

不十分な計測情報に基づく実物体の写実的な画像合成

山崎 俊太郎 (指導教官: 池内 克史)
東京大学

実世界の風景や物体を計測した情報を利用して、仮想空間で物体の画像を写実的に合成するためには、対象物の三次元形状(幾何情報)とテクスチャ(光学情報)が一定量以上必要である。ところが、実際には多くの場合、計測装置の精度の制限や、計測コストや容量等の現実的な制約から、必ずしも十分な計測情報が利用できるとは限らない。そのため、計測が困難な対象物や環境では写実的な再現を行うことができず、手法の適用範囲を制限してきた。本論文では、このような問題を克服するために、三つの相補的な手法を提案する。

Photo-realistic Rendering of Real-world Objects Based on Insufficient Measurement

Shuntaro Yamazaki (Supervisor: Katsushi Ikeuchi)
University of Tokyo

In order to display real scenes and objects in a virtual space by synthesizing photo-realistic virtual views based on the measured data sets, sufficient amount of geometric and photometric information of the objects is required. In practice, however, it is often unfeasible to take sufficient measurements to achieve photo-reality due to the limitation on sensor accuracy, storage size and other practical restrictions. In this dissertation, we present three mutually complementary methods designed to solve the problem described above.

1 はじめに

1.1 研究の背景

本論文では、実世界の風景や物体を仮想空間で再現することを目的とし、計測によって得られた情報を利用して、対象の仮想3次元視点画像を写実的に合成するための手法を提案する。このような研究の応用は、電子商取引における電子カタログ、遠距離通信における視覚情報伝達、仮想現実感モデル、娯楽など、非常に広い分野が含まれる。

計測情報に基づく描画法は過去に数多く行われてきたが、実際に利用する際には次に挙げるような問題点からその利用範囲が制限されてきた。まず第一に、実際に計測を行う場合の現実的な問題として、計測回数の制限から計測情報の量が不足し、描画が困難になる場合がある。多くの場合、対象物の計測は一度しか行うことができず、描画結果に問題が発生したとしても、追加の計測を行うことができない。また、記憶容量の制限から計測回数を制限せざるを得ない状況も考えられ、このような場合には描画の解像度や描画視点の自由度、写実性が犠牲となる。また第二に、計測の技術的な限界として、対象物によっては正確な計測ができず、これにより正確なモデルを構築することができない場合がある。例えば、CCD(Charge-Coupled Devices) 素子を用いた光学セ

ンサーを用いて計測を行う場合、毛髪などの微細な形状を計測する時には解像度が不足するため正確な計測ができない。あるいは同様に、透明物体の形状を通常の光学センサーで計測することは困難である、などの問題が挙げられる。

1.2 論文の構成

1.1節で述べた問題を克服するために、三つの相補的な手法を提案する。

一つ目の方法である自動モーフィングに基づく視点画像補間法(2節)では、描画対象の幾何情報が全く利用できない場合に、形状の再構築を行うことなく、画像のモーフィング処理によって自由視点画像を合成する。まず、事前知識なしに、与えられた2次元画像間の密な対応を推定し、この対応点に基づき画像を変形・混合することによって自由視点画像を生成する。画像を対応付けするための特徴量の抽出手法として、人間の初期視覚で用いられているものと類似した、回転偏微分ガウシフィルタを用いたフィルタバンクを提案する。また、このようにして得られた画像特徴の対応付けを行うために、局所解を避けながら目的関数を最小化する手法を提案する。提案法では、透明物体やテクスチャのない均質な物体に対しても、写実的な視点画像補間が可能である。

二つ目の手法である Pop-up Light Field 法 (3 節) では、一つ目の手法と同様に幾何形状が利用できず、さらに入力画像から付加的な情報を得ることが難しい場合、利用者による対話的な幾何モデル生成を用いて、写実的な表示を実現する。提案法では、取得した計測情報を基に Light Field 表示を行い、現実感を損なっている部分に対して層状構造 (Coherent Layers) 化処理を行うことによって描画画質を改善する。この処理は利用者が対話的に行うことが可能で、そのためのユーザーインターフェースを提案する。また、Layer 化によって生じる、対象物の輪郭線付近におけるエイリアスを防ぐために、視点画像間で一貫性のあるアンチエイリアシングを実現する Coherence Mating 法を提案する。提案法では、Layer 化を繰り返すことにより、利用者の目的に応じた写実性を持つ自由視点画像の合成が可能である。

三つ目の手法である Microfacet Billboarding 法 (4 節) では、光学情報に加えて粗い幾何情報も利用し、既存の計測装置では取得が困難な微細な形状を持つ対象物に対しても写実的な描画を実現する。まず、得られた幾何情報の信頼度を計算し、精度に応じて対象物の形状を、大域的、局所的 (視点依存) モデルの組み合わせとして表現する。大域的形状に基づき視点依存の微小面集合を用いて対象物の形状を近似表現し、視点依存距離マップ、視点依存テクスチャマップを用いて詳細な描画を行う。また、テクスチャ画像の透過度推定を行うことにより、光学センサの解像度以下の形状を扱うことができ、本手法は、樹木や毛髪といった形状計測が困難な物体に対しても適用可能である。

1.3 関連研究

仮想 3 次元視点画像を描画するためには、描画対象に関して 2 つのモデルを構築する必要がある。ひとつは幾何情報を表す形状モデルであり、計測情報を用いる際には、レーザー光を用いた形状計測、ステレオ視や視体積交差法などによる画像からの立体復元などによって得られる情報が利用される。もうひとつは色 (光学情報) を表す光学モデルであり、カメラ計測された 2 次元カラーが用いられることが多い。

これらの 2 種類の情報量に関して、写実的な描画を行うために必要な量が Chai らによって理論的に解析されており [9]、その下限は幾何情報と光学情報の空間において図 1.3 に示すような曲線で表されることが知られている。写実的な描画を実現するためには、利用可能な情報量が、グラフ中で曲線の上側にある必要があり、このとき様々な入力の組み合わせに適した描画法が提案されてきた。今論文では紙面の都合上、個々の手法に関する説明は省略する。

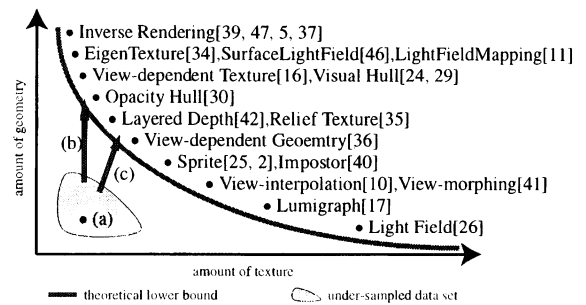


図 1: 計測情報に基づく描画法の関連研究。横軸に、幾何 (形状) 情報、縦軸に光学 (写真) 情報の量を示し、写実的な描画に必要な情報量の下限をグラフで示す。本論文では、計測情報が不足している場合に、(a) 自動画像モーフィングにより見た目に違和感の少ない自由視点画像を合成する手法、(b) 利用者の簡単な作業により幾何情報を増やす方法、(c) 視点依存モデルと透過度推定を用いた描画法、の 3 つの相補的な手法を提案する。

2 自動画像補間による視点画像補間

2.1 導入

単一物体を複数視点から計測した画像が入力として与えられたとき、画像の各画素間における対応関係が分かると、対象物に遮蔽がない限り補間によって中間視点における仮想視点画像を合成することが可能である。このような複数の画像間の対応点の計算 (対応点問題) は、コンピュータビジョンの主要な研究課題のひとつである。

対応点問題に対する既存の多くの研究では、対象物の特徴 (色分布の線形性など)、画像間の関係 (カメラの外部パラメータが既知、対象物の変化が微小など)、計測カメラの特性 (カメラの内部パラメータが既知、反応曲線が同一など) が既知であるという拘束を用いて対応点を計算する。従って、例えば、計測カメラがノイズを含む場合、レンズ歪みが未知の場合、対象物が特徴を持たない場合などには、対応点を計算する事が困難である。本節では、2 枚の画像間の補間画像を生成することを目的とし、画像に対する事前知識無しに非拘束で対応点を計算する手法を提案する。

入力画像中の物体が未知の場合、画像間の対応点計算を行うためには次の 2 つが問題となる。第一に、画像間の対応を決定するための特徴量を定義する必要がある。特に、画像の対応点間での輝度が同一であるという保証がない場合、“自然な”対応を定義する必要がある。第二に、対応点計算の際に、局所解を避けるための工夫が必要である。本節では、自然な対応を得るために人間の初期視覚で用いられているものと同様のフィルタを用いて画像の

特徴を抽出する。また局所解を避けるために、対応写像の幾何的な拘束を考慮した非線形最適化手法を提案する。

2.2 関連研究

対応点計算の手法は、様々な対象物、計測環境の組み合わせに対して数多く提案されてきた。この章では、画像間の密な対応を計算する手法を簡単に紹介する。

Optical flow 画像間の視差が小さく、また対応点間で画像輝度に変化しない時、Optical flow を計算することによってカメラまたは対象物の移動を検出できる [3]。Optical flow はその原理から、画像間の視差が小さく、また輝度変化が線形近似できる場合にのみに適用可能な手法である。また、画像輝度に変化しないという拘束により、光源の変化を伴う場合や、画像がノイズを含む場合に推定が不正確となる。

画像に関する知識を利用する方法 対象物や計測カメラに関する情報があらかじめ得られている場合、対応点の計算を低い次元で行うことが可能である。例えば、ステレオ視による3次元復元的手法では、カメラの姿勢からエピソード拘束を導くことにより各対応点の探索範囲を1次元にすることが可能である [6]。また、顔画像の対応付け [14] や、医療画像の位置あわせ [27] 手法では、各画像が類似の形状を持っており、輝度分布に一定の規則性を仮定できることから、テンプレートの当てはめによって問題を簡単化することが可能である。本手法ではこのような事前知識なしに対応を計算する。

非拘束画像マッチング法 画像の輝度に関する拘束のみを使って対応点を計算する手法として、非拘束画像マッチング法が品川らによって提案されている [43]。この手法では Critical Point Filter (CPF) と呼ばれる非線形フィルタを用いて画像特徴を抽出し、多重解像度計算によって目的関数を最小化することで対応点を計算する。この方法は画像間の輝度分布が似ている場合には有効であるが、実際の画像では必ずしもその仮定は満たされないため、対応の精度が低い。また、入力画像の視差が大きい場合には、画像の一部分において対応点計算が局所解に陥り、それによって周辺への対応が大きく歪むという問題がある。また、CPF の特徴として画像輝度のピークに非常に強く反応するため、ノイズを含む画像では結果が大きく歪む。提案法は、輝度の変化やノイズに対して頑健な特徴抽出法を提案し、局所解に陥りにくい計算手法を提案することによって、視点画像補間に適した非拘束画像マッチング法を提案する。

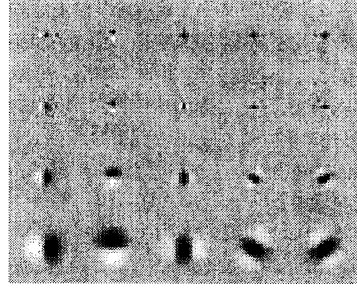


図2: 特徴抽出に用いたフィルタバンク。

2.3 対応点問題の定式化

\mathbb{D} を画像座標空間、 \mathbb{V} を画像輝度空間、 $I: \mathbb{D} \rightarrow \mathbb{V}$ を画像の輝度分布関数とする。このとき、画像間の対応点問題は、画像 I_{src} から画像 I_{dst} への写像 $m: \mathbb{D} \rightarrow \mathbb{D}$ のうち、 $I_{src}(m(\cdot))$ と $I_{dst}(\cdot)$ との差 ϵ_D^2 が最小となるようなものを見つけること、として定式化できる。

多くの場合、画像は離散的な画素の2次元集合として与えられるが、画素を補間して再サンプルすることにより、 \mathbb{D} をさらに高解像度で定義できる。また、写像 m は全単射とする。即ち、 m^{-1} は \mathbb{D} の各点で一意的に定義できるものとする。

以降では、議論を単純化するため、2次元のグレースケール画像の対応点計算の手法を説明する。提案法は、より高次の、例えばカラー2次元画像、またはグレースケール3次元画像(ポリウム)などへの適用も、自明な拡張によって容易に実現可能である。カラー画像への拡張は2.6節で述べる。

2.4 目的関数の構成

2.4.1 特徴抽出

本研究では、多くのコンピュータビジョンの手法の仮定と異なり、画像中の輝度は必ずしも画像間で一定ではない状況を想定している。そのため、画像輝度を直接、特徴量とすることはできない。その上でさらに、対応点は人間の視覚的に違和感のないものである事が望ましい。

このような考察から、人間の初期視覚で用いられる特徴抽出フィルタと類似した、回転偏微分ガウスフィルタによって特徴抽出を行う。このフィルタのカーネルは次の式で定義される。

$$K_{\sigma, n, \theta}(x, y) = G_{\sigma}^{(n)}(x \cos \theta - y \sin \theta) \times G_{\sigma}^{(0)}(x \cos \theta + y \sin \theta) \quad (1)$$

ここで $G_{\sigma}^{(n)}$ は分散 σ^2 の1次元ガウス関数の n 階微分である。

Malik らはこのフィルタを組み合わせるフィルタバンクを構成して画像から特徴量を抽出する手法を提案している [28]. ただし複数のフィルタが線形従属性を持つ場合、フィルタの反応に冗長性が生じて計算効率が下がる。そこで、多くの異なる θ, σ, n の組み合わせに対するカーネルに対し主成分分析を適用し、カーネル空間の次元を決定して無駄のないフィルタバンクを構成する。実験では、 n 階の偏微分フィルタは $n+1$ 次元空間を構成するという結果 [20] に基づき、図 2 に示すような 4 種類の異なる分散を持つ、1 階および 2 階フィルタから構成されるフィルタバンクを用いた。

画像の対応計算の際には、まず I_{src}, I_{dst} に対し、 N_f 種類 (実験では $N_f = 20$) の異なるカーネル $K_{\sigma, n, \theta}$ から構成されるフィルタバンク $F = \{F_i\} (i = 1, \dots, N_f)$ を適用する。このフィルタの反応は N_f 次元のベクトル画像であり、これを特徴ベクトル画像と呼ぶ。以降の処理では、特徴ベクトル画像のみを用いて計算を行う。

2.4.2 目的関数の構成

I_{src} から I_{dst} への最適写像 m は、 $I_{src}(m(\cdot))$ と $I_{dst}(\cdot)$ の類似度を m の滑らかさを保ちながら最大化することにより計算する。そのために、特徴ベクトル画像間の類似度と m の微分から構成される目的関数を構成する。

まず、類似度に関する基準として、次の ϵ_D^2 を用いる。

$$\epsilon_D^2 = \int_{p \in D} \sum_{i=1}^{N_f} \rho_D \left((F_i * I_{src})(p) - (F_i * I_{dst})(m(p)) \right) \quad (2)$$

ここでオペレータ $*$ は畳み込み積分であり、 ρ_D は M 推定量 [18] である。入力画像の視差が大きい時、2.5 節で述べる反復計算における初期対応点の精度は低いことが予想される。M 推定量を用いることによって、精度の低い初期解から始めたときに局所解に陥る可能性を減らすことができる。

実験では、M 推定量として次の式で表される Lorentzian を用いた。

$$\rho(x) = \log \left(1 + \frac{x^2}{\sigma^2} \right) \quad (3)$$

ただし ρ_D は、 $F_i * I$ の各要素に対する $\sigma = \sigma_D$ の Lorentzian の和である。

次に写像の滑らかさの基準として、次の ϵ_S^2 を用いる。

$$\epsilon_S^2 = \int_{p \in D} \sum_{d=1}^{dimD} \rho_S \left(\nabla m_d(p) \right) \quad (4)$$

ここで ρ_S は、 ∇m_d の各要素に対する $\sigma = \sigma_S$ の Lorentzian の和である。入力画像に遮蔽がある場合、対応写像には不連続性が生じる。M 推定量を用いることにより写像の不連続性を扱うことが可能となる。

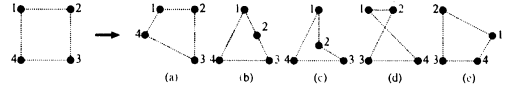


図 3: 隣接 4 点からなる四辺形の対応候補。

式 (2) と式 (4) を重み付け係数 α^2 を用いて線形結合することにより、最終的に最小化する目的関数 e^2 は次のようになる。

$$e^2 = \epsilon_D^2 + \alpha^2 \epsilon_S^2 \quad (5)$$

本手法で利用している特徴量は N_f 次元の情報を持つため、多くの場合 ϵ_D^2 だけを用いて対応点の候補を絞る事が可能である。 ϵ_S^2 は特徴ベクトルから写像を決定できない場合、即ち画像輝度分布が一定の領域における写像を確定するために用いられる。従って、一般に α^2 は非常に小さい値でよい。また、抽出された特徴ベクトルは対応点を決定する唯一の情報であるため、 σ_S は大きめに取る必要がある。 σ_D は実験的に決定する必要があり、実装では画像サイズの 5% となるように設定した。

2.5 目的関数の最小化

入力画像間の差が大きい場合には、初期推定の精度が低くなるために、最適な α^2 を選択することは非常に困難である。この章では、写像が凸性を保つという拘束を考慮することによって、計算する課程で写像の大きな歪みを防ぎ、比較的小さい α^2 でも滑らかな拘束が得る方法を提案する。

I_{src} の隣接する 4 格子点なる正方形が、写像 m によって I_{dst} に向って変形するとき、 m の全単射性から、変形後の 4 格子点の位置関係は図 3 の通りである。四辺形内部における全単射性を満たすためには、(c) と (d) を候補から除外する必要がある。現実には、これらの写像は、例えば画像が遮蔽領域を持つ場合に存在しうるが、四辺形が非常に小さくつぶれる場合、大きく広がる場合を考慮することによって、(c) と (d) の場合を無視しても画像間の対応は矛盾なく決定できる。次に、(b) の変形が存在すると、図中の頂点 2 を含む隣接四辺形における可能な写像が制限され、全体として写像が局所解に陥りやすい。そこで (b) も目的関数の最小化の際に候補から除外する。以上より、推定の各段階において (a) の場合の写像のみを考慮することにより、目的関数とは別に、最小化の過程において写像の滑らかさを保証することができる。この拘束を凸写像拘束と呼ぶ。

全体として凸写像拘束を満たす写像 m を決定するためには、局所的な写像の推定の際に凸写像拘束が満たされるようにすれば十分である。これは次のようにして実現できる。図 4 のように、 I_{src} 中の点 $p_{x,y}$ が、 m に

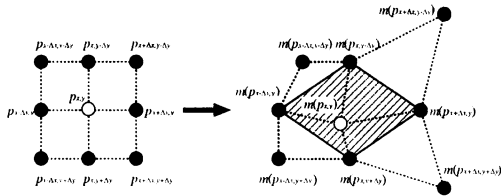


図 4: 凸写像拘束を満たすための対応点の探索範囲.

よって I_{dst} に対応付けられる場合、 p の対応点の候補を $m(p_{x+\Delta x, y}), m(p_{x, y+\Delta y}), m(p_{x-\Delta x, y}), m(p_{x, y-\Delta y})$ からなる四辺形の内側に限定すればよい.

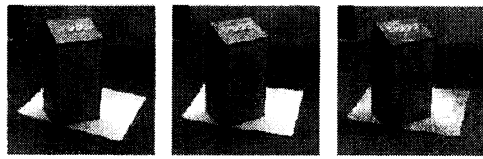
2.6 実装

凸写像拘束によって導入された写像の隣接対応点の幾何的な依存関係のため、画像中の各点の写像を同時に推定することは困難である。そこで、我々は品川らの手法 [43] を拡張し、反復法による逐次最適化を用いる。まず I_{src} 中の一点 p に対し、 I_{dst} 中で $m(p)$ として選択可能な全ての点に関して目的関数を計算し、最も関数値が小さい点を $m(p)$ とする。この操作を I_{dst} 中の全ての点に関して繰り返すことにより、画像全体に対する写像 m を推定する。また、 p の選択順序による写像の偏りを防ぐために、 p は乱数を用いて選択する。

さらに、局所解を防ぐために、目的関数 e^2 の最小化を多重解像度計算で行う。対応点計算をする前に、特徴ベクトル画像をダウンサンプルして解像度を下げ、ピラミッド構造を構築しておく。写像の推定時には 2×2 サイズの画像から計算を開始し、各解像度で目的関数が最小化されたときの写像を、一つ上の解像度における推定の初期値とする。これを必要な解像度になるまで繰り返す。実装では、各解像度における推定には反復法を用い、最初の局所最適解をその解像度における最適解とした。

また画像がカラー情報を持つ場合には、2.4.1 節で述べた特徴抽出フィルタを各カラーチャネルに対して適用して、特徴ベクトル画像を生成した。即ち、RGB 画像の場合、R, G, B の各チャネルに対し N_f 次元のフィルタバンクを適用し、全体として $3N_f$ 次元の特徴ベクトル画像を生成した。

現在の実装では、 e^2 の積分計算は、 \mathbf{D} を離散的にサンプリングして和を取ることで行っている。サンプリング幅は画像の画素サイズと一致する必要はなく、解像度を上げると精度が向上するが、計算時間が増す傾向にある。実験では、画素幅を 1.0 としたとき、サンプル幅を 0.1 から 0.5 程度に設定した。



(a) I_{src} (b) I_{dst} (c) 対応点

図 5: 非拘束画像マッチングの例



(a) mean (b) CPF (c) 提案法

図 6: 画像輝度の変化に対する頑健性の比較

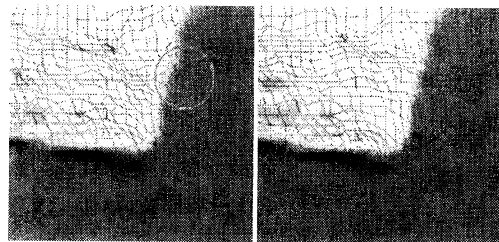
2.7 実験結果

2.7.1 非拘束画像マッチング

図 5 は、回転している物体の画像間における推定写像の結果である。画像解像度は 128×128 である。(a) と (b) が入力画像で、(c) が提案法で推定された写像である。

図 6 では画像の輝度に変化ある場合の推定の頑健性を比較した。安価なカメラで自動ホワイトバランスで計測を行う場合には、このような揺らぎを考慮する必要がある。実験では図 5(a) と、輝度を 10% 明るくした (b) の間の対応点を推定した。平均値フィルタを使った推定法、品川らの手法 [43] による推定法、提案法による推定法で対応点を計算し、trilinear 補間によって中間視点画像を生成して結果を比較した。輝度の変換性を仮定する既存手法では、わずかな輝度変化で写像が大きく歪むのに対し、提案法では中間視点画像で写実性が失われていないことが分かる。

図 7 では、写像の滑らかさに関して比較を行った。図 5



CPF 提案法

図 7: 推定写像の滑らかさの比較

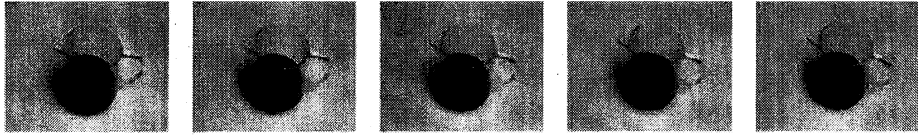


図 8: 透明物体に対する視点画像補間の結果

で用いた画像間の推定写像において、品川らの手法 [43] で推定した写像は部分的に裏返っているのに対し、提案法では滑らかさが維持されていることが分かる。

2.7.2 視点画像補間

推定された対応写像を基に、入力画像を trilinear 補間して中間画像を生成した結果を示す。

図 9 は 2 枚の画像を用いた視点画像補間の結果である。(a) と (b) が入力画像、(c) が画像モーフィングによって生成された中間視点画像である。入力画像の内部、外部カメラパラメータは未知である。対応点は 256×256 解像度で推定した。(d) は、(c) と同一の視点で実際に計測された画像であり、(c) の合成には用いていない。合成画像はほとんどの部分で、実際の計測画像と似ているが、遮蔽が存在する部分では対応する点が存在しないために、補間画像中に本来存在しない色が現れている。この問題を改善するためには、推定した対応写像中で、隣接対応点が近い場所では I_{src} を、離れている場合には I_{dst} を重み付けして補間画像を生成するなどの工夫が考えられる。本節では対応点計算の手法の提案を主眼としており、補間手法に関しては紙面の都合上これ以上踏み込まないものとする。

図 8 に、透明物体に対する視点画像補間の例を示す。この物体は、透明で、テクスチャがほとんど存在せず、強い鏡面反射成分を持ち、また蛍光灯下で自動ホワイトバランスを使って安価なカメラで撮影した。このような画像に対して、特徴点抽出に基づくカメラ姿勢の推定を行うことは困難である。また、カメラパラメータが未知であるため、ステレオ視による立体復元は現実的ではない。提案法では、画像の対応に基づく視点画像補間を行うため、このような画像に対しても写実性のある中間視点画像を生成することが可能である。ここで、推定される対応点は幾何的に正しい対応となっている保証はないが、合成される中間視点画像は人間の視覚にとって正しく見ることが期待できる。

2.7.3 性能評価

本節で提案した手法を Pentium3 1GHz、主記憶 768MB の PC で実装し、性能を比較した。表 1 は計算速度と対応

表 1: 性能評価。計算時間の単位は秒、誤差は画像輝度差の二乗和 ($\sum_p |I_{src}(p) - I_{dst}(m(p))|^2$) である。

data (size)	[43]		proposed method	
	time	error	time	error
boxes (128^2)	2.1	6.0×10^6	8.5	5.0×10^6
textures (128^2)	2.0	3.8×10^8	10.7	1.6×10^8
faces (256^2)	8.2	1.3×10^8	51.3	1.1×10^8

点の輝度の自乗差である。非拘束画像マッチング法の一つである品川らの手法 [43] と比較した場合、特徴量の次元が高いこと、また推定に反復法を利用していることから計算時間は増加している。一方、得られた写像はより画像輝度差の小さい対応を得ることができていることから、より大域的な最適解に近い推定ができていることが推定される。

2.8 まとめ

本節では、完全に自動で、画像に関する知識無しに、画像間の密な写像を推定する非拘束画像マッチング手法を提案した。提案法では回転偏微分フィルタを用いた新しい特徴量を提案し、滑らかな写像を推定するための凸写像拘束を提案した。既存手法と比べ、推定写像の精度が向上していることを示し、様々な物体に対して視点画像補間を行い手法の有効性を実証した。

3 Pop-Up Light Field 法

3.1 はじめに

2次元画像の補間によって仮想3次元視点画像を合成する場合、入力として大量の画像が必要となる。密な画像集合に基づく描画手法である光線空間法 [26] では、対象物の幾何形状が全くない状況下でも立体的な描画が可能であるが、一方で、入力画像が少ない場合には補間が不正確となり、合成された画像中にエイリアスと呼ばれる不具合が生じて写実性が損われる。

一般に、入力画像が不十分な場合、画像の空間解像度を下げるか [26]、2節で述べた方法や、ステレオ視や視

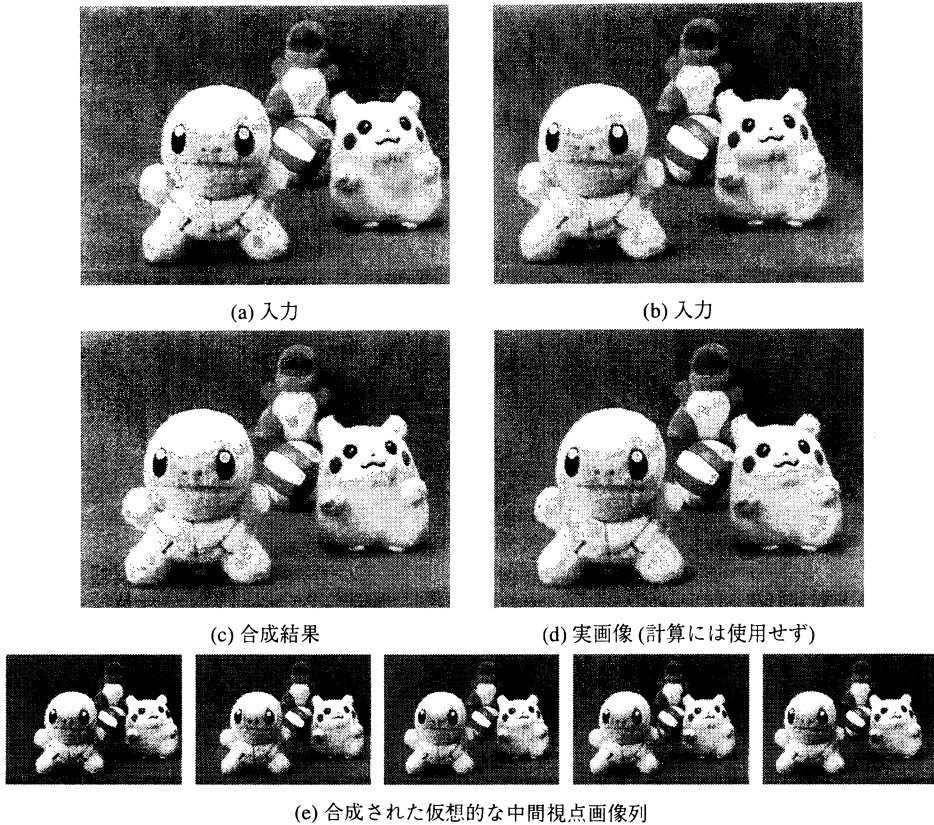


図 9: 視点画像補間の結果

体積交差法を用いて画像から対象物の形状を復元することで補間精度を向上させる [17, 7] などの工夫が必要である。ところが、画像の対応点計算に基づく補間では遮蔽があるときに精度が悪くなるという問題があり、必ずしも写実性の高い画像を合成できるとは限らない。このことが、実際にこれらの手法を用いる際の不安要素となり、現実的には手法の応用範囲を狭めている。

そこで、本節では、利用者による対話的な操作を用いることによって、疎な画像集合からエイリアスのない自由視点画像を合成するためのシステムを提案する。この手法では、画像からの形状復元では扱うことが困難であった画像の遮蔽領域 [23, 21] においても写実性を損なうことなく 3次元描画が可能である。

提案する手法の背景にある基本的な考え方は、疎な画像列が与えられた時、写実的な描画を行うためには、対象物の完全な形状を復元する必要はないという事である。提案法では、形状を復元する代わりに、対象物を Coherent Layer と呼ばれる多層構造に分解する。Coherent Layer は既存の Layer 構造 [25, 42, 2] と次のような点で

異なる。(a) Layer 数を利用者が対話的に決定する。(b) Layer 境界を、一部のキーフレーム (入力画像) で指定する。(c) Layer 形状として単純な平面を用いる。(d) Layer 間で α 混合を用いたアンチエイリアス処理を行う。

3.2 関連研究

対話的モデル生成 対話的なモデル生成を考慮した多くのシステムでは、1枚の画像を入力とし、対象物が単純な幾何構造を持つことを仮定している。例えば、Tour Into Pictures 法 [19] では透視投影における消失点が画像中に存在することを仮定し、クモの巣状のメッシュ構造のパラメータを利用者が決定する。また、Photo editing 法 [32] では、利用者が、画像の各画素に対して深度情報を割り当てる必要がある。これらの手法は単一画像に対しては適用可能であるが、複数画像が入力となる疎な光線空間に対して適用することが困難である。複数画像を用いて対話的モデル生成を行う方法として、Debevec らは Facade 法を提案している [16]。このシステムでは対象

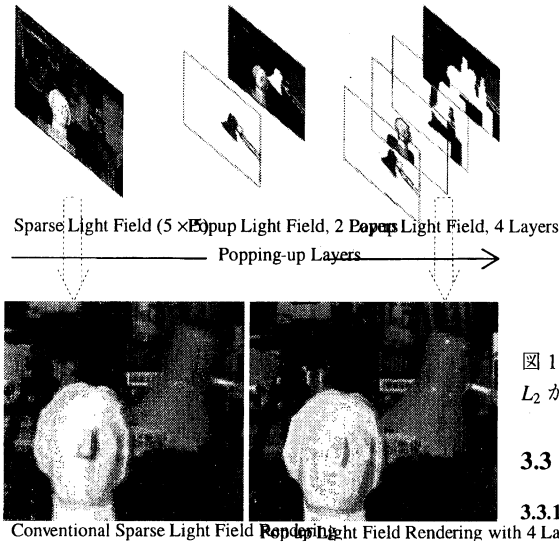


図 10: Pop-up Light Field 描画の例.

物を平面の集合として近似するが、全ての入力画像に対して平面境界を指定する必要があるため、モデル生成のコストが非常に高い。

Layer によるモデル化・描画 画像に基づく描画法は対象物の形状を Layer 構造を用いて表現することで、効果的に実現できる [25, 2, 42]. Layered Depth Image 法では、深度情報を持った Layer を効果的に表示するための 2 段階描画法を提案している。また、このような Layer を自動生成する手法について、コンピュータビジョンの分野で多くの研究がなされてきた [45, 2, 22]. これらの手法を用いても正しく Layer 境界を決定できない場合も考慮し、提案法では利用者による対話的な Layer 生成を利用する。

透過度推定 Layer 表現では、その境界においてアンチエイリアス処理を行うために、正確な透過度を推定する必要がある [42]. 画像から透過度を推定する最も古典的な方法は、固定の背景色 (例えば青) を用いる方法である [44]. 近年、任意の背景を持つ画像に対して透過度推定を行う手法が提案され [31, 4, 38], Chuang らによって Bayes 推定を使った精度のよい方法に拡張されている。また、この手法を複数画像に適用するために Video Matting 法 [12] が提案されている。ただしこの手法ではフレーム間の透過度の一貫性が考慮されていないという問題があり、自由視点画像合成での利用には適していない。

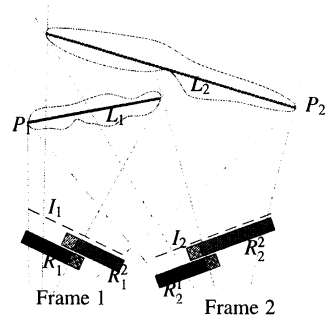


図 11: 画像 I_1 と I_2 から構成される Light Field を L_1 と L_2 から成る Coherent Layer で表現している。

3.3 Pop-up Light Field 表現

3.3.1 多層構造

光線空間のサンプル密度が不十分な場合、古典的な Light Field 法 [26] による描画ではエイリアスと呼ばれる不具合が発生する。図 10 上は 5×5 枚の画像から構成される光線空間を 1 つの補間平面を用いて描画した結果である。エイリアスによる画像の二重化現象が発生している。図 10 下は 4 つの平面で表現された多層構造を用いることによりアンチエイリアス描画を実現した例である。対象物を複数の層に分割することにより、各層における深度変化を小さくすることができ、画像の二重化現象を抑えることができる。

Pop-up Light Field を構成する Layer を *Coherent Layer* と呼ぶ。Coherent Layer の数や形状は対象の複雑度や入力画像の計測密度に応じて必要なだけ、利用者によって対話的に Pop-up されて生成される。

3.3.2 Coherent layer

Coherent Layer L_j は、光線空間を構成する各画像 I 中の対応する Layer 領域 R_j^I の集合として表現される。図 11 中の例では、これらの領域の幾何形状は単純な固定平面 P_j として表現されている。より複雑な対象を表現する場合には、各画像 I に対して定義された視点依存の平面 P_j^I を使うことも可能である。

Pop-up Light Field における Layer は次の二つを満たすときに Coherent と呼ぶ。

1. 各 Layer が単独でエイリアスなく描画可能。
2. 各 Layer が背後の Layer に対してエイリアスなく描画可能。

第一の条件を満たすためには、Plenoptic Sampling 理論 [9] から、Layer 内の深度分布は十分に小さくなければ

ばならない。これを満たすために、提案法では利用者による対話的なモデル生成を導入する。

第二の条件を満たすためには、Layer境界が画像間で正確でなければならない。Layer境界を画素サイズ以下の精度で表現するためには、境界における透過度の推定が必要である。また視点移動に伴い境界部分の画素を補間したときにエイリアスが生じないように、フレーム間で透過度は一貫している必要がある。これを満たすような透過度を推定する手法を3.3.3節で述べる。図11にCoherent Layerの例を示す。

3.3.3 Coherence matting

Coherent Layerの各画像中における領域境界での透過度を推定する単純な方法として、画像列に対してvideo matting法[12]を適用することが考えられる。Video Matting法はBayesian Matting法[13]に基づく透過度推定法であり、次に示す事後確率最大化(MAP)を行うことにより推定を行う。

$$\begin{aligned} & \arg \max_{F, B, \alpha} P(F, B, \alpha | C) \quad (6) \\ & = \arg \max_{F, B, \alpha} L(C|F, B, \alpha) + L(F) + L(B) + L(\alpha) \quad (7) \end{aligned}$$

ただし $L(\cdot) = \log P(\cdot)$ は対数尤度、 C は画像中で観測されたRGB色、 F, B, α はそれぞれ前景色、背景色、透過度である。Video Matting法やBayesian Matting法では $L(\alpha)$ は一様分布であるとして式(7)から省略する。

Video Matting法では、既知の F, B を持つ領域マップ(Trimap)を隣接画像へと伝播させることによって、透過度推定の処理の半自動化を可能にしている。ところが推定される透過度は各画像中の色分布のみによって決定されるため、画像中に遮蔽などが存在すると、隣接する画像の対応する点における透過度が大きく変化し、連続的に変化させた場合に視覚的な不具合を生じる。

この問題を解決するために、我々はCoherence Mattingと呼ぶ新たな透過度推定法を提案する。この処理の流れを図12に示す。まず、利用者が指定した領域境界を各入力視点画像に伝播させる。次に、各画像中で、境界線付近のがその集合を得ることによって透過度を推定する必要のある領域を確定する。続いて、隣接視点画像を可能な限り利用して、ある視点では遮蔽領域にあるため観測できなかった画素を、モザイク化によって埋める。これにより B を確定することが可能であり、式(7)から $L(B)$ を省略できる。最後に、式(7)を解くことによって F, α を推定する。このとき α がアンチエイリアスのための透過度であることから、Feathering関数

$$L(\alpha) = -(\alpha - \alpha_0)^2 / \sigma_\alpha^2 \quad (8)$$

を用いて α の事前確率を推定結果に反映させる。ただし



図12: Coherence Mattingの処理の流れ。(a)利用者が領域境界を指定。(b)透過度推定が必要な領域を前景と背景の近傍に追加。(c)別の視点画像を用いて背景モザイクを生成。(d)前景色と透過度を推定。

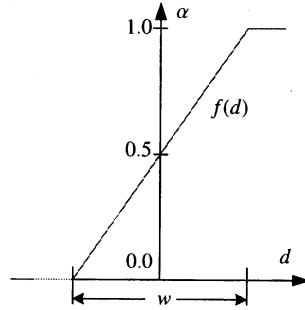


図13: coherence matting で用いる feathering 関数

$\alpha_0 = f(d)$ は d に関するFeathering関数で、 σ_α^2 は α の分散である。 d はLayer境界からの画素の距離である。実装ではfeathering関数として $f(d) = (d/w) * 0.5 + 0.5$ を用い、 w をパラメータとして調整した(図13)。

尤度関数 $P(C), P(F)$ がガウス関数で表現できるとすると、 C, F の局所性を仮定することにより近傍での値を用いて次のような式を構成することができる。

$$L(C|F, B, \alpha) = -\|C - \alpha F - (1 - \alpha)B\|^2 / \sigma_C^2 \quad (9)$$

$$L(F) = -(F - \bar{F})^T \Sigma_F^{-1} (F - \bar{F}) \quad (10)$$

ここで σ_C^2 は C の分散、 \bar{F} は近傍における前景色の重み付き平均、 Σ_F は同様に重み付き共分散行列である。式(7)の F と α に関する偏微分を0とおくと、次の式を得る。

$$F = \frac{\Sigma_F^{-1} \bar{F} + C\alpha / \sigma_C^2 - B\alpha(1 - \alpha) / \sigma_C^2}{\Sigma_F^{-1} + I\alpha^2 / \sigma_C^2} \quad (11)$$

$$\alpha = \frac{(C - B) \cdot (F - B) + \alpha_0 \cdot \sigma_C^2 / \sigma_\alpha^2}{\|F - B\|^2 + \sigma_C^2 / \sigma_\alpha^2} \quad (12)$$

式(11)と式(12)を使い、 F と α の一方を固定し、他方に関して最小化を行う、という処理を交互に繰り返すことによりMAP推定を行う。初期値として $\alpha = \alpha_0$ とした。

図14に推定された透過度の視点間における一貫性の比較を示す。図は α に関するEPI画像(α -EPI)である。通常のEPI[1]が特定の空間方向と時間方向に対する2次元の輝度変化を表示するのにに対し、 α -EPIでは特定の空間方向(図では水平方向)と視点変化の方向(図では水平方向)に対する透過度変化を表示している。

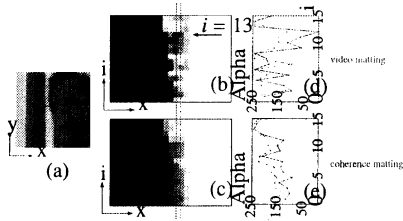


図 14: video matting と coherence matting の比較. (a) Plaza データ (図 18) の一部. (b,c) (a) 内の線分に対応する α -EPI. (d) video matting の結果. (e) coherence matting の結果.

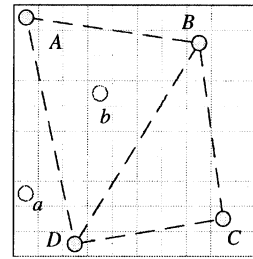


図 16: 近傍の視点画像の選択.

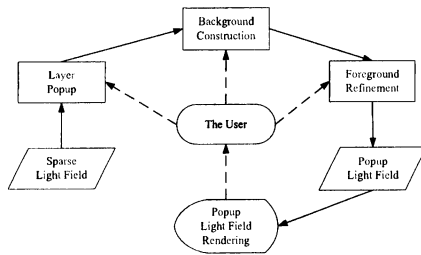


図 15: Flowchart of pop-up light field construction UI.

3.4 Pop-up light field の構築

Pop-up light field を生成するために、簡単に使える UI (User Interface) を設計した.

3.4.1 処理の流れ

Figure 15 は UI の概観である. 利用者は、入力である疎な光線空間を基に描画を行い、エイリアスの発生により写実性が損なわれている部分があった場合に、次の 4 つの処理を行うことで画質を改善する.

Layer pop-up: 利用者は光線空間を構成する画像集合の中からキーフレームを選択し、エイリアスの発生する画像領域の輪郭線、及び Layer 平面の幾何的なパラメータを決定する. この輪郭線を別のフレーム画像に伝播させることによって全ての入力画像で領域分けを行う. 詳しくは

Background construction: Coherence matting に必要な背景モザイクを生成する. 背景モザイクは通常、Layer の輪郭線の近傍でのみ生成すればよい. 平面による形状の近似精度がよい場合、この処理は自動化が可能である.

Foreground refinement: Coherence matting により前景色、及びその透過度を推定する. 透過度の性質上、視点から遠い Layer から順にこの処理を行う.

Rendering feedback: 以上の全ての変更は処理が終わり次第すぐに反映され、その結果が表示される. 表示された結果に対して描画視点移動を行うことによってエイリアスの発生を確認し、画像全体で写実性が十分に高まるまで以上の処理を繰り返す.

3.4.2 Layer 境界の生成

Pop-up Light Field の生成で最も複雑で時間のかかる部分は Layer 境界の設定である. そこで本節でその詳細を述べる.

まず、Layer の輪郭線は複数の制御点をつないだ環状の連結線分として表現する. この表現は、構造が直感的であると共に、輪郭線の変更が容易であるという特徴がある. また、同一の Coherent Layer の輪郭は、同じ位相をもつ、即ち同数の制御点から構成され、フレーム間で制御点が一対一に対応するものとする.

制御点の設定、及び編集はいくつかのキーフレーム中でを行い、他のフレームへは内挿、及び外挿補間によって伝播させる. 利用者によって変更された制御点はキー制御点と呼び、各フレームの輪郭線の制御点は、一部または全てがキー制御点であるとする.

あるフレーム画像中の制御点がキー制御点でないとき、その点の座標は近傍フレームの、対応するキー制御点の座標から計算する. Light Field が 1 次元のカメラ配列の場合、近傍フレームとして、最近傍 2 つのフレームを利用する. 2 次元のカメラ配列の場合は、キー制御点を含むフレーム画像の位置を Delaunay 三角形分割し、三角形の頂点にあるフレームのキー制御点を使って三角形内部にあるフレームの制御点を生成する. 図 16 では、 A, B, D が b の近傍フレームとなる. Delaunay 三角形の外にあるフレーム I_P の制御点を生成する際には、 $\angle I_0 I_P I_1$ が最小になる I_0, I_1 を選択して外挿する. 図 16 では、 A, D が a の近傍フレームとなる.

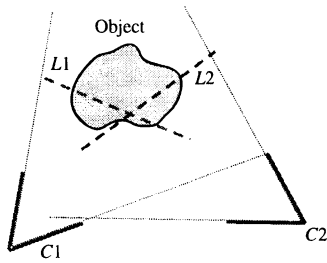


図 17: 視点依存 Layer

表 2: 描画性能

Data	Resolution (w × h × #cameras)	Orig. LF size (MB)	Pop-up LF size (MB (#layers))	FPS
Tsukuba	384 × 288 × 25	8.3	19.4 (5)	62.5
Plaza	640 × 486 × 16	14.9	38.3 (16)	58.8
Pokemon	512 × 384 × 81	47.8	117.8 (5)	31.3
Fur	1136 × 852 × 23	66.8	140.4 (5)	46.9
Statuette	1136 × 852 × 43	124.8	161.2 (4)	37.1

3.5 実時間描画システム

編集中の Pop-up Light Field を待ち時間なく表示するために、グラフィクスハードウェアを用いた実時間描画システムを構築した。具体的には、Unstructured Lumigraph Rendering 法 [8] を各 coherent Layer の描画に適用し、描画始点から遠い Layer から順に、 α 混合を行いながら描画を行うことで Pop-up Light Field を描画する。

3.6 実験結果

複数の光線空間データに対し Pop-up Light Field を生成し描画した結果を以下に示す。

提案するシステムを OpenGL、及びその拡張機能であるマルチテクスチャ、ピクセル単位照光を用いて実装した。実験で用いた PC は CPU が 660 MHz、主記憶が 768 MB、グラフィクスハードウェアが NVidia GeForce4 128MB であり、描画速度は表 2 の通りであった。表 2 に示すように、全ての描画は実時間で行うことができた。モデル生成に要する時間は対象の複雑さに依存し、実際のデータでは表 3 に示すとおりであった。

図 10 に 5 × 5 の Tsukuba データ¹ に対し 4 つの Coherent Layer を生成して描画した結果を示す。実験では、Coherent Layer を 7 つにすることによって 3 × 3 の画像ではほぼ同等のエイリアスのない描画が実現できることが分かった。

図 19 は比較的示唆の大きい 9 × 9 の Pokemon データに対する描画結果である。5 つの Coherent Layer を用い、3 つを個々の物体、2 つを床面と手前の面とした。背面

¹提供: Prof. Ohta of University of Tsukuba

表 3: 利用者のモデル生成に要した時間

Data	Tsukuba	Plaza	Pokemon	Fur	Statuette
# frame	25	16	81	23	43
# layer	4	16	5	5	4
points/frame	129	379	90	167	47
key points/frame	6.3	62.4	8.6	16.1	12.3
Time (hours)	≈ 0.5	≈ 4	≈ 1	≈ 1	≈ 1.5



図 18: 遮蔽やモーションブラーを含む対象物の描画結果。

は一番後方の物体と同じ layer で描画している。

図 20 は、視点に依存して Coherent Layer の方向を変化させる例である。この実験で用いた Statuette データはカメラの計測範囲が広いいため、固定面を用いると、ある方向で正しく表示できたとしても、別の方向でエイリアスが発生することがある (図 20(a)(b))。視点依存の平面を用いることによりこのような不具合を負は具ことができる (図 20(c))。

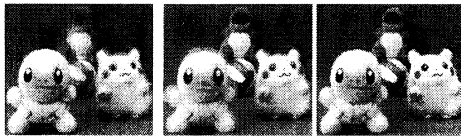
図 18 は “time-frozen カメラ” によって撮影された 16 枚の画像から生成された Plaza データ² に対する実験結果である。対象物は多くの遮蔽領域を持ち、モーションブラーなどの複雑な現象を含んでいるが、提案法によってエイリアスのない描画を実現できる。

3.7 まとめ

本節では、利用者による対話的なモデル生成を利用して疎な画像列で構成された光線空間に対し、エイリアスのない描画を実現する手法を提案した。

提案法は、利用者がモデルを作りこむことによって写実性を高めて行くことが可能であるが、入力が疎であることから、例えば鏡面反射などの視点依存して見え方が変わる物体の描画は困難である。また形状の近似モデルとして平面を利用しているため、例えば樹木のように平面による境域分けが困難な物体に対しては適用が難しい。また視線によって領域境界の位相が変化するような状況

²提供: Dayton Taylor



手前の物体に焦点 奥の物体に焦点 pop-up LF

図 19: Light Field 法と提案法の比較.



(a) (b) (c)

図 20: 示唆の大きいカメラ配置で計測した疎な Light Field. (a) 視線に垂直な固定面による描画. (b) (a)を別の視点から見た場合. (c) (b)の視点で視点依存面を用いて描画した結果.

に関しても提案法では対策していない.

4 視点依存の微小面を用いた表示手法

4.1 はじめに

コンピュータグラフィクスで3次元物体を表示する際に最も広く使われる手法は、物体表面を境界面表現し、表面の属性をテクスチャで与えて表示するモデルベース描画法である。この方法では、物体表面の幾何形状を精度良く計測するために、レーザーレンジファインダや、パターン光投影法等が用いられる。しかし幾何形状の取得は必ずしも容易ではなく、例えば毛髪や樹木のような細かく入り組んだ形状を持つ物体の場合、計測精度が不足するため、得られる形状は不正確、不完全である。また計測精度が撮影対象の表面特性に影響されやすく、ノイズを多く含むため描画に利用することができない。

このような対象物の描画の際には、幾何形状を粗く近似し、画像ベース描画法を用いて描画することで、表示精度を向上させることが可能である。View-dependent texture mapping 法 [16], Unstructured Lumigraph 法 [8], Opacity-Hull 法 [30] などでは、視点非依存(固定)の近似表面形状に対して視点依存テクスチャマッピングを行うことで描画を行う。しかしこれらの方法では、近似した表面と実際の表面のずれの方向と視線のなす角が大きくなる描画視点において、形状の近似誤差が描画結果に大きく影響し、現実感が著しく失われる問題がある(図 21, 左図)。

本節では、視線と常に垂直になるように視点に依存し

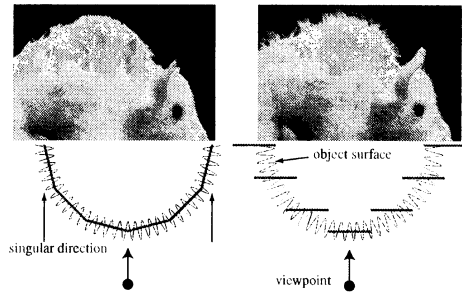


図 21: 幾何形状を微小面近似した場合の結果の比較. 左図: 固定微小面で近似する場合. 右図: 視点依存微小面で近似する場合.

て生成される微小平面の集合を用いて幾何形状を近似し、視点依存のテクスチャマッピングを行うことで描画を行う手法を提案する(図 21, 右図)。提案手法は、画像を用いた表示手法の特徴として、任意の対象に適用可能であると同時に、利用できる幾何形状を最大限に活用することにより、入力画像中に含まれていない視点での描画を精度よく行うことができる。また幾何形状の近似を多重解像度化することが容易であり、視点依存の詳細度制御が可能である。

4.2 視点依存微小面によるモデル生成と描画

4.2.1 計測データの取得

本手法では、センサーによる計測やステレオ視などによって得られた物体の粗い幾何形状と、複数視点から撮影されたカラー画像を入力とする。一般に、幾何形状が正確であるほど、またカラー画像の撮影位置の数が多いほど最終的な描画結果の画質は向上する。幾何形状とカラー画像の撮影位置の相対関係、およびカメラパラメータは既知とする。

4.2.2 幾何形状のボクセル近似

まず、与えられた幾何形状をポリウム空間で再サンプリングし、テクスチャを利用した描画処理に適したデータ構造に変換する。そのために、対象物体を完全に含むポリウム空間を考え、各ボクセルにおいて入力の幾何形状と交差判定を行って二値ポリウムを生成する。

4.2.3 微小面の生成

生成された二値ポリウム中で、表面と交差するボクセルの位置に微小面を発生させて描画を行う。生成する

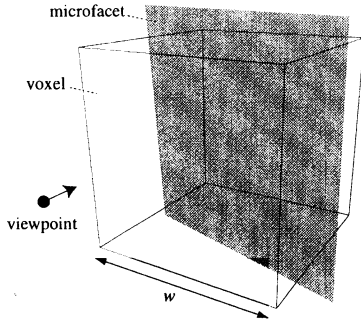


図 22: 物体表面と交差するボクセル内に微小面を発生させる。

微小面は表示対象の表面形状の近似であり、表示における方向依存性をなくすため、描画視点の変更に伴って方向と位置を変更する。

提案手法では視点からボクセル中心に向かう方向を視線方向とすると、微小面を視線方向と垂直になるように、視線に対して一定の間隔で生成する(図 22)。即ち、微小面の発生間隔を l とするとき、各表面ボクセルに対して、視点から nl (n は整数) の位置に、視線が面の中心と交差するように微小面を生成する。また、任意の視点で微小面がボクセル全体を被覆するように、ボクセルの幅を w とするとき、微小面は一辺の長さが $\sqrt{3}w$ の正方形とする。

4.2.4 テクスチャ画像

微小面は入力のカメラ画像を射影テクスチャマッピング (Projective texture mapping) [15] して描画される。カメラが複数ある場合には、視線方向とカメラの撮影方向のなす角 ϕ が小さいカメラから順に、カメラ画像中に微小面の占める領域が完全に含まれるものを選択し、画像を投影する。この処理はボクセル毎に行い、ボクセルに対する視線としては、視点からボクセル中心に向かうベクトルを用いる。

選択されるカメラは視点の変更とともに描画時に動的に決定されるため、特にカメラの撮影間隔が疎な場合、視点移動の際にカメラ画像が切り替わることによる画像の変化が大きく、これによって描画結果の現実感が損なわれる。そこで先の ϕ を用いてカメラ画像を近いほうから順に複数枚選択し、それらを角度に応じて内挿し、補間によって得られた画像をテクスチャとして投影することで、カメラ画像が切り替わる際の描画を滑らかに行うことができる。

また画素サイズ以下の幾何情報を表現するため、3 節で述べた手法と類似の透過度推定をあらかじめ全ての画

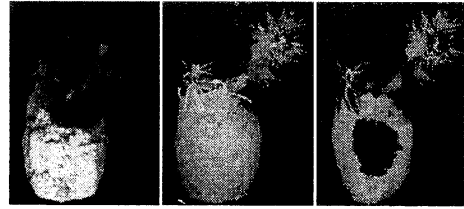


図 23: 入力画像(左図), 対応する距離画像(中央図), および距離に応じたテクスチャの切り抜きの例(右図)。

像に対して適用しておき、透過度も補間する。

4.2.5 距離情報を用いたテクスチャの切り抜き

微小面に対するテクスチャ画像の選択は、視線方向とカメラの撮影方向から決定される。そのため、複数の微小面で同一のテクスチャ画像が選択される可能性があり、これによりテクスチャ上の一つの画素が複数回描画されてしまう可能性がある。テクスチャの各画素は対象の一つの点から得られた情報であるから、これは描画時に同一の物体が複数現れる現象 (double imaging) として観測され、描画結果の現実感を損なう。

そこで、カラー画像中の各画素に対してあらかじめ描画対象までの距離を持たせておき、各画素の距離と描画対象の微小面の距離を比較して、それらが一定の閾値内に含まれる場合のみ表示することでこの問題を解決する。図 23 にその一例を示す。カメラ画像(左図)の各画素に、色情報の他に対象物体までの距離を記録しておく(中央図)、描画時にその値を微小面の位置と比較し、距離が一致する領域だけを切り抜いて表示する(右図)。

4.2.6 詳細度制御

描画の速度を決定するもっとも大きな要因のひとつは、微小面の数である。この数を減らすことにより描画速度を向上させることが可能であるが、同時に結果画像の画質が悪化することが考えられる。そこで、描画面上で対象物体が占める領域が小さく、画質への影響が無視できる範囲で微小面の数を減らし、最終的な描画効率を上げるといふ、視点依存の詳細度制御ができることが望ましい。

微小面の数を減少させる手段は 2 種類あり、ひとつはボクセル分割数を下げる方法であり、もう一つは微小面の発生間隔を上げる方法である。ただし後者に関しては、微小面の発生間隔を一定値以上に設定すると、微小面が生成されないボクセルが発生して結果画像に穴が観測される問題が生じる。

そこで、描画に必要な精度の変化に応じて、微小面の発生間隔を変更すると同時に、ボリュームの解像度も変更

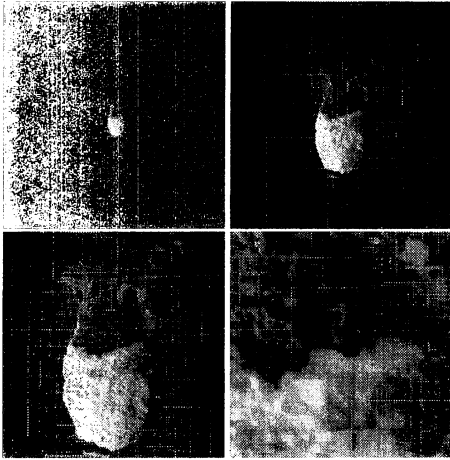


図 24: 視点移動に応じた微小面生成の詳細制御。

表 4: 描画速度

Size of volume	32 ³	64 ³	128 ³	256 ³
# facets	2460	11207	50560	208289
FPS	>30	>30	26.3	8.1

して詳細制御することを考える。ボリュームの解像度が $i \times i \times i$ の時のボクセル幅を w_i とすると、微小面の発生間隔 l を変更する際に、

$$l \in \left(\frac{1}{2} w_i \cos \phi, w_i \cos \phi \right] \quad (13)$$

を満たすような解像度 i のボリュームを用いて微小面を発生させることで、結果画像の現実感を損ねることなく詳細制御できる。

4.3 実験結果

実験に用いる幾何形状、カメラ画像はミノルタ社製 VIVID900 を用いて計測した。この装置を使用することにより、対象物体の距離画像とカラー画像を同一視点で計測でき、両者の相対的な位置合わせを行う必要がない。計測して得られた距離画像を三角形メッシュの集合に変換し、Neugebauer らの手法 [33] を用いて相対位置合わせを行うことで、カメラの相対姿勢を決定した。描画処理の実験に使った計算機環境は、PentiumIII 1GHz、主記憶 512MByte、グラフィクスカード GeForce4 Ti4600、ビデオメモリ 64MByte の PC である。

実験結果を図 25 に示す。毛で覆われた物体を回転台上に設置し、物体の周囲から 10° 間隔で、合計 36 回の計測を行った。計測したカラー画像と距離画像の解像度は共

に 640×480 である。入力幾何形状を 4.2.2 節で述べたアルゴリズムでボリューム化して得られた 64×64×64 サイズの符号付距離ボリュームを図の左列に示す。図中、赤で示されている部分が符号付距離が正の領域、緑が負の領域を意味し、図示のために不透明度を部分的に変化させて表示している。このボリューム中で表面と交差するボクセルに対して視点依存の微小面を発生させて幾何形状を近似し、対象物体を表示した結果が図の中央列、最終的な描画結果が図の右列である。ここに示されている画像は、隣接する 2 つの入力画像のカメラ視点の中間の位置で表示した結果であり、それら 2 つの画像と近似幾何形状から合成されている。また詳細制御は行っていない。結果画像中において、物体境界付近の幾何形状が不完全な部分も正確に表示できていることがわかる。また、この実験における描画速度は表 4 の通りであり、ほぼ対話的な速度で描画処理を行うことが出来る。

図 26 は入力カメラ位置の中間に視点がある場合の見え方を従来法と比較した結果である。データの計測は対象物体の周囲から 10° 間隔で合計 36 回行い、それらを半分に間引いた 18 組の距離画像とカラー画像を用いて表示した結果と、そこに含まれていないカラー画像と比較した。左図が、結果描画の合成には用いられていないカラー画像である。同一の視点において、従来法の一つである固定のポリゴンモデルに対して視点依存テクスチャマッピングを行って表示した結果を中央図、提案手法で表示した結果を右図に示す。従来法では、微小面の近似誤差が描画結果に大きく影響しているが、視点依存の微小面を使うことにより現実感ある描画が可能になることが分かる。

4.4 まとめ

本節では、層状に生成される視点依存の微小面と視点依存テクスチャマッピングを利用して、複雑な表面を持つ実物体を表示する手法を提案した。この手法は正確な幾何形状を取得、または表現することが困難な物体に対して適用可能であり、毛皮、樹木など従来手法では扱うことが難しかった対象物を精度良く描画することができる。

提案手法では、対象物の幾何形状を不連続な微小面の集合で近似することから、特に入力画像が少ない場合には結果画像に不連続性が現れる可能性がある。これを防ぐ手法としては、テクスチャの切り抜き幅を大きめに設定する、隣接するカメラ画像を用いて亀裂を埋める、などの方法が有効であるが、このことにより逆に結果画像中の不連続性が大きくなり、double imaging が増加する可能性がある。また、遮蔽などの理由により表示する領域のテクスチャが存在しない場合は、結果画像中に穴が現れる。

視点依存の詳細制御に関しては、ボリュームを階層

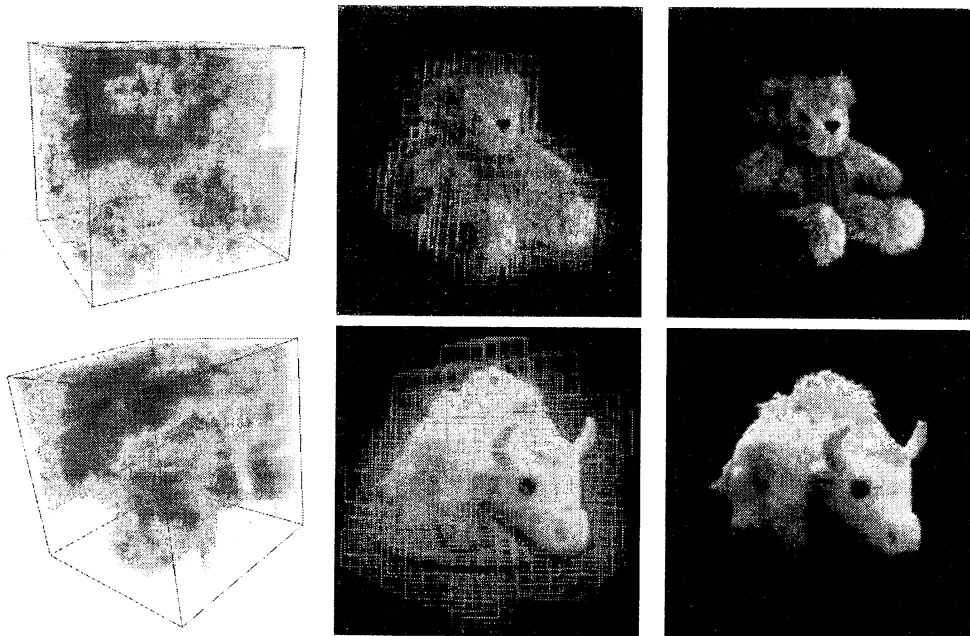


図 25: 描画結果: (左列) ボリューム化された幾何形状, (中央) 視点依存微小面, (右列) 最終結果.

化することで実現したが、これをオクツリーに拡張することで、入力形状の部分毎に必要な詳細度を評価し、微小面の発生密度を適応的に変化させることが可能である。また、描画結果に寄与しない微小面を描画しないように工夫することにより、描画効率をさらに上げることも期待できるが、これらは将来の課題とする。

5 考察

本論文では、計測情報に基づき、実物体の仮想 3 次元画像を合成する際に、計測が不十分であっても写実性の高い表示を行うための手法について、3つの基礎的な方法を提案した。今後は、Pop-up Light Field の Coherent Layer 上での画像補間、あるいは Microfacet Billboarding 法の Microfacet 上での画像補間など、各手法の統合化、またはこれら 3 手法を統一的に扱う枠組みの構築が課題である。

参考文献

- [1] H.H. Baker and R.C. Bolles. Generalizing epipolar-plane image analysis on the spatiotemporal surface. *International Journal of Computer Vision*, 3(1):33–39, 1989.
- [2] S. Baker, R. Szeliski, and P. Anandan. A layered approach to stereo reconstruction. In *Proc. CVPR '98*, pages 434–441, 1998.
- [3] Steven S. Beauchemin and John L. Barron. The computation of optical flow. *ACM Computing Survey*, 27(3):433–467, 1995.
- [4] A. Berman, A. Dadourian, and P. Vlahos. Method for removing from an image the background surrounding a selected object. In *U.S. Patent 6,134,346*, 2000.
- [5] Samuel Boivin and Andre Gagalowicz. Image-based rendering of diffuse, specular and glossy surfaces from a single image. In *Proc. SIGGRAPH 2001*, pages 107–116, 2001.
- [6] Myron Z. Brown, Darius Burschka, and Gregory D. Hager. Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):993–1008, 2003.
- [7] Chris Buehler, Michael Bosse, and Leonard McMillan. Non-metric image-based rendering for video stabilization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2001.
- [8] Chris Buehler, Michael Bosse, Leonard McMillan, Steven J. Gortler, and Michael F. Cohen. Unstructured lumigraph rendering. In *Proc. SIGGRAPH '01*, pages 425–432, 2001.
- [9] Jinxiang Chai, Xin Tong, Shing-Chow Chan, and Heung-Yeung Shum. Plenoptic sampling. In *Proc. SIGGRAPH 2000*, pages 307–318, 2000.
- [10] Shenchang Eric Chen and Lance Williams. View interpolation for image synthesis. In *Proc. SIGGRAPH '93*, pages 279–288, 1993.
- [11] Wei-Chao Chen, Jean-Yves Bouguet, Michael H. Chu, and Radek Grzeszczuk. Light field mapping: Efficient representation and hardware rendering of surface light fields. *ACM Transactions on Graphics*, 21(3):447–456, 2002.
- [12] Yung-Yu Chuang, Aseem Agarwala, Brian Curless, David H. Salesin, and Richard Szeliski. Video matting of complex scenes. In *Proc. SIGGRAPH 2002*, pages 243–248, 2002.
- [13] Yung-Yu Chuang, Brian Curless, David H. Salesin, and Richard Szeliski. A bayesian approach to digital matting. In *Proc. CVPR 2001*, volume 2, pages 264–271, 2001.
- [14] Timothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor. Active appearance models. In *Proc. ECCV '98*, pages 484–498, 1998.
- [15] Paul E. Debevec, George Borshukov, and Yizhou Yu. Efficient view-dependent image-based rendering with projective texture-mapping. In *Proc. the 9th Eurographics Workshop on Rendering*, pages 105–116, 1998.

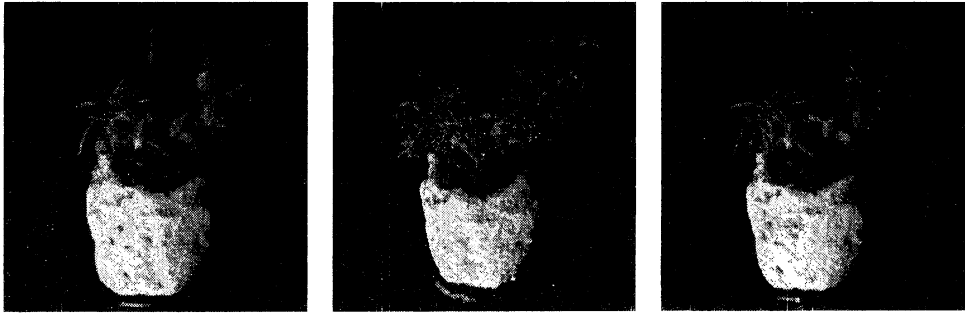


図 26: 入力画像のカメラ位置の中間から見たときの描画結果. 左: 実際にその視点で撮影した画像. 中央: 固定微小面に視点依存テクスチャマッピングした結果. 右: 本手法の結果.

- [16] Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In *Proc. SIGGRAPH '96*, pages 11–20, 1996.
- [17] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen. The luminograph. In *Proc. SIGGRAPH '96*, pages 43–54, 1996.
- [18] Frank R. Hampel, Elvezio M. Ronchetti, Peter J. Rousseeuw, and Werner A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. John Wiley, 1986.
- [19] Youichi Horry, Ken-ichi Anjyo, and Kiyoshi Arai. Tour into the picture: using a spidery mesh interface to make animation from a single image. In *Proc. SIGGRAPH 1997*, pages 225–232, 1997.
- [20] David Jones and Jitendra Malik. A computational framework for determining stereo correspondence from a set of linear spatial filters. In *Proc. ECCV '92*, pages 395–410, 1992.
- [21] Sing Bing Kang, Richard Szeliski, and Jinxiang Chai. Handling occlusions in dense multi-view stereo. In *Proc. CVPR 2001*, pages 1:103–110, 2001.
- [22] Qifa Ke and Takeo Kanade. A robust subspace approach to layer extraction. In *IEEE Workshop on Motion and Video Computing*, pages 233–242, 2002.
- [23] Vladimir Kolmogorov and Ramin Zabih. Multi-camera scene reconstruction via graph cuts. In *Proc. European Conference on Computer Vision*, pages 82–97, 2002.
- [24] Aldo Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):150–162, 1994.
- [25] Jed Lengyel and John Snyder. Rendering with coherent layers. In *Proc. SIGGRAPH 1997*, pages 233–242, 1997.
- [26] Marc Levoy and Pat Hanrahan. Light field rendering. In *Proc. SIGGRAPH '96*, pages 31–42, 1996.
- [27] J. B. Antoine Maintz and Max A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.
- [28] Jitendra Malik, Serge Belongie, Jianbo Shi, and Thomas Leung. Textons, contours and regions: Cue integration in image segmentation. In *Proc. International Conference on Computer Vision '99*, pages 918–925, 1999.
- [29] Wojciech Matusik, Chris Buehler, Ramesh Raskar, Steven J. Gortler, and Leonard McMillan. Image-based visual hulls. In *Proc. SIGGRAPH '00*, pages 369–374, 2000.
- [30] Wojciech Matusik, Hanspeter Pfister, Addy Ngan, Paul Beardsley, Remo Ziegler, and Leonard McMillan. Image-based 3D photography using opacity hulls. In *Proc. SIGGRAPH 2002*, pages 427–437, 2002.
- [31] Y. Mishima. Soft edge chroma-key generation based upon hexoctahedral color space. In *U.S. Patent 5,355,174*, 1993.
- [32] Byong Mok, Max Chen, Julie Dorsey, and Frédo Durand. Image-based modeling and photo editing. In *Proc. SIGGRAPH 2001*, pages 433–442, 2001.
- [33] Peter Neugebauer. Geometrical cloning of 3D objects via simultaneous registration of multiple range images. In *Proc. Int. Conf. on Shape Modeling and Application*, pages 130–139, 1997.
- [34] Ko Nishino, Yoichi Sato, and Katsushi Ikeuchi. Eigen-texture method: Appearance compression and synthesis based on a 3D model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1257–1265, 2001.
- [35] Manuel M. Oliveira, Gary Bishop, and David McAllister. Relief texture mapping. In *Proc. SIGGRAPH 2000*, pages 359–368, 2000.
- [36] Paul Rademacher. View-dependent geometry. In *Proc. SIGGRAPH 1999*, pages 439–446, 1999.
- [37] Ravi Ramamoorthi and Pat Hanrahan. A signal-processing framework for inverse rendering. In *Proc. SIGGRAPH 2001*, pages 117–128, 2001.
- [38] Mark A. Ruzon and Carlo Tomasi. Alpha estimation in natural images. In *Proc. CVPR 2000*, pages 24–31, 2000.
- [39] Yoichi Sato, Mark D. Wheeler, and Katsushi Ikeuchi. Object shape and reflectance modeling from observation. In *Proc. SIGGRAPH '97*, pages 379–387, 1997.
- [40] Gernot Schaufler. Dynamically generated impostors. In *Proc. GI Workshop on Modeling, Virtual Worlds, Distributed Graphics*, pages 129–135, 1995.
- [41] Steven M. Seitz and Charles R. Dyer. View morphing. In *Proc. SIGGRAPH '96*, pages 21–30, 1996.
- [42] Jonathan Shade, Steven Gortler, Li wei He, and Richard Szeliski. Layered depth images. In *Proc. SIGGRAPH 1998*, pages 231–242, 1998.
- [43] Yoshihisa Shinagawa and Toshiyasu L. Kunii. Unconstrained automatic image matching using multiresolutional critical-point filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(9):994–1010, 1998.
- [44] Alvy Ray Smith and James F. Blinn. Blue screen matting. In *Proc. SIGGRAPH 1996*, pages 259–268, 1996.
- [45] J. Wang and E. Adelson. Representing moving images with layers. In *IEEE Transactions on Image Processing*, volume 3, pages 233–242, 1994.
- [46] Daniel Wood, Daniel Azuma, Wyvern Aldinger, Brian Curless, Tom Duchamp, David Salesin, and Werner Stuetzle. Surface light fields for 3D photography. pages 287–296, 2000.
- [47] Yizhou Yu, Paul Debevec, Jitendra Malik, and Tim Hawkins. Inverse global illumination: recovering reflectance models of real scenes from photographs. In *Proc. SIGGRAPH '99*, pages 215–224, 1999.