A Practical Digital Video Database Based on Language and Image Analysis

Yiqing Liang

Howard University
Department of Systems and Computer Science
Howard University, Washington, D.C. 08544
yliang@scs.howard.edu

*ADVANCE Inc.
2200 Wilson Blvd., Suite 700
Arlington, VA 22201
liangy@advco.com

The Project

- *Supported by*
  - *DARPA's Image Understanding (IU) program under "Video Retrieval Based on Language and Image Analysis" project*
  - *DARPA's Computer Assisted Education and Training Initiative program (CAETI)*
- *Objective: Develop practical systems for automatic understanding and indexing of video sequences using both audio and video tracks*

The View

- *We do not know how users would like to access digital video libraries contents yet*
- *Users should expect the most flexible way to achieve their access to video contents*
- *Internet-based Web Page could be a preferred Interface*

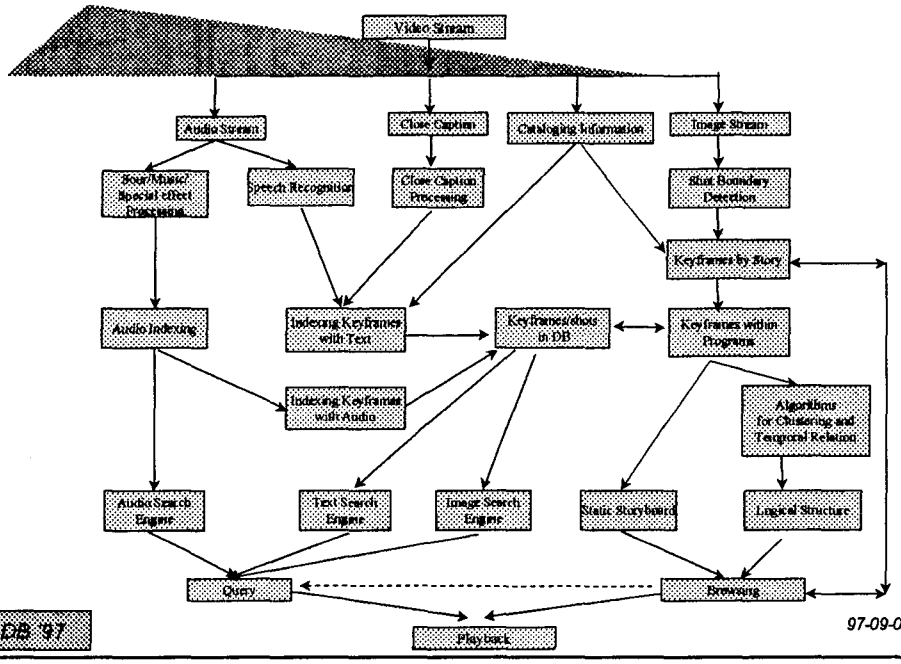
The View (cont'd)

- *Multiple Modalities Embedded in Video Should make Access Points for Users*
  - *Image*
    - *Audio Track*
    - *Speech*
  - *Special Effects (music, laughing, gunfire, explosion)*
  - *Closed Captioning*
  - *Caption*
  - *Motion*
  - *Temporal Information*

Complete Access

- *Access Video Contents at Different Levels*
  - *Video Programs/Clips*
  - *Video Scene*
    - *a collection of one or more adjoining shots that focus on an object or objects of interest*
  - *Video Shot*
    - *an unbroken sequence of frames from one camera operation*
    - *boundary changes: cut, fade, dissolve, and wipe*
  - *Video Objects*

A Digital Video Database Block Diagram

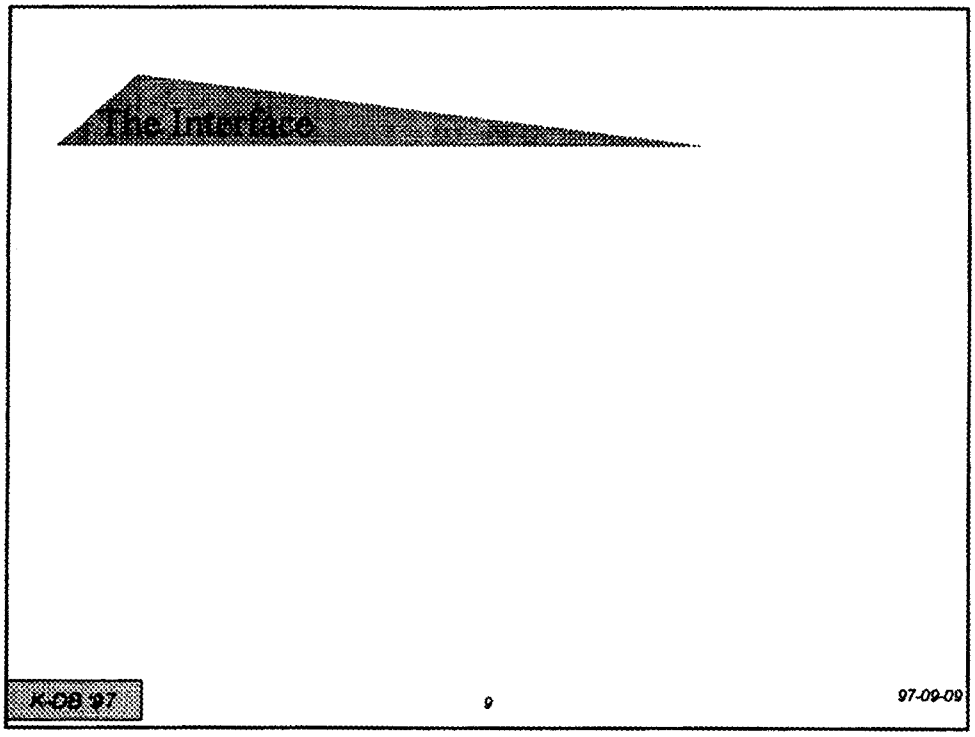


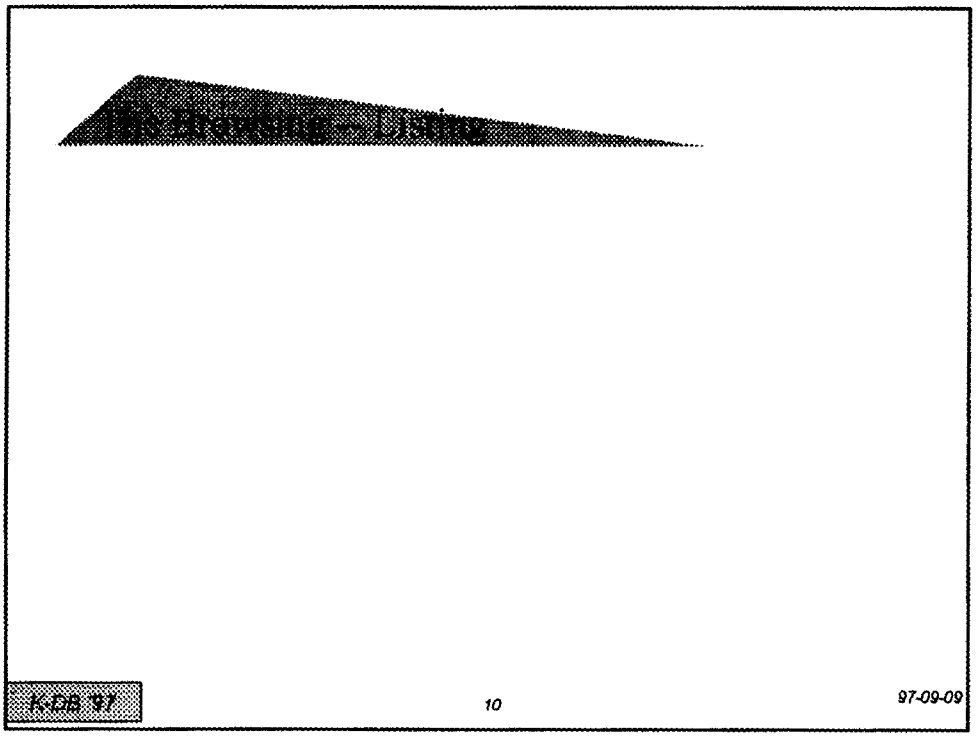
Access Methodologies

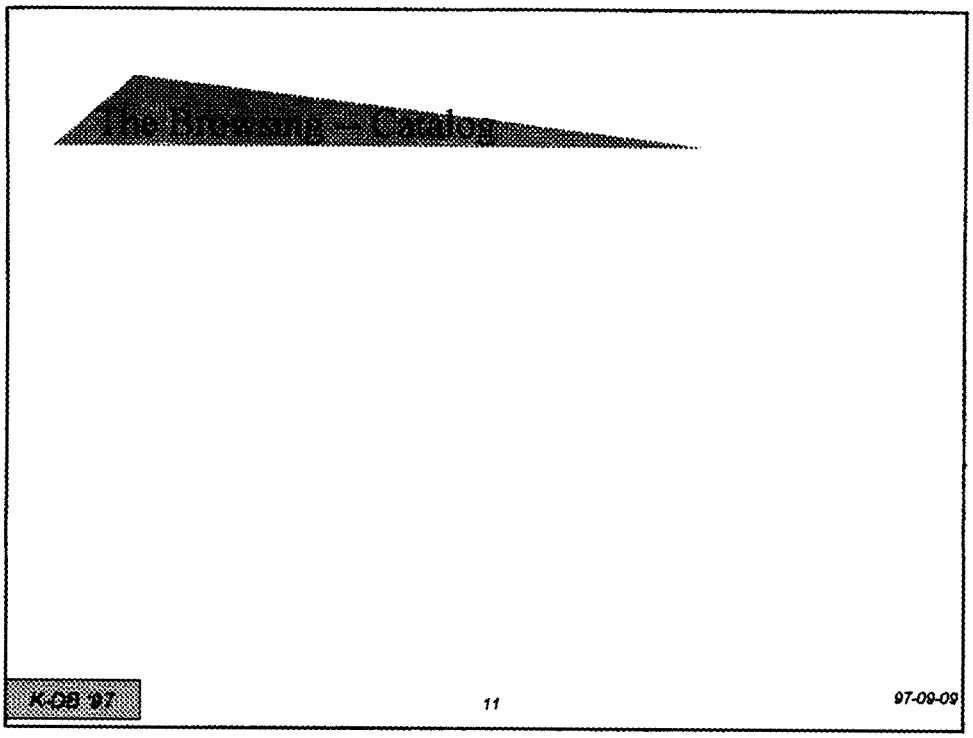
- *Access to Video Contents employing Different Methodologies:*
  - *Text-based: using text to annotate video shot/segments/programs as retrieval points*
  - *Icon-based*
  - *Content-based*
  - *Multiple Modalities Based*
    - *The combination of image, audio, speech, annotation*
    - *The combination of text-based and content-based*

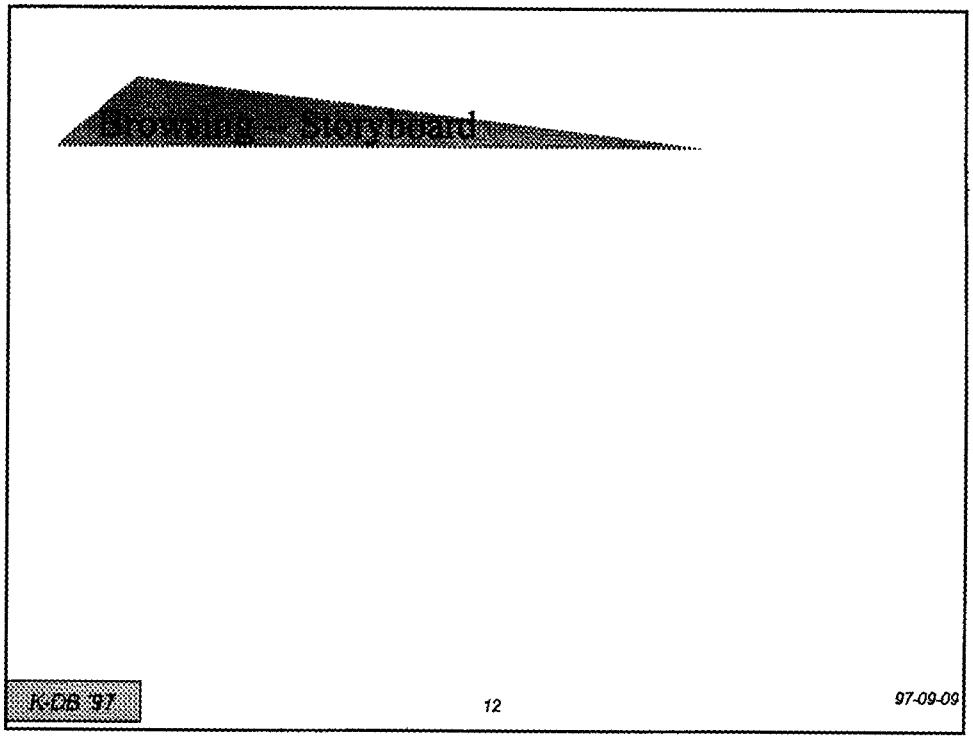
Access Approaches

- *Different Approaches for Content-based Access to Video Contents*
  - *Structured Approach -- Browsing*
  - *Non-linear Access -- Search*
  - *Internet-based Search and Browsing*


The Interface
K-DB '97
9
97-09-09


The Browser -- Listing
K-DB '97
10
97-09-09


The Browsing – Catalog
K-DB '97
11
97-09-09


Browsing – Storyboard
K-DB '97
12
97-09-09

Content-based Search

- *Similarity vs. Match*
  - *Similarity-based Search*
  - *Subject-based Search*
  - *Text-based Search -- Match*
- *Interactive Search vs. Fully Automated Search*
  - *Interactive -- Similarity-based*
  - *Fully Automated -- Subject-based*

Search -- Methodologies (con't)

- *Search based on:*
  - *The difference -- levels*
    - *Segment/scene*
    - *shot*
  - *The difference -- motion information*
    - *Shots*
    - *Keyframes*
  - *The basic -- shot boundary detection*
  - *Application-specific; Trade-off between computing time and rich-features for indexing*

Shot Boundary Detection

- *Absolute Frame Difference*
- *Histogram-based*
  - *Simple Histogram Difference*
  - *Weighted Histogram Difference*
  - *Histogram Difference after Equalization*
  - *Intersection of Histograms*
  - *Shared Histogram Difference*
- *Based on Moment Invariants*
- *Based on the range of pixel-value changes*
- *Based on Edge Detection*
- *Based on Encoded Information*

Search by Match -- Text Search

- *Information Acquisition*
  - *Speech -- Speech Recognition*
  - *Closed Caption -- Closed Caption equipment*
  - *Caption -- Algorithms*
- *Information Association*
  - *Text associated with shots or scenes by calculation based on the time synchronization in the video*
  - *The associated text can be treated as a document*
  - *Document Retrieval is a "Information Retrieval" Problem*
  - *Retrieving desired text is equal to the associated shots or scenes*

Information Retrieval Techniques

- *Traditional Text Retrieval*
  - *Full text scanning*
  - *Signature files*
  - *Inversion*
  - *Vector model and clustering -- Salton's SMART system*
- *Using Semantic Information*
  - *Natural Language Processing*
  - *Latent Semantic Indexing*
  - *Neural Networks*

Search Quality based

Search -- Subject-based

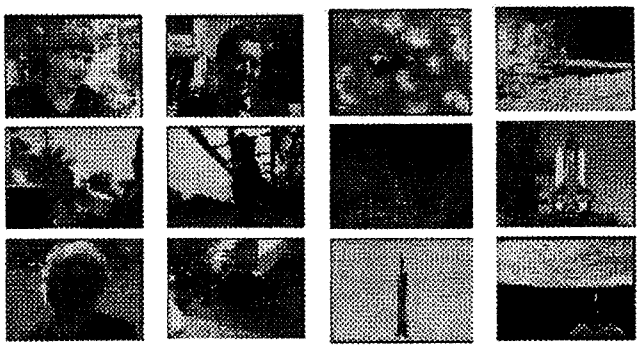
- *A sample keyframe*
- *Segment the keyframe*

Hong-heather Yu and Wayne Wolf, "A Visual Search System for Video and Image Database, IEEE Multimedia, 1997

Search -- Subject-based (con't)

Result from searching subject "sky"

Hong-heather Yu and Wayne Wolf, "A Visual Search System for Video and Image Database, IEEE Multimedia, 1997



sky & foliage human & outside sky

Shot-Based Search

- *Video Masaicing*
  - *Spatial extent is increased by panning the camera while mosaicing (e.g., by making a panorama)*
  - *Spatial resolution is increased by zooming the camera and by combining overlapping frames from different viewpoints*
  - *Implemented through coordinate transform for each pair of frames, using many different algorithms*

Shot-based Search — Mosaicing

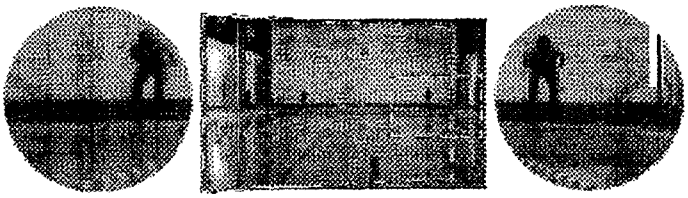
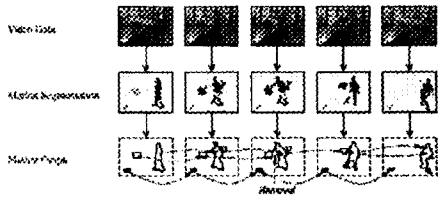


Figure 1: *Image mosaic made from three images, one taken looking straight ahead (outlined in a solid line) one taken panning to the left (outlined in a dashed line) and the third taken panning to the right with underwater zoom-in (outlined in a dot-dash line). The second two have undergone a coordinate transformation to get them into the same coordinates as the one outlined in a solid line (which we call the reference frame). This image mosaic, made from NTSC-resolution images, occupies about 3000 pixels across and, in places, shows good detail down to the pixel level. Note increased sharpness in regions visited by the zooming-in, compared to other areas. (See magnified portions of the picture at right.) This figure only shows the result of combining three images, but in the final production, many more images were used, resulting in a high resolution full-color image showing most of the room.*

S. Mann and R.W. Picard, Video Orbits of the Projective Group: A New Perspective on Image Mosaicing, MIT Media Lab Perceptual Computing Section Technical Report No. 338

Shot-based Search -- Motion

- *Video Motion*
  - *Counting motion information embedded in video*



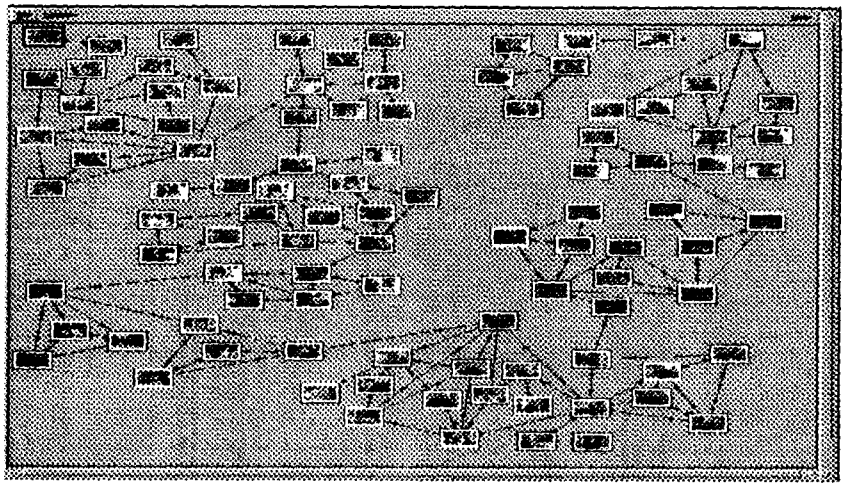
relation between video data, motion segmentation information, and the symbolic link graph

Courtney, Jonathan "Automated Video Indexing via Object Motion Analysis", Pattern Recognition, April, 1997

Image Understanding

- *Access to Video Contents based on Image Understanding of*
  - *Syntactic*
    - *Image global information -- pixels*
    - *Color, histogram, texture, draw, etc.*
  - *Semantics*
    - *logical structure*
    - *Model based -- Zhang, Jain*
    - *objects and their motion*
      - *human Face*
      - *objects of Potential Interest*

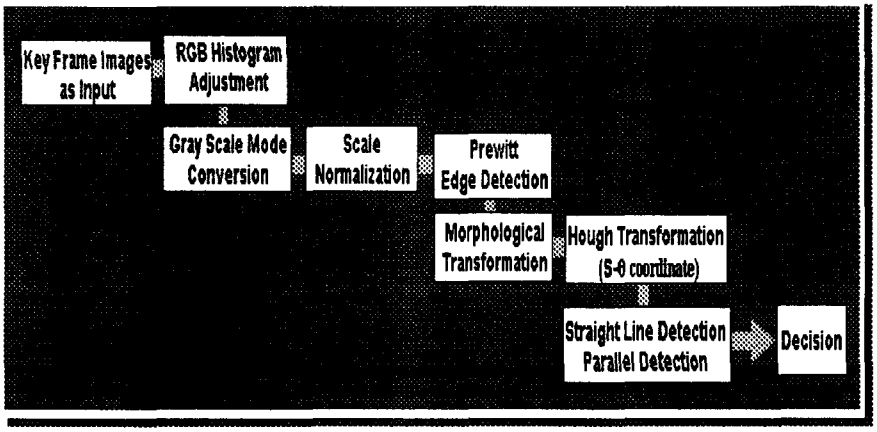
Logical Structure -- STG



Objects of Potential Interest for Battlefield

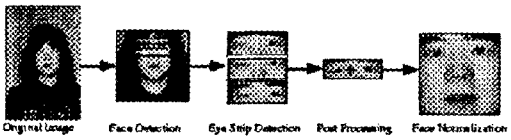
- *UAV Surveillance Video*
  - *Huge Amount of Information*
  - *Only Part of them is of Interest*
  - *Short Time to Process*
  - *Real Time Transmission*
- *Objects of Potential Interest are Artificial*
  - *Artificial Shapes -- Straight Lines*
  - *Artificial Texture*
  - *Artificial Color*
- *Detect Based on Keyframes or Shots*

Methodology



Face Recognition in Image Database

- *Face Recognition in Static Image Database*
  - *FERET project at Army Research Lab*
  - *Face Image Database Testbed*



  - *Based on Neural Network classification and hybrid algorithms*

Face Recognition in Video Database

- *Keyframe-based*
  - *Based on Static image -- keyframes*
  - *Challenges*
    - *Complex background in video database vs clean background*
    - *A few face poses vs complete set of face poses*
    - *Detection of which shots/keyframes containing certain faces*

Face Recognition in Video Database

- *Shot-based and Video Sequence-based – motion*
  - *Scenarios: Static Background vs moving human faces -- Surveillance*
    - *One approach is (Jeffrey Huang:)*
      - *Video break to locate face frames using optical flow*
      - *Face detection to locate and normalize using DT*
      - *Authetification using radial basis function (RBF) network*
  - *Scenaris: Changing background and moving human faces -- Video Database*
    - *Above approach does not apply because motion can not be used effectively to locate face frames*

The Database

- *Built a Digital Video Database*
  - *24 + 38 Video Stories*
  - *1 hr and 5 minutes +12.75 hr or about 13.85 hours*
  - *About 8.5 GB*
  - *In addition, 3 hours' digitized court TV*
- *Built with Real Data*
  - *Predator Video*
  - *News About US Troops in Bosnia*
  - *Video Archives about American Presidents*
  - *NASA Video Archives*
  - *Republican Primary 1996*

System Components

- *Digitization Subsystem*
- *Closed Captioning Extraction Component*
- *Segmenting Video into Shots*
  - *Operation on Compressed Video*
  - *Operation on Uncompressed Video*
- *Extracting Keyframes to Represent Shot*

System Components (con't)

- *Associating Closed Captioning to Keyframes*
- *Summarizing keyframes into storyboard*
- *Scene Transition Graph (STG)*
- *Keyword Search for Catalog Information*
- *Image Search Engine - IBM QBIC*
- *Text Search Engine - BellCore LSI*

System Tools

- *Video Editing Subsystem*
  - *For real MPEG files*
  - *cut, copy, past, delete, insert, and play MPEG files and frames*
  - *special effects: fade from black, fade to black, cross fade in, cross fade out, rotate, variable rotate, edge trace, emboss, and wipe*
  - *color changes: black, invert color, suppress color, invert color, brightness boost/cut, color boost/cut, and gamma convert*
  - *file and bookmark manipulations*

System Tools (con't)

- *Netscape Plugins*
  - *Allows playing back of video clips from specific starting frame to specific stopping frame*
- *CGI program as database search tool*

List of Future Features

- *Algorithms for Different Applications*
- *Semantic Indexing and Retrieval (Shape, Motion, Face, Object, etc.)*
- *Video Logical Structure, its Indexing and Retrieval*
- *Novel Browsing Approach*
- *Annotation*
- *Watermark*

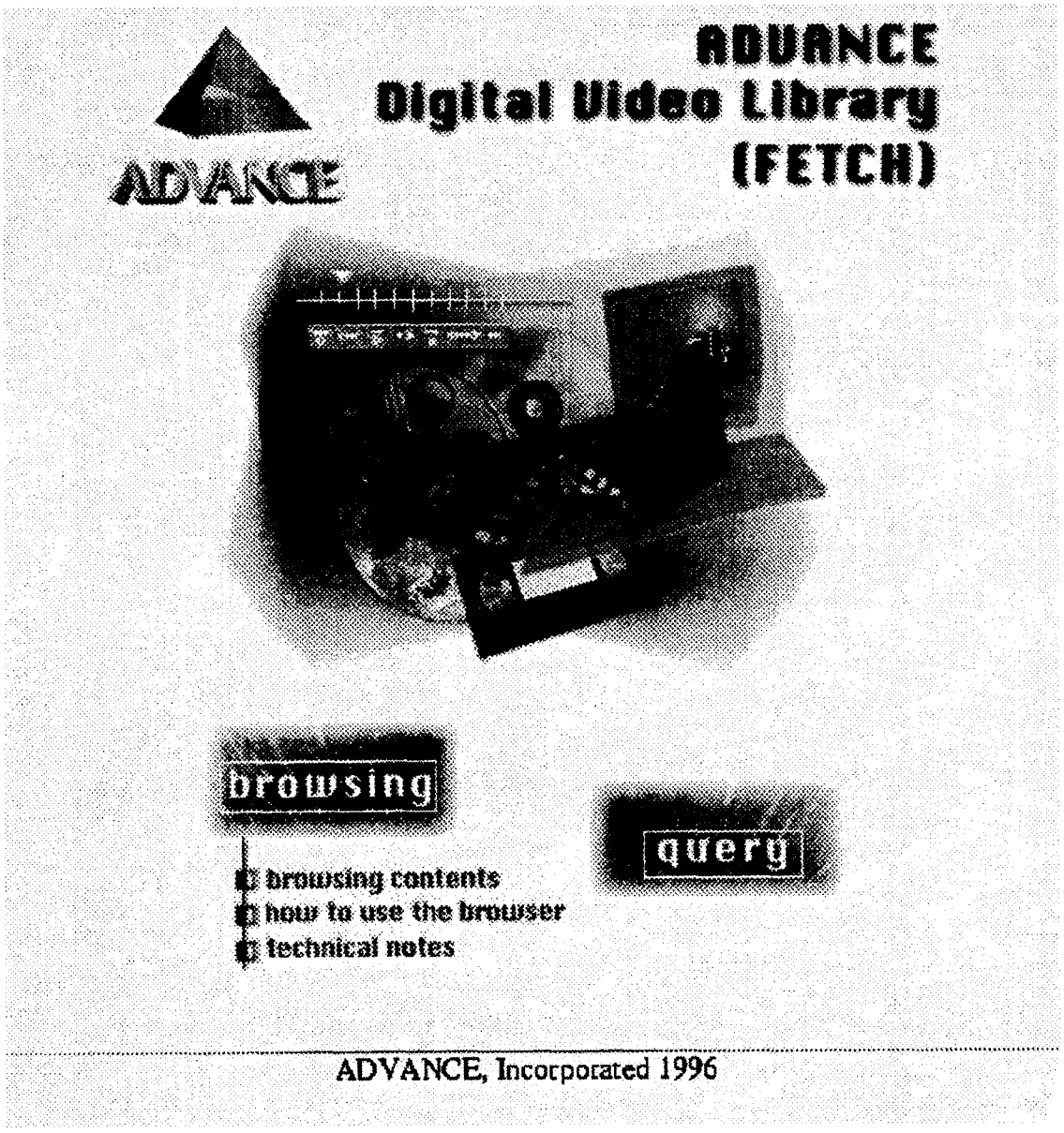
List of Future Features (con't)

- *Using Data Fusion to Help Image Retrieval*
- *Special Effect (music, gunfire, applause)*
- *Caption*
- *Speech Recognition*
- *Integration with Video Editing Systems*

List of Future Features (con't)

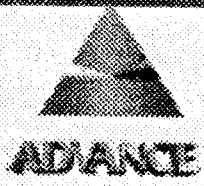
- *Database Support*
- *Billing System*
- *System Security*
- *Improved Image Search Engine*
- *Intelligent Text Search Engine*

The Interface



K-Db '97

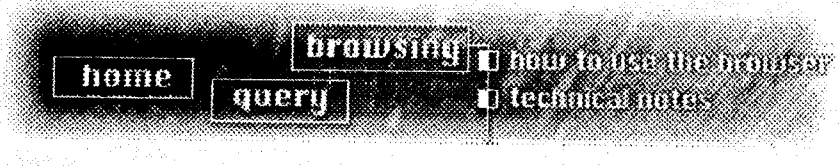
Browsing -- List



ADVANCE Digital Video Library (FETCH)

This is a searchable index. Enter search keywords:

Contents



The Browsing -- Catalog

ADVANCE Digital Video Library for DARPA

MAE UAV Predator Deployment

Summary

Title: MAE UAV Predator Deployment
Author: Unmanned Aerial Vehicle Joint Project
Date: 1995
Provenance: Battle Space
Keywords: predator deployment
Description: Predator has scored great success in accomplishing many missions and its deployment in Europe has been completed.

Storyboards:

- Weapon Only
- Complete program
- Scene Transition Graph

Playback options:

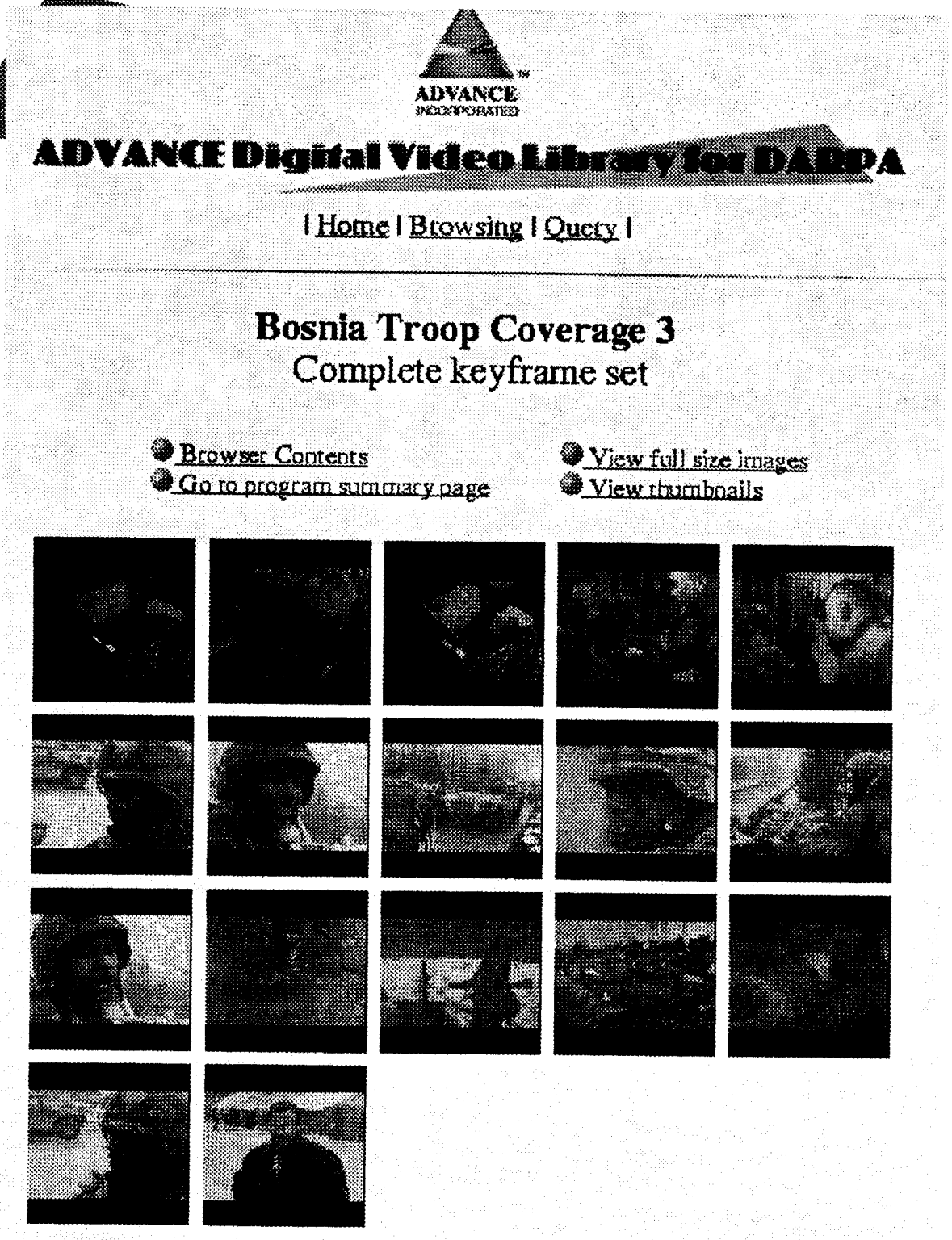
- video + audio
- audio only

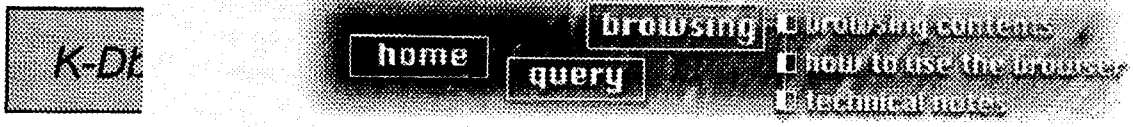
Browsing Contents

Home

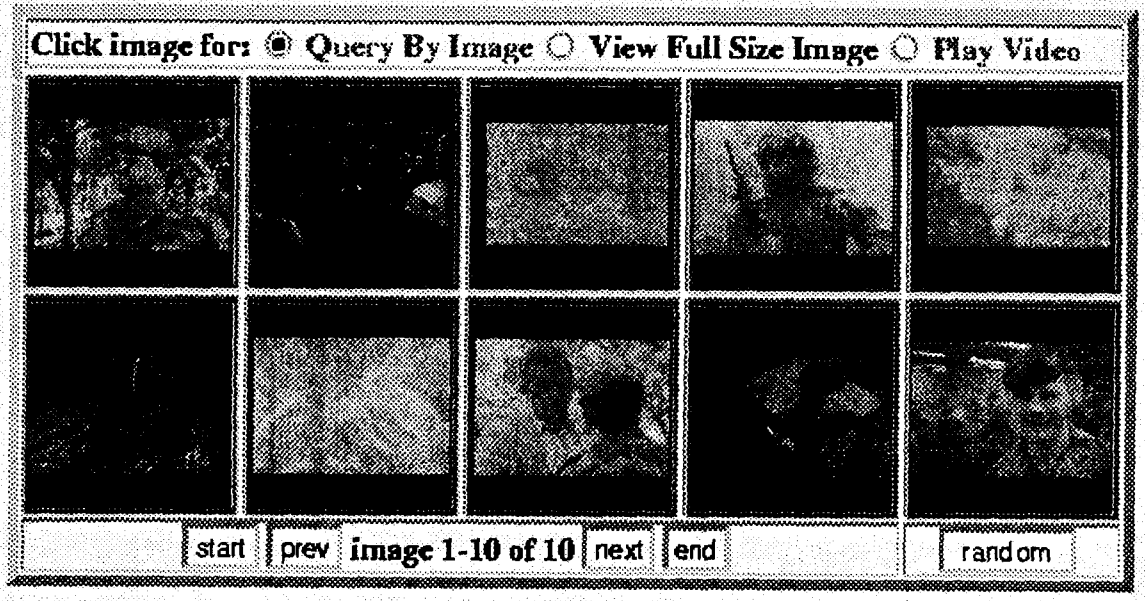
K-Db '97

Browsing -- Storyboard





Search Based on Similarity



Weight	Query Method
1	**COLOR**: average color of an image
1	**COLOR HISTOGRAM**: percentages of the different colors
1	**DRAW**: color layout of an image
1	**TEXTURE**: average texture of an image
1	**Keywords:** Keyword Only
1	**Concept Text Search:** Text Only
	There have been 4584 queries

K-Db '97