Near-Lossless Compression for Large Traffic Networks

Muhammad Tayyab Asif, *Student Member, IEEE,* Kannan Srinivasan, *Member, IEEE,* Nikola Mitrovic, *Student Member, IEEE,* Justin Dauwels*, *Senior Member, IEEE,* and Patrick Jaillet

***Abstract*—With advancements in sensor technologies, intelligent transportation systems (ITS) can collect traffic data with high spatial and temporal resolution. However, the size of the networks combined with the huge volume of the data puts serious constraints on the system resources. Low-dimensional models can help ease these constraints by providing compressed representations for the networks. In this study, we analyze the reconstruction efficiency of several low-dimensional models for large and diverse networks. The compression performed by low-dimensional models is lossy in nature. To address this issue, we propose a near-lossless compression method for traffic data by applying the principle of lossy plus residual coding. To this end, we first develop low-dimensional model of the network. We then apply Huffman coding in the residual layer. The resultant algorithm guarantees that the maximum reconstruction error will remain below a desired tolerance limit. For analysis, we consider a large and heterogeneous test network comprising of more than 18000 road segments. The results show that the proposed method can efficiently compress data obtained from a large and diverse road network, while maintaining the upper bound on the reconstruction error.**

***Index Terms*—Low-dimensional models, near-lossless compression.**

I. Introduction

Advancements in sensor technologies have enabled intelligent transportation systems (ITS) to collect traffic information with high spatial and temporal resolution [1]–[3]. This has led to the development of data driven algorithms for many traffic related applications such as sensing, traffic prediction, estimation, and control [4]–[14]. However, the huge size of collected data poses another set of challenges. These challenges include storage and transmission of these large data sets in an efficient manner. Moreover, nowadays many ITS applications need to transfer information to remote mobile devices [15], [16]. These devices usually have limited storage capacity and may also have limited bandwidth. In such scenarios, efficient data compression can be considered as a major requirement for optimal system operation.

We propose to utilize spatial and temporal patterns for efficient compression of traffic data sets. Traffic parameters tend to exhibit strong spatial and temporal correlations [7], [17]. These spatial and temporal relationships can help to obtain a low-dimensional representation for a given network. The low-dimensional models have been previously considered for applications related to feature selection [7], [17], [18], missing data imputation [4], [12] and estimation [3], [14]. Low-dimensional models provide compressed state for a given network. Hence, they are naturally suited for data compression.

Previous studies related to traffic data compression have mostly focused on data from a single road or small subnetworks [19]–[22]. Yang et al. compressed flow data obtained from individual links using 2D wavelet transform [19]. They did not provide any numerical analysis regarding the compression efficiency of the method. Shi et al. performed compression of data obtained from a single loop detector using wavelets [21]. Qu et al. applied principal component analysis (PCA) to compress data obtained from a small road network. They analyzed the compression efficiency of PCA for traffic flow data [20]. However, they did not provide any comparison with other low-rank approximation techniques. Hofleitner et al. proposed non-negative matrix factorization (NMF) to obtain low-dimensional representation of a large network consisting of around 2626 road segments [3]. They used NMF to extract spatial and global patterns in the network. However, no analysis in terms of reconstruction efficiency of NMF was provided. In a related study, we compared the reconstruction efficiency of different low-dimensional models which were obtained from matrix and tensor based subspace methods for a network comprising of around 6000 road segments [23].

The previous studies related to low-dimensional models such as in [3], [13], [19], [20], [22]–[26] only consider lossy compression. In lossy compression, there is no bound on the maximum absolute error (MAE). In terms of traffic data sets, lossy compression does not provide any bound on the loss of information during individual time instances. For instance, consider $x(t)$ to be the traffic speed on a certain road at time t. Let us denote the reconstructed speed value, as a result of lossy compression, by $\hat{x}(t)$. Lossy compression schemes provide no guarantee that the absolute reconstruction error $\mid x(t) - \hat{x}(t) \mid$ will remain below a certain threshold δ.

In summary, we can state that previous studies related to traffic data compression have focused on data collected from either individual roads or small networks [19]–[22]. These studies do not typically compare the reconstruction efficiency of various low-dimensional models for a given road network [3], [13], [19], [20]. Furthermore, the proposed methods do not provide any upper bound on the maximum loss of information.

Muhammad Tayyab Asif, K. Srinivasan, Nikola Mitrovic and Justin Dauwels, are with the School of Electrical and Electronic Engineering, College of Engineering, Nanyang Technological University, Singapore (e-mail: muhammad89@e.ntu.edu.sg; jdauwels@ntu.edu.sg).

Patrick Jaillet is with the Department of Electrical Engineering and Computer Science, School of Engineering, and also with the Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. He is also with the Center for Future Urban Mobility, Singapore-MIT Alliance for Research and Technology, Singapore (e-mail: jaillet@mit.edu).

*Justin Dauwels is the corresponding author.

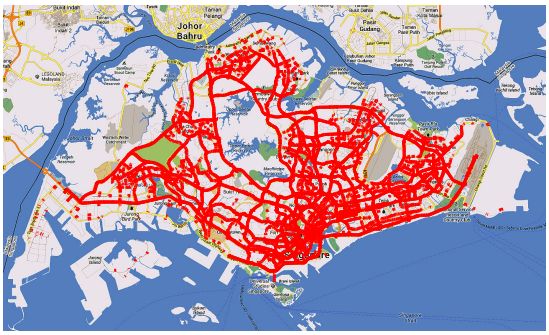


Fig. 1: Test network consisting of 18101 road segments from the country wide road network of Singapore.

CATA	CATB	CATC	Slip Roads	Other
2167	8674	2078	1832	3350

TABLE I: Categories of road segments. The test network comprises of 18101 road segments.

In this study, we compress speed data obtained from a large test network comprising of around 18000 road segments in Singapore. The network consists of a diverse set of roads such as expressways, region around the Changi Airport, industrial, residential, arterial roads etc. We consider different subspace methods such as singular value decomposition (SVD), discrete cosine transform (DCT), wavelet transform and NMF for low-dimensional representation. We compare their reconstruction efficiency for a large and diverse road network. To limit the maximum reconstruction error below a certain threshold δ, we propose a two-step compression algorithm. We first perform lossy compression by developing low-dimensional model for the network. Then we encode the residuals by applying Huffman coding and keeping quantization error of residuals below the specified tolerance limit. The resulting compression algorithm guarantees that the maximum loss of information at any time instance and for any road segment will remain below a certain tolerance level. We analyze the performance of the proposed algorithm for different road categories as well as during weekdays and weekends.

The rest of the paper is structured as follows. In section II, we explain the data set and different performance measures. In section III, we discuss several low-dimensional models for traffic data compression. In section IV, we explain the idea of two-step encoding for traffic data sets. We discuss the results in section V. In section VI, we summarize the contributions and conclude the paper.

II. Data Set and Performance Measures

In this section, we explain the data set used in this study. We also provide different performance measures that we will use for analysis in later sections.

A. Data set

We represent the road network shown in Fig. 1 by a directed graph $G=(N,E)$, where the set E consists of p road segments such that $E=\{s_i\}_{i=1}^{p}$. Here s_i represents a road segment. The set N contains the list of nodes in the graph. For this study, we consider average speed data. Let $z(s_i,t_j)$ represent the average speed on the road segment s_i during the time interval $(t_j-\Delta t,t_j)$ where $\Delta t=5$ minutes. For each road s_i, we create speed profile $\mathbf{a}_i\in\mathbb{R}^n$ where $\mathbf{a}_i=[z(s_i,t_1)...z(s_i,t_n)]^T$. These speed profiles are stacked together to obtain the network profile $\mathbf{A}\in\mathbb{R}^{n\times p}$ where $\mathbf{A}=[\mathbf{a}_1...\mathbf{a}_p]$. Fig. 2 shows the example of a low-dimensional representation $\hat{\mathbf{A}}$ of the network profile matrix $\mathbf{A}$. For the analysis, we consider continuous speed data from the month of August, 2011. The data was provided by Land Transportation Authority (LTA) of Singapore.

The test network is composed of a diverse set of road segments belonging to roads from different categories as well as different regions (residential, industrial, downtown, around the airport etc.). Table I shows the number of roads belonging to each category as defined by LTA. Expressways are assigned to category A (CATA), where as major and minor arterial roads belong to CATB and CATC, respectively. The primary access and local access roads are referred as other in Table I.

B. Performance measures

Now, let us consider different performance measures. We calculate the relative error between the actual network profile $\mathbf{A}$ and the estimated network profile $\hat{\mathbf{A}}$ as:

$$\text{Relative Error}=\frac{\|\mathbf{A}-\hat{\mathbf{A}}\|_F}{\|\mathbf{A}\|_F}, \tag{1}$$

where $\|\mathbf{A}\|_F$ is the Frobenius norm of matrix $\mathbf{A}$ and calculated as:

$$\|\mathbf{A}\|_F=\left(\sum_{i=1}^{n}\sum_{j=1}^{p}a_{ij}^2\right)^{1/2}. \tag{2}$$

The relative error provides a measure of loss of signal energy $\mathbf{A}-\hat{\mathbf{A}}$ as compared to the original network profile $\mathbf{A}$. We also calculate mean absolute percentage error (MAPE) as follows:

$$\text{MAPE}=\frac{1}{np}\sum_{i=1}^{n}\sum_{j=1}^{p}\left|\frac{a_{ij}-\hat{a}_{ij}}{a_{ij}}\right|\times 100\%, \tag{3}$$

where $\hat{a}_{ij}$ is the reconstructed speed value $\hat{z}(s_j,t_i)$ for link s_j at time t_i. We will use these measures to compare the reconstruction efficiency of different subspace methods.

We consider the following measures to analyze the efficiency of the proposed near-lossless compression algorithm. We calculate the maximum absolute error (MAE) for the reconstructed network profile $\hat{\mathbf{A}}$ as:

$$\text{MAE}(\mathbf{A},\hat{\mathbf{A}})=\max_{i,j}\ |a_{ij}-\hat{a}_{ij}|. \tag{4}$$

We also calculate the peak signal-to-noise ratio (PSNR) as:

$$\text{PSNR}=20\cdot\log_{10}\left(\frac{2^B-1}{\sqrt{\text{MSE}}}\right), \tag{5}$$

where B is the resolution of the data set in bits. The mean square error (MSE) is calculated as:

$$\text{MSE}=\frac{1}{np}\|\mathbf{A}-\hat{\mathbf{A}}\|_F^2. \tag{6}$$

PSNR is commonly used in the domain of image processing to evaluate the performance of compression algorithms [27].

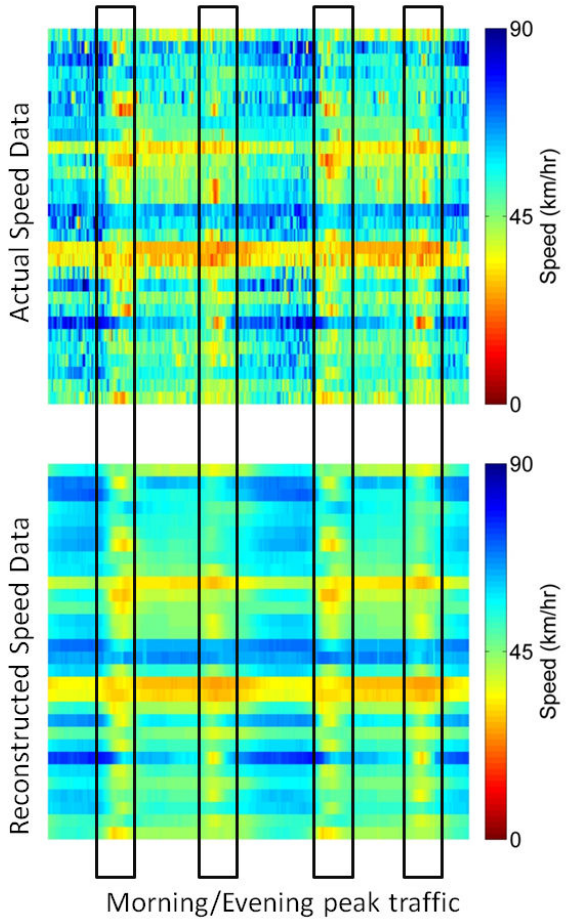


Fig. 2: Actual and reconstructed speed profiles of a subset of road segments from $\mathbf{A}^T$ and $\hat{\mathbf{A}}^T$, respectively. The colorbars represent the colors corresponding to different speed values. A row in the image represents the speed values (actual $\mathbf{a}_i^T$ and reconstructed $\hat{\mathbf{a}}_i^T$) of an individual link for 1st and 2nd August, 2011.

III. Low-Dimensional Models

In this section, we briefly discuss various subspace methods to obtain low-dimensional representations for large road networks. To this end, we will consider the following methods: singular value decomposition (SVD), 2D discrete cosine transform (DCT), 2D wavelet transform and non-negative matrix factorization (NMF). We compare their reconstruction efficiency in terms of the number of elements Θ required to reconstruct a particular low-dimensional representation $\hat{\mathbf{A}}$. We define the element ratio as:

$$\text{Element Ratio(ER)} = \frac{np}{\Theta}, \tag{7}$$

where np represents the total number of elements in the network profile matrix $\mathbf{A}$.

A. Singular Value Decomposition

Singular value decomposition based methods have found applications in many ITS applications including missing data imputation [4], [12] and estimation [13]. By applying SVD, network profile matrix $\mathbf{A}$ can be represented as $\mathbf{A} = \mathbf{USV}^T$, where the columns of matrix $\mathbf{U} \in \mathbb{R}^{n\times n}$ and matrix $\mathbf{V} \in \mathbb{R}^{p\times p}$ are called the left singular vectors and the right singular vectors of $\mathbf{A}$ respectively. The matrix $\mathbf{S} \in \mathbb{R}^{n\times p}$ is a diagonal matrix containing $min(n,p)$ singular values of the network profile matrix $\mathbf{A}$.

The left singular vectors can be obtained by performing eigenvalue decomposition of $\mathbf{AA}^T$ such that $\mathbf{AA}^T = \mathbf{U}\Lambda\mathbf{U}^T$. The matrix Λ contains eigenvalues of $\mathbf{AA}^T$ where $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. Similarly the right singular vectors can be obtained by performing eigenvalue decomposition of $\mathbf{A}^T\mathbf{A}$ such that $\mathbf{A}^T\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^T$. The singular values $\{\sigma_i\}_{i=1}^{min(n,p)}$ are calculated as $\{\sigma_i = \sqrt{\lambda_i}\}_{i=1}^{min(n,p)}$, where λ_i is the i^{th} diagonal entry of Λ. Furthermore, we can obtain the rank-r $(r \leq min(n,p))$ approximation of $\mathbf{A}$ as:

$$\hat{\mathbf{A}} = \sum_{i=1}^{r} \sigma_i \, \mathbf{u}_i \otimes \mathbf{v}_i, \tag{8}$$

where $\mathbf{u}_i \otimes \mathbf{v}_i = \mathbf{u}_i\mathbf{v}_i^T$ and σ_i is the i^{th} diagonal entry of $\mathbf{S}$. The vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ are the columns of matrices $\mathbf{U}$ and $\mathbf{V}$ respectively.

For traffic related applications, the matrix $\mathbf{A}^T\mathbf{A}$ can be interpreted as the covariance matrix for the road segments $\{s_i\}_i^p$ in the network G. Consequently, if the traffic patterns $\{\mathbf{a}_i\}_{i=1}^p$ between the road segments $\{s_i\}_{i=1}^p$ are highly correlated then the network profile $\mathbf{A}$ can be compressed with high efficiency. We perform lossy compression, by storing an appropriate low-rank approximation obtained from (8). To this end, we need to store r columns each from the matrices $\mathbf{U}$ and $\mathbf{V}$ and r elements from the matrix $\mathbf{S}$. Hence, the total number of stored elements will be $\Theta = (n+p+1)r$.

B. 2D Discrete Cosine Transform

In SVD based decomposition, we obtain the basis vectors $\{\mathbf{v}_i\}_{i=1}^p$ from the covariance matrix $\mathbf{A}^T\mathbf{A}$ of road segments in the network. Consequently, we need to store the matrices $\mathbf{U}$ and $\mathbf{V}$ along with the singular values $\{\sigma_i\}_{i=1}^r$ for decompression. In 2D DCT, we consider the cosine family as the basis set and use these basis functions to transform the network profile $\mathbf{A}$ into the so called frequency domain with matrix $\mathbf{Y} \in \mathbb{R}^{n\times p}$ containing the frequency coefficients. For the transformation, let us represent the speed $z(s_{j+1}, t_{i+1})$ for link s_{j+1} at time t_{i+1} as m_{ij} such that $m_{ij} = z(s_{j+1}, t_{i+1})$. We can then calculate the transformed coefficients $y_{k_1k_2}$ as:

$$y_{k_1k_2} = \alpha_{k_1}\alpha_{k_2} \sum_{i=0}^{n-1}\sum_{j=0}^{p-1} m_{ij} \cos\left(\frac{k_1(2i+1)\pi}{2n}\right) \cos\left(\frac{k_2(2j+1)\pi}{2p}\right), \tag{9}$$

where $0 \leq k_1 < n$, $0 \leq k_2 < p$. The factors α_{k_1} and α_{k_2} are defined as:

$$\alpha_{k_1} = \begin{cases} \sqrt{\frac{1}{n}} & k_1 = 0 \\ \sqrt{\frac{2}{n}} & k_1 = 1, \ldots n-1, \end{cases} \tag{10}$$

$$\alpha_{k_2} = \begin{cases} \sqrt{\frac{1}{p}} & k_2 = 0 \\ \sqrt{\frac{2}{p}} & k_2 = 1, \ldots p-1. \end{cases} \tag{11}$$

As the basis functions are orthonormal, the inverse transform can be easily calculated as:

$$\hat{m}_{ij} = \sum_{k_1=0}^{n-1} \sum_{k_2=0}^{p-1} \alpha_{k_1} \alpha_{k_2} y_{k_1 k_2} \cos\Big(\frac{k_1(2i+1)\pi}{2n}\Big) \cos\Big(\frac{k_2(2j+1)\pi}{2p}\Big), \tag{12}$$

where $\hat{m}_{ij}$ is the estimated speed value $\hat{z}(s_{j+1}, t_{i+1})$ for link s_{j+1} at time t_{i+1}. Traffic parameters such as speed and flow tend to be highly correlated across the network [12], [28]. Therefore, we expect that most of the information contained in the network profile $\mathbf{A}$ can be represented by considering a small number of frequency components Θ and we can discard the rest of the frequency components $\{y_{k_1 k_2} = 0\}_{k_1 k_2 \notin \Omega}$. Here Ω is the set of the indices of the frequency components used for reconstruction of traffic data [23]. As the basis functions are pre-specified, we only need to store the elements belonging to the set Ω to reconstruct the network profile $\hat{\mathbf{A}}$.

C. Wavelets

Wavelet transforms have been widely used in compression related applications including images [29], [30] and medical data sets such as electroencephalogram (EEG) [31], [32]. Similar to DCT, wavelets also perform compression using a pre-specified basis set. Wavelet based methods have also been applied for compression of traffic related data sets [19], [21], [22], [25]. However, these studies only analyze data obtained from either individual links or small networks [19], [22]. Moreover, these studies have not compared the performance of wavelets with other subspace methods for traffic data sets. In this study, we apply 2D wavelet transform to compress speed data obtained from a large road network.

To choose an appropriate wavelet type, we consider 5 commonly used wavelets and compare their reconstruction efficiency for a large road network. The different wavelet types we consider are *Near Symmetric*, *Bi-orthogonal 3/5*, *Discrete Meyer Daubechies* and *Coiflets* wavelets. We calculate the element ratio for the wavelet transform by taking the ratio of the total number of elements np in the network profile $\mathbf{A}$ and the number of wavelet coefficients Θ used for the reconstruction.

Fig. 3 shows the reconstruction performance of different wavelet types. For our analysis we will consider *Near Symmetric* wavelet as it provides the best reconstruction efficiency.

D. Non-negative matrix factorization

The non-negative matrix factorization (NMF) provides low-rank approximation of a given matrix by constraining the factors to be non-negative. It has found applications in many fields including text mining [33] and transportation systems [3], [17]. In the field of transportation systems, NMF has been applied to applications related to estimation as well as prediction [3], [17]. In this study, we will focus on the reconstruction efficiency of NMF for large-scale networks. For the network profile $\mathbf{A}$, the NMF optimization problem is

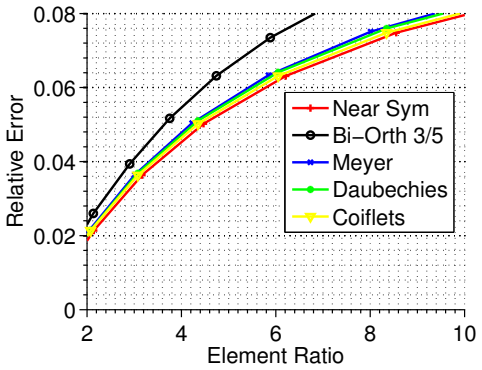


Fig. 3: Reconstruction efficiency of different wavelet types.

defined as:

$$\begin{aligned} \min \;\; & f(\mathbf{W},\mathbf{H}) = \frac{1}{2}\|\mathbf{A} - \mathbf{W}\mathbf{H}\|_F^2, \\ \text{s.t}: \;\; & w_{ij} \geq 0 \;\; \forall\, i,j\,, \\ & h_{ij} \geq 0 \;\; \forall\, i,j\,, \end{aligned} \tag{13}$$

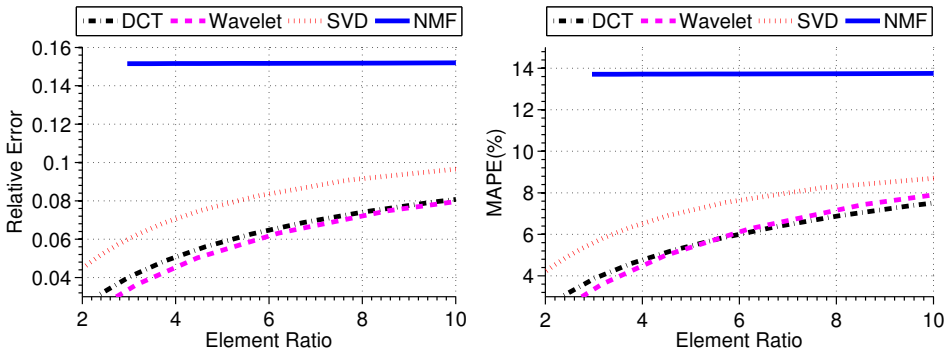
where w_{ij} and h_{ij} are the elements of matrices $\mathbf{W} \in \mathbb{R}^{n \times r}$ and $\mathbf{H} \in \mathbb{R}^{r \times p}$ respectively. Furthermore, matrices $\mathbf{H}$ and $\mathbf{W}$ both have rank r, where $r \leq \; min(n,p)$. Since traffic parameters such as speed, flow and travel time take on non-negative values, NMF may prove suitable for obtaining row-rank approximations for traffic data sets. The non-negative nature of factors $\mathbf{W}$ and $\mathbf{H}$ can potentially provide better interpretability for the underlying model $\hat{\mathbf{A}} = \mathbf{W}\mathbf{H}$ [3], [34]. The optimization problem in (13) is typically solved using the multiplicative update algorithm [3]. To reconstruct a given compressed state, we need to store $\Theta = (n+p)r$ elements.

E. Performance Comparison

In this section, we compare the reconstruction efficiency of the four low-dimensional models presented in the previous section. Fig. 4 shows the performance of these low-dimensional models for different element ratios (ER). DCT and wavelets have comparable performance in terms of both relative error and MAPE. For DCT and wavelets, we only need to store the transformed coefficients for reconstruction. For SVD, we need to store matrices $\mathbf{U}$, $\mathbf{V}$ as well as singular values. Hence, SVD has a lower element ratio for a particular threshold of relative error.

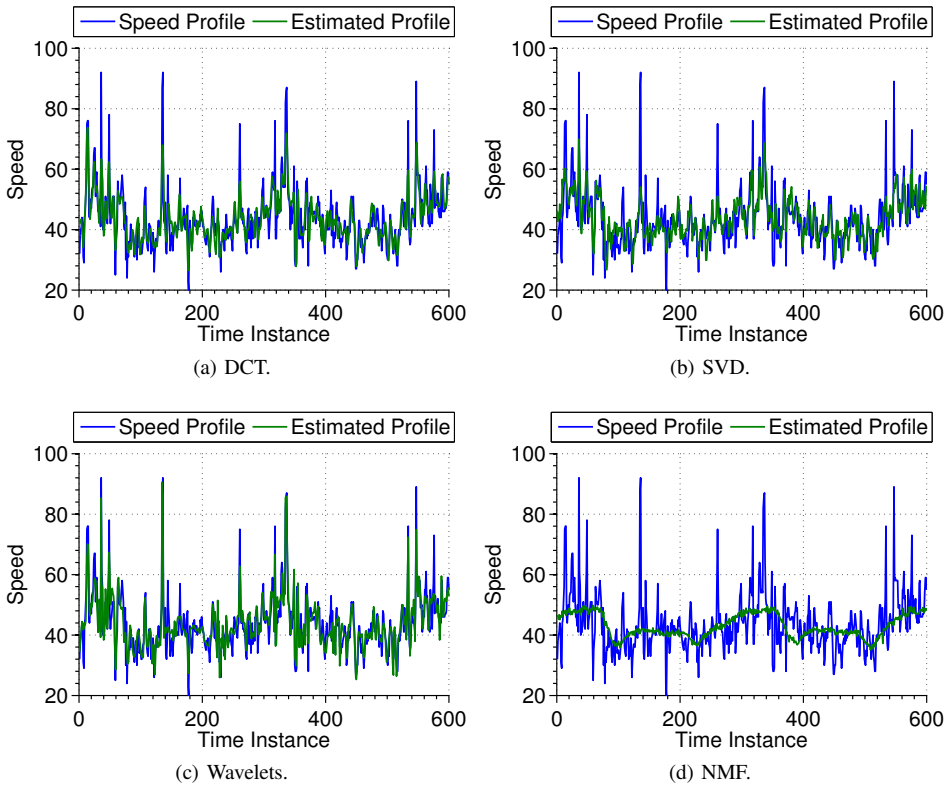
To visualize the reconstruction patterns for these low-dimensional models, we show the data from a typical link in Fig. 5. The figure shows around two days of actual speed data from the link along with the reconstructed speed profile. The ER for this particular representation $\hat{\mathbf{A}}$ is around 9.5 for the four methods.

NMF provides an interesting case. The low-dimensional model obtained from NMF was able to capture the dominant trend in the speed profile. However, the model failed to incorporate localized variations in the speed profile (see Fig. 5d). Moreover, the curve representing the reconstruction error of NMF remains flat for different element ratios (see Fig. 4). We observe that although, NMF can provide reasonable



(a) Relative error. (b) MAPE.

Fig. 4: Comparison of the reconstruction efficiency of different low-dimensional models.



(a) DCT. (b) SVD.

(c) Wavelets. (d) NMF.

Fig. 5: Reconstructed speed profile of a typical road segment using different low-dimensional models. The units for speed values are km/h. The speed data in this figure is taken from 1st and 2nd August, 2011.

reconstruction performance with a small number of factors, it cannot model the local variations efficiently.

Lossy reconstruction does not provide any guarantee that the maximum loss of information $|z(s_i,t_j) - \hat{z}(s_i,t_j)|$ will remain below a certain tolerance limit δ. For instance, although it may seem from the example that the maximum error for DCT, wavelets and SVD should be small. However, all four methods reported MAE in excess of 60 km/h for the reconstructed network profile $\hat{\mathbf{A}}$. In the next section, we propose a two-step compression algorithm to mitigate this issue by performing near-lossless compression for traffic data sets.

IV. Near-Lossless Data Compression

In this section, we discuss a two-step algorithm for near-lossless compression of traffic data obtained from large networks. Fig. 6 shows the steps involved during the encoding and decoding phases. We now briefly discuss the design of encoder and decoder for near-lossless compression of traffic data sets.

A. Encoder

The encoder consists of two stages: (1) lossy compression and (2) residual coding to keep the maximum reconstruction error below the tolerance limit δ. We start by creating the

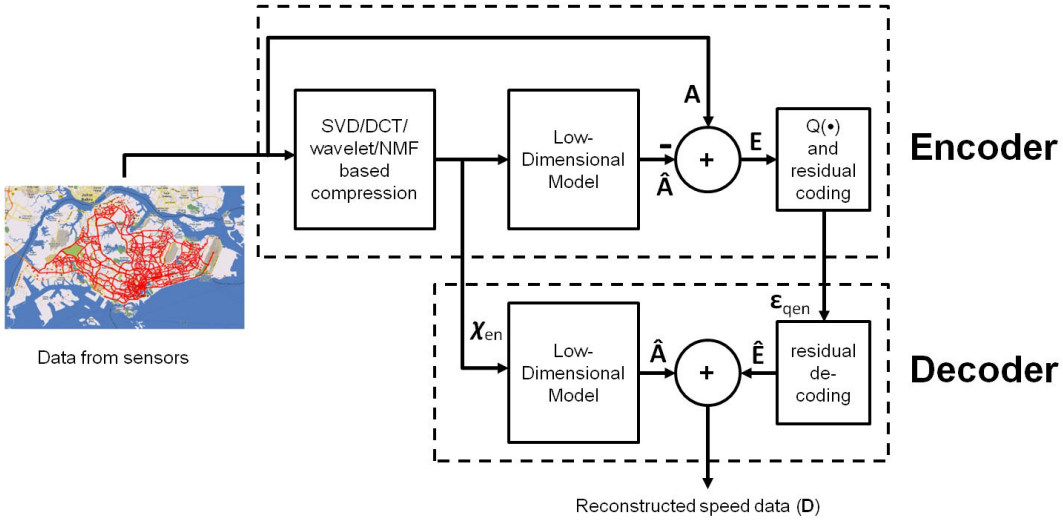


Fig. 6: Schematic block diagram of the near-lossless compression of traffic data obtained from large road networks. $Q(\cdot)$ refers to the quantization scheme used for residuals (see (15)).

network profile $\mathbf{A}$ for traffic speed data obtained from the set of links $\{s_i | s_i \in E\}$. In the first step, we perform lossy compression by means of different subspace methods such as DCT, SVD, wavelets, and NMF. The bitstream χ_{en} represents the compressed representation obtained from these methods. For SVD the factors $\{(\mathbf{u}_i, \sigma_i, \mathbf{v}_i)\}_{i=1}^{r}$ can be encoded in straight forward manner. The same goes for factors $(\mathbf{W}, \mathbf{H})$ obtained from NMF. For DCT and wavelets, techniques such as run length encoding or sparse representation can be used to store the appropriate set of coefficients and their positions Ω. Many ITS applications such as feature selection, estimation and prediction only require transformed set of variables instead of the complete data set [13], [17], [18], [35]–[38]. These applications can directly use the compressed state χ_{en} instead of actual network profile $\mathbf{A}$ or the decompressed representation $\hat{\mathbf{A}}$. However, by only storing χ_{en}, we cannot control the maximum loss of information $|z(s_j,t_i) - \hat{z}(s_j,t_i)|$. In the second step, we encode the residual error so that the maximum reconstruction error remains below the tolerance limit δ. The elements e_{ij} of the residual matrix $\mathbf{E}$ are calculated as:

$$e_{ij} = z(s_j,t_i) - \hat{z}(s_j,t_i), \tag{14}$$

where $\hat{z}(s_j,t_i) = \hat{a}_{ij}$ represents the speed value obtained from the low-dimensional representation $\hat{\mathbf{A}}$. We then obtain the quantized residuals $\hat{e}_{ij} = Q(e_{ij}, \delta)$ as follows:

$$\hat{e}_{ij} = \begin{cases} 0 & |e_{ij}| \leq \delta \\ \lfloor e_{ij} \rceil & \text{otherwise.} \end{cases} \tag{15}$$

where $\lfloor \cdot \rceil$ rounds the argument to the nearest integer value. In this quantization scheme, we discard the residuals e_{ij} whose absolute value is equal to or less than the tolerance limit. For the rest of the residuals $\{e_{ij}\}_{|e_{ij}|>\delta}$, we round them to the nearest integer value. We apply Huffman coding to encode the quantized residuals $\hat{e}_{ij}$ in a lossless manner [39]. We represent the resultant bitstream by ε_{qen}. Let us now briefly discuss the decoder design.

Low-Dim model	DCT	NMF	SVD	Wavelets	HC
CR	**1.66**	1.57	1.65	1.62	1.2

TABLE II: Compression ratios of different methods for $\delta = 0$. HC refers to the case, in which Huffman coding is directly applied to perform lossless compression.

B. Decoder

At the decoder end, we decompress the streams χ_{en} and ε_{qen} to obtain the matrices $\hat{\mathbf{A}}$ and $\hat{\mathbf{E}}$ respectively. Decompressing the bitstream χ_{en} will yield the low-dimensional representation $\hat{\mathbf{A}}$. Decompressing ε_{qen} will provide the quantized residuals stored in matrix $\hat{\mathbf{E}}$. Consequently, the reconstructed network profile will be $\mathbf{D} = \hat{\mathbf{A}} + \hat{\mathbf{E}}$. The decompressed speed value d_{ij} for a link s_j at time instance t_i will be:

$$d_{ij} = \hat{z}(s_j,t_i) + \hat{e}_{ij}. \tag{16}$$

The maximum absolute error (MAE) between the network profile $\mathbf{A}$ and the reconstructed profile $\mathbf{D}$ can be calculated as:

$$\begin{aligned} \text{MAE}(\mathbf{A},\mathbf{D}) &= \max_{i,j} |a_{ij} - d_{ij}|, \\ &= \max_{i,j} |(a_{ij} - \hat{a}_{ij}) - (d_{ij} - \hat{a}_{ij})|, \\ &= \max_{i,j} |e_{ij} - \hat{e}_{ij}|, \\ &\leq \delta. \end{aligned} \tag{17}$$

Selecting different tolerance limits will result in different compression ratios. We calculate the compression ratio (CR) as:

$$\text{CR} = \frac{\text{L}_{\text{org}}}{\text{L}_{\text{comp}}}, \tag{18}$$

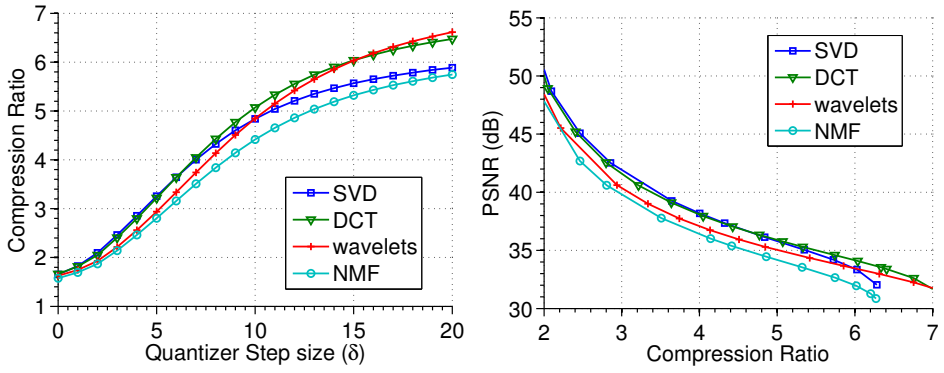
where L_{org} and L_{comp} represent the bitstream lengths of the original and compressed sources respectively. For calculating CR, we set the resolution of speed data to be $B = 8$ *bits*.

TABLE III: Near-lossless compression performance of low-dimensional models for different tolerance levels $\delta \in \{1\,\text{km/h}, \ldots, 15\,\text{km/h}\}$.

Low-dimensional Model	$\delta = 1$		$\delta = 2$		$\delta = 3$		$\delta = 4$		$\delta = 5$	
	CR	PSNR	CR	PSNR	CR	PSNR	CR	PSNR	CR	PSNR
DCT	1.82	54.23	2.05	48.93	2.40	45.20	2.80	42.55	3.22	40.59
SVD	1.83	53.95	2.10	48.69	2.46	45.07	2.86	42.53	3.26	40.66
NMF	1.69	54.56	1.87	49.29	2.14	45.46	2.46	42.69	2.81	40.61
Wavelets	1.75	54.64	1.93	49.37	2.21	45.52	2.56	42.72	2.94	40.62
Low-dimensional Model	$\delta = 6$		$\delta = 7$		$\delta = 8$		$\delta = 9$		$\delta = 10$	
	CR	PSNR	CR	PSNR	CR	PSNR	CR	PSNR	CR	PSNR
DCT	3.64	39.10	4.05	37.95	4.43	37.05	4.77	36.34	5.07	35.76
SVD	3.64	39.26	4.00	38.18	4.32	37.34	4.60	36.67	4.84	36.14
NMF	3.16	39.01	3.51	37.77	3.84	36.80	4.14	36.02	4.42	35.39
Wavelets	3.34	39.01	3.74	37.74	4.14	36.74	4.51	35.94	4.85	35.30
Low-dimensional Model	$\delta = 11$		$\delta = 12$		$\delta = 13$		$\delta = 14$		$\delta = 15$	
	CR	PSNR	CR	PSNR	CR	PSNR	CR	PSNR	CR	PSNR
DCT	5.33	35.30	5.55	34.91	5.74	34.59	5.90	34.32	6.04	34.09
SVD	5.04	35.71	5.21	35.35	5.35	35.05	5.47	34.79	5.57	34.57
NMF	4.65	34.88	4.86	34.45	5.04	34.10	5.19	33.80	5.32	33.55
Wavelets	5.15	34.78	5.42	34.35	5.65	33.98	5.85	33.68	6.03	33.41

TABLE IV: PSNR for different road categories.

Low-dimensional Model	CR	**Tol.** δ	CATA	CATB	CATC	Slip Roads	Other
DCT	3.22	5	41.09	40.41	40.42	40.81	40.77
SVD	3.26		41.59	40.45	40.41	40.84	40.75
NMF	2.81		40.80	40.45	40.47	40.88	40.84
Wavelets	2.94		40.60	40.49	40.50	40.89	40.90
Low-dimensional Model	CR	**Tol.** δ	CATA	CATB	CATC	Slip Roads	Other
DCT	5.07	10	37.20	35.70	35.47	35.50	35.47
SVD	4.84		37.95	36.12	35.77	35.78	35.72
NMF	4.42		35.54	35.41	35.32	35.22	35.36
Wavelets	4.85		35.62	35.31	35.22	35.11	35.24



(a) Compression ratio and quantizer step size δ (km/h). (b) PSNR for different compression ratios.

Fig. 7: Performance of near-lossless compression methods based on different low-dimensional models (SVD/DCT/wavelets/NMF).

In the next section, we analyze the performance of the proposed near-lossless compression method for different tolerance limits.

V. Results and Discussion

In this section, we analyze the performance of the proposed algorithms for the test network. We also compare the compression efficiency of these algorithms for different road categories, weekdays, and weekends. All the compression ratios are calculated by keeping the resolution of field data to $B = 8$ *bits*. Considering higher resolution for field data will automatically result in higher albeit inflated CR. Table II shows the compression ratios for different subspace methods by setting $\delta = 0$. If we ignore the rounding off error due to (15), then this scenario can be considered as the lossless case. If we directly apply Huffman coding (HC) on the network profile **A**, we obtain a CR of 1.2. Hence, the two step compression method straight away provides around 30 % to 38 % (CR: $1.57 - 1.66$) improvement in CR. Table III shows the CR and PSNR achieved by the four low-dimensional models for different tolerance levels. For tolerance level of 5 km/h, the algorithms yield a CR of around 3. In this case, the SVD based method provides the highest compression ratio and PSNR. The other three methods also report similar PSNR. DCT has comparable CR to that of SVD. The other two methods (wavelets and NMF) report slightly smaller CR for this particular tolerance level. For tolerance level of 15 km/h, DCT and wavelets achieve CR of more than 6 (see Table III). While SVD has the best PSNR for this tolerance level, it reports slightly smaller CR (5.57) as compared to DCT (6.04) and wavelets (6.03).

To observe these trends in more detail, we plot the compression ratios achieved by these algorithms against different tolerance levels in Fig. 7a. All the algorithms report similar CR for tolerance levels close to zero. For higher tolerance levels, DCT and wavelets achieve higher CR as compared to matrix decomposition methods (SVD, NMF). For low tolerance levels, residual coding can easily compensate for the inefficiency of a particular low-dimensional model. For higher CR, such inefficiencies become prominent. The tolerance level (MAE) alone does not provide all the information about the overall reconstruction error. Two methods with the same MAE can still have different mean square error. To this end, we plot PSNR against the compression ratios for these methods. PSNR values for different compression ratios are shown in Fig. 7b. The plot shows that SVD provides similar PSNR to DCT for a given CR. However, SVD yields higher maximum error as compared to DCT (see Fig. 7a).

Table IV shows the category wise analysis of different near-lossless compression methods for tolerance levels of 5 km/h and 10 km/h. Expressways (CATA roads) achieve higher PSNR as compared to roads from CATB and CATC for all four algorithms. This difference in PSNR is more visible for $\delta = 10$ as opposed to $\delta = 5$. For $\delta = 10$, expressways have the highest PSNR in comparison with CATB, CATC, slip roads and other categories for each algorithm. Nonetheles, the PSNR gain for CATA as compared to other categories varies from one algorithm to another. These values have some intuitive sense as well. Normally, traffic on expressways behaves much more smoothly in comparison with other roads. Therefore, subspace methods can model traffic conditions on expressways with high accuracy.

TABLE V: PSNR for different methods during weekdays and weekends for the month of August, 2011.

Low-dimensional Model	CR	**Tol.** δ	Weekdays	Weekends
DCT	3.22	5	40.60	40.57
SVD	3.26		40.66	40.66
NMF	2.81		40.63	40.54
Wavelets	2.94		40.65	40.54
Low-dimensional Model	CR	**Tol.** δ	Weekdays	Weekends
DCT	5.07	10	35.73	35.86
SVD	4.84		36.12	36.21
NMF	4.42		35.38	35.43
Wavelets	4.85		35.28	35.38

Table V shows the PSNR values for weekdays and weekends for different algorithms. We observe that reconstructed data for both weekdays and weekends has similar PSNR. This trend is observed for all the compression algorithms. Weekdays and weekends tend to have distinct traffic patterns. However, the similar compression performance indicates that traffic patterns across different time periods (from one week to another) remain quite similar. Consequently, traffic data sets can be easily compressed (see Tables III and V).

Various studies [40]–[42] have identified management and storage of traffic data as one of the major big data problems currently faced by ITS. The proposed compression methods may prove to be useful for efficient storage and management of traffic data for systems dealing with large-scale networks. Furthermore, many ITS applications such as feature selection, estimation and prediction only require information related to transformed variables instead of the actual data. These applications can directly utilize the compressed data streams obtained from the first step.

VI. Conclusion

In this paper, we proposed a near-lossless compression algorithm to compress traffic data obtained from a large and diverse network. In the first step, we developed low-dimensional model of the network. Then, we encoded the residuals by applying Huffman coding. The resulting algorithm guarantees that the loss of information, for any link at any given time, will remain below a pre-specified threshold. To develop the low-dimensional model, we considered different subspace methods including 2D wavelets, 2D DCT transform, SVD and NMF. These methods have been previously applied to many transportation related problems such as prediction, estimation and feature selection. However, these studies typically do not compare the reconstruction performance of different methods for a given network.

In this study, we analyzed the reconstruction efficiency of different subspace methods for a large and diverse network which consisted of more than 18000 road segments. We then applied these methods to perform near-lossless compression of the speed data. We also analyzed the performance of the proposed algorithm for different road categories, weekdays and weekends by considering different low-dimensional methods. We conclude that low-dimensional based near-lossless compression methods can efficiently compress traffic data sets obtained from large and diverse road networks.

ACKNOWLEDGMENT

This work was supported in part by the Singapore National Research Foundation through the Singapore Massachusetts Institute of Technology Alliance for Research and Technology (SMART) Center for Future Mobility (FM).

LIST OF ABBREVIATIONS

CR	Compression ratio
DCT	Discrete cosine transform
EEG	Electroencephalogram
ER	Element ratios
HC	Huffman coding
ITS	Intelligent transportation systems
LTA	Land transportation authority
MAE	Maximum absolute error
MAPE	Mean absolute percentage error
MSE	Mean square error
NMF	Non-negative matrix factorization
PCA	Principal component analysis
PSNR	Peak signal-to-noise ratio
SVD	Singular value decomposition

REFERENCES

[1] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12, no. 4, pp. 1624–1639, 2011.

[2] P. J. Bickel, C. Chen, J. Kwon, J. Rice, E. Van Zwet, and P. Varaiya, "Measuring traffic," *Statistical Science*, vol. 22, no. 4, pp. 581–597, 2007.

[3] A. Hofleitner, R. Herring, A. Bayen, Y. Han, F. Moutarde, and A. de La Fortelle, "Large-scale estimation of arterial traffic and structural analysis of traffic patterns from probe vehicles," in *Transportation Research Board 91st Annual Meeting*, no. 12-0598, 2012.

[4] M. T. Asif, N. Mitrovic, L. Garg, J. Dauwels, and P. Jaillet, "Low-dimensional models for missing data imputation in road networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 3527–3531.

[5] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, "Using mobile phones to determine transportation modes," *ACM Transactions on Sensor Networks (TOSN)*, vol. 6, no. 2, p. 13, 2010.

[6] A. Hofleitner, R. Herring, P. Abbeel, and A. Bayen, "Learning the dynamics of arterial traffic from probe data using a dynamic bayesian network," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 13, no. 4, pp. 1679–1693, 2012.

[7] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 606–616, 2011.

[8] M. T. Asif, J. Dauwels, C. Y. Goh, A. Oran, E. Fathi, M. Xu, M. Dhanya, N. Mitrovic, and P. Jaillet, "Spatiotemporal patterns in large-scale traffic speed prediction," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 15, no. 2, pp. 794–804, April 2014.

[9] C. Y. Goh, J. Dauwels, N. Mitrovic, M. T. Asif, A. Oran, and P. Jaillet, "Online map-matching based on hidden markov model for real-time traffic sensing applications," in *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*, 2012, pp. 776–781.

[10] Z. Li, Y. Zhu, H. Zhu, and M. Li, "Compressive sensing approach to urban traffic sensing," in *Distributed Computing Systems (ICDCS), 2011 31st International Conference on*, 2011, pp. 889–898.

[11] L. Li, D. Wen, and D. Yao, "A survey of traffic control with vehicular communications," *Intelligent Transportation Systems, IEEE Transactions on*, vol. PP, no. 99, pp. 1–8, 2013.

[12] L. Qu, L. Li, Y. Zhang, and J. Hu, "Ppca-based missing data imputation for traffic flow volume: A systematical approach," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 10, no. 3, pp. 512–522, 2009.

[13] T. Djukic, G. Flotterod, H. van Lint, and S. Hoogendoorn, "Efficient real time od matrix estimation based on principal component analysis," in *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*. IEEE, 2012, pp. 115–121.

[14] R. Herring, A. Hofleitner, P. Abbeel, and A. Bayen, "Estimating arterial traffic conditions using sparse probe data," in *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*. IEEE, 2010, pp. 929–936.

[15] J. Ribeiro Junior, M. Mitre Campista, and L. Costa, "Cotrams: A collaborative and opportunistic traffic monitoring system," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 15, no. 3, pp. 949–958, June 2014.

[16] J. Rodrigues, A. Aguiar, F. Vieira, J. Barros, and J. P. S. Cunha, "A mobile sensing architecture for massive urban scanning," in *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*, Oct 2011, pp. 1132–1137.

[17] Y. Han and F. Moutarde, "Analysis of network-level traffic states using locality preservative non-negative matrix factorization," in *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*. IEEE, 2011, pp. 501–506.

[18] C. Furtlehner, Y. Han, J.-M. Lasgouttes, V. Martin, F. Marchal, and F. Moutarde, "Spatial and temporal analysis of traffic states on large scale networks," in *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*. IEEE, 2010, pp. 1215–1220.

[19] Y. Xiao, Y.-m. Xie, L. Lu, and S. Gao, "The traffic data compression and decompression for intelligent traffic systems based on two-dimensional wavelet transformation," in *Signal Processing, 2004. Proceedings. ICSP'04. 2004 7th International Conference on*, vol. 3. IEEE, 2004, pp. 2560–2563.

[20] Q. Li, H. Jianming, and Z. Yi, "A flow volumes data compression approach for traffic network based on principal component analysis," in *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*. IEEE, 2007, pp. 125–130.

[21] X.-f. Shi and L.-l. Liang, "A data compression method for traffic loop detectors' signals based on lifting wavelet transformation and entropy coding," in *Information Science and Engineering (ISISE), 2010 International Symposium on*. IEEE, 2010, pp. 129–132.

[22] J. Ding, Z. Zhang, and X. Ma, "A method for urban traffic data compression based on wavelet-pca," in *Computational Sciences and Optimization (CSO), 2011 Fourth International Joint Conference on*. IEEE, 2011, pp. 1030–1034.

[23] M. T. Asif, S. Kannan, J. Dauwels, and P. Jaillet, "Data compression techniques for urban traffic data," in *Computational Intelligence in Vehicles and Transportation Systems (CIVTS), 2013 IEEE Symposium on*, 2013, pp. 44–49.

[24] G. Ke, J. Hu, L. He, and Z. Li, "Data reduction in urban traffic sensor network," in *Intelligent Vehicles Symposium, 2009 IEEE*, June 2009, pp. 1021–1026.

[25] F. Qiao, H. Liu, and L. Yu, "Incorporating wavelet decomposition technique to compress transguide intelligent transportation system data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1968, no. 1, pp. 63–74, 2006.

[26] ——, "Intelligent transportation systems data compression using wavelet decomposition technique," Tech. Rep., 2009.

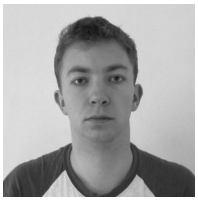
[27] N. Thomos, N. V. Boulgouris, and M. G. Strintzis, "Optimized transmission of jpeg2000 streams over wireless channels," *Image Processing, IEEE Transactions on*, vol. 15, no. 1, pp. 54–67, 2006.

[28] T. Djukic, J. van Lint, and S. Hoogendoorn, "Exploring application perspectives of principal component analysis in predicting dynamic origin-destination matrices," in *Transportation Research Board 91st Annual Meeting*, no. 12-2702, 2012.

[29] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *Image Processing, IEEE Transactions on*, vol. 1, no. 2, pp. 205–220, 1992.
[30] R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *Image Processing, IEEE Transactions on*, vol. 8, no. 12, pp. 1688–1701, 1999.
[31] K. Srinivasan, J. Dauwels, and M. Reddy, "Multichannel EEG compression: Wavelet-based image and volumetric coding approach," *Biomedical and Health Informatics, IEEE Journal of*, vol. 17, no. 1, pp. 113–120, 2013.
[32] J. Dauwels, K. Srinivasan, M. Reddy, and A. Cichocki, "Near-lossless multichannel EEG compression based on matrix and tensor decompositions," *Biomedical and Health Informatics, IEEE Journal of*, vol. 17, no. 3, pp. 708–714, 2013.
[33] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 155–173, 2007.
[34] D. Donoho and V. Stodden, "When does non-negative matrix factorization give a correct decomposition into parts?" in *Advances in Neural Information Processing Systems*, 2004, pp. 1141–1148.
[35] X. Jin, Y. Zhang, and D. Yao, "Simultaneously prediction of network traffic flow based on pca-svr," in *Advances in Neural Networks–ISNN 2007*. Springer, 2007, pp. 1022–1031.
[36] T. T. Tchrakian, B. Basu, and M. O'Mahony, "Real-time traffic flow forecasting using spectral analysis," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 13, no. 2, pp. 519–526, 2012.
[37] X. Jiang and H. Adeli, "Dynamic wavelet neural network model for traffic flow forecasting," *Journal of Transportation Engineering*, vol. 131, no. 10, pp. 771–779, 2005.
[38] Y. Zhang, Y. Zhang, and A. Haghani, "A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model," *Transportation Research Part C: Emerging Technologies*, 2013.
[39] K. Sayood, *Introduction to data compression*. Access Online via Elsevier, 2012.
[40] T. Qu, S. T. Parker, Y. Cheng, B. Ran, and D. A. Noyce, "Large-scale intelligent transportation system traffic detector data archiving," in *Transportation Research Board 93rd Annual Meeting*, no. 14-5448, 2014.
[41] I. Vilajosana, J. Llosa, B. Martinez, M. Domingo-Prieto, A. Angles, and X. Vilajosana, "Bootstrapping smart cities through a self-sustainable model based on big data flows," *Communications Magazine, IEEE*, vol. 51, no. 6, pp. 128–134, June 2013.
[42] H. Jagadish, J. Gehrke, A. Labrinidis, Y. Papakonstantinou, J. M. Patel, R. Ramakrishnan, and C. Shahabi, "Big data and its technical challenges," *Communications of the ACM*, vol. 57, no. 7, pp. 86–94, 2014.

Muhammad Tayyab Asif (S'12) received the B. Sc degree in Electrical Engineering from the University of Engineering and Technology, Lahore, Pakistan. He is currently working toward the Ph.D. degree in the School of Electrical and Electronic Engineering, College of Engineering, Nanyang Technological University, Singapore. Previously, he was with Ericsson as a Design Engineer in the domain of mobile packet core networks. His research interests include sensor fusion, network optimization, and modeling of large-scale networks.

K Srinivasan (M'06) received the B.E degree in electronics and communication from Bharathiar University, Coimbatore, India, in 2004, the M.E. degree in medical electronics from Anna University, Chennai, India, in 2006. From 2007, he was working towards the Ph.D. degree at biomedical engineering group, Department of Applied Mechanics, Indian Institute of Technology (IIT) Madras, and obtained the degree in 2012. He worked as a Visiting Research Scholar at Nanyang Technological University (NTU), Singapore, from August 2010 to April 2011. He is currently working as a Research Fellow at NTU, Singapore. His research interests are biomedical signal compression, EEG signal processing, and interpretation.



Nikola Mitrovic (S'14) received the bachelor degree in traffic engineering from the University of Belgrade, Serbia, in 2009. He obtained masters degree at Department of civil engineering at Florida Atlantic University, USA, in 2010. He is currently a PhD student with the department of Electrical and Electronic engineering at Nanyang Technological University. His research topics are traffic modeling, intelligent transportation systems, and transportation planning.

Justin Dauwels (M'09-SM'12) is an Assistant Professor with School of Electrical & Electronic Engineering at the Nanyang Technological University (NTU) in Singapore. His research interests are in Bayesian statistics, iterative signal processing, and computational neuroscience. He obtained the PhD degree in electrical engineering at the Swiss Polytechnical Institute of Technology (ETH) in Zurich in December 2005. He was a postdoctoral fellow at the RIKEN Brain Science Institute (2006-2007) and a research scientist at the Massachusetts Institute of Technology (2008-2010). He has been a JSPS postdoctoral fellow (2007), a BAEF fellow (2008), a Henri-Benedictus Fellow of the King Baudouin Foundation (2008), and a JSPS invited fellow (2010,2011). His research on Intelligent Transportation Systems (ITS) has been featured by the BBC, Straits Times, and various other media outlets. His research on Alzheimer's disease is featured at a 5-year exposition at the Science Centre in Singapore. His research team has won several best paper awards at international conferences. He has filed 5 US patents related to data analytics.

Patrick Jaillet received the Ph.D. degree in operations research from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1985. He is currently the Dugald C. Jackson Professor of the Department of Electrical Engineering and Computer Science, School of Engineering, and a Codirector of the Operations Research Center, Massachusetts Institute of Technology. His research interests include algorithm design and analysis for online problems, real-time and dynamic optimization, network design and optimization, and probabilistic combinatorial optimization.