Video Enlargement | Noise Reduction | Video Service Quality Improvement

Technology Reports

Super-resolution Video Reconstruction Technology

The demand to view high quality video on mobile terminals is increasing. To enable higher quality video viewing, we developed technology for high quality video enlargement. This technology is effective even for video content that has a lot of coding distortion as can be seen on video sharing sites.

Research Laboratories
Akira Fujibayashi
Boon Choong Seng

1. Introduction

Video services such as video sharing sites and video on demand have been attracting interest in recent years as the bandwidth of communication has been increasing. In mobile environments, the groundwork for enjoying video services and One Seg broadcasting has been established by the increasing sophistication of mobile terminals, the implementation of high speed communication techniques such as High Speed Downlink Packet Access (HSDPA)[*1] and flat rate packet charging schemes.

According to a survey done in May, 2008 [1], both the frequency and length of time spent in viewing video on mobile terminals has increased since a year ago. In particular, the browsing of video sharing sites has been increasing, and it is becoming common to enjoy video on mobile terminals.

According to another survey [2], there is a large user demand for viewing the content of video sharing sites at "higher quality" and "larger screen size" (**Table 1**). Similar demand also exists for One Seg broadcasting (**Table 2**) [3].

A high resolution display is one of the features that is related to a better video viewing experience on mobile terminals. The high-specification newer devices appearing after the FOMA 905i series are equipped with displays capable of resolution higher than Video Graphics Array (VGA)[*2] and allow the viewing of large images. Mobile terminals are also equipped with video quality enhancement technology that was originally developed for large-size liquid crystal TVs by various manufacturers, allowing high-quality video display on mobile terminals.

On the other hand, the image quality provided in current video service is not high. The quality of video that can be seen on video sharing sites is relative low, and the original size of most of the video provided in current video services

Table 1 Expectations for video sharing sites

Points where improvement is expected	Response (%)
Video quality	55.4
Screen size	37.8
Communication speed	38.2

Table 2 Important points for One Seg broadcast viewing

Important points	Response (%)
Clear video	46.6
Good reception sensitivity	40.8
Large screen	34.1

*1 **HSDPA**: A high-speed downlink packet transmission system based on W-CDMA. Maximum downlink transmission speed under the 3GPP standard is approximately 14 Mbit/s. Optimizes the modulation method and coding rate according to the radio reception status of the mobile terminal.

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

is only about 1/4 of the display capability of mobile terminals.

This situation creates a need for technology that allows a better viewing experience of high quality enlarged video, even for low resolution video, to take advantage of the high resolution display capabilities of mobile terminals.

We therefore took up the challenge, in collaboration with DOCOMO USA Labs, Communication Device Development Department and Morpho, Inc., of developing technology for enlarging low resolution video to high quality video (images) to allow users to enjoy video at higher quality.

In this article, we first focus on increasing video resolution for a better viewing experience and summarize the current situation of video services and the issues regarding them. We then describe the features and issues of conventional technology for high resolution video enlargement. Finally, we explain the technology that can solve those issues and report on its performance.

2. Current Situation of Video Services

The limited transmission bandwidth in a mobile environment makes it essential to adopt video compression technologies to reduce the amount of data that has to be transmitted for receiving. The video that can be viewed on video sharing sites such as YouTube®*3 or Niconico Douga must also be compressed for efficient server operation.

For video compression, videophones and i-motion use Moving Picture Experts Group phase 4 Visual (MPEG-4 Visual)*4, One Seg uses H.264/Advanced Video Coding (AVC)*5, and video sharing sites use the Sorenson Spark*6, On2 VP6*7 or other such video codec in Flash®*8 Video (FLV) format. Such compression technology allows efficient use of bandwidth and reduces server load by reducing the amount of video data that is handled.

On the other hand, use of compression technology means loss of some information from the original video, and that results in block distortion*9, flicker near sharp edges (mosquito distortion), blur or other such artifacts that are called coding distortion.

An example of coding distortion is shown in **Figure 1**. The compressed image is blurred over the entire screen, and a lot of block distortion and mosquito distortion can be observed. These kinds of distortion are very common in the video available on video sharing sites in particular, and there is a lot of low-quality content.

The most common resolution currently used in video services is lower than VGA because of bandwidth limitation and server management policies. The video size of One Seg and i-motion is Quarter Video Graphics Array (QVGA) at 320×240 pixels and the video size of videophones is Quarter Common Intermediate Format (QCIF) at 176×144 pixels. The video on video sharing sites generally has a resolution close to QVGA, which is low compared to the display capabilities of mobile terminals, that is, the video size itself is small.

Therefore, video displayed as is becomes small compared to the display

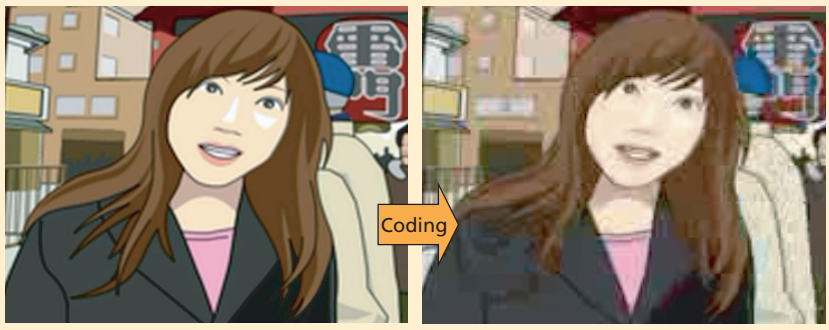


Figure 1 Example of coding distortion

*3 **YouTube®**: A registered trademark of Google, Inc.

*4 **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).

*5 **H.264/AVC**: A video encoding method standardized by the Joint Video Team (JVT) - a partnership between MPEG and the Joint Technical Committee 1 (JTC1) 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.

*6 **Sorenson Spark**: One codec used for the Adobe FLV video format.

*7 **On2 VP6**: A video codec used by Flash Player 8 and originally developed by On2 Technologies Inc.

*8 **Flash®**: A trademark or registered trademark of Adobe Systems Inc. in the United States and other countries.

size, and current full-screen displayed image tends to blur because of simple interpolation processing. It is thus not possible to enjoy enlarged video at high quality, even when using full-screen display.

These kinds of issues where the video quality and resolution are restricted in current video services point to the need for technologies that provide comfortably viewable video at high quality that matches the high resolution display capability of mobile terminals.

3. Higher Video Resolution on Mobile Terminals

Improving the quality and resolution of the video offered by a video service requires higher transmission bandwidth, changes in the system, and the preparation of large size content, measures that are difficult to implement immediately.

We therefore developed technology for high quality enlargement of videos on the mobile terminal to enable viewing of video at larger size and higher quality.

If it were possible to enlarge the received video with a quality (higher resolution) that matches the display performance of the mobile terminal, even video whose quality and resolution are low because of various constraints could be viewed at high quality (**Figure 2**). It would also be possible to output video content acquired with a mobile terminal to a large external display.

3.1 Conventional Video Enlargement Technology

Conventional video enlargement techniques include pixel interpolation and super resolution technology (**Figure 3**).

Pixel interpolation is a video enlargement technique that calculates pixel values that originally do not exist in the image from surrounding pixel values, as shown by the triangle marks in Fig. 3. Typical pixel interpolation methods include the bi-linear method, the nearest-neighbor method[*10], and the bi-cubic method[*11].

The bi-linear method (Fig. 3(a)), for example, performs a linear interpolation between adjacent pixels. Pixel interpolation is generally fast and features low computational load, but the pixel values tend to change smoothly, so this technique does not improve sharpness and results in blurring in the enlarged image.

Super resolution technology [4][5] on the other hand produces sharper enlarged video. Super resolution techniques generally use advanced computation. They generate pixel values that are not in the original images from multiple images in the video to generate video of very high sharpness (Fig. 3 (b)).

However, processing that requires advanced computation is difficult on devices that have limited computational power, such as mobile terminals. Furthermore, these conventional techniques emphasize the coding distortion

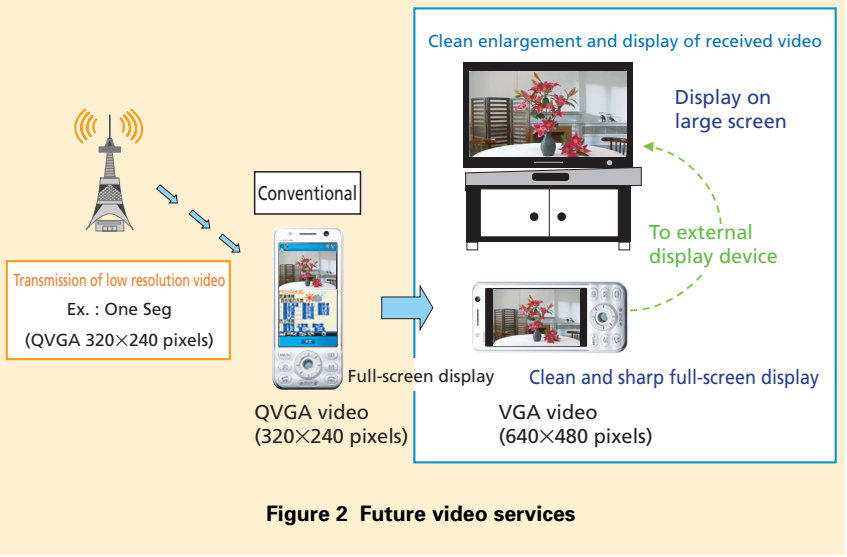


Figure 2 Future video services

*9 **Block distortion**: A kind of distortion in which mosaics and other block-shaped boundaries can be seen. It is particularly visible in screen switching and scenes that involve a lot of motion, etc.

*10 **Nearest-neighbor method**: A pixel interpolation method that copies the values of the nearest pixels as the interpolation value.

*11 **Bi-cubic method**: A pixel interpolation method in which the interpolation value is generated by polynomial computation using the surrounding 16 pixels.

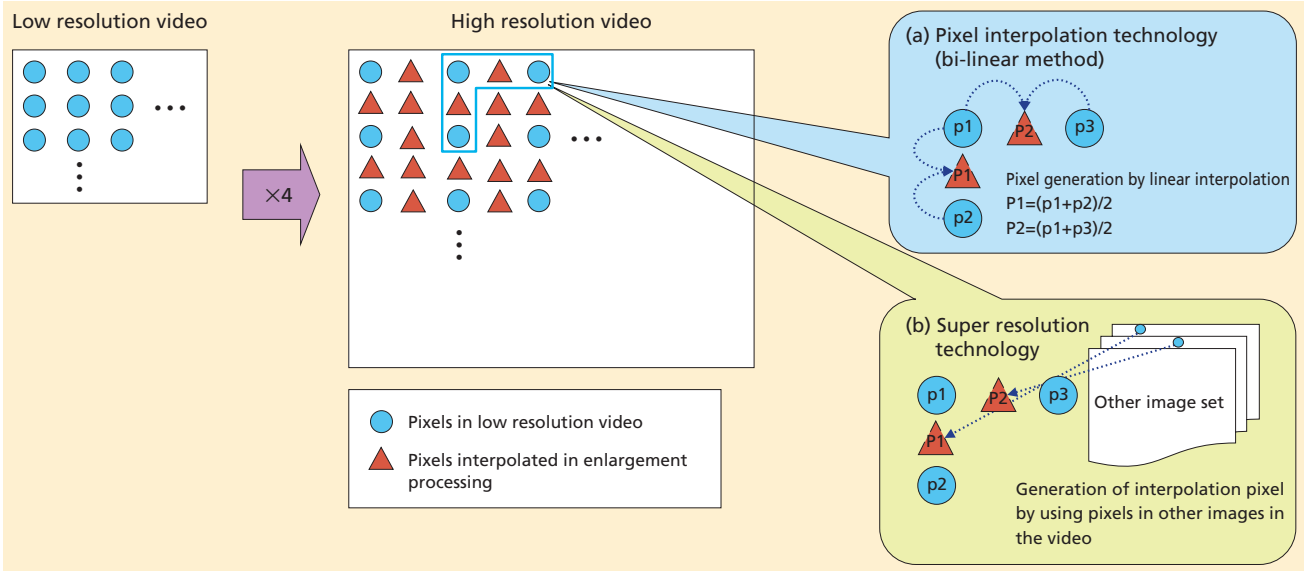


Figure 3 Example of conventional resolution increasing techniques

in the enlargement process. The result is that distortion not noticeable in the smaller video size becomes obvious upon enlargement and these regions in the image are not seen as high quality.

3.2 Original Video Enlargement Technology

We developed technology for enlarging video at high quality and sharpness by reducing video coding distortion and suppressing the aliasing noise[*12] and blur caused by pixel interpolation. The processing of this technology can be executed on single image, so the processing load can be less than that of super resolution technology, which uses multiple images [6][7].

3.3 Overview of the Original Technology

This technology combines processing for "distortion reduction," "pixel interpolation," and "image sharpening" to generate enlarged video with high quality (**Figure 4**).

There are two profiles for this technology. One is a low-computation profile that is mainly for improving the quality of video that has a lot of distortion. The other, the full profile, uses the maximum capabilities of the technology to produce video of greater sharpness. The low-computation profile is capable of real-time processing of One Seg broadcasts with software on a mobile terminal.

Of the three kinds of processing, the distortion reduction processing and image sharpening processing, which are original NTT DOCOMO technologies, are both based on a method known as the over complete transform.

3.4 Over Complete Transform

The over complete transform uses local similarity of the pixel values that constitute the image (**Figure 5**). If we enlarge the area framed in red in Fig. 5(a), we can see that it includes pixel areas A (brown) and B (flesh tone) and an edge, as in Fig. 5 (b). A certain pixel P (indicated by the x mark in the figure) has a value close to another pixel that originally belongs in area B, but if coding distortion is added to Fig. 5 (a), such as seen in the green-framed area in the image on the right, then pixel P may take on a value that is greatly different

*12 **Aliasing noise**: Appears as jagged lines and edges and clumping artifacts in an image.

from the values of other pixels that belong to area B.

Therefore, the distortion can be reduced by taking various small areas into consideration and identifying areas in which the local similarity is high and then using a small region that has high similarity to pixel P to approximate the value of pixel P.

For example in Fig. 5 (b), we consider the different small areas that include pixel P in Fig. 5 (c) and Fig. 5 (d). The blue-framed small area in Fig. 5 (c) contains pixels from areas A and B, which have greatly different values, so pixels in that small area do not have a high similarity. In Fig. 5 (d) on the other hand, all pixels in the small area belong to area B and have high similarity.

The similarity here is judged according to the transform coefficient for the frequency transform (**Figure 6**). In Fig. 6, the horizontal axis represents spatial frequency and the vertical axis represents the magnitude of the transform coefficient. If the pixel similarity is high, as in Fig. 5 (d) and a frequency transform is done for that area, then the number of transform coefficients that have large values is small, as in Fig. 6 (a), and they are represented by transform coefficient $X(k)$.

If that small area contains distortion, $X(k)$ is changed to $c(k)$ ($=X(k)+noise$), which additionally includes small transform coefficients due to the distortion, as shown in Fig. 6 (b). Therefore, distortion can be reduced and $c'(k)$ ($\fallingdotseq X(k)$), which is close to coding-distortion-free $X(k)$, can be obtained by removing only the transform coefficients that have small values in $c(k)$, as in Fig. 6(c), thus raising the similarity of $c(k)$.

In this way, the processing to improve the similarity in frequency regions is executed for each pixel in multiple small areas, and the results are added as a weighting function to match the pixel similarity.

The result of this processing is

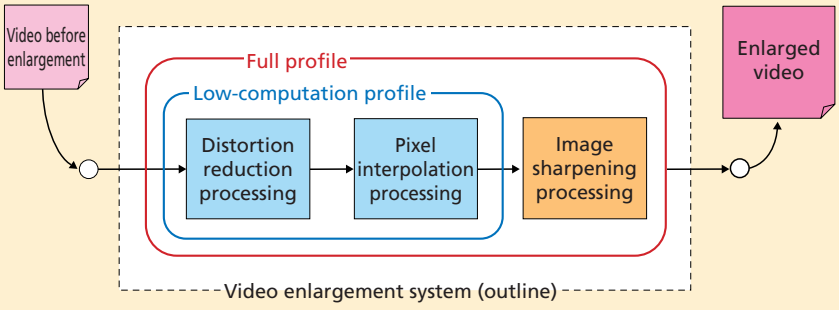


Figure 4 The developed technology

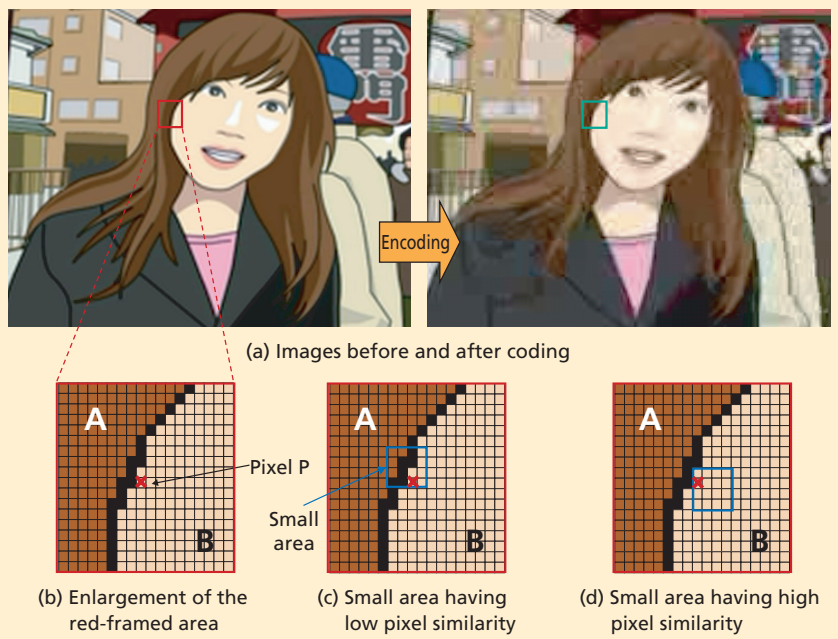


Figure 5 Local pixel similarity

reduction of the coding distortion present in the image before enlargement as well as the aliasing noise and signal broadening that accompany pixel interpolation in the enlarging process. Thus, the distortion in video enlarged by this technique is reduced and subjectively clear edges can be viewed.

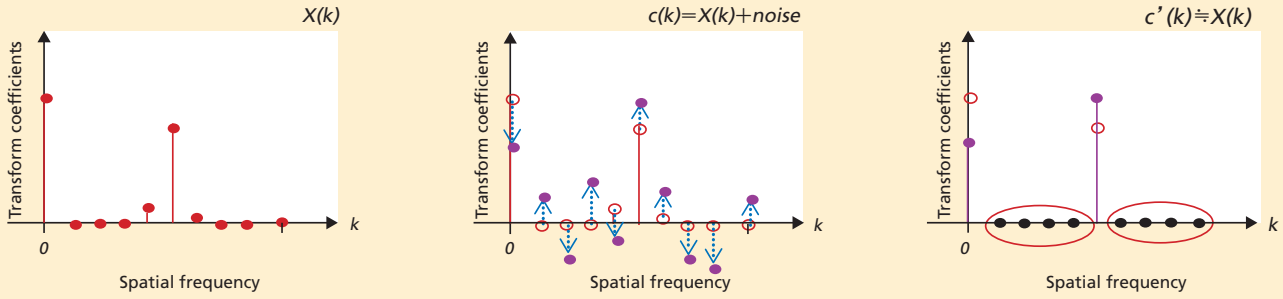
3.5 Performance of the Original Technology

The difference in quality of video enlarged by the conventional technology and by our original technology is shown in **Figure 7** and **Figure 8**. In both figures, (a) is the image before enlargement, (b) is the image enlarged to twice the size with the conventional bi-linear method, and (c) is the image enlarged with the original technology. In both figures, we can confirm that the conventional technique makes the coding distortion noticeable, and it cannot be seen as high quality. This kind of distortion is perceived as flicker in video, and leads to a decrease in video quality.

With our original technology on the other hand, the distortion around people and written characters is reduced and the edges from the original image are clear, allowing a more enjoyable video viewing.

4. Conclusion

In this article, we described the current state of video services and the issues faced. We also presented an overview of video enlargement technol-



(a) Transform coefficients for the region that has high similarity

(b) With noise included in $X(k)$

(c) Transform coefficients after noise removal

Figure 6 Improvement of similarity

(a) Image before enlargement

(b) Image enlarged by the bi-linear method

(c) Image enlarged by our original technology

Figure 7 Comparison of image quality 1

(a) Image before enlargement (b) Image enlarged by the bi-linear method (c) Image enlarged by our original technology

Figure 8 Comparison of image quality 2

ogy designed to provide large size video at higher quality to take advantage of the display capabilities of mobile terminals and explained the technique used and its performance. This technology incorporates a mechanism for reducing coding distortion and blur, and is thus capable of enlarging video, including the low resolution video of less than high quality submitted to sharing sites, with greater sharpness and higher quality. Currently, various manufacturers and vendors are investigating the implementation of our technology for mobile terminals to be sold after the end of 2009.

Future work includes construction of cooperative relationships with vendors and continuing to verify and improve the video enlargement performance for various types of applications (effective use of video assets and high video quality enlargement and display on DVD players and media players) for commercial development of this technology.

References

[1] Mobile Research: "Survey on Cell Phone Viewing of Motion Picture Content" (In Japanese).

[2] iShare, Inc.: "Survey on Cooperation between Video Sharing Sites and Television" (In Japanese).

[3] Net Research DIMSDRIVE Questionnaire: "Questionnaire on One Seg Devices" (In Japanese).

[4] Y. Altunbasak, A. Patti and R. Mersereau: "Super-Resolution Still and Video Reconstruction From Mpeg-Coded Video," IEEE Trans. Circuits Systems for Video Tech., Vol. 12, 2002.

[5] S. Farsiu, D. Robinson, M. Elad and P. Milanfar: "Fast and robust multi-frame super-resolution," IEEE Trans. on Image Processing, Oct. 2004.

[6] S. Kanumuri, O. Guleryuz and M. R. Civanlar: "Sparse directional transforms for super resolution reconstruction of mobile video," in Proc. SPIE, San Diego, Aug. 2007.

[7] S. Kanumuri, O. Guleryuz, M. R. Civanlar, A. Fujibayashi and C. S. Boon: "Temporal Flicker Reduction and Denoising in Video using Sparse Directional Transforms," in Proc. SPIE, San Diego, Feb. 2008.