SSMS 2006 Lecture: Semantic Multimedia Analysis

Jan Nesvadba, Philips Research Eindhoven, The Netherlands
jan.nesvadba@philips.com

Agenda

- Part 1:
  Real-Time and Distributed Multimedia Analysis System (~90 min)
- Break (15 min)
- Part II:
  Multimedia Content Analysis Service Units in Distributed Content Analysis System (~75 min)

SSMS 2006 Lecture: Semantic Multimedia Analysis

Part 1: Real-Time and Distributed Multimedia Analysis System

Jan Nesvadba, Philips Research Eindhoven, The Netherlands
jan.nesvadba@philips.com

Agenda

- Evolution of Technology and Consumer Trends
- Media Management with AmbientDB
- Real-Time and Distributed Multimedia Analysis System
- Conclusions

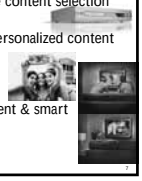
SSMS 2006 Lecture: Semantic Multimedia Analysis

Part 1a: Evolution of Technology and Consumer Trends

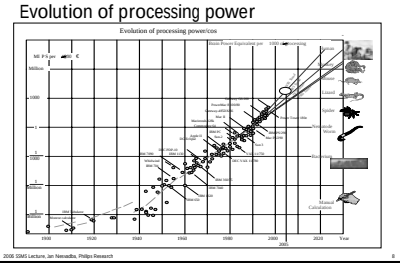
Jan Nesvadba, Philips Research Eindhoven, The Netherlands
jan.nesvadba@philips.com

Evolution of Technology and Consumer Trends

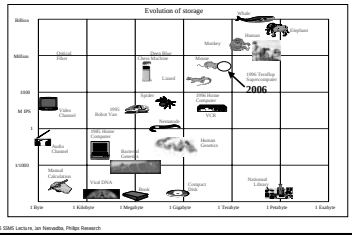
- Evolution of processing / storage / connectivity
- Trend 1: Passive consumption to active content selection
- Trend 2: From uniform reception to personalized content creation and individual consumption
- Trend 3: From passive devices to ambient & smart environments and smart content



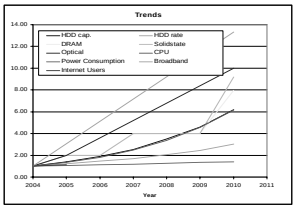
Evolution of processing power



Evolution of storage



Evolution of connectivity



The burden of too much…

Home Theater of the Future?

Trend 1: Passive consumption to active content selection

Trend 2: From uniform reception to personalized content creation ..

.. and individual consumption

Trend 3: From passive devices to ambient & smart environments and smart content

Smart environment and smart content

Vision

Ambient Intelligence

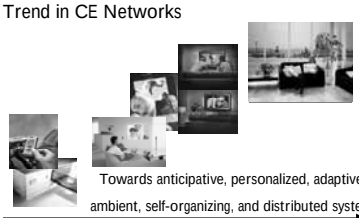
- Sensitive, adaptive, and responsive to consumers and anticipatory to consumer desires
- Required:
  - Smart, personalized and interconnected systems
  - Ubiquitous and transparent

Towards Realizing the Vision

connectedplanet

A world where consumers enjoy great entertainment experiences and services whenever and wherever they want

Trend in CE Networks



Towards anticipative, personalized, adaptive, ambient, self-organizing, and distributed systems

time

Architecture of a
very solid & very general

anticipative, personalized, adaptive, self-organizing, distributed and generic

audiovisual processing and storage system

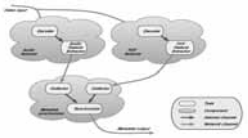
100b neurons each 25k connections, neuroarchitecture, 1.5 kg, 1300-1600 cm³, 20% energy of body

Specialized Units at (neo)cortex, thalamus, cerebellum and Hippocampus e.g. Visual Analysis at Visual Cortex

Van Essen Nature 1992

Distributed multimedia content analysis system

- Component-based Streaming Framework
  - Specialized expert Service Units
  - Aim: generic / domain agnostic: in-home, security, mobile, automotive and medical (care) applications



SSMS 2006 Lecture: Semantic Multimedia Analysis

Part 1b: Media Management with AmbientDB

Jan Nesvadba, Philips Research Eindhoven, The Netherlands
jan.nesvadba@philips.com

Media Management with AmbientDB

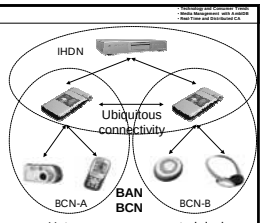
- Vision of AmbiDB
- AmbiDB Requirements
- From Physical to Logical Approach
- Current implementations
- Conclusions for AmbiDB

CE Networks

Device Class	Storage Capacity
Stationary *(e.g. e-Hub or Home Server)*	100s GB
Smart Mobile *(e.g. Personal Infotainment Companion, Mobile Server)*	10s GB
Mobile Peripheral	10s-100s MB

⇧ Storage growth

➔ **Need for integrated media management solutions**



Heterogeneous connected devices (stationary, mobile, sensors)

IHDN = In-Home Digital Network
BCN = Body Centric Network
BAN = Body Area Network

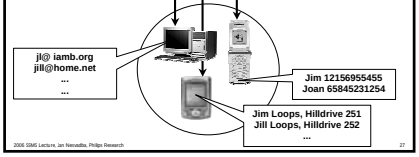
Vision AmbientDB

- Integrated Media Management
  → Transparent across heterogeneous platforms (legacy & future devices and sensors)
  → high-level P2P (meta)data management functionality in distributed middleware layer
  ➔ global DB abstraction layer over ad-hoc networks of heterogeneous peers

➔ Applications / consumers can transparently accessing media (meta) data on all devices

Integrated Media Management example

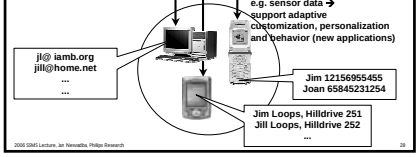
Jim Loops	; Hilldrive 251	; 12156955455	; jl@iamb.org
Jill Loops	; Hilldrive 252	;	; jill@home
Joan	;	; 65845231254	;



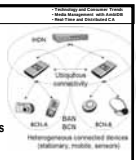
Integrated Media Management example

Jim Loops	; Hilldrive 251	; 12156955455	; jl@iamb.org	; home
Jill Loops	; Hilldrive 252	;	; jill@home	; Jims
Joan	;	; 65845231254	;	

e.g. sensor data ➔ support adaptive customization, personalization and behavior (new applications)



Distributed Memory AmbientDB: **Requirements**



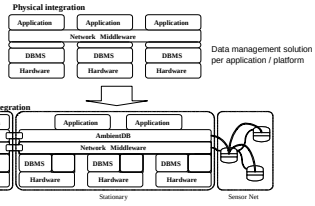
Distributed Memory Faculties
→ **Need for integrated media / metadata management solutions**

- P2P self-organization
- Distributed query processing
- Adaptive synchronization incl. on- / off-line media management ('Snapshot')
- Scalability (mobile – stationary)
- Robust media / metadata management across heterogeneous data sources

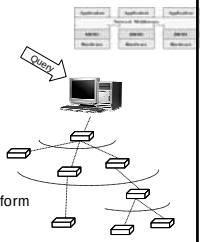
→ **Everything transparent to the application**

AmbientDB: Distributed P2P DBMS

Traditional approach

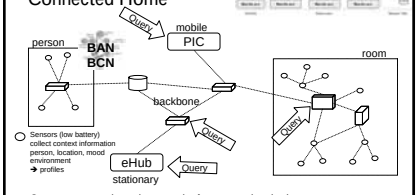


Traditional Networks



- Single root (query source)
  - Base station
  - Global queries
- Many nodes
- Homogeneous network
- Power optimizations
- One (meta)data management solution per application / platform

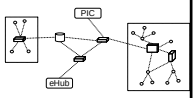
Networks in Connected Home



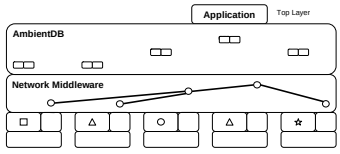
• Query can enter through any node → no central authority

Networks in Connected Home

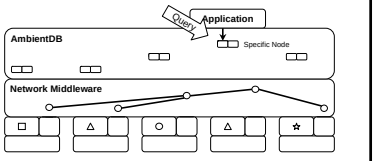
- Fewer nodes
- Heterogeneous system
- Complex ontology
- Complex network topology
- Dynamic
- Data Flow optimized



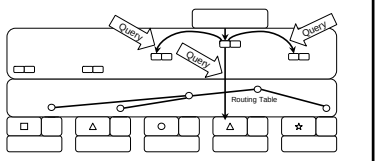
Distributed query processing



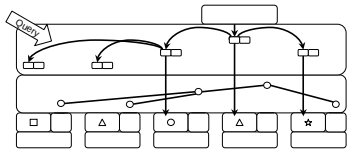
Distributed query processing



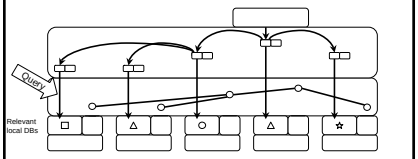
Distributed query processing



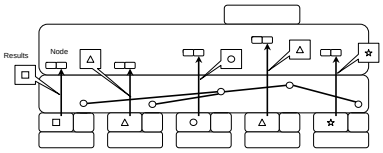
Distributed query processing



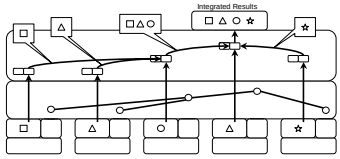
Distributed query processing



Distributed query processing



Distributed query processing



Goal with AmbientDB

Provide user with experience
of having all digital AV media available
at any time, in any place,
regardless of connection availability
in heterogeneous environment

Example Metadata Store

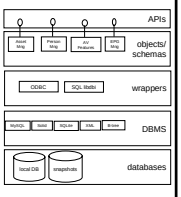
Schema-dependent metadata abstraction layer

Object-oriented interfaces
(to support metadata standards DIDLite, MPEG-7)
(API standards: ODBC, JDBC)
Object-oriented data models
(performs mapping onto relational model)

Relational data models and interfaces
(e.g. SQL API)

Transparent usage of any DBMS

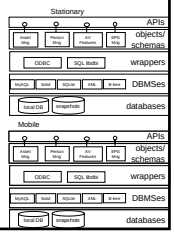
Transparent (meta)data management on local (on-line) and remote datasets
('Snapshot' – off-line) incl. smart sync



DBMS = DataBase Management System

Example Metadata Store

- Transparent Data Management
- Operation in heterogeneous environments
  - Multiple protocols
  - Active and passive devices
  - Mobile and stationary



Conclusions for AmbiDB

- Transparent distributed media (meta)data management is crucial for future Networks and Ambient Intelligence
  - Disconnect applications from hardware and network protocols
  - Connect all (relevant) heterogeneous (meta)data sources to present a single view

SSMS 2006 Lecture: Semantic Multimedia Analysis

Part 1c: Real-Time and Distributed Content Analysis System

Jan Nesvadba, Philips Research Eindhoven, The Netherlands
jan.nesvadba@philips.com

Real-Time and Distributed Content Analysis System

- Introduction to content analysis
- PC based content analysis feature assessment
- A content analysis system based on PC technology
- Real-Time and Distributed Content Analysis System
- Conclusions

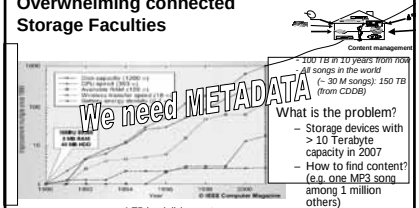
Introduction to content analysis

- **Contents**
  - Background / rationale
  - Example application
  - "How it works"

Content Management in Consumer Systems

- The future:
  - Distributed storage capacity of >10 Tbyte in home networks
- The problem:
  - Easy categorizing, searching, managing and retrieving AV content
- Required:
  - Automatic generation of metadata (content descriptors)
  - System solutions for distributed content analysis and fast prototyping of applications
  - Self-organizing smart media management (e.g. AmbiDB)

Overwhelming connected Storage Faculties



- *100 TB in 10 years from now*
- *All songs in the world (~ 30 M songs): 150 TB (from CDDB)*

What is the problem?
- Storage devices with > 10 Terabyte capacity in 2007
- How to find content? (e.g. one MP3 song among 1 million others)

1 TB hard disk can store:
200 DVDs / 1000 VCDs
2000 CDs / 200.000 MP3 songs @ 5 MB per MP3 song
10.000.000 HiRes JPEGS @ 100 kB per JPEG

Introduction to content analysis

- **How to get the metadata?**
  - From content provider
  - Do it yourself
    - **CONTENT ANALYSIS**

Introduction to content analysis

- **Why content analysis?**
  - When no metadata is available
  - To enrich the provided metadata
  - To enable searching within the content
  - To modify the content to improve the viewing experience / navigation
  - To enable new features

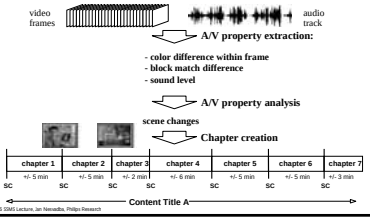
Introduction to content analysis

- **Example application / feature**
  - Intelligent chaptering for DVD+RW

- Create chapter markers intelligently
  - Not every 5 minutes, but
  - at every scene boundary, but
  - not more often than 1 per minute

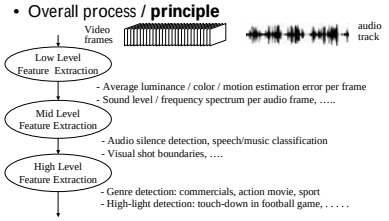
Introduction to content analysis

- Example "Intelligent chaptering": **how it works**



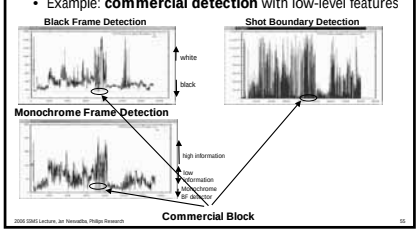
Introduction to Content Analysis (CA)

- Overall process / **principle**



Introduction to content analysis

- Example: **commercial detection** with low-level features



PC technology for feature assessment

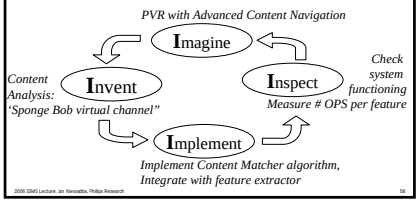
- **Contents**
  - Background / Rationale
  - Basic principles
  - The network as the system bus

PC technology for feature assessment

- **Background: What do you want?**
  - Fast evaluation of content analysis features
  - Fast product-concept assessment

PC technology for feature assessment

- **How**



Requirements for MCA Feature Assessment

- **P**roblem
  - Various expert teams (disciplines) , various solution spaces (languages)
  - Frequent changes (updates) and CPU performance & memory hungry
- Solution **requirements**
  - Minimal modification to feature implementations (no optmization)
  - Minimal dependencies between features & control (decoupled)
  - Standardized, easy "add-on" communication
- Solution **space**
  - Heterogeneous environments
    - Linux, Windows, TCP/IP, UPnP, SDE's

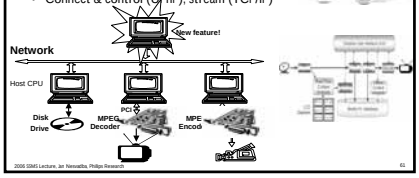
PC technology for feature assessment

- **Contents**
  - Background / Rationale
  - **Basic principles**
  - The network as the system bus

Requirements for MCA Feature Assessment

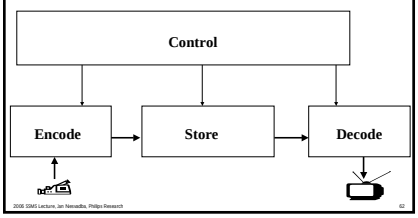
Basic principle

- Integration: minimal effort
- Pack into IP networked, executable component
- Connect & control (UPnP), stream (TCP/IP)



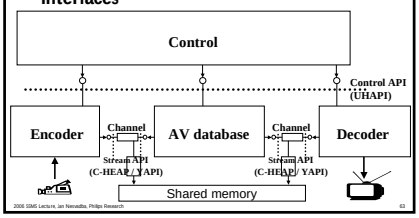
PC technology for feature assessment

- **Basic streaming application example**



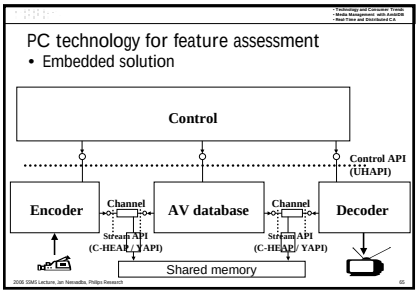
PC technology for feature assessment

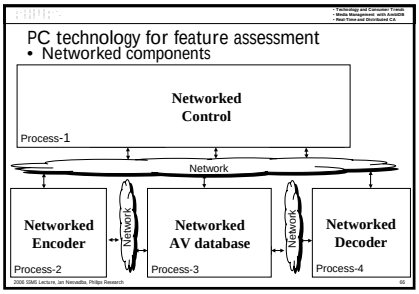
- **Software architecture with standardized interfaces**

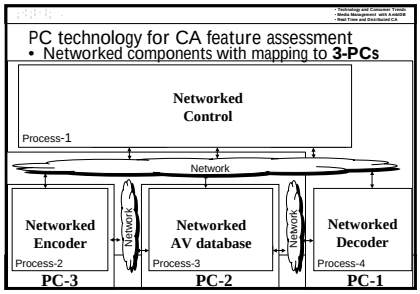


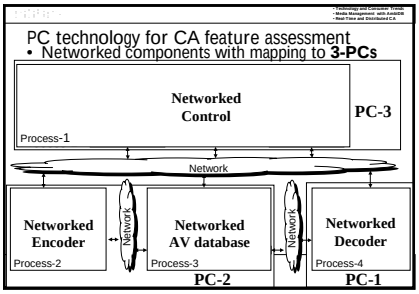
PC technology for feature assessment

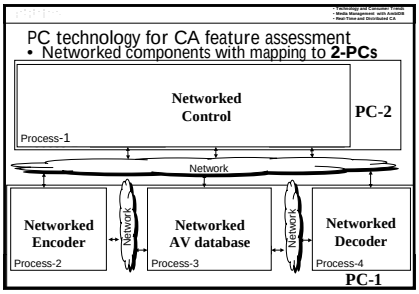
- **Contents**
  - Background / Rationale
  - Basic principles
  - **The network as the system bus**
    - All components communicate via network
      - Takes advantage of fast developments in computer industry: HW & SW standards
      - Scalable
      - Minimal effort for integration

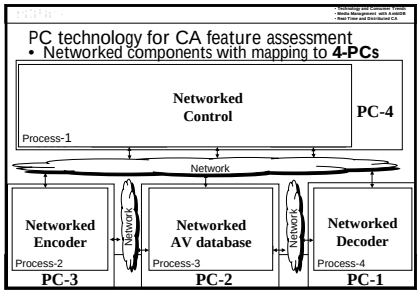

PC technology for feature assessment
• Embedded solution
Control
Control API (UHAPI)
Encoder
Channel
AV database
Channel
Decoder
Stream API (C-HEAP / YAPI)
Stream API (C-HEAP / YAPI)
Shared memory

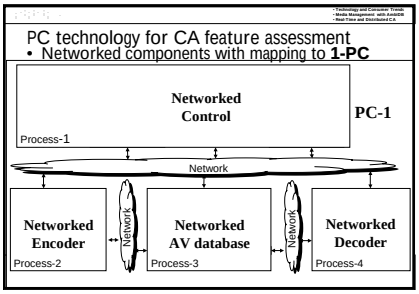

PC technology for feature assessment
• Networked components
Networked Control
Process-1
Network
Networked Encoder
Process-2
Network
Networked AV database
Process-3
Network
Networked Decoder
Process-4

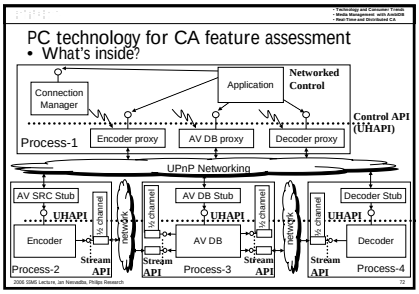

PC technology for CA feature assessment
• Networked components with mapping to 3-PCs
Networked Control
Process-1
Network
Networked Encoder
Process-2
Network
Networked AV database
Process-3
Network
Networked Decoder
Process-4
PC-3
PC-2
PC-1

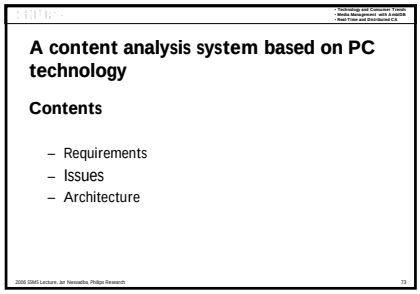

PC technology for CA feature assessment
• Networked components with mapping to 3-PCs
Networked Control
PC-3
Process-1
Network
Networked Encoder
Process-2
Network
Networked AV database
Process-3
Network
Networked Decoder
Process-4
PC-2
PC-1


PC technology for CA feature assessment
• Networked components with mapping to 2-PCs
Networked Control
PC-2
Process-1
Network
Networked Encoder
Process-2
Network
Networked AV database
Process-3
Network
Networked Decoder
Process-4
PC-1


PC technology for CA feature assessment
• Networked components with mapping to 4-PCs
Networked Control
PC-4
Process-1
Network
Networked Encoder
Process-2
Network
Networked AV database
Process-3
Network
Networked Decoder
Process-4
PC-3
PC-2
PC-1


PC technology for CA feature assessment
• Networked components with mapping to 1-PC
Networked Control
PC-1
Process-1
Network
Networked Encoder
Process-2
Network
Networked AV database
Process-3
Network
Networked Decoder
Process-4


PC technology for CA feature assessment
• What's inside?
Connection Manager
Application
Networked Control
Control API (UHAPI)
Process-1
Encoder proxy
AV DB proxy
Decoder proxy
UPnP Networking
AV SRC Stub
UHAPI
Encoder
Stream API
Process-2
½ channel
network
AV DB Stub
UHAPI
AV DB
Stream API
Process-3
Decoder Stub
UHAPI
Decoder
Stream API
Process-4


A content analysis system based on PC technology
Contents
– Requirements
– Issues
– Architecture

Content Analysis System based on PC technology

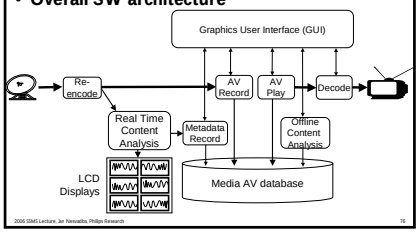
- **Requirements for content analysis demonstrator**
  - Show all AV content analysis research in one real-time demonstrator
    - Advanced CA algorithms from different disciplines, running in parallel
    - Real-time display of CA results (GUI)
  - Show offline navigation and playback of stored content based on content analysis results
    - Storage of content descriptors in a media database (e.g. mySQL)
    - Storage of A/V content in Real Time File System
    - Link with GUI for navigating & playback of stored content.

Content Analysis System based on PC technology

- **Issues**
  - Mixed environment (linux-windows)
  - Multiple PC + dedicated HW
    - (x86/pnx7300/pnx1300/pnx1500/..)
  - Real-time streaming of data between >30 modules
  - Synchronisation of audio-video content analysis data
  - Collaboration of several projects & people on different locations
  - System should be built from scratch and be ready in less than 3 months!!!

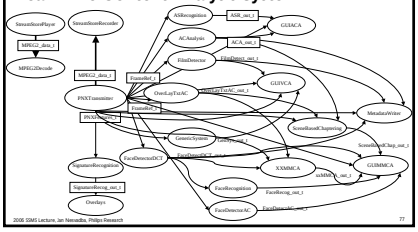
Content Analysis System based on PC technology

- **Overall SW architecture**



Content Analysis System based on PC technology

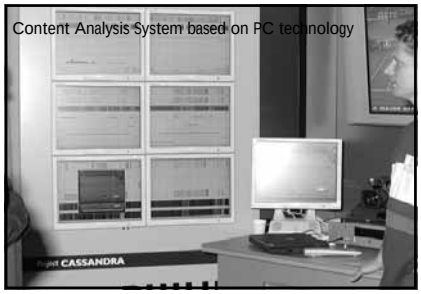
Real-Time Content Analysis System



Content Analysis System based on PC technology

- **Feature Visualization**

Real-time visualization framework
- Extendable and flexible
- Easy configuration

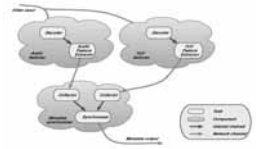


Real-Time and Distributed Content Analysis System

- Cassandra Framework
  - Real-time and distributed streaming system
- AV Content Analysis algorithms
  - E.g. AV segmentation, commercials, speech, faces
- Content management application
  - Easy and intuitive navigation and management of stored or received audio or audiovisual content.

Real-Time and Distributed Content Analysis System

- Streaming framework for rapid prototyping
  - For in-home, in-vehicle and medical applications
  - Example: distributed video content analysis:



Real-Time and Distributed Content Analysis System
Features of distributed content analysis system

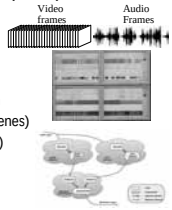
- Extensible, Scalable, Flexible, Upgradeable
  - Modularization of SW/HW components e.g. distributed Service Units
  - Standardized protocols / interfaces / API's (e.g. UPnP, XML/MPEG-7)
  - Easy extensibility – scalable architecture,
  - (Function) Upgradeability
  - Developed for heterogeneous networks / environments (of-the-shelf HW)
- Easy-to-use software architecture
  - *Service Units* encapsulate functionality (components, black box)
  - *Service Units* functionalities exposed on the network by standardized APIs
  - *Applications* as compositions of *Service Units* (e.g. CBD applications)
- Efficiency
  - Grid computing (resources shared, processing farms)
  - Auto load balancing (usage of idle unused processing)

Real-Time and Distributed Content Analysis System
Features of distributed content analysis system

- Reliability
  - No interference of unstable components
    - Auto-Error detection, Auto-recovery
  - → Self-organizing, autonomous system
- Connectivity:
  - TCP/IP backbone (future WiFi / QoS)
- Fusion and management of data sources
  - Common DB, Auto-Sync

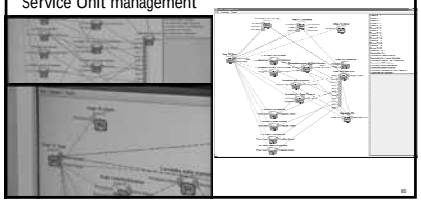
Content-Awareness Creation
Multimedia Content Analysis (MCA) Features

- Colour information
- Camera motion
- Video Segmentation (shots)
- Audio Classification
- Automatic Speech Recognition
- Semantic AV segmentation (scenes)
- AV Genre Classification (Adds)
- Face detection
- Emotion Recognition

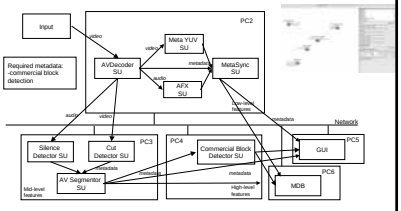


Self-organizing distributed system

Service Unit management

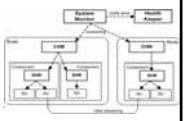


Example: commercial block detector



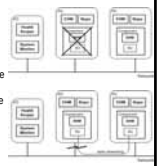
Error detection in Cassandra Framework

- Motivation
  - Services may disappear
  - More points of failure: no centralized maintenance of a big number of services (over 50)
  - Integration of 3rd party MCA components
- Efficiency parameters
  - Detection time
  - Overhead
  - Accuracy
- Fault model
  - Service / Component / host breakdowns
  - Communication (network) failures
- Fault model
  - push/pull (heartbeats / liveness requests) units

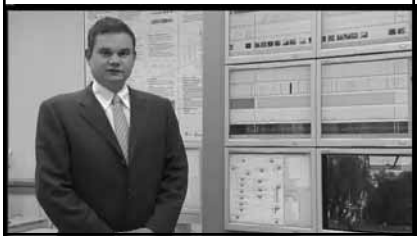


Error treatment in Cassandra Framework

- Performed by the Health Keeper
- Crash scenarios
  - Crash of a Service Unit
  - Crash of a component
  - Node disappearance
  - Any of the above combined with a network failure
- Treatment
  - Health Keeper rolls-back the system to a use case (restores connections graph)
  - Health keeper looks for new Service Units in Repositories (redundancy)
  - Resource-load is considered when choosing between same SUs in different Repositories
  - Treatment is preformed transparently for applications
  - Selection of the most appropriate treatment strategy from a pre-defined set.

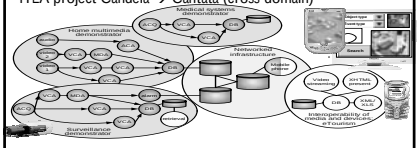


Content Analysis Service Units



Real-Time and Distributed Content Analysis System

ITEA project Candela → Cantata (cross-domain)



Basic: Generic domain-unaware →
Split: Domain-specific (narrow-scope solutions)

Conclusions

- Real-Time and Distributed Content Analysis System
  - Modularized approach → Upgradeable, Scalable, Flexible, Extendable
  - Enables efficient feature and application development
  - Standardized interfaces and protocols
    - → enables seamless collaboration of heterogeneous expert teams
    - → distributed content analysis in heterogeneous CE networks

→ One of the few ways to easily realize a

anticipative, personalized, adaptive,

ambient, self-organizing, and distributed system

Publications for further information about Part I

- F de Lange, J. Nesvadba, 'Applying PC network technology to assess new multimedia content analysis applications for future consumer electronics storage devices', 4th Int. Conf. On Intelligent Multimedia Computing and Networking (IMMCN), Salt Lake City, USA, July 21-26, 2005.
- J. Nesvadba, et al., 'Real-Time and Distributed AV Content Analysis System for Consumer Electronics Networks', Proc. Int. Conf. for Multimedia and Expo (ICME 2005), pp 1549-1552, Amsterdam, The Netherlands, June 6-8, 2005.
- W. Fontijn, J. Nesvadba, A. Sinitsyn, 'Integrating Media Management towards Ambient Intelligence' (doc), Journal Lecture Notes in Computer Science 'Title: Adaptive Multimedia Retrieval: User, Context, and Feedback', Springer-Verlag, ISSN: 0302-9743, ISBN: 3-540-32174-8, Vol. 3877 / 2006, pp. 102 - 111, 2006E.
- Jaspers, R. Wijnhoven, R. Albers, J. Nesvadba, A. Sinitsyn, J. Lukkien, X. Desurmont, P. Pietarila, R. Truyen, J. Palo, 'CANDELA – Storage, Analysis and Retrieval of Video Content in Distributed Systems', Int. Workshop on Adaptive Multimedia Retrieval (AMR 2005), Glasgow, UK, July 28-29, 2005, Published book: "Adaptive Multimedia Retrieval: User, Context and Feedback", pp 116 - 131, Springer.
- J. Nesvadba, F. d. Lange, A. Sinitsyn, J. Lukkien, A. Korostelev, 'Distributed and Adaptive Multimedia Content Analysis Prototyping Framework for Consumer Electronics', IEEE International Conference on Consumer Electronics (ICCE 2005), Las Vegas, USA, January 7 - 11, 2006.
- A. Korostelev, J. Lukkien, J. Nesvadba, 'Error Detection in Service-Oriented Distributed Systems', IEEE Int. Conf. on Dependable Systems and Networks (DSN 2006), Philadelphia, USA, June 25 - 28, 2006.
- F de Lange, J. Nesvadba, 'Early Evaluation of Future Consumer AV Content Analysis Applications with PC networks', Journal Multimedia Tools and Applications (Special Issue of MTAP).

SSMS 2006 Lecture: Semantic Multimedia Analysis

Part II: Multimedia Content Analysis Service Units in Distributed Content Analysis System

Jan Nesvadba, Philips Research Eindhoven, The Netherlands
jan.nesvadba@philips.com

Agenda

- Semantic Multimedia Analysis – Film Grammar
- AV Content Analysis – Service Units
- Conclusions

SSMS 2006 Lecture: Semantic Multimedia Analysis

Part IIa: Semantic Multimedia Analysis – Film Grammar

Jan Nesvadba, Philips Research Eindhoven, The Netherlands
jan.nesvadba@philips.com

Semantic Multimedia Analysis

Human's audio and video analysis system extracts relevant information from incoming audio and video signal
→ matching with memorized data
→ interpretations and conclusions based on experiences from the past
➔ non-uniform interpretation of signals by various individuals
(➔ personalization) , BUT ..

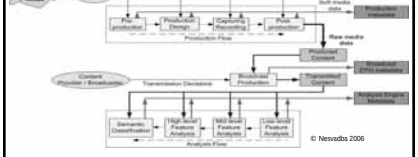
Semantic Multimedia Analysis – Film Grammar

➔Domain specific rules applicable:

Human language → grammatical rules → words (dictionaries) → sentences → chapters → books → series
Standardized rules enable transparent communication

Audiovisual content production → film grammar rules → shot sequences → scenes (chapters) → content items → series
➔ Standardized rules enable production, but also analysis

Semantic Multimedia Analysis –Film Grammar



Production Metadata – Input: Storyboard & Script

Production Metadata – Cinematographic rules

Dialogue A-B-A-B with Eye Line

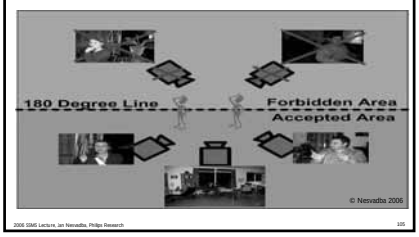
Production Metadata – Shot Classification

Production Metadata – Camera Position

Production Metadata – Camera Motion

Production Metadata – Camera Motion

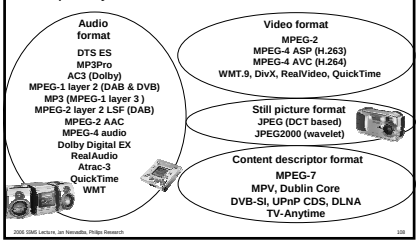
Production Metadata – 180º Rule



Film Grammar – Cross-cuttings

Film Grammar – Shot-Reverse-Shot

Complexity: Audio Video standards



Production Metadata Conclusions

- Semantic powerful production metadata
  → currently lost due to unresolved business model
- Production follows rules: Film Grammar
  → prior knowledge of rules can be applied in case of domain specific solutions
- Metadata standardization:
  objective low-level / mid-level standardized to some extend (e.g. MPEG-7) and ongoing,
  semantic (also subjective) metadata standardization difficult, but ongoing

SSMS 2006 Lecture: Semantic Multimedia Analysis

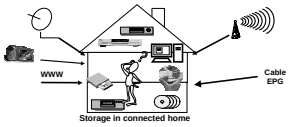
Part IIb: AV Content Segmentation – Service Units

Jan Nesvadba, Philips Research Eindhoven, The Netherlands
jan.nesvadba@philips.com

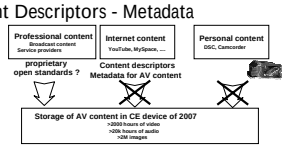
Why do we need Multimedia Analysis?

Due to
- increased storage capacity of CE devices,
- greater digital compression of A/V content and
- increasing number of A/V sources into the home

→consumers do need guidance through the stored A/V jungle using content descriptors – metadata



Content Descriptors - Metadata



Question: How to get information about the content?
Broadcaster? Services?

Alternative: By local real-time multimedia content analysis
Alternative or in combination as enhancement

The Multimedia Analysis Pyramid

Service Units for Semantic Multimedia Analysis

• **Audio:**
Low-level audio features (power, zero-crossing, MFCC, ..)
Mid-level audio features (audio classification, speaker gender, ..)
High-level audio features (audio emotion, ASR, ..)

• **Video:**
Low-level video features (color, motion, texture, ..)
Mid-level features (shot boundaries, face localization, ..)
High-level features (scene boundaries, genre classification, ..)

• **Text:**
Close Caption, EPG data, Script data

➔**Multimedia:**
combination of multiple media (audio, speech, video, text)
➔ multimedia content analysis

Audio content analysis

- **Automatic audio classification**
(speech, music, noise, crowd, silence, highlights, ..)
- **Automatic music classification**
(music genres: pop, classical, jazz; tempo, beat, rhythm, ..)
- **Automatic speech analysis**
(speaker ID, speaker change, gender, speech-to-text,..)

Automatic audio classification

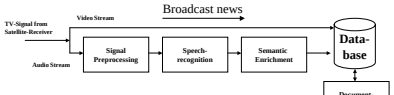
- Audio classes:
1. Cheering and applause
2. Environmental noise
3. Speech (female and male)
4. Classical music
5. Popular music
- Feature vectors:
  - Low-level features
  - MFCC features
  - Psycho-acoustic features

Based on single 743-ms frames

Automatic music classification

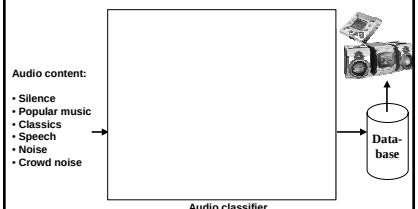
- Popular music classes:
1. Rap
2. Rock
3. R&B
4. Electronica
5. Folk
6. Pop
7. Jazz
- Features:
1. **Rhythm**
2. **Beat**
3. **Tempo**
4. **Swing**

Automatic speech analysis



- Topic detection and speaker gender
- Content-focused transcription
- ASR: high-level semantic data
(e.g. named entities or locations)

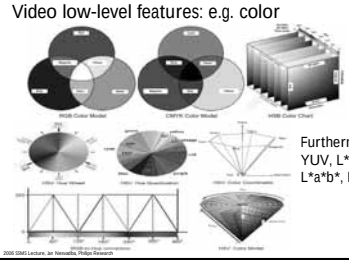
Demo General Sound Classifier



Video content analysis

- **Automatic low-level features** (color, motion, texture, ..)
- **Automatic mid-level features** (shot boundaries, face localization and pose estimation, text localization and extraction, ..)
- **Automatic high-level features and applications** (genre classification, face and emotion recognition, automatic summary, ..)

Video low-level features: e.g. color



Furthermore: YUV, L*u*v*, L*a*b*, Lhs, ..

Video mid-level features: shot boundary detector

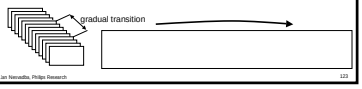
Video shot

- Definition:
  - a series of interrelated consecutive frames taken contiguously by a single camera and representing a continuous action in time and space
- Meaning for content analysis:
  - Partitioning a video sequence into shots is an important step toward video content analysis
  - VS are considered to be the primitives for higher level content analysis, indexing and classification
  - Reliable shot boundary detector is required

Shot boundary detector - Shot transitions

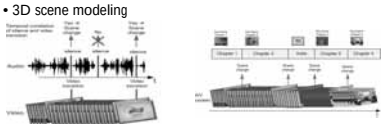
- Abrupt transitions
  (hard cuts / transitions)

- Gradual (progressive) transitions
  - E.g. video editing special effects: dissolve, fade-in, fade out, wipe



Shot boundary detector - Applications

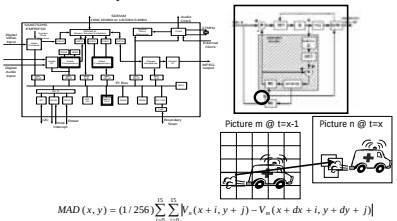
- Compression trans/encoding
- Discrimination interlaced / progressive
- Identify and trace objects
- 3D scene modeling



Shot boundary detector - Feature extraction and metrics

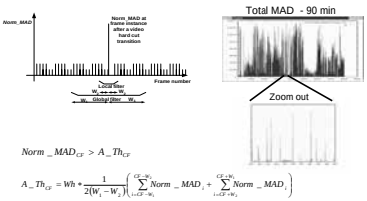
- Different approaches
  - pixel, histogram, edges …
  - averaging filters, motion compensation, various color spaces …
    - most of visual content differences between consecutive frames in the same shot are caused by camera motion and lighting changes
- Segmentation of frames
  - insensitive to lighting changes
  - motion compensation can be used against camera motion sensitivity

Shot boundary detector - MB-correlation-factor



$$MAD(x,y) = (1/256)\sum_{i=0}^{15}\sum_{j=0}^{15}\left|V_n(x+i,y+j) - V_m(x+dx+i,y+dy+j)\right|$$

Shot boundary detector - MB-correlation-factor



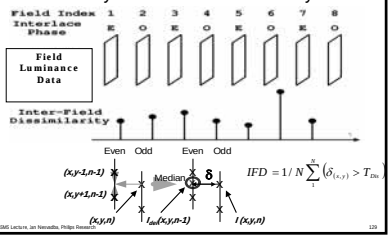
$$Norm\_MAD_{CF} > A\_Th_{CF}$$

$$A\_Th_{CF} = Wh * \frac{1}{2(W_1 - W_2)}\left(\sum_{i=CF-W_1}^{CF-W_2} Norm\_MAD_i + \sum_{i=CF+W_2}^{CF+W_1} Norm\_MAD_i\right)$$

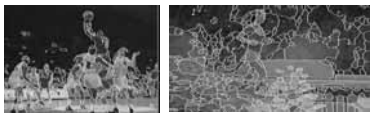
Shot boundary detector - Pixel-similarity-based

- Evaluates a history of differences: "Difference significantly greater than recent history"
- Differences calculated between consecutive fields
- De-interlace method: Vertical Temporal Median

Shot boundary detector - Pixel-similarity-based

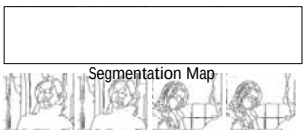


Shot boundary detector - Color segmentation



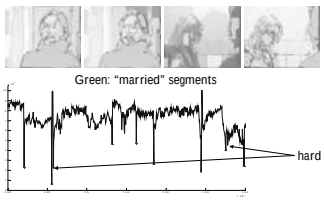
- "Division of image in regions of constant RGB"
- Can be considered as "content adaptive macroblocks"
- NOT intended to be object segmentation!

Shot boundary detector - Segmentation around SB



Properties of segmentation:
- –Segments typically part of same object
- –Consistent over time: segmentations of subsequent frames in same shot look similar

Shot boundary detector - Segmentation around SB



Shot boundary detector - Measures for benchmark

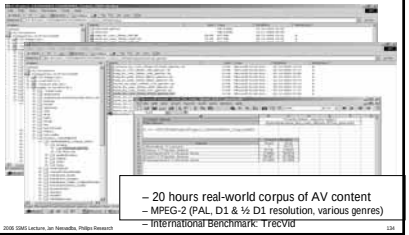
- Recall: Do we find all the shot boundaries?

$$R = \frac{N_{correct}}{N_{correct} + N_{missed}} \cdot 100\%$$

- Precision: Don't we give false alarms?

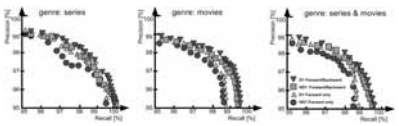
$$P = \frac{N_{correct}}{N_{correct} + N_{false}} \cdot 100\%$$

Shot boundary detector – Benchmark Corpus



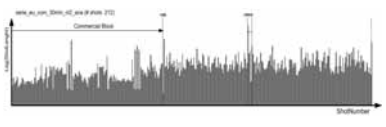
– 20 hours real-world corpus of AV content
– MPEG-2 (PAL, D1 & ½ D1 resolution, various genres)
– International Benchmark: Trecvid

Shot boundary detector - Cut Detection Results



Genre	Total SBs (CDs)	Correct CDs	False CDs	Missed CDs	Recall [%]	Precision [%]
Movies	4942	4855	134	87	98,2	97,3
Series	1591	1551	24	40	97,5	98,5
M&S	6533	6406	158	127	98,1	97,6

Shot boundary detector – Segmentation into shots

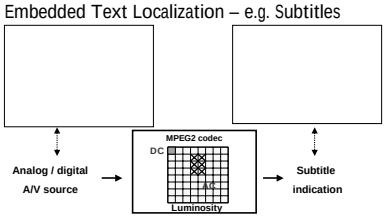


Shot boundary detector - MPEG-7 representation of cuts

```
<MultimediaContent xsi:type="AnalyticEditedVideoType">
  <AnalyticEditedVideo xsi:type="EditedVideoType">
    <AnalyticEditingTemporalDecomposition>
      <GlobalTransition>
        <MediaTime>
          <MediaRelIncrTimePoint mediaTimeUnit="PT1N90000F">
            10
          </MediaRelIncrTimePoint>
        </MediaTime>
        <EvolutionType href="urn:mpeg:mpeg7:cs:EvolutionTypeCS:2001">
          <Name xml:lang="en">Cut</Name>
        </EvolutionType>
      </GlobalTransition>
    </AnalyticEditingTemporalDecomposition>
  </AnalyticEditedVideo>
</MultimediaContent>
```


Shot boundary detector - MPEG-7 representation into DB

Video mid-level features: Embedded Text Localization – e.g. Subtitles



Video mid-level features: Scrolling Text Localization – e.g. Credits

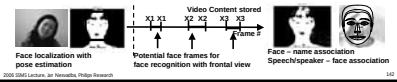
Input: Embedded Text Localization →

Video mid-level features: Embedded Text Extraction

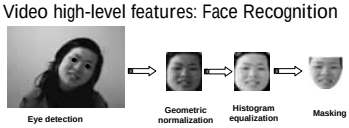
Input: Embedded Text Localization → OCR

Video mid-level features: Omni-directional Face Detector, Pose Estimator

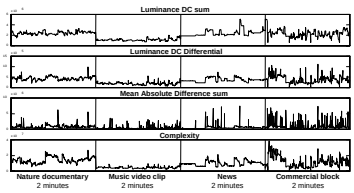
Face localization (Omni-directional Face Detector → 15 poses, Output: XML/MPEG-7) and identification



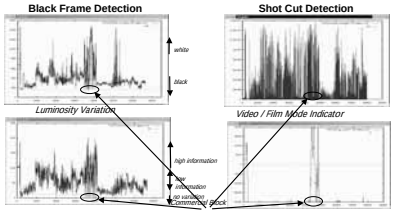
Video high-level features: Face Recognition



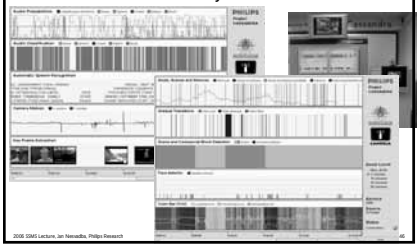
Video high-level features: Genre classification



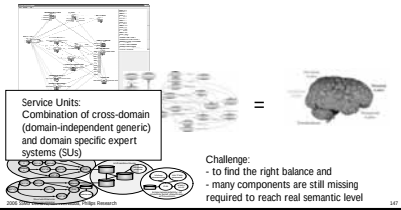
Genre classification: Adds



Semantic Multimedia Analysis



Semantic Multimedia Analysis – Merge of Multiple Modalities



Service Units:
Combination of cross-domain (domain-independent generic) and domain specific expert systems (SUs)

=

Challenge:
- to find the right balance and
- many components are still missing required to reach real semantic level

Demonstrator: Analysis Results based on aggregated data from various SUs

Publications for further information about Part II

- P. Fonseca, J. Nesvadba, 'Face Tracking in the Compressed Domain', Journal: EURASIP Journal on Applied Signal Processing, Information Mining from Multimedia Databases, Volume 2006, Article ID 59451, Pages 1–11, DOI: 10.1155/ASP/2006/59451.
- A. Hanjalic, J. Nesvadba, J. Benois-Pineau, 'Moving away from narrow-scope Solutions in Multimedia Content Analysis', European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies, London, UK, November 30 - December 1, 2005.
- J. Nesvadba, A. Hanjalic, P. Fonseca, B. Kroon, H. Celik, E. Hendriks, 'Towards a real-time and distributed system for face detection, pose estimation and face-related features', Proc. Int. Con. on Methods and Techniques in Behavioral Research, Wageningen, The Netherlands, 2005.
- J. Nesvadba, F. Ernst, J. Perhavc, J. Benois-Pineau, L. Primaux, 'Comparison of Shot Boundary Detectors', Int. Conf. for Multimedia and Expo (ICME 2005), pp 788-791, Amsterdam, The Netherlands, June 6-8, 2005.
- J. Nesvadba, N. Louis, J. Benois-Pineau, M. Desainte-Catherine, M. Klein Middelink, 'Low-level cross-media statistical approach for semantic partitioning of audio-visual content in a home multimedia environment', Proc. IEEE Int. Workshop on Systems, Signals and Image Processing (IWSSIP'04), pp. 235-238, Poznan, Poland, September 13-15, 2004.
- J. Nesvadba, P. M. Fonseca, R. Kleihorst, H. Broers, J. Fan, 'Face Related Features in Consumer Electronic (CE) device environments', Proc. IEEE Int'l Conf. on Systems, Man, and Cybernetics (IEEE SMC 2004), Special Session on Automatic Facial Expression Recognition, vol. 1, pp 641-646, ISBN 0-7803-8567-5, The Hague, Netherlands, 2004.
- P. Miguel Fonseca, J. Nesvadba, 'Face Detection in the Compressed Domain', IEEE Proc. Int. Conf. on Image Processing (IEEE ICIP 2004), vol. 3, pp.2015-2018, Singapore, October 2004, ISBN 0-7803-8554-3, Published by the IEEE, (IEEE Catalog Number: 04CH37580), 2004.
- L. Agnihotri, N. Dimitrova, T. McGee, S. Jeannin, D. Schaffer, J. Nesvadba, 'Evolvable Visual Commercial Detector', Proc. 2003 IEEE Conf. on Computer Vision and Pattern Recognition, Madison, Wisconsin, Vol 2, pp. II-79 – II-84, ISSN: 1063-6919, 18-20 June, 2003.

PHILIPS