CONTEXT-BASED CONCEPT FUSION WITH BOOSTED CONDITIONAL RANDOM FIELDS

*Wei Jiang** *Shih-Fu Chang* *Alexander C. Loui*

Columbia University
Dept. Electrical Engineering

Eastman Kodak Company
Kodak Research Labs

ABSTRACT

The contextual relationships among different semantic concepts provide important information for automatic concept detection in images/videos. We propose a new context-based concept fusion (CBCF) method for semantic concept detection. Our work includes two folds. (1) We model the inter-conceptual relationships by a Conditional Random Field (CRF) that improves detection results from independent detectors by taking into account the inter-correlation among concepts. CRF directly models the posterior probability of concept labels and is more accurate for the discriminative concept detection than previous statistical inferencing techniques. The Boosted CRF framework is incorporated to further enhance performance by combining the power of boosting with CRF. (2) We develop an effective criterion to predict which concepts may benefit from CBCF. As reported in previous works, CBCF has inconsistent performance gain on different concepts. With accurate prediction, computational and data resources can be allocated to enhance concepts that are promising to gain performance. Evaluation on TRECVID2005 development set demonstrates the effectiveness of our algorithm.

Index Terms–image classification, image object detection

1. INTRODUCTION

Recognition of semantic information from visual content has been an important goal for research in image/video indexing. In recent years, NIST TRECVID video retrieval evaluation has included a task in detecting high-level features, such as locations, objects, people, and events from videos. Such high-level features, termed *concepts* in this paper, have been found to be very useful in improving quality of retrieval results in searching broadcast news videos [10].

Semantic concepts usually do not occur in isolation - knowing the contextual information (e.g. "outdoor") of an image is expected to help detection of other concepts (e.g. "car"). Based on this idea, several context-based concept detection methods have been proposed, which can be classified into two categories. The first category tries to segment an image into object regions (e.g. "building" or "road") by considering object relationships. In [4] a hidden *scene* is detected, and the correlation between global scene context and local objects is modeled to help object detection. In [9] a Conditional Random Field (*CRF*) is used whose graph nodes are pixels in the image. The graph structure is discriminatively learned through the LogitBoost algorithm. These methods have one-layer structures, where the system input is low-level representations of images (e.g. color features) and the output is the probabilities of object assignment for every pixel. The second category aims at detecting concepts in the whole images/videos [5, 6, 7] in a two-layer structure. In the first layer independent concept detectors are applied to get posteriors of class labels on a given image, and then in the second layer detection results of each individual concept is updated through a context-based model by taking into account detection confidence of other concepts. We refer to this kind of approach as Context-Based Concept Fusion (*CBCF*), which is the main issue we explore in this paper.

Several CBCF methods have been proposed. The Multinet method [5] represents inter-conceptual relations with a factor graph where co-occurrences statistics of concepts are used as compatibility potentials. Posterior probabilities of concepts are updated by loopy probability propagation. In [6], models based on Bayesian Networks are used to capture the statistical interdependence among concepts. Such techniques, though intuitive and effective in some cases, require a large amount of data to estimate joint statistics and interdependence of concepts. This makes the technique impractical in many implementations. To avoid the difficulty of estimating generative distributions, the Discriminative Model Fusion (*DMF*) method [7] uses support vector machine (*SVM*) as the context-based model. A model vector comprising of detection scores of independent detectors is fed to SVM to refine the detection result of each concept. However, results reported so far [1, 8] have indicated that not all concepts benefited from CBCF learning. The lack of consistent performance gain could be attributed to several reasons: 1) insufficient data for learning reliable relations among concepts, and 2) unreliable detectors.

In this paper we model the inter-conceptual relationships by a CRF [3] (as shown in Fig.1). For each image $\mathbf{I}$, CRF takes as input the detection results, $\mathbf{h_I}$=[$\hat{P}(y_\mathbf{I}^1$=1|$\mathbf{I}$), ..., $\hat{P}(y_\mathbf{I}^M$ =1|$\mathbf{I}$)]^T ($y_\mathbf{I}^i$ is the label for concept C_i), from M independent concept detectors, and produces updated marginal probabilities $P(y_\mathbf{I}^i$=1|$\mathbf{I}$) of each concept C_i. CRF directly models the conditional distribution $P(\mathbf{y_I}|\mathbf{h_I})$ of class label $\mathbf{y_I}$ given input

*The work was supported by Kodak graduate fellowship.

observation $\mathbf{h_I}$, while the generative methods (e.g. Multinet [5]) model the joint distribution $P(\mathbf{y_I}, \mathbf{h_I})$. When the training set is limited, the discriminative CRF can better use the sample resources to model the distribution relevant to the discriminative concept detection task than the generative approach.

To avoid the difficulty of designing compatibility potentials in CRF, a discriminative objective function aiming at class separation is directly optimized. The Boosted CRF framework [9] is incorporated, and the Real AdaBoost algorithm[2] is adopted to iteratively improve concept detection. SVM is used as weak learner for boosting because of its excellent performance found in TRECVID concept detection so far [10].

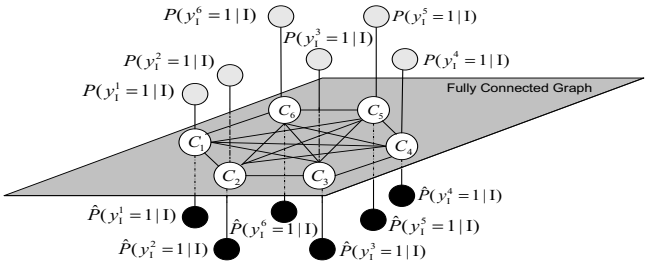

Fig. 1. The CRF that models relationships among M concepts $C_1, \ldots, C_M$. Concepts are related to each other, so the graph is full connected. The CRF takes as input (black nodes) detection results, $\hat{P}(y_{\mathbf{I}}^1{=}1|\mathbf{I}), \ldots, \hat{P}(y_{\mathbf{I}}^M{=}1|\mathbf{I})$, from M independent concept detectors, and produces updated marginal probabilities $P(y_{\mathbf{I}}^i{=}1|\mathbf{I})$ (gray nodes) of each C_i. $\mathbf{y_I}{=}[y_{\mathbf{I}}^1, \ldots, y_{\mathbf{I}}^i]^T$ is the vector of concept labels.

In addition, a simple but effective criterion is proposed to predict which concepts will benefit from CBCF, based on both information theoretic and heuristic rules. This criterion takes into consideration both the strength of relationships between a concept and its neighborhood and the robustness of detections of this neighborhood. In our experiment the prediction accuracy is 81% when 26 (out of 39) concepts are selected. The accurate prediction scheme allows us to use CBCF in practice, applying it only when it is likely to be effective.

The proposed algorithm is called Boosted CRF–Concept Fusion (*BCRF-CF*). As will be shown in Sec.2 traditional DMF [7] corresponds to the initial stage of BCRF-CF. We will also show that the extended iterative steps introduced in our method further improve the performance. Experiments are carried out over TRECVID 2005 development set [10] (80 hours, 137 video clips). Out of 39 concepts, 26 are automatically chosen to use context-based fusion among which 21 indeed get noticeable performance gain. Compared with independent detectors, BCRF-CF improves the MAP by 6.8% on a relative basis, and performance gains for several concepts are significant, e.g., 1221% for "office".

2. BOOSTED CRF CONCEPT FUSION

We start by defining notations. Let $C_1, \ldots, C_M$ be M concepts and $\mathcal{D}$ be the set of training data $\{(\mathbf{I}, \mathbf{y_I})\}$. Each $\mathbf{I}$ is an image and $\mathbf{y_I} = \{y_{\mathbf{I}}^1, \ldots, y_{\mathbf{I}}^M\}$ is the vector of concept labels, where $y_{\mathbf{I}}^i = +1$ or -1 denoting the presence or absence of concept C_i in $\mathbf{I}$ respectively. In the CBCF scenario, for each $\mathbf{I}$ the observations (system inputs) are the posteriors $\mathbf{h_I}{=}[h_{\mathbf{I}}^1, \ldots, h_{\mathbf{I}}^M]^T$, $h_{\mathbf{I}}^i{=}\hat{P}(y_{\mathbf{I}}^i = 1|\mathbf{I})$, generated by independent concept detectors. Our goal is to feed these inputs into an inferencing model to get improved posterior probility $P(\mathbf{y_I}|\mathbf{I})$ by taking into account inter-conceptual relationships.

The posterior $P(\mathbf{y}|\mathbf{h_I})$ can be modeled by a CRF [3] as:

$$P(\mathbf{y_I}|\mathbf{h_I}) = \frac{1}{Z} e^{\sum_{i=1}^M \phi_i(y_{\mathbf{I}}^i, \mathbf{h_I}) + \sum_{i=1}^M \sum_{j=1, j\neq i}^M \psi_{ij}(y_{\mathbf{I}}^i, y_{\mathbf{I}}^j, \mathbf{h_I})}$$

Z is a normalizing constant; $\phi_i(y_{\mathbf{I}}^i, \mathbf{h_I})$ and $\psi_{ij}(y_{\mathbf{I}}^i, y_{\mathbf{I}}^j, \mathbf{h_I})$ are the local and compatibility potentials respectively. One issue of CRF modeling is the design of potential functions. $\phi_i(y_{\mathbf{I}}^i, \mathbf{h_I})$ is a local decision term which influences the posteriors of concept C_i independent of its neighbors. Compatibility potentials $\psi_{ij}(y_{\mathbf{I}}^i, y_{\mathbf{I}}^j, \mathbf{h_I})$ are generally used to specify heuristic constraints for relationships between pairs of nodes, e.g. spatially smoothing constraints in image segmentation [3]. However in our problem it is unclear what kind of relationship among concept nodes we should adopt, and it is difficult to define appropriate compatibility potentials. In this paper we incorporate the Boosted CRF framework proposed in [9] which directly optimizes a discriminative objective function based on CRF and avoid the design of compatibility potentials. In the next subsections we will introduce the Boosted CRF framework [9], followed by our BCRF-CF algorithm.

2.1. Boosted CRF

After the inference with CRF the belief $b_{\mathbf{I}}^i$ on each node C_i is used to approximate the posterior: $P(y_{\mathbf{I}}^i{=}{\pm}1|\mathbf{I}) \approx b_{\mathbf{I}}^i(\pm 1)$. The aim of CRF modeling is to minimize the total loss J for all concepts over all training data:

$$J = -\prod_{\mathbf{I}\in\mathcal{D}} \prod_{i=1}^M b_{\mathbf{I}}^i(+1)^{(1+y_{\mathbf{I}}^i)/2} b_{\mathbf{I}}^i(-1)^{(1-y_{\mathbf{I}}^i)/2} \quad (1)$$

Eqn(1) is an intuitive function: the minimizer of J favors those posteriors close to training labels. Moreover we have [9]:

$$\log J = \sum_{\mathbf{I}\in\mathcal{D}} \sum_{i=1}^M \log\left[1 + e^{-y_{\mathbf{I}}^i (F_{\mathbf{I}}^i + G_{\mathbf{I}}^i)}\right] = \sum_{i=1}^M \log \tilde{J}_i \quad (2)$$

where $\log\tilde{J}_i{=}\sum_{\mathbf{I}\in\mathcal{D}}\log[1{+}e^{-y_{\mathbf{I}}^i(F_{\mathbf{I}}^i+G_{\mathbf{I}}^i)}]$; $F_{\mathbf{I}}^i$ is a discriminant function (e.g. a logistic regression stump) taking input $\mathbf{h_I} = [h_{\mathbf{I}}^1, \ldots, h_{\mathbf{I}}^M]^T$. $G_{\mathbf{I}}^i$ is a discriminant function whose input is belief $\mathbf{b}_{\mathbf{I}}^i{=}[b_{\mathbf{I}}^1(+1), \ldots, b_{\mathbf{I}}^{i-1}(+1), b_{\mathbf{I}}^{i+1}(+1), \ldots, b_{\mathbf{I}}^M(+1)]^T$, where:

$$b_{\mathbf{I}}^i(+1) = 1/(1 + e^{-(F_{\mathbf{I}}^i + G_{\mathbf{I}}^i)}) \quad (3)$$

In [9], by assuming additive models: $F_{\mathbf{I}}^i(T){=}\sum_{t=1}^T f_{\mathbf{I}}^i(t)$, and $G_{\mathbf{I}}^i(T){=}\sum_{t=1}^T g_{\mathbf{I}}^i(t)$, LogitBoost is used to iteratively optimize $\log J$, with $\mathbf{b}_{\mathbf{I}}^i$ being updated in each iteration. Logistic regression stumps are used as weak learners for $f_{\mathbf{I}}^i(t)$ and $g_{\mathbf{I}}^i(t)$.

2.2. Boosted CRF–Concept Fusion

Motivated by Ref.[9] we avoid designing compatibility potentials (which are very difficult to obtain in our problem as described earlier) by optimizing the discriminative objective function Eqn(2) with a BCRF-CF algorithm. Our BCRF-CF are different from the original Boosted CRF [9] in two aspects.

First, SVM classifiers are used instead of logistic regression stumps because of the following three reasons. (1) As discussed in [9] linear regression would work well when the

graph was densely connected, i.e., there were a large number of nodes (pixels in [9]). But the number of nodes (concepts) in our graph is small. Thus the linear approximation of the discriminant function $G^i_{\mathbf{I}}$ made in [9] may be not valid anymore. More complex function should be assumed for $G^i_{\mathbf{I}}$, e.g., the nonlinear discriminant function from kernel-based SVM. (2) In our CBCF problem, the training data is usually highly biased (the positive training samples for a concept are much less than the negative ones), and SVM is more adaptive to this biased classification problem than logistic regression because of the use of support vectors. (3) Previous literatures indicate that SVM generally performs well in semantic concept detection for TRECVID data set [1, 8], because it shows good generalization ability with limited training data.

Second, the Real AdaBoost algorithm is adopted instead of LogitBoost in [9]. LogitBoost uses logistic regression stumps as weak learners, but instead we adopt the general Real AdaBoost [2] so that we can use other weak learners, including SVM. Specifically, the solution of minimizing $\log \tilde{J}_i$ coincides with the solution of minimizing the following Q_i [2]:

$$Q_i = \sum\nolimits_{\mathbf{I}\in\mathcal{D}} e^{-y^i_{\mathbf{I}}\Gamma^i_{\mathbf{I}}}, \quad \Gamma^i_{\mathbf{I}} = \left(F^i_{\mathbf{I}} + G^i_{\mathbf{I}}\right)/2 \tag{4}$$

Q_i is exactly the objective function of Real AdaBoost [2] with the following additive model: $\Gamma^i_{\mathbf{I}}(T) = \sum_{t=1}^{T} \gamma^i_{\mathbf{I}}(t)$, $\gamma^i_{\mathbf{I}}(t) = \left(f^i_{\mathbf{I}}(t) + g^i_{\mathbf{I}}(t)\right)/2$. That is, during each iteration t, $f^i_{\mathbf{I}}(t)$ is the discriminant function generated based on input $\mathbf{h}_{\mathbf{I}}$; $g^i_{\mathbf{I}}(t)$ is the discriminant function generated based on the current beliefs $\mathbf{b}^i_{\mathbf{I}}(t)$. $\gamma^t_{i,\mathbf{I}}$ is the overall discriminant function, obtained by averaging $f^i_{\mathbf{I}}(t)$ and $g^i_{\mathbf{I}}(t)$.

The detailed BCRF-CF algorithm is given in Fig.2. The initial step of BCRF-CF is exactly the DMF approach proposed in [7]. As we will see in the experiments, this DMF method gets performance improvement in some concepts while degrading performance in many other concepts, and our boosting process can avoid this problem and achieve more consistent improvements.

3. WHICH CONCEPTS TO UPDATE

Not all concepts benefit from CBCF. As shown in [1], only 8 out of 17 concepts gained performance. Although experiments in [8] showed improvements on 80 (out of 101) concepts, our baseline independent detectors are relatively stronger than theirs, e.g., our baseline detector get 61% average precision on "car", while theirs get only 25%. Our strong independent detectors make it difficult to show improvements from CBCF.

Intuitively, two reasons may cause performance deterioration using CBCF: (1) the concept has weak relations with other concepts; (2) the related concepts have poor independent detectors. This suggests an intuitive criterion: concept C_i should use CBCF learning when C_i is strongly related to other concepts, and the average performance of detectors of the related concepts is strong. In other words, when a concept has a very strong independent detector and very poor neighborhood, it will not use CBCF. Specifically, the relationship between C_i and C_j can be measured by their mutual information $I(C_i;C_j)$, which reflects how much information one concept can get from another. The detection error E_i of the independent detector for C_i is used to estimate its robustness. Our criterion for applying CBCF learning to concept C_i is:

$$E_i > \lambda \text{ or } \frac{\sum_{j:j\in\mathcal{N}_i} I(C_i;C_j)E_j}{\sum_{j:j\in\mathcal{N}_i} I(C_i;C_j)} < \beta \tag{5}$$

The co-occurrence statistics of concepts C_i and C_j in the training set is calculated to approximate probability $P(C_i,C_j)$, based on which $I(C_i;C_j)$ is computed. Note that concept co-occurrence is a very rough approximation of $P(C_i,C_j)$, especially with limited samples. Lack of accurate estimation of joint statistics $P(C_i,C_j)$ is often the main reason that prior methods of concept fusion (e.g., [5]) fail, since the estimation error may accumulate during the iterative inferencing process. However such approximation may be sufficient for the purpose of simple concept prediction, where estimations of joint probabilities, though approximate, are not used in any iterative fusion process. Empirical experiments also verify the effectiveness of Eqn(5).

Input: training set $\mathcal{D}$; posteriors $\mathbf{h}_{\mathbf{I}}$ from independent detectors.

- Initialization:
  For each concept C_i:
  - Train SVM classifier H^0_i based on $\mathbf{h}_{\mathbf{I}}$; get $p^0(y^i_{\mathbf{I}}=1|\mathbf{I})$
  - Set $\gamma^i_{\mathbf{I}}(0) \leftarrow \frac{1}{2}\log\frac{p^0(y^i_{\mathbf{I}}=1|\mathbf{I})}{1-p^0(y^i_{\mathbf{I}}=1|\mathbf{I})}$; $\Gamma^i_{\mathbf{I}}(0)=\gamma^i_{\mathbf{I}}(0)$; $b^i_{\mathbf{I}}(+1,0)=1/(1+e^{-\Gamma^i_{\mathbf{I}}(0)})$; $w^i_{\mathbf{I}}(0)=\exp[-y^i_{\mathbf{I}}\gamma^i_{\mathbf{I}}(0)]$
- For $t=1,\ldots,T$
  For each concept C_i:
  - Form a new training data set $\tilde{\mathcal{D}}$ with size $|\mathcal{D}|$ by sampling the original set $\mathcal{D}$ according to sample weights $w^i_{\mathbf{I}}(t-1)$.
  - Train SVM classifiers $H^i_f(t)$ and $H^i_g(t)$ based on $\tilde{\mathcal{D}}$, with $\mathbf{h}_{\mathbf{I}}$ and $\mathbf{b}^i_{\mathbf{I}}(t)$ respectively. Get the corresponding class probability estimation $p^t_f(y^i_{\mathbf{I}}=1|\mathbf{I})$ and $p^t_g(y^i_{\mathbf{I}}=1|\mathbf{I})$, and also get $p^t(y^i_{\mathbf{I}}=1|\mathbf{I})=\left(p^t_f(y^i_{\mathbf{I}}=1|\mathbf{I})+p^t_g(y^i_{\mathbf{I}}=1|\mathbf{I})\right)/2$
  - $f^i_{\mathbf{I}}(t)+g^i_{\mathbf{I}}(t) \leftarrow \log\frac{p^t(y^i_{\mathbf{I}}=1|\mathbf{I})}{1-p^t(y^i_{\mathbf{I}}=1|\mathbf{I})}$; $\gamma^i_{\mathbf{I}}(t)=(f^i_{\mathbf{I}}(t)+g^i_{\mathbf{I}}(t))/2$
  - Update $\Gamma^i_{\mathbf{I}}(t)=\Gamma^i_{\mathbf{I}}(t-1)+\gamma^i_{\mathbf{I}}(t)$; $b^i_{\mathbf{I}}(+1,t)=1/(1+e^{-2\Gamma^i_{\mathbf{I}}(t)})$; $w^i_{\mathbf{I}}(t)\leftarrow w^i_{\mathbf{I}}(t-1)\ e^{(-y^i_{\mathbf{I}}\gamma^i_{\mathbf{I}}(t))}$

Fig. 2. The BCRF-CF algorithm.

4. EXPERIMENTS

Experiments are carried out on TRECVID 2005 development set [10], which contains 137 broadcast news videos and has been labeled with 39 concepts from LSCOM-Lite ontology [10]. It is separated into 2 training sets $\mathcal{T}$, $\mathcal{D}$ and 1 test set, as shown in Table 1. Independent concept detectors are SVM-based classifiers over simple image features such as grid color features and texture, extracted from key frames of a video subshot. Such classifiers have been shown to be effective for detecting generic concepts [1]. Outputs of SVMs are converted into probabilities through a standard sigmoid function.

4.1. Performance Evaluation

First, we empirically set λ=0.95, β=0.7 in the concept prediction criterion Eqn(5), and 26 concepts are automatically selected (shown in Fig.3) to use BCRF-CF learning. Fig.3 gives the MAP (the averaged AP over all the selected concepts, and AP is the official TRECVID performance metric

Table 1. The data sets for experiments

Name	Size	Usage
training set $\mathcal{T}$	41837 subshots	train independent detectors
training set $\mathcal{D}$	13547 subshots	train Boosted CRF model
test set	6507 subshots	evaluation

which related to the multi-point average precision value of a precision-recall curve) comparison of BCRF-CF, DMF, and original independent detectors, through 10 iterations over the selected 26 concepts. The figure indicates that both CBCF methods, i.e., BCRF-CF and DMF, improve the overall MAP. DMF corresponds to the initial stage of our BCRF-CF and can achieve 2.6% performance gain. Our BCRF-CF can further enhance the initial DMF performance in successive boosting iterations with 4.2% MAP gain after 10 iterations. The performance improvement in MAP obtained by our BCRF-CF is 6.8% compared with the baseline independent detectors.

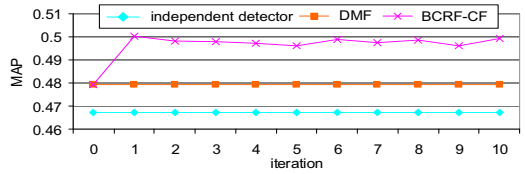

Fig. 3. The MAP comparison, averaged over 26 selected concepts.

Fig.4 gives individual AP of our BCRF-CF, the DMF and the independent detectors (after 10 iterations) over the selected 26 concepts. DMF obtains performance improvements over 15 concepts, while degrading detection results on the other 11 concepts. The performance deterioration over several concepts are severe, e.g., 8.1% in "vegetation" and 23.6% in "walking-running". Our BCRF-CF algorithm can achieve performance improvement over 21 concepts and avoid significant performance degradation over many concepts. For example, BCRF-CF improves the performance of DMF by 13% and 39% for "vegetation" and "walking-running" respectively. Generally speaking, BCRF-CF can further improve the performance of DMF, and detection results of BCRF-CF is more stable than DMF. Compared with independent detectors, significant AP improvement is achieved by BCRF-CF for some concepts, e.g., 1221% for "office".

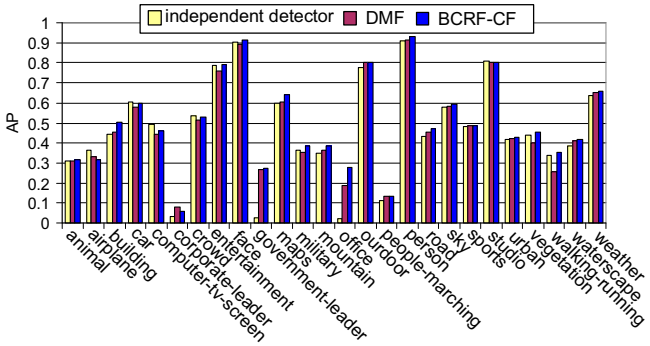

Fig. 4. The individual AP comparison.

We have also compared Real AdaBoost with LogitBoost (used in the original Boost CRF method [9] as discussed in Sec.2.1). Results confirmed the superiority of the adopted Real AdaBoost method, with 55% performance difference in terms of MAP over the 39 concepts.

4.2. Evaluation with Different Parameters

Here we evaluate the performance of detection with different β in Eqn(5), since we find β is the most sensitive parameter for concept prediction. By varying β different concepts are selected to use BCRF-CF. Let's define the *precision* of the concept selection as the percentage of concepts that actually have performance gain among selected concepts. Table 2 shows the precision and MAP gain of BCRF-CF after 10 iterations over the selected concepts with $\beta = 1$, 0.7, 0.4 respectively. Such results are promising and can be used to achieve the highest performance-cost ratio tradeoff. If we don't do concept selection, only 59% of concepts get performance improvement. When 26 concepts are selected, 21 concepts (86%) indeed get performance improvement with a gap of 6.8% in MAP. With concept selection, computational resources can be allocated to enhance concepts that have best chance to gain performance improvement.

Table 2. Performance of detection with different β

β	precision	MAP gain	# of selected concepts
1	59%	2.2%	39 (*all the concepts*)
0.7	81%	6.8%	26
0.4	91%	10.4%	11

5. CONCLUSIONS

We propose to model the inter-conceptual relations by a CRF which takes as input detection results from independent detectors and computes updated marginal probabilities as improved detection results. A modified Boosted CRF framework over SVM classifiers is incorporated to optimize the discriminative objective function and avoid the difficulty of designing compatibility potentials. A simple but effective concept selection criterion is developed to predict which concepts will benefit from CBCF. Experimental results on TRECVID 2005 development set proves the effectiveness of our BCRF-CF method and concept prediction method.

6. REFERENCES

[1] A. Amir et al. IBM research trecvid-2003 video retrieval system. *Proc. NIST TRECVID Video Retrieval Evaluation Workshop*, 2003.

[2] J. Friedman, et al. Additive logistic regression: a statistical view of boosting. *Dept. Statistics, Stanford Univ. Technical Report*, 1998.

[3] J. Lafferty et al. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *Proc. ICML*, pp.282–289, 2001.

[4] K. Murphy, et al. Using the forest to see the trees: a graphical model relating features, objects, and scenes. *Advances in NIPS*, 2003.

[5] M. Naphade et al. A factor graph framework for semantic video indexing. *IEEE Trans. on CSVT*, 12(1):40–52, 2002.

[6] S. Paek and S.F. Chang. Experiments in constructing belief networks for image classification systems. *IEEE Proc. ICIP*, 3:46-49, 2000.

[7] J. Smith et al. Multimedia semantic indexing using model vectors. *IEEE Proc. ICME*, 2:445–448, 2003.

[8] C.G.M. Snoek et al. The mediamill trecvid 2005 semantic video search engine. *Proc. 3rd TRECVID Workshop*, Gaithersburg, USA, Nov. 2005.

[9] A. Torralba et al. Contextual models for object detection using boosted random fields. *Advances in NIPS*, 2004.

[10] TRECVID. Trec video retrieval evaluations. *in http://wwwnlpir.nist.gov/projects/trecvid/.*