Call for Proposals

Media Technology

HIRP OPEN 2017

Application Deadline: 09:00 A.M., 16th June, 2017 (Beijing Standard Time, GMT+8).

If you have any questions or suggestions about HIRP OPEN 2017, please send Email (innovation@huawei.com). We will reply as soon as possible.

Catalog

HIRPO2017070301: High-precision sound source detection and separation

1 Theme: Media Technology

2 Subject: Human-Machine Speech Interaction

3 Background

Automatic Speech Recognition has been widely researched and deployed in the last decade. Good recognition rate has been achieved in recent years with the rapid growth of deep learning technologies. However, this is only limited to use cases where speakers are close to microphone(s). When speakers are further away from the microphone(s), speech recognition performance will be significantly degraded due to environmental sound sources which would interfere with the desired speech. Therefore, it is extremely important that one can monitor the sound field, detect the sound sources and separate one source from the others with high precision. So each source with high fidelity could then be fed into recognition module to maintain a high recognition rate.

4 Scope

1) Research on high-precision sound source localization: develop new and improved technologies for identifying the number of sound sources in the surrounding of microphone(s) and their respective locations. The location of the sound source could be used to facilitate the sound source separation;

2) Research on high-precision sound source separation: develop new and improved technologies for separating one sound source from others with high fidelity. Preferably, the technology would be based on a microphone array. An important application of this technology would be to separate each speaker

from several concurrent speakers and understand each speaker with high recognition rate. If a microphone array is used, the size of the array and the number of microphones should be commercially friendly targeting for household devices.

5 Expected Outcome and Deliverables

Microphone array hardware (if a microphone array is used);

C code of the algorithms;

Technical reports describing the algorithm developed and the evaluation methods and results;

6 Acceptance Criteria

Project proposal is accepted by the evaluation team, Huawei.
Project deliverables are accepted by the evaluation team, Huawei.
Performance of the algorithm meet the requirement agreed among Huawei and the applicant.

7 Phased Project Plan

Phase1 (~1 month): survey the state of the art of sound source localization and separation technologies, outputs the related technical report.
Phase2 (~3 months): research on high-precision sound source localization, outputs the related deliverables.
Phase3 (~8 months): research on high-precision sound source separation, outputs the related deliverables.

Click here to back to the Top Page

HIRPO2017070302: Binaural Room Impulse Response (BRIR) characterization and modeling

1 Theme: Media Technology

2 Subject: Virtual Reality Audio

List of Abbreviations

AR: Augmented Reality

BRIR: Binaural Room Impulse Response

VR: Virtual Reality

3 Background

In Virtual Reality (VR) and Augmented Reality (AR) audio reproduction, the acoustic nature of the virtual, or rendered, room and environment that the user is in should be mimicked as faithfully as possible in the sounds that are heard by the user.

It is well understood that as sound sources move and the user moves throughout a virtual, or rendered, space, the acoustics will alter due to e.g. shadowing behind physical or virtual objects or obstructions in the room, proximity to reflective or absorbent structures in the room or simply due to the relative reflectivity or absorbency of the walls, ceiling or floor of the room producing reverberation. In binaural rendering of 3D audio these phenomena are often characterized by the Binaural Room Impulse Response (BRIR).

This project aims to characterize rooms and objects in order to derive BRIR's that may be applied in VR and AR 3D audio reproduction. Ideally a method to construct BRIR's from individually reconfigurable components is envisaged.

4 Scope

1) Research BRIR Human Perception: investigate the human perception of different BRIR representations and establish how well and what features of the BRIR's are required to achieve perceptual equivalence;

2) Research on Parameterization of BRIR's: investigate ways in which BRIR's may be parameterized for different features, rooms and locations to achieve perceptual equivalence.

5 Expected Outcome and Deliverables

Technical reports of BRIR's characterization and modeling;

The algorithm and code of human perception of the accuracy of BRIR's for transparency or equivalence;

6 Acceptance Criteria

Technical report describes BRIR's characterization and modeling clearly.
There is no bug of the new algorithm code and the algorithm has good immersive perception to most of the listeners.

7 Phased Project Plan

Phase1 (~2 months): Survey the state of the art of BRIR characterization and provide a related technical report.
Phase2 (~6 months): Conduct experiments on the human perception of different BRIR representations and research how well and what features of the BRIR's are required to achieve perceptual equivalence and provide a related technical report.

Phase3 (~4 months): The algorithm and code about parameterization for different features, rooms and locations to achieve perceptual equivalence.

Click here to back to the Top Page

HIRPO2017070303: Effective flexible VR audio capturing algorithm

1 Theme: Media Technology

2 Subject: VR Audio Capturing System

3 Background

Currently, the structure of VR audio capturing device is almost fixed, it is spherical harmonic structure, or TETRA structure and etc., the structures are all integrated and fixed, so they cannot be used to some mobile devices flexibly.

More and more people want to capture panoramic audio and video by the portable devices, but the shape of the portable devices may be not the same, such as, Smartphone, hexahedral or spherical VR camera, or HMD devices and etc, obviously, the shape of them are different. So what is the minimal number of the Microphone and what is the best structure of the Microphones in the different devices, meanwhile, how to model and parameterize the flexible structures of the Microphones, so that the portable devices can capture the entire sound field, and can output the standardized FOA or HOA coefficients.

4 Scope

1) **The minimal number of Microphones:** what is the minimal number of the microphones based on the flexible structures, so that the entire sound filed can be captured well.

2) **Modeling the flexible Microphone structures:** modeling and parameterizing the flexible microphone structures, so that the standardized FOA or HOA coefficients can be outputted.

5 Expected Outcome and Deliverables

1. Technical report of the minimal number, the flexible structure of the microphones, and modeling and parameterization of the flexible structure of the microphones.
2. The algorithm and code of the modeling, and the front/back and height information and left/right can be perceptive well when the outputted FOA or HOA coefficients of the algorithm are inputted an existing rendering algorithm.

6 Acceptance Criteria

The technical report can describe the number and the flexible structure of the microphones, and the algorithm in detail clearly.

The front/back and height information and left/right can be perceptive well when the outputted FOA or HOA coefficients of the algorithm are inputted an existing rendering algorithm.

7 Phased Project Plan

Phase1 (~2 months): Technical report, describes the current status of the flexible Microphone structure.

Phase2 (~6 months): the basic modeling of the flexible microphone structure, the basic performance of the algorithm can be achieved.

Phase3 (~4 months): The optimized algorithm, the performance can be improved further.

Click here to back to the Top Page

HIRPO2017070304: Dynamic 3D face reconstruction on mobile

1 Theme: Media Technology

2 Subject: 3D reconstruction

3 Background

Obtaining a user-specific 3D face surface model is useful for a variety of applications such as face recognition, video editing, avatar and more. In the past two decades, methods for acquiring 3D facial performances can be categorized to multi-view stereo, photometric stereo ， structured light based approaches and 3D Model based method.

One approach to capturing 3D dynamic faces is multi-view image-based 3D model reconstruction, which captures the 3D geometry of a face and produces fine-scale facial details, but the recovered geometry could be scale-ambiguous and not metrically correct. More importantly, this kind of system is limited to capturing static facial geometry.

Photometric stereo-based approaches jointly estimate the surface normal, albedo, lighting conditions, and pose angles. It aims to identify a single representative face from the entire collection, which is challenging given the expression variation among images. However, there are still major limitations in photometric stereo-based reconstruction. One is that they require a sufficiently large collection of photos for reconstruction. Theoretically, only few images are necessary if they are in perfect correspondence, but in practice the approaches use over one hundred images.

Structured light based approaches are capable of capturing 3D models of dynamic faces in real time, and become popular recently with the rise of

consumer depth cameras (such as the Microsoft Kinect and Asus Xtion). Some researchers capture dynamic depth maps and fit a smooth template to the captured depth maps using embedded deformation techniques. However, structured light sensors cannot match the spatial resolution of static face scans or the acquisition speed of marker-based systems. Recently some non-rigid object reconstruction systems appear, for example, Dynamic Fusion. Since the sensors are capable of delivering depth maps at real-time rates, a particular focus of recent systems is to perform online surface reconstruction.

To the end of dynamic face surface reconstruction, we consider it a challenging process that varies significantly depending on the nature of input (expression, accessories, etc.), output (mesh, volume, etc.), hardware platforms (mobile, pc, etc.) and types of background environment (indoor, outdoor, etc.). Specifically, we are looking for the solution to dynamic 3D face reconstruction on mobile phones. As the new mobile phones are equipped with high-resolution stereo cameras and high-performance processors, it is possible to dynamically reconstruct human face on mobile devices.

4 Scope

1) Research on high quality 3D face reconstruction on mobile device.

2) Research on high quality dynamic face reconstruction on mobile device.

5 Expected Outcome and Deliverables

Technical reports of dynamic 3d face reconstruction on mobile phone.

3D face reconstruction on mobile phone

3D dynamic face reconstruction on mobile phone

1~2 Invention/patents;

6 Acceptance Criteria

Project proposal is accepted by the evaluation team, Huawei.

Project deliverables are accepted by the evaluation team, Huawei.

The average of Distance of error is less than 0.05mm

7 Phased Project Plan

Phase1 (~3 months): Survey the state-of-the-art works in 3D face recognition. Define system setup and propose solution to the project. Give the baseline system.

Phase2 (~5 months): motion and geometry combination non-rigid registration and multi-scale face reconstruction and Shading-based enhancement of face enhancement, High details of face enhancement through learning method.

Phase3 (~5 months): face expression and animation demo.

Click here to back to the Top Page

HIRPO2017070305: Neural network based representation of a visual content for efficient image/video compression

1 Theme: Media Technology

2 Subject: Video Compression Technology

3 Background

The essential problem of image/video compression is to get a compact description of a visual content representation at a given quality level at the encoder side, and reconstruct the visual content at the decoder side. Advantages of neural networks have been shown for visual content recognition and representation. Preliminary research reveals the possibility of achieving better image quality compared with the state-of-the-art image/video) compression standard in low bitrate scenario [1][2]. Especially, previous work has shown that the deep network models can have the ability of generating scalable representation of images, i.e. the generated bitstream can be truncated (to a certain degree), partial bitstream is able to decode and reconstructs images with lower quality.

It is a challenging topic to build a generic neural network model that can be used for efficient representation of image and video signal at different quality level, and design an image/video codec based on the model for large-scale deployment in practical image/video compression and communication systems for various applications.

[1] Gregor K, Besse F, Rezende D J, et al. Towards conceptual compression. NIPS 2016: 3549-3557

[2] Toderici G, Vincent D, Johnston N, et al. Full Resolution Image Compression with Recurrent Neural Networks. arXiv:1608.05148, 2016

4 Scope

1) Neural network model for efficient representation of image/video content: investigate neural network models that could be potentially used for image/video compression, and tune the parameters of the model for efficient representation of image/video content. The model must be able to produce representations at a wide range of quality levels.

2) Image/video codecs based on neural network models: based on the above neural network models, design and implement a codec for image/video compression. Common functionalities for a codec such as rate/quality control and parallel processing are required.

5 Expected Outcome and Deliverables

- A survey document on neural network based image and video compression
- A technical document with an initial design of a proposed model and the corresponding image/video codec
- A neural network model for efficient representation of image/video content
  - Software implementation of the model
  - A technical document describing the model in detail, including model structure, training, and testing
- An image/video codec based on the neural network model
  - Software implementation and testing of the codec
  - A technical document for describing the codec in detail, including encoder operations, decoder operations, and the structure of the compressed image/video stream

6 Acceptance Criteria

- Survey documents
  - Covers all related techniques and related public resources since 2015
- Technical documents
  - Has details that can be used to reproduce the technique by an engineer in this field
- Software implementation of the proposed neural network based codec
  - Can be used to reproduce the simulation data in the technical document
- Proposed neural network model
  - Has advantage of over existing models
- Proposed image/video codec based on the proposed model
  - Has comparable compression performance with the best of the existing codecs of the same kind
  - Has potential better compression efficiency over H.265/HEVC

7 Phased Project Plan

Phase1 (~2 months): survey the state-of-the-art of neural network model based image and video compression designs, and finish an initial design of the proposed model(s) and the image (and optionally video) codec.

Phase2 (~4 months): produce a detailed design and finish the SW implementation of the proposed model(s) for image (and optionally video) compression, advantage of the proposed model(s) over existing models being verified. One patent application is filed by the end of this phase.

Phase3 (~6 months): produce a detailed design and finish the SW implementation of the image (and optionally video) codec using the proposed model(s), comparative performance evaluation of the proposed codec with both existing codec of this kind being conducted and the state-of-the-art image

(and optionally video) compression standard, potential better compression efficiency being verified. One patent application is filed by the end of this phase.

Click here to back to the Top Page

HIRPO2017070401: Punctuation prediction for Automatic Speech Recognition

1 Theme: Media Technology

2 Subject: Punctuation prediction for Automatic Speech Recognition

3 Background

Punctuation plays an important role in the understanding of the text semantically. However, the automatic speech recognition system is usually lack of the ability to predict or insert punctuation. Without punctuation, it will be very difficult for users to understand the text of speech recognition results. And it is also difficult to do further natural language processing. Therefore, it is very important to add the punctuation information in the speech recognition result accurately.

4 Scope

Problem to be resolved:

Break sentences and add punctuation (,.?!) in the results of speech recognition.

Improve the user experience of the online speech recognition system.

Provide the necessary semantic segmentation and emotion information for subsequent natural language processing.

5 Expected Outcome and Deliverables

The Expected outcome of this project is a set of algorithms which can be run in

Smartphone to predict the punctuation of ASR. It includes:

Detailed technical report.

C/C++ source code.

Possible patents

6 Acceptance Criteria

The prediction accuracy (F1-measure) of the algorithm should be more than 80%.

7 Phased Project Plan

Expected project Duration (year): one year

Phase1 (~6 months): Work plan for Stage 1: develop a PC version of hot words update method.

Phase2 (~4 months): Work plan for Stage 2: Convert the PC version of the algorithm to ARM version which keeps the accuracy.

Phase3 (~2 months): Work plan for Stage3: optimize the performance in order to meet the product requirement.

Click here to back to the Top Page

HIRPO2017070402: Hot word update for language model

1 Theme: Media Technology

2 Subject: Hot word update for language model

3 Background

In recent years, with the rapid growth of the Internet, people have more opportunities to express their views, feelings, experiences and stories if they can easily access the Internet. The evolution of the language has been greatly promoted by the convenience of the Internet. New expressions and words have been constantly created, and quickly spread all over the world as the hot words. But boost of new words and expressions bring a lot of challenges to the automatic speech recognition system.

The speech recognition dictionary and language models are trained from the existing large-scale language corpus. Therefore, in general, the dictionary or language model can easily cover the existing language phenomenon. But for the new emerging words and expressions, the performance becomes worse. We have to design a new way to update these dictionaries or language models in order to include the hot words. The new method should take the frequent occurrence of new hot words into account, and update the dictionary and language automatically and timely.

4 Scope

Problem to be resolved:

Automatically update dictionaries and language models when hot words are known.

Correcting the word probability by the hot words without completely re-training the language model.

5 Expected Outcome and Deliverables

The Expected outcome of this project is a set of algorithms which can be run in Smartphone to update the hot words. It includes:

Detailed technical report.

C/C++ source code.

Possible patents

6 Acceptance Criteria

The accuracy of the recognition of hot words should be enhanced by more than 10%.

7 Phased Project Plan

Expected project Duration (year): one year

Phase1 (~6 months): Work plan for Stage 1: develop a PC version of hot words update method.

Phase2 (~4 months): Work plan for Stage 2: Convert the PC version of the algorithm to ARM version which keep the accuracy.

Phase3 (~2 months): Work plan for Stage3: optimize the performance in order to meet the product requirement.

Click here to back to the Top Page

HIRPO2017070403: Sensor fusion based new innovative applications for Smartphone

1 Theme: Media Technology

2 Subject: Sensor fusion Innovations for Smartphone

3 Background

Smartphone is going to be more functions and miscellaneous shapes, which looks much more intelligent and fashionable, but one the other hand, brings more challenges to the sensors technologies and applications.

Convenient and accurate sensing between user and device become the most important factor to good user experience.

Innovative sensor technologies can be a good way to make better sensing application and support new ID design of future smartphones.

As more and more sensors deployed in the smartphone, combining different sensors and any other useful information in the phone to refine new innovative applications will bring better user experience.

4 Scope

The scope of this project includes but not limited to:

Prototypes of new sensor and model for the smartphones.

Implementation of sensor and any other information fusion technologies to aware the user behaviors.

5 Expected Outcome and Deliverables

One patent;

One technical report with respect to the design and implementation;

Prototype for the proposed methods.

6 Acceptance Criteria

Fail: No patents /prototype are delivered.

Pass: 1 patents pass Huawei's review AND 1 detailed technical report AND corresponding prototype

Excellent: More than 1 patents are delivered.

7 Phased Project Plan

Phase1 (~4 months): deliver a patent and a survey on Sensor Innovations for Smart phone with curved or full screen.

Phase2 (~4 months): deliver first version of prototype.

Phase3 (~4 months): deliver optimized prototype and one technical report

Click here to back to the Top Page