Video Collaborative Annotation Forum: Establishing Ground-Truth Labels on Large Multimedia Datasets

Ching-Yung Lin, Belle L. Tseng and John R. Smith

IBM T. J. Watson Research Center
19 Skyline Dr.
Hawthorne, NY 10532, USA
{chingyung, belle, jsmith}@us.ibm.com

ABSTRACT

We developed a new version of The VideoAnnEx, a.k.a. IBM MPEG-7 Annotation Tool, for collaborative multimedia annotation task in a distributed environment. The VideoAnnEx assists authors in the task of annotating video sequences with MPEG-7 metadata. Each shot in the video sequence can be annotated with static scene descriptions, key object descriptions, event descriptions, and other lexicon sets. The annotated descriptions are associated with each video shot or regions in the keyframes, and are stored as MPEG-7 XML file. We proposed a forum to collaboratively annotate semantic labels to the NIST TRECVID 2003 development set. From April to July 2003, 111 researchers from 23 institutes worked together to associate 198K of ground-truth labels (433K after hierarchy propagation) to 62.2 hours of videos. This large set of valuable ground-truth data is publicly available to the research community, especially for multimedia indexing and retrieval, semantic understanding, and supervised machine learning fields.

1. INTRODUCTION

The growing amount of digital video is driving the need for more effective methods for indexing, searching, and retrieving of video based on its content. While recent advances in content analysis, feature extraction, and classification are improving capabilities for effectively searching and filtering digital video content, the process to reliably and efficiently index multimedia data is still a challenging issue. Besides, in order to learn audio-visual concept models, supervised learning machines also require ground truth labels being associated with training videos.

We implemented a *VideoAnnEx* MPEG-7 annotation tool to allow authors to semi-automatically annotate video content with semantic descriptions [9][19]. It is one of the first MPEG-7 annotation tools being made publicly available. The tool explores a number of interesting capabilities including automatic shot detection, key-frame selection, automatic label propagation, and template annotation propagation to similar shots, and importing, editing, and customizing of ontology and controlled term lists. In Feb. 2003, we released the *VideoAnnEx* v2.0, which is an MPEG-7 annotation system, including clients that are similar to the previous stand-alone versions and administrative web interfaces for ontology management, user management, group management, and annotation task management.

Given the lexicon and video shot boundaries, visual annotations can be assigned to each shot by a combination of label prediction and human interaction. Labels can be associated to a shot or a region on the keyframe. Regions can be manually selected from the keyframe or injected from the segmentation module. Annotation of a video is executed shot by shot without permuting their time order, which we consider an important factor for human annotators because of the time-dependent semantic meanings in videos. Label prediction utilizes clustering on the keyframes of video shots in the video corpus or within a video. By the time a shot is being annotated, the system predicts its labels by propagating the labels from the last shot in time within the same cluster. Annotator can accept these predicted labels or select new labels from the hierarchical controlled-term lists. All the annotation results and descriptions of ontology are stored as MPEG-7 XML files.

Other MPEG-7 annotation tools are available publicly. *MovieTool* is developed by Ricoh for creating video content descriptions conforming to MPEG-7 syntax interactively [14]. While the use of MPEG-7 in *VideoAnnEx* is transparent to the users, *MovieTool* requires users to be familiar with MPEG-7 and edit the XML files directly using MPEG-7 tags. The Know-Center released a MPEG-7 based annotation and retrieval tool for digital photos [16]. The IBM Multimedia Mining Project released a *Multimodal Annotation Tool*, which is derived from an earlier version of *VideoAnnEx* with special features with audio signal graphs and manual audio segmentation functions [2].

Some other media annotation systems, including collaborative annotations, have been developed for various purposes. Bargeron et. al. developed an Microsoft Research Annotation System (MRAS), which is a web-based system for annotating multimedia web content [4]. Annotations include comments and audio in the distance learning scenario. Comparing with *VideoAnnEx,* MRAS does not make use of lexicon, shots nor personalized management system. Steves et. al. developed a Synchronous Multimedia and Annotation Tool (SMAT) [15]. SMAT is used to annotate images. There is no granularity for video annotations nor controlled-term labels. Nack and Putz developed a semi-automated annotation tool for audio-visual media in news [12]. This is a stand-alone application. Users have to specify shots manually an. It does not use controlled-term items, either. The European Cultural heritage Online (ECHO) is developing a multimedia annotation tools which allows people to work collaboratively on a resource and to add comments to it [6].

Year	Data	# of Annotators	Labels	Source
2001	11 hrs	5 -- IBM	85 visual: 8 events, 28 scene, 49 objects	NASA, BBC
2002	23 hrs (~13K shots)	8 – IBM, 4 – Tsing-Hua U.	123 visual: 28 events, 36 scenes, 51 objects	Internet Movie Archive (1940s – 1970s)
2003	62 hrs (46K shots)	111 -- Accenture, CMU, CLIPS, Columbia U., CWI, Dublin, EPFL, EURECOM, Fudan U., IBM, Intel, KDDI, Tsing-Hua U., U. Singapore, TUT, UCF, U. Chile, UniDE, U. Geneva, U. Glasgow, U. Mass, UNC, U. Oulu	133 – audio & visual: 35 A&V events, 38 visual scenes, 11 sounds, 49 visual objects	CNN & ABC news, (1998) C-SPAN (1998, 2000)

Table 1: Completed Annotation Tasks using the VideoAnnEx System

Table 1 shows a list of completed annotation tasks using the VideoAnnEx system. In 2001, 5 researchers in IBM annotated 11 hours of video with 85 controlled-term concepts. In 2002, 123 visual concepts were annotated on 23 hours of video. These annotated labels were served as the foundation of IBM's TREC Video Retrieval Systems in 2001 and 2002 [18][1]. This tool is further applied in the video collaborative annotation forum in 2003 to establish 433K of semantic labels on 62 hours of video.

Overview of the Video Collaborative Annotation Forum

In the wrap-up discussions on TREC 2002 conference, many participants agreed with the importance of common ground truth for system development and evaluation. Such a large set of ground truth labels should benefit semantic concept. Therefore, in March 2003, we proposed a forum to collaboratively annotate semantic labels to the NIST TRECVID 2003 development set using *VideoAnnEx* annotation system. The objective of this forum is to establish ground-truth labels on large video datasets as common assets to research society. They are meant to promote progress in video content modeling, understanding, indexing and retrieval researches and simplify evaluation across systems.

The first phase of the forum was to annotate labels on the NIST TREC Video Retrieval Evaluation 2003 (TRECVID) development video data set. This development video data is part of the TRECVID 2003 video data set which includes:

- ~120 hours (241 30-minute programs) of ABC World News Tonight and CNN Headline News recorded by the Linguistic Data Consortium from late January through June 1998 and
- ~13 hours of C-SPAN programming (~ 30 mostly 10- or 20-minute programs) about two thirds 2001, others from 1999, one or two from 1998 and 2000. The C-SPAN programming includes various government committee meetings, discussions of public affairs, some lectures, news conferences, forums of various sorts, public hearings, etc.

The total TRECVID 2003 video set is about 104.5 GB of MPEG-1 videos, that includes the development set (51.6 GB, 62.2 hours including 3390 minutes from ABC & CNN, 340 minutes from C-SPAN) and the test set (52.9 GB, 64.3 hours including 3510 minutes from ABC & CNN, 350 minutes from C-SPAN).

TRECVID 2003 participants have the option to join the Video Collaboration Annotation Forum, which establishes the common annotation set that all forum participants agree to contribute annotations. The set of resulting common annotations was available to everyone participating in the forum. Based on these common development set and common annotation set, forum participants can develop Type 1 (as specified by NIST) feature/concept extraction system, search system or donation of extracted features/concepts. This set of common annotation was available to the public after the TRECVID 2003 workshop [11].

2. OVERVIEW OF VIDEOANNEX COLLABORATIVE ANNOTATION SYSTEM

VideoAnnEx v2.0 allows collaborative annotation among multiple users through the Internet (see Figure 1). Users of the collaborative *VideoAnnEx* are assigned user IDs and passwords to access a central server, called the *VideoAnnEx CA* (collaborative annotation) *Server*. The *VideoAnnEx CA Server* centrally stores the MPEG-7 data files, manages the collaboration controls, and coordinates the annotation sessions. For collaborative annotation, there are three categories of user access to the *VideoAnnEx CA Server*, and they are: (1) project manager, (2) group administrator, and (3) general user. The project manager sets up the project on the *VideoAnnEx CA Server*, creates the different groups' IDs and allocates video responsibilities to groups. The group administrator coordinates the annotations of the assigned videos and distributes the annotation tasks among the individual general users. The general users are the end users who actually perform the annotation task on the *VideoAnnEx v2.0* Client.

There are four major components in the *VideoAnnEx* clients. First, video segmentation is performed to cut up the video sequence into smaller video units. Second, semantic lexicon is defined in order to regulate the video content descriptions. In the collaborative annotation environments, the first two steps may be replaced by downloading a shot segmentation MPEG-7 file and an MPEG-7 lexicon file from the *VideoAnnEx CA Server*. Third, an annotator labels the video segments with the semantic. An automatic annotation-learning component can be used to speed up the annotation task. Fourth, the MPEG-7 descriptions of the annotation process are directly outputted from the *VideoAnnEx*. The goal of the video annotation is to categorize the semantic content of each video unit or regions in the keyframes and output the MPEG-7 XML description file. In the collaborative annotation mode, the users can check in the annotated XML to the server, which controls the versions of annotations. Some additional functions such as template matching and label editing were added to the *VideoAnnEx v2.0 client*. In the following subsections, we first introduce the user interface and then describe the main client components in further detail. The label editing function includes copying, pasting and deleting annotation labels of an individual shot or groups of shots. This is similar to general operations in the common word editing tools, that we will not show more details.

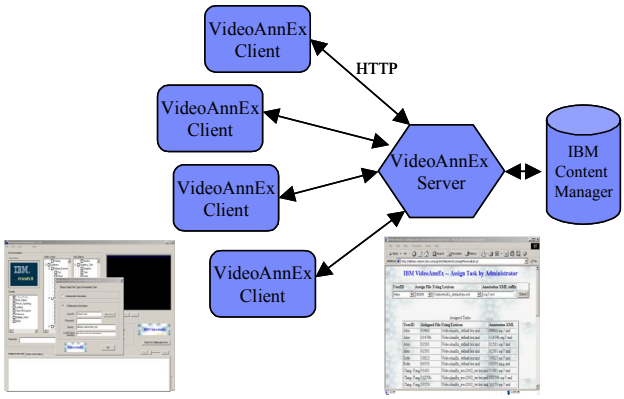

Figure 1: VideoAnnEx Collaborative Annotation System

2.1 Graphical User Interface

The *VideoAnnEx* is divided into four graphical sections as illustrated in Figure 2. On the upper right-hand corner of the tool is the *Video Playback* window with shot information. On the upper left-hand corner of the tool is the *Shot Annotation* with a key frame image display. On the bottom portion of the tool is two different *Views Panel* of the annotation preview. A fourth component, not shown in Figure 2, is the *Region Annotation* pop-up window for specifying annotated regions. These four sections provide interactivity to assist authors of the annotation tool.

The *Video Playback* window displays the opened MPEG video sequence. As the video is played back in the display window, the current shot information is given as well. The *Shot Annotation* module displays the defined semantic lexicons and the key frame window. The key frame is a representative image of the video shot segment, and thus offer an instantaneous recap of the whole video shot. This is the region where the annotator selects the descriptions for the video segment. The *Views Panel* displays two different previews of representative images of the video. The *Frames in the Shot* shows all the I-frames as representative images of the current video shot, while the *Shots in the Video* view (as in the bottom of Figure 2) shows all the key frames of each shot as representative images over the entire video. As the annotator labels each shot, the descriptions are displayed below the corresponding key frames in the *Shots in the Video* view. Furthermore after the MPEG-7 descriptions are saved into an XML file, anyone can load and review these files at a later time by previewing the annotations at this views panel. The *Region Annotation* window allows the author to associate a rectangular region with a labeled text annotation. After the text annotations are identified on the *Shot Annotation* window, each description can be associated with a corresponding region on the selected key frame of that shot. More details are shown in [17].

2.2 Video Shot Segmentation

A short video clip can be simply annotated by describing its content in its entirety. However when the video is longer, annotation of its content can benefit from segmenting the video into smaller units. A video shot is defined as a continuous camera-captured segment of a scene, and is usually well defined for most video content. Given the shot boundaries, the annotations are assigned for each video shot.

The *VideoAnnEx* Shot Segmentation component is based on the frame differencing of the color and motion histogram. This algorithm uses sampled RGB color histograms in the I- and motion histograms in the P-frames of video sequences. Heuristic rules are designed to make the algorithms robust to flashes and noises. Shot segmentation process is executed in the background thread. Thus, users can start annotating videos right after they open an MPEG-1 file. Shot segmentation information can be saved or loaded in the MPEG-7 XML. An example of MPEG-7 shot segmentation file can be found in [17].

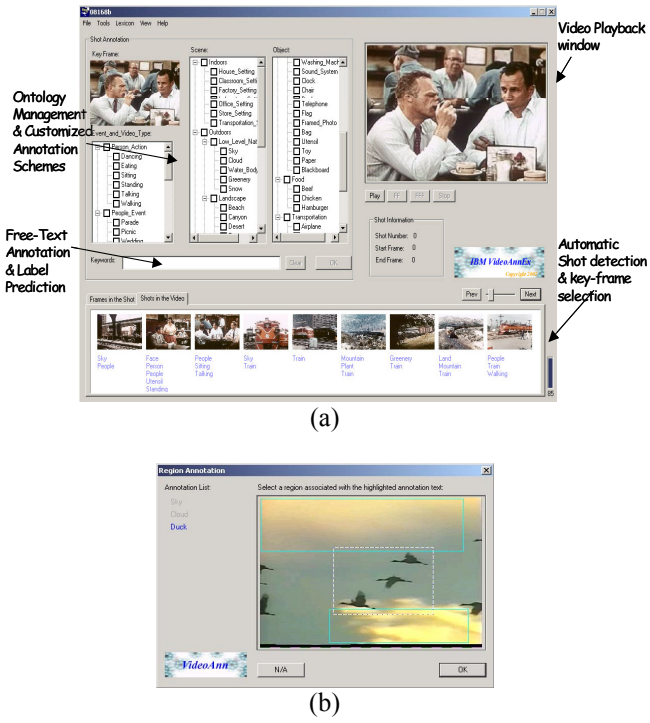

Figure 2: Graphic Interface of VideoAnnEx.

2.3 Ontology Editor and Controlled Item List

Given the segmentation of video content into video shots, the second step is to define the semantic lexicon in which to label the shots. A video shot can fundamentally be described by three attributes. The first is the background surrounding of where the shot was captured by the camera, which is referred to as the *static scene*. The second attribute is the collection of significant subjects involved in the shot sequence, which is referred to as the *key object*. Lastly, the third attribute is the corresponding action taken by some of the key objects, which is referred to as the *event*. These three types of lexicon define the vocabulary for our video content.

Using the defined vocabulary for static scenes, key objects, and events, the lexicon is imported into *VideoAnnEx*. Note that the set of lexicon as well as the category attributes are dependent on the application, and can be easily generated and modified using *VideoAnnEx*. Details of this ontology-editing component can be seen in [10].

2.4 Annotation Learning

Annotation Learning is a characteristic that helps speed up the annotation speed. Right before the user annotates a video shot, predicted labels would have been shown on the "keyword" field of the *VideoAnnEx*. The prediction functionality on the current public-release version of *VideoAnnEx* v. 1.5 propagates labels from the visually most similar annotated shot. When *VideoAnnEx* opens a video, a background thread calculates the feature-space distances between shots in the video. A distance combining both the feature space distance and the temporal space difference of shots are calculated to decide the visually closest shot. This propagation mechanism has been shown quite effective and helpful in speeding up the annotation task. A new mechanism of incorporating pre-trained models is under development.

2.5 MPEG-7 Video Segment Description

The ISO standardized MPEG-7 defines the compatible scheme and language to represent semantic meaning of multimedia content. Our MPEG-7 output is the Video Segment Description Scheme. In MPEG-7, each video shot is defined as a Video Segment. Furthermore, the embedded <SpatioTemporal Decomposition> tag allows us to specify the region location and the corresponding text annotation in a key frame. An example of the output XML file can be found at [19].

2.6 Template Matching

We developed a template matching mechanism to help users to detect text, logo regions in the shots with similar texts/logos in the same locations. Users first select a region from a shot. Then the client tool will automatically detect the similarity of the same region in other shots of the video and propagates the labels. We used color and edge features for

template matching. Only the regions that correspond to the location of templates are tested, and the result S is a binary decision on the test frames.

$$S = \delta(S_C > \tau'_C) \,\&\, \delta(S_E > \tau'_E)$$

and

$$S_C = \frac{1}{N}\sum_n \delta(d(P_C, P_{MC}) > \tau_C)$$

$$S_E = \frac{1}{N}\sum_n \delta(d(P_E, P_{ME}) > \tau_E)$$

where C represents the color features and E represents the edge features. Four thresholds $\tau_C, \tau_E, \tau'_C, \tau'_E$, were used. $\delta()$ is the binary decision function, and d() represents the Euclidean distance of the test regions in the feature space. N is the number of pixels in that region. After binary decisions were made to the individual shots in a video, two consecutive temporal median filters were used to eliminate randomly false classified shots. .The window size of both median filters is five shots. This template matching functions has also been applied as a news/commercial detector [3].

3. OPERATIONS OF COLLABORATIVE ANNOTATION SYSTEM

The *VideoAnnEx CA Server* provides a web interface for administrators and users to coordinate registration activities and manage annotation assignments. In the initial stage, the project manager will assign to each group administrator a group ID and password to manage the group configurations. Afterwards, each group administrator is responsible for the coordination of its individual users through the *VideoAnnEx CA Server* web interface. The general users also access the *VideoAnnEx CA Server* to perform registration and follow up on their annotation tasks. In this section, we will describe how the *VideoAnnEx CA Server* is used by the group administrators and the general users. These steps described below are advised to be followed in the prescribed order.

Figure 3 describes how a user uses the *VideoAnnEx Client* for the annotation task. She first logs in the system using the client interface, then selects project and gets assigned lexicon and downloads the previous annotations. This finishes the check out process. After a video is checked out, it will be locked in the server so that no other annotator can annotates that video until this user checks in her annotation. She can annotate the video by saving the videos at local corpus and annotates video off-line. After the annotation is done, then she checks in the video to the server. This will unlock the video so it can be annotated by other users. More detailed description as well as example screen shots can be seen at [10].

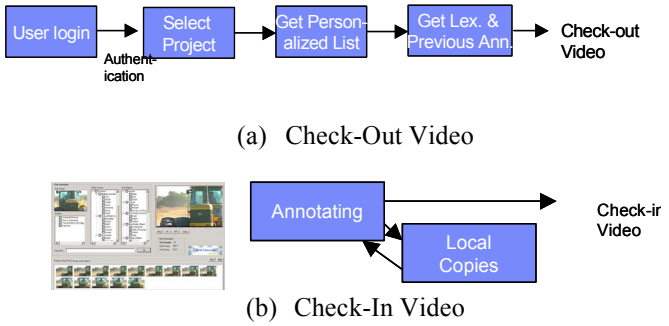

Figure 3: Client Interaction with Annotation Server

3.1 Registration

After the project manager assigns to each group administrator its group ID and password, the administrator goes to the *VideoAnnEx CA Server* home web page to register the group. Thus during the first visit, the group administrator selects the "New User Registration" link to start the group registration. At the user registration page, the group administrator creates a new user ID for herself and click the submit button. After the new user ID is accepted by the *VideoAnnEx CA Server*, the individual must enter a user profile, which requires the full name, password, email address, and affiliation. After completing the user profile, submit the form to the server.

When a user finishes setting up her user ID, password and user profile, each user must select the corresponding project, group, and role. These advanced selections allow the user to designate the specific project and responsibilities. Project denotes the collaborative project that the user is participating in. For example, there are currently two projects, TREC 2002 and TREC 2003. Group specifies the local group community that the user belongs to. Role refers to the responsibility of the user. There are two roles to choose from, Administrator and General User. Subsequently, the group administrators should choose Administrator, and the end users choose General User. Finally, a registration password is required to validate the new user. The group administrators will be receiving passwords from the project manager. The general users will receive passwords from their corresponding group administrators.

Finally after the user registration is completed, users are welcomed with a congratulations page with a summary of the project and group selections. Also, a link for the newly registered user to log in to the VideoAnnEx CA Server is provided.

3.2 User and Administrator Login

After a user completes the registration process described in the previous section, the user can return to the VideoAnnEx *CA Server* home web page to login. After the *VideoAnnEx CA Server* has verified the registered user, the individual must choose the appropriate collaborative annotation project they wish to work on. As soon as a user enters a collaborative project, the assignment management view is show to the user.

3.3 Assignment Management

After a user is registered at the VideoAnnEx CA Server, she can login at the home page and enter a collaborative project where the assignment management view is displayed, as shown in Figure 4. The group administrator and the general users will get a slightly different view. Figure 4 illustrates the view that a group administrator will see, which includes additional access features. In the assignment management page, both the administrators and users will see the assignment list for their entire group. This assignment list will include entries for user names, their assigned video files, corresponding lexicon files, resulting annotation XML files, and status of their latest activities. The activities can be in one of the following annotation states: (1) no action, (2) checked out, (3) updated, and (4) completed annotation, which have corresponding color coded highlighting. A summary of the entire group's annotation status is displayed in the bottom row of Figure 4 called Status Statistics. Note that the group summary status statistics is also viewable by all users of the project.

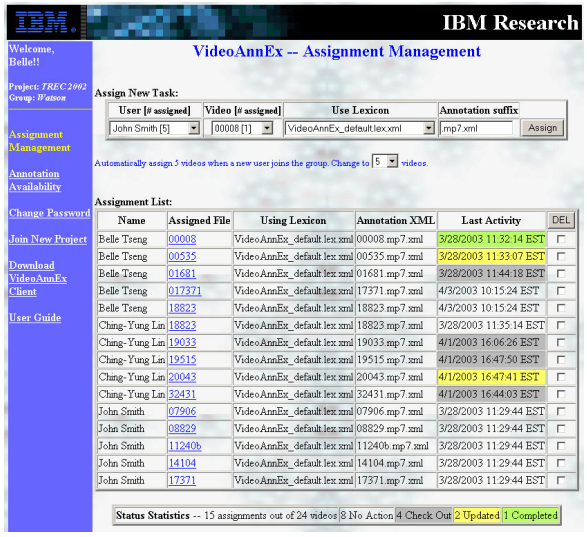

Figure 4: Interface for Assignment Management on the VideoAnnEx CA Server

In addition to the assignment listing and status statistics, the group administrator has additional functions. In the assignment list, the administrator has an additional column called "DEL", which allow the administrator to delete the corresponding video annotation assignment. The administrator is given the power to reallocate the group annotation with this delete functionality. Furthermore, the administrator can allocate additional annotation tasks by using the "Assign New Task" table. Using drop down menu selections, the administrator can assign new videos to users in her group. Another useful feature is to automatically assign a fixed number of video annotation tasks to each newly registered member in the group. This can be performed by selecting that fixed number. In Figure 4, when a new user joins the group, the *VideoAnnEx CA Server* will automatically assign 5 videos to that user.

Using the Assignment Management page of the *VideoAnnEx CA Server*, general users can track their annotation status and group administrators can manage the group annotation responsibilities. The Assignment Management allows flexibility in allocating video annotation tasks while keeping track of everyone's progress.

3.4 Annotation Download

When a group has finished their allocated annotation tasks, the group is permitted to download all the complete project annotations. On the annotation availability page of the *VideoAnnEx CA Server*, one can see the group task status list by the different groups of the project. Entries include the group name, administrator's name, their allocated video assignments, and the annotation status. Whenever a group finishes all their assigned annotations, they will be able to download the annotations.

3.5 Collaborative Annotation Client

The *VideoAnnEx CA Server* provides a web interface for group administrators/users to register themselves and monitor their annotation responsibilities. The *VideoAnnEx Client* tool allows the actual annotations of video sequences and the registration of general users. When an user open the *VideoAnnEx v2.0 Client*, the mode selection window pops up to ask the user to choose the annotation mode. There are two modes, independent annotation and collaborative annotation.

In the Collaborative annotation mode, we need to specify the user ID, password, the URL of *VideoAnnEx CA Server*, and the local video corpus directory. A new general user can click on "New User" to register his information on the server. Note that the local video corpus is the working directory of your video data. It can be a mapped network drive or a directory on your PC. This working directory should be used to contain the videos. If videos are not in any of the local or mapped-drive directories, then there will be a selection appeared in later session which allows users to download or copy video files to this directory.

After login, user can select a project to work with, and then will get an assignment page from the *VideoAnnEx Server*. The annotator can choose a file to annotate. He can also see the story board of the video via the links under "Assigned File". *VideoAnnEx Client* will check the availability of the video. If the selected video is not in the specified local directory, then the annotator can choose to download it from other directory. Finally, the *VideoAnnEx Client* will download both lexicon and annotation MPEG-7 XML files from the server and allow the annotator to start or resuming annotating the selected video. After these steps, then users can start the annotation task. Detailed instruction on the annotation steps and tips can be seen in [10].

4. VIDEO COLLABORATIVE ANNOTATION FORUM

The objective of the video collaborative annotation forum is to establish ground-truth labels on large video datasets as common assets to research society. They are meant to promote progress in video content modeling, understanding, indexing and retrieval researches and simplify evaluation across systems.

The total TRECVID 2003 video set is about 104.5 GB of MPEG-1 videos, that includes the development set (51.6 GB, 62.2 hours including 3390 minutes from ABC & CNN, 340 minutes from C-SPAN) and the test set (52.9 GB, 64.3 hours including 3510 minutes from ABC & CNN, 350 minutes from C-SPAN).

Based on these common development set and common annotation set, forum participants can develop Type 1 (as specified by NIST) feature/concept extraction system, search system or donation of extracted features/concepts.

4.1 Phases of the Annotation Forum

We built an MPEG-7 Annotation Tool to facilitate multimedia annotation tasks for general users. Use of MPEG-7 is transparent to users so that no prior knowledge on MPEG-7 is required. Various features, such as shot segmentation, ontology editing, storyboard generation, etc., are provided. In the next phase, we are developing a new version for collaborative multimedia annotation task in a distributed environment.

There were five steps on the development of this collaborative annotation forum: In TREC 2002 conference, many participants agreed with the importance of common ground truth for system development and evaluation. Thus, **from Dec. 2002 to Feb. 2003**, we extended our existing stand-alone VideoAnnEx annotation into a collaborative annotation system (*VideoAnnEx v 2.0*). As discussed in Section 2, this system provides a web interface for administrators and users to coordinate registration activities and manage lexicon and annotation assignments. **From March 2003 to May 2003**, we initialized discussions, made proposal, provided testing environments, accepted group signed-in, and discussed the 1st draft of controlled-term lexicon. We revised *VideoAnnEx* from v 2.0 to v 2.1.2 according to user feedback. We added several editing functionality and the multi-region concept annotation functionality in the tool. Twenty-Three groups signed in this annotation forum.

From May 2003 to June 2003, we assigned 37 sample videos to groups, debugged/improved the client tool, finalized the lexicon, assisted some groups to get videos and set up experimental environment, and checked the validity of annotation results.

In the next step, **from June 2003 to July 2003**, we assigned 106 videos to groups. In this step, forum participants completed the annotation of the TRECVID 2003 development set. We cleaned the annotated XMLs, corrected some typo and some irregular MPEG-7 XML files. The final set was released to the forum participants in July 14, 2003.

In **October and November 2003**, we sent a questionary survey to the participants, collected their opinions on the forum, and presented the report in the TRECVID conference. The annotation result was released to public after the conference.

4.2 Lexicon

The lexicon used in this annotation task was drafted by IBM Research TREC Video team. It was finalized after the forum participants test annotating 37 example videos and then finalized by the common agreement of forum participants. A draft of this lexicon was first developed by IBM for the annotation task of TREC Video Retrieval Benchmarking 2001 [18]. We categorized the lexicon items into event, scene and objects. In 2001, we looked at the content of 11 hours of NASA and BBC videos and developed a lexicon consisted of 85 visual labels. These label items were hierarchically organized. In 2002, we expanded the original lexicon by looking at the training examples that are movies from 1940s to 1970s. Some part of 2001 lexicon was deleted, e.g., outer space planets. And, more life-related items were added to the lexicon. 123 visual labels were used in our TREC 2002 video annotation [1]. In 2003, we looked at the training video shots and added audio labels and more events. This 133-item lexicon is consisted of 35 audio and visual events, 38 visual scenes, 11 sounds and 49 visual objects. A list of these 133 items as well as their hierarchy is shown in Figure 5.

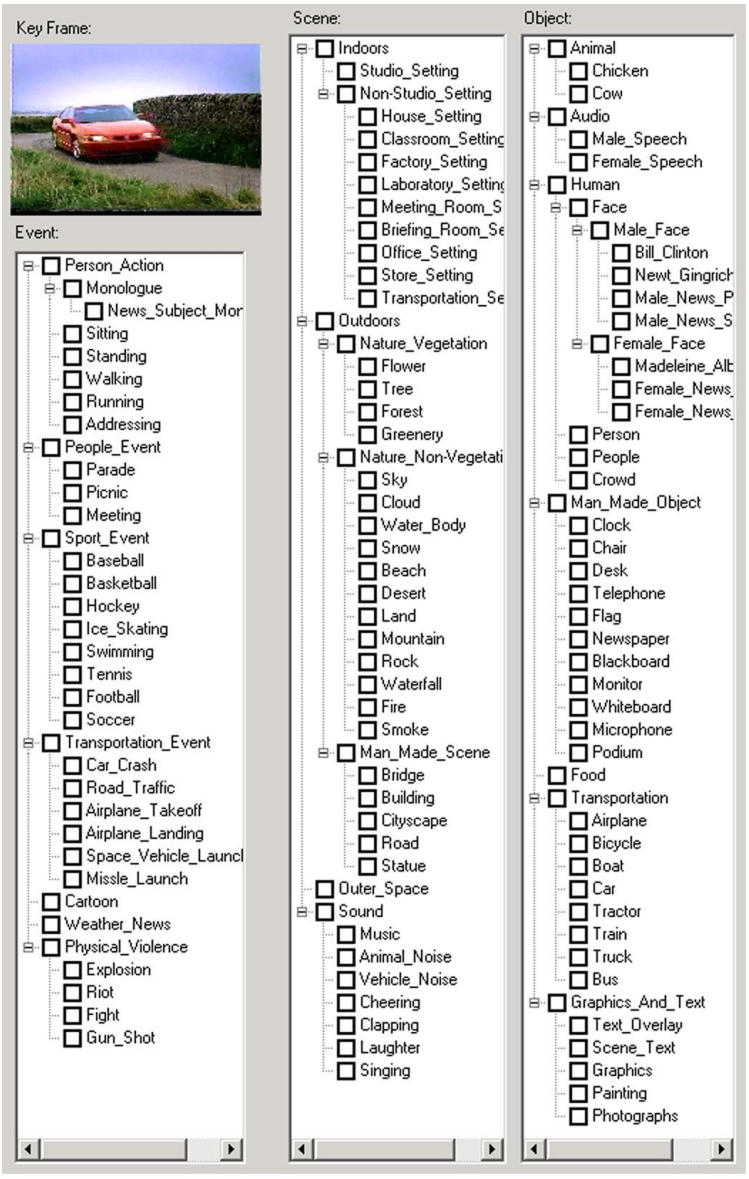

Figure 5: The Taxonomy used in the annotation forum

There were not specific definitions or descriptions on individual lexicon items. However, NIST defined some items for the purpose of serving as a guideline for high-level feature (concept) detection. These descriptions are meant to be clear to humans, e.g., assessors/annotators creating truth data and system developers attempting to automate feature detection. They are not meant to indicate how automatic detection should be achieved. If the concept is true for some frame (sequence) within the shot, then it is true for the shot; and vice versa. A list of NIST defined lexicon items is shown in the Appendix.

4.3 Annotation Guidelines

These guidelines were enacted at the beginning of the forum. They served as a common agreement among forum annotators.

- Common shot boundaries and key frames of the development video set are provided by volunteer TRECVID 2003 participants. These information will be stored on the VideoAnnEx CA Server.

- Because the automatically shot boundaries & key frames may not be perfect, annotators can/should manually improve the accuracy of shot boundaries and key frames of shots using the VideoAnnEx Client.

- For each shot, labels are associated on the whole shot and on the rectangular regions of the key frame of the shot.
- (Updates on VideoAnnEx v2.1) Multiple regions can be selected on a key frame of the shot using the same label.
- (Updates on VideoAnnEx v2.1) Templates can be used to automatically annotate logos and overlay text areas for the videos.
- Annotators can specify additional keywords, if that are not covered by the lexicon.
- Annotators only need to select child label items in the lexicon hierarchy. Individual system of participants should automatically propagate labels to their parent nodes.
- Lexicon is designed for the appropriate description of the high-level feature (concept) in this video set.

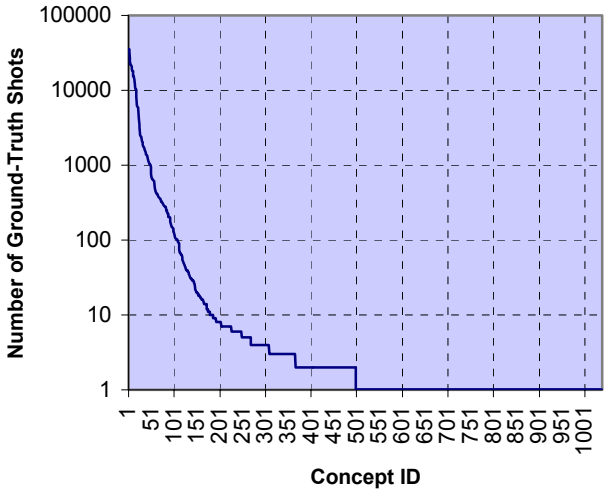

Figure 6: Concepts and their numbers of positive examples in the development set

- The final common lexicon will be mutually decided upon by the forum participants.

4.4 Results of Forum Annotation Task and User Studies

From April to July 2003, 111 researchers from 23 institutes worked together to associate 197,822 of ground-truth labels (433,338 after hierarchy propagation) at 62.2 hours of videos. 1038 different kinds of labels were annotated on 46,305 manually aligned shots. These videos are in the MPEG-1 format. The total file sizes are 51.6GB, with 6,707,286 video frames. A list of the histogram distribution of annotation labels is shown in Appendix B. Figure 6 shows the histogram of the annotated concepts. We can see that 107 concepts have more than 100 examples. Only 185 concepts have at least 10

Questionnaire for Video Collaborative Annotation Forum 2003

Q1. After you were familiar with the VideoAnnEx Annotation Tool, in average, how long did you need to annotate a 30 min news video? (Please don't count your rest time!!)

Q2. Did you use the following functions? (Please select all you've used and indicate whether they were useful or not useful)
a.) Template Matching, e.g., propagate text labels, logos, etc.
b.) Annotation Learning, i.e., learned from previous annotation and propagated labels to the nearest neighbors in the feature domain
c.) Label Editing -- copy, paste, delete, clear, etc.

Q3. If there is another annotation task in the future, what kind of new functionality do you expect can make the annotation task more efficient?

Q4. Do you think the lexicon we used (133 terms, including events, scene and objects) can cover most general concepts in the videos you annotated? How many percent of concepts do you think the lexicon covers?

Q5. Do you agree we should use a larger lexicon for the annotation task? Given the limitation of the size of display window, how many label items do you think are reasonable and practical?

examples.

After the annotation task was finished, we sent a questionnaire survey to the 111 forum participants. Within two weeks, we received 38 effective replies. The questions are listed in the sidebar "Questionnaire for Video Collaborative Annotation Forum 2003."

In **Question 1**, we asked the users of the average annotation time for a 30 min video. The statistics of annotation time is shown in Figure 6. In average, the annotators use 3.39 hour per 30-min video. This is corresponding to 6.8x of the real time speed. The annotation efforts include shot boundary alignments (split or merge), keyframe adjustment, visual global annotation, audio annotation, and visual region annotation. The maximum one is 9 hours and the minimum is 1 hour. Although the annotation time varies from 2x to 18x, we could not observe apparent difference on the annotated labels contributed by the annotators who spend the most and the fewest time. We randomly select 12 videos that are annotated by one of these two categories of annotators and use VideoAnnEx to check these annotated labels. But, we could not observe apparent difference between them in terms of accuracy and completeness.

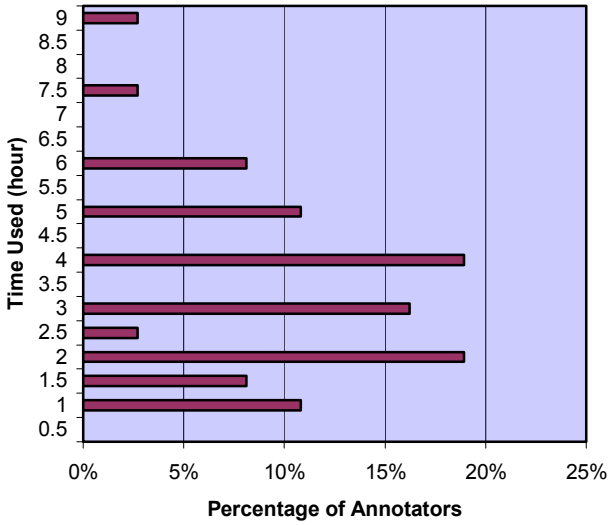

Figure 7: Distribution of Time used by Annotators on a 30 Minutes Video

Question 2 is a survey of the usability of three VideoAnnEx functionalities: Template Detection, Annotation Learning and Editing. The result is shown in Figure 8. Template Detection is designed for text region and logo detection. 55% of the annotators considered this feature helpful. Annotation learning is a label propagation feature which automatically annotates a new shot with the labels of its nearest neighborhood shot in weighted time and feature space. About 1/3 of the users consider this function useful. Nearly 80% of the users use the editing function, which facilitates the copy, paste, delete and clear of annotation labels.

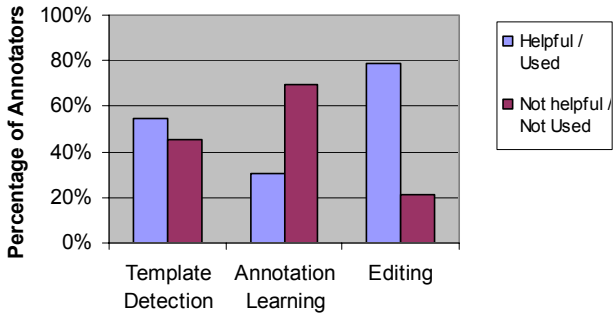

Figure 8: Subjective Usefulness Study of VideoAnnEx Features

Question 3 is an open question to the annotators. We asked for their opinions on the future improvement of the annotation forum as well as the annotation system. Their opinions can be classified into these four categories: *Interface, Efficiency, Stability* and *Ontology*. Here, we excerpt several representative suggestions. For **Interface**, the users suggest these features may be useful: (1) Adding a Help: tool-tip which includes built-in annotation and lexicon instructions, and sample annotations; (2) Speech Interface which allows users to annotation via speech recognition or adding speech comment; (3) No lexicon scrolling to increase the efficiency; (4) Playing the video and audio faster; and (5) Automatic detect the existence of audio concept. Among these suggestions, we found that users are inclined to having more audio/speech interfaces to the system. For suggestion (3), this shall depend on the number of lexicon labels in the ontology or the way user selects a label. It may involve more complicated lexicon selection interface design.

Users made some suggestions to improve the **efficiency** of the system: (1) Reduce time on shot alignment: with better shot segmentation and better correcting tools; (2) Annotate large groups of shots at once; (3) Rules for region annotation (e.g., propagation [people => person], non-regional concepts [indoors, outdoors, audios, …]) and (4) Initial detection for specific domain videos (e.g., sports, movie, ..). Among these suggestions, we think (1) can be improved as the performance of shot boundary detection improves each year in TREC VID benchmarking. VideoAnnEx can import shot segmentations from other algorithms via MPEG-7. (2) and (3) can be improved by additional interface design. If we observe the high-

feature detection result of TRECVID 2003, we see participants can detect several genres of videos in a very good accuracy (e.g., weather, sports, etc.). Thus, some initial detection genre detector results may be considered to import into the system for the future annotation task.

Another user concern is the **stability** of the system. The users hope to have (1) Less crashing and bugs; and (2) Regular automatic saves of annotation. From the users' feedback, about 10% of the client system may crash when the annotator processes the annotation task after playing and annotating 250 shots. This may due to memory management issues in the MS operating system, while we have not clearly identified the cause yet. Because of this reason, some users suggest a regular automatic saves should be useful.

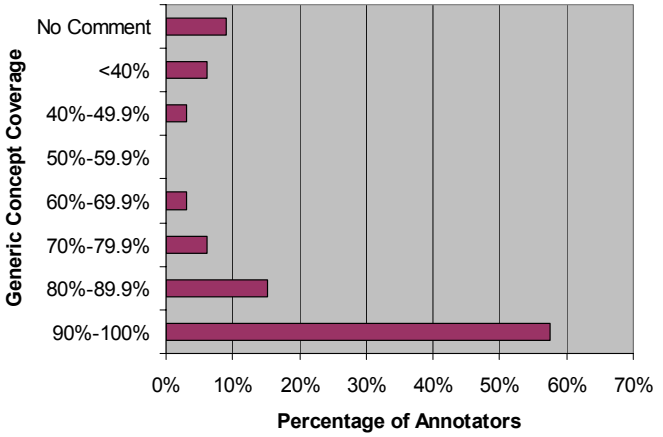

Figure 9: Subjective Opinion on the Completeness of Lexicon Generic Concept Coverage

Users suggest the **ontology** (concept description) can be improved to allow annotating more semantic meanings. They suggested to (1) Allow associate semantic relations between labels on the objects, e.g., a man is speaking *in front of* an U.S. flag; and (2) Have a built-in automatic hierarchy propagation, which is an interface design issue.

In **Question 4**, we tried to ask annotators' subjective opinion on the generic concepts that had been covered by this lexicon. This research is meant to explore users' experience on the number of generic concepts as well as the completeness of the lexicon. We knew that the answer of this question may be highly related to the purpose of annotation and a lack of concrete definition of "*general concept*". However, we purposely not to specify the context of this question in order to receive a statistics of general intuition from the annotators' subjective opinions., Overall, the annotators consider the lexicon had already covered 81% (in average) to 90% (in median value) of the concepts they would like to annotate on those news videos. 58% of the annotators thought the lexicon covered at least 90% of the concepts. We may assume these annotators answer this question under the context of TRECVID concept detection and search retrieval benchmarking. We noted that 9% of the annotators chose not to answer this question directly, because of their concern on the ambiguity of context of this question, such as the purpose of annotation, the scope of annotation, the details of annotation, etc. A statistics of users' subjective opinion on the lexicon completeness is shown in Figure 9.

Question 5 is another subjective question to the annotators. We tried to ask annotators' opinions on whether a larger lexicon should be used. Among the effective answers, 61% of the annotators thought the current number of lexicon labels is adequate. 21% suggested a larger lexicon, while 6% suggested to trim down the lexicon. The distribution of users' opinions is shown in Figure 10. In the current design of *VideoAnnEx*, the lexicon is organized in hierarchy and is selected based on user's mouse selection. In annotation task, users need have a rough memory on what labels are in this lexicon as well as their locations in the hierarchy. Although label trees can be collapsed or opened, users may sometimes need to scroll the bars to find out exact labels. Therefore, there is a limit on the lexicon to be used in practical issues. For instance, we tried to convert the Thesaurus for Graphic Materials I (TGM I) of Library of Congress [13] into MPEG-7 format and imported it into the annotation tool. TGM I provides a controlled vocabulary for describing a broad range of subjects depicted in such materials, including activities, objects, types of people, events, and places. This lexicon has 16,736 terms. Although this

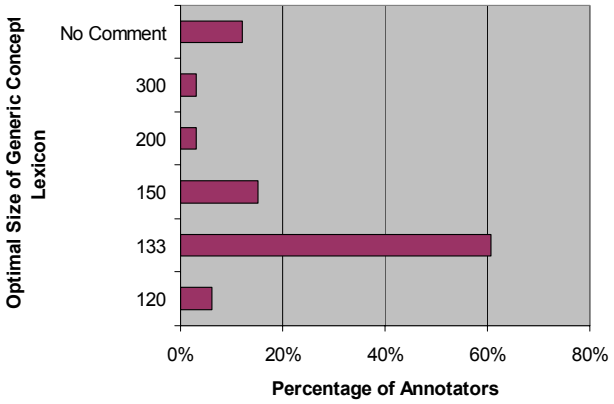

Figure 10: Subjective Opinion on the Optimal Size of Generic Concept Lexicon

lexicon provides more complete descriptive terms in assisting indexers in selecting terms for indexing and helping researchers find appropriate terms with which to search for pictures, finding the lexicon labels themselves becomes a problem in such a big lexicon. In the three years' use of *VideoAnnEx* for TREC video annotation, we tried to constraint the number of controlled terms to the number of 100 -150, and leave an open keyword section for free annotation. Keywords may be organized in some sort of hierarchy afterwards. This strategy might be useful in practical use. However, to our knowledge, no rigorous study on the limit of lexicon number and user interface design issues have been studied before. In our opinion, how to effectively adopt a large lexicon in an annotation task is still an open issue.

In addition to the user studies, we also tried an experiment to study the completeness of annotations on different annotators. Because this annotation task is totally done by human annotation, we can assume the false positive of the annotation accuracy is zero. To study the statistics of miss in annotations, we randomly selected 10 videos and assign each video to two different persons in the annotation task. In other words, these 10 videos were annotated twice and 20 persons are involved in this experiment. In Figure 11, we show a comparison of the annotation results of these annotators. We assume the union of these two annotators is the complete ground-truth labels of shots in each video. Annotator 1 is the annotation result of the "better" annotator, who annotates more labels, for each video. The results are shown as the average number in each video. In average, among these 20 annotators, each annotator labels about 68% of the assumed ground-truth. 78.7% of the ground-truth was annotated by the "better" annotator. This statistics may provide a hint of the completeness of annotations.

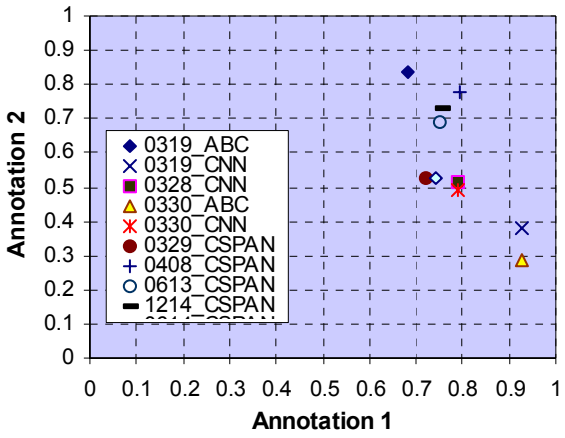

Figure 11: Test of completeness of annotations by different annotators

5. SUMMARY

We built an MPEG-7 Annotation Tool to facilitate multimedia annotation tasks for general users. Use of MPEG-7 is transparent to users so that no prior knowledge on MPEG-7 is required. Various features, such as shot segmentation, ontology editing, storyboard generation, etc., are provided. We developed a new version for collaborative multimedia annotation task in a distributed environment. We proposed a forum to collaboratively annotate semantic labels to the NIST TRECVID 2003 development set. From April to July 2003, 111 researchers from 23 institutes worked together to associate 200K of ground-truth labels (433K after hierarchy propagation) at 63 hours of videos. 1038 different kinds of labels were annotated on 46K manually aligned shots. This large set of valuable ground-truth data should be very useful for the research community in the years to come.

6. ACKNOWLEDGEMENTS

This forum could not success without the help and support of researchers around the world. We would like to thank to all the forum contributors. Especially, we want to thank these individuals who helped to make this annotation forum possible: Paul Over (NIST), Alan Smeaton (DCU), and Wessel Kraaij (TNO) lead the NIST TRECVID 2003 benchmarking task; Georges Quenot (CLIPS) provided the common shot boundaries for the second round of annotation; Arnon Amir (IBM Almaden) provided the common shot boundaries for the first round of annotation; Ronald Murray (Library of Congress) helped boosted the original thought of collaborative annotation; Milind Naphade, Harriet Nock, Giri Iyengar, and Arnon Amir (IBM) helped us to draft the initial lexicon; Apostol Natsev, Matt Hill and Alex Jaimes (IBM) helped with the initial system design.

We also want to thank to 21 group administrators (of 23 groups) who assigned annotations to their members and monitored progress, which is critical to the success of this forum -- Valery A. Petrushin and Gang Wei (Accenture Labs.), Alexander Hauptmann, Mark Egerman and W. Drozd (Carnegie Mellon University), Georges Quenot (CLIPS-IMAG), Arjen P. de Vries and Tzveta Ianeva (CWI), Georgina Gaughan (Dublin City University), Datong Chen (EPFL), Bernard Merialdo

(EURECOM), Feng Zhe (Fudan University), Belle Tseng (IBM Research, Columbia University, the University of Chile), Rainer Lienhart (Intel Labs)., Keiichiro Hoashi (KDDI R&D Labs), Shang-Hong Lai (National Tsing-Hua Univ.) , Yunlong Zhao (National University of Singapore), Esin Guldogan (Tampere University of Technology), Zeeshan Rasheed (the University of Central Florida), Uri Lurgel (the University of Duisburg Essen), S. Marchand-Maillet, N. Moenne-Loccoz and B. Janvier (the University of Geneva), Mark Baillie, Chih-Tsung Lu (the University of Glasgow), Jiwoon Jeon (the University of Massachusetts), Gary Marchionini and Meng Yang (the University of North Carolina at Chapel Hill), and Mika Rautiainen (the University of Oulu).

Most important of all, we are very grateful to all the 111 researchers around the world who spent their time to take part in the collaborative annotation of the entire TRECVID development set.

7. REFERENCES

[1] B. Adams, A. Amir, C. Dorai, S. Ghosalx, G. Iyengar, A. Jaimes, C. Lang, C.-Y. Lin, A. Natsev, M. Naphade, C. Neti, H. Nock, H. Permuter, R. Singh, J. R. Smith, S. Srinivasan, B. L. Tseng, Asshwin T. V., and D. Zhang. "IBM Research TREC-2002 Video Retrieval System," NIST *TREC-2002 Text Retrieval Conference*, Gaithersburg, MD, November 2002.

[2] W. H. Adams, C.-Y. Lin, G. Iyengar, B. L. Tseng and J. R. Smith, "IBM Multimodal Annotation Tool," IBM Alphaworks, August 2002.

[3] A. Amir, M. Berg, S.-F. Chang, W. Hsu, G. Iyengar, C.-Y. Lin, M. Naphade, A. Natsev, C. Neti, H. Nock, J. R. Smith, B. L. Tseng, Y. Wu, D. Zhang, "IBM Research TRECVID-2003 Video Retrieval System," NIST *TREC-2003 Text Retrieval Conference*, Gaithersburg, MD, November 2003.

[4] D. Bargeron, A. Gupta, J. Grudin, , E. Sanocki, "Annotations for Streaming Video on the Web: System Design and usage Studies," ACM 8th Conference on World Wide Web, 1999.

[5] B. Chandrasekaran, J. R. Josephson and V. R. Benjamins, "Ontology of Tasks and Methods," In Proceedings of the 11th Knowledge Acquisition Modeling and Management Workshop, KAW'98, Banff, Canada, April 1998

[6] European Cultural Heritage Online (ECHO) , http://www.mpi.nl/echo/.

[7] J. Hunter, "Adding Multimedia to the Semantic Web – Building an MPEG-7 Ontology," In International Semantic Web Working Symposium (SWWS), Stanford, July 2001.

[8] C. Jorgensen, "Indexing Images: Testing an Image Description Template," ASIS 1996 Annual Conference Proceedings, October 1996.

[9] C.-Y. Lin, B. L. Tseng and J. R. Smith, "IBM MPEG-7 Annotation Tool," IBM Alphaworks, http://alphaworks.ibm.com/tech/videoannex, July 2002.

[10] C.-Y. Lin and B. L. Tseng, "VideoAnnEx Collaborative Annotation System" http://mp7.watson.ibm.com/VideoAnnEx.

[11] C.-Y. Lin and B. Tseng, "Video Collaborative Annotation Forum", http://mp7.watson.ibm.com/projects/VideoCAforum.html

[12] F. Nack, W. Putz, "Semi-automated Annotation of Audio-Visual Media in News," GMD Report 121, December 2000.

[13] Prints and Photographs Division, Library of Congress, "Thesaurus for Graphic Materials I: Subject Terms (TGM I)", http://www.loc.gov/rr/print/tgm1/, 1995

[14] Ricoh Movie Tool, http://www.ricoh.co.jp/src/multimedia/MovieTool/

[15] M. P. Steves, M. Ranganathan, E. L. Morse, "SMAT: Synchronous Multimedia and Annotation Tool", Hawaii International Conference on System Sciences (HICSS-34), , Maui, Hawaii, January 2001

[16] http://www.know-center.at/en/divisions/div3demos.htm

[17] http://www.research.ibm.com/VideoAnnEx

[18] J. R. Smith, S. Srinivasan, A. Amir, S. Basu, G. Iyengar, C.-Y. Lin, M. Naphade, D. Ponceleon and B. L. Tseng, "Intergrating Features, Models, and Semantics for TREC Video Retrieval," *NIST TREC-10 Text Retrieval Conference,* Gaithersburg, MD, Nov. 2001.

[19] B. L. Tseng, C.-Y. Lin and J. R. Smith, "Video Summarization and Personalization for Pervasive Devices," Proceedings of SPIE Storage and Retrieval for Media Databases, Vol. 4676, pp. 359-370, San Jose, January 2002

8. APPENDIX

8.1 Specific Lexicon Items Defined by NIST

These 17 labels in the lexicon are defined by NIST for the purpose of high-level feature (concept) detection:

1. **Outdoors**: segment contains a recognizably outdoor location, i.e., one outside of buildings. Should exclude all scenes that are indoors or are close-ups of objects (even if the objects are outdoors).
2. **News subject face**: segment contains the face of at least one human news subject. The face must be of someone who is not an anchor person, news reporter, correspondent, commentator, news analyst, nor other sort of news person.
3. **People**: segment contains at least THREE humans.
4. **Building**: segment contains a building. Buildings are walled structures with a roof.
5. **Road**: segment contains part of a road - any size, paved or not.
6. **Vegetation**: segment contains living vegetation in its natural environment
7. **Animal**: segment contains an animal other than a human
8. **Female speech**: segment contains a female human voice uttering words during and the speaker is visible.
9. **Car/truck/bus**: segment contains at least one automobile, truck, or bus exterior.
10. **Aircraft**: segment contains at least one aircraft of any sort.
11. **News subject monologue**: segment contains an event in which a single person, a news subject not a news person, speaks for a long time without interruption by another speaker. Pauses are ok if short.
12. **Non-studio setting**: segment is not set in a tv broadcast studio
13. **Sporting event**: segment contains video of one or more organized sporting events
14. **Weather news**: segment reports on the weather
15. **Zoom in**: camera zooms in during the segment
16. **Physical violence**: segment contains violent interaction between people and/or objects
17. **Person x**: segment contains video of person x (x = Madeleine Albright)

8.2 List of the statistics of annotated labels (433,338 labeled items; 1,038 labels; 46,305 shots)

35571	Graphics_And_Text
35403	Human
26757	Audio
24224	Sound
22066	Outdoors
21737	Male_Speech
20532	Text_Overlay
18971	Face
18140	Person_Action
18120	Indoors
15546	Music
15293	Male_Face
14043	Non-Studio_Setting
11687	Female_Speech
11350	Monologue
10268	Man_Made_Object
10230	Female_Face
8278	Person
6289	Nature_Non-Vegetation
5991	Graphics
5910	People
5668	Man_Made_Scene
4173	Scene_Text
3450	Transportation
3172	Studio_Setting
2523	Nature_Vegetation
2428	Sport_Event
2400	Sky
2312	News_Subject_Monologue
2062	House_Setting
2035	Building
1875	Car
1784	Male_News_Subject
1727	Sitting
1625	People_Event
1603	Standing
1551	Greenery
1457	Crowd
1401	Walking
1357	Female_News_Person
1347	Male_News_Person
1212	Animal
1212	Monitor

1124	Road
1074	Food
1069	Office_Setting
1024	Tree
1017	Water_Body
949	Commercial
761	Female_News_Subject
675	Meeting_Room_Setting
655	Basketball
653	Transportation_Event
633	Singing
618	Microphone
608	Desk
605	Cityscape
491	Vehicle_Noise
458	Chair
450	Snow
431	Photographs
416	Cartoon
414	Cloud
412	CNN_Text_Overlay
391	Briefing_Room_Setting
374	Flag
374	Running
364	GraphicsText_Overlay
360	Mountain
351	Addressing
347	Meeting
326	Bill_Clinton
325	Forest
322	Physical_Violence
314	Land
300	Airplane
292	Truck
287	Store_Setting
286	Weather_News
281	Cheering
277	Abc_Logo
276	Flower
248	Factory_Setting
246	Beach
236	Dog
225	Telephone

222	Boat
205	Transportation_Setting
203	Hockey
202	Classroom_Setting
200	Rock
176	Baseball
162	Road_Traffic
155	Laboratory_Setting
151	Painting
147	Animal_Noise
145	Clapping
143	Podium
129	Fire
123	Desert
114	Smoke
112	Bridge
106	Newspaper
104	Blank
104	Explosion
102	Hand
101	Horse
96	Laughter
95	Parade
93	Outer_Space
91	Ice_Skating
74	Bus
68	40th_Anniversary_Of_The_Freedom_Rides
67	Tractor
64	Football
63	Car_Crash
63	Train
59	Clock
54	Journalist_Discussion
52	Cat
49	Dancing
47	Fight
47	Peter_Jennings
45	Missle_Launch
43	Bicycle
42	Zoom_In
40	Fish
40	Airplane_Takeoff
40	Waterfall

38	Gun_Shot
38	Blackboard
37	Interview
35	Madeleine_Albright
34	Chicken
32	Golf
31	Cut
31	Swimming
31	Powerpoint
30	Newt_Gingrich
29	Riot
29	Players
29	Book
28	Statue
27	Noise
26	Tennis
23	Sun
22	Map
21	Space_Vehicle_Launch
20	Picnic
20	Whiteboard
20	Fencing
19	Eating
19	Airplane_Landing
18	Human_Hand
18	Jacques_Nasser
18	Cinema_Setting
17	Gun
17	Skiing
17	Guitar
16	Camera
16	Keyboard
16	Glasses
16	Solar_Eclips
15	Underwear_Model
15	Cow
14	Saddam_Hussein
14	Store
14	Soccer
14	Foot
14	Dance
14	Helicopter
12	Gymnasium_Setting
12	Chocolate_Production

12	Congress
11	Chocolate
11	CNN_Splash_Page
11	Princess_Diana
11	Olympics
10	LCI
10	Stock_Exchange
10	Aliens
10	Car_Part
10	Bird
10	Porch_Setting
10	Sea
9	Bottle
9	Restaraunt_Setting
9	Mop
9	House
9	Children
9	Rocket
8	Bra
8	Computer
8	News_Person_Monologue
8	Cigarette
8	Shoe
8	Lamp
8	Acupuncture
8	Aligator
8	Movie
8	Flowers
8	Wrecked_Car
8	NBA
7	Zoom_Out
7	Vacuum_Cleaner
7	Chairman
7	Girl
7	TV
7	6_Abc_Logo
7	Dining_Room_Setting
7	Children_Playing
7	Medicine
7	Tapir
7	James_Greenwood
7	Camera_Movement
7	Senator
7	Women's_Volleyball

7	Racing
7	Sadam_Hussein
7	Giraffe
7	Ice_Hockey
7	Ship
7	Astronaut
7	Water
7	Can
6	Wind_Noise
6	Yeltsin
6	FBI
6	Space_Station
6	Congress_Setting
6	John_Dingell
6	CGM_Realty_Fund
6	Women
6	Plant
6	Kid
6	Kids
6	Billy_Tauzin
6	Cable_Car_Tragedy
6	Pencil
6	Baby
6	Nelson_Madela
6	Russia
6	Flood
6	Table
6	Grass
6	Tank
6	Pen
6	Castle
5	Picture
5	Reporter
5	Player
5	Soldier
5	School
5	Sheep
5	Stilted
5	Drug
5	Bed
5	Surfing
5	Lottery
5	$
5	Night
5	Ball
5	Gas_Station

5	Basketball_Player
5	Boy
5	Mexico
5	Jumping
5	3
4	Crying
4	Party
4	Advil_Commercial
4	Box_Of_Chocolates
4	Policeman
4	Window
4	Monica_Lewinsky
4	Coach
4	Play_Of_The_Day
4	Mouth
4	Hospitdal_Setting
4	Lobster
4	US_Dolar
4	Basketball_Court_Setting
4	Sofa
4	I_Love_Lucy
4	Doll
4	Saddam
4	Students
4	Clothe_Store
4	Court_Room
4	Famale_Speech
4	Mouse
4	Underwear
4	Car_Setting
4	Buttery
4	Sports_News
4	Iraq
4	Girls
4	Wires
4	Dinner
4	Accident
4	Cereal
4	Playing_Instruments
4	Butter
4	Pills
4	Little_Girl
4	Speaker
4	Puppet
4	Intercom
3	Deer

3	Shell
3	Capital_Hill
3	Dime
3	Lottery_Drawing
3	Chicago_Bulls
3	Albright
3	Cookie
3	Surfboard
3	Shooting
3	Poles
3	Phone
3	Cereal_Commercial
3	Pothole
3	Motorbike
3	Doctor
3	Bmv
3	Cruise_Ship
3	Door
3	Ballon
3	Worker
3	Kofi_Annan
3	Book_Shelf
3	WaterSound
3	Old_Women
3	Glass
3	New_Item
3	Carton
3	Tug_Of_War
3	Jar
3	Parking
3	Michael_Jordan
3	Shopping
3	Elephant
3	Missile
3	Barn
3	Video_Transition
3	Rowing
3	Anthrax
3	Zebra
3	Driving
3	Floor
3	Model
3	Woman
3	Tiger
3	40th_Anniversary_Of_Freedom_Rides
3	Couple

3	Commcercial
3	Toothbrush
3	Finger
3	Gym
3	Old_People
3	Mug
3	School_Boy
3	TV_Screen
3	Crime_Lab
3	Program_Guide
2	Smoking
2	Cushion
2	Stairs
2	Black_Frame
2	Kitchen_Appliance
2	Pressing
2	Bowl
2	Parachute
2	Dance_Studio
2	Shoes
2	Logo
2	Parrot
2	Gorilla
2	Parking_Lot
2	Sign
2	Packet_Of_Food
2	Dog_Running
2	Candy_Map
2	Skier
2	Smashed_Potato
2	Shirt
2	Swimming_Pool
2	Tablets
2	Patient
2	US_Weather_Map
2	Paper
2	Family
2	River
2	Ski_Slope
2	Computer_Animation
2	New_York
2	Golf_Ball
2	Rhinoceros
2	Texas
2	Van
2	Computer_Voice

2	Toll_Station
2	School_Girl
2	Skeleton
2	Shout
2	Playing_The_Piano
2	Junk
2	Fashion
2	Monkey
2	Courtroom
2	Brushing_Teeth
2	Speaking
2	Lewisky
2	Dog_Food
2	Siren
2	Chicken_Broth
2	Fruit
2	Cellphone
2	Hockey_Players
2	Shelf
2	Actionboard
2	Fast_Motion
2	Snack_Bar
2	Astronauts
2	Clothes
2	NBA_Players
2	Concert
2	School_Shooting
2	MIR_Space_Station
2	Press_Conference
2	Watch
2	Medcine_Commercial
2	Maalox
2	Monks
2	Hotel_Room
2	FilmProductionFloor_Setting
2	Smile
2	Vacuum
2	Hat
2	Internet
2	Dead_Bodies
2	Broke_The_Record
2	Beano
2	Bag
2	Soldiers
2	Restaurant

2	Temple
2	FiberCon
2	Feet
2	Television_Crew
2	Lottery_Balls
2	Watching_TV
2	FedEx
2	Toy
2	Court
2	Street
2	Stadium
2	Martin_Luther_King
2	Tellurion
2	Death_Of_Hollywood_Star
2	Airbag
2	Basketball_Court
2	Lottery_Ball
2	Pole
2	Drinking
2	Warehouse
2	Camera_Man
2	Boys
2	NBA_Scores
2	Clothe
2	Eagle
2	Video
2	FilmProduction_Setting
2	Dow_Jones
2	Paint_Can
2	Reporters
2	Cup
2	Storm
2	Calculator
2	William_Cohen
2	Scholl_Shooting
2	Garden_Tool
2	Referee
2	Dog_Barking
2	Telephone_Ringing
2	Bond
2	Chocolates
2	FilmProduction_Studio
2	Fashion_Show
2	Stock_News
2	Basketball_Hoop

2	Iraqi_Exile
2	Golf_Flag
2	Finance
2	Television
2	Margarine_Commercial
2	Broccoli
2	Handle
1	Shadows
1	Chevy_Venture
1	Mom_And_Baby
1	Guns
1	World_Globe
1	Sports
1	Crossword_Puzzle
1	Squash
1	Chess
1	Webpage
1	Nuclear_Weapons
1	US_And_Mexico
1	Playing_Violion
1	Ashma_Research_Petri_Dish_Breathing_Apparatus
1	Leg
1	David_Robinson
1	Mower
1	40th_Anniversary_Of_The_Freedom_Riders
1	Lottery_Advertisement
1	Stock
1	Playing
1	Tennis_Racket
1	Anchor_Person
1	Direct_TV
1	Excercise
1	Mom
1	Fishes
1	Aerobic
1	Rescure
1	Iraq_Congress
1	President_Clinton
1	Charles_Schwab_Commercial
1	Knocking_The_Door
1	Moon
1	Coaming_The_Liquid
1	School_Bell_Noise

1	MonoloFemale_Speech
1	Eye_Glasses
1	Hole
1	Sail
1	Drum
1	Cheryl_Watson
1	Sun_Spotters
1	Llovd_Bridges
1	Envelope
1	Mobile
1	Embedded_Scenes
1	Globe
1	Bar
1	Towel
1	Children_Television_Program
1	Steps
1	Tylenol_Commercial
1	Cambridge_Business wear
1	Paper_And_Dust
1	American_Loans
1	Crash
1	Italy
1	Cloth
1	Dog_Eating
1	Basketball_Area
1	Airplane_Crash
1	Undersea
1	Leopard
1	FiberCon_Commercial
1	Pearle_Vision_Commercial
1	Door_Handle
1	Gas_Statoin
1	School_Boys
1	Tying_A_Shoe
1	New_York_City_Middle_School
1	UN
1	Restaraunt_Settting
1	Doctor's_Website
1	Pope
1	Police_Man
1	Mandela
1	NASA
1	Iaqie_Exile
1	Raising_Money

1	Reading
1	Pregnant
1	Musicians
1	Black
1	Court_Room_Setting
1	Airport
1	Oil_Commercial
1	40th_Anniversary_Of_Freeom_Riders
1	Bottle_And_Cup
1	Snack_Bar_Commercial
1	Teethbrush
1	News_About_Marijuana
1	Film_Set
1	Axe
1	Notebook
1	Platinum_Coin
1	American_Soldier
1	Titanic
1	Health
1	Wrestling
1	Penguin
1	Players_Falling
1	Accupunctures_Needles
1	Tv_Set
1	Treadmill
1	Kitchen
1	Helmet
1	Wheel_Chair
1	SUV
1	Ladder
1	Internet_Sales
1	Pesticized
1	Telephone_Number
1	Starfish
1	Advil_Commercia
1	Discus_Athletics
1	Signal_Noise
1	Oil
1	Drug_Smuggling
1	Lottery_Balls_And_Wild_Card_Ball
1	Statistics
1	South_Africa
1	Writing
1	Lemon

1	Coffee
1	Oil_Rig
1	Snowman
1	Lotter_Drawing
1	Pail
1	Chevy
1	Edit_Effect
1	Exiting_Car
1	Iraqi_Soldiers
1	Ahmad_Chalabi
1	JAMA_Article_Breast_Surgery
1	Injury
1	Junks
1	Alarm
1	Officer
1	Talking
1	Computer_Monitor
1	Brummeal_And_Brown
1	Platinum_Coins
1	Reagans
1	Garden_Tools
1	Israel_And_Palestin
1	Plug_In
1	American_Century_Commercial
1	Acrobatics
1	Nancy_Reagan
1	Advil
1	Mandala
1	Crops
1	Steven_Kings
1	Toilet
1	Wedding
1	Mom_And_Daughter
1	Reagans_Airplane
1	Hockey_Rink
1	Drug_Enforcement_Administrator
1	Shrimp
1	Manhole_Cover
1	Drug_Smuggling_In_Mexico
1	Spade
1	Lake
1	Paint_Cans
1	Anchor_Intro_And_Reporter_Voiceover_For_Ken_Starr_And_Car

1	Preparation_Cream
1	Pens
1	NHL_Scores
1	Box
1	Wires_Flashing
1	Weapon_Instruction
1	Bagel
1	Statue_Of_Liberty
1	Carpet_;_Vacuum_Cleaner
1	Stock_Index
1	5
1	Hugging
1	Personal_Income
1	Treasure
1	DC
1	Hotel
1	Sonics
1	Class_Reunion
1	Drug_Dealing
1	Injection
1	Babies
1	Hand_Shake
1	Underwear_Commercial
1	Universal_Studio
1	Park
1	America_And_Cuba
1	Human_Arm
1	Legs
1	Travolta
1	Hospital
1	Television_Children_Program
1	Cheers
1	Shelf_;shoes
1	Flashing_Light
1	Passport
1	Drug_Smuggling_In_Mexica
1	Titanic_Wreck
1	Committee_Of_Concerned_Journalists
1	Senators
1	Box_Of_Chocolate
1	Travel_To_Ireland
1	Balls
1	Turning_Off_The_Lamp
1	Bar_Setting

1	Space_Vehicle_Interior
1	Falling
1	Rowboat
1	Whistle
1	Sprint_Commercial
1	Skyscraper
1	Yogout
1	Lewinski_Case
1	Wire_Broken
1	Wheelchair
1	Celebrity_Ads
1	Cry
1	Hall__Way
1	Splashing
1	Computer_Room_Setting
1	Iraq_Weapon_Instruction
1	New_York_City
1	Boat_Flouder
1	Library
1	Lottery_Candidates
1	Canola_Oil
1	Biscuit
1	Sand
1	Martain_Luther_King
1	Violin
1	Tennis_Court
1	Hall_Way
1	Western_Movie
1	WNBA
1	Mashed_Potato
1	Bottles_Of_Medicine
1	Father_And_Son
1	Restaurant_Setting
1	Church
1	Starr_Gets_In_Car_Clinton_Hugs_Lewinski
1	Salmon
1	Knife
1	Nuclear_Training_Saftey_Center
1	Waters
1	Green
1	Karate
1	UN_Secretary_General
1	Celebrity

1	Nicolas_Cage
1	Television_Program
1	Vegetable
1	Sicentific_Evidence
1	MVP
1	Killed
1	Animations
1	Music_Instrument
1	Water_Flashing
1	Siebel_Commercial
1	Rush_Inlow
1	Plate
1	Stature_Of_Liberty
1	Roots
1	Iraqi_National_Congress
1	Creadit_Card
1	Sleep
1	Laptop
1	Shelves
1	Oil_Platform
1	Tragedy
1	Grecian_Formula_16
1	Patato
1	Surgeon_General'
1	Bombs
1	Moose
1	Wall
1	Israel
1	Whale
1	Advertisement
1	Waste_Managment
1	Lewinsky
1	CIA
1	Wind
1	Man_Riding_Horse
1	Washington
1	MonoloNon-Studio_Setting
1	Papers
1	Car_Interior
1	Saving_Endangered_Species
1	Radar
1	Workers
1	Playing_Music
1	Champagne_Bottle
1	Flooding_Georgia

1	Travel_Book
1	Chidren's_Voice
1	Exotic_Bird
1	Making_Snack_Bar
1	Flashing
1	Organizer
1	Glass_Reflection
1	Chinese_American_Community
1	Earth
1	Westernn_Movie
1	Arrestment
1	Receipt
1	Dropping
1	Chalabi
1	Secratory_Of_State
1	Flying
1	Russian
1	Betty_Curry_Driving_Car
1	Orange
1	Lobster_Pizza
1	News_SubTree
1	Medcine
1	Security_Door
1	Greclan_Formula
1	Mickey_Mantle
1	Margarine
1	Long_Distance_Commercial
1	Ray
1	Showroom
1	Clinton_Walking_Starr_Walking
1	Cushlin_Gel_Commercials
1	Cimena_Setting
1	Laughing
1	Supermarket
1	Trash_Can
1	Hollywood_Star
1	Doing_Homework
1	Telephone_Interview
1	Hight_Moon
1	Acupuncture_Foot
1	Press-photographer
1	Poultry
1	Manhattan_Hotel
1	Flock_Of_Birds

1	Cameraman
1	UN_Weapon_Inspection
1	Noodle
1	Superman
1	Rain
1	Starr_Walking
1	Lizard
1	Bill_Clinton_Face
1	Lecture
1	Microscope
1	FBI_Agent
1	Falling_Down
1	Nordstrom
1	Sitting_On_The_Roof
1	Mirror
1	Incident_Investigation
1	Human_Face
1	Nutrition_Facts
1	Pill
1	School_Gilr
1	Rabbit
1	Animal_Eyes
1	Drugs
1	Envolope
1	Space_Vechile_Interior
1	Handshake
1	Hourse
1	Butterfly
1	Children_Screaming
1	Skating
1	Dumping
1	Surgery
1	Book_Shelves
1	Bread
1	Hockey_Rink_Setting
1	Broken_Wire
1	Gloves
1	Balloon
1	Jet_Skating
1	Cutting_Butter
1	Calendar
1	Clinton_Leaving_Airplane
1	Shark
1	John_McCain

1	Screaming
1	Throwing
1	Audience
1	Restaruant_Setting
1	Text_OverlNon-Studio_Setting
1	JAMA_Breast_Cancer_Surgery
1	Guard
1	Celebrity_Endorsements
1	Light
1	Boxes_Of_Chocolate
1	ICG
1	Courtoom
1	Boris_Yeltsin
1	DirectTV_Commercial
1	Elevator
1	Justic_Department_Officer
1	Bottlle
1	Sigh
1	Hockey_Player
1	Disability_Work_Legislation
1	Tamato
1	Brand
1	Investigation
1	Space_Shuttle_Setting
1	Prison
1	StandingTree
1	Program_Schedules
1	Lighter
1	Astronaut_Helmet

1	Testing_Setting
1	Elephont
1	Chopping_The_Tree
1	Lottery_Draw
1	Birds
1	Boat_Race
1	Blur
1	Inteview
1	Lwinski
1	Jumping_Rope
1	Los_Angelos
1	California
1	Intimate_Apparel
1	Cocain
1	Cliff
1	Teeth_Brush
1	British_Government
1	Florida
1	Oil_Machine
1	Headline_Sports
1	Home_Loan
1	Plane
1	Cell_Phone
1	Plates
1	Spinning
1	FBI_Investigation
1	Flipping_Coin
1	Rocks
1	Eclipse
1	Radio
1	Walking_Exercise
1	Siscus_Athletic_Commercial

1	Nelson_Mandela
1	Space_Shuttle
1	Body
1	Girl_Making_A_Phone
1	Vaccine
1	Fish_Tales
1	Female
1	San_Franscico
1	Butterflies
1	Area
1	Lawsuit
1	Flooding
1	Ice_Hockey_Rink
1	Duck
1	Animal_Running
1	Julia_Roberts
1	Video_Game
1	Ronald_Hill
1	Monk
1	Chessboard
1	White_House
1	Car_Plate
1	Webpages
1	Strecher
1	Travel_Guide
1	6
1	Insect_Noise
1	Parking_Garage_Setting
1	Child
1	Bathing
1	Telephone_Poles

1	El_Nino
1	Turkey_And_Ham
1	Operating_Room
1	Crying_People
1	Phone_Service
1	Open_Book
1	Cyber_Cafe
1	Baking
1	Riding_A_Horse
1	Drink
1	Chefs
1	Shop_Setting
1	Kids_Playing
1	Vegetables
1	Hand_Place_Item
1	Politics_Reform
1	Commerce
1	Brain
1	Charles_Schwab
1	Eating_Cereal
1	Water_Bra
1	Parliament
1	Red_Lobster
1	Bench
1	Yankee
1	Research
1	Eye
1	Landslide
1	Acupunction
1	Cape_Town
1	Israel_And_Palestine
1	Dump

1	Moscow
1	Missle
1	Tylenol_PM
1	Surgeon
1	Cooking
1	Reagan_Walking
1	Ashma
1	Dinning_Table
1	Video_Camera
1	Basketballs
1	Ahmed_Chalabi
1	Agassi
1	Traffic_Light
1	Gym_Setting
1	Endangered_Species
1	Pig
1	Mexican_Drug_Runner
1	Luge
1	Boris_Jeltsin
1	Fishing
1	Spoon
1	Firing
1	Total_Return
1	Telescope
1	Xerox
1	Womenn_Dancing
1	Cough
1	Text_Overlcut
1	Joseph_Rothenberg