

デジタルシネマを想定した解像度変換システム

石川 孝明[†] 渡辺 裕[†]

[†] 早稲田大学 大学院 国際情報通信研究科
〒 367-0035 埼玉県本庄市西富田大久保山 1011

E-mail: †takaxp@fuji.waseda.jp

あらまし JPEG 2000 の空間スケーラビリティを用いた、デジタルシネマの映像解像度変換システムを提案する。従来手法は、低域画像とデシメーションフィルタを組み合わせることで有理数比の解像度変換を実現しているが、デジタルシネマの空間解像度をソースとした場合や、現行の表示デバイスに焦点を絞った検討が不十分であった。そこで本稿では、デジタルシネマの空間解像度から、HDTV、NTSC、PAL などの空間解像度で表示するための解像度変換手法の検討を行い、従来手法と同等の処理量でより高品質の映像解像度変換を行うシステムを提案する。シミュレーションによりその有効性を示した。

キーワード デジタルシネマ, Motion JPEG 2000, 空間スケーラビリティ, 解像度変換, ローパスフィルタ

An Image Resolution Conversion System for Digital Cinema

Takaaki ISHIKAWA[†] and Hiroshi WATANABE[†]

[†] Graduate School of Electronics, Information and Communication Engineering Studies, Waseda
University
1011 Okuboyama, Nishi-Tomida, Honjo, Saitama, 367-0035 Japan

E-mail: †takaxp@fuji.waseda.jp

Abstract We propose a resolution conversion system, which can generate sequences of different resolutions, such as HDTV, SDTV, CIF, QVGA and QCIF, from Digital Cinema-size video sequences. The simulation results show an improvement in PSNR of about one or two dB as compared to traditional resolution conversion system, without extra calculation cost. It is implemented by combining a decimation filter with the LL sub-band of the source image encoded by JPEG 2000. We address the problem of pixel aspect ratio conversion, and also propose a method to combine LL sub-band with a decimation filter.

Key words Digital Cinema, Motion JPEG 2000, Spatial Scalability, Resolution Conversion, Low-pass filter

1. まえがき

近年、映像符号化のスケラビリティに注目が集められており、その具体的なアプリケーションとしてデジタルシネマがある。デジタルシネマの配信用フォーマットは、DCI (Digital Cinema Initiatives) と SMPTE (Society of Motion Picture and Television Engineers) および ISO/IEC JTC 1/SC 29/WG1 (JPEG) の協力により標準化が進められている。符号化方式としては、符号化データの空間スケラビリティを持つ JPEG 2000 の利用が内定している。空間スケラビリティを利用することで、異なる空間解像度のプロジェクタを所有する劇場に配信するビットストリームを容易に生成することが可能となる。

しかし、デジタルシネマを頂点とした“ワンソース・マルチユース”を実現し、現状の様々な表示デバイスに向けた映像を作成するためには、技術的な課題が残されている。特に、デコード処理のみでは現行のテレビや様々な表示デバイスに対する映像ソースを取得することが困難であることが指摘されている [1] [2]。

そこで本研究では、今後の普及が期待されるデジタルシネマをターゲットとした、空間スケラビリティを利用した映像解像度変換システムを提案する。提案システムは、Motion JPEG 2000 の空間スケラビリティで取得可能な低域画像とデシメーションフィルタを組み合わせる映像解像度変換手法により、デジタルシネマの最大空間解像度である $4096 \times 2160[\text{pel}]$ から、HDTV や SDTV などの複数のデバイスを対象とした、アスペクト比の違いを反映した解像度変換を行う。また、デシメーションフィルタの演算処理を定式化し、それを元にコスト関数を定義することで、従来手法と比較して同等かそれ以下の処理量でより高品質な映像を取得することが可能であることを示し、シミュレーションを通して提案システムの有効性を確認する。

2. 従来手法

画像の縮小処理において、エイリアシングの有無が出力画像の品質を大きく左右する。そのため、マルチレート信号処理に基づく縮小処理においては、図 1 のようにローパスフィルタとダウンサンプラを組み合わせたデシメーションフィルタが利用される。

2.1 空間スケラビリティを用いた解像度変換法

JPEG 2000 の空間スケラビリティを利用した効率的な解像度変換法として、文献 [2] の手法がある。図 2 に示した従来システムは、まず、Motion JPEG 2000 のビットストリームから、空間スケラビリティ

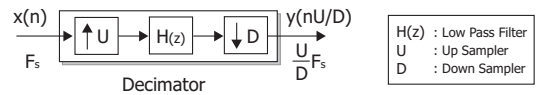


図 1 A typical structure of decimator (U/D)

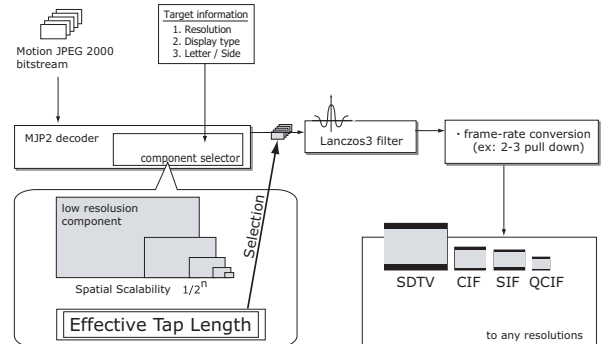


図 2 Traditional multi-decoding system

を利用して縮小画像を得る。JPEG 2000 で用いられている空間スケラビリティは、Wavelet 変換による周波数帯域のオクターブ分割により与えられるため、取得可能な縮小画像は $1/2^n$ に制限される。そこで、各低域成分に対しデシメーションフィルタを用いて画素を間引くことで、有理数比の縮小率を実現し、任意の解像度の映像を生成する事ができる。

従来手法では、実際のシステムフローを想定したピクセルアスペクト比の違いが考慮されていないため、SDTV 用コンテンツや CIF サイズの携帯向けコンテンツを生成するための変換式を定義する必要がある。

2.2 3-Lobed Lanczos Window Sinc Function

従来システムでは、デシメーションフィルタとして式 (1) で定義される Lanczos フィルタにおいて、 $N = 3$ の場合のフィルタが用いられている。

$$LanczosN(x) = \begin{cases} \frac{\sin(\pi x)}{\pi x} \frac{\sin(\frac{\pi x}{N})}{\frac{\pi x}{N}}, & |x| < N \\ 0, & otherwise \end{cases} \quad (1)$$

N: Lanczos フィルタのローブ数

式 (1) のようなデシメーションフィルタを用いる場合、要求されるカットオフ周波数は、 π/D ではなく、 $\pi U/D$ である。特に Lanczos フィルタのタップ数は、カットオフ周波数により決定されるため、低レートの場合にはより多くのタップ数が必要とされる。

2.2.1 Effective Tap Length

図 2 のシステムフローにおいて、空間スケラビリティにより取得可能な縮小画像の種類は、最終的

にシステムが出力する画像の解像度が低ければ低いほど、多様性を増す。そこで、最高品質の出力画像を得るために、従来手法では、式(2)から式(4)に示す定義式(“Effective Tap Length”)を、最高品質の出力画像を選択するための指標として定義している。

$$\begin{aligned}
 TAPS_{DWT}(n, t) &= 2 \sum_{k=0}^{n-1} \left(\frac{t-1}{2} \right)^{2^k} + 1 \\
 &= (t-1)(2^n - 1) + 1 \\
 &\quad (n > 0) \tag{2}
 \end{aligned}$$

t : DWT フィルタのタップ数, n : 分解数

奇数タップの FIR フィルタとして定義される各 Wavelet フィルタは、対称性を利用して式(2)のように、分解数 n と、そのタップ数 t (5/3-Integer フィルタは 5tap, 9/7-Daubechies フィルタは 9tap) で表される。また、Lanczos3 フィルタのタップ数は、縮小率 r に対し、式(3)で定義される。

$$TAPS_{LANC3}(r) = \lfloor \frac{1}{r} \cdot 3 \rfloor \cdot 2 + 1 \tag{3}$$

このように、Lanczos3 フィルタの特性は、フィルタのタップ数が縮小率により決定される点にある。

以上より、最終的にシステムがフィルタリング処理で利用するフィルタの有効タップ数、すなわち “Effective Tap Length” は、式(2)および式(3)を用いて、式(4)のように定義される。

$$\begin{aligned}
 TAPS_{(r,n,t)} &= TAPS_{DWT} + TAPS_{LANC3} \\
 &= \begin{cases} \lfloor \frac{1}{r} \cdot 3 \rfloor \cdot 2 + 1 & (n = 0) \\ \lfloor \frac{1}{r} \cdot 3 \rfloor \cdot 2 + (t-1)(2^n - 1) + 2 & (n \geq 1) \end{cases} \tag{4}
 \end{aligned}$$

しかし、自然画像を入力とした場合には、必ずしも正確な指標として用いることができないことがすでに明らかにされている [3]。

3. 提案システム

DCI の提案するデジタルシネマ用のファイルフォーマットである DCDM (Digital Cinema Distribution Master) は、9/7-Daubechies フィルタを用いる JPEG 2000 符号化方式と MXF (Material eXchange Format) の組み合わせにより実現される [4]。DCI のデジタルシネマは、主に劇場向けアプリケーションであり、現状のマルチメディアコンテンツの視聴環境までを考慮していない。そのため、様々な環境に対して映画コンテンツを流通させるための映像解像度変換システムが必要である。

図 3 に本提案システムの概要図を示す。以下に、JPEG 2000 で符号化されたデジタルシネマの映像ソースから、現行の様々な出力デバイスにおける映像ソースを取得するまでのシステムフローについて述べる。

表 1 Parameter values of DAR and PAR

	Horizontal[pe]l	Vertical[pe]l	DAR	PAR
HDTV	1920	1080	1.778	1.000
SDTV(NTSC)	720	480	1.333	0.889
SDTV(PAL)	720	576	1.333	1.067
VGA	640	480	1.333	1.000
QVGA	320	240	1.333	1.000
CIF	352	288	1.333	1.067
QCIF	176	144	1.333	1.067

3.1 アスペクト比変換

現在利用されている様々な映像表示デバイスは、デジタルシネマをソースとする場合に、空間解像度だけではなくピクセルアスペクト比も異なる場合がある。特に、現行の SDTV を対象とした映像のピクセルアスペクト比は、NTSC と PAL とで異なる。本提案システムでは、画面アスペクト比とピクセルアスペクト比を次のように定義する。

- DAR (Display Aspect Ratio) 画面アスペクト比
- PAR (Pixel Aspect Ratio) ピクセルアスペクト比

DAR および PAR は、垂直方向を基準とした場合の水平方向の大きさを表している。すなわち、DCDM の 4K は、 $DAR = 1.896 (= 4096/2160)$ と $PAR = 1$ であり、NTSC 方式の SDTV は、 $DAR = 1.33 (4:3)$ と $PAR = 0.89 (8:9)$ である。表 1 に、現行の出力デバイスに対する DAR と PAR の値を示す。

ピクセルアスペクト比の変化を考慮する場合、出力画像の表示方法には主に 4 種類が用いられる。

- レターボックス 上下に黒帯を入れ、ソースの映像情報は変化しない手法。
- サイドカット 左右を切り取り、全画面を利用する手法。
- フルサイズ(スクイーズ) 縦方向を基準として、横方向に圧縮する手法。
- 14:9 上下黒帯を付けると共に、左右の一部を切り取る手法。

本システムでは、レターボックス方式とサイドカッ

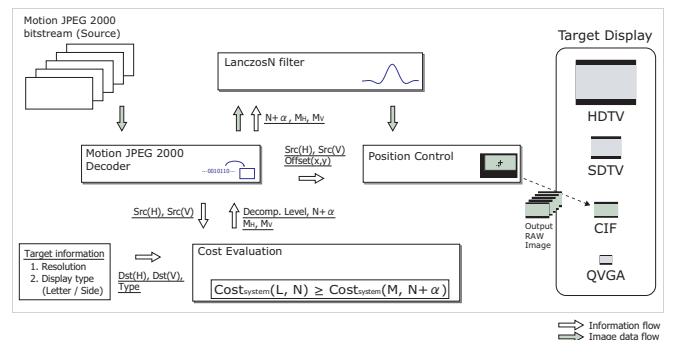


図 3 Proposed resolution conversion system

ト方式を用いる．レターボックス方式は，映画などのコンテンツ．サイドカット方式は，スポーツなどのコンテンツに対して適していると言われている [5]．

以下に，レターボックス方式とサイドカット方式，それぞれにおける解像度変換後に必要とされる有効画素領域の解像度を求めるための関係式を示す．式 (5) は，レターボックス方式における垂直解像度，式 (6)(7) は，サイドカット方式における水平解像度および左右それぞれにおいて切り捨てるピクセル数を表している．

$$Dst(V) = (Dst(H)/DAR) \times PAR \quad (5)$$

$$Dst(H) = (Dst(V) \times DAR)/PAR \quad (6)$$

$$Cut(H) = (Dst(H) - cut)/2 \quad (7)$$

cut: 符号化対象画像領域における水平画素数，*Dst(H)*: 出力映像の水平画素数，*Dst(V)*: 出力映像の垂直画素数，*DAR*: 入力映像の画面アスペクト比，*PAR*: 出力映像のピクセルアスペクト比，

3.2 最大分割数とデシメーションフィルタの縮小率

出力映像の PAR と表示タイプを元に，各組み合わせで用いる Lanczos フィルタの水平・垂直方向それぞれの縮小率 $r_{Lanczos}$ を決定する．提案システムにおいて，Lanczos フィルタに入力される画素数は，Wavelet 変換の分割数 L を用いて，式 (8) で表される．したがって，Lanczos フィルタの縮小率は，式 (9) で表される．システム全体の縮小率 r_{system} は， $r_{system} < 1$ を満たす．そのため，システムが選択可能な最大の入力シーケンスの低域分割数は，式 (10) を満たす最大の整数値である．

$$\begin{aligned} & \text{Lanczos フィルタへの入力画素数} \\ & \quad \text{システムへの入力画素数} \\ & = \frac{\quad}{2^L} \end{aligned} \quad (8)$$

$$\begin{aligned} r_{Lanczos} & = \frac{\text{システムの出力画素数}}{\text{Lanczos フィルタへの入力画素数}} \\ & = \frac{\text{システムの出力画素数}}{\text{システムへの入力画素数}} \times 2^L \\ & = r_{system} \times 2^L \end{aligned} \quad (9)$$

$$L < -\log_2(r_{system}) \quad (10)$$

3.3 演算コスト

Lanczos フィルタのローブ数と演算量の関係は，式 (1) において N の値を無限大にすれば，Lanczos フィルタが sinc 関数と等しくなり，理想的な出力信号に収束する．一方で，演算量が無限に発散することも

明らかである．そこで，従来システムよりも低い演算量での処理を保証するために，デシメーションフィルタの処理コストを定式化する．以下に，デシメーションフィルタをポリフェーズ構成により最適化した場合の演算コストを示す．この場合，デシメーションフィルタは，デシメータが実現する有理数比縮小率の分子の値により次のように回路構成に違いがある．

• U が偶数の場合

U が偶数の場合，デシメータを構成するチャンネルが偶数個存在し， $phase = 0$ と π に対応するチャンネルでは対称性を利用できる． $phase = 0$ では奇数タップの対称性， $phase = \pi$ では，偶数タップの対称性を利用できる．

• U が奇数の場合

U が奇数の場合，デシメータを構成するチャンネルが奇数個存在し， $phase = 0$ に対応するチャンネルで，奇数タップの対称性を利用できる．

本来，ポリフェーズ構成の回路では，各チャンネルからの出力を統合する加算器が存在するが，アップサンブラと遅延器をコンピュータで代用することで加算器を省略する．これにより，出力信号の 1 サンプルあたりに動作するのは，ポリフェーズ分解により分割された各チャンネルのうち 1 つのみであると言える．

また，入力信号 Src に対し，出力信号 Dst を出力するためにそれぞれのチャンネルが動作する回数は，式 (11) の関係から各チャンネルごとに U と D の最大公約数である $\gcd(U, D)$ 回で動作するため，全入力信号に対するチャンネルごとの動作回数は等しいことが分かる．

$$\begin{aligned} \frac{Dst}{Src} & = \frac{U}{D} \cdot \frac{\gcd(Src, Dst)}{\gcd(Src, Dst)} \\ \frac{Dst}{U} & = \gcd(Src, Dst) \end{aligned} \quad (11)$$

以上のような性質を利用し，ポリフェーズ構成による最小回路の演算コストを式 (12) ~ 式 (15) のように定義する．

• U が偶数の場合

$$\begin{aligned} Cost_{multi-polyphase} & = XY(2(\lfloor \frac{D_H}{U_H} N \rfloor + 1)(1 - \frac{1}{U_H}) \\ & \quad + 2(\lfloor \frac{D_V}{U_V} N \rfloor + 1)(1 - \frac{1}{U_V})) \end{aligned} \quad (12)$$

$$Cost_{add-polyphase} = XY((2\lfloor \frac{D_H}{U_H} N \rfloor + 1) - \frac{1}{U_H} + 2(\lfloor \frac{D_V}{U_V} N \rfloor + 1) - \frac{1}{U_V}) \quad (13)$$

• U が奇数の場合

$$Cost_{multi-polyphase} = XY((\lfloor \frac{D_H}{U_H} N \rfloor + 1)(2 - \frac{1}{U_H}) + (\lfloor \frac{D_V}{U_V} N \rfloor + 1)(2 - \frac{1}{U_V})) \quad (14)$$

$$Cost_{add-polyphase} = XY((2\lfloor \frac{D_H}{U_H} N \rfloor + 1) - \frac{1}{U_H} + 2(\lfloor \frac{D_V}{U_V} N \rfloor + 1) - \frac{1}{U_V}) \quad (15)$$

N : Lanczos フィルタのローブ数,
 X : 出力 2 次元信号の水平画素数,
 Y : 出力 2 次元信号の垂直画素数,
 U_H, U_V : 水平および垂直の縮小率, 分子,
 D_H, D_V : 水平および垂直の縮小率, 分母,

3.4 コスト関数

演算コスト評価部で用いるコスト関数を, 回路の乗算回数によって定義する. まず, 式 (12) と (14) で表される Lanczos フィルタの縮小率は, 式 (9) および, 式 (11) を用いることで, システムの入出力画素数である $Dst(H), Src(H)$ および $Dst(V), Src(V)$ とローブ数, Wavelet ステージの分割数 L を用いて, 式 (16) および式 (17) のように表現される.

$$Cost_{system}(L, N) = Dst(H)Dst(V)(Cost_{sub}(H, N) + Cost_{sub}(V, N)) \quad (16)$$

$$Cost_{sub}(p, N) = \begin{cases} 2(\lfloor \frac{Dst(p)2^L N}{Src(p)} \rfloor + 1)(1 - \frac{1}{U}), & U = \text{Even Number} \\ (\lfloor \frac{Dst(p)2^L N}{Src(p)} \rfloor + 1)(2 - \frac{1}{U}), & U = \text{Odd Number} \end{cases}$$

$$\text{where } U = \frac{Dst(p)}{\gcd(Dst(p), Src(p)/2^L)} \quad (17)$$

L : Wavelet ステージの分割数

N : Lanczos フィルタのローブ数,

H : 入力水平画素数, V : 入力垂直画素数,

ここで, 分割数として $L < M$ を満たす M を導入する. 分割数 M の場合には, 使用する Lanczos フィルタのローブ数を $N + \alpha$ とする. α は正の整数であ

表 2 Maximum value of alpha

L \ M	1	2	3	4
0	3	9	21	43
1		2	9	20
2			3	8
3				2

り, ロープ数の増加数を表す変数である. 分割数 L と分割数 M の場合を比較すると, 式 (18) に示される不等式を満たす α が存在するならば, 分割数が M の場合は, L の場合と比較し, 同等かもしくは低い演算量で済むと同時に Lanczos フィルタのローブ数が増加し, 出力画像の品質が向上することが予想される.

$$Cost_{system}(L, N) \geq Cost_{system}(M, N + \alpha) \quad (18)$$

表 2 に, $4096 \times 2160[\text{pel}]$ のデジタルシネマサイズからレターボックス形式で CIF サイズ ($176 \times 100[\text{pel}]$) への解像度変換における最大ローブ増加数 α の値を示す. ただし, $N = 3$ とする.

4. シミュレーション

入力画像の画像サイズをデジタルシネマで使われる, DCDM フル画素 ($4096 \times 2160[\text{pel}]$), Cinema Scope ($4096 \times 1714[\text{pel}]$), American Vista ($3996 \times 2160[\text{pel}]$) の 3 種類. 出力形式をレターボックス方式とサイドカット方式とし, 出力画像の種類として, HDTV, VGA, QVGA, SDTV (NTSC), SDTV (PAL), CIF, QCIF の計 7 種類の場合における提案システムの有効性について検証する.

4.1 実験条件

デジタルシネマに対応した 3 種類の入力画像と出力画像サイズにおける有効画素領域のサイズの対応関係は, 式 (5) および式 (6) を用いて求めることができる. ただし, DFT における理想解像度変換の出力結果と PSNR を比較するために, 解像度変換後の有効画素領域のサイズは偶数値に四捨五入して算出している.

入力画像としては, 標準画像の harbor, wool, bride, bottles の 4 種類を用いる [6]. ただし, ビット深度は $8[\text{bit}/\text{pel}]$ とする.

4.2 実験結果

実験結果を代表し, harbor における DCDM フル画素を入力とした場合の出力結果を表 3 と表 4 にまとめる. ただし, CIF の場合の結果は, 紙面の都合上省略している. 各表において, 太字で示された PSNR が各出力サイズにおける最も高い値を示した低域成分との組み合わせである.

4.3 考察

実験結果より, 提案手法による低域画像と Lanc-

表 3 PSNR(harbor, DCDM, Letter Box)

	HDTV	VGA	QVGA	NTSC	PAL	QCIF
L=0	42.279	35.618	33.812	35.648	35.979	32.448
M=1	41.180	36.962	35.037	36.949	37.315	33.558
M=2		37.861	36.262	37.659	37.765	34.665
M=3			36.112			35.013
M=4						34.238

[dB]

表 4 PSNR(harbor, DCDM, Side Cut)

	HDTV	VGA	QVGA	NTSC	PAL	QCIF
L=0	43.071	36.818	34.669	37.070	37.598	33.110
M=1	41.241	38.084	35.979	37.544	38.820	34.297
M=2		36.311	37.002	35.219		35.218
M=3			34.139			35.505

[dB]

zosN との組み合わせにより、HDTV を除く出力解像度において従来手法よりも画質が改善していることが分かる。これらの結果は、映像ソースの解像度が、American Vista の場合と Cinema Scope の場合とで共通する結果であった。

一方、提案手法において最高品質を得ることのできる低域成分と Lanczos フィルタの組み合わせは、選択可能な最大分割数の低域成分である場合と、そうでない場合の 2 つが存在している。そこで、いずれの組み合わせが最良であるかを明らかにするために、システム全体の周波数特性を分析した。その結果、周波数特性のカットオフ周波数以降の最大ピークの減衰量が、カットオフ周波数付近の減衰量を上回る周波数特性を示す場合に、画質低下が発生することが分かった。具体的な違いについて、図 4 と図 5 に、出力サイズがサイドカット形式の VGA である場合の例を示す。前者が、 $M = 1$ の場合であり後者が $M = 2$ の場合である。これらの性質は、harbor 以外の標準画像においても同様であった。

以上より、提案システムを用いることで、従来システムよりも低い演算量で、1~2[dB] の画質改善を実現することを示した。

5. まとめ

本研究では、JPEG 2000 で符号化されたデジタルシネマを起点とし、現行の様々な表示デバイスに向けた映像を取得するための解像度変換システムを提案した。提案システムは、ポリフェーズ構成により最適化されたデシメーションフィルタの演算量を定式化し拘束条件とすることで、従来手法よりも低い演算量で、アスペクト比の違いを考慮した解像度変換を行うことを可能とした。シミュレーションにより、提案手法の有効性を確認し、最高品質を与え

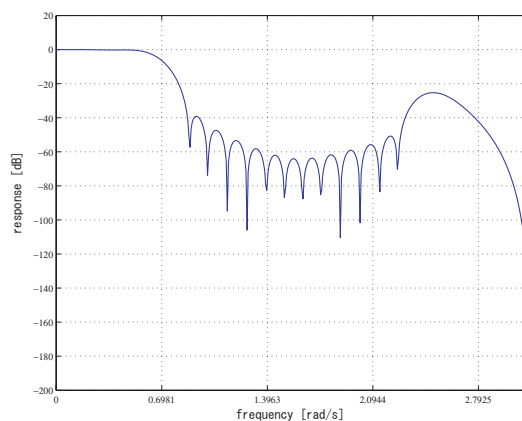


図 4 Frequency response (VGA, Side Cut, M=1)

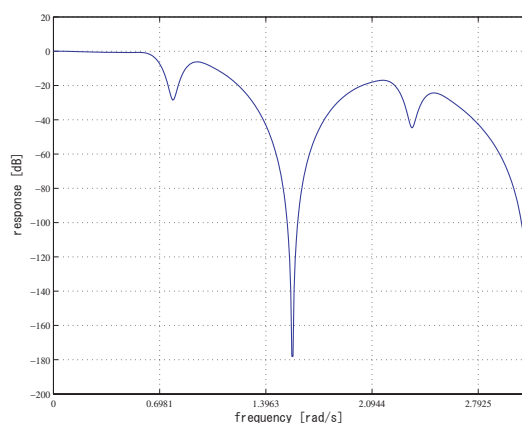


図 5 Frequency response (VGA, Side Cut, M=2)

る低域画像とデシメーションフィルタの組み合わせが、システム全体の周波数特性に依存していることを示した。

謝辞

この研究は、情報通信研究機構 (NICT) 委託研究課題“通信ネットワーク利用放送技術の研究開発”によって行われた。

文献

- [1] 富永英義, 川島正久, 沼田文彦, “Mバンド拡張 QMF バンクを用いた動画像の帯域分割符号化,” 電子情報通信学会論文誌, vol.J74-B-I, no.12, pp.1065-1073, December 1991.
- [2] 田邊集, 渡辺裕, 富永英義, “Motion JPEG 2000 における最適解像度変換法,” 画像符号化シンポジウム PCSJ2003, vol.P-1.02, November 2003.
- [3] 石川孝明, 渡辺裕, “Motion JPEG 2000 を利用したマルチ解像度変換システム,” 情報処理学会 AVIM 研究会, December 2004.
- [4] Digital Cinema Initiatives, LLC, “Digital Cinema System Specification v4.3,” December 2004.
- [5] 神原康伸, “アップ/ダウンコンバートの基礎知識,” ビデオ, pp.52-64, June 2003.
- [6] 画像電子学会, “高精細 XYZ・CIELAB・RGB 標準画像 (SHIPP),” April 2000.