Fraunhofer

#WeKnowMedia

Digital Media

In the Business Area Digital Media five Fraunhofer Institutes are cooperating to provide technologies and solutions for the media industry.
Fraunhofer IIS, HHI, FOKUS, IAIS, IDMT

www.digitalmedia.fraunhofer.de

Follow us on LinkedIN

Trends and Technologies in Digital Media

Pioneers in immersive digital media technologies

issue 12

Topics

Preface

Welcome to our latest edition of our trend broschure!

Cloud-defined workflows are currently a big deal in the industry. From the past until today, the post-production was kept on private servers. Cloud-based workflows provide a more efficient alternative to that due to scalable processing and resource allocation, and by eliminating the need to maintain dedicated hardware. Fraunhofer is working on media processing micro services to enable such workflows.

Artificial intelligence (AI) has a big influence on media and technology. Chat GPT and other implementations of AI revolutionized the dealing with language in the industry. We at the Fraunhofer Business Area Digital Media are keeping up with the rapid progress of media technology with our research.

In terms of speech-to-text and text-to-speech conversion, the use of AI allows natural language understanding and speaking. The automatic transcription and subtitling of movies or videos from voice to text is one of our research-fields. Here AI will make the deciding difference. It can capture the speech and translate movies into various languages, making them barrier-free for any audience, without too much effort.

Through AI, we are optimizing picture-based as well as audio-based search with our technologies. AI is a key-factor in accelerating search-processes. Therefor these can be faster and more effiecently applied by the user. This has a special impact on the applications, for example the automatic ad insertion in television-programs.

The virtualization of reality is also what drives our research. With volumetric video and light-field technology, our teams are experimenting with the potential of immersive experiences. The goal is to realize a close-to-reality experience in a virtual environment with our research. For example we are using light-field technology to portray a show on stage realistically so that the feeling of being on site is brought anywhere by the use of VR.

And with our audio and video codecs like MPEG-H, xHE-AAC, VVC, JPEG XS etc. for broadcast, production and distribution we provide best performance solutions suited for any use. Enjoy reading! We look forward to seeing you at IBC 2023 in Amsterdam!

Prof. Dr. Siegfried Foessel
Spokesman Fraunhofer Business Area Digital Media

JPEG XS Integration for Professional Workflows and Devices – made easy

The new JPEG XS codec is a revolutionary milestone in bringing professional high-quality video transfer to an all-IP workflow without neither limiting image resolution nor latency. With the Fraunhofer JPEG XS SDK you now receive a free Application Support Package that guide and assist system integrators and developers in the fast and convenient implementation of JPEG XS applications.

JPEG XS is the premium choice for ultra-low latency data transfer in broadcasting environments

Integrating an SDK is more often a tricky start, time- and money-consuming and costs a few sleepless nights. It´s not about a lack of knowledge, but what appears obvious in the code holds many pitfalls before it gets started successfully. The video coding experts of the Fraunhofer Institute for Integrated Circuits IIS decided to shorten this time by providing an Application Support Package that is included in the Fraunhofer JPEG XS SDK.

"We have focused on code examples for the most common encoding and decoding functionalities in typical workflows such as camera image to stream conversion or stream to display image" explains Siegfried Foessel, Head of the department Moving Picture Technologies at the Fraunhofer IIS.

JPEG XS – High performing for the Broadcaster´s choice

In broadcasting a variety of video frame grabber and playout cards, from standard to ultra-low latency, are used for data transfer. Having these in mind, the Application Support Package for the JPEG XS SDK from Fraunhofer IIS supports the three currently most used I/O cards for SDI and HDMI – the most common for each segment – with sample code implementations.

But, what about already built-in internal computer interfaces for webcameras, X11 system or the GPU? For these use cases the Fraunhofer experts developed software code for interfacing with Video4Linux, X11 and the Direct Rendering Manager (Linux only). And, for cross-platform use, software code for integrating the SDL2 library is available.

Transport streams and file formats

Additional functionality is provided by Fraunhofer code for encapsulating code streams into file formats and transport streams. Supported file formats are *.JXS, *.JXSC, MXF in AS-02 style (single-essence component), MXF in AS-11 style (multi-essence component), *.MP4.

Many of these formats are primarily used or required in broadcasters´ workflows or equipment in order to take full advantage of the low-latency and high quality of JPEG XS. For transport streams, the Application Support Package also contains code for packaging and unpacking of JPEG XS code streams into RTP (RFC 9134), enabling ST2110 workflows and MPEG-TS (ISO/IEC 13818-1).

Improved error resilience for transport streams

In order to achieve a smooth presentation of the video streams in the event of network errors, the Fraunhofer code adds additional error resilience functions to avoid interrupted or broken video streams. Even better error resilience can be reached by using the special Fraunhofer functions for Forward Error Correction (FEC) according to ST2022-5. Our SDK customers therefor also find a sample code to integrate these amazing JPEG XS supporting features in their workflows and devices.
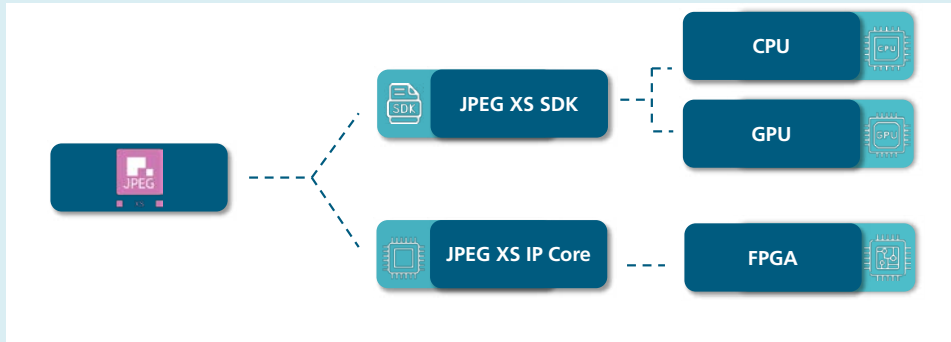
3 ways to easily integrate JPEG XS in your workflow

Fraunhofer JPEG XS inside – realized in KVMs of IHSE

Together with the IHSE GmbH, the leading developer and manufacturer of advanced KVM devices in global demand, the integration of JPEG XS into the Draco Con App provides a high performance real-time video playout on computer monitors. This realizes a software based KVM over IP as virtual KVM. FPGA-based KVMs are the next step in the development.

Subframe latency support – balancing the core workload

To improve performance, Fraunhofer IIS provides a smart thread allocation in its SDK. This allows users to achieve Ultra-Low-Latency by evenly distributing the processing workload across all cores. The Fraunhofer code is available for parallel and interleaved processing of image slices and results in only 180 lines latency for a purely software based coding and transport system according SMPTE ST2110.

High Speed conversion from 422 YUV to 444 RGB

A certain plus for broadcasters are this high speed conversion from standard broadcast 422 to high resolution displays of 444. A real-time display for broadcast streams on standard monitors e.g. in the camera control room are possible without additional expense.

For further information on JPEG XS

Contact

Maseeh Habibzay
Fraunhofer IIS
jpeg-xs-business@iis.fraunhofer.de

VVC Video Streaming:
Smooth Transitions during Resolution Jumps

When streaming a video and the bandwidth goes down and the resolution of the video image therefore drops, this happens in most of the times abruptly. The new international video coding standard VVC however, ensures pleasantly perceived transitions.


From one moment to the next, summer seems to be over, temperatures fall, rain clouds appear wherever you look. Many people are troubled by such sharp weather changes. If, on the other hand, the summer is fading away and the autumn weather is slowly taking over, we usually only notice this on the sidelines. It's a similar story with other abrupt changes, such as when streaming video: If the connection deteriorates, the resolution of the video image collapses – and in a jerky manner. Not necessarily a pleasant effect. The

reason for this: If the bandwidth is no longer sufficient for 4K or HD quality, the system switches to smaller resolutions like SD and thus benefits from lower data rates. However, the encoding always has to be restarted when the resolution is changed which causes an abrupt change in the video's picture quality. Moreover, restarting encoding comes at the expense of efficiency: after all, encoding is at its best when the image currently being encoded can be inferred or predicted from parts of already encoded, previous images. However, if the resolution changes, the previous images referred to will have a different size and they can no longer be easily used for prediction.

Smooth transitions thanks to VVC

This is not the case with the new international compression standard Versatile Video Coding or H.266/VVC, in the development of which the Fraunhofer Heinrich Hertz Institute HHI played a major role. Compared to the predecessor standard H.265/HEVC, it

H.266 / VVC
Codec Support
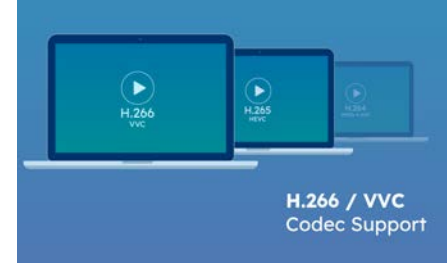
not only saves half the bit rate at the same visually perceived quality, but also makes it possible to continue encoding without a hard drop in quality when the resolution changes. "Via the `reference picture resampling' technique, we can scale the previous pictures used for prediction to the same size and thus continue to use them for encoding as well", says Benjamin Bross, group leader 'Video Coding Systems' at Fraunhofer HHI. "In this way, the content can be efficiently compressed even when the resolution changes, which again saves bandwidth." This also offers advantages in terms of viewing pleasure: Instead of changing jerkily, the resolution of the image slowly gets worse or better – as a viewer, you therefore hardly notice the change.

Video streaming with different resolutions

How big the perceived gain is when streaming video with different resolutions can be experienced by two Fraunhofer HHI showcases . On the one hand, at a cloud-based 'video on demand' solution: "Here we can show the entire multi-resolution streaming workflow with our partner Bitmovin: The video data is encoded in different resolutions in the Bitmovin Cloud with our open-source VVC encoder VVenC, and the playback is done via the mobile

Bitmovin Player, in which our VVC decoder VVdeC is integrated", explains Bross. For visitors, this means they can easily switch between high and low resolutions on a tablet – and admire the amazing video quality at very low bit rates.

In addition to video-on-demand, live encoding will also be on display: SpinDigital's VVC Live Encoder can simultaneously encode 8K, 4K as well as HD and smaller resolutions. The encoded video sequences can be seen on an 8K TV, using a compact Apple Mac Studio computer for playback. This runs the Elmedia software player, which also integrates Fraunhofer HHI's VVdeC decoder. "Our VVdeC Decoder is specially optimized for processors based on the ARM design. These are as powerful as they are energy-efficient and are integrated, for example, in all new Apple computers", says Adam Wieckowski, co-group leader 'Video Coding Systems' at Fraunhofer HHI. Visitors can switch between the old and new coding methods and experience the improvement first-hand.

Contact

Benjamin Bross
Fraunhofer HHI
benjamin.bross@hhi.fraunhofer.de

Volumetric Video

If someone sits opposite to us via a screen, the impression of the person remains two-dimensionally flat. Not so, however, with Fraunhofer HHI's »Interactive Volumetric Video« technology. Group manager Dr. Oliver Schreer explains what is behind this term and what possibilities the new technology offers.

Interactive Volumetric Video technology

Mr. Schreer, what is meant by Interactive Volumetric Video?

Usually, people captured on video remain for the spectator as two-dimensional impression. With Volumetric Video, however, the viewer has the feeling of being directly opposite the person. Through the VR glasses, they see themselves standing in front of them just as much threedimensionally as being in front of a real person, and they can walk around them – and even keep eye contact. We achieve this by subsequently aligning the head orientation in the direction of the viewer's current position.

This makes sense, for example, with eyewitnesses of the Holocaust: In a VR experience that is now completed, Eva Umlauf shares her memories as the youngest survivor of the Holocaust of the Auschwitz-Birkenau concentration camp, in three scenes of 30 minutes in total. This video is enriched

by additional media such as an interactive photo book, videos and audio material. At IBC we will present this VR experience. The official launch will take place on November 8, 2023 in Berlin at the Science Tech Space@HHI – together with Eva Umlauf.

With contemporary witnesses such as Eva Umlauf, it is a matter of preserving memories. Is Volumetric Video technology also suitable for real-time applications?

Yes, for sure. Let´s take an example in the form of the Toolbox for Hybrid Event Formats in the ToHyVe project, a current joint project of the German Federal Ministry for Economic Affairs and Climate Action, BmWK together with nine partners. In this project, components, services and tools are developed in order to better connect the virtual and real worlds for virtual and real worlds events – be it for hybrid events, workshops or trade fairs. How can face-to-face and online participants be brought together as well and as realistically as

possible? One example is the standing table, where two real participants can talk to an online participant shown on a display – with the feeling of really being there together. Incidentally, this also applies to eye contact. Other examples include showcase presentations, where both, real and virtual exhibits can be presented, including a complete virtual event space where people can meet and hold conferences.

With new Augmented Reality technologies, there are likely to be assumed numerous issues regarding the safety, accessibility, and ethics of the applications. How do you deal with these concerns?

We are addressing such issues in the EU project XR4Human, which is funded in the Horizon Europe research framework program. A total of twelve partners, including Fraunhofer HHI, are working on the usability of Augmented Reality user

experience and inclusion, as well as ethical and legal issues, particularly with regard to privacy and data security. Standardization and interoperability aspects are parts of this project, too. The goal is: We want to develop guidelines that assist research and development as well as politics and users in the development and application of augmented reality.

Contact

Dr. Oliver Schreer
Fraunhofer HHI
oliver.schreer@hhi.fraunhofer.de

The Stage in the Living Room

SocialSTAGE-VR - cultural and social participation through a joint virtual experience of real stage events

The conversations in the hall fall silent, the lights dim, the last audience members find their seats, the stage lights come on. The curtain opens. There is much more to an evening at the opera than just the show on stage. Being part of the impressive atmosphere – that is what makes theater great. An evening at the opera is not always possible nor for everyone. There are various reasons, for instance no need to travel that can prevent from such an event. And, the theaters are able to multiply the possibility to be part of a theatrical performance beyond the seating capacity of the theater. What if all these obstacles could be overcome?

Dr. Joachim Keinert

That is exactly what the project SocialSTA-GE-VR tries to realize: a realistic and social experience at the opera from home by using virtual reality (VR), light-field techno-logy and sound technologies to provide an immersive experience close-to-reality for the audience. In an interview with Dr. Joachim Keinert from the Computa-tional Imaging group of Fraunhofer IIS we were able to sketch the challenges and the possibilities.

How are you realizing such a close-to-reality experience for the audience?

For the audience to feel like they are really on-site, it is important to virtualize the whole spatiality of the theater. That is done by using the light-field technology. By that, we capture the live act with a camera array to be able to provide different perspectives. Based on these images the light-field can be partially recreated in a virtual space. The benefit of those light-field images is that the audience can naturally move their torso without getting distorted images. By these means, the solution is particularly tailored for immersive video experiences where the audience is essentially seating or standing on a dedicated place.

What are the major challenges?

One of the most challenging working steps is the extraction of depth-information from the captured images in order to build a complete light-field picture. The necessary algorithms that are able to filter all needed information to create such a photorealistic light-field image are developed by specialists for computational imaging at Fraunhofer IIS.

Dr. Keinert, what is the current status of SocialSTAGE-VR?

Currently we have just finished a study in which we tested the immersive experien-ce for potential users. We invited various experts and persons to test and try out our application. Our focus was on the close-to-reality-feeling for the users in terms of the social experience. We tried different possi-bilities of implementation for the VR-appli-cation to find out, how we can realize our claim of an immersive experience. The study has shown that for a convincing experien-ce, it is important to consider a variety of aspects, such as camera calibration, image sharpness and proper color reproduction, depth budgets and human perception, as well as depth map quality and rendering technology.

What are the next steps?

The project is supported by the German Federal Ministry of Education and Research and coordinated by the Chemnitz University of Technology. Together with our project partners rooom AG, Die Etagen GmbH, pointomega AG and YOUSE GmbH, the next step is now to transform all the lessons learned from our first trial into a second version of the application for an improved demonstration. Based on the study outcomes, we target best expe-riences for current user expectations with today's available technology. Moreover, we will of course continue our research efforts in making light-field rendering better than ever.

Contact

Dr. Joachim Keinert
Fraunhofer IIS
joachim.keinert@iis.fraunhofer.de

Energy-efficient and sustainable video streaming

Private households in Munich consume around 2.5 terawatt hours of electricity per year. The electricity consumption of the Internet – i.e., the operation of networks and data centers – amounts to 13 terawatt-hours per year in Germany. Video streaming accounts for a large share of the total online data traffic, with the CO_2 emissions for one hour of video streaming ranging from 42.7g CO_2 to 400g CO_2 in recent studies. The partners of the "Green Streaming" project are investigating how the power consumption and thus CO_2 emissions of streaming can be reduced.

In 2027, around 3.7 billion people worldwide will use video on demand services and the delivery of live content via OTT platforms is growing steadily. Streaming is the predominant technology that delivers those digital video content to viewers and their screens. However, to provide streaming services, video content must first be produced, encoded, packaged, and transmitted before it can be played on a suitable device.

All steps within this streaming supply chain require energy and thus contribute to the emission of greenhouse gases. The specific energy demand vary by component and applied settings like the video codec used, as well as the resolution and quality of the video content. In addition, the distribution of the content and how it is used on different end devices impacts the total energy demand. Despite the importance of these interrelationships for the environment,

there is still a lack of possibilities for precisely measuring the specific energy consumption of the involved components along the streaming supply chain and their interactions with one another. The "Green Streaming" project aims to change this: the project partners have set themselves the goal of increasing energy efficiency in video streaming and using existing resources intelligently. To achieve this, they are relying on hybrid cloud production environments and artificial intelligence (AI)-based, context-aware video encoding.

With the help of a "green digital twin", a digital image of the real streaming workflow is created. The "digital twin" makes it possible to model optimizations within the individual steps of the streaming workflow in a time-efficient and resource-saving manner and to apply these to real workflows. The project aims to develop, test, and apply a

measurable, assessable and sustainable digital media value chain that enables the seamless, holistic, and demonstrably "green" provisioning of streaming content.

Challenges

There are several challenges to overcome in order to achieve this goal. First, the complexity of the processing chain must be analyzed and understood. Various factors involved in encoding, storage, and distribution affect the resulting data traffic and ultimately the energy consumption. Machine learning will be used to analyze these various factors to ensure that the viewer is provided with the best possible quality video stream while consuming as little energy as possible. A key goal here is to ensure a balance between a high-quality user experience and environmental sustainability.

Along the streaming value chain various components are responsible for preparing, processing, storing, and distributing media. To make the most use of hardware and

software resources in this infrastructure, an intelligent and dynamic provisioning, orchestration, and utilization of available computing power, storage and workflow components is essential. Machine learning will be used to determine and optimize dynamic resource requirements based on the "green digital twin", historical data, and incorporated real-time metrics. This approach will ensure that the required resources are used efficiently and sustainably to minimize environmental impact.

Another essential challenge is to reduce the energy consumption of the playback devices while maintaining the Quality of Experience (QoE). For this purpose, a green streaming mode will be developed that individually determines how much energy can be saved at the end device without compromising the user experience. By using quality measurements and user behavior data, "Green Streaming" attempts to find an adequate trade-off between energy savings of devices and viewing quality. However, this requires suitable measurement points along the

entire streaming chain, which are critical to accurately assess energy consumption and make corresponding adjustments. The determination of these measurement points represents a central challenge in the project.

Green Digital Twin for Streaming

With the "Green Digital Twin", an image of reality is being created that ideally replicates the entire streaming value chain. The "Green Digital Twin" is initially modeled based on study results and synthetic data. Using the previously identified measurement points along the streaming value chain, further real data is collected step by step.

As an example: power consumption during streaming is measured - both during delivery in data centers and during playback on user devices. Supported by artificial intelligence, the "Green Digital Twin"

continuously learns and improves its digital representation of real streaming workflows. The data collected under real conditions is then modeled and the subcomponents are adjusted in terms of energy-efficient and sustainable streaming. In this way, the "Green Digital Twin" behaves increasingly similar to the physical world. Based on this digital image, the project partners can analyze and adopt the components and their parameters to optimize the workflow. During the project, time- and resource-intensive optimizations, such as measures to minimize power consumption, can thus be simulated, and tested more efficiently.

Subsequently, the modified parameters are tested in real workflows, and recommendations for actions or adjustments to the streaming workflow can be developed. Thus, the "Green Digital Twin" fulfills two important functions in the project: on the one hand, it is the central tool for optimizing the streaming chain, and on the other hand, it is a clear "best practice" communication tool for presenting the project results.

The results achieved by the project will be tested in practice by the participating partners through prototypes and demonstrators with a focus on energy consumption, sustainability and environmental friendliness.

The use of open-source software (OSS) and open standards ensures that the solutions developed by the project are accessible for broad application supporting a sustainable and environmentally friendly future in the domain of digital streaming.

Contact
—

Robert Seeliger
Fraunhofer Fokus
robert.seeliger@fokus.fraunhofer.de

Delivering MPEG-H Audio to Creatives Around the Globe

Immersive and personalized sound experiences are the future of audio. A growing demand for them makes versatile, high-quality technologies like MPEG-H Audio indispensable. Thanks to new strategic partnerships, producing in MPEG-H Audio has become easier and more accessible than ever before.

The feeling of being immersed in crystal-clear sound transforms any listening environment into a home theater or a front-row seat at the concert. Thanks to advanced 3D positioning functionalities, creatives can easily place the listener in the middle of the action so that they and can enjoy a truly immersive experience.

Personalized sound is tailored to individual hearing preferences – just like a custom-made audio system is set up to meet a person's taste in every respect. In addition, accessibility features such as dialogue enhancement in TV broadcast support listeners with hearing and visual impairments.

A Variety of Needs, One Compact Solution

The MPEG-H Audio system by Fraunhofer IIS is a comprehensive solution that provides creatives in the music, streaming and broadcast industries with a rich toolset of functionalities to meet the demands of their sophisticated audience. The Next Generation Audio technology makes it possible to define personalization options during production and even to allow listeners to customize some parameters within a predefined frame. This feature is the basis of advanced accessibility features such as dialogue enhancement and audio presets in TV broadcast as well as of sophisticated interactive music productions. In addition, sound sources such as vocals, chorus, and instruments can be positioned in a three-dimensional space to perfectly match the creative and artistic intent.

Music producers and sound engineers do not only value MPEG-H Audio for its remarkable sound quality, binaural headphone rendering, and extraordinary immersive playback experience. Its unique personalization options make it the first choice for those who want to explore exciting, innovative opportunities for artists and fans to engage with each other and those who value the chance to deliver content that is accessible for an even broader audience.

MPEG-H Audio for Personalized Immersive Sound

Broadcasters and streaming providers around the globe already count on MPEG-H Audio for their services. For example, MPEG-H Audio outperformed numerous candidates and was selected as the only mandatory audio system for the upcoming Brazilian TV 3.0 standard. In South Korea, terrestrial ATSC 3.0 broadcasting is already on air, making MPEG-H Audio the world's first commercialized Next Generation Audio technology.

**The Industry's Favorite Workstations: Now Powered by MPEG-H Audio**

Thanks to Fraunhofer's strategic partnerships with Avid and Steinberg, producing MPEG-H Audio content has become easier than ever before. Avid now provides a free plugin for all Pro Tools Ultimate customers that enables creators to produce personalized and immersive audio.

Steinberg has also integrated the world's most advanced Next Generation Audio system seamlessly in upcoming Nuendo

updates. Without having to deviate from their customary workflows, producers are now able to create unique sound experiences. These developments open up unmatched creative possibilities for all users of these industry-leading DAWs.

GitHub – the Road to Simpler and More Diverse Playback

The final puzzle piece to making MPEG-H Audio widely accessible is ensuring there are plenty of playback options to choose from. Fraunhofer IIS has now released their decoder software implementation on Github to make it easier for all developers to include MPEG-H Audio playback into their applications. This facilitates the development of comprehensive solutions for personalized and immersive sound for software players and client devices. Releasing the decoder software on GitHub is the next step on the road to deliver MPEG-H Audio to creatives as well as sophisticated listeners around the globe.

MPEG-H AUDIO
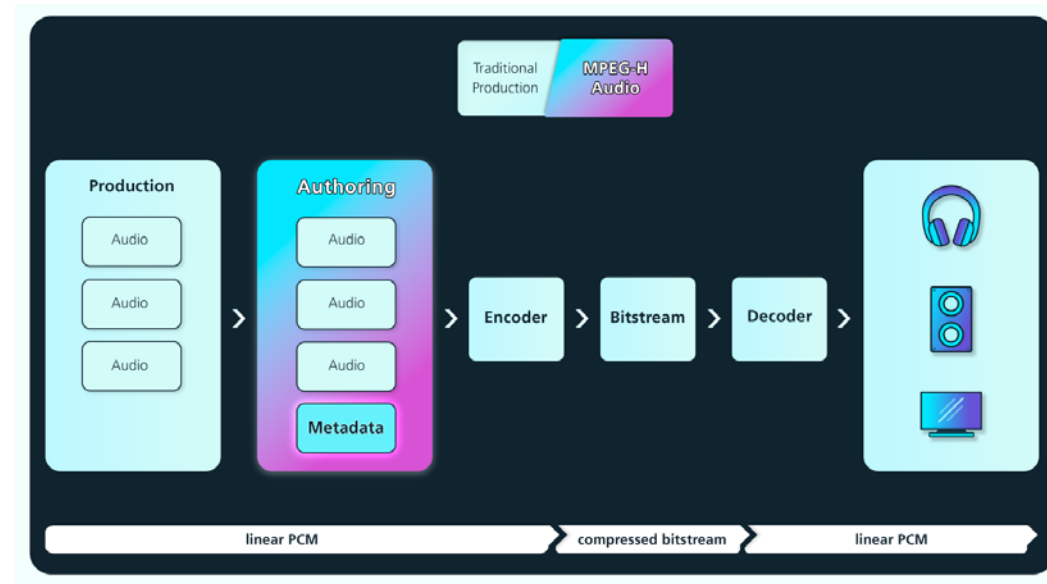
Contact

Mandy Garcia
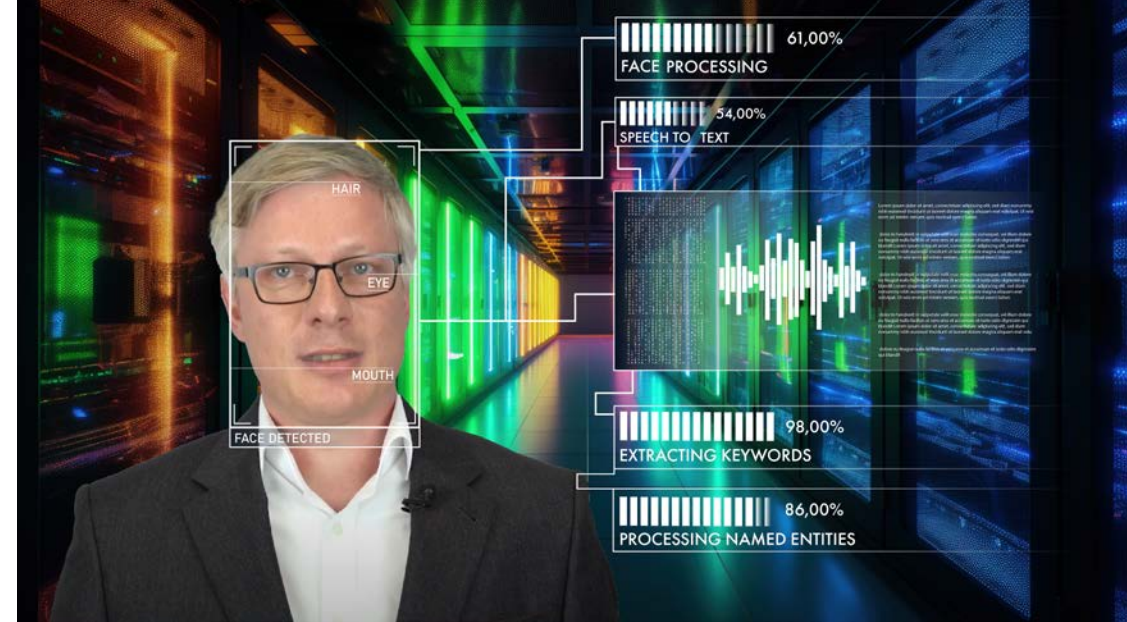Fraunhofer IIS
audio-info@iis.fraunhofer.de

Mining Platform: Automated AI-based Insight Extraction at Scale

The Fraunhofer IAIS Mining Platform is a container-based cloud application for automatically extracting insights from text, audio, image, and video at scale. Many AI-based services for analyzing media assets in different modalities are available out-of-the-box: text documents can be processed with named-entity recognition, keyword extraction, topic modeling, or smart keywording services. Speech can be transcribed using the Fraunhofer IAIS Audio Mining solution, and faces can be detected and recognized in both image and video files.

Automatic extractions from audio, text, video at scale

Thanks to its modular structure, additional services can be integrated easily, as the services only need to implement a standardized REST API and need to be deployable as a Docker container. AI-based services for object recognition, concept detection, cut and scene detection, and semantic keyframe extraction are ready for deployment resp. close to completion. All AI services offered by Fraunhofer IAIS come with readymade models trained on openly available datasets, but it is also possible to train and use customer-specific models. Third-party AI services can also be easily integrated, by implementing an adapter that mediates between Mining Platform and the third-party AI services. This approach has already been used to integrate a text translation as well as a face recognition service with the Mining Platform.

The individual steps for analyzing media assets are governed by a workflow component, which makes large-scale insight extraction robust and fail-safe. Explicitly orchestrating the workflow steps also allows to prioritize the analysis of media assets depending on the use case. For example, breaking news-related media files can be given precedence over archival footage, enabling the Mining Platform to serve the information needs of different user groups.

The workflow component also allows modeling complex media analysis workflows, which facilitates combining and enriching results obtained with different AI services. A first example is an entity linking workflow, which first disambiguates named entities extracted from an input text, and then links the named entities

with various knowledge graphs and databases, hence providing richer information than simple named entities.

Ongoing and Future Developments

We are currently integrating the updated Audio Mining components into the Mining Platform, which will later serve as the middleware of the next Audio Mining system.

A very important upcoming feature is support of external workflow definitions, which would allow customers to define their own workflows. In combination with the simple integration of AI services via a standardized REST-API, this will make the Mining Platform even more flexible and adaptable to the needs of its customers.

Contact

Dr. Jens Fisseler
Fraunhofer IAIS
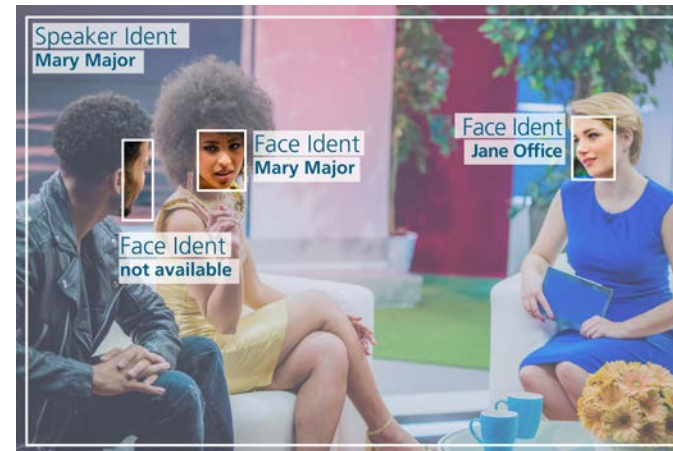jens.fisseler@iais.fraunhofer.de

Audiovisual Identity Suite

In today's digital media era, the pressing question for many is: How can vast volumes of media content be swiftly and smartly analyzed, particularly in relation to the individuals portrayed? The Fraunhofer Institute for Digital Media Technology IDMT has developed a solution that quantifies and visualizes the media footprint of individuals or groups based on video and audio analysis. Target groups are media companies, broadcasters and public authorities.

Automatic AI-based insight extraction at scale

Critics often remark on Chancellor Olaf Scholz's lack of television presence. But is that perception grounded in reality? To address such queries, large amounts of video content have to be analyzed with regard to the people shown.

The researchers at the Fraunhofer IDMT specialize in this very domain, dissecting huge audiovisual datasets. "We have already conducted studies with more than 23,000 hours of video content," confirms Dr. Uwe Kühhirt, head of Audiovisual Systems at Fraunhofer IDMT. However, not only the size of the data volume to be analyzed is so special, but also the possibility of combining video and audio analysis and thus optimizing the informative value and quality of the results.

Person analysis in words and pictures

Such comprehensive analyses have multifaceted applications. On the one hand, such an analysis can be used to examine which persons were most present in a program. The results are presented in an organized tabular layout, accompanied by thumbnail images of the person for example, together with specific broadcasting times and durations. It is also possible to identify specific people. We might question, for instance, when the Chancellor was visually or audibly present in specific clips from a talk show. "Such a combination of face and speaker recognition is especially of interest, e.g. when a specific person speaks in talk shows, but other people of the conversation group or from the audience are faded in", explains Christian Rollwage, head of Audio Signal Enhancement.

The results can be displayed in "heat maps" that visually represent when and where a person appeared, or was audible, across various TV channels over customer-specified durations. Such analyses are profitable for program design and planning: Planners can, for example, select individual programs and broadcast times, such as the main news program of July 23, 2023, and receive a list of who was featured and their speaking lengths.

Support for balanced reporting

In addition to this people-based analyses, it is possible to search for spoken keywords in the program and to determine the perceived gender of the people portrayed. This faciliates gender-based statistical analyses, answering questions like: How often are women seen or heard in the program, and how often are men? It is also possible to examine the speech intelligibility of entire programs or individual segments.

Licenses and services

Fraunhofer IDMT offers the Audiovisual Identity Suite as a bespoke service, including data capture and visualization on request. They aggregate media content on behalf of customers, enrich it with metadata, and tailor evaluations to specific needs. In order to use the data and the analysis, customers do not need any access to a cloud – a big advantage in terms of data security. Moreover, clients have the option to license selected analysis components for integration into their existing systems. Surprisingly, the tangible outcomes of such evaluations often defy public perceptions: Contrary to public perception, Olaf Scholz appeared quite frequently on the screen.

Contact

Dr. Uwe Kühhirt
Fraunhofer IDMT
uwe.kuehhirt@idmt.fraunhofer.de

Fraunhofer Business Area DIGITAL MEDIA

The cooperation of Fraunhofer Institutes within the business area Digital Media provides innovative solutions and products for the digital age of motion picture.
We provide technological innovations for digital media workflows and for immersive viewing and sound experiences. Benefit from our expertise in research and development as well as in standardization. The institutes offer research and development in the areas of production, audio systems, data compression, post processing, transmission, projection, distribution and digital archiving.

As an one-stop competence center for digital media we provide for our customers scientific know-how and the development of solutions that can be integrated in workflows and optimize process steps.

The members of the Digital Media Business Area are actively working in renowned organizations and bodies like International Standardization Organization ISO, ISDCF (Inter-Society Digital Cinema Forum), SMPTE (Society for Motion Picture and Television Engineers), FKTG (German Society for Broadcast and Motion Picture), and in the EDCF (European Digital Cinema Forum). We are also a partner of the 3IT, the Innovation Center for Immersive Imaging Technologies and the Fraunhofer Digital Media Technologies, Fraunhofer USA, Inc. These contributions enable research and development activities based on international standards.

The Fraunhofer Institute members are

- Integrated Circuits IIS, Erlangen
- Telecommunications, Heinrich-Hertz-Institut HHI, Berlin
- Open Communication Systems FOKUS, Berlin
- Intelligent Analysis and Information Systems IAIS, St. Augustin
- Digital Media Technologie IDMT, Ilmenau

Impressum

Contact

Fraunhofer
Business Area DIGITAL MEDIA
Angela Raguse M.A.
Phone +49 9131 776-5105
contact-digitalmedia@iis.fraunhofer.de
www.digitalmedia.fraunhofer.de