THE SJTU 4K VIDEO SEQUENCE DATASET

Li Song[1], Xun Tang[2], Wei Zhang[1], Xiaokang Yang[1], Pingjian Xia[2]
[1]Institute of Image Communication and Network Engineering, Shanghai Jiao Tong University
[2]National Engineering Research Center of Digital Television, Shanghai, 200125 China

ABSTRACT

This short paper presents a set of 15 new 4K resolution ultra-high definition (UHD) video sequences for catering the requirement of active UHD video quality assessment algorithm in coming years, as well as help to fully evaluate coding efficiency of latest HEVC (High Efficient Video Coding). The details of processing procedures and sequences characteristics are illustrated.

***Index Terms*—**UHD, video dataset, video quality assessment, HEVC

1. INTRODUCTION

As high definition (HD) video applications have been becoming increasingly popular, UHD applications also are becoming a trend in the future. The rising demands for an enhanced end-user experience with high fidelity content especially for home theater and premium events like sports push the development of the related technology.

Last year, the International Telecommunication Union (ITU) has officially approved UHDTV as a standard, conjointly standardizing both 4K and 8K resolutions in [1]. Preserving the 16/9 picture aspect ratio, 4K is defined as 3840×2160 pixels and 8K as 7680×4320 pixels. Associated with the UHD video applications, a new coding standard, HEVC under development by ISO MPEG and ITU-T VCEG, was also approved by the ITU-T on 25th January 2013, which was originally designed and optimized for better-perceived visual quality on large image hence targeting Ultra High Definition coding.

Before a video application coming into service, a great number of researches need to be conducted for obtaining ideal results, including performance evaluation for the systems and algorithms. The evaluations of the video quality assessment algorithms and coding algorithms are the most common example. So far, there have been many datasets already widely available, and most related information were collected in [2][3], including LIVE video database[4], TUM HD video dataset [5] and so on. However, except that elemental recently released some clips recorded by the RED Epic 4K camera [6], the majority of these datasets focused on smaller resolutions, no more than HD resolution. Along this direction, we contribute a set of new 4K@30fps video sequences with diverse characteristics. Our shot and post-processing team are from Shanghai Jiao Tong University (SJTU) and National Engineering Research Center of Digital Television (NERC-DTV). We expect it can help coming evaluation of UHD video quality, HEVC/H.265 encoding performances and so forth.

2. 4K VIDEO SEQUENCES

2.1. Shooting and processing

Camera manufacturers which were involved in cinema technologies or 4K displays, have recently launched a few cameras designed for 4K content creation. F65, the masterpiece of Sony, was designed for premium cinema productions recording. All of the UHD video sequences included in our dataset were shot using Sony F65 camera. The photoelectric signals of images received from 8K CMOS sensor were stored in the disk arrays. Then the signal was converted to digital format with utmost care, guaranteeing that the video sequences are distortion free. The raw data of the images quantified with 10 bit were exported at resolution of 3840x2160 and frame rate of 30fps with DPX format. The resulted picture quality provides the actual highest fidelity of contrast and colors. The DPX image files were combined and converted into uncompressed YUV files using FFMPEG software. Both YUV 4:4:4 color sampling, 10 bits per sample and YUV 4:2:0 color sampling, 8 bits per sample formats were offered in our dataset. The former retains the original information and the later is the most common file format for compressing. All video sequences presented in our dataset are of 300 frames.

2.2. Sequences characteristics description

Considering that video content plays a key role in the related researches, we attempt to shoot video sequences which can be representative of a wide variety of content types. All scenes are chosen in Shanghai, China. The factors such as image texture, image detail, movement speed of the object in the image, light intensity, and the camera lens stretching, panning are taken into account. Some sequences are shot from an overlooking angle by using tripod. The first frame of each video sequences is showed in our project web page. The detail descriptions for video sequences in our dataset are shown as follows.

- **Bund Nightscape**---Shows nightscape of the gallery of International Architecture in Bund.

- **Campfire Party**---Shows flames before the crew of the NERC-DTV in a campfire Party.
- **Construction Field**---Shows some excavator arms in a construction area.
- **Fountains**---Shows the artificial fountains jet water into the air in front of a tall building
- **Library**---Shows sparsely moving students in the front of the new library.
- **Marathon**---Shows the scene of the early stages of 2012 Shanghai International Marathon Race.
- **Residential Building**---Shows an residential building. The dense texture of the outside wall and intricate lines are presented.
- **Runners**---Shows many runners in midway of 2012 Shanghai International Marathon Race.
- **Rush Hour**---Shows many students on the ways to canteen or dormitory after classes.
- **Tall Buildings**--- Shows high buildings in Lujiazui, Pudong New District in Shanghai.
- **Traffic and Building**---Shows many cars on the road and distant buildings.
- **Traffic Flow**---Shows roadways with the automobiles moving in a different direction.
- **Tree Shade**---Shows the shade under a large tree.
- **Scarf**---Shows exhibition of an accessory shop and reveals diverse contrasts and colors.
- **Wood**---Shows a wood in the campus of SJTU. Sunlight pierce through the woods

Generally, the spatial and temporal information were used as representing the video content. The spatial and temporal perceptual information of the scenes are critical parameters. These parameters play a crucial role in determining the amount of video compression and exert an important influence on the corresponding researches to some extent.

The analysis of content classification has been performed by computing the SI and TI indexes on the luminance component of each video sequences according to [7]. And the calculation process can be represented in equation form as:

$$SI = max_{time}\{std_{space}[Sobel(F_n)]\} \quad (1)$$

$$TI = max_{time}\{std_{space}[F_n(i,j) - F_{n-1}(i,j)\} \quad (2)$$

Here $Sobel()$ stand for *Sobel* filter and $F_n(i,j)$ is the pixel at the ith row and jth column of nth frame in time. All SI and TI indexes of the dateset are indicated in Figure 1.

The size of all 15 sequences is about 270GB, with both YUV 4:4:4 color sampling, 10 bits per sample and YUV 4:2:0 color sampling, 8 bits per sample formats. All sequences can be downloaded from our public server through the following link:
http://medialab.sjtu.edu.cn/web4k/index.html

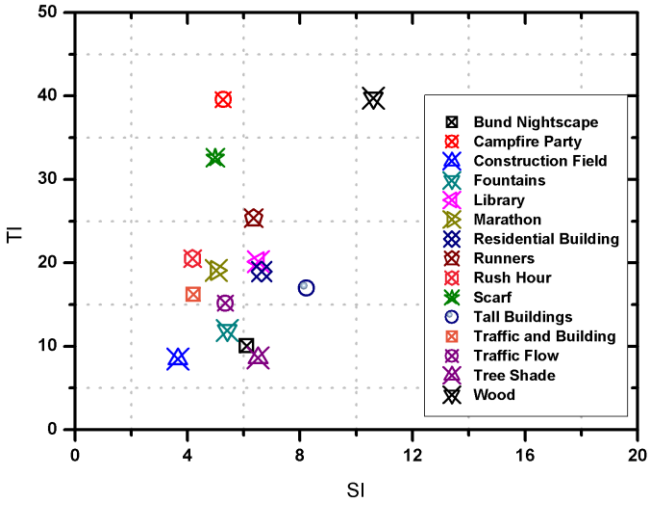

Figure 1. Spatial Information (SI) and Temporal Information (TI) indexes of the 4K video sequences

3. SUMMARY

This paper presents a new batch of 4K video sequences which will be opened to related research circles. The details of processing procedures and sequences characteristics are illustrated.

4. ACKNOWLEDGEMENTS

This work was supported by National 863 project (2012AA011703), National Key Technology R&D Program of China (2013BAH53F04), NSFC (61221001, 61271221), the 111 Project (B07022) and the Shanghai Key Laboratory of Digital Media Processing and Transmissions.

5. REFERENCES

[1] ITU Recommendation BT.2020, "Parameter values for ultra-high definition television systems for production and international program exchange," 2012.

[2] S. Winkler, "Image and video quality resources," 2012 [Online]. Available: http://stefan.winkler.net/resources.html

[3] S. Winkler: "Analysis of public image and video databases for quality assessment." *IEEE Journal on Selected Topics in Signal Processing*, vol. 6, no. 6, Oct. 2012.

[4] K. Seshadrinathan, R. Soundararajan, A. C. Bovik *et al.* "LIVE video quality database," 2010 [Online]. Available:http://live.ece.utexas.edu/research/quality/live_video.html

[5] C. Keimel, A. Redl, K. Diepold, "The TUM high definition video datasets," *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on* , vol., no., pp.97-102, 5-7 July 2012

[6] http://www.elementaltechnologies.com/resources/4K-test-sequences

[7] ITU-T Recommendation ITU-R P910, "Subjective video quality assessment methods for multimedia applications". Sep. 1999.