IJICTR
International Journal of Information & Communication Technology Research
ITRC

IJICTR International Journal of Information & Communication Technology Research

ITRC

Volume 8 - Number 2- Spring 2016

Topics of Interest

Information Technology

- Information Systems
- IT Applications & Services
- IT Platforms: Software & Hardware Technology
- IT Strategies & Frameworks

Communication Technology

- Communication Devices
- Communication Theory
- Mobile Communications
- Optical Communications
- Satellite Communications
- Signal / Image / Video Processing

Network Technology

- Computer & Communication Networks
- Wireless Networks
- Network Management
- Network Security
- NGN Technology
- Security Management

IJICTR
International Journal of Information & Communication Technology Research
ITRC
Volume 8- Number 2- Spring 2016

IJICTR International Journal of Information & Communication Technology Research

Volume 8- Number 2- Spring 2016(1-7)

Learning Motion Patterns in Traffic Scenes by Improved Group Sparse Topical Coding

Parvin Ahmadi
Department of Electrical Engineering
Sharif University of Technology
Tehran, Iran
electronic_iut82@yahoo.com

Iman Gholampour
Department of Electrical Engineering
Sharif University of Technology
Tehran, Iran
imangh@sharif.ir

Mahmoud Tabandeh
Department of Electrical Engineering
Sharif University of Technology
Tehran, Iran
tabandeh@sharif.ir

Received: April 14, 2015- Accepted: February 23, 2016

Abstract—**Analyzing motion patterns in traffic videos can directly lead to generate some high-level descriptions of the video content. In this paper, an unsupervised method is proposed to automatically discover motion patterns occurring in traffic video scenes. For this purpose, based on optical flow features extracted from video clips, an improved Group Sparse Topical Coding (GSTC) framework is applied for learning semantic motion patterns. Then, each video clip can be sparsely represented by a weighted sum of learned patterns which can further be employed in very large range of applications. Compared to the original GSTC, the proposed improved version of GSTC selects only a small number of relevant words for each topic and hence provides a more compact representation of topic-word relationships. Moreover, in order to deal with large-scale video analysis problems, we present an online algorithm for improved GSTC which can not only deal with large video corpora but also dynamic video streams. Experimental results show that our proposed approach finds the motion patterns accurately and gives a meaningful representation for the video.**

Keywords—Motion patterns, Group Sparse Topical Coding, traffic scene

I. INTRODUCTION

In many surveillance scenarios, such as monitoring vehicles traffic at intersections, crowded video scenes with various motions may be involved. In these scenes, some typical activities, called motion patterns, occur regularly and periodically. It is highly desired to analyze the motion patterns and extract a high-level interpretation of the video contents. Discovering such motion patterns would directly lead to a semantic scene model that could further facilitate the task of scene analysis [1]. This is a very challenging task since for complex or crowded scenes the performance of most conventional analysis tools is highly degraded [1]. Traditional methods [2-4] for analyzing a traffic scene are based on trajectory data. These methods need an accurate detection and tracking of vehicles that are very difficult in crowded videos due to the noise, change in lighting and weather conditions, shadows, and occlusion [5].

To handle these issues, some researchers [1, 6-11] have applied low-level motion features such as optical flow, which can be easily calculated, and focused on developing more complex methods such as topic models. Wang et al. [6] and Kuettel et al. [7] characterized typical activities by topic models, such as Latent Dirichlet Allocation (LDA) and Hierarchical Dirichlet Process (HDP). Song et al. [8] trained a two-level LDA topic model first for single-agent motion, which is the input to second level LDA for multi-agent interactions. A two-staged cascaded LDA model was formulated by Li et al. in [9] in which, the first stage learns regional behavior and the second stage learns

global context over the regional models. Varadarajan et al. in [10] introduced a sequential topic model for mining recurrent activities from long term video logs. Rana et al. [11] used Fast rank-1 robust PCA for foreground detection with counts of pixels in blocks used as input for Dirichlet process mixture model (DPMM) learning, which enables incremental learning and inference. Fu et al. [1] used Sparse Topical Coding (STC) for efficient learning and representation of the topic model.

In this paper, we focus on automatic learning of the semantic motion patterns from a traffic video. The learning is done by using low-level features and applying topic models. For this purpose, an improved Group Sparse Topical Coding (GSTC) framework is proposed for learning a motion pattern dictionary. This improved version of GSTC considers the sparsity of words which construct a topic. By enforcing this sparsity, GSTC gains the ability to automatically select the most relevant words for each latent topic. This makes improved GSTC more suitable for modeling motion patterns occurring in video scenes. By learning semantic dictionary, the video scene can be represented as a sparse summation of bases. This representation can further be applied to scene analysis such as rule mining, abnormal event detection, etc. Moreover, in order to deal with large data collections and dynamic data streams, we have developed the online improved GSTC, which learns the topical dictionary via an online algorithm. Various experiments show that our improved GSTC achieves high performance in discovering scene patterns.

II. BACKGROUND THEORY

Probabilistic Topic models (PTMs) such as PLSA [12], LDA [13] and HDP [14] were first developed to capture latent topics in a large collection of textual documents and then utilized by researchers for video analysis. In 2011, Zhu and Xing [15] presented a Non-Probabilistic topic Model (NPM) called Sparse Topical Coding (STC) which assigns a sparse set of topics to each document. Bai et al. [16] in 2013 proposed a novel non-probabilistic topic model for discovering sparse latent representations of large text corpora, referred as group sparse topical coding (GSTC). This model enjoys both the advantages of the PTMs and NPMs. On the one hand, GSTC can derive document-level admixture proportions in topic simplex like PTMs. On the other hand, GSTC can directly control the sparsity of the inferred representations by relaxing the normalization constraint like NPMs [16]. Moreover, compared to STC, since GSTC does not need to model the document codes θ, it has fewer variables to be estimated and therefore requires less training data.

Suppose that a collection of D documents $\{\mathbf{w}_1,\ldots,\mathbf{w}_D\}$ is given which contains words from a vocabulary $\mathbf{v}$ with size N. A document is simply represented as a $|I|$-dimension vector $\mathbf{w}=\{w_1,\ldots,w_{|I|}\}$, where I is the index set of words that appear and the nth entry w_n ($n\in I$) denotes the number of appearances of the specific word in the document. Let $\beta\in\mathbb{R}^{K\times N}$ be a dictionary with K bases, where each base is assumed to be a topic base, i.e. a unigram distribution over $\mathbf{v}$. For the dth document $\mathbf{w}_d$, GSTC projects $\mathbf{w}_d$ into a semantic space spanned by a set of automatically learned topic bases β and directly obtain the un-normalized word code $s_{d,.n}\in\mathbb{R}^K$ for each individual word in document $\mathbf{w}_d$. The admixture proportion of the entire document $\mathbf{w}$ then be derived from the learned word code set $s=\{s_{.1},\ldots,s_{.|I|}\}\in\mathbb{R}^{K\times|I|}$ and the topic bases β. GSTC solves the optimization problem (1).

$$\begin{aligned}
&\min_{\{s_d\}_{d=1}^D,\beta}\sum_{d=1}^{D}\sum_{n=1}^{|I_d|}\left(s_{d,n.}^T\beta_{.n}-w_{d,n}\ln(s_{d,n.}^T\beta_{.n})\right)\\
&+\sum_{d=1}^{D}\lambda\|s_d\|_{2,1}+C=\\
&\sum_{d=1}^{D}\sum_{n=1}^{|I_d|}\left(\sum_{k=1}^{K}s_{d,kn}\beta_{kn}-w_{d,n}\ln(\sum_{k=1}^{K}s_{d,kn}\beta_{kn})\right)+\\
&\sum_{d=1}^{D}\sum_{k=1}^{K}\lambda\|s_{d,k.}\|_2+C\\
&s.t.\ \ s_{d,.n}\geq 0,\forall d,n\in I_d;\ \beta_{k.}\geq 0,\sum_{n=1}^{N}\beta_{kn}=1,\forall k
\end{aligned}\tag{1}$$

The first term in (1) is equivalent to minimizing an un-normalized KL-divergence between observed word counts $w_{d,n}$ and their reconstructions $s_{d,n}^T\beta_{.n}$. The second term is a group-lasso [17], i.e. a mixed ℓ1/ ℓ2 norm, for the matrix of reconstruction coefficients. This leads to the desired document-level sparsity.

The document-level admixture proportion can be derived with the learned word codes. Let θ be the topic proportion vector of document $\mathbf{w}$, the kth topic proportion will be [16]:

$$\theta_k=\frac{\sum_{n=1}^{|I|}s_{kn}\beta_{kn}}{\sum_{n=1}^{|I|}\sum_{k=1}^{K}s_{kn}\beta_{kn}}\tag{2}$$

III. PROPOSED METHOD

A. Video Representation

Given an input video, we first divide it into a sequence of clips without overlapping. Each clip is considered as a document.

We utilize Shi and Tomasi [18] corner detector to find the key points and use these features to extract the optical flow using Lucas–Kanade method [19] from each pair of consecutive frames. To remove noise, a threshold is applied to the amplitude of optical flow vectors. In order to generate the vocabulary, the optical flow vectors are quantized into discrete visual words. Optical flow vectors are denoted by (x, y, α). The positions (x, y) are quantized to the nearest position on a grid with spacing of 10 pixels and the angles of flow vectors, α, are quantized into 8 directions. Finally a fixed vocabulary is formed namely $\mathbf{v}=\{v_1,\ldots,v_N\}$ with N total flow words, in which each word contains information about position and motion direction.

Flow words are accumulated over the frames of each video clip. Then a clip of video is represented by a vector $\mathbf{w}=\{w_1,\ldots,w_{|I|}\}$, where I is the set of word indexes and w_n denotes the number of occurrence of word n in the clip. Using a word-document topic model, optical flow words with high co-occurrence frequencies in a video clip make a motion pattern. Motion patterns are represented as dictionary β whose rows show the typical topics in the video which are a distribution over the vocabulary $\mathbf{v}$.

B. Learning Motion Patterns

We try to learn a sparse representation for the number of topics in each document. That means a video clip is interpreted by only a few motion patterns. To our knowledge, GSTC has not yet been used for video analysis. We formulate the problem of learning motion patterns based on an improved version of GSTC that is more suitable for video scene understanding.

a) Formulation of Improved GSTC

Our proposed formulation solves the optimization problem in (3).

$$\min_{\{s_d\}_{d=1}^D,\beta} \sum_{d=1}^{D}\left(\left\|\mathbf{w}_d-\mathrm{diag}(s_d^T\beta)\right\|_2^2+\lambda\|s_d\|_{2,1}\right)+\rho\|\beta\|_1 = \sum_{d=1}^{D}\Big(\sum_{n=1}^{|I_d|}(w_{d,n}-s_{d,n.}^T\beta_{.n})^2+\lambda\sum_{k=1}^{K}\|s_{d,k.}\|_2\Big)+\rho\sum_{n=1}^{N}\|\beta_{.n}\|_1$$
$$s.t. \quad s_d\geq 0,\forall d;\ \ \beta\geq 0 \qquad (3)$$

where (λ, ρ) are non-negative hyper-parameters that must be set by users.

In the proposed formulation, unlike GSTC which minimizes KL-divergence, for simplicity we minimize $\ell 2$ norm of the reconstructions error. Furthermore, we improve the GSTC model to learn a better dictionary from the videos by adding $\|\beta\|_1$ term to the objective function.

Whereas only a sparse set of words contribute in making each topic, the term $\|\beta\|_1$ is added to impose this sparsity on the dictionary. That also reduces the overlapping between different topics. To this end, we relax the normalization constraint in GSTC on the rows of dictionary (we do not impose the constraint $\sum_{n=1}^{N}\beta_{kn}=1,\ \forall k$ in our model).

b) Optimization

The objective function in (3) is bi-convex. That is, convex over either $\{s_d\}_{d=1}^D$ or β when the other is fixed. Furthermore, the feasible set is a convex set. Therefore, a typical solution to this bi-convex problem is the coordinate descent algorithm [20] which alternatively performs the optimization over $\{s_d\}_{d=1}^D$ and β as shown in Algorithm 1. After learning the dictionary of topics β through training phase, it can further be used for finding the word code of a test video clip as shown in Algorithm 4. The topic proportion θ of the training or test clips can also be calculated based on (2).

Optimization over $\{s_d\}_{d=1}^D$:
This step aims to find the word codes $\{s_d\}_{d=1}^D$ when dictionary β is fixed via optimization (4).

$$\min_{\{s_d\}_{d=1}^D}\sum_{d=1}^{D}\left(\left\|\mathbf{w}_d-\mathrm{diag}(s_d^T\beta)\right\|_2^2+\lambda\|s_d\|_{2,1}\right)= \sum_{d=1}^{D}\Big(\sum_{n=1}^{|I_d|}(w_{d,n}-s_{d,n.}^T\beta_{.n})^2+\lambda\sum_{k=1}^{K}\|s_{d,k.}\|_2\Big) \qquad (4)$$
$$s.t. \quad s_d\geq 0,\forall d$$

Similarly to [16], we perform this optimization for each document separately and only focus on one group, e.g. the kth group. The objective function can then be written as:

$$\ell(s_{k.})=\sum_{n=1}^{|I|}(w_n-s_{n.}^T\beta_{.n})^2+\lambda\|s_{k.}\|_2 \qquad (5)$$

Equation (5) is strictly convex with respect to s_{kn}. Therefore, s_{kn} can be obtained by setting the gradient equal to zero:

$$\nabla_{s_{kn}}\ell(s_{k.})=-2(w_n-s_{n.}^T\beta_{.n})\beta_{kn}+\lambda\frac{s_{kn}}{\|s_{k.}\|_2}= -2(w_n-s_{kn}\beta_{kn}-\sum_{i\in I,i\neq k}s_{in}\beta_{in}) \qquad (6)$$
$$+\lambda\frac{s_{kn}}{\sqrt{s_{kn}^2+\sum_{i\in I,i\neq n}s_{ki}^2}}=0$$

The encoding algorithm is presented as Algorithm 2.

Optimization over β:
After inferring all the latent word codes of the collection, the dictionary β is updated by minimizing (7).

$$\min_{\beta}\sum_{d=1}^{D}\left(\left\|\mathbf{w}_d-\mathrm{diag}(s_d^T\beta)\right\|_2^2\right)+\rho\|\beta\|_1 = \sum_{d=1}^{D}\sum_{n=1}^{|I_d|}(w_{d,n}-s_{d,n.}^T\beta_{.n})^2+\rho\sum_{n=1}^{N}\|\beta_{.n}\|_1 \quad s.t. \ \ \beta\geq 0 \qquad (7)$$

By assigning zero values to $w_{d,n}$ and $s_{d,.n}$ for $n\notin|I_d|$, we can replace the $|I_d|$ with N. Then (7) can be written as:

$$\sum_{n=1}^{N}\left(\sum_{d=1}^{D}(w_{d,n}-s_{d,n.}^T\beta_{.n})^2+\rho\|\beta_{.n}\|_1\right) \quad s.t. \ \ \beta\geq 0 \qquad (8)$$

Based on the idea of iteratively re-weighted least squares (IRLS) [21], the $\ell 1$ norm is written as (9).

$$\|\beta_{.n}\|_1=\sum_{k=1}^{K}|\beta_{kn}|=\sum_{k=1}^{K}q_{kn}\beta_{kn}^2=\beta_{n.}^T\mathbf{Q}\beta_{.n} \qquad (9)$$

where $q_{kn}=1/(|\beta_{kn}|+\varepsilon)$ and $\mathbf{Q}=\mathrm{diag}(q_{1n},\ldots,q_{Kn})$. In q_{kn}, ε is much smaller than the absolute value of the non-zero entries of β. According to (9), for each nth column of β, we have:

$$\underline{\beta}_{.n} = \arg\min_{\underline{\beta}} \{\sum_{d=1}^{D} (w_{d,n} - s_{d,n.}^{T}\underline{\beta})^2 + \rho\underline{\beta}^{T}\mathbf{Q}\underline{\beta}\} \quad (10)$$

It is a convex optimization problem that can be efficiently solved by setting the gradient equal to zero:

$$-\sum_{d=1}^{D}\left(s_{d,n} \times (w_{d,n} - s_{d,n.}^{T}\underline{\beta})\right) + \rho\mathbf{Q}\underline{\beta} = 0$$
$$\underline{\beta} = \left(\rho\mathbf{Q} + \sum_{d=1}^{D} s_{d,n} s_{d,n.}^{T}\right)^{-1}\left(\sum_{d=1}^{D} w_{d,n} s_{d,n}\right) \quad (11)$$

As $\mathbf{Q}$ depends on β, an alternative optimization on β and $\mathbf{Q}$ is performed, in which in each alternation, one of them is fixed and the other one is updated. The topic learning procedure is described in Algorithm 3.

c) Online version of improved GSTC

The above batch mode algorithm requires a full pass through the video collection at each gradient descent step in dictionary learning. A full pass over a very large dataset would be very expensive in terms of both memory and efficiency. Furthermore, the batch gradient descent for dictionary learning can be inefficient in utilizing the redundancy information of a large dataset [22]. To overcome such inefficiency, we propose the online version of improved GSTC model which uses a sample mode learning algorithm to learn the dictionary β. Our online algorithm is nearly as simple as the batch algorithm, but converges much faster for large datasets. The online learning algorithm is described in algorithm 5.

IV. EXPERIMENTAL RESULTS

We evaluated the performance of our proposed method on QMUL Junction video [23] which was captured at 25 frames per second from a traffic junction. The video file has been divided into 12s-length non-overlapping clips, with a frame size of 360×288 pixels. 73 clips have been considered for training and 39 clips for the test. Training clips are used for learning the usual motion patterns. The hyper-parameters have been found experimentally for the best visual results to be (λ=1, ρ=1) and the number of topics K=20.

Algorithm 1: Training Phase – Batch Mode (offline)

Inputs: training video clips $\{\mathbf{w}_d\}_{d=1}^{D}$, the number of topics K, the hyper parameters (λ, ρ)
Outputs: dictionary β, training word codes $\{s_d\}_{d=1}^{D}$

Initialize $\beta \in \mathbb{R}^{K\times N}$ to a random matrix with positive elements
Initialize $\{s_d\}_{d=1}^{D} \in \mathbb{R}^{D\times K\times N}$ to random matrices with positive elements
repeat
for d=1:D
$s_d \leftarrow$ Algorithm 2
end
$\beta \leftarrow$ Algorithm 3
until convergence

Algorithm 2: Sparse Coding

for k=1:K
if $\sum_{n=1}^{|I_d|}(w_{d,n} - s_{d,n.}^{T}\beta_{.n})^2 \beta_{kn}^2 \le \frac{\lambda}{2}$
$s_{d,k.} \leftarrow 0$
else
for n=1:$|I_d|$
$a_{kn} = \sum_{i\in I_d, i\neq k} s_{d,in}\beta_{in} - w_{d,n}$
$c_{kn} = \sum_{i\in I_d, i\neq n} s_{d,ki}^2$
$s_{d,kn} \leftarrow$ roots of equation:

$$\beta_{kn}^4 s_{d,kn}^4 + 2a_{kn}\beta_{kn}^3 s_{d,kn}^3 + (c_{kn}\beta_{kn}^4 + a_{kn}{}^2\beta_{kn}^2 - \frac{\lambda^2}{4})$$
$$s_{d,kn}^2 + 2c_{kn}a_{kn}\beta_{kn}^3 s_{d,kn} + c_{kn}a_{kn}{}^2\beta_{kn}^2 = 0$$

$s_{d,kn} \leftarrow \max(s_{d,kn}, 0)$
end
end

Algorithm 3: Dictionary Learning

for n=1:N
repeat
$\mathbf{Q} = \mathrm{diag}(\frac{1}{|\beta_{1n}|+\varepsilon}, \ldots, \frac{1}{|\beta_{Kn}|+\varepsilon})$
$\beta_{.n} = \left(\rho\mathbf{Q} + \sum_{d=1}^{D} s_{d,n} s_{d,n.}^{T}\right)^{-1}\left(\sum_{d=1}^{D} w_{d,n} s_{d,n}\right)$
until convergence
end

Algorithm 4: Test Phase

Input: a test video clip $\mathbf{w}$, the dictionary learned before β
Output: test word code s

Initialize $s \in \mathbb{R}^{K\times|I|}$ to a random matrix with positive elements
$s \leftarrow$ Algorithm 2

Algorithm 5: Sample Mode (online)

repeat
for d=1:D
$s_d \leftarrow$ Algorithm 2
for n=1:N
repeat
$\mathbf{Q} = \mathrm{diag}(\frac{1}{|\beta_{1n}|+\varepsilon}, \ldots, \frac{1}{|\beta_{Kn}|+\varepsilon})$
repeat
$g = -s_{d,n}(w_{d,n} - s_{d,n.}^{T}\beta_{.n}) + \rho\mathbf{Q}\beta_{.n}$
$\beta_{.n} \leftarrow \beta_{.n} - \mu g$
until convergence
until convergence
for k=1:K
$\beta_{kn} \leftarrow \max(\beta_{kn}, 0)$

end
end
end
until convergence

For quantitative comparison of different models, we define and calculate three measures:

Topics sparsity: It is defined as the sparsity ratio of learned topics and is calculated based on proportion of zero entries in the dictionary $\beta \in \mathbb{R}^{K \times N}$, i.e.:

$$\text{Topics sparsity} = \frac{\neq \text{zeros of } \beta}{K \times N} \quad (12)$$

Words sparsity: It is defined as the sparsity ratio of learned word codes and is calculated based on proportion of zero entries in the D word codes $s_d \in \mathbb{R}^{K \times |I_d|}$, i.e.:

$$\text{Words sparsity} = \frac{\neq \text{zeros of } \{s_1, \ldots, s_d, \ldots, s_D\}}{K \times \sum_{d=1}^{D} |I_d|} \quad (13)$$

Topics similarity: It is defined as the similarity of discovered topics and is calculated based on average correlation between every two different topics, i.e.:

$$\text{Topics similarity} = \frac{\sum_{i=1}^{K} \sum_{j=1, j \neq i}^{K} \frac{\beta_{i.} \beta_{.j}^{T}}{\|\beta_{i.}\|_1 \|\beta_{j.}\|_1}}{K \times (K-1)} \quad (14)$$

We computed (12), (13) and (14) at various number of topics for PLSA, LDA, STC, original GSTC and our improved GSTC. The results are shown in figures 1, 2 and 3. According to Fig. 1, the improved GSTC achieves fewest words that contribute in topics construction. This improvement is due to imposing the sparsity constraint on the dictionary. Fig. 2 depicts that both original GSTC and improved GSTC can discover the most sparse word codes. This means that each word belongs to just a sparse set of topics. According to Fig. 3, both original GSTC and improved GSTC models achieve high performance in terms of topics similarity.

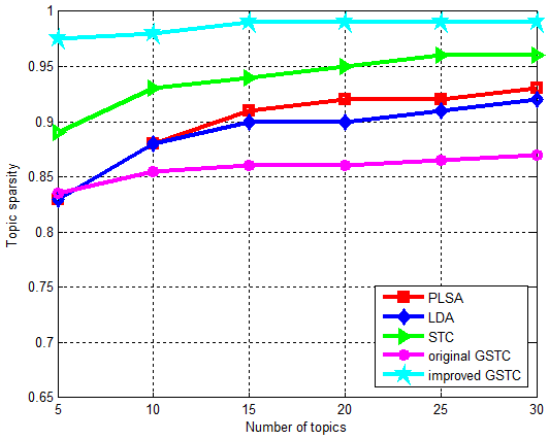

Fig. 1. Topics sparsity vs. the number of topics

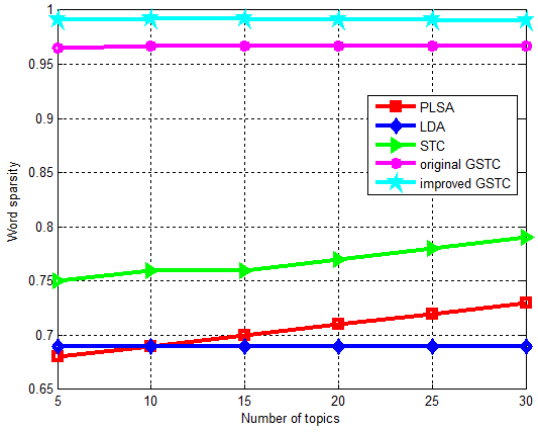

Fig. 2. Words sparsity vs. the number of topics

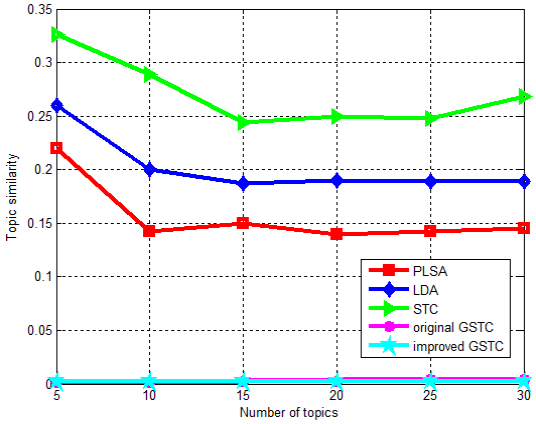

Fig. 3. Topics similarity vs. the number of topics

For qualitative comparison, the top four most probable motion patterns, according to their frequencies of occurrences, learned by PLSA, LDA, STC and our improved GSTC are shown in Fig. 4. As it can be seen, each motion pattern learned by different models has an explicit semantic meaning. For example the upward traffic flow, the downward traffic flow, the leftward traffic flow, and so on. Besides, some motion patterns are comprised of other simpler patterns. With the composite patterns, fewer bases can be utilized to reconstruct a complex scene i.e. a sparse set of topics is assigned to each document.

V. CONCLUSION

In this paper, discovering semantic motion patterns for traffic videos has been formulated in improved group sparse topical coding framework. Improved GSTC could provide a more precise interpretation of topic-word relations by selecting a small number of relevant words for each latent topic. Based on semantic topical representation learned by the model, each video clip can be sparsely reconstructed. Experimental results have shown the advantages of our approach by meaningful sparse representations of videos. The method can be employed further in scene analysis applications.

REFERENCES

[1] W. Fu, J. Wang, Z. Li, H. Lu, and S. Ma, "Learning semantic motion patterns for dynamic scenes by improved sparse topical coding", *IEEE International Conference on Multimedia and Expo* (*ICME*), pp. 296–301, 2012.

[2] D. Vasquez, T. Fraichard, and C. Laugier, "Incremental learning of statistical motion patterns with growing hidden Markov models", *IEEE Trans. Intell. Transp. Syst.* 10(3), 403–416, 2009.

[3] B. T. Morris and M. M. Trivedi, "Trajectory learning for activity understanding: unsupervised, multilevel, and long-term adaptive approach", *IEEE Trans. Pattern Anal. Mach. Intell.* 33(11), 2287–2301, 2011.

[4] N. Noceti and F. Odone, "Learning common behaviors from large sets of unlabeled temporal series", *Image Vis. Comput.* 30(11), 875–895, 2012.

[5] B. T. Morris, M. M. Trivedi, "Understanding vehicular traffic behavior from video: a survey of unsupervised approaches", *Journal of Electronic Imaging* 22, no. 4, pp. 041113-041113, 2013.

[6] X. Wang, X. Ma, E. Grimson, "Unsupervised activity perception by hierarchical Bayesian models", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2007.

[7] D. Kuettel, M. Breitenstein, L. Van Gool, V. Ferrari, "What's going on? Discovering spatio-temporal dependencies in dynamic scenes", *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1951–1958, 2010.

[8] L. Song, F. Jiang, Z. Shi, A. Katsaggelos, "Understanding dynamic scenes by hierarchical motion pattern mining", IEEE *International Conference on Multimedia and Expo* (*ICME*), pp. 1–6, 2011.

[9] J. Li, S. Gong, T. Xiang, "Learning behavioural context", *International Journal of Computer Vision* 97(3), pp. 276–304, 2012.

[10] J. Varadarajan, R. Emonet, J.-M. Odobez, "A sequential topic model for mining recurrent activities from long term video logs," *Int. J. Comput. Vis.* 103(1), pp. 100–126, 2012.

[11] S. Rana, D. Phung, S. Pham, S. Venkatesh, "Large-scale statistical modeling of motion patterns: a Bayesian nonparametric approach", *Indian Conference on Computer Vision, Graphics and Image Processing*, 2012.

[12] T. Hofmann, "Probabilistic latent semantic analysis", *UAI*, pp. 289–296, 1999.

[13] D.M. Blei, A.Y. Ng, M.I. Jordan, J. Lafferty, "Latent Dirichlet Allocation", *Journal of Machine Learning Research* (3), pp. 993–1022, 2003.

[14] Y.W. Teh, M.I. Jordan, M.J. Beal, D.M. Blei, "Hierarchical Dirichlet processes", *Journal of the American Statistical Association,* 101 (476), pp. 1566–1581, 2006.

[15] J. Zhu and E. Xing, "Sparse topical coding", *Proceedings of the Twenty-Seventh Conference Annual Conference on Uncertainty in Artificial Intelligence* (*UAI*), pp. 831–838, 2011.

[16] L. Bai, J. Guo, Y. Lan, X. Cheng, "Group sparse topical coding: from code to topic", *Proceedings of the sixth ACM international conference on Web search and data mining*, pp. 315-324. ACM, 2013.

[17] M. Yuan, Y. Lin, Y. Lin, "Model selection and estimation in regression with grouped variables", *Journal of the Royal Statistical Society*, Series B, 68:49–67, 2006.

[18] J. Shi and C. Tomasi, "Good features to track", *Proc. of IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), pp. 593–600, Seattle, Washington, June 1994.

[19] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision", *Proceedings of imaging understanding workshop*, 1981.

[20] H. Lee, A. Battle, R. Raina, and A. Y. Ng., "Efficient Sparse Coding Algorithms", *NIPS*, pp. 801–808, 2006.

[21] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in IEEE ICASSP, 2008.

[22] A. Zhang, J. Zhu, B. Zhang, "Sparse online topic models", *Proceedings of the 22nd international conference on World Wide Web*, pp. 1489-1500, 2013.

[23] http://www.eecs.qmul.ac.uk/~ccloy/downloads_qmul_junction.html

Parvin Ahmadi received her degrees of B.Sc. and M.Sc. both in electrical engineering from Isfahan University of Technology. She is currently a PhD candidate at Sharif University of Technology, Tehran, Iran. Her research interests are mainly in the areas of signal and image processing, computer vision and pattern recognition.

Iman Gholampour received his B.Sc. , M.Sc. and Ph.D. degree in Electrical Engineering, all from Sharif University of Technology Tehran, Iran. His current interests include signal processing, machine vision, and image and video analysis for intelligent traffic systems. He is now an assistant professor at Electronics Research Institute, Sharif University of Technology.

Mahmoud Tabandeh received his B.Sc. , M.Sc. , Ph.D. degrees from INSA (France), LSU and university of California, Berkeley (UCB), respectively. He is currently an associate professor in the School of Electrical Engineering, Sharif University of Technology, Tehran, Iran. His research interests include digital systems, hardware and software, and image processing.

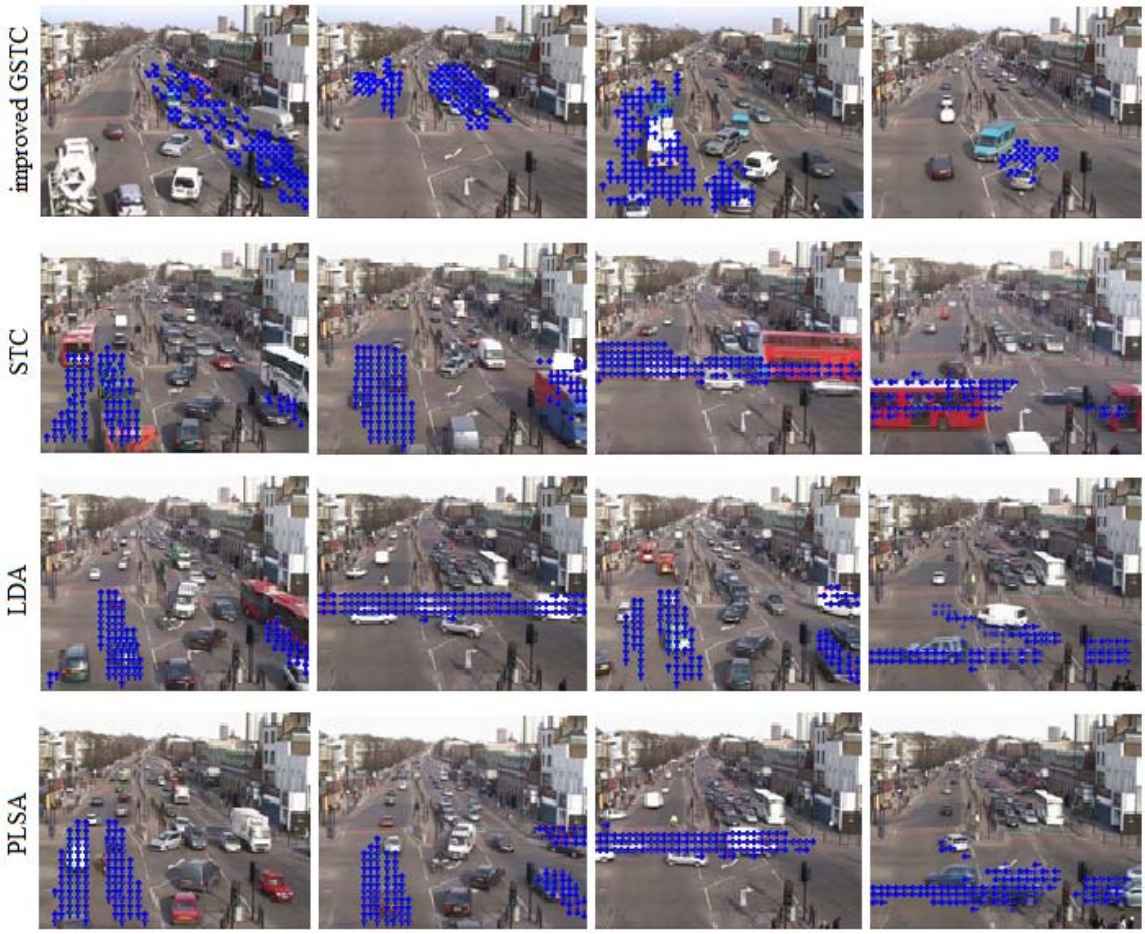

Fig. 4. Four most probable motion patterns discovered by pLSA, LDA, STC and our improved GSTC

This Page intentionally left blank.