Focus-Aid Signal for Super Hi-Vision Cameras

1. Introduction

Super Hi-Vision (SHV) is a next-generation broadcasting system with sixteen times (7,680x4,320) the number of pixels of Hi-Vision. Cameras for capturing SHV video have been developed, including a four-chip camera using the pixel offset method and four eight-megapixel complementary metal-oxide semiconductor (CMOS) image sensors\(^1\) and a full-resolution three-chip camera with three 33 megapixel CMOS image sensors\(^2\).

It is very important to focus SHV and other high-resolution cameras accurately, but as the image resolution increases, the size of their optical systems also increase, the size of the pixels in the image sensors decrease, and the depth of field becomes narrower. These features make it significantly more difficult to focus on an SHV camera than to focus a conventional camera. There are also no camera viewfinders (VF) with a 4,320 television line (TVL) resolution, so it is especially difficult for camera operators to focus accurately while looking through the VF. Because of this, a video engineer (VE) has to adjust the focus remotely while viewing the video on a high-resolution monitor located, for example, in a production truck. This not only increases the workload of the VE, it also prevents the camera operator from adjusting the focus in relation to his/her camera work. Future compact, hand-held SHV cameras will need to allow the camera operator to adjust the focus by him or herself, as is done in current broadcasting production.

One way to increase focusing accuracy would be to use an autofocus (AF). Some Hi-Vision broadcast cameras have AFs that can adjust quickly while the operator shoots the scene\(^3\). However, these AFs require special optical systems that would be difficult to apply to current SHV camera systems. On the other hand, although the contrast detection AF schemes\(^2\) used by commercial digital cameras do not require special optical systems, they have slow response times, which make it difficult to capture images that would be normally found in the productions. In their current form, these methods would be too difficult to apply to SHV cameras.

Another method is to add a focus-aid signal to the VF video. This method was devised when the transition was being made from standard television to Hi-Vision. Currently, most Hi-Vision cameras are equipped with a function that adds an edge enhancement signal called focus peaking to the VF\(^4\), enabling areas that are in focus to be seen clearly when they come into focus. Another method\(^5\) that makes edges flicker when in focus has also been proposed.

However, in SHV cameras, the resolution of the VF is much lower than that of the SHV image, so the conventional focus-aid signals cannot be used as they are.

In this article, we propose a new focus-aid signal for SHV cameras. In Section 2, we discuss the difficulty of focusing SHV cameras using a low-resolution VF by examining the modular transfer function (MTF)\(^2\) characteristics of the video. In Section 3, we describe how the proposed focus-aid signal is generated, and in Section 4, we show the effectiveness of the proposed method through simulations. In Section 5, we describe experiments done using a prototype that confirm the effectiveness of the proposed method.

2. Resolution Characteristics of VF Images

2.1 VF Video for SHV Cameras

The VF resolutions of major commercial Hi-Vision broadcasting cameras are shown in Table 1 (according to materials published by the manufacturers). Note that for cathode ray tubes (CRTs), this is the center resolution. The resolution of the VFs used in Hi-Vision cameras ranges from 400 to 650 TVL, which is approximately half that of Hi-Vision video (1,080 TVL) and one eighth that of SHV video (4,320 TVL).

This shows that the VF resolution is low even with Hi-Vision cameras, meaning that it is difficult for an operator to focus while viewing only the VF video. However, since the difference in resolution is small, it is possible to focus using a peaking signal generated from the video signal as a focus aid.

On the other hand, as shown in Figure 1, a down-converted signal is input to the VF for SHV cameras. This means the high-frequency components of SHV video would not be present in the VF video, making it even more difficult to focus on the VF. One way to solve this problem is to use a focus-aid signal that can be seen clearly on the VF.

\(^{2}\) A function that expresses the spatial frequency transfer characteristics of the elements of the image capturing system (lenses, image sensors, etc.). The MTF corresponds to the decrease in contrast (amount of blurring) in images passing through the elements under consideration, equaling 1 if there is no loss of contrast, and zero if shades of black and white have completely disappeared.

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Type</th>
<th>Resolution (TVL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2-in, B/W</td>
<td>600</td>
</tr>
<tr>
<td></td>
<td>6.3-in, Color</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td>9-in, Color</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>11-in, Color</td>
<td>540</td>
</tr>
<tr>
<td>B</td>
<td>2-in, B/W</td>
<td>600</td>
</tr>
<tr>
<td></td>
<td>5-in, B/W</td>
<td>650</td>
</tr>
<tr>
<td></td>
<td>9-in, Color</td>
<td>480</td>
</tr>
<tr>
<td>C</td>
<td>2-in, B/W</td>
<td>600</td>
</tr>
<tr>
<td></td>
<td>7.9-in, Color</td>
<td>450</td>
</tr>
</tbody>
</table>

\(^{1}\) Schemes that adjust the focus to maximize contrast in a particular area of the image, such as the center.
harder to focus. We will show theoretically that it is
difficult to focus SHV video by comparing the resolution
characteristics of VF video and SHV video.

2.2 Comparing the Resolution Characteristics of SHV Video and VF Video

The MTF is useful for comparing the resolution
characteristics of SHV video and VF video. The MTF of
an overall imaging system incorporates the MTFs of all
elements comprising the system, such as the lenses and
image sensors. It depends on spatial frequency, and the
resolution increases as the spatial frequencies with large
MTF values increase.

As shown in Equation (1), the MTF of SHV video can be
expressed as the product of the MTF of the lens (MTF<sub<lens></sub>)<sup>6)</sup>
and the MTF of the image sensor (MTF<sub<sens></sub>)<sup>7)</sup> (see Appendix
1).

\[ MTF_{\text{SHV}}(v) = MTF_{\text{lens}}(v) \cdot MTF_{\text{sens}}(v) \] (1)

Here, \( v \) is the spatial frequency.

Figure 2 shows the positional relationship of the
object, lens, and image sensor. When the image sensor
is located where the image of the object is formed, it will
be in focus, and when the device is positioned a distance
\( d_z \) from the focus position in the depth direction, it will be
blurred (defocused). We computed the MTF of SHV video
in terms of \( d_z \) by using Equation (1).

The MTF of SHV video is shown in Figure 3. In this
case, the F-number of the lens was 4, wavelength was
550 nm, image sensor pixel aperture was 3.5 \( \mu m \), pixel
pitch was 3.8 \( \mu m \), the SHV video limiting resolution was
4,320 TVL, and that of VF video was 540 TVL. Note that
in this figure, spatial frequency, \( v \), is converted into TVL.

As \( d_z \) increases in Figure 3, MTF tends to decrease more
quickly. For example, for \( d_z = 150 \mu m \), MTF is zero for
resolutions of 1,100 TVL and higher; thus, SHV video
components at 1,100 TVL or higher are completely
defocused.

The VF limiting resolution is 540 TVL, and even at
\( d_z = 150 \mu m \), when SHV video components over 1,100 TVL are

---

**Figure 1:** SHV camera VF video signal system

**Figure 2:** Positional relationship of object, lens, and imaging sensor

**Figure 3:** SHV video MTF
completely defocused, the MTF is approximately 0.6. The MTF at the focus \((d_f = 0 \mu m)\) for 540 TVL is approximately 0.95. This means a low-resolution VF \(^3\) can be used to make the contrast adjustment, but it will not be easy to use one to maximize the contrast and focus accurately. Because of this, we considered that the method of overlaying focus-aid signal on the VF video would be effective to show the drop in MTF for high-frequency components of the SHV video.

3. Focus-Aid Signal

A low-pass filter (LPF) is normally used to eliminate SHV video aliasing, and the simplest way to make a focus-aid signal is to sample with or without an LPF or to use an LPF that leaves a small amount of aliasing and use that aliasing component as the focus aid. This aliasing component can be viewed while adjusting the focus, but it is difficult to focus accurately because the VF resolution is only one eighth that of SHV video and aliasing also appears in areas that are not in focus.

Thus, we devised another focus-aid signal that enables accurate focusing. Figure 4 shows how this signal is generated and overlaid. In the signal generating section, a finite impulse response (FIR) filter is used to extract the high-frequency components from the SHV video. Depending on the type of FIR filter, the extracted high-frequency components could contain negative values, so the absolute value is taken in order to make the components all have positive values. The resulting positive signal also contains random noise and unneeded contours outside of the focal range. The level of these components is low compared with that of the high-frequency components within the focal range, so they can be eliminated by setting a threshold and setting the values below the threshold to zero. Note that a variety of filming environments and objects can be handled by varying this threshold value. Finally, the resolution of the focus-aid signal is converted into that of the VF video signal. Generally, the focus-aid signal extracted by the FIR filter contains the peak values while focusing, so it is desirable to maintain these peak values while converting the resolution. Accordingly, we sub-sample after applying a maximum-value\(^4\) process to the image over \(M \times M\) pixel regions.

The value of \(M\) is the ratio between the SHV limiting resolution, \(R_{SHV}\), and the VF video limiting resolution, \(R_{VF}\), or:

\[
M = \frac{R_{SHV}}{R_{VF}} \quad \quad (2)
\]

This procedure generates a focus-aid signal without aliasing and also preserves peak values.

In the down-converter section, the SHV video is first passed through an LPF to eliminate aliasing. It is then sub-sampled at a ratio of \(M:1\). This converts \(M \times M\) pixels of SHV video to one pixel of VF video.

The focus-aid signal and down-converted video signal are then combined into the VF video signal. The VFs of most handheld cameras are monochrome, so normally the luminance signal is used to generate the focus-aid signal, but if the VF is color, the focus-aid signal can also be a color one.

4. Simulated Generation of the Focus-Aid Signal

4.1 Video used for Simulation

Figure 5 shows the filming conditions used for the simulations. The SHV video had 7,680 x 4,320 pixels and a frame frequency of 59.94 Hz. Three test charts, A, B, and C, were placed at fixed distances from the SHV camera.

\(^3\) Strictly speaking, the MTF of the VF video should be computed, but it will not differ much from the 540 TVL MTF for SHV video.

\(^4\) A process of selecting the maximum value.
camera, and video was captured while moving the focus continuously at a mostly steady, low speed from close to far away. An example of the captured images is shown in Figure 6; here, the camera is focused on test chart A, and test charts B and C are blurred. The VF video is 960x540 pixels, and the focus-aid signal was generated from the G (green) component of the SHV video. For convenience, numbers from zero to 200 were assigned to the captured video frames, with the distance of the focus point from the camera increasing as the frame number increase.

4.2 Simulation of Generating the Focus-aid Signal

We used a two-tap, first-order differential filter, which can be created in hardware using a small circuit, as the high-pass FIR filter. The amplitude-frequency characteristic of this filter is shown in Figure 7. As can be seen in the figure, components under 850 TVL are attenuated by 10 dB or more. Eight-bit quantization with signal levels from zero to 255 were used in the simulations. To eliminate undesired components such as noise, three threshold values, $T_{TH}$, of 16, 32, and 64 were established, an 8x8 maximum-value filter was used, and a 960x540 pixel image, the same resolution as the VF, was generated. Images of the focus-aid signals are shown in Figure 8(a) to (c). It is clear that as the threshold value increases, the noisy and defocused areas are eliminated more thoroughly. The best threshold value depends on the filming conditions, such as brightness of the object, so it is desirable to be able to set the best threshold value.

4.3 Quantitative Evaluation of the Focus-aid Signal

When the desired object is brought into focus, the contrast of that object increases. In so doing, the standard deviation of the pixels surrounding it increases, and this means the standard deviation can be used as an index for focusing. We used the SHV video signal as a reference and compared the standard deviations of the VF video signals with and without the focus-aid signal overlaid. For these simulations, we used the threshold value, $T_{TH}$, of 64. We assumed a scenario with the focus on Chart C and used the region in the box in Figure 9 for the evaluation.

If we use an evaluation region of $N$ pixels horizontally and $M$ pixels vertically, indicating the pixel coordinates within the region as $(k, l)$, and the pixel value as $x_{k,l}$ (with $k = 1, 2, ..., N$; $l = 1, 2, ..., M$), and then take the average pixel value, $\bar{x}$, within the region, the standard deviation, $\sigma$, is given by Equation (3).

$$\sigma = \sqrt{\frac{1}{NM} \sum_{k=1}^{N} \sum_{l=1}^{M} (x_{k,l} - \bar{x})^2}$$

.....(3)
Figure 10 shows the results of calculating the standard deviations of the SHV video signal (for reference), the VF video signal with the focus-aid signal superimposed, and the VF video signal without the focus-aid signal. The horizontal axis is the frame number, and the vertical axis is the standard deviation. The standard deviation of the SHV video signal reaches a maximum at frame number 114, indicating that the focus is best at this frame. We will call this the focused frame.

Regardless of whether the focus-aid signal is present or not, the standard deviation of the VF video reaches a peak near the focused frame, but the peak is very gentle for the case without the focus-aid signal, and the change is small. On the other hand, for the case with the focus-aid signal, there is a large peak near the focused frame, suggesting that it will be easier to adjust the focus.

5. Focusing Experiments

We prototyped equipment to generate a focus-aid signal and conducted focusing experiments with it. Figure 11 shows the block diagram of the prototype.

The equipment allows selection of either color or monochrome for the low-resolution video directly down-converted from the SHV video. It also allows selection of one of seven colors (white, red, green, blue, cyan, magenta, or yellow) for the focus-aid signal generated from the green (G) signal. When a color VF can be used, a green focus-aid signal can be overlaid on the monochrome video signal as shown in Figure 12, so the operator can select the combination that works the best for him/her.

The focusing experiments covered the following three cases:
1) Remotely controlling the camera while viewing a high-resolution monitor (conventional focusing method)
2) Operating the camera while viewing the VF video without the focus-aid signal
3) Operating the camera while viewing the VF video with the focus-aid signal

For each case, three different operators performed 30 repetitions of a focusing operation, and we studied the distribution of the resulting focus values. For case (1), we used a 22-inch, 3,840x2,160-pixel high-resolution monitor, cropping images so as to display the center section at full size. For cases (2) and (3), we used an 11-inch, 960x540 pixel color VF, displaying video down-converted from SHV to Hi-Vision size in monochrome. All three operators used a green focusing-support signal, as shown in Figure 12.

The positional relationship between the camera and object in the focusing experiments is shown in Figure 13. The distance from the lens to the object was 2.15 m, the focal length of the lens was 31 mm, the F-number was 4, and the image sensor pixel size was 3.8 μm. The depth of field...
The experimental results are shown in Table 2(a) to (c). Note that focus accuracy in the table indicates the probability that the focusing value is within the depth-of-field of the object (see Appendix 2). There tended to be a variation in focusing by the operators, but accuracy was the highest for all operators in case (3), when using the focus-aid signal. Case (1) was second, when using the high resolution monitor. Focusing accuracy was very poor in case (2), the case without a focus-aid signal; that means focusing was generally very difficult in this case.

The results in Table 2 show that the focus-aid signal helps to make the focusing accuracy as good as or better than the conventional method using a high-resolution monitor, even when using a low-resolution VF.

6. Conclusion

We proposed a focus-aid signal for SHV cameras that allows direct focusing using a low resolution VF. We conducted simulations on this method and confirmed that it is effective. We also prototyped equipment to generate a focus-aid signal and confirmed that with it, a camera operator could maintain a focusing accuracy as good as or better than those of conventional methods using a high-resolution monitor. The proposed equipment has already been built into an SHV camera system and is being used as a way to focus using the VF.

The proposed method can, in principle, be used for all kinds of manual-focus cameras, not just SHV cameras. As the resolution of cameras increases, focus-aid signals can be expected to play an increasingly important role during production in order to alleviate problems caused by the difference in resolution and screen size between the shot image and the viewfinder.

In the future, we intend to accumulate operating time data from SHV cameras, clarify any problem points, and continue to make improvements. We will continue to develop a single-chip SHV camera with the goal of reducing its size and compact signal processing equipment to be built into a handheld SHV camera. We are also studying the possibility of applying the method described here to an AF system.

This article was written and edited with reference to the following paper appearing in the ITE Journal.


(Ryohei Funatsu)

References


The MTF of SHV video \( MTF_{\text{SHV}} \) is given by the product of the lens MTF \( MTF_{\text{lens}} \) and the imaging sensor MTF \( MTF_{\text{sens}} \), as shown in Equation (A1).

\[
MTF_{\text{SHV}}(v) = MTF_{\text{lens}}(v) \cdot MTF_{\text{sens}}(v) \quad \text{......... (A1)}
\]

Here, \( v \) is the spatial frequency.

The lens MTF can be expressed as shown in Equation (A2)\(^6\).

\[
MTF_{\text{lens}}(v) = \frac{4}{na} \cos \left( \frac{av}{2} \right) + \sum_{n=1}^{\infty} \frac{\sin(2Na)}{2N} \left( J_{2N+1}(a) - J_{2N+1}(a) \right)
\]

\[
- \frac{4}{na} \sin \left( \frac{av}{2} \right) \sum_{n=0}^{\infty} \frac{\sin(2N+1)a}{2N+1} \left( J_{2N}(a) - J_{2N+1}(a) \right) \quad \text{......... (A2)}
\]

Here, \( J \) is a Bessel function of the first kind. Defining \( \lambda \) as the wavelength of light, \( F \) as the F-number, and \( d \) as the deviation from the focus position, \( s, a, \) and \( \beta \) can be expressed as in Equation (A3). We assume that the lens is sufficiently far from the object that the light rays entering the lens can be approximated as parallel.

\[
s = 2\lambda F_w \nu
\]

\[
a = \frac{\pi v}{F_w} \frac{d}{\lambda}
\]

\[
\beta = \cos^{-1}(\lambda F_w \nu)
\]

The MTF of the image sensor can be expressed as in Equation (A4), in terms of the pixel aperture, \( \omega \), and the pixel sampling interval, \( p \)\(^7\).

\[
MTF_{\text{sens}}(v) = \frac{\sin(\pi \omega)}{\pi \omega} \cdot \frac{\sin(\pi p)}{\pi p} \quad \text{......... (A4)}
\]

Substituting Equations (A2) and (A4) into Equation (A1) gives the MTF characteristic shown in Figure 3.