

ICA Mixture Hidden Conditional Random Field Model for Sports Event Classification

Xiaofeng Wang, Xiao-Ping Zhang

Dept. of Electrical and Computer Engineering, Ryerson University
350 Victoria Street, Toronto, Ontario, Canada, M5B 2K3

xwang, xzhang@ee.ryerson.ca

Abstract

In this paper, a new hidden conditional random field (HCRF) model with independent component analysis (ICA) mixture feature functions is developed for sports event classification. Unlike Hidden Markov Model (HMM), HCRF is a discriminative model without conditional independence assumption of observations, which is more suitable for video content analysis. According to the non-Gaussian property of sports event features, a new feature function using the likelihood of ICA mixture component is proposed to further enhance the HCRF model. The discriminant power of HCRF and representation power of ICA mixture for non-Gaussian distribution are combined. The new model is applied to challenging bowling and golf event classification. The simulation results prove our analysis that the new ICA mixture HCRF outperforms the existing mixture HMM models in term of classification accuracy.

1. Introduction

In the past several years, the field of multimedia content and computer vision analysis is dominated by content-based image and video indexing and retrieval systems. But most systems have limited performance by using only a few low level features such as color, texture, shape, and motion. The reason is that there is a huge semantic gap between low level features and high level semantic meanings. The videos have rich structural information that could be explored for the usage of indexing and retrieval. The video content analysis is to find meaningful structure and patterns from visual data for the purpose of efficient indexing and mining of videos. Video analysis tasks include video parsing, content indexing and abstraction, and representation. The video parsing is to segment video to different levels of segments. The early works focus on low level parsing, the video shot boundary detection [18][5]. After the shots are segmented, the next task is to classify these shots to differ-

ent categories. It is to label video shots and give meaningful names to these shots. The video event classification is to classify shots to different events. The event classification in sports video [3][5] is a popular research topic in recent years which could be addressed by using probabilistic graphical models such as Hidden Markov Model (HMM) and Hidden Conditional Random Field (HCRF) model.

The HMM is widely used in many video analysis algorithms [16][8][1][2][9]. In [16], unsupervised classification based on color ratio and motion in soccer domain is discussed and the observation model is Gaussian mixture. In [8], the audio features such as applause and cheering are modeled as HMM. In [1], baseball highlights are modeled as HMM using various kinds of features. It is extended to maximum entropy model [2] which puts several shot features together for classification and does not use the useful temporal graph information. Hierarchical HMM presented in [9] is a more complex HMM model. In [20], based on the non-Gaussian property of visual features the ICA mixture [15] observation model is applied in HMM for golf video event classification. As mentioned in the HMM tutorial paper [11], there are certain limitations of HMM, the conditional independence of observations, the form of observation distribution and the Markov chain interaction.

The CRF [7] is first introduced by Lafferty to address limitations of HMM. It is successfully used in applications such as speech [7], image labeling [6] and object recognition [13]. But the fully labeling of training sequence states prevents it from applying to event analysis. To solve this problem HCRF is proposed by Quattoni recently in [10]. HCRF is a general extension of HMM which relaxes the independent observation and generative assumption. It is applied to phone classification [4], gesture recognition [14] and meeting segmentation [12]. In this paper we present a new HCRF model for sports events classification.

To bridge the semantic gap of indexing and retrieval we present a new ICA mixture HCRF (ICAMHCRF) model for sports video analysis. The new model takes advantage of discriminant power of HCRF and the representation power

of nonstructural ICA mixture model. In [4] and [12], the Gaussian mixture equivalent is employed in HCRF. Since the non-Gaussian properties of the observation distributions such as sports video features, it is more suitable to use the ICA mixture model. The likelihoods of ICA mixture components are used as feature functions in the new model. To test the effectiveness of the new model, it is applied to bowling and golf event detection. The new model gets better performance of both event classification tasks, which proves our analysis. The ICAMHCRF model is better than existing HMM models in sports video event classification.

This paper is organized as follows. First, a brief introduction of CRF and HCRF is given in section 2. Then we introduce the new HCRF model based on ICA mixture local observation in section 3. After that in section 4 the new system of the video content analysis using the ICAMHCRF model is presented and the comparison between HCRF and HMM is discussed. In section 5, the new model is applied to two kinds of sport video analysis, bowling and golf and the performance is given. We conclude the paper with summaries and future research directions in section 6.

2. An Introduction of Hidden Conditional Random Field Model

HCRF Model is a relative new graphical model for event and object recognition. It is an extension of CRF model.

2.1. Formulation of CRF

The CRF model as shown in Fig. 1 is a commonly used structural statistical model in computer vision areas such as image labeling. The dash circles are the observed features at the node. The empty circles are the labels, which are unknown for the testing data and known for the training data. The interactions between these random variables are shown as edges in the figure. It is a graphical model defined on

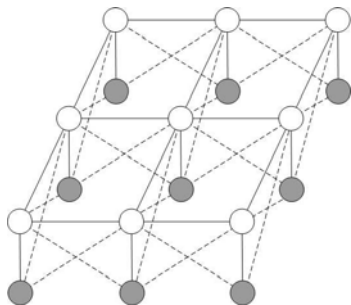


Figure 1. An example of 2D CRF model.

an undirected graph $G(V, E)$ with vertexes V and edges E . We denote $\mathbf{x} = \{x_i\}_{i \in V}$ the observations from the input

data and V a set of sites which could be pixels in image labeling and frames in video analysis. The observation at the site i , $x_i \in \mathcal{X}$ is a vector of observation features. The set \mathcal{X} is all possible observations. The features could be any meaningful filter response of the site such as color, texture and shape. The observation variable \mathbf{x} has an associated labels $\mathbf{h} = \{h_i\}_{i \in V}$ where $h_i \in \mathcal{H}$ is the label for site i . The set of all possible labels is \mathcal{H} . The labeling problem is to infer the underlying labels \mathbf{h} given the image features \mathbf{x} and parameters of the model. The probabilistic expression of the problem is $P(\mathbf{h}|\mathbf{x})$. A CRF defines a conditional probability of the following form,

$$P(\mathbf{h}|\mathbf{x}, \boldsymbol{\theta}) = \frac{1}{Z(\mathbf{x})} \exp\{\Psi(\mathbf{h}, \mathbf{x}; \boldsymbol{\theta})\}, \quad (1)$$

$$Z(\mathbf{x}; \boldsymbol{\theta}) = \sum_{\mathbf{h}'} \exp\{\Psi(\mathbf{h}', \mathbf{x}; \boldsymbol{\theta})\}, \quad (2)$$

$$\Psi(\mathbf{h}, \mathbf{x}; \boldsymbol{\theta}) = \sum_{i \in V} \varphi(h_i, \mathbf{x}) + \sum_{(i,j) \in E} \psi(h_i, h_j, \mathbf{x}), \quad (3)$$

where $\boldsymbol{\theta}$ is the parameter vector of the model, φ is the association potential between the observation data and the label of site i , and ψ is the interaction potential between current site i and its neighboring site j given the observed features. The function Z is a normalization factor. Usually as in Fig. 1, φ represents the prediction of the label h_i based on the local feature vector x_i at site i and ψ predicts the label h_i based on local compatibility between neighboring labels and features. For example, CRF image labeling is to infer the most probable labels given an input image and the model parameters, which are learned from the training set. The video event detection is to estimate the probability of a label for a whole sequence. But CRF needs a label for every node, which prevents it from applying to video content analysis directly.

2.2. Hidden Conditional Random Field Model

For video event detection usually the states are hidden. It is a troublesome work to label all states manually. The same as HMM widely used in event detection, the HCRF is developed for object recognition. The formulation is as follows,

$$P(y|\mathbf{x}, \boldsymbol{\theta}) = \frac{1}{Z(\mathbf{x}; \boldsymbol{\theta})} \sum_{\mathbf{h}} \exp\{\Psi(y, \mathbf{h}, \mathbf{x}; \boldsymbol{\theta})\}. \quad (4)$$

Here $y \in \mathcal{Y}$ is a label for a whole sequence and \mathcal{Y} is a set of all possible labels. For example the binary event detection problem, $\mathcal{Y} = \{-1, 1\}$, 1 represents the existence of the event and -1 nonevent. Unlike the node labels \mathbf{h} in CRF, The unknown hidden states \mathbf{h} is summed out in the equation. The normalization factor is,

$$Z(\mathbf{x}; \boldsymbol{\theta}) = \sum_{y'} \sum_{\mathbf{h}} \exp\{\Psi(y', \mathbf{h}, \mathbf{x}; \boldsymbol{\theta})\}, \quad (5)$$

where y' is a possible label for the sequence.

Here we choose a restricted form of function Ψ in the following form,

$$\begin{aligned}\Psi(y, \mathbf{h}, \mathbf{x}; \boldsymbol{\theta}) &= \sum_{i \in V} \varphi(y, h_i, \mathbf{x}) + \sum_{(i,j) \in E} \psi(y, h_i, h_j, \mathbf{x}) \\ &= \sum_{i \in V} \phi(y, x_i) \theta(h_i) + \sum_{i \in V} \theta(y, h_i) \\ &\quad + \sum_{(i,j) \in E} \theta(y, h_i, h_j),\end{aligned}\quad (6)$$

where $\phi(y, x_i)$ is an observation feature function with the label y at site i , $\theta(h_i)$ is a parameter vector for associate potential of the hidden state h_i , $\theta(y, h_i)$ is a compatibility indicator of the sequence label and the hidden state h_i and $\theta(y, h_i, h_j)$ is a compatibility indicator of the label and the interaction edges. The first two terms are a simplification of φ and the third term is an implementation of ψ in Equation (3). HCRF training is to learn the parameters $\boldsymbol{\theta} = [\theta(h_i) \ \theta(y, h_i) \ \theta(y, h_i, h_j)]$ and the inference is to find the label for a given input using these parameters.

3. ICA Mixture Hidden Conditional Random Field Model

In traditional form, the feature vector x_i is directly used as the observation function $\phi(y, x_i)$ at site i in HCRF. It usually includes hundreds of features which makes the learning process slow and the algorithm may not find the optimal value in a reasonable time period. In image analysis, mixture models are widely used nonstructural classifiers. The use of mixture models as observation functions for HCRF is not widely investigated except the Gaussian mixture mentioned in [4][12]. We develop a new ICA mixture HCRF (ICAMHCRF) model for video event classification.

The observation feature function in previous section is a feature vector or feature statistics vector $\phi(y, x_i) = [f_1(y, x_i) \dots f_d(y, x_i) \dots f_D(y, x_i)]$ which is weighted by parameter vector $\theta(h_i) = [\theta_1(h_i) \dots \theta_d(h_i) \dots \theta_D(h_i)]$ in HCRF model. Here D is the total number and d is the index of the feature function. The functions f could be features themselves or functions of features. Note that we only consider the local observation x_i and the sequence label y in feature functions. In this paper we define f as log likelihood of the feature x_i belonging to a mixture model component. Since the log likelihood carries certain probabilistic meanings, the function could better reflect the local observation model. Gaussian and Gaussian mixtures are commonly used kernel functions. Mixture model is more expressive than non mixture model. Mixture means the observation could be divided into mutual exclusive components. It is true for video frames since they are often in interlaced manner. Though mixture of Gaussian could approach nearly

any distribution, if the observation is non-Gaussian, Gaussian generally needs more mixture components to approach the distribution. In our case video frame observation features usually show non-Gaussian characteristics. It is also approved by our experiment. The distribution can be decomposed into ICA mixture components. The ICA mixture distributed feature could be expressed as follows,

$$x_i = \mathbf{M}_k \mathbf{S}_k + \boldsymbol{\mu}_k \quad x_i \in C_k, \quad (7)$$

where \mathbf{M}_k is mixing matrix, $\boldsymbol{\mu}_k$ is bias and \mathbf{S}_k is independent sources for k th component of mixture. We denote C_k the k th component of the mixture. The probability of observation x_i could be expressed as follows,

$$\begin{aligned}p(x_i|y) &= \sum_{k=1}^K p(x_i|y, C_k) p(C_k|y) \\ &= \sum_{k=1}^K p(C_k|y) \exp(\log p(\mathbf{S}_k) - \log(\det \mathbf{M}_k)).\end{aligned}\quad (8)$$

The log likelihood of each observation belongs to a mixture component is chosen as feature function,

$$f_d(y, x_i) = \log p(C_k|y) p(x_i|y, C_k). \quad (9)$$

Here x_i is represented by $p(x_i|y)$ locally and the feature functions are computed using components of mixtures. And k is the index of the mixture component and K is the number of components. Note that the number of feature functions D could be more than the number of components K . The reason is we could have several groups of features in feature vector x_i such as color and texture and each feature group has K feature functions. In our experiment, only one group of feature is used so the number of feature functions D is equal to the number of mixture components. The probability $p(C_k|y)$ is mixture coefficient for certain component. During training the class label y is given, the parameters of ICA, \mathbf{S}_k , \mathbf{M}_k and $\boldsymbol{\mu}_k$ and mixture components $p(C_k|y)$ and $p(x_i|y, C_k)$ are learned by modified standard ICA algorithm as in [15].

The new ICAMHCRF model derived in this paper provides a new way to model both the local and temporal interaction for sequence labeling task. ICA mixture is used in both the training and testing. Unlike the Gaussian assumption in other works [4][12], we use a non Gaussian model as local feature function for observations. The new function is a better representation of complex feature distribution such as video frames. Since the real scene such as sport video consists of non Gaussian components which could not be represented by Gaussian mixture with reasonable number of components, the proposed ICA mixture feature function is more suitable for video content analysis. When ICA mixture is combined with HCRF model, it could adapt to spatial and temporal probabilistic structure of the data simultaneously.

4. ICA Mixture Hidden Conditional Random Field Model for Sports Event Detection

The task of video content analysis is to assign the chunks of digital video data to content categories such as sports highlight, news anchor and snow mountain landscape. Graphical models such as HMM and HCRF could be applied to address this problem. Unlike gesture and meeting segmentation in which the backgrounds are simple, sports events with real scene setting are more challenging. But the sports video also consists of a set of predefined actions in a certain order which fits the requirement of linear chain model, HMM and HCRF.

4.1. A Comparison between HMM and HCRF Event Detection

As shown in Fig. 2, for video event detection using HMM, a specific model should be set up for each specific event y . For example the golf event detection, there are three events, the full swing, the non-full swing and other events ($y \in \mathcal{Y} = \{1, 2, 3\}$). There are three models corresponding to three events. During the training, the parameters are learned for each model. The class label for the testing sequence is inferred by finding the most probable model for the sequence. In HCRF video event analysis, there is only one model and the weights of different factors serve as the discriminate coefficients to classify the sequences. During the training the weights θ is learned from training sequences. It is used to label the events in the testing process. There are several differences between two model. First, there are direct links between y and hidden states sequence in HCRF, while HMM does not have this useful structure. Second, links of HMM have direction. This is the generative nature of the model. Observations are children of the states. They are generated by the states. So fully observation is needed for the training. The HCRF relaxes this assumption by getting rid of the direction of the links. Third, in HMM the observations are independent and only depend on their own state. HCRF could have links between current observation and other states beside its current state. Note that in our experiment only the link between the local observation and current state is used. These differences set HCRF in a better position to deal with complex video event detection task.

4.2. Steps of ICAMHCRF for Event Detection

The new ICAMHCRF video event detection system is shown in Fig. 3. The training and testing are as follows. We divide the video to shots as preprocessing step using shot boundary detection technique. In the training part, the features of the shot frames are extracted. The ICA based dimension reduction is used to reduce the computational complexity. After that the compact features are modeled as ICA

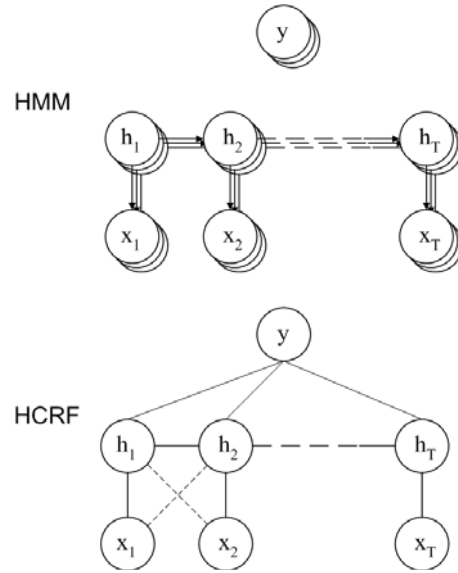


Figure 2. An illustration of HMM and HCRF model structure. Here T is the length of the sequence.

mixture which is done by using ICA algorithm in [15] with Laplacian source assumption. The ICA mixture parameters are learned. We compute the log likelihood of each feature or feature group belonging to a mixture component as the feature function in HCRF. The HCRF parameters are learned based on the following maximum likelihood objective function,

$$\mathcal{L}(\mathcal{T}, \theta) = \sum_{(x, y) \in \mathcal{T}} \mathcal{L}(\theta | x, y) - \frac{\|\theta\|^2}{2\delta^2}, \quad (10)$$

$$\begin{aligned} \mathcal{L}(\theta | x, y) &= \log P(y | x; \theta) \\ &= \log \frac{\sum_{\mathbf{h}} \exp\{\Psi(y, \mathbf{h}, \mathbf{x}; \theta)\}}{Z(\mathbf{x}; \theta)}. \end{aligned} \quad (11)$$

Here \mathcal{T} is the training data set. The objective function $\mathcal{L}(\mathcal{T}, \theta)$ is the summation of log-likelihood of all training data minus regulation factor. The first term $\mathcal{L}(\theta | x, y)$ is the log-likelihood of one training data belonging to the model with parameter θ . The second term is a regulation factor which we suppose the parameters are Gaussian distributed with variance δ^2 . There are several method available for HCRF training. The gradient descent method [10][14] is used in the experiment. The optimal parameter θ^* is,

$$\theta^* = \arg \max_{\theta} \mathcal{L}(\mathcal{T}, \theta). \quad (12)$$

In the testing part, the compact features are computed and log likelihood feature functions are calculated using the parameters from ICA mixture learning in the training part.

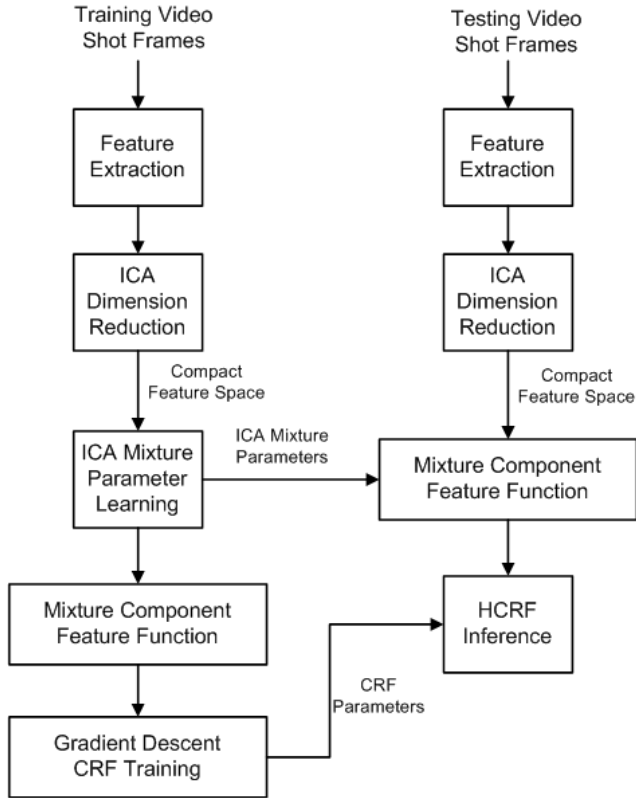


Figure 3. The flowchart of ICAMHCRF model for video event classification.

Then using belief propagation method [17] the most probable class label y^* of the testing sequence is inferred as follows,

$$y^* = \arg \max_{y \in \mathcal{Y}} P(y|x, \theta^*). \quad (13)$$

5. Simulation Results

The new ICAMHCRF system is applied to two kinds of video content analysis tasks. One is bowling activity recognition and the other is golf video event analysis. It is compared with ICAMHMM and Gaussian mixture HMM (GMHMM) in both cases. Note choosing the number of mixture components of ICA mixture and the number of hidden states of HMM and HCRF are both non-trivial. However both of them could be optimized using training or validation set. In our experiment, the numbers are initially chosen in the range from 2 to 4 and the numbers, which maximize the classification accuracy in both methods, are selected. It is also worth to mention that because of the existence of hidden states, the optimization is no longer convex. We choose the best result with random parameter initialization. The videos are first segmented to shots as in [19] and then ICA is applied to 256 illumination-invariant color his-

toграм of frame features to reduce the feature vector to 2 dimension for each frame.

5.1. Bowling Activity Detection

ICAMHCRF is tested to recognize the bowling shot event and compared with ICAMHMM and GMHMM. A 30 minutes professional bowling TV program is used in our experiment. The video is divided to 232 video shots. There are 65 bowling shot event and the others are comments, commercials, player's preparation and players after the shot. An example bowling shot event sequence is shown in Fig. 4 (a), which features following activities, bowler preparing to release his ball toward the pins, bowler dropping the ball on the lane, the ball striking the pins and finally the camera turning back to the player. The ICA mixture parameters and HCRF parameters are learned from one training shot of both events. Two hidden states and two mixture components are used in the experiment. The two mixture components of an bowling shot event is shown in Fig. 5. The final event detection receiver operating characteristic (ROC) curve is shown in Fig. 6. The confusion matrix and the detection accuracy are given in Table 2 and Table 1. The accuracy is defined as the ratio between the correctly labeled event and the total number of events. The performance of ICAMHCRF is about 7% better than the ICAMHMM and 9.5% better than the GMHMM.

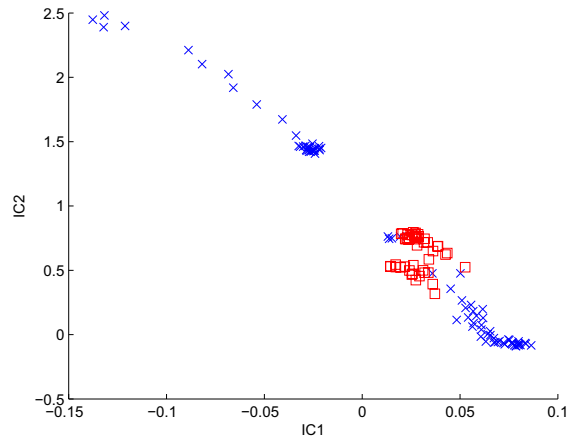


Figure 5. Two ICA mixture components for bowling shot event. The axes are two features of compacted feature space.

Method	Accuracy
ICAMCRF	85.28%
ICAMHMM	78.45%
GMHMM	75.86%

Table 1. Classification accuracy of bowling event classification.



Figure 4. An example selected frames of bowling events. There are two events, (a) bowling shot event and (b) other event (an advertisement event is shown).

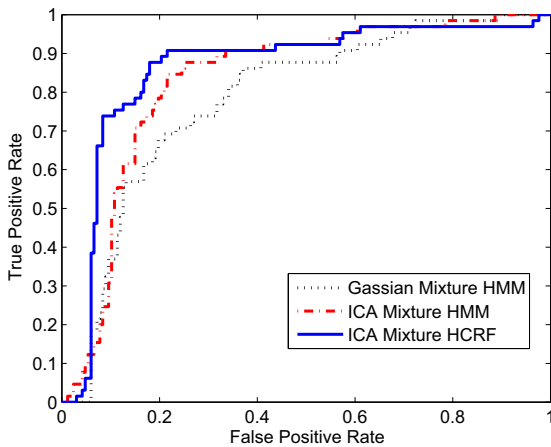


Figure 6. ROC performance of bowling shot classification.

	bowling shot	other event
bowling shot	49 (36)	16 (29)
other event	18 (21)	149 (146)

Table 2. The confusion matrix for bowling event classification using the ICAMHCRF. Row labels are the true classes and column labels the predicted classes. The ICAMHMM results are shown in parentheses for comparison.

5.2. Golf Event Classification

For golf video event detection, an hour professional golf video from the authors of [20] is used. The same procedure as bowling is applied except that one event consists

Method	Accuracy
ICAMCRF	73.28%
ICAMHMM	70.79%
GMHMM	56.93%

Table 3. The classification accuracy of golf event classification.

three shots for fair comparison with ICAMHMM and better representation of golf event. Three example events are shown in Fig. 7. These three events are used for training the model. The total number of events is 202. It is manually annotated to three categories, full swing, non-full swing and other irrelevant events. The event is very recognizable with recurrent patterns as in Fig. 7. The golf shot includes activities as follows. Player prepares for the shot. Then player hits the ball. After that the camera follows the ball quickly with high activity. Finally the scene features the golf court and/or players with low activity. The mixture components of a full swing shot are displayed in Fig. 8. The confusion matrix is shown in Table 4. The ICAMHCRF is better than ICAMHMM in both full swing and non-full swing classification. For the other event it is not better than ICAMHMM. The only one training sample from the other event may not be representative for the other event class. It is used for fair comparison with the results of ICAMHMM in paper [20]. We expect better result when more training samples are used. The overall accuracy of ICAMHCRF is 2.5% better than ICAMHMM as shown in Table 3.

5.3. Discussion

In both event classification cases, the new ICAMHCRF exhibits higher classification accuracy than HMM models.

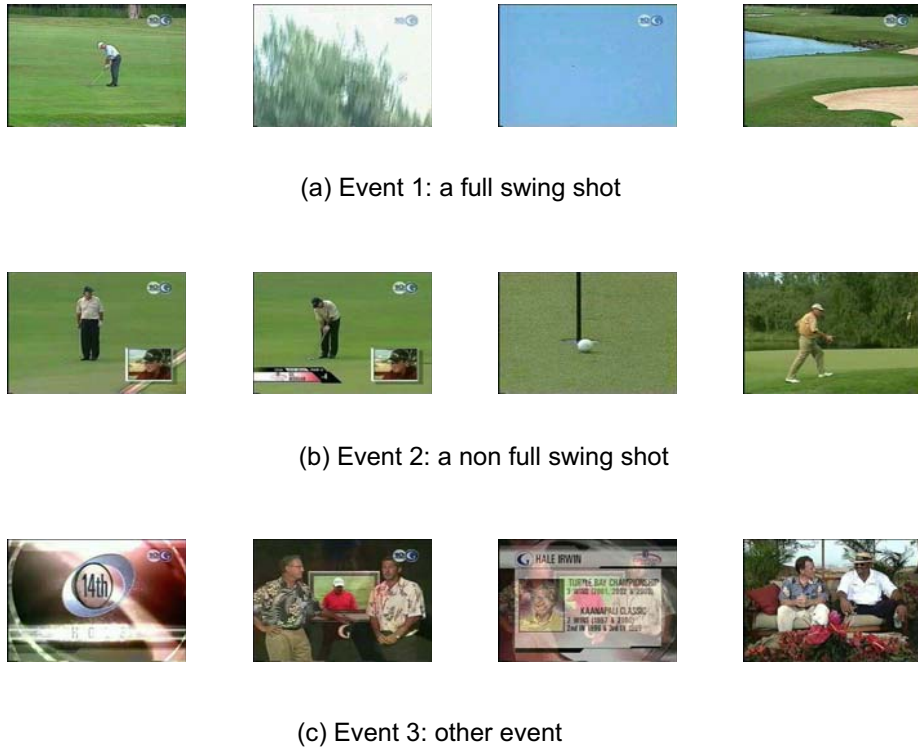


Figure 7. An example selected frames of golf events. There are three events, (a) full swing, (b) non-full swing, (c) other event.

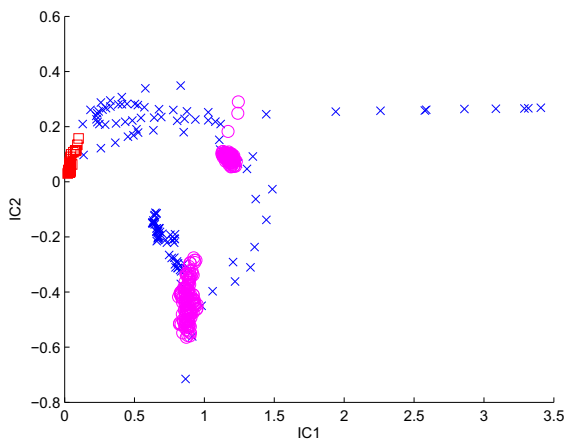


Figure 8. Three ICA mixture components of full swing event in golf video. The axes are two features of compacted feature space.

There are two main reasons. First, the ICA mixture can approach the non-Gaussian distribution of compacted fea-

	full swing	non full swing	others
full swing	33 (26)	20 (27)	1 (1)
non full swing	16 (23)	109 (104)	7 (5)
others	2 (2)	8 (1)	6 (13)

Table 4. The confusion matrix for golf event classification using the ICAMHCRF. Row labels are the true classes and column labels the predicted classes. The ICAMHMM results shown in parentheses for comparison are from paper [20].

tures of video frames. As shown in Fig. 5 and Fig. 8, a strong non-Gaussian character of compacted video features is observed. We have reason to believe the use of ICA mixture likelihood as feature function in HCRF is a more reasonable solution for video event. Second, comparing with HMM the relaxed assumption of HCRF model do generate good result with limited training data. The feature distribution is characterized by ICA mixture and the chain temporal

information is captured by HCRF. The new ICAMHCRF combines the good properties of the two and shows superior performance in two event detection tasks over existing HMM models.

6. Conclusion

A new ICAMHCRF model is developed for sports event classification. With non-Gaussian property, the local observations of each event category can be modeled as ICA mixtures. By introducing a new kind of feature function we successfully combine ICA mixture with HCRF. It is proved by experiments with bowling and golf event classification that the new model has better discriminant power than other HMM based methods. The results also demonstrate the advantage of using ICA mixture over Gaussian mixture for non-Gaussian features. Future work may include extending the method to multi-modality and other kinds of features and adding links between current observation and other hidden states.

References

- [1] P. Chang, M. Han, and Y. Gong. Extract highlights from baseball game video with hidden markov models. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 1, pages I-609–I-612 vol.1, 2002. 1
- [2] Y. Gong, M. Han, W. Hua, and W. Xu. Maximum entropy model-based baseball highlight detection and classification. *Comput. Vis. Image Underst.*, 96(2):181–199, 2004. 1
- [3] Y. Gong and W. Xu. *Machine Learning for Multimedia Content Analysis*. Springer-Verlag, 2007. 1
- [4] A. Gunawardana, M. Mahajan, A. Acero, and J. C. Platt. Hidden conditional random fields for phone classification. In *Interspeech*, pages 1117–1120, 2005. 1, 2, 3
- [5] A. Hanjalic. *Content-based analysis of digital video*. Kluwer Academic Publishers, 2004. 1
- [6] S. Kumar and M. Hebert. Discriminative random fields. *International Journal of Computer Vision*, 68(2):179–201, 2006. 1
- [7] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. 18th Int'l. Conf. Machine Learning*, pages 282–289, 2001. 1
- [8] Z. Liu, Z. L. Jincheng, and Y. Wang. Classification of tv programs based on audio information using hidden markov model. In *IEEE Signal Processing Society Workshop on Multimedia Signal Processing*, pages 27–32, 1998. 1
- [9] M. Naphade and T. Huang. Extracting semantics from audiovisual content: The final frontier in multimedia retrieval. *IEEE Trans. on neural network*, 13(4):793–810, 2002. 1
- [10] A. Quattoni, S. Wang, L.-P. Morency, M. Collins, and T. Darrell. Hidden conditional random fields. *IEEE Trans. Pattern Anal. Machine Intell.*, 29(10):1848–1852, 2007. 1, 4
- [11] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proc. of IEEE*, 77(2):257–286, 1989. 1
- [12] S. Reiter, B. Schuller, and G. Rigoll. Hidden conditional random fields for meeting segmentation. In *IEEE International Conference on Multimedia and Expo*, pages 639–642, 2007. 1, 2, 3
- [13] A. Torralba, K. P. Murphy, and W. T. Freeman. Contextual models for object detection using boosted random fields. In *Advances in Neural Information Processing Systems 17*, pages 1401–1408, 2005. 1
- [14] S. B. Wang, A. Quattoni, L.-P. Morency, D. Demirdjian, and T. Darrell. Hidden conditional random fields for gesture recognition. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2:1521–1527, 2006. 1, 4
- [15] T. won Lee, M. S. Lewicki, and T. J. Sejnowski. Ica mixture models for unsupervised classification and automatic context switching. *IEEE Trans. Pattern Anal. Machine Intell.*, 22(10):1078–1089, 2000. 1, 3, 4
- [16] L. Xie and S. fu Chang. Structure analysis of soccer video with hidden markov models. In *Pattern Recognition Letters*, pages 767–775, 2002. 1
- [17] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Generalized belief propagation. In *Advances in Neural Information Processing Systems 13*, pages 689–695, 2001. 5
- [18] H. Zhang, A. Kankanhalli, and S. Somaliar. Automatic partition of full-motion video. *Multimedia Systemes*, 1(1):11–28, 1993. 1
- [19] J. Zhou and X.-P. Zhang. Video shot boundary detection using independent component analysis. In *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages 541–544, March 2005. 5
- [20] J. Zhou and X.-P. Zhang. An ica mixture hidden markov model for video content analysis. In *Circuits and Systems for Video Technology, IEEE Transactions on*, volume 18, pages 1576–1586, 2008. 1, 6, 7