeke, "A system for real-time fire detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, NY, Jun 1993, Donmar Limited, IEEE.

[3] N.Ahuja C.B Liu, "Vision based fire detection," in *International Conference on Pattern Recognition*. Beckman Institute, Aug 2004, IEEE.

[4] Y.Dedeoglu B. Toreyin and E.Cetin, "Wavelet based real-time smoke detection in video," in *Proceedings of the 13th European Signal Processing Conference*, Antalya, Turkey, Sept 2005, Bilkent University.

[5] B.U. Toreyin and A.E. Çetin, "Online Detection of Fire in Video," in *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07*, 2007, pp. 1–5.

[6] W.Jinjun Y.Chunyu, F.Jun and Z.Yongming, "Video fire smoke detection using motion and color features," *Fire Technology*, vol. 46, no. 3, pp. 651–663, Jul 2010.

[7] Y. Brenier, "Polar factorization and monotone rearrangement of vector-valued functions," *Com. Pure Appl. Math.*, vol. 64, pp. 375–417, 1991.

[8] W. Gangbo and R. McCann, "The geometry of optimal transportation," *Acta Math.*, vol. 177, pp. 113–161, 1996.

[9] J.D Benamou and Y. Brenier, "A computational fluid mechanics solution to the monge-kantorovich mass transfer problem," *Numerische Mathematik*, vol. 84, pp. 375–393, 2000.

[10] S.Haker S. Angenent and A.Tannenbaum, "Minimizing flows for the monge-kantorovich problem," *SIAM J. Math. Analysis*, vol. 35, pp. 61–97, 2003.

[11] P.Hart R. Duda and D.Stork, Eds., *Pattern Classification*, John Wiley and Sons,Inc, New York, NY, 2001.

[12] C.M. Bishop et al., *Pattern recognition and machine learning*, Springer New York:, 2006.