AUT Journal of Electrical Engineering

AUT J. Elec. Eng., 50(2) (2018) 177-186
DOI: 10.22060/eej.2018.12366.5065

Traffic Scene Analysis using Hierarchical Sparse Topical Coding

P. Ahmadi[1*], I. Gholampour[2], M. Tabandeh[3]

[1] IT Research Faculty, Iran Telecommunication Research Center, Tehran, Iran
[2] Electronics Research Institute, Sharif University of Technology, Tehran, Iran
[3] Electrical Engineering Department, Sharif University of Technology, Tehran, Iran

ABSTRACT: Analyzing motion patterns in traffic videos can be exploited directly to generate high-level descriptions of the video contents. Such descriptions may further be employed in different traffic applications such as traffic phase detection and abnormal event detection. One of the most recent and successful unsupervised methods for complex traffic scene analysis is based on topic models. In this paper, a two-level Sparse Topical Coding (STC) topic model is proposed to analyze traffic surveillance video sequences which contain hierarchical patterns with complicated motions and co-occurrences. The first level STC model is applied to automatically cluster optical flow features into motion patterns. Then, the second level STC model is used to cluster motion patterns into traffic phases. Experiments on a real world traffic dataset demonstrate the effectiveness of the proposed method against conventional one-level topic model based methods. The results show that our two-level STC can successfully discover not only the lower level activities but also the higher level traffic phases, which makes a more appropriate interpretation of traffic scenes. Furthermore, based on the two-level structure, either activity anomalies or traffic phase anomalies can be detected, which cannot be achieved by the one-level structure.

Review History:

Received: 5 January 2017
Revised: 1 January 2018
Accepted: 16 January 2018
Available Online:11 February 2018

Keywords:
Traffic Phase Detection
Topic Model
Sparse Topical Coding
Temporal Video Segmentation
Anomaly Detection

1- Introduction

Along with developments of intelligent video surveillance systems, dynamic scene analysis is a hot topic that has attracted significant attention. It remains a challenging problem due to complex video surveillance scenes with multiple activities occurring simultaneously [1]. In many surveillance scenarios, such as the analysis of a crowded traffic scene, various motions are involved. It is highly desirable to discover the motion patterns and obtain some high-level interpretation of the semantic content. For example, in a video monitoring intersection without any prior knowledge about traffic rules in the specific scene, it is useful to discover typical vehicle behavior and dependency involved in this scene and to detect anomalous motions for security concerns [2].

Typically, many vehicles are involved in the traffic video scene. Motion patterns of these vehicles usually have a hierarchical nature. At low level, the motion of vehicles might follow some regular flows. At a higher level, the co-occurrence of multiple flows at the same time might also be subject to the constraints which define the traffic phases. For instance, in a traffic intersection, motion patterns are all regular paths going through the intersection which are called “activities”. Besides, the mixture motion patterns are possible combinations of paths determined by traffic lights which are called “traffic phases” [2].

Considering the hierarchical nature of motion patterns, most methods used for scene understanding and motion pattern discovery work based on hierarchical modeling. One common approach relies on the long term object trajectory-based motion analysis. However, such trajectory analysis is still nontrivial under difficult conditions for the lack of reliable and persistent multi-object tracking algorithms. In addition, rapid adaptation to sudden changes in movement is often problematic. Object tracking methods need an accurate object detection, recognition and tracking, and due to occlusions, they face serious problems in complex or crowded scenes. To improve robustness, topic model based methods have been developed. These methods avoid tracking and are directly performed on low-level features such as object location and intensity gradient. The most popular low-level feature is optical flow that contains abundant local motion information. Probabilistic topic models such as Probabilistic Latent Semantic Analysis (PLSA) [3] and Latent Dirichlet Allocation (LDA) [4], Fully Sparse Topic Models (FSTM) [5] and also non-probabilistic topic models such as Sparse Topical Coding (STC) [6] were first introduced to discover latent topics in large text corpora and then utilized by researchers for video analysis. The topic model has shown a great success to represent the low level motion features (words) in a lower dimensional motion patterns (topics) space, especially in complicated surveillance scenes. Low-level features are considered as visual words in video sequences which are treated as documents. Motion patterns can be discovered as topics (groups of visual words) shared by all documents.

Among different topic models, STC has demonstrated its effectiveness in finding meaningful motion patterns (activities) and discovering the abnormal events in traffic videos [7]. In this paper, we propose a novel two-level motion pattern analysis method based on the STC model. Without prior knowledge of the traffic rules, the two-level STC detects the activities as a mixture of optical flow features and discovers

Corresponding author, E-mail: p.ahmadi@itrc.ac.ir

the traffic phases as a mixture of activities according to the traffic signals. Traffic phases discovered by the two-level STC can be further applied to scene analysis such as temporal video segmentation and anomaly detection. Experiments will show that, besides the activities, our two-level STC can successfully detect co-occurrence of activities which are traffic phases and shared among video clips. Moreover, the two-level STC can be effectively employed for temporal video segmentation and anomaly detection. Throughout the paper, the clauses of "motion pattern", "activity" and "topic" imply the same meaning. Also, all of the clauses of "mixture motion pattern, "mixture of activities", "clusters of topics" refer to "traffic phase".

The rest of the paper is organized as follows. In section 2, a brief survey of the related works is presented. STC model is explained in Section 3. In Section 4, our two-level STC method for motion pattern mining and traffic phase discovery, temporal video segmentation and anomaly detection is introduced. Experimental results are shown in Section 5, and finally the paper is concluded in Section 6.

2- Related Work

The topic models can model relationships through the co-occurrence of simple features at different hierarchical levels. Following this hierarchical interpretation, researchers have proposed some models with multi-level framework. The authors of [8] adopted the Markov Clustering Topic Model (MCTM) which was built on LDA and Markov chain for activity learning and video clip clustering. A model was proposed in [9] that relied on a Dirichlet Process (DP) to discover activities and their occurrences. In [10], Diffusion Maps were used to embed the words into a lower dimensional space and to cluster them into motion patterns while video clips were clustered to determine co-occurring motion patterns.

The work in [11] relied on hierarchical PLSA to identify abnormal activities and repetitive cycles. The authors of [12] used a two level Hierarchical Dirichlet Process (called dual-HDP) to detect usual and unusual activities in traffic scenes. For this purpose, the Dual-HDP model was employed to cluster all the moving pixels into activities. Moreover, the video clips were clustered into interactions. This two-level motion pattern analysis provides a good representation of hierarchical nature of video scenes. The authors of [13] further extended the dual-HDP with the delta dual-HDP structure for jointly learning both normal and abnormal behavior using weakly supervised training examples.

A Markov clustering topic model with a three-level structure was presented in [14] which contained pixel motion, single object activity and co-occurring activities. The authors of [15] proposed a Mixed Event Relationship Model (MERM) that employed a binary activity matrix and discovered temporal relationships between activity pairs as well as global rules. Activities did not need to be dependent on the current state. Neither did they depend on any activity in the past.

A two-level LDA topic model was used to learn the scene behaviors and to detect the anomalies in [16]. The first level of this model learns the single-agent motions and the second level is exploited to learn the interactions. Such a method was also used in [17]. In [18], a two-stage cascaded LDA (Cas-LDA) model was formulated for automatic discovering and learning of behavioral context. In the first level, regional behavior and context are learned by LDA, and in the second, global context over the regional models are discovered. This behavioral context was further used for video based complex behavior recognition and anomaly detection. The authors of [19] proposed a novel two-level HDP model to discover both activities and traffic phases.

The method proposed in [20], called Dual-Sparse Topic Model (DsparseTM) and used for text analysis, is completely different from our proposed method. The DsparseTM [1] addresses the sparsity in both the topic mixtures and the word usage. By applying a "Spike and Slab", prior to decoupling the sparsity and smoothness of the document-topic and topic-word distributions, it allows individual documents to select a few focused topics and a topic to select focused terms, respectively.

Despite applying different topic models to traffic video analysis, simultaneous usage of the advantages of two-level structure and non-probabilistic topic models (e.g. STC) has not been investigated in previous research works. In this paper, we develop an extension of STC employing the two-level structure for traffic video analysis.

3- STC

STC [5] relaxes the normalization constraints made in probabilistic topic models. Such a relaxation makes STC enjoy nice properties, such as direct control on the sparsity of discovered representations, efficient learning algorithm, and seamless integration with a convex loss function for learning predictive latent representations. In STC, each individual input feature (e.g., a word count) is reconstructed from a linear combination of a set of bases, where the coefficient vectors (or codes) are un-normalized, and the representation of an entire document is derived via an aggregation strategy (e.g., truncated averaging) from the codes of all its individual features.

Suppose 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 $\boldsymbol{\beta}\in\mathbb{R}^{K\times N}$ be a dictionary with K bases, where each base is assumed to be a topic base, i.e. a unigram distribution over $\mathbf{v}$. We will use $\boldsymbol{\beta}_{.n}$ to denote the nth column of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_{k.}$ to denote the kth row of $\boldsymbol{\beta}$. Let P be a (N-1)-simplex, then $\boldsymbol{\beta}_{k.}\in P$. For the dth document $\mathbf{w}_d$, STC projects $\mathbf{w}_d$ into a semantic space spanned by a set of automatically learned topic bases $\boldsymbol{\beta}$ and achieves a high-level representation of the entire document jointly. STC is a hierarchical latent variable model, where $\boldsymbol{\theta}_d\in\mathbb{R}^K$ is the document code of document d and $\boldsymbol{s}_{d,n}\in\mathbb{R}^K$ is the word code of word n. Let $\boldsymbol{\Theta}=\{\boldsymbol{\theta}_d,\boldsymbol{s}_d\}_{d=1}^D$ denote the codes for a collection of documents $\{\mathbf{w}_d\}_{d=1}^D$. STC solves the optimization problem [5]:

$$\begin{aligned}&\min_{\boldsymbol{\Theta},\boldsymbol{\beta}}\sum_{d=1}^{D}\sum_{n=1}^{|I_d|}\left(\boldsymbol{s}_{d,n}^T\boldsymbol{\beta}_{.n}-w_{d,n}\ln(\boldsymbol{s}_{d,n}^T\boldsymbol{\beta}_{.n})\right)+\lambda_1\sum_{d=1}^{D}\|\boldsymbol{\theta}_d\|_1\\&+\lambda_2\sum_{d=1}^{D}\sum_{n=1}^{|I_d|}\|\boldsymbol{s}_{d,n}\|_1+\lambda_3\sum_{d=1}^{D}\sum_{n=1}^{|I_d|}\|\boldsymbol{s}_{d,n}-\boldsymbol{\theta}_d\|_2^2\\&s.t.\quad\boldsymbol{\theta}_d\geq 0,\forall d;\ \boldsymbol{s}_{d,n}\geq 0,\forall d,n;\ \boldsymbol{\beta}_{k.}\in P,\forall k\end{aligned}\tag{1}$$

The first part of (1) is equivalent to minimizing an un-normalized KL-divergence between observed word counts $w_{d,n}$ and their reconstructions $\boldsymbol{s}_{d,n}^T\boldsymbol{\beta}_{.n}$. The $\ell 1$-norm will bias

towards finding sparse codes. STC can learn meaningful topical bases, identify sparse topical senses of words and assign a sparse number of topics to each document.

4- Two-level STC model

To make a finer perception of the hierarchical nature in traffic scene, activities and traffic phases (which are the co-occurrence of activities generated by the traffic signals) need to be discovered automatically. In other words, our goal is to model not only topics but also the clusters of topics, both of which are shared among documents. Consequently, the two-level STC model is proposed. As named, it contains two levels. The original STC model introduced in section 2 is used in both levels.

The proposed method includes two levels of motion pattern mining. At each level, the STC model, with different definitions of words and topics, is used to discover the frequent motion patterns that exist in video data. The flowchart of our method is shown in Fig. 1. At the first level STC modeling, video sequences are divided into short clips, which are regarded as documents. Motion features are regarded as visual words. Activities (topics) are the motion patterns represented as a mixture of visual words. At the second level, we keep the same video clips as documents but consider the activities discovered by the first level STC as words. Due to the second level STC modeling, traffic phases (clusters of topic) are discovered, which are represented as a mixture of activities.

As shown in Fig. 1, firstly, based on the visual words extracted from training video clips, the first level and second level STC models are applied to learn the topics and traffic phases, respectively. Then, the learned topics are employed to define the topic proportion of the test video clip through first level STC modeling. Using the topic proportions as input for the second level STC, the learned traffic phases are employed to define the traffic phase proportion of the test video clip. Finally, the test video clip is assigned to a traffic phase with highest value in the traffic phase proportion.

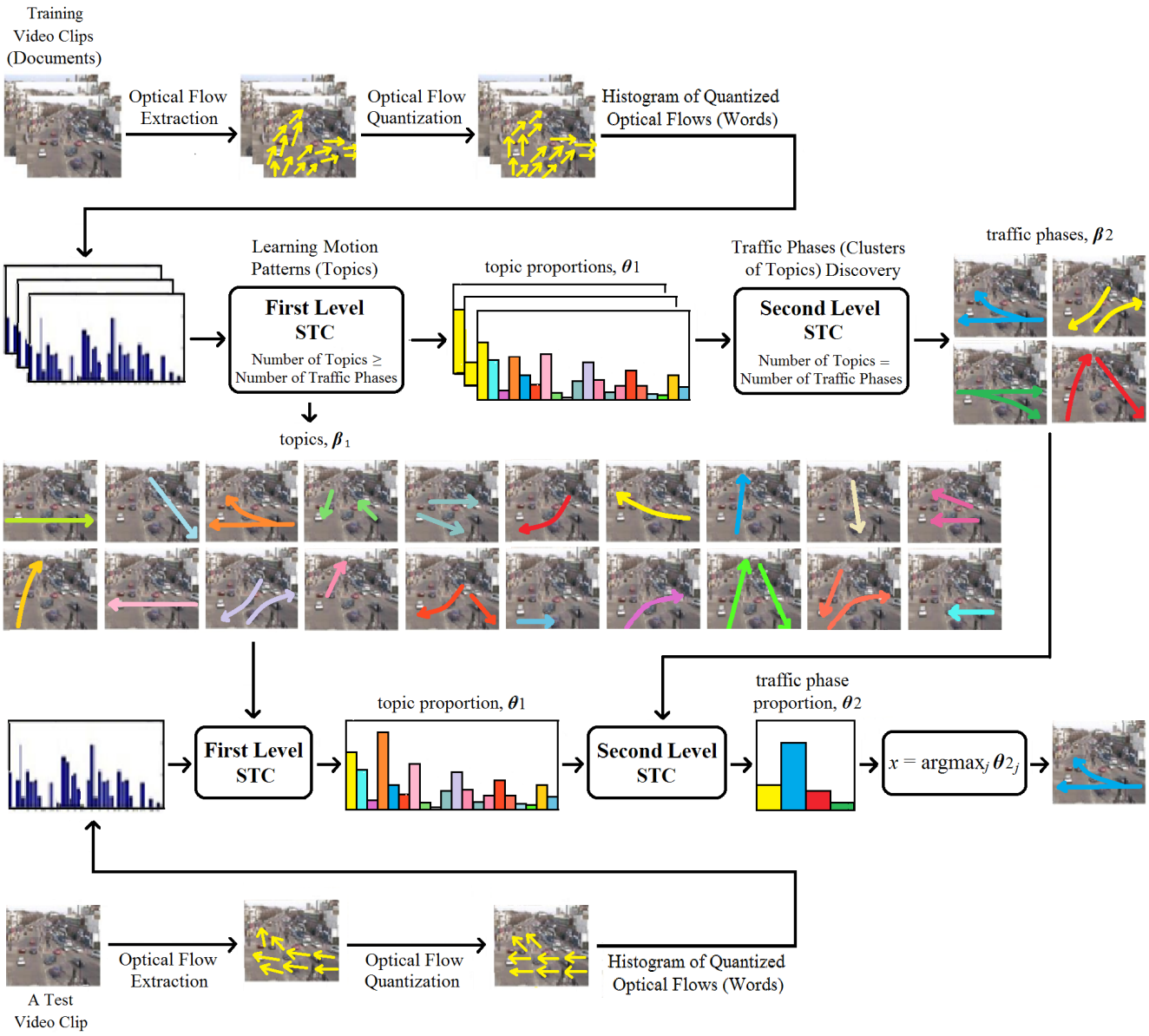

Fig. 1. A graphical representation of two-level STC method for traffic phase discovery

4- 1- Activity Learning by the First Level STC

Given an input video, the whole video sequence is segmented into D non-overlapping short video clips. Each clip is considered as a "document" in STC context.

We utilize Shi and Tomasi corner detector [21] to find the key points and use these features to extract the optical flow using Lucas–Kanade method [22] 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. Each quantized optical flow vector is considered as a "word" in STC context. Optical flow vectors are denoted by (x, y, α). The positions (x, y) are quantized to the nearest position on a grid of 10-pixels spacing, and the angles of flow vectors, α, are quantized into 8 directions. Finally a fixed vocabulary is formed as $\mathbf{v}=\{v_1,\ldots,v_N\}$ with N total flow words, in which each word contains information about position and motion direction.

A video clip is represented as a vector $\mathbf{w}=\{w_1,\ldots,w_N\}$, where w_n denotes the number of appearances of word n in the clip. Using a word-document topic model, flow words with high co-occurrence frequencies in a video clip make a motion pattern (activity). Each motion pattern is considered as a "topic" in STC context. Motion patterns are represented as dictionary, $\boldsymbol{\beta}1$, whose rows show the typical topics in the video which are a mixture of words from the vocabulary $\mathbf{v}$.

In the first level STC model, the observations are words in documents, i.e. the quantized optical flow word in short clips, and the latent variables to be modeled are topics shared among documents, i.e. the motion patterns (activities) which are shared among all video clips. Activities are modeled as a mixture of flow words, and clips are modeled as a mixture of activities. The graphical model of the first level STC is shown in Fig. 2. In this figure, $\boldsymbol{\beta}1\in\mathbb{R}^{K\times N}$ is the dictionary of K topics $\boldsymbol{\beta}1_k$; document code $\boldsymbol{\theta}1_d$ is the topic proportion of document d; $w_{d,n}$ denotes the number of appearances of word n in document d; word code $\boldsymbol{s}1_{d,n}$ denotes the proportion of topics assigned to the word n in document d; D is the number of documents and N is the total number of words in the vocabulary (I_d is the number of words that appear in the document d); K is the number of topics.

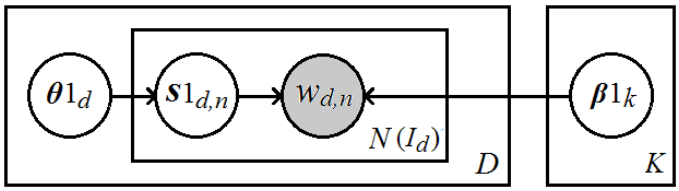

Fig. 2. The graphical model of the first level STC

4- 2- Traffic phase detection by Second Level STC

At the second level STC, our goal is to find out the mixture motion patterns defined by certain combinations of activities occurring at one time. The same video clips are denoted as documents, but the observations are the topics discovered by the first level STC, and the latent variables are clusters of topics which are also shared among documents. Therefore the traffic phases (cluster of topics) are discovered, which are modeled as a mixture of motion patterns (topics).

By performing the second level STC, the co-occurring activities are discovered and considered as traffic phase. The second level STC model is shown in Fig. 3. In this figure, $\boldsymbol{\beta}2\in\mathbb{R}^{L\times K}$ is the dictionary of L traffic phases $\boldsymbol{\beta}2_l$; $\boldsymbol{\theta}2_d$ is the traffic phase proportion of document d; $\boldsymbol{\theta}1_{d,k}$ is the proportion of topic k in document d obtained from the first level STC; $\boldsymbol{s}2_{d,k}$ denotes the proportion of traffic phases assigned to the topic k in document d; D is the number of documents and K is the total number of topics (J_d is the number of topics that appear in document d); L is the number of traffic phases.

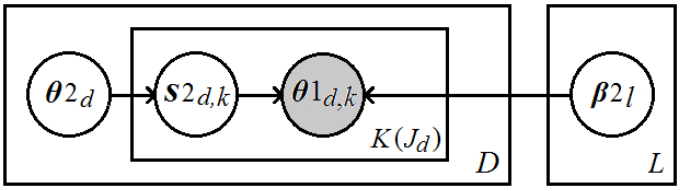

Fig. 3.The graphical model of the second level STC

4- 3- Temporal Video Segmentation

Given a long video sequence, it can be segmented based on different traffic phases. Our method provides an unsupervised manner to complete this task, by automatically assigning the video clips into traffic phases. From the two-level STC, we obtain a hierarchical representation of the dynamics contained in the video: the activities modeled as a mixture of visual words and the traffic phases modeled as a mixture of activities. Based on the STC model, we first use some video clips to train the model, detect the activities and discover the traffic phases. Then, each clip is modeled as the mixture of activities and is assigned to the traffic phase which has the maximum value in the traffic phase proportion. Therefore, the video can be segmented based on the assignments. The final result is a video, segmented into different traffic phases.

4- 4- Anomaly Detection

With the two-level motion pattern discovery, videos can be interpreted by the following hierarchical structure: motion features (visual words), activities (topics), and traffic phases (clusters of topic).

Specifically, every visual word at each clip can be assigned to a certain activity and every activity can be assigned to a certain traffic phase. Therefore, motion anomalies can be detected at two levels:

- *Activity anomaly*: visual words do not belong to any of the activities;
- *Traffic phase anomaly:* activities cannot coexist with others in that clip according to the corresponding traffic phase.

Using the first level STC with assuming a dictionary $\boldsymbol{\beta}$, for a video clip d with document code $\boldsymbol{\theta}_d$ and word codes $\boldsymbol{s}_d$, a sparse reconstruction cost (SRC) is [7]:

$$f1_{SRC} = \sum_{n=1}^{|I_d|}\left(\boldsymbol{s}1_{d,n}^T\boldsymbol{\beta}_{,n} - w_{d,n}\ln(\boldsymbol{s}1_{d,n}^T\boldsymbol{\beta}1_{,n})\right) + \lambda_1\left\|\boldsymbol{\theta}1_d\right\|_1 + \lambda_2\sum_{n=1}^{|I_d|}\left\|\boldsymbol{s}1_{d,n}\right\|_1 + \lambda_3\sum_{n=1}^{|I_d|}\left\|\boldsymbol{s}1_{d,n} - \boldsymbol{\theta}1_d\right\|_2^2 \quad (2)$$

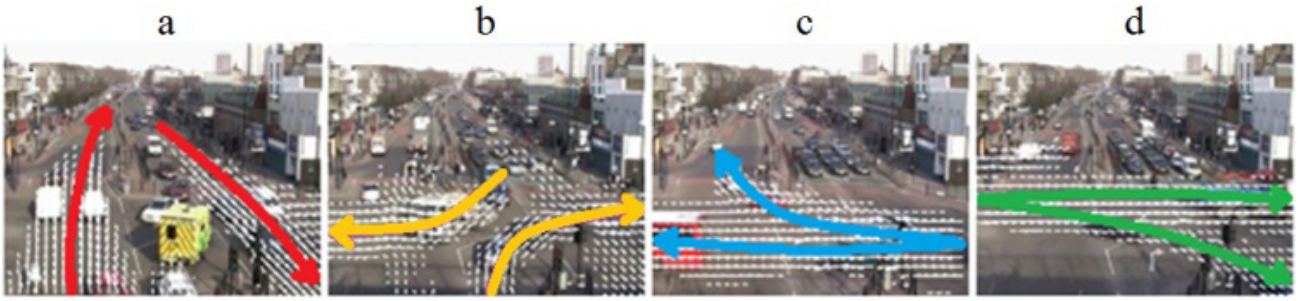

Fig. 4. Traffic phases in QMUL Junction dataset

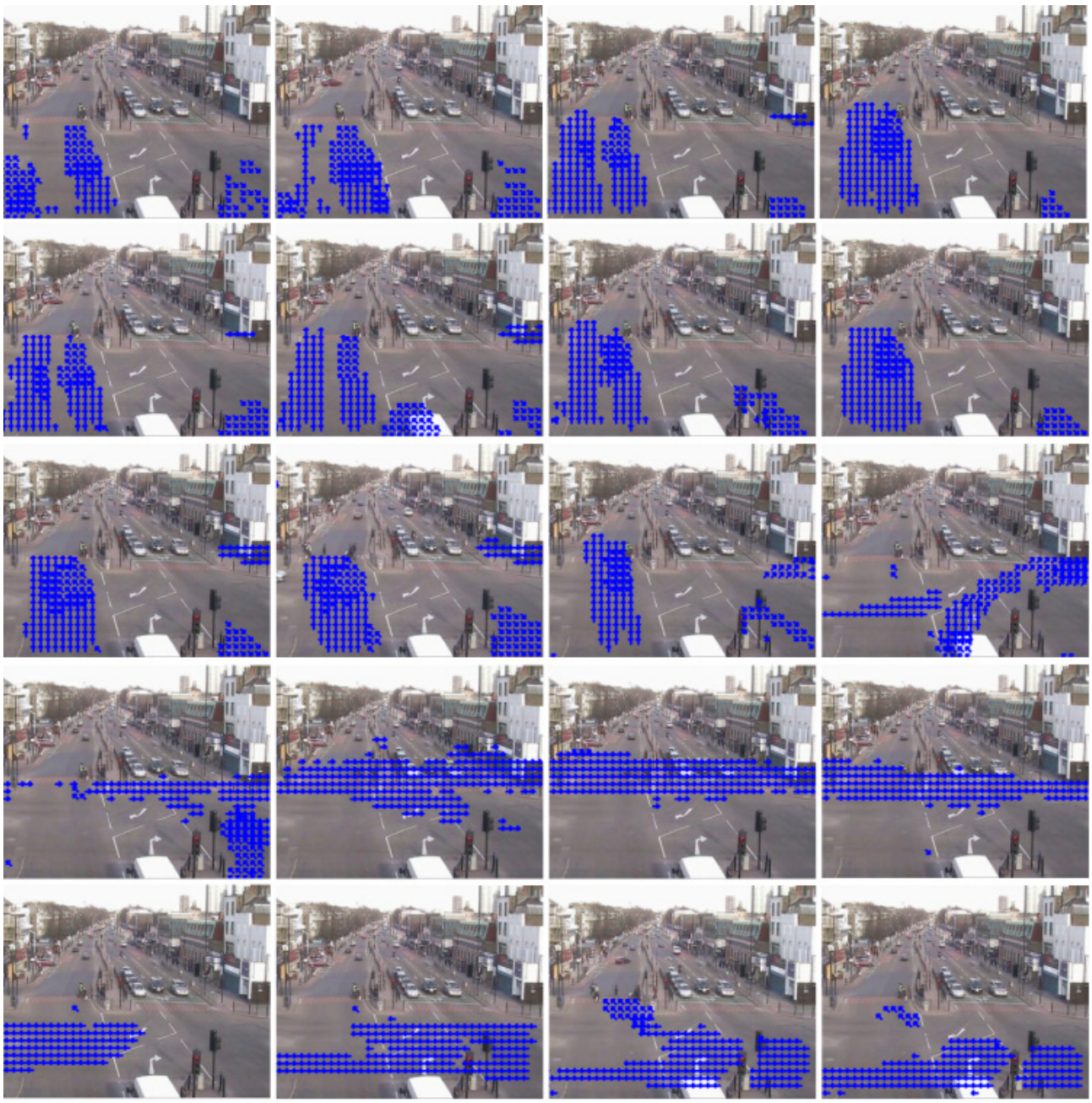

Fig. 5. The motion patterns discovered by the first level STC

and the SRC of clip d using the second level STC is:

$$f2_{SRC} = \sum_{k=1}^{|J_d|} \left(s2_{d,k}^T \beta 2_{.k} - \theta 1_{d,k} \ln(s2_{d,k}^T \beta 2_{.k}) \right) + \lambda_1 \|\theta 2_d\|_1 + \lambda_2 \sum_{k=1}^{|J_d|} \|s2_{d,k}\|_1 + \lambda_3 \sum_{k=1}^{|J_d|} \|s2_{d,k} - \theta 2_d\|_2^2 \quad (3)$$

Obviously, a clip with a high f_{1SRC} or f_{2SRC} tends to be with abnormalities. We consider $f_{1SRC} + f_{2SRC}$ as the SRC for abnormality detection using our proposed two-level STC.

5- Experimental Results

We evaluated the performance of our proposed method on the widely used QMUL Junction dataset [23]. The dataset has been recorded at 25 frames per second from a busy traffic intersection. The video files have been divided into 12-second non-overlapping clips, with a frame size of 360×288 pixels. The Junction dataset contains 73 video clips for training and 39 video clips for testing.

QMUL Junction is governed by traffic lights, dominated by four types of traffic phases as illustrated in Fig. 4. Specifically, Flow "a" corresponds to traffic in vertical directions. Flows "b", "c" and "d" are regarded as traffic flows in horizontal directions. In particular, Flow "b" represents left-turning and right-turning traffics with some vertical traffic. Flow "c" corresponds to rightward traffic and Flow "d" corresponds to leftward traffic.

There are 8 abnormal activities in the 39 test clips of the Junction dataset. The abnormal activities occur in the test clip numbers of 4, 8, 18, 22, 27, 28, 30 and 31. The abnormal events defined in these clips are dangerous driving, traffic rule violations, interrupting the traffic flow, and rare maneuvers such as U-turns.

We set the hyper parameters of STC as λ_1=0.5, λ_2=0.2, λ_3=0.2 through our experiments.

5- 1- Activity Learning by the First Level STC

Fig. 5 shows the non-zero motion patterns (activities), discovered by the first level STC model. The number of topics is set to 20. As it can be seen, each motion pattern has a clear semantic meaning. For example, the upward traffic lane, the downward traffic lane, the leftward traffic, and so on.

5- 2- Traffic phase detection by Second Level STC

Fig. 6 shows the traffic phases (a, b, c and d), discovered by the second level STC model. The number of topics is set to the number of possible traffic phases, which is equal to 4 in QMUL Junction dataset. As illustrated in this figure, compared to the traffic phases shown in Fig. 4, the two-level STC method has found meaningful traffic phases.

5- 3- Temporal Video Segmentation

The segmentation result is evaluated by comparing the recognized traffic phase of each clip with the traffic phase provided by the ground-truth. The ground-truth has been created by manually labeling the whole QMUL Junction dataset into 4 traffic phases. To find the correspondence between the traffic phases from ground-truth data (the actual phases) and the traffic phases recognized by the two-level STC (the learnt phases), we employed Kuhn-Munkres algorithm [24,25]. Using Kuhn-Munkres algorithm, the learnt phases are matched with the actual phases such that the accuracy with regards to the training data is maximized. The video segmentation results are shown by the bar graph in Fig. 7 for the training and test video clips. All video clips are labeled as phase 1 (phase "a"), 2 (phase "b"), 3 (phase "c") or 4 (phase "d"). As it can be seen in this figure, our traffic phase detection results for both training and test video clips are very close to the ground truth.

The accuracy of temporal video segmentation for the training and test video clips using the two-level STC are reported in Table 1. For comparison, we also report the accuracy values of the one-level PLSA, LDA, STC and FSTM (which their codes are available online at [26-29]) and also the two-level PLSA, LDA and FSTM in Table 1. As it can be seen in this table, the accuracy of the two-level STC is 83.56% and 76.92% for the training and test video clips, respectively, which is higher than other two-level topic models. However, compared to the one-level topic models, the two-level topic models often lead to lower performance in traffic phase detection. This shows that performing topic modeling and traffic phase detection in one level presents higher performance compared to separating them into two levels. The reason is that traffic phase detection and topic modeling can mutually promote each other. Information on the traffic phase of video clips helps to solve the ambiguity of motion features in discovering the motion patterns, and vice versa. Thus, coupling traffic phase detection and topic modeling into a unified framework produces superior performance than separating them into two procedures.

5- 4- Anomaly Detection

Based on the two-level structure, we can detect either activity anomalies or traffic phase anomalies, which cannot be achieved by the one-level structure. Fig. 8 shows the ROC curve for the two-level STC. The ROC curve has been plotted through changing the abnormality threshold over $f1_{SRC}+f2_{SRC}$. The SRC values above the threshold are regarded as abnormality. For comparison, the ROC curve for the one-level STC is also shown in Fig. 8. As it can be seen in this figure, compared to the one-level STC, the two-level STC

Fig. 6. Four discovered traffic phases in QMUL Junction dataset by the second level STC

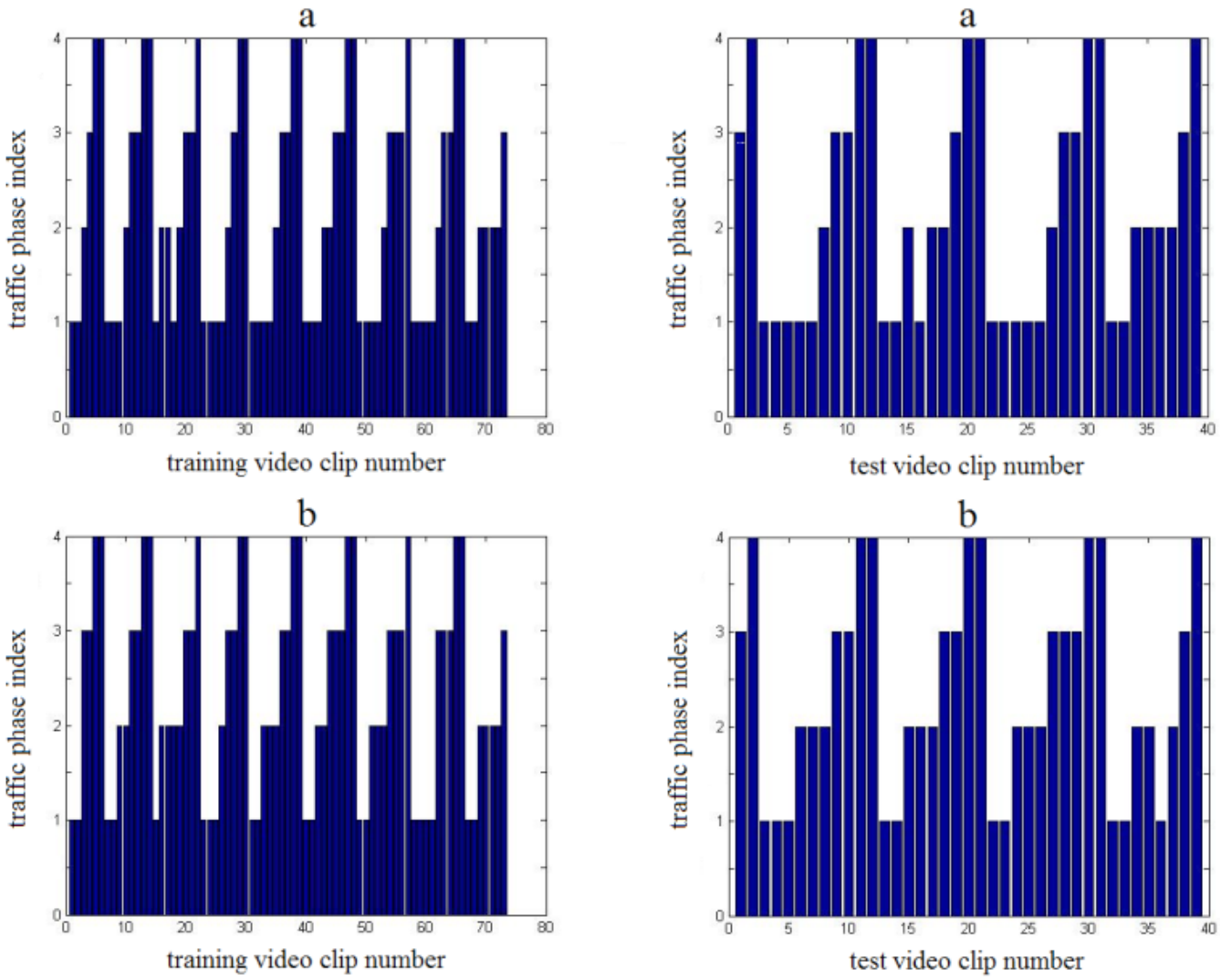

Fig. 7. Results of temporal video segmentation for: left) training video clips, right) test video clips; a) ground truth, b) the two-level STC result

achieves higher performance in abnormality detection.

The Area Under ROC (AUROC), True Positive Rate (TPR) and False Positive Rate (FPR) values for abnormality detection using the two-level STC are reported in Table 2. For comparison, we also report the AUROC, TPR and FPR values of the one-level STC, PLSA, LDA and FSTM and also the two-level PLSA, LDA and FSTM in Table 2. As it can be seen in this table, compared to the one-level topic models, the two-level topic models lead to better performance in abnormal event detection. Among different topic models, the best result is achieved by using the two-level STC with the values of AUROC=74% and FPR=26.4% (TPR=75%). This achievement is because of the two-level structure that finds the abnormalities in two levels, activity anomalies and traffic phase anomalies, which cannot be achieved by a one-level structure. Moreover, advantages of the non-probabilistic topic model, STC, lead to achieve higher performance in abnormality detection, compared to the probabilistic topic models (PLSA, LDA and FSTM).

5- 5- Computational Complexity

Our experiments have been performed on an Intel CoreTM i7-4790 3.6 GHz CPU with 32 GB RAM, running Linux (Ubuntu 14.04). Using Python implementation, our two-level STC method takes about 7 minutes for training model on 73 12-second 360×288-pixel clips in QMUL Junction dataset.

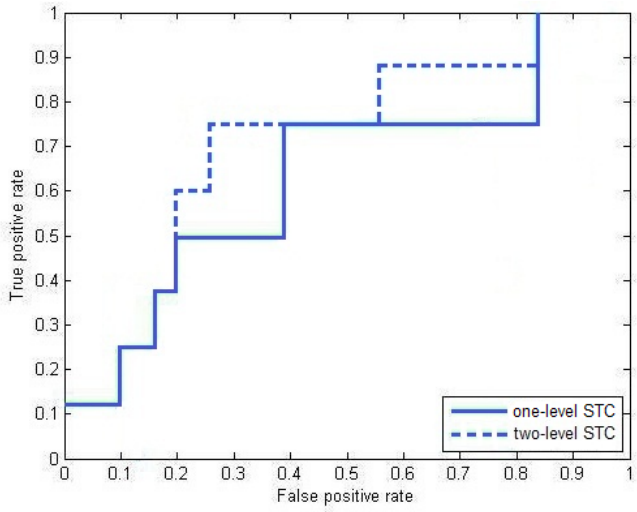

Fig. 8. The ROC curves for the one-level and two-level STC

Table 1. The accuracy of temporal video segmentation for the training and test video clips in QMUL Junction dataset

	Topic Model	Training Data	Test Data
one-level topic models	PLSA [3]	78.08	76.92
	LDA [4]	87.67	74.36
	STC [6]	86.30	**89.74**
	FSTM [5]	**90.47**	89.34
two-level topic models	PLSA	71.30	72.54
	LDA	80.82	**76.92**
	STC	**83.56**	**76.92**
	FSTM	78.08	71.79

Table 2. The AUROC values, TPR and FPR values (%) for abnormal event detection in QMUL Junction dataset

	Topic Model	AUROC	FPR (TPR=75)	FPR (TPR=87.5)	FPR (TPR=100)
one-level topic models	PLSA [3]	0.68	64.5	87.1	90.3
	LDA [4]	**0.71**	51.6	87.1	90.3
	STC [6]	0.69	38.7	83.9	83.9
	FSTM [5]	0.68	54.8	93.5	100.0
two-level topic models	PLSA	0.70	61.8	87.1	90.3
	LDA	0.73	39.7	69.3	90.3
	STC	**0.74**	26.4	56.5	83.9
	FSTM	0.70	53.2	90.3	100.0

Testing 39 12-second clips with the same resolution and the same dataset, for spatial abnormal event detection, requires about 3 seconds. For one-level STC, training and testing takes about 6 minutes and 2.5 seconds, respectively. This shows that the computational overhead of our two-level approach over one-level one is very low.

6- Conclusion

In this paper, we propose a hierarchical motion pattern mining method to interpret a dynamic surveillance video scene. A two-level STC model is introduced to discover both activities and traffic phases in the video. In the first level STC, the clusters of topics are modeled as a mixture of topics. After that, all the documents are modeled as a mixture of the clusters of topics by the second level STC. Experiments on a real surveillance videos show that the two-level STC can discover not only the lower level activities but also the higher level traffic phases, which makes a more appropriate interpretation. Furthermore, temporal video segmentation and anomaly detection may also be achieved based on the results.

References

[1] O.P. Popoola, K. Wang, Video-based abnormal human behavior recognition—a review, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on., 42(6) (2012) 865-78.

[2] L. Song, F. Jiang, Z. Shi, R. Molina, A.K. Katsaggelos, Toward dynamic scene understanding by hierarchical motion pattern mining, Intelligent Transportation Systems, IEEE Transactions on., 15(3) (2014) 1273-85.

[3] T. Hofmann, Probabilistic latent semantic analysis, UAI, (1999) 289-296.

[4] D.M. Blei, A.Y. Ng, M.I. Jordan, J. Lafferty, Latent Dirichlet allocation, Journal of Machine Learning Research, (3) (2003) 993-1022.

[5] K. Than, T.B. Ho, Fully Sparse Topic Models, Machine Learning and Knowledge Discovery in Databases, 7523 (2012) 490-505.

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

[7] W. Fu, J. Wang, H. Lu, S. Ma, Dynamic scene understanding by improved sparse topical coding, Pattern Recognition., 46(7) (2013) 1841-50.

[8] T. Hospedales, S. Gong, T. Xiang, A Markov Clustering Topic Model for Mining Behaviour in Video, IEEE International Conference on Computer Vision, Kyoto, Japan, (2009) 1165-1172.

[9] R. Emonet, J. Varadarajan, J. Odobez, Extracting and Locating Temporal Motifs in Video Scenes Using A Hierarchical Non Parametric Bayesian Model, IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, USA., (2011) 3233-3240.

[10] Y. Yang, J. Liu, M. Shah, Video Scene Understanding Using Multi-scale Analysis, IEEE International Conference on Computer Vision, Kyoto, Japan, (2009) 1669-1676.

[11] J. Li, S. Gong, T. Xiang, Global behaviour inference using probabilistic latent semantic analysis, British Machine Vision Conference, 3231 (2008) 3232.

[12] X. Wang, X. Ma, E.L. Grimson, Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(3) (2009) 539-555.

[13] T. Haines, T. Xiang, Delta-dual hierarchical Dirichlet processes: a pragmatic abnormal behaviour detector, Proc. IEEE Int. Conf. Computer Vision, Barcelona, Spain, (2011) 2198-2205.

[14] T.M. Hospedales, S. Gong, T. Xiang, Video Behaviour Mining using A Dynamic Topic Model, International Journal of Computer Vision, 98(3) (2012) 303-323.

[15] J. Varadarajan, R. Emonet, J.M. Odobez, Bridging The Past, Present and Future: Modeling Scene Activities From Event Relationships and Global Rules, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2012) 2096-2103.

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

[17] Y. Fan, S. Zheng, Dynamic Scene Analysis Based on the Topic Model, 2nd IEEE International Symposium on Instrumentation and Measurement, Sensor Network and Automation (IMSNA), (2013) 436-439.

[18] J. Li, S. Gong, T. Xiang, Learning behavioural context, International Journal of Computer Vision, 97(3) (2012) 276-304.

[19] L. Song, L. Mei, Z. Liu, H. Duan, N. Liu, J. Wang, C. Hu, Motion perception for traffic surveillance, IEEE 17th International Conference on Intelligent Transportation Systems (ITSC), (2014) 1298-1303.

[20] T. Lin, W. Tian, Q. Mei, H. Cheng, The Dual-Sparse Topic Model: Mining Focused Topics and Focused Terms in Short Text, Proceedings of the 23rd international conference on World wide web, Seoul, Korea, (2014) 539-550.

[21] J. Shi, C. Tomasi, Good features to track, Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, Washington, (1994) 593-600.

[22] B.D. Lucas, T. Kanade, An Iterative Image Registration Technique with an Application to Stereo Vision, Proceedings of imaging understanding workshop, (1981) 674-679.

[23]http://www.eecs.qmul.ac.uk/~sgg/QMUL_Junction_Datasets/Junction/Junction.html

[24]H.W. Kuhn, The hungarian method for the assignment problem, Naval research logistics quarterly, 2(1-2), (1955) 83-97.

[25] J. Munkres, Algorithms for the assignment and transportation problems, Journal of the Society of Industrial and Applied Mathematics, 5(1) (1957) 32-38.

[26] PLSA: http://lear.inrialpes.fr/~verbeek/code/plsa.tar.gz

[27] LDA: http://www.cs.columbia.edu/~blei/lda-c/

[28] STC: http://bigml.cs.tsinghua.edu.cn/~jun/stc.shtml

[29] FSTM: https://www.jaist.ac.jp/~s1060203/codes/fstm

Please cite this article using:

P. Ahmadi, I. Gholampour, M. Tabandeh, Traffic Scene Analysis using Hierarchical Sparse Topical Coding, *AUT J. Mech. Eng.*, 50(3) (2018) 177-186.

DOI: 10.22060/eej.2018.12366.5065