Data Compression Techniques for Urban Traffic Data

Muhammad Tayyab Asif*, Srinivasan Kannan*, Justin Dauwels* and Patrick Jaillet †‡
*School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 639798
†Laboratory for Information and Decision Systems, MIT, Cambridge, MA, 02139
‡Center for Future Urban Mobility, Singapore-MIT Alliance for Research and Technology, Singapore

***Abstract*—With the development of inexpensive sensors such as GPS probes, Data Driven Intelligent Transport Systems (D^2ITS) can acquire traffic data with high spatial and temporal resolution. The large amount of collected information can help improve the performance of ITS applications like traffic management and prediction. The huge volume of data, however, puts serious strain on the resources of these systems. Traffic networks exhibit strong spatial and temporal relationships. We propose to exploit these relationships to find low-dimensional representations of large urban networks for data compression. In this paper, we study different techniques for compressing traffic data, obtained from large urban road networks. We use Discrete Cosine Transform (DCT) and Principal Component Analysis (PCA) for 2-way network representation and Tensor Decomposition for 3-way network representation. We apply these techniques to find low-dimensional structures of large networks, and use these low-dimensional structures for data compression.**

I. Introduction

Data Driven Intelligent Transport Systems (D^2ITS) have increasingly found greater role in applications such as traffic management, sensing, route guidance, and prediction [1]–[8]. This has been made possible due to availability of large amount of data, collected by sensors such as GPS probes and loop detectors. D^2ITS require traffic data with high spatial and temporal resolution. Large road networks are typically divided into thousands of road segments. Data acquisition systems usually have temporal resolution of 5 minutes. While the sensors provide detailed information regarding the state of the network, the large volume of data puts strain on resources of data management systems [9]. The huge amount of information also tends to hinder the scalability of many data driven applications dealing with route guidance and prediction [10], [11]. For D^2ITS dealing with large road networks, these problems can potentially compound into serious bottlenecks [12].

Traffic parameters such as speed, flow and travel time usually exhibit strong spatial and temporal correlations [5]. Moreover, there exist certain temporal and spatial trends in traffic data [9], [13]. These relationships have previously been utilized for applications such as traffic prediction [5], sensing [4], data imputation [14] and even data compression [9], [15]. The previous studies related to traffic data compression have mainly focused on a few intersections [9], [13], [15]. Practical implementations require data compression algorithms for large and diverse urban networks. In this study, we examine whether data from a large and diverse urban network can be compressed without losing a lot of information. To this end, we consider a road network in Singapore from Outram park to Changi, comprising of more than 6000 road segments. The network consists of expressways as well as roads in the downtown area, and around Singapore Changi Airport. We propose to exploit the spatial and temporal relations between traffic parameters to construct low-dimensional models for large road networks using different subspace methods. We then use these low-dimensional models for data compression. Singapore Land Transportation Authority (LTA) provided data for experimental purposes. It consists of space averaged speed values for all the road segments for August 2011. In this study, we do not deal with the problem of missing data imputation. We only considered those road segments for which adequate data was available.

We employ methods such as Discrete Cosine transform (DCT), Principal Component Analysis (PCA) and Higher Order Singular Value Decomposition (HO-SVD), to find low-dimensional representations for large networks. Using these low-dimensional representations, we perform data compression. We compare their compression efficiency by calculating reconstruction error for different compression ratios.

The paper is structured as follows. In section II, we briefly explain the data set. We propose different data compression techniques for large networks in section III. In section IV, we compare the performance of proposed methods. In the end (section V), we summarize our contributions and suggest topics for future work.

II. Data Set

We define the road network under study, as a directed graph $G = (N,E)$. $E = \{s_i | i = 1,...p\}$ is the set of road segments (links), where a road segment is denoted by s_i. We represent the weight of each link s_i by $z(s_i,t_j)$, which is a time varying function, representing the average traffic speed during the interval $(t_j - t_0, t_j)$.

For data compression, we consider a large sub network from Outram park to Changi in Singapore (see fig. 1). The network G consists of road segments from three expressways, which are Pan Island Expressway, East Coast Parkway and Kallang Paya Lebar Expressway. It also includes the region around Changi Airport and roads in the downtown area of Singapore carrying

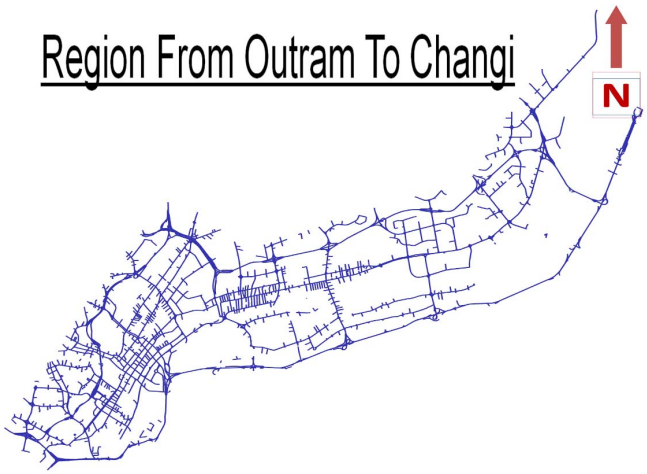

Fig. 1: Road network in Singapore (Outram to Changi).

significant volumes of traffic. The road segments comprising the setup have different speed limits, lanes, and capacities.

We use space averaged speed data provided by LTA from August, 2011. For this study, we consider kilometer/hour (km/hr) as units for traffic speed. Sampling interval for data t_0 is 5 minutes.

III. Data Compression Methods

In this section, we propose different data compression methods for large networks. We consider two different formulations for methods dealing with 2-way (matrix) and 3-way (tensor) compression.

For methods dealing with matrix compression, we consider a network profile $\mathbf{M_G} \in \mathbb{R}^{d\times p}$ for network G. Each link s_i is represented by a speed profile $\mathbf{m}_i$ such that $\{\mathbf{m}_i = [z(s_i,t_1)z(s_i,t_2)...z(s_i,t_d)]^T\}_{s_i\in E}$. Hence, the network profile $\mathbf{M}_G = [\mathbf{m}_1\mathbf{m}_2...\mathbf{m}_p]$ contains speed data for p links from time t_1 to t_d. We use PCA and DCT to obtain low-dimensional representation $\mathbf{X}_G$, of network profile $\mathbf{M}_G$.

Traffic parameters tend to exhibit certain daily, repetitive patterns [14]. To observe the effect of these correlations on compression efficiency, we consider a 3-way structure for network profile. For tensor compression, network profile is defined as a 3-way tensor $\mathscr{M}_G \in \mathbb{R}^{d_t\times p\times w}$, such that speed values $\{z(s_i,t_j^k)\}_{k=1}^{w}$, observed at same time t_j, during different days $\{l_k\}_{k=1}^{w}$, are stacked as entries $\{\mathfrak{m}_{ijk}\}_{k=1}^{w}$ of $\mathscr{M}_G$. We set $d = d_t \times w$, to keep the total information in $\mathbf{M}_G$ and $\mathscr{M}_G$ same. We use HO-SVD to find compressed representation $\mathscr{X}_G$ of network profile $\mathscr{M}_G$.

Based on these definitions, we develop following methods for network data compression.

A. Discrete Cosine Transform

DCT is the workhorse behind JPEG, which is the industry standard for image compression [16], [17]. The network profile $\mathbf{M}_G$, carrying information regarding network G, can also be considered as an image. In this case, speed value $z(s_j,t_i)$ for link s_j at time t_i can be thought of as a pixel such that $\{m_{i'j'} = z(s_j,t_i)\}_{(i,j)=(1,1)}^{(d,p)}$, where $i' = i-1$ and $j' = j-1$. Using DCT, we create transformed representation $\mathbf{N}_G$ of $\mathbf{M}_G$ in the following manner:

$$n_{i''j''} = \alpha_{i''}\alpha_{j''}\sum_{i'=0}^{d-1}\sum_{j'=0}^{p-1} m_{i'j'}\cos(\frac{(2i'+1)\pi i''}{2d})cos(\frac{(2j'+1)\pi j''}{2p}), \tag{1}$$

where $0 \le i'' < d$, $0 \le j'' < p$. $\alpha_{i''}$ and $\alpha_{j''}$ are defined as:

$$\alpha_{i''} = \begin{cases} \frac{1}{\sqrt{d}} & i'' = 0 \quad (2a) \\ \sqrt{\frac{2}{d}} & i'' = 1,...d-1, \quad (2b) \end{cases}$$

$$\alpha_{j''} = \begin{cases} \frac{1}{\sqrt{p}} & j'' = 0 \quad (3a) \\ \sqrt{\frac{2}{p}} & j'' = 1,...p-1. \quad (3b) \end{cases}$$

Fig. 5 shows the magnitude of frequency components $n_{i''j''}$, which are stored in $\mathbf{N}_G$ and plotted on the log scale. The figure shows that there is a sizable portion of frequency components, with small magnitude, which can be neglected. As an example, we consider a small subset of frequency components Ψ from $\mathbf{N}_G$ to reconstruct a low-rank network profile $\mathbf{X}_G$ (see Fig. 3) of $\mathbf{M}_G$ (see Fig. 2), using inverse DCT. For convenience, we show plots for $\mathbf{N}_G^T$, $\mathbf{M}_G^T$ and $\mathbf{X}_G^T$ ($\mathbf{A}^T$ represents transpose of $\mathbf{A}$) in Fig. 5, 2 and 3 respectively. $\mathbf{X}_G$ provides a much smoother picture of the network as compared to $\mathbf{M}_G$. This implies that $\mathbf{X}_G$ failed to capture small variations in speed, which are present in $\mathbf{M}_G$. Fig. 4 shows ${\mathbf{H}_G}^T$, whose entries h_{ij} represent reconstruction error such that $\{h_{ij} = \frac{|\, m_{ij} - x_{ij}\,| \times 100}{m_{ij}}\}_{i=1,j=1}^{i=d,j=p}$. h_{ij} provides us with an estimate of reconstruction error for each speed value $z(s_j,t_i)$. The error distribution in Fig. 4 also reinforces the idea that ${\mathbf{X}_G}^T$ needs more frequency components to capture smaller and relatively local variations in speed profiles.

Nonetheless, $\mathbf{X}_G$ still preserves the main traffic trends such as peak hours and off peak hours. The scales in Fig. 2 and 3, represent colors corresponding to the respective speed values in the figures. The color bar in Fig. 4 represents colors corresponding to the reconstruction percentage error.

We use the following formula to calculate Compression Ratio (CR):

$$CR = \frac{Total\quad number\quad of\quad elements\quad (\Theta)}{number\quad of\quad storage\quad elements\quad (\theta)}. \tag{4}$$

For DCT compression, $\Theta_{DCT} = d \times p$ and $\theta_{DCT} = \Psi$, hence $CR_{DCT} = \frac{d\times p}{\Psi}$. Another advantage offered by DCT is in terms of computational complexity. DCT can be computed in $O(N\log N)$ time, using Fast Fourier Transform (FFT), where $N = n \times m$ for $\mathbf{A} \in \mathbb{R}^{n\times m}$ [17].

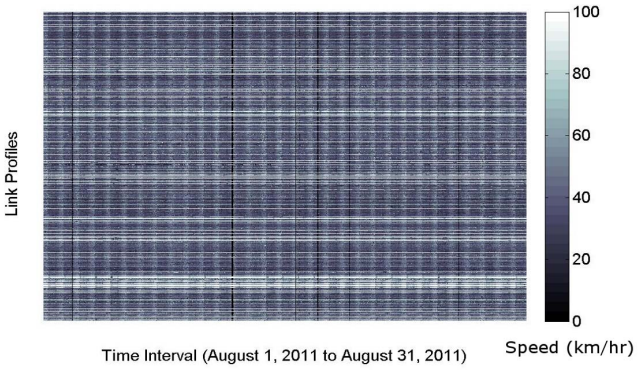

Fig. 2: Speed profiles of links in $\mathbf{M}_G^T$. Color bar represents colors corresponding to different speed values.

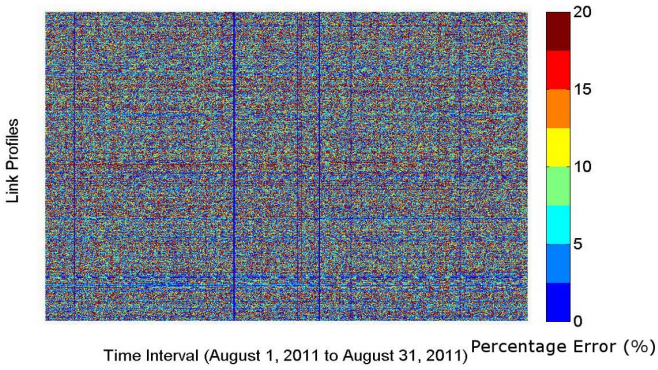

Fig. 4: Error profiles $\mathbf{H}_G^T$ for the links in network G. Color bar represents colors corresponding to the reconstruction percentage error.

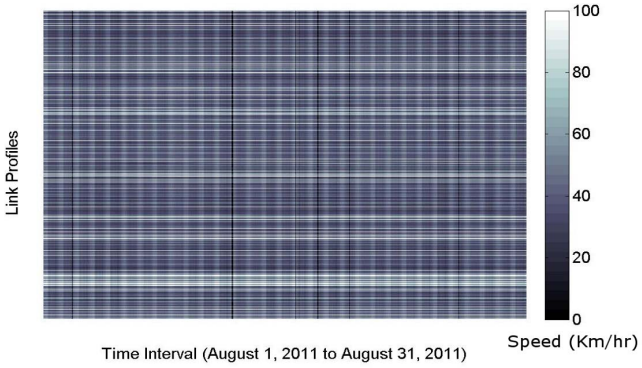

Fig. 3: Reconstructed speed profiles of links in $\mathbf{X}_G^T$ using DCT compression. Color bar represents colors corresponding to different speed values.

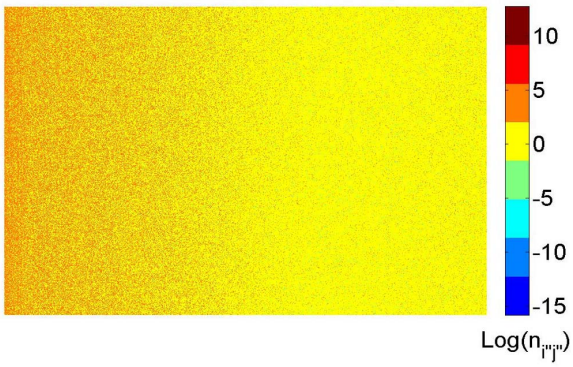

Fig. 5: Frequency Components stored in $\mathbf{N}_G^T$. Color bar represents colors corresponding to *log* of magnitude of frequency components $n_{i''j''}$.

B. Principal Component Analysis

PCA is a particular Karhunen-Loeve (KL) Transform, which has been employed in many D^2ITS applications including data compression [9] and missing data imputation [14]. The previous studies involving PCA in relation to traffic data compression have been limited to a few intersections [9]. In this study, we analyze the compression efficiency of PCA for large and diverse networks. Similar to DCT, PCA transformation tends to de-correlate the components of data and concentrates energy to small number of basis vectors. Unlike DCT, however, in case of PCA, the basis vectors are obtained by eigenvalue decomposition of the covariance matrix. To calculate covariance matrix Σ_s, we consider links $\{s_i\}_{i=1}^p$ as variables with observations taken during different time intervals. Speed profile $\mathbf{m}_i$ contains speed values observed on the link s_i during the time interval (t_1, t_d). Let us consider a variant of network profile $\mathbf{M}_{\mu G} = [\mathbf{m}_{\mu 1} ... \mathbf{m}_{\mu p}]$ obtained by centering all the speed profiles in $\mathbf{M}_G$ around their mean values, such that $\{\mathbf{m}_{\mu i} = \mathbf{m}_i - \mu_{si}\mathbf{q}\}_{i=1}^p$, where $\mathbf{q} \in \mathbb{R}^d$ and $\mathbf{q} = [11...1]^T$. Speed values across different links, exhibit strong spatial correlations [5]. This results in a non-diagonal co-variance matrix Σ_s for variables $\{s_i\}_{i=1}^p$, calculated as

$$\Sigma_s = \frac{1}{d}\mathbf{M}_{\mu G}^T \mathbf{M}_{\mu G}. \tag{5}$$

Eigenvalue decomposition of $\Sigma_s = \Phi_s \Delta_s {\Phi_s}^T$ will yield orthonormal vectors $\Phi = [\phi_{s1}\phi_{s2}...\phi_{sp}]$, such that $\{{\phi_{si}}^T \phi_{sj} = 0\}_{i \neq j}$, ${\phi_{si}}^T \phi_{si} = 1$. The diagonal matrix Δ_s contains eigenvalues for Σ_s. Using the basis vectors $\{\phi_{si}\}_{i=1}^p$, we can obtain an uncorrelated subspace$\{\chi_i\}_{i=1}^p$. The resultant transformed network profile will be $\mathbf{M}_{\chi G} = \mathbf{M}_{\mu G}\Phi_s$. In this case, the principal components are the transformed speed profiles

$\{\mathbf{m}_{\chi i}\}_{i=1}^{p}$. The 1^{st} principal component will be along the direction of maximum variance (not the historic mean) in the data and so forth. This transformation allows relatively small number of principal components to explain most of the variance in the data. We obtain low-dimensional representation $\mathbf{X}_{\mu G} = \mathbf{M}_{\chi Gr}\Phi_{sr}^{T}$ (mean centered) of the network G, using these small number of principal components $\{\mathbf{M}_{\chi Gr} = [\mathbf{m}_{\chi 1}, \mathbf{m}_{\chi 2}, ...\mathbf{m}_{\chi r}]\}_{r<p}$, such that $\{\Phi_{sr} = [\phi_{s1}, \phi_{s2}, ...\phi_{sr}]\}_{r<p}$.

It is also interesting to observe the traffic behavior between different time periods. For this case, let us consider each time instance $\{t_j\}_{j=1}^{d}$ as a variable, with speed observations $\{z(s_i,t_j)\}_{s_i\in E}$ taken across the network. Σ_t is the covariance matrix in this case such that $\Sigma_t = \Phi_t \Delta_t {\Phi_t}^T$. Δ_t and Δ_s provide a measure of how much variance the corresponding principal components can explain. It may seem from Fig. 7 and 8 that Σ_t will provide better compression as the 1^{st} principal component explains around 90% of variance in the data. However, this might be misleading. If we consider Singular Value Decomposition (SVD) of $\mathbf{M}_G$, we get:

$$\mathbf{M}_G = \mathbf{U}_G \Lambda_G \mathbf{V}_G^T, \tag{6}$$

where columns of $\mathbf{U}_G$ and $\mathbf{V}_G$ are left-singular vectors and right-singular vectors of $\mathbf{M}_G$ respectively [18], [19] and:

$$\mathbf{M}_G \mathbf{M}_G^T = \mathbf{U}_G \Lambda_G \Lambda_G^T \mathbf{U}_G^T, \tag{7}$$
$$\mathbf{M}_G^T \mathbf{M}_G = \mathbf{V}_G \Lambda_G^T \Lambda_G \mathbf{V}_G^T. \tag{8}$$

The representations in (7) and (8) are quite similar to (5) and $\sum_t$, baring the scaling factor and mean subtraction. Hence, compression achieved by either using $\sum_s$ or $\sum_t$ will provide similar results [18](see Fig. 6).

Nonetheless, It can be shown that the basis vectors $\{\Phi_\omega\}_{\omega\in\{s,t\}}$ obtained by eigenvalue decomposition of $\{\Sigma_\omega\}_{\omega\in\{s,t\}}$ provide the optimal linear transformation to rotate space ($s_i \in E$) or time($t_i \in (t_1, t_d)$) variables to uncorrelated space [18].

For data reconstruction using PCA, both principal components $\mathbf{M}_{\chi Gr}$ and basis vectors Φ_{sr} need to be stored. So, $\theta_{PCA} = d \times r + p \times r$, $\Theta_{PCA} = d \times p$ and CR will be

$$CR_{PCA} = \frac{d \times p}{d \times r + p \times r}. \tag{9}$$

C. Data Compression using Tensor Decomposition

Tensor based methods have found applications in many fields including Chemometrics [20], telecommunications [21] and neuroscience [22], [23]. In this section, we propose a tensor based compression method for traffic data. Techniques such as PCA and DCT consider a 2-way representation of the network G. However, observations taken at the same time on different days may also correlate strongly [9]. To utilize these daily patterns for compression, we consider a 3-way network profile $\mathscr{M}_G \in \mathbb{R}^{d_t \times p \times w}$ of G. We consider Higher Order Singular Value Decomposition (HO-SVD) [24]–[26] for data compression, using 3-way representation of the network.

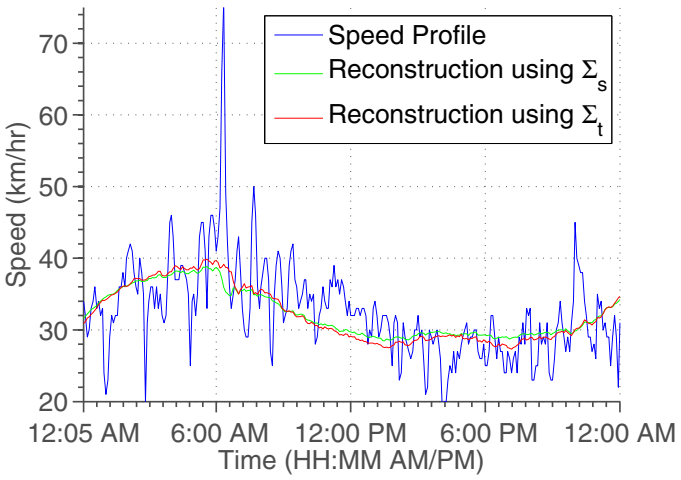

Fig. 6: Data reconstruction for a typical link, using just the 1^{st} principal components obtained from Σ_s and Σ_t. Principal components corresponding to basis vectors ϕ_{s1} and ϕ_{t1} provide similar performance.

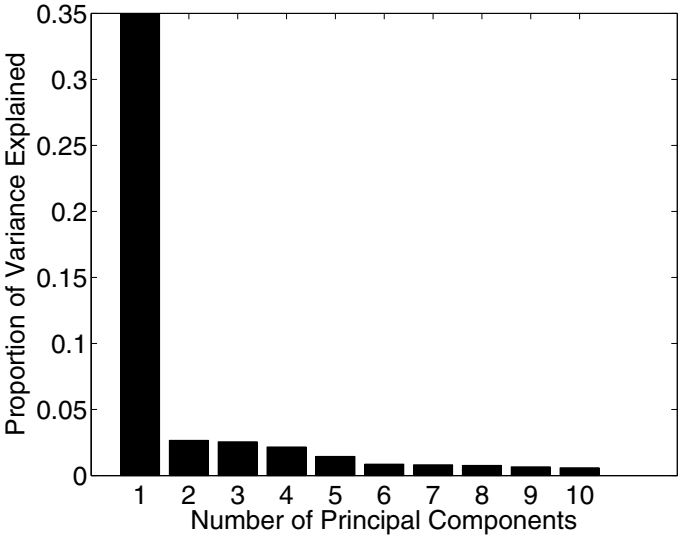

Fig. 7: Proportion of variance in Σ_s explained by principal components.

Using HO-SVD, we can decompose the network profile $\mathscr{M}_G$ as [24], [25]:

$$\mathscr{M}_G = \mathscr{C}_G \times_1 \mathbf{Y}^{(1)} \times_2 \mathbf{Y}^{(2)} \times_3 \mathbf{Y}^{(3)}, \tag{10}$$

where $\mathscr{C}_G \in \mathbb{R}^{d_t \times p \times w}$ is called core tensor and matrix $\{\mathbf{Y}^{(n)}\}_{n=1}^{3}$ contains n-mode singular vectors of $\mathscr{M}_G$ [25]. $\mathscr{F} = \mathscr{A} \times_n \mathbf{B}$ is called the n-mode product between the tensor $\mathscr{A} \in \mathbb{R}^{I_1 \times I_2 \ldots I_n \ldots I_N}$ and the matrix $\mathbf{B} \in \mathbb{R}^{J_n \times I_n}$. Entries for $\mathscr{F} \in \mathbb{R}^{I_1 \times I_2 \ldots I_{n-1} \times J_n \ldots I_N}$ are calculated as:

$$f_{i_1 i_2 \ldots i_{n-1} j_n \ldots i_N} = \sum_{i_n} a_{i_1 i_2 \ldots i_{n-1} j_n \ldots i_N} b_{j_n i_n}. \tag{11}$$

For data compression, we estimate $\mathscr{M}_G$ by a low-rank tensor $\mathscr{X}_G$ using the low rank core tensor $\mathscr{C}_{Gr} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$ and

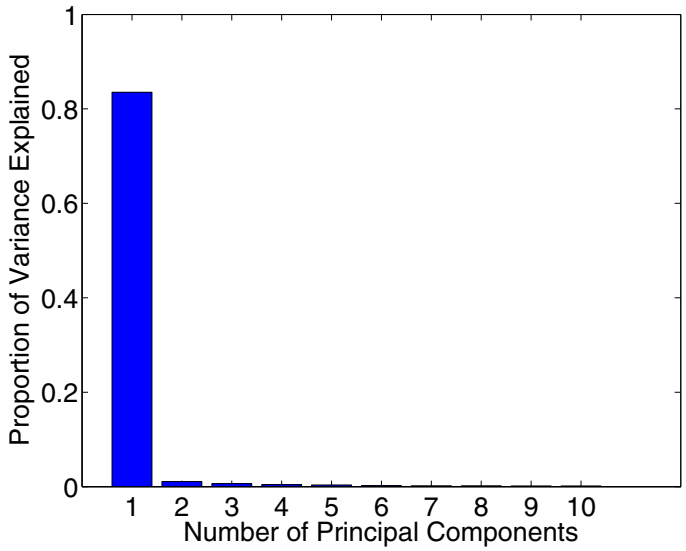

Fig. 8: Proportion of variance in Σ_t explained by principal components.

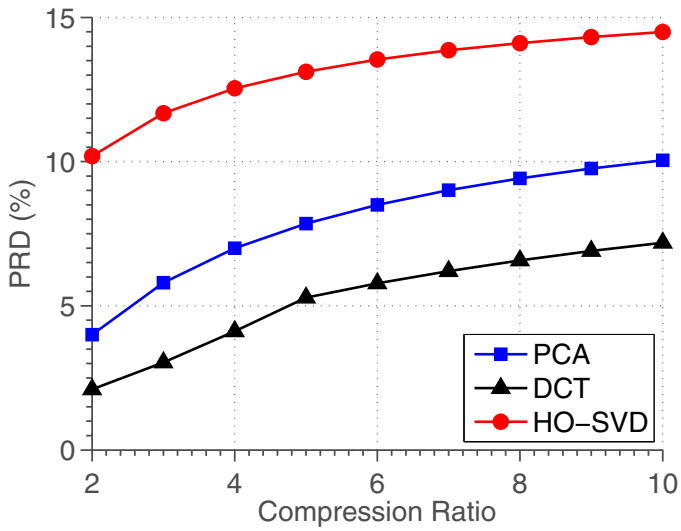

Fig. 9: Compression efficiency of different techniques.

corresponding matrices $\mathbf{W^{(1)}} \in \mathbb{R}^{d_t \times r_1}$, $\mathbf{W^{(2)}} \in \mathbb{R}^{p \times r_2}$ and $\mathbf{W^{(3)}} \in \mathbb{R}^{w \times r_3}$ containing singular vectors for their respective modes [25].

For reconstruction, we need to store $\mathscr{C}_{Gr}$ as well as $\mathbf{W^{(1)}}$ $\mathbf{W^{(2)}}$ and $\mathbf{W^{(3)}}$. So $\Theta_{HOSVD} = d_t \times p \times w$, $\theta_{HOSVD} = r_1 \times r_2 \times r_3 + d_t \times r_1 + p \times r_2 + w \times r_3$ and CR will be

$$CR_{HOSVD} = \frac{d_t \times p \times w}{r_1 \times r_2 \times r_3 + d_t \times r_1 + p \times r_2 + w \times r_3}. \quad (12)$$

D. Compression Efficiency

We use percent root-mean-square distortion (PRD) [23], [27] to compare compression efficiency of proposed methods. PRD is also often referred as Relative error [28], [29]. It is a commonly used performance metric for low-dimensional estimation of matrices and tensors [23], [27]–[29]. PRD can be considered as a measure of energy loss in the reconstructed network profile $\mathbf{X}_G$ ($\mathscr{X}_G$) as compared to the original profile $\mathbf{M}_G$ ($\mathscr{M}_G$). PRD for matrices is calculated as:

$$PRD_{\mathbf{M}}(\%) = \frac{\| \mathbf{M}_G - \mathbf{X}_G \|_F}{\| \mathbf{M}_G \|_F} \times 100, \quad (13)$$

and for tensors it is

$$PRD_{\mathscr{M}}(\%) = \frac{\| \mathscr{M}_G - \mathscr{X}_G \|_F}{\| \mathscr{M}_G \|_F}, \quad (14)$$

where Frobenius norms for the matrix $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2}$ and the tensor $\mathscr{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots I_N}$ are defined as

$$\| \mathbf{A} \|_F = \left(\sum_{i_1} \sum_{i_2} a_{i_1 i_2}^2 \right)^{1/2}, \quad (15)$$

$$\| \mathscr{A} \|_F = \left(\sum_{i_1} \sum_{i_2} \cdots \sum_{i_N} a_{i_1 i_2 \ldots i_N}^2 \right)^{1/2}. \quad (16)$$

IV. Results and Discussion

In this section, we compare the compression efficiency of the proposed methods. Fig. 9 shows PRD for the proposed techniques for different CR. We chose CR from 2:1 up till 10:1 for performance evaluation.

DCT consistently provides highest reconstruction accuracy for different error ratios. Even for CR of $10:1$, DCT has PRD of around 7%.

HO-SVR has the highest PRD for different values of CR. For CR of 10:1 PRD for HO-SVR is 14%, which about double the error suffered by DCT.

PCA provides comparable performance to DCT, with PRD of 10% for CR of 10:1. The value is still quite high, considering that PCA is supposed to be the optimal linear transformation. However, the reason for sub-optimal performance of PCA is due to extra overhead in data storage. As evident in (9), for data reconstruction using PCA, we need to store both transformed entries, as well as basis vectors. There is no such requirement in DCT as it uses predefined basis vectors. This greatly improves the compression efficiency of DCT. PCA will need extra $p \times r$ memory units, to store same number of transformed entries as DCT. Consequently, DCT can achieve higher CR for given error tolerance, even if basis vectors used by DCT are not optimal for $\mathbf{M}_G$ [17]. For HO-SVD, storage overhead cost is even higher. In this case, apart from core tensor, we need to store singular vectors for all n-modes. These overheads in case of both PCA and HO-SVR, cannot be neglected, as without these basis vectors data reconstruction will not be possible.

Nonetheless, we can conclude that subspace methods can be effectively employed for data compression for large and diverse road networks. PCA and HO-SVR can provide CR of $10:1$ with error tolerance of around 10% and 14% respectively. DCT can deliver the same performance with even tighter error tolerance of around 7%.

V. Conclusion

Nowadays, due to advances in the sensor technology, D^2ITS systems have large volumes of data at their disposal. While, availability of data helps to enhance the performance of these systems, the size of data tends to hinder their scalability, especially for large networks. In this paper, we propose different subspace methods such as PCA, DCT and HO-SVD to deal with the problem of data compression for large and diverse road networks. We considered a large network consisting of expressways as well as urban arteries carrying significant traffic volumes. The network consisted of more than 6000 road segments. DCT provided best compression ratio for a given error tolerance. PCA provided comparable compression efficiency, where as HO-SVR suffered due to higher storage overhead.

In the future, traffic prediction and route guidance algorithms can use these compressed states of large networks, and obtain greater scalability. Consequently, more data intense algorithms related to traffic prediction, estimation and modeling can be made to run in real time by using these compressed network states. Many D^2ITS applications can benefit from the use of tensor based methods. In particular, for data compression applications, their efficiency can potentially be improved by considering other traffic parameters such as flow and travel time along side speed in multi modal structures.

Acknowledgment

The research described in this project was funded in part by the Singapore National Research Foundation (NRF) through the Singapore MIT Alliance for Research and Technology (SMART) Center for Future Mobility (FM).

References

[1] J. Zhang, F. Wang, K. Wang, W. 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] C. Goh, J. Dauwels, N. Mitrovic, M. 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*, sept. 2012, pp. 776 –781.

[3] D. Work, S. Blandin, O. Tossavainen, B. Piccoli, and A. Bayen, "A traffic model for velocity data assimilation," *Applied Mathematics Research eXpress*, vol. 2010, no. 1, pp. 1–35, 2010.

[4] 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*. IEEE, 2011, pp. 889–898.

[5] 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.

[6] E. Schmitt and H. Jula, "Vehicle route guidance systems: Classification and comparison," in *Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE*. IEEE, 2006, pp. 242–247.

[7] S. Lim, H. Balakrishnan, D. Gifford, S. Madden, and D. Rus, "Stochastic motion planning and applications to traffic," *The International Journal of Robotics Research*, vol. 30, no. 6, pp. 699–712, 2011.

[8] M. T. Asif, J. Dauwels, C. Y. Goh, A. Oran, E. Fathi, M. Xu, M. M. Dhanya, N. Mitrovic, and P. Jaillet, "Unsupervised learning based performance analysis of n-support vector regression for speed prediction of a large road network," in *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*, sept. 2012, pp. 983 –988.

[9] 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.

[10] M. Castro-Neto, Y. Jeong, M. Jeong, and L. Han, "Online-svr for short-term traffic flow prediction under typical and atypical traffic conditions," *Expert systems with applications*, vol. 36, no. 3, pp. 6164–6173, 2009.

[11] X. Jin, Y. Zhang, and D. Yao, "Simultaneously prediction of network traffic flow based on pca-svr," *Advances in Neural Networks–ISNN 2007*, pp. 1022–1031, 2007.

[12] W. Dong, M. Li, Q. B. Vo, and H. Vu, "Reliability in stochastic time-dependent traffic networks with correlated link travel times," in *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*, sept. 2012, pp. 1626 –1631.

[13] 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.

[14] L. Qu, J. Hu, L. Li, and Y. Zhang, "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.

[15] Y. Xiao, Y. 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.

[16] Y. Fisher, B. Bielefeld, A. Lawrence, and D. Greenwood, "Fractal image compression." DTIC Document, Tech. Rep., 1991.

[17] G. Strang, "The discrete cosine transform," *SIAM review*, vol. 41, no. 1, pp. 135–147, 1999.

[18] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2005.

[19] G. Strang, *Introduction to linear algebra*. Wellesley Cambridge Pr, 2003.

[20] T. Kolda and B. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[21] N. Sidiropoulos, G. Giannakis, and R. Bro, "Blind parafac receivers for ds-cdma systems," *Signal Processing, IEEE Transactions on*, vol. 48, no. 3, pp. 810–823, 2000.

[22] E. Acar, C. Aykut-Bingol, H. Bingol, R. Bro, and B. Yener, "Multiway analysis of epilepsy tensors," *Bioinformatics*, vol. 23, no. 13, pp. i10–i18, 2007.

[23] J. Dauwels, S. Kannan, R. Ramasubba, and A. Cichocki, "Near-lossless multi-channel eeg compression based on matrix and tensor decompositions," *Biomedical and Health Informatics, IEEE Journal of*, vol. PP, no. 99, p. 1, 2012.

[24] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, 2000.

[25] L. De Lathauwer and J. Vandewalle, "Dimensionality reduction in higher-order signal processing and rank-(r_1,r_2,...,r_N) reduction in multilinear algebra," *Linear Algebra and its Applications*, vol. 391, pp. 31–55, 2004.

[26] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-(r 1, r 2,..., rn) approximation of higher-order tensors," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1324–1342, 2000.

[27] Y. Zigel, A. Cohen, and A. Katz, "The weighted diagnostic distortion (wdd) measure for ecg signal compression," *Biomedical Engineering, IEEE Transactions on*, vol. 47, no. 11, pp. 1422–1430, 2000.

[28] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, "Scalable tensor factorizations for incomplete data," *Chemometrics and Intelligent Laboratory Systems*, vol. 106, no. 1, pp. 41–56, March 2011.

[29] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and bregman iterative methods for matrix rank minimization," *Mathematical Programming*, vol. 128, no. 1, pp. 321–353, 2011.