Zeitera

AUDIO-VIDEO CONTENT FINGERPRINTING FOR SMART TV AND SYNCHRONOUS MOBILE CONTENT IDENTIFICATION

MIHAILO STOJANCIC

JUNE 28, 2011

Contents

- Zeitera overview
- Emerging new television viewing model
- Content fingerprinting based ACR
- Audio-Video content fingerprinting technologies
- System requirements/robustness
- SmartTV applications
- Synchronous Mobile applications
- Summary and conclusions

Scope of Discussions

- Fingerprinting technology and current and future use cases for the automatic content recognition (ACR) and synchronous mobile devices
- Opportunity for novel ways of audience engagement
- Applications include:
  - Interactive advertising
  - Multi-screen viewing environment
  - Synchronous mobile applications
  - Immersive social networking apps, and more…

Zeitera Overview

Founding	Founded in April 2006
Business Overview	Audio and Video content identification company enabling the discovery, identification, management and monetization of video content
Team	Experienced management team with extensive industry and start-up experience. World-class technology team with deep backgrounds in video, audio, and search system development
Offices	Mountain View, California
Funding	Funded by private technology investors

Zeitera Overview (continued)

- Zeitera is recognized as a leading provider of digital audio and video fingerprinting technology for ACR applications
- Zeitera's patented audio-video content identification and search system enables applications for smartTVs, smart phones and tablets
- 18 patent filings
- Major use cases are in:
  - ACR for interactive, targeted advertizing
  - Synchronous mobile-based TV content analysis with enhanced second screen user experience
  - Video-Audio data base analysis and management.

Emerging new television model

- Over the past a few years the definition of television has been continuously evolving
- Today television has become a combined home theater, video-audio store, Internet portal, gaming platform, shopping mall…
- SmartTVs now offer an interactive medium for advertisers, allowing for targeted, personalized ads, and accurate audience measurement
- SmartTV and smart phone applications are enabling new forms of engaging audience in highly interactive way providing freedom in innovative advertising that was not possible before.

Hybrid Broadcast /Internet vs. linear broadcast television model

- SmartTVs and smart portable devices are becoming ever more powerful and pervasive
- New model of television has emerged as a hybrid broadcast-Internet based device with interactive applications
- Instead of viewing and absorbing television content viewers are engaged and participate actively in the television experience
  - Searching additional information about people and places discussed on TV
  - Checking out additional product information and special offers
  - Voting in real-time
  - Participating in polls and surveys, etc.

Example: Web Connected TV - SmartTV

Web-connected apps for your TV

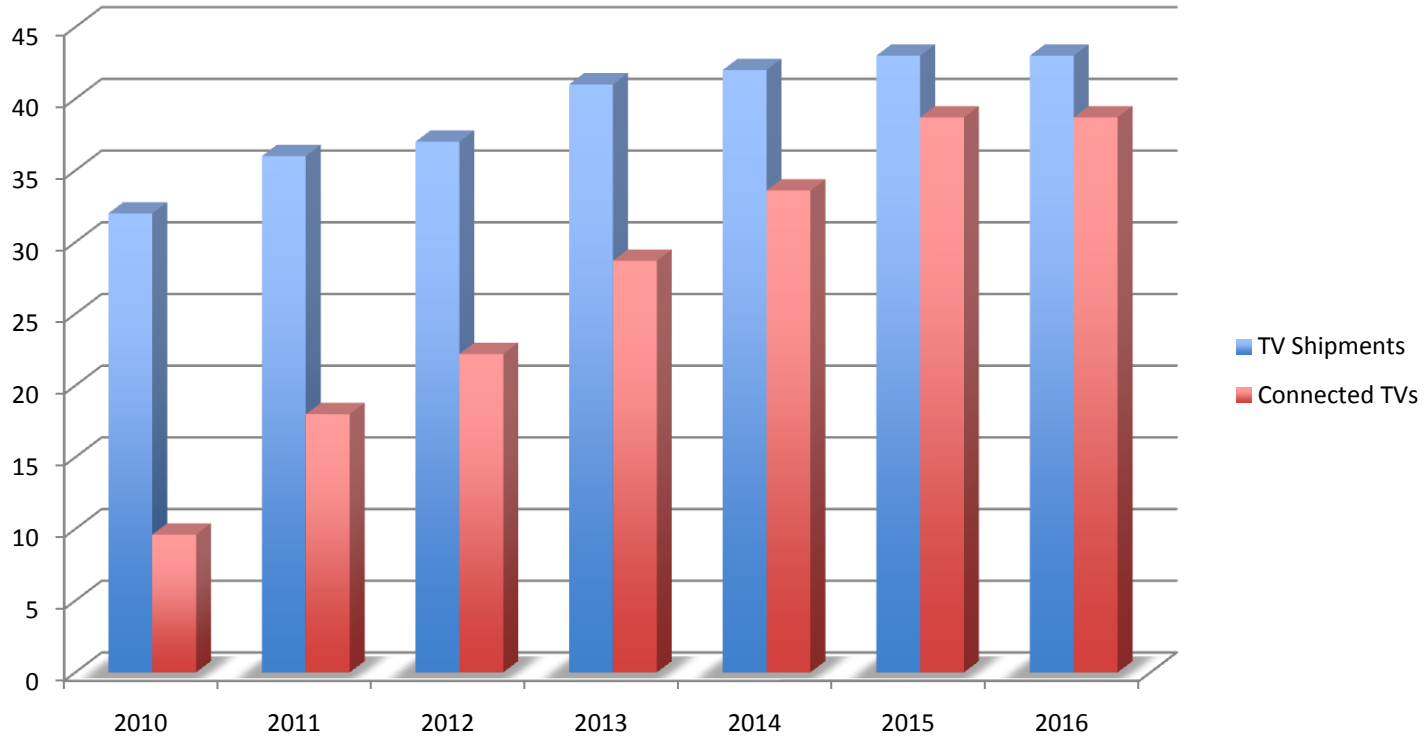
Samsung Apps

Samsung Apps allows users to get the best of the web right from their TV screen.

Users can choose from a gallery of apps built for their TV that let you stream video, play games, view pictures and more...

Example: Yahoo Connected TV

SmartTV Market is Growing



Michael Collette 2011

Automatic Content Recognition (ACR) Technology

- ACR is an emerging technology that plays crucial role in the development of interactive features of smartTVs, smart phones and tablets
- It allows for automatic recognition of programs and commercials on multiple screens
- ACR is a key, strategic technology in the current and future television development

ACR Technology (cont.)

- An ACR interactive application may be embedded in a TV device, smart phone or tablet, allowing for real time identification of played content
- Embedded ACR application allows recognition of specific content a viewer is watching at any given time thus providing a fine granularity information on viewer's behavior and viewing habits
- This information allows advertiser to directly connect with their targeted audience

ACR Technology (cont.)

- ACR is essential to making interactivity attractive, more engaging, and a large part of the viewing experience
- Although linear TV will continue to dominate the television viewing world in the near future, smartTVs and over the top systems will establish a long term dominance
- In the future all devices used to watch and interact with video will have some variation of embedded content recognition technology

Audio-Video content identification

- Audio-Video content fingerprinting is at the core of any ACR system
- When deciding on fingerprinting algorithms many tradeoffs need to be considered, including:
  - accuracy, robustness
  - signature size
  - signature rate
  - computational requirements
  - overall system cost

Audio-Video content identification (cont.)

- Recently many different algorithmic schemes and systems have been introduced targeting different applications
- General application area is wide and includes:
  - audio-video identification in consumer electronics
  - copyright protection (antipiracy)
  - content management, database de-duplication
  - A-V sync, etc.

Two major content identification approaches

- Fingerprinting
  - Technique that doesn't require any modification of the content
  - Requires a reference data base, and search of the reference database, either in the cloud or on a local device
  - Can be very fast, will indentify content based on small query clips
  - Can be deployed independently in smartTVs, smart phone/tablet without a need to engage third parties
  - No additional hardware requirements for deployment in the broadcast workflow
  - Allows advertisers and application developers a freedom to approach and engage viewers in a unique, creative way

Two major content identification approaches (cont.)

- Watermarking
  - Inserts signal into either the video or audio content
  - Requires pre-processing of content at some point in the broadcast workflow
  - Doesn't requires search of a reference database
  - Allows unique ID of each piece of content
  - Can be set below visible or audible threshold, but at the cost of amount of bits inserted and time of recovery

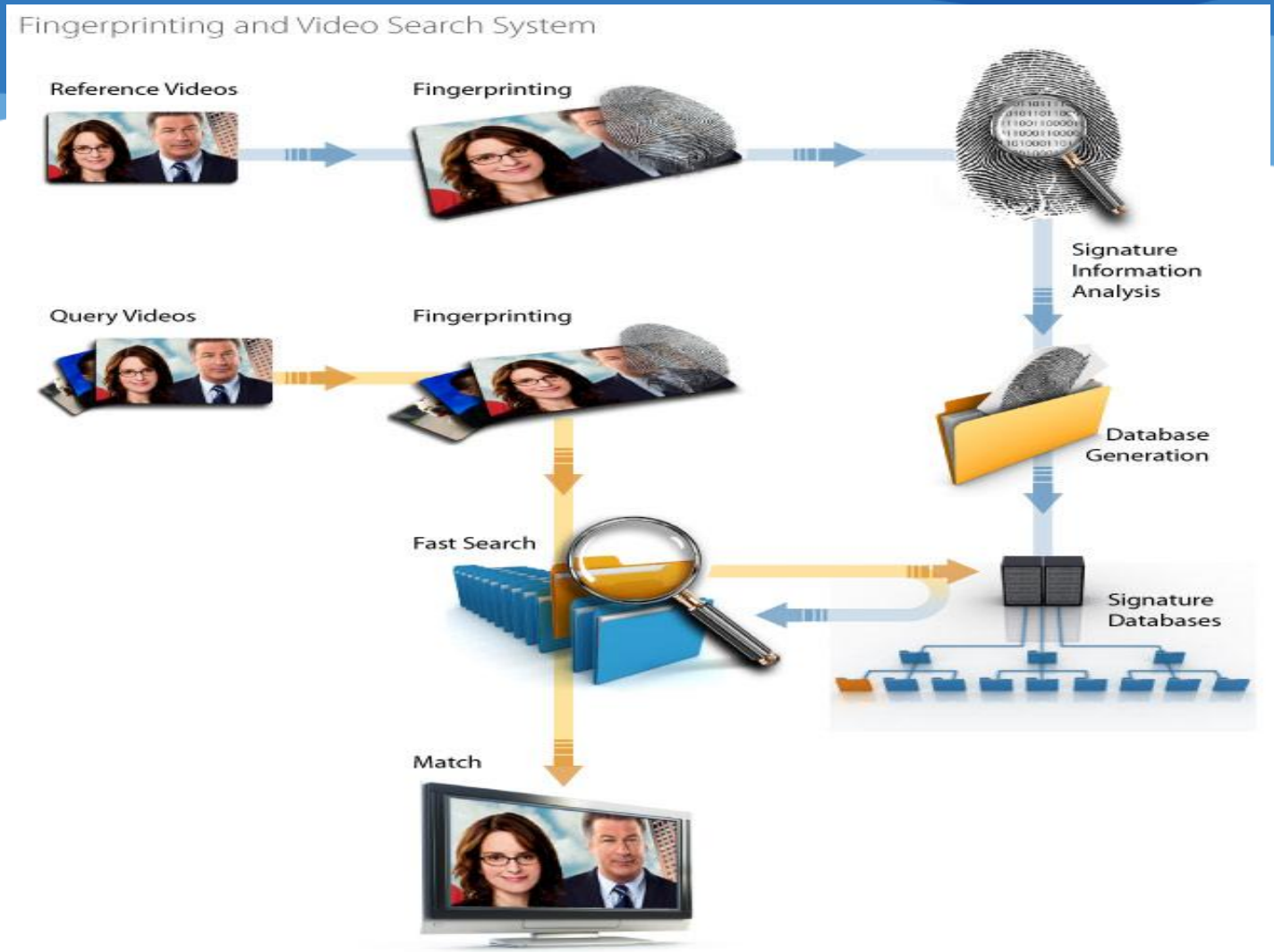
Audio-Video content fingerprints

- Content fingerprinting is a technology allowing identification and encoding of content the way human perceptual system operates
- It exploits fine-grained, multiple point of view measurements of the content
- These observations are translated and encoded as small fingerprints representing an audio signal or a video signal
- During the content identification/search process, fingerprints are matched to a reference database of fingerprints
- This is done in real-time, identifying the content as it is being played.

Characterictics of Fingerprinting

- Robust to distortions
  - Can be made robust to many distortions: Rescaling, low bit rate encoders, aspect ratio changes, rotations, camcorder, pixelation, etc.
- Flexibility
  - Content databases can be created at any monitoring point – therefore overall identification of content can be very broad.
  - Deals well with unmanaged content like commercials and promotions
  - Identifies any version of a particular piece of content, not just a specific one
- Accuracy and performance
  - Can reliably recognize content based on very small query clips and run much faster than real time
- No change to content
  - Does not affect or alter content in any way.

Fingerprinting and Search



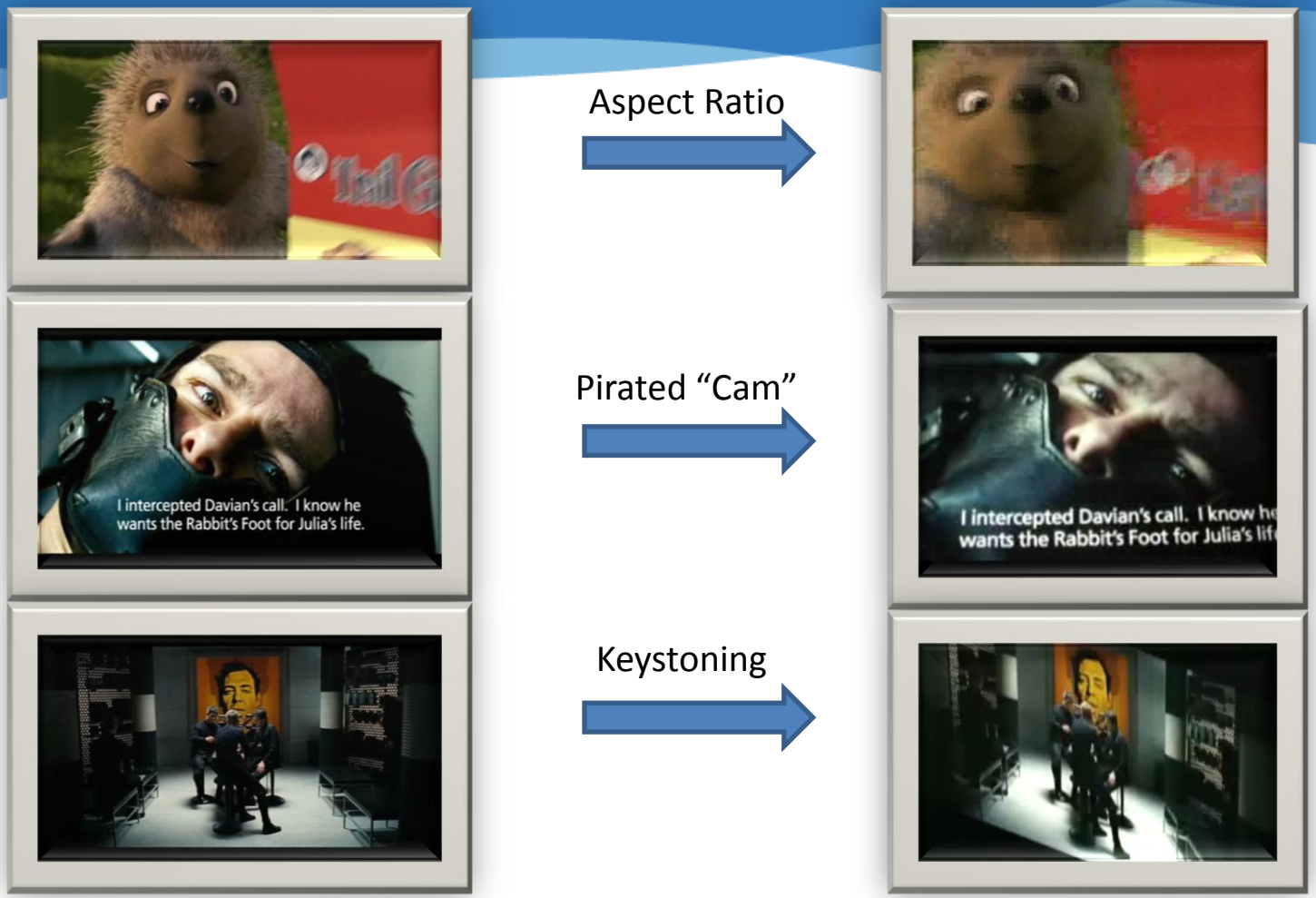
Video Fingerprinting Technologies

- Matching a video sequence at an observation point to a reference video requires reliable matching of two digital images (video frames)
- Many video frames in both original and query video sequences may exhibit statistical similarity
- Also many features within a single video frame may show statistical similarity in both optical and geometric sense
- A precise, discriminative, and robust video frame feature characterization is desirable

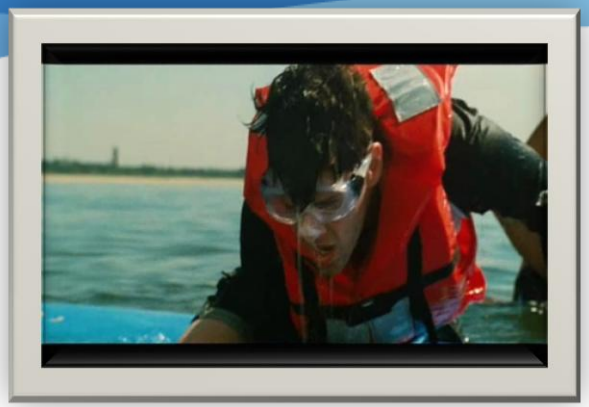
Zeitera Video Fingerprint Types

- Multiple types of video signatures based on spatial-temporal content information extraction
  - Localized detection and description of scale (affine) invariant interest points
  - Spatial derivative weighted pixel orientation (grid, histogram)
  - Spatial Intensity/Color distribution (grid, histogram)
  - Region based segmentation with contour description
  - Spatio-temporal optical flow vectors
- Short video signatures for DB clustering

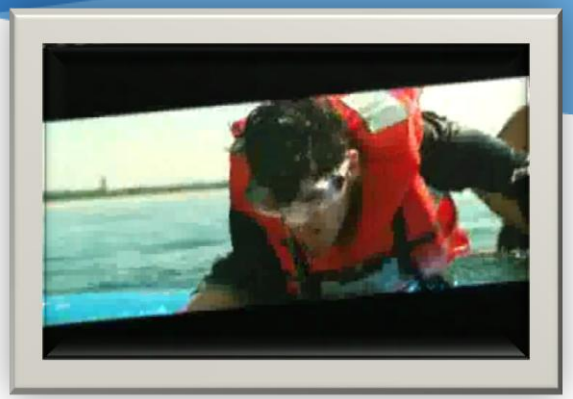
Example Content Distortions

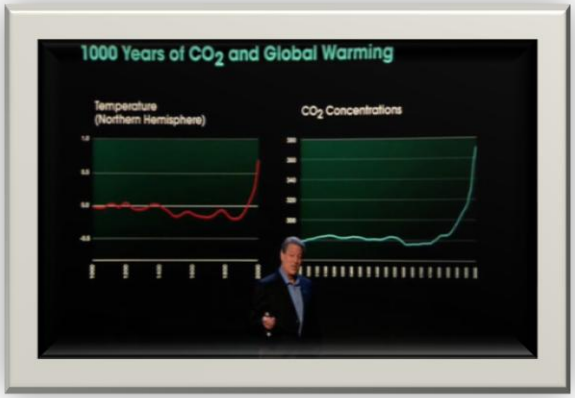


Example Content Distortions



Rotation

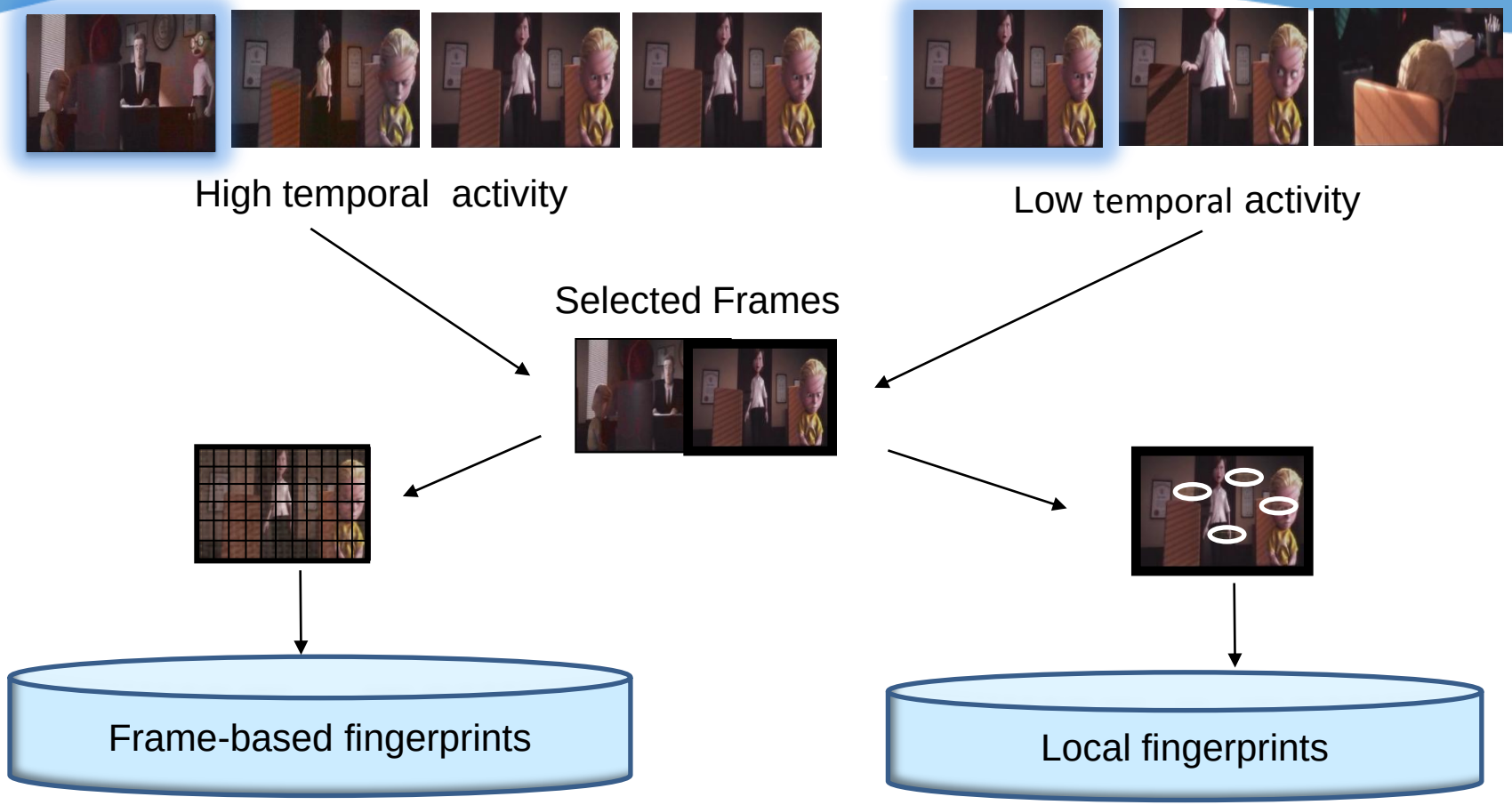




UGC/YouTube

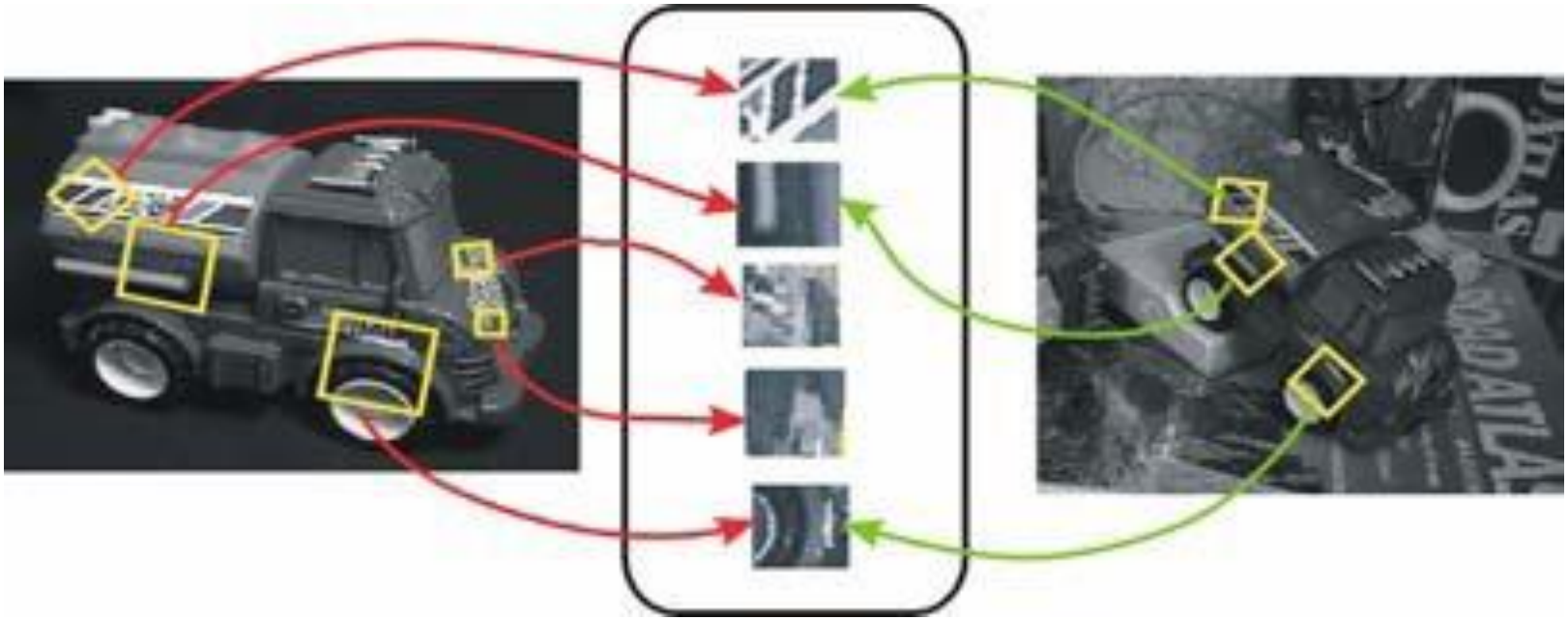
Other distortions: zoom, mash-ups, frame rate, short clips, color/contrast, brightness

Top level temporal structuring Frame selection



Scale Invariant Interest Point Detection

- Scale space representation of images
  - Local features invariant to translation, scale, rotation
  - Highly distinctive, robust against cropping, occlusion and clutter
  - Robust to changes in illumination/contrast



Interest Point Detector

$$g(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

$$L_{yy}(\mathbf{z}, \sigma) = \frac{\partial^2 g(\mathbf{z}, \sigma)}{\partial^2 y} * I(\mathbf{z})$$

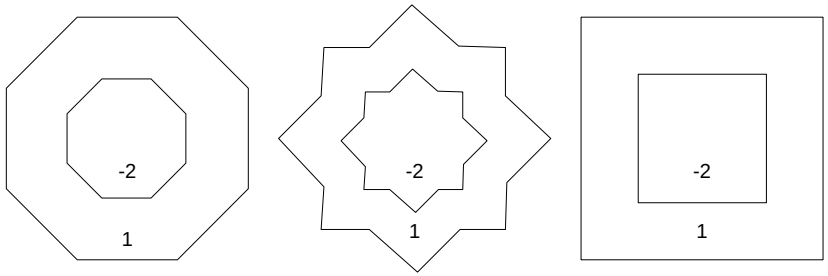
$$L_{xx}(\mathbf{z}, \sigma) = \frac{\partial^2 g(\mathbf{z}, \sigma)}{\partial^2 x} * I(\mathbf{z})$$

Laplacian-of-Gaussian (LoG) - Extrema in scale-space represent interest points

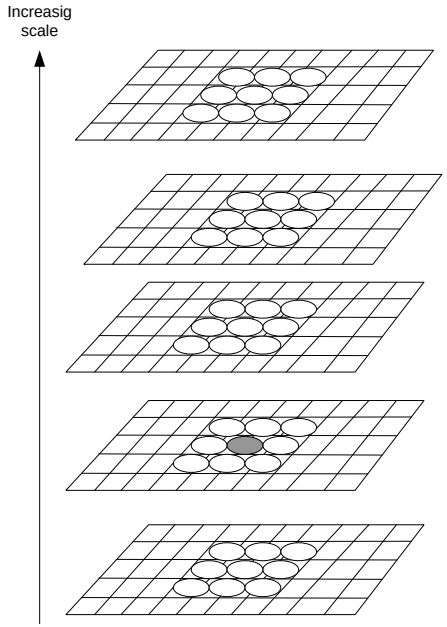
Fast computation with bi-level Gaussian second order partial derivative filters

Determine preliminary interest points (regions) to be refined in the second stage

$$max_{\sigma} |\sigma^2 (L_{xx}(\mathbf{z}, \sigma) + L_{yy}(\mathbf{z}, \sigma))|$$



Octagon Star Box

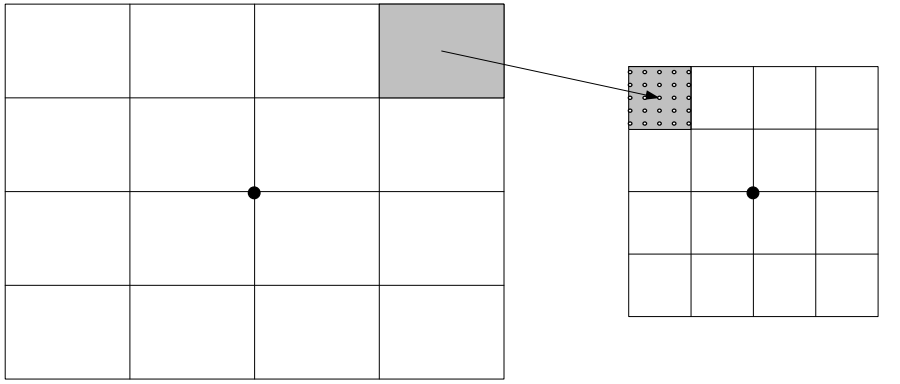


Laplacian of Gaussian response Images for 5 scales (s_0, s_1, s_2, s_3, s_4); Example of 45 pixel 3-dimensional scale-space neighborhood shown

Interest Region Formation, Descriptor Generation

Possible IR descriptor based on pixel intensity gradient vector

64-dimension Descriptor/Signature generated for each IR with 16 5x5 blocks

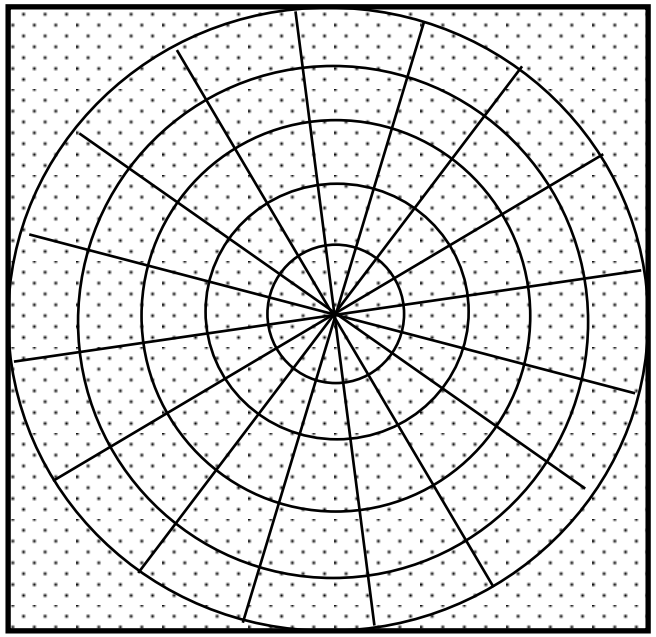


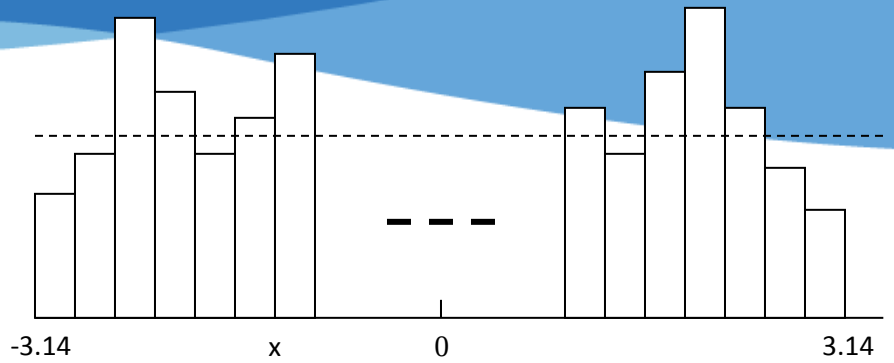
Ns_x x Ms_y rectangular box drawn around an interest point at the center; The box is divided into 16 sub-regions

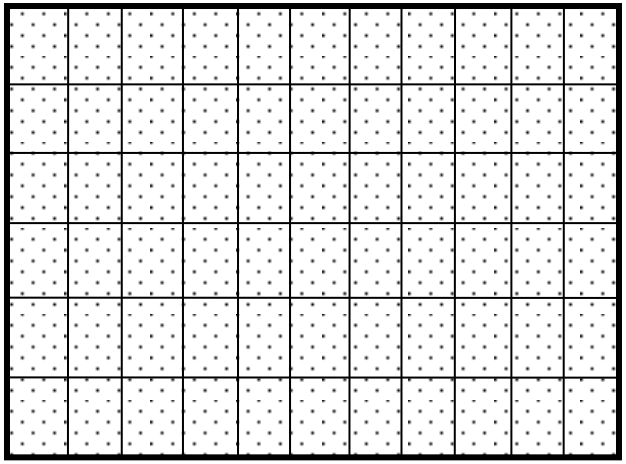
Re-sampled rectangular box with 5x5 re-sampled pixel sub-regions

Global Spatial Video Signature Generation

Intensity gradient and orientation (phase angle) computed for each pixel







Resultant $\Omega_k = \Sigma_n (G_p \theta_p) / \Sigma_n G_p$ values for each bin compared to a Ω value computed for the entire functional space to derive a multi-dimensional signature

Audio content fingerprinting technologies

- Many different types of audio signatures have been developed in the past several years
- They are robust and computationally efficient for audio identification in the presence of considerable distortion and/or noise
- In general, the problem of mapping high-dimensional audio input data into lower-dimensional feature vectors containing sufficient relevant information is at the core of audio fingerprinting and identification systems.

An example of audio fingerprinting method

Algorithmic base: Mel Frequency Cepstral Coefficients with coefficient quantization and signature generation

- The classic MFCC algorithm is well understood
- Used for speech analysis, also music
- TV audio mostly speech

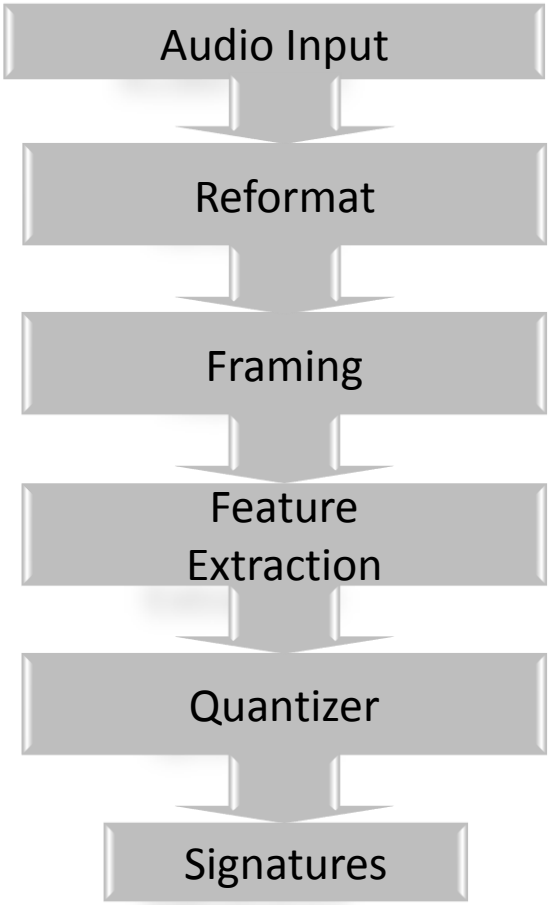
System Parameters

- Filterbank
- Quantizers
- Number of output parameters

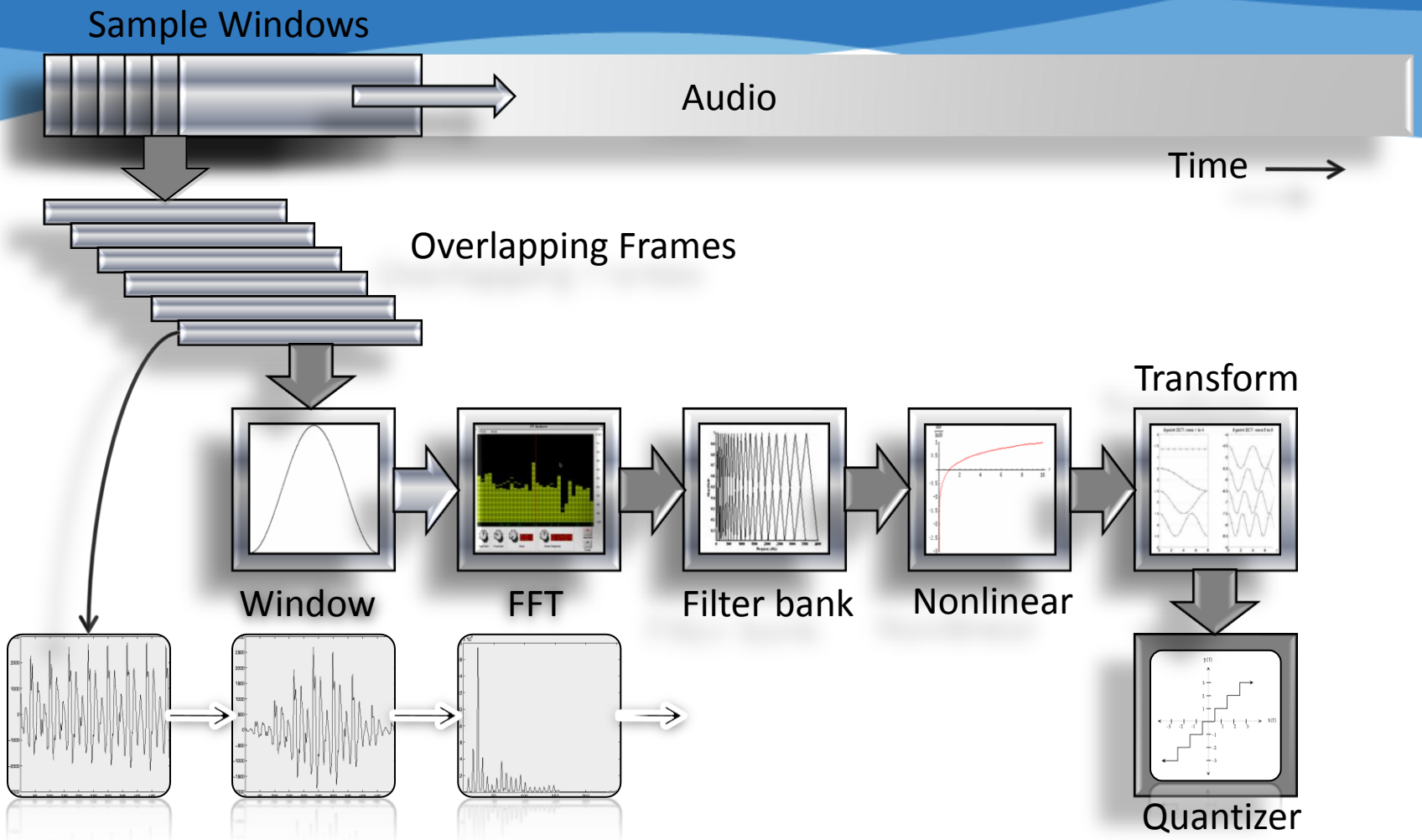
Benefits of parametric spectral representation

- Principal trends => few parameters
- Parameters independent of number of filter bands
- Automatic volume extraction
- Lower bit rate

General steps in audio fingerprinting



Example Audio Fingerprint Extraction

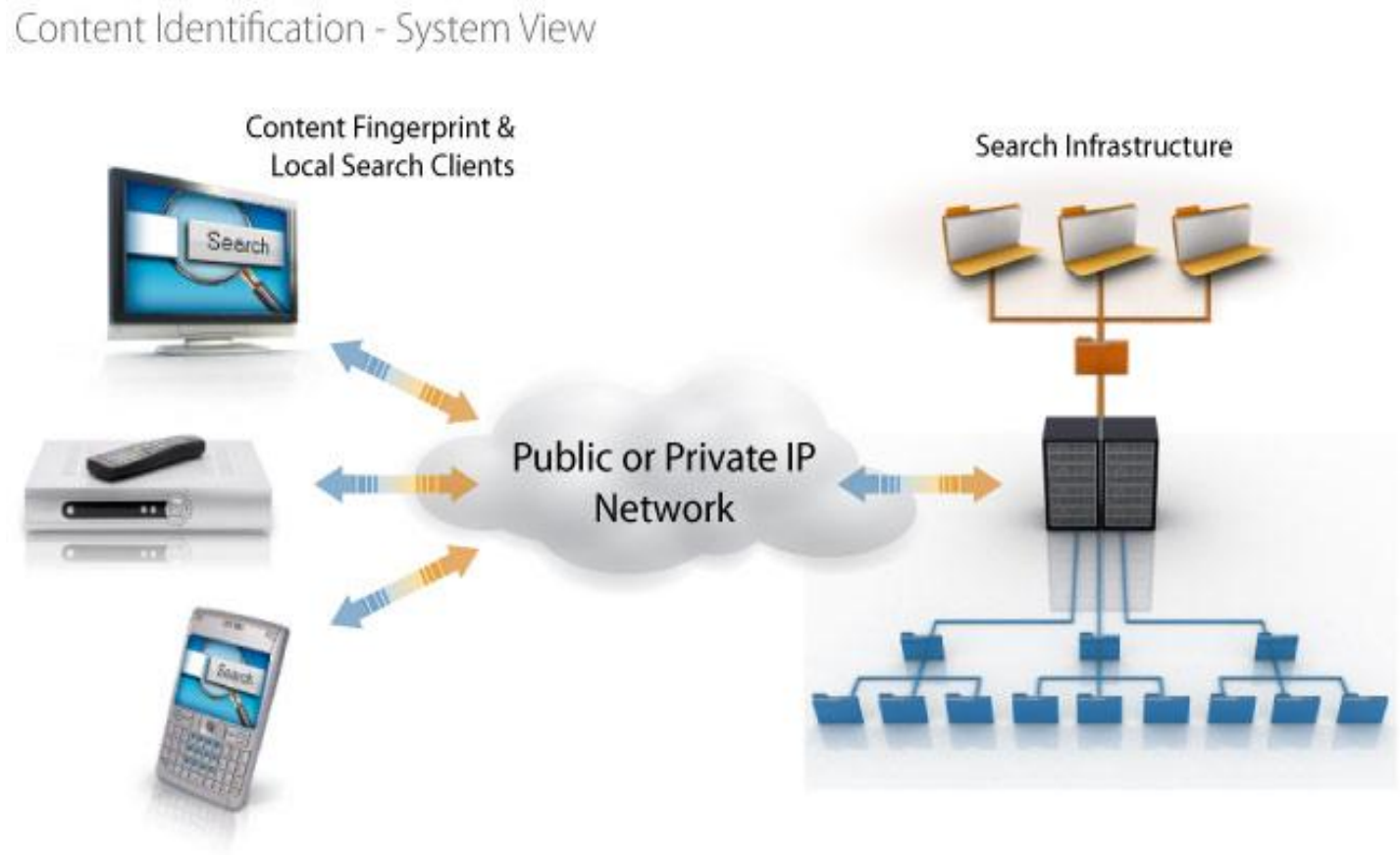


ACR – Fingerprinting Applications

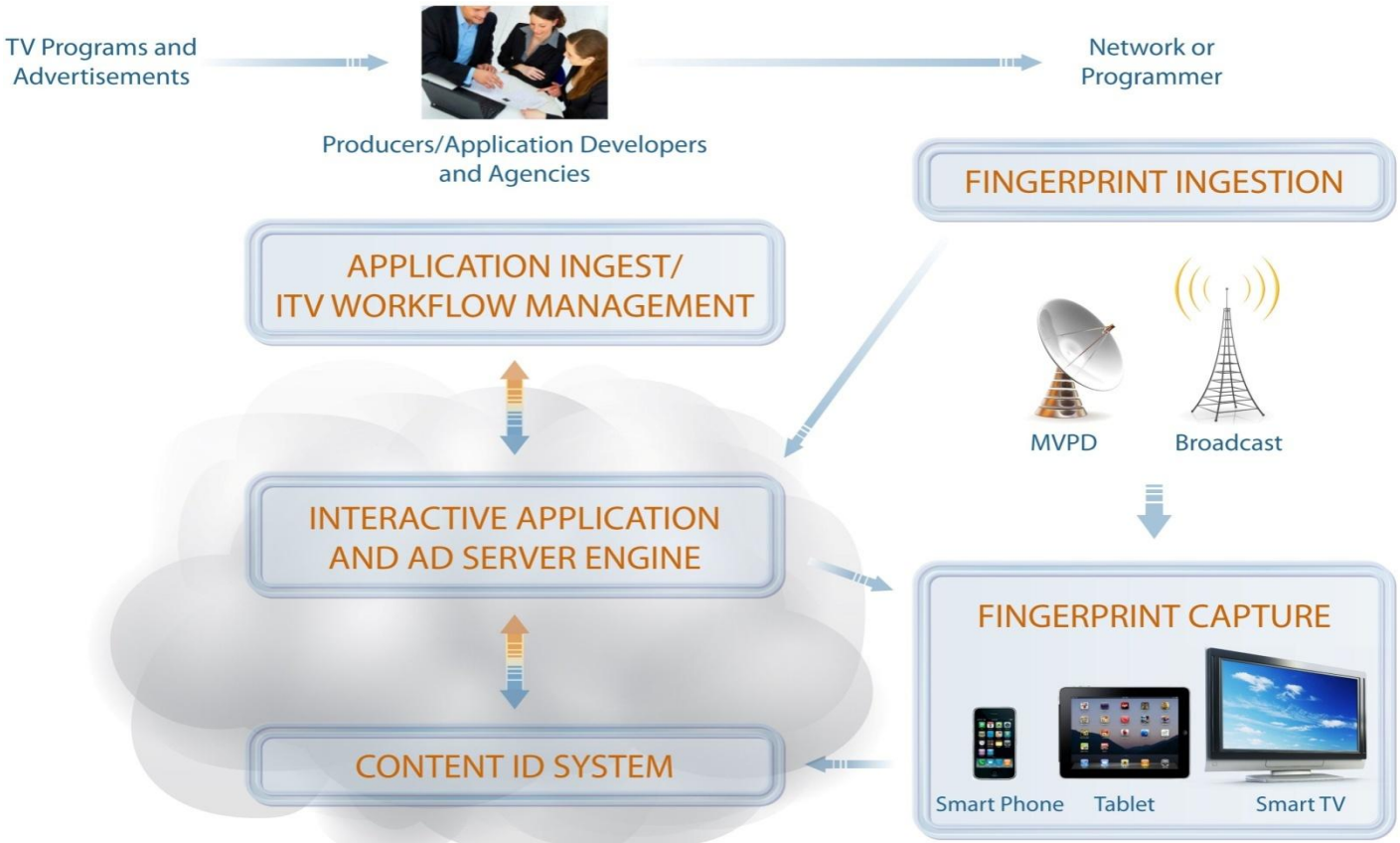
- SmartTV Applications
  - Allows SmartTV apps to interact with TV content (including commercials)
  - Interactive advertising applications
  - Targeted ads
  - Coupon capabilities - for local and national advertisers
  - Commercial monitoring/localization/replacement

- Synchronous Mobile Applications
  - Smart phone 2 screen interaction with video content
  - Immersive social networking applications
  - 2nd Screen Applications
  - Direct Check-in
  - Program guide correlation with rich meta-data

A typical ACR system



Cloud based Ad serving - SmartTV



SmartTV Applications

Interactive advertising

POST PRODUCTION

Create fingerprint and ad overlay

DATA TRANSMISSION

Push fingerprint and code for ad overlay to local database on device, such as set top box (STB)

AD BROADCAST

Ad captured by device (STB).

FINGERPRINT RECOGNITION

As broadcast is received, content is identified using local database on device.

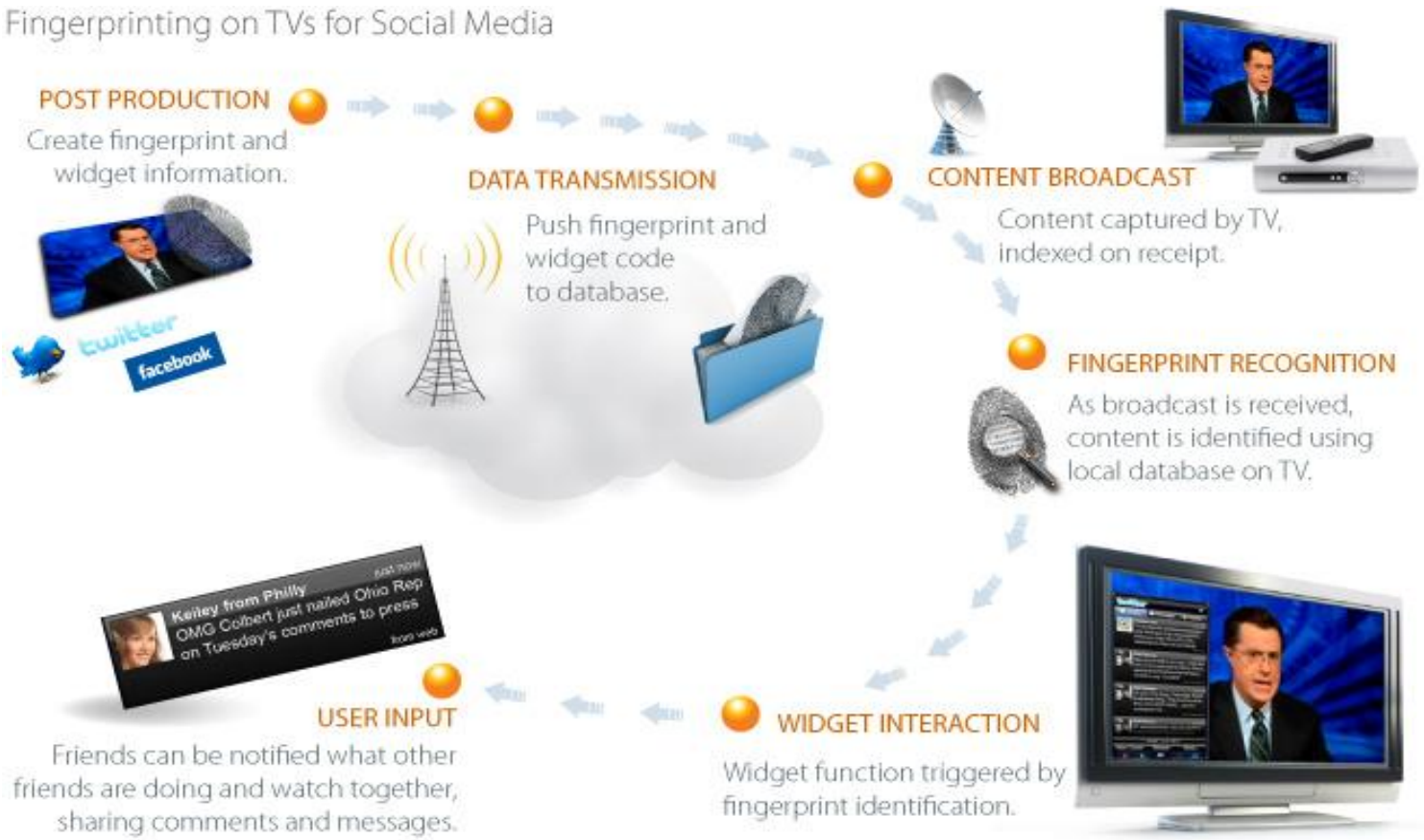
CALL TO ACTION

Ad overlay triggered by fingerprint identification.

USER INPUT

Call to action/ special promotion invites response from viewer

Social Media

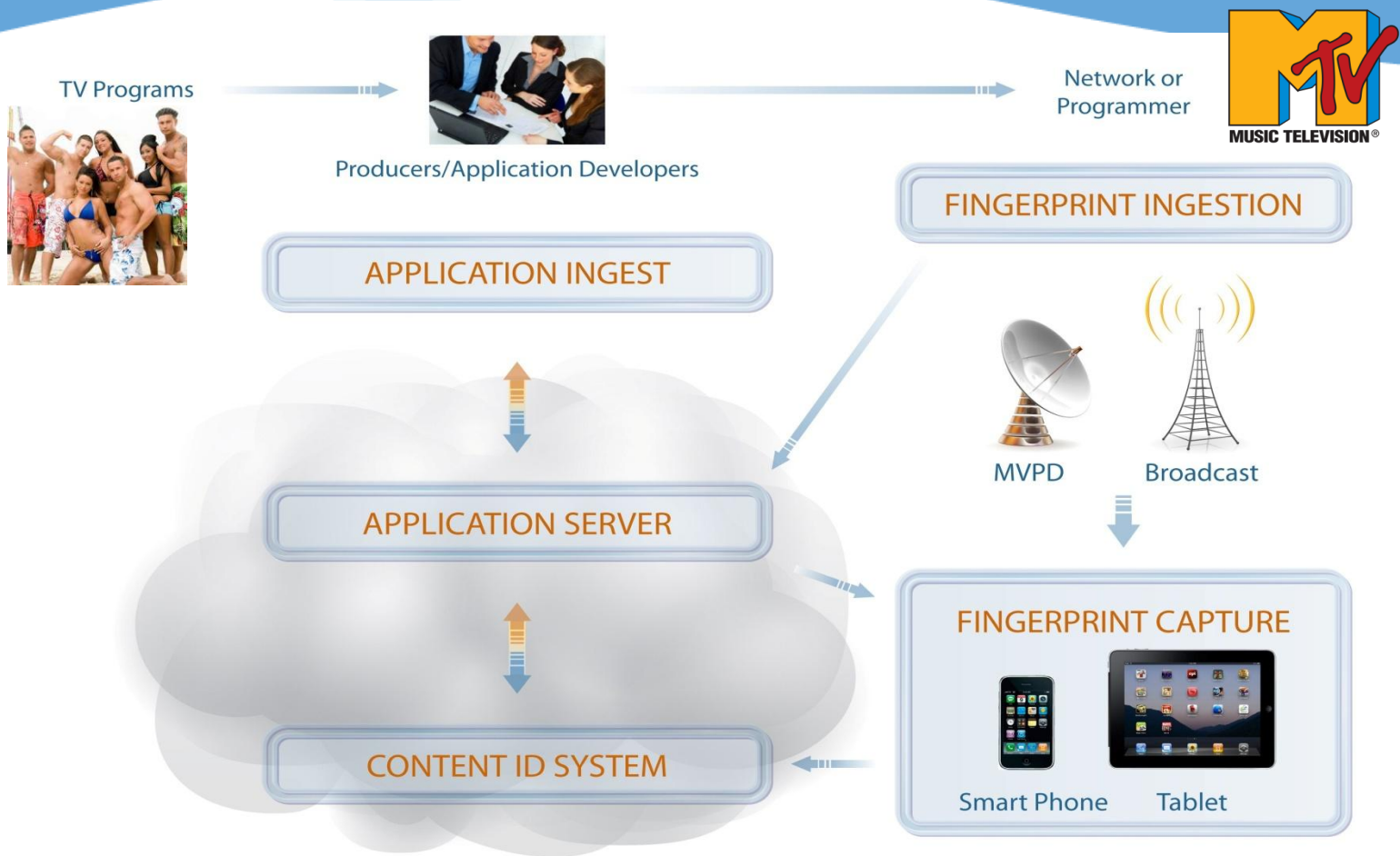


Synchronous 2 Screen Applications

Consumers multi-task while watching TV

No direct relationship with the audience

Synchronous 2 Screen Application



What's on TV?

① Launch Vvid App when watching TV
② Vvid Background Listens for 5-10 sec
③ IDs TV show, Movie, Commercial
④ Time Aligns App with Media

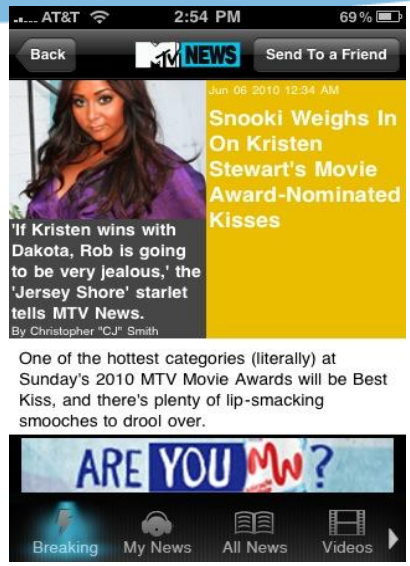
Coupon Serving – Synchronous 2 Screen

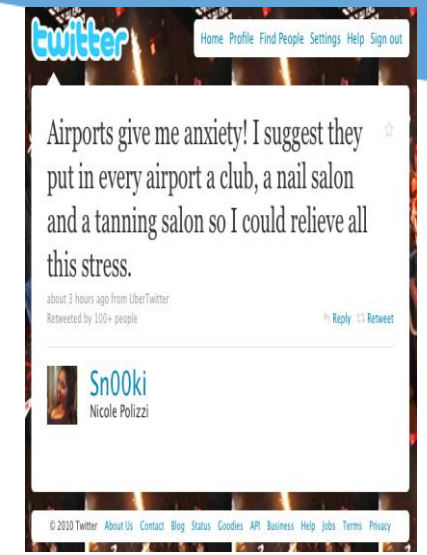
Second Screen App Enhancement

Synchronous Application Features

- ID program to call up application
- Enhanced content delivery tied to program slots
- Interactive – Interest, Coupons, RFIs
- Direct feedback from local audience (polling)
- Advanced advertising – sponsorship, measurement

Linear synchronous behavior





Social View

Vvid

Beginning of program 0:30

Sequence with extended viewing 15:16

Twitter – Facebook Updates 45:32

MUSIC TELEVISION®

Zeitera's Technology Advantage

Sound Algorithms for Content Recognition and Search

- Find virtually any video clip in an enormous database at very high speeds
- Virtually no False Positives, and near 100% True Positives at scale
- Deals with video or audio content

Excellent Scaling Characteristics

- Client runs on embedded processors in mobile phones, TVs and set-top-boxes to laptops and multi-core server PC platforms
- Search technology performance scales sub-linearly from hundreds of hours up to hundreds of thousands of hours

Good Economics for Consumer Electronics Deployments

- Large search infrastructure can be deployed with very-low operating cost
- Client runs on embedded processors and supports local databases
- Field tested, proven system with large networks and CE manufacturers

Zeitera's Software Offering

Product Offering

Audio-Video signature technology for broadcast and broadband video delivery from infrastructure to devices

Vvid Client

- Client deployed on TV, Smart Phone, or PC (Also STB/DVR)

Vvid Ingest

- Broadcast quality signature creation and analysis software (File based, Streaming –SDI or as library for integration)

Vvid Search

- Search infrastructure for storing and matching signatures (Managed Service or Technology License)

Target Customers

Programmers and Broadcast Networks, CE Companies - TV, Phone and Tablets, MVPDs, Advertisers and Ad Distributors, Semiconductor Vendors, Content owners, Content publishers and other technology/service providers to broadcast industry.

Summary

- Rapid convergence of the traditional TV and Internet is introducing fundamental transformation in home video and mobile entertainment environment
- It is paving a way to enhance and enrich passive TV viewing experience
- New technologies are emerging to support new interactive applications that will make user experience more entertaining and more engaging
- New “smart devices” have been developed, including SmartTVs, smart phones, and tablets

Summary (continued)

- ACR is an enabling software technology that goes hand-in-hand with innovative hardware
- It allows for a specific content to be identified, generate actionable event and launch a host of new interactive applications
- Content providers and advertisers are equipped now with new ways to make interactive programming and advertising more attractive and more profitable