THESIS

LOCALIZED ANOMALY DETECTION VIA HIERARCHICAL INTEGRATED ACTIVITY DISCOVERY

Submitted by

Thiyagarajan Chockalingam

Department of Electrical and Computer Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2014

Master's Committee:

Advisor: Sanjay Rajopadhye
Co-Advisor: Chuck Anderson

Sudeep Pasricha
Wim Bohm

ABSTRACT

LOCALIZED ANOMALY DETECTION VIA HIERARCHICAL INTEGRATED ACTIVITY DISCOVERY

With the increasing number and variety of camera installations, unsupervised methods that learn typical activities have become popular for anomaly detection. In this thesis, we consider recent methods based on temporal probabilistic models and improve them in multiple ways. Our contributions are the following: (i) we integrate the low level processing and the temporal activity modeling, showing how this feedback improves the overall quality of the captured information, (ii) we show how the same approach can be taken to do hierarchical multi-camera processing, (iii) we use spatial analysis of the anomalies both to perform local anomaly detection and to frame automatically the detected anomalies. We illustrate the approach on both traffic data and videos coming from a metro station. We also investigate the application of topic models in Brain Computing Interfaces for Mental Task classification. We observe a classification accuracy of up to 68% for four Mental Tasks on individual subjects.

ACKNOWLEDGMENTS

This thesis is a culmination of inspiration and knowledge obtained during my collaboration with research groups at CSU and Idiap Research Institute. The foundation for this thesis was developed during my internship at Idiap Research Institute. I would like to thank my supervisors Dr. Remi Emonet and Dr. Jean Marc Odobez from Idiap Research Institute for their thought provoking discussions and the internship opportunity. I would also like to thank Remi Emonet for his continued support and guidance after the internship and for proof reading this thesis.

My advisor at CSU Dr. Chuck Anderson provided me the freedom to work on areas that are not part of his core BCI research group. His insight and depth of knowledge in the field of BCI helped me to apply probabilistic modeling to BCI. I would also like to thank my Advisor Dr. Sanjay Rajopadhye and committee members Dr. Wim Bohm and Dr. Sudeep Pasricha.

I would also like to thank my parents for their support in various decisions and having faith in my judgements. I also like to thank my friends at CSU, in particular Bala, Bhavesh, Karthik, Sridhar and Vamshi for making my stay at CSU a memorable one.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

An increasing number of camera networks are being deployed to ensure safety and abnormal event detection through visual surveillance. Even if some applications can afford systematic human monitoring, this is surely impossible when the number of cameras in the network is huge. It has become of prime importance to design algorithms able to handle this vast amount of data, filter out typical activities and show the most abnormal parts to human operators. In this thesis, we improve over recent approaches to do anomaly detection and video abnormality characterization. There are two main kinds of approaches for anomaly detection. The main distinction between them lies in modeling abnormalities or modeling usual activity.

The first kind explicitly model and learn to recognize abnormal events. From a well specified event type, one can build dedicated detectors [3] that usually perform well. These approaches have the drawbacks that abnormal events have to be defined in advance, and a variety of training data have to be gathered for these events. These approaches are thus not adequate in large camera networks where no supervision is expected. Even if these approaches allow to specify in a direct way what is to be detected, this actually has some drawbacks: abnormal events have to be defined in advance, and a variety of training data have to be gathered for these events. These steps require human intervention and might need to be re-executed for new cameras or view points. These constraints limit the application of these approaches in large camera networks where no supervision of the algorithms is expected at camera installation.

1.1 Related Work

Given the above limitations, unsupervised methods have gained interest recently. As these approaches cannot rely on pre-defined abnormality classes, they rather learn what is a normal activity and they consider as an anomaly anything that deviates from these normal activities. Different features have been used to characterize videos. In the context of public spaces, person tracking with person re-identification across cameras provides an effective solution to abnormal behavior detection [2, 30, 33]. However, robust tracking requires sufficient resolution and frame-rate, and, often surveillance cameras have low resolution, low quality (dirty, blurry, etc.) and low frame-rate (e.g., 5 frames per second). This profile of cameras explain the growing interest in relying on lower level features such as background subtraction information [22] or localized motion in the form of tracklets [15, 16] or optical flow [10].

Probabilistic methods have been shown very effective in handling these low level features in a principled way. Originally designed for text semantic analysis and after their success in many domains, various Topic Models has been proposed and applied for activity modeling [16, 20]. In this thesis, we build upon the Probabilistic Latent Sequential Motifs (PLSM) model that have been proposed in [26, 28]. The main advantage of PLSM is its capacity of automatically (with no supervision) extracting motifs (temporal patterns) that capture strong temporal information in temporal documents represented by $word \times time$ count matrices. Applied to traffic or metro station videos, the motifs are shown to capture the typical activities (related to trajectories) observed in a scene, as illustrated in Fig. 5.2. PLSM has been used for anomaly detection in surveillance video [10].

In previous works, PLSM was applied to documents built from an intermediate representation learned by dimensionality reduction of the low-level features. This intermediate representation had been learned in advance which had two drawbacks: it made it possible to create artifacts for PLSM, and the learning ignored temporal information, and thus was not benefiting by the temporal structure that PLSM can provide. Also, when used for anomaly

detection, PLSM was not considering the semantic of the intermediate representation and thus it was for example ignoring the spatial layout of anomaly in the scene. In this thesis we explore solutions to these two restrictions in PLSM usage.

1.2 Contributions

The overall aim of this thesis is to improve the anomaly detection approach using PLSM [26] by (i) jointly learning the dimensionality reduction representation along with the PLSM temporal model, and (ii) reasoning about the spatial distribution of anomalies in the image. We also investigate the application of PLSM to Brain Computing Interfaces.

We achieve the integration of the dimensionality reduction step (PLSA) with the PLSM model through Dirichlet prior feedback. The feedback encodes the relevance of the information (topics) captured by PLSA with respect to a sequential pattern. Topics which are not useful for temporal modeling are discouraged by this feedback, hence allowing PLSA to capture better topics. We provide some qualitative evidence on surveillance cameras in Chapter 5 which illustrates the improvement of temporal patterns with feedback.

We achieve localized abnormality detection through blocking. Blocking divides the frame of a video into sub-frames. The reconstruction errors for individual blocks are calculated and then the group of blocks that are most abnormal are identified.

We investigate the applicability of PLSA and PLSM for the first time for mental task Classification. Here we consider the frequency domain representation of brain waves from multiple electrodes as our vocabulary. This vocabulary representation is useful for finding the frequencies and channels that define each mental task.

1.3 Thesis organization

We now provide a brief outline of each of the chapters in this thesis. Chapter 2 reviews the necessary statistical concepts that are utilized throughout this thesis. The chapter begins

by introducing the principles of parameter estimation methods like MLE and MAP. We then describe the Expectation Maximization (EM) algorithm which is a technique used for parameter estimation of incomplete data. The chapter then proceeds to discuss conjugate priors and finally topic modeling approaches.

The datasets used for testing our models is described in Chapter 3. We explain the traffic and metro surveillance videos and also motivate the choice of these videos. The description of the Brain Computing Interface (BCI) dataset is deferred until Chapter 8.

Chapter 4 motivates the need for jointly learning the dimensionality reduction representation along with the PLSM temporal model by describing the previous approach to modeling. Chapter 5 and Chapter 6 present the two main thesis contributions: The introduction of the Integrated PLSM model and its extension to hierarchical multi-camera processing. Results are shown in Chapter 7 for traffic data and for public space data (metro station).

The application of PLSA and PLSM to BCI is discussed in Chapter 8. Here we describe the dataset used, feature extraction for vocabulary construction and results on nine subjects. We also illustrate and interpret the distributions obtained from PLSA applied to frequency domain transformed EEG signals recorded when subjects were performing mental tasks. Finally, Chapter 9 concludes the work.

Chapter 2

Background

In this chapter, the concepts required to better understand this thesis are briefly described. Sections 2.1 and 2.2 describe the parameter estimation problem and explains estimation of parameters on incomplete data using the Expectation Maximization algorithm. Section 2.3 explains the conjugate priors for binomial distribution and generalizes it to a Multinomial distribution. The final section 2.4 describes the topic models, Probabilistic Latent Semantic Analysis (PLSA) and Probabilistic Latent Sequential Motifs (PLSM).

2.1 Parameter Estimation

Parameter estimation [12] is the problem of finding the parameters θ for a set of distributions that best explains the observations X. The dataset $X = \{x_i\}_{i=1}^{|X|}$ can be considered as a set of observations generated independently and identically distributed realizations of a random variable. The parameters θ, depends on the distributions considered, for example, Multinomial distribution, $\theta = \{p_i\}_{i=1}^{i=D}$, where D is the cardinality of the possible outcomes.

The joint distribution $P(X, \theta)$ describes the probability of the observations and the vector θ. Bayes' theorem gives the relationship between the probabilities X and θ as below:

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}, P(X) = \Sigma_\theta P(X, \theta). \tag{2.1}$$

The interpretation of the distributions in Eq. 2.1 is given below:

$$posterior \ \propto \ likelihood \times prior \tag{2.2}$$

The proportionality is due to the fact that $P(X)$ is a marginal over θ and hence a constant for any prior $P(\theta)$. Below, we explain some of the methods for parameter estimation. We

will start with simple Maximum Likelihood Estimation (MLE) and then describe how prior belief can be included in the estimation.

2.1.1 Maximum Likelihood Estimation

MLE is the method of finding the parameters of the model that maximizes the probability of the observations (likelihood) under the resulting distribution. This approach only estimates the parameters with respect to the observations and doesn't incorporate prior belief, hence $P(\theta)$ is not considered. The likelihood is,

$$\ell(\theta|X) = P(X|\theta) = \prod_{i=1}^{|X|} P(x_i|\theta). \tag{2.3}$$

Because of the product, it is often mathematically convenient to express the likelihood, ℓ, as the log-likelihood,

$$L(\theta|X) = \sum_{i=1}^{|X|} log(P(x_i|\theta)). \tag{2.4}$$

The MLE can then be formulated as,

$$\theta_{ML} = \underset{\theta}{\text{argmax}}(L(\theta|X)). \tag{2.5}$$

The parameter θ then can be estimated by solving the Eq. 2.5 as follows:

$$\frac{\partial}{\partial\theta_d} L(\theta|X) = 0; \forall\theta_d \in \theta \tag{2.6}$$

As an example, consider a set X of N Bernoulli experiments of an unfair coin toss with unknown parameter θ. The probability of the event x, for a single experiment, for the random variable X_i is,

$$P(X_i = x|\theta) = \theta^x \cdot \theta^{1-x}. \tag{2.7}$$

where $x = 1$ is heads and $x = 0$ is tails. The MLE for θ can be found by solving Eq. 2.5,

$$\begin{aligned} L &= \sum_{i=1}^{N} (log(P(X_i = x|\theta))) && (2.8)\\ &= (n^1 log(P(X_i = 1|\theta))) + n^0 log(P(X_i = 0|\theta))), && (2.9) \end{aligned}$$

where n^1 denotes the number of heads and n^0, the number of tails.

$$\frac{\partial}{\partial\theta}(L) = \frac{n^1}{\theta} - \frac{n^0}{1-\theta} = 0, \tag{2.10}$$

$$\theta_{ML} = \frac{n^1}{n^1 + n^0} = \frac{n^1}{N}. \tag{2.11}$$

which is the ratio of heads to the total number of samples. It can be seen from Eq. 2.11, the MLE estimates parameters which best explains the observations. If the observations or the sample dataset (subset of sample space) is not a good representative of the population (sample space) then the MLE estimate approach over-fits the parameters to the sample dataset.

2.1.2 Maximum a Posteriori Estimation

Maximum a posteriori (MAP) estimation is similar to MLE but also incorporates a mechanism to add prior belief in the form of a prior distribution. In MAP, the parameters of the model are obtained by maximizing the posterior distribution in Eq. 2.1 with respect to the model parameters θ

$$\theta_{MAP} = \underset{\theta}{\text{argmax}}(P(X|\theta)P(\theta)) \quad | \; P(X) \neq f(\theta) \tag{2.12}$$

$$= \underset{\theta}{\text{argmax}}(\sum_{i=1}^{N} log(P(x_i|\theta)) + log(P(\theta))) \tag{2.13}$$

$$= \underset{\theta}{\text{argmax}}(L + log(P(\theta))), \tag{2.14}$$

Continuing with the coin example as in MLE, the prior distribution $P(\theta)$ can be represented by the Beta distribution (explained in Section 2.3) with hyperparameters α and β as below:

$$P(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)} \quad | \; B(\alpha,\beta) = \text{beta, function}, \tag{2.15}$$

$$\frac{\partial}{\partial\theta} log(P(\theta)) = \frac{\alpha-1}{\theta} + \frac{\beta-1}{1-\theta}, \tag{2.16}$$

$\frac{\partial}{\partial\theta}L$ is same as Eq. 2.11. Substituting Eq. 2.11 and 2.16 in 2.14 and simplifying, we obtain,

$$\theta_{MAP} = \frac{n^1 + \alpha - 1}{N + \alpha - 1 + \beta - 1}. \tag{2.17}$$

From the MAP estimate in Eq. 2.17, we can see that, the addition of prior distribution is just including past experimental results or belief. The addition of prior belief acts like regularization to the MLE estimate

2.2 Expectation Maximization

The Expectation Maximization (EM) algorithm [8] is an iterative approach for parameter estimation for an incomplete dataset. The missing values in the data corresponds to unobserved variables which are also known as hidden or latent variables. Each iteration of the EM algorithm consists of two steps: the E-step and the M-step. In the E-step, the missing data is estimated with the current estimate of the parameters. In the M-step, the parameters are estimated using the MLE. Consider a set X of observed data, a set Z of unobserved data and let θ be a vector of unknown model parameters. The log-liklihood of the observed data is given below:

$$\begin{aligned} L(\theta|X) &= \sum_i log(\sum_{z^i} P(x^i, z^i|\theta)) \\ &= \sum_i log(\sum_{z^i} \frac{P(x^i,z^i|\theta)Q_i(z^i)}{Q_i(z^i)}) \\ &\geq \sum_i \sum_{z^i} Q_i(z^i) log \frac{P(x^i,z^i|\theta)}{Q_i(z^i)}. \end{aligned} \tag{2.18}$$

The last step in the above equations is obtain by applying Jensen's Inequality to the concave log function. Q^i could be any set of distributions. Since we know that the distribution should sum to,

$$\sum_z Q_i(z) = 1. \tag{2.19}$$

We can choose the Q_i to be $P(z^i|x^i, \theta)$, the posterior distribution given the data and the parameters. To calculate this posterior distribution we should have some initial estimate for the parameters θ (could be random). Using this estimated posterior distribution the parameters can be estimated by maximizing Equ. 2.18 w.r.t θ. Hence we have the following steps of the EM algorithm:

E-Step:

$$P(z^i|x^i,\theta^n) = \frac{P(x^i,z^i|\theta^n)}{P(x^i,\theta^n)} \tag{2.20}$$

M-Step:

$$\theta^{n+1} = \arg\max_{\theta}(\sum_i \sum_{z^i} P(z^i|x^i,\theta^n))log\frac{P(x^i,z^i|\theta)}{P(z^i|x^i,\theta^n)}) \tag{2.21}$$

The above two steps are iterated until convergence. The EM algorithm is guaranteed to increase the liklihood in each iteration [8]

2.3 Conjugate Priors

In Section 2.1.1 it was shown that the MLE estimate leads to overfitting the parameters if the sample dataset is not a good representation of the population. The MAP estimate overcomes this problem by introducing a prior distribution $P(\theta)$ over the parameter θ. The prior distribution is chosen such that it has a simple interpretation and useful analytical properties. The posterior distribution is proportional to the product of liklihood and prior. If the posterior and prior have the same functional form then they are known as conjugate pairs.

2.3.1 Beta Distribution

The beta distribution is a continuous distribution in the interval $[0,1]$ parameterized by two positive shape parameters α, β.

$$Beta(\theta|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1} \tag{2.22}$$

The parameters α and β are also called hyperparameters because they control the distribution of the θ parameter. Γ is gamma function ($\Gamma(n) = (n-1)!$). The variance of the beta distribution is governed by the value of the hyperparameters as illustrated in Fig. 2.1. The beta distribution has similar form as the binomial distribution. The likelihood function in case of a dataset sampled from a binary random variable takes the form $\theta^l(1-\theta)^m$ (l

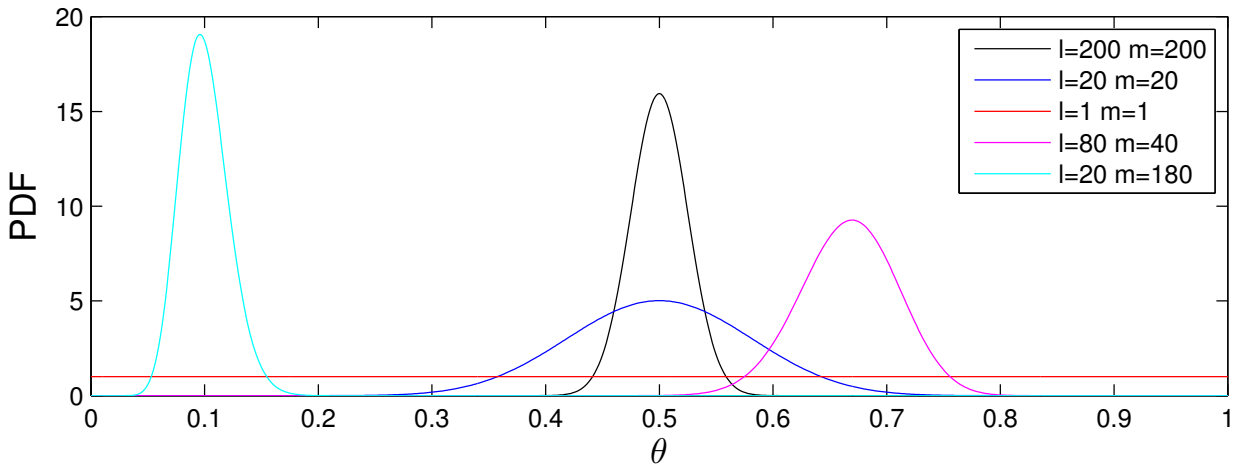


Figure 2.1: Beta Distribution: Depicts the variance (confidence of θ) of beta distribution for various hyperparameters

successes, m failures). By using the beta distribution as prior, the posterior can then be calculated using the Baye's rule in 2.1.

$$posterior(\theta|l,m,\alpha,\beta) = \frac{\Gamma(\alpha+\beta+l+m)}{\Gamma(\alpha+l)\Gamma(\beta+m)}\theta^{\alpha+l-1}(1-\theta)^{\beta+m-1} = Beta(\theta|\alpha+l,\beta+m) \quad (2.23)$$

As seen from Eq. 2.23, the posterior distribution takes the form of the beta distribution. This same form is the interesting property of Beta Distribution or in general conjugate distributions which provides convenient mathematical form of including prior knowledge for parameter estimation.

2.3.2 Dirichlet Distribution

In case of a Multinomial likelihood function, the beta distribution can be generalized from 2 to K dimensions.

$$Dir(\vec{\theta}|\vec{\alpha}) = \frac{\Gamma(\sum_{k=1}^{K}\alpha_k)}{\prod_{k=1}^{K}\Gamma(\alpha_k)}\prod_{k=1}^{K}\theta_k \quad (2.24)$$

Similar to the Beta distribution, the posterior distribution in case of Multinomial likelihood function can be shown to take the form of a Dirichlet distribution.

2.4 Topic Modelling

Topic models [4, 24] are a suite of algorithms that aim at expressing documents as a mixture of topics. A topic is a distribution over words. Topic models are generative models: a document is generated by first choosing a distribution over topics, then randomly choosing a topic from this distribution and drawing a word from that topic. Using standard inference techniques the parameters for the distributions can be inferred. Here we will be discussing Probabilistic Latent Semantic Analysis (PLSA) and Probabilistic Latent Sequential Motifs (PLSM). PLSA has better statistical significance than LSA. The interpretation of topics is also straight forward. They are also good at capturing polysemy. PLSA does not make any assumptions about the order of the words in the document. This is known as the bag of words approach to modeling text. On the other hand, PLSM tries to capture the order of occurrence of the words.

2.4.1 PLSA

Latent semantic analysis (LSA) [7] represents documents in a lower dimensional space called the latent semantic space by linear projection of term-document matrix through singular value decomposition. The term-document matrix represents the frequency of words from a well defined vocabulary in a document. LSA has deficits, like capturing polysems, and interpretation is difficult due to its unsatisfactory statistical foundation. Probabilistic Latent Semantic Analysis (PLSA) [13] introduces a statistical foundation to LSA, since it is based on a likelihood principle and defines a generative model of documents pertaining to a dataset [13].

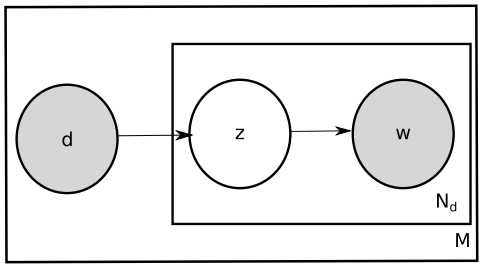


Figure 2.2: PLSA generative model: d is the document variable, z is the topic variable dependent on d and w is the word variable independent of d given z. d and w are observed variables. N_d is the document length and M is the number of documents

PLSA adopts the aspect model to model the joint probability of each co-occurrence of a word $w \epsilon W = \{w_1, w_2, \cdots, w_V\}$($V$ is the size of the vocabulary) in a document $d \epsilon D =$

$\{d_1, d_2, \cdots, d_M\}$ by associating a latent class variable z. The generative process of the model in Fig. 2.2 can be explained as below :

- pick a document d with probability $P(d)$.
- pick a topic with probability $P(z|d)$
- pick a word from topic z with probability $P(w|z)$

The joint probability model of PLSA is given by the expression:

$$P(w,d) = P(d)P(w|d), \tag{2.25}$$

$$P(w|d) = \sum_z P(w|z)P(z|d). \tag{2.26}$$

PLSA is a mixture model. This is based upon the conditional independence assumption that given a topic, the choice of word is independent of the document. The parameters $P(z|d)$, $P(w|z)$ and $P(d)$ can be estimated by maximizing the log-likelihood function using the EM algorithm.

$$L = \sum_d \sum_w n(w,d) log(P(w,d)). \tag{2.27}$$

The E-step can be obtained by using Baye's rule:

$$P(z|d,w) = \frac{P(z|d)P(w|z)}{\sum_z P(z|d)P(w|z)}. \tag{2.28}$$

By standard calculations, the equations for M-step can be obtained:

$$P(z|d) = \frac{\sum_w n(w,d)P(z|w,d)}{\sum_w \sum_{z'} n(w,d)P(z'|w,d)}, \tag{2.29}$$

$$P(w|z) = \frac{\sum_d n(w,d)P(z|w,d)}{\sum_w \sum_d n(w',d)P(z|w',d)}, \tag{2.30}$$

$$P(d) = \frac{\sum_z \sum_w n(w,d)P(z|w,d)}{\sum_{d'} \sum_w \sum_z n(w,d')P(z|w,d')}. \tag{2.31}$$

2.4.2 PLSM

Topic models like PLSA have been shown to be successful in capturing scene level activity patterns by co-occurrence analysis of low level features (words). These models fail to

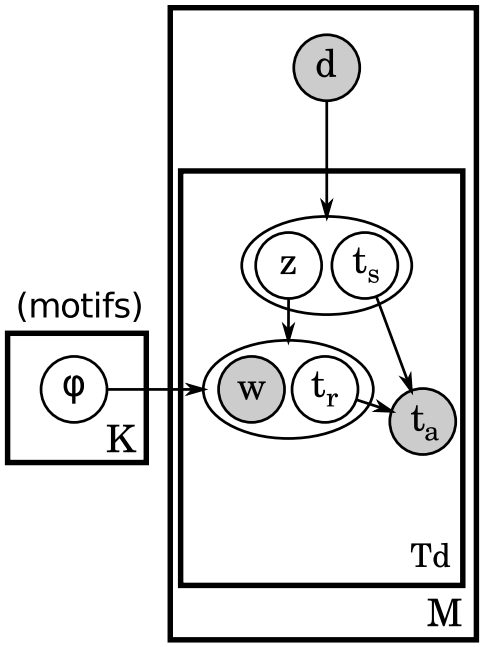


Figure 2.3: Bayesian network representing PLSM generative model: d is the document variable, z is the topic variable dependent on d and w is the word variable independent of d given z. d and w are observed variables. Td is the document length and M is the number of documents

capture the temporal order of co-occurrence of words. These topic models have been adapted (like HMM on top of LDA) [14] to discover temporal patterns but have drawbacks [26, 28] like manual segmentation of videos to synchronize with the start of the cycle for effective topic discovery, discovering multiple patterns occuring at the same time. Probabilistic latent sequential motifs (PLSM) [26, 28] addresses some of these drawbacks by capturing spatio-temporal co-occurrences of words in a temporal window called motifs and the starting time of occurrence of these motifs. Fig. 2.3 shows a graphical model [19] representation of the generative process of PLSM. Let $D = \{d_1, d_2, d_3, \cdots, d_M\}$ represent the set of documents (like videos from a surveillance camera), $W = \{w_1, w_2, \cdots, w_V\}$ represents the vocabulary, $Ta = \{t_1, t_2, \cdots, t_{Td}\}$ represents the time of occurrence of words in the document d. Then according to our model, the documents can be described by a set of motifs $Z = \{z_1, z_2, \cdots, z_K\}$ which have a duration of Tz i.e $tr = \{t_1, t_2, \cdots, t_{Tz}\}$. Each motif can occur at any time $ts = \{t_1, t_2, \cdots, t_{Tds}\}$ in a document d. The generative process of the PLSM model can be described as follows:

· Pick a document d from $P(d)$

· Pick topic z and its starting time ts from $P(z, ts|d)$

· Pick a word w and relative time tr from $P(w, tr|z)$

· Set $ta = ts + tr$ or equivalently $P(ta|ts, tr)$ is a Dirac function at ta

The joint probability distribution $P(w, ta, d, z, ts, tr)$ can be obtained from the model as below:

$$P(w, ta, d, z, ts, tr) = P(d)P(z, ts|d)P(w, tr|z)P(ta|tr, ts), \tag{2.32}$$

$$= \begin{cases} P(w, z, ts, tr, d), & \text{if } ta = ts + tr \\ 0, & \text{otherwise} \end{cases} \tag{2.33}$$

Given a corpus of documents C in the form of a term frequency matrix $n(w, ta, d)$, the likelihood of the data is given by the expression:

$$P(C) = \prod_{w,ta,d} P(w, ta, d)^{n(w,ta,d)} \tag{2.34}$$

The motifs $P(w, tr|z)$ and their start times $P(z, ts|d)$ which form the parameters of the model can be inferred from the observations. The inference is performed by maximizing the log-likelihood of the data. The inference is also guided towards estimating a sparse distribution of $P(z, ts|d)$ which is motivated in [26, 28]. The sparsity constraint is incorporated on $P(z, ts|d)$ using KullbackLeibler (KL) divergence with a uniform distribution (U).

$$L(D|\theta) = \sum_{w,ts,tr,d} n(w, ts + tr, d) log(\sum_{z,ts} P(w, tr, d, z, ts)) + KL(U||P(z, ts|d)) \tag{2.35}$$

Since the data is partially observed, parameters are estimated using the expectation maximization algorithm.

$$E[L] = \sum_{w,ts,tr,z,d} n(w, ts + tr, d)P(z, ts|w, ts + tr, d) log(P(w, ts + tr, d, z, ts)) - \sum_{z,ts,d} \frac{\lambda_d}{K \cdot Tds} log(P(z, ts|d)) \tag{2.36}$$

The E-Step can be obtained using Baye's rule

$$P(z, ts|w, ta, d) = \frac{P(z, ts|d)P(w, tr|z)}{\sum_{z,ts} P(z, ts|d)P(w, tr|z)} \tag{2.37}$$

The M-step expressions can be obtained by standard calculations

$$P(z, ts|d) \propto \max(\epsilon, \sum_{w,tr} n(w, ts + tr, d)P(z, ts|w, ts + tr, d) - \frac{\lambda_d}{K \cdot Tds}) \tag{2.38}$$

$$P(w, tr|z) \propto \sum_{ts,d} n(w, ts + tr, d)P(z, ts|w, ts + tr, d) \tag{2.39}$$

The term λ_d as in [26, 28] is defined as $\lambda\, n_d$ where n_d is the number of words in the document d and λ indicates the sparsity level and can take any positive real number as its value. ϵ is a very small probability used to add additional sparsity constraint. The probability of terms which are lower than $\frac{\lambda_d}{K \cdot Tds}$ are set to this probability.

Chapter 3

Dataset

The dataset used in this thesis consists of videos from traffic cameras and metro stations. The traffic cameras usually exhibit structured patterns and a set of valid and invalid traffic flow patterns. The set of valid and invalid patterns for each traffic camera is described here. The metro station on the other hand contains loosely constrained movements and hence is more noisy, hence only the scene for each camera in the metro network is discussed.

3.1 Traffic Cameras

The traffic dataset consists of three cameras capturing unrelated scenes, meaning that the cameras are unrelated. Below we will discuss the scene captured by these cameras, their sources and the motivation for choosing them.

The Roundabout contains 60 minutes of video at a resolution of 360 × 288 at 25fps. The traffic movements in the roundabout signal are restricted to only certain driving directions as illustrated in Fig. 3.1. This video can be downloaded from `http://www.eecs.qmul.ac.uk/~jianli/Roundabout.html`. This dataset is interesting and used in this work because it exhibits well-defined and constrained traffic flow patterns, thus forming a good validation set for integrated PLSM and anomaly detection.

QMUL Junction contains a 60 minute video at a resolution of 360 × 288 at 25fps. In this signal, a traffic cycle consists of 4 different traffic flows as depicted in Fig. 3.2. This video can be downloaded from `http://www.eecs.qmul.ac.uk/~ccloy/downloads_qmul_junction.html`. This dataset is interesting due to its busy traffic cycles.

Figure 3.1: QMUL Roundabout: The arrows depict driving flow directions which are allowed.

A traffic cycle

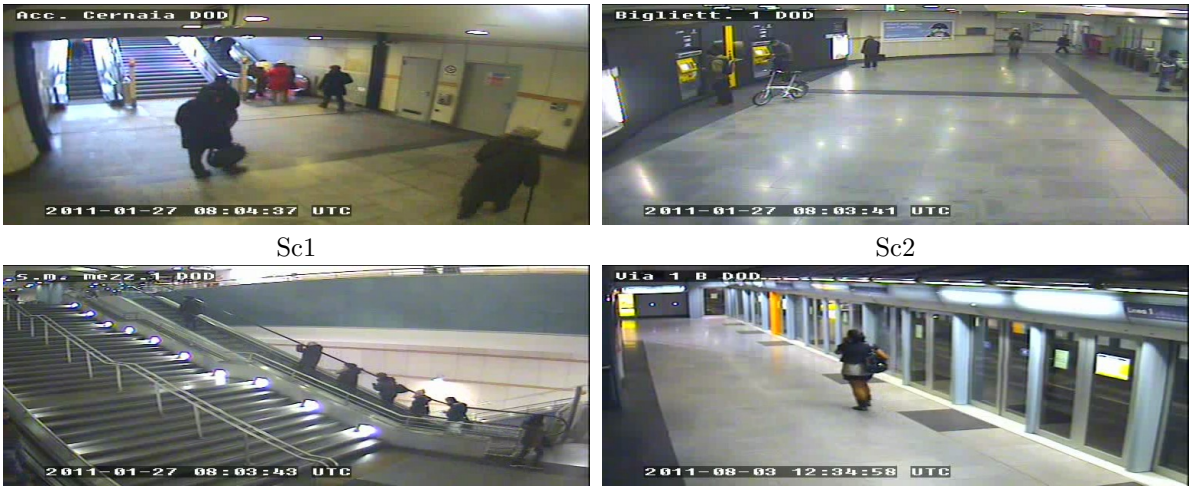
Figure 3.2: QMUL Junction: The arrows depict driving flow directions which are allowed.

The Traffic Junction dataset introduced in [29] consists of a video which is controlled by traffic lights. The length of the video is 44.12 minutes at 25fps at resolution of 360 × 288. The activities present in the video are: cars stopping at the red light, pedestrians waiting for crossing. This dataset in particular contains a lot of pedestrian activity that is more loosely constrained than the cars. A scene from the camera is depicted in Fig. 3.3

3.2 Metro Cameras

The metro camera dataset consists of cameras monitoring a metro network. There are in total 18 cameras in this dataset, only 4 of which have been used in this thesis. These cameras were chosen based on their spatial connectivity and motion constraints. (A scene depicting an escalator will have well-defined pattern and very low probability for anomaly over other scenes). A brief description of the scene shown in Fig. 3.4 captured by these 4 cameras will be discussed here. This dataset is a property of IDIAP Research Institute and not yet available for public downloads at the time of this work.

Figure 3.3: Traffic Junction: A scene from the traffic camera.



Sc1 Sc2

Sc3 Sc4

Figure 3.4: Metro Camera Dataset .

In Scene 1 (Sc 1), the camera captures the entrance/exit to the subway network. The scene contains an elevator, stairs and two escalators.

In Scene 2 (Sc 2), the camera overlaps with Sc 1. It contains a hallway and turnstiles connecting the metro network.

In Scene 3 (Sc 3), the camera is in the neighbourhood of camera 2. It contains an escalator moving up towards Sc 2 and a walk way leading to Sc 4.

In Scene 4 (Sc 4), the camera shows a platform where passengers get on and off the train.

Chapter 4

PLSM on top of PLSA

In this chapter, the feature extraction process is described in brief and then we describe how PLSA is used as a dimensionality reduction step to pre-process the input to PLSM. As our model consists of multiple layers, we will systematically use the superscript ll to denote the lowest layer

4.1 Feature Extraction

For each video we extract optical flow features (motion features) using Lucas-Kanade algorithm [25] from a dense image grid. We keep only pixels where some motion is detected and we quantize the motion into 8 directions and the 9^{th} direction indicates slow moving pixels. We obtain low level words w^{ll} defined by a position in the image and a direction of motion. We apply a sliding window of 1 second, without overlap to obtain a histogram $n^{ll}(w^{ll}, d^{ll}_{ta})$. Here d^{ll}_{ta} represents the low level document obtained by the sliding window at time ta.

4.2 PLSA model

PLSA (Fig. 2.2) as explained in Chapter 2 is a minimal topic model. Given a set of documents made of word counts and summarized in a count matrix $n^{ll}(w^{ll}, d^{ll})$, it extracts "topics" capturing sets of words that often co-occur in the documents. Each topic is actually a distribution over words $\phi^{ll}_{z^{ll}} = p(w^{ll}|z^{ll})$. Topics in our context represent spatially co-occurring set of pixels in an image/frame or a group of frames in a video. Fig. 4.1 illustrates some of the topics obtained from $Sc2$ of the metro camera. Each document d^{ll} is also

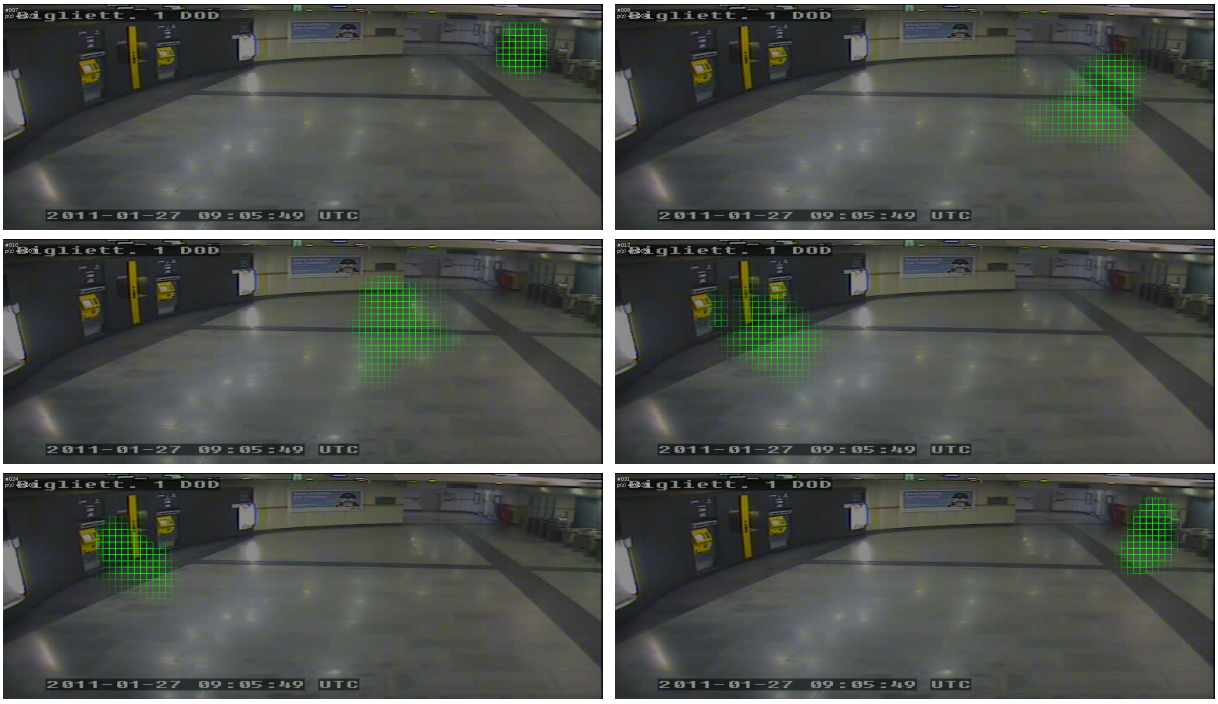


Figure 4.1: PLSA Topics: Six topics of $Sc2$ from metro station camera is shown. Each topic represents a person spatially located at different regions of the scene.

decomposed as a mixture $\theta^{ll}_{d^{ll}} = p(z^{ll}|d^{ll})$ of these automatically learned topics. PLSA usually has a non-informative prior α^{ll} on the θ^{ll} weight vectors. This means that the model is not encouraged to give any special shape to θ^{ll} distributions and thus it can arbitrarily choose the best ones that explains the data. We will exploit this prior as a mechanism for feeding higher level information to PLSA.

4.3 PLSM Model

PLSM (Fig. 2.3) as explained in Chapter 2 adds time to PLSA: it is a topic model which automatically finds temporal and spatial co-occurrences of words. More precisely, it takes as input a count matrix $n(w, t_a, d)$ indicating for each document d (video clips), the number of times the word w occurs at time t_a. By describing the documents as mixtures of temporal motifs, PLSM learns two sets of distributions, similarly to PLSA but adding time: a set of motifs z, each represented by a distribution $\phi_z = p(w, t_r|z)$ denoting the probability that a word w occurs at a relative time t_r since the start of the motif. In our application, a

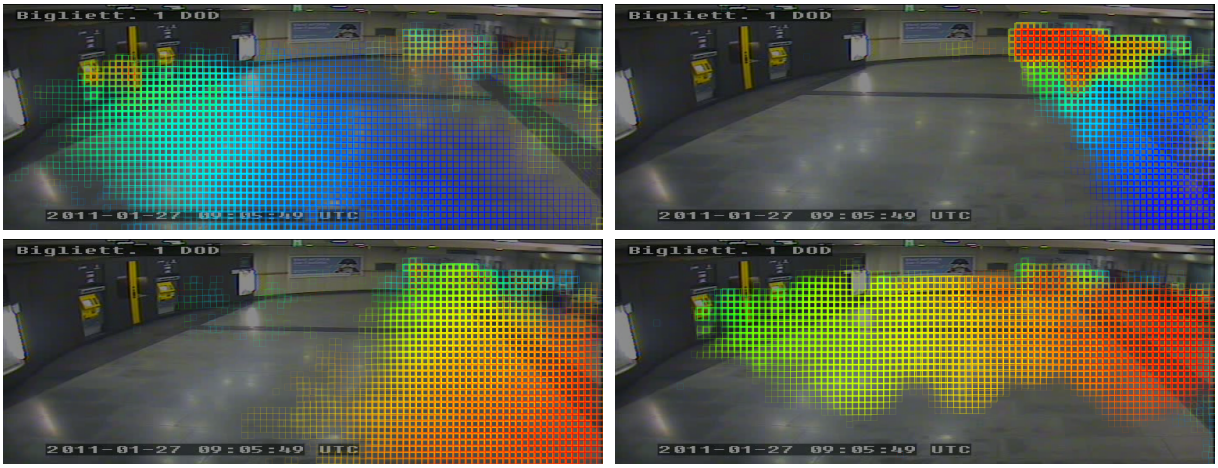


Figure 4.2: PLSM Motifs: Four Motifs of $Sc2$ from metro station camera is shown. Each Motif represents a temporal pattern with time progression indicated by color: blue(start), green(middle) and red(end).

motif captures spatially and temporally co-occurring words in the document as illustrated in Fig. 4.2 and the distributions $p(z, t_s|d)$ which indicate when the motifs occur, i.e., the probability that a motif z starts at time t_s.

4.4 PLSM on Top of PLSA

One modeling issue with PLSM is how to define the count matrix $n(w, t_a, d)$. A first possibility would consist in ordering the low level documents $d^{ll}_{t_a}$ of the video clip d according to time t_a to obtain the temporal document $n^{ll}(w^{ll}, t_a, d)$. However, as the number of low-level features is quite high, the learning of PLSM can be time-consuming. To overcome this issue, the PLSA topic model can be applied as a dimensionality reduction pre-processing step using as input the un-ordered low-level documents. Fig. 4.3 illustrates this approach with blue arrows. PLSA results in a set of topics z^{ll} which captures the frequently co-occurring words in the video which often correspond to local spatial clusters of words, and the distribution of topics $p(z^{ll}|d^{ll}_{ta})$ within each document. By assimilating these low-level topics z^{ll} as the words w of PLSM, we can build the temporal document for PLSM as:

$$n(w = z^{ll}, ta = d^{ll}, d) = p(z^{ll}|d^{ll}_{ta}) \sum_{w^{ll}} n^{ll}(w^{ll}, d^{ll}_{ta}) \tag{4.1}$$

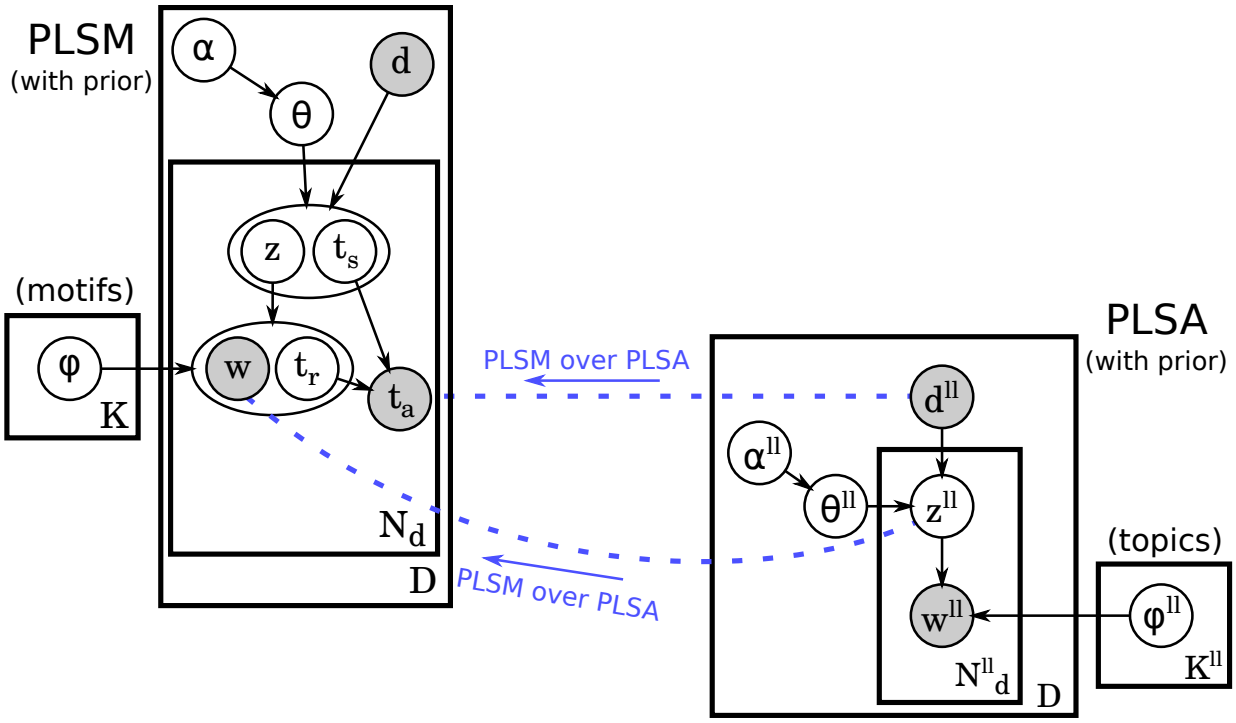


Figure 4.3: PLSM over PLSA, from [26] and illustration of PLSM applied on PLSA topics without feedback).

where the topic distributions $p(z^{ll}|d^{ll}_{ta})$ are weighted by the mass of the document (number of low-level words at time ta) to account for the overall amount of activity at each time step. This temporal document is then fed as input to PLSM to learn the temporal motifs and their starting times. This method is illustrated in Fig. 4.4. The PLSA topics $1-6$ are arranged in time to form the PLSM motif. The PLSA topics thus captures spatial co-occurrences and PLSM captures the temporal ordering of these spatial co-occurrences, hence the quality of motifs recovered by PLSM are only as good as the topics discovered by PLSA.

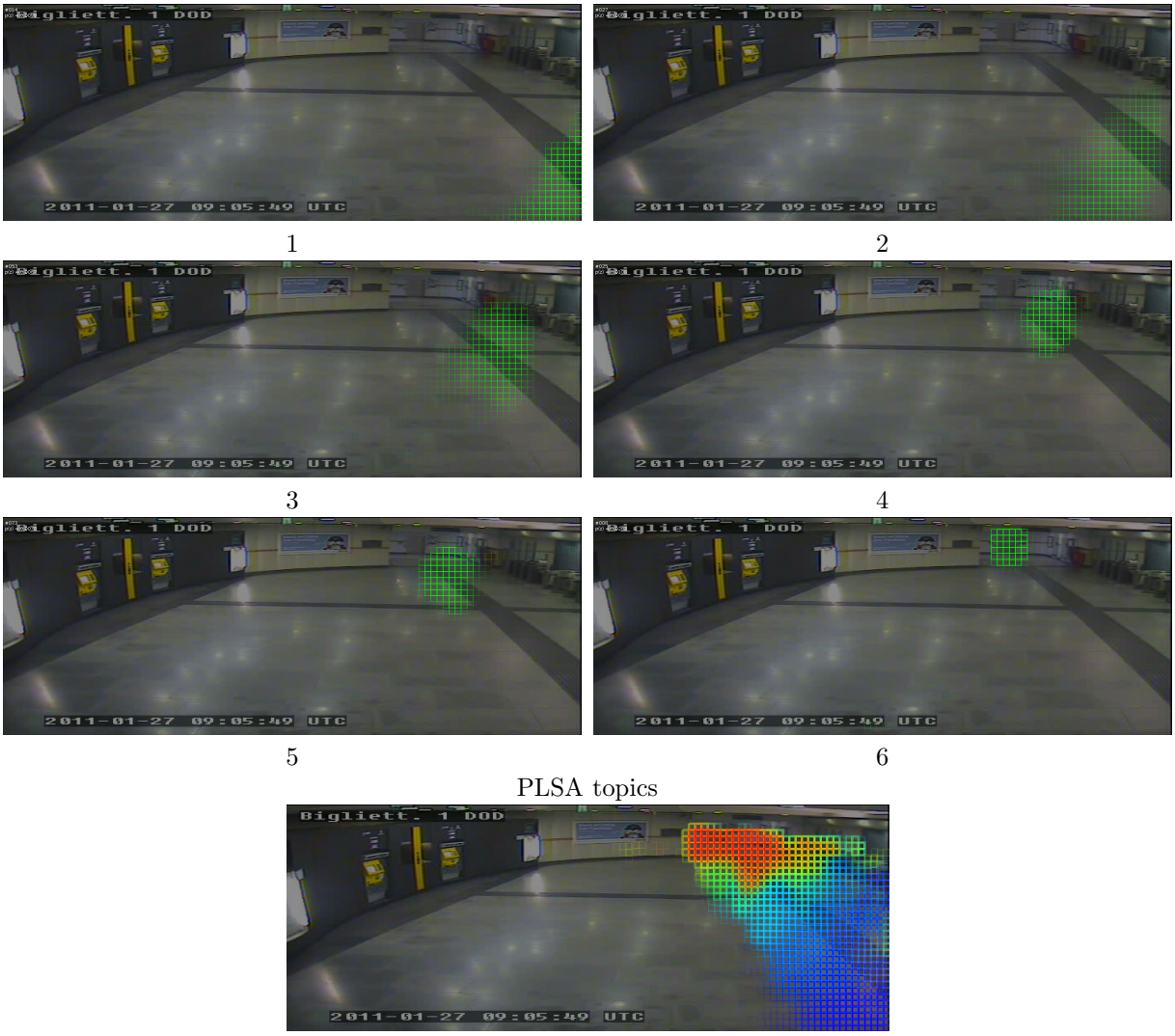


Figure 4.4: PLSM on top of PLSA: PLSA topics and PLSM motif obtained by temporal ordering of topics.

Chapter 5

Integrated PLSM and Hierarchial PLSM

In this chapter, we will formulate the integration of PLSA dimensionality reduction step with PLSM. We call this approach Integrated PLSM (IPLSM), which overcomes the disadvantages of using PLSM on top of PLSA. In the following sections, we introduce this model and give an idea of the inference process. We then show how the principle can be generalized to a hierarchical model consisting of three levels of topic modeling, capturing temporal patterns across cameras. As the models involve multiple layers, we will systematically use the superscript ll to denote lowest level elements, cl to denote combined elements and i to denote the camera index. The camera index denotes a specific layer in a stack of IPLSM. The hierarchical model consists of three levels of topic modeling.

5.0.1 Integrated PLSM

As explained in Chapter 4, from the distributions ϕ and θ, PLSM is fully able to reconstruct an updated version of its input that takes into account temporal co-occurrence captured in the motifs. The reconstruction of the input is done following the PLSM equations (with $t_r = t_a - t_s$):

$$\tilde{n}^{rec}(w, t_a, d) = \sum_z \sum_{t_s} p(w, t_r|z)p(z, t_s|d) \tag{5.1}$$

By applying PLSM on top of PLSA as presented above, PLSM captures temporal patterns of occurrences of PLSA topics. Given that PLSM tries to explain the documents with temporal motifs, the reconstructed input $\tilde{n}^{rec}$ (normalized word count) corresponds to the original input but updated to exhibit better temporal coherence. This motif-constrained

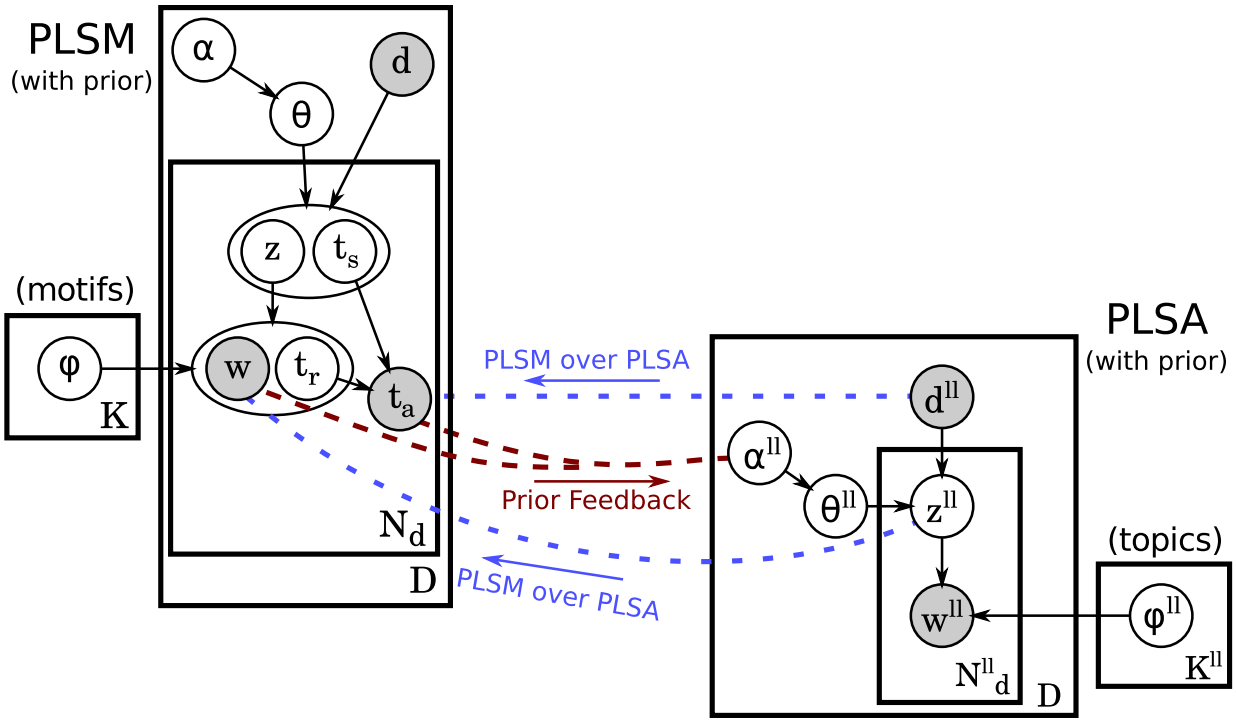


Figure 5.1: IPLSM, from [26] and illustration of the prior feedback (in red). The otherwise uninformative prior on the topic weights in each document $\theta^{ll}_{d^{ll}}$ is replaced by a prior coming from the time-aware higher-level PLSM model.

reconstruction motivated [10] to use the difference between the input and n^{rec} (unnormalized) as a measure of temporal anomaly. We discuss more on anomalies in the upcoming chapters. In case of no anomaly, all the topics captured by PLSA have temporal significance, otherwise the topics which are not well reconstructed are temporally less significant. To increase the quality of the PLSA topics, we propose an integrated model [6] shown in Fig. 5.1. The goal of IPLSM [6] is to have the temporal structure of the data (captured by PLSM) impact on the image-level topics (PLSA level). We do it by jointly learning the PLSA and PLSM models. When a scene is crowded, it is often difficult for PLSA to capture clean topics. When getting temporal information from PLSM, PLSA is able to capture cleaner topics. In practice, we exploit two facts: PLSM can reconstruct its input with some added temporal constraints, and a modified PLSA can accept a prior probability of what topics occur in which document. Using IPLSM, the temporal information allows to disambiguate the instantaneous information and the captured low level topics contain fewer artifacts. In

practice, the inference of the IPLSM model is done in an iterative manner, iterating multiple Expectation Maximizations.

5.1 Inference by Prior-Feedback

The overall goal of IPLSM is to use higher-level temporal structure (motifs) to improve lower level structures (topics). As exact inference is intractable (even just for PLSA), we use an iterative approximate inference method and formulate the feedback of higher levels as a prior for lower levels. The document reconstructed by PLSM, n^{rec} will be used after re-weighting as a prior α^{ll} for PLSA. This process goes on iteratively.

More precisely, the prior α^{ll} for the document at time t_a (α^{ll} is used as an alias for $\alpha^{ll}_{d^{ll}_{t_a}}$)is computed as follows:

$$\alpha^{ll}(z^{ll}) = n^{rec}(z^{ll}, t_a, d) \times S + Uniform \tag{5.2}$$

For a given low level document, α^{ll} represents the hyper parameters of a Dirichlet prior over θ^{ll}. Adding the uniform distributions encodes the inherent uncertainty of the PLSM feedback. This helps in Expectation Maximization algorithm to obtain a better local maximum during the learning phase for PLSA and PLSM. S is a constant and is a parameter to the model: the bigger the S, the more the information coming from PLSM will be trusted. An S value greater than 1 indicates a very strong prior and would not allow the motifs to change much as it requires more observations than the training documents to change the belief. So a prior less than one is more meaningful in this context.

As previously mentioned, this is an iterative approach, the necessary number of iterations is also a parameter to the model. The number of iterations is camera specific, a metro camera may require more iterations than a traffic camera because of loosely constrained motion. An example of the evolution of topics on a single traffic camera is shown in Fig. 5.2. In this example, the motif 3 in iteration 0 captures two patterns which depicts traffic flow. According to this motif the occurrence of one pattern leads to the occurence of the other

pattern. In actuality, these two patterns are independent of each other. In iteration 4, the motif has evolved to capture this fact and captures one of the two patterns with higher probability. The lower probability pattern in motif 3 is captured in motif 5. Motif 2 in iterations 4 has evolved to capture a completely new traffic flow compared to iteration 0 which was very similar to motif 3. Some motifs like motif 1 in iteration 4 and motif 4 in unchanged motifs are quite similar. This may be due to the fact that the motif 4 more frequently than not exhibits such a traffic flow in the video.

5.1.1 Multi-Camera with Hierarchical IPLSM

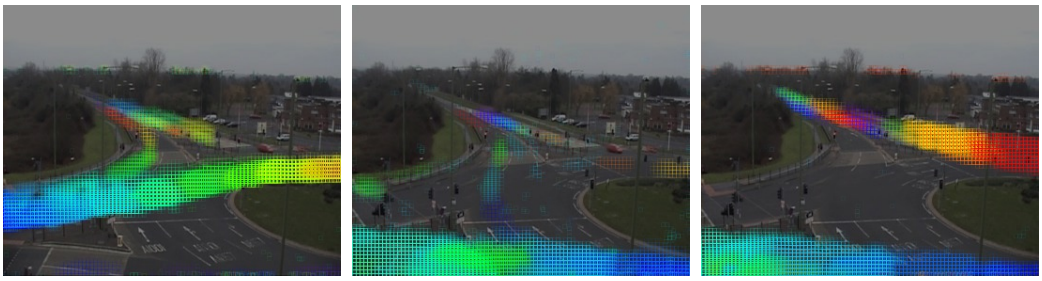
Following the idea of integrating PLSA with PLSM, layers of PLSM can be stacked and integrated. The prior feedback can be used on the α parameter of PLSM. This approach becomes especially interesting when considering multiple cameras as illustrated in Fig. 5.3.

The idea is to have an IPLSM for each camera and a higher-level PLSM working on their combined outputs and capturing motifs of per-camera motifs. The motivation for feedback from hierarchical layer to single camera IPLSM layer is to capture motifs that are relevant (in time) across cameras as depicted in Fig 5.3. The process is explained in more detail below.

For modeling recurrent activities across multiple related cameras, IPLSM can be used for each camera to learn a distribution $P^i(z, ts|d)$, i denoting the camera index. We generate temporal documents by multiplying this distribution by the document mass. This document $n^{i'}(w, ta, d)$ represents a dimensionality reduced version of the input document to PLSM in the IPLSM (1st level PLSM). The equation is shown below

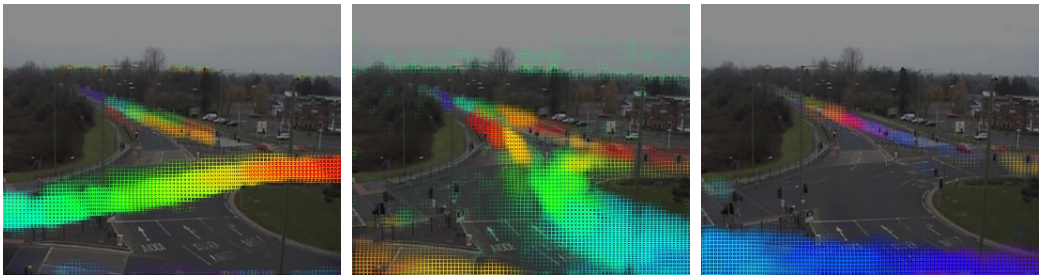
$$n^{i'}(w, ta, d) = \sum_{w} \sum_{ta} n^i(w, ta, d) P^i(z, ts|d) \tag{5.3}$$

The limits on the summation depends on the number of topics chosen and the document length for camera i. The temporal documents obtained from the above step are combined to form $n^{cl}(w^{cl}, ta^{cl}, d^{cl})$. The length of the combined document and the number of words



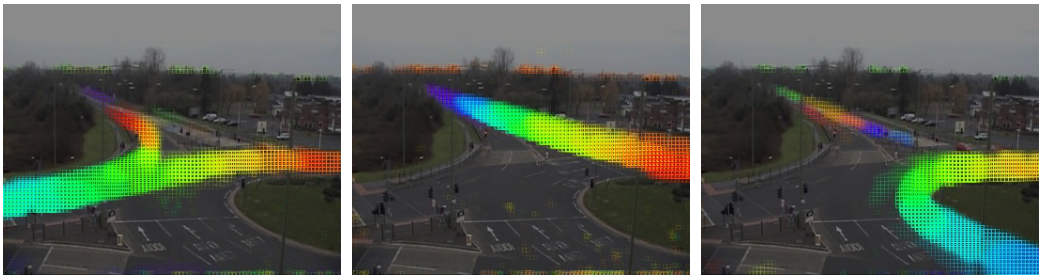
motif 1 motif 2 motif 3

iteration 0 - no feedback



motif 1 motif 2 motif 3

iteration 4 - iterative feedback



motif 4 motif 5 motif 6

unchanged motifs

Figure 5.2: QMUL Roundabout dataset. Example of evolution of motifs during iterative IPLSM learning. Parameters: S=0.75, iteration=5, PLSA topics=80, PLSM motifs=10 with motif length=12. The color gradient represents time from blue (start) to red (12s). Motifs 1-3 evolve during the iterative process from having similar temporal patterns with respect to other motifs in iteration 0 to more distinct patterns in iteration 4. Motifs 4-6 don't evolve as they all represent distinct patterns. Motifs 7-10 are not shown.

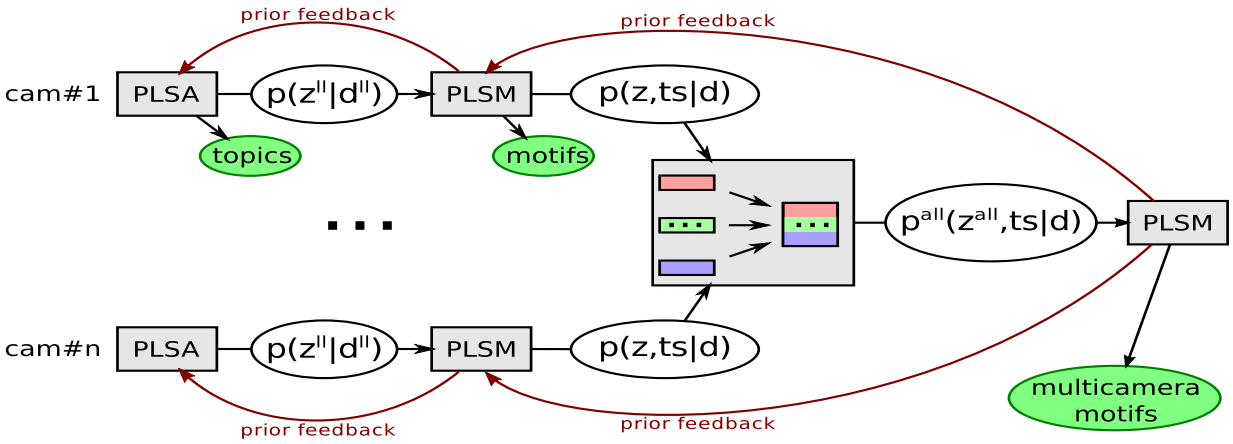


Figure 5.3: Hierarchical Integrated PLSM Model. Cam#i refers to a camera view i. The process captures information at three levels: topics (image level), motifs (per-camera temporal patterns) and multi-camera motifs. Each level iteratively feeds back information as a prior to the previous level.

are given below

$$W^{cl} = \sum_i K^i \tag{5.4}$$

$$Ta^{cl} = \max_i Ta^i \tag{5.5}$$

K^i and Ta^i denotes the number of motifs and document length for camera i. Applying PLSM on the combined temporal document we obtain the distribution $P^{cl}(w^{cl}, tr|z^{cl})$ which represents the frequently co-occurring sequential patterns across cameras. We also obtain a distribution $P^{cl}(z^{cl}, ts|d^{cl})$. The prior for individual IPLSM from the third level PLSM is similar to the prior construction explained previously. Fig. 5.4 illustrates motif evolution for a combination of three cameras from the metro dataset. $Sc1$ and $Sc2$ are overlapping cameras and hence contain temporally related information. $Sc4$ is spatially well separated from the other two cameras and hence contain information that may not be relevant to the regions covered by the previous two cameras. Iteration-0 illustrates two motifs which captures patterns across the three cameras. Any pattern occurring in the third camera with respect to the other two cameras is a coincidence and doesn't correspond to any temporal continuity between cameras. Iteration-4 captures this fact and captures motifs only in the

motif 1 motif 2

iteration 0 - no feedback

motif 1 motif 2

iteration 4 - with feedback

Figure 5.4: QMUL Roundabout dataset. Example of evolution of motifs during iterative Hierarchical-IPLSM learning. Parameters: S=0.75, iteration=5, IPLSM = 80 topics: motifs=20 with motif length=10, combined PLSM= motifs=20 with motif length=15 The color gradient represents time from blue (start) to red (10s).

third camera. The activities corresponding to the overlapping regions are captured in the other motifs.

Chapter 6

Abnormality Detection

As described in the previous chapter, PLSM fits its input with some motifs and it can be used to reconstruct a corrected version of the input. Intuitively, the reconstructed input n^{rec} is the same as the original input when the motifs explains perfectly the input.

Following the above intuition [10], the difference between the input and n^{rec} can be used as an anomaly index. One limitation of this approach is that it ignores the semantics of the words used as input of PLSM. In practice, these words correspond to PLSA topics and thus to patches of localized motion in the image. This chapter explains the lower-level document reconstruction followed by description of the abnormality measure and finally, explains the intuition behind abnormality detection at different levels of the model. Fig 6.1 depicts the hierarchical model described in the previous chapter along with abnormality detectors at different levels.

We will use the notations in Table 6.1 throughout this chapter

Table 6.1: explains notations

P^{ll}_{wz} P^{ll}_{zd}	$P(w^{ll}\|z^{ll})$ $P(z^{ll}\|d^{ll})$	level-1
P_{wtz} P_{ztd}	$P(w,ta\|z)$ $P(z,ta\|d)$	level-2
P^{cl}_{wtz} P^{cl}_{ztd}	$P(w^{cl},ta\|z^{cl})$ $P(z^{cl},ta\|d^{cl})$	level-3
$mat(A,B)$	Matrix multiplication of A and B	

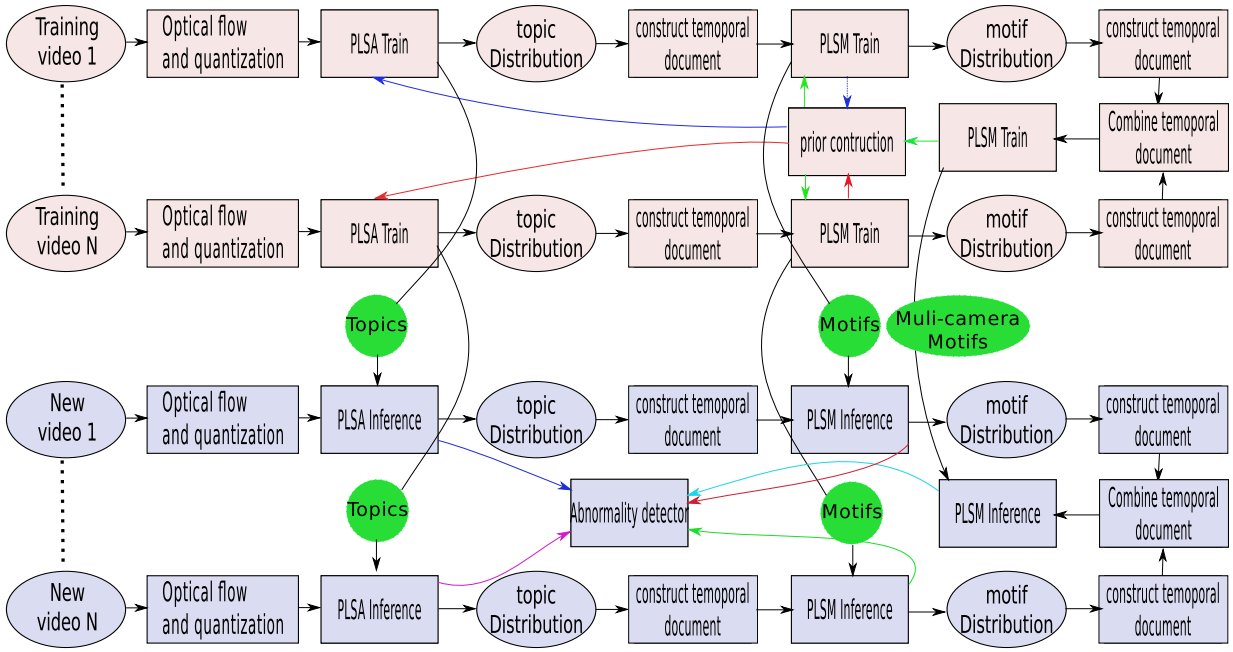


Figure 6.1: Hierarchical Integrated PLSM Model with abnormality detectors. Video#i refers to a camera view i. The process captures information at three levels: topics (image level), motifs (per-camera temporal patterns) and multi-camera motifs. Each level iteratively feeds back information as a prior to the previous level. Using the captured information at the three levels different forms abnormality detection can be performed.

6.1 Low-Level Document Reconstruction

We explain the lower-level document reconstruction from the pattern learned at each level in the following sections. The document reconstruction is the first step towards detecting abnormalities at each level. The reconstruction process (Inference) tries to explain the documents based on the captured information during learning (topics for PLSA and motifs for PLSM).

6.1.1 Level-1

From the distributions obtained at level-1 the lower-level document for each camera can be reconstructed by unrolling the lower level topics (z^{ll}) with the word distribution Pwz.

$$\tilde{n}^{rec\,ll}(w^{ll}, d^{ll}) = mat(P^{ll}_{wz}, P^{ll}_{zd}). \tag{6.1}$$

6.1.2 Level-2

From the distributions $Pwtz$ and $Pztd$ we can obtain $n^{rec}(w, t_a, d)$ as shown in Chapter 5. Once we obtained $n^{rec}(w, t_a, d)$, as w corresponds to z^{ll} (PLSA topic) it is possible to use the PLSA topics to reconstruct the low level documents ordered in time. By unrolling the equations, we obtain:

$$n^{rec\,ll}(w^{ll}, d^{ll}_{t_a}) = \sum_{z^{ll}} n^{rec}(z^{ll}, t_a, d) P^{ll}_{wz}. \tag{6.2}$$

6.1.3 Level-3

From the information learned at the combined level, we can obtain $n^{rec\,cl}(w^{cl}, t_a, d^{cl})$ as seen earlier. As w^{cl} corresponds to motifs of a specific camera at the first-level, it is possible to use these motifs to obtain documents $n^{rec}(w, t_a, d)$ pertaining to the first level. We can finally obtain the low-level document as seen in level-1.

6.2 Localized Abnormality Measure

We can obtain abnormalities [5] from the reconstructed document by using the distance measure proposed in [10]. However, this measure does not take into account the spatial locality of the anomaly. We thus compute anomaly by first extracting anomaly in blocks and then finding the most abnormal group of blocks.

We achieve localized abnormality by dividing a frame of video into $h \times w$ sub-frames where h and w are parameters to the model. Another advantage of blocking is we can detect anomalies in presence of high normal activity. This fact is illustrated in Fig. 6.2. We also normalize the blocks with respect to the total activity in the block. The reason for normalizing can be better understood by a scenario as in Fig. 6.3. Both scenes have no abnormality but the scene which is crowded might be considered abnormal due to the additive effect of the reconstruction error per low-level word. We compute the reconstruction

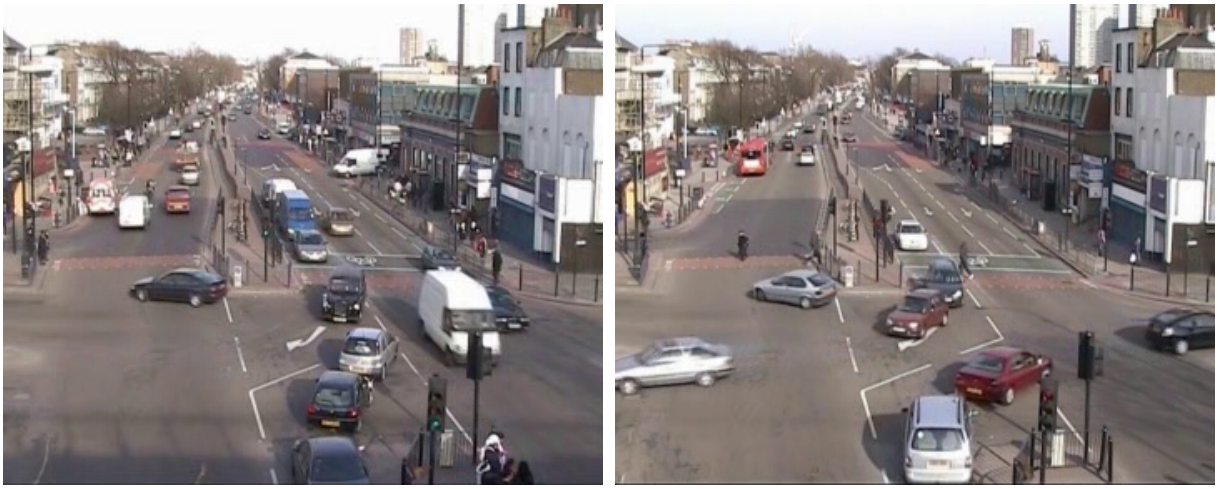


Figure 6.2: localized anomaly: Traffic scene depicts the presence of a single vehicle making a U-turn (anomaly) in otherwise normal scene.

Low Density High Density

Figure 6.3: Density Correction: Scene from the metro camera.

error measure to each sub-frame as:

$$abn(ta, x, y, d) = \sum_{w^{ll} \in R_{xy}} \left| n^{rec\, ll}(w^{ll}, d^{ll}_{t_a}) - n^{ll}(w^{ll}, d^{ll}_{t_a}) \right| \tag{6.3}$$

where R_{xy} represents all the words mapping to the sub-frame x, y. We also normalize the reconstruction error in a sub-frame by diving it by the mass of the document corresponding to the sub-frame.

$$normabn(ta, x, y, d) = \frac{abn(ta, x, y, d)}{\sum_{w^{ll} \in R_{xy}} n^{ll}(w^{ll}, d^{ll}_{t_a})} \tag{6.4}$$

We then use Kadane's algorithm for the **maximum 2D sub-array** problem on $normabn(ta, ., ., d)$ to obtain the abnormality measure for the whole frame and its spatial locality. The crux of Kadane's 2D algorithm is Kadane's 1D algorithm which involves a scan through the array values, computing the maximum sum up to that position.

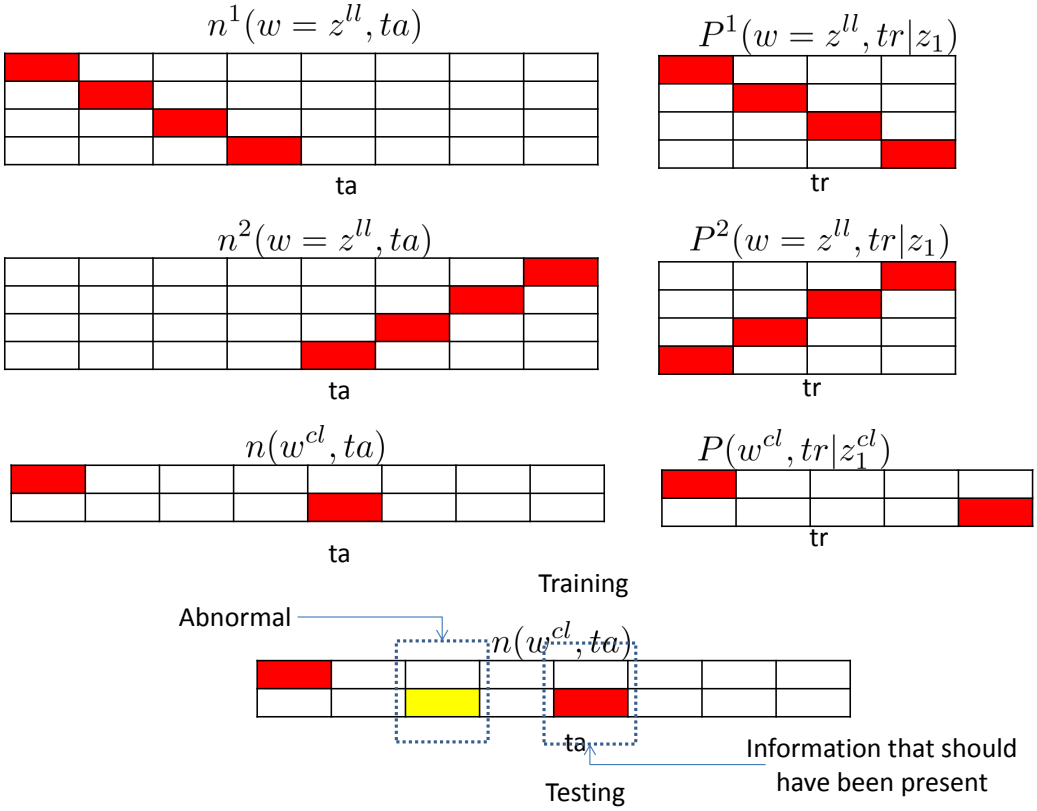


Figure 6.4: Abnormality Detection at the combined-level. The figure illustrates patterns captured at the 2-level and combined-level on their respective training documents and the testing phase shows an abnormal region arising due to bad reconstruction from the learned patterns at the combined level.

6.3 Abnormality Detection by Different Levels

The model can be used to detect specific kinds of abnormalities based on the level of the topic modeling. At the PLSA level, the spatial abnormalities can be detected.At the second-level abnormalities that are related to time can be detected. At the combined-level, where the cameras are related, the patterns which are valid in individual cameras but abnormal in time relative to the cameras can be detected. We illustrate this fact by the following example.

Fig. 6.4 shows the patterns learned on the training documents from two cameras which are spatially related and hence have temporal correlation. The pattern $P^2(w = z^{ll}, tr|z_1)$ occurs

in the second camera at the completion of $P^1(w = z^{ll}, tr|z_1)$ in the first camera. The pattern learned at the combined-level PLSM captures this fact. If this temporal restriction is violated as seen in the testing document, the reconstruction will be poor at the combined-level and hence anomaly can be detected.

Chapter 7

Results

In this chapter we test whether our model can detect activity patterns not learned by the model. Below, we will describe in brief the datasets, the anomalies they contained, and provide some quantitative and qualitative evaluation on traffic and metro camera datasets. Finally, we summarize the contributions made and provide future directions for research.

7.1 Experiments on traffic camera dataset

To test whether our model can detect activity patterns not learned by the model, we used two different traffic scenes: QMUL Roundabout and QMUL Junction. As parameters for IPLSM, we used 80 PLSA topics (based on hierarchical Dirichlet process), S=0.75, and 4 feedback iterations for learning. The frame segmentation parameter $h \times w$ for the anomaly detector is 24×24. The number of motifs and motif length is specified individually for datasets.

7.1.1 QMUL Roundabout

This dataset contains 60 minutes of video at a resolution of 360×288 at 25fps. The traffic movements in the roundabout signal are restricted to only certain driving directions as illustrated in Fig. 7.1. The single type of anomaly present in this dataset is indicated by a red arrow in 7.1 and corresponds to driving straight ahead on a right only lane. Annotation of these events was conducted on 10 minutes of the dataset.

The IPLSM model was trained using either 10 or 20 motifs on 15 minutes of video ensured to contain low instances of the abnormality we wanted to detect. As the longest duration for a vehicle to cross the roundabout was around 12 seconds, we choose a motif

QMUL-Junction QMUL-Roundabout

Figure 7.1: The arrows depict driving flow directions which are allowed (in green) or not (in red).

Table 7.1: Roundabout abnormality results

Abnormalities	GT	IPLSM	
		10 motifs	20 motifs
Incorrect direction	15	10	12
False alarms	0	10	4

length of 12 seconds. Examples of extracted motifs are shown in Fig. 5.2. We applied our method to the test data and compared the results to the ground-truth (Gt) (a detected event was considered to match the Gt if it overlapped with it). The results are summarized in Table 7.1.

The false alarm rate was lower when we used higher number of motifs as it could better capture the different traffic patterns variations due to speed, density and type of vehicles in the traffic. We also observed that small vehicles were more difficult to detect in general and would require to set a lower threshold for their detection. Fig. 7.2 provides some examples of the regions detected by the system.

Figure 7.2: Localized anomaly regions detected by our approach. Note that the regions are large as they encompass all the regions with unusual temporal activity (including the regions were activity should have occurred in the normal situation).

7.1.2 QMUL Junction

This dataset is explained in [16]. In this case, the valid driving trajectories are illustrated in Fig. 7.1. The abnormalities in this dataset were defined as U-turn and disruptions, where U-turn denotes driving back around the road center, and disruptions indicate interruption of normal flow of traffic by a fire-engine, police or an ambulance. We have not considered Jay-walking as an abnormality because our system is not currently designed to detect abnormalities which reason about the validity of motif occurrences in the context of a cycle. Modeling cycles could be done by adding an HMM on top of motifs occurrences [27].

To evaluate the approach, we trained our approach on 45 minutes of videos (that included the abnormal events) using different parameterization. In one case, we considered motifs of 10s maximum duration, which is more or less the maximum that a vehicle takes to cross the junction, and of 80s duration, which is the duration of a full traffic cycle. Examples of extracted motifs are shown in Fig. 7.3

The result are summarized in Table 7.2. In practice, we observed that traffic disruptions (which often occur out of sync from the traffic cycle) required higher motif lengths able to capture full cycles and provide the necessary context. Also, we noticed that U-turns could

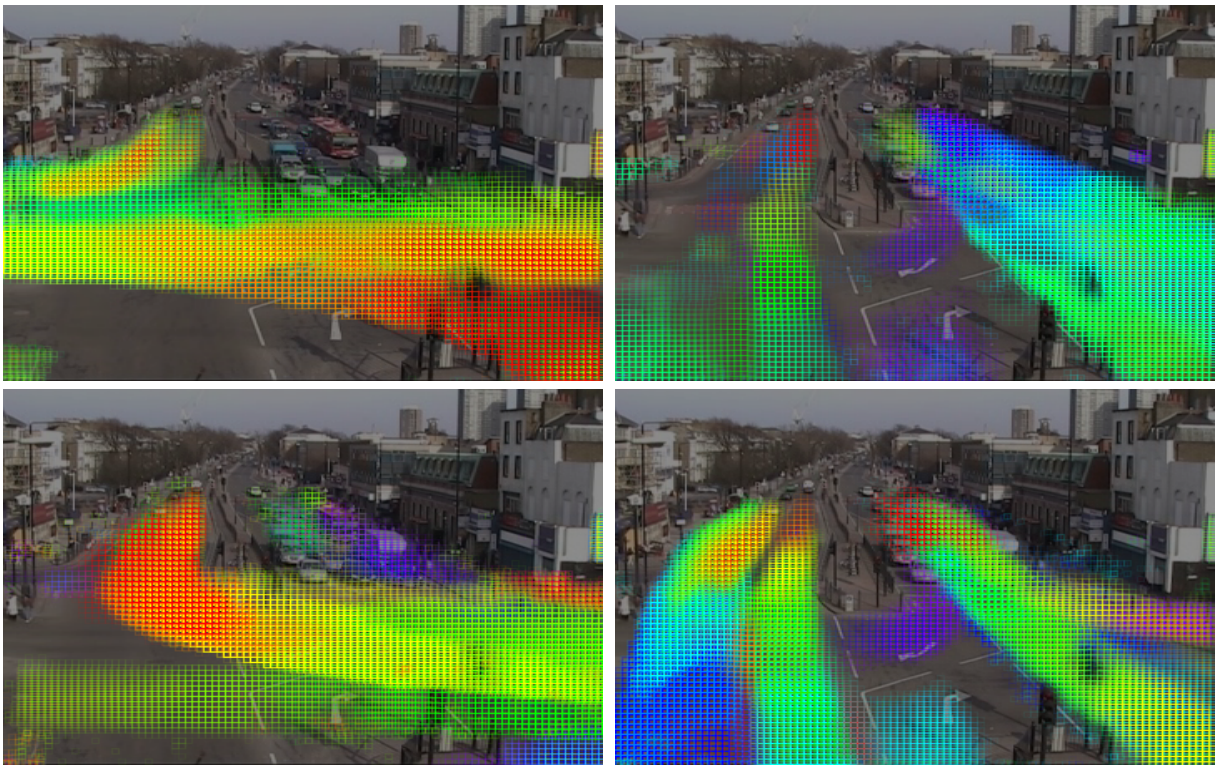


Figure 7.3: Motifs learned on Junction dataset for a M=14 and ML=10. Four Motifs with the highest probabilities among the 14 are shown here.

Table 7.2: QMUL Junction abnormality results. M denotes the number of motifs, ML their maximum length.

Abnormalities	GT	IPLSM	
		M=4, ML=80	M=14, ML=10
U-turn	10	7	8
Disruptions	6	4	1

not be detected well in very dense traffic, as the generated abnormalities were considered negligible as compared to the global activity. Examples of detections are shown in Figure 7.4.

7.1.3 Traffic Junction

This dataset was introduced in [29] consists of a video which is controlled by traffic lights. The dataset is explained in detail in Chapter 3. The unusual events present in the video are: vehicle stopping after the stop line (VSAS), people crossing the road away from the zebra crossing (ZC), jay walking and car stopping in the pedestrian area (PA). We have not considered Jay-walking for reasons already explained.

The IPLSM model was trained using 16 motifs on 20 minutes of video ensured to contain low instances of the abnormality we wanted to detect. As the longest duration for a vehicle to cross the junction was around 8 seconds, we choose a motif length of 8 seconds. Examples of extracted motifs are shown in Fig. 7.5. We applied our method to the test data (the entire video) and compared the results to the ground-truth (a detected event was considered to match the Gt if it overlapped with it). The results are summarized in Table 7.3.

In practice, we observed that pedestrian movements were difficult to model. This is one of the reasons for choosing a higher number of motifs. The false alarm rate could vary greatly depending on the pedestrian activity. Examples of detections are shown in Fig. 7.6.

U-turn disruption

Figure 7.4: Sample of correct detections (a u-turn and a disruption). Notice that in the disruption case, the vehicles on the left should have closely followed the vehicles on the right so that there are 'missing' cars in the middle.

Table 7.3: Traffic Junction abnormality results. M denotes the number of motifs, ML their maximum length.

Abnormalities	GT	IPLSM
		M=16, ML=8
VSAS	5	5
ZC	14	9
PA	1	0

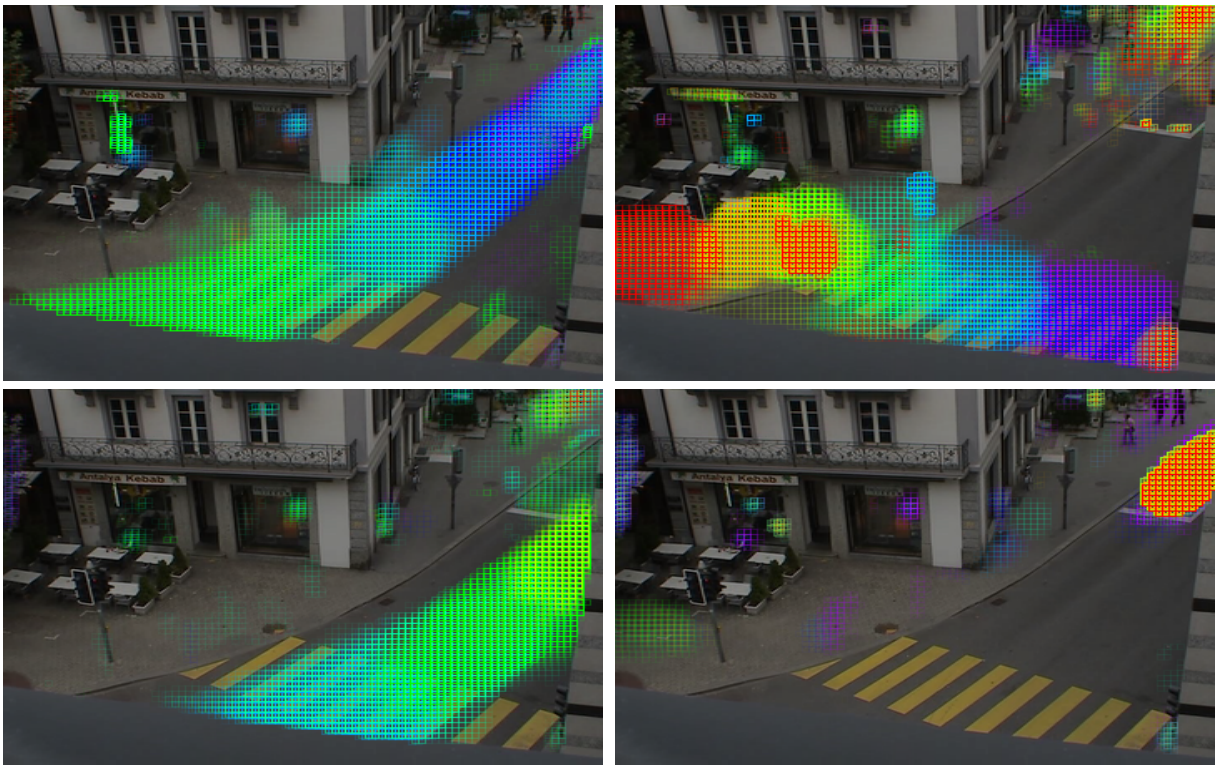
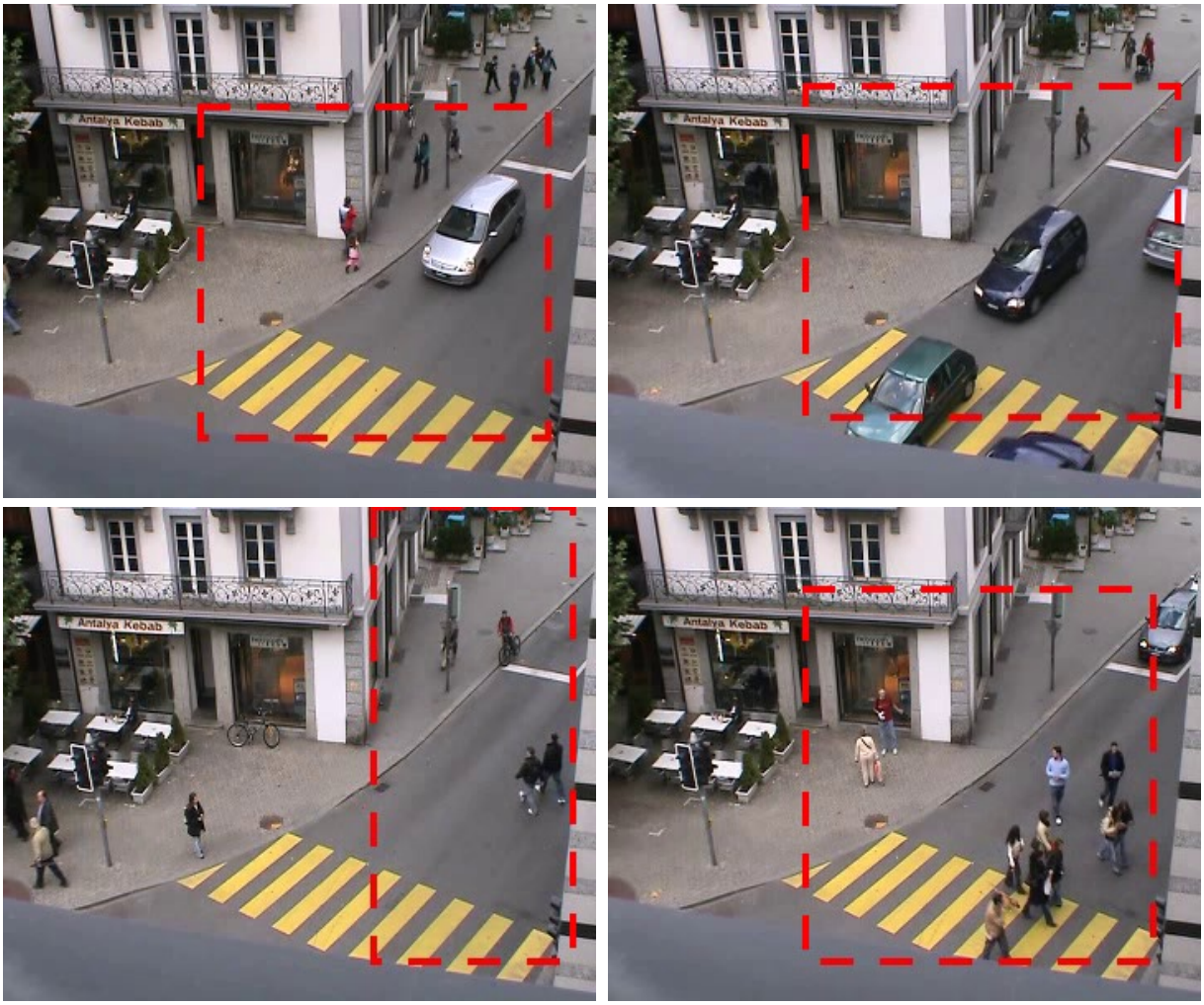


Figure 7.5: Motifs learned on Traffic Junction dataset for a M=16 and ML=8. Four Motifs with the highest probabilities among the 16 are shown here.



Car stopping after stop line

People crossing away from zebra crossing

Figure 7.6: Sample of correct detections (VSAS and ZC).

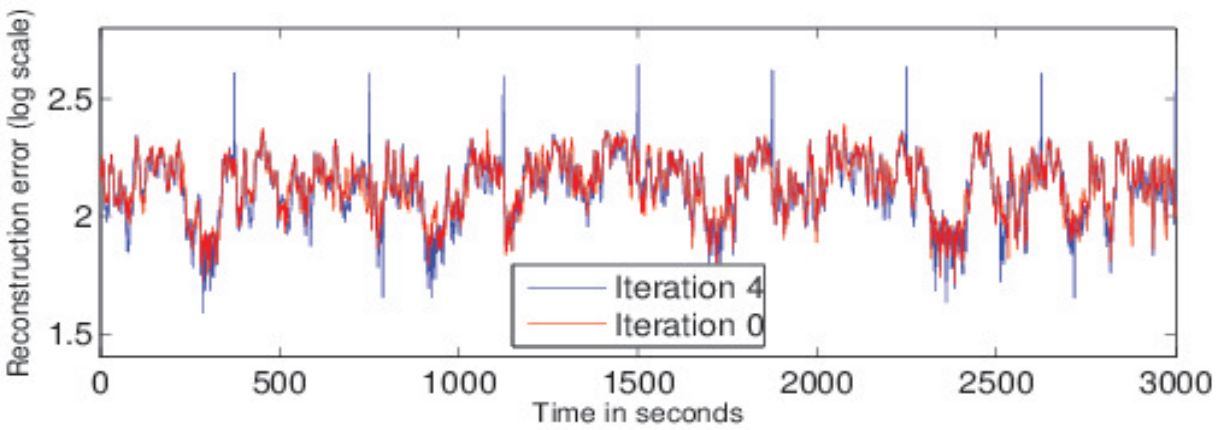
7.2 Experiments on Metro Camera Dataset

This dataset consists of two overlapping cameras, recording neighbouring areas: a stairs, escalator ($Sc1$) and a hall way ($Sc2$). $Sc1$ acts as the entrance into the metro station. $Sc2$ has a ticket counter and turnstiles to the metro network. Below we discuss the effect of feed-back on data reconstruction and briefly state the abnormalities detected.

We trained the hierarchical model with 10 motifs with motif length of 4 seconds for 2nd level PLSM and 12 motifs with motif length of 8 seconds for third level PLSM on 720 seconds of video. The rest of the parameters are the same as used for the traffic dataset. The motifs learned are not shown here. We performed inference on the trained video by splitting the video into 8 parts using the model obtained at iteration 0 (no feedback) and iteration 4 (with 4 iterations of feedback). We reconstructed the data from both the models from the 3^{rd} level PLSM and obtained a plot of the reconstruction error as shown in Fig. 7.7. The reconstructed data from the 8 partial videos were merged together to form the plot for better clarity of information. The reconstruction of the data may be poor at the boundaries where the video was split. This may be because of an activity in this region and also due to the lack of information at the boundaries about how the video might proceed. As seen from the plot, this fact is better captured by iteration 4, as it is better able to identify this temporally disruption due to less over-fitting, hence higher reconstruction error than iteration zero. We showed in Chapter 5 how the prior feedback improved the sequential patterns obtained. Better motifs should constitute better reconstruction of temporally significant data, hence lower reconstruction error. We performed inference on a new video from the same cameras. The abnormalities detected by the system include, unusual density of crowd, people blocking each other and disrupted trajectories. Fig. 7.7 shows two of these abnormalities. The abnormality detector for the 3^{rd} level PLSM would detect regions in the most abnormal camera while the detectors in the 2^{nd} and 1^{st} level would detect abnormalities pertaining to a single camera.

Unusual crowd density

disruption in trajectory



Reconstruction error

Figure 7.7: Abnormalities for Metro dataset at 3^{rd} and 2^{nd} level PLSM and effect of feedback on reconstruction error

Chapter 8

Applications to BCI

This chapter discusses the application of topic models, in specific PLSA and PLSM introduced in Chapter 2 to find spatial correlation (inter channels) and temporal structure in EEG data. First, we will briefly describe EEG signals with respect to Brain Computing Interfaces (BCI) and discuss its spectral characteristics and motivate the purpose of using PLSA and PLSM. We then describe the dataset followed by feature extraction process and then finally we discuss preliminary results on 9 subjects.

8.1 Introduction

Our brain acts as a central processing unit which processes information and acts as the control center for all our body functions like walking, talking, vision. Our brains are filled with neurons which are like electrical conductors which communicate these control signals to various parts of the body. They also carry external input to the brain. These electrical signals produced by the brain are called EEG or brain waves.

BCI are devices which read the EEG signals and communicate it to an external device usually to establish new methods of communication previously not possible. The EEG signals can be recorded by invasive and non-invasive techniques. Invasive techniques [17] implant the BCI device directly onto the grey matter surgically. This technique provides the highest quality of signals and has the highest signal to noise ratio. Non-invasive techniques usually consist of placing electrodes spatially along the scalp. These recordings are highly corrupted by noise and EEG artefacts. EEG artefacts are signals produced by non cerebral activity like eye blinks, pacemakers, and muscle movements. Their power spectral density (PSD) is

usually higher than the PSD of the EEG signal representing the activity [18, 23].

In this work we try to solve the problem of pattern detection from EEG signals recorded while subjects were performing four mental tasks described later. The EEG signals were recorded using a non-invasive technique. The problem of pattern detection can be addressed using the actual time domain representation of the EEG signals [11] or using the frequency domain representation [21]. We will be using the frequency domain representation (Fourier transforms). This can be motivated by the fact that different brain activities are attributed to certain spectral bands and electrodes in EEG. Typical spectral bands in EEG [32] are the following.

- Delta waves are in the range 1-4hz. They can be located in frontal lobe in adults. These are the slowest waves and have the highest amplitude. They have been found during continuous attention tasks and during sleep.

- Theta waves are in the range 4-7hz. They can be located in hippocampus region. They are usually found during idling, arousal and drowsiness

- Alpha waves are in the range 7-14hz. They can be located in frontal and central regions. They are normally found when relaxed or reflective.

- Beta waves are in the range 15-30hz. They can be located symmetrically on both sides but most evident on the frontal region. These are low amplitude waves and are found during alertness, anxiety and active concentration.

- Gamma waves are in the range 30-100hz. They can be located in somatosensory cortex. They are normally noted during task like sound perception and visual processing.

PLSA learns topics that can best define the mental tasks. As explained above each mental task is characterized by activities in specific regions of the brain, the topics learned by PLSA may capture this information. These topics define the mental task irrespective of their time of occurrence. PLSM can be used to reason about the order in which these activities occur

(if any) in accomplishing a mental task. The spectral components for any activity is subject dependent [21, 23]. Hence the topics models provide a data driven dimensionality reduction or basis vector generation.

8.2 Dataset

We use the g.Tec g.GAMMAsys dataset from `http://www.cs.colostate.edu/eeg/main/data/2011-12_BCI_at_CSU` which consists of EEG signals from subjects recorded at the CSU BCI lab and impaired subjects home. The EEG was recorded using g.Tec g.GAMMAsys system with eight active electrodes (channels) and a sampling frequency of $256hz$. The 8 channels are $F3, F4, C3, C4, P3, P4, O1, O2$ and their positions on the scalp is depicted in Fig. 8.1. The dataset contains recordings from 12 subjects out of which four were impaired. Subject 11, 13, 15 and 16 had severe motor impairments and recording was done at home. Subjects 20-28 recordings took place in a laboratory. Each subject performed four mental tasks: silently sing a song (S), visualize a rotating cube (R), imagine right hand clenching (F) and counting backwards from 100 in steps of three (C). Each task was performed for 10 seconds. Up to six trials were performed, each trial consisting of all four tasks. We will only be using data from Subjects 11, 13 and 20-27. The rest of the subject's data was not used either due to insufficient number of trials or due to corrupted data.

8.3 Feature Construction for PLSA and PLSM

We use the frequency domain representation of the EEG signal by transforming the time-domain signal from each channel to its frequency domain using Short Time Fourier transform. The transform uses a window length (WL) of 256 which represents a signal of one second with an overlap of 230 points. The motivation for choosing the particular WL is the fact that a subject is unlikely to switch activity within this period.

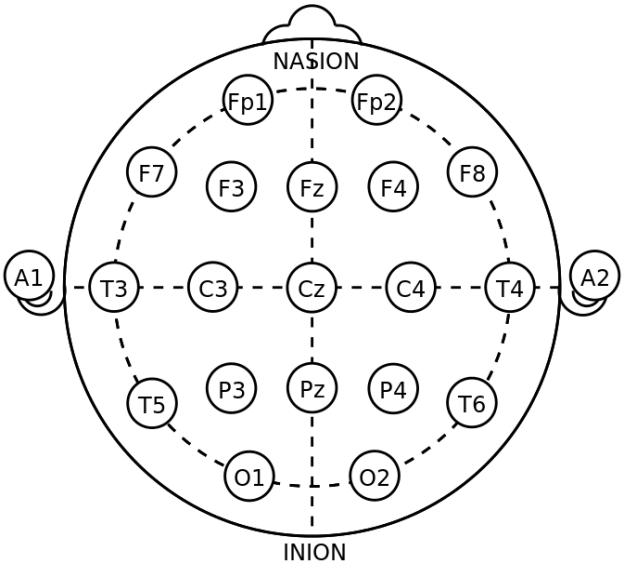


Figure 8.1: 10-20 BCI system: Depicts the electrode positions in 10-20 BCI system [31]

8.3.1 Vocabulary Construction for PLSA

We are interested in capturing co-occurrence of frequencies and channels which represent a mental task. PLSA can be used to capture this spatial correlation provided this information is encoded in its vocabulary. To encode this information, we combine all the windows of the channels in a particular time to form a combined WL of 1032. Each combined window represents a document d and its spectral components represents the words w. Each mental task is thus represented by a total of 89 documents. We form the term-document frequency matrix $n(d, w)$ by concatenating all the mental tasks from all the trials for a subject. The ordering of the documents maintains the order of time within each mental activity (the number of rows (documents) is a multiple of 89). This is essential for PLSM.

After applying PLSA, we obtain the distributions $P(w|z)$ and $P(z|d)$ described in Chapter 2. $P(w|z)$ represents the co-occurrence frequency and channel distribution among different topics. $P(z|d)$ gives the mixture components or weights of the topics that represents a document. Essentially $P(z|d)$ represents mental activities in reduced space (dimensionality reduction)

8.3.2 Applying PLSM

$P(z|d)$ describes windowed mental tasks in terms of activities. The activities capture the frequencies and channels that are most likely to co-occur for a given mental task. We hypothesize that mental task are described by sequential structure that occurs at some repeated intervals. We use PLSM to capture this sequential structure. The sequential structure (motifs) will represent a pattern consisting of activities captured by PLSA ordered in time. $P(z|d)$ is processed as described in Chapter 4 to obtain $n(w, ta, d)$. $P(z|d)$ is reordered at intervals of 89 windows which correspond to different tasks, this represents the documents d for PLSM. The 89 windows are ordered in time ta.

8.4 Experiments and Results

In this section we will discuss the experimental setup which involves the parameter selection for PLSA and PLSM, the partition of data into train and test and the evidence accumulation algorithm used to improve the classification accuracy. We also discuss the impact of classification accuracy on the size of the training data.

8.4.1 PLSA

The parameters of PLSA are estimated using MLE technique. As previously explained, MLE has the drawback of over fitting the parameters to the input data. Since the number of trials for each subject is small in number, the impact of over fitting would be more evident in our case. To estimate the impact of over fitting on classification accuracy, we estimate the parameters of PLSA for each subject under two different experimental set-ups.

- $SET-1$ consists of all the 9 subjects and we use only the first 5 trials in the dataset. One trial was designated as test data and other 4 trials were training data. The experiment was repeated for all 5 combinations of test and train set.

- $SET-2$ consists of only 7 unimpaired subjects (impaired subjects didn't perform the sixth trial) and we use all the 6 trials in the dataset. One trial was designated as

test data and other 5 trials were training data. The experiment was repeated for all 6 combinations of test and train set.

The topic distribution $P(z|d)$ obtained from PLSA is provided as input to a support vector machine for classification. As previously described each 10 seconds mental task is represented by 89 documents. Hence an SVM can make approximately 9 *decisions/sec*. Since it is unlikely that a person will switch between tasks so fast, the classification accuracy can be improved by using the M-ary Sequential Probability Ratio Test (MSPRT) algorithm [1, 9] that incrementally calculates the joint probability of each class given an increasing window of EEG samples. The MSPRT makes a decision when the accumulated log probability crosses a threshold and a specific time interval has elapsed between decisions. A threshold of 0.8 and a time interval of 9 decisions (one second) was used as parameters for the MSPRT algorithm. A maximum of 9 decisions can be made for mental task spanning 10 seconds.

The average classification accuracy (CA) for different number of topics for $SET-1$ and $SET-2$ is compared in Fig. 8.2, 8.3, 8.4, 8.5. It can be seen that $SET-2$ has better CA. This can be attributed to PLSA finding better topics that can represent the training data.

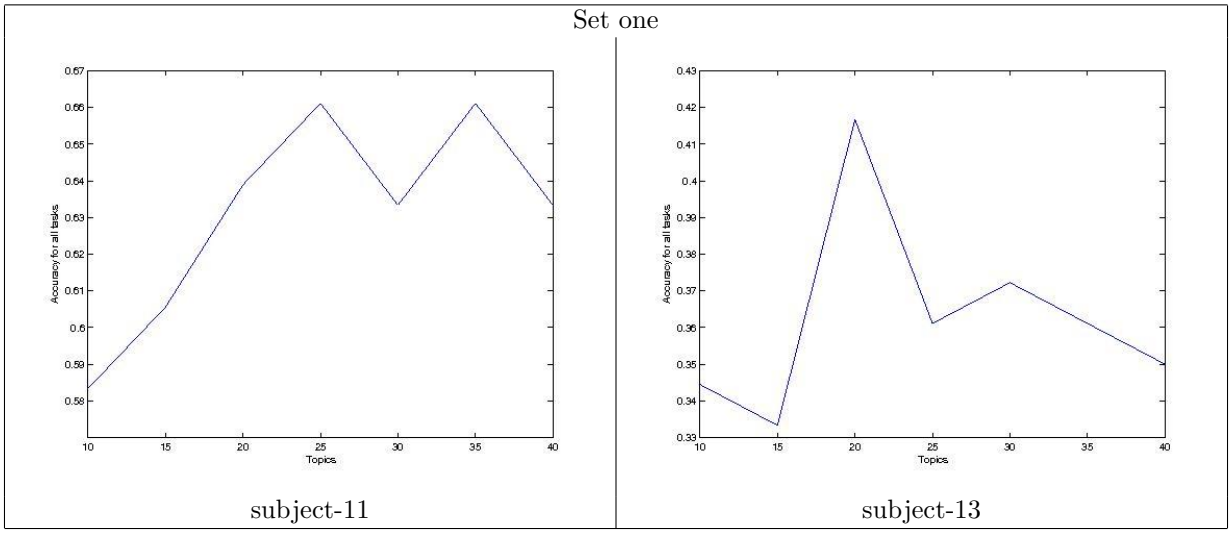


Figure 8.2: Average classification accuracy for subjects 11 and 13 : The figure provides a comparison of the average classification accuracy on test set for topics in the range 5-40 for PLSA with an SVM classifier.

The BCI system is intended to be used by the physically challenged. Hence the topic and word distribution of Subject-11 which has the highest CA is discussed in greater detail. The CA for subject-11 is highest for 25 topics and Figure 8.6 shows the classification rate for four mental tasks for all combinations of the test set. Trial four has the highest classification accuracy and the tasks C and F have perfect classification rate. Thus the distribution $P(w|z)$ should have discriminative aspects and certain topics should correspond to the mental tasks and others should capture the noise and artefacts. Figure 8.7 shows the topic distribution. The topics 12, 20 and 21 in task C have high probability and are persistent throughout the task. But the topics 11, 21 and 20 are also present with high probability in the other three tasks hence doesn't provide any discriminative power for representing a specific mental task. The word distribution of this topic in-fact indicates that this may be noise. Topic-11, though, could possibly represent the Task C because this word distribution was seen to occur in other subjects and this distribution represents delta wave activity and is seen during high concentration tasks. Topics 8 and 16 represent activity in the left, topics 9 and 14 in the right portion of the brain. These four topics were only captured when the classification rate was above 0.4 for tasks R and S across all subjects. Though this cannot be directly concluded based on the topic distribution in Fig. 8.7 as it is fairly dense. Topic 3 may represent eye blinks because the power spectrum decreases in power from the frontal to the parietal region. The above mentioned topics were also identified in other subjects.

8.4.2 PLSM on top of PLSA

We train PLSA, PLSM and SVM classifier on individual subjects. The number of topics for PLSA was chosen based on the classification accuracy of the test set. We use some of the subjects from $SET - 1$ and Fig. 8.9 shows the CA for these subjects for a range of motifs and motif lengths. A motif Length of 5 provides the highest CA. A motif length of 5 means 5 documents of PLSA are considered in time. The 5 PLSA documents would correspond to 100 discrete samples adding up to a total time of less than 0.5 second. This is indicative of non existence of long temporal patterns which captures the changes in frequency for a motif

that repeats in time. The CA is better than a random classifier (25% accuracy) but the accuracy is poor compared to using PLSA and SVM without PLSM.

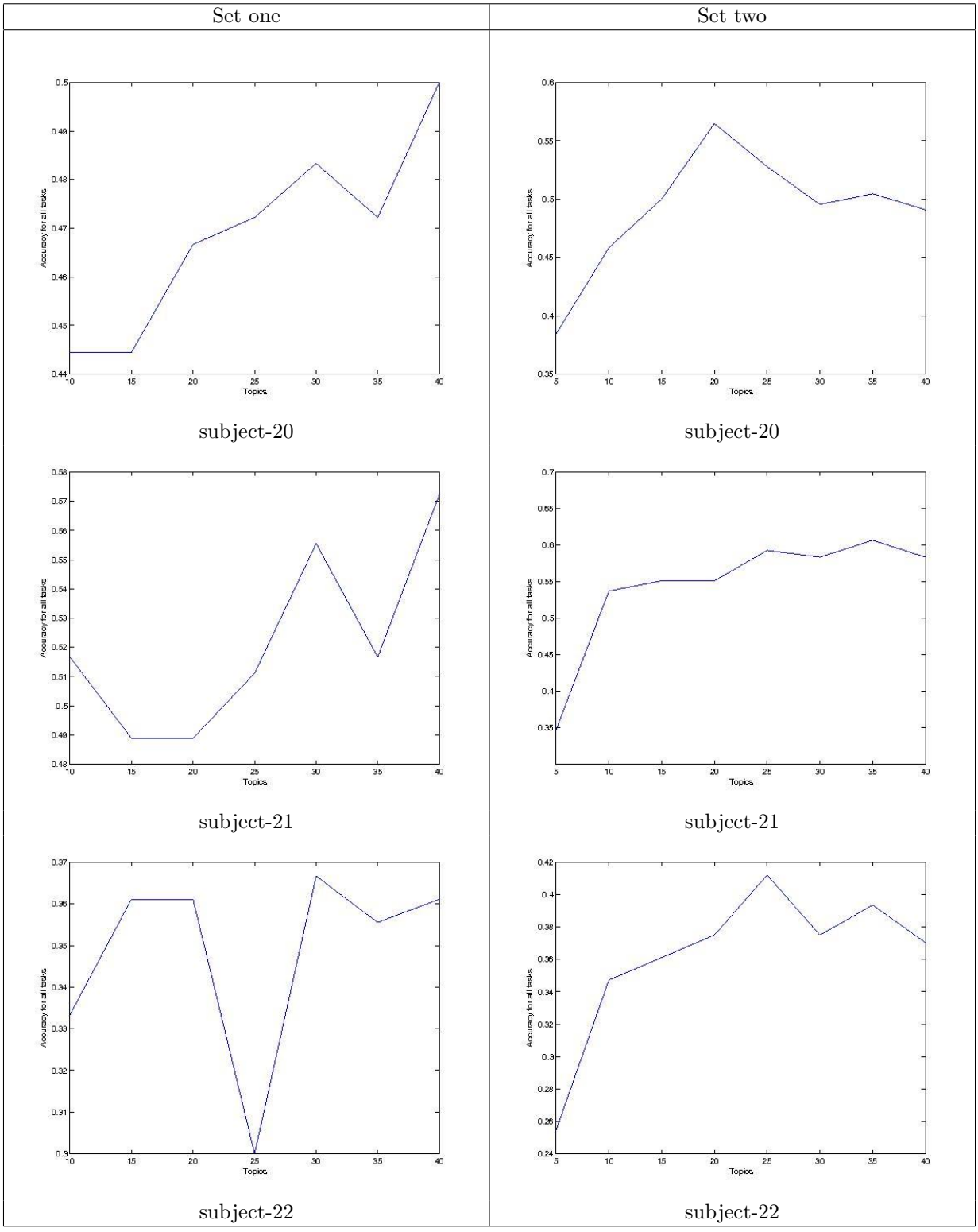


Figure 8.3: Average classification accuracy for subjects 20, 21 and 22 : The figure provides a comparison of the average classification accuracy on test set for topics in the range 5-40 for PLSA with an SVM classifier.

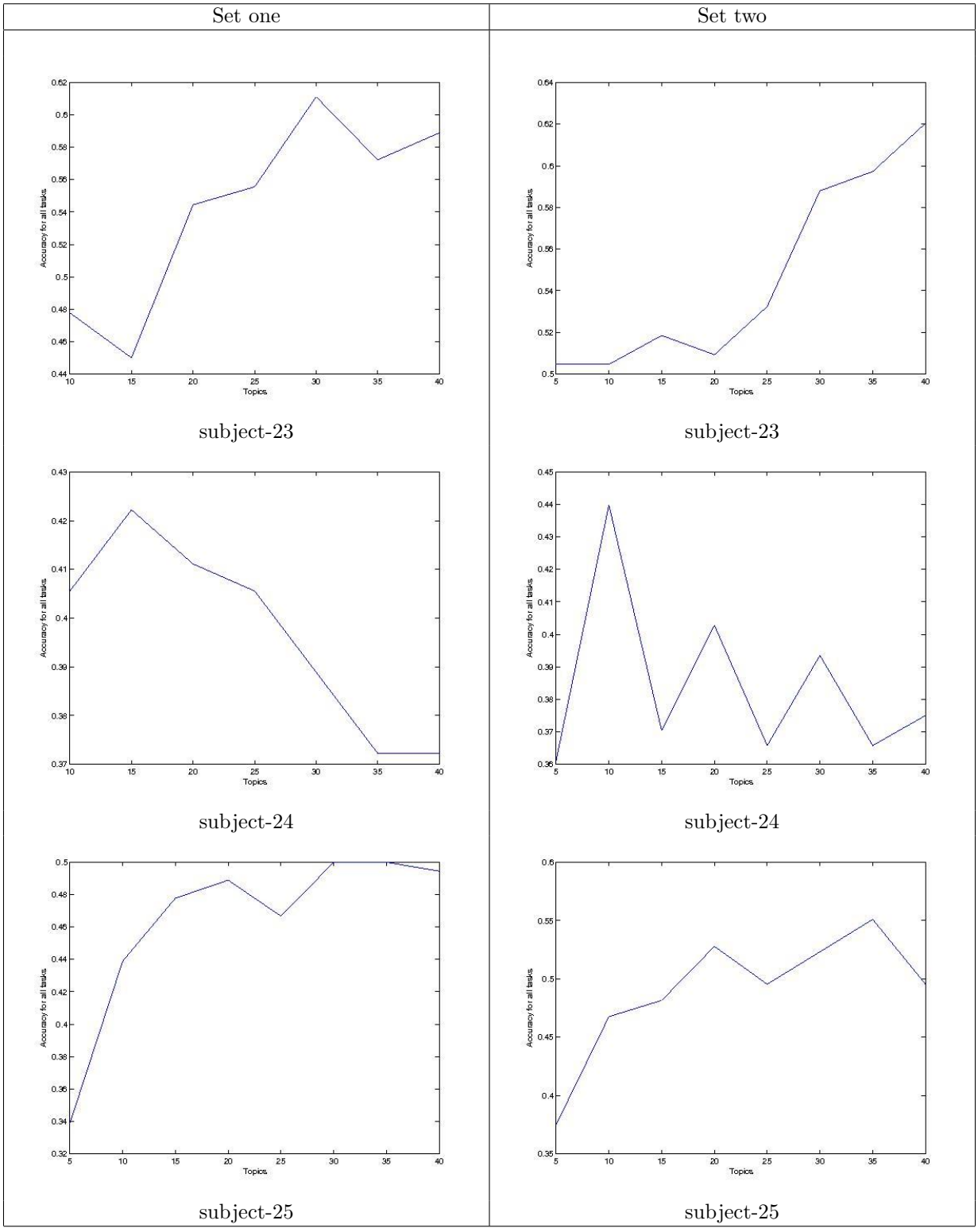


Figure 8.4: Average classification accuracy for subjects 23, 24 and 25 : The figure provides a comparison of the average classification accuracy on test set for topics in the range 5-40 for PLSA with an SVM classifier.

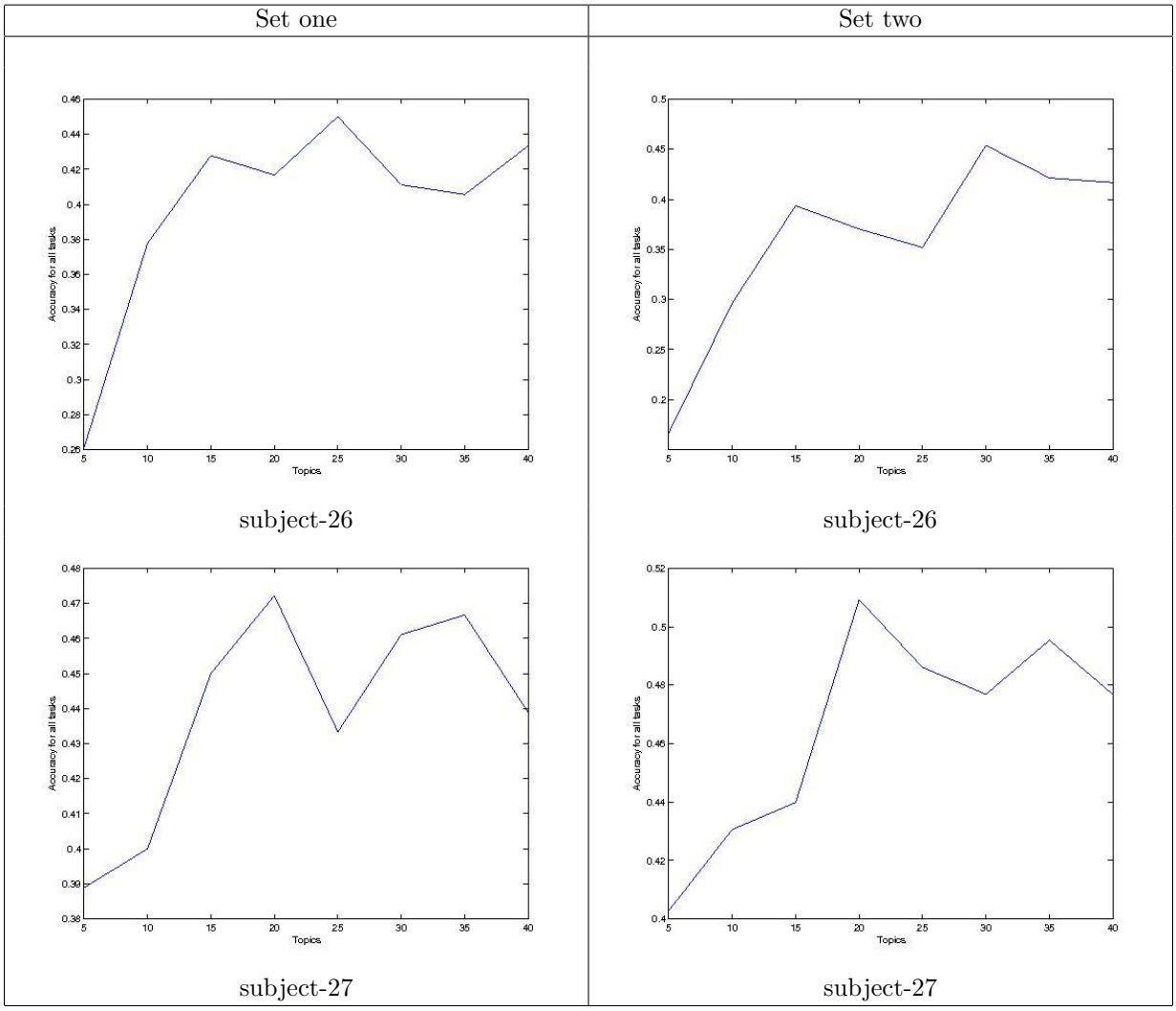


Figure 8.5: Average classification accuracy for subjects 26 and 27 : The figure provides a comparison of the average classification accuracy on test set for topics in the range 5-40 for PLSA with an SVM classifier.

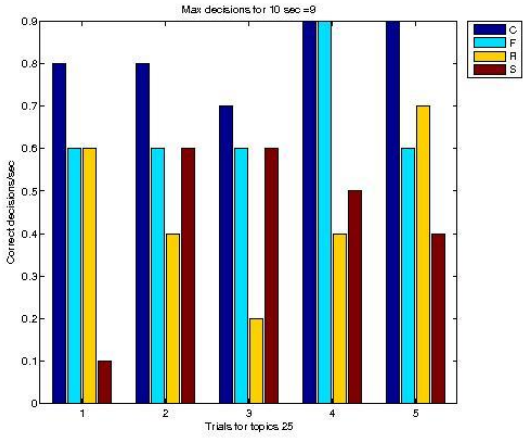


Figure 8.6: Classification rate for all 5 combinations of test set for Subject-11 with 25 topics as parameter to PLSA

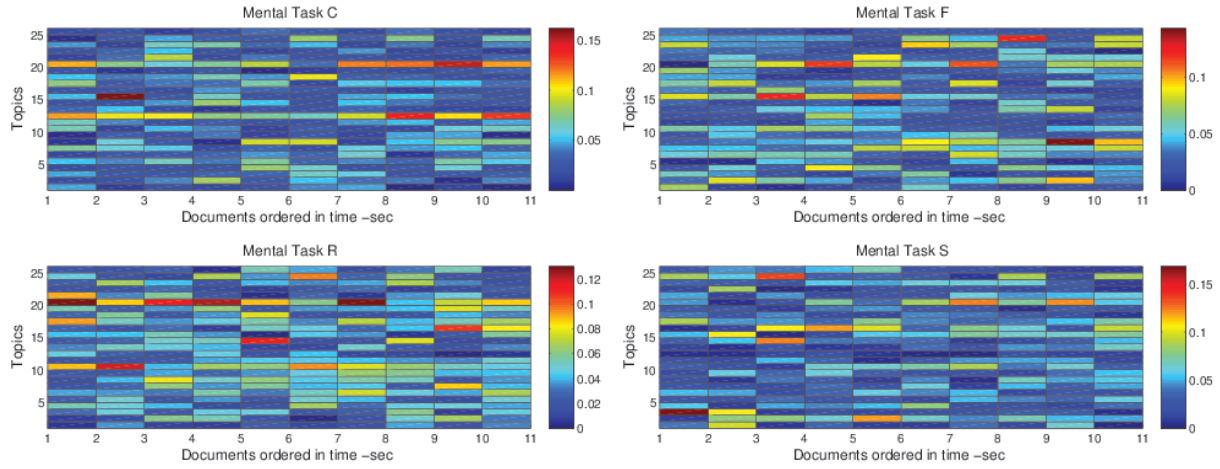


Figure 8.7: Topic distribution: Shows the $P(z|d)$ distribution for subject-11 for test-set trial 4 and 25 topics

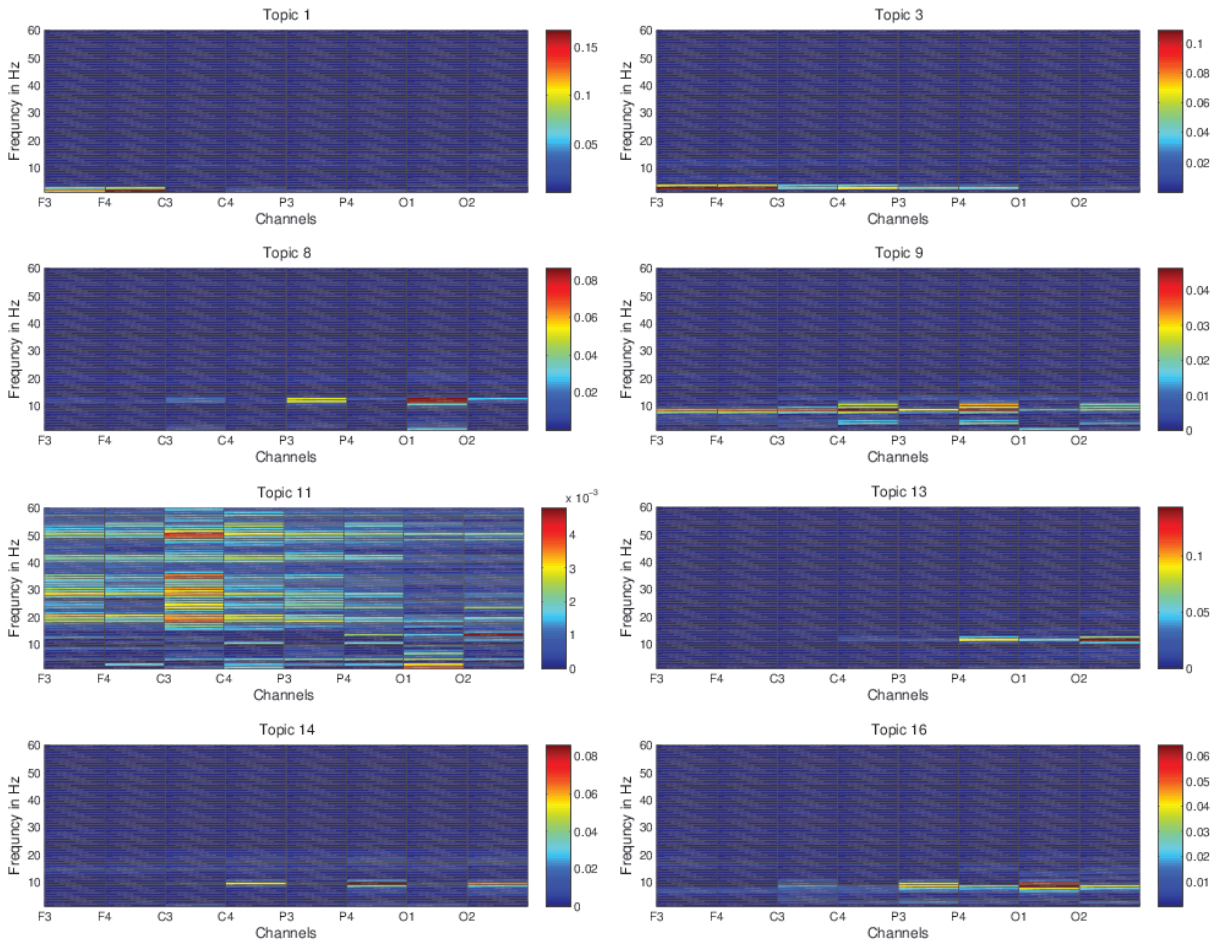


Figure 8.8: Word distribution: Shows the $P(w|z)$ distribution for subject-11 for test-set trial 4 and 25 topics. The frequency band 60-128hz is not shown due to lack of activity.

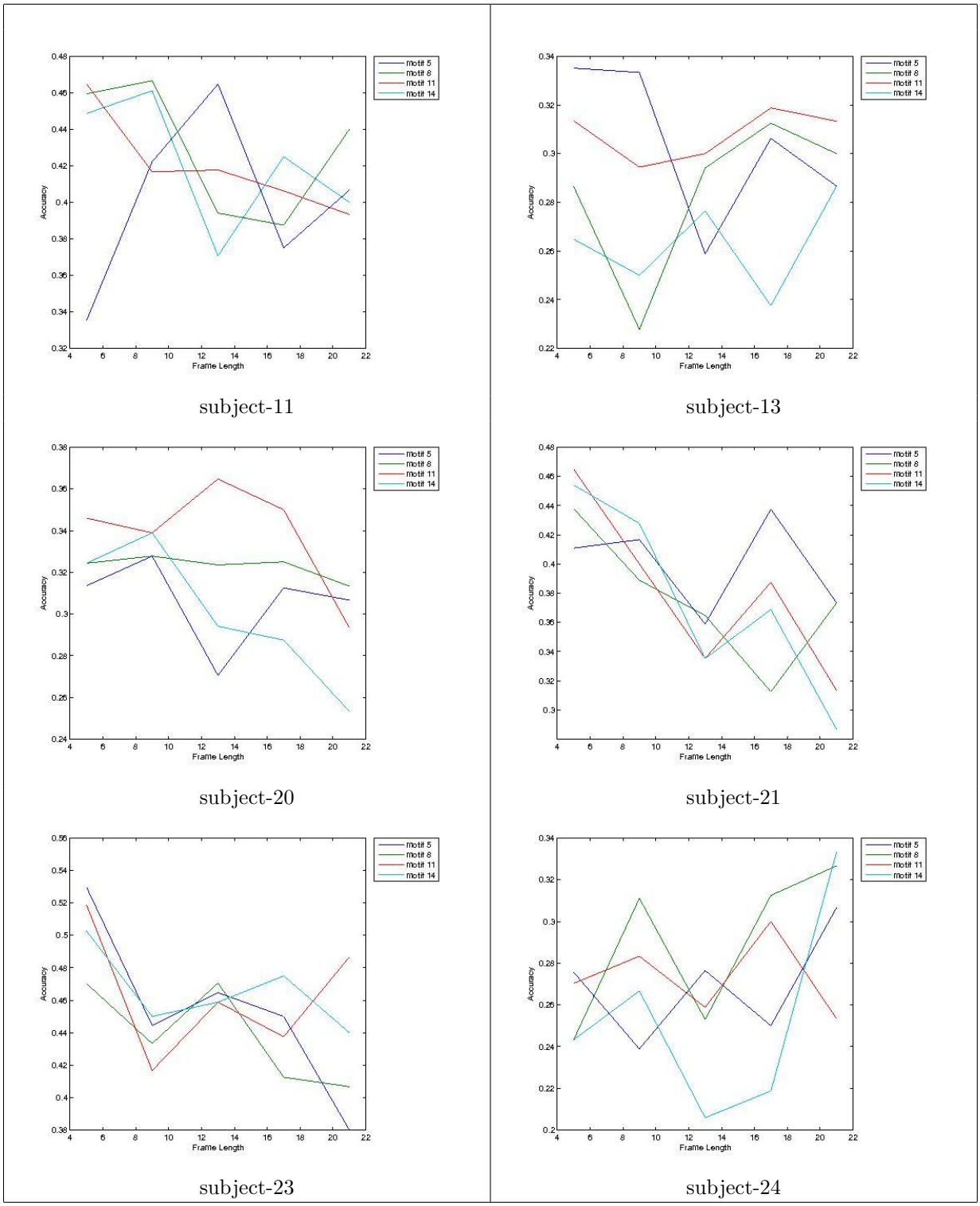


Figure 8.9: Average Classification accuracy: The figure provides a comparison of the average classification accuracy on test set for motifs in the range 5-14 and motif length 5-14 for PLSM on top of PLSA with an SVM classifier

Chapter 9

Conclusion and Future work

In this work, we have addressed the problem of abnormality detection in large camera networks by learning normal activities through hierarchical unsupervised modeling technique. The activities were learned through a temporal modeling approach called PLSM. Previous works applied PLSM on documents built from an intermediate representation learned by dimensionality reduction of low-level features. These low level features were learned in advance and hence they did not benefit from the temporal structure PLSM could provide. PLSM was also ignoring the spatial locality of the anomaly as it was ignoring the semantics of the intermediate representation. These two limitations of PLSM were addressed in this work.

The main contributions we made are the following. We formulated the IPLSM model which integrates PLSA into PLSM and detect spatially localized abnormalities. We achieved the integration by formulating the inter-level feedback in terms of a Dirichlet prior. This feedback was constructed and used in an iterative process by IPLSM. IPLSM was shown to improve the detected sequential patterns compared to the previous approach. We then showed how the model can be extended to mine activities from multiple cameras. The hierarchical model was shown to capture temporally significant patterns in a multi-camera setting. For cameras that share spatial information, the pattern detected is independent of any activity in the other unrelated cameras. Spatial anomalies were identified by splitting a frame in the video into rectangular blocks and then identifying the most abnormal blocks using Kadane's algorithm. We also tested the model on real datasets like traffic surveillance cameras and showed that it can detect abnormalities and localize it on the frame as well.

We also investigated the application of PLSA and PLSM for mining activities from EEG signals. The vocabulary construction was performed using Fourier transforms. Over-fitting issue with PLSA on a small corpus of data was explored by comparing the classification accuracy by creating two different sets of data with one set containing higher number of trials than the other. Analysis of the topic and word distribution was also performed. PLSM was applied on dimensionality reduced documents using PLSA. Mental task classification was performed using an SVM classifier on information received from PLSM and PLSA. We noticed that the parameters of the model and classification rate was subject dependent. The classification accuracy of PLSA with SVM was found be higher compared to PLSM on top of PLSA with SVM

9.1 Limitations and Future Work

As previously discussed, our model cannot reason about the co-occurrence of motifs or its time of occurrence. This can be addressed using an HMM over our model to detect cycles in which these motifs should occur (Jay walking is a typical example). When the number of cameras increases, there may be multiple spatial anomalies in the same or different cameras. Our model can easily be extended to detect multiple anomalies in a frame. In the future IPLSM can be applied on EEG data and the results can be compared to existing work. HMM can also be used for decision filtering to improve the classification accuracy of mental tasks. We also noticed that the topic distribution was fairly dense, this made the distinction of contributions of topics to specific mental tasks difficult. This limitation can be solved by taking a discriminative approach to parameter estimation using Fisher kernels instead of the MLE approach.

References

[1] C. Anderson, E. Forney, D. Hains, and A. Natarajan. Reliable identification of mental tasks using time-embedded eeg and sequential evidence accumulation. *Journal of Neural Engineering*, 8(2), 2011.

[2] N. Anjum and A. Cavallaro. Trajectory association and fusion across partially overlapping cameras. *Proceedings of the Sixth IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, pages 201–206, 2009.

[3] A. Avanzi, F. Bremond, C. Tornieri, and M. Thonnat. Design and assessment of an intelligent activity monitoring platform. *EURASIP Journal on Appl. Signal Proc.*, pages 2359–2374, January 2005.

[4] D. M. Blei. Introduction to probabilistic topic models. *Communications of the ACM*, 55(4):77–84, April 2011.

[5] S. Calderara, U. Heinemann, A. Prati, R. Cucchiara, and N. Tishby. Detecting anomalies in people's trajectories using spectral graph analysis. *Journal of Computer Vision and Image Understanding*, 115(8):1099–1111, Aug. 2011.

[6] T. Chockalingam, R. Emonet, and J.-M. Odobez. Localized Anomaly Detection via Hierarchical Integrated Activity Discovery. In *Proceedings of the IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, pages 51–56, 2013.

[7] S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science and Technology*, 41(6):391–407, 1990.

[8] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, SERIES B*, 39(1):1–38, 1977.

[9] V. Dragalin, A. G. Tartakovsky, and V. V. Veeravalli. Multihypothesis sequential probability ratio tests, part ii: Accurate asymptotic expansions for the expected sample size, 1998.

[10] R. Emonet, J. Varadarajan, and J.-M. Odobez. Multi-camera Open Space Human Activity Discovery for Anomaly Detection. In *Proceedings of the IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, pages 218–223, 2011.

[11] E. M. Forney and C. W. Anderson. Classification of EEG during imagined mental tasks by forecasting with Elman Recurrent Neural Networks. In *Proceedings of the 2011 International Joint Conference on Neural Networks (IJCNN)*, pages 2749–2755, 2011.

[12] G. Heinrich. Parameter estimation for text analysis. Technical report, 2004.

[13] T. Hofmann. Probabilistic latent semantic analysis. In *In Proc. of Uncertainty in Artificial Intelligence, UAI99*, pages 289–296, 1999.

[14] T. Hospedales, S. Gong, and T. Xiang. A Markov Clustering Topic Model for Mining Behaviour in Video. In *Proceedings of the 12th IEEE International Conference on*

Computer Vision (ICCV), pages 1165–1172, 2009.

[15] E. Jouneau and C. Carincotte. Mono versus Multi-view tracking-based model for automatic scene activity modeling and anomaly detection . In *Proceedings of the IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, pages 95–100, 2011.

[16] E. Jouneau and C. Carincotte. Particle-based tracking model for automatic anomaly detection. In *Proceedings of the 2011 18th IEEE International Conference on Image Processing (ICIP)*, pages 513–516, 2011.

[17] W. JR, B. N, H. WJ, M. DJ, P. PH, S. G, D. E, Q. LA, R. CJ, and V. TM. Braincomputer interface technology: A review of the first international meeting. *IEEE Transactions on Rehabilitation Engineering*, 8(2):164–173, 2000.

[18] T.-P. Jung, S. Makeig, M. Westereld, J. Townsend, E. Courchesne, and T. J. Sejnowski. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Journal of International Federation of Clinical Neurophysiology*, 111(8):1745–1758, June 2000.

[19] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

[20] D. Kuettel, M. D. Breitenstein, L. V. Gool, and V. Ferrari. What's going on? discovering spatio-temporal dependencies in dynamic scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1951–1958, 2010.

[21] H. Lee, A. Cichocki, and S. Choi. Kernel nonnegative matrix factorization for spectral eeg feature extraction. *Journal of Neurocomputing*, 72(13-15):3182–3190, Aug. 2009.

[22] C. C. Loy, T. Xiang, and S. Gong. Multi-camera activity correlation analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1988 – 1995, 2009.

[23] P. Manoilov. Eye-blinking artefacts analysis. In *Proceedings of the 2007 International Conference on Computer Systems and Technologies*, CompSysTech '07, pages 52:1–52:6, New York, NY, USA, 2007. ACM.

[24] M. Steyvers and T. Griffiths. Probabilistic topic models. In T. Landauer, D. Mcnamara, S. Dennis, and W. Kintsch, editors, *Latent Semantic Analysis: A Road to Meaning*. Laurence Erlbaum, 2006.

[25] C. Tomasi and T. Kanade. Detection and tracking of point features. Technical report, International Journal of Computer Vision, 1991.

[26] J. Varadarajan, R. Emonet, and J.-M. Odobez. Probabilistic latent sequential motifs: Discovering temporal activity patterns in video scenes. In *Proceedings of the British Machine Vision Conference*, pages 117.1–117.11. BMVA Press, 2010. doi:10.5244/C.24.117.

[27] J. Varadarajan, R. Emonet, and J.-M. Odobez. Bridging the Past, Present and Future: Modeling Scene Activities From Event Relationships and Global Rules. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2096–2103, 2012.

[28] J. Varadarajan, R. Emonet, and J.-M. Odobez. A Sequential Topic Model for Mining Recurrent Activities from Long Term Video Logs. *International Journal of Computer Vision*, 103(1):100–126, 2012.

[29] J. Varadarajan and J.-M. Odobez. Topic Models for Scene Analysis and Abnormality Detection. In *Proceedings of the 12th IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 1338–1345, 2009.

[30] Y. Wang, L. He, and S. Velipasalar. Real-time distributed tracking with non-overlapping cameras. In *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP)*, pages 697–700, 2010.

[31] Wikipedia. 10-20 system (eeg) — wikipedia, the free encyclopedia, 2013. [Online; accessed 13-December-2013].

[32] Wikipedia. Electroencephalography — wikipedia, the free encyclopedia, 2013. [Online; accessed 17-November-2013].

[33] E. E. Zelniker, S. Gong, and T. Xiang. Global Abnormal Behaviour Detection Using a Network of CCTV Cameras. In *Proceedings of the The International Workshop on Visual Surveillance (VS)*, 2008.