Remarks

1. Claims 1-4, 6-10, 12-18, and 20 are pending in this application.
2. Independent claims 1, 8, and 15 have been amended based on the subject matter of claims 2, 3, 5, 9, 10, 12, 16, 17, and 19.
3. The dependent claims 5, 11, and 19 have been cancelled.
4. No new subject matter has been added.

Further, the Applicant submits the following response for the Non-Final Office Action dated August 01, 2023:

Response to the rejection Under 35 U.S.C. § 102

The office action on page 2 assets that *(c)laim(s) 1-4, 7-11, 14-18 are rejected under 35 U.S.C. 102(a)(2) as being anticipated by Ninan (U.S. 2023/0215129).*

The amended claim 1 recites:

> "(a) method for providing content in a bandwidth constrained environment, the method comprising: receiving, by a processing unit, a content along with audio inputs received during rendering of the content and provided to one or more users in a bandwidth constrained environment; identifying, by the processing unit, at least one object of interest within the content and based on the received audio inputs, wherein identifying the at least one object of interest associated with the audio inputs comprises: processing the audio inputs using a Natural Language Processing (NLP) model associated with the processing unit, to derive context related to the at least one object of interest; and identifying the at least one object in the display of the content based on the derived context; determining, by the processing unit, one or more regions of interest based on the at least one object of interest, such that the one or more regions of interest include the at least one object of interest, in the bandwidth constrained environment; modifying, by the processing unit, bitrate for rendering the content, based on the determined one or more regions of interest, to obtain a modified content for the bandwidth constrained environment; and providing, by the processing unit, the modified content to be rendered in the bandwidth constrained environment." (emphasis added)

Thus, the present invention relates to a method designed for delivering content in environments where bandwidth is limited, especially in immersive settings like virtual reality or augmented reality. In such contexts, delivering high-quality content can be challenging due to the constraints of available data transfer rates. The method involves a multi-step process: first, content and associated audio inputs are received; then, objects of interest within the content are identified. The method determines specific regions of interest based on these objects. The integration of Natural Language Processing (NLP) techniques adds a layer of sophistication, allowing the system to process audio inputs and derive context related to the objects, enhancing the precision of object identification. Subsequently, the content's bitrate is modified, allocating higher data rates to the regions of interest, and lower rates to other areas, thus optimizing the use of limited bandwidth. For example, consider at an instant of time, the commentator provides audio inputs as "PLAYER 1 IS APPROACHING THE BALL AND AIMING TO HIT A GOAL!!". The context may indicate that the audio inputs are referring to PLAYER 1 amongst the players of the game and/or BALL, such that the object of interests may be identified as the PLAYER 1 and/or the BALL on the display of the content. Accordingly, the present invention determines a region of interest on the display that contains the identified object to enhance its quality and lower the quality of other regions to reduce the overall bitrate of the video. Further, such that region of interest may be accordingly highlighted in the succeeding frames without doing the analysis again (Reference: Paragraph [0049] of the as-filed specification).

This method demonstrates a holistic approach, combining audio-visual processing, user interaction, and adaptive bitrate modification, tailored specifically for immersive environments where an enriched and interactive user experience is crucial despite bandwidth limitations.

Further, the office action on pages 2-4, asserts that,

> *"(a)s per claim 1, Ninan discloses a method for providing content in a bandwidth constrained environment, the method comprising:*
> *receiving, ...*
> *identifying, by the processing unit, at least one object of interest within the content and associated with the audio inputs ([0025] where "The image data encoded in the saliency video stream for the saliency region may depict one or more of: a* ***highly interesting object*** *and/or zero or more less interesting*

objects. a bounded region, a ***region of interest or ROI****, an image area pre-designated by a content creator or director, an image area determined through viewer statistical information gathering and analyses. an image area generated through padding to an image object recognized through computer vision techniques, and so on");*
determining, by the processing unit, one or more regions of interest, including the at least one object of interest, in the bandwidth constrained environment ([0025] where "a ***region of interest or ROI****, an image area pre-designated by a content creator or director, an image area determined through viewer statistical information gathering and analyses");*
modifying...
providing... to a viewer" (emphasis added)

Applicant respectfully disagrees with this assertion and submits that Ninan, in the cited paragraphs or elsewhere, discloses a method for streaming volumetric video content by identifying specific saliency regions within the overall video. Next, the method generates corresponding video streams that focus on such saliency regions to form a saliency stream-based representation of the volumetric video. During transmission to a video streaming client, these saliency video streams are prioritized based on their saliency ranks. Notably, if there is a reduction in the available data rate, the system dynamically adjusts the streaming process. Specifically, lower-ranked saliency video streams, which are assigned a lower saliency rank, such as a second saliency video stream, can be temporarily removed from the transmission set. This adaptive approach ensures optimal streaming quality and content delivery, especially under varying network conditions.

However, both Ninan and the present invention share a common goal of optimizing content delivery in bandwidth-constrained environments, specifically in immersive settings such as virtual reality or augmented reality, but Ninan utilizes a different approach than the present invention i.e., emphasizing the identification and tracking of saliency regions within the video for prioritizing and adaptive removal of certain video stream based on their saliency ranks in response to any reduced data rates.

Further, Ninan is not particularly clear about how the saliency region is identified and is dependent on an image metadata being provided with the volumetric video stream, such image metadata includes geometry information of a saliency region to indicate its size, shape, zoom,

scaling factor, spatial resolution, etc. and is utilized to saliency region identification (Reference: Paragraph [0068] and [0115] of the specification). Specifically, Ninan fails to teach about identifying an object of interest based on the received audio to determine one or more regions of interest including such identified objects. Thus, the Ninan is completely silent on the following features of the present invention:

identifying at least one object of interest within the content and based on the received audio inputs,

wherein identifying the at least one object of interest associated with the audio inputs comprises: processing the audio inputs using a Natural Language Processing (NLP) model associated with the processing unit, to derive context related to the at least one object of interest; and identifying the at least one object in the display of the content based on the derived context, and

determining one or more regions of interest based on the at least one object of interest, such that the one or more regions of interest include the at least one object of interest.

Thus, the amended claim 1 is allowable under 35 U.S.C. 102(a)(2). Accordingly, the Applicant requests the Ld. Examiner to take the amended claim 1 on records and withdraw the rejection.

Further, the Applicant submits that the dependent claims 2-4, and 6-7 are allowable at least by virtue of their dependency over the allowable claim 1.

Furthermore, the office action on page 7 asserts that independent claims 8 and 15 are analogous to claim 1, thus the claims 8 and 15 are also allowable based on the similar arguments as stated above for claim 1

Accordingly, the dependent claims 9-10, 12-14, 16-18, and 20 are allowable at least by virtue of their dependency over the allowable claims 8 and 15.

Thus, the Applicant respectfully submits that the amended claims 1-4, 6-10, 12-18, and 20 are not anticipated by Ninan (U.S. 2023/0215129). Reconsideration and withdrawal of the rejections under 35 U.S.C. § 102(a)(2) are respectfully requested.

Response to the rejection Under 35 U.S.C. § 103

The office action on page 8 assets that *(c)laim(s) 5, 12, and 19 are rejected under 35 U.S.C. 103 as being unpatentable over Ninan (U.S. 2023/0215129) as applied to claim 1 above, and further in view of Smith (U.S. 2022/0353465).*

Further, the office action on pages 2-4, asserts that,

> *"(a)s per claim 5, Ninan demonstrated all the elements as disclosed in claim 1. wherein identifying the at least one object of interest associated with the audio inputs comprises:*
> *processing the audio inputs using a Natural Language Processing (NLP) model associated with the processing unit, to derive context related to the at least one object of interest; and identifying the at least one object in the display of the content based on the derived context. However, this is known in the art as taught by Smith. Smith discloses a method of identifying a ROI in a video capture based on conversational context ([0111])."*

Applicant respectfully disagrees with this assertion and submits that Smith, in the cited paragraphs or elsewhere, discloses an operating method for a conference gallery view intelligence system that determines regions of interest for display. Smith, in the cited paragraphs or elsewhere, discloses determining a conversational context within the conference room based on the direction of the audio and locations of the conference participants in the conference room. Such determined conversational context is then used to determine a first region of interest within the conference room and correspond to a region within a field of view of received input video stream which includes the one or more conference participants who are part of the conversational context. For example, the field of view of the video includes four conference participants and a conversational context based on the determined directions of audio and the locations of the detected conference participants indicates that two of those four conference participants are actively participating in a conversation, the region of interest may correspond to only that portion within the field of view of the video capture device in which those two conference participants are located within the conference room (Reference: Paragraph [0111] of the specification).

Although, Smith teaches about determining the conversation context for determining the region of interest, but this conversational context is based on the direction of audio to capture the participants of such conversations in the field of view of the video. On the other hand, the present invention teaches about processing the audio inputs using a Natural Language Processing (NLP) to derive context related to at least one object of interest and then identifying such object in the display of the content based on the derived context, such as by using an

image processing technique. For example, consider at an instant of time, the commentator provides audio inputs as "PLAYER 1 IS APPROACHING THE BALL AND AIMING TO HIT A GOAL!!". The context may indicate that the audio inputs are referring to PLAYER 1 amongst the players of the game and/or BALL, such that the object of interests may be identified as the PLAYER 1 and/or the BALL on the display of the content. Accordingly, the present invention determines a region of interest on the display that contains the identified object to enhance its quality and lower the quality of other regions to reduce the overall bitrate of the video. Further, such that region of interest may be accordingly highlighted in the succeeding frames without doing the analysis again (Reference: Paragraph [0049] of the as-filed specification).

Thus, it may be apparent to a person skilled in the art that though the term "context" is used both by Smith and the present invention, the meaning of the terms in each of them is different. Specifically, Smith utilizes the audio data to determine "who is speaking" to display a video including associated speakers, and the present invention utilizes the audio data to determine "what is talked about" to enhance the quality of the display area around the associated object.

Accordingly, Smith fails to disclose the following features of the present invention:

identifying at least one object of interest within the content and based on the received audio inputs,

wherein identifying the at least one object of interest associated with the audio inputs comprises: processing the audio inputs using a Natural Language Processing (NLP) model associated with the processing unit, to derive context related to the at least one object of interest; and identifying the at least one object in the display of the content based on the derived context, and

determining one or more regions of interest based on the at least one object of interest, such that the one or more regions of interest include the at least one object of interest

Therefore, Ninan and Smith, individually or in combination fail to disclose at least the following features of the amended claim 1:

> "...identifying at least one object of interest within the content and based on the received audio inputs,
>
> wherein identifying the at least one object of interest associated with the audio inputs comprises: processing the audio inputs using a Natural Language Processing (NLP) model associated with the processing unit, to derive context

> related to the at least one object of interest; and identifying the at least one object in the display of the content based on the derived context; and determining one or more regions of interest based on the at least one object of interest, such that the one or more regions of interest include the at least one object of interest…"

Accordingly, the Applicant respectfully asserts that the amended claim 1 is not obvious over Ninan (U.S. 2023/0215129) in view of Smith (U.S. 2022/0353465).

The Applicant further submits that the subject matter of amended claim 1 is analogous to amended claim 8 and amended claim 15. Therefore, the Applicant's response for the amended claim 1 applies mutatis mutandis to the amended claims 8 and the amended claim 15. Accordingly, the Applicant respectfully asserts that the amended claims 8 and 15 are also not obvious over Ninan (U.S. 2023/0215129) in view of Smith (U.S. 2022/0353465).

Further, the dependent claims 2-4, 6-7, 9-10, 12-14, 16-18, and 20 are also not obvious over Ninan (U.S. 2023/0215129) in view of Smith (U.S. 2022/0353465) at least by virtue of their respective dependencies on either one of the independent claims 1, 8, and 15 that have been proven non-obvious above.

For at least the foregoing reasons, independent claims 1, 8, 15, and claims depending thereon are allowable over the cited references. Reconsideration and withdrawal of the rejections under 35 U.S.C. § 103 are respectfully requested.

Conclusion

In view of the foregoing, Applicant believes the pending application is in condition for allowance.

Dated: 10/27/2023

Respectfully submitted,

BY: /RIYON HARDING/
RIYON RAE HARDING
REGISTRATION NO. 58365
ATTORNEY FOR APPLICANT