Mobile Multispectral Video Streaming

Linsen Chen
Nanjing University
njucls@gmail.com

Cc Dong
Nanjing University
dong_chen@smail.nju.edu.cn

Du Chen
Nanjing University
njuchendu@gmail.com

Han Li
Nanjing University
hanli1994nju@gmail.com

Yuanyuan Zhao
Nanjing University
Yuan_square@smail.nju.edu.cn

Xun Cao
Nanjing University
caoxun@nju.edu.cn

Zhan Ma*
Nanjing University
mazhan@nju.edu.cn

ABSTRACT

The Multi-Spectral (MS) image provides 10× or 100× extra information compared with the RGB content. It plays an important role in various applications such as material identification, object tracking, environment monitoring, etc. However, complex and bulky optical subsystem of the acquisition setup and ultra high volume of the spectral data have severely hindered the widespread adoption of MS image/video based applications. This paper presents a mobile MS video streaming system, including the compact hand-held MS video acquisition, transmission, and rendering in the back-end servers. All of these subsystems are realized using low-cost off-the-shelf optical and electrical components, resulting in a small form factor mobile device. It enlarges the potential of applying the MS videos in different areas. Towards this purpose, we have demonstrated two real-time simulations, including environment monitor using material discrimination and criminal investigation using human-face tracking, by leveraging the spectral characteristics of the MS video that are not contained in conventional RGB content.

KEYWORDS

Multi-spectral image/video, acquisition, compression, analysis and mobile system

1 INTRODUCTION

The spectrum is a characteristic distribution of the electromagnetic radiation in the full frequency range that a particular object absorbs or emits. Traditional RGB imaging only obtain three spectral channels (i.e., Red, Green, Blue), which are the down-sampled visible spectrum ranging from 400 to 700nm. MS video is a four-dimensional cube, including two spatial dimensions, one spectral dimension and one temporal dimension [14, 30], as shown in Fig. 1.

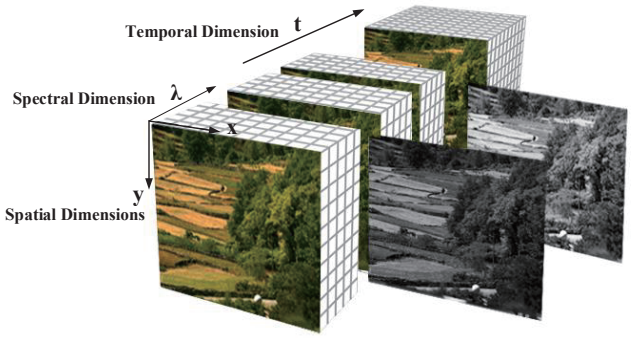

Figure 1: An illustration of the four-dimensional MS video cube, including two spatial dimensions, one spectral dimension and one temporal dimension.

Compared with the RGB content, the MS video has numerous times channels in the spectral dimension, with which the inherent material characteristics of different scenes or objects can be precisely obtained. Therefore, the spectral data is widely used in material identification, biological sciences, criminal investigation, environmental monitoring and many other fields [6, 11, 12, 16, 28].

Constrained by the spectral sampling bandwidth, the traditional spectrometers generally use the spatial scanning or spectral filtering scheme for MS image acquisition. The typical systems are illustrated as follows: A spectral filtering spectrometer commonly utilize numerous narrow bandpass color filters or a liquid-crystal tunable device, and the images at different spectral bands can be recorded by successively switching the filters [20]; A spatial scanning imager, such as whisk-broom [15] or push-broom spectrometer [12], scans entire scene to obtain the distinct intensity at each wavelength. These techniques generally sacrifice temporal resolution for high spectral resolution. Generally, these scanning-based systems fail to capture the dynamic scenes [24].

Recently, with the development of compressive sensing theory, some down-sampling and computational reconstruction based spectral acquisition systems have been developed, such as Computed-Tomography Imaging Spectrometer (CTIS) [13], Coded Aperture Snapshot Imagers (CASSI) [25] and Image Mapping Spectroscopy (IMS) [16]. CTIS uses a diffraction grating to generate multiple dispersed 2D projections and reconstructs a 3D spectral datacube

*Corresponding author.

ThematicWorkshops'17, October 23–27, 2017, Mountain View, CA, USA.
 ISBN 978-1-4503-5416-5/17/10…$15.00
DOI: https://doi.org/10.1145/3126686.3126762

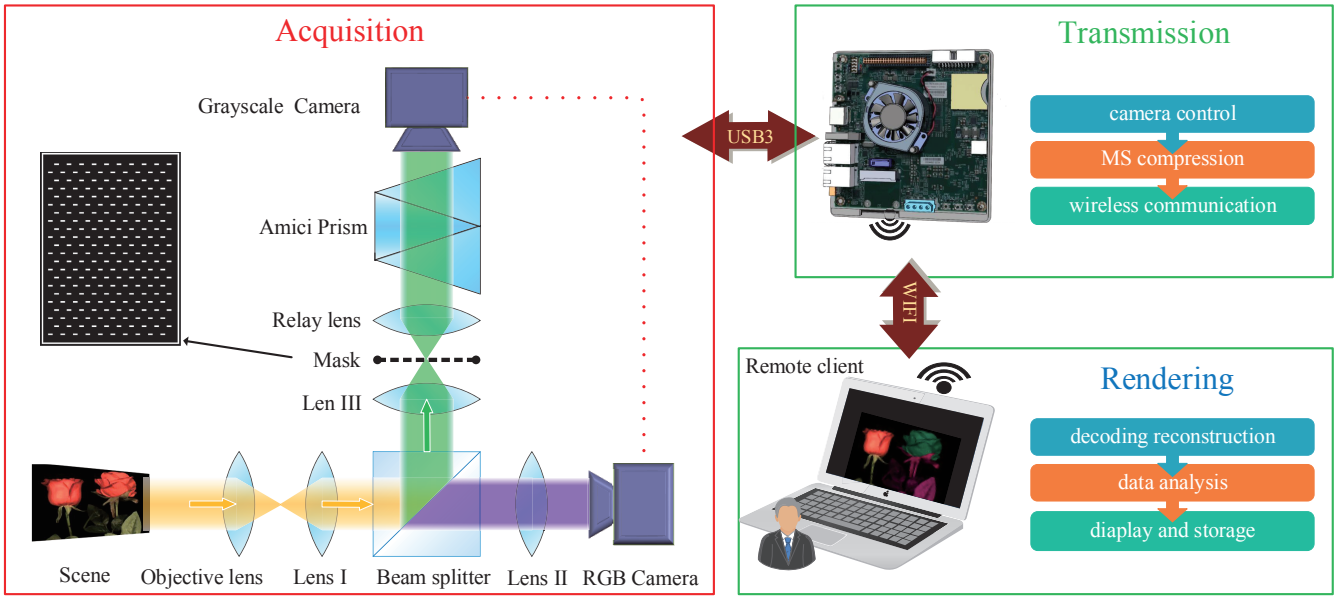

Figure 2: The schematic diagram of the proposed mobile MS video streaming system.

from these 2D projections. Without any filters, CTIS has advantages in light conversion efficiency. While only a limited number of 2D projections can be recorded in CTIS, it faces a inevitably ill-posed reconstruction problem, which is commonly known as missing cone problem. CASSI which uses one or more well-designed coded apertures and diffractive prisms or gratings, can estimate instantaneous 3D spectral data cube from the 2D coding and compression of CASSI measurements, but the spectral reconstruction algorithm is quite computational cost. IMS uses an array of densely packed tiny mirror facets and map the neighboring image zones to the isolated area on the CCD, which can achieve MS images at frame rates of 5.2 frames per second. These systems aforementioned have capacity of 3D multispectral image cube snapshot with the sacrifice of high-cost, complicated optical designs. On the other hand, the data throughput of the MS video, e.g. 2048×1536 spatial pixels, 140 spectral channels, 15 frame rate and 8 bit per pixel, is about 52.848Gbps. How to capture, transmit and render such large volume data has become the bottleneck for MS video being widely adopted in the market.

Main contributions of this paper are summarized as follows:

- We propose a flexible MS video streaming framework, consisting of MS video acquisition, transmission and rendering subsystems, where we can adapt resolutions (i.e., spatial resolution, temporal resolution) to afford various applications, such as real-time tracking, or high-resolution reconstruction;
- We build a **compact light-weight dual-camera system** ($210mm \times 190mm \times 40mm$, 1.3kg) for MS video acquisition at high spatial resolution (2048×1536), high spectral resolution ($400 - 1000$nm, 140 channels) and high temporal resolution (15 fps), which does not require specially-made optical devices and is easy to build with just a few off-the-shelf components;
- We develop appropriate algorithms to compress and transmit the raw videos from the dual-camera to the back-end in a **wireless fashion** (5Mbps bit rate for MS video transmission);
- We also demonstrate the environment monitor using material discrimination and criminal investigation using human-face tracking simulations by leveraging the spectral characteristics of the MS video.

The rest of the paper is organized as follows: the MS video acquisition, transmission and rendering subsystems will be provided in Section 2. Experimental applications are presented in Section 3 to demonstrate the effectiveness of our proposed system. Finally, the conclusion is given in Section 4.

2 SYSTEM ARCHITECTURE AND IMPLEMENTATION

Rather than capturing the 3D MS Image cube, we build up the dual-camera acquisition system to record two projections of the 3D data cube, namely, the high-spatial-resolution low-spectral-resolution RGB images and the low-spatial-resolution high-spectral-resolution MS images. Then, we conduct the well-designed compression algorithm on the two projections, through which the MS video data volume can be reduced sharply while preserving high spectra fidelity. Finally, reconstruction and rendering are conducted in the back-end servers (i.e., possibly in the cloud).

2.1 MS video Acquisition

2.1.1 System Configuration. Inspired by recent MS video acquisition works [8, 9], we redesign the optical system and build a compact, light-weight mobile MS video acquisition system as shown in

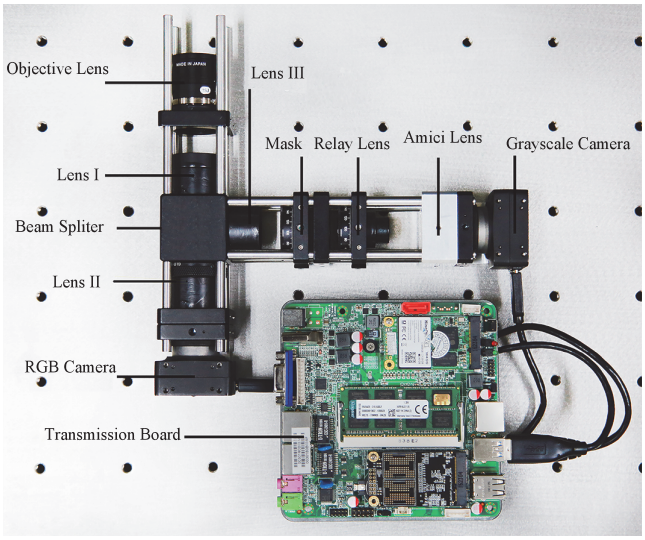

Figure 3: The compact dual-camera MS video acquisition and transmission subsystems.

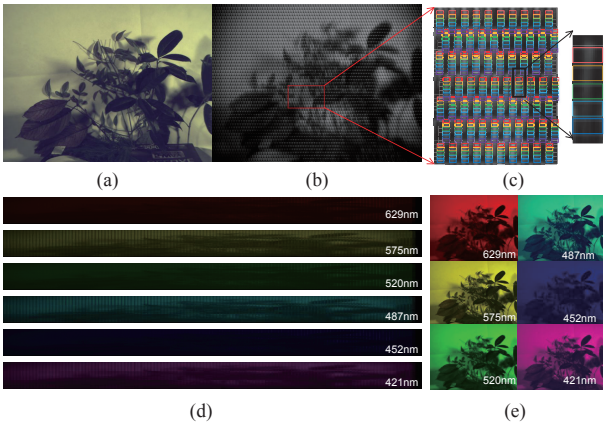

Figure 4: (a) The RGB video frame captured by the MS video acquisition system; (b) The grayscale video frame captured by the MS video acquisition system; (c) Stripes in the grayscale image. Light with different wavelength are dispersed vertically within the stripes and collected by different pixels; (d) Single-spectral-channel images. Pixels that collect light at the same wavelength are spliced together to form single-spectral-channel images; (e) The high-spatial-resolution MS video frame which are reconstructed from the RGB and low-spatial-resolution spectral images.

Fig. 2. The incoming light is collected by the objective lens, collimated by achromatic Lens I, and then divided by the beam splitter, which reflects half of the light along violet path and transmit the remainder along the green path. In the violet path, the transmitted light is converged by achromatic Lens II and collected by the RGB camera to form a high-spatial-resolution RGB image. In the green path, the reflected light is converged to the mask surface by achromatic Lens III, then it is spatially sub-sampled by the occlusion mask. The sub-sampled light transmits through the relay lens and then dispersed by the Amici prism, finally collected by the grayscale camera, which measures numerous spectral channels of the scene at a low spatial resolution. To ensure that the entire optical components are fixed along the same optical axis, the RGB and grayscale cameras share the same Field of View (FoV). On the other hand, the triggering signal of the two cameras is synchronized so that the high spatial RGB and the low spatial spectral videos can be simultaneously captured.

2.1.2 System Implementation. As shown in Fig. 3, the objective lens are off-the-shelf commercial C-mount lens, so that it can be easily replaced with other C-mount lens to meet different FoV demand. The other lenses used in our system are all well-designed and have a high transmittance percentage across the spectral range from 400nm to 1000nm. The grayscale camera is PointGrey CM3-U3-31S4M-CS, which can capture 55fps grayscale video at the spatial resolution of 2048×1536. It has a high QE (Quantum Efficiency) value at the infrared wavelength. The RGB camera is PointGrey CM3-U3-31S4C-CS, which can capture 55fps RGB video at the spatial resolution of 2048 × 1536. The occlusion mask is configured as shown in Fig. 2, light passes through the white rectangle holes and dispersed by the Amici prism. The **light throughput** is decided by the size of each white rectangle, thus too large size leads to the spectrum blur. Therefore, in the proposed system, each white rectangle is mapped to 2 × 3 pixels on the detector. The **spectral resolution** is decided by the ratio between the distance of mask - relay lens and relay lens - detector, which can be easily adjusted for various application fields. We then can get the the aligned RGB and grayscale video streams after necessary calibration and registration.

2.2 MS Video Transmission

In this section, the MS video compression algorithm is described first. Then, we propose a novel compression assessment metric, based on which we discuss how to choose the optimal parameters in practice.

2.2.1 RGB and Grayscale Video Compression. We simultaneously obtain the grayscale and the RGB video streams from the front-end acquisition subsystem, as shown in Fig. 4(a) and (b). Leveraging their intrinsic features, we propose the separate compression pipeline as follows.

Compression on grayscale video stream: Each incoming light ray is sub-sampled by the occlusion mask, dispersed by the Amici prism, and finally form numerous continuous spectral stripes on the grayscale camera as shown in Fig. 4 (b). Our compression algorithm is based on two basic ideas: (1) The cross-channel redundancy generally exists along the spectral dimension; (2) Pixels between adjacent stripes are not useful in the spectral extraction process. Therefore, we only extract the pixels in the stripes, as illustrated in Fig. 4 (c), and generate the single-spectral-channel images as shown in Fig. 4 (d). Then, the single-spectral-channel images are compressed using H.264 [27] with the implementation of the Intel-QSV technology to accelerate coding process. As known, Intel Quick Sync Video (QSV) is dedicated video encoding and decoding hardware core. Unlike video encoding on a CPU or a GPU,

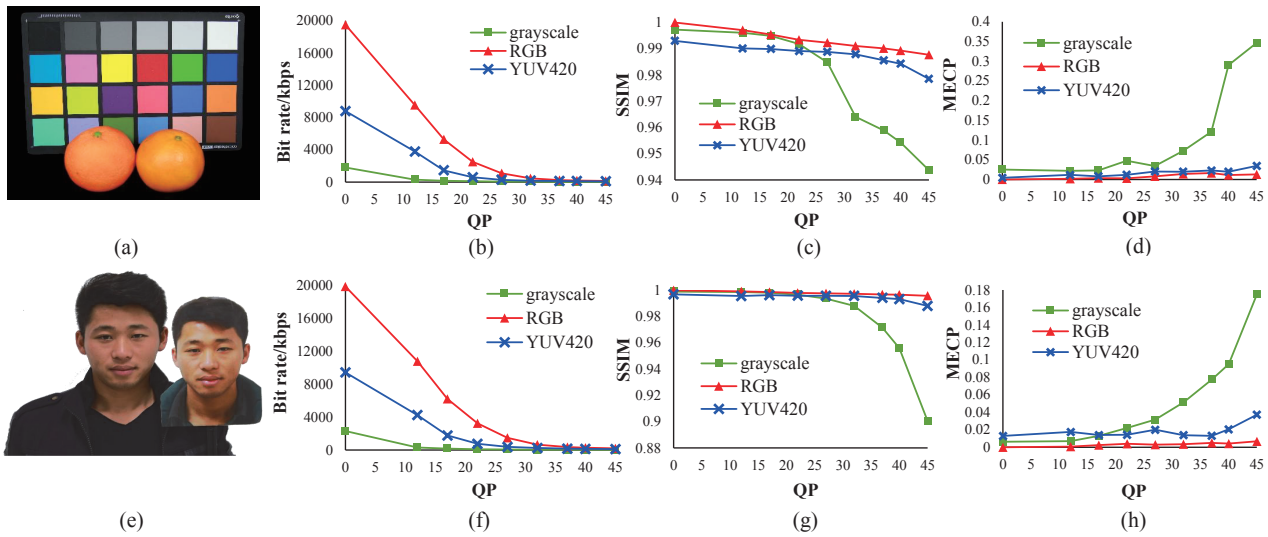

Figure 5: Illustration of compression with different QPs.

the Quick Sync is a dedicated hardware core on the processor die, which allows for a much more power efficient video processing. The QSV implementation supports several hardware-based codecs, including H.264 and VP8 [7][1]. Meanwhile, they provide quite a few video preprocessing (VPP) operations, such as color space conversion, resizing, and de-interlacing. The encoders provided by QSV only support the NV12 pixel format. The original pixel format of the raw grayscale video frame is RAW8 (8 bit depth), we use its VPP feature to convert the original pixel format to NV12. We assign the 8 bit grayscale value to the Y channel of NV12. The other two channels U and V are both set to zeros.

Compression on RGB video stream: For simplicity, we utilize the integrated open source library FFmpeg (with x264 [17]) [22] to apply the H.264 for RGB video stream compression. Additionally, we convert RGB to YUV420 [3] and then compress video stream using YUV420 source format. This is because YUV420 format is widely used for compression that are massively deployed in many video applications, such as streaming like YouTube, Netflix, etc, conferencing like FaceTime, SnapChat, etc.

We simultaneously use software codec for RGB video and hardware codec for grayscale video to maximize the performance of the resource constrained mobile transmission board, so that the MS video can be captured and processed in real-time.

2.2.2 Performance Assessment Metrics. We adapt the QP (quantization parameter) from 0 (lossless) to 45 (lossy with heavy compression) to encode respective RGB and grayscale videos. Often distortion for conventional RGB content is measured using Peak Signal-to-Noise Ratio (PSNR) [2] or Structural Similarity (SSIM) [26]. However, MS video based applications typically measure specific spectral bands using the spectral characteristic peaks. For example, human skin has a W-shaped protrusion at 559nm [4], which can be used to distinguish the real or fake human skin even with the same RGB value. To better understand the compression induced distortion of spectral characteristic peaks, we propose another performance metric: the Mean Error of Characteristic Peaks (MECP), which is denoted as $D(G, C)$ in this paper.

The MECP of a specific MS image pixel is calculated as follows:

$$\Delta d_i = s_i - s_{i-1}, \tag{1}$$

where, s_i denotes the intensity at spectral channel i, Δd_i refers to the difference between two adjacent channels.

We define the sign function Δf_i as follows:

$$\Delta f_i = \begin{cases} 1, & \Delta d_i > 0 \\ 0, & \Delta d_i = 0 \\ -1, & \Delta d_i < 0. \end{cases} \tag{2}$$

Then, the adjacent difference of Δf_i is calculated as:

$$L_i = \Delta f_i - \Delta f_{i-1}. \tag{3}$$

.

Therefore, we can get the i index where the spectral characteristic peaks locate (when $L_i = 2, or -2$). Then, the mean error of spectral characteristic peaks between the ground truth and the compressed spectrum is calculated as

$$D(G, C) = \frac{1}{m} \sum_{i=1}^{m} (G_i(L) - C_i(L)), \tag{4}$$

where m is the number of spectral characteristic peaks, G stands for the ground truth, and C is the compressed spectral signature.

Fig. 5 shows two examples of compressed performances. We can see that the two curves (in red and blue), which are quantified by compressing the RGB and YUV420 video streams, are basically smooth and close to each other when QP ranges from 0 to 45. Notice that the error caused by compressing YUV420 video stream is slightly larger than that of RGB video stream compression due to

[1]We choose H.264 because of its popularity.

the inevitable loss in the YUV to RGB format decoding procedure. In grayscale video compression (as shown in green curve), we find that the MECP value (D) begins increasing sharply when QP equals to 27, which means that some spectral characteristic peaks have been removed by the compression.

According to the simulations, we select QPs at 17 and 22 for grayscale and RGB videos to balance the transmission bandwidth and reconstruction quality. Various applications will be used to demonstrate the effectiveness of the compression method in Section 3.

2.3 MS Video Rendering

On the back-end, the received dataflow is decoded back to the raw RGB and grayscale video frames. From the grayscale frame, the low-spatial-resolution spectral image can be read out directly according to the previous calibrations as mentioned in [8]. For some real-time applications, such as the human-skin tracking experiment mentioned below, we can use the low-spatial-resolution spectral image for real-time object discrimination and use the pseudo color to mark out the target on the corresponding RGB frame. For some high-spatial-resolution applications, the multispectral video is consequently reconstructed using the bilateral filtering algorithm [9, 23], in which both the spatial proximity and color consistency are used to guide the propagation process. And, it is illustrated as

$$s_{ij} = \sum_{c\in r,g,b} \frac{\sum_{k\in\Omega} G_{\sigma_r}(d_k^{RGB}) G_{\sigma_s}(d_k^{x}y) \rho_k^c (w^c \bigotimes s_k)}{\sum_{k\in\Omega} G_{\sigma_r}(d_k^{RGB}) G_{\sigma_s}(d_k^{x}y)}, \quad (5)$$

where s_{ij} denotes the spectral signature vector of pixel (i, j), $k \in \Omega$ indexes the pixels within a neighborhoods centered on (i, j), $G_{\sigma_s}()$ represents the Gaussian operator with zero mean and variance σ, and d_k^{RGB} and d_k^{xy} denote the Euclidean distance between the pixels (i, j) and k in RGB space and (x, y) space. The factor ρ_k represents the ratio of a given color channel value at k to the corresponding value at (i, j).

We also implement many typical spectra analysis methods, including the the pseudo color rendering and real-time object tracking, as described in detail in Section 3.

3 EXPERIMENTS

In this section, the performance of the proposed MS video system is evaluated through the material discrimination and Human-face Tracking experiments. In the MS video acquisition subsystem, the objective lens is 25mm focal length, and spectral acquisition range is from 450nm to 900nm. In the MS video transmission and rendering subsystems, specifications of transmission board and the server are shown in Table 1.

3.1 Material Discrimination

3.1.1 Subjective evaluation. In this part, we conduct a yes/no subjective evaluation experiment to verify the performance of the proposed system. As shown in Fig. 8, the everyday-use items, including real and artificial roses, eggplants, and bell peppers, are utilized in our experiment. Tested images are captured under the halogen illumination source.

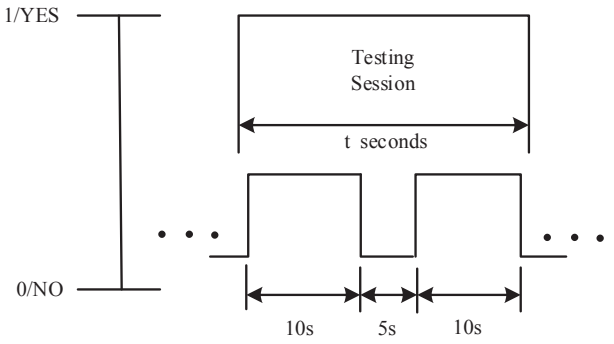

Figure 6: An illustration of subjective assessment protocol.

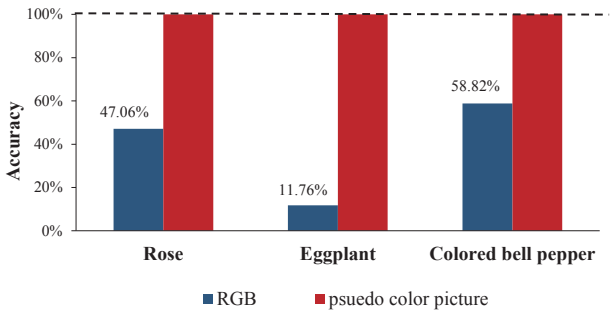

Figure 7: The ratio that volunteers can correctly discriminate the real or artificial items.

We perform the Single Stimulus Continuous Quality Evaluation (SDSCE) index [19] to implement the subjective assessment. 17 volunteers, including 9 males and 8 females, are randomly selected from different majors in campus. As illustrated in Fig. 6, we sequently display testing image samples on the 86 inch SeeWo screen at a distance of $4.5m$ [18] and let each volunteer gives his/her yes/no ratings in 10 seconds, and the interval time is 5 seconds.

Discrimination on RGB images As is shown in Fig. 7, about 50% volunteers can figure out the the real rose and bell pepper through their RGB images, which is consistent with the probability of random distribution, and it indicates that human eye completely cannot discriminate the real and fake ones. It is quite interesting that only 11.76% volunteers can figure out the real eggplant, which indicates that the artificial one looks too 'real'.

Discrimination on pseudo-color images In this section, objects are rendered into pseudo-color images using the MS data. Take the real and artificial rose for example. The spectral signatures of the real and artificial ones are shown in Fig. 9. For intuitive discrimination, the scaled intensity gradients between four distinct wavelength (615nm, 640nm, 665nm and 705nm) are calculated successively and then assigned to Red, green and blue value respectively, as is shown in Fig. 8 (d) (e) (f). We also implement the subjective assessment on the pseudo-color images, and the ratio that volunteers can correctly discriminate the real and artificial items arise to 100%, as is shown in Fig. 7.

Table 1: Specifications of Server and Transmission Board

Hardware	Server	Transmission Board
Processor	Intel Core i7-6700k 4.00GHZ	Intel Core i7-4500U 1.8GHZ
Operating systems	Windows 7 Pro 64 bit	Linux Ubuntu 14.04
RAM memory	32 GB DDR4 2400MHZ	8 GB DDR3 1600MHZ
GPU	Nvidia GeForce GTX 1070 8GB	Haswell-ULT Integrated Graphics Controller
Hard Disks	MSATA SSD 240G	SSD 240G,HD 2T

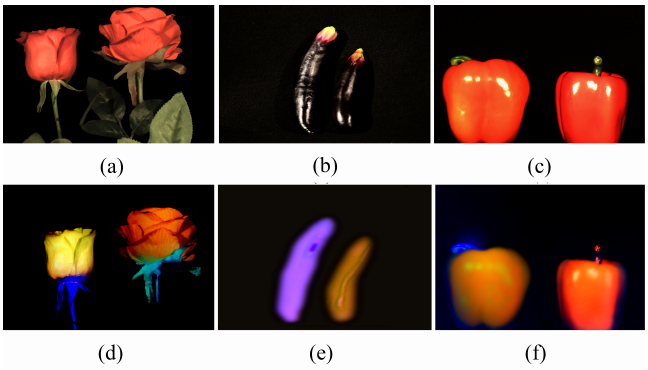

Figure 8: RGB and rendered pseudo-color images of the items. In pseudo-color images, the scaled intensity gradients between four distinct wavelength (615nm, 640nm, 665nm and 705nm) are calculated successively and then assigned to Red, green and blue value respectively.

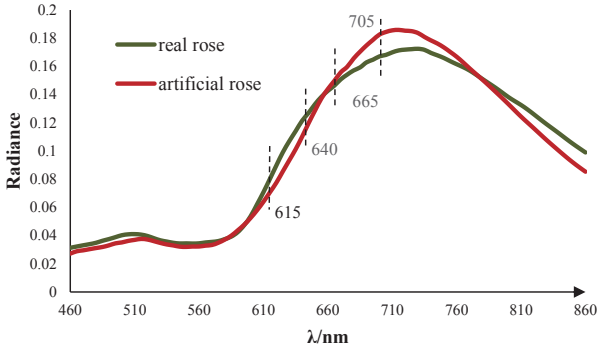

Figure 9: The spectral curve of real and artificial roses.

3.2 Human-face Tracking

Face recognition has been widely used in various AI applications, such as intelligent authentication, security check, entertainment and etc. However, the state-of-the-art human face recognition algorithms, which generally utilize an RGB camera as the input, can hardly discriminate the real or man-made face masks. In the following part, we'll demonstrate the usage of MS video camera to achieve accurate performance in human face discrimination and tracking.

Tracking algorithm: Spectral Angle Mapping (SAM) is a kind of spectral classification methods. The spectral signature is considered as a N-dimensional vector (N denotes the number of spectral channels). The similarity θ between the detected spectrum of a certain pixel $A_r\vec{(\lambda)}$ and reference spectrum $Re\vec{f}(\lambda)$ is calculated as

$$\theta = \frac{A_r\vec{(\lambda)} \cdot Re\vec{f}(\lambda)}{\|\vec{A_r}\|_2 \|Re\vec{f}(\lambda)\|_2}. \quad (6)$$

In our experiment, we manually extract the characteristic spectral channels of the detected object/scene beforehand; Then, Equ. 6 is used to calculate the similarity between the detected and reference spectrum; At last, the target with a higher θ value than the threshold is selected and marked out using the pseudo-color rectangle in the corresponding RGB frame.

Table 2: Quantitative Comparison of Human-face Tracking with 5 State-of-the-art Methods.

Method	DFT	MIL	CT	STC	CVT	Ours
CLE/pixels	580	274	250	374	284	70
DP	0.0130	0.8312	0.6234	0.5065	0.6620	0.9870

Table 3: Quantitative Comparison of Human-face Tracking with 6 Different QPs.

QP	0	12	17	22	27	32
CLE/pixels	70	95	130	192	318	338
DP	0.9870	0.9870	0.9610	0.7922	0.4675	0.5325
Bit Rate (kbps)	23716	2387	1309	693	539	462

3.2.1 Comparison with the state-of-the-art trackers (RGB). : We have conducted a series of spectral acquisition and tracking experiments on challenging sequences, and made comparisons with some of the state-of-the-art RGB methods. The sequences used in our experiments pose challenging situations such as motion blur, in-plane and out-of-plane rotations, scale variation, heavy occlusions, out of view and deformation. The five RGB trackers we compare with are: Distribution Fields for Tracking (DFT) [21] tracker, Multiple Instance Learning (MIL) [5] tracker, Compressive tracking (CT) [32], Spatio-Temporal Context (STC) [31] and Coloring Visual Tracking (CVT) [10]. The parameters of the proposed algorithm are fixed for all the experiments.

In this work, we use the precision rate for quantitative analysis. The results are presented using two evaluation metrics: Center Location Error (CLE) [29], Distance Position (DP) [29]. CLE is computed as the average Euclidean distance between the estimated center location of the target and the manually labeled ground-truth of each frame. DP is the relative number of frames in the sequence

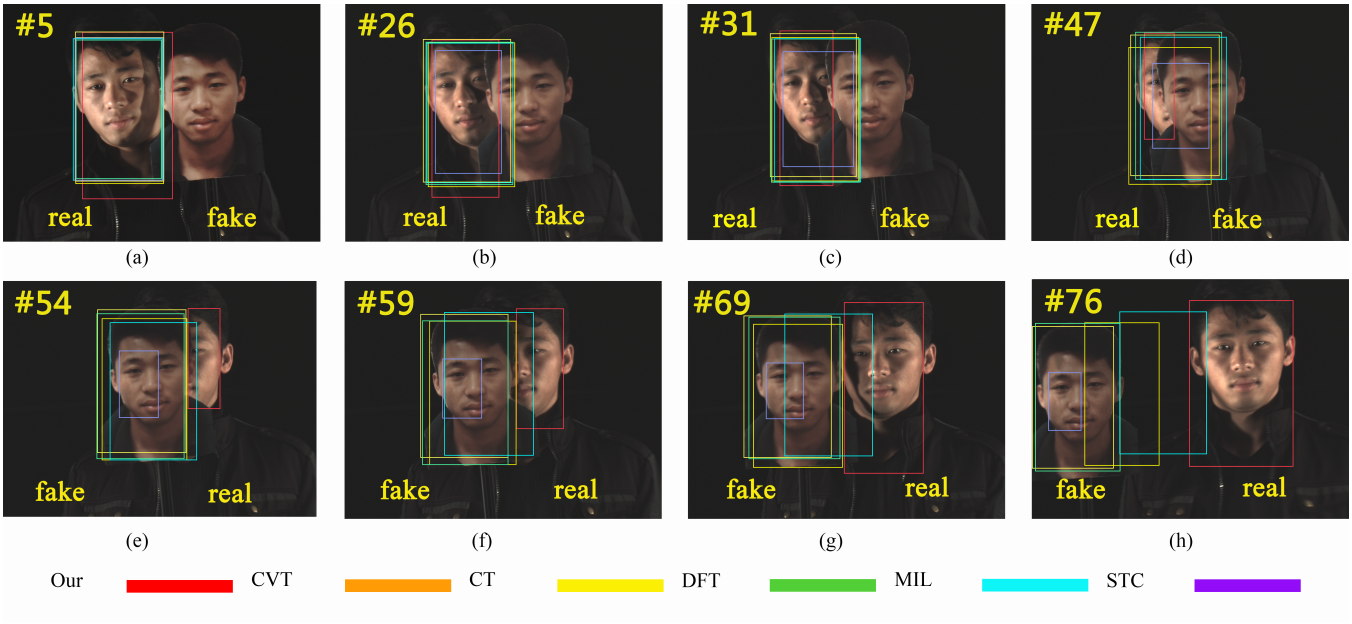

Figure 10: Screenshots of human-face tracking results.

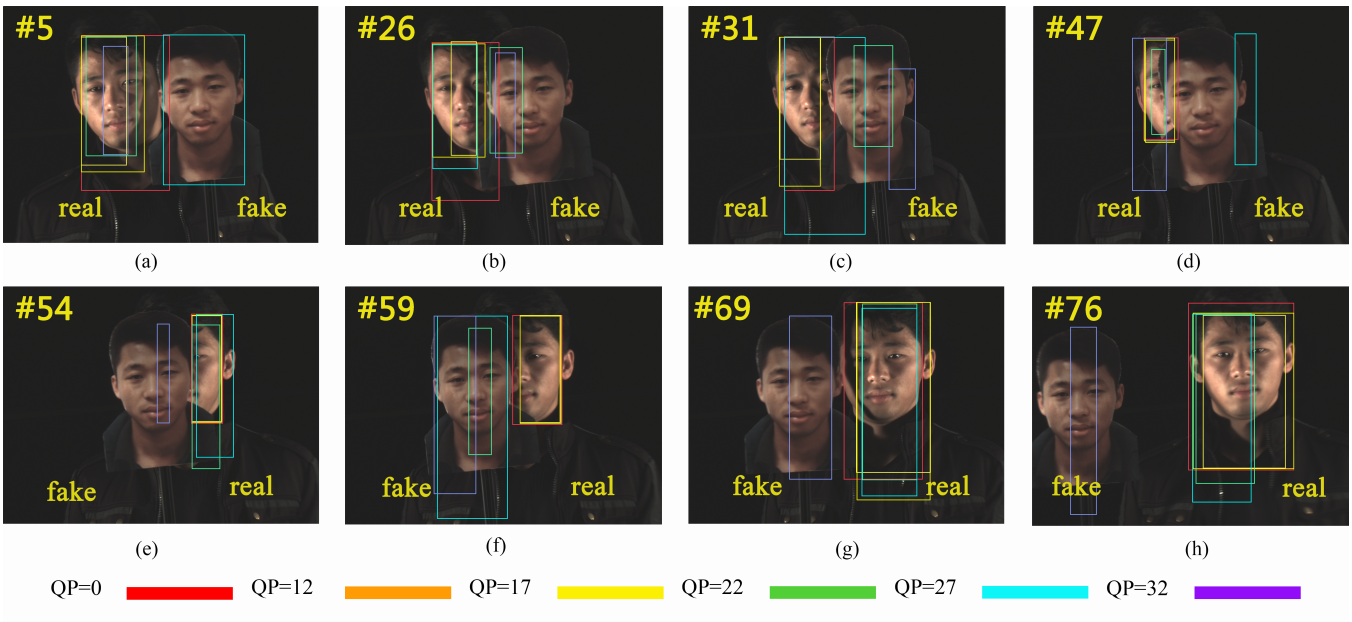

Figure 11: Screenshots of human-face tracking results with different QPs.

where the center location error is smaller than a certain threshold which is set to be 300 pixels there (Considering the spatial resolution of per picture is 2048×1536 pixels). For the trackers involving randomness, we repeat the experiments 10 times on each sequence and report the averaged results.

In the experiments, a printed artificial face is used to occlude the real face at times. Table 2 shows the quantitative results in which our system achieves the best performance both in terms of center location error and distance position. The screenshots intuitively show that our system can accurately identify and track the real face, perfectly avoiding the confusion of artificial face, but the traditional method of RGB tend to mismatch target (See #54, #59, #69 and #76 of the sequence in Fig. 10).

3.2.2 Tracking with different compression bit rate. : To test the robustness, we have conducted experiments with different QP values for the grayscale compression. Also, the QP value for the RGB

compression is fixed at 22. The performance assessment metrics and the bit rate are shown in Table 3.

When QP for grayscale compression is smaller than 17, our system can achieve a stable and accurate tracking performance (See Fig. 11), and the bit rate is reduced to 1.309Mbps. At the same time, the bit rate of RGB video stream is not lager than 4Mbps when QP value is fixed at 22, as shown in Fig. 5. Therefore, the total bandwidth of the grayscale and RGB video streams is about 5.309Mbps. It is noted that out proposed system can satisfy the UAV (Unmanned Aerial Vehicle) wireless transmission speed.

4 CONCLUSIONS

We have presented a compact and light-weight mobile MS video streaming device that is capable of capturing, transmitting and rendering the high spatial and high spectral video in real-time. This is achieved by re-designing the dual-camera optics and implementing the cost-efficient parallel compression of image data from both cameras on a mobile transmission board. Our system has demonstrated the comparable dynamic spectral acquisition precision with world-class spectrometer [1], but with significant cost reduction by using the off-the-shelf optical and electrical components and with much smaller form factor about 210 mm × 190 mm × 40 mm. Leveraging the efficient compression algorithm, we have managed to capture and stream the high fidelity multi-spectral video wirelessly. All the complex rendering and further application oriented analysis can be offloaded to the edges servers. This certainly has broadened the potential of such mobile MS streaming device to be deployed in many areas, such as criminal investigation (cf. human-face tracking example in Section 2), environment monitoring (cf. material discrimination example in Section 2), etc.

ACKNOWLEDGMENTS

Thanks volunteers to participate our field tests. This work is supported by the National Natural Science Foundation of China (61627804, 61371166, 61422107, 61571215, 61671236), Natural Science Foundation for Young Scholar of Jiangsu Province (BK20140610, BK20160634).

REFERENCES

[1] *Cubert UHD285.* http://cubert-gmbh.com/product/uhd-285-raccoon/.
[2] *PSNR.* https://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio.
[3] *YUV420.* https://en.wikipedia.org/wiki/YUV.
[4] Elli Angelopoulou. 2001. Understanding the color of human skin. In *Photonics West 2001-Electronic Imaging*. International Society for Optics and Photonics, 243–251.
[5] Boris Babenko, Ming-Hsuan Yang, and Serge Belongie. 2009. Visual tracking with online multiple instance learning. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 983–990.
[6] V Backman, Michael B Wallace, LT Perelman, JT Arendt, R Gurjar, MG Müller, Q Zhang, G Zonios, E Kline, T McGillican, and others. 2000. Detection of preinvasive cancer cells. *Nature* 406, 6791 (2000), 35–36.
[7] Jim Bankoski, Paul Wilkins, and Yaowu Xu. 2011. Technical overview of VP8, an open source video codec for the web. In *Multimedia and Expo (ICME), 2011 IEEE International Conference on*. IEEE, 1–6.
[8] Xun Cao, Hao Du, Xin Tong, Qionghai Dai, and Stephen Lin. 2011. A prism-mask system for multispectral video acquisition. *IEEE transactions on pattern analysis and machine intelligence* 33, 12 (2011), 2423–2435.
[9] Xun Cao, Xin Tong, Qionghai Dai, and Stephen Lin. 2011. High resolution multispectral video capture with a hybrid camera system. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 297–304.
[10] Martin Danelljan, Fahad Shahbaz Khan, Michael Felsberg, and Joost Van de Weijer. 2014. Adaptive color attributes for real-time visual tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1090–1097.
[11] W Debskia, P Walczykowskia, A Klewskia, and M Zyznowskib. 2004. Analysis of usage of multispectral video technique for distinguishing objects in real time. In *20th ISPRS Congress*. Istanbul: ISPRS.
[12] Stephanie Delalieux, Annemarie Auwerkerken, Willem W Verstraeten, Ben Somers, Roland Valcke, Stefaan Lhermitte, Johan Keulemans, and Pol Coppin. 2009. Hyperspectral reflectance and fluorescence imaging to detect scab induced stress in apple leaves. *Remote sensing* 1, 4 (2009), 858–874.
[13] Michael Descour and Eustace Dereniak. 1995. Computed-tomography imaging spectrometer: experimental calibration and reconstruction results. *Applied Optics* 34, 22 (1995), 4817–4826.
[14] Jiao Feng, Xiaojing Fang, Xun Cao, Chenguang Ma, Qionghai Dai, Hongbo Zhu, and Yongjin Wang. 2014. Advanced hyperspectral video imaging system using Amici prism. *Optics express* 22, 16 (2014), 19348–19356.
[15] Robert O Green, Michael L Eastwood, Charles M Sarture, Thomas G Chrien, Mikael Aronsson, Bruce J Chippendale, Jessica A Faust, Betina E Pavri, Christopher J Chovit, Manuel Solis, and others. 1998. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). *Remote Sensing of Environment* 65, 3 (1998), 227–248.
[16] Robert T Kester, Noah Bedard, Liang Gao, and Tomasz S Tkaczyk. 2011. Real-time snapshot hyperspectral imaging endoscope. *Journal of biomedical optics* 16, 5 (2011), 056005–056005.
[17] Loren Merritt and Rahul Vanam. 2006. x264: A high performance H. 264/AVC encoder. *online] http://neuron2. net/library/avc/overview_x264_v8_5. pdf* (2006).
[18] Yen-Fu Ou, Zhan Ma, Tao Liu, and Yao Wang. 2011. Perceptual quality assessment of video considering both frame rate and quantization artifacts. *IEEE Transactions on Circuits and Systems for Video Technology* 21, 3 (2011), 286–298.
[19] Yen-Fu Ou, Yuanyi Xue, and Yao Wang. 2014. Q-star: a perceptual video quality model considering impact of spatial, temporal, and amplitude resolutions. *IEEE Transactions on Image Processing* 23, 6 (2014), 2473–2486.
[20] Yoav Y. Schechner and Shree K. Nayar. 2002. Generalized mosaicing: Wide field of view multispectral imaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 10 (2002), 1334–1348.
[21] Laura Sevilla-Lara and Erik Learned-Miller. 2012. Distribution fields for tracking. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 1910–1917.
[22] Suramya Tomar. 2006. Converting video formats with FFmpeg. *Linux Journal* 2006, 146 (2006), 10.
[23] Carlo Tomasi and Roberto Manduchi. 1998. Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*. IEEE, 839–846.
[24] Freek van der Meer, Steven De Jong, and Wim Bakker. 2002. Imaging spectrometry: basic analytical techniques. In *Imaging spectrometry*. Springer, 17–61.
[25] Ashwin A Wagadarikar, Nikos P Pitsianis, Xiaobai Sun, and David J Brady. 2009. Video rate spectral imaging using a coded aperture snapshot spectral imager. *Optics express* 17, 8 (2009), 6368–6388.
[26] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13, 4 (2004), 600–612.
[27] Thomas Wiegand, Gary J Sullivan, Gisle Bjontegaard, and Ajay Luthra. 2003. Overview of the H. 264/AVC video coding standard. *IEEE Transactions on circuits and systems for video technology* 13, 7 (2003), 560–576.
[28] Gerald Wong. 2009. Snapshot hyperspectral imaging and practical applications. In *Journal of Physics: Conference Series*, Vol. 178. IOP Publishing, 012048.
[29] Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang. 2013. Online object tracking: A benchmark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2411–2418.
[30] Masahiro Yamaguchi, Hideaki Haneishi, Hiroyuki Fukuda, Junko Kishimoto, Hiroshi Kanazawa, Masaru Tsuchida, Ryo Iwama, and Nagaaki Ohyama. 2006. High-fidelity video and still-image communication based on spectral information: Natural vision system and its applications. In *Electronic Imaging 2006*. International Society for Optics and Photonics, 60620G–60620G.
[31] Kaihua Zhang, Lei Zhang, Qingshan Liu, David Zhang, and Ming-Hsuan Yang. 2014. Fast visual tracking via dense spatio-temporal context learning. In *European Conference on Computer Vision*. Springer, 127–141.
[32] Kaihua Zhang, Lei Zhang, and Ming-Hsuan Yang. 2012. Real-time compressive tracking. In *European Conference on Computer Vision*. Springer, 864–877.