UHD World Association
世界超高清视频产业联盟

General Technical Specifications for Intelligent Video Processing Systems

（Version NO. 1.0）

Release Time
2022-02-10

UHD World Association (UWA)
T/UWA 010-2022

Contents

General Technical Specifications for Intelligent Video Processing Systems

1 Scope

This document specifies the general technical requirements for intelligent video processing systems in terms of smart video quality improvement, smart video restoration, smart video editing, smart video encoding, and supported video formats, and describes corresponding test methods.

This document is applicable to the design, development, testing, and use of intelligent video processing systems.

2 Normative References

The content of the following documents constitutes essential provisions of this document in the form of normative references. In terms of dated reference documents, only versions with those dates are applicable to the document. In terms of undated reference documents, their latest versions (including all modifications) are applicable to this document.

GB/T 9813.1 General specification for computer - Part 1: Desktop microcomputer

GB/T 9813.2 General specification for computer - Part 2: Laptop microcomputer

GB/T 9813.3 General specification for computer - Part 3: Server

GY/T 155 Video parameter values for the HDTV standard for production and programme exchange

GY/T 307 Parameter values for ultra-high definition television systems for production and programme exchange

GY/T 315 Image parameter values for high dynamic range television for use in production and programme exchange

GY/T 340-2020 Subjective method for evaluating image quality of ultra-high-definition television and dual-stimulus continuous quality scale method

3 Terms and Definitions

The following terms and definitions apply to this document.

3.1

Intelligent video processing systems

A system that uses AI to optimize videos.

Note: Video optimization includes but is not limited to video quality improvement, restoration of video taken from old film stock, video editing efficiency improvement, and video compression efficiency improvement. The system may be implemented on a cloud platform or a terminal.

3.2

Content adaptive encoding

An encoding method that is used to generate encoding configuration with the optimal bit rate and corresponding resolution, based on analysis of video content features.

Note: Adopt an AI model or a machine learning model to predict the optimal encoding configuration to achieve the lowest encoding bit rate for a given video.

3.3

Intelligent coloring

A method of color rendering involving the application of AI technologies to video.

3.4

Scratch

Scratches on physical media (such as film stock) that cause long, narrow blemishes on parts of the film. Scrathes can also occur when abnormal contact of magnetic heads damages the film, causing long, narrow blemishes on parts of the film, or they can be the result of other factors.

3.5

Noise

Granular image damage on the video.

Note: Noise includes snow noise, salt and pepper noise, Gaussian noise, artifact, block effect, blur, speckle, etc.

3.6

Specific subimage

An area of a video image that has a specific meaning.

4 Abbreviations

The following abbreviations appear in this document.

CRF: Constant Rate Factor

CMAF: Common Media Application Format

DASH: Dynamic Adaptive Streaming over HTTP

HDR: High Dynamic Range

HLS: HTTP Live Streaming

ROI: Region Of Interest

VMAF: Video Multimethod Assessment Fusion

5 Overview

An intelligent video processing system integrates AI-based video processing algorithms into conventional video processing. It is mainly used for video content reproduction and the restoration of video taken from old film stock. In terms of video processing workflow, the functional architecture of an intelligent video processing system mainly includes smart video quality improvement, smart video restoration, smart video editing, smart video encoding, and video formats. In terms of system logic implementation, technologies are deployed at the basic layer, the processing layer, the platform layer, and the interaction layer. This document standardizes the relevant technical requirements and test methods along the workflow of video processing and logic implementation. The mapping between test items and technical requirements shall be based on Appendix A. For each test item, the test system configuration shall be based on Appendix B, the source video configuration based on Appendix C, and the formula for calculating the technical indicators based on Appendix D.

Figure 1 shows the functional architecture of an intelligent video processing system with five main functional modules: smart video quality improvement, smart video restoration, smart video editing, smart video encoding, and supported video formats. The smart video quality improvement module mainly enables smart super resolution, smart enhancement, smart frame interpolation, and smart HDR conversion to improve the comprehensive picture quality. Smart video restoration is mainly a matter of scratch removal, noise removal, and intelligent coloring to comprehensively restore videos taken from old film stock. The smart video editing module enables black bar cropping, smart landscape/portrait conversion, smart captioning, and the removal of specific subimages. The smart video encoding module mainly performs content adaptive encoding and ROI encoding. Supported video formats mainly specify the input encoding format, output encoding format, input container format, and output container format.

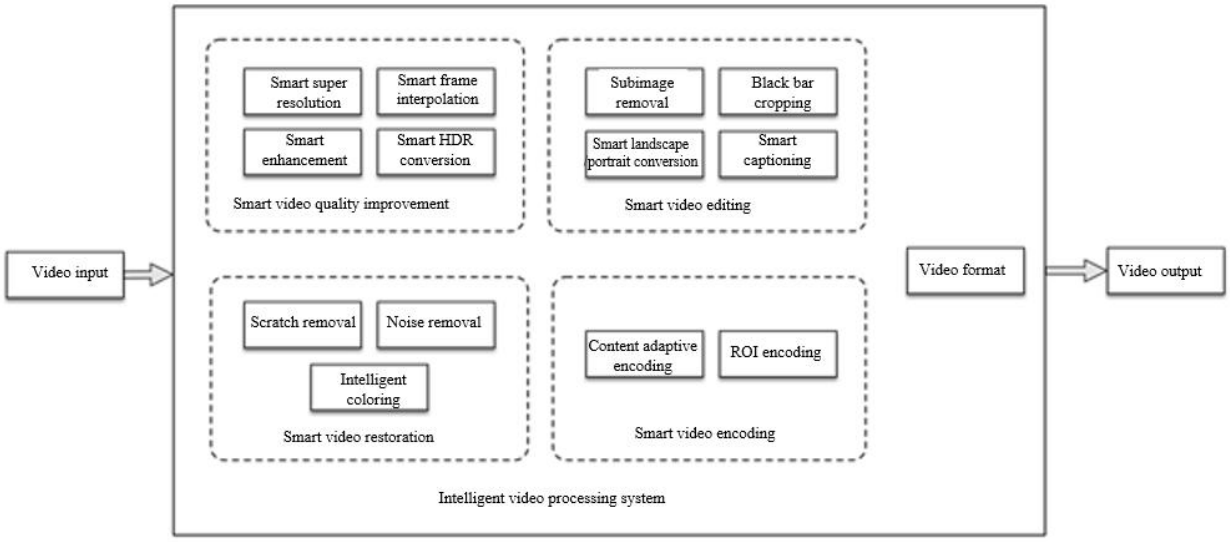

Figure 1 Functional architecture of an intelligent video processing system

The technical architecture of an intelligent video processing system is shown in Figure 2. It comprises four main technical integration layers: the basic layer, the processing layer, the platform layer, and the interaction layer. The basic layer is the layer of infrastructure such as computing, storage, network, and containers. The main components of the processing layer are processing modules such as video editing and synthesis, video analysis and segmentation, video processing and transcoding, and video joining and encapsulation. The platform layer is mainly made up of platform components such as platform access, media asset management, task queue, and task scheduling. The interaction layer is the layer at which user interaction occurs, through consoles and APIs.

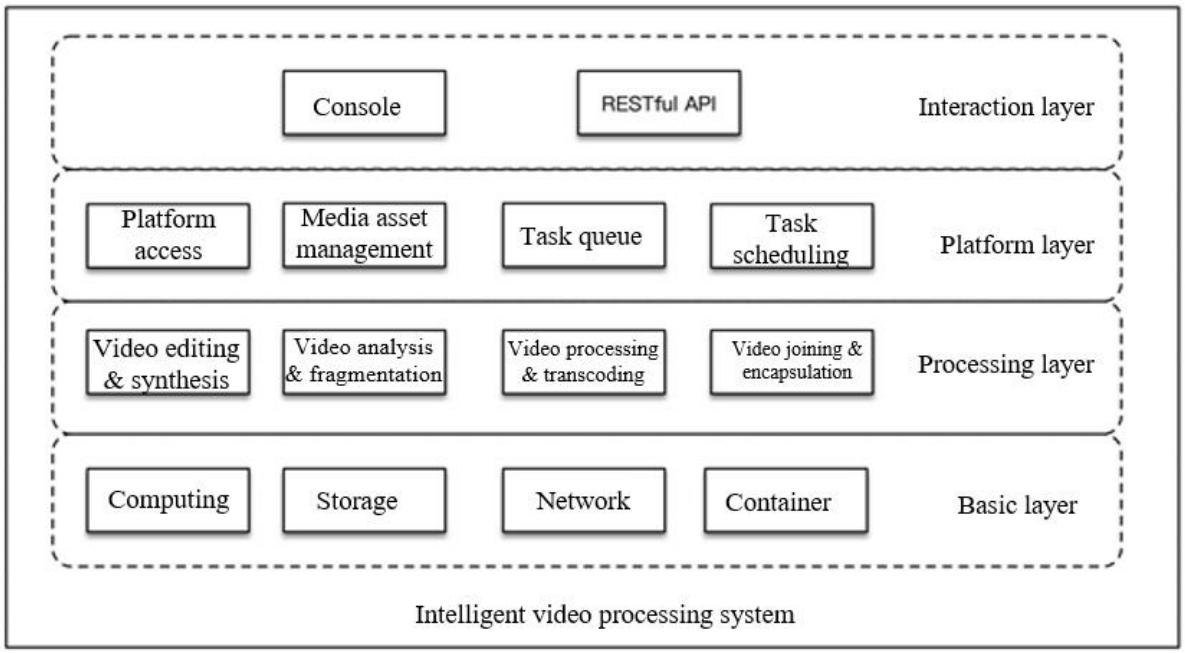

Figure 2 Technical architecture of an intelligent video processing system

6 Technical Requirements

6.1 Smart Video Quality Improvement

6.1.1 Smart Super Resolution

The intelligent video processing system should meet the following smart super resolution requirements:

a) The system should support conversion of SD (720 x 576) videos to HD (1920 x 1080) videos.
b) The system should support conversion of HD (1920 x 1080) videos to UHD 4K (3840 x 2160) videos.
c) It is recommended that the system support conversion of HD (1920 x 1080) videos to UHD 8K (7680 x 4320) videos.
d) The system should support at least 2x resolution improvement.
e) It is recommended that the system support 4x resolution improvement.

6.1.2 Smart Enhancement

The intelligent video processing system should meet the following smart enhancement requirements:

a) The system should support smart enhancement of image edge textures in videos.
b) The system should support smart enhancement of text edge textures in videos.
c) It is recommended that the system support adaptive enhancement of specific ROI areas, such as faces.
d) The system should support de-interlacing of interlaced videos.

6.1.3 Smart Frame Interpolation

The intelligent video processing system should meet the following smart frame interpolation requirements:

a) The system should support at least 2x frame interpolation.
b) It is recommended that the system support 4x frame interpolation.

6.1.4 Smart HDR Conversion

The intelligent video processing system should meet the following smart HDR conversion requirements.

a) The system should support two HDR conversion curves, PQ and HLG.
b) The system should support dynamic conversion from GY/T 155 to GY/T 315.
c) The system should support color space conversion from GY/T 155 to GY/T 307.
d) The system should support conversion from 8 bit depth to 10 bit depth.
e) The system should support one of the following coding formats: H.265, AVS2, or AVS3.
f) The system should support writing HDR metadata in coding.
g) It is recommended that conversion between different HDR standards be provided, for example, conversion from HLG to HDR10 or HDR Vivid.

6.1.5 Comprehensive Image Quality Improvement

After applying one or more of the following processing methods: smart super resolution, smart enhancement, smart frame interpolation and smart HDR conversion, the image quality of the output target video should be improved by no less than 10% compared with the source video. The image quality improvement ratio shall be calculated according to formula (D.1) in Appendix D.

6.2 Smart Video Restoration

6.2.1 Scratch Removal

The intelligent video processing system should meet the following scratch removal requirements.

a) The system should support smart scratch detection. The recall rate of the detection should be no less than 70%. The recall rate shall be calculated according to formula (D.2) in Appendix D.
b) The system should support scratch removal, and the parameters for removal intensity should be adjustable.
c) It is recommended that a visualized environment be provided to support manual adjustment of scratch detection results.

6.2.2 Noise Removal

The intelligent video processing system should meet the following noise removal requirements.

a) The system should support smart noise detection.
b) The system should support noise removal, and the parameters for removal intensity should be adjustable.
c) It is recommended that a visualized environment be provided to support manual adjustment of noise detection results.

6.2.3 Intelligent Coloring

The intelligent video processing system should support intelligent coloring.

6.2.4 Comprehensive Restoration and Improvement

After applying one or more of the following processing methods: scratch removal, noise removal, and intelligent coloring, the image quality of the output target video should be improved by no less than 10% compared with the source video. The image quality improvement rate is calculated according to formula (D.1) in Appendix D.

6.3 Smart Video Editing

6.3.1 Specific Subimage Removal

The intelligent video processing system should meet the following subimage removal requirements.

a) The system should support intelligent detection of subimages. The recall rate of the detection shall not be lower than 95%. The recall rate is calculated according to formula (D.2) in Appendix D.
b) The system should support intelligent removal of subimages in a specified area.

6.3.2 Black Bar Cropping

The intelligent video processing system should meet the following black bar removal requirements:

a) The system should support intelligent detection of black bars in the left and right areas of the video image. The recall rate of the detection should be no lower than 95%. The recall rate is calculated according to formula (D.2) in Appendix D.
b) The system should support intelligent cropping of black bars in a specified area.

6.3.3 Smart Landscape/Portrait Conversion

The intelligent video processing system should meet the following requirements for smart landscape/portrait conversion.

a) The system should support conversion of horizontal videos to vertical videos (9:16) based on the main picture.
b) When there are multiple main pictures, the system should support smart selection of the main picture for cropping.
c) When there is frequent switching between video scenes, or there is cropping of multiple main pictures, the output picture should be stable and free of jitter.

6.3.4 Smart Captioning

The intelligent video processing system should meet the following smart captioning requirements.

a) The system should support automatic speech recognition and captioning of videos. The word error rate should not be higher than 10%. The word error rate is calculated according to formula (D.3) in Appendix D.
b) The system should support caption files of ASS or SRT format.

6.4 Smart Video Encoding

6.4.1 Content Adaptive Encoding

The intelligent video processing system should meet the following adaptive encoding requirements:

a) The system should support smart content adaptive coding that distinguishes between different video scenes and encodes them based on their characteristics.
b) When tested using the method described in 7.4.1, the average VMAF score of each output target video clip should be no less than 91 points and no more than 94 points, and the difference in average VMAF score between different output target video clips should be no more than 2 points.
c) When tested using the method described in 7.4.1, no more than 8% of single frames in each output target video clip should have a VMAF score 5% lower than the average score for frames in that clip.
d) When tested using the method described in 7.4.1, the average VMAF score of each output target video clip with adaptive content coding should be at least 3 points higher than that of every output target video clip with CBR transcoding. The standard deviation of the VMAF scores of the output target video clips with adaptive content coding should be at least 40% lower than that of the VMAF scores for the output target video clips with CBR transcoding. The formula used to calculate the standard deviation of VMAF scores can be found in Appendix D.4.

6.4.2 ROI Encoding

The intelligent video processing system should meet the following ROI encoding requirements.

a) The system should support smart ROI encoding.
b) The average bit rate change of ROI encoding in CRF encoding mode should be lower than 5%. The average bit rate change is calculated according to formula (D.5) in Appendix D.

Note: Common ROI areas include faces, human bodies, texts, etc.

6.5 Video Formats

6.5.1 Input Encoding Formats

The intelligent video processing system should meet the following requirements regarding support for input encoding formats.

a) The system should support at least two input encoding formats used for content delivery, such as H.264, H.265, AVS2, etc..
b) The system should support at least two input encoding formats used for post-production, such as Apple ProRes 422, DNxHD-DNxHR, XAVC-I Class 300/480, etc.

6.5.2 Output Encoding Formats

The intelligent video processing system should support at least two output coding formats, such as H.264, H.265, AVS2, etc..

6.5.3 Input Container Formats

The intelligent video processing system should meet the following requirements regarding support for input container formats.

a) The system should support at least five input container formats used for content delivery, such as MP4, MOV, TS, FLV, AVI, MKV, 3GP, etc..
b) The system should support at least two input container formats used for post-production, such as Mov, MXF, AVI, etc..

6.5.4 Output Container Formats

The intelligent video processing system should meet the following requirements regarding support for output container formats.

a) The system should support the MP4 and TS output container formats.
b) It is recommend to provide support for the FLV and MOV output container formats.
c) It is recommend to provide support for DASH protocol, HLS protocol, or CMAF protocol.

7 Test Method

7.1 Smart Video Quality Improvement

7.1.1 Smart Super Resolution

The procedure for testing smart super resolution is as follows:

a) Apply super resolution to an SD source video converted from a 4K video as specified in Appendix C.5, and output a HD (1920 x 1080) target video. The output target video should retain the clarity of the SD source video.
b) Apply super resolution to a 25fps HD source video converted from a 4K video as specified in Appendix C.3, and output a UHD (3840 x 2160) target video. The output target video should retain the clarity of the 25fps HD source video.
c) Apply super resolution to a 25fps HD source video converted from a 4K video as specified in Appendix C.3, and output a UHD 8K (7680 x 4320) target video. The output target video should retain the clarity of the 25fps HD source video.
d) Use the media analysis tool specified in Appendix B to analyze the above target videos, and check whether the three videos meet the requirements in 6.1.1 a), b), and c) respectively.
e) If the output target video from step b) meets expectations, the intelligent video processing system is deemed to satisfy requirement of 2X super resolution.
f) If the output target video from step c) meets expectations, the intelligent video processing system is deemed to satisfy the requirements of 4X super resolution.

7.1.2 Smart Enhancement

The procedure for testing smart enhancement is as follows:

a) Apply smart enhancement to a 25 fps HD source video converted from a 4K video that satisfies the requirements specified in Appendix C.3, and the output target video should retain the sharp edges and texture of the source video and reproduce the subjective experience of a human observer.
b) De-interlace an interlaced source video that meets the requirements specified in C.7. The output target video should have no obvious interlacing.

7.1.3 Smart Frame Interpolation

The procedure for testing smart frame interpolation is as follows:

a) Apply smart frame interpolation to a 25 fps HD source video converted from a 4K video as specified in C.3, and output a 50fps target video.
b) Apply smart frame interpolation to a 15 fps HD source video converted from a 4K video as specified in C.4, and output a 60 fps target video.
c) Use the media analysis tool specified in Appendix B to analyze the target videos created in steps a) and b). The target videos should retain the smoothness of their respective source videos. The target videos should be free of jitter, with no visible flickering.

7.1.4 Smart HDR Conversion

The procedure for testing smart HDR conversion is as follows:

a) Apply smart HDR conversion to the source video, and output a UHD 4K output target video (3840 x 2160; dynamic range: GY/T 315; color space: GY/T 307; bit depth: 10 bits; HDR format: HDR10; encoding standard: H.265, AVS2, or AVS3). The output target video should highlight bright and dark details and have higher color saturation.
b) Use the media analysis tool in Appendix B to analyze whether the output target video from step a) satisfies the requirements in 6.1.4 b), c), d), and e).
c) Apply HLG-to-HDR10 or HLG-to-HDR Vivid conversion to the output target video created in step a), and output a video. The HDR10/HDR Vivid output target video should highlight bright and dark details and have higher color saturation.
d) Use the media analysis tool specified in Appendix B to analyze the output target video from step c). The metadata of the video should be HDR 10 or HDR Vivid.

7.1.5 Comprehensive Picture Quality Improvement

Output the output target video after performing smart super resolution, smart enhancement, smart frame interpolation, and/or smart HDR conversion on the source video.

Compare the output target video with the source video. The test phase, test image display, and scoring scale of the subjective image quality improvement evaluation should comply with the provisions of sections 5.5, 5.6, and 5.7 in GY/T 340-2020. The result analysis and result description should comply with the provisions of sections 5.8 and 5.9 in GY/T 340-2020.

Obtain the score of the source video images and that of the output target video images using the Double Stimulus Continuous Quality Scale method, and calculate the picture quality improvement ratio with formula (D.1) in Appendix D.

Note: The following factors should be taken into account in the scoring process:

a) Clarity, including: clarity of image, obvious blur, respiratory effects
b) Cleanness, including: obvious noise, pixelation or block distortion (not including artistic block effects)
c) Fidelity, including: obvious color loss, improvements in brightness and contrast
d) Smoothness, including: obvious stalling, jitter, ghosting, image lag, improvements in smoothness
e) Improvements in image sharpness
f) Subtitle quality, including: subtitle blur, synchronization of subtitles and images
g) Improvements in overall video quality

7.2 Smart Video Restoration

7.2.1 Scratch Removal

The scratch removal test procedure is as follows:

a) Set the scratch intensity, detect and remove scratches on the source video, and output the target video. The scratch removal effect should be obvious in the target video.
b) Use the media analysis tool specified in Appendix B to detect scratches in the target video.
c) Calculate the Intersection over Union (IoU) of scratch detection using formula (D.6) in Appendix D. If the result is greater than 0.1, the detection is deemed correct.
d) Based on the result of step c), calculate the recall rate of scratch detection using formula (D.2) in Appendix D.

7.2.2 Noise Removal

The noise removal test procedure is as follows:

a) Set the noise intensity, detect and remove noise from the source video, and output the target video. The output target video should preserve image details and clarity should be perceptibly improved.
b) Use the media analysis tool specified in Appendix B to detect noise in the target video.

7.2.3 Intelligent Coloring

Test the performance of the intelligent coloring function, perform intelligent coloring on the source video. The output target video should not flicker or jitter, and the color should be stable.

7.2.4 Comprehensive Restoration and Improvement

Output the output target video after performing scratch removal, noise removal, and/or intelligent coloring on the source video.

Compare the output target video with the source video. The test phase, test image display, and scoring scale of the subjective image quality evaluation should comply with the provisions of sections 5.5, 5.6, and 5.7 in GY/T 340-2020. The result analysis and result description should comply with the provisions of sections 5.8 and 5.9 of GY/T 340-2020.

Obtain the score of the source video images and that of the output target video images using the Double Stimulus Continuous Quality Scale method, and calculate the picture quality improvement ratio using formula (D.1) in Appendix D.

Note: The following factors should be taken into account in the scoring process:

a) Clarity, including: clarity of image, obvious blur, respiratory effects
b) Cleanness, including: significant scratches and noise, the impact of scratches and noise on the subjective experience

c) Fidelity, including: obvious color loss, distortion, improvements in brightness and contrast
d) Smoothness, including: obvious stalling, jitter, ghosting, or image lag, and improvements in smoothness
e) How close the displayed color is to natural color
f) Improvements in overall video quality

7.3 Smart Video Editing

7.3.1 Specific Subimage Removal

The specific subimage removal test procedure is as follows:
a) Remove specific subimages from the source video, and output the target video.
b) Compare the source video with the target video, and record watermark detection and removal data.
c) Calculate the Intersection over Union (IoU) of specific subimage detection using formula (D.6) in Appendix D. If the result is greater than 0.5, the detection is deemed correct.
d) Based on the result of step c), calculate the recall rate of specific subimage detection using formula (D.2) in Appendix D.

7.3.2 Black Bar Cropping

The black bar cropping test procedure is as follows:
a) Remove black bars from the source video and output the target video.
b) Compare the source video with the target video, and record black bar detection and cropping data.
c) Calculate the Intersection over Union (IoU) of black bar detection using formula (D.6) in Appendix D. If the result is greater than 0.95, the detection is deemed correct.
d) Based on the result of step c), calculate the recall rate of black bar detection using formula (D.2) in Appendix D.

7.3.3 Smart Landscape/Portrait Conversion

The test procedure for smart landscape/portrait screen conversion is as follows:
a) Apply the smart landscape-to-portrait conversion to the source video and output the target video.
b) Use the media analysis tool specified in Appendix B to confirm that the video display aspect ratio is 9:16.
c) When the video is switched to a new scene, the system automatically switches to the main picture of the new scene without jitter.

7.3.4 Smart Captioning

The smart captioning test procedure is as follows:
a) Generate subtitles in the ASS or the SRT format automatically from the source video.
b) Compare the subtitles automatically generated with the source subtitles, and calculate the word error rate of the automatically generated subtitles using formula (D.3) in Appendix D.

7.4 Smart Video Encoding

7.4.1 Content Adaptive Encoding

The test procedure for content adaptive encoding is as follows:
a) Set the target resolution to HD (1920 x 1080), perform content adaptive encoding on the source video, and output the target video.
b) Inspect the output target video with the media analysis tool specified in Appendix B. Calculate the average VMAF score of each output target video clip, the differences between the average VMAF scores of the clips, the VMAF standard deviation, and the proportion of single-frame VMAF scores lower than the average VMAF score by more than 5%. The calculation results should meet the technical requirements specified in 6.4.1 b) and c).
c) Set the target resolution to HD (1920 x1080) and the target bit rate to be the same as that in step a), perform CBR encoding on the source video, and output the target video.
d) Inspect the output target video with the media analysis tool specified in Appendix B. Calculate the average VMAF score and VMAF standard deviation for each output target video clip encoded during

step c).

e) Calculate the increased average VAMF score and decreased VMAF standard deviation of each output target video clip output in step a) compared with that output in step c). The calculation results should meet the technical requirements specified in 6.4.1 d).

7.4.2 ROI Encoding

The ROI encoding test procedure is as follows:

a) Set the CRF value to 23 in the CRF mode, perform ROI encoding on the source video, and output the target video.
b) Perform non-ROI encoding on the same source video and output the target video.
c) Use the media analysis tool specified in Appendix B to detect the average bit rates of the two target videos.
d) Calculate the average bit rate change of ROI encoding in the CRF mode using formula (D.5) in Appendix D.

7.5 Video Formats

7.5.1 Input Encoding Formats

7.5.1.1 The test procedure for the input encoding formats used for content delivery is as follows:

a) Upload source videos encoded in H.264, H.265, and AVS2 formats to the system, transcode them into the MP4 format, and output the target videos.
b) Check the transcoding duration and format of the target videos with the media analysis tool specified in Appendix B. The target videos should have the same duration as the source videos and be in the expected format.

7.5.1.2 The test procedure for the input encoding formats used for post-production is as follows:

a) Upload source videos encoded in Apple ProRes 422, DNxHD/DNxHR, XAVC (Intra) Class 300/480 formats to the system, transcode them into the MP4 format, and output the target videos.
b) Check the transcoding duration and format of the target videos with the media analysis tool specified in Appendix B. The target videos should have the same duration as the source videos and be in the expected format.

7.5.2 Output Encoding Formats

The test procedure for the output encoding formats is as follows:

a) Upload the source video encoded in the MP4 format to the system, transcode it into H.264, H.265 and AVS2 formats, and output the target videos.
b) Check the transcoding duration and format of the target videos with the media analysis tool specified in Appendix B. The target videos should have the same duration as the source video and be in the expected formats.

7.5.3 Input Container Formats

7.5.3.1 The test procedure for the input container formats used for content delivery is as follows:

a) Upload source videos encoded in MP4, MOV, TS, FLV, AVI, MKV and 3GP formats to the system, transcode them into the MP4 format, and output the target videos.
b) Use the media analysis tool specified in Appendix B to check the transcoding duration and format of the target videos. The target videos should have the same duration as the source videos and be in the expected format.

7.5.3.2 The test procedure for the input container formats used for post-production is as follows:

a) Upload source videos encoded in Mov, MXF, and AVI formats to the system, transcode them into the MP4 format, and output the target videos.
b) Use the media analysis tool specified in Appendix B to check the transcoding duration and format of the target videos. The target videos should have the same duration as the source videos and be in the expected format.

7.5.4 Output Container Formats

The test procedure for the output container formats is as follows:

a) Upload the source video encoded in the MP4 format to the system, transcode it into MP4, TS, FLV, MOV, DASH, HLS, and CMAF formats, and output the target videos.
b) Use the media analysis tool specified in Appendix B to check the transcoding duration and format of the target videos. The duration of the target videos should be the same as that of the source video, and be in the expected formats.

Appendix A
(Informative)
Test Items and Technical Requirements

The testing of intelligent video processing systems includes subjective evaluation and objective assessment, which are complementary to each other. During the subjective evaluation, evaluators provide subjective comments on the video quality and the improvement of video quality after processing. In objective assessment, technical indicators of system functions are measured. See Table A.1 for the mapping between test items and technical requirements.

Table A.1 Test Items and Technical Requirements

No.	Test Item	Section No. of Technical Requirements
1	Smart super resolution	6.1.1
2	Smart enhancement	6.1.2
3	Smart frame interpolation	6.1.3
4	Smart HDR conversion	6.1.4
5	Comprehensive picture quality improvement	6.1
6	Scratch removal	6.2.1
7	Noise removal	6.2.2
8	Intelligent coloring	6.2.3
9	Comprehensive restoration and improvement	6.2
10	Specific subimage removal	6.3.1
11	Black bar cropping	6.3.2
12	Smart landscape/portrait conversion	6.3.3
13	Smart captioning	6.3.4
14	Content adaptive encoding	6.4.1
15	ROI encoding	6.4.2
16	Input encoding formats	6.5.1
17	Output encoding formats	6.5.2
18	Input container formats	6.5.3
19	Output container formats	6.5.4

Appendix B
(Normative)
Test System Configurations

The test system mainly comprises hardware such as display and computer, and software such as media analysis tool and VMAF tool. The test system configuration of each test item is shown in Table B.1. The display device shall comply with the requirements in Table 2 of GY/T 340-2020. The computer shall comply with the requirements of GB/T 9813.1, GB/T 9813.2, or GB/T 9813.3.

Table B.1 Test System Configurations

No.	Test Item	Display	Computer	Media Analysis Tool	VMAF tool	Subjective Evaluation Personnel
1	Smart super resolution	▻	▻	▻		
2	Smart enhancement	▻	▻			
3	Smart frame interpolation		▻	▻		
4	Smart HDR conversion		▻	▻		
5	Comprehensive picture quality improvement	▻	▻			▻
6	Scratch removal		▻	▻		
7	Noise removal		▻	▻		
8	Intelligent coloring		▻			
9	Comprehensive restoration and improvement	▻	▻			▻
10	Specific subimage removal	▻	▻			
11	Black bar cropping	▻	▻			
12	Smart landscape/portrait conversion	▻	▻	▻		
13	Smart captioning	▻	▻			
14	Content adaptive encoding	▻	▻	▻	▻	
15	ROI encoding	▻	▻	▻		
16	Input encoding formats	▻	▻	▻		
17	Output encoding formats	▻	▻	▻		
18	Input container formats	▻	▻	▻		
19	Output container formats	▻	▻	▻		
Example 1: Media analysis software such as MediaInfo Example 2: VMAF tools such as FFMPEG						

Appendix C
(Normative)
Source Video Configurations

C.1 General Provisions

The testing of an intelligent video processing system requires 11 source videos with different test data sets, including 4K video, 4K to 25fps HD video, 4K to 15fps HD video, 4K to SD video, content adaptive encoding video, interlaced video, video converted from old film stock, video with specific subimages, video with black bar, video with smart landscape/portrait conversion, and video with subtitles. The 4K source video is the basis on which the other videos are developed. The configuration of each source video used in the test is shown in Table C.1.

Table C.1 Source Video Configurations

No.	Test Item	Source Video										
		4K	4K to 25 fps HD	4K to 15 fps HD	4K to SD	Content adaptive encoding	Interlaced	Old film stock	Specific subimage	Black bar	Landscape/portrait conversion	Subtitles
1	Smart super resolution		▻		▻							
2	Smart enhancement		▻				▻					
3	Smart frame interpolation		▻	▻								
4	Smart HDR conversion		▻									
5	Comprehensive picture quality improvement		▻									
6	Scratch removal							▻				
7	Noise removal							▻				
8	Intelligent coloring							▻				
9	Comprehensive restoration and improvement							▻				
10	Specific subimage removal								▻			
11	Black bar cropping									▻		
12	Smart landscape/portrait conversion										▻	
13	Smart captioning											▻
14	Content adaptive encoding					▻						
15	ROI encoding		▻									
16	Input encoding formats		▻									
17	Output encoding formats		▻									
18	Input container formats		▻									
19	Output container formats		▻									

C.2 4K video

The 4K video shall meet the following requirements:

a) The video format is UHD 4K (3840×2160).
b) The total number of segments is not less than 20.
c) The content of each video segment is dynamic and lasts for 10 to 15 seconds.
d) The content includes both images with rich details and textures and images with text.
e) The content includes images of human bodies and faces with different skin colors.
f) The content includes images of objects in motion and the frame rate is 50 fps.

C.3 4K to 25 fps HD video

The 4K to 25 fps HD video shall meet the following requirements:

a) It is an HD (1920×1080) video converted from the UHD 4K (3840×2160) video in C.2.
b) The color conversion space of the HD (1920×1080) video is BT.709, the bit depth is 8 bits, the frame rate is 25 fps, and the dynamic range is SDR.
c) The source video image should score 3 points on the double-stimulus continuous quality scale for subjective evaluation.

C.4 4K to 15 fps HD video

The 4K to 15 fps HD video shall meet the following requirements:

a) It is an HD (1920×1080) video converted from the 4K source video in C.2.
b) The color conversion space of the HD (1920×1080) video is BT.709, the bit depth is 8 bits, the frame rate is 15 fps, and the dynamic range is SDR.

C.5 4K to SD video

The 4K to SD video shall meet the following requirements:

a) It is an SD (720×576) video converted from the UHD 4K (3840×2160) video in C.2.
b) The color conversion space of the SD (720×576) video is BT.709, the bit depth is 8 bits, the frame rate is 25 fps, and the dynamic range is SDR.

C.6 Content adaptive encoding video

The content-adaptive encoding video shall meet the following requirements:

a) The total number of segments is not less than three.
b) Each video segment lasts for at least five seconds.
c) Each video segment contains at least five scenes, and the content includes animation, sports, games, animals, outdoor environments, human faces, movies, advertisements, and music videos.

C.7 Interlaced video

The interlaced video shall meet the following requirements:

a) The total number of segments is not less than five.
b) The content of each segment is dynamic and lasts for 10 to 15 seconds.
c) The resolution of the source video is at least SD (720×576).
d) The video uses interlaced rendering.

C.8 Old film stock

The video converted from old film stock shall meet the following requirements:

a) The total number of segments is not less than 20.
b) The content of each segment is dynamic and lasts for 10 to 15 seconds.
c) Images with scratches are included.
d) Noisy images are included. (Noises may include snow noise, salt and pepper noise, Gaussian noise, artifacts, block effects, blurring, and speckles.)
e) Black and white video is included.

C.9 Video with specific subimages

The video with specific subimages are provided by the vendor and shall meet the following requirements:

a) The total number of segments is not less than 50.

b) The content of each segment is dynamic and lasts for 10 to 15 seconds.
c) The resolution of the video is at least SD (720×576).
d) The content contains 20 different specific subimages that have been included in the system.
e) The specific subimages persist for the duration of the video.

C.10 Video with black bars

The video with black bars shall meet the following requirements:
a) The total number of segments is not less than 20.
b) The content of each segment is dynamic and lasts for 10 to 15 seconds.
c) The resolution of the video is at least SD (720×576).
d) 10-20% of the video samples have no black bars.
e) Samples with a black bar area of less than 5% are included.
f) Samples with black bars on the left and right are included.
g) Samples with asymmetric black bars are included.

C.11 Video with landscape/portrait conversion

The video with landscape/portrait conversion shall meet the following requirements:
a) The total number of segments is not less than 20.
b) The content of each segment is dynamic and lasts for 10 to 15 seconds.
c) The resolution is at least SD (720×576).
d) The image is in landscape orientation with a 4:3 or 16:9 aspect ratio.
e) Human figures are the main subjects of the video segments.
f) Multiple subjects appear in the images.
g) Video content includes scene switching.

C.12 Video with subtitles

The video with subtitles shall meet the following requirements:
a) The total number of segments is not less than 20.
b) The content of each segment is dynamic and lasts for 10 to 15 seconds.
c) The resolution is at least SD (720 x 576).
d) The video contains Chinese subtitles, placed in the bottom fifth of the image.
e) The subtitles are clear and easily readable on the display.
f) Subtitles of each segment contain at least 10 Chinese characters.

Appendix D
(Normative)
Formula of Technical Indicator

D.1 Image quality improvement ratio

The image quality improvement ratio is calculated according to formula (D.1).

$$E=(\mathrm{b}-\mathrm{a})/a\times 100\% \quad \text{(D.1)}$$

Where:
E: image quality improvement ratio;
a: subjective evaluation score of the source video;
b: subjective evaluation score of the target video.

D.2 Recall rate

The recall rate is calculated according to formula (D.2).

$$\mathrm{Recall}=\frac{S_d}{S_l} \quad \text{(D.2)}$$

Where:
Recall: recall rate;
S_d: number of objects correctly detected;
S_l: number of objects in the video.

D.3 Word error rate

The word error rate is calculated according to formula (D.3).

$$\mathrm{WER}=\frac{S+D+I}{S+D+C} \quad \text{(D.3)}$$

Where:
WER: word error rate;
S: number of words incorrectly replaced;
D: number of words not recognized;
I: number of words not present in the identified verification data;
C: number of words correctly identified.

D.4 VMAF standard deviation

The standard deviation of VMAF is calculated according to formula (D.4).

$$v_{std}=\sqrt[2]{\frac{\sum_{i=1}^{n}(v_i-\bar{v})}{n-1}} \quad \text{(D.4)}$$

Where:
*i:*sequence number of each image frame in the video segment;
v_{std}: VMAF standard deviation of the video segment;
v_i: VMAF of each iamge frame in the video segment;
$\bar{v}$: average VMAF of the video segment.

D.5 Average bit rate change of ROI encoding in CRF mode

The average bit rate change of ROI encoding in CRF mode is calculated according to formula (D.5).

$$\mathrm{Change}=\frac{ABS(B_{ROI}-B_{normal})}{B_{normal}} \quad \text{(D.5)}$$

Where:
Change: bit rate change;
B_{ROI}: average bit rate of ROI encoding;
B_{normal}: average bit rate of normal encoding.

D.6 Intersection over Union

The intersection over union is calculated according to formula (D.6).

$$IoU=\frac{P_d\cap P_l}{P_d\cup P_l} \quad \text{(D.6)}$$

Where:
IoU: intersection over union;
P_d: the collection of detected object pixels;
P_l: the collection of real object pixels.

T/UWA 010-2022

UHD World Alliance Standards

General technical specifications for intelligent video processing systems
T/UWA 010-2022

*

Released by China Ultra HD Video Industry Alliance

*

Sheet: 880 × 1230 1/16

Printed sheet: 1 $\frac{1}{2}$

Number of words: 36,000

First edition in February 2022

First print in February 2022

Number of prints: 200 volumes