

Noise Characteristics of a High Dynamic Range Camera with Four-Chip Optical System

Takayuki YAMASHITA^{1,2} and Yoshihiro FUJITA²

1. NHK Engineering Department, Tokyo 150-8001, Japan

2. The Department of Electrical and Electronic Engineering and Computer Science, Ehime University, Ehime 790-8577, Japan

Abstract: HDR (high dynamic range) imaging has recently been applied to video systems, including the next-generation UHD TV (ultrahigh definition television) format. This format requires a camera with a dynamic range of over 15 f-stops and a signal to noise SN-ratio that is the same as that of HDTV systems. However, current UHD TV cameras cannot satisfy these conditions. We proposed a four-chip capturing method that combines three-chip and single-chip systems. A prism divides incident light into two rays. Most of the incident light is directed to the three-chip capturing block; the remainder is directed to a single-chip capturing block, avoiding saturation. High quality HDR video can then be obtained by synthesizing the images obtained from the two systems. In this paper, we detail this image synthesis method, and analyze the differences between the three-chip and single-chip capturing systems. In particular, we focus on noise analysis of the single chip system. Most of the noises do not affect to image quality but those at around the switching point degrade the visual picture quality. To improve the degradation we propose a new switching method for smooth matching between different noise characteristics of the two systems. Finally, we discuss the results of a simulation used to verify its effectiveness.

Key words: HDR, UHD TV, camera, image processing, single-chip imaging.

1. Introduction

In addition to being applied to still images, HDR (high dynamic range) imaging has recently been applied to videos. A significant feature of the HDR system for still images is that the dynamic range of the display is expanded on both the lower and brighter ends of the spectrum. For display devices, the CRT (cathode-ray tube) and LCD (liquid crystal display) equipment used in HDTV broadcast program production has a peak luminance range of from about 100 cd/m² to several times that, a black luminance of 0.1 cd/m², and a dynamic range of several thousand to one.

Standards for next-generation television systems include the HDR video and ultrahigh definition (UHD TV) formats. In the case of HDR systems, a dynamic range of over 15 f-stops is required for the camera. In addition, a better SN-ratio than that of a

standard dynamic range system is required. However, current UHD TV cameras cannot satisfy the aforementioned conditions, because the full well capacity of UHD TV cameras is lower than that of HDTV cameras.

To address these issues, many methods have been proposed. For example, one method expands the dynamic range of the image sensor itself. This approach includes a method that utilizes MOSFET's sub-threshold characteristics [1-3]. Another method uses the non-destructive read-out function of an active pixel sensor [4].

We proposed an approach using a four-chip capturing method that combines the ordinary three-chip capturing system used in broadcasting applications with a single-chip capturing system [5, 6]. The incident light is divided into two light rays by a prism, with intensities in the ratio $m:1$. Then, the low light refracted by the prism is directed into a single-chip capturing block to prevent saturation in the case of high-exposure video shooting. By synthesizing

Corresponding author: Yoshihiro FUJITA, professor, research fields: media technology and applications.

the high-quality picture from the three-chip capturing system with the low-saturation picture from the single-chip capturing system, we obtain an HDR video of high quality.

In this paper, we discuss the proposed capturing system and picture synthesizing method in detail and derive the best parameter values to use for synthesizing. Then, we analyze the picture quality of the synthesized image by calculating the noise at the switching point of the system and show a new switching method. Finally, we discuss the results of a simulation used to verify its effectiveness.

2. Basic Concept of HDR Video Capturing System with Four Image Sensors

2.1 Optical System

In general, the signal-to-noise ratio is degraded when the light input level of the camera is set low enough to obtain the reference output and still not saturate under intense input light. Nevertheless, it would be possible to achieve both high signal-to-noise ratio and HDR if it was possible to switch to the video image from a camera system for which the exposure was reduced for highlights to prevent saturation. Here, we describe an HDR camera based on that principle. Fig. 1 shows optical block of the proposed system. The camera comprises two camera systems and outputs an image that is synthesized from the images of the two systems. The basic approach is the same as method for implementing still-image HDR photography. That method is to combine images for which the exposure level is high to obtain a good signal-to-noise ratio and images for which the exposure level is low to avoid saturation in highlights. The proposed method involves splitting the input light into two paths, one which has a higher exposure level and one which has a low exposure level. As the incident light travels through the prism block (Fig. 1), it is reflected by the first prism surface (the reflected amount is $1/(m+1)$, where m is the exposure ratio) before reaching the single-chip color image sensor. The remaining amount, $m/(m+1)$,

is further reflected by the second and third surfaces as it travels through the prism block, dividing it into red, blue, and green before reaching the respective image sensors. HDR is achieved by using the output of the three-chip camera as the main image and switching to the image from the single-chip camera for the highlight parts. The major feature of the system is that most of the light is input to the three-chip camera, so the signal-to-noise ratio and sensitivity of the output signal basically satisfies the requirements for broadcasting use [7, 8]. At the same time, the remaining light is input to the single-chip camera at a low exposure level, so the imaging element has a high saturation level. The dynamic range can be extended by switching the output of each camera block just before reaching the level of saturation for the three-chip camera.

2.2 Method of Image Synthesizing [9]

We consider a scene comprising mostly low-intensity light, with high-intensity light being assigned to the minor portion. The principle of image synthesis is to replace the saturated highlight portion of high-quality images obtained from the three-chip color imaging with high-intensity light images from the single-chip color imaging. To synthesize two images, the switching point is set in the area of high-intensity light near the reference white level.

Here, let us suppose that each pixel of a single-chip color imager has the same maximum charge capacity as each imager in a tri-color imaging system does. We further suppose that the spectrum characteristic of the

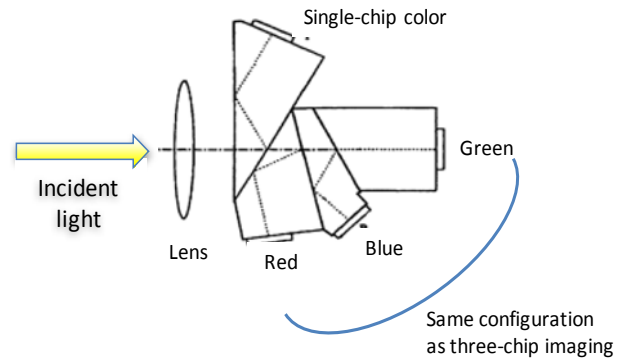


Fig. 1 Optical block in this system.

color filter array is the same as that of the three-chip prism.

Fig. 2 shows the input-output characteristics, where L_{in} is the level of the incident light, L_H is the output level on the high-exposure side, and L_L is the output level on the low-exposure side. If the input-output conversion gain is one, L_H and L_L are expressed as follows:

$$L_H = \frac{m}{m+1} L_{in} \quad (1)$$

$$L_L = \frac{1}{m+1} L_{in} \quad (2)$$

where L_S is the incident light level at which L_o saturates, and the amount of incident light on the high-exposure side at which L_H saturates are given by $(m+1)/m L_S$, given that the amount of incident light becomes $m/(m+1)$ after separation, as shown in Eq. (1). In the same manner, the amount of incident light at which L_L saturates is given by $(m+1) \cdot L_S$. To synthesize two images, L_H and L_L are multiplied by the amplification factors “ $m+1/m$ ” and “ $m+1$,” respectively, before switching from the high-exposure images to the low-exposure images at the saturation point of L_H . In this way, the incident light will remain linear until $(m+1) \cdot L_S$, yielding a dynamic range of $(m+1)$ times. The expansion rate of the dynamic range increases with the ratio of exposure, m .

There are two HDR standards in video industries, namely PQ (perceptual quantization) [10] and HLG (hybrid log-gamma) [11]. Both standards are included in the international program exchange format of recommendation ITU-R [12].

A feature of PQ is that the digital codes are related in terms of the absolute value of display luminance. On the other hand, in the HLG standard, the digital codes are related in terms of the relative levels of image sensor output. The HLG design succeeds that of the current SDR (standard dynamic range) system.

The OETF (opt-electronic transfer function) of HLG in Rec. ITU-R BT. 2100 [12] is represented as follows:

$$E' = \begin{cases} \sqrt{E}/2, & 0 \leq E < 1 \\ a \cdot \ln(E - b) + c, & 1 \leq E \end{cases} \quad (3)$$

($a = 0.17883277$, $b = 0.28466892$, $c = 0.55991073$), where E is the signal for each color component, proportional to the scene linear light and scaled by camera exposure, normalized to the range [0:12]. Then, E' is the resulting non-linear signal in the range [0:1]. From this equation, it is found that the maximum dynamic range is 1,200%.

Fig. 3 shows the chart of this OETF. Given that the proposed system is applied to HLG, the value of m is equal to 11.

When this system used in SDR video system, as the white level is customarily suppressed by knee processing for broadcasting cameras, SN ratio is better than this. It is necessary to take these into consideration in determining the separation ratio m .

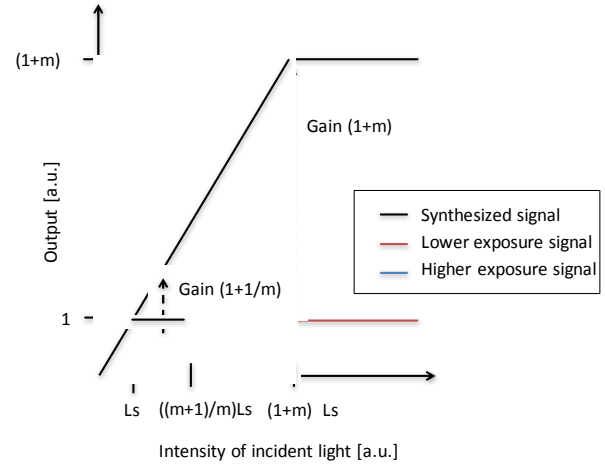


Fig. 2 Transfer function characteristics.

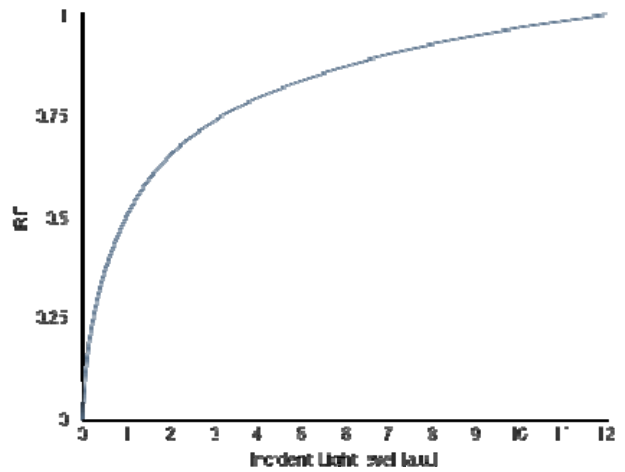


Fig. 3 HLG OETF.

Let consider that this system used in HDR video system. For simplicity, assuming that the image sensor has a 12-bit output and the noise is less than 1 LSB, the SN ratio is 72 dB. When $m = 11$ derived earlier, gain increase by 12 times is achieved, so S/N degradation of -22.3 dB is obtained, which is equivalent to 50 dB. Fig. 4 shows the measured values of the image noise detection source and the group of different dashed and solid lines showing the relationship between the brightness of the display device [13], but nearly the same result is derived. From this chart, it can be seen that if SN ratio is 50 dB or more, noises are not detected in an image of 100 rlx. BT. 2100 [12] specifies that a mastering monitor with a peak of 1,000 cd/m^2 should be used.

Therefore, when calculating the incident light by calculating (without total system gamma) using inverse OETF as follows, a 100 rlx of the luminance level on a monitor corresponds to 50% of a level of incident light, 100% of a level of the incident light as the switching has sufficient margin.

3. Noise Analysis of the Proposed System

3.1 Design Issue of HDR Camera System

Generally, the dynamic range of a camera can be increased by changing internal settings, but there is a trade-off with the noise characteristic. If the sensitivity and SN-ratio values for a given dynamic range setting are known in advance, the values for when the setting is changed can be calculated the sensitivity^{*)} and signal-to-noise ratio are denoted as F_1 and SN_1 when the dynamic range is set to $D_1\%$ and as F_2 and SN_2 when the dynamic range is set to $D_2\%$ and the sensitivity is 2,000 lx. F_2 and SN_2 can be calculated with Eqs. (4) and (5).

$$F_2 = F_1 \sqrt{\frac{D_2}{D_1}} \quad (4)$$

^{*)} The notation of the sensitivity of the camera is given by the aperture value which gives the video level of 100% when a white board with a reflectance of 89.9% is taken under 2,000 lx illumination.

$$SN_2 = SN_1 + 20 \log \left(\frac{D_1}{D_2} \right) \quad (5)$$

Measurements of the changes in actual 4 K and 8 K cameras for next-generation broadcasting have been reported [14]. The results of sensitivity and signal-to-noise ratio measurements for changes in the dynamic range setting for an 8 K camera (sensitivity at 2,000 lx F5.0; S/N: 51 dB; dynamic range: 600%) are presented in Fig. 5. When the dynamic range was changed to 1,200% (standard HDR), the sensitivity was F6.9, and the SN-ratio was 45 dB at 2,000 lx. We can see that the sensitivity increased, but the signal-to-noise ratio decreased for the higher dynamic range setting. Compared with the 60 dB SN-ratio for

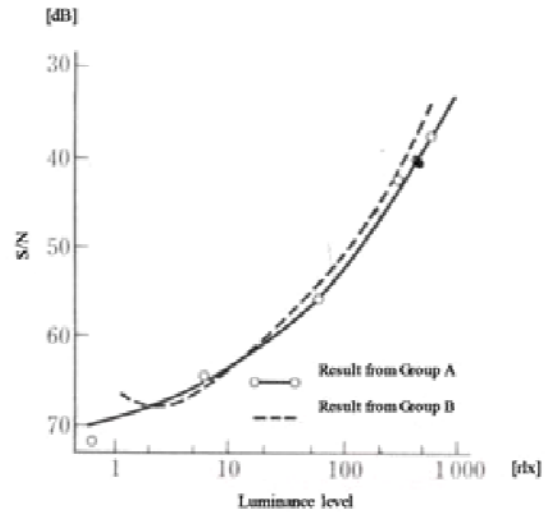


Fig. 4 Relationship between detection limit on S/N and luminance on monitor [13].

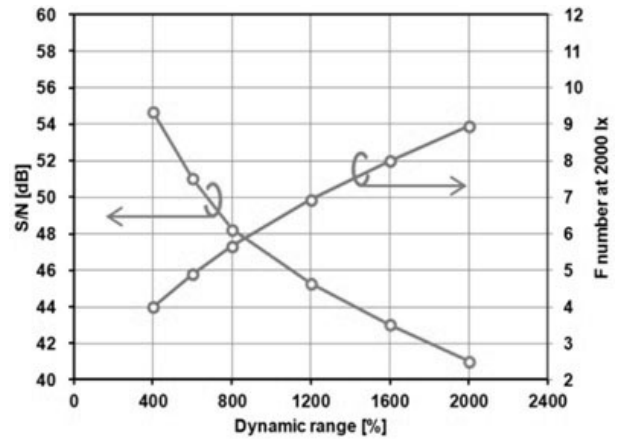


Fig. 5 Relationship between dynamic range, SN-ratio, and sensitivity in an 8 K camera [1, 4].

current HDTV cameras, the noise characteristic is not satisfactory. For this reason, noise is the most important problem to solve in implementing an HDR camera.

In particular, the primary degradation factor in the proposed camera is noise from the single-chip camera, so a noise analysis in the vicinity of the switching between the three-chip cameras in the single-chip camera is important.

3.2 Noise Analysis at Switching Point

As the value of m increases, the dynamic range of the camera becomes higher, but its SN-ratio deteriorates, because the low-exposure signal is amplified. To understand this issue further, we examine the calculation of noises at the switching point from dark noise (a portion not dependent on the incident light) and photon shot noise on the low-exposure signal.

In the following discussion, N_{dk} and N_{st} represent dark noise and photon shot noise, respectively, at one imager. When the incident light is divided into $m:1$, the dark noise N_{ldk} and photon shot noise N_{lst} on the low-exposure side are given by:

$$N_{ldk} = N_{dk} \quad (6)$$

$$N_{lst} = \frac{1}{\sqrt{m+1}} N_{st}. \quad (7)$$

When the incident light is $1/(m+1)$, the photon shot noise deteriorates in proportion to its square root, and we obtain Eq. (7). The incident light is amplified by $(m+1)$ when the images are combined, which means Eqs. (6) and (7) then become:

$$N'_{ldk} = (m+1)N_{ldk} = (m+1)N_{dk} \quad (8)$$

$$N'_{lst} = (m+1)N_{lst} = \sqrt{m+1}N_{st} \quad (9)$$

We can see that the dark noise increases in proportion to $m+1$, and that the photon shot noise increases in proportion to $\sqrt{m+1}$. By combining these two noises, we obtain the total amount of noise on

the low-exposure signal, as follows:

$$\begin{aligned} N_{Total_low} &= \sqrt{N'_{ldk}{}^2 + N'_{lst}{}^2} \\ &= \sqrt{\{(m+1)N_{dk}\}^2 + (m+1)N_{st}^2} \end{aligned} \quad (10)$$

We can calculate the total amount of noise on the high-exposure signal in the same way:

$$\begin{aligned} N_{Total_high} &= \sqrt{\left(\frac{m+1}{m}N_{dk}\right)^2 + \frac{m+1}{m}N_{st}^2} \end{aligned} \quad (11)$$

The noise degradation Nd at the switching point is derived from Eqs. (10) and (11), as follows:

$$\begin{aligned} Nd &= \frac{N_{Total_low}}{N_{Total_high}} \\ &= \frac{\sqrt{\{(m+1)N_{dk}\}^2 + (m+1)N_{st}^2}}{\sqrt{\left(\frac{m+1}{m}N_{dk}\right)^2 + \frac{m+1}{m}N_{st}^2}} \end{aligned} \quad (12)$$

Here, N_{st} varies depending on the intensity of the incident light. Fig. 6 shows how Nd changes with respect to Eq. (12) in response to changes in N_{st} . Here, we set $N_{dk} = 5$ and $m = 11$.

Note that N_{st} is proportional to the square root of the incident light quantity. If $N_{st} \gg N_{dk}$ is assumed when the intensity of the incident light at the switching point is sufficient, then we have

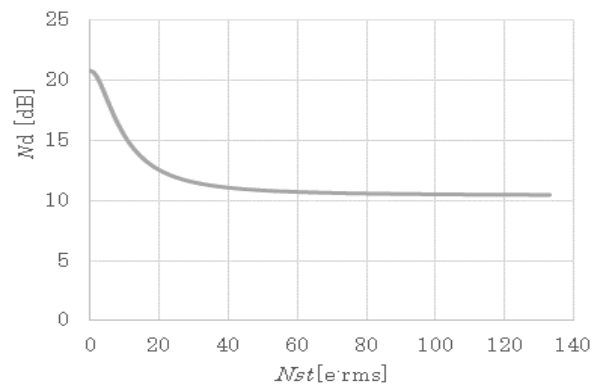


Fig. 6 Noise degradation at the switching point.
($m = 11$, $N_{dk} = 5$).

$$Nd|_{N_{st} \gg N_{dk}} \cong \frac{\sqrt{(m+1)N_{st}^2}}{\sqrt{\frac{m+1}{m}N_{st}^2}} = \sqrt{m} \quad (13)$$

From the above equation, Nd would be 10 dB when $m = 11$.

Generally, 10 dB dark noise remarkably degrades the picture quality. However at the switching point in this case shot noise has a high luminance level (about 0.5 IRE) on the display so the perceived noise becomes lower. This reduction effect of high light detection is about 3 dB (Fig. 5). Noise is less noticeable on the screen at higher luminance values, but the presence of level dependant noise can be a factor in degradation of

perceived image quality. Particularly in the case of moving images, noise differences may be perceived as false contours where there is a gradual change in luminance. We show a simulated image of this noise degradation in the following section.

3.3 Simulation Images

Fig. 7 is a model of the imaging system simulating the optical system. Noise is added, and because the lower exposure side corresponds to a single plate, the Bayer treatment is added.

On the other hand, a model for synthesis is shown in Fig. 8. A gain function is applied to the lower and the higher exposure side image, which is subjected to

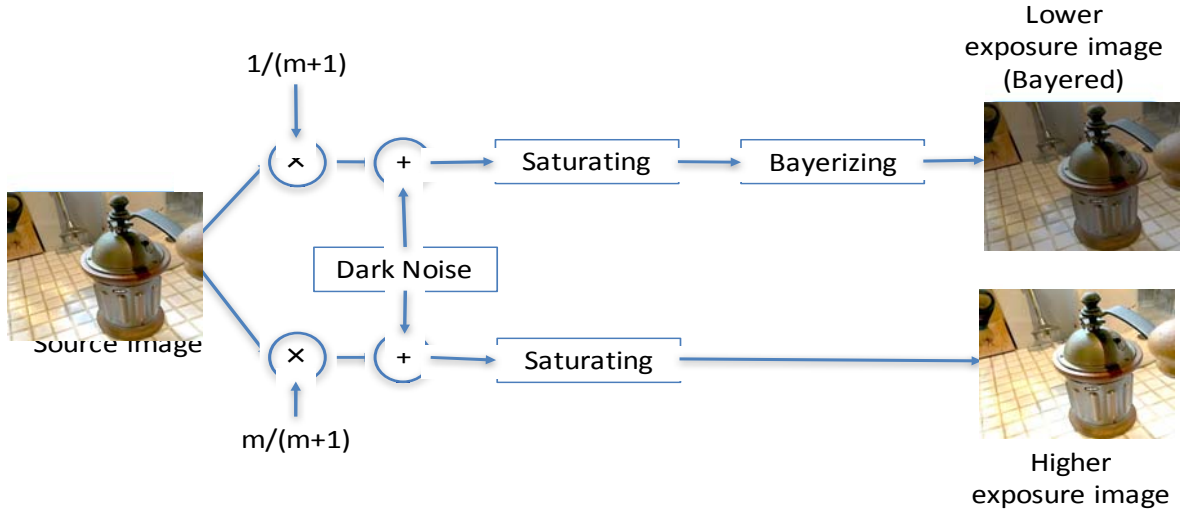


Fig. 7 Acquisition model for simulation.

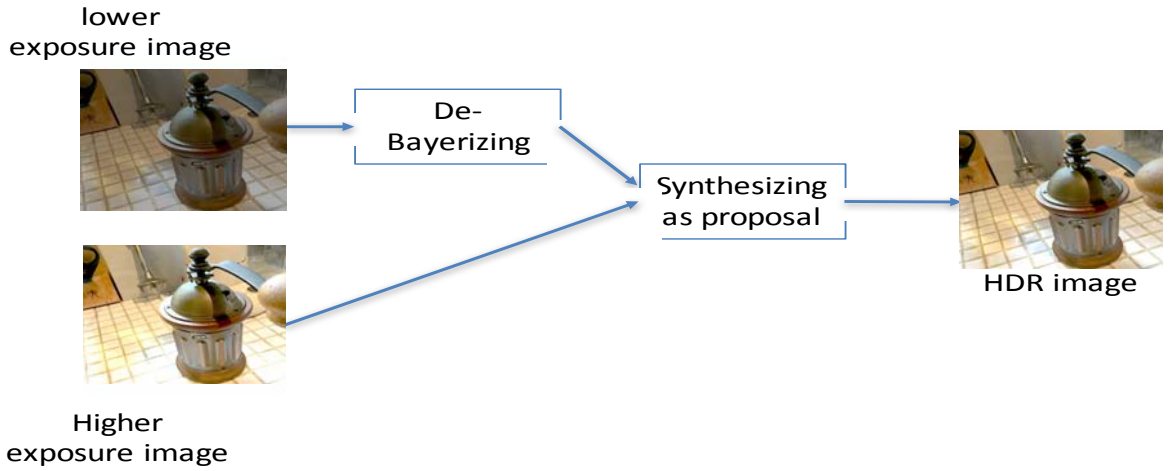


Fig. 8 Synthesizing model.

de-Bayer processing. For the switching point between the low- and the high-exposure sides, the synthesized image is output as the level at which the higher exposure side saturates.

For this simulation, the Radiance.hdr format image is used (see Fig. 10). Images on the low-exposure side and high-exposure side during synthesis are also shown in Figs. 11 and 12, respectively. Comparing Figs. 11 and 12, we recognize tones are reproduced in the brighter area in Fig. 9. Fig. 13 shows a part of images degraded at the switching point. Level dependent noise on the wall inside the room is noticeable.

To reduce the visual impact caused by this degradation, we applied a new method to synthesize the high-exposure and low-exposure signals with weighting and by creating an overlap period at switching.



Fig. 9 Synthesized HDR image.



Fig.10 Source HDR image ("Dani_belgium_oC65", Image courtesy of Dani Lischinski).



Fig. 11 Image on low-exposure side (Gained).



Fig. 12 Image on high-exposure side.



Fig. 13 Image example at the switching point (fixed level switching method applied).

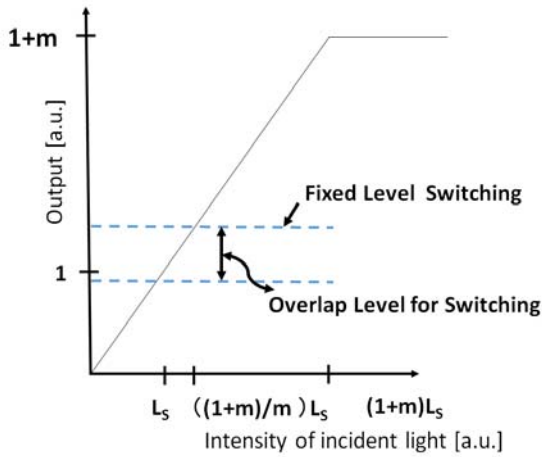


Fig. 14 Fixed and overlap level switching.



Fig. 15 Image example at the switching point (overlap level switching method applied).

by creating an overlap period. Both the fixed level and overlap level are shown in Fig. 14. Fig. 15 shows an image example when we applied this method. Noise was less visible compared with that in Fig. 13. We found this method to be effective.

4. Conclusion

We proposed an HDR acquisition system that combines three-chip color imaging, single-chip color imaging, and a video frame-synthesizing method. We analyzed the noise, increasing at the switching point.

An appropriate m value was derived from the HDR-TV standard. We also have showed some noise degradation of the single chip camera. A new switching

method to reduce visual degradation was described. We further confirmed that the proposed system and a new switching method were effective in obtaining high-quality pictures with a high dynamic range.

References

- [1] Chamberlain, G., and Lee, J. P. 1984. "A Novel Wide Dynamic Range Silicon Photodetector and Linear Imaging Array." *IEEE Journal of Solid-State Circuits* sc-19. 1: 41-8.
- [2] Kavadias, S., Dierickx, B., and Scheffer, D. 1999. "On-chip Offset Calibrated Logarithmic Response Image Sensor." *1999 IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, 68-71.
- [3] Loose, M., Meier, K., and Schemmel, J. 1999. "R27 Self-calibrating Logarithmic CMOS Image Sensor with Single Chip Camera Functionality." *1999 IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, 191-4.
- [4] Shimamoto, H., Mitani, K., and Fujita, Y. 2000. "Dynamic Range Expansion Method for a CMD Imager." *The Journal of the Institute of Image Information and Television Engineers* 12: 1781-7.
- [5] Yamashita, T., and Fujita, Y. 2017. "Video Frame Synchronizing Method for HDR Video Capturing System with four Image Sensors." IPAS-086, *Proceedings of Electronic Imaging*.
- [6] Yamashita, T., et al. 2002. "Wide-Dynamic-Range Camera Using a Novel Optical Beam Splitting System." *Proc. SPIE 4669, Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications III*, 82.
- [7] Sugawara, M., Mitani, K., Saitoh, T., Fujita, Y., and Suetsugu, K. 1994. "Four-Chip CCD Camera for HDTV." *SPIE Proceedings* 2173: 122-9.
- [8] Fujita, Y., Sugawara, M., Mitani, K., and Saitoh, Y. 1995. "A Compact, High-Performance HDTV Camera with Four-CCD Chips." *IEEE Transaction on Broadcasting* 41 (2): 76-82.
- [9] Yamashita, T., and Fujita, Y. 2017. "HDR Video Capturing System with Four Image Sensors." *ITE Trans on MTA* 5 (4): 141-6.
- [10] SMPTE ST 2084. 2016. "High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays." *Society of Motion Picture & Television Engineers*.
- [11] ARIB STD-B67. 2015. "Essential Parameter Values for the Extended Image Dynamic Range Television (EIDRTV) System for Programme Production." *Association of Radio Industries and Businesses*.

- [12] Recommendation ITU-R BT. 2100. "Image Parameter Values for High Dynamic Range Television for Use in Production and International Programme Exchange." International Telecommunication Union (2016).
- [13] Nishida, Y. et al. 1985. "Noise in CCD Image Sensor and Consideration for Future System." *ITEJ Technical Report* 9 (30): 1-5.
- [14] Funatsu, R., Kitamura, K., Yasue, T., Koide, D., and Shimamoto, H. 2016. "8 K HDR Camera with Hybrid Log-Gamma." *ITE Technical Report* 40 (23): 25-8.



Takayuki Yamashita is currently a doctoral student at Ehime University and a senior manager in the Engineering Department at NHK (Japan Broadcasting Corp.). He is working on the research, development, and standardization of ultrahigh-definition television systems. His research fields include the development of camera systems and high-bandwidth digital signal processing. He joined NHK in 1995 and has been engaged in research of HDTV camera systems since 1999. He is a member of the Society of Motion Picture and Television Engineers (SMPTE) and the Institute of Image Information and Television Engineers of Japan (ITE).



Yoshihiro Fujita is currently a professor at the Dept. of Electrical and Electronics and Computer Science at Ehime University. He earned this position in 2011. From 1976 to 2011, he had been with Japan Broadcasting Corporation (NHK), where he conducted research on advanced imaging systems for HDTV and ultra-high definition TV (UHDTV). He received his B.E. and Ph.D. degrees in 1976 and 1998 from the University of Tokyo, and he is a fellow of the IEEE and the Institute of Image Information and Television Engineers of Japan (ITE).
