Detailed Information about the IVP Subjective Quality Video Database

Image and Video Processing (IVP) Lab

The Chinese University of Hong Kong

1. Source Videos

Ten uncompressed, high definition (1920×1088) videos are used as source videos. One video which contains animations were rendered directly from 3D models, while other videos which contains natural scenes were shot with professional, high end equipment and converted to digital format with utmost care. The videos shot by us were recorded in a progressive RAW format at 25 fps without any compression. Before generating the distortions, the source video was transformed from RAW formant to YUV 4:2:0. The videos show diverse contents as listed in Table 1. Each video has duration about 10 sec and no audio component is appended.

Table 1. Source videos.

Video name	Content features	Ownership
Laser	Sparkling instruments in a dark background	IVP lab
Bus	With a fade-in shot transition	IVP lab
Overbridge	Head-on pedestrians under shadow	IVP lab
Robot	Cartoon animation appended with a text watermark	IVP lab
Shelf	Panning and then defocusing/refocusing	IVP lab
Square	Passersby encountering on an open square	IVP lab
Toy_calendar	Colorful scene	VQEG
Tractor	Tracking a tractor by telephoto shooting and then zooming out	VQEG
Train	Tracking a train by a wide-angle shooting	IVP lab
Tube	Zooming into a bright and structural instrument	IVP lab

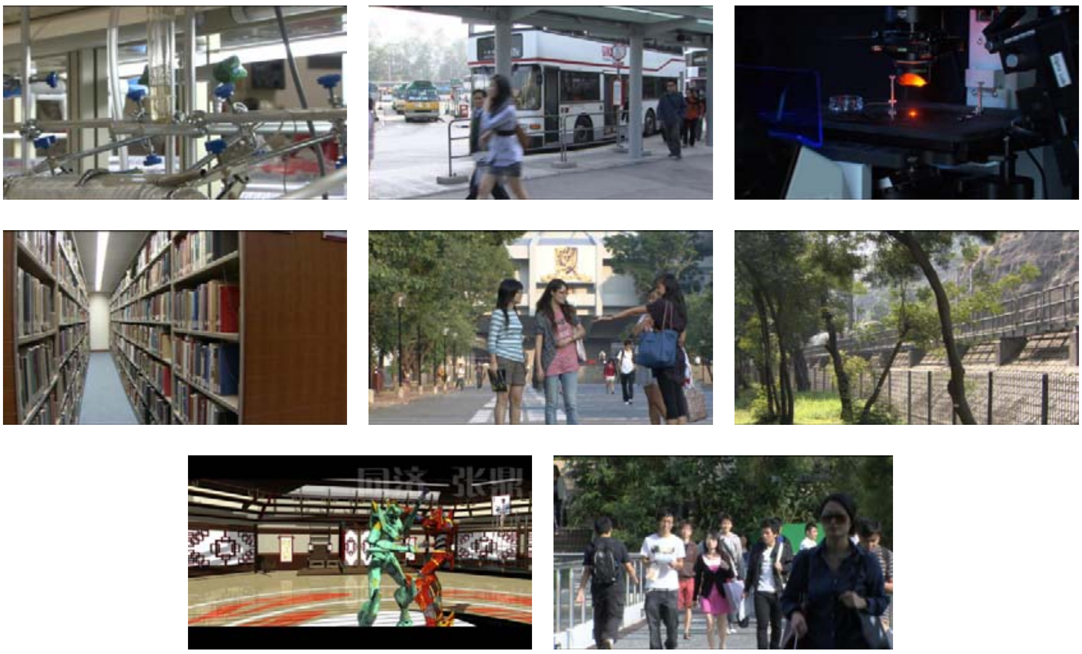

Figure 1. Sample images of source video contents.

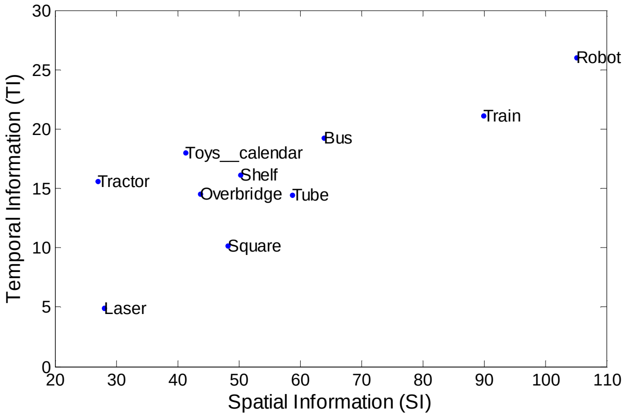

Figure 2. The distribution of content complexity of source video contents. The SI and TI are calculated according to ITU-T P.910.

2. Distorted Videos

Distorted sequences are generated using 4 types of distortions: MPEG 2 compression, Dirac wavelet compression, H.264 compression and packet loss on the H.264 streaming through IP networks. MPEG 2 is a traditional standard in broadcasting society, H.264 is rapidly gaining popularity due to its superior compression efficiency, and Dirac wavelet is also competitive codec employed by BBC (British Broadcast Corporation). Videos compressed by MPEG 2 or H.264 exhibit typical coding artifacts such as blocking, blur, ringing and motion compensation mismatches around object edges. Videos compressed by Dirac wavelet coding (Intra mode only) have few block artifacts but show ringing graininess and blotchiness. Different from the uniform distortions in compressed videos, network loss leads to the transient and uneven distortions. To ensure an enough span of video qualities, the distortion strengths were diversified.

(1) Coding*:* We compressed each source video at five bit rate levels including 1, 2, 4, 7, and 10 Mbps by each codec, among which three, videos are compressed by MPEG, three by Dirac-wavelet and four by H.264 codec.

(2) Packet loss. Firstly, each source video (except *bus*, *train*, and *toys_calendar*) was compressed and encapsulated to packets by H.264 with a fixed QP of 20, where each frame of a sequence is divided into a fixed number of slices and the number may vary from 8 to 140 for different sequences. Then, packet loss was simulated at a rate of 0.1%, 0.5%, 1%, 3% or 5%, where the error patterns to guide the packet loss were obtained from real-world experiments on congested networks and are recommended by the VCEG (Video Coding Experts Group) to simulate the Internet backbone performance for video coding experiments. (To ensure successful decoding, we did not drop the first several packets containing coding parameters). Finally, for each source video (except *bus*, *train*, and *toys_calendar*), four lossy streams were selected to be decoded with default error concealment mechanism of JM software and flexible macroblock ordering (FMO).

3. Subjective Tests

We conducted the subjective evaluation in IVP Lab of CUHK. The evaluation was performed in a studio room with lighting condition satisfying the lab environment requirement of the ITU-R BT.500 standard. The display monitor is a 65’’ Panasonic plasma display (TH-65PF9WK) and the viewing distance is 3 times of the picture height. Background illumination has a D65 chromaticity. We used a high-performance workstation stored and displayed all videos in a format of raw TIF sequences. This configuration can avoid any errors in displaying the HD video such as latencies or frame drops.

42 paid observers participated in the subjective test, including 11 females and 31 males. Among them, 25 observers did not work in video processing related jobs and were not involved in any video quality evaluation within the past four months, and thus were regarded as “non-experts”. The other 17 observers were from IVP Lab, who did research on image processing and were regarded as “experts”. Their eyesight was either normal or had been corrected to be normal with spectacles. In this database, we provide three files of the subjective scores, which are derived from non-experts, experts, and all observers, respectively.

Each observer assessed 10 source videos and 128 distorted videos. The single-stimulus quality scale test method (ACR) was used, where each video (including the reference) occurred once in a random order yet the two successive videos must come from different source videos so as to remove contextual and memory effects in quality evaluation. Between the presentations of two videos, a mid-gray video in 5 second was displaying, and meanwhile the evaluation can be reported on the five-point scale: 5-Excellent, 4-Good, 3-Fair, 2-Poor, and 1-Bad. The evaluation consisted of two sessions, between which the observers had a ten-minute rest. A source video and the distorted video generated from it must be contained in the same session. At the beginning of each session, five videos were arranged as the training videos to stabilize the observers’ opinion.

Figure 3. Subjective viewing test environment.

4. Post-processing of subjective scores

The post-processing of subjective scores includes three steps. Firstly, every group of scores which are rated by each subject in each session (a total of two sessions) are mapped to the Z scores, such that each group of scores have zero mean and unit variance; secondly, subject rejection by β_2 test is carried on as suggested in BT. 500; Thirdly, all the unrejected Z scores are mapped back by linear regression (i.e. inverse Z transform). DMOS and their standard deviations are calculated finally. The procedures above are conducted on experts’ scores, non-experts’ scores and all observers’ scores respectively. Correspondingly, 2 out of 17 experts, 2 out of 25 non-experts, and 7 out of all observers are rejected.