# Content Structure Discovery in Educational Videos Using Shared Structures in the Hierarchical Hidden Markov Models

Dinh Q. Phung[1], Hung H. Bui[2], and Svetha Venkatesh[1]

[1] Department of Computing, Curtin University of Technology,
GPO Box U 1987, Perth, Western Australia
{phungquo,svetha}@computing.edu.au
[2] Artificial Intelligence Center, SRI International
333 Ravenswood Ave Menlo Park, CA 94025, USA
bui@ai.sri.com

**Abstract.** In this paper, we present an application of the hierarchical HMM for structure discovery in educational videos. The HHMM has recently been extended to accommodate the concept of shared structure, ie: a state might multiply inherit from more than one parents. Utilising the expressiveness of this model, we concentrate on a specific class of video – educational videos – in which the hierarchy of semantic units is simpler and clearly defined in terms of topics and its sub-units. We model the hierarchy of topical structures by an HHMM and demonstrate the usefulness of the model in detecting topic transitions.

## 1 Introduction

The discovery of structure in video data is an important problem. Solution to this problem will form the core of multimedia indexing and browsing systems. The discovery of structure is important as it enables the partitioning of the information into meaningful sub-units and to build a hierarchy of such units in increasing levels of detail. Such hierarchies are naturally used in other media, for example, the table of contents in a book. In the case of video, construction of such a hierarchy is equally meaningful and will allow users to browse the media using a table-of-contents style. The difficult question includes not only the construction of the hierarchy, but also the understanding of the sub-units used in the hierarchy.

In this paper, we concentrate on a specific class of video, the educational video, in which the hierarchy of units is simpler and clearly defined in terms of topics and sub-topics. We propose the use of the HHMM to segment this class of video. We first modify the parameter estimation to allow multiple inheritance in hierarchic structures. This is because a video has shared sub-structures and the model needs to accommodate this fundamental aspect of topic organisation. For example, all topics generally start with some introduction shots. The novelty of this work is in the content structure discovery of educational videos where the shared 'concepts' are utilised and incorporated into the model. This is important because without the ability to model shared structures, the shared units will have to be repeated, increasing the state space and thus make the process computationally inefficient. This is particularly relevant when dealing with very long observation sequences such as a full video.

A. Fred et al. (Eds.): SSPR&SPR 2004, LNCS 3138, pp. 1155–1163, 2004.

## 2 Related Background

Discovering structure of videos has been a rapidly growing area, in particular as a subfield of multimedia content management. In these systems, the central task is effectively building units of indexation, possibly at different levels of abstractions, to simplify the process of retrieving and browsing, and also to enrich the viewing experience from the end-users. There have been many systems proposed for specific video genres. Partition and classification of broadcast videos into meaningful sections have attracted significant attention [1–8]. In [7], Liu *et al.* segment news reports from other categories based on both audio and visual information. Low-level features are combined with the concept of shot syntax in [2] to identify and label different narrative structures such as anchor shots, voice-over segments and interview sections found in news programs. Unsupervised groupings of news stories according topical content with only audio information was studied in [8]. Research into the domain of lecture videos has also been found in [9]. In their work, visual events are detected from the visual stream and then incorporated with audio information in a probabilistic framework to detect topic transitions. The domain of entertainment film has also been targeted lately [10–13]. In [10], for example, Adams *et al.* formulate an algorithmic solution for the computation of movie *tempo*, a high-level construct, and later utilise this function to segment a movie into story units. Wang *et al.* [12] attempt to detect *scenes* in film using the similarity in visual information and further improve the results with guidance from cinematic grammar.

The HHMM is a powerful stochastic model, first introduced in [14], in which the HHMM is viewed as a form of probabilistic context free grammar (PCFG), and the inference algorithm and parameter learning procedures are constructed based on the inside-outside algorithm. In [15], the HHMM is converted to a DBN, and applies general DBN inference to the model to achieve complexity linear in time $T$, but exponential in the depth $D$ of the model. The same analysis applied is [16], ie: the HHMM is 'flatened' into regular HMM with a very large state space for inference purpose. Their work [17, 16] aims to detect structures of soccer videos in an unsupervised manner. The model selection is first carried out using the MCMC to determine the structure parameters for the model, followed by a feature selection procedure. Finally, the HHMM is used to detect two semantic concepts, namely *play* and *break* in soccer videos. As there is little hierarchy at this level, the power of the hierarchic probabilistic model is not used. The HHMM is also applied in other domains other than multimedia such as in hand-written recognition [14], robot navigation [18], behaviour recognition [19] and information retrieval [20].

## 3 Model Definition and EM for the HHMM

The discrete HHMM and its extension to accommodate shared structured has been addressed in our previous work [21]. Here we refocus our attention to elucidate the idea of the shared structures and briefly discuss the EM algorithm for parameter estimation. We then discuss the case when the emission probability is modeled as mixture of Gaussian in the context of the hierarchical HMM.

A HHMM is formally defined by a topological structure $\zeta$ and a set of parameter $\theta$ attached to the topology. The general form of DBN representation is shown in Fig. 1(a). The depth $D$ and the number of states available at each level $Q^d$, for $d = 1, .., D$, are specified by $\zeta$. Level 1 is the root level and is always fixed to have only a single state. Furthermore, the topological structure reveals the 'parent-children' relationship of states between two consecutive levels[1]. A state $p$ at level $d$ is assigned to a set of children, ch$(p)$, at level $d+1$. A state $i$ at level $d+1$ therefore *might multiply inherit* from more than one parents at level $d$. For example, the HHMM defined in Fig. 2 has a depth $D = 3$ with $1, 3, 4$ are the number of states respectively at level $d = 1, 2, 3$. The set of children for state 3 at level 2 is $\{2, 3, 4\}$ (at lower level 3). The set of parents for state 2 at level 3 is $\{1, 2, 3\}$, which, in this case, is 'shared' by all states at level 2.

Given such a topological structure $\zeta$, the parameter $\theta$ of the HHMM is specified in the following way. For each level $d \in \{1..D-1\}$, $p \in Q^d$, $i, j \in$ ch$(p)$, where ch$(p)$ denote the children set of $p$:

- $\pi_i^{d,p} \triangleq \Pr(q_t^{d+1} = i \mid \bullet q_t^d = p)$ : is the initial probability of the child $i$ given the parent is $p$ at level $d$.
- $\mathrm{A}_{i,j}^{d,p} \triangleq \Pr(q_{t+1}^{d+1} = j \mid e_t^d = 0, q_t^{d+1} = i, q_t^d = p)$ : is the transition probability from child $i$ to child $j$ given that both are children of $p$.
- $\mathrm{A}_{i,\mathsf{end}}^{d,p} \triangleq \Pr(e_t^d = 1 \mid q_t^d = p, q_t^{d+1} = i)$: is the probability that state $p$ terminates at level $d$ given its current child is $i$.

where the dot in front of $q_t^d$ represents the event $q_{t-1}^d = 1$ (ie: $q_t^d$ is started at $t$), and the dot after $q_t^d$ represents the event $q_t^d = 1$ (ie: $q_t^d$ is ended at $t$). The constraints of stochastic processes requires that $\sum_i \pi_i^{d,p} = 1$, $\sum_j \mathrm{A}_{i,j}^{d,p} = 1$, and $\mathrm{A}_{i,\mathsf{end}}^{d,p} \leq 1$. Finally, at the lowest level $D$, an observation probability matrix $B$ is specified in the discrete observation case, or a set of $\{\mu_{im}, \Sigma_{im}\}$ are given when the observation values are continous and modeled as a mixture of Gaussians.

Given an observed data set $\mathcal{O}$ and some initial parameters, the EM algorithm iteratively re-estimates a new parameter $\hat{\theta}$, hill climbing in the parameter space which is guaranteed to converge to a local maxima. As shown in [21], doing EM parameter re-estimation reduces to first calculating the expected sufficient statistics (ESS) $\bar{\tau} = \mathrm{E}_{\mathcal{V} \setminus \mathcal{O}} \tau$, and then set the re-estimated parameter $\hat{\theta}$ to the normalized value of $\bar{\tau}$. The ESS for parameter $\left\{ \mathrm{A}_{i,j}^{d,p} \right\}$, for example, is calculated as:

$$\bar{\tau}(\mathrm{A})_{i,j}^{d,p} = \mathop{\mathrm{E}}_{\mathcal{V} \setminus \mathcal{O}} \tau(\mathrm{A})_{i,j}^{d,p} = \sum_{t=1}^{T-1} \xi_t^{d,p}(i,j) \Big/ \Pr(\mathcal{O}) \tag{1}$$

where the auxiliary variable $\xi_t^{d,p}(i,j)$ is defined as the probability $\Pr(q_{t+1}^{d+1} = j, q_t^{d+1} = i, q_{t+1}^d = p, e_t^{d:d+1} = 01, \mathcal{O})$. Readers are referred to [21] for further details on computation of the auxiliary variables and other expected sufficient statistics. In the rest of this section we will discuss the case when the emission probability is modeled as a mixture of Gaussians.

[1] Note that the original HHMM[14] assumes that a state has a only a single parent and therefore the topology reduces strictly to a tree.

In general, modeling the observation probability as a mixture of Gaussians for the hierarchical HMM is similar to the regular HMM. The DBN structure at level $D$ is modified as in Fig 1(b), where a mixture variable $z_t$ is added. For simplicity, thereafter in this section we will drop the index $D$. Let $M$ be the number of mixtures and $N$ be the num-

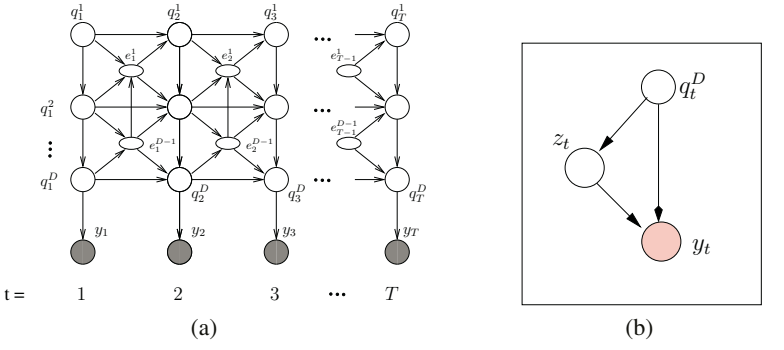



**Fig. 1.** (a) DBN representation for the discrete HHMM; (b) Mixture component $z_t$ at level $D$.

ber of states at level $D$. The observation matrix $B$ in the discrete case is replaced by the mixing weight matrix $\{\varepsilon_{im}\}$ and a set of means and covariance matrices $\{\mu_{mi}, \Sigma_{mi}\}$ for $i = 1, \ldots, N$ and $m = 1, \ldots, M$. Given observed data $\mathcal{O}$, expressing the expected complete log-likelihood and discarding terms irrelevant to $z_t$ and $y_t$ we have:

$$\langle \ell(\theta; \mathcal{O}) \rangle = \sum_{\substack{1 \le i \le N \\ 1 \le m \le M}} \left[ \sum_{t=1}^{T} \left\langle \mathrm{I}_{m,i}^{z_t, q_t} \right\rangle \log \mathcal{N}(y_t, \mu_{mi}, \Sigma_{mi}) + \sum_{t=1}^{T} \left\langle \mathrm{I}_{m,i}^{z_t, q_t} \right\rangle \log \varepsilon_{mi} \right]$$

where $\mathrm{I}_{m,i}^{z_t, q_t}$ is the identity function and $\triangleq 1$ if $\{z_t = m\} \cup \{q_t = i\}$ ; $\triangleq 0$ otherwise; and its expected value is calculated as:

$$\begin{aligned} \left\langle \mathrm{I}_{m,i}^{z_t, q_t} \right\rangle &= \Pr(z_t{=}m, q_t{=}i \mid \mathcal{O}) = \Pr(z_t{=}m \mid q_t{=}i, y_t) \Pr(q_t{=}i \mid \mathcal{O}) \\ &= \frac{\Pr(y_t \mid z_t{=}m, q_t{=}i) \Pr(z_t{=}m \mid q_t{=}i)}{\Pr(y_t \mid q_t{=}i)} \times \frac{\Pr(q_t{=}i, \mathcal{O})}{\Pr(\mathcal{O})} \\ &= \frac{\varepsilon_{mi} \mathcal{N}(y_t, \mu_{mi}, \Sigma_{mi})}{\sum_{m=1}^{M} \varepsilon_{mi} \mathcal{N}(y_t, \mu_{mi}, \Sigma_{mi})} \times \frac{\gamma_t^D(i)}{\Pr(\mathcal{O})} \end{aligned}$$

where[2] the auxiliary variable $\gamma_t^D(i)$ is defined as the probability $\Pr(q_t^D{=}i, \mathcal{O})$ and can be computed directly from horizontal transition probability $\xi_t^{d,p}(i,j)$ and vertical transition probability $\chi_t^{d,p}(i)$ (*see* [21]). Finally, maximising the expected complete log-likelihood $\langle \ell(\theta; \mathcal{O}) \rangle$ with respect $\varepsilon_{im}$ and $\mu_{mi}, \Sigma_{mi}$ respectively. Introducing the Lagrange multipliers for $\varepsilon_{im}$; and setting derivatives to zero for the case of $\mu_{mi}, \Sigma_{mi}$. The set of re-estimated parameters is given as:

[2] We put back hierarchic index $D$ for clarity here.

$$\hat{\varepsilon}_{mi} = \frac{\sum_{t=1}^{T} \left\langle \mathrm{I}_{m,i}^{z_t,q_t} \right\rangle}{\sum_{m=1}^{M} \sum_{t=1}^{T} \left\langle \mathrm{I}_{m,i}^{z_t,q_t} \right\rangle}, \qquad \hat{\mu}_{mi} = \frac{\sum_{t=1}^{T} \left\langle \mathrm{I}_{m,i}^{z_t,q_t} \right\rangle y_t}{\sum_{t=1}^{T} \left\langle \mathrm{I}_{m,i}^{z_t,q_t} \right\rangle}$$

$$\hat{\Sigma}_{mi} = \frac{\sum_{t=1}^{T} \left\langle \mathrm{I}_{m,i}^{z_t,q_t} \right\rangle (y_t - \mu_{mi})(y_t - \mu_{mi})^{\mathsf{T}}}{\sum_{t=1}^{T} \left\langle \mathrm{I}_{m,i}^{z_t,q_t} \right\rangle}$$

When multiple observation sequences are given, the set of above equations can be adjusted by simply adding a summation over the number of sequences. This corresponds to 'counting' over all sequences.

## 4 Elucidating Structures in Educational Videos

An intrinsic functionality of educational videos[3] is to 'teach' [22], and therefore structuralizing the content and building meaningful indices are important to improve the learning experience. Materials delivered in an educational video might vary widely to suit different purposes; however, when restricted to instructional and safety videos, the content organization is relatively simple. In this paper, we are interested in this particular type of video and observe that linear presentation is generally chosen to present the content. Subjects are arranged into a sequence of topics started with a few introduction shots. Literature in this field [22] offers further insight into how a topic is constructed. Generally, there are three presentational styles: (1) *direct* instruction, (2) *on-screen* instruction, and (3) *illustrative* instruction (*see* Fig. 2). In *direct* instruction, the videomaker choose to present a topic by means of text captions and voice over. In *on-screen* instruction, s/he decides to directly appear on the camera to talk directly to the viewers. Lastly in *illustrative* instruction, illustrative examples are the major mode of presentation to convey the subject with possible appearance of the anchors. These presentational styles shares certain similar semantic concepts at the lower levels such as introduction shot, or direct appearance of anchor(s) (Fig. 2).

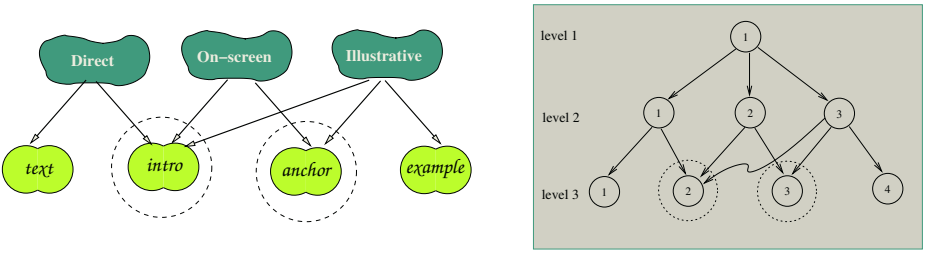



**Fig. 2.** Structure of topic generating process with assumed hidden 'styles'; and its mapping to a topology for the HHMM. Shared structures are identified with extra dotted circle.

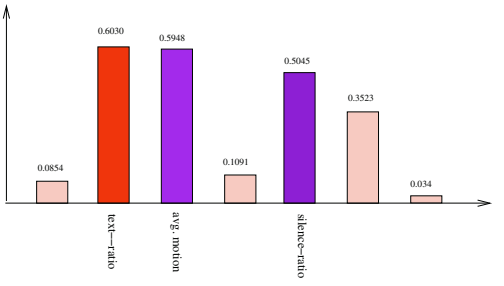
[3] The class of educational videos discussed in this work is of professional productions, excluding hand-held recorded videos such that lecture videos recorded in the classroom.

## 5 Experimental Results

Our aim is to apply the HHMM model, taking advantage of the shared units, to segment an educational video into high-levels of abstraction – ie: detection of topic transitions in this case. We construct a 3-level HHMM as follows. The root level represents the entire video, followed by three states at the next level, each of which corresponds to one topic presentational style. Our assumption here is that the topic content organisation strictly follows the styles outlined in Sec. 4. The production level includes four states, corresponding to four semantic concepts at the shot level: (1) the introduction, (2) instruction delivered by mean of captioned texts, (3) instruction delivered directly by the presenter and (4) illustrative example (Fig.2). Given a set of training $N$ videos, we extract features from each video and use them as input observation sequences to the EM parameter learning algorithm to estimate a new model parameter. This new parameter set will be used in the second phase to segment a video based on results from the generalised Viterbi decoding algorithm. The data set includes eight instructional and safety videos, whose topics span a variety of subjects such as how to exercise safety at home, in office, or at workplace. Shot indices are assumed to be available, which is first detected by a commercial software and errors are manually corrected. In the training phase, each video yields an observation sequence with each shot-based feature vector $\mathbf{o}$ is a column vector of seven elements $\mathbf{o} = [o_1, o_2, o_3, o_4, o_5, o_6, o_7]^t$. From the visual stream, we extract three features including the face-content-ratio, text-content-ratio and average motion based on camera pan and tilt $(o_1, o_2, o_3)$. The other four features namely music-ratio, speech-ratio, silence-ratio and non-literal sound ratio $(o_4, o_5, o_6, o_7)$ are from the audio track. Feature music-ratio, for example, is calculated as the ratio of number of clips classified as music to the total number of audio clips in the shot. Readers are referred to [23] for further details on the computation of these features.

At the production level of the trained model, the estimated matrices $\hat{\mu}$ and $\hat{\Sigma}$ can be examined to get an idea about the semantics . Fig. 3(a), for instance, shows the estimated mean value for different features with respect to state 2, which is intended to model the 'style' of shots that used to introduce a new topic. As can be seen, this state is 'sensitive', ie: will yield a high probability, to shots with displayed captioned texts and no audio. When compared with the ground-truth, we observe that this is indeed a major kind of shots that demarcate topics.

To evaluate the detection performance, we manually watch and segment each video into topics. In some cases, this information is available directly from the video manuals. This results in a total of 75 indices. We use two well-known metrics, namely, recall and precision to measure the performance of the detection. To perform segmentation, we first run the Viterbi algorithm to get the time indices for which a state at topic level (ie: level 2) make the transition. Let $\tau$ be such an index, we then examine the state $x_\tau^3$, which is the corresponding state at the production level being called. If this state coincides with the introduction shot (ie: $= 2$), then $\tau$ is recorded as a topic transition. The entire segmentation results are reported in Fig. 3. Calculation yields a recall of $77.3\%$ and a precision of $70.7\%$. Given that the segmentation has been done in an completely unsupervised manner, ie: there is no hints in the training data as to what is a topic boundary, the result demonstrate the validity of the HHMM-based detection scheme.




(a) $\hat{\mu}_{21}$

| Video | GT | TP | Errors | |
|---|---|---|---|---|
| | | | FN | FP |
| 1. | 10 | 8 | 2 | 4 |
| 2. | 9 | 6 | 3 | 3 |
| 3. | 8 | 7 | 1 | 0 |
| 4. | 5 | 3 | 2 | 4 |
| 5. | 7 | 6 | 1 | 10 |
| 6. | 19 | 15 | 4 | 2 |
| 7. | 10 | 7 | 3 | 0 |
| 8. | 7 | 6 | 1 | 1 |
| **Total** | **75** | **58** | **17** | **24** |

(b) detection results

**Fig. 3.** (a) estimated $\mu$ vector for state 2 (introduction shot), (b) detection results for 8 videos – GT : number of ground-truth indices, TP : number of correct detection, FN : number of miss, FP: number of over segmented indices.

This result is comparative with the probabilistic detection framework developed in [24] with a slight degradation in performance.

## 6 Discussion

Fig. 3 reveals that over segmentation is the major source of error causing a degradation in precision; and the high number of 'miss' (false negatives) causing a low recall rate. The resulting false negatives is not surprising since a topic is introduced in numerous ways, but the estimated model has learned only a subset of these methods of introduction. To overcome this, we obviously need a more complex model structure, which will be considered in our future work. The fact that the detector usually over segments a video (eg: video 5) is worth further discussion. A close analysis discloses that while these (over-segmenting) indices do not match the ground-truth, they frequently map to the lower level of sub-structures within a topic such as segments emphasising a safety message (for example, this happens many times in video 5). Fig. 4 draws an insight into the structure of an educational video and illustrate this problem. The vertical solid lines have been the target of our detection, while dashed-lines correspond to the over-segmented indices from the detector. This fact suggests that the model might be utilised to exploit further structure in topics, which will also be considered in our future work.

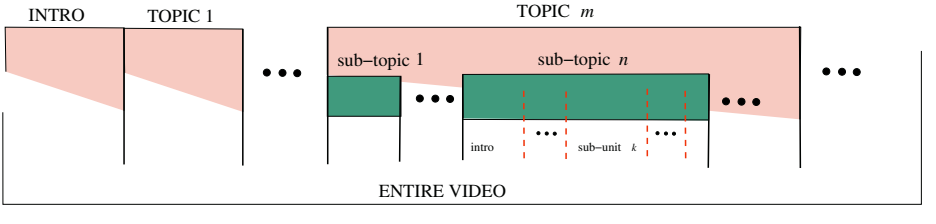



**Fig. 4.** Structure of a Video.

## 7 Conclusion

We have presented a framework for topic structure discovery using the HHMM in this paper. An important aspect of video data when considering its content organisation is the shared structures embedded in the data. This suggests a natural mapping to the hierarchical HMM, which we have utilised to model the topic structures in educational videos. We have briefly addressed the issue of parameter learning in the HHMM, in particular when the emission probability is modeled as a mixture of Gaussians. Finally, the experimental results have demonstrated the usefulness of the detection scheme.

## References

1. Ariki, Y., Shibutani, A., Sugiyama, Y.: Classification and retrieval of TV Sports News by DCT features. In: IPSJ International Symposium on Information System and Technologies for Network Society. (1997) 269–272
2. Shearer, K., Dorai, C., Venkatesh, S.: Incorporating domain knowlege with video and voice data analysis (2000) MDM/KDD 2000, Workshop on Multimedia Data Minning, Aug 20-23, Boston, USA.
3. Bertini, M., Bimbo, A.D., Pala, P.: Content based annotation and retrieval of news videos. In: International Conference on Multimedia and Expo. (2000) 479–482
4. Eickeler, S., Müller, S.: Content-based video indexing of TV broadcast news using Hidden Markov Model. In: Proceedings of IEEE International on Acoustics Speech and Signal Processing. Volume 6., Phoenix (1999)
5. Huang, Q., Liu, Z., Rosenberg, A.: Automated semantic structure reconstruction and representation generation for broadcast news. In: Proc. IS&T/SPIE Conference on Storage and Retrieval for Image and Video Databases VII. Volume 3656. (1999) 50–62
6. Liu, Z., Huang, J., Wang, Y.: Classification of TV programs based on audio information using hidden markov model. In: IEEE Signal Processing Society 1998 Workshop on Multimedia Signal Processing. (1998) 27–32
7. Liu, Z., Huang, Q.: Detecting news reporting using audio/visual information. In: International Conference on Image Processing, Kobe, Japan (1999) 24–28
8. Walls, F., Jin, H., Sista, S., Schwartz, R.: Topic detection in broadcast news. In: Proceedings of the DARPR Broadcast News Workshop. (1999) 193–198
9. Seyeda-Mahmood, T., Srinivasan, S.: Detecting topical events in digital video. In: ACM Multimedia. (2000) 85–94
10. Adams, B., Dorai, C., Venkatesh, S.: Novel approach to determining movie tempo and dramatic story sections in motion pictures. In: 2000 International Conference on Image Processing,. Volume II., Vancouver, Canada (2000) 283–286
11. Adams, B., Dorai, C., Venkatesh, S.: Role of shot length in characterizing tempo and dramatic story sections in motion pictures. In: IEEE Pacific Rim Conference on Multimedia 2000, Sydney, Australia (2000) 54–57
12. Wang, J., Chua, T.S., Chen, L.: Cinematic-based model for scene boundary detection. In: The Eighth Conference on Multimedia Modeling, Amsterdam, Netherland (2001)
13. Sundaram, H., Chang, S.F.: Video scene segmentation using audio and video features. In: International Conference on Multimedia and Expo. (2000)
14. Fine, S., Singer, Y., Tishby, N.: The hierarchical hidden markov model: Analysis and applications. Machine Learning **32** (1998) 41–62
15. Murphy, K., Paskin, M.: Linear time inference in hierarhical hidden markov models. In: Proceedings of Neural Information Processing Systems, Vancouver, Canada (2001)

16. Xie, L., Chang, S.F., Divakaran, A., Sun, H.: Unsupervised discovery of multilevel statistical video structures using hierarhical hidden markov models. In: IEEE International on Multimedia and Expo, Baltimore, USA (2003) III.29 – III.32
17. Xie, L., Chang, S.F., Divakaran, A., Sun, H.: Learning hierarhical hidden markov models for unsupervised structure discovery from video. Technical report, Columbia University, New York (2002)
18. Theocharous, G., Mahadevan, S.: Learning the hierarchical structure of spatial environments using multiresolution statistical models. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). (2002)
19. Luhr, S., Bui, H.H., Venkatesh, S., West, G.: Recognition of human activity through hierarchical stochastic learning. In: International Conference on Pervasive Computing and Communication (PERCOM-03). (2003)
20. Skounakis, M., Craven, M., Ray, S.: Hierarchical hidden markov models for information extraction. In: Proceedings of the Eighteen International Joint Conference on Artificial Intelligence (IJCAI-03). (2003)
21. Bui, H.H., Phung, D.Q., Venkatesh, S.: Hierarchical hidden markov models with general state hierarchy. In: *to appear in* The Nineteenth National Conference on Artificial Intelligence (AAAI-04), San Jose, California USA (2004)
22. Herman, L.: Educational Films: Writing, Directing, and Producing for Classroom, Television, and Industry. Crown Publishers, INC., New York (1965)
23. Phung, D.Q., Venkatesh, S., Dorai, C.: On extraction of thematic and dramatic functions in educational films. In: IEEE International Conference on Multimedia and Expo, Baltimore, New York, USA (2003) 449 – 452
24. Phung, D.Q., Dorai, C., Venkatesh, S.: High level segmentation of instructional videos based on the content density function. In: ACM International Conference on Multimedia, Juan Les Pins, France (2002) 295–298