Video Compression - Hw2

D11921B15 程煒倫

2. study VMAF paper and summarize your finding and understanding.

VMAF (Video Multi-Method Assessment Fusion) is a tool developed by Netflix for evaluating video quality. It combines various image quality metrics to predict video quality.

Visual Information Fidelity (VIF): VIF simulates the efficiency of the human visual system by analyzing and comparing the efficiency of visual information transfer in images. VIF is based on natural scene statistics (NSS) and analyzes specific patterns and structures in images. It assesses to what extent processed images retain essential visual information from the original images. For example, even if an image is compressed, it checks whether important details, textures, and colors are still discernible. These patterns and structures align with what the human eye expects under natural viewing conditions. VIF takes into account information loss due to image processing and viewing conditions, providing a quantitative assessment of visual quality. VIF scores are calculated by comparing the efficiency of information transfer between the original and processed images. A higher score indicates that the processed image retains most of the information from the original image, implying lower visual quality loss.

- **Detail Loss Metric (DLM):** This metric is used to assess the loss of detail information in images. It evaluates image quality by analyzing the loss of detail and texture information.

- **Mean Co-located Pixel Difference (MCPD):** MCPD is an indicator that measures color variations in images. It compares pixel color differences between the original and compressed or processed images. By combining these metrics, VMAF can provide a comprehensive assessment of video quality. This holistic evaluation is more accurate in reflecting human perception of video quality compared to individual metrics.

- **Feature Fusion through SVM Regression Model:** VMAF utilizes features including VIF, DLM, MCPD, and other quality metrics. These metrics offer different dimensions of visual quality information. An SVM (Support Vector Machine) regression analysis is employed to effectively blend these metrics into an integrated quality score. Unlike classification models, SVM regression models are used to predict continuous numerical values rather than classification labels.

- **Difference Mean Opinion Score (DMOS):** DMOS is a method to evaluate the overall quality of videos. It calculates by taking the arithmetic average of VMAF scores for each frame in a video sequence. This approach reflects the overall quality of the video, not just the quality of individual frames.

- **Custom Models and Adaptation to Mobile Screens:** VMAF supports custom models to adapt to different viewing conditions, such as mobile screens or UHD televisions. Custom models can be trained to adjust scoring criteria based on factors like screen size and resolution. This means that videos optimized for mobile viewing can achieve higher quality scores, even when watched on small screens. The same content may yield higher quality scores if it is optimized for specific playback conditions, such as mobile screen optimization. Custom models tailored for specific viewing conditions can contribute to higher quality scores for such videos.

By calculating the average scores of all frames, VMAF can provide a comprehensive score reflecting the overall video quality, rather than relying solely on the evaluation of individual frames. VMAF offers a relatively accurate and comprehensive assessment of video quality by combining multiple image quality metrics using machine learning methods.

Research Findings:

Performance Analysis: Research indicates that VMAF outperforms traditional metrics (such as SSIM and MS-SSIM) in some cases but may not always be optimal on specific datasets. SSIM, MS-SSIM, and VMAF are all algorithms used for evaluating video or image quality, with each focusing on different characteristics and evaluation methods.

✓ **Key Focus**: SSIM and MS-SSIM primarily concentrate on image structure, brightness, and contrast, while VMAF considers a

broader range of quality metrics, including but not limited to structural similarity.

- ✓ **Complexity**: MS-SSIM builds upon SSIM by adding multiscale considerations for more comprehensive evaluation. VMAF is more complex as it combines multiple metrics (such as VIF, DLM, MCPD, etc.) and incorporates SVM-based regression analysis for video quality assessment.

- ✓ **Applicability**: SSIM and MS-SSIM are suitable for basic image quality assessment, particularly in image processing and compression. VMAF is suitable for a broader range of video quality evaluation, especially in the fields of streaming media and broadcasting.

Adaptation to Different Viewing Conditions: VMAF's ability to adapt to specific viewing conditions enables it to provide accurate video quality assessments on various devices and in different environments.

Future Developments: Netflix plans to further improve VMAF, including the addition of a time-aware model, capturing the effects of perceptual optimizations, incorporating chromatic features, and better expressing the perceptual advantages of HDR/WCG videos.

- ✓ **Adding a Time-Aware Model:** Currently, VMAF primarily focuses on the quality of individual frames. However, when watching videos in practice, the variations between consecutive frames (such as motion blur or subtle differences between frames) are also essential. A "time-aware model" means assessing not only the quality of individual frames but also considering factors that change over time, such as motion smoothness and stability in dynamic scenes.

- ✓ **Capturing the Effects of Perceptual Optimizations:** Evaluating how different encoding and processing techniques affect video quality. For instance, certain encoding techniques may maintain high visual quality while reducing file sizes. VMAF will be able to assess the actual effects of these techniques, aiding in selecting the optimal encoding settings.

- ✓ **Incorporating Chromatic Features:** Chromaticity refers to the quality of colors. While VMAF primarily focuses on luminance (i.e., the brightness of images), future versions will give more

consideration to color accuracy and saturation, which are crucial for high-quality videos.

- ✓ **Better Expressing the Perceptual Advantages of HDR/WCG Videos:** HDR (High Dynamic Range) and WCG (Wide Color Gamut) are critical technologies for enhancing video quality. HDR provides richer details in brightness, and WCG offers a broader range of colors. VMAF's updates will better evaluate the improvements in video quality due to HDR and WCG technologies, such as more realistic colors and finer variations in lighting.

Detailed explanation:

1. Resolution and Aspect Ratio Adjustment

- Resampling and Scaling: NTSC typically has a resolution of 720x480, while PAL has 720x576. A standard computer display, like mentioned (1280x1024), has a higher resolution. Therefore, video content needs to be upsampled to match or be close to the monitor’s resolution. This process involves resampling and scaling the image. Resampling is required to transform the image data from the original resolution to a new one, while scaling adjusts the size of the image to fit the screen properly.

- Maintaining Aspect Ratio: TV standards (NTSC/PAL) usually have a 4:3 aspect ratio, whereas some computer screens might have different aspect ratios (e.g., 5:4 for 1280x1024). To prevent image distortion, the original video's aspect ratio needs to be maintained. This might result in black bars on the sides of the video during playback.（防止圖像變形，兩側有黑邊）

2. Scan Type Conversion

- From Interlaced to Progressive Scan: TV video (especially NTSC/PAL) typically uses interlaced scanning, while computer screens use progressive scanning. For optimal quality display on computer screens, the video needs to be converted from interlaced to progressive scan. This process, known as deinterlacing, can be achieved through various algorithms like Bob, Weave, or Motion-Adaptive Deinterlacing.

3. Refresh Rate Compatibility

- Refresh Rate Adjustment: The frame rate of TV standards (approximately 29.97 Hz for NTSC and 25 Hz for PAL) differs from the refresh rate of computer monitors. Although most modern monitors and video playback software can handle frame rate differences automatically, frame rate conversion might be necessary to adapt to the 72 Hz refresh rate of the computer display, ensuring smooth playback.

4. Color Space Conversion

- Color Space Adjustment: TV video typically uses the YUV color space, while computer monitors operate in the RGB color space. The playback software or video processing software needs to convert the video from YUV to RGB color space.

5. Appropriate Playback Software

- Use of Video Playback Software: Modern video playback software, such as VLC or MPC-HC, has built-in features to handle these conversions automatically. They can adjust resolution, aspect ratio, scan type, and color space to optimally display the video on a computer screen.