Video Services | Visual Characteristics | Video Encoding Techniques

Technology Reports

Special Articles on Technology toward Further Diversification of Life-Style Mobile

A Model of Visual Characteristics for the Implementation of High-quality Video Services

The opportunity to watch video content on mobile terminals is gradually becoming a daily life. With the aim of substantially improving the efficiency of video encoding, which is essential for video services in mobile environments, this article introduces a model of visual characteristics based on motion in video, for which there have hitherto been few applications, and describes the effects of applying this model.

Research Laboratories *Akira Fujibayashi*
Choong Seng Boon

1. Introduction

In recent years, the increasing availability of broadband communication has led to an increase in the popularity of video services such as video sharing sites and on-demand video delivery services. In mobile environments, the provision of video services has been facilitated by advances in mobile terminals and the introduction of high speed communication techniques such as HSDPA (High-Speed Downlink Packet Access)[*1], and has generated a new market for video delivery services. NTT DoCoMo recently launched the 10 MB i-motion service and introduced the FOMA 905i series which incorporates, as a standard feature, a high resolution display equivalent to VGA (Video Graphics Array)[*2]. This technology allows users to enjoy high-quality video on mobile terminals. As shown in **Table 1**, a recent survey [1] showed that 35.6% of users make use of video viewing functions including videophone, showing that video viewing in mobile terminal is becoming a part of everyday life. This leads to increased expectations for improved quality in video content, making it necessary to develop technologies that can meet these expectations.

In video services that operate in mobile environments, the limited transmission bandwidth makes it essential to adopt efficient video compression techniques to reduce the amount of data that has to be transmitted. Systems such as videophones, V-live, and i-motion use MPEG-4 visual (Moving Picture Experts Group phase 4 visual)[*3], while One Seg uses the state-of-the-art H.264/AVC

Table 1 Functions available on mobile terminals

Function	Availability (%)
Camera	87.2
Applications (games, etc.)	42.7
Barcode reader	25.4
★Video file playback	20.9
Music playback	13.6
★Videophone	8.5
★TV broadcast reception	6.2
GPS	8.0
Video playback (total of items marked ★)	35.6

GPS : Global Positioning System

*1 **HSDPA**: A high-speed downlink packet transmission system based on W-CDMA. Used by NTT DoCoMo in the "FOMA High-Speed" service.

*2 **VGA**: A video display resolution of 640×480 dots.

*3 **MPEG-4 visual**: The video encoding part of the MPEG-4 video format specification aimed at the delivery of high-quality video over low-speed circuits defined by International Organization for Standardization (ISO).

(Advanced Video Coding)[*4]. Despite using the latest video compression techniques, however, the video of One Seg programs, for example, may contain visible block distortion and blurring artifacts as shown in **Photo 1**. It is therefore not always possible to provide high-quality videos services. In particular, for highly popular content such as sports events and horse races, in which the whole scene changes instantaneously due to very large motion (hereinafter referred to as "large motion") it is very challenging to reduce the amount of data even with H.264 encoding while providing high-quality video.

Therefore, with the aim of implementing a video encoding technique that substantially improves upon conventional techniques in terms of the quality of videos that include large motion, we have conducted research on the utilization of visual characteristics based on motion in video sequences, which has hitherto had few applications.

In this article we first describe the illusions[*5] perceived by humans as a result of motion in video sequences. In particular, we will focus on an illusion called Motion Sharpening (MS), and discuss our efforts to construct a model by clarifying how this illusion works from the viewpoint of visual masking effects. The effects of applying this model to video encoding will also be described.

Photo 1 Distortion seen on services such as One Seg

2. Visual Characteristics Based on Motion in the Video

2.1 Illusions Caused by Motion in the Video

Humans perceive video sequences through the eyes. The impression that people get from watching a video sequence changes according to the physical characteristics of the video. In general, a video consists of a wide range of spatiotemporal frequency components[*6]. If there are high spatial frequencies included in the image then this image is perceived to be sharper, while conversely if these components are not included then the image looks blurred. Thus the perception of images operates in response to the input of spatial frequency characteristics of the images into the eyes. However, human eyes do not always perceive things exactly as they are. For example, you may have experienced how recorded video sequences look blurred when you press the pause button, but appear sharp again when you resume playback. This is due to the phenomenon called MS.

Figure 1 shows the sharpening effect of a video sequence based on motion in the video. The MS phenomenon acts in such a way that, even if all the images constituting the video are blurred as shown in Fig. 1(2), they appear sharper when played back as a moving image than when they are static. The MS phenomenon has been studied for over 20 years in the field of visual science. It is said that

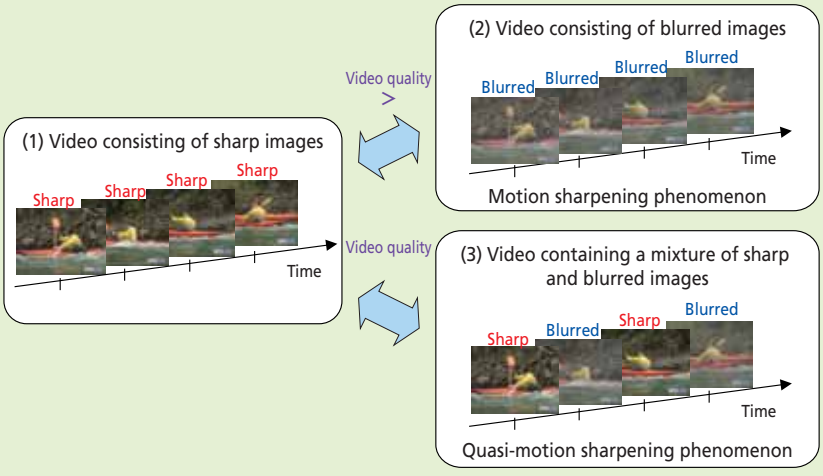


Figure 1 Image sharpening effect based on motion

*4 **H.264/AVC**: A video encoding method standardized by the Joint Video Team (JVT) — a partnership between MPEG and the JTC 1 (Joint Technical Committee 1) of the ISO/International Electrotechnical Commission (IEC). It achieves approximately twice the compression efficiency of earlier compression methods such as MPEG-2, and is used as the standard video format in services such as One Seg broadcasting.

*5 **Illusion**: A phenomenon where the characteristics of an object as perceived by the eye do not match the objectively measured characteristics.

*6 **Spatiotemporal frequency component**: A representation in the frequency domain of the differences between pixels of a picture in the horizontal and vertical directions (spatial directions) and between pixels at the same position in each frame (temporal direction).

MS is caused by a nonlinear mechanism of the initial visual system, but the explanation of how MS occurs remains inconclusive [2].

There is a special case of the MS phenomenon [2]. In general, the MS phenomenon increases its apparent sharpness even if all the images constituting the video are blurred as shown in Fig. 1(2), but, it still looks blurry compared with a video composed of images with uniform sharpness as shown in Fig. 1(1) . It is sometimes possible to obtain the same apparent level of quality as the video of Fig. 1(1) by mixing together sharp and blurred images as shown in Fig. 1(3). This means that, when some blurred images are mixed in a video sequence, it is possible to maintain the same subjective quality as when none of the images are blurred.

Hereafter we will refer to the phenomenon of Fig. 1(3) as "Quasi-Motion Sharpening" (Quasi-MS). In the following section, we construct a hypothesis in order to clarify the mechanism behind this phenomenon.

2.2 Mechanism of the Quasi-MS Phenomenon

In Fig. 1(3), the Quasi-MS phenomenon can be seen as the result brought about by the periodic appearance of sharp images. The sharp images partially negate the ambiguity of the blurred images, which hinders the correct perception of the blurred images. This perceptual effect is called visual masking. Visual masking is the phenomenon whereby the perception of a weak stimulus (target stimulus) is obstructed by a temporally or spatially adjacent strong stimulus (masker stimulus). Specifically, we considered that the Quasi-MS phenomenon shown in **Figure 2** occurs because the sharp images act as the masker stimulus and obstruct the perception of the blurred images, which serve as the temporally adjacent target stimulus. This leads to a level of perceived quality similar to that for the same sequence of images but with uniform sharpness.

So what properties of the video play a role in the mechanism of masking in video images? We considered the following three properties:

1) Motion Property

There is a strong relationship between the MS phenomenon and the size of motion in video [2]. The motion property is the temporal changes or the displacement of pixels between the masker image and target image, and in this study as an initial investigation, we concentrated on global motion of the video.

2) Average Luminance

This is the brightness of the video. As the viewing ability of video sequences depends on the brightness, this property has a strong effect on the visual characteristics.

3) Power Spectrum[*7] of Image

This property is strongly related to visual masking. Masking is determined by the magnitude relationship of the target stimulus and masker stimulus at each of the spatial frequency components of the image, and its effect is proportional to the power of the masker stimulus [3].

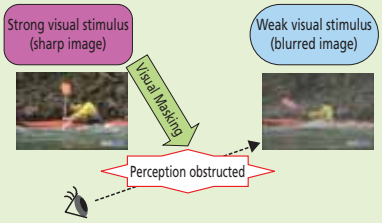


Figure 2 Clarification of masking mechanism

In Chapter 3, we clarify the conditions for maintaining subjective video quality by the masking effect based on changes of these three properties of the video.

3. Construction of a Quasi-MS Model

3.1 Subjective Assessment Tests

By performing subjective assessment tests using simple video sources such as a sine-wave grating[*8], which is commonly used in surveys of visual characteristics, we clarified the quantitative relationships between changes in the three properties and changes in the spatial frequency characteristics of the target stimulus [4][5] such as the power spectrum and bandwidth.

The resulting relationships are shown below:

Relationship (1): As the motion in the video increases, the bandwidth and power spectrum of the target image can be restricted more, and hence the target image can be made more blurry.

Relationship (2): When the average

*7 **Power spectrum**: A representation of the strength of a signal in each frequency component.

*8 **Sine-wave grating**: An image in which the brightness varies sinusoidally.

luminance is low, it is harder to restrict the spatial frequency characteristics of the target image.

Relationship (3): The amount of change in the power component of each spatial frequency in the target image bears a proportional relationship to the power component of the same spatial frequencies in the masker image.

If we assume that the amount of change in the power component contributes to the masking effect, relationship (3) is thought to be identical to the characteristics of visual masking as described in Chapter 2. Therefore this result supports the validity of our hypothesis in explaining the mechanism of the Quasi-MS phenomenon from the viewpoint of visual masking.

An example of the experimental results is shown here. **Figure 3** shows the relationship between the amount of motion and the bandwidth of the target image for maintaining the subjective video quality. The horizontal axis shows the amount of motion in the video (pixels per frame), the vertical axis shows the bandwidth of the spatial frequencies in the target image in cycles per degree (cpd)[*9], and the lines show the difference in average luminance. From this figure it can be seen that the bandwidth of the spatial frequencies of the target image can be restricted as the motion in the video becomes larger. With the restriction of bandwidth, each image appears more blurred, but from relationship (1) there is no loss in subjective video quality

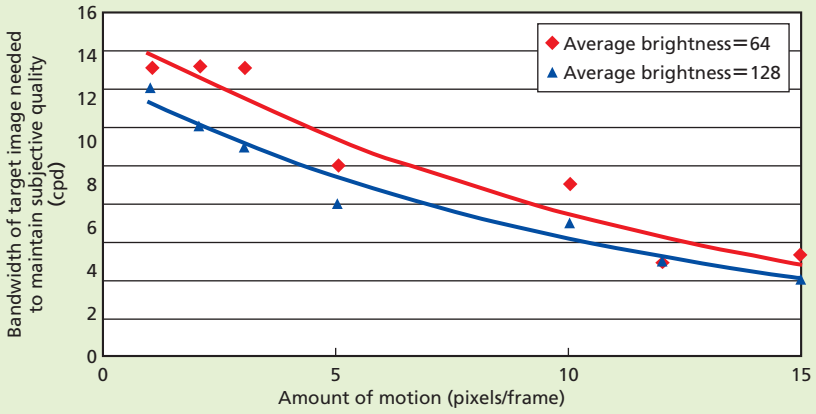


Figure 3 Relationship between amount of motion and bandwidth of target image needed to maintain subjective quality

Furthermore, the results obtained at low luminance (red line) occupy a higher position than the results obtained at high luminance (blue line). This means that the bandwidth cannot be restricted if the average luminance is low, thus proving relationship (2).

3.2 Quasi-MS Model

From the experimental results described in Section 3.1, we can construct a model of the Quasi-MS phenomenon. Breitmeyer et al. proposed that visual sensitivity is composed of two sensitivity channels known as the "transient" and "sustained" channels[*10] [3]. By hypothesizing a structure with these two channels, they successfully explained a wide range of masking phenomenon [3]. Based on this idea, we constructed a Quasi-MS model that approximately expresses the three properties discussed in Section 2.2 and the corresponding spatial frequency characteristics of the target images.

The equation of our model is shown below.

$$P_{TH}(f)=\frac{\alpha}{1+(\beta\times f/9)^8}+\gamma\times\exp\left[-\left(\frac{f-\omega}{\tau}\right)^2\right]$$

The above equation uses five coefficients (α, β, γ, ω, τ) determined on the basis of the three properties of video (motion property, average luminance of sharp images and power component of the spatial frequency f of sharp images) to calculate the power component P_{TH} of the spatial frequency f of the target image necessary for maintaining the subjective video quality.

The correlation coefficient between the data calculated with this model and the experimental data attains a value of more than 0.93, confirming that our model is very accurate [6].

*9 **cpd**: The number of periods of an image signal of a wave pattern presented on a plane that occupy 1° of the visual field. An image with a larger cpd value is perceived as a pattern with a greater level of detail.

*10 **Transient channel / sustained channel**: A transient channel is a sensitivity channel whose characteristics are short-lived or are highly sensitive to large changes in a temporal stimulus, and conversely have low sensitivity to stimuli representing the overall state of a picture. On the other hand, a sustained channel is a sensitivity channel whose characteristics are highly sensitive to stimuli that change little with time and to stimuli that express the details of a picture.

3.3 Application to Natural Video

We evaluated the applicability of this model to typical natural video by using it to process alternate frames in four video sequences. The three properties of each video sequence were analyzed. The power spectrum of the target image was then restricted on the basis of the Quasi-MS model using these three properties. For verification of the video quality obtained with the Quasi-MS model, subjective assessment tests were conducted using natural video sequences containing large motion. Results showed that our model was applicable to three out of the four sequences. In the sequences where the subjective quality was not maintained, there were stationary regions in parts of the sequence. In such cases, a model constructed under the assumption of global motion may not be suitable.

Photo 2 shows the image quality obtained when this model was applied to one scene from the "American football" sequence. Photo 2(a) is the original image, and photo 2(b) is the image obtained after applying the model. From photo 2, it can be seen that for a still image, features such as the numbers on the players' backs are perceived as being blurred, but when viewed as a moving sequence, the video appears to be sharp.

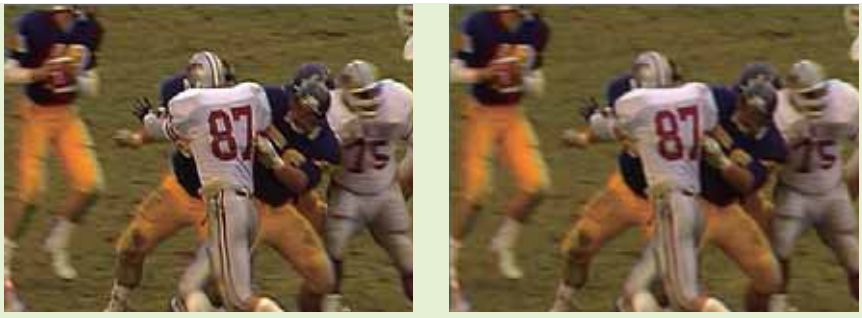
4. Effects of Applying This Model

When applying Quasi-MS model to video encoding schemes, the power spectrum of the target image is restricted, and irrelevant information for perception of the video can be eliminated while maintaining the subjective quality. Here, we present the results of encoding the standard test sequences[*11] provided by MPEG in an encoding environment in accordance with the specifications of i-motion.

Using original video source material for four different sports (American football, canoeing, rugby and F1 racing), we compared the results of encoding using H.264 to those of the Quasi-MS model. The results showed that it was possible to improve the encoding efficiency as shown in **Table 2** while maintaining the same level of subjective quality even when viewed by video experts. As Table 2 shows, for video containing large motion, for which it was hitherto difficult to reduce the amount of information, improvement in the encoding efficiency by 23-33% can be achieved. The proposed method also produces higher quality video with the same amount of data as conventional video encoding.

5. Future Work

The model will be extended to cover a wider range of video conditions, such as variance in luminance between images and partial motion within an image. In addition, since the model was initially constructed for use in a mobile environment, and since human visual characteristics vary with the viewing factors, the model will be adapted for use with various screen sizes and viewing distances. The application of this model in other fields will also be investigated.



(a) Original image (b) Image processed by model

Photo 2 Change in picture quality resulting from application of the Quasi-MS model

Table 2 Comparison of bit rates

	H.264 (kbit/s)	Quasi-MS (kbit/s)	Improvement of coding efficiency (%)
American football	664.2	532.9	27.00
Canoeing	707.2	510.6	28.05
Rugby	713.9	548.8	23.71
F1 racing	658.5	451.6	33.99

*11 **Standard test sequence**: A common group of videos that are used for the evaluation of methods proposed for standard techniques defined by the standards organization.

6. Conclusion

In this article, we discussed the Quasi-MS phenomenon as an application of visual characteristics based on motion in the video with the aim of improving the quality of video. We also clarified the mechanism of this illusion and constructed a model of this mechanism. By applying this model to video encoding techniques, we have shown that it can improve the encoding efficiency as well as the subjective quality of video with a large motion. We plan to integrate this encoding technique with other signal processing techniques and the like and to develop an encoding technique that enables more effective use of the Quasi-MS phenomenon.

REFERENCES

[1] Dentsu Communication Institute Inc.: "Information media white paper 2007," Diamond, Inc., 2007.

[2] T. Takeuchi and K.K. De Valois: "Sharpening image motion based on the spatio-temporal characteristics of human vision," Human Vision and Electronic Imaging X, 2005.

[3] Vision Society of Japan: "Visual Information Processing Guidebook," Asakura Publishing Co., Ltd., 2000.

[4] A. Fujibayashi, S. Kato, C.S. Boon, S. Hangai and T. Hamamoto: "Clarification of the relationship between video movements and motion sharpening effects," IEICE General Conference, 2006.

[5] A. Fujibayashi, S. Kato and C.S. Boon: "Modeling of Perceptual Sensitivity towards Motion Sharpening Phenomenon of Video Sequences," Proc. Second International Workshop on Image Media Quality and Applications, 2006.

[6] A. Fujibayashi and C.S. Boon: "A Masking Model for Motion Sharpening Phenomenon in Video Sequences," IEICE English Journal, Special Section on Image Media Quality (to be published), 2008.