Classification of Traffic Video Based on a Spatiotemporal Orientation Analysis

Konstantinos G. Derpanis and Richard P. Wildes
Department of Computer Science and Engineering
York University
Toronto, Ontario, Canada
{kosta,wildes}@cse.yorku.ca

Abstract

This paper describes a system for classifying traffic congestion videos based on their observed visual dynamics. Central to the proposed system is treating traffic flow identification as an instance of dynamic texture classification. More specifically, a recent discriminative model of dynamic textures is adapted for the special case of traffic flows. This approach avoids the need for segmentation, tracking and motion estimation that typify extant approaches. Classification is based on matching distributions (or histograms) of spacetime orientation structure. Empirical evaluation on a publicly available data set shows high classification performance and robustness to typical environmental conditions (e.g., variable lighting).

1. Introduction

1.1. Motivation

Traffic congestion is a serious issue confronting many urban centres. To address this issue, traditional solutions have mainly consisted of increasing the supply of roads; however, such remedies are becoming less feasible due to the prohibitive costs involved and the scarcity of suitable land. Instead, contemporary solutions focus on optimizing the throughput of existing roads. Here, methods for gathering real-time information on traffic flows are key. Early traffic monitoring systems relied on inductive-loop detectors [38], which are buried underneath roadways, to count vehicles traveling over them. More recently, the use of camera networks have shown promise for monitoring traffic. In contrast to loop detectors, video based systems are less disruptive, less costly to install and allow for a more detailed understanding of traffic flow patterns.

In response to the above motivations, this paper presents a novel approach to representing and classifying traffic congestion scenes as captured in video. In this work, traffic scenes are treated as instances of a dynamic texture [9], i.e., as spatiotemporal image patterns best characterized in terms of the aggregate dynamics of a set of constituent elements, rather than in terms of the individuals (cf. spatial texture [36]). In particular, traffic patterns will be classified directly in terms of measures of their dynamics aggregated over regions of image spacetime, (x, y, t), rather than via the analysis of individual vehicles. Toward that end, an approach is developed that is based solely on observed dynamics (i.e., excluding purely spatial appearance cues). For such purposes, local spatiotemporal orientation is of fundamental descriptive power, as it captures the first-order correlation structure of the data irrespective of its origin (i.e., irrespective of the underlying visual phenomena), even while allowing for the discrimination of pattern differences (e.g., levels of congestion). Correspondingly, each traffic scene will be associated with a distribution (histogram) of measurements that indicates the relative presence of a particular set of 3D orientations in visual spacetime, (x, y, t), as captured by a bank of spatiotemporal filters and recognition will be performed by matching such distributions.

1.2. Related work

Most extant approaches to classifying vehicular traffic videos use a combination of segmentation and tracking, e.g., [24, 25, 3, 28, 35, 11, 23, 22, 30, 29, 6]. The general procedure consists of the following three steps: (i) motion detection, (ii) tracking of the individual vehicles and (iii) combining trajectory information to derive an overall description of traffic flow. Problems associated with these approaches include, (i) segmentation issues due to varying environmental conditions (e.g., lighting), occlusions and low-resolution imagery resulting in small pixel support of vehicle targets and (ii) tracking issues related to correspondence problems and occlusions.

Alternatively, several approaches have attempted to recover a holistic representation (macroscopic view) of traffic flow information directly, thereby avoiding the need for detecting individual moving objects. For example, statistics of optical flow taken over the roadway have been used to

characterize traffic flow [41, 27, 32, 31, 26]. A drawback of these approaches is that extracting reliable measurements of flow is difficult in traffic scenarios due to environmental conditions and are subject to noise in cases where there are many vehicles in the scene. Other work has proposed that traffic flows could be modeled directly as dynamic textures [7], defined as an autoregressive (AR) stochastic process with spatial and temporal components [15]. The utility of this representation was demonstrated in the context of a traffic congestion classification application. A drawback of this approach is the large computational load in fitting the model, as a result analysis is limited to relatively small image patches and might make this approach impractical for application to real-time traffic monitoring.

Similar to [7], the proposed approach adopts previous work directed at general dynamic texture classification [13] for the purpose of classifying traffic congestion scenes. Previous empirical evaluation of the proposed approach on a standard image data set consisting of a wide range of dynamic textures has shown significant classification improvement over alternative state-of-the-art approaches. Spatiotemporal oriented energy filters serve in defining the representation employed. In addition to the analysis of dynamic textures [13], previous efforts have used similar operators in the analysis of image sequences for various other purposes, e.g., enhancement [17, 20], motion estimation [21, 18, 34] and activity recognition [10, 14, 12]. Significantly, it appears that no previous work has used the filter outputs to support traffic congestion classification, as shown here.

1.3. Contributions

In the light of previous research, the contributions of the present work are as follows. First, a system is presented for classifying traffic congestion scenes depicted in videos. The system is based on a particular spatiotemporal filtering formulation for measuring spatiotemporal oriented energy that is used for representing and recognizing traffic congestion scenarios based solely on their underlying dynamics. Second, empirical evaluation on a publicly available data set demonstrates that the proposed system achieves state-of-the-art performance, while at the same time being amenable to computationally efficient realization.

2. Technical approach

The local spacetime orientation of a visual pattern captures significant, meaningful aspects of its dynamic structure [1, 40, 13]. As examples, it can provide the basis for characterizing image motion as well as more general pattern dynamics, even while exhibiting robustness to illumination as well as purely spatial appearance variation. Correspondingly, a spatiotemporally oriented decomposition of a visual pattern provides a useful basis for local representation of image dynamics. By extension, aggregate measures of orientation over a region of visual spacetime are of use in characterizing the region's spacetime texture. The characteristics of this approach are well matched to the analysis of traffic video: The dynamics of the pattern are captured, robustness to illumination provides ability to operate consistently over a wide range of natural and artificial lighting conditions and robustness to purely spatial appearance allows traffic patterns to be characterized independent of the exact composition of the vehicles involved.

2.1. Representation: Distributed spacetime orientation

The desired spacetime orientation decomposition is realized using a bank of broadly tuned 3D Gaussian third derivative filters, $G_{3_{\hat{\theta}}}(x, y, t)$, with the unit vector $\hat{\theta}$ capturing the 3D direction of the filter symmetry axis. The responses of the image data to this filter are pointwise rectified (squared) and integrated (summed) over a spacetime region, Ω, that covers the entire traffic scene sample under analysis, to yield the following energy measurement for the region

$$E_{\hat{\theta}} = \sum_{(x,y,t)\in\Omega} (G_{3_{\hat{\theta}}} * I)^2, \tag{1}$$

where $I \equiv I(x, y, t)$ denotes the input imagery and $*$ convolution. Notice that while the employed Gaussian derivative filter is phase-sensitive, summation over the support region ameliorates this sensitivity to yield a measurement of signal energy at orientation θ. More specifically, this follows from Parseval's theorem [5] that specifies the phase-independent signal energy in the frequency passband of the Gaussian derivative:

$$E_{\hat{\theta}} \propto \sum_{(\mathbf{k},\omega_t)} |\mathcal{F}\{G_{3_{\hat{\theta}}} * I\}(\omega_x, \omega_y, \omega_t)|^2, \tag{2}$$

where $(\omega_x, \omega_y, \omega_t)$ denotes the spatiotemporal frequency coordinate and $\mathcal{F}$ the Fourier transform[1].

Each oriented energy measurement, (1), is confounded with spatial orientation. Consequently, in cases where the spatial structure varies widely about an otherwise coherent dynamic region (e.g., single motion across a region with varying spatial texture), the responses of the ensemble of oriented energies will reflect this behaviour and thereby are spatial appearance dependent; whereas, a description of pure pattern dynamics is sought. To remove this difficulty, the spatial orientation component is discounted by "marginalization" of this attribute, as follows.

In general, a pattern exhibiting a single spacetime orientation (e.g., image velocity) manifests itself as a plane

[1] Strictly, Parseval's theorem is stated with infinite frequency domain support on summation.

through the origin in the frequency domain [39]. Correspondingly, summation across a set of x-y-t-oriented energy measurements consistent with a single frequency domain plane through the origin is indicative of energy along the associated spacetime orientation, independent of purely spatial orientation. Since Gaussian derivative filters of order $N = 3$ are used in the oriented filtering, (1), it is appropriate to consider $N+1 = 4$ equally spaced directions along each frequency domain plane of interest, as $N+1$ directions are needed to span orientation in a plane with Gaussian derivative filters of order N [19]. Let each plane be parameterized in terms of its unit normal, $\hat{\mathbf{n}}$; a set of equally spaced $N+1$ directions within the plane are given as

$$\hat{\theta}_i = \cos\left(\frac{2\pi i}{N+1}\right)\hat{\theta}_a(\hat{\mathbf{n}}) + \sin\left(\frac{2\pi i}{N+1}\right)\hat{\theta}_b(\hat{\mathbf{n}}),\ 0 \leq i \leq N, \tag{3}$$

with

$$\hat{\theta}_a(\hat{\mathbf{n}}) = \hat{\mathbf{n}} \times \hat{\mathbf{e}}_x / \|\hat{\mathbf{n}} \times \hat{\mathbf{e}}_x\| \quad \hat{\theta}_b(\hat{\mathbf{n}}) = \hat{\mathbf{n}} \times \hat{\theta}_a(\hat{\mathbf{n}}) \tag{4}$$

where $\hat{\mathbf{e}}_x$ denotes the unit vector along the ω_x-axis[2].

Now, energy along a frequency domain plane with normal $\hat{\mathbf{n}}$ and spatial orientation discounted through marginalization, is given by summation across the set of measurements, $E_{\hat{\theta}_i}$, as

$$\tilde{E}_{\hat{\mathbf{n}}} = \sum_{i=0}^{N} E_{\hat{\theta}_i}, \tag{5}$$

with $\hat{\theta}_i$ one of $N+1 = 4$ specified directions, (3), and each $E_{\hat{\theta}_i}$ calculated via the oriented energy filtering, (1). In the present implementation, 27 different spacetime orientations, as specified by $\hat{\mathbf{n}}$, are made explicit, corresponding to static (no motion/orientation orthogonal to the image plane), slow (half pixel/frame movement), medium (one pixel/frame movement) and fast (two pixel/frame movement) motion in the directions leftward, rightward, upward, downward and diagonal, and flicker/infinite vertical and horizontal motion (orientation orthogonal to the temporal axis); although, due to the relatively broad tuning of the filters employed, responses arise to a range of orientations about the peak tunings.

Finally, the marginalized energy measurements, (5), are confounded by the local contrast of the signal and as a result increase monotonically with contrast. This makes it impossible to determine whether a high response for a particular spacetime orientation is indicative of its presence or is indeed a low match that yields a high response due to significant contrast in the signal. To arrive at a purer measure of spacetime orientation, the energy measures are normalized by the sum of consort planar energy responses,

$$\hat{E}_{\hat{\mathbf{n}}_i} = \tilde{E}_{\hat{\mathbf{n}}_i} / \left(\sum_{j=1}^{M} \tilde{E}_{\hat{\mathbf{n}}_j} + \epsilon\right), \tag{6}$$

where M denotes the number of spacetime orientations considered and ϵ is a constant introduced as a noise floor. Conceptually, (1) - (6) can be thought of as taking an image sequence, $I(x, y, t)$, and carving its power spectrum into a set of planes, with each plane corresponding to a particular spacetime orientation, to provide a relative indication of the presence of structure along each plane.

2.2. Representation properties

The constructed representation enjoys a number of attributes worth emphasizing. First, owing to the bandpass nature of the Gaussian derivative filters (1), the representation is invariant to additive photometric bias in the input signal. Second, owing to the divisive normalization (6), the representation is invariant to multiplicative photometric bias. These first two invariances combine to provide a level of robustness to illumination variation, which is important in the analysis of traffic video, if it is to provide consistent results across diurnal and other anticipated lighting variations. Third, owing to the marginalization (5), the representation is invariant to changes in appearance manifest as spatial orientation variation. Overall, these three invariances allow abstractions to be robust to pattern changes that do not correspond to dynamic pattern variation (i.e., spatial appearance), even while making explicit local orientation structure that arises with temporal variation. Robustness to purely spatial appearance is of importance for traffic congestion classification to provide consistent estimates independent of the particular composition of vehicles that are present in a given scene. Finally, the representation is efficiently realized via linear (separable convolution, pointwise addition) and pointwise non-linear (squaring, division) operations; thus, efficient computations are realized [19], including real-time realizations on GPUs [42]. The issue of computational complexity is important since a system may consist of an array of hundreds of video cameras including the potential need for real-time analysis.

Overall, each of the normalized oriented energies can be viewed as expressing the evidence for the presence of a particular, spacetime oriented structure. Taken as an ensemble (distribution), they provide the relative contribution of each spacetime orientation in the decomposition of the traffic scene signal under consideration.

2.3. Recognition: Spacetime orientation distribution similarity

An ensemble of (normalized) energy measurements, $\hat{E}_{\hat{\mathbf{n}}_i}$, is taken as a distribution with spatiotemporal orien-

[2]Depending on the spacetime orientation sought, $\hat{\mathbf{e}}_x$ can be replaced with another axis to avoid the case of an undefined vector.

Algorithm 1: Traffic congestion recognition.

Input: Q: Query traffic video, D: Database containing labeled traffic congestion videos
Output: c: Classification label

`Step 1: Compute spacetime oriented energy representation (Sec. 2.1)`
1. Initialize 3D G_3 steerable basis.
2. Compute normalized spacetime oriented energies for Q and D, Eq. (1) - (6).

`Step 2: Recognition (Sec. 2.3)`
3. Compute nearest-neighbour of Q in D using the Bhattacharyya measure, (7).
4. Assign label of nearest-neighbour in D to c.

tation, $\hat{\mathbf{n}}_i$, as variable. (In practice, these measurements are maintained as histograms.) Given the spacetime oriented energy distributions of an input query and database with entries represented in like fashion, the final step of the approach is recognition. In the present application, the database is composed of a set of spatiotemporal oriented energy distributions labeled according to the level of congestion that were derived from a set of exemplar traffic videos; the queries derive from traffic video that is to be classified and are likewise represented in terms of their spatiotemporal oriented energy distributions. In general, to compare two distributions, denoted $\mathbf{x}$ and $\mathbf{y}$, there are several standard similarity measures in the literature that are applicable [33]. In the present application, the Bhattacharyya measure was employed as it was empirically demonstrated to yield superior classification performance over other popular measures for dynamic texture recognition [13]. (In the following, individual entries in the employed histogram representation of the distributions are specified via subscripting, e.g., x_i, and summations are taken across all bins.) In particular, the Bhattacharyya coefficient (similarity on hyper-sphere) [4] is defined as

$$s_B(\mathbf{x}, \mathbf{y}) = \sum_i \sqrt{x_i y_i}. \tag{7}$$

Finally, for any given distance measure, a method must be defined to determine the classification of a given probe relative to the database entries. In this work, a standard Nearest-Neighbour (NN) classifier [16] was used in the experiments to be presented. Although not state-of-the-art, the NN classifier has been shown to yield competitive results relative to the state-of-the-art Support Vector Machine (SVM) classifier [37] for dynamic texture classification [8] and thus provides a useful lower-bound on performance.

To recapitulate, the proposed system for traffic congestion recognition is given in algorithmic terms in Algorithm 1.

Traffic Condition	Description	Number of Videos
light	traffic around the speed limit	165
medium	reduced speed	45
heavy	slow or stop and go speeds	44

Table 1. UCSD traffic data set summary.

		Classified		
		light	*medium*	*heavy*
Actual	*light (total 165)*	**163**	1	1
	medium (45)	1	**40**	4
	heavy (44)	1	4	**39**

Table 2. Traffic congestion confusion matrix. Cumulative confusion matrix for traffic congestion classification for all four testing trials using the proposed system.

3. Empirical evaluation

3.1. UCSD traffic data set

The UCSD traffic video data set[3] consists of video sequences of daytime highway traffic in Seattle, Washington, totalling 20 minutes of video footage [7]. The videos contain a variety of traffic congestion patterns and weather conditions (e.g., raining, overcast and sunny). Each video has a resolution of 320×240 pixels with 42 to 52 frames captured at 10 frames per second. The data set provides representative 48×48 video patches for training and testing, which were manually selected over the area with the "most activity". Also, hand-labeled ground truth is provided that describes the amount of traffic congestion in each sequence. In total there are 254 video sequences, grouped into three classes of traffic congestion, light, medium and heavy; see Table 1 for summary. Example frames from the data set are shown in Fig. 1.

3.2. Traffic congestion classification

In previous work using the USCD traffic video data set [7], reported traffic congestion classification results were computed as the average classification rate taken over four trial runs. Each trial consisted of splitting the data differently with 75% of the video samples reserved for training and 25% for testing. The training and test data splits for each trial are provided with the data set. The empirical results described next are based on the same testing protocol.

The proposed system achieved an overall classification rate of 95.28%. By way of comparison, the best previous result reported on this data set resulted in an overall classification result of 94.5%. Table 2 provides the cumulative confusion matrix for all four testing trials using the pro-

[3] Available at: www.svcl.ucsd.edu/projects/traffic

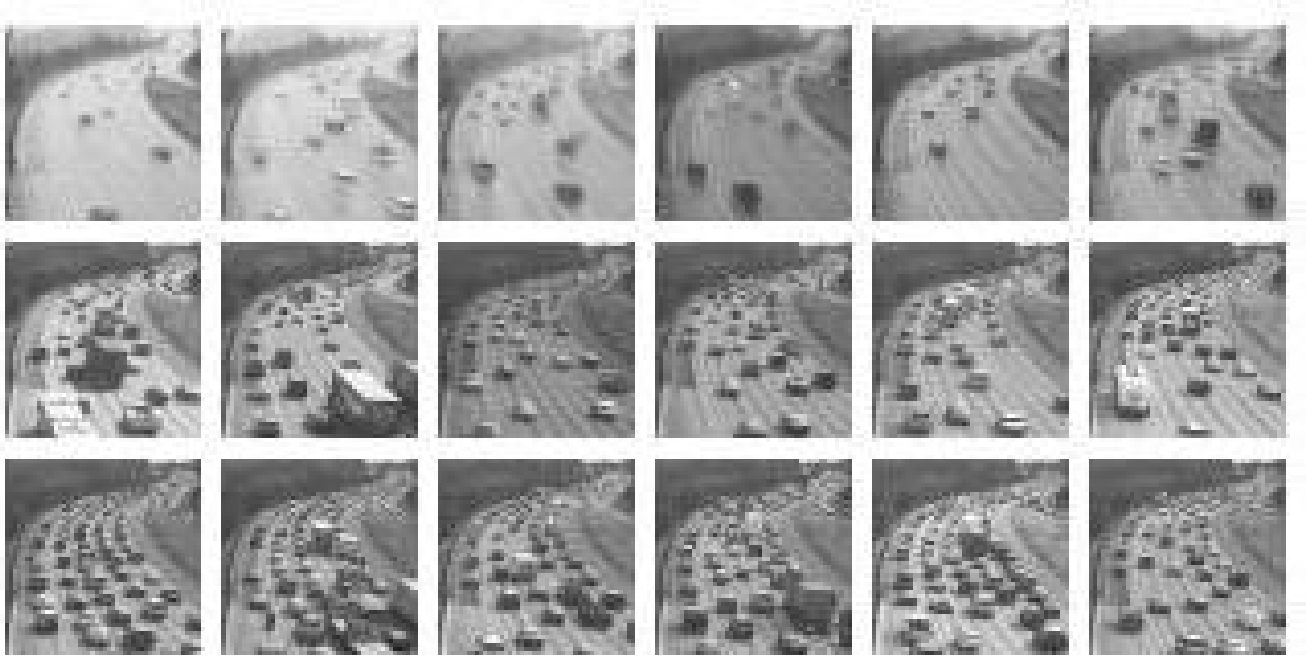


Figure 1. Example frames from the UCSD traffic video data set. The sample frames depict various traffic congestion conditions, coarsely categorized as light (top row), medium (middle row) and heavy traffic (bottom row).

posed system. Figure 2 shows a correct classification result for each of the three traffic congestion conditions.

As shown in Table 2, the majority of misclassifications (ten cases in total) occur between neighbouring classes (i.e., light vs. medium and medium vs. heavy). Given the indeterminate nature of category boundaries and the corresponding ambiguities of generating ground truth, such matches are reasonable. Figure 3 (a) shows an example of this ambiguity where the input depicting heavy traffic is matched closest to an instance of medium traffic. The remaining two misclassifications are related to confusions between light and heavy traffic (i.e., non-neighbouring classes). In both instances, the light traffic sequences largely depict the background (i.e., few cars are present), while the matched heavy traffic sequences depict cars that are virtually at a standstill (see Fig. 3 (b) for an example). From the viewpoint of dynamics, both scenes are similar and thus the matches are reasonable.

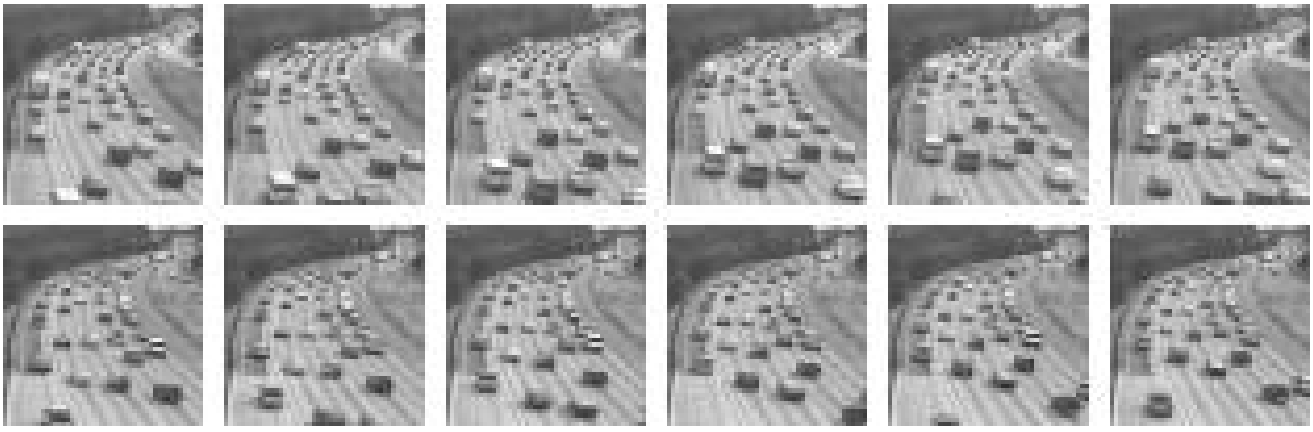
As noted above, in previous experiments with the UCSD database [7], an alternative approach based on dynamic texture analysis yielded a comparable overall recognition rate (94.5%); however, this result was based on the computationally intensive estimation of an autoregressive stochastic model of dynamic texture, which, to date, precludes efficient computation and real-time applications. In contrast, the spatiotemporal oriented energy model of dynamic texture that is demonstrated here is amenable to real-time estimation [42]. Ability to analyze traffic video in an efficient fashion impacts the manner in which the results can be deployed, e.g., if traffic video is to be analyzed for congestion to advise drivers and otherwise control traffic online, then real-time operation is critical.

4. Discussion and summary

The main contribution in this paper is a system for recognizing traffic congestion scenarios from video based on their observed visual dynamics. Dynamic information is encapsulated by a set of spacetime orientation measurements realized by a particular spatiotemporal filtering formulation for measuring spatiotemporal oriented energy. As compared to most extant approaches, the proposed system has the following key advantages: (i) does not rely on segmentation, (ii) does not rely on tracking, (iii) does not rely on optical flow estimation, (iv) can accommodate variable appearance, such as lighting variation and (v) is amenable to efficient computation.

There are several possible directions for future work. First, the current approach has limited ability to distinguish between completely stopped traffic and an empty (stationary) roadway because it is based on scene dynamics (both stopped traffic and the empty roadway are static). Other approaches based on dynamics also exhibit this limitation. A straightforward way to extend the current approach to make this distinction would be to incorporate spatial appearance

(a) light congestion

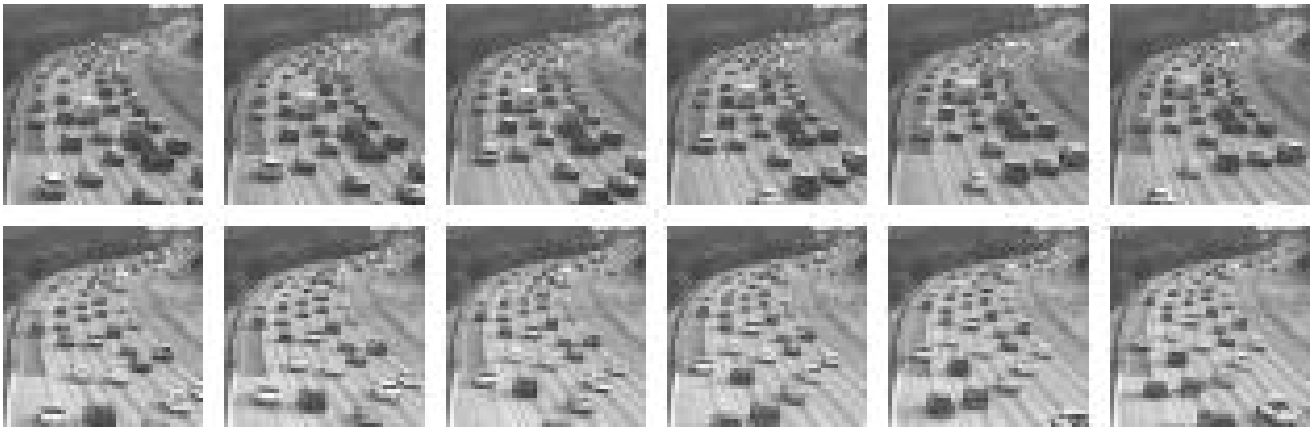


(b) medium congestion

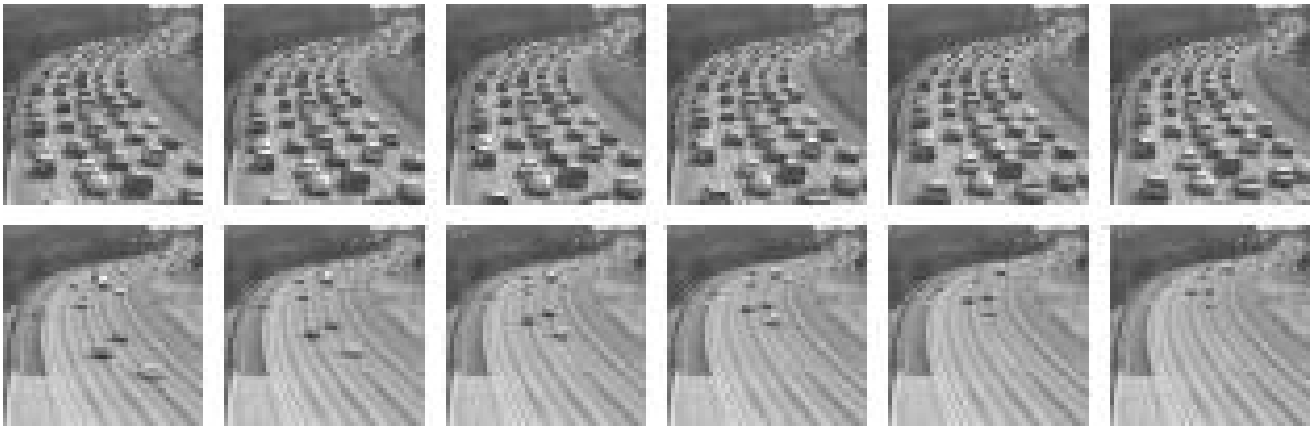
(c) heavy congestion

Figure 2. Example correct classifications for traffic congestion categorization. In each subfigure, the first row shows several frames from an input sequence and the second row shows the corresponding nearest match in the database.

information in the form of a background image that would distinguish the presence vs. absence of vehicles. Second, extensions related to detecting anomalous traffic behaviour, such as detecting the presence of vehicles moving in the wrong direction (cf. [31]), is another interesting direction for future research. Third, attention in the current system has been limited to vehicle-based scene elements. The analysis of the movement of pedestrians and crowds is an inter-



(a) heavy congestion misclassified as medium



(b) heavy congestion misclassified as light

Figure 3. Example misclassifications for traffic congestion categorization. In each subfigure, the first row shows several frames from an input sequence and the second row shows the corresponding nearest match in the database.

esting domain for further application of the proposed system (cf. [2]).

In summary, this paper has presented a system for recognizing traffic congestion scenarios based on the underlying pattern dynamics. The approach is based on a distributed characterization of visual spacetime in terms of 3D, (x, y, t), spatiotemporal orientation. Empirical evaluation on a publicly available data set assembled from real world data shows that the proposed system achieves state-of-the-art performance, while being amenable to computationally efficient realization.

Acknowledgements

Portions of this research were funded by an NSERC Discovery Grant to R. Wildes.

References

[1] E. Adelson and J. Bergen. Spatiotemporal energy models for the perception of motion. *JOSA-A*, 2(2):284–299, 1985.

[2] E. Andrade, S. Blunsden, and R. Fisher. Modelling crowd scenes for event detection. In *ICPR*, pages I: 175–178, 2006.

[3] D. Beymer, P. McLauchlan, B. Coifman, and J. Malik. A real time computer vision system for measuring traffic parameters. In *CVPR*, pages 495–501, 1997.

[4] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distribution. *Bull. Calcutta Math. Soc.*, 35:99–110, 1943.

[5] R. Bracewell. *The Fourier Transform and Its Applications*. New York, New York: McGraw-Hill, 2000.

[6] A. Cavallaro, O. Steiger, and T. Ebrahimi. Tracking video objects in cluttered background. *CirSysVideo*, 15(4):575–584, 2005.

[7] A. Chan and N. Vasconcelos. Classification and retrieval of traffic video using auto-regressive stochastic processes. In *IVS*, pages 771–776, 2005.

[8] A. Chan and N. Vasconcelos. Classifying video with kernel dynamic textures. In *CVPR*, 2007.

[9] D. Chetverikov and R. Peteri. A brief survey of dynamic texture description and recognition. In *CORES*, pages 17–26, 2005.

[10] O. Chomat and J. Crowley. Probabilistic recognition of activity using local appearance. In *CVPR*, pages II: 104–109, 1999.

[11] R. Cucchiara, M. Piccardi, and P. Mello. Image analysis and rule-based reasoning for a traffic monitoring system. *Trans. ITS*, 1(2):119–130, 2000.

[12] K. Derpanis, M. Sizintsev, K. Cannons, and R. Wildes. Efficient action spotting based on a spacetime oriented structure representation. In *CVPR*, 2010.

[13] K. Derpanis and R. Wildes. Dynamic texture recognition based on distributions of spacetime oriented structure. In *CVPR*, 2010.

[14] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie. Behavior recognition via sparse spatio-temporal features. In *PETS*, pages 65–72, 2005.

[15] G. Doretto, A. Chiuso, Y. Wu, and S. Soatto. Dynamic textures. *IJCV*, 51(2):91–109, 2003.

[16] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. New York, New York: Wiley, 2001.

[17] M. Fahle and T. Poggio. Visual hyperacuity: Spatio-temporal interpolation in human vision. *Proceedings of the Royal Society of London - B*, 213(1193):451–477, 1981.

[18] D. Fleet. *Measurement of Image Velocity*. Norwell, Massachusetts: Kluwer, 1992.

[19] W. Freeman and E. Adelson. The design and use of steerable filters. *PAMI*, 13(9):891–906, 1991.

[20] G. Granlund and H. Knutsson. *Signal Processing for Computer Vision*. Norwell, Massachusetts: Kluwer, 1995.

[21] D. Heeger. Model for the extraction of image flow. *JOSA-A*, 2(2):1455–1471, 1987.

[22] Y. Jung, K. Lee, and Y. Ho. Content-based event retrieval using semantic scene interpretation for automated traffic surveillance. *Trans. ITS*, 2(3):151–163, 2001.

[23] S. Kamijo, Y. Matsushita, K. Ikeuchi, and M. Sakauchi. Traffic monitoring and accident detection at intersections. *Trans. ITS*, 1(2):108–118, 2000.

[24] D. Koller, K. Daniilidis, and H. Nagel. Model-based object tracking in monocular image sequences of road traffic scenes. *IJCV*, 10(3):257–281, 1993.

[25] D. Koller, J. Weber, T. Huang, J. Malik, G. Ogasawara, B. Rao, and S. Russell. Towards robust automatic traffic scene analysis in real-time. In *ICPR*, pages I:126–131, 1994.

[26] J. Lee and A. Bovik. Estimation and analysis of urban traffic flow. In *ICIP*, pages 1157–1160, 2009.

[27] X. Li and F. Porikli. A hidden Markov model framework for traffic event detection using video features. In *ICIP*, pages V: 2901–2904, 2004.

[28] A. Lipton, H. Fujiyoshi, and R. Patil. Moving target classification and tracking from real time video. In *WACV*, pages 8–14, 1998.

[29] D. Magee. Tracking multiple vehicles using foreground, background and motion models. *IVC*, 22(2):143–155, 2004.

[30] B. Maurin, O. Masoud, and N. Papanikolopoulos. Monitoring crowded traffic scenes. In *ICITS*, pages 19–24, 2002.

[31] G. Monteiro, M. Ribeiro, J. Marcos, and J. Batista. Wrong-way drivers detection based on optical flow. In *ICIP*, pages V: 141–144, 2007.

[32] F. Porikli and X. Li. Traffic congestion estimation using HMM models without vehicle tracking. In *IVS*, pages 188–193, 2004.

[33] Y. Rubner, J. Puzicha, C. Tomasi, and J. Buhmann. Empirical evaluation of dissimilarity measures for color and texture. *CVIU*, 84(1):25–43, 2001.

[34] E. Simoncelli. *Distributed Analysis and Representation of Visual Motion*. PhD thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 1993.

[35] T. Tan, G. Sullivan, and K. Baker. Model-based localization and recognition of road vehicles. *IJCV*, 27(1):5–25, 1998.

[36] M. Tuceryan and A. Jain. Texture analysis. In C. Chen, L. Pau, and P. Wang, editors, *Handbook of Pattern Recognition and Computer Vision (2nd Edition)*. River Edge, New Jersey: World Scientific Publishing, 1998.

[37] V. Vapnik. *The Nature of Statistical Learning Theory*. New York, New York: Springer, 1995.

[38] *Traffic Detector Handbook*, page 338. Institute Transportation Engineers, 1990.

[39] A. Watson and A. Ahumada. A look at motion in the frequency domain. In *Motion Workshop*, pages 1–10, 1983.

[40] R. Wildes and J. Bergen. Qualitative spatiotemporal analysis using an oriented energy representation. In *ECCV*, pages 768–784, 2000.

[41] X.-D. Yu, L.-Y. Duan, and Q. Tian. Highway traffic information extraction from skycam MPEG video. In *ITS*, pages 37–42, 2002.

[42] A. Zaharescu and R. Wildes. Anomalous behaviour detection using spatiotemporal oriented energies, subset inclusion histogram comparison and event-driven processing. In *ECCV*, 2010.