Real-Time Enhancement of Low-Quality Video for Constrained Camera Systems

Wangyu Choi
wangyu92@hanyang.ac.kr
Hanyang University
Ansan, Republic of Korea

Jongwon Yoon*
jongwon@hanyang.ac.kr
Hanyang University
Ansan, Republic of Korea

Abstract

Deploying high-spec cameras in video systems often falls short of user expectations. Leveraging advancements in deep learning, we propose a mobile, lightweight, real-time video enhancement system. Our approach adopts cutting-edge models and introduces novel optimization techniques for real-time streaming, improving low-resolution, grayscale, and low frame-rate videos. Preliminary evaluations show significant improvements in PSNR and SSIM, while visual assessments confirm substantial quality enhancements while maintaining real-time processing requirements.

CCS Concepts

• **Computer systems organization → Real-time systems**.

Keywords

Real-time video enhancement, Video camera system.

ACM Reference Format:
Wangyu Choi and Jongwon Yoon. 2024. Real-Time Enhancement of Low-Quality Video for Constrained Camera Systems. In *The 30th Annual International Conference On Mobile Computing And Networking (ACM MobiCom '24), November 18–22, 2024, Washington D.C., DC, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3636534.3697451

*We would like to thank Pablo Canella and Yenni Tadjer for their help in implementing the model and measurements. This work was partly supported by Basic Science Research Program through the National Research Foundation of South Korea (NRF) funded by the Ministry of Education NRF-2022R1A2C1008743 and Innovative Human Resource Development for Local Intellectualization program through the Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (IITP-2024-RS-2020-II201741, 50%). Jongwon Yoon is the corresponding author.

ACM MobiCom '24, November 18–22, 2024, Washington D.C., DC, USA

ACM ISBN 979-8-4007-0489-5/24/11
https://doi.org/10.1145/3636534.3697451

1 Introduction

With advancements in technologies such as networks, mobile devices, and the Internet of Things (IoT), video camera systems have become pervasive in our daily environments. Surveillance cameras monitor security in smart homes, dash cams and vehicle black boxes capture real-time footage for accident investigations, and remote cameras allow wildlife observation without disturbance. The ability to remotely monitor and manage locations using smartphones significantly enhances convenience and security.

Deploying video camera systems in suboptimal conditions poses significant challenges. The compact hardware design limits component size, and low-light environments or reliance on infrared sensors capturing only grayscale images reduce quality. Power supply issues and size restrictions on large image sensors further complicate deployment. In such conditions, the shortcomings of low-resolution, grayscale cameras become evident, necessitating improvements such as upscaling and transitioning to high-resolution, color imaging. Users accustomed to high-resolution, color footage often find these limitations unacceptable, highlighting the need for advanced image quality enhancements to meet expectations while balancing hardware constraints.

Remarkable advancements in artificial intelligence and deep learning over the past decade have enabled tasks like super-resolution, colorization, and video frame interpolation to transform low-quality videos into high-quality ones. We explore enhancing low-quality video footage as a more effective alternative to using high-quality camera modules from the outset. However, these models require substantial computational resources and time, posing a challenge for real-time processing. Real-time capability is crucial in video camera systems, but the computational demands of these tasks often limit their practicality for live video feeds, necessitating more efficient solutions for real-time applications.

To address this challenge, we propose a mobile, lightweight, real-time video enhancement system for video camera applications. The system ensures real-time capability and

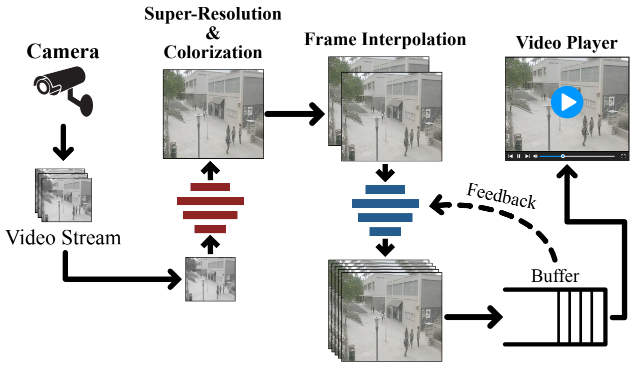

Figure 1: System overview. The red and blue stripes represent deep learning models: red for super-resolution and colorization, and blue for frame interpolation.

high-quality output by employing efficient deep-learning models and optimized algorithms. By minimizing computational load and latency, our approach significantly enhances video quality with minimal processing delay, maintaining superior video quality and the responsiveness required for real-time applications. This mobile, lightweight solution underscores the importance of portability and efficiency in achieving real-time video enhancement.

2 System Design

The goal of the system is to enhance low-quality video–characterized by low resolution, grayscale, and low frame rate–into high-resolution, color, and high frame rate video in real-time. This work addresses the increasing user expectations driven by advancements in network and display technologies, contrasted with the difficulty of deploying high-spec camera systems in constrained environments.

To bridge this gap, we employ three deep learning models: super-resolution, colorization, and frame interpolation as depicted in Figure 1. Although these models have achieved remarkable performance, they face the challenge of processing video at a rate of one second per second to maintain real-time performance. To meet this constraint, we implement various optimization techniques, ensuring our system delivers high-quality video enhancements within the strict limits required for real-time applications.

Core enhancement models. To achieve real-time video enhancement, we adopt three state-of-the-art deep learning models: RIFE [1] for frame interpolation, a deep learning-based model [4] for colorization, and SRGAN [2] for super-resolution. These models are selected for their proven efficacy and advanced methodologies. RIFE efficiently generates temporally coherent intermediate frames, enabling smooth motion in low frame-rate videos. The deep learning colorization model transforms grayscale images into high-quality colorized ones using a fully convolutional architecture. SRGAN enhances image details and texture, producing photo-realistic high-resolution images from low-resolution inputs. Integrating RIFE, the colorization model, and SRGAN allows us to comprehensively enhance video quality in real time, meeting modern user expectations and overcoming the limitations of constrained camera systems.

Optimization for a real-time system. In a real-time video enhancement system, the stringent requirement is to process one second of video within one second. Achieving this with all three enhancement modules—super-resolution, colorization, and frame interpolation—is challenging. For example, a 30 FPS video requires each frame to be processed within approximately 33 milliseconds. This tight deadline means any delays can interrupt video playback, especially with varying frame arrival times or processing times due to content complexity.

To address these challenges, we adopt state-of-the-art techniques and propose novel optimizations. One optimization is quantization, converting computations to 16-bit floating-point operations to reduce processing time without sacrificing quality. Additionally, we integrate super-resolution and colorization into a single, efficient model to achieve shorter inference times. To mitigate interruptions from processing delays, as illustrated in Figure 1, we introduce an enhancement buffer, ensuring a continuous flow of video frames even with varying individual processing times. We also dynamically adjust the final output frame rate based on the buffer level, consistently meeting real-time processing deadlines without noticeable disruptions to the user.

The adjustment process involves monitoring the buffer's fill level and dynamically modifying the output frame rate to prevent overflow or underflow. When the buffer is nearly full, indicating that frames are being processed faster than they are consumed, the system slightly increases the output frame rate, allowing more frames to be displayed per second. Conversely, when the buffer level drops, indicating slower processing, the system decreases the output frame rate to avoid exhausting the buffer and causing playback interruptions. This adaptive frame rate adjustment ensures a smooth viewing experience, maintaining synchronization between processing speed and playback.

3 Preliminary Evaluation and Conclusion

Experimental setup. For the experimental setup, we use the original configurations and pretrained models of the three enhancement models as specified in their respective papers. We evaluate the proposed system by downloading various videos from YouTube and collecting surveillance videos from the VIRAT dataset [3]. To simulate low-spec video camera conditions, we convert these videos to lower-quality settings:

Table 1: Quantitative evaluation of the proposed system compared to the original one.

Metric	PSNR (dB)	SSIM
Low quality video	21.34 ± 1.24	0.7163 ± 0.0532
Enhanced video	26.71 ± 2.23	0.8501 ± 0.0459

120p and 320p resolutions, grayscale, and frame rates of 1 FPS or 5 FPS. The 120p videos are upscaled to 480p, and the 320p videos to 1280p. We assess the enhanced videos using the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM).

Inference time. We measure the time taken, in milliseconds, to enhance videos from 120p or 320p grayscale at 1FPS or 5FPS to 480p or 1280p resolution, color, and frame rates of 10, 15, 20, and 30FPS. These results are illustrated in Figure 2. A key observation is that converting a 320p video to 1280p at 10 FPS meets the 1000 ms (1 second) deadline. For 120p videos, the increase in inference time with higher output FPS is negligible. These results demonstrate the feasibility of our approach in real-time environments. This evaluation highlights that the system effectively enhance low-quality video to meet the stringent real-time processing requirements, validating its practical applicability.

Effectiveness of super-resolution, colorization. To quantitatively evaluate the performance of the system, we measure the PSNR and SSIM of the low-quality videos and the enhanced videos against the original videos. The results are presented in Table 1. The results indicate a significant improvement in video quality. The PSNR of the enhanced video is 26.71dB, compared to 21.34dB for the low-quality video, indicating a substantial reduction in distortion. Similarly, SSIM of the enhanced video is 0.8501, compared to 0.7163 for the low-quality video, demonstrating a notable enhancement in structural similarity and visual quality. These improvements underscore the effectiveness in enhancing video quality, validating its potential for real-world applications.

Visual quality. Figure 3 illustrates frames from a video in the surveillance video dataset, showcasing the visual quality improvement. Despite the input being a very low-quality 120p video, the output at 480p demonstrates a significant enhancement in clarity and detail. Similarly, when enhancing 320p video to 1280p, there is a notable quality improvement. The effectiveness of the proposed system is particularly pronounced when enhancing lower-quality (120p) videos, highlighting its capability to substantially improve video clarity. These results indicate that our system is highly suitable for the low-spec camera video systems, providing considerable visual quality improvements in real-time applications.

Conclusion. In this work, we propose a real-time video enhancement system utilizing three state-of-the-art deep learning models: super-resolution, colorization, and frame interpolation. Our system effectively transforms low-quality video into high-resolution, color, and high frame rate output, meeting the demands of modern users and overcoming the limitations of constrained camera systems. Experimental results demonstrate significant improvements in video quality, validating the feasibility and effectiveness of our approach for practical applications.

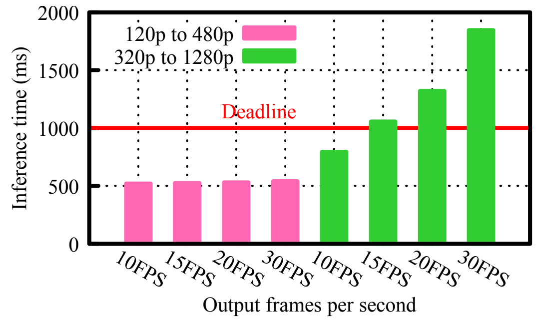

Figure 2: Inference time of the proposed system for various video resolutions and frame rates.

Figure 3: Visual quality comparison between low-quality video and enhanced video.

References

[1] Zhewei Huang, Tianyuan Zhang, Wen Heng, Boxin Shi, and Shuchang Zhou. 2020. Real-Time Intermediate Flow Estimation for Video Frame Interpolation. In *ECCV*. 624–642.

[2] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. 2017. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In *CVPR*. 4681–4690.

[3] Sangmin Oh, Anthony Hoogs, Amitha Perera, Naresh Cuntoor, Chia-Chih Chen, Jong Taek Lee, Saurajit Mukherjee, J K Aggarwal, Hyungtae Lee, Larry Davis, Eran Swears, Xioyang Wang, Qiang Ji, Kishore Reddy, Mubarak Shah, Carl Vondrick, Hamed Pirsiavash, Deva Ramanan, Jenny Yuen, Antonio Torralba, Bi Song, Anesco Fong, Amit Roy-Chowdhury, and Mita Desai. 2011. A Large-scale Benchmark Dataset for Event Recognition in Surveillance Video. In *CVPR*. IEEE, 3153–3160.

[4] Richard Zhang, Phillip Isola, and Alexei A Efros. 2016. Colorful Image Colorization. In *ECCV*. 649–666.