INTERNATIONAL TELECOMMUNICATION UNION
TELECOMMUNICATION STANDARDIZATION SECTOR
STUDY PERIOD 2022-2024

SG16-TD240/WP3
STUDY GROUP 16
Original: English

Question(s): 8/16 Rennes, 15-26 April 2024

TD

Source: Editor H.ILE-3DINR

Title: H.ILE-3DINR "Framework and requirements of 3D reconstruction systems based on implicit neural representation for immersive live experience (ILE) services" (New): Initial draft (Rennes, 15-26 April 2024)

Contact: Hang Zou
China Telecom, China
Tel: +86 18811783596
E-mail: zouh3@chinatelecom.cn

Contact: Zekun Wang
China Telecom, China
Tel: +86 18964957670
E-mail: wangzk2@chinatelecom.cn

Contact: Qing Liu
China Telecom, China
Tel: +86 18910171773
E-mail: liuqing6@chinatelecom.cn

Contact: Xinhui Ding
CAICT (MIIT), China
Tel: +86 18310905691
E-mail: dingxinhui@caict.ac.cn

Abstract: This TD contains the initial text of draft new Recommendation ITU-T H.ILE-3DINR " Framework and requirements of 3D reconstruction systems based on implicit neural representation for immersive live experience (ILE) services" based on the result of the discussion on the contribution SG16-C592-R1 in Q8/16 sessions, Rennes, 15-26 April 2024.

NOTE – Editorial adjustments performed by TSB, including in the title. Issues needing editor's attention are identified in WinWord comments. Editors are requested to carefully read the Author's Guide for drafting ITU-T Recommendations, https://www.itu.int/oth/T0A0F000004/en.

This initial new draft Recommendation is based on the discussion of contribution SG16-C592-R1 at this SG16 meeting Q8/16 sessions, Rennes, 15-26 April 2024.

DOC	Source	Title	Agreements
SG16-C592-R1	China Telecom, MIIT (China)	New: H.3D-INR: Proposed new work item on "Framework and Requirements of 3D Reconstruction Systems Based on Implicit Neural Representation"	The contribution is accepted with modifications as follows: 1. Change the title of this new work item to H.ILE-3DINR: "Framework and requirements of 3D reconstruction systems based on implicit neural representation for ILE service". 2. Add a new clause 6 and Figure 1 to provide the overall framework of the 3D reconstruction system support ILE service. 3. Modify the summary, scope, Figure 2, and some content to focus on the ILE service support. 4. Add sub-clauses of clause 7 to separate real-time and non-real-time 3D-INR systems to support ILE service. 5. Add more real-time use cases of 3D-INR for ILE service in Appendix I.

CONTENTS

Page

1 Scope 4

2 References 4

3 Definitions 5
3.1 Terms defined elsewhere 5
3.2 Terms defined in this Recommendation 5

4 Abbreviations and acronyms 5

5 Conventions 5

6 Overview of 3D reconstruction system for ILE service 6

7 Overall Framework of 3D reconstruction system based on INR for ILE services 6
7.1 General concept and overview 6
7.2.1 Real-time 3D-INR system for ILE service 7
7.2.2 Non-real-time 3D-INR system for ILE service 7
7.2 Data Acquisition 7
7.3 Data Processing 7
7.4 3D Reconstruction 7
7.5 Rendering Processing 7

8 Requirements of 3D Reconstruction System Based on INR for ILE Service 7
8.1 Data type 7
8.2 Data transmission 7
8.3 Rendering processing 7
8.4 Support for ILE Services 7

Appendix I Use cases of 3D reconstruction system based on implicit neural representation for ILE service 9
I.1 Real-time 3D-INR system for ILE service 9
I.2 Non-real-time 3D-INR system for ILE service 10

Bibliography 12

List of Tables

Page

No table of figures entries found.

List of Figures

Page

Figure 1 – Overview of the 3D reconstruction system for ILE service 6
Figure 2 – Overview of the 3D reconstruction with INR-based system for ILE 6
Figure I.1 – Real-time RE-NeRF support ILE service with VR device 9
Figure I.2 – Real-time Instant-NGP support ILE service with VR device 10
Figure I.1 – Fox scene rendered by NeRF with Unity engine. [ImmersiveNeRF] 10
Figure I.3 – INR-based MobileR2L interaction on mobile devices [Realtime_IEEE] 11

Draft New Recommendation ITU-T H.ILE-3DINR

Framework and requirements of 3D reconstruction systems based on implicit neural representation for immersive live experience (ILE) services

Summary

The live 360° virtual reality service of single point view, multi-angle viewing service, and live streaming service in inaccessible areas in H.430.3 of ILE service are all based on the 3D reconstruction technology. These services could need 3D reconstruction of the object or scene. Solutions based on implicit representation include, but are not limited to, real time rendering, the usage of normal camera, without usage of specific 3D data format or other specific equipment, such as depth camera and Lidar. As one of the technologies to realize these services, INR has the value to enhance quality of rendering image and performance of 3D construction processes.

This Recommendation provides the framework and requirements of 3D reconstruction system based on implicit neural representation for ILE service. The Recommendation describes the overview of the 3D-INR system and provides the framework and requirements for supporting the ILE service with 3D reconstruction needs, and the use cases are provided in ~~the~~ Appendix I.

Commented [TSB1]: Editor: be specific in the cross-references.

Keywords

3D reconstruction; implicit neural representation; requirements

1 Scope

This Recommendation identifies the framework and requirements of 3D reconstruction system based on implicit neural representation for ILE service of live 360° virtual reality service of single point view, multi-angle viewing service, and live streaming service in inaccessible areas. The use cases are provided in the Appendix.

The scope of this Recommendation includes:

– The overview of the 3D-INR system.

– The framework and requirements of the 3D-INR system to support ILE service with 3D reconstruction including live 360° virtual reality service of single point view, multi-angle viewing service, and live streaming service in inaccessible areas.

The use cases are provided in Appendix I.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T H.430.1] Recommendation ITU-T H.430.1 (2018), *Requirements for immersive live experience (ILE) services*.

[ITU-T H.430.3] Recommendation ITU-T H.430.3 (2023), *Service scenario of immersive live experience (ILE)*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 Immersive Live Experience (ILE) [ITU-T H.430.1]: A shared viewing experience that stimulates emotions within audiences at both the event site and the remote sites, as if the audience at the remote sites had wandered into substantial event venue and had actually watched the events taking place in front of them. This impression is due to high-realistic sensations provided by a combination of multimedia technologies such as sensorial information acquisition, media processing, media transport, media synchronization and media presentation.

[TBD]

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

[TBD]

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ILE	Immersive Live Experience
INR	Implicit Neural Representation
NeRF	Neural Radiance Fields
VR	Virtual Reality

[TBD]

5 Conventions

In this Recommendation:

– The keywords "**is required**" indicate a requirement which must be strictly followed and from which no deviation is permitted if conformance to this Recommendation is to be claimed.

– The keywords "**is recommended**" indicate a requirement which is recommended but which is not absolutely required. Thus, this requirement needs not be present to claim conformance.

– The keywords "**can optionally**" indicate an optional requirement which is permissible, without implying any sense of being recommended. These terms are not intended to imply that the vendor's implementation must provide the option and the feature can be optionally enabled by the network operator/service provider. Rather, it means the vendor may optionally provide the feature and still claim conformance with the specification.

– The keywords "**is prohibited from**" indicate a requirement which must be strictly followed and from which no deviation is permitted, if conformance to this Recommendation is to be claimed.

– The keywords "**is not recommended**" indicate a requirement which is not recommended but which is not specifically prohibited. Thus, conformance with this Recommendation can still be claimed even if this requirement is present.

6 Overview of 3D reconstruction system for ILE service

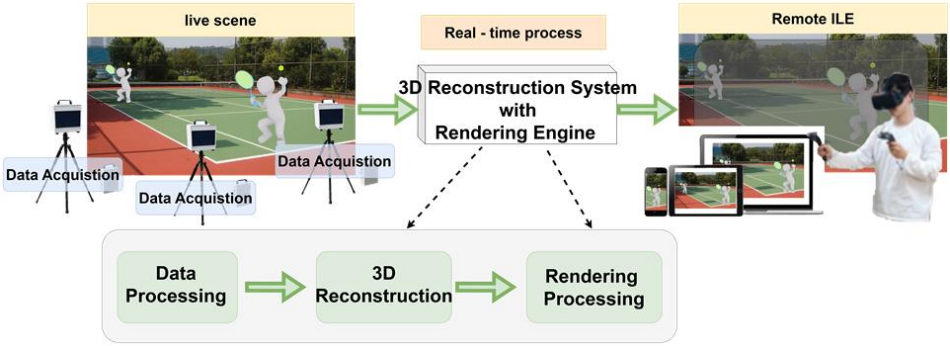

Figure 1 – Overview of the 3D reconstruction system for ILE service

[Editor notes: this is the overview of the 3D reconstruction system support ILE service. It includes three main parts: Live scene, 3D reconstruction system, and remote ILE service. Data are acquired from the live scene with devices and transmitted to the system. And then processed with three parts and finally used to support the ILE service.]

[TBD]

7 Overall Framework of 3D reconstruction system based on INR for ILE services

7.1 General concept and overview

The 3D reconstruction system based on INR are formed with several blocks, including Data acquisition, Data processing, 3D Reconstruction, and Rendering processing. The data acquisition block acquisition images, videos, and other data from devices and then transmits them into the data processing block. The data processing block processes the data with algorithms to ensure the data matches the requirements of the 3D-INR system and then transmits the data into the 3D reconstruction block. The 3D reconstruction block learns the implicit 3D object from the data with the neural network. The rendering processing block renders images or videos from the implicit 3D object and then transmits them to support ILE service.

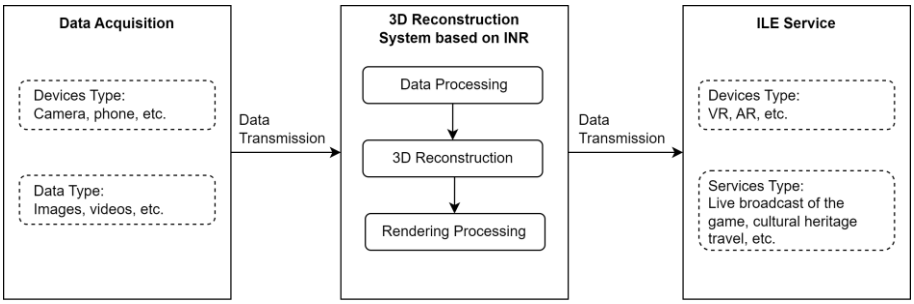

Figure 2 – Overview of the 3D reconstruction with INR-based system for ILE

[TBD]

7.2.1 Real-time 3D-INR system for ILE service

[Editor notes: this part should provide the framework that how real-time 3D-INR system support ILE service]

[TBD]

7.2.2 Non-real-time 3D-INR system for ILE service

[Editor notes: this part should provide the framework that how non-real-time 3D-INR system support ILE service. It could reconstruct the background of the live scene and combine with other technology to support the ILE services]

[TBD]

7.2 Data Acquisition

The data acquisition block acquisition images, videos, and other data from devices and then transmits them into the data processing block.

[TBD]

7.3 Data Processing

The data processing block processes the data with algorithms to ensure the data matches the requirements of the INR system and then transmits the data into the 3D reconstruction block.

[TBD]

7.4 3D Reconstruction

The 3D reconstruction block learns the implicit 3D object from the data with the neural network.

[TBD]

7.5 Rendering Processing

The rendering processing block renders images or videos from the implicit 3D object and then transmits them to support applications such as ILE services and others.

[TBD]

8 Requirements of 3D Reconstruction System Based on INR for ILE Service

8.1 Data type

[TBD]

8.2 Data transmission

[TBD]

8.3 Rendering processing

[TBD]

8.4 Support for ILE Services

[TBD]

Appendix I
Use cases of 3D reconstruction system based on implicit neural representation for ILE service

(This appendix does not form an integral part of this Recommendation)

I.1 Real-time 3D-INR system for ILE service

[RE_ReND] introduced a method which called RE-ReND for rendering nerf in real time across devices. It is designed to achieve real-time performance by transforming NeRF into a representation that can be efficiently processed through a standard graphics pipeline. Re-ReND is able to render pre-trained NeRF methods in real time on a variety of devices with limited computing resources, such as AR/VR headsets and mobile phones. Figure I.1 shows how it support ILE service with Unity engine. Re-ReND preserves remarkable photo-metric quality even when rendering at over 1,013 FPS on a desktop browser, or at the capped 74 FPS of a VR headset. In particular, it extracts NeRF by extracting the learned density into a grid, and breaks the learned color information into a set of matrices, effectively breaking down the light field of the scene. As a result, Re-ReND is able to render NeRF in real time, making it deployable on a variety of devices. When rendering challenging infinitely realistic scenes, it proposed approach achieves a speed improvement of more than 2.6x while maintaining comparable quality.

Commented [TSB2]: Editor: not the right tag style for Bibliographic references. See the ITU author's guide ([b-...]). Multiple occurrences, please fix.

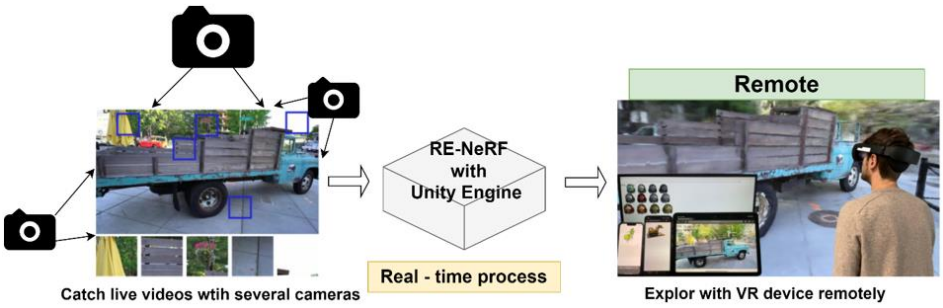

Figure I.1 – Real-time RE-NeRF support ILE service with VR device

Nvidia's Instant-NGP [Instant-NGP] [Instant-git] can achieve the performance of real-time reconstruction of a 3D scene with some 2D videos/pictures shot from the scene, which could support ILE services in real-time. The 2D pictures should be distinct enough when reconstructing and the reality of the reconstructed 3D scene increases with the sharpness of input 2D pictures. And if there are more time for reconstruction, the 3D scene will be clearer and more realistic.

Figure I.2 shows an example, the real-time construction of a statue. It's necessary to take live videos from different directions with variant camera poses. Then Instant-NGP reconstructs the 3D scene of this statue as soon as the images are transmitted into the system. Finally, users could enjoy the ILE services, which include seeing the scene with their VR device and exploring the real-time reconstructed 3D scene by themself.

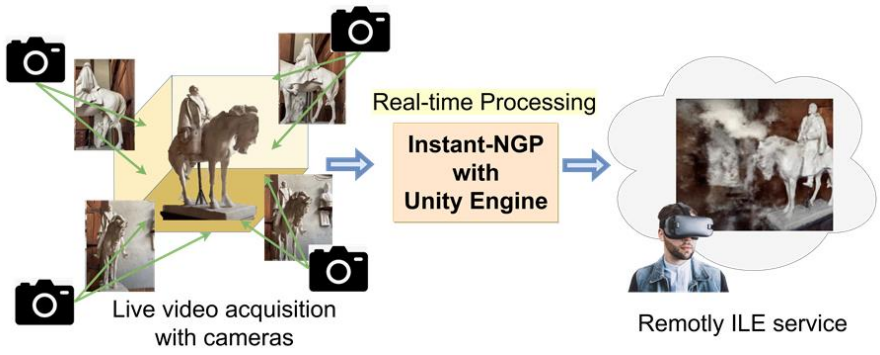

Figure I.2 – Real-time Instant-NGP support ILE service with VR device

I.2 Non-real-time 3D-INR system for ILE service

[Editor notes: this part should revise to give the use cases that how non-real-time 3D-INR system support ILE service]

Implicit Neural Representation-based 3D reconstruction system is capable of rendering scenes for immersive services with VR. It allows users to freely move their heads to explore complex scenes with the VR devices.

Figure I.1 shows an example, it is a scene with a fox head. The first line is three images of this scene with different angles shot by the normal camera. Then use the INR technology to reconstruct this scene with these images, and render with the Unity engine. Finally, users could see the scenes with the VR device and explore it by moving their heads as shown in the second line of Figure I.1. It shows how the INR technology with Unity engine could provide a reality immersive experience based on the VR device.

Commented [TSB3]: Editor: make sure ITU receives a license to use the pictures in this document.

Figure I.1 – Fox scene rendered by NeRF with Unity engine. [ImmersiveNeRF]

Users can also create their scenes and view them in their VR headsets with free software powered by NeRF. [VR_NeRF]Figure I.2 shows examples of deploying Real-time INR-based MobileR2L on

mobile devices for real-time interaction with users. it can build neural rendering applications where users interact with 3D objects on their devices, enabling various applications such as virtual try-on. [Realtime_IEEE] [Realtime_Git]

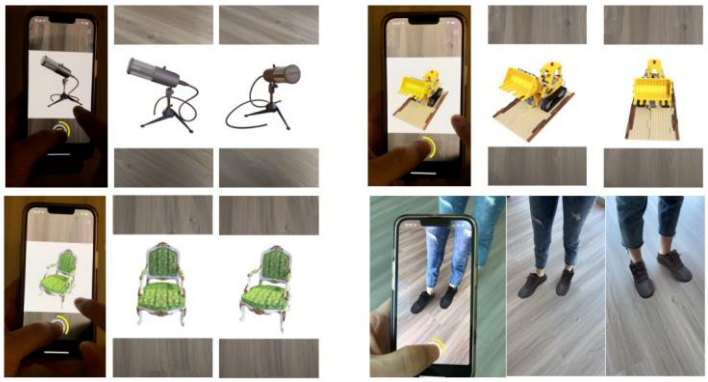

Figure I.2 – INR-based MobileR2L interaction on mobile devices [Realtime_IEEE]

[TBD]

Bibliography

[RE_ReND] Sara Rojas, Jesus Zarzar, Juan C. Pérez, Artsiom Sanakoyeu, Ali Thabet, Albert Pumarola, Bernard Ghanem; Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 3632-3641

[Instant-NGP] Müller T, Evans A, Schied C, et al. Instant neural graphics primitives with a multiresolution hash encoding[J]. ACM transactions on graphics (TOG), 2022, 41(4): 1-15.

[Instant-git] add description, https://github.com/NVlabs/instant-ngp (visited: yyyy-mm-dd)

[ImmersiveNeRF] Li, Ke, et al. "Immersive Neural Graphics Primitives." arXiv preprint arXiv:2211.13494 (2022).

[VR_NeRF] How to create and see NeRFs in a VR headset, https://mixed-news.com/en/nerf-guide-virtual-reality/

[Realtime_IEEE] Cao, Junli, et al. "Real-time neural light field on mobile devices." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

[Realtime_Git] https://gitcode.com/snapresearch/MobileR2L/overview?utm_source=csdn_blog_hover

Commented [TSB4]: Editor: not the right tag style for Bibliographic references. See the ITU author's guide ([b-...]).

Commented [TSB5]: Editor: Add description of the resource liked to this URL (page, topic, etc) and date visited. Also for the last reference.