Resampling HD Images with the Effects of Blur and Edges for Future Musical Collaboration

Mauritz Panggabean and Leif Arne Rønningen
Department of Telematics
Norwegian University of Science and Technology (NTNU)
O.S. Bragstads plass 2B, N-7491, Trondheim, Norway
{panggabean,leifarne}@item.ntnu.no

Abstract

Image down/upsampling can give significant bit rate reduction with hardly noticeable quality degradation. This is attractive to our vision of real-time multiparty collaboration from distributed places with video data at HD resolution that must guarantee maximum end-to-end delay of 11.5ms to enable musical synchronization. Based on the Distributed Multimedia Plays architecture, this paper compares the performances of bicubic and Lanczos techniques as well as one recently proposed for image upsampling in terms of computing time, level of blur, and objective image quality in PSNR and SSIM. The effects of image blur and edges to resampling are also examined. The results show that the classic Lanczos-2 and bicubic techniques perform the best and thus are suitable for the vision due to their potential for efficient parallel processing in hardware. We also show that composite images with different downsampling factors can achieve 4dB increase in PSNR.

1 Introduction

The fields of electronics and communication technology have been growing rapidly and opening new and more creative ways of real-time multiparty collaborations limited only by time and space. Our vision is to realize such collaboration in the future with *near-natural* quality of experience using networked *collaboration spaces* that enable seamless integration of virtual (taped) and live scenes from distributed sites on the continents despite different technical specifications. Figure 1 depicts a simple exemplary scenario where an audience in Oslo (A) are enjoying a concert featuring two opera singers in a specially designed room, namely a *collaboration space*. The quality of experience is expected to be *near-natural*, meaning that they hardly realize that the two singers are performing live from two different cities, say Trondheim (B) and Tromsø (C), each in their own collaboration space. As both simultaneous performances are so harmonious with life-like multimedia quality, the audience feel that they are enjoying a live opera concert and being in the very same room with the two singers. Furthermore each opera singer singing live also experiences performing together with the other one displayed in his or her own collaboration space, as if they were on the same stage.

This paper was presented at the NIK-2010 conference; see http://www.nik.no/.

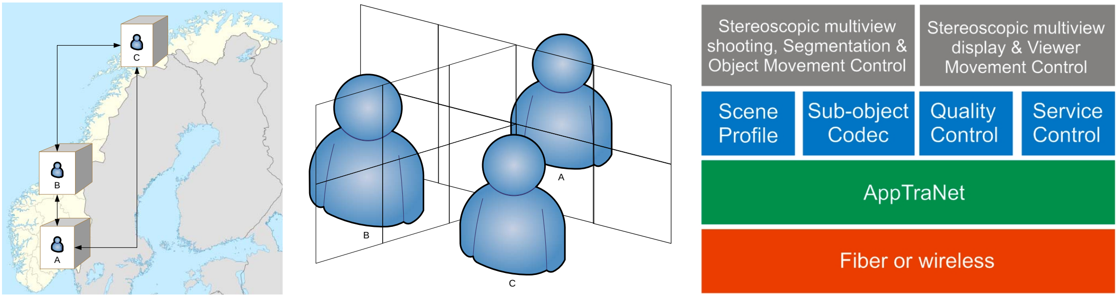
C
B
A
B
A
C
Stereoscopic multiview shooting, Segmentation & Object Movement Control
Stereoscopic multiview display & Viewer Movement Control
Scene Profile
Sub-object Codec
Quality Control
Service Control
AppTraNet
Fiber or wireless

Figure 1: A simple example of the envisioned collaboration (left) and the combined networked collaboration spaces (middle). All surfaces of the collaboration space consist of arrays of multiview autostereoscopic 3D display, dynamic cameras, speakers and microphones. The multimedia output data is handled by the proposed three-layer DMP architecture shown from a user's perspective (right).

Table 1 lists the main technical requirements on the important aspects for the envisioned collaborations. Guaranteeing the maximum end-to-end delay $\leq$ 10-20ms to enable good synchronization in musical collaboration is the most challenging. Chafe et al. [2] reported experimental results on the effect of time delay on ensemble accuracy by placing pairs of musicians apart in isolated rooms and asking them to clap a rhythm together. Longer delays were reported to produce increasingly severe tempo deceleration while shorter ones yield a modest yet surprising acceleration. The study found that the observed optimal delay for synchronization is 11.5ms which equates with a physical radius of 2,400 km (assuming signals traveling at approximately 70% the speed of light and no routing delays).

Table 1: The main technical requirements for the envisioned collaborations derived from the aimed quality of experience.

Nr.	Main technical requirements
1.	Guaranteed maximum end-to-end delay $\leq$ 10-20ms (cf. 400ms for videoconference [1])
2.	Near-natural video quality
3.	Autostereoscopic multi-view 3D vision
4.	High spatial and temporal resolution due to life-size object dimension (mostly humans)
5.	Accurate representation of physical presence cues e.g. eye contact and gesture
6.	Real 3D sound
7.	Quality allowed to vary with time and network load due to different technical specifications and behaviors among collaboration spaces
8.	Minimum quality guarantee using admission control
9.	Graceful quality degradation due to traffic overload or failure
10.	Privacy provided by defined security level

Realizing such collaborations with very high quality and complexity is possible only if all requirements can be fulfilled within 11.5ms. We observe that current standards are still unable to realize this vision and it leads to our proposal of the three-layer Distributed Multimedia Plays (DMP) architecture [3], as illustrated in Figure 1. Along with the proposed architecture, we also introduce the concept of Quality Shaping built upon the concept of Traffic Shaping [4] that degrades the video quality in graceful manner when traffic overloads or system malfunctions occur [3].

The traffic generated from near-natural scenes at high-definition (HD) resolution from the arrays of cameras is estimated to be extremely high which may be up to 10^3 – 10^4 higher than that of today's videoconferencing systems. Near-natural quality requires an extreme resolution with data rates in Gbps, even to represent a human face alone. Applying block-based video coding to the video data will incur more processing delays particularly for encoding that exploits interframe redundancies. Therefore DMP system favors full independence between frames and between objects within a frame in the video data to facilitate fully parallel video processing in hardware. Novel techniques for that purpose are under current investigation in our research. However, as additional phase prior to that, it is attractive to downsample HD images at the transmitter and upsample them at the receiver to considerably reduce more bit rate with hardly noticeable quality reduction. The resampling techniques must show promising potential for parallel hardware implementation for ultrafast operations.

This paper presents our study of the latter resampling idea based on the two following objectives. First, we compare four resampling techniques considered promising from literature in four comparison criteria: computational time, objective image quality assessment using Peak-Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) index [5], and quantified resulting blur. SSIM is used due to its close approximation to subjective image quality assessment and its widespread use. The four techniques are Lanczos technique with two and three lobes, bicubic technique [6] and that proposed by Shan et al. [7]. In the rest of this paper, the first three and the latter are respectively called the classic techniques and the new technique. Second, as blur and edges are often present in major regions in natural images, we examine their effects to the best performing techniques from the experiments on the first objective. Our goal is to exploit blur and edges in applying the resampling techniques to attain further bit saving for transmission without remarkable video-quality reduction.

The organization of the paper is as follows. Section 2 briefly presents the examined resampling techniques and elaborates the setup of the experiments. Section 3 presents the experimental results with evaluations that lead to our conclusions in Section 4.

2 Image Resampling Techniques and Experimental Setup

An in-depth formal explanation on image interpolation and resampling can be found for example in [8] where the authors define interpolation as a model based recovery of continuous data from discrete data within a known range of abscissa. Among many interpolation and resampling techniques for images, the Lanczos-windowed sinc functions offer the best compromise in terms of reduction of aliasing, sharpness, and minimal ringing [9]. Lanczos kernels with two and three lobes are widely used, as shown in Figure 2. They are respectively called Lanczos-2 and Lanczos-3 in the subsequent text. It is also a common fact that bicubic technique is widely used for most images due to its good trade-off between accuracy and speed. These explain why we select these techniques in this study.

In addition to these, most recent years have seen proposals of novel techniques for upsampling natural images, for example [7] and [10]. Although the output visual quality in the latter is promising, it is unacceptably slow for HD images. The first claims better visual quality than that of the latter and also presents possibilities for parallel hardware implementation, yet without clear indication of the required computing time. Thus it is included in our study only for image upsampling.

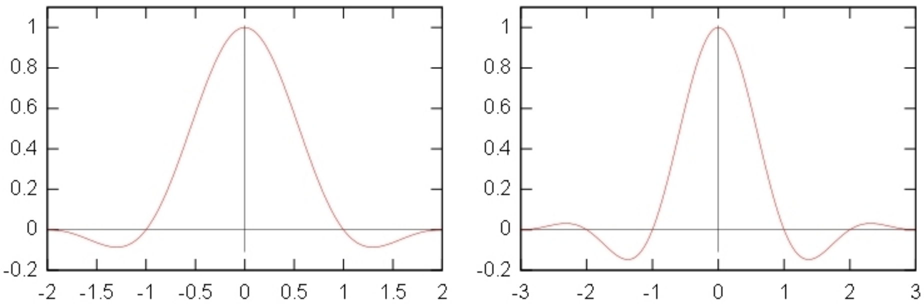

Figure 2: The kernels of Lanczos-2 (left) and Lanczos-3 (right) techniques.

Table 2: *DF*s and the output resolutions relative to 1920×1080 pixels.

DF	Output resolutions	*DF*	Output resolutions
1.2	1600×900	4.0	480×270
1.5	1280×720	5.0	384×216
2.0	960×540	6.0	320×180
2.5	768×432	7.5	256×144
3.0	640×360	8.0	240×135

To achieve the two objectives aforementioned, two experiments namely Experiment A and B are conducted in Matlab on a PC with 3GHz processor and 3.46GB RAM, respectively. As one exception, the image upsampling by the new technique employs the application provided online [7]. The details of the experiments are as follows. In Experiment A we assume that the test images are captured by a digital camera at 1920×1080 resolution and then they are downsampled with a number of downsampling factors (*DF*s). Afterward the output images from the downsampling are upsampled with the same *DF* by using the four techniques. The same classic technique is used for both downsampling and upsampling. For simplicity, in the rest of the text, the process of downsampling continued with upsampling the result with a *DF* is called *down/upsampling*. As the resolutions of the downsampled images must be integers, the selected *DF*s in Experiment A are 1.2, 1.5, 2.0, 2.5, 3.0, and 4.0, as shown in Table 2 with the resulting downsampled resolutions. *DF*s can denote approximate bit savings from the original test images. Figure 3 illustrates the comparison of the downsampled resolutions of images for transmission to show clearly the magnitude of the possible data reduction. The blur caused by resampling with increasing *DF* will be quantified by means of an objective blur metric [12, 13].

Experiment B examines the effects of image blur and edges to resampling. The level of blur is quantified as in Experiment A. The objective image quality will be presented as a function of *DF*s and the blur metric. The used *DF*s include those in Experiment A with extension to 4.0, 5.0, 6.0, 7.5 and 8.0, as listed in Table 2 with the corresponding downsampled resolutions. Different *DF*s are also applied in down/upsampling an image with clear and blurred regions. The quality of the resulting images will be quantified as a function of increasing *DF*s. Exemplary resulting images from both experiments will be provided for subjective assessment by the readers.

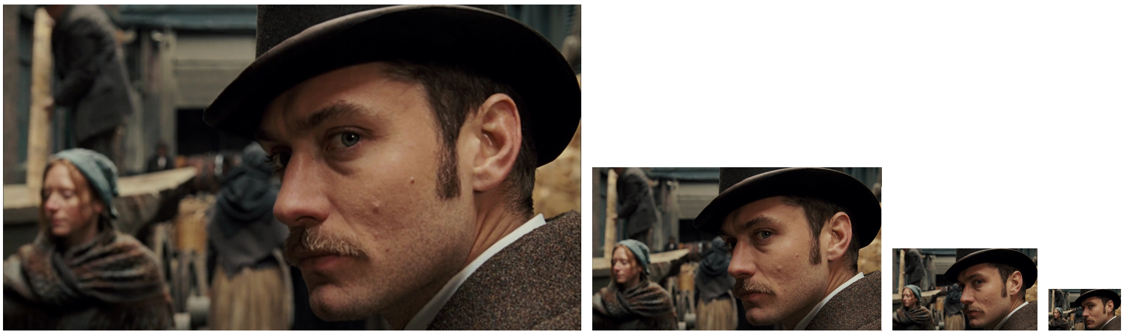

Figure 3: A sample image captured from [14] with original resolution (left) and, next to the right, those downsampled with *DF* equals 2.0, 4.0 and 8.0, respectively, to graphically illustrate the magnitude of the data reduction achieved by resampling.

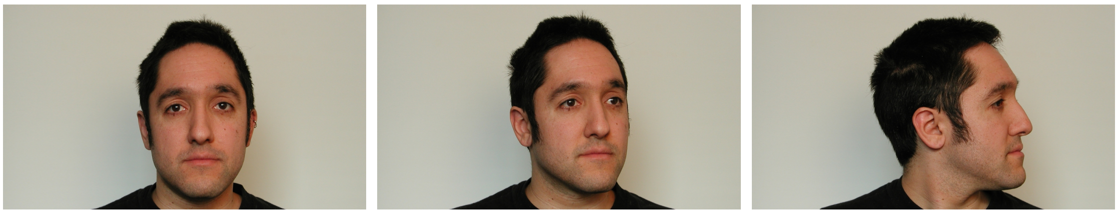

Figure 4: Typical test images with frontal (left) and non-frontal sides.

3 Experimental Results and Evaluations

This section presents more details on the two experiments with the results and evaluations as the main contributions from this work.

Experiment A: Comparison of Resampling Techniques

We used HD images of human faces from [11] in this experiment since collaboration spaces support live collaboration between humans and thus human faces are the most important object to process. All test images show human faces either from frontal or non-frontal sides, as shown in Figure 4. This suits the design of the collaboration space which arrays of cameras can capture the face of the persons within from both sides.

Our experimental results indicate that each resampling technique shows relatively constant performance in all comparison criteria not only in all test images, despite different positions of the faces, but also in natural images such as depicted in Figure 3. Thus it suffices to present the results from any test image as a general representation to compare the performances of the four resampling techniques. Figure 5 exhibits the experimental results for the test image at left in Figure 4. Sample images of the area around the right eye of the man in the test image are shown in Figure 6 to illustrate the effects of resampling for subjective assessment by the readers. This area is selected because it contains regions with and without many edges and high frequency contents. All images for assessment are to be seen on screen for best viewing quality. All the test images have plain background that fits our assumption that human faces and other important objects in a collaboration space can be efficiently segmented prior to downsampling.

All diagrams in Figure 5 confirm that Lanczos-2 and bicubic techniques deliver very similar performances in all comparison criteria. Although Lanczos-3 technique yields better image quality and causes less blur, they come at the expense of more processing

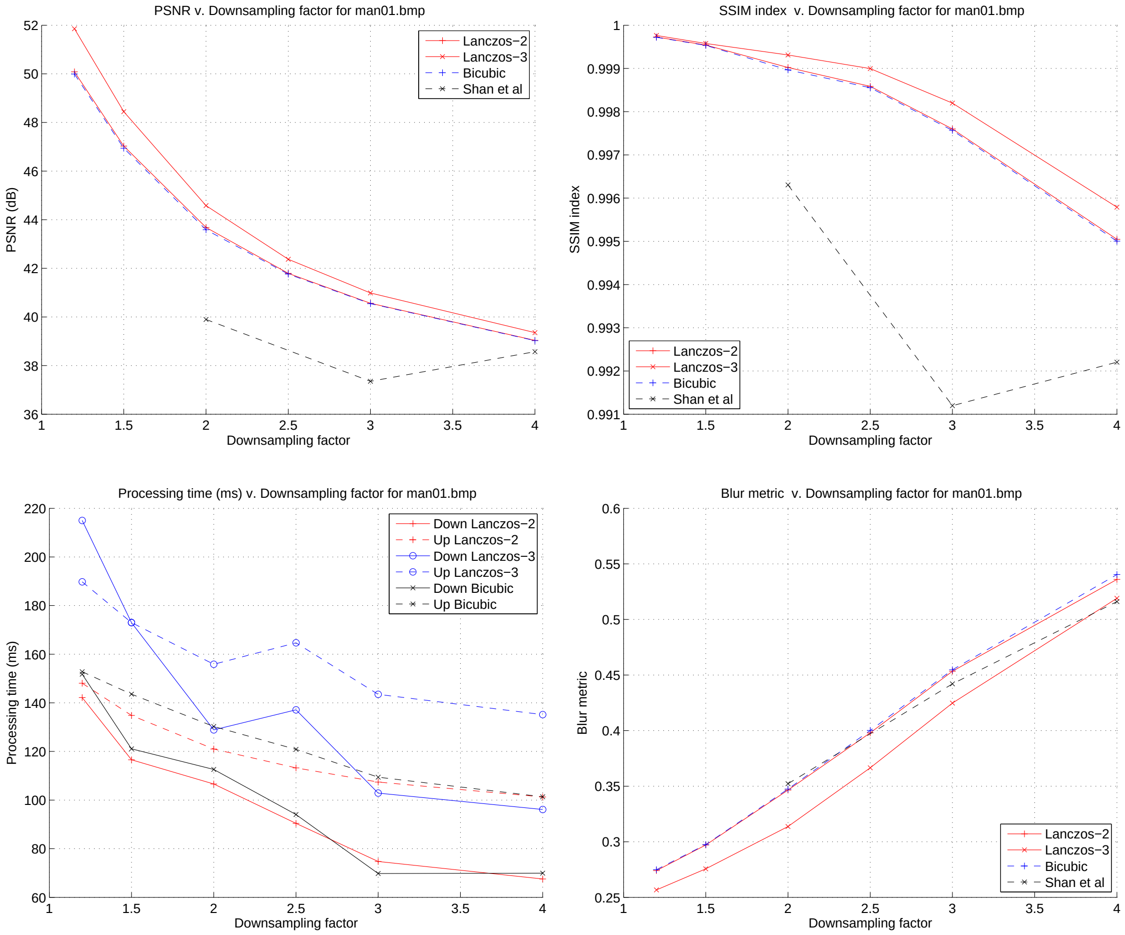

Figure 5: Image quality in PSNR (top left) and SSIM (top right), processing times (bottom left) and blur metrics (bottom right) for the test image at left in Figure 4.

time, particularly for upsampling process. Surprisingly, the worst performance in image quality and computing time goes to the new technique. Despite the high-end computing power for this experiment, the technique fails to work for $DF < 2.0$. For $DF \geq 2.0$ the technique operates within several minutes and it clarifies the absence of the technique in the diagram on computing time. At $DF = 3.0$ the output image quality of the technique behaves strangely and we are yet unable to explain the phenomena. Nevertheless it is clear from the results that Lanczos-2 and bicubic techniques perform the best.

There might be a question: what if the image is already captured by the camera at the downsampled resolution? If the output is then upsampled by the resampling techniques, what is the effect in terms of the four criteria? These are interesting questions because if the camera can do that, no downsampling is necessary after image acquisition which thus saves more time. We attempted to examine this by comparing the images sequentially captured in different resolutions at the same aspect ratio by a high-end camera with a remote control. Despite the use of remote control in the image acquisitions, shifts of pixels always occur in those images which cause considerable reduction in image quality when comparing an image at a low resolution and that downsampled to that resolution from a higher one. However those images look almost the same from subjective point of view. This study is saved for further work as it requires a more advanced device that can capture a scene and save it into images in different resolutions with the same aspect

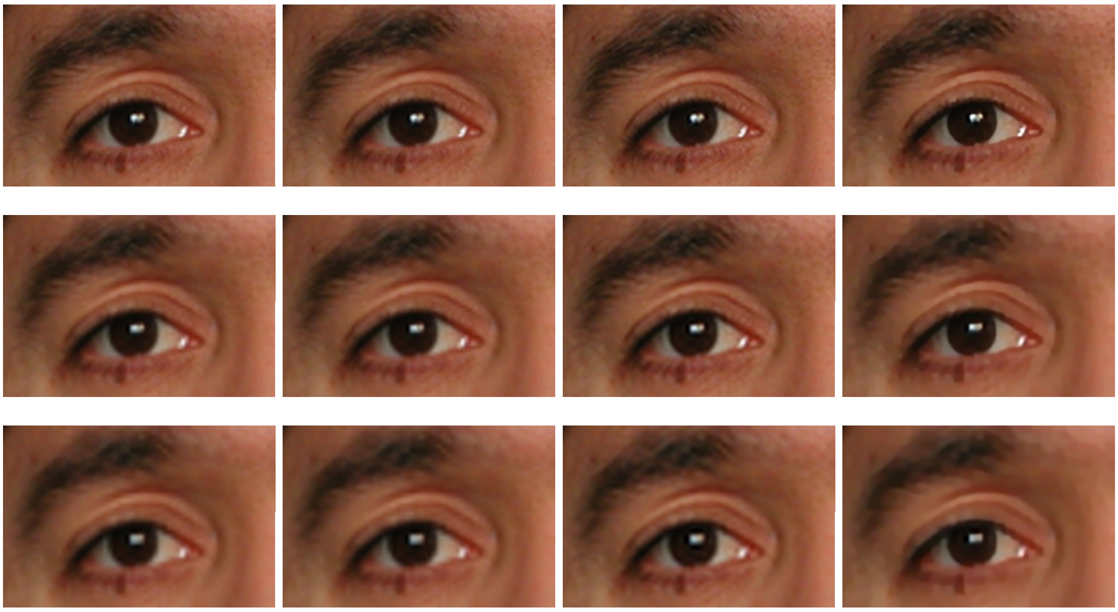

Figure 6: Sample images from the test image at left in Figure 4. The top, middle and bottom rows respectively refer to *DF* equals 2.0, 3.0, and 4.0. The columns from left to right refer to bicubic, Lanczos-2, Lanczos-3 and the new techniques, respectively. Images are to be seen on screen for best perception quality.

ratio in one go. It is aligned to our research since *dynamic* cameras in the collaboration space imply that their parameters, including image resolution and aspect ratio, can be fully controlled on the fly by the DMP system.

Experiment B: The Effects of Blur and Edges to Resampling

Artifacts caused by image upsampling have been categorized as ringing, aliasing, blocking, and blurring [8]. The latter is more apparent in sample images shown in Figure 6 as the *DF*s are increased. However, if we take the reverse direction, an interesting question arises: what will be the effect to the result after down/upsampling if the original image is more blurred? Experiment B attempts to provide a quantitative answer to the question.

We induce more blur into the test images from [11] by applying a Gaussian low-pass filter with standard deviation 10 and increasing size. Then we down/upsample the blurred test images using Lanczos-2 technique with all ten *DF*s in Table 2. As in Experiment A, the resulting image quality in PSNR and SSIM from the test images are typically similar, as shown in Figure 7 for the same test image. The legend in Figure 7 denotes the level of blur objectively quantified in the blurred test images where higher values indicates more blur. Figure 8 depicts sample images of the right eye area of the man as in Figure 6, cropped from the blurred test images.

Comparing the diagrams in Figure 7 and the sample images in Figure 8 shows some interesting insights. First, the more blurry the original image, the higher the *DF* for down/upsampling the image with unobjectionable quality of the output image. This is clear as images with more blur have lighter tails in SSIM than those of clearer ones. The initial values of PSNR and SSIM for $DF = 1.2$ are also higher for test images with more blur. Second, ringing artifact will appear when the *DF* is already too high and this occurs more likely to clearer images. This artifact is mainly caused by overshoot and oscillations

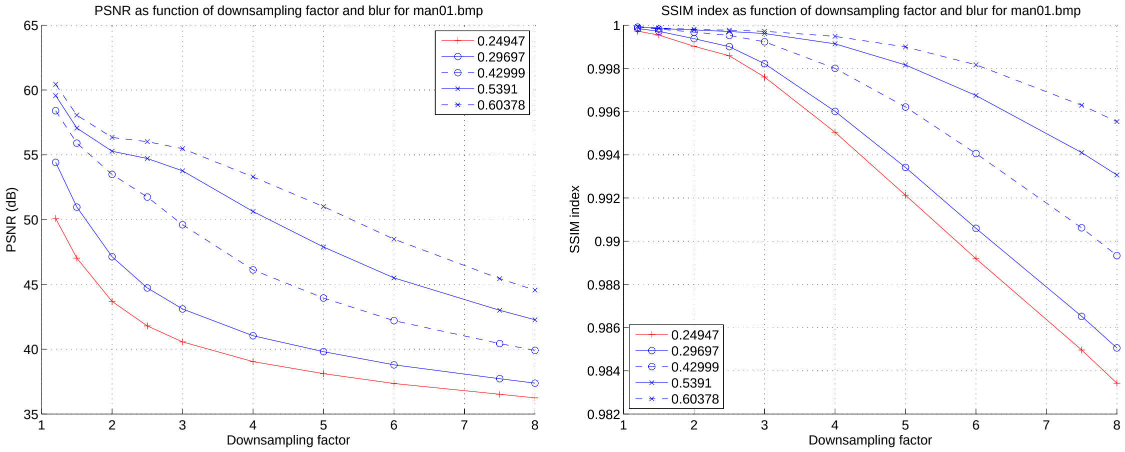

Figure 7: Image quality in PSNR (left) and SSIM (right) for the test image at left in Figure 4 using Lanczos-2 technique.

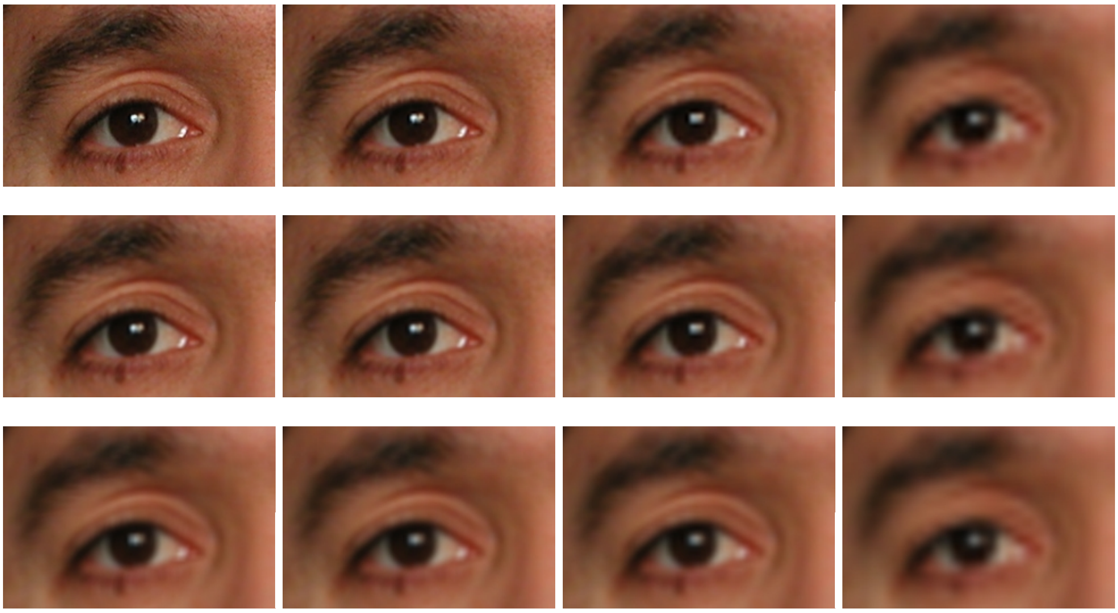

Figure 8: Sample images from the test image at left in Figure 4. The top, middle and bottom rows respectively refer to blur index 0.24947, 0.42999, and 0.60378. The left column presents the original test images with initial blur, while those next to the right are the output images with DF equals 2.0, 4.0, and 8.0, respectively. Images are to be seen on screen for best perception quality.

by the lobes of the Lanczos-2 filter. The presence of ringing artifact can be potentially exploited to indicate the limit of acceptable DFs. Third, the PSNR values for two most blurred images at $DF > 2.0$ are very interesting for further analysis.

It is common that an image has regions of different levels of blur, for example the background is much more blurry than the clear focused object in the foreground. The above results lead us to another interesting question: how will the result look like when such different regions in an image are down/upsampled with different DFs as a function of their level of blur? We attempt to answer it by exercising the idea to natural HD images of selected scenes captured from [14].

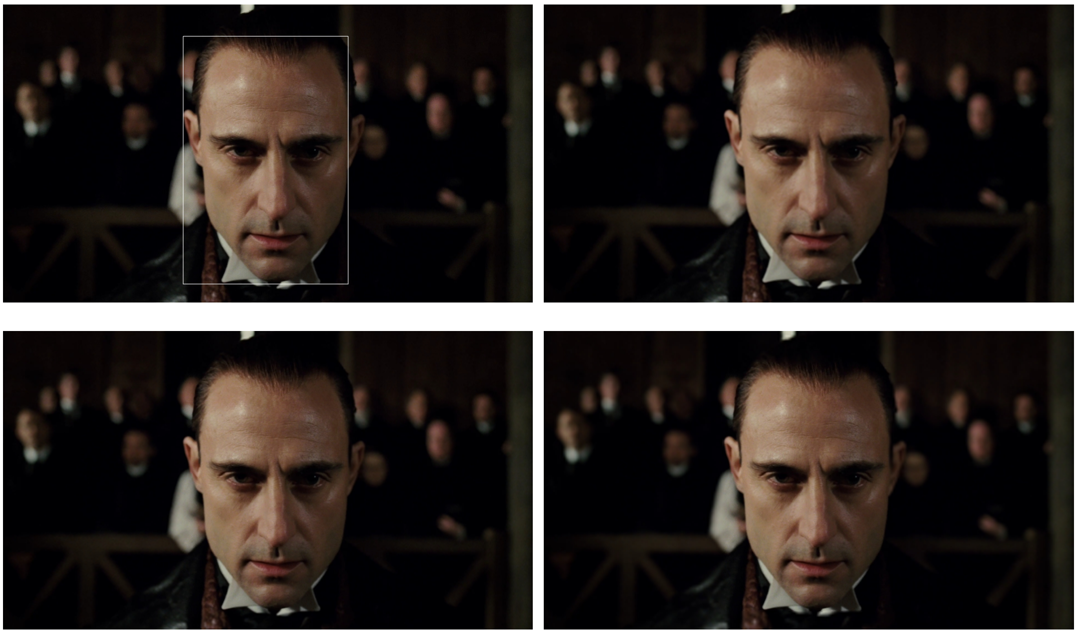

Figure 9: Original image (top left), overall image with $DF = 4.0$ (top right), composite image which ROI and overall images are down/upsampled with DF equals 2.0 and 4.0, respectively (bottom left), and that with DF equals 2.0 and 8.0, respectively (bottom right). The ROI is 26% of the image. Images are to be seen on screen for best quality.

Table 3: Objective quality of the sample images in Figure 9 (PSNR in dB).

	ROI			Overall			Composite	
Image	DF_{ROI}	PSNR	SSIM	DF_I	PSNR	SSIM	PSNR	SSIM
2	4	41.3662	0.97507	4	45.3855	0.99791	45.3528	0.99782
3	2	44.8048	0.99563	4	45.3855	0.99791	47.2685	0.99917
4	2	44.8048	0.99563	8	41.6313	0.9906	45.5645	0.99765

Let us assume a test image with a region of interest (ROI) defined in a bounding box. First, the test image and the ROI are down/upsampled with downsampling factors DF_I and DF_{ROI}, respectively, where $DF_{ROI} \leq DF_I$. The image resulting from down/upsampling the test image with DF_I is denoted as *overall* image while the *ROI* image refers to the resulting ROI which quality is compared with the luminance values of the ROI in the test image. Finally, the resulting ROI is patched to the same location in the *overall* image and the eventual output image is called a *composite* image.

Figure 9 shows an original HD image and three output images from three configurations of DFs using Lanczos-2 technique as detailed in the figure caption. In the original test image, the background is very blurry without strong edges and the face in the foreground is very clear and rich in details. The objective quality of the different regions in PSNR and SSIM from Figure 9 is presented in Table 3. Another test image which background region is also blurry but with many strong edges is examined with the same configurations and the results are shown in Figure 10 and Table 4.

Images (3) in Tables 3 and 4 clearly have the best quality relative to the others of the same scene. The improvement in the quality of the composite images with respect to

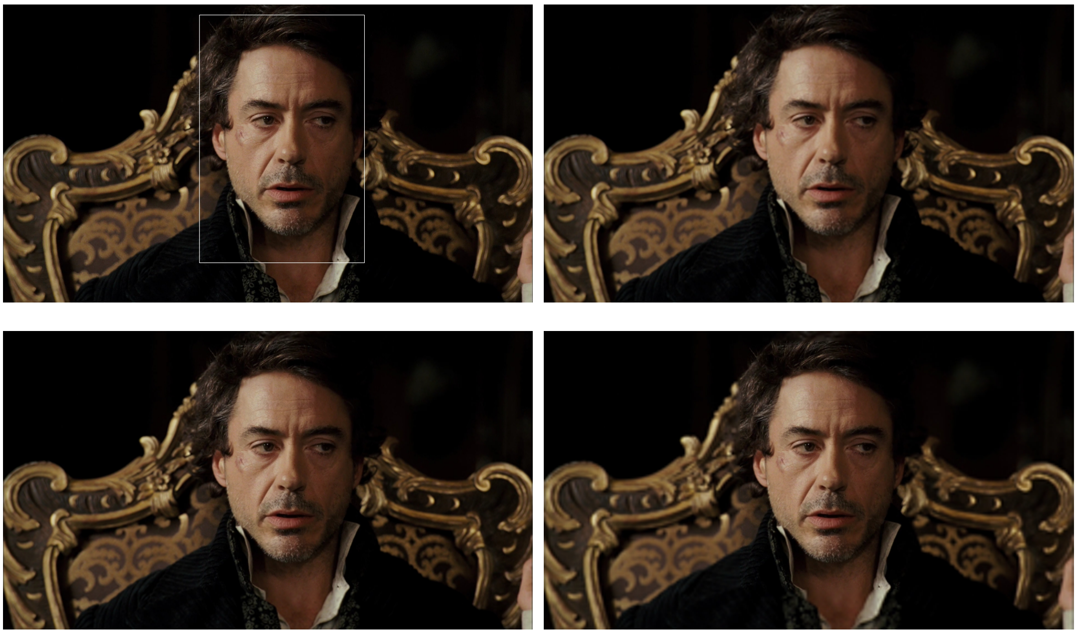

Figure 10: Original image (top left), overall image with $DF = 4.0$ (top right), composite image which ROI and overall images are down/upsampled with DF equals 2.0 and 4.0, respectively (bottom left), and that with DF equals 2.0 and 8.0, respectively (bottom right). The ROI is 26% of the image. Images are to be seen on screen for best quality.

Table 4: Objective quality of the sample images in Figure 10 (PSNR in dB).

	ROI			Overall			Composite	
Image	DF_{ROI}	PSNR	SSIM	DF_I	PSNR	SSIM	PSNR	SSIM
2	4	37.3468	0.95513	4	40.0571	0.99584	40.0403	0.99585
3	2	41.8998	0.99405	4	40.0571	0.99584	41.6801	0.9982
4	2	41.8998	0.99405	8	35.6731	0.97292	37.468	0.98599

that of the corresponding overall images is contributed by that of the ROI. Considerable 2-4db and 1.5-2dB increase in PSNR can be achieved by composite images (3) and (4) in Figures 9 and 10, respectively. Although composite image (4) in Figure 9 and Table 3 still has acceptable quality shown by very high PSNR and SSIM values, the quality of composite image (4) in Figure 9 and Table 3 is the worst for that scene. This shows the effect of edges in an image to how far an image can be down/upsampled.

As obvious in composite image (4) in Figure 9, unacceptable DF causes not only more blurring but also ringing artifacts, particularly in the areas where the edges are strong, as depicted in Figure 11 and also by the images in the fourth column of Figure 8. Nevertheless, the quality of composite image (3) in Figure 10 is still acceptable, as shown by the values of PSNR and SSIM. These results lead to a safe recommendation that for an image with clear region as ROI and blurry region, the first can be down/upsampled with $DF = 2.0$ and the latter with $DF = 4.0$, although strong edges are present in the blurry region. The level of blur must certainly be quantified beforehand. Composite images are expected to perform better in quality and bit rate than overall images. If the density of the edges in the blurry region can be quantified such as in [15], more bits can be saved by applying $DF > 4$ to the blurry area.

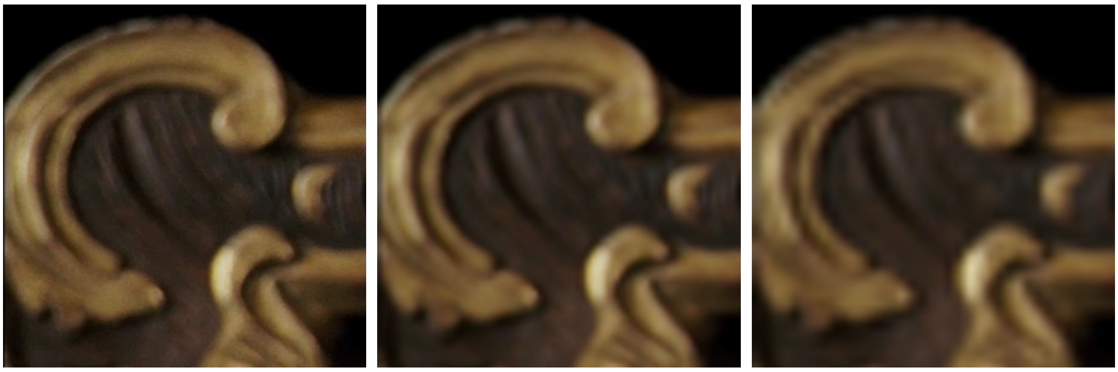

Figure 11: Images extracted from Figure 9: original (left), with $DF = 4.0$ causing blur (middle), and with $DF = 8.0$ causing more blur and ringing artifact (right). Images are to be seen on screen for best perception quality.

4 Conclusion

We have presented our examination on four resampling techniques for down/upsampling HD images. Our experimental results show that the performances of each technique are relatively constant not only in the tested images showing human faces but also in other natural images. Lanczos-2 and bicubic techniques comparatively perform the best in terms of computing time, objective image quality and the level of introduced blur. Computing time is very critical in our comparison since this study is driven by our vision of future musical collaboration with maximum end-to-end delay of 11.5ms to guarantee good synchronization between collaborating musicians. Lanczos-2 and bicubic techniques are based on convolutions that are suitable for parallel implementation in hardware, for example using systolic approach in field-programmable gate arrays (FPGAs) [16, 17], making them good candidates for next pursuit in our research.

The results from our second experiment reveal that blur and edges are important factors when down/upsampling HD images. Higher downsampling factors (DFs) can be applied to an image or a region in an image that is more blurry than the others with unobjectionable quality of the output image. Ringing artifact in the result is potential to be exploited as an indicator when a used DF is already too high. If an image has major clear and blurry regions, usually as foreground and background, respectively, the latter can be down/upsampled with higher DF than that applied in down/upsampling the first. Combining the results from the two down/upsampling processes yields a composite image that is expected to give a better image quality with less bit rate than that from applying a DF to down/upsample the entire image. Our experimental results point that the improvement of image quality in PSNR can reach up to 4dB in composite images. This increase is attributed to the clear region down/upsampled with a lower DF.

As edges limit the (DFs) in down/upsampling due to unwanted ringing artifact, our safe recommendation is to down/upsample the clear and blurry regions with $DF = 2.0$ and $DF = 4.0$, respectively, assuming strong edges present in the blurry region. This should be preceded by applying available metrics to quantify the level of blur and the edge density in the blurry region. If the region is very blurry with few weak edges detected, it can be down/upsampled with higher DFs, even up to 8.0 without making the resulting quality reduction very apparent. However if the down/upsampling is applied to the whole HD image without ROI, using $DF = 2.0$ is a safe recommendation.

References

[1] ITU F.702, Multimedia Conference Services.

[2] C. Chafe, M. Gurevich, G. Leslie and S. Tyan. .Effect of time delay on ensemble accuracy. In Proc. International Symposium on Musical Acoustics. 2004.

[3] L. A. Rønningen. The DMP system and physical architecture. Technical report, Department of Telematics, Norwegian University of Science and Technology, 2007.

[4] L. A. Rønningen. Input Traffic Shaping. In Proc. International Teletraffic Congress, Montreal, 1982.

[5] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600-612, 2004. Available online at http://www.ece.uwaterloo.ca/˜z70wang/research/ssim/.

[6] R. Keys. Cubic convolution interpolation for digital image processing. *IEEE Transactions on Signal Processing, Acoustics, Speech, and Signal Processing*, 29(6):1153-1160, 1981.

[7] Q. Shan, Z. Li, J. Jia, and C-K. Tang. Fast image/video upsampling. *ACM Transactions on Graphics (SIGGRAPH ASIA)*, 2008. Available online with the implementations at http://www.cse.cuhk.edu.hk/ ˜leojia/projects/upsampling/.

[8] P. Thevenaz, T. Blu, and M. Unser. Image interpolation and resampling. in *Handbook of medical imaging*, pp. 393-420, Academic Press, 2000.

[9] K. Turkowski and S. Gabriel. Filters for common resampling tasks. in A. S. Glassner. *Graphics Gems I*, pp. 147-165, Academic Press, 1990.

[10] R. Fattal. Image upsampling via imposed edges statistics. *ACM Transactions on Graphics (SIGGRAPH)*, 26(3), 2007.

[11] Face recognition database. Center for Biological and Computational Learning, MIT, 2005. Available online at http://cbcl.mit.edu/software-datasets/heisele/facerecognition-database.html.

[12] F. Crete, T. Dolmiere, P. Ladret, and M. Nicolas. The blur effect: perception and estimation with a new no-reference perceptual blur metric. In Proc. SPIE Human Vision and Electronic Imaging, 2007.

[13] Q. B. Do. Matlab implementation of image blur metric [12]. Available online at http://www.mathworks.com/matlabcentral/fileexchange/24676-image-blur-metric.

[14] G. Ritchie. Sherlock Holmes. Trailer in 1080p resolution available online at http://www.digital-digest.com/movies/Sherlock_Holmes_1080p_Theatrical_Trailer.html.

[15] S. L. Phung and A. Bouzerdoum. Detecting People in Images: An Edge Density Approach. In Proc. IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP), 2007.

[16] H. T. Kung. Why systolic architectures. *Computer Magazine*, vol. 15, pp. 37-45, 1982.

[17] M. A. Nuno-Maganda and M. O. Arias-Estrada. Real-time FPGA-based architecture for bicubic interpolation: an application for digital image scaling. In Proc. International Conference on Reconfigurable Computing and FPGAs, 2005.