Association and Temporal Rule Mining for Post-Filtering of Semantic Concept Detection in Video

Ken-Hao Liu, Ming-Fang Weng, Chi-Yao Tseng, Yung-Yu Chuang, *Member, IEEE,* and Ming-Syan Chen, *Fellow, IEEE*

***Abstract*—Automatic semantic concept detection in video is important for effective content-based video retrieval and mining and has gained great attention recently. In this paper, we propose a general post-filtering framework to enhance robustness and accuracy of semantic concept detection using association and temporal analysis for concept knowledge discovery. Co-occurrence of several semantic concepts could imply the presence of other concepts. We use association mining techniques to discover such inter-concept association relationships from annotations. With discovered concept association rules, we propose a strategy to combine associated concept classifiers to improve detection accuracy. In addition, because video is often visually smooth and semantically coherent, detection results from temporally adjacent shots could be used for the detection of the current shot. We propose temporal filter designs for inter-shot temporal dependency mining to further improve detection accuracy. Experiments on the TRECVID 2005 dataset show our post-filtering framework is both efficient and effective in improving the accuracy of semantic concept detection in video. Furthermore, it is easy to integrate our framework with existing classifiers to boost their performance.**

***Index Terms*—Semantic concept detection, association rule mining, temporal rule mining, post-filtering, content-based video retrieval and mining.**

I. Introduction

With rapidly increasing capturing, storage and delivery capabilities, a vast number of video data are available. While enjoying the luxury of a plenitude of videos, people often find that videos accessible to them are more than they can absorb and it is difficult to efficiently retrieve relevant ones. Therefore, effective video retrieval and mining has become a research focus to address this need. To facilitate effective video retrieval and mining, automatic semantic concept detection [1], [2], i.e. finding video shots that match specific concepts such as *outdoor*, *face*, *office* and *nature*, plays an important role because it bridges the gap between low-level features and high-level human interpretation.

The concept detection problem can typically be formulated as a pattern classification problem where multiple classifiers based on visual, audio and text features are trained from videos and a set of annotations using machine learning methods. Most work in this area focuses on learning the mapping between the low-level features extracted from videos and the corresponding high-level concept annotations. Unfortunately, due to the gap between low-level features and high-level semantic interpretation [3]–[5], semantic concepts are still often difficult to be accurately detected even after we utilize multi-modal features and various fusion techniques [6]. Therefore, effective and efficient semantic concept detection in video remains a challenging problem to be solved.

Learning for semantic concept detection often requires a large set of ground truth annotations which demand a tremendous manual effort. Although such annotations are precious, most approaches only utilize annotations for learning a mapping between low-level features and a concept at a time. However, annotations actually contain more information that we can explore to improve concept detection performance. For example, the co-occurrence of several semantic concepts in a shot could imply the presence of other concepts. For instance, the presence of the concept *building* likely implies the presence of the concept *outdoor*. Thus, we could discover such inter-concept associations from the annotations and use them to improve detection accuracy. In addition, a video is often visually smooth and semantically coherent. Thus, the presence of a semantic concept generally spans multiple consecutive shots. For example, the presence of the concept *sports* in a shot indicates that the same concept is likely present in its previous and next few shots in the same video sequence. Therefore, the presence of a semantic concept for the current shot could be inferred from detection results of neighboring shots. Such inter-shot temporal dependency can also be learned from annotations.

Motivated by the observations that a video shot usually is annotated with multiple correlated concepts and that a semantic concept usually spans multiple shots, this paper proposes a general post-filtering framework that infers the presence of a semantic concept from both inter-concept association relationships and inter-shot temporal dependency. We use the association analysis [7], [8] and temporal rules to enhance the performance of semantic concept detection for video data. To exploit inter-concept association relationships, based on concept annotations of video shots, we discover the hidden association between concepts, i.e., frequent concept patterns, which are sets of concepts frequently appearing together within a shot. The concept association rules that define implication relationships between concepts are used to

Ken-Hao Liu, Chi-Yao Tseng and Ming-Syan Chen are with the Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan. (e-mail: kenliu@arbor.ee.ntu.edu.tw, cytseng@arbor.ee.ntu.edu.tw, mschen@cc.ee.ntu.edu.tw).

Ming-Fang Weng and Yung-Yu Chuang are with Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan. (e-mail: mfueng@cmlab.csie.ntu.edu.tw, cyy@csie.ntu.edu.tw).

improve the detection accuracy by integrating the associated concept detectors using our combined ranking scheme. For exploring inter-shot temporal coherence, the temporal rules that model the temporal dependency among neighboring shots are used to aggregate results from neighboring shots to predict the current shot with respect to a concept. We explore several design options and propose an effective smoothing scheme that exploits temporal coherence to correct the (mis-)prediction for a shot using its adjacent shots.

Although some previous work shares similar spirit of using inter-concept relationship or temporal coherence to improve concept detection, most of them integrated such ideas into their classifiers using graphical models or other means [9]–[16]. The complexity of modeling the relationships among all concepts often grows exponentially as the number of concepts increases. Therefore, large amount of training data is needed to effectively learn the relations among concepts and such models are coupled with specific sets of classifiers. The reported improvement on concept detection accuracy using these methods is often limited due to the usually unreliable outputs from single concept detectors. In addition, such methods are often difficult to be integrated with classifiers using different approaches. Our post-filtering framework does not require separate training data and uses an efficient data-driven approach to obtain inter-concept association rules and inter-shot temporal dependency. Furthermore, our framework is universally applicable to any given set of independent classifiers to boost their performance. Finally, our post-filtering scheme is extremely efficient in both learning and detection.

The rest of this paper is organized as follows. Related work is discussed in Section II. The semantic concept detection framework is introduced in Section III. Concept association analysis and a combined ranking scheme are described in Section IV. Section V presents temporal rule mining. Experiments are discussed in Section VI, followed by conclusions in Section VII.

II. Related Work

Association classification [17], [18] has been proposed in recent studies in data mining to achieve higher classification accuracy than traditional rule-based classifiers such as C4.5 [19]. However, these approaches generate the association rules between features and class labels for prediction and do not consider the association between different classes. Such techniques have also been applied to web image clustering [20]. Analysis of concept annotation data has been proposed by Kender and Naphade [21] to track news stories. However, their focus is on clustering of video episodes into new stories with low-level features instead of improving semantic concept detection. Xie and Chang tested different mining schemes on annotations for a fixed lexicon and showed that discovered patterns can indicate semantics beyond the lexicon for annotations [22]. Frequent itemsets are defined on the concept annotations and their consistency is verified on two different sets of concept lexicons. However, they did not use them for visual concept detection.

Various multi-concept relational learning approaches via graphical models, such as Bayesian network, restricted Boltzman machines, Markov random fields, and conditional random fields, have been proposed by researchers [9]–[11] to capture the relationship between outputs of independent concept classifiers. Because of their complexity, large amount of training data is often needed for effective learning. The semantic pathfinder [12] also considers concepts in context. It utilizes the Discriminative Model Fusion (DMF) method [13] which uses an extra layer of contextual SVM classifiers that take the detection scores of independent detectors to further refine the detection results. Boosted conditional random field [14] is used to improve the results of DMF by combining the power of boosting with Conditional Random Field (CRF). On the contrary, our post-filtering framework efficiently explores inter-conceptual association rules and inter-shot temporal dependency within training data. Furthermore, our framework is more easily applied to other classifiers.

Ebadollahi *et al.* detected novel visual events by modeling them as stochastic temporal processes in the semantic concept space [15]. They used Hidden Markov Model (HMM) to map the concept score evolution patterns to a visual event. However, they did not consider inter-shot temporal dependency to refine concept scores. Yang and Hauptmann studied the effects of temporal consistency on video retrieval and proposed to use active learning with temporal sampling strategies to improve accuracy of concept detectors [16]. They also concluded that linear smoothing did not have any significant improvement. However, they did not consider the posterior probability of positive and negative results. Thus, the smoothed score of a shot is simply a weighted combination of likelihood scores of three neighboring shots. The impact of temporal window size was not considered in their work and the filter weights were estimated only by logistic regression which may suffer from outliers or noise in the original prediction scores. On the contrary, our post-filtering framework explores more effective smoothing schemes by estimating proper temporal filtering window sizes and weights based on annotations and statistical measurements.

III. Semantic Concept Detection

Users often input queries to a video database to retrieve videos corresponding to specific high-level concepts. Due to the large amount of video data, a general approach for semantic concept detection is needed to automatically annotate large-scale video archive based on a fixed concept lexicon to facilitate such queries [6], [23]. Let $C = \{c_1, c_2, ...c_M\}$ be the concept lexicon, i.e. the set of M concepts that the system attempts to detect. For semantic concept detection, a video is first segmented into a sequence of scenes; each scene is segmented into shots; and each shot is comprised of a set of keyframes. Shots are the commonly-used basic semantic units for annotation and retrieval. To train concept detectors, some shots are annotated manually to create the ground truth. Let $S = \{s_1, s_2...s_t, ..., s_N\}$ be the training set of N shots and $\{A_1, A_2, ...A_N\}$ be the set of corresponding annotations, in which A_t is the annotation for the t-th shot s_t. Because multiple concepts could simultaneously be present in a shot, the annotation A_t is a subset of C. We could use a binary

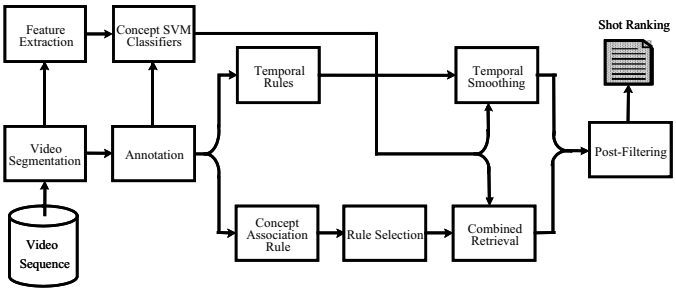

Fig. 1. Our post-filtering framework for semantic concept detection in video with concept association rules and temporal rules.

variable $l_t^i = |A_t \bigcap \{c_i\}|$, where $|.|$ indicates the number of elements in a set, to represent whether the concept c_i is present in the shot s_t. Each keyframe within a shot is processed to extract a set of features characterizing the visual properties of the annotated concept. These visual features could include color, texture, motion, structure, color moments and so on. Audio and speech information could also be included to enhance the performance. Finally, let $\{\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_N\}$ be the set of features for classification, where $\mathbf{x}_t$ is the feature associated with the shot s_t.

Given the feature vectors extracted from the video data and the corresponding annotations given by users, a typical approach to the semantic concept detection task is to use supervised learning. Classification techniques such as support vector machines (SVMs) find patterns associated with a specific concept in the features of the video data. The SVM classifier d_i for each of the semantic concepts c_i can be trained from the manually annotated training data. Platt's conversion method [24], [25] can be used to convert the output margin of the SVM method into a posterior probability. Thus, for each concept c_i, each concept classifier d_i provides a *prediction value* within $[0, 1]$ as the probability measurement $P(l_t^i|\mathbf{x}_t)$ for the presence of the concept c_i in the test shot s_t given s_t's associated feature vector $\mathbf{x}_t$. The retrieval result is often presented to the user as a ranked list of all shots in the order of their prediction values.

Due to the semantic gap, discrepancy between low-level features and high-level semantic interpretation [3], some semantic concepts may be difficult to detect solely based on the concept classifiers. In this paper, we propose a post-filtering technique to incorporate context knowledge (both inter-concept and inter-shot) to further improve the accuracy of semantic concept detection in video. Figure 1 shows our post-filtering framework for semantic concept detection using concept association rules and temporal rules. During the training phase, these rules are discovered from manual shot annotations without any extra training data and are independent of the types of classifiers. Concept association rules capture the inter-concept relationships between multiple concepts while temporal rules model the temporal intra-concept dependency among multiple neighboring shots. At the detection stage, given only the prediction values for shots, our rule-based post-filtering module uses the learned association and temporal rules to re-rank the test shots.

IV. Association Rule Mining

The co-occurrence of semantic concepts in a shot represents a context that can be used to discover hidden relationships between semantic concepts. Such context can be modeled as concept association rules that can be used to infer the presence of a concept based on the presence of other associated concepts. In Section IV-A, we first present formal definitions of concept association rules to clearly show what we aim to discover from the annotation data, followed by efficient algorithms to discover frequent patterns and generate these rules. We then present a combined ranking scheme in Section IV-B for post-filtering of semantic concept detection results based on the discovered concept association rules.

A. Concept Association Rules

Let A and B be two annotations containing concepts from C. We say that annotation A contains B if and only if $B \subseteq A$.

Definition (**Concept Association Rule**) A concept association rule is an implication of the form $A \Longrightarrow B$, where $A \subset C$, $B \subset C$, and $A \cap B = \phi$.

Definition (**Support**) The support of a concept association rule, $A \Longrightarrow B$, is the percentage of annotations that contain $A \cup B$.

Definition (**Confidence**) The confidence of a concept association rule, $A \Longrightarrow B$, is the percentage of annotations containing A that also contain B.

Intuitively, a concept association rule $A \Longrightarrow B$ means the co-occurrence of concepts in set A in a shot implies the presence of concepts in set B in that shot. The support of the concept association rule measures how often such an association occurs in the ground truth and the confidence indicates how likely such an implication happens when concepts in A co-occur. For example, the rule, *building⇒outdoor*, indicates that appearance of the concept *building* implies it is likely that the concept *outdoor* also appears in the same shot. The *support* and *confidence* represent the interestingness of a discovered rule. A support of 2% means that 2% of the annotations of all shots show that these two concepts appear together. A confidence of 60% means that 60% of the shots whose annotations contain *building* also contain *outdoor*. Typically, we are interested in association rules that satisfy both a user-given minimum support threshold *min_supp* and a minimum confidence threshold *min_conf*.

Example 1 The following table shows an example of an annotated training video dataset. $\{aircraft\} \Longrightarrow \{sky\}$ is an example of a concept association rule with support of 2/5 and confidence of 2/3. ■

Shot	Annotation
s_1	A_1={*aircraft*, *sky*}
s_2	A_2={*urban*, *people*, *outdoor*}
s_3	A_3={*aircraft*, *outdoor*}
s_4	A_4={*aircraft*, *sky*, *outdoor*}
s_5	A_5={*people*, *walking_running*, *military*}

To discover concept association rules, ground truth annotations are analyzed to find hidden frequent patterns, or *itemsets*, reflecting which concepts are frequently associated or appear together in a shot. These patterns can then be used to discover concept association rules. The Apriori algorithm [26] is an iterative approach to perform a level-wise search for frequent itemsets, where frequent k-itemsets, i.e., itemsets that contain exactly k distinct items, are used to generate frequent $(k+1)$-itemsets. The level-wise search space can be reduced effectively by using the following property:

(Apriori Property) *All nonempty subsets of a frequent itemset must also be frequent.*.

For example, if an itemset I is not frequent, i.e., $Support(I) < min\_supp$, then the itemset with another added item A, $I \cup A$ cannot occur more frequently than I, *i.e.*, $Support(I \cup A) < min\_supp$. The algorithmic form of the Apriori algorithm is as follows.

Algorithm 1 APRORI. *Given an annotation data set* $\{A_1, A_2, ...A_N\}$ *and the minimum support threshold, min_supp, find frequent itemsets.*

```
1: procedure APRORI({A_1, A_2, ...A_N}, min_supp)
2:     F_1 = the set of frequent 1-itemset
3:     for (k=2; F_{k-1} ≠ φ; k++) do
4:         C_k = candidates generated from F_{k-1}
5:         for each annotation A_t
6:             increment the count of all sets c ∈ C_k that are
                   subsets of A_t
7:         end for
8:         F_k = {c ∈ C_k | Support(c) ≥ min_supp}
9:     end for
10:    return F = ∪_k F_k
11: end procedure
```

The association rules can then be generated from the frequent itemsets discovered with the Apriori algorithm by enumerating nonempty subsets and testing the confidence against the minimum confidence threshold, *min_conf*. For semantic concept association rules, we are only interested in rules that have a single concept on their right-hand sides. That is, we only generate rules of the form $\{c_1, ..., c_{k-1}\} \Rightarrow c_k$, where c_1, ..., $c_{k-1}, c_k \in C$. We are only interested in the rules with one-concept right-hand side because a rule with n concepts on its right-hand side, $A \Rightarrow \{b_1, \cdots, b_n\}$, can be equivalently captured using n rules, $A \Rightarrow \{b_k\}, k = 1..n$. Thus, it is enough to only consider the rules with a single concept on their right-hand sides. Specifically, the semantic concept association rules are generated using Algorithm 2.

Example 2 Using the annotation dataset in Example 1, suppose the minimum support count is 2 and minimum confidence is 50%. The Apriori algorithm first obtains F_1: {*aircraft* <3>, *sky* <2>, *outdoor* <3>, *people* <2>} by scanning through the table. Then, the Apriori algorithm generates C_2: {{*aircraft*, *sky*} <2>, {*aircraft*, *outdoor*} <2>, {*aircraft*, *people*} <0>, {*sky*, *outdoor*} <1>, {*sky*, *people*} <0>, {*outdoor*, *people*} <1>} and obtains the corresponding support counts in the brackets. Therefore, we have F_2: {{*aircraft*, *sky*}, {*aircraft*, *outdoor*}}. Before we continue to generate C_3, note that in order for {*aircraft*, *sky*, *outdoor*} to be frequent, {*sky*, *outdoor*} needs to be frequent, but it is not. Therefore, we have obtained all the frequent itemsets and can proceed to generate the concept association rules and calculate their confidence values as follows: *aircraft* $\Longrightarrow$ *sky* $<\frac{2}{3}>$, *sky* $\Longrightarrow$ *aircraft* $<\frac{2}{2}>$, *aircraft* $\Longrightarrow$ *outdoor* $<\frac{2}{3}>$, and *outdoor* $\Longrightarrow$ *aircraft* $<\frac{2}{3}>$. Note that all these rules are valid since their confidence values are all larger than the minimum confidence threshold. ■

Algorithm 2 RuleGen. *Given the frequent itemset F and the minimum confidence threshold, min_conf, generate association rules.*

```
1: procedure RULEGEN(F, min_conf)
2:     for each concept c_i in frequent itemset F
3:         generate two subsets {F − c_i}, {c_i}
4:         if Support(F) / Support(F−{c_i}) ≥ min_conf then
5:             Output the rule "F − {c_i} ⟹ c_i"
6:         end if
7:     end for
8: end procedure
```

The Apriori algorithm finds a complete set of rules based on a user-given minimum support and a minimum confidence threshold. It is often the case that we have multiple rules which all imply the same concept association. Redundant rules are pruned by testing if the left-hand side of the rule is a subset of the left-hand side of a more general rule. After rule pruning, the best rule for each inferred concept is selected based on the confidence and support values, in that order. Specifically, given two rules R_1 and R_2 that both infer the same concept, i.e., both rules have the same right-hand side. R_1 is selected over R_2 if and only if (1) R_1 is not redundant with respect to R_2; and (2) one of the following conditions holds: $confidence(R_1) > confidence(R_2)$, or $support(R_1) > support(R_2)$ if $confidence(R_1) = confidence(R_2)$.

B. Combined Ranking

Based on the concept association rules, we can integrate the output from an ensemble of concept detectors corresponding to the left-hand side of a discovered association rule. Given a shot s_t with a feature vector $\mathbf{x}_t$, assume that the concept detector for concept c_i outputs a prediction value $p_t^i = P(l_t^i|\mathbf{x}_t) \in [0, 1]$, where $i = 1, 2, \ldots, M$ and $t = 1, 2, \ldots, N$. This value indicates the likelihood the concept detector regards the presence of concept c_i in shot s_t. The discovered association rules are used to combine the prediction values of the associated concept detectors and generate the combined ranking.

The distribution of prediction values over all the shots differs from one concept classifier to another. Note that a high/low prediction value means that the classifier is more certain about the presence/absence of the corresponding concept. In this paper, we propose to use the entropy function H to transform prediction values p into recommendation values

$R \in [-1, 1]$.

$$H(p) = -p\log_2(p) - (1-p)\log_2(1-p),$$
$$R(p) = \begin{cases} 1 - H(p) & \text{for } 0.5 \le p \le 1 \\ H(p) - 1 & \text{for } 0 \le p < 0.5 \end{cases}.$$

The entropy function is in essence a measure of the uncertainty [27]. A recommendation value R is positive when the concept is more likely to be present, i.e., $p > 0.5$, and is negative when the concept is more likely to be absent, i.e., $p < 0.5$. In addition, the absolute value of a recommendation value reflects certainty on the detection output. For example, when the prediction value of a video shot from a concept classifier is 0.5, the value of the entropy function, $H(p) = 1$, is the highest since we are most uncertain about the outcome of this video shot. Thus, its recommendation value is 0 and this concept classifier will not have any contribution to the final combined ranking for this shot. On the other hand, a prediction value closer to 1.0 or 0.0 will have a higher recommendation value and more contribution to final ranking results. From our experiments, we found that combined ranking with recommendation values gives better performance than prediction values because it combines results from associated classifiers in a uniform and normalized way that considers certainty on both the presence and the absence of the corresponding concepts.

Consider the rule $\{c_1, c_2, \cdots, c_{k-1}\} \Longrightarrow \{c_k\}$ with confidence f, the recommendation metrics for the associated classifier which outputs $\{p_1, p_2, ..., p_{k-1}, p_k\}$ are combined as follows:

$$R_{\text{combined}}(c_k) = \frac{1}{K-1}\left(\sum_{i=1}^{k-1} R(p_i)\right) * f + R(p_k),$$

The combined recommendation value increases the original recommendation value for the right-hand side concept (implied concept) by an amount of the average recommendation value of the left-hand side concepts (associated concepts). Since the association rule has a confidence value f on such an implication relationship, the increase on the recommendation value is adjusted by multiplying with f. We are in effect exploiting associated concept detectors to infer the presence of the implied concept and re-rank shots. Therefore, such a combined ranking scheme can be more effective and robust than ranking solely based on a single concept detector.

V. Temporal Rule Mining

Videos exhibit temporal continuity in both visual content and semantics. This section attempts to exploit this coherence to improve the performance of detectors by learning temporal association rules from the ground truth annotations. We first explore several measurements for testing whether temporal dependence among neighboring shots are statistically significant. Next, we present our design of the temporal filter for effective temporal smoothing of the prediction values with respect to a concept.

A. Temporal Dependency Test

Recall that $l_t^i \in \{0, 1\}$ is a binary random variable indicating whether a shot s_t is relevant to a semantic concept c_i. In this section, we only consider the temporal consistency between neighboring shots for a concept c_i at a time. For simplicity, we can drop the index i without ambiguity. We first estimate the conditional probabilities from annotations. The conditional probabilities of the shot s_t being relevant to the concept c given that its neighboring shot of a temporal distance k, s_{t-k}, is relevant or irrelevant to c are calculated as:

$$P(l_t = 1|l_{t-k} = 1) = \frac{\#(l_t = 1, l_{t-k} = 1)}{\#(l_{t-k} = 1)} \text{ and}$$
$$P(l_t = 1|l_{t-k} = 0) = \frac{\#(l_t = 1, l_{t-k} = 0)}{\#(l_{t-k} = 0)},$$

where $\#(l_{t-k}=1)$ and $\#(l_{t-k}=0)$ are equivalent to the total numbers of relevant and irrelevant shots in the training dataset, respectively; $\#(l_t=1, l_{t-k}=1)$ is the total number that two shots are k shots apart and both relevant to the concept c; and $\#(l_t=1, l_{t-k}=0)$ is the total number that shot s_t is relevant to c but its k-shot-preceding shot s_{t-k} is irrelevant.

Next, we present several statistical measurements for testing dependency between random variables, chi-square test, likelihood ratio, mutual information and pointwise mutual information [28].

Chi-square test. Chi-square test is a statistical test for dependency. For our temporal dependency test, it is used to compare the observed frequencies in the following 2-by-2 table,

	$l_{t-k}=0$	$l_{t-k}=1$
$l_t=0$	$\zeta_{00}=\#(l_t=0, l_{t-k}=0)$	$\zeta_{01}=\#(l_t=0, l_{t-k}=1)$
$l_t=1$	$\zeta_{10}=\#(l_t=1, l_{t-k}=0)$	$\zeta_{11}=\#(l_t=1, l_{t-k}=1)$

and the χ^2 value is then calculated by

$$\chi_k^2 = \frac{N(\zeta_{00}\zeta_{11} - \zeta_{01}\zeta_{10})^2}{(\zeta_{00}+\zeta_{01})(\zeta_{00}+\zeta_{10})(\zeta_{01}+\zeta_{11})(\zeta_{10}+\zeta_{11})}.$$

A high χ^2 value means two events are likely associated. One disadvantage of using χ^2 values is that they are not intuitively interpretable. A table lookup is necessary to convert them into confidence values for the dependency hypothesis.

Likelihood ratio. Likelihood ratio is used to tell us how much more likely one of the following two hypothesis is than the other.

- Hypothesis 1 (a formulation of independence): the occurrence of the concept c in the shot s_t is independent to the occurrence in the shot s_{t-k}. Thus,

$$P(l_t = 1|l_{t-k} = 1) = p = P(l_t = 1|l_{t-k} = 0)$$

- Hypothesis 2 (a formulation of dependence):

$$P(l_t = 1|l_{t-k} = 1) = p_1 \neq p_2 = P(l_t = 1|l_{t-k} = 0).$$

The probabilities p, p_1 and p_2 are estimated as

$$p = \frac{\#(l_t = 1)}{N} = \frac{\xi_1}{N},$$
$$p_1 = \frac{\#(l_t = 1, l_{t-k} = 1)}{\#(l_{t-k} = 1)} = \frac{\xi_{12}}{\xi_2},$$
$$p_2 = \frac{\#(l_t = 1) - \#(l_t = 1, l_{t-k} = 1)}{N - \#(l_{t-k} = 1)} = \frac{\xi_1 - \xi_{12}}{N - \xi_2},$$

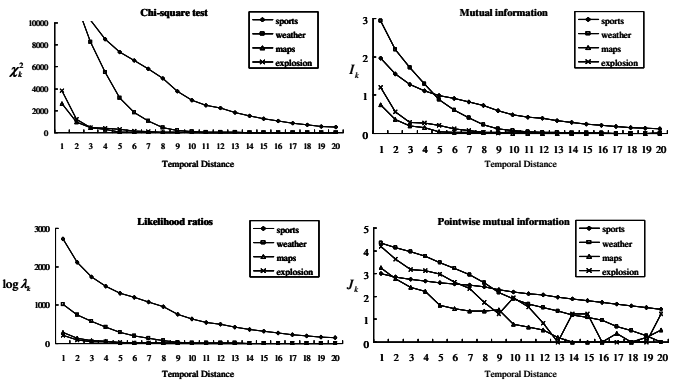

Fig. 2. Temporal dependency of shots at different temporal distances for four different concepts, *Sports*, *Weather*, *Maps* and *Explosion*, evaluated using four different statistical measurements. These concepts show that the temporal dependency is highly concept dependent. Some concepts such as *Sports* have stronger temporal dependency than others like *Maps*.

where $\xi_1 = \#(l_t = 1)$, $\xi_2 = \#(l_{t-k} = 1)$ and $\xi_{12} = \#(l_t = 1, l_{t-k} = 1)$. Assuming a binomial distribution, the likelihood ratio of hypothesis 2 over hypothesis 1 is then calculated as

$$\lambda_k = \frac{L(\xi_{12}, \xi_2, p_1) L(\xi_1 - \xi_{12}, N - \xi_2, p_2)}{L(\xi_{12}, \xi_2, p) L(\xi_1 - \xi_{12}, N - \xi_2, p)},$$

where $L(m, n, q) = q^m (1-q)^{n-m}$. The likelihood ratio λ_k means that hypothesis 2 is λ_k times more likely than hypothesis 1, meaning the chance that shots s_t and s_{t-k} are associated is λ_k times larger than the one that they are independent.

Mutual information. Mutual information is the entropy difference between two random variables, in our case, l_t and l_{t-k}. It is thus defined as

$$I_k = \sum_{\alpha,\beta \in \{0,1\}} P(l_t = \alpha, l_{t-k} = \beta) \log \frac{P(l_t = \alpha, l_{t-k} = \beta)}{P(l_t = \alpha) P(l_{t-k} = \beta)},$$

and tells us the reduction of uncertainty of one variable due to knowing about the other.

Pointwise mutual information. Pointwise mutual information measures dependency between two particular events, instead of random variables. In our case, we are most concerned about how much the fact that the concept c is present in s_{t-k} reduces the uncertainty of the event that c is present in s_t. Thus, we define the pointwise mutual information as

$$J_k = \log \frac{P(l_t = 1 | l_{t-k} = 1)}{P(l_t = 1)}.$$

Note that Yang and Hauptmann also used pointwise mutual information to measure temporal dependency [16].

Figure 2 shows these dependency measurement values of various temporal distances (from 1 to 20) for different concepts (*Sports*, *Weather*, *Maps* and *Explosion*). It is obvious that the temporal dependency varies a lot among concepts. For example, concepts like *Sports* and *Weather* show temporal dependency over a relatively large range of temporal distances while those like *Explosion* and *Maps* only show temporal dependence over a relatively short range of temporal distances.

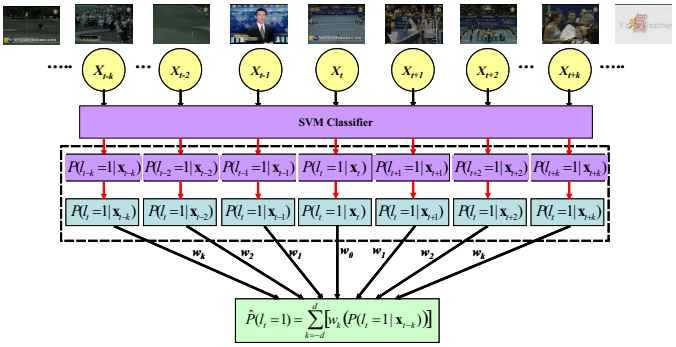

Fig. 3. Temporal smoothing by a weighted combination of inference values from the neighboring shots.

B. Temporal Smoothing

We can exploit temporal coherence to "smooth" the prediction of a shot with respect to a concept by a weighted combination of the *inference values* of its neighboring shots. Note that we use the inference values that are estimated with prior probabilities and prediction values instead of using prediction values directly.

We define the temporal neighborhood distance d for a shot with respect to a concept as the maximum temporal distance within which shots will have impacts on predicting the result for the current shot. Given the temporal neighborhood distance d, our temporal smoothing filter for a concept can thus be defined as follows:

$$\begin{aligned}
\hat{P}(l_t = 1) &= \sum_{k=-d}^{d} w_k P(l_t = 1 | \mathbf{x}_{t-k}) \\
&= \sum_{k=-d}^{d} w_k [P(l_t = 1 | l_{t-k} = 1) P(l_{t-k} = 1 | \mathbf{x}_{t-k}) \\
&\qquad + P(l_t = 1 | l_{t-k} = 0) P(l_{t-k} = 0 | \mathbf{x}_{t-k})] \\
&= \sum_{k=-d}^{d} w_k [P(l_t = 1 | l_{t-k} = 1) P(l_{t-k} = 1 | \mathbf{x}_{t-k}) \\
&\qquad + P(l_t = 1 | l_{t-k} = 0)(1 - P(l_{t-k} = 1 | \mathbf{x}_{t-k}))],
\end{aligned}$$

where $\mathbf{x}_{t-k}$ is the visual features extracted from the shot s_{t-k}, $P(l_t = 1 | l_{t-k} = 1)$ and $P(l_t = 1 | l_{t-k} = 0)$ are prior probabilities estimated from the annotations, $P(l_{t-k} = 1 | \mathbf{x}_{t-k})$ is the prediction value given by the detector indicating how likely concept c is present in shot s_{t-k}, and w_k is a concept-dependent weighting coefficient that measures the contribution from the shot that is temporally k shots apart from s_t. The sum of w_k equals one. We call the term, $P(l_t = 1 | \mathbf{x}_{t-k})$, inference value because it infers the prediction value $P(l_t = 1)$ by using the feature vector $\mathbf{x}_{t-k}$ of shot s_{t-k}. It can be taken as a posterior probability since it takes both likelihood $P(l_{t-k} | \mathbf{x}_{t-k})$ and prior $P(l_t | l_{t-k})$ into account. On the contrary, Yang and Hauptmann used directly the prediction values $P(l_{t-k} = 1 | \mathbf{x}_{t-k})$ of s_{t-k} for logistic regression [16].

Figure 3 shows an illustration of the temporal filter for temporal smoothing. Given this framework, to design a temporal smoothing filter, we have to determine two sets of parameters for each concept: (1) the temporal neighborhood distance and

(2) a set of distance-dependent weighting coefficients. Proper thresholding on statistical measurements could be used to determine the temporal distance for each concept. However, after extensive experiments on the results of the training set, we have empirically decided to use the chi-square test with confidence level at 99.9% to determine the temporal neighborhood distance and thus reject the shots whose χ^2 value is less than 10.82. In addition, we set a maximal temporal distance at 20, since we observe that temporal dependency beyond that distance is negligible. For determining the weighting coefficients, the values of statistical measurements at different distances are directly used.

VI. EXPERIMENTS

A. Experimental Setting

To evaluate the performance of our proposed approach, we have tested our approach on the TRECVID 2005 dataset. TRECVID is an annual video retrieval evaluation event organized by National Institute of Standards and Technology (NIST) to promote progress in content-based retrieval from digital video via open, metrics-based evaluation [23]. The TRECVID 2005 training corpus consists of 85 hours of Arabic, Chinese and US broadcast news video sources. Since the TRECVID 2005 test data does not have ground truth, we only used the TRECVID 2005 training data in our experiments. We partitioned the TRECVID 2005 training set into a training data set of 30,993 shots and a test data set of 12,914 shots, exactly in the same way as that in MediaMill, so as to evaluate our performance. We used the ground truth annotations from MediaMill with a lexicon of 101 semantic concepts [6]. Association rules and temporal rules were learned from the annotations of the training set only. Performance was then evaluated on the test set.

We use the classifications of MediaMill [6], MM, as one of the baselines for comparison. These classifiers are learned from visual feature extraction described in Snoek *et al.*'s paper [6]. Specifically, a set of predefined regions in a key frame image is labeled with similarity scores for a total of 15 low-level visual concepts, like road, water body and so on. The sizes of the predefined regions were adjusted to obtain a total of 8 concept occurrence histograms. We have also generated another optimized classifier, NTU classifier, based on the same features supplied by MediaMill as another classifier for comparison. Since parameters of SVMs have significant influence on performance of detectors, we adopt Gaussian kernels and use libSVM [25] to obtain classifiers with optimal gamma parameters in kernel function and misclassification penalty cost, selected via five-fold cross validation. The NTU classifier has better classification performance (MAP=0.285) than the MM baseline (MAP=0.216). We use the NTU classifier to show that our post-filtering method helps improve performance of classifiers with different accuracy.

B. Performance Metrics

To evaluate the performance of the proposed post-filtering framework, we compare the detection performance using average precision, which is adopted by NIST [23] to measure the accuracy of a ranked concept detection result. Average precision is proportional to the area under a recall-precision curve and favors highly ranked relevant shots. Let S be the size of the test set and R the number of relevant shots. At any given index j, let R_j be the number of relevant shots in the top j shots. Let $I_j = 1$ if the j^{th} shot is relevant and 0 otherwise. The average precision is then defined as

$$AP = \frac{1}{R} \sum_{j=1}^{S} \frac{R_j}{j} * I_j.$$

concept association rule	confidence
{*crowd*, *face*,*government_leader*}$\Longrightarrow${*people*}	100%
{*military*, *outdoor*, *people*, *walking_running*}$\Longrightarrow${*violence*}	100%
{*car*, *outdoor*, *people*}$\Longrightarrow${*vehicle*}	100%
{*building*, *sky*, *urban*}$\Longrightarrow${*outdoor*}	100%
{*male*, *people*}$\Longrightarrow${*face*}	100%
{*face*, *people*, *studio*}$\Longrightarrow${*indoor*}	100%
{*military*, *outdoor*, *people*, *violence*}$\Longrightarrow${*walking_running*}	100%
{*anchor*, *face*, *indoor* ,*overlayed_text*, *people*}$\Longrightarrow${*studio*}	100%

TABLE I
SAMPLES OF NON-TRIVIAL CONCEPT ASSOCIATION RULES.

C. Concept Association Rules

The training dataset consists of annotations for 101 concepts, in which each shot is annotated with a subset of the given concept lexicon. We have observed that the average number of annotated concepts per shot is roughly 4. Then we performed the Apriori algorithm with *min_supp*=2% and *min_conf*=80% on the annotations. As a result, we have found 32 concepts that have statistically significant rules for inference. Among them, some of the discovered concept association rules are intuitive, such as $\{car\} \Longrightarrow \{vehicle\}$, while others represent frequent patterns that may otherwise remain hidden due to the large number of shot annotations given by the users, for example, {*military, outdoor, people, violence*} $\Longrightarrow$ {*walking_running*}. Table I shows examples of association rules that are not trivial.

Baseline classifiers are used to obtain the posterior probability scores $p(l_i|\mathbf{x}_j)$ for each concept i in the lexicon and each shot j in the test dataset. For the concept with association rules, we re-rank all the shots based on the combined ranking algorithm in Section IV-B. Figure 4 shows the performance of our combined re-ranking based on the MM baseline results for the 32 concepts that have association rules. Our re-ranking improves performance for 24 concepts. Among them, 40% have improvements more than 5% in average precision values. Overall, we observe 3.3% and 2.0% improvement over the MM baseline and the NTU classifier respectively in terms of mean average precision.

D. Temporal Smoothing

In this experiment, we test the performance of our temporal filtering scheme. We first perform experiments on the effectiveness of different dependency measures. Figure 5 compares the mean AP of the 101 concepts using these measures on both the MM baseline and the NTU classifier. Overall, we observe

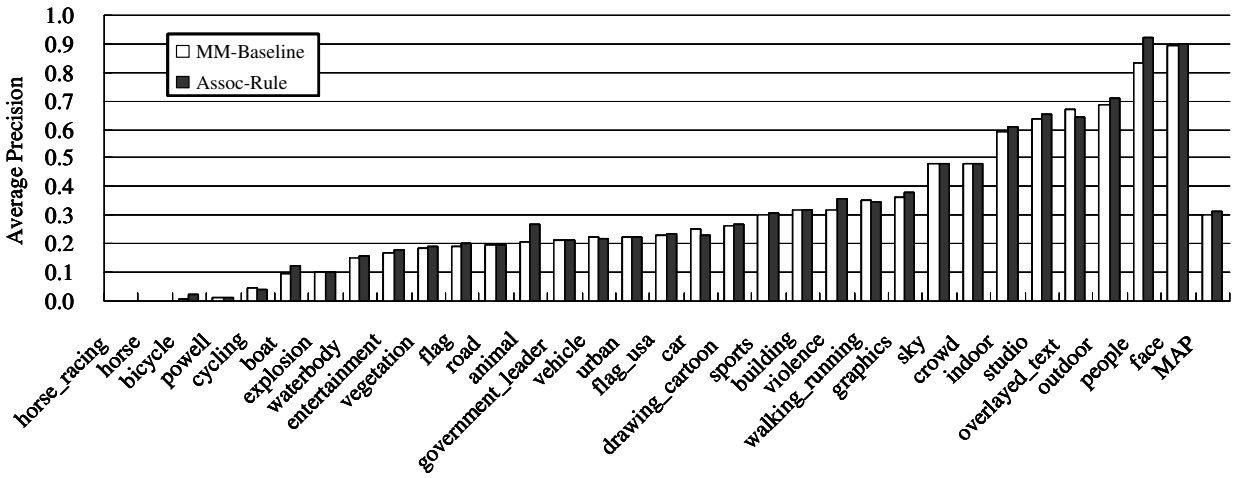

Fig. 4. Performance of our combined re-ranking using inter-concept association rules on the MM baseline classifications [6] for the 32 concepts which are found to have association rules. For these 32 concepts, combined re-ranking improves accuracy for 24 of them. The average performance gain is 3.3%.

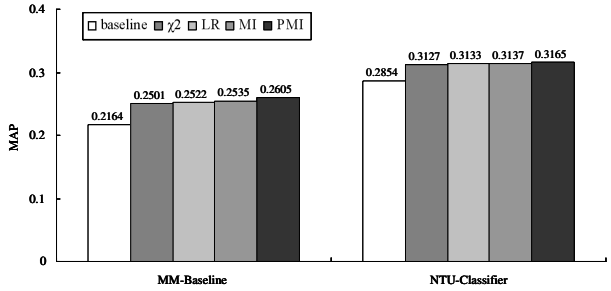

Fig. 5. Mean average precision of 101 concepts on both the MM and the NTU baseline classifiers for temporal smoothing using four different dependency measures. Pointwise mutual information consistently outperforms others. With temporal filtering, the overall performance gains for the MM baseline and the NTU classifier are 15% and 10% respectively.

that temporal filtering is effective on improving accuracy for both classifiers (around 15% and 10%, respectively) as shown in Figure 5. It shows that post-filtering is useful for classifiers with different performance. Although not much, pointwise mutual information consistently outperforms other measures. The problem with χ^2 and likelihood ratio is that it is difficult to find a proper normalization factor because of their extremely high values for strong dependency. We suspect that mutual information does not perform as well as pointwise mutual information because it also measures less relevant dependencies other than $P(l_t = 1|l_{t-k} = 1)$.

Figure 6 shows the average precision for 101 concepts using the MM baseline results [6], temporal logistic regression on prediction values [16] and our temporal smoothing on inference values. The performance gain of the temporal filter varies among concepts, ranging from -87% to 394%. Temporal filtering improves performance for more than 85% of the concepts. Overall, 72% and 59% of the concepts with improvement have more than 5% and 10% improvement respectively. Our temporal filtering outperforms logistic regression in 77 concepts. Overall, our proposed method improves the mean average precision by 20.4% and 10.9% for the MM baseline and the NTU classifier respectively. It is, not surprisingly, especially effective for the concepts that have strong temporal dependency. Among these 101 concepts, there are totally 46 concepts whose average pointwise mutual information values are larger than 3. For these concepts, the average performance gains are respectively 40% and 14% for the MM baseline and the NTU classifier. For those 20 concepts whose average PMIs are larger than 4, the average performance gains significantly reach 58% and 16%.

Yang and Hauptmann suggested that linear smoothing does not work well on improving performance of concept detection [16]. We observe the contrary. We think that there are two main reasons. First, we use the inference values instead of predication values. We notice that the performance of using inference value improves when increasing the temporal window. On the other hand, regression with prediction values leads to performance degradation when the temporal window size grows. Second, they only tested for the first order neighboring shots while we include more neighbors. Yang and Hauptmann suggested that temporal smoothing might not work because it can't pick up a missed shot at some distance. Because we consider neighbors of higher orders, we can overcome this problem. As they suggested, a missed shot is often very close to the decision boundary. Thus, a little contribution from its positive neighboring shots is often enough for it to become positive.

E. Combined Post-Filtering

We also performed experiments to evaluate the performance of the combination of both association rules and temporal filtering. We perform combined re-ranking and temporal smoothing separately first. Then, the scores of both methods are normalized to have zero-valued mean and unit standard deviation. The normalized scores are then averaged to give the final score. We applied the combined post-filtering to the 32 concepts with association rules. Figure 7 shows average precision using combined post-filtering for these concepts. The results for mean average precision for all 101 concepts and the 32 concepts that have association rules are shown in Figure 8

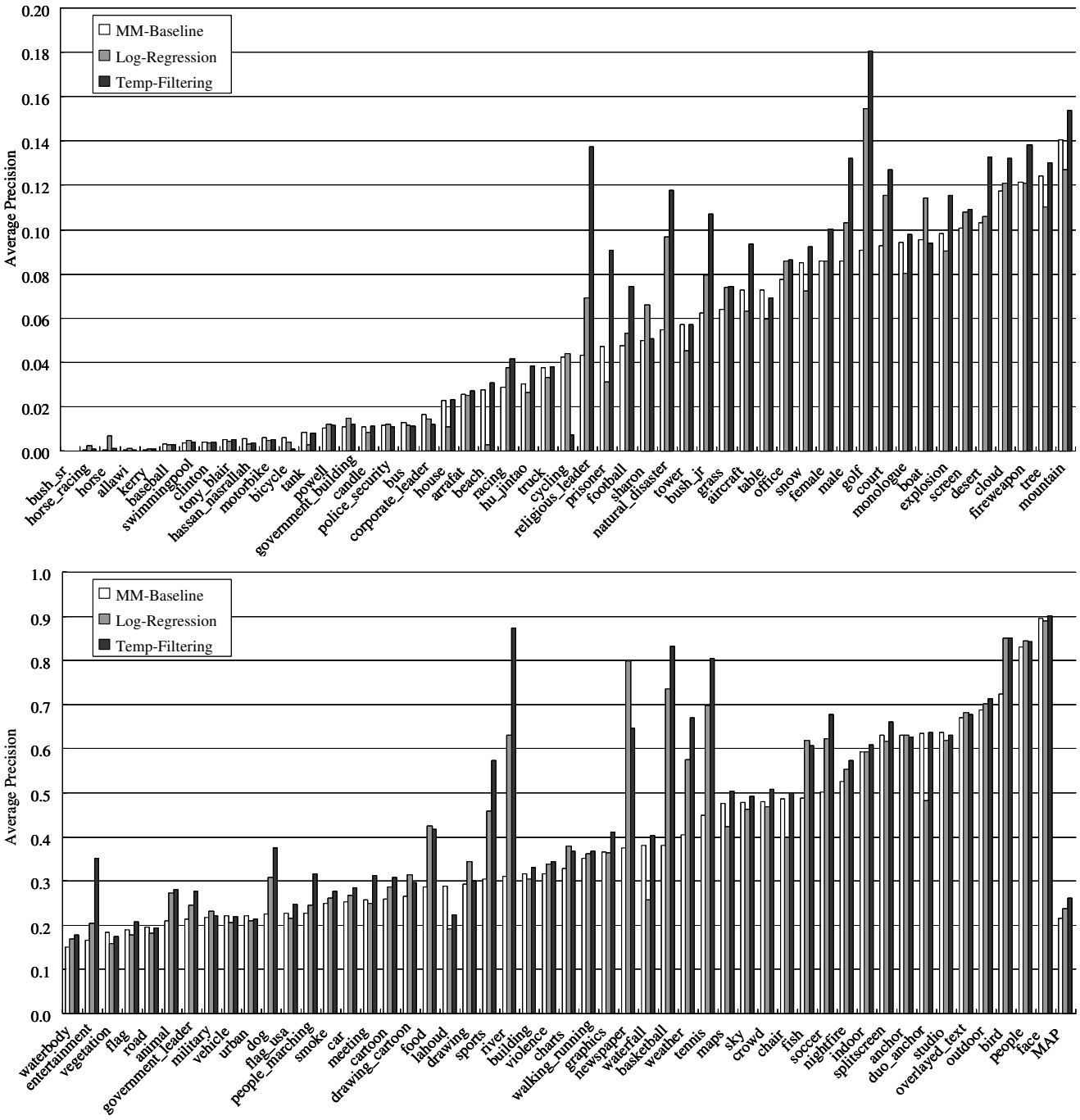

Fig. 6. Average precision for the 101 concepts from the MM baseline results [6], temporal logistic regression [16] and our proposed temporal filtering. The performance gains vary among concepts, ranging from -87% to 394%. Overall, temporal filtering improves accuracy for 85 of these 101 concepts.

and Figure 9 respectively. We observe that the combination of inter-concept association rules and inter-shot temporal filters can further improve classification.

In our training dataset, there are many concepts with few positive examples and it often leads to moderate or inferior performance for the corresponding classifiers. Figure 10 shows the mean average precision for concepts with different percentages of annotated examples. We observe that our combined post-filtering approach improves performance even for the concepts that have very few positive examples, less than 0.2% of shots annotated (i.e., around only 60 annotations). Significant performance improvement is found in both the MM baseline and the NTU classifier regardless of their annotation percentages. This shows the effectiveness of our post-filtering framework with association rules and temporal smoothing filters.

VII. Conclusion

This paper proposes a general post-filtering framework to improve performance of semantic concept detection by

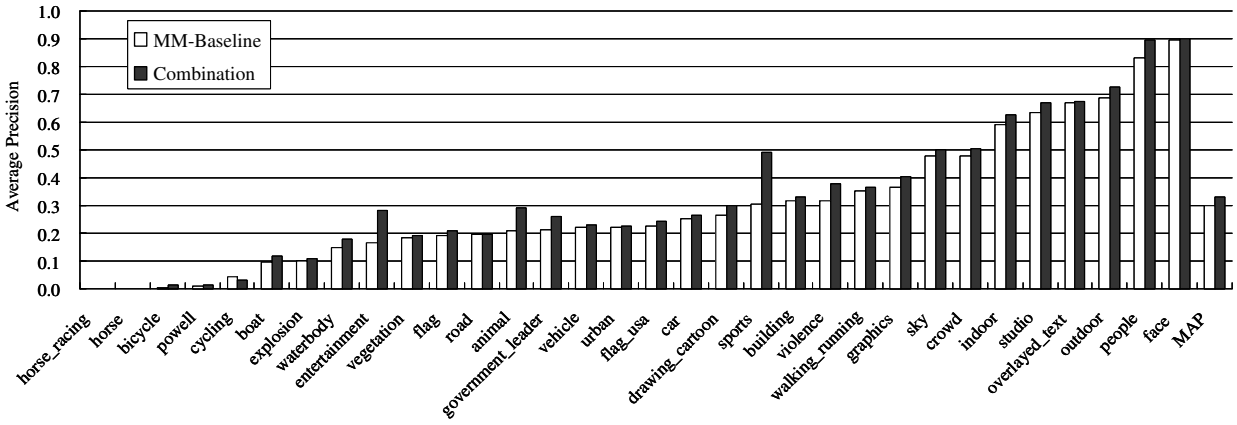

Fig. 7. A performance comparison of the MM baseline [6] and our combined post-filtering on the 32 concepts with association rules.

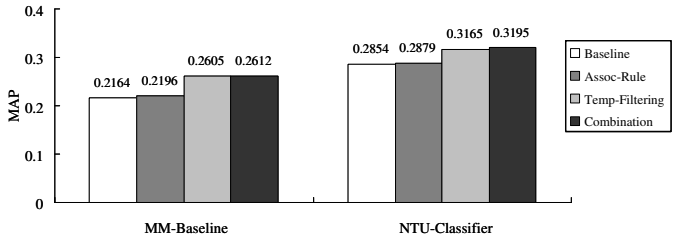

Fig. 8. Mean average precision of 101 concepts using our combined post-filtering framework.

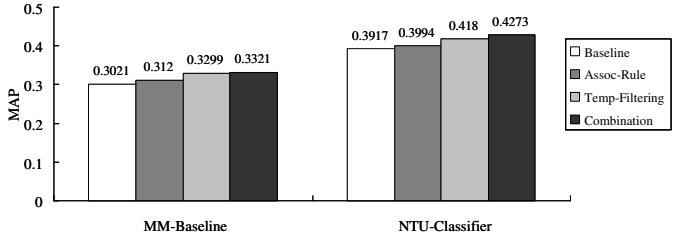

Fig. 9. Mean average precision of the 32 concepts, found to have association rules, using our combined post-filtering framework.

Ann. Ratio	Num. of Concepts	MM-Baseline		NTU-Classifier	
		Orig. MAP	Imp. Ratio	Orig. MAP	Imp. Ratio
< 0.2%	*25*	0.1306	**+32.04%**	0.2019	**+12.27%**
0.2% - 0.5%	*18*	0.1538	**+35.23%**	0.2509	**+10.41%**
0.5% - 1%	*20*	0.1925	**+31.25%**	0.2587	**+20.18%**
1% - 5%	*20*	0.1994	**+19.11%**	0.2583	**+11.15%**
5% - 10%	*8*	0.3071	**+7.94%**	0.3435	**+10.68%**
> 10%	*10*	0.5532	**+6.13%**	0.6173	**+7.15%**

Fig. 10. Performance of our combined post-filtering for concepts with different annotation percentages. (*Ann. Ratio* represents the annotation ratio. *Orig. MAP* means the original MAP and *Imp. Ratio* represents the improvement ratio.)

using association mining and temporal filtering for context knowledge discovery. To exploit inter-concept association, we have discovered non-trivial hidden association rules between concepts and proposed a re-ranking scheme to combine the associated concept detectors to improve performance. To perform inter-shot temporal dependency mining, we have proposed an effective temporal filter to integrate the predictions of neighboring shots. The combination of association rules and temporal filters can further improve the accuracy for concept detection. In addition, our post-filtering methods can be universally applied to any classifier. Our experiments on the annotated TRECVID 2005 corpus demonstrate that our framework can significantly improve the accuracy of concept detection and enhance effectiveness for concept-based video retrieval and mining.

ACKNOWLEDGEMENT

The authors would like to thank the reviewers for their insightful comments and helpful suggestions. The work was supported in part by the National Science Council of Taiwan, R.O.C., under contracts NSC95-2752-E-002-006-PAE and NSC95-2622-E-002-018.

REFERENCES

[1] C. G. M. Snoek and M. Worring, "Multimodal video indexing: A review of the state-of-the-art," *Multimedia Tools and Applications*, vol. 25, no. 1, pp. 5–35, 2005.

[2] M. R. Naphade and T. S. Huang, "Extracting semantics from audiovisual content: The final frontier in multimedia retrievel," *IEEE Trans. Neural Netw.*, vol. 13, no. 4, pp. 793–810, 2002.

[3] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. on Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, 2000.

[4] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys*, to appear.

[5] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: State-of-the-art and challenges," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 2, no. 1, pp. 1–19, 2006.

[6] C. G. M. Snoek, M. Worring, J. C. van Gemert, J.-M. Geusebroek, and A. W. M. Smeulders, "The challenge problem for automated detection of 101 semantic concepts in multimedia," in *Proceedings of ACM Multimedia*, 2006, pp. 421–430.

[7] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 2nd ed. Morgan Kaufmann, 2006.

[8] M.-S. Chen, J. Han, and P. S. Yu, "Data mining: An overview from a database perspective," *IEEE Trans. Knowl. Data Eng.*, vol. 8, no. 6, pp. 866–883, 1996.
[9] M. R. Naphade and T. S. Huang, "A probabilistic framework for semantic video indexing, filtering and retrieval," *IEEE Trans. Multimedia*, vol. 3, no. 1, pp. 141–151, 2001.
[10] M. R. Naphade, I. V. Kozintsev, and T. S. Huang, "Factor graph framework for semantic video indexing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 1, pp. 40–52, 2002.
[11] R. Yan, M.-Y. Chen, and A. G. Hauptmann, "Mining relationship between video concepts using probabilistic graphical model," in *Proceedings of IEEE Int'l Conf. Multimedia and Expo*, 2006, pp. 301–304.
[12] C. G. M. Snoek, M. Worring, J.-M. Geusebroek, D. C. Koelma, F. J. Seinstra, and A. W. M. Smeulders, "The semantic pathfinder: Using an authoring metaphor for generic multimedia indexing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1678–1689, 2006.
[13] J. R. Smith, M. R. Naphade, and A. Natsev, "Multimedia semantic indexing using model vectors," in *Proceedings of IEEE Int'l Conf. Multimedia and Expo*, 2003, pp. 445–448.
[14] W. Jiang, S.-F. Chang, and A. C. Loui, "Context-based concept fusion with boosted conditional random fields," in *IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing*, 2007, pp. 949–952.
[15] S. Ebadollahi, L. Xie, S.-F. Chang, and J. R. Smith, "Visual event detection using multi-dimensional concept dynamics," in *Proceedings of IEEE Int'l Conf. Multimedia and Expo*, 2006, pp. 881–884.
[16] J. Yang and A. G. Hauptmann, "Exploring temporal consistency for video retrieval and analysis," in *Proc. of 8th ACM SIGMM Int'l Workshop on Multimedia Information Retrieval*, 2006, pp. 33–42.
[17] X. Yin and J. Han, "CPAR: Classification based on predictive association rules," in *Proceedings of the Third SIAM Int'l Conf. on Data Mining*, 2003, pp. 369–376.
[18] W. Li, J. Han, and J. Pei, "CMAR: Accurate and efficient classification based on multiple class-association rules," in *Proceedings of the 2001 IEEE Int'l Conf. on Data Mining*, 2001, pp. 369–376.
[19] J. R. Quinlan, *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1993.
[20] H. H. Malik and J. R. Kender, "Clustering web images using association rules, interestingness measures, and hypergraph partitions," in *Proceedings of the 6th Int'l Conf. on Web Engineering*, 2006, pp. 48–55.
[21] J. R. Kender and M. R. Naphade, "Visual concepts for news story tracking: Analyzing and exploiting the NIST TRECVID video annotation experiment," in *Proceedings of IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 2005, pp. 1174–1181.
[22] L. Xie and S.-F. Chang, "Pattern mining in visual concept streams," in *Proceedings of IEEE Int'l Conf. Multimedia and Expo*, 2006, pp. 297–300.
[23] A. F. Smeaton, P. Over, and W. Kraaij, "Evaluation campaigns and TRECVid," in *Proceedings of the 8th ACM Int'l Workshop on Multimedia Information Retrieval*, 2006, pp. 321–330.
[24] J. C. Platt, "Probabilities for SV machines," *Advances in Large Margin Classifiers*, pp. 61–74, 2000.
[25] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," 2001, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[26] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases," in *Proceedings of Int'l Conf. on Very Large Data Bases*, 1994, pp. 487–499.
[27] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.
[28] C. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.

Ken-Hao Liu received the B.S. degree in electrical engineering and the Ph.D degree in computer science from the National Taiwan University, Taipei, Taiwan, in 2001 and 2007, respectively. His research interests include data clustering, data streams management systems and multimedia data mining.

Ming-Fang Weng received the B.S. degree and M.S. degree in computer science and information engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1998 and 2000, respectively. He spent five years as an engineer working for Institute for Information Industry. Currently, he is a Ph.D. student in Department of Computer Science and Information Engineering, National Taiwan University. His research interests include computer vision, machine learning and semantic computing for multimedia content and information system.

Chi-Yao Tseng received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan in 2004, and now is the Ph.D. candidate in Network Database Lab, led by professor Ming-Syan Chen, of Graduate Institute of Electrical Engineering at National Taiwan University, Taipei, Taiwan. His current research interests include sequential pattern mining, email spam detection, and multimedia data mining.

Yung-Yu Chuang received his B.S. and M.S. from National Taiwan University in 1993 and 1995 respectively, Ph.D. from University of Washington at Seattle in 2004, all in Computer Science. He is currently an assistant professor in the Department of Computer Science and Information Engineering, National Taiwan University. His research interests include multimedia data mining, computer vision, digital photography and real-time rendering. He is a member of the IEEE and a member of the ACM.

Ming-Syan Chen received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, and the M.S. and Ph.D. degrees in Computer, Information and Control Engineering from The University of Michigan, Ann Arbor, MI, USA, in 1985 and 1988, respectively. Dr. Chen is currently a professor in Electrical Engineering Department, National Taiwan University, Taipei, Taiwan. He was a research staff member at IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA from 1988 to 1996. His research interests include database systems, data mining, mobile computing systems, and multimedia networking, and he has published more than 200 papers in his research areas. In addition to serving as program committee members in many conferences, Dr. Chen served as an associate editor of IEEE Transactions on Knowledge and Data Engineering (TKDE) from 1997 to 2001, is currently on the editorial board of Very Large Data Base (VLDB) Journal, Knowledge and Information Systems (KAIS) Journal, Journal of Information Science and Engineering, and International Journal of Electrical Engineering, and is a Distinguished Visitor of IEEE Computer Society for Asia-Pacific from 1998 to 2000, and also from 2005 to 2007 (invited twice). He served as the international vice chair for INFOCOM 2005, program chair of PAKDD-02 (Pacific Area Knowledge Discovery and Data Mining), program co-chair of MDM-03, program vice-chair of IEEE ICDE-06, IEEE ICDCS-05, ICPP-03, and VLDB-2002, and many other program chairs and co-chairs. He received the Outstanding Innovation Award from IBM Corporate in 1994 for his contribution to parallel transaction design and implementation for a major database product, and numerous awards for his research, teaching, inventions and patent applications. Dr. Chen is a Fellow of IEEE and a Fellow of ACM.