TYPE Editorial
PUBLISHED 05 January 2024
DOI 10.3389/frsip.2023.1357892

Editorial: Signal processing in computational video and video streaming

Anil Kokaram[1]*, Anil Anthony Bharath[2]* and Feng Yang[3]*

[1]Trinity College Dublin, Dublin, Ireland, [2]Imperial College London, London, England, United Kingdom, [3]Google Research, Mountain View, CA, United States

KEYWORDS

computational video, video streaming and broadcasting, DNNs, Perceptual video quality assessment, audiovideo editing, rate distortion

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Frederic Dufaux,
Université Paris-Saclay, France

*CORRESPONDENCE
Anil Kokaram,
✉ akokaram@gmail.com
Anil Anthony Bharath,
✉ a.bharath@imperial.ac.uk
Feng Yang,
✉ fengyang@google.com

RECEIVED 18 December 2023
ACCEPTED 20 December 2023
PUBLISHED 05 January 2024

CITATION
Kokaram A, Bharath AA and Yang F (2024), Editorial: Signal processing in computational video and video streaming.
Front. Sig. Proc. 3:1357892.
doi: 10.3389/frsip.2023.1357892

Editorial on the Research Topic
Signal processing in computational video and video streaming

Digital Signal Processing and especially Image and Video Processing has been disrupted by the emergence of Deep Learning as a ubiquitous tool. This is particularly the case in pixel level manipulation of video and images, e.g., denoising. Nevertheless, the traditional topics of DSP still have a role to play in the development of efficient processing pipelines, and in creating and supporting explainable systems. Video quality in capture and display has also never been higher, and devices are pushing the boundaries of brightness (displays), pixels (resolutions), and speed/battery life. Technologies for live streaming video production related to video conferencing and entertainment was accelerated by the pandemic and represents a major challenge to efficiency in terms of the balance of computational load and video quality.

The papers in this Research Topic reflect the extent to which our claims about “DSP inside” video pipelines remain true today. Two papers by teams at Google and Utexas Austin consider the problem of Video Quality Assessment and one from Trinity College Dublin considers the video encoding application itself. The final paper from the team at the University of Galway considers the emerging technology of “Video dubbing” in the streaming production pipeline. In all cases the holy grail of research right now is the design of an end-to-end Neural Network approach which consumes raw video and audio and outputs a final useable output or measurement directly without any pre or post processing steps. It is only because of the computational challenge posed by large neural network systems that aspects of the traditional DSP toolkit are co-opted to strike the balance between compute load and overall performance. In the following paragraphs we summarise the context and contributions of each paper.

In the first of our two papers on Video Quality Assessment (VQA), “*Perceptual video quality assessment: the journey continues!*”, Saha and Pentapti et al. provide a survey of the state of the art in the area. Importantly, they consider VQA for a wide range of video formats including immersive video and high dynamic range content. Measuring the perceptual quality difference between two image or video examples is the only way that we can advance pixel manipulation techniques. Their survey charts the course from the early work in DSP and model based approaches (exemplified by SSIM and PSNR) through to the recent activity in DNNs for measuring quality. What is interesting is that the proliferation of work in Deep Learning VQA is directly proportional to the availability of large databases and associated human ground truth measurements. Hence we see Deep Learning VQA applied to standard dynamic range video and gaming video genres but not as much to High Dynamic Range video material because of the lack of datasets. Many competitive VQA techniques, rather than end-to-end Neural networks, instead use the backbone of other DNN architectures to generate features which are articulated under some other ML framework.

Our second paper in this Research Topic on VQA is "*MRET: Multi-resolution transformer for video quality assessment*", Ke and Zhang et al. The authors present computational considerations in the use of Transformers for VQA that motivate a classical multiscale approach. In this work, frames are represented as modified multiscale pyramids. Rather than create classical oversampled pyramids (e.g., Gaussian pyramids from the 1990s), only a selection of rectangular windows within each frame are represented in this way. The spatial extent of those windows is selected as that which results in a complete tiling of the coarse level image decomposition. This gives the effect of a global attention layer (at the coarse level) directing the extraction of sparsely sampled details at the finer scales. The result is a transformer based framework which can handle the high resolution image sizes typical in User Generated Content uploads. They report state of the art performance over a variety of metrics on standard definition content.

Our third paper in this Research Topic "*The disparity between optimal and practical Lagrangian multiplier estimation in video encoders*", Ringis and Vibhoothi et al., considers a well known engine within existing video compression schemes: the rate distortion (RD) control. The pioneering use of Lagrange multipliers for constrained optimisation in this problem was adopted as the strategy for RD control in the late 1990's. In this work the authors pose the question "How much compression gain is there available if we find the optimal Lagrange multiplier for each clip in a corpus?". The authors present a direct optimisation scheme that treats the entire encoder like a black box with a single parameter of interest. They show that in fact compression gains of more than 20% are possible for particular clips with at least 2% gain on average. What is interesting here is that this work is a necessary preamble for then exploring the use of ML strategies to predict the optimal multiplier from analysis of the content. Because the optimisation process proposed is based on iterations of encoding, it is necessarily computationally heavy. A suitably chosen ML technique applied to a video content feature set that is cheap to calculate (e.g., frame textures and motion differences) could yield a faster prediction albeit with some loss of optimality. Instead the authors explore a more pipeline oriented approach by learning the optimal multiplier from a proxy version of the content (low resolution version) and a proxy version of the encoder (a fast preset configuration). This approach yields more than 10x computational load reduction depending on the encoder.

Our final paper "*Multilingual video dubbing—a technology review and current challenges*, Bigioi and Corcorcan" considers a newly emerging part of the video content production pipeline. This is usually in the final conforming step in which new audio versions are laid over the video content for different languages or different audiences. The authors survey the emergence of deep-fake technology applied to audio/video synthesis and synchronisation. In both cases the idea is to synthesise convincing facial video expressions that match the new audio track. Both face-landmark based strategies and end-to-end strategies are surveyed. Animation of a facial skeleton based on manipulation of topological landmarks through a DNN synthesis process remains competitive. This is an example of a streaming video pipeline component which was not feasible to automate before the recent exploration of DNN approaches to the problem.

Overall then, we see that DSP still finds a place within the emerging DNN infrastructure in media streaming pipelines. This appears to be more motivated from a computational point of view. It remains to be seen whether end-to-end approaches to solving problems in this domain with Neural Network frameworks will succeed on both fronts.

Author contributions

AK: Conceptualization, Project administration, Writing–original draft, Writing–review and editing. AB: Conceptualization, Writing–original draft, Writing–review and editing. FY: Conceptualization, Writing–original draft, Writing–review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.