Multimedia Understanding through Semantics, Computation and Learning

Kıvanç Köse
Bilkent University
Turkey

Outline

- **Introduction to MUSCLE**
- **MUSCLE**
  - Why the project initiated, what are the objectives ?
  - What the project deals with ?
    - Creating a showcase for content analysis and audio-visual search
    - Visual saliency, focus of attention and habituation
    - Detecting and recognizing humans and human behaviour
  - How the project deals with ?
    - Objectives and motivation
    - Planned activities
- **Accomplishments**
- **Future Plans**

Introduction to **MUSCLE**

- Supported by the European Commision under Framework Programme 6
- Network of Excellence (NoE)
- Coordinator: Nozha Boujemaa, INRIA France
- Manager: ERCIM – European Research Consortium for Informatics and Math.
- 41 partners, 8 work-packages
  - Turkey,France,Greece,UK,Austria,Netherlands,Finland,Sweden, Spain,Czech Republic,Israel,Ireland,Hungary
- www.muscle-noe.org
- Budget : 6.9 Million Euro

I think we should drop this slide

Educational Side of the Project

- The biggest deficiency of European Universities is at MSc and PhD degrees
- One of the goals of the NoE is to obtain U.S. level PhD education in Europe
- The research groups in Europe are small.
- NoE projects gather several researchers from different institutes and universities. This brings out collaborative work. → Good Quality PhD education.

Why **MUSCLE** initiated?

- Priority list for next 10-15 years in Networked and Electronic Media (NEM)
  - How can people face with incredible amount of data accessible in various places
  - How relevant content from media libraries can be extracted with minimal assistance
- State of the Art not sufficiently advanced to realize NEM
  - Semantic search algorithms should be integrated in audio-visual applications
  - Digital content can be indexed in an efficient and standardized manner manually, semi automatically or automatically

What are the Objectives

- *Scientific* : Improving content-based multimedia retrieval based on semantic metadata by:
  - Improving single modality processing
  - Combining different modalities to improve robustness
  - Harnessing machine learning to (semi)automate extraction manipulation and understanding of semantic metadata
- *Integration* : Establish durable integration between key scientific and industrial players

Example Applications of **MUSCLE**

- Security and watching systems
  - Alarming automatically from the video
  - Human detection
  - Suspicious object detection
- Multimedia databases
  - Image retrieval
  - Image-Sound matching (News Reports)
- Patient and Old people tracking
  - Any body fell or faint of crawling in the scene
- Automation in the factories
  - Defective car part, agricultural product, circuit boards

What & How **MUSCLE** deal with?

- Creating a showcase for content analysis and audio-visual search
- Visual saliency, focus of attention and habituation
- Detecting and recognizing humans and human behaviour

Creating a showcase for content analysis and audio-visual search

- This is the brainstorming part
- Media data are collected by each contributor
  - News reports, music clips, etc…
- This data is integrated and shared among all the participants.
- Everyone participating this part will try to come up with semantic extraction methods.
  - Ex: Face, fire and smoke, text, logo, etc… detection
- The ambition in this part is to act as an inspiration pump and stimulating collaboration (no TRECVID contest!)
  - By applying all those proposed techniques will lead the team to new methodologies. See pros and cons!!

Creating a showcase for content analysis and audio-visual search

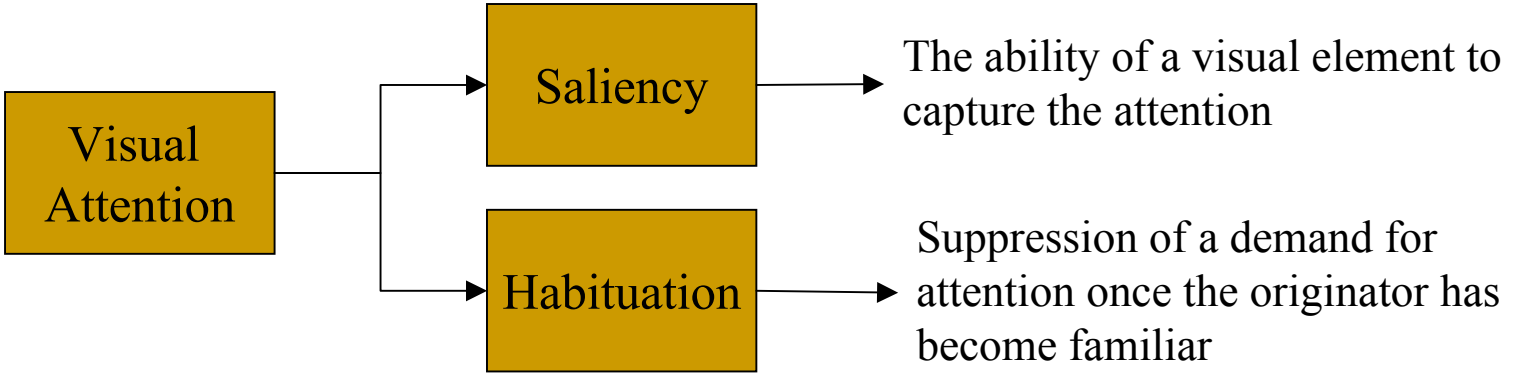
- The result will be:
  - A video to be used for analysis, serving as a common data set.
  - See the know-how of the consortium.
  - A set of different features for audio, images, and text extracted from this video.
    - Can be used later by other partners in MUSCLE
  - A comprehensive set of annotations performed on this video using the variety of Machine Learning techniques employed by MUSCLE members

State-of-the-art showcase of the capabilities and competences on feature extraction and semantic annotation within MUSCLE !!!

What & How **MUSCLE** deal with?

- Creating a showcase for content analysis and audio-visual search
- Visual saliency, focus of attention and habituation
- Detecting and recognizing humans and human behaviour

Visual saliency, focus of attention and habituation



- Accurate computational models for saliency should be found.
  Applications:
  - Foreground-background segmentation, compression, similarity scoring and retrieval, visual inspection, etc…
- Because of the importance of the saliency, number of different approaches will be explored
  - Probabilistic Probing, Geometric configuration of local interest points, Spatiotemporal visual attention, Habituation.

What & How **MUSCLE** deal with?

- Creating a showcase for content analysis and audio-visual search
- Visual saliency, focus of attention and habituation
- Detecting and recognizing humans and human behaviour

Detecting and recognizing humans and human behaviour

- The subject of the audio-visual data is mostly human activities and so the main actor is human
  - Automatic recognition of humans and human behaviors will make the data easier to comprehend & interpret
- Different MUSCLE groups will set up a coordinated action
  - Ties in closely with the Content Analysis Showcase

Detecting and recognizing humans and human behaviour

- Planned Activities and contributions
  - Body Detection
    - Use background learning tech. for object detection
    - Single or Multi cam
    - Context Knowledge exists or unconstrained environment
  - Body Analysis
    - Human body model will be extracted and systems are trained according to those models.
  - Person Tracking in Crowded Scenes
  - Adaptive Detection
    - Choosing among different techniques for different situations
    - Robustness and applicability
  - Emotion Recognition
    - Besides what is said, in which mood it was said is also important
    - Face recognition, gesture recognition

MUSCLE-Bilkent grubu

- 4 ogretim uyesi: 2 EE, 2 CS: A. Enis Cetin, L. Onural, O. Ulusoy, U. Gudukbay
- 4 doktora ogrencisi
- Benzer konularda iki Ingiliz sirketi icin yazilim gelistiriyoruz

Accomplishments

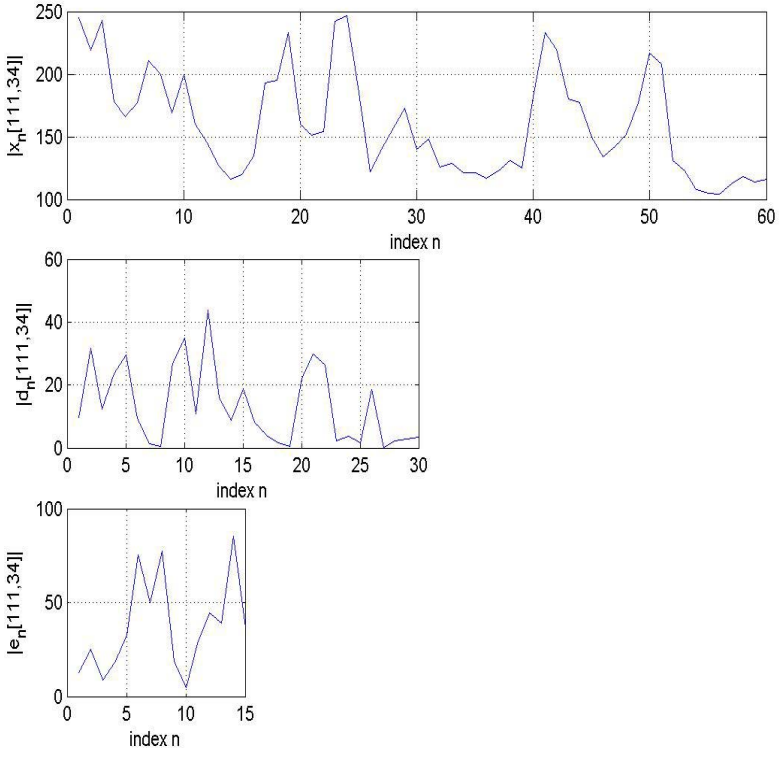
- More than 10 post-doc fellowships (jointly supervised by 2 labs) and counting…
- Exchange and collaboration visits sponsored by MUSCLE mobility grants for students and researchers
- Joint summer-school with DELOS, lecture videos online
- Joint research projects involving teams of several network participants
- Software, publication and books
- Showcases, see www.muscle-noe.org

Some examples of MUSCLE research

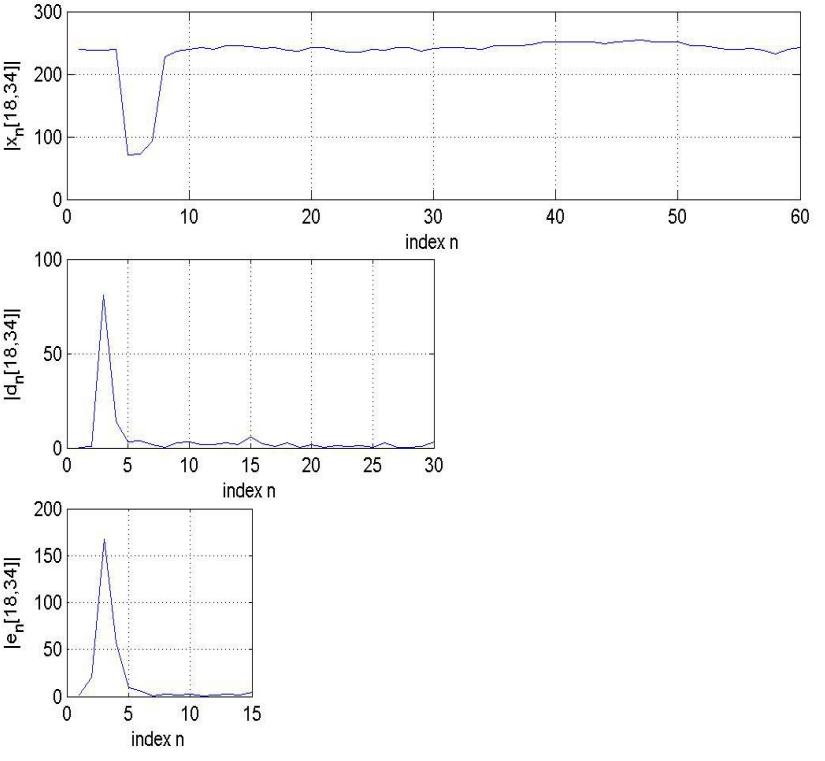
- Fire and Smoke detection, Recognizing humans and human behaviour in video: detecting falling people (Enis Cetin, Bilkent U.)
- Automatic interpretation and annotation of sports videos (W. Christmas, U. Surrey)
- Alternative Interfaces to Large Music Collections (Andi Rauber, TUVienna)

Temporal analysis of Pixels for fire Detection

•Flame pixel history

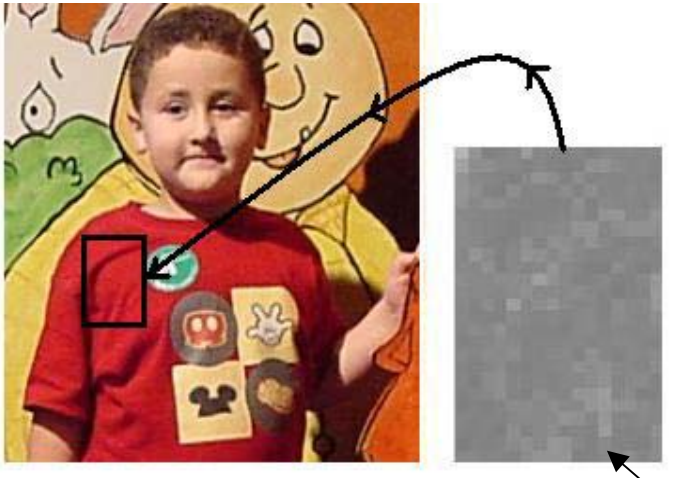


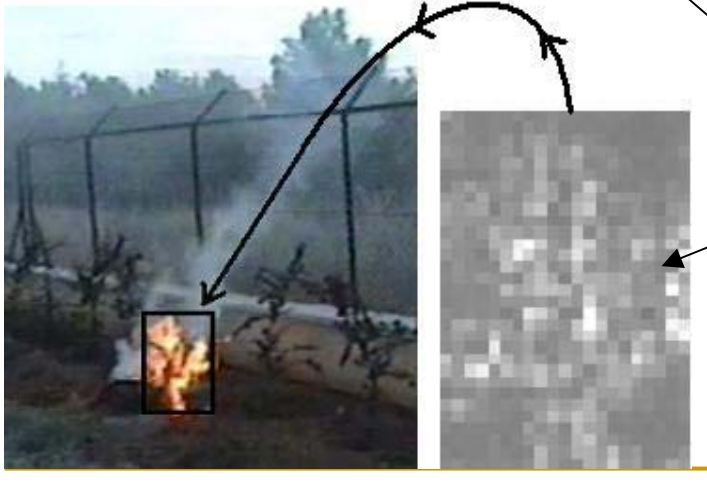
•Moving pixel history



• Flicker is detected by analyzing the temporal variation of the red channel image pixel values: HMMs can be trained in the wavelet domain: Robust state definitions

Spatial Analysis for Fire Detection

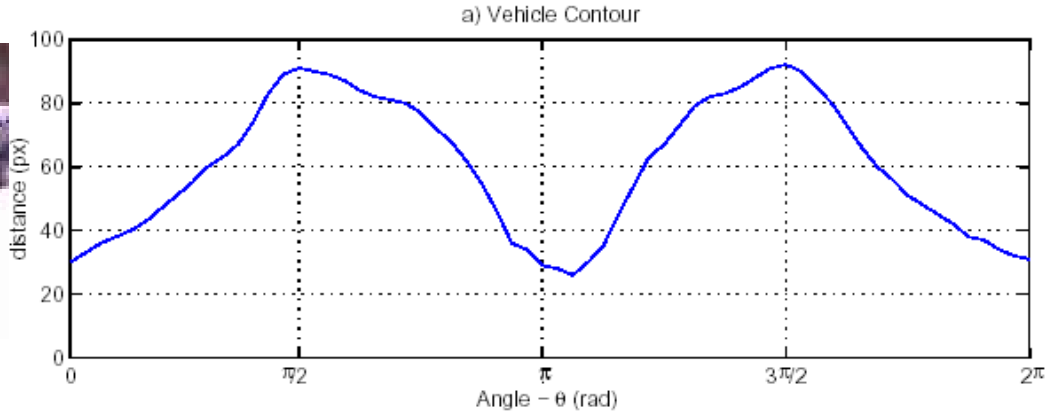


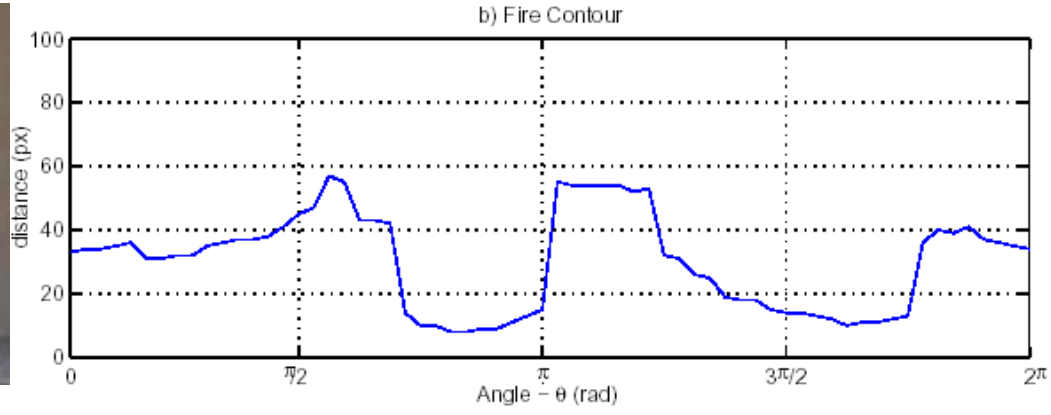


- Flame pixels exhibit a similar spatial variation in their chrominance and luminosity values

$$|x_{lh}[k, l]|+|x_{hl}[k, l]|+|x_{hh}[k, l]|$$

Contour Analysis



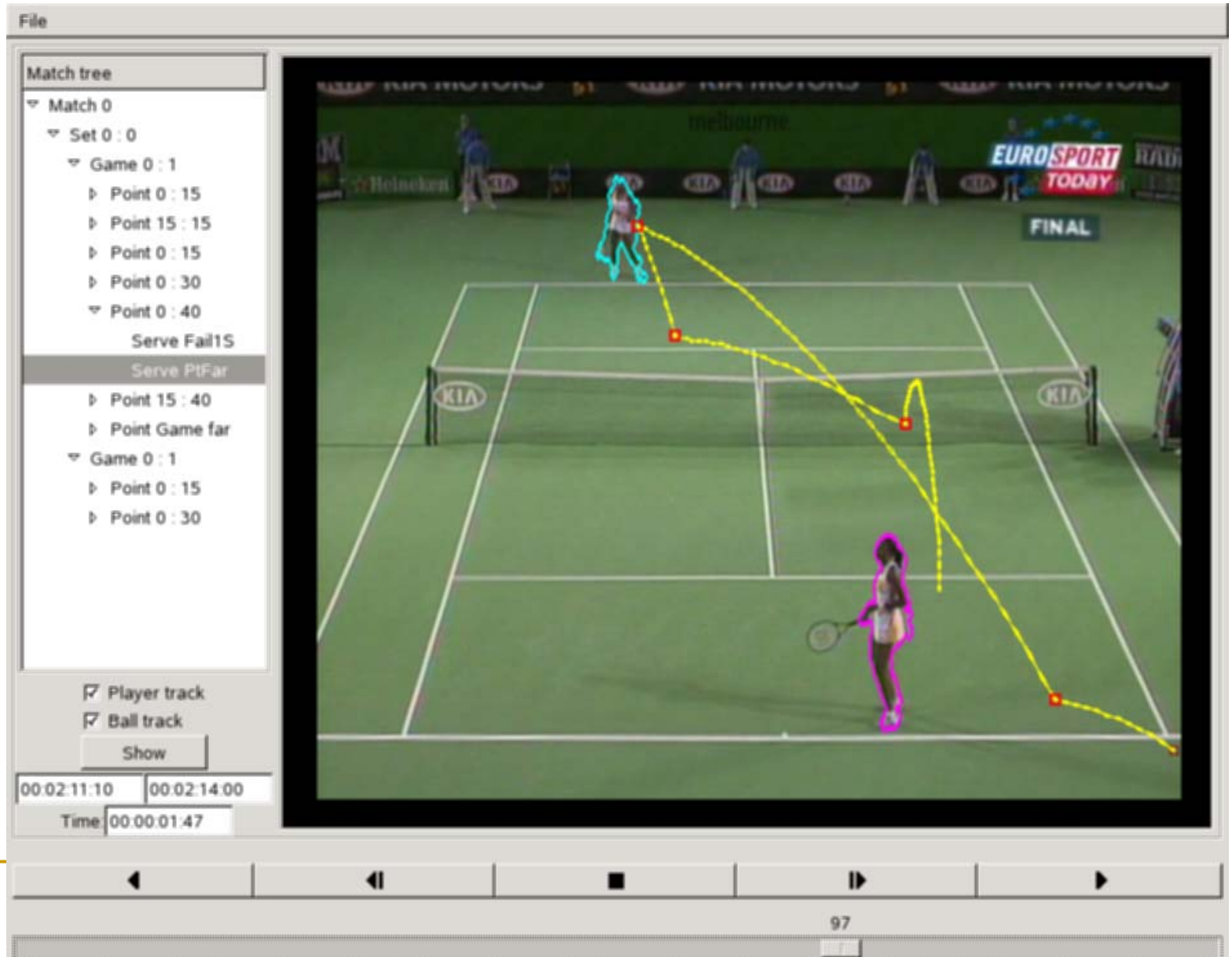


Fire Smoke Detector

- The smoke detector part is sold to a power plant in Singapur

- For more information
  - http://signal.ee.bilkent.edu.tr/VisiFire/
  - Contact : ugur@ee.bilkent.edu.tr, cetin@ee.bilkent.edu.tr

Automatic Annotation of Tennis Videos



Some examples of MUSCLE research

- Detecting Falling People
- Human Detection in Difficult Scenarios by Combining Motion and Appearance
- An Online Learning Framework for Object Detection, Tracking and Recognition.
- Eyetracking improves image retrieval
- Visual Tags for Mobile Services
- Automatic classification of coins
- Accurate delineation of piecewise smooth optical flow
- RETIN: An Interactive Content-Based Image Retrieval System

Future Plans

- Continue contribution to high quality training of young researchers (exchange visits, mobility grants, fellowships)
- Organization of Summerschools and online dissemination
- Organization and participation in **benchmarking events** (e.g. ImageCLEF, Coin Image Classification Contest,…)
- More emphasis on integration of components into systems: e.g. how can single modality features (developed by different groups) be combined into high performance content retrieval systems. This will be stimulated by sponsored integration workshops.
- The above-mentioned systems will used to foster closer links with industry