AUTOMATIC VIDEO INDEXING

Itxaso Bustos

Maite Frutos

TABLE OF CONTENTS

- Introduction

- Methods

  - Key-frame extraction

  - Automatic visual indexing

  - Shot boundary detection

  - Video OCR

- Index in motion

- Image processing algorithms

TABLE OF CONTENTS

- Types of knowledge

- Indexing tools

- Incremental galleries

- Video searching

INTRODUCTION

- Browsing, searching and manipulating video documents
  - → Index describing the video content is required.

- Until now → Carried out by documentalists

- Manual indexing → Expensive and time consuming task.

INTRODUCTION

- Therefore, Automatic Classification of video contest is necessary →

  Video Index: *"The process of automatically assigning content-based labels to video documents"*

INTRODUCTION

- "Raw" video streams → Structured and indexed database-driven information entities

- Maximize effectiveness of Web Contents.

- One of the major applications being recently addressed in international market.

INTRODUCTION

- Video indexing is comparable to text indexing or bookmarking.

- However, implementation is very different and much more difficult than the query mechanism for a textual database.

INTRODUCTION

- In a video document, three information channels are considered:
  - Visual modality.
  - Auditory modality.
  - Textual modality.

INTRODUCTION

- Two main techniques of Video Indexing:

    - Scene change detection technique can be applied for:

        - Scene Browsing.

        - Automatic and intelligent video indexing of video sequences for video databases.

    - Content-based indexing mechanisms such as:

        - Object texture, shape, color, motion, etc.

INTRODUCTION

- Video usually contains an enormous amount of visual information → Very important indexing consideration.

- Only keep a representative image from a long scene → This process is called "Keyframe extraction".

KEY-FRAME EXTRACTION

- Algorithm that provides summarization of a video by selecting a small number of frames, able to catch the visual and semantic properties of a video or video-shot.

- Problem:
  - Redundancy clutter → Most used systems today rely on the power of human observation to validate out coming data.

KEY-FRAME EXTRACTION

- In short:
  - An efficient and modular methodology for video indexing, presents a technological challenge.

AUTOMATIC VISUAL INDEXING

- QBIC (Query By Image Content)

  – Indexing method based on color, texture, shape etc.

  – Still images → Each video clip is segmented into small units (shots) then one or more keyframes in each unit are selected.

  – Each keyframe is automatically indexed using its visual characteristics.

AUTOMATIC VISUAL INDEXING

- Problems:

  – One clip usually contains more than one keyframe.

  – The number & density of keyframes in a clip can be problematic :

    - Long video.

    - Short clip.

OTHER METHODS

- Other supplementary methods for digital video libraries include:

  - Shot Boundary Detection.

  - Keyframe extraction.

  - Speech recognition (Using Sphinx speech recogniser).

  - Face detection.

  - Video OCR
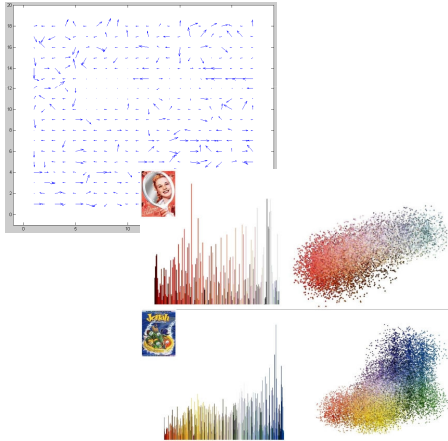
  - Image search based on color histogram features.

SHOT BOUNDARY DETECTION

- Major techniques used for shot boundary detection:

  - Pixel differences.

  - Statistical differences.

  - Edge differences.

Edged Image
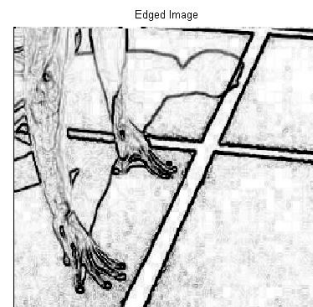
SHOT BOUNDARY DETECTION

- Major techniques used for shot boundary detection:

  – Motion vectors.

  – Histogram comparisons.

VIDEO OCR

- OCR – Optical Character Recognition

  – Mechanical or electronic translation of scanned images.

  – Search engine.

  – Field of research in pattern recognition, artificial intelligence and computer vision.

VIDEO OCR

- What kind of text occurrences should be considered?

Scene Text

OverlayText

VIDEO OCR

- In what kind of media data?

  – Treating video as a set of independent images.

  – Same text line occurs in videos for some time → Multiple instances of the same text line → Better detection.

- Font Attributes

VIDEO OCR

- Usages:

  - Video Indexing: Tolerates pixel errors in the localization and segmentation steps.

  - For object-based video encoding: Minimize the errors in pixel classification.

  - Visual removal of text from videos → Automatic translation into another language.

VIDEO OCR - DEMOS

- http://www.youtube.com/watch?v=fn1ZJ0HeeOc&NR=1

- http://www.youtube.com/watch?v=zkjBqea8RPM&NR=1

- http://www.youtube.com/watch?v=BLG28-xNoMU&NR=1

- Software: OpenCV

  - Library of programming functions for real time Computer Vision

INDEX IN MOTION

- The user can query by video clips' moving characteristics.

- Techniques:

  – Feature extraction.

  – Object tracking across frames.

- Early stage: Small number of experimental systems.

IMAGE PROCESSING ALGORITHMS

- **Camera Motion Estimation and Compensation:**

  - Recover parametric estimates of the camera motion between consecutive frames in the video.

  - Parameters used:

    – to locate independently moving objects

    – to label the sequence in terms of camera motion operations.

# IMAGE PROCESSING ALGORITHMS

- **Object Segmentation and Tracking Techniques:**

  – Objects of interest are segmented from the background based on color or motion attributes.

  – Tracked from across the different frames.

IMAGE PROCESSING ALGORITHMS

- **Line Detection:**

  – Used to detect structures of interest in the scene.

  – The algorithms incorporate operations like edge-detection, edge-linking and edge-thinning.
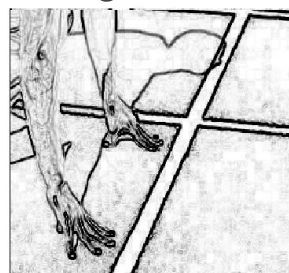
TYPES OF KNOWLEDGE

- **Physical Knowledge:**
  - Includes constraints derived from the physical environment.

- **Cinematic Knowledge:**
  - Includes the details on how the particular event is filmed and produced.

TYPES OF KNOWLEDGE

- **Semantic Knowledge:**
  - Includes the knowledge about the actual sport, the temporal structure of the game, the rules of the games and other high level information about the sport.

INDEXING TOOLS

- **Key-frame detection**

  – Is based on the similarity measures. A key-frame is a frame which contains the most information and is the most distinct from the other frames within a shot.
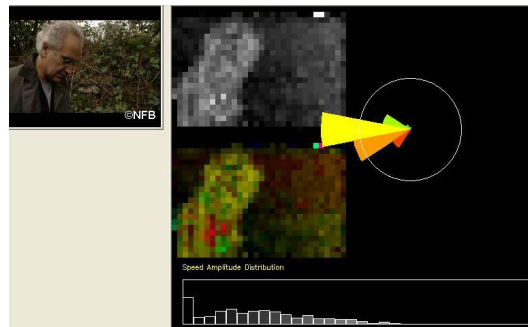
INDEXING TOOLS

- **Key-frame detection**

  – Two functions are used:

    • the "utility function": based on the entropy of the color distribution

    • the "frame distance function": based on the Bhattacharyya distance.

INDEXING TOOLS

- **Motion characterization**

  - Motion characterization is based on which calculates the optic flow on automatically detected salient points.

  - Visual interface provides local information through intensity and direction motion maps.

INDEXING TOOLS

- **Key-face detection**

  - Useful for automatic cast summarization and automatic extraction of face image training sets for face recognition.

INDEXING TOOLS

- **Key-face detection**

  - Two main steps:

    - frontal face detection: detected on each frame with a cascade of boosted classifiers

    - face image dissimilarity assessment: checks if the currently processed frame contains a face image similar to the last detected image face.

INDEXING TOOLS

- **Face recognition**

  - Face candidates are detected on each frame.

  - Spatial likelihood is assigned to each face candidates detected. The likelihood is used to increase face recognition performance in complex scenes.

  - Face recognition of a face candidate is done with an Hidden Markov Model encoding procedure.
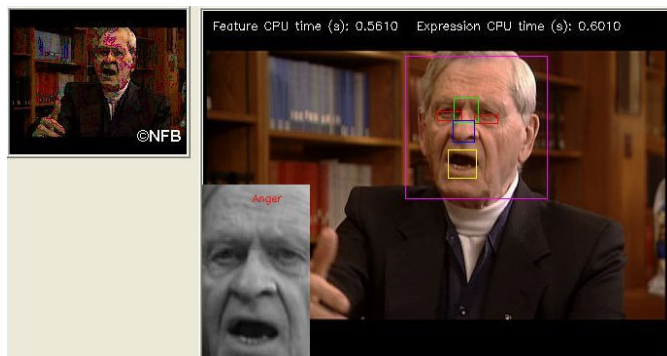
INDEXING TOOLS

- **Facial characterization**

  - Provides information regarding the facial expression of almost frontal faces.

    - Face feature detection

    - Face normalization: consists in using the position of features to scale and orient the face image according to a normalized pose.

INDEXING TOOLS

- **Facial characterization**

  - Facial expression recognition: seven basic emotions (anger, disgust, fear, happiness, sadness, neutral and surprise)

  - Emotions are detected globally on a frame by frame basis

INDEXING TOOLS

- **Text spotting**

    – Scanning the frame with a variable-size window.

    – First stage: simple classifier that discards as many false alarms as possible.

    – Second stage: the output of the cascade is a decision map showing regions of interest that may contain text, and on which a commercial OCR will be applied.

INCREMENTAL GALLERIES

- **Personal gallery is adapted by**

    – updating an entry of the gallery with the input (known person case)

    – adding the input as a new entry of the gallery (unknown person case)

VIDEO SEARCHING

- Searching by person, event, topic, date…

- Inconvenients:

  – Very high cost of maintenance with big and dynamic databases.

  – Creation of a hierarchy of categories which makes easy localization of documents to a large number of users.

REFERENCES

- www.daedalus.gr/prdinfo1.html

- www.ieee.org/

- ieeexplore.ieee.org

- Una panorámica de las Telecomunicaciones (Anibal R. Figueiras)

REFERENCES

- Related Papers:

  - Application of Content-Based Video Indexing to Computer-Assisted Descriptive Video

  - Automatic Video Indexing with Incremental Gallery Creation

  - Semantic Video Indexing: Approach and Issues

  - A Web-Enabled Video Indexing System

  - Automatic Text Segmentation and Text Recognition for Video Indexing