Video AI: What’s in your content?

Martin Wahl
Principal Program Manager
Microsoft Azure Media Services
Redmond, WA
mawahl@microsoft.com
Azure is the most trusted and certified Cloud ...

**Standards and Guidelines**
- **Production and Broadcast**
- **Content Protection Standard v3.0**
- **Cloud & Application Security Guidelines**
- **Federation Against Copyright Theft**

**+ Relevant Frameworks**

...for media and entertainment workflows
Azure Media Services

Azure Media Services is an extensible, multi-tenant platform that enables end-to-end video workflows in the Azure public cloud and relies on Azure Storage, Azure Compute, Azure Networking, and other Azure Cloud Infrastructure Services.
Imagine if you could...

- Automatically create all of the metadata (spoken text, written text, faces, places, objects etc) for any video, including from their archives – flag any adult or objectionable content
- Relate advertisements dynamically to the content of the program for increased viewer relevance, value, and response
- Collaboratively upload & edit videos in real-time and through the Cloud to speed up time to air for new stories
- Enable viewers to take advantage of deep search capabilities – search by Name, Search by Keyword, Search by Topic – and view content in any language even if different from source
- Distribute playable clips of live video content internally or externally without long download or encode times
- Understand which part of your broadcasts were most popular/interesting to viewer and then add Interactive Elements into the viewing experience
- Create automated summaries or highlight reels of video content based on scene detection, specific people, and motion/actions within the video
Microsoft Cognitive Services
Give your apps a human side

**Vision**
From faces to feelings, allow your apps to understand images and video

**Speech**
Hear and speak to your users by filtering noise, identifying speakers, and understanding intent

**Language**
Process text and learn how to recognize what users want

**Knowledge**
Tap into rich knowledge amassed from the web, academia, or your own data

**Search**
Access billions of web pages, images, videos, and news with the power of Bing APIs

**Labs**
An early look at emerging Cognitive Services technologies; discover, try & give feedback on new technologies before general availability
### Video AI features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech-to-Text</td>
<td>Convert audio to text based on acoustic language models</td>
</tr>
<tr>
<td>Face detection</td>
<td>Find when each face appears in the video</td>
</tr>
<tr>
<td>Video stabilization</td>
<td>Create smooth videos from videos captured by moving camera</td>
</tr>
<tr>
<td>Video OCR</td>
<td>Extract text that appears in videos as overlay, slides or background</td>
</tr>
<tr>
<td>Face/Image redaction</td>
<td>Detect faces and choose which ones you want to redact</td>
</tr>
<tr>
<td>Motion detection</td>
<td>Detect when motion has occurred in videos</td>
</tr>
<tr>
<td>Face emotion detection</td>
<td>Recognize the emotion of a person or crowd based on facial expressions</td>
</tr>
<tr>
<td>Video summarization</td>
<td>Create summaries of long videos to enable quick previews</td>
</tr>
<tr>
<td>Content moderation</td>
<td>Detect and prevent explicitly visual or objectionable content</td>
</tr>
<tr>
<td>Object detection</td>
<td>Detect objects based on a pre-defined object model</td>
</tr>
</tbody>
</table>

Available as individual Media Processors via APIs
Video Indexer | Bundling 20 AI Features Together

**Linguistic Transcript**
Convert speech to text for 10 languages

**Noise canceling**
Eliminate background noise for call recordings using Skype filters

**Face detection**
Find when does each face appears in the video

**Keywords Extraction**
Find out the keywords discussed in each segment

**Contextual Search**
Understand the context of search results

**Spoken Language ID**
Detect spoken language to support multi language content

**Call recording enhancement**
Optimized ingestion for calls

**Face grouping**
Identify multiple appearances of the same person

**Sentiment Analysis**
Compare levels of positive vs negative spoken or written moments over the timeline

**In-place Editing**
Make manual fixes for errors detected

**Custom Vocabulary**
Fit to Industry, market and domain specific terms

**Annotations**
Tag objects such as cat, table, car, ball etc when they appear

**Identification**
See name, job and biography of celebrities and ordinary people

**Content Moderation**
Detect explicit visuals or text in audio or overlay

**Sub-Clipping**
Source video is stored once for multiple playlists of video segments

**Speaker Diarization**
Map and understand who spoke when

**OCR**
Extract text that appears in video as overlay, slides or background

**Shot Detection**
Detect when a shot starts/ends based on visual analysis

**Recommendations**
Find more videos with similar people discussing similar topics

**Programmatic APIs**
Index / Search API and UI widgets enable embedding in other website/apps

**Translation**
Translate source to 54 languages – text or voice

**Brand detection**
Track brand mentions in speech or on screen overheads

**Emotion sensing**
Detect emotions expressed in speech, vocal signals and facial expressions

**Logos**
Identify visual logos that appear on screen

**Live Analytics**
Analyze content coming from a live broadcast source

* Coming Soon
3 steps of using Video Indexer

Extract
Extract a set of pre-defined cognitive metadata

Curate
Review and curate the metadata using a web portal

Publish
Embed player and insights widgets in your own apps

Customer app
Microsoft Video AI Demo

http://www.videoindexer.ai
Video Indexer | Building Blocks

Re-envisioned classic experiences:

1. Gallery: In-video Search
2. Player: Insights Panel
3. Player: Content based Recommendations
4. Editor: Cognitive timeline
5. Editor: On the fly preview
6. Editor: Segments Playlist

APIs & Widgets

- Delete Breakdown
- Get Get Processing State
- Get Get Vtt Url
- Get Search
- POST Upload Video

https://aka.ms/viapi

AVI Platform

Search
Streaming
Indexing
Example: Customized News Services

I have a video with the latest volleyball match point

User A
Subscribed to “Sports:volleyball”

User B
Subscribed to “Business”

User C
Subscribed to “Sports:swimming”
Example: User Generated Content Filtering

User

Upload video
Encode and publish video

Content workflow

User

Checks video content
Coarse language
Nude scenes

Content is OK
Content is NOK

Notify user
Quarantine Content

Moderator workflow

"Your content has been rejected"
Example: Deliver Content to multiple partners

I have a list of partners who need to receive specific content files in different formats and qualities.

I also have a list of assets which need to be distributed to those partners.

If I add/remove an asset, the system needs to distribute/remove it to/from the right partners in the corresponding formats.

If I add/remove a partner, the system must distribute or delete the right assets to/from the destination.

Asset DB

Asset

Partner Table

Select partners interested in asset

Encode asset in multiple formats

Open FTP Connections and upload content

Partner A
Format 1

Partner B
Format 2
Example: Auto Generate subtitles and translate

As an operator, I want to generate automatically the subtitles of my content.

I want to remain in control of the quality of these subtitles.

I want those subtitles to be available in multiple languages.

Content workflow

Quality workflow

Asset

Speech-To-Text

Add subtitles to asset

Verify subtitles quality

Bing Translate

n times
Big Brother Going Into The Cloud

Adopting this ground-breaking technology will completely revolutionize the way we produce our global formats and opens up an unprecedented level of creative freedom. This is the first of many innovations we are working on that will optimize our productions around the world, giving our audiences across multiple platforms a much richer experience. [This] isn’t about cost cutting or optimization, this is actually about being able to tell great stories in totally different ways.”

— Lisa Perrin
CEO, Creative Networks, EndemolShine
Using Cloud & Video AI within a Production Environment

Multi-camera Production Output

- Low-res Live stream (RTMP)
- High-res Live stream
- 5 min MXF segments

Azure Media Services

- Subclipping and stitching
- Output copy
- Asset cleaning

Logic App

- Every minute
  - Subclip
  - Then run analytics
  - Face redaction
  - Speech-to-text
  - Motion Detection

Cosmos DB

- Face Redaction
- Speech
- Motion Detection

Video AI Services:
- Face API
- Speech Analytics

Logger

Upload

Azure Storage

- Upload

Logic App

- Logic App

Auto-generated

Logic App

- Logic App

AVID

- Import to on-premise

Editor

- Export to social media

YouTube

- Export to social media

Facebook

- Export to social media

Legend

- Low resolution flow
- High resolution flow
- Scene download request