ADAPTIVE DOWNSAMPLING FOR HIGH-DEFINITION VIDEO CODING

Jie Dong and Yan Ye

InterDigital Communications, San Diego, CA

ABSTRACT

Previous research has shown that downsampling prior to encoding and upsampling after decoding can improve the rate-distortion (R-D) performance compared with directly coding the original video using standard coding technologies, e.g., JPEG and H.264/AVC, especially at low bit rates. This paper proposes a practical algorithm to find the optimal downsampling ratio that balances the distortions caused by downsampling and coding, thus achieving the overall optimal R-D performance. Given the optimal sampling ratio, dedicated filters for down- and up-sampling are also designed. Simulations show this algorithm improves the R-D performance over a wide range of bit rates, e.g., from 1.0 dB at high bit rates to 2.5 dB at low bit rates.

1. INTRODUCTION

Previous research has shown that downsampling prior to encoding and upsampling after decoding (see Fig. 1(a)) can improve the rate-distortion (R-D) performance compared with using standard coding technology, e.g., JPEG or H.264/AVC, directly on the original input, especially at low bit rates [1]-[6]. In [1] and [2], downsampling and upsampling were performed at the macroblock (MB) level. The so-called critical bit rate was studied, below which an MB is downsampled in the horizontal, vertical, or both directions with a fixed ratio of 2:1 before being coded, and upsampled to the original size after reconstruction. However, performing such operations at the MB level no longer conforms to the coding standards. To maintain conformance to the coding standards, the downsampling and upsampling should be performed as pre- and post-processing, respectively. In [3], a theoretical model to determine the optimal downsampling ratio was studied, which analytically explained the advantage of downsampling an image prior to applying JPEG and upsampling the decoded image. The benefits of using this paradigm for video coding were reported in [4], where a set of downsampling ratios were tried over a wide range of bit rates. Using the optimal ratio at a given bit rate, it is shown that 30% and 50% bit rate reduction can be achieved for H.264/AVC and MPEG-2, respectively. Many other related literatures in this context focus on the sampling filter design. Following [3], Tsaig *et al*. proposed an image-dependent algorithm to find optimal filters for decimation and interpolation [5]. In [6], an interpolation-dependent image downsampling method was proposed, where, given a specific interpolator, the downsampler that minimizes the difference between the input image and the output image was designed.

This paper proposes a practical algorithm to find the optimal downsampling ratio for the system in Fig. 1(a), which can theoretically be any rational number represented by A/B ($A \geq B$). The proposed algorithm decomposes the overall distortion to two types: one caused by downsampling and the other by coding, and estimates them separately. Based on these estimates, the proposed algorithm then finds the optimal downsampling ratio that makes the best balance of the two distortions, thus achieving the best overall R-D performance. Given the downsampling ratio, the dedicated downsampling and upsampling filters are designed. Experimental results show that this algorithm efficiently improves the R-D performance over a wide range of bit rates, e.g., from 1.0 dB at high bit rates to 2.5 dB at low bit rates.

2. OPTIMAL RATIO FOR VIDEO DOWNSAMPLING

2.1. Problem statement

In Fig. 1(a), the video input, denoted as f, is downsampled with the ratio M to generate d_1; d_1 is coded and decoded to reconstruct d_2; d_2 is upsampled with the ratio M to generate the full-resolution output video f_1. The overall MSE between f and f_1 is called overall error, denoted as σ_A^2. The system in Fig. 1(a) can be decomposed to the sampling part (Fig. 1(b)) and the coding part (Fig. 1(c)). In the sampling part, for the input video f, upsampling with a ratio M is applied right after downsampling to generate f_2; that is, the MSE between f and f_2 is caused only by sampling and is called downsampling error, denoted as σ_D^2. In the coding part, the MSE between the input and output, denoted as d_1 and d_2, is caused only by coding and is denoted as the coding error σ_C^2.

Given an input video signal and a bit rate, the downsampling ratio M is optimal, when the overall error σ_A^2 reaches the minimum, as shown in (1).

$$M = \arg\min_M \sigma_A^2 \tag{1}$$

As σ_A^2 is jointly determined by σ_D^2 and σ_C^2, the relationship among σ_A^2, σ_D^2, and σ_C^2 is analyzed as below.

$$\begin{aligned}\sigma_A^2 &= E[(f-f_1)^2] \\ &= E[(f-f_2+f_2-f_1)^2] \\ &= E[(f-f_2)^2+(f_1-f_2)^2-2(f-f_2)(f_1-f_2)]\end{aligned} \tag{2}$$

As $(f-f_2)$ and (f_1-f_2) are uncorrelated, $E[(f-f_2)(f_1-f_2)]$ is equal to zero. The difference between f_1 and f_2, i.e., (f_1-f_2), is the upsampled version of (d_1-d_2), and has the same energy as (d_1-d_2), because the upsampling filter gain is equal to M. Then, (2) can be re-written as in (3).

$$\sigma_A^2 = E[(f-f_2)^2] + E[(d_1-d_2)^2] = \sigma_D^2 + \sigma_C^2 \tag{3}$$

(3) shows that the overall error is the summation of the downsampling error and the coding error, which is confirmed by our experimental results. Therefore, the optimization problem in (1) is re-written as in (4).

$$M = \arg\min_M(\sigma_D^2 + \sigma_C^2) \tag{4}$$

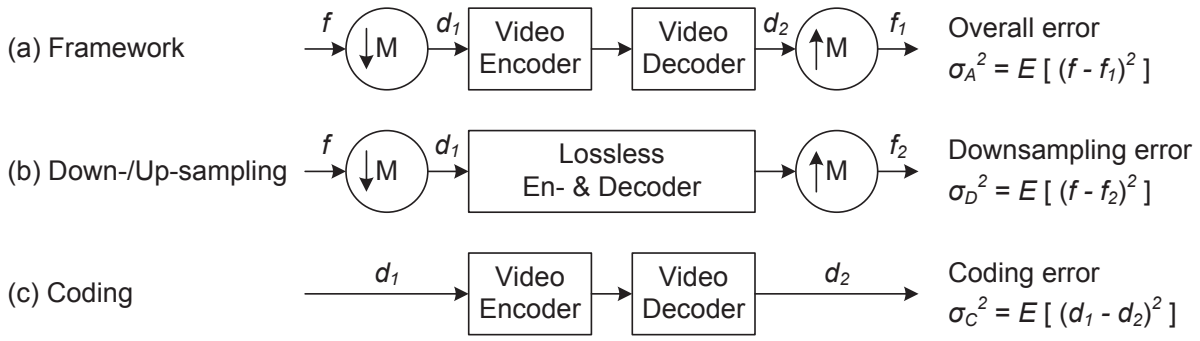


Fig. 1 (a) Coding system with down- and up-sampling is decomposed to (b) the sampling part and (c) the coding part.

2.2. Downsampling error estimation

The most straightforward and accurate way to estimate σ_D^2 is to follow the paradigm in Fig. 1(b) and to calculate σ_D^2 by definition. However, the complexity is too high. In this paper, σ_D^2 is measured in the frequency domain by the energy of the high frequency components that exist in f but are lost in f_2. The energy distribution in the frequency domain is modeled by its Power Spectral Density (PSD). PSD of f is estimated by periodgram, as in (5),

$$\hat{S}_{xx}(\omega_1,\omega_2)=\frac{1}{WH}\left|\sum_{w=0}^{W-1}\sum_{h=0}^{H-1}x[w,h]e^{-j\omega_1 w-j\omega_2 h}\right| \quad (5)$$

where W and H are the width and height of the sequence f, and $x[w,h]$ is one frame in f. When the entire sequence f consists of consistent content without scene change, $\hat{S}_{xx}(\omega_1,\omega_2)$ calculated based on one typical frame $x[w,h]$, e.g., the first frame, can well represent the energy distribution of the whole sequence f. When the sequence f contains scene changes, $\hat{S}_{xx}(\omega_1,\omega_2)$ can be calculated by averaging the PSDs of frames from different scenes.

Let the ratio M be represented by A/B ($A \geq B$), then the downsampled video has the resolution $\left(\frac{B}{A}\mathrm{W}\times\frac{B}{A}\mathrm{H}\right)$. In other words, the proportion of the reduced resolution in either direction is equal to (1-B/A). In the frequency domain, the lost components all have high frequency, of which the proportion is also (1-B/A), if the anti-aliasing filter applied to f has a sharp cut-off frequency at $\pm\frac{B}{A}\pi$. In this ideal case, all the high frequency components of f_2 in the bands $\left[-\pi,-\frac{B}{A}\pi\right]$ and $\left[\frac{B}{A}\pi,\pi\right]$ are lost. We can estimate the PSD of f_2, denoted as $\hat{S}_{yy}(\omega_1,\omega_2)$, from $\hat{S}_{xx}(\omega_1,\omega_2)$ by setting the values in the aforementioned bands to zero, as in (6).

$$\hat{S}_{yy}(\omega_1,\omega_2)=\begin{cases}\hat{S}_{xx}(\omega_1,\omega_2), & \text{if } \omega_1,\omega_2\in\left[-\frac{B}{A}\pi,\frac{B}{A}\pi\right]\\ 0, & \text{otherwise}\end{cases} \quad (6)$$

When the horizontal and vertical directions have different downsampling ratios, i.e., $M_h=A_h/B_h$ and $M_v=A_v/B_v$, $\hat{S}_{yy}(\omega_1,\omega_2)$ is estimated as in (7).

$$\hat{S}_{yy}(\omega_1,\omega_2)=\begin{cases}\hat{S}_{xx}(\omega_1,\omega_2), & \text{if } \omega_1\in\left[-\frac{B_h}{A_h}\pi,\frac{B_h}{A_h}\pi\right]\,\&\,\omega_2\in\left[-\frac{B_v}{A_v}\pi,\frac{B_v}{A_v}\pi\right]\\ 0, & \text{otherwise}\end{cases} \quad (7)$$

In practice, estimating $\hat{S}_{yy}(\omega_1,\omega_2)$ as in (6) and (7) is only a close approximation of the true PSD of f_2, because it is impossible for the anti-aliasing filter to have an ideal sharp cut-off frequency.

After estimating the PSD of f in (5) and f_2 in (6) or (7), one can calculate the downsampling error σ_D^2 by (8).

$$\sigma_D^2=\frac{1}{WH}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left[\hat{S}_{xx}(\omega_1,\omega_2)-\hat{S}_{yy}(\omega_1,\omega_2)\right]d\omega_1 d\omega_2 \quad (8)$$

2.3. Coding error estimation

The coding error σ_C^2 can be estimated by an empirical R-D model, as shown in (9),

$$\sigma_C^2=\frac{\beta}{r^\alpha} \quad (9)$$

where r is the average number of bits allocated to each pixel, i.e., bits per pixel (bpp), and can be calculated as in (10).

$$r=\frac{R\times M_h\times M_v}{F\times W\times H} \quad (10)$$

In (10), F is the frame rate, and R is the target bit rate for the video sequence. This model has two parameters, α and β, whose values vary according to the content, the resolution of the sequence, the encoder implementation and configurations, and so on. Given a certain encoder configured with certain settings, one may encode a given sequence at a range of bit rates $\{R_0, R_1, \ldots, R_{N-1}\}$, and get a corresponding set of distortions $\{d_0, d_1, \ldots, d_{N-1}\}$ measured by MSE. The target bit rates $\{R_0, R_1, \ldots, R_{N-1}\}$ are then normalized to bpp $\{r_0, r_1, \ldots, r_{N-1}\}$ using (11),

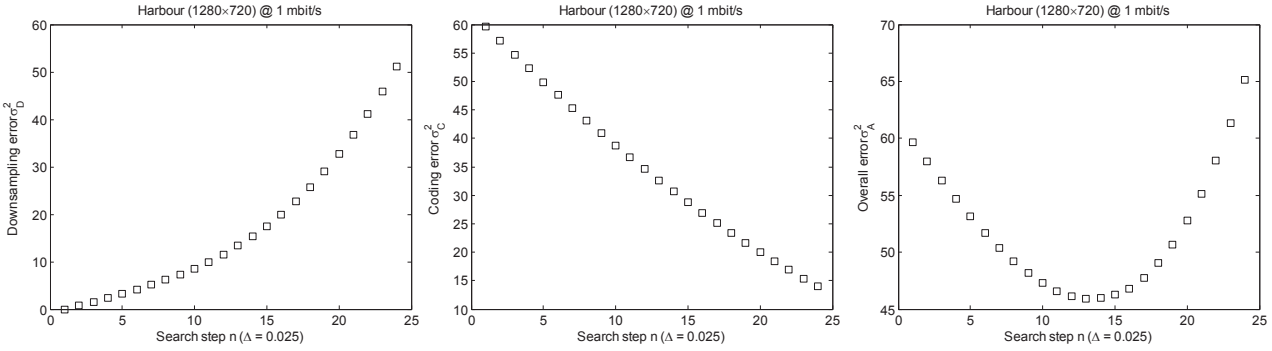
$$r_i=\frac{R_i}{F\times W\times H} \quad (11)$$

The pairs of normalized bit rate and distortion $[r_i, d_i]$ ($0\leq i<N$) are plotted as an R-D curve; any commonly used numerical optimization algorithm can be used to fit the R-D curve. The optimal values of α and β are found by solving the problem in (12).

$$[\alpha_{opt},\beta_{opt}]=\arg\min_{\alpha,\beta}\sum_{i=0}^{N-1}\left(d_i-\frac{\beta}{r_i^\alpha}\right)^2 \quad (12)$$

2.4. Optimal ratio determination

Given the target bit rate R, bpp increases with respect to the downsampling ratio M (see (10)), and therefore coding error σ_C^2 monotonically decreases with M (see (9)), whereas downsampling error σ_D^2 monotonically increases with M, as shown in Fig. 2 (a) and (b). The optimal ratio balances the two types of distortions, such that the overall error reaches the minimum. Fig. 2 shows an example of searching for the



(a) (b) (c)

Fig. 2 Example of serching the optimal ratio for *Harbour* at 1 mbit/s. (a) Downsampling error; (b) coding error; (c) overall error.

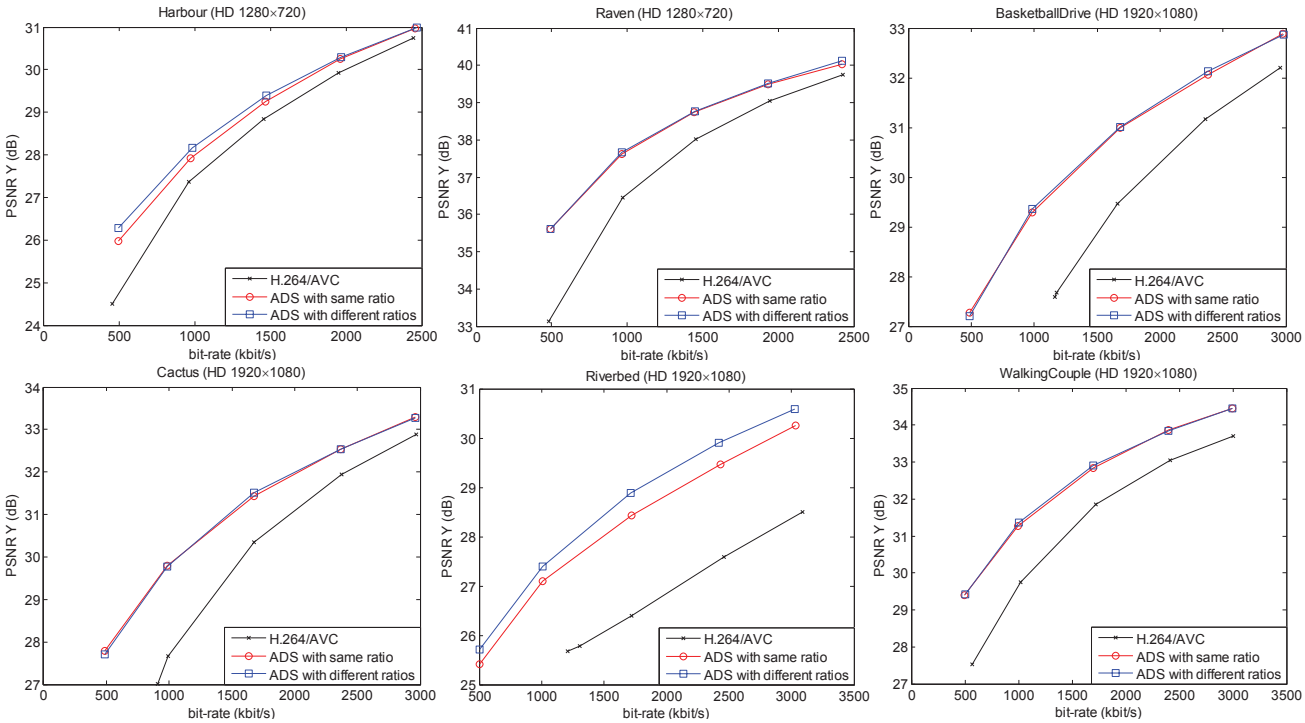


Fig. 3 Operational R-D curves provided by H.264/AVC, ADS with same ratio, and ADS with different ratios.

optimal ratio for the 720*p* sequence *Harbour*, given the target bit rate of 1 mbit/s. In search step n (n≥1), the downsampling ratio M is equal to $1+(n-1)\Delta$, where Δ is an increment of the ratio in each step, and σ_D^2 and σ_C^2 are obtained using the algorithms introduced in Sections 2.2 and 2.3, respectively. In the first a few steps, i.e., n is small, σ_D^2 increases slowly, and is well compensated by better reconstruction in the coding part. Therefore, σ_A^2 decreases. As M becomes large, the increase of σ_D^2 outweighs the decrease of σ_C^2, and σ_A^2 goes up with each search step. The optimal ratio M is determined by finding the global minimum of the overall error σ_A^2 in Fig 2 (c). For videos with different energy distributions in horizontal and vertical directions, using different downsampling ratios to apply heavier downsampling along the direction with less high frequency energy further improves the performance. In this case, the plots in Fig. 2 form convex surfaces, of which the optimal (M_h, M_v) makes σ_A^2 reaches the global minimum.

3. DOWN- AND UP-SAMPLING FILTER DESIGN

Once the optimal ratio is determined, the dedicated down- and up-sampling filters are designed on-the-fly. The filters are 2-D separable, and without the loss of generality, only the derivation of horizontal filters is explained here.

Suppose the horizontal sampling ratio is A_h/B_h. In the downsampling step, the rows of the original frames are upsampled to B_h times the width by zero-insertion, filtered by the anti-aliasing filter $f_{d,h}$ in (13) with cut-off frequency $\pm\frac{\pi}{A_h}$ and filter gain B_h, and then decimated by a factor of A_h.

$$f_{d,h}(n) = \frac{1}{2\pi}\int_{-\frac{\pi}{A_h}}^{\frac{\pi}{A_h}} B_h e^{jn\omega}\, d\omega = \frac{B_h}{A_h}\mathrm{Sinc}(\frac{\pi}{A_h}n) \qquad (13)$$

In the upsampling step, the rows of the downsampled frames are upsampled to A_h times the downsampled frame width by zero-insertion, filtered by the anti-aliasing filter $f_{u,h}$ in (14)

Table 1 Optimal ratios obtained by the proposed algorithm

HD Sequences	Bit Rate (mbit/s)	Same Ratio for Hor. & Ver.	Different Ratio for Hor.	Different Ratio for Ver.
Harbour (1280×720)	0.5	40/20	40/22	40/13
	1.0	40/21	40/23	40/15
	1.5	40/22	40/24	40/17
	2.0	40/23	40/24	40/18
	2.5	40/23	40/24	40/18
Raven (1280×720)	0.5	40/18	40/18	40/18
	1.0	40/21	40/20	40/22
	1.5	40/22	40/21	40/24
	2.0	40/23	40/21	40/26
	2.5	40/24	40/22	40/28
Cactus (1920×1080)	0.5	60/24	60/29	60/21
	1.0	60/30	60/33	60/26
	1.7	60/34	60/36	60/31
	2.4	60/37	60/38	60/34
	3.0	60/39	60/39	60/38
BasketballDrive (1920×1080)	0.5	60/23	60/22	60/26
	1.0	60/28	60/26	60/31
	1.7	60/32	60/30	60/35
	2.4	60/36	60/32	60/41
	3.0	60/40	60/36	60/42
Riverbed (1920×1080)	0.5	60/16	60/8	60/24
	1.0	60/17	60/9	60/26
	1.7	60/22	60/11	60/28
	2.4	60/24	60/13	60/30
	3.0	60/24	60/13	60/39
WalkingCouple (1920×1080)	0.5	60/22	60/18	60/25
	1.0	60/25	60/21	60/29
	1.7	60/28	60/23	60/33
	2.4	60/30	60/24	60/36
	3.0	60/31	60/25	60/38

with cut-off frequency $\pm(\pi/A_h)$ and filter gain A_h, and decimated by a factor of B_h.

$$f_{u,h}(n) = \frac{1}{2\pi}\int_{-\frac{\pi}{A_h}}^{\frac{\pi}{A_h}} A_h e^{jn\omega}\, d\omega = \mathrm{Sinc}(\frac{\pi}{A_h}n) \qquad (14)$$

The filters in (13) and (14) have infinite sizes and need to be truncated by appropriate window functions before being used for interpolation. Here, Gaussian window is used, which is empirically better than other window functions, such as rectangular, triangular, and Hanning windows.

4. EXPERIMENTAL RESULTS

Our simulations are based on 6 HD sequences, each coded at 5 target bit rates with 3 methods: 1) H.264/AVC, 2) the proposed Adaptive Down-Sampling (ADS) algorithm with the same horizontal and vertical ratio, and 3) ADS with different horizontal and vertical ratios. The test sequences are selected to have diverse content, including rapid and irregular motions (*Cactus*, *BasketballDrive*, and *Riverbed*), sharp edges and rich textures (*Harbour* and *WalkingCouple*), and smooth areas (*Raven*). For the video encoder, we used x264 implementation of H.264/AVC [7] with the default settings. Table 1 summarizes the test sequences, the target bit rates, and the optimal downsampling ratios found by the proposed ADS methods. Fig. 3 shows the operational R-D curves provided by the three methods. The PSNR values for the two ADS methods (blue and red curves) are calculated by comparing the upsampled decoded video to the original one.

For all the sequences, the ADS schemes significantly outperform H.264/AVC (black curve), not only at low bit rate as shown by the previous research, but also at relatively high bit rates. The gains range from 1.0 to 2.5 dB. Though not shown here due to space limitation, the visual quality is also improved significantly. Interested readers are referred to http://jdong.avsx.org/icip2012.zip for details. Based on Table 1, the optimal ratio depends on bit rate and the video content. The proposed algorithm selects larger downsampling ratio at lower bit rates than it does at higher bit rates. That is because given the increment of bpp σ_C^2 is better suppressed at lower bit rates, i.e., the slope in Fig. 2(b) is steeper. Adding constant σ_D^2 to σ_C^2, one needs more steps to reach the minimum, meaning larger ratio.

Compared with ADS with the same horizontal and vertical ratio, ADS with different ratios jointly optimizes the ratios for two directions, and therefore can further improve the R-D performance. For some cases, the improvement is negligible, as the horizontal and vertical ratios are close to each other, due to similar energy distributions in the two directions in that given sequence. For other cases (*Harbour* and *Riverbed*), the improvements are more remarkable. For example, in *Harbour*, vertical edges (masts) dominate the content. Since the vertical direction is smoother than the horizontal direction, a bigger downsampling ratio can be applied in the vertical direction without causing much information loss. In contrast, *Riverbed* is dominated by horizontal textures (waves) moving vertically; a bigger downsampling ratio can be applied to the horizontal direction.

5. CONCLUSION

This paper proposes an ADS algorithm. For given sequence and target bit rate, it finds the optimal sampling ratio that minimizes the overall distortion by balancing the downsampling distortion and the coding distortion, thereby achieving the best R-D performance for the overall system. Simulations show the proposed ADS algorithm improves the coding efficiency over a wide range of bit rates, i.e., from 1.0 dB at high bit rates to 2.5 dB at low bit rates.

6. REFERENCES

[1] W. Lin and L. Dong, "Adaptive downsampling to improve image compression at low bit rates," *IEEE Trans. Image Process.*, vol. 15, no. 9, pp. 2513-2521, Sept. 2006.

[2] V.-A. Nguyen, Y.-P. Tan, and W. Lin, "Adaptive downsampling/upsampling for better video compression at low bit rate," *IEEE Int'l Symp. Circuits Syst. 2008*, Seattle, USA, May 2008.

[3] M. Bruckstein, M. Elad, and R. Kimmel, "Down-scaling for better transform compression," *IEEE Trans. Image Process.*, vol. 12, no. 9, pp. 1132-1144, Sept. 2003.

[4] A. Segall, M. Elad, P. Milanfar, R. Webb, C. Fogg, "Improved high-definition video by encoding at an intermediate resolution," *Visual Comm. Image Process. 2004*, San Jose, USA, Jan. 2004.

[5] Y. Tsaig, M. Elad, and P. Milanfar, "Variable projection for near-optimal filtering in low bit-rate block coders," *IEEE Trans. Circuits Syst. Video Technol.*, vol.15, no.1, pp.154-160, Jan. 2005.

[6] Y. Zhang, D. Zhao, J. Zhang, R. Xiong, and W. Gao, "Interpolation-dependent image downsampling," *IEEE Trans. Image Process.*, vol. 20, no. 11, pp. 3291-3296, Nov. 2011.

[7] http://www.videolan.org/developers/x264.html