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 11

Preface

Welcome to our latest edition of our trend broschure!

One of the influencing aspects for our developments and projects in the current situation is the corona pandemic, which has led to a focus on public and private communication. The last weeks especially taught us what it means when events, concerts and most of our public life have to take place in hybrid or digital environments: The learning is that there is still some way to go until we will have all necessary tools available. It´s still a workload to develop and implement that enable hybrid media events and make them feel similar to on-site ones.

In 2021, the Institutes of the Fraunhofer Business Area Digital Media carried out a lot of developments, evaluations and tests in cooperation with the media and creative industry to find new ways to add additional experiences and values to existing audio and video streaming, collaborative workflows and immersive media events.

With our project Virtual LiVe hosted by 3IT we started this year an unprecedented initiative where we tried to set-up a tool box with Fraunhofer technologies from codecs to volumetric video and from 3D audio to immersive 360 degree streaming.

In validation projects together with industry, media and creative partners we tested different technologies for use cases like hybrid concerts, trade shows and concerts. In December e.g. there will be a live concert in the Berlin Kesselhaus, that is at the same time streamed to the planetarium in Bochum and can be watched via internet.

In this issue of our trend and technology brochure you will find insights about these technologies and more about developments of Fraunhofer Digital Media like 3D MPEG-H Audio, Versatile Video Coding VVC, Volumetric Video Streaming, xHE-AAC, JPEG XS coding, DCP/IMF generation and Digital Radio.

We are happy to share the recent news with you on our website and on our LinkedIn channel and to meet you in person at one of our upcoming events.

Stay healthy and enjoy reading our trend topics.

Sincerely, yours

Prof. Dr.-Ing. Siegfried Foessel

Spokesman Fraunhofer Business Area Digital Media

Topics

Hybrid live concerts: Bringing artists and music fans closer together

Live concerts? We've had those forever. But picture this: Live concerts in which the musicians, instead of standing together on a single stage, are connected online and able to play their music without any perceptible delay. Concertgoers at different venues could participate in the action in person, and artists would hear their fans' applause in real time.
Now, that would truly be something new.

With cinemas and theaters closed down and long-awaited concerts cancelled, the COVID-19 pandemic and subsequent lockdown have made life difficult for event managers and artists alike. Researchers from three institutiones, Fraunhofer FOKUS, Fraunhofer HHI, and Fraunhofer IIS are working together to get the industry back on track and lend it as much support as possible. To do this, they established the Virtual LiVe project, funded by Fraunhofer's "KMU akut" program, which focuses on research for small and medium-sized companies burdened by the corona pandemic. The aim of the project is to make it easier for SMEs to access cutting-edge technologies, all while fostering the growth of new and hitherto unimaginable formats that offer long-term benefits across the industry.

"We are combining a range of existing Fraunhofer technologies in a single toolkit," explains Christian Weissig from Fraunhofer HHI. SMEs can access the tools they require and combine these with their own technologies in any way they need. The research team began by conducting workshops with companies and external partners to get a better understanding of the demands of the SMEs. "In further stages we plan to carry out feasibility studies, and validation projects in which we work together with the SMEs to test the toolkit in specific applications," reports Prof. Siegfried Foessel from Fraunhofer IIS. Companies also have access to a monthly colloquium where they can network and share information about their technologies.

A hybrid concert in Berlin's Kesselhaus

»The best way to understand the Virtual-LiVe project is to consider one of the initiative's validation projects, in this case, the hybrid concert at the Kesselhaus in Berlin. «, explains Stephan Steglich from Fraunhofer FOKUS.
On December 11, 2021, Billy Andrews, also known as The Dark Tenor, will hold a concert at the venue. This event will have three very special features. First, Billy Andrews will be accompanied live by the Queenz of Piano. Unlike Andrews, however, the Queenz of Piano will not be playing at the Kesselhaus – instead, the performers will be connected via the Internet. Second, the concert is set to be streamed live to the Planetarium Bochum, where the audience can enjoy a 360-degree audio-visual experience – almost like the live show itself. Audio and video will not only stream from the Kesselhaus to the planetarium; data will be transmitted in both directions, allowing performers at the Kesselhaus in Berlin to hear synthesized applause from the audience in Bochum. This synthesized applause is generated from emojis that participants can send via a web app. Cinemas may serve as additional locations, and initial tests are already underway. Finally, the concert will also be streamed live to and for viewers at home. Here, the idea is to offer remote viewers the possibility to enjoy the concert in a way that comes as close as possible to the live experience. Viewers will be able to interact with the artist and, with a little luck, even be invited on stage by Andrews himself. It may even be possible to sell special online tickets for these kinds of extra features in the future.

Finally, completing the toolkit is MPEG-H, an audiovisual coding standard. MPEG-H 3D audio, which is developed in large part by Fraunhofer IIS, is used to produce three-dimensional sound that improves upon existing surround sound technologies. "This 3D sound is not only perfectly suited for the planetarium in Bochum – it is also ideal for home sound systems or headphones when you're on-the-go," explains Dr. Ulli Scuda from Fraunhofer IIS, adding, "and the audio only needs to be produced once."
The MPEG-H decoder integrated in commercially available AV receivers renders the signal for the specific end device in each application, producing high-quality playback in any setting.

The virtual concert venue

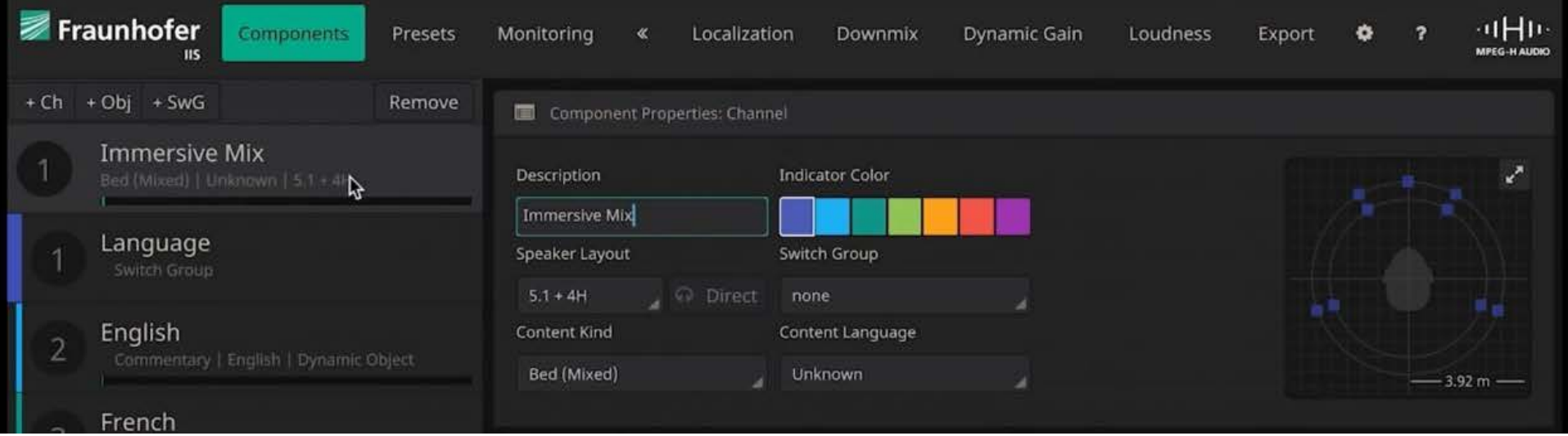
In a second validation project, the Fraunhofer research team plans to take another major step forward – this time, joined by Radar Media and High Road Stories, the companies behind FANTAVENTURA, a virtual reality experience created for the German hip-hop ensemble Die Fantastischen Vier. In this validation project, titled "The Other Room: Martin Kohlstedt," the team makes a volumetric recording of a concert, for which the necessary processing steps are performed offline. The pre-processed recording can then be streamed in real time to remote viewers. These viewers may experience a live performance of an artist inside a virtual space through which they are able to move freely. Viewers wearing a VR headset will be able to take a few steps forward, backward, or to the side and turn their heads in different directions. The necessary video data is transmitted to the VR headset, offering viewers what experts refer to as "six degrees of freedom," or 6DoF for short. In this field test, between 50 and 100 viewers shall be able to simultaneously experience the virtual concert, which is complemented by an immersive, spatial audio playback. "We want to use MPEG DASH to send scalable three-dimensional data so that participants can even use smartphones as end devices, for instance," says Steglich. In the long run, the team's aim is to establish a complete live end-to-end volumetric workflow from capture to playback.

Among other technologies, the team plans to incorporate the MPEG-I standard MPEG Immersive video (MIV), which is currently in development, as an alternative way to transmit the required video data. The advantage of MIV is that the decoder comes with a built-in renderer which can then generate any novel view needed. Instead of transmitting the entire 3D model and rendering it on the viewer side as usual, only a rendering from 2D videos would be required. The researchers at Fraunhofer IIS are currently exploring whether the MPEG-I MIV standard is suitable for this type of application and how many virtual cameras would be needed to be placed throughout the virtual scene. This all begs the question: What can we achieve using current technology? Virtual LiVe aims to find the answer.

Project website:
https://www.3it-berlin.de/projects/virtuallive/

Contact

Christian Weissig
Fraunhofer HHI
info@3it-berlin.de



Producing customizable immersive sound with the MPEG-H Authoring Suite 4.0

Immersive audio experiences are on the rise and are becoming increasingly common in private media consumption. The MPEG-H Audio system delivers enveloping sound on every kind of playback device – from home theaters to 3D soundbars to mobile devices. With its unparalleled range of customization options, it also helps producers and broadcasters meet modern audiences' demand for personalized and accessible media reception. It enables users to adapt the content to their personal needs and preferences with fully adjustable dialogue levels, customizable audio description, multiple languages, and even interactive object positioning.

To supports creators in the production of content which includes MPEG-H metadata, so-called MPEG-H Masters, Fraunhofer IIS provides the MPEG-H Authoring Suite (MAS) and recently released an updated version 4.0. It is a set of tools that make the production of MPEG-H Masters easier, faster, and available for everybody. The upgraded version streamlines and accelerates MPEG-H production workflows. It can be used in all areas of the field from radio and TV broadcast to immersive music streaming and supports authoring and export in production environments running at 48 and 96kHz sample rates.

It comprises:

- The MPEG-H Authoring Plug-in (MHAPi) that takes users through all steps of creating object- or channel-based MPEG-H Audio productions inside a VST3- or AAX-enabled Digital Audio Workstation (DAW).
- The MPEG-H Authoring Tool (MHAT) that enables users to create MPEG-H metadata with existing audio material without the need of a Digital Audio Workstation (DAW).
- The MPEG-H Conversion Tool (MCO) that converts MPEG-H compliant content masters between different file formats. It also serves as an interface to the MPEG-H Audio ecosystem and supports the import and export of MPEG-H Master files.
- The MPEG-H Production Format Player (MPF Player) that enables quality control of files containing video and associated MPF audio. It supports a variety of video formats, such as ProRes422 in a MOV or H264/H265 in an MP4 container.

Learn more about MAS 4.0 and download it on https://www.iis.fraunhofer.de/mas

Contact

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

xHE-AAC – test service and new licensees

Millions of hours of music, movies, or TV series are streamed over the internet daily. For the best user experience on any device and in any listening environments, delivering an uninterrupted stream with best possible audio quality is key. xHE-AAC is an audio codec that meets exactly those demands.

xHE-AAC offers maximum coding efficiency with a usable bit rate range that spans from 12 kbit/s to 500 kbit/s and above for stereo services. This helps video and audio streaming providers to overcome audio bandwidth constraints and thereby facilitates an enhanced, more reliable consumer experience. Created with adaptive streaming via DASH or HLS in mind, xHE-AAC delivers transparent quality under good network conditions and, when necessary, seamlessly switches to the bit rates and quality levels that a congested network can support.

xHE-AAC also comes with mandatory MPEG-D loudness and dynamic range control metadata, which can be used on the playback side to achieve a consistent loudness level and optimal dynamic range for any playback device and environment. In a living room environment, a film can be enjoyed with the full dynamic range and on a mobile device in a noisy environment the loudness management ensures intelligible dialog.

Fraunhofer is offering a web-based test service that developers and manufacturers can use to validate their implementations of the xHE-AAC audio codec for compliance with MPEG standards. The service, which is available exclusively at https://test.xhe-aac.com, is free to use upon registration with Fraunhofer IIS and will test both encoders and decoders.

xHE-AAC is included in mobile devices running Android, Fire OS, or Apple operating systems and has also been licensed to Microsoft. Hence it is only logical that the codec is being gradually adopted by social media and streaming services. With Facebook and Netflix, two major players in their respective fields got on board in 2021.

For further information on xHE-AAC, go to https://www.iis.fraunhofer.de/en/ff/amm/broadcast-streaming/xheaac.html

Contact

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

Receive DRM Digital Radio on Android smartphones anytime and anywhere

Starwaves and Fraunhofer partner for an Android app enabling DRM reception on mobile devices. The Starwaves DRM SoftRadio App upgrades mobile devices to receive undistorted DRM Digital Radio anytime and anywhere.

The new SoftRadio App by Starwaves enables Android phones and tablets to receive entertainment, text information, and emergency warnings via DRM Digital Radio – without costly data plans, independent from cell phone network availability, and based on innovative Fraunhofer technology.

The app was developed in close cooperation with Fraunhofer IIS. Its goal is to ensure easy access to innovative DRM radio services for everybody. It is available since summer 2021 in the Google and Amazon Android app stores. The app provides listeners with access to all the essential features of the DRM digital radio standard, across all transmission bands from DRM on long wave to FM band and VHF band-III.

The app also supports more DRM features such as the Emergency Warning Functionality (EWF), image slideshows, station logos, and service descriptions including Unicode support for worldwide application. To provide all these services, the app only requires a standard off-the-shelf SDR RF dongle that is attached to the device's USB port.

In order to meet the needs of everyday radio listeners and to clearly separate this app from the engineering-driven approaches of the past, usability was a primary development objective from day one. With only a few clicks on the clutter-free interface, users select their preferred radio service, navigate through the clearly structured menus, and gain instant access to the various advanced information services that DRM provides. By supporting multiple user interface languages, the app ensures optimized usability in many countries around the globe.

Contact

Alexander Zink
Fraunhofer IIS
audio-info@iis.fraunhofer.de

The next level in video education

Video-based classes and seminars are still a far cry from the real deal. With novel technology developed by Fraunhofer HHI, professors will soon be transported directly into dorm rooms – virtually, that is. And they will even be able to interact with students.

So, the professor will be standing in the middle of the student's room teaching the class? It may sound like an exclusive private tutoring session, but this technique stands to become a reality for students throughout the entire semester. It is all possible thanks to technology developed by the Fraunhofer Heinrich Hertz Institute, or Fraunhofer HHI. "The student wears a mixed reality headset," explains Dr. Cornelius Hellge, head of group at Fraunhofer HHI. "That way, they see both their room and the professor, who appears to be standing in the middle of the space, looking at the student and explaining the material." To bring this vision to life, HHI is working on novel techniques for the generation and streaming of interactive virtual humans

Interactive virtual humans

The first step creating this type of visualization is to produce a volumetric recording of the professor. "To do this, we use 32 cameras arranged in stereo pairs, with 20 centimeters distance between the cameras of each pair," explains Ingo Feldmann, head of group at Fraunhofer HHI. These stereo pairs are distributed evenly throughout the recording space and the recordings are processed into 3-dimensional meshes, creating a volumetric video stream. New hybrid animation techniques, which combine elements from classical computer graphics as well as AI-based methods allow animation of the volumetric representation of the person. Students can not only walk around the professor; they can also maintain "eye contact" at all times. "To enable this additional level of immersion, the volumetric data is dynamically animated to have the professor's gaze follow the student," explains Wieland Morgenstern, researcher in the Computer Vision & Graphics group at Fraunhofer HHI. New AI-based animation techniques will also allow for the animation of gestures and facial expressions and speech that were not part of the original recording. In the coming months, the avatar's lips will be animated as well to match the contents of the text, and within the next year, animation of the entire body could follow, to have the teacher dynamically moving around in the space.

Rendering in the edge cloud

Achieving such a feat of animation requires considerable computing power – more than a smartphone can handle – so the rendering process takes place in the edge cloud, a computer located near the end user.
And for the data to be transmitted to the headset, it must first be compressed.
The raw data is far too large to stream – several gigabytes to be exact.
"We render the volumetric data in the cloud and send it to the headset via 5G as a 2D video stream with standard formats and bitrates," says Dr. Hellge. Simultaneously, the headset sends data about the direction of the user's gaze back to the edge cloud, which positions the render camera based on the movement of the user and moves the head of the avatar to match. Of course, this process has to be fast.
"It's a whole new ballpark. Other streaming services are still working with latencies of around a second. We need less than 60ms," clarifies Dr. Hellge. This technological leap was made possible by combining the Web Real-Time Communications standard, or WebRTC, which is normally used for video conferences, with an extremely fast video encoder and optimizing all components of the streaming chain.
Finally, the mixed reality headset ensures that the 2D image is perfectly integrated in its 3D setting, so that the professor is not seen holding a lecture while standing in the middle of an armchair.

Contact

Maria Ott
Fraunhofer HHI
maria.ott@hhi.fraunhofer.de

Versatile Video Coding is coming of age

A growing number of videos are being streamed and sent electronically. The new Versatile Video Coding (VVC) compression standard is designed to manage this flood of data while offering excellent image quality. There are already open source encoders and decoders available for this new standard.

A child's first steps, a new puppy, scenes from a beach vacation – more and more often now, people are choosing to send videos – rather than photos – to their friends and family. At work, due in part to the COVID-19 pandemic, video conferences have replaced many phone calls and in-person meetings, while at home, people often elect to spend their evenings on the sofa streaming their favorite series rather than submitting to predefined television broadcasting schedules. And, of course, videos are a fundamental part of the gaming world, too. Managing this mounting tide of video data requires a continuous supply of new compression technology.

The new, international VVC video standard

The latest international video data compression standard is called H.266 or VVC, which is short for Versatile Video Coding, and Fraunhofer HHI played an integral role in its development. VVC requires only half the bit rate of its predecessor to achieve the same level of quality. The new standard is particularly important for ultrahigh resolution 8K videos and 360-degree videos in virtual reality applications. In fact, VVC yields a wide array of opportunities for many kinds of applications.

Encoders and decoders as open source software

Take cloud video streaming and video processing, for example. When you are streaming a video, your video normally stops playing if the network no longer has the necessary bandwidth to accommodate the stream. With "adaptive streaming," on the other hand, your video continues to play with reduced image quality. To allow for adaptive streaming, videos in the cloud are encoded and uploaded in different bit rates and resolutions, and the best version is selected and streamed based on the end user's current available bandwidth. This is where the new standard truly shines. To pave the way for VVC to enter the media landscape, Fraunhofer HHI developed the VVenC open source encoder, which is optimized specifically for cloud applications. The VVenC encoder already won an Open Source Award at the IEEE International Conference on Multimedia and Expo (ICME) 2021. It is currently being tested by the German public broadcaster WDR and the multimedia technology company Bitmovin. In short, just 18 months after the standard was developed, the technology needed for VVC encoding is already available on the market.

We are honored Fraunhofer Heinrich Hertz Institute has chosen to work closely with Bitmovin to integrate the new VVC codec into our Bitmovin Encoding solution," said Stefan Lederer, CEO Bitmovin. «

As for decoding, the first chips designed to be installed in end devices are scheduled for market release by the end of 2021. Until this hardware solution becomes widely available, the VVdeC developed by Fraunhofer HHI serves as a software-based VVC decoding solution. The VVdeC source code is available for free and can be integrated into web browsers and other applications.

Contact

Maria Ott
Fraunhofer HHI
maria.ott@hhi.fraunhofer.de

What you should know about light-field

A light-field is defined by the light rays within a specific area. If all light rays are captured within a scene, it is possible to generate a perspective from any position. This is a major advantage when editing scenes during the postproduction process. Researchers in the Moving Picture Technologies department at Fraunhofer IIS are already demonstrating how scenes can be optimized using a set-up with professional cameras in a planar arrangement that capture slightly different scene perspectives. This opens up new creative opportunities. In a recent interview Joachim Keinert, head of the "Computational Imaging" group, shares his insights.

What are the differences between classic media production and light-field production?

Joachim Keinert: It depends on what is meant by classic media production. Considering how many videos are produced today using compositing, 3D modelling or even volumetric capturing where you can have quite sophisticated virtual camera movements that are close to VR experiences. Light-field captures a scene in such a way that you can virtually move it afterwards and add the photorealistic effects corresponding to the real situation on-site, not as an assumption of the scenic conditions. For example if you want to capture the shininess of your surfaces or specular highlights to really makes the scene or object feel realistic.

What are the key facts and what makes this technology so appealing for media production?

Joachim Keinert: To avoid confusion, there are two kinds of light-fields. There are light-field cameras based on micro lenses inside the camera. It gives you the option to refocus but they don´t give you a lot of perspective change. So, when we talk about light-field we really mean several perspectives and several cameras. For video this means that you take multiple cameras and mount them on a rig. For static scenes you can take a single camera and capture your scenes sequentially. Then, the next step is to take all the data you processed – this is where our expertise comes into play – to recover geometry or colors. Once you are done with these steps you get the great benefit that you have recorded your scene as it is with high photorealism for good memorization.

In which part of the media production workflow does light-field open new possibilities?

Joachim Keinert: Light-field can do the most in an application where you really want to reproduce a scene as it is and don´t want to change it too much. Let´s take an example: Imagine you have a nice apartment for sale and you want to reach as many interested parties as possible. You want to give them the possibility to tour it beforehand in a virtual way. For such an application it is important that you stick as close as possible to the original to not disappoint your customers because your 3D model did not reflect the true situation. You don´t want to add or change anything of the real-life place. With the use of light-fields, you really want to reproduce your apartment as it is.

What are the challenges and opportunities?

Joachim Keinert: A challenge definitively is the large number of views or captures you have to take and the large amount of data that is coming along with the process. But once you are through with this, it gives you the opportunity to memorize and preserve previous scenes or objects for later on. Because you can virtually watch them from different perspectives and you have a much better realistic impression compared to alternative technologies that are available .

Can you give us a number of cameras for light-field recording?

Joachim Keinert: This obviously depends on the scene complexity. For video it´s much smaller than for static scenes. To give you an order of magnitude for static scenes: It´s similar to photogrammetry. In photogrammetry you need hundreds or thousands of captures and for light-field it´s similar. But compared to photogrammetry you are able to recover a better color impression by looking to the shininess of your surface and on small reflections. This is how light-field tops it off.

Would you rate light-field as a top technology?

Joachim Keinert: Definitively. When you want to reproduce the scene as it is, you need to observe your objects from different perspectives and this is exactly where light-field is exceptional.

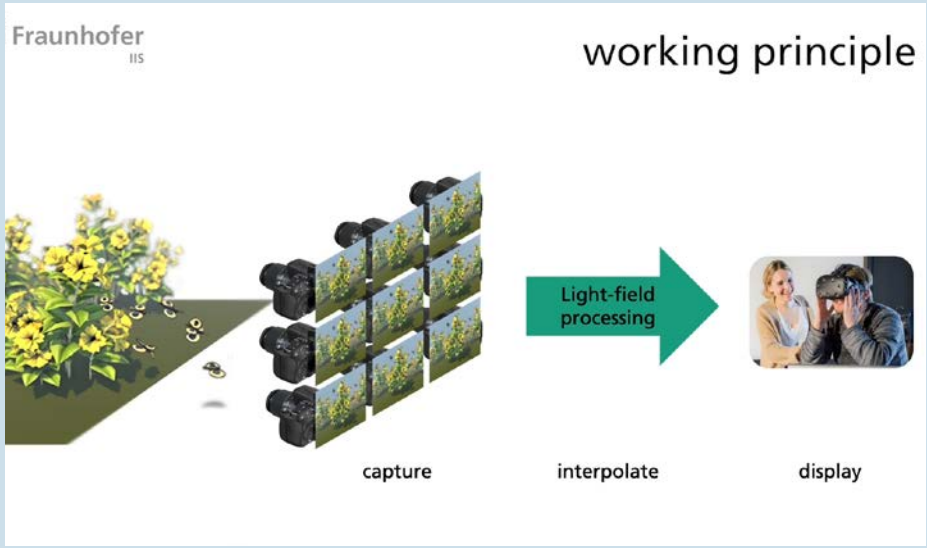
Joachim, what is your approach to light-field?

Joachim Keinert: My team started with video arrays for light-field. Today, we are concentrating more on static scenes because it is possible to capture all the views you need with this sort of set-up. Our current focus is on reproducing a scene in a faithful way. In the future, we could of course imagine using also video or taking all the data t we have captured, for example all the reflections. This would allow us to estimate material properties and once you know what kind of material your object is composed of, you can relight it much better. This would obviously be a big step forward.

In comparison to 360 degree virtual walk-arounds, 3D navigation or 3D renderings, what is exceptional about light-field?

Joachim Keinert: Let´s start with the 360 degree photos or videos because they also capture reality. It´s a very good comparison because they have a similar goal but they lack in capturing geometry. You don´t know the distances of your object instead you assume that everything is located on the sphere situated around you and if you move you get distortions because of the lacking distances. Light-field follows the same goal - reproduction of reality - but it gives the correct geometric perspectives. Another technique that we should mention are game engines. You can move around in amazing worlds but you have to create your photorealistic models first which is a lot of work and requires a lot of expertise, in particular, if you want to reproduce reality. Light-field technology gives you the advantage to cover all the different perspectives so the user can choose which one he wants to have and of course different users can choose different perspectives.

Thank you for this interview.



Contact

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

Audio Mining & Live-Subtitling

TV and Radio Archives Searched in a Flash. Need to find a specific original quote in the radio or television archive? Until now, this has been difficult. A smart Fraunhofer speech technology transcribes every broadcast and delivers the desired broadcasts and time codes in mere seconds.

Where and when did Joe Biden utter his now-famous words "Build Back Better"? Finding these and similar original recordings from radio or video is a tedious business for journalists and editors: it is estimated that only ten percent of broadcasters' archive material includes detailed, manually inserted annotations, and only if the archivist considered them important at the time.

In the case of all other materials, one has only the title information, which permits little conclusion on the specific contents.

Find the right original sound clip in just one click …

The Fraunhofer IAIS Audio Mining system makes this procedure much easier. "If a radio or television recording is archived, our tool uses deep learning to transcribe any spoken language into a text," explains Dr. Christoph Schmidt, Head of Business Unit "Speech Technologies" at Fraunhofer IAIS. "Every broadcast is thus available as a text file in which individual search terms can be found in fractions of a second. For each word, the time markers are also stored in the broadcast – so you can mark the desired position in the text and cut out the audio snippet you are looking for."

For the editors this means that if, for example, they are searching for the original sound clip of Joe Biden's statement, they can enter the right words in the search field of the user screen and receive a list of all broadcasts including the exact time in which this excerpt can be heard. The system automatically segments broadcasts according to the speaking persons; the researchers call this "speaker clustering." Here, the persons speaking in a broadcast are numbered consecutively – if you have listened briefly to which speaker belongs to which digit, you can choose only to listen to the answers of the person being interviewed, for example.

With the speaker recognition function, the scientists even go one step further: the system recognizes the exact speaker, for example a certain politician. It is therefore able to answer more complex queries such as: "Joe Biden´s statements on the Coronavirus disease" – or can jump to the contribution of a specific person in a talkshow with a click. This involves teaching the computer how different people sound. With speech snippets of one to two seconds this is still quite difficult, but given a talk time of 30 seconds, speaker recognition already works very reliably."

This is a great way to search through huge archives," says Schmidt. A second usage scenario involves interviews or other recordings that are directly transcribed live as this facilitates the production of programs. The Fraunhofer IAIS live recognition system is already in use at the Saxon state parliament; other state parliaments in Germany and Austria have already shown interest.

The usage in regional parliaments has also shown that the system is robust against dialectal language. In the medium term, it is also conceivable that the tool could be used in real time for automatic subtitling of television or video programs. While this already works quite well for the news, for speakers with strong dialects or accents, or for rare technical terms – such as those used in astrophysics – some research work is necessary to reliably recognize all words. Already in use at ARD and ZDF The Fraunhofer IAIS Audio Mining tool analyzes 2000 hours of audio and video material daily for Germany's ARD public broadcasting stations. It is also in use at the broadcaster ZDF. The system is currently mainly used in the context of archives. "It is conceivable that, in the medium term, stations will not only use it for their archives, but also for their media libraries and automatic subtitling and for working with raw material in the editorial offices," Schmidt explains.

Contact

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

Mining Platform: Automated Multimedia Content Analysis at Scale

The Fraunhofer IAIS Mining Platform is a Microservice-based system for automatically extracting insights from text, audio, image, and video at scale. A large number of AI-based services for analyzing the different modalities have already been integrated: 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.

Thanks to its modular structure, additional services can easily be integrated, and AI-based services for object recognition, concept detection, cut and scene detection, and semantic keyframe extraction are currently under development. All services come with readymade models trained on openly available datasets, but it is also possible to train and use customer-specific models, an advantage not offered by other systems so far.

The individual steps for analyzing media files 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 files depending on urgency. For example, Mining Platform for the daily business of news editors and journalists.

The workflow component also allows modeling more complex media analysis workflows. For example, a video could be processed using face recognition (and other visual analysis services), while Audio Mining would transcribe its speech content and recognize the speakers. The extracted insights–the persons recognized in the video and the detected speakers–could then be linked together, which would allow to detect when a person is shown in the video while also talking. The transcript could then be further analyzed using text mining services to identify topics, as well as people, places, or institutions that are mentioned.

Future Developments

To focus our developments and to benefit our customers, the Audio Mining components will be integrated into our Mining Platform. As the Mining Platform can be customized to need, such an integrated platform can serve both customers interested in automatic speech recognition services as well as customers in need of a versatile media analysis platform.

The availability of video mining services will facilitate the development of more complex analysis workflows, which allow to comprehensively analyze videos, and combine the generated insights to gain a more extensive and deeper understanding of videos. In the field of speech recognition, we plan to extend our systems to other European languages and support more dialects and accents.

Contact

Dr. Christoph Andreas Schmidt
Fraunhofer IAIS
christoph.andreas.schmidt@iais.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

Publication Information

Fraunhofer Business Area Digital Media
c/o Fraunhofer Institute for
Integrated Circuits IIS
Am Wolfsmantel 33
91058 Erlangen, Germany

Concept and Publisher
Angela Raguse
Fraunhofer Business Area Digital Media

Editors
Dr. Janine van Ackeren
Falk Jäger
Angela Raguse
Florian Meister

Translation: Saskia McDonagh

Layout and production
Ariane Ritter

Photo acknowledgements
Cover picture: © stock.adobe.com
Page 2: Anja Menge/HFF München
Page 5-7: Paul Hahn, Fraunhofer FOKUS
Page 9: Fraunhofer IIS
Page 11: © Tierney, Adobe Stock/Fraunhofer IIS
Page 13: © Starwaves/Fraunhofer IIS
Page 17-19: Fraunhofer FOKUS
Page 23: Fraunhofer IIS
Page 25: © stock.adobe.com

Contact

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