2008－CVIM－162 (37)
2008／3／10

多視点画像技術の信号処理的枠組みに向けて

高橋 桂太† 苗村 健†

† 東京大学大学院情報理工学系研究科
〒113-8656 東京都文京区本郷 7-3-1

あらまし 本稿では，コンピュータビジョンの文脈で語られることの多かった多視点画像技術を，信号処理の立場から再解釈する我々の試みを紹介する．まず，左右 2 視点の中間視点画像を合成する場合を例に，視差の精度と合成画像の品質の関係を SN 比として定量化する理論モデル（視点補間の SN 比モデル）を紹介する．次に，多視点画像からの奥行き推定を，多次元信号空間において離散信号の補間を最適化する問題に置き換える理論的枠組み（エイリアシング分離理論）を紹介する．

キーワード 視点補間，エイリアシング分離，多眼ステレオ法, light field

Toward Signal Processing Framework of Multi-view Image Technologies

Keita TAKAHASHI† and Takeshi NAEMURA†

† Graduate School of Information Science and Technology, The University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8656 Japan

Abstract This paper presents our recent studies trying to reinterpret multi-view imaging technologies from the viewpoint of siginal processing, which were often discussed in the context of computer vision. We first introduce a theoretical SNR model of view interpolation that quantifies the quality of a view synthesized at the center of two given views as a function of the accuracy of disparities. We then discuss a theoretical framework of aliasing separation, in which depth estimation from multi-view images is redefined as an optimization problem of signal interpolation in the domain of ray space.

Key words view interpolation, aliasing separation, multi-view stereo, light field

1. まえがき

近年，多数の視点位置から取得されたカメラ映像を用いて，自由視点映像などの三次元的な映像効果を実現する技術が注目されている．従来，これらの技術は，コンピュータビジョンの文脈で語られることが多かった [1]〜[7]．典型的な考え方は，多視点カメラの映像をもとに，まず被写体の幾何学的な形状モデルを推定し，そのモデルに適切な着色処理をすることで自由視点映像を得る，というものである．明示的な形状を必ずしも要請せず視点間の射影幾何に基づいて画像を補間する手法も，基盤技術の分類という観点ではこのアプローチに含められよう．このアプローチの長所の一つは，コンピュータビジョン分野における数十年の技術的蓄積を直接活用できることである．カメラキャリブレーションや形状復元はコンピュータビジョンにおける伝統的なテーマであり，完成度の高い手法が数多く提案されているからである．また，幾何形状モデルによる表現は，コンピュータグラフィックス技術と親和性が高く，コンテンツの再編集などの取り扱いが容易という利点も指摘されている．本稿ではこのような立場を「コンピュータビジョン的」アプローチと呼ぶことにする．

その一方で，三次元の本質を光線とみなし，空間を伝播するあらゆる光線を多次元信号空間上でパラメータ付けして取り扱うアプローチもある [8]〜[12]．人の目が知覚しているのは，被写体の形状ではなく，被写体から放射されている光線群にほかならない．したがって，空間を伝播する光線群を記録・再現することで，三次元的な映像効果が実現できる．例えば，自由視点映像を合成する場合には，あらかじめ取得された光線群データから，所望の視点位置に到達するそれぞれの光線を読み出せばよい．一般にこのアプローチでは，「多視点」といえば，多

数のカメラが比較的高密度に配置されていることを意味することが多い．視点間隔が密になれば，隣り合う画像を別個の画像として考える代わりに，多視点画像全体を一つの多次元信号として取り扱うことが可能になるからである．本稿では，このような立場に立つアプローチを，「信号処理的」アプローチと呼ぶことにする．

一般的に，合成映像の品質においては，信号処理的アプローチのほうが優れていると言われている．しかし，それは単に入力映像の量の違いから来るものだという指摘もある．さらに，信号処理的アプローチでは，充分に高密度なカメラ配置を実現することが難しいことから，ある程度は形状復元に頼らざるを得ないことがほとんどであり，多くの場合，この部分に技術的な工夫が最も必要となる．結果として，信号処理的な立場から構成した技術であっても，単に「コンピュータビジョン応用」の一例として理解されることが多い．

筆者らはこれまで，やや信号処理寄りの立場から自由視点画像生成について研究を行なってきており [13], [14]，最近では，64 眼カメラアレイを用いたオンライン自由視点画像合成システムの開発に成功した [15], [16]．これまでの取り組みから，多視点画像技術の分野では，コンピュータビジョンのみならず，信号処理に端を発する理論体系や方法論を見直すことにも意義があると考えている．本稿では，多視点画像技術を信号処理の立場から再解釈する筆者らの最近の取り組みを紹介する．まず，左右 2 視点の中間視点画像を合成する場合を例に，視差の精度と合成画像の品質の関係を SN 比として定量化する理論モデル（視点補間の SN 比モデル）[17], [18] を紹介する．次に，多視点画像からの奥行き推定を，多次元光線空間における離散信号補間の最適化問題に置き換える理論的枠組み（エイリアシング分離理論）[19], [20] を紹介する．いずれの理論も，信号処理の分野における伝統的なトピックス（動画像符号化，離散信号の補間）に端を発するが，コンピュータビジョン分野で議論されてきた技術（視点補間，奥行き推定）に新たな知見を与えるものである．

2. 視点補間の SN 比モデル

本章では，平行撮影されたステレオ画像から視差に基づいて中間視点画像を合成する問題を考え，視差精度と合成品質との関係を定量化する新しい理論モデルについて述べる．中間視点画像の合成は，自由視点画像合成の最も基本的な設定と位置づけられる．視差精度についての定量化は，特に被写体空間を奥行き方向に層状に分割したモデル（レイヤモデル）で近似する場合，レイヤの配置間隔を決定するのに有用である．

視差精度（レイヤの配置間隔）について論じた先行研究がいくつかある．Chai ら [21] は，光線空間の周波数解析に基づくアンチエイリアス条件から，「必要な視差精度は真の視差に対して ±1 画素以内」という結論を導いた．Lin ら [22] は，幾何学的な解析に基づいて合成画像上で二重像を防ぐ条件を定式化し，Chai らと同等の結論を導いた．いずれにおいても，視差精度が不充分な場合には，プレフィルタによって高周波成分を除去することで等価的に解像度を下げる必要があるとされている．

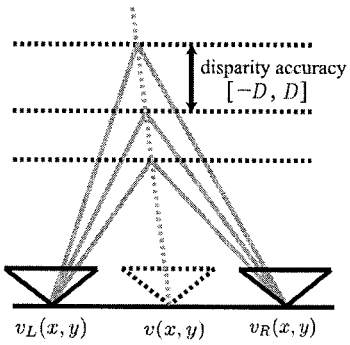


図 1　視点補間の設定
Fig. 1　Configuration for view interpolation.

一方で，我々の理論は，動画像符号化に関する理論モデル [23], [24] からヒントを得たものであり，単に視差精度の限界値を求めるのではなく，視差精度の変化に伴う合成品質の連続的な変化を定量的に表すものである．このモデルに基づいて，合成画像の誤差電力を最小化する，より優れたプレフィルタが導出される．さらに本章では，理論を検証する数値シミュレーションに加えて簡単な実験結果についても報告する．

2.1 理論モデル

2.1.1 中間視点画像合成の理論モデル

図 1 に示すように，平行に配置された左右のカメラ画像を $v_L(x,y)$, $v_R(x,y)$，合成するべき中間視点画像を $v(x,y)$ とする．簡単のため，被写体が奥行き一定の平面で構成されると仮定する．実際の画像には様々な奥行きの被写体が写っているが，局所的には上記の仮定が近似的に成り立つと考えられる．左右のカメラ画像における被写体の視差を d 画素とする．中間視点を基準に考えると，左右の視点への視差は $\mp d/2$ 画素となり，以下の関係が成り立つ．

$$v_L(x,y) = v\left(x - \frac{d}{2}, y\right)$$
$$v_R(x,y) = v\left(x + \frac{d}{2}, y\right) \qquad (1)$$

上記のモデルにおいては，オクルージョンや非ランバート反射成分の影響を無視している．さらに左右のカメラの特性差やセンサ雑音の影響も考慮していない．これらの仮定は，Chai ら [21] や Lin ら [22] の理論においても適用されており，特殊なものではない．

中間視点画像は，左右の画像に対して視差を補償して平均を取ることで合成できる．このとき，視差の真値が

正確には得られず，誤差を含む値（$d+\Delta d$）を用いると仮定し，得られる合成画像を $\hat{v}(x,y)$ とする．

$$\hat{v}(x,y)=\frac{1}{2}\left\{v_L\left(x+\frac{d+\Delta d}{2},y\right)+v_R\left(x-\frac{d+\Delta d}{2},y\right)\right\} \tag{2}$$

合成の誤差 $e(x,y)$ は以下のように定義される．

$$e(x,y)=v(x,y)-\hat{v}(x,y) \tag{3}$$

(1), (2) 式を (3) 式に代入し，フーリエ変換する．

$$E(\omega_x,\omega_y)=\left\{1-\cos\left(\frac{\Delta d\cdot\omega_x}{2}\right)\right\}V(\omega_x,\omega_y) \tag{4}$$

ここで，$E(\omega_x,\omega_y)$，$V(\omega_x,\omega_y)$ は，それぞれ $e(x,y)$，$v(x,y)$ のフーリエ変換である．さらに，(4) 式の両辺の絶対値を取って二乗し，$|E(\omega_x,\omega_y)|^2=\Phi_{ee}(\omega_x,\omega_y)$，$|V(\omega_x,\omega_y)|^2=\Phi_{vv}(\omega_x,\omega_y)$ と表記する．

$$\Phi_{ee}(\omega_x,\omega_y)=\left\{1-\cos\left(\frac{\Delta d\cdot\omega_x}{2}\right)\right\}^2\Phi_{vv}(\omega_x,\omega_y) \tag{5}$$

$\Phi_{ee}(\omega_x,\omega_y)$，$\Phi_{vv}(\omega_x,\omega_y)$ は，それぞれ，$e(x,y)$, $v(x,y)$ の電力スペクトル密度関数である．

最後に，Δd を $[-D,D]$ で一様分布する確率変数として考え，(5) 式を Δd について平均する．これは，一般の画像において，多様な値を取る視差 d を $2D$ のステップで量子化した場合に対応付けられる．

$$\begin{aligned}\overline{\Phi_{ee}}(\omega_x,\omega_y) &= \frac{1}{2D}\int_{-D}^{D}\Phi_{ee}(\omega_x,\omega_y)d(\Delta d)\\ &= G_D(\omega_x)\cdot\Phi_{vv}(\omega_x,\omega_y)\end{aligned} \tag{6}$$

$$G_D(\omega_x)=\frac{3}{2}+\frac{1}{2}\cdot\frac{\sin(D\omega_x)}{D\omega_x}-2\cdot\frac{\sin(D\omega_x/2)}{D\omega_x/2} \tag{7}$$

(6) 式によれば，誤差の電力スペクトル密度関数は，元画像の電力スペクトル密度関数にフィルタ $G_D(\omega_x)$ を施したものに他ならない．$G_D(\omega_x)$ は信号と誤差の電力比を表すため，これを誤差のゲインと呼ぶことにする．(6) 式を全周波数帯域で積分すると，合成誤差の平均電力（平均二乗誤差，MSE）が得られる．

$$MSE=\frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\overline{\Phi_{ee}}(\omega_x,\omega_y)d\omega_x d\omega_y \tag{8}$$

MSE と信号電力との比が SN 比となる．したがって，このモデルを用いれば，視差精度 D をパラメータとして，合成画像の品質を SN 比として計算できる．

ここまでに述べたモデルは，動画像符号化モデル [23], [24] における時系列方向の動きを，視点間の視差に置き換えたものと解釈できる．しかし，動きがランダムな振る舞いをするのに対して，視差は被写体の奥行きと明確に対応するものであり，(1) 式のような左右対称の関係が成り立つ．

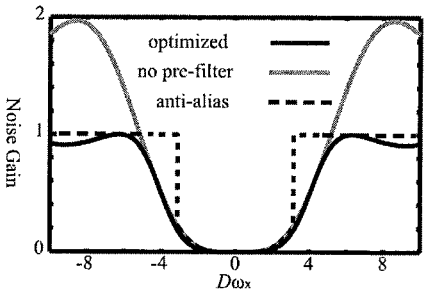


図 2　元信号に対する誤差のゲイン
Fig. 2　Noise gain.

2.1.2　誤差を最小化するプレフィルタ

本節では，(6) 式第一項のフィルタ（$G_D(\omega_x)$）の性質に着目する．Fig. 2 の "no pre-filter" は，横軸 $D\omega_x$ として $G_D(\omega_x)$ の波形を示したものである．$D\omega_x$ の定義域は $[-D\pi, D\pi]$ である．$G_D(\omega_x)$ の振幅は，$\omega_x\approx 0$ ではほぼ零に近いが，$|D\omega_x|\approx 5.16$ において 1 を超え，それ以降は振動しながら 3/2 に収束する．$G_D(\omega_x)\geqq 1$ の場合，その周波数帯においては，元の信号電力が「増幅されて」誤差電力となる．

そこで，左右の視点画像にあらかじめフィルタリングを施す（プレフィルタ）により，誤差電力を最小に抑えることを考える．すなわち，左右の画像にあらかじめ $p(x)$ を畳み込み，

$$\begin{aligned}v'_L(x,y)&=v_L(x,y)\circ p(x)\\ v'_R(x,y)&=v_R(x,y)\circ p(x)\end{aligned} \tag{9}$$

とする．以後，$v_L(x,y)$，$v_R(x,y)$ の代わりに $v'_L(x,y)$，$v'_R(x,y)$ を用いて，(2)–(8) 式と同じ過程をたどることで，プレフィルタを施した場合の平均誤差電力を導出できる．$p(x)$ の周波数波形を $P(\omega_x)$ とすると，(7) 式の $G_D(\omega_x)$ に相当する項 $G'_D(\omega_x)$ は以下のように求められる．

$$\begin{aligned}G'_D(\omega_x)=\ & 1+\frac{1}{2}\left(1+\frac{\sin(D\omega_x)}{D\omega_x}\right)\|P(\omega_x)\|^2\\ & -2\cdot\frac{\sin(D\omega_x/2)}{D\omega_x/2}\mathrm{Re}\{P(\omega_x)\}\end{aligned} \tag{10}$$

なお，$P(\omega_x)=1$ はプレフィルタを施さない場合に相当し，そのとき (10) 式は (7) 式と等しくなる．雑音を最小化するためには，すべての ω_x に対して (10) 式を最小化すればよい．したがって，$P(\omega_x)$ の絶対値，偏角に対する偏微分:

$$\frac{\partial G'_D}{\partial|P(\omega_x)|}=0,\quad \frac{\partial G'_D}{\partial(\arg(P(\omega_x)))}=0 \tag{11}$$

より，$P(\omega_x)$ の最適値 $P_{opt}(\omega_x)$ を以下のように得る．

$$P_{opt}(\omega_x)=\frac{4\sin(D\omega_x/2)}{D\omega_x+\sin(D\omega_x)} \tag{12}$$

このプレフィルタを施したときの $G'_D(\omega_x)$ の波形を，Fig.

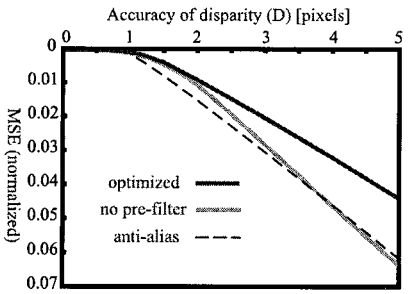


図 3　数値シミュレーション
Fig. 3　Numerical simulations.

2 の "optimized" に示す．"no pre-filter" と比較して，全帯域で振幅が抑制されていることがわかる．

一方，Chai ら [21] および Lin ら [22] の定式化に基づくアンチエイリアス条件を満たすプレフィルタは，以下の通りである（添え字は aa とする）．

$$P_{aa}(\omega_x) = \begin{cases} 1 & (|\omega_x| < \pi/D) \\ 0 & (|\omega_x| \geq \pi/D) \end{cases} \tag{13}$$

ω_x の定義域は $[-\pi, \pi]$ なので，このプレフィルタは $D < 1$（視差の誤差が 1 画素未満）のときには作用せず，$D \geq 1$（視差の誤差が 1 画素以上）のときには，エイリアシングを引き起こす高周波成分を遮断する．注目するべき点は，このとき，エイリアシング成分と同時に信号成分も遮断されることである．$P_{aa}(\omega_x)$ を用いた場合の $G'_D(\omega_x)$ の波形は，図 2 の "anti-alias" である．$\pi \leq |D\omega_x| \leq 5.16$ では "no pre-filter" よりも振幅が大きくなっており，この範囲では誤差が増大していることがわかる．

2.2　シミュレーションと実験

2.2.1　数値シミュレーション

(6) 式の第二項 $\Phi_{vv}(\omega_x, \omega_y)$ について，隣接画素間の相関係数 ρ をパラメータとする以下のモデルを用いる．

$$\Phi_{vv}(\omega_x, \omega_y) = \frac{2\pi}{{\omega_0}^2}\left(1 + \frac{{\omega_x}^2 + {\omega_y}^2}{{\omega_0}^2}\right)^{-\frac{3}{2}}$$
$$\omega_0 = -\ln(\rho) \tag{14}$$

自然画像においては ρ の値は 0.90–0.98 程度とされており，本実験では $\rho = 0.93$ を用いた．

(8) 式に基づいて MSE の数値計算を行った結果を Fig. 3 に示す．縦軸は信号電力で正規化した MSE，横軸は視差精度 D [pixels] である．2.2 の議論と対応して，プレフィルタを施さない場合（"no pre-filter"），最適なプレフィルタを用いた場合（"optimized"），およびアンチエイリアス条件 [21], [22] に基づく帯域制限フィルタを用いた場合（"anti-alias"）の三通りについて曲線を描いた．

上記の三通りのいずれにおいても，$D \leq 1$ では MSE の増加は緩やかだが，$D > 1$ ではほぼ線形の増加となる．"no pre-filter" を例に取ると，$D = 1$ では MSE=0.001 (30dB) だが，$D = 2$ では MSE は 0.01 (20dB) 程度にまで増大する．逆に，$D \leq 1$ では，視差の精度を高めて（D を小さくして）も，MSE の低減はごく限られたものである．この点は，視差精度を ±1 画素以内で充分とする従来の理論 [21], [22] とも一致しており，興味深い結果である．

さらに，三つの曲線を比較すると，プレフィルタの効果を確認できる．"optimized" では，すべての D において MSE を小さく抑える結果が得られ，有効性が確認できた．一方，"anti-alias" [21], [22] では "no pre-filter" よりも MSE が大きくなる場合がある．2.2 でも述べたように，このプレフィルタは誤差電力の抑制には適さないことがわかる．

2.2.2　実　　験

画像を用いて実際に視点補間を行い理論を検証した．まず，CG ソフトウェアで斜めに配置された 2 枚の平面から成るシーンを合成し，3 視点（左，中央，右視点）から画像をキャプチャした．次に，左右の視点画像から中央視点画像を合成した．最後に，合成された中央視点画像と，CG から直接キャプチャされた中央視点画像を比較して品質を評価した．ここでは奥行き推定そのものを問題とはしないため，視点補間の際には CG データから生成したデプス情報を直接用いた．デプスを量子化することで視差精度のパラメータ D を制御した．

例として，$D = 20/3$ のときの結果を示す．図 4 は，"no pre-filter"，"anti-alias"，および "optimized" に対応する誤差のゲインの実測値をプロットしたものである．図 2 と比較すると，定性的に一致した波形が得られていることがわかる．図 5 は合成画像である．プレフィルタを施さない場合（左, MSE=0.00530），画像上に二重像が発生していることがわかる．アンチエイリアスフィルタを施した場合（中央，MSE=0.0431），二重像は抑えられているが，画像全体がぼけている．一方，最適なプレフィルタを施した場合（右，MSE=0.0383），二重像を抑えると同時に，ある程度高周波成分が保たれていることがわかる．このとき MSE の値も三つのうち最小となっており，理論の妥当性が確認された．

3.　エイリアシング分離理論から多眼ステレオ法へ

ステレオ法とは，複数の視点位置から撮影された画像間で「対応点」を探索し，その対応関係に基づいて各点に対する視差（奥行き）を推定する手法であり，一般に空間的な幾何学の問題として説明される．一方で本章では，多眼画像を光線空間における信号とみなし，周波数解析の観点から奥行き推定の問題を考えることにより，より一般的な多眼ステレオ法の枠組みを示す．

多眼画像は，光線空間を離散的にサンプリングしたデータとみなせる．本章の議論では，奥行き推定は，離散デー

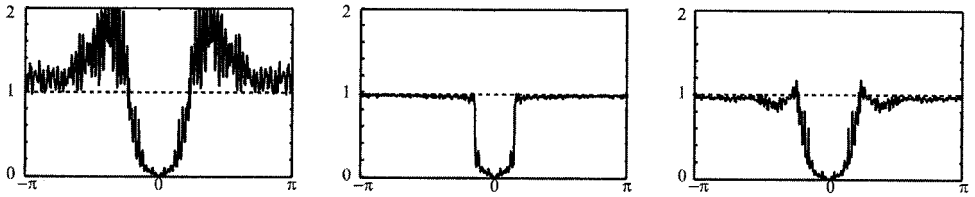


図 4　誤差ゲインの実測値：左より右に，プレフィルタなし，アンチエイリアスフィルタ，最適フィルタ．

Fig. 4　The noise gains meausured from experiments with no prefilter, the anti-alias filter, and the optimized filter.

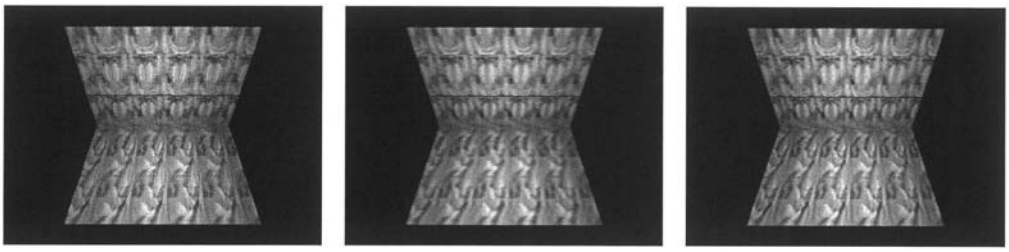


図 5　補間画像：左より右に，プレフィルタなし，アンチエイリアスフィルタ，最適フィルタ．

Fig. 5　The resulting images with no prefilter, the anti-alias filter, and the optimized filter.

タの補間に対する最適化問題に置き換えられる．すなわち，再構成される連続信号においてエイリアシング成分を最小化するように補間のパラメータを最適化することにより，間接的に奥行き推定が実現される．画像同士の単純なマッチングを基本とする従来のステレオ法は，提案する一般化された枠組み中のある特殊な場合に相当することが示される．

3.1　光線空間とエイリアシング分離理論

Fig. 6 に示すように，多数のカメラが直線上に平行・等間隔に配置されている場合を考える．カメラの位置を s，各カメラ上の画素位置を u で表すことにより，2 次元の光線空間を定義する（画像は本来 2 次元であるが，ここでは簡単のため 1 次元で考える）．$l(s,u)$ は，パラメータ (s,u) で表される光線の輝度（色）を表す．$l(s,u)$ のフーリエ変換を $L(\omega_s,\omega_u)$ と表記することにする．光線空間そのものを連続空間として定義した場合，実際の多眼画像データは，光線空間をカメラ間隔 Δs，画素間隔 Δu で離散化したものとみなせる．その空間波形および周波数波形を，それぞれ，$l_d(s,u)$，$L_d(\omega_s,\omega_u)$ と表記する．

多眼画像データの周波数成分 $L_d(\omega_s,\omega_u)$ の分布の様子を図 7 に示す．図中で，$\Omega_s = 2\pi/\Delta s$，$\Omega_u = 2\pi/\Delta u$ である．Chai ら [21] によると，スペクトルの形状は，カメラの焦点距離 f，および対象とするシーンの奥行きの最大値（$z_{\max}$）と最小値（$z_{\min}$）で決まる．中央の点線で囲んだ部分は連続信号のスペクトル成分 $L(\omega_s,\omega_u)$ に相当し，他のスペクトルは離散化に起因するエイリアシング成分を表す．

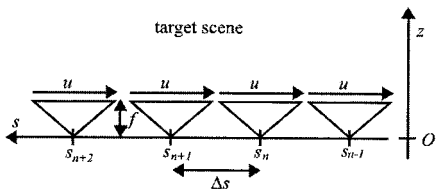


図 6　エイリアシング分離理論の設定．

Fig. 6　Configuration of aliasing separation theory.

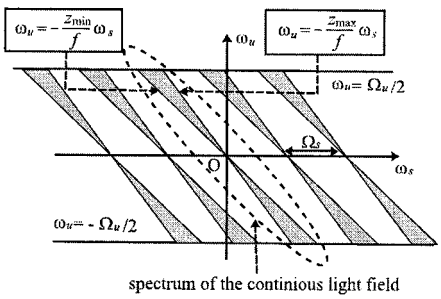
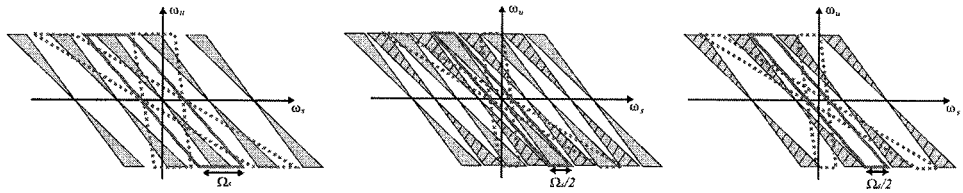


図 7　多眼画像のスペクトル分析．

Fig. 7　Spectral analysis of multi-view image data.

この離散データから連続信号を再構成するためには，Fig. 8(a) に示すような，平行四辺形状の通過帯域を持つ理想低域フィルタを施せばよい．ここで鍵となるのはフィルタの傾きである．傾きが適切な場合（図中実



(a) original ($L_d(\omega_s, \omega_u)$)　(b) subsampled ($L'_d(\omega_s, \omega_u)$)　(c) aliasing ($L_d{}^*(\omega_s, \omega_u)$)

図 8　エイリアシング分離と再構成フィルタの最適化.

Fig. 8　Aliasing separation for optimizing reconstruction filter.

線）には，連続信号が正しく再構成されるが，不適切な場合（図中破線）には，エイリアシング成分が含まれてしまう．このフィルタの傾きは，$-z_f/f$ で表され，$1/z_f = (1/z_{\min} + 1/z_{\max})/2$ を満たすときに最適である．ここまでの議論は，Chai ら [21] の文献に示されている．

では，$z_{\min}$，$z_{\max}$ が未知の場合には，フィルタの傾きをいかに最適化したらよいのだろうか．フィルタの傾きが奥行きと結びつけられることから，これは奥行き推定の問題とも捉えられる．一つの解として筆者らが提案したのが，エイリアシング分離である．この手法では，まず，与えられた離散データ $L_d(\omega_s, \omega_u)$ を 1/2 にサブサンプリングする（図 8(b): $L'_d(\omega_s, \omega_u)$）．続いて，次式に示すように，振幅を揃えて両者を引き算することで，サブサンプリングによって追加された成分のみを分離する（図 8(c): $L_d{}^*(\omega_s, \omega_u)$）．

$$L_d{}^*(\omega_s, \omega_u) = L_d(\omega_s, \omega_u) - 2L'_d(\omega_s, \omega_u) \tag{15}$$

図 8(c) のように，$L_d{}^*(\omega_s, \omega_u)$ に対して再構成フィルタを施す場合には，フィルタの傾きが最適であれば再構成される信号は零となる．したがって，次のような評価関数を z_f に対して最小化する問題を考えればよい．

$$E(z_f) = L_d{}^*(\omega_s, \omega_u) \cdot H_{z_f}(\omega_s, \omega_u) \tag{16}$$

ただし H_{z_f} は傾き $-z_f/f$ の理想的低域フィルタを表す．すなわち，この評価関数は，分離されたエイリアシング成分から再構成される信号を表しており，その信号を最小化するように再構成フィルタのパラメータを調整することが奥行き推定に対応付けられる．

ここで，$z_{\min} = z_{\max} = Z_0$ の場合を考える．このとき，図 8(c) において，各スペクトル成分は傾き $-Z_0/f$ の直線に収束し，これらの成分がフィルタの通過域に含まれないときに $E(z_f) = 0$ となる．その条件は以下のように表される．

$$\left|\frac{1}{z_f} - \frac{1}{Z_0}\right| \leqq \frac{\Omega_s}{2f\Omega_u} = \frac{1}{f\Delta s} \cdot \frac{\Delta u}{2} \tag{17}$$

(17) 式は奥行き推定の精度を表しており，推定される奥行き z_f と真の値 Z_0 の誤差が，隣合うカメラで ± 0.5 画素 ($\Delta u/2$) 以内の視差に収まることを意味している．従来のステレオ法では，推定される奥行きの曖昧さは，テクスチャの不足や画像の雑音に起因するものである．したがって，理想的な条件のもとで，小数画素精度でマッチングを行えば，原理的には真の奥行きが一意に推定できるとされている [25]．一方，(16) 式による評価では，同じ理想的な条件においても，(17) 式で表される範囲で奥行きの曖昧さが許容される．この精度は，多眼画像データから連続的な光線空間信号を再構成する目的に対しては原理的に充分であり，2 章の議論から実用上も充分であると言える．

3.2　一般化された多眼ステレオ法の導出

3.1 で周波数領域において展開した議論を，空間領域に戻して考える．まず，(16) 式をフーリエ逆変換する．

$$e(s, u, z_f) = l_d{}^*(s, u) \circ h_{z_f}(s, u) \tag{18}$$

ここで，$\circ$ は畳み込み積分を表す．$h_{z_f}(s, u)$ は，sinc 関数 ($\mathrm{sinc}(x) = \sin(x)/x$) を用いて以下のように表される．

$$h_{z_f}(s, u) = \mathrm{sinc}\left(\frac{\Omega_s s}{4}\right) \cdot \mathrm{sinc}\left(\frac{\Omega_u(u - fs/z_f)}{2}\right) \tag{19}$$

ここで，第 n 番目のカメラ画像（連続信号）を $c_n(u)$ と表記する．さらに，このカメラ画像を，s を基準とし，奥行き z_f を仮定して視差補償した画像を $c_n(u; s, z_f)$ とする．

$$c_n(u) := l(n\Delta s, u) \tag{20}$$

$$c_n(u; s, z_f) = c_n\left(u - \frac{f(s - n\Delta s)}{z_f}\right) \tag{21}$$

(18) 式は (19)-(21) 式を用いて次のように表される．

$$e(s, u, z_f) = \sum_n \{c_{2n}(u; s, z_f) - c_{2n+1}(u; s, z_f)\} \circ \mathrm{sinc}\left(\frac{\Omega_s s}{4}\right) \tag{22}$$

繰り返しになるが，この評価関数は，分離されたエイリアシング成分から再構成される信号を表しており，これを最小化する z_f を求めることが奥行き推定に対応する．

(22) 式が，従来の基本的なステレオ法を包含する，より一般的なステレオ法となっていることを示す．例とし

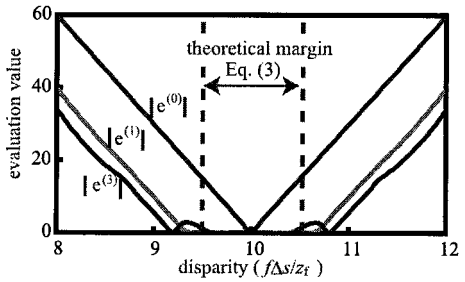


図 9　多眼ステレオの実験結果
Fig. 9　Experimental result of multi-view stereo.

て，$s = 1/2$ の場合を考える．一般に，sinc 関数は厳密には実装できないため近似を用いる．よく用いられる近似として，零次ホールド（最近傍補間），一次ホールド（線形補間），および 3 次 B スプライン関数(注1)を当てはめた場合の評価関数をそれぞれ，$e^{(0)}(.)$，$e^{(1)}(.)$，$e^{(3)}(.)$ とする（注：スペースの都合上，引数を省略して表記する）．

$$e^{(0)}(.) = c_0(.) - c_1(.) \tag{23}$$

$$e^{(1)}(.) = -\frac{1}{4}c_{-1}(.) + \frac{3}{4}c_0(.) - \frac{3}{4}c_1(.) + \frac{1}{4}c_2(.) \tag{24}$$

$$\begin{aligned} e^{(3)}(.) = &-\frac{1}{384}c_{-3}(.) + \frac{27}{384}c_{-2}(.) - \frac{121}{384}c_{-1}(.) \\ &+ \frac{235}{384}c_0(.) - \frac{235}{384}c_1(.) + \frac{121}{384}c_2(.) \\ &- \frac{27}{384}c_3(.) + \frac{1}{384}c_4(.) \end{aligned} \tag{25}$$

(23) 式は 2 枚の画像間の差分を評価するものであり，従来のステレオ法は基本的にこれと等価である．このような画像の組み合わせを複数同時に評価するのが，従来の多眼ステレオ法 [25] である．さらに (22) 式からは，(24)(25) 式のように，複数の画像を重み付けして評価する関数も導出される．次数が高くなるほど計算が複雑になるが，より確実に高周波成分を遮断するようになる．

(23)–(25) 式で示した評価関数の性質を比較する予備実験を行った．この実験では，奥行き一定の平面に自然画像を貼り付けた CG シーンを，視点位置を移動しながらキャプチャすることにより，多眼画像データを生成した．サブピクセルレベルの視差補償では線形補間を用いて画素値を生成した．それぞれの評価関数の変化の様子を Fig. 9 に示す．縦軸は 10^4 画素に対する評価関数の平均値（各画素の RGB 各成分についての評価値を単純に平均したもの）を表す．横軸には奥行き z_f を隣接カメラ間の視差に換算した値を用いた．参考のため，(17) 式で与えられる奥行きの許容範囲を図中に描き入れた．

これらの結果より，従来のステレオ法（$e^{(0)}$）では，真の奥行き（$z_f = Z_0$：視差 10.0 に相当）に対してのみ評価値が零となるが，$e^{(1)}$，$e^{(3)}$ では，より広い範囲で零に近い値を取ることがわかる．これは，$e^{(1)}$，$e^{(3)}$ が理想低域フィルタをより正確に近似するためであり，3.1 の議論を支持する結果である．これは，一次以上の近似を取り入れると，奥行き推定の精度が粗くなることを意味していると捉えられる．しかし同時に，真値付近の比較的広い範囲で評価関数が零に近くなることから，奥行き推定の安定性の向上が期待できる．

例えば，被写体がカメラの撮像面に対して厳密に平行平面にならない場合を考える．このような場合，通常のステレオ法では，ブロックベースでマッチングコストを計算すると，奥行き推定の安定性が低下することが知られている．これは，ブロック内で画素ごとにわずかに奥行きが異なるため，最も妥当な奥行きにおいてもマッチングコストが充分に小さくはならないためである．一方，一次以上の近似を取り入れたコスト関数は真値付近で平坦な特性を持つため，このような問題に対して効果を発揮することが期待できる．また，本稿とは問題設定が異なるが，文献 [26] では，画素のサンプリングの影響に対応することを目的として，別の方法で同様の性質を持つ評価関数を実現しており，奥行き推定の性能向上が報告されている．

4. ま と め

本稿では，多視点画像技術を信号処理的な立場から再解釈する筆者らの取り組みを大きく 2 つに分けて紹介した．これらの取り組みを通して，信号処理とコンピュータビジョンを関連付ける新たな接点が見出されたことに加え，両者の融合によって初めて可能となる新たな技術的知見も得られつつある点は収穫と言えよう．これらの研究をさらに発展させ，多視点画像技術の信号処理的枠組みの構築に結び付けていきたい．

謝辞　本研究の一部は，文部科学省「先端融合領域イノベーション創出拠点の形成：少子高齢社会と人を支える IRT 基盤の創出」の支援を受けました．熱心に議論してくださった東京大学の原島 博 教授に感謝いたします．

(注1)：本稿では 3 次 B スプライン関数を直接畳み込んでいるが，より厳密な補間のためには特殊なプレフィルタ処理が必要である．このプレフィルタと 3 次 B スプラインを組み合わせたものは，cardinal cubic spline と呼ばれる．

文　献

[1] T. Kanade, P. Rander, and P.J. Narayanan, "Virtualized Reality: Constructing Virtual Worlds from Real Scenes," IEEE Multimedia, Immersive Telepresence, 4, 1, pp. 34–47, 1997.

[2] W. Matusik, C. Buehler, R. Raskar, S.J. Gortler, and L. McMillan: "Image-Based Visual Hulls," Proc. ACM SIGGRAPH 2000, pp. 369–374, 2000.

[3] R. Yang, G. Welch, G. Bishop,"Real-Time Consensusbased Scene Reconstruction Using Commodity Graphics Hardware," Proc. Pacific Graphics, pp.

225–235, 2002.

[4] T. Matsuyama, X. Wu, T. Takai, S. Nobuhara, "Real-Time 3D Shape Reconstruction, Dynamic 3D Mesh Deformation, and High Fidelity Visualization for 3D Video," Intl. J. Computer Vision and Image Understanding, 96, 3, pp. 393–434, 2004

[5] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski: "High-Quality Video View Interpolation using a Layered Representation," Proc. ACM SIGGRAPH'04, pp. 600–608, 2004.

[6] N. Inamoto, H. Saito: "Virtual Viewpoint Replay for a Soccer Match by View Interpolation From Multiple Cameras," IEEE Trans. Multimeida, 9, 6, pp. 1155–1166, 2007.

[7] 鍋島, 上田, 有田, 谷口: "実時間自由視点映像生成のフレームレート安定化 –形状復元の多重解像度処理–", 信学論, J90-D, 12, pp. 3221–3233, 2007.

[8] 藤井, 金子, 原島: "光線群による 3 次元空間情報の表現とその応用", テレビ誌, 50, 9, pp. 1312–1318, 1996.

[9] M. Levoy, P. Hanrahan: "Light Field Rendering," Proc. ACM SIGGRAPH'96, pp. 31–42, 1996.

[10] T. Naemura, M. Kaneko, H. Harashima: "Orthographic Approach to Representing 3-D Images and Interpolating Light Rays for 3-D Image Communication and Virtual Environment", Eurasip J. Signal Process.: Image Commun., 14, pp. 21–37, 1998.

[11] M. Tanimoto, "FTV (Free Viewpoint Television) Creating Ray Based Image Engineering," Proc. IEEE ICIP'05, pp. 25–28, 2005.

[12] J. Berent, P.L. Dragotti: "Plenoptic Manifolds – Exploiting Structure and Coherence in Multi-View Images," IEEE Signal Processing Magazine, 24, 6, pp. 34–44, 2007

[13] K. Takahashi, A. Kubota, T. Naemura: "A Focus Measure for Light Field Rendering," Proc. IEEE ICIP'04, 4, pp. 2475–2478, 2004.

[14] K. Takahashi, T. Naemura, "Layered Light-Field Rendering with Focus Measurement," Eurasip Signal Processing: Image Communication, 21, 6, pp. 519–530, 2006.

[15] Y. Taguchi, K. Takahashi, T. Naemura, "Real-Time All-in-Focus Video-Based Rendering Using Network Camera Array," submitted to 3DTV-conference 2008.

[16] 田口, 高橋, 苗村, "ネットワークカメラアレイを用いた実時間全焦点自由視点画像合成システム", 信学技報 PRMU, Mar. 2008.

[17] 高橋, 苗村: "Plenoptic Sampling 再考 –視差精度が視点補間の品質に与える影響について–", 映像メディア処理シンポジウム, pp. 27–28, Nov. 2007.

[18] K. Takahashi, T. Naemura, "Theoretical Model and Optimal Prefilter for View Interpolation," submitted to IEEE ICIP08.

[19] K. Takahashi, T. Naemura, "A theory of aliasing separation for light field data," Proc. IEEE ICIP'06, pp. 1201–1204, 2006.

[20] 高橋, 苗村, "多眼ステレオ法の周波数解析的な解釈と一般化", 映像メディア処理シンポジウム, pp. 99–100, Nov. 2006.

[21] J.-X. Chai, S-.C. Chai, H.-Y. Shum, X. Tong, "Plenoptic sampling," Proc. ACM SIGGRAPH'00, pp. 307–318, 2000.

[22] Z. Lin, H.-Y. Shum, "A Geometric Analysis of Light Field Rendering," IJCV, 58, 2, pp. 121.138, 2004.

[23] B. Girod, " The efficiency of motion-compsensating prediction for hybrid coding of video sequences, " IEEE Journal SAC, SAC-5, 7, pp. 1140.1154, 1987.

[24] 酒井, 吉田, "映像情報符号化" オーム社, ヒューマンコミュニケーション工学シリーズ, 2001.

[25] M. Okutomi, T. Kanade: "A multiple baseline stereo," IEEE Trans. PAMI, 15, 4, pp. 353–363, 1993.

[26] S. Birchfiled, C. Tomasi: "A Pixel Dissimilarity Measure That Is Insensitive to Image Sampling," IEEE TPAMI, 20, 4. pp. 401–406, 1998.