

LI\_001

# 遮蔽領域の再現性を向上させた視点依存信頼度マッピング法

## View-Dependent Probability Mapping Method Improving Fidelity of Occluded Regions

國田 豊† 上野 雅浩† 田中 敬二†

Yutaka Kunita Masahiro Ueno Keiji Tanaka

### 1 はじめに

本稿では、多視点画像からシングルかつロバストに高画質な任意視点画像を生成する手法を提案する。

信頼度マッピング法 [1, 2] では、被写体は多層の2次元レイヤにより表現され、レイヤの各画素は被写体表面の色と存在確率の情報を持つ。この手法は、高速性と画質のバランスが良く、撮影から提示までの実時間処理が実証されている。しかし、被写体がオクルージョン(遮蔽領域)を含むような場合、画質が著しく低下するという問題があった。グローバルな座標系におけるボクセルの色と存在確率を推定し、ボリュームレンダリングを施す方法も提案されている [3, 4] が、やはり同様の欠点を持つと考えられる。

そこで本稿では、信頼度マッピング法を発展させ、オクルージョン領域でも再現性が高い新たな画像生成方法を提案する。

### 2 信頼度マップ

#### 2.1 原理

まず、最も基本的な画像生成の原理は次のような式で表される。

$$c_V = \sum_{c \in \Gamma} p(c)c. \quad (1)$$

ここで、 $c_V$  は目的となる画素の色ベクトルであり、 $[r, g, b]^T$  で表される。 $\Gamma$  は  $c_V$  がとりうる色ベクトルの集合であり、 $p(c)$  は  $c_V = c$  となる確率を示す。すなわち式 (1) は、目的となる色が、とりうる色の集合における期待値により表されることを示している。ここで  $\Gamma$  は

$$\Gamma_{\text{all}} = \{[r, g, b]^T | 0 \leq r \leq 1, 0 \leq g \leq 1, 0 \leq b \leq 1\}. \quad (2)$$

のようにすべての色空間とすることもできるが、以下で述べるようにカメラの幾何的な拘束を利用すると、より限定することができる。

図1に示すように、元となるカメラを  $\{C_i | i \in N\}$ 、その撮影画像を表す色ベクトルを  $I_i(m)$  とする。 $m \equiv [u, v]^T$  はデジタル座標で、 $\tilde{m} \equiv [u, v, 1]^T$  はその拡張ベクトルである。そして、 $u = u'/w', v = v'/w'$  として  $I_i(m)$  と  $I_i(\tilde{m})$  とは同じとみなす。同様に、目的となる仮想カメラを  $C_V$ 、注目する画素を  $m_V$ 、生成する画像を  $I_V(m) (= I_V(\tilde{m}))$  とする。被写体上の点  $P$  を撮影している  $m_V$  の対応点は  $C_i (i \in N)$  におけるエピポーラ線上にあり、これを  $l_i (i \in N)$  とすると色の候補はこの線上にあると考えられる。

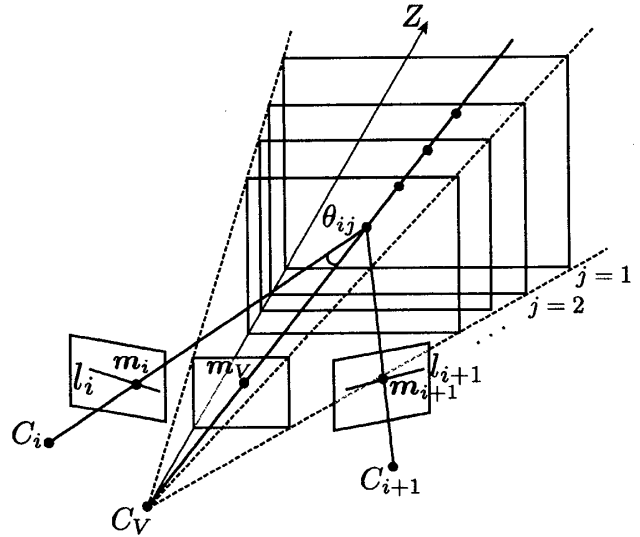


図1 提案手法における座標系の設定

すると、式 (1) における  $\Gamma$  は仮想カメラの位置  $m_V$  の画素に対応するエピポーラ線上の色の集合に限定でき、

$$\Gamma = \{I_i(m_i) | m_i \in l_i, i \in N\}. \quad (3)$$

さらに、被写体上の点は、離散的な平面  $\{L_j | j \in M\}$  上にあると仮定する。すると対応点が特定でき、

$$\Gamma = \{I_i(m_i) | \tilde{m}_i = H_j^{V_i} \tilde{m}_V, i \in N, j \in M\}. \quad (4)$$

ただし  $H_j^{V_i}$  はレイヤ  $L_j$  により規定されるホモグラフィ行列であり、 $m_V$  を  $m_i$  に写す。

一般に、レイヤ  $L_j$  を介してカメラ  $C_a$  の点  $m_a$  をカメラ  $C_b$  の点  $m_b$  に写すホモグラフィ行列は、以下のようにならされる。

$$H_j^{ab} = A_b \left( R + \frac{tn^T}{d_j} \right) A_a^{-1}. \quad (5)$$

ここで、カメラ  $C_a$  の座標系において、レイヤ  $L_j$  の正規ベクトルを  $n$ 、原点からの距離を  $d_j$  で表している。 $A_a$  と  $A_b$  はそれぞれカメラ  $C_a, C_b$  の内部行列である。 $R, t$  はカメラ  $C_a$  の座標系における点  $X_a$  をカメラ  $C_b$  の座標系における点  $X_b$  に変換し、 $X_b = RX_a + t$  なる関係がある。

ここで、レイヤ  $L_j$  を介した  $m_V$  のカメラ  $C_i$  における対応点を  $c_{ij}$  とすると、 $c_{ij} = I_i(H_j^{V_i} \tilde{m}_V)$ 。そして、 $m_V$  における色が  $c_{ij}$  である確率  $p(c_{ij})$  は、被写体がレイヤ  $L_j$  上にある確率  $p(j)$  と、被写体がレイヤ  $L_j$  上にある条件でのカメラ  $C_i$  の寄与率  $p(i|j)$  の積として以下のように表される。

$$p(c_{ij}) = p(j)p(i|j). \quad (6)$$

† 日本電信電話(株)NTTサイバースペース研究所, NTT Cyber Space Laboratories, Nippon Telegraph and Telephone Corporation

すると、式(1)は以下のように表される。

$$c_V = \sum_{j \in M} \sum_{i \in N} p(j) p(i|j) c_{ij}. \quad (7)$$

さらに、式(7)における画素単位の表現を書き直し、画像単位の表現とすることで、以下の式を得る。

$$I_V(\mathbf{m}) = \sum_{j \in M} \sum_{i \in N} q_j^V(\mathbf{m}) w_{ij}^V(\mathbf{m}) I_{ij}^V(\mathbf{m}). \quad (8)$$

ここで式(7)中の  $p(j)$ ,  $p(i|j)$ ,  $c_{ij}$  は式(8)中ではそれぞれ  $q_j^V(\mathbf{m})$ ,  $w_{ij}^V(\mathbf{m})$ ,  $I_{ij}^V(\mathbf{m})$  に対応する。またこのとき  $I_{ij}^V$  はレイヤ  $L_j$  に投影され、ターゲットカメラ  $C_V$  により撮影されたカメラ  $C_i$  の画像であり、

$$I_{ij}^V(\mathbf{m}) = I_i(\mathbf{m}_i), \quad \tilde{\mathbf{m}}_i = H_j^{V_i} \tilde{\mathbf{m}}. \quad (9)$$

なお、 $I_{ij}^V(\mathbf{m})$ ,  $q_j^V(\mathbf{m})$ ,  $w_{ij}^V(\mathbf{m})$  の右上の添字 "V" はカメラ  $C_V$  の画像座標系における関数であることを示している。

このように、色と確率が世界座標系で統一の表現を持たず、視点依存性を持つことが、本手法の特徴の1つとなっている。

## 2.2 信頼度マップとカメラ寄与率の推定

$q_j^V(\mathbf{m})$  はレイヤ  $L_j$  上に被写体が存在する確率を表す信頼度マップである。この値は、ソースカメラの対応点の一致度により推定される。

まず、 $[r, g, b]^T$  のすべての色要素における分散値ベクトルを以下のように計算する。

$$\delta_{ij} = I_{ij}^V(\mathbf{m}) - \frac{1}{\text{Num}(N)} \sum_{i \in N} I_{ij}^V(\mathbf{m}). \quad (10)$$

そして、要素の合計を計算し、以下のスカラー値を得る。

$$\sigma_j^2 = \frac{1}{\text{Num}(N)} \sum_{i \in N} \delta_{ij}^T \cdot \delta_{ij}. \quad (11)$$

ここで、 $\text{Num}(N)$  は集合  $N$  に含まれる要素数を表すものとする。

値を平滑化し、ノイズの影響を除去するために、以下のフィルタ処理をする。

$$\tilde{\sigma}_j^2 = \frac{1}{\text{Num}(B)} \sum_{\mathbf{m}' \in B} \sigma_j^2(\mathbf{m} + \mathbf{m}'), \quad (12)$$

ここで  $B$  はフィルタのブロック領域である。 $\tilde{\sigma}_j^2$  はカメラの対応点の色が一致するほど小さい値をとる。

$q_j(\mathbf{m})$  は一致度が高いほど大きな値をとり、その値は0から1までの値であり、さらに重なり合う点の合計値は1となる ( $\sum_j q_j(\mathbf{m}) = 1$ )。このような  $q_j(\mathbf{m})$  は、例えば以下のように得ることができる。

$$q_j(\mathbf{m}) = \text{norm} \frac{1}{\sum_j (\tilde{\sigma}_j^2)^n + \epsilon} \quad (13)$$

ここで  $\text{norm}$  は正規化演算を表し、 $\text{norm}_j f_j(x) = f_j(x) / \sum_j f_j(x)$ 。  $n$  は  $q_j(\mathbf{m})$  の急峻さを決定する調整用パラメータであり、 $\epsilon$  は0で除算することを防ぐための微小値である。

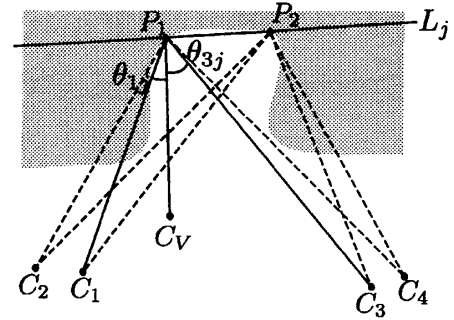


図2 オクルージョンを含むシーン

次に、 $w_{ij}^V(\mathbf{m})$  を推定する。 $w_{ij}^V(\mathbf{m})$  は、視点依存テクスチャマッピング [5] のカメラ重みパラメータに相当する。近接する光線ほどその色情報が近いという前提により、 $PC_V$  と  $PC_i$  のなす角度  $\theta_{ij}$  により重みが決定し、例えば

$$w_{ij}^V(\mathbf{m}) = \text{norm} \exp(-\tau \theta_{ij}^V(\mathbf{m})), \quad (14)$$

となる。ここで  $\tau$  は調整用パラメータである。

## 3 オクルージョンの対策

### 3.1 方針

ここまで、すべてのカメラ  $C_i (i \in N)$  を用いて画像を生成する方法について述べた。しかし、シーンが不連続な奥行きを持つ場合、あるカメラからは撮影できて、あるカメラからは撮影できない箇所がある。このような現象は一般にオクルージョンと呼ばれ、例えば図2において、カメラ集合  $\{C_1, C_2, C_3, C_4\}$  を用いてある1つの奥行きを特定する奥行きマップを用いる場合、奥行きを推定を失敗する可能性がある。

そのような場合でも、奥行きが多義性を許す信頼度マップを用いると、奥行きマップを用いる場合よりは滑らかな画像が生成されることが期待できる。しかしそれでも、奥行きが不連続な箇所では生成画像にぼけが生じ、画像が劣化する。

また、たとえ正しい奥行きが推定できた場合であっても、 $P_1$  における色を決定する際に、カメラ  $C_1, C_2$  における対応点の色を用いると、色の再現性が悪くなる。

そこで、カメラを部分集合に分け、それぞれの部分集合ごとに視点依存の推定を行い、最終的に仮想視点で統合することでより信頼性の高い推定を行うことを提案する。

例えば図2において、 $P_1$  はカメラ部分集合  $\{C_1, C_2\}$  を用いて推定すると信頼度が低いが、カメラ部分集合  $\{C_3, C_4\}$  を用いると信頼度が高まる。逆に、 $P_2$  はカメラ部分集合  $\{C_1, C_2\}$  を用いて推定すると信頼度が高く、カメラ部分集合  $\{C_3, C_4\}$  を用いると信頼度が低い。そして、仮想視点  $V$  からの信頼度に統合し、それぞれのカメラ部分集合の推定のうち、信頼度の高い部分の影響が大きくなるようにする。

### 3.2 部分信頼度マップの推定と統合

以下では、カメラ部分集合ごとの信頼度の推定法について述べる。カメラ部分集合はそのメンバの中に基準となるカメラを持ち、基準カメラの座標系においてレイヤ

$L_j$  における信頼度  $q_j^k(m)$  の推定を行う。  $C_k$  を基準とするカメラ部分集合は一意で、これを  $\Xi_k = \{C_i | i \in N_k\}$  とおく。

図2の場合、 $\Xi_1 = \{C_1, C_2\}$  と  $\Xi_3 = \{C_3, C_4\}$  なる2つのカメラ部分集合があり、 $C_1$  と  $C_3$  がそれぞれ基準カメラとなる。すべての部分集合における信頼度マップは共通のレイヤ  $L_j$  上で推定される。

まず、カメラ部分集合  $\Xi_k$  における基準カメラと他のカメラとの不一致度は、SSAD(Sum of Square Absolute Differences) により、以下のように計算される。

$$\text{SSAD} = \sum_{i \in N_k} \sum_{m' \in B} |I_{ij}^k(m + m') - I_k(m + m')|. \quad (15)$$

ここで、 $I_k$  は基準カメラ  $C_k$  における画像、 $I_{ij}^k$  は平面  $L_j$  を投影面としてカメラ  $C_i$  の画像を基準カメラ  $C_k$  の座標系に投影変換した画像を示す。すなわち、

$$I_{ij}^k(m) = I_i(m_i), \quad \tilde{m}_i = H_j^{ki} \tilde{m}. \quad (16)$$

そして、カメラ部分集合  $\Xi_k$  により推定される基準カメラ  $C_k$  における信頼度  $q_j^k$  は、

$$q_j^k(m) = \text{norm}_j \frac{1}{(\text{SSAD})^n + \epsilon}. \quad (17)$$

ここで、 $|x|$  は  $x$  の L1 ノルムをあらわす。また、 $n$  急峻さを調整するパラメータであり、 $\epsilon$  は 0 で除算するのを防止する微小値である。

すべてのカメラ部分集合  $\Xi_k, k \in K$  において信頼度マップ  $q_j^k$  を求めたら、次に仮想カメラ  $C_V$  における統合処理を行う。F まず、仮想視点への座標変換を以下のように行い、仮想カメラ  $C_V$  から見たカメラ部分集合  $\Xi_k$  の信頼度マップ  $q_{kj}^V$  を得る。

$$q_{kj}^V(m) = q_j^k(m_k), \quad \tilde{m}_k = H_j^{V_k} \tilde{m}. \quad (18)$$

次に、図3に示すように、全ての部分集合の信頼度マップ  $q_{kj}^V$  を掛け合わせて、さらにすべての投影面について正規化処理を行い、統合的な信頼度マップ  $q_j^V$  を得る。すなわち、

$$q_j^V(m) = \text{norm}_j \prod_{k \in K} \{q_{kj}^V(m) + \text{bias}\}. \quad (19)$$

ここで  $\text{bias}$  は  $\{q_{kj}^V | k \in K\}$  のどれかが極めて 0 に近い数である場合に、掛け算結果が 0 に丸まらないためのバイアスに相当する値である。

### 3.3 カメラ寄与率マップの決定

このようにして  $q_j^V(m)$  を求めたあとは、カメラ寄与率マップ  $w_{ij}(m)$  を求める。式(14)のように、光線の角度をもとにして求めることも可能ではあるが、オクルージョンのあるシーンの場合、2つの光線  $C_i P$  と  $C_V P$  のなす角度が小さくても、2つの光線の間に遮蔽物がない保証はない。図2に示すように、 $\theta_{1j} < \theta_{3j}$  であるが、 $C_1$  からは  $P$  は  $Q$  に遮られており、画像を撮影できていないため、 $w_{1j}$  を大きくするのは誤りである。そこで、

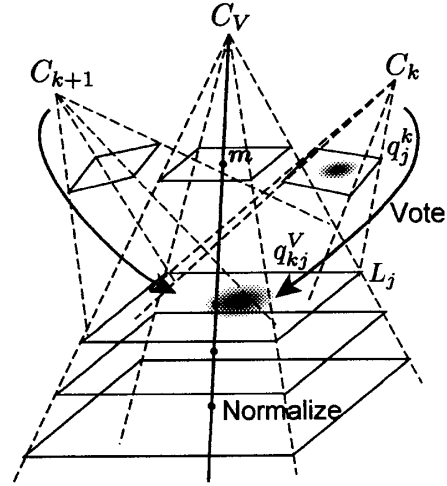


図3 視点依存信頼度の投票・正規化処理

着目する点において、高い信頼度を持つカメラ部分集合の基準カメラほど高い寄与率を割り振る。

$$w_{ij}^V(m) = \begin{cases} \text{norm}_k q_{kj}^V(m) & k = i, \\ 0, & \text{if } i \notin K \end{cases} \quad (20)$$

式(20)について更に論ずる。ある点において信頼度が高いことはその点が存在し、かつ遮蔽されずに撮影している可能性が高いことを示している。また逆に、信頼度が低いことは、その点が遮蔽されているかそもそも点が存在しない可能性が高い。図2において、 $q_j^3(m_3)$  は光線  $l_3$  上にて正規化されており、 $P$  においては極大値をとることが予想される。一方で  $q_j^1(m_1)$  は光線  $l_1$  上にて正規化されており、 $Q$  において極大値を持ち、 $P$  においては小さな値をとることが予想される。よって、式(20)によって  $P$  における重みは  $C_3$  が大きく、 $C_1$  が小さく割り振られる。つまり、 $C_1$  について言えば、手前の  $Q$  における大きな値の存在により  $P$  における重み  $w$  が抑制されていると言え、これは一種の可視性チェックとして働く。

### 3.4 レンダリング

このように、 $q_j^V(m)$  および  $w_{ij}^V$  が求めれば、式(8)により仮想視点  $V$  からの画像が生成できる。

式(21)のように重みと射影画像を掛け合わせることで、

$$T_j^V(m) \equiv \sum_{i \in N} w_{ij}^V(m) I_{ij}^V(m), \quad (21)$$

すると、式(8)は以下ようになる。

$$I_V(m) = \sum_{j \in M} q_j^V(m) T_j^V(m). \quad (22)$$

ここで、 $T_j^V(m)$  はレイヤ  $L_j$  の色を表すと考えられる。

また、以下のように特定のレイヤに効果を与えることもできる。

$$T_j^V(m) = \text{Filter}(T_j^V(m)) \quad (23)$$

例えばモザイク (Pixelize) フィルタを用いると、特定のレイヤの解像度を落とし、情報の詳細度を制御することができる。

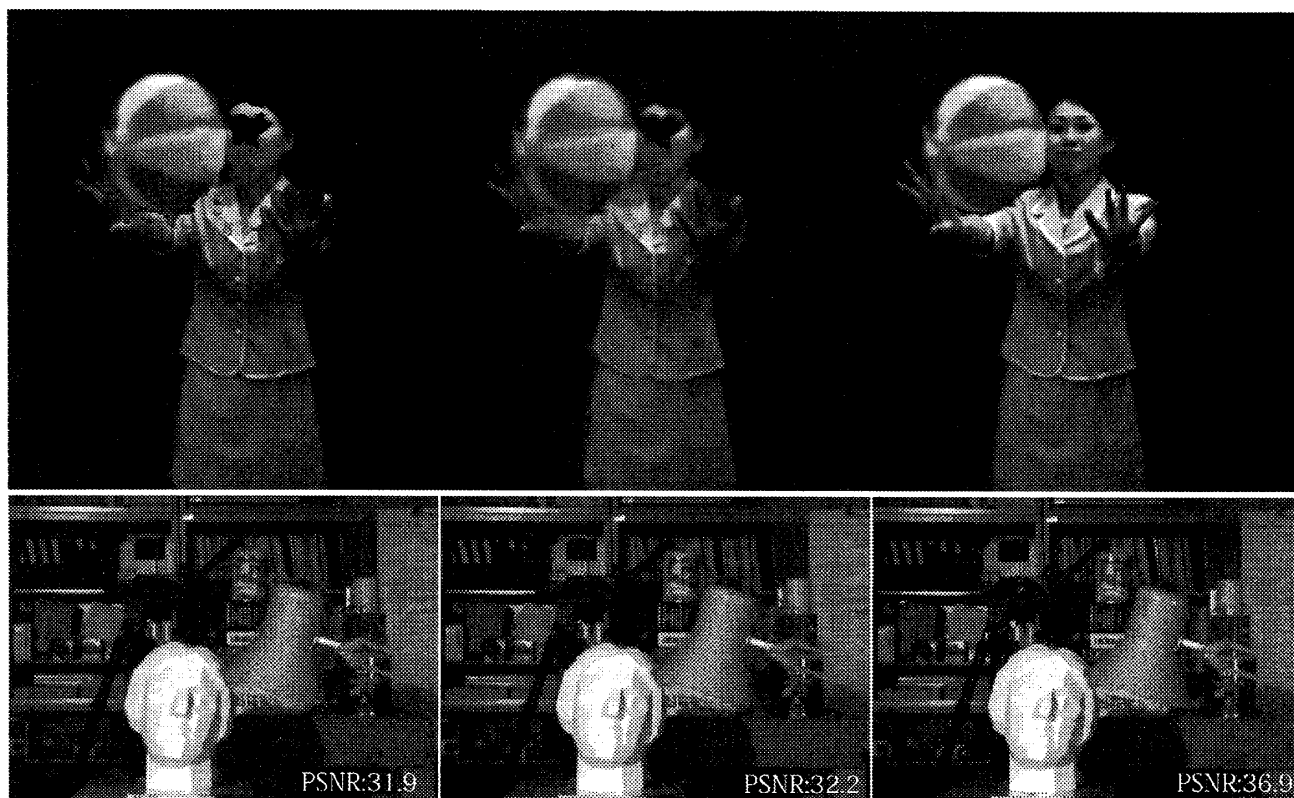


図4 画像生成結果. 左列:すべてのカメラをいちどに用いて奥行きマップを求めた場合. 中央列:すべてのカメラをいちどに用いて信頼度マップを求めた場合. 右列:視点依存信頼度マッピング法(提案手法)

#### 4 画像生成結果

まず、半径 12cm の円周上に配置した 8 台のカメラにより、中心の位置から見た画像を生成した。図 4 上列にその結果を示す。シーン中、人物の手前に紙風船が存在し、すべてのカメラから人物の顔が撮影できているわけではない。そのため、すべてのカメラをいちどに用い、奥行きマップ(左)や信頼度マップ(中央)を求めた場合は画像に欠損が生じていることがわかる。一方で、近接する 5 つのカメラをカメラ部分集合として本稿で提案した手法を用いると、オクルージョン領域においても非常に高い再現性で画像生成できることを確認した。

次に、筑波大学による多視点画像データセット”ohta”を用い、定量的な比較を行った。上下左右 4 つの元画像から真ん中の画像を合成し、実際の画像との差を PSNR で評価した。視点依存信頼度マッピングにおいては、近接する 3 つのカメラを部分集合としている。奥行きマップを用いた方法が 31.9dB、従来の信頼度マッピング法が 32.2dB に対し、本稿で提案した方法は 36.9dB と、大幅な画質の改善を達成した。

#### 5 まとめと将来展望

視点依存信頼度マッピングにより高品位な任意視点画像が生成できることを示した。一方で、条件分岐や反復的な演算を含まないシンプルなアルゴリズムであるため、高速処理に向いていると言える。そのため、今後は実時間処理を可能とするシステムを実装したい。

#### 参考文献

- [1] 國田, et al. 信頼度を利用した多層平面レンダリング法の提案. 3次元画像コンファレンス 2004, pp. 135-138.
- [2] 國田, et al. 多層信頼度マップを利用した3次元映像の実時間生成システム. *IMPS2005*, pp. 61-62.
- [3] J. S. De Bonet and P. Viola. Poxels: Probabilistic voxelized volume reconstruction. *ICCV '99*, pp. 418-425.
- [4] 山本, 奥富. ボリュームレンダリングを利用したイメージベースレンダリングの試み. 画像電子学会誌, Vol. 30, No. 4, pp. 371-378, 2001.
- [5] P. E. Debevec, et al. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. *SIGGRAPH '96*, pp. 11-20.