View Reviews

Print

Paper ID	899
Paper Title	Optimizing Immersive Video Streaming Using Deep Learning Approaches: A Case Study on TMIV
Track Name	Main Track

Reviewer #1

Not Submitted

Reviewer #2

Questions

1. [Paper Summary] What is the paper about? Please, be concise (3 to 5 sentences).
This paper attempts to generate better TMIV configurations, and focuses on two parameters of TMIV configurations: the number of passes and the number of view per pass.

The paper proposes two deep-learning based algorithms: a CNN based one and a deep reinforcement one.

The paper then describes an experiment where these algorithms are compared to the default TMIV configuration and to the optimal configuration found by exhaustive search: there is no statistically significant difference in the quality of the output, however, the configurations of the proposed approach lead to fewer views, and thus, better performance in terms of network and execution time.

The paper finally describes a subjective experiment where users are asked to rate the outputs quality, which confirms that there is no significant difference in terms of perceived quality.

2. [Relevance] Is this paper relevant to an audience to ACM Multimedia? Please check https://2020.acmmm.org/call-for-paper.html.
Of limited interest to an audience

3. [Significance] Are the results significant?
Significant

4. [Novelty] Are the problems or approaches or applications/systems novel?
Novel

5. [Evaluation] Is the idea proposed in this paper well supported by theoretical analysis or experimental results?
Sufficient

6. [Paper Strengths] Please discuss. Justifying your comments with the appropriate level of details about the strengths of the paper (i.e. novelty, theoretical approach and/or technical correctness, adequate evaluation, clarity, applications, etc.). For instance, a theoretical paper may need no experiments, while a paper with a new approach or application may require comparisons to existing methods.
The paper clearly states a novel problem it tries to solve and proposes two possible solutions.

The paper compares the proposed solutions to the default and optimal solutions: it provides an in-depth analysis of the many metrics of the algorithm (number of required views, video quality, running time, utility function value, inference time) as well as a subjective evaluation where 23 subjects were recruited to compare the algorithms.

The key message of the paper is that the proposed methods give similar outputs, but require less resources and does a good job at demonstrating it.

The two methods have different behaviours and the paper gives recommendation about when to use which method depending on the context.

7. [Paper Weaknesses] Please discuss. Justifying your comments with the appropriate level of details about the weaknesses of the paper (i.e. lack of novelty – given references to prior work, lack of novelty, technical errors, or/and insufficient evaluation, etc.). Note: If you think there is an error in the paper, please explain why it is an error. It is not appropriate to ask for comparisons with unpublished papers and papers published after the ACM Multimedia deadline. In all cases, please be polite and constructive.
My main concern is that the inference time of the system is described very quickly in this paper. The paper states that the CNN takes 37ms and DRL takes 51ms on average to compute the configuration. However, it is not clear how many times those algorithms are run during a session, and while I agree that the paper demonstrates that the systems allow for network resources saving, it is unclear on how much computing resources are saved by running an algorithm that consumes computing resources to infer a configuration that will save computing resources later.

8. [Preliminary Rating] Please rate the paper according to one of the following choices.
Poster

9. [Rebuttal Requests] Please pose questions you want to be answered in the rebuttal. Please do NOT ask the author(s) to include any new results (e.g., experiments and theorems) in the rebuttal.
I would like to know how many times the algorithms are run during a typical streaming session, and if there is an estimation of how much computing resources are spent during this step.

10. [Confidence]
Not Confident

Reviewer #3

Questions

1. [Paper Summary] What is the paper about? Please, be concise (3 to 5 sentences).
The paper proposes two NN algorithms to help find the best source videos for view synthesis for a target viewpoint. The algorithms can be used in free-viewpoint video applications, or TMIV (Test Model for Immersive Video) as named by MPEG and this paper. The evaluation shows the proposed algorithms delivering similar or better view quality compared to the default algorithm defined in MPEG spec while using much less time (fast inference).

2. [Relevance] Is this paper relevant to an audience to ACM Multimedia? Please check https://2020.acmmm.org/call-for-paper.html.
Relevant to researchers in subareas only

3. [Significance] Are the results significant?
Moderately significant

4. [Novelty] Are the problems or approaches or applications/systems novel?
Novel

5. [Evaluation] Is the idea proposed in this paper well supported by theoretical analysis or experimental results?
Somewhat weak

6. [Paper Strengths] Please discuss. Justifying your comments with the appropriate level of details about the strengths of the paper (i.e. novelty, theoretical approach and/or technical correctness, adequate evaluation, clarity, applications, etc.). For instance, a theoretical paper may need no experiments, while a paper with a new approach or application may require comparisons to existing methods.
The paper tries to propose a better algorithm to select reference views. It is probably a niche problem to the current audience but will be an important one if free-viewpoint video or volumetric video becomes popular in the future. The proposed algorithms use the similar idea from video coding and applies well. The algorithms are evaluated with both objective and subjective experiments. A lot of effort but there are some major issues with the evaluation as I will point out next.

7. [Paper Weaknesses] Please discuss. Justifying your comments with the appropriate level of details about the weaknesses of the paper (i.e. lack of novelty – given references to prior work, lack of novelty, technical errors, or/and insufficient evaluation, etc.). Note: If you think there is an error in the paper, please explain why it is an error. It is not appropriate to ask for comparisons with unpublished papers and papers published after the ACM Multimedia deadline. In all cases, please be polite and constructive.
The first section needs a better explanation. Maybe adding a drawing to illustrate the relationship between source views and synthesized views.

I originally thought the paper is about view synthesis for systems like free-viewpoint video. But the experiments in the paper using 360 video to simulate different source views puzzles me. Using different view angles in a 360 video doesn't address the real problems in free-viewpoint video, like occlusion exposure issue, or insufficient sampling.

Another problem is also with evaluation. It looks to me that you are using all three sample videos in your training and then use the same video in testing. Does that mean your system needs to train the test content first before being used on it?

8. [Preliminary Rating] Please rate the paper according to one of the following choices.
Borderline Reject

9. [Rebuttal Requests] Please pose questions you want to be answered in the rebuttal. Please do NOT ask the author(s) to include any new results (e.g., experiments and theorems) in the rebuttal.
Your paper referred in many places to the technical report that you submit together. I find it hard to get a full picture of your work if I don't read the technical report. Is your technical report published anywhere? or how do you plan to resolve this if your paper is actually accepted?

Also, what is "View FoV" in table 1 of your technical report? Is that the FoV of the your simulated camera view? But the FoV in Fig 10 is just a normal FoV.

I am also not entirely sure what does 256x256x(7+7+21+21) mean in Fig 4. 7 for texture, 7 for depth, 21 for position and 21 for orientation? bytes?

10. [Confidence]
Confident

Go Back