

レート-歪み理論から見た多眼画像の間引きと符号化 Subsampling and Coding of Multi-View Images from Rate-Distortion Theory

高橋 桂太[†]

Keita Takahashi

苗村 健[†]

Takeshi Naemura

1. はじめに

本稿では、多数のカメラを配列して取得される多眼カメラ画像を対象とした符号化の問題に注目する。筆者らの知る限りにおいて、この分野の研究は20年以上の歴史を持つ[1]。近年においても、さらなる符号化効率の向上を目指して、視差補償予測や幾何学モデルベース符号化などの多眼画像に特化した要素技術について多くの工夫が積み重ねられており[2, 3, 4, 5]、実用化に向けた標準化活動も行われている[6]。

その一方で、多眼カメラの配置密度と情報量の関係については、必ずしも十分に議論されてきたとは言えない。例えば、近年では、百台以上のカメラを配列した画像取得方式も検討されている[7, 8, 9]。仮にカメラを一つおきに間引いて台数を半分(二次元配列の場合は1/4)にしたとしても、何らかの補間処理により、間引かれた画像をある程度復元できる可能性がある。では、間引きによって削減できるビット量と失われる情報量のバランスはどのようなものだろうか?この問題は、「対象空間の情報を取得するのに、そもそも、それほど多くのカメラが必要だったのだろうか?」という、より根本的な問いにも結びつく。

本稿では、この問題に対して、レート-歪み理論[10, 11]を用いて数値的な解析を試みた結果を報告する。本稿の解析により、対象とする被写体の幾何学的・光学的な複雑さや、推定される視差の精度に応じて、カメラを間引きが有利/不利になる場合の双方あり得ることが示される。本稿の構成は以下の通りである。2章では、背景として、画像を対象としたレート-歪み理論の概要を述べる。3章で、多眼画像の間引きと符号化効率を結びつける理論モデルを提案する。4章では、提案モデルを用いた数値シミュレーションの結果を示し、5章で本稿を締めくくる。

2. 背景

符号化対象の二次元画像を $v(x, y)$ 、その電力スペクトル密度関数を $\Phi_{vv}(\omega_x, \omega_y)$ と表記する。許容される雑音レベルを決める正のパラメータ θ に対して、一画素あたりのレート $R(\theta)$ (bit) および歪み $D(\theta)$ の理論的限界は以下のように与えられる[10, 11]。

$$R(\theta) = \frac{1}{8\pi^2} \int_{\omega_x} \int_{\omega_y} \max \left\{ 0, \log_2 \frac{\Phi_{vv}(\omega_x, \omega_y)}{\theta} \right\} d\omega_x d\omega_y \quad (1)$$

$$D(\theta) = \frac{1}{4\pi^2} \int_{\omega_x} \int_{\omega_y} \min \{ \theta, \Phi_{vv}(\omega_x, \omega_y) \} d\omega_x d\omega_y \quad (2)$$

歪み D の最大値を $D(\theta)$ に制限した場合、レート R の最小値は $R(\theta)$ となる。 θ が小さいほど、高レートかつ高品質(歪みが小さい)になる。

静止画像の電力スペクトル密度関数としては、隣り合う画素同士の相関係数 ρ (自然画像の場合、 $\rho = 0.90-0.98$ 程度)のみをパラメータとする以下のようなモデルがよく用いられる[12, 13]。

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

$$\omega_0 = -\ln(\rho) \quad (4)$$

この関数の波形は、直流成分 $\omega_x = \omega_y = 0$ でピークを持ち、高域に行くにしたがって減衰する。これは、自然画像では、低周波に電力が集中することに対応する。

動画像や多眼画像を対象とする予測符号化においては、(1), (2) 式で $\Phi_{vv}(\omega_x, \omega_y)$ の代わりに、予測残差画像 $e(x, y)$ の電力スペクトル密度関数 $\Phi_{ee}(\omega_x, \omega_y)$ を代入することにより、その画像に対するレート-歪み限界を定式化できる[11, 12, 13]。

3. 多眼画像の間引きと符号化の理論モデル

3.1 問題設定

本稿では、水平方向に平行かつ等間隔に多数のカメラを配置して取得された多眼画像を対象とし、以下の二つのシナリオを比較する。

1. すべてのカメラ画像を同一の品質で符号化・伝送する。
2. 送信側では、カメラを一つおきに間引き、半数のカメラ画像のみを同一の品質で符号化・伝送する。受信側では、受信した画像を復号するとともに、復号した画像をもとに視差推定または形状推定を行い、間引かれた画像も推定して復元する。

1. と比較して、2. では伝送される画像数が半分となるため、レートが削減できる。しかし、間引かれた画像を完全に復元することはできないため、全体の歪みも増大する。どのような場合に、削減できるレートに対して、失われる情報量が「割に合う」のだろうか?

筆者らは文献[14]において、カメラ間の画像の相関を用いた符号化方式を想定し、レート歪理論と plenoptic sampling 理論[15]を組み合わせた理論モデルを構築した。しかし、問題の複雑さのため、同一の θ に対して1. と2. を比較した場合の、レートの変化量および歪みの変化量を導出するにとどまった。本稿では、各カメラで独立に画像を符号化する(各カメラでの時系列方向の予測符号化も考えない)、より単純なケース

[†]東京大学 大学院情報理工学系研究科 電子情報学専攻
Dept. Inform. & Commun. Engineering, School of Inform. Science and Technology, The University of Tokyo

を想定することにより、上記の二つのシナリオに対するレートおよび歪みの限界を明示的に導出し、 θ に対する動的な特性を解析する。

3.2 レート-歪み特性の導出

本節では、3.1 で述べた二つのシナリオのレート-歪み特性を定式化する。まず、最初のシナリオでは、各画像は同等の品質で独立に符号化される。したがって、レート $R_1(\theta)$ 、および歪み $D_1(\theta)$ は、それぞれ、(1)、(2) 式によって直接与えられる。

2 番目のシナリオでは、送信側でカメラを一つおきに間引き、残りの画像を同等の品質で独立に符号化する。したがって、間引いた画像の分まで考慮した全体のレート $R_2(\theta)$ は $R_1(\theta)$ の半分となる。

$$R_2(\theta) = \frac{1}{2} R_1(\theta) \quad (5)$$

次いで歪みについて考える。伝送される画像には $D_1(\theta)$ の歪みが生じる。間引かれる画像は、受信側で推定して復元されるため、その予測誤差に相当する歪みが生じる。したがって、全体の歪み $D_2(\theta)$ は、これらの2項の平均として与えられる。

$$D_2(\theta) = \frac{1}{2} \left\{ D_1(\theta) + \frac{1}{4\pi^2} \int_{\omega_x} \int_{\omega_y} \Phi_{ee}(\omega_x, \omega_y) d\omega_x d\omega_y \right\} \quad (6)$$

ただし、 $\Phi_{ee}(\omega_x, \omega_y)$ は、間引かれた画像の予測誤差の電力スペクトル密度関数である。本章の残りの部分では、この予測誤差成分を定式化する。

3.3 間引かれた画像の予測誤差のモデル化

いま、第 i 番目の画像 $v_i(x, y)$ が送信側で間引かれるとする。受信側では、左右の隣接するカメラ画像を視差補償して平均を取るることによって、予測画像 $\hat{v}_i(x, y)$ を合成すると仮定する。 $v_i(x, y)$ と $\hat{v}_i(x, y)$ の差分が予測誤差であり、この信号の電力スペクトル密度関数を導出することが本節の目的である。本節の議論は、文献 [11] に示されている議論の拡張であるが、いくつか独自のパラメータを導入しているため、以下、順を追って説明することにする。

3.3.1 予測画像生成のモデル

まず、予測対象のカメラ画像 $v_i(x, y)$ と、左右の隣接するカメラ画像 $v_{i-1}(x, y)$ 、 $v_{i+1}(x, y)$ とを以下のようなモデルで関連付ける。

$$v_{i-1}(x, y) = v_i(x - d, y) + n_{i-1}(x, y) \quad (7)$$

$$v_{i+1}(x, y) = v_i(x + d, y) + n_{i+1}(x, y) \quad (8)$$

上記で、 d は、推定対象の画像 $v_i(x, y)$ と隣接する画像 $v_{i\pm 1}(x, y)$ の視差を表す。隣接する2枚の画像は、推定対象の画像を基準に左右対称の位置にあるため、それぞれの視差は正負を逆にした関係になる。ここで、 d を位置 x に依存しない量として定義することは、被写体を奥行き一定の平面と仮定していることに等しい。議論の一般性を高めるため、上記の式では、さらに雑音成分 $n_{i\pm 1}(x, y)$ を加算している。この項は、表面の

微小な凹凸、非ランバート反射、オクルージョン、カメラのノイズなどに起因する雑音成分など、視差補償によって予測不可能な成分すべてを含む。実際の多眼画像に対しても、局所的に考えれば上記のモデルを当てはめることができる。

受信側で実際に得られる画像 $v'_{i\pm 1}(x, y)$ は、符号化・復号化の過程を経るため、 $v_{i\pm 1}(x, y)$ に対して符号化に起因する雑音成分 $n'_{i\pm 1}(x, y)$ を加算したものとなる。

$$v'_{i-1}(x, y) = v_{i-1}(x, y) + n'_{i-1}(x, y) \quad (9)$$

$$v'_{i+1}(x, y) = v_{i+1}(x, y) + n'_{i+1}(x, y) \quad (10)$$

最後に、 $v'_{i-1}(x, y)$ 、 $v'_{i+1}(x, y)$ に対して視差補償を行い、両者の平均を取るることによって $v_i(x, y)$ の予測画像 $\hat{v}_i(x, y)$ を合成する[†]。ここでは、伝送されてきた画像 $v'_{i-1}(x, y)$ 、 $v'_{i+1}(x, y)$ の間でマッチング演算などを行うことにより、視差の値 d を推定すること想定している。ただし、一般に視差の推定精度には限界（原理的な限界または実用上の計算量的な限界）があるため、誤差 Δd が加算された視差 $d + \Delta d$ で視差補償すると仮定する。

$$\hat{v}_i(x, y) = \frac{v'_{i-1}(x + (d + \Delta d), y) + v'_{i+1}(x - (d + \Delta d), y)}{2} \quad (11)$$

Δd は確率変数であり、後で平均を取る。

3.3.2 予測誤差の定式化

(7)–(11) 式を用いて、元の画像 $v_i(x, y)$ と予測された画像 $\hat{v}_i(x, y)$ との誤差 $e(x, y)$ を計算する。

$$e(x, y) = v_i(x, y) - \hat{v}_i(x, y) = \left\{ v_i(x, y) - \frac{v_i(x + \Delta d, y) + v_i(x - \Delta d, y)}{2} \right\} - \frac{n_{i-1}(x + d + \Delta d, y) + n_{i+1}(x - (d + \Delta d), y)}{2} - \frac{n'_{i-1}(x + d + \Delta d, y) + n'_{i+1}(x - (d + \Delta d), y)}{2} \quad (12)$$

上式で、第1項は視差の誤差（奥行き推定の誤差）に起因する成分、第2項は、被写体の複雑さやカメラ雑音などに起因する成分、第3項は予測元画像の符号化歪みに起因する成分である。(12) 式のフーリエ変換によって、誤差信号のスペクトル $E(\omega_x, \omega_y)$ を得る。

$$E(\omega_x, \omega_y) = V_i(\omega_x, \omega_y) \cdot \left(1 - \frac{e^{j\Delta d\omega_x} + e^{-j\Delta d\omega_x}}{2} \right) - \frac{N_{i-1}(\omega_x, \omega_y)P_{i-1}(\omega_x) + N_{i+1}(\omega_x, \omega_y)P_{i+1}(\omega_x)}{2} - \frac{N'_{i-1}(\omega_x, \omega_y)P_{i-1}(\omega_x) + N'_{i+1}(\omega_x, \omega_y)P_{i+1}(\omega_x)}{2} \quad (13)$$

上記では、フーリエ変換の空間シフトの性質を用いている。 $V_i(\omega_x, \omega_y)$ は、 $v(x, y)$ のフーリエ変換を表す。 $N_{i\pm 1}(\omega_x, \omega_y)$ 、 $N'_{i\pm 1}(\omega_x, \omega_y)$ についても同様である。 $P_{i\pm 1}(\omega_x)$ は位相変位の項を表し、 $P_{i\pm 1}(\omega_x) = \exp(\mp j(d + \Delta d)\omega_x)$ とする。

[†] 雑音成分 $n_{i\pm 1}$ 、 $n'_{i\pm 1}$ の電力が等しく、かつ v_i に対して独立であると仮定すれば、 $v'_{i\pm 1}$ のどちらか一方の画像を用いるよりも、両者の平均を取ったほうが、予測誤差が小さくなる。

3.3.3 電力スペクトル密度関数の導出

いよいよ、予測誤差信号の電力スペクトル密度関数 $\Phi_{ee}(\omega_x, \omega_y)$ を計算する。 $v_i(x, y)$, $n_{i\pm 1}(x, y)$, $n_{i\pm 1}^\theta(x, y)$ がそれぞれ統計的に独立であると仮定すると、誤差信号の電力は、(13) 式の各項ごとの電力の和となる。

$$\Phi_{ee}(\omega_x, \omega_y) = \|E(\omega_x, \omega_y)\|^2 = C(\omega_x) \cdot \Phi_{vv}(\omega_x, \omega_y) + \Phi_{nn}(\omega_x, \omega_y) + \Phi_{n\theta n\theta}(\omega_x, \omega_y) \quad (14)$$

上記において、 $\Phi_{vv}(\omega_x, \omega_y) = \|V_i(\omega_x, \omega_y)\|^2$ は画像の電力スペクトル密度関数そのものであり、(3) 式によって与えられる。 $C(\omega_x)$, $\Phi_{nn}(\omega_x, \omega_y)$, $\Phi_{n\theta n\theta}(\omega_x, \omega_y)$ はそれぞれ以下のように求められる。

$$C(\omega_x) = (1 - \cos(\Delta d \omega_x))^2 \quad (15)$$

$$\Phi_{nn}(\omega_x, \omega_y) = \frac{\|N_{i-1}(\omega_x, \omega_y)\|^2}{4} + \frac{\|N_{i+1}(\omega_x, \omega_y)\|^2}{4} \quad (16)$$

$$\Phi_{n\theta n\theta}(\omega_x, \omega_y) = \frac{\|N_{i-1}^\theta(\omega_x, \omega_y)\|^2}{4} + \frac{\|N_{i+1}^\theta(\omega_x, \omega_y)\|^2}{4} \quad (17)$$

(16), (17) 式では、 $i-1$ 番目の画像と $i+1$ 番目の画像の雑音成分同士も相互に独立であると仮定している。

ここで、視差の誤差 Δd (確率変数) について (14) 式の平均を求めたものを $\overline{\Phi_{ee}(\omega_x, \omega_y)}$ と置く。

$$\overline{\Phi_{ee}(\omega_x, \omega_y)} = \overline{C(\omega_x)} \cdot \Phi_{vv}(\omega_x, \omega_y) + \Phi_{nn}(\omega_x, \omega_y) + \Phi_{n\theta n\theta}(\omega_x, \omega_y) \quad (18)$$

3.4 画像を間引いた場合の歪みの導出

最後に、(6) 式の $\Phi_{ee}(\omega_x, \omega_y)$ に (18) 式を代入して、2 番目のシナリオにおける歪み $D_2(\theta)$ を計算する。

$$\begin{aligned} D_2(\theta) &= \frac{1}{2} \left\{ D_1(\theta) + \frac{1}{4\pi^2} \iint_{\omega_x, \omega_y} \overline{\Phi_{ee}(\omega_x, \omega_y)} d\omega_x d\omega_y \right\} \\ &= \frac{1}{2} D_1(\theta) + \frac{1}{2} \cdot \frac{1}{4\pi^2} \iint \overline{C(\omega_x)} \Phi_{vv}(\omega_x, \omega_y) d\omega_x d\omega_y \\ &\quad + \frac{1}{2} \cdot \frac{1}{4\pi^2} \iint \Phi_{nn}(\omega_x, \omega_y) d\omega_x d\omega_y \\ &\quad + \frac{1}{2} \cdot \frac{1}{4\pi^2} \iint \Phi_{n\theta n\theta}(\omega_x, \omega_y) d\omega_x d\omega_y \quad (19) \end{aligned}$$

(19) 式の第2項で、 $\overline{C(\omega_x)}$ については、 Δd が区間 $-\beta/2 \leq \Delta d \leq \beta/2$ で一様分布すると仮定して、以下のように求める。 β は視差 (奥行き推定) の精度を表すパラメータとみなせる。

$$\begin{aligned} \overline{C(\omega_x)} &= \frac{1}{\beta} \int_{-\beta/2}^{\beta/2} C(\omega_x) d(\Delta d) \\ &= \frac{3}{2} + \frac{1}{2} \frac{\sin(\beta \omega_x)}{(\beta \omega_x)} - 2 \frac{\sin(\beta \omega_x/2)}{\beta \omega_x/2} \quad (20) \end{aligned}$$

横軸を $\beta \omega_x$ とした $\overline{C(\omega_x)}$ の波形を図1に示す。 $\overline{C(\omega_x)} > 1$ となる場合、その周波数成分は予測によって増幅される。ただし、通常 β は数画素以内、かつ $-\pi \leq \omega_x \leq \pi$ であることから、実際に使うのは $\beta \omega_x$ が零付近の区間であり、増幅の影響は小さい。

(19) 式の第3項は、視差補償によって予測不可能な信号成分に相当する。 $\Phi_{nn}(\omega_x, \omega_y)$ の周波数波形を具

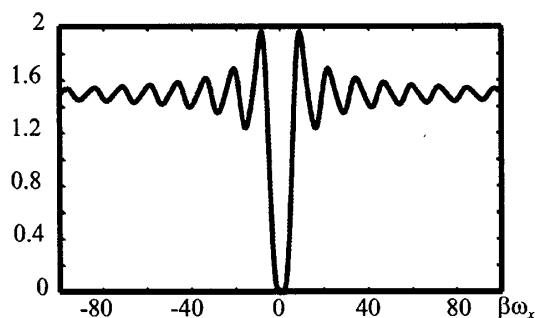


図1: $\overline{C(\omega_x)}$ の波形

体的にモデル化するのは困難だが、これを積分して得られる電力については、以下のように、画像の信号電力の α 倍としても一般性を失わない。 α は多眼画像の複雑さ (被写体の複雑さと各カメラに固有の雑音を含む) を示すパラメータと考えることができる。

$$\iint \Phi_{nn}(\omega_x, \omega_y) d\omega_x d\omega_y = \alpha \iint \Phi_{vv}(\omega_x, \omega_y) d\omega_x d\omega_y \quad (21)$$

(19) 式の第4項は、符号化に起因する歪み成分である。(17) 式において、定義から

$$\frac{1}{4\pi^2} \iint \|N_{i\pm 1}^\theta(\omega_x, \omega_y)\|^2 d\omega_x d\omega_y = D_1(\theta) \quad (22)$$

が成立することから、以下の関係が求まる。

$$\frac{1}{2} \cdot \frac{1}{4\pi^2} \iint \Phi_{n\theta n\theta}(\omega_x, \omega_y) d\omega_x d\omega_y = \frac{1}{4} D_1(\theta) \quad (23)$$

4. 数値シミュレーション

3章で構築した理論モデルを用いて、数値計算ソフトウェア MATLAB 上でシミュレーションを行った。(3) 式の相関係数 ρ は 0.93 とした。(21) 式の、多眼画像の複雑さを表すパラメータ α については、0.00, 0.01, 0.05 の3つの値を、視差の精度を表すパラメータ β については、0, 1, 2, 5 の4つの値を用いた。 $\alpha=0.00$ は、被写体が極めて単純 (奥行き一定のランバート反射平面) の場合に、 $\beta=0.0$ は、視差が完璧に推定できる理想的な場合に相当する。

符号化パラメータ θ を変化させながら、レートと歪みを計算することにより、図2(a)-(c) のようなグラフを得た。これらのグラフにおいて、横軸はビットレート (bit/pixel)、縦軸は SN 比 (dB) を表す。“Ref” は、画像の間引きを行わない場合 (3.1 の1番目のシナリオ) のレート-歪み曲線であり、(1), (2) 式によって求められた。その他の曲線は、画像を間引いた場合 (3.2 の2番目のシナリオ) のレート-歪み曲線であり、(5), (19) 式に基づいて得られた。なお、(c) では、 $\beta=0$ と $\beta=1$ の曲線がほとんど重なっている。これらの曲線が、“Ref” の曲線を上回る場合には、カメラの間引きによって符号化効率が向上することを意味する。

(a)-(c) に共通して、低レートでは間引きをしたほうが、高レートでは間引きをしないほうが SN 比が高くなる傾向が見出された。したがって、平均的な歪み

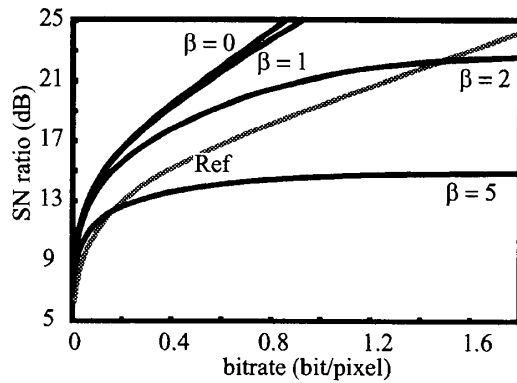
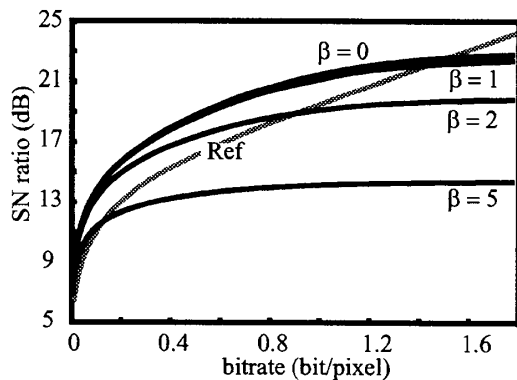
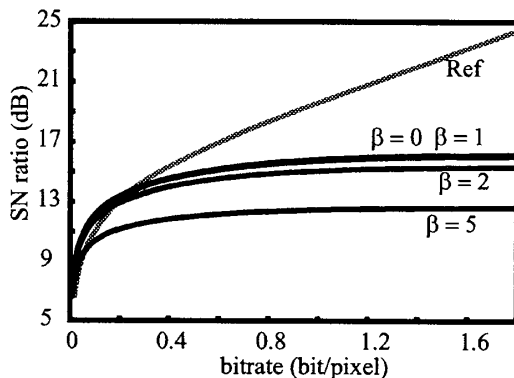
(a) $\alpha = 0.00$ (b) $\alpha = 0.01$ (c) $\alpha = 0.05$

図 2: 画像の間引きとレート-歪み特性

を最小化する観点では、レートに応じてカメラ数を削減することが有効である。ただし、カメラを間引きが有利/不利になる分岐点は、多眼画像の性質を表すパラメータ α 、および視差の精度を表すパラメータ β に依存する。 α 、 β のどちらも、小さいほど、間引きが有利になる方向に働く。逆に、特に α が大きい場合には、間引きが有利になるのは、超低レートのごく狭い範囲に限られる。

5. まとめ

本稿では、レート-歪み理論に基づき、多眼画像の間引きと符号化特性の関係を示す一理論モデルを提案した。水平方向に等間隔に配置された多眼カメラを想定し、すべてのカメラ画像を符号化・伝送する場合と、

カメラを一つおきに間引いて符号化・伝送し、受信側で間引かれた画像を予測・復元する場合を比較した。数値シミュレーションにより、被写体の形状・光学的性質が単純な場合、および視差推定の精度が高い場合には、間引きによって低レートでの符号化特性が向上する場合があることが示された。これは、別の見方をすると、帯域や被写体の複雑さに応じて、最適なカメラの配置密度が異なることを意味している。

今後は、本稿の議論を画像間の予測を用いる符号化方式にも拡張するとともに、実写画像を用いた検証実験を実施したい。また、本稿の議論を敷衍することで、多眼画像の取得と画像群の符号化を不可分なものとして捉え、空間情報そのものの符号化と考える、新しい枠組みの提唱に結び付けたい。

謝辞: 本研究を進めるに際して熱心な議論をしてくださった、東京大学の原島 博 教授に謝意を表します。

参考文献

- [1] M. E. Lukacs, "Predictive coding of multi-viewpoint image sets," IEEE ICASSP'86, pp. 521-524, 1986.
- [2] T. Takano et al., "3D space coding using virtual object surface," Systems and Computers in Japan, 32, 12, pp. 47-59, 2001.
- [3] M. Magnor et al., "Multi-view coding for image-based rendering using 3-D scene geometry," IEEE TCSVT, 13, 11, pp. 1092-1106, 2003.
- [4] K. Yamamoto et al., "Multi-view video coding using view-interpolated reference images," Proc. Picture Coding Symposium (PCS2006), 2006.
- [5] K. Mueller et al., "Multi-view video coding based on H.264/MPEG4-AVC using hierarchical B pictures," Proc. Picture Coding Symposium (PCS2006), 2006.
- [6] "Description of core experiments in mvc," ISO/IEC JTC1/SC29/WG11 N7798, Jan. 2006.
- [7] B. Wilburn et al., "High performance imaging using large camera arrays," Proc. ACM SIGGRAPH 2005, pp. 765-776, 2005.
- [8] 高橋ほか, "空間共有通信のための多眼カメラレイ構築に向けた基礎検討", VR 大会, pp. 475-476, 2005.
- [9] 藤井ほか, "大規模実世界データベース構築のための多元多点計測装置の開発", 3次元画像コンファレンス, pp. 41-44, 2006.
- [10] T. Berger, "Rate distortion theory," Englewood Cliffs, NJ: Prentice-Hall, 1971.
- [11] 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.
- [12] B. Girod, "The efficiency of motion-compensating prediction for hybrid coding of video sequences," IEEE Journal SAC, SAC-5, 7, pp. 1140-1154, 1987.
- [13] P. Ramanathan, B. Girod, "Rate-distortion analysis for light field coding and streaming," EURASIP Journal SP:IC, 21, 6, pp. 462-475, 2006.
- [14] K. Takahashi, T. Naemura, "How does subsampling of multi-view images affect the rate-distortion performance?," submitted to IEEE ICIP2007.
- [15] J.-X. Chai et al., "Plenoptic sampling," Proc. ACM SIGGRAPH'00, pp. 307-318, 2000.