筑波大学大学院博士後期課程

システム情報工学研究科博士論文

単視点画像からの対話的な
3次元シーンモデル生成に関する研究

飯塚 里志

(コンピュータサイエンス専攻)

指導教員　三谷純

2015年3月

概要

コンピュータグラフィクスやコンピュータビジョンの分野において，画像からその 3 次元シーンモデルを生成する手法は盛んに研究されてきた．作成された 3 次元空間の視点を移動することにより，ユーザはその空間を立体的に感じることができる．これは印象的な視覚効果をユーザに与えることから，空間ナビゲーションなどに用いられる．しかしそのような 3 次元シーンを作成するには，視点をずらしながら対象物を撮影した多くの写真が必要となる．また，モデリング関連の専門的な知識や多くの編集作業が必要となる場合もある．そこで，単視点の画像を入力として簡単なユーザ入力のみで 3 次元シーンモデルを生成するための研究を行った．本研究では，幅広いシーンに対応するため，景観画像の地面の境界線にもとづき 3 次元シーンを構築する手法と，少数のデプス入力によって滑らかな表面形状をもつ 3 次元モデルを生成する手法を提案する．また，それらの手法が 3 次元情報を利用したコンテンツ制作に幅広く応用できることを示す．まず入力画像を広い景観画像と仮定し，地面と物体の境界にもとづく奥行き推定および前景物の対話的な抽出，背景領域の合成による 3 次元シーンモデリング手法について提案する．続いて，ユーザが局所的にデプスを指定することで，物体領域の不連続箇所や遮蔽領域を考慮した滑らかな表面形状をもつ 3 次元モデルを生成する手法について提案する．この 2 種類の手法を入力画像に応じて使い分けることで，ユーザは様々な画像の 3 次元シーンを効率よく生成することができる．最後に，これらの手法によって算出したシーンの奥行きや物体領域の情報を利用することで，遠近を考慮した画像の構図編集や 3 次元映像の生成など，様々な画像コンテンツ制作を行えることを示す．本論文で提案するそれぞれの手法について複数の結果を示し，既存手法と比較することで評価を行った．これにより，簡単なユーザ操作のみで良好な 3 次元シーンを生成できることや，奥行き情報を利用した様々な画像編集に応用可能であることを確認した．本研究の成果および今後の発展により，3 次元映像に関わるコンテンツ産業がさらに発展していくことを期待する．

目 次

図 目 次

第1章　序論

1.1　はじめに

今日，デジタルカメラやカメラ付き携帯電話の普及によって，デジタル画像は我々の世界を記録し表現する媒体として，非常に身近なものになっている．これらは自ら撮影するだけでなく，多くの人が撮影した様々な場所・シーンの写真を Flickr[28] などのウェブサービスなどを通じて共有することができる．写真は立体構造をもつ現実世界を 2 次元に投影したものであるため，我々はこれを一視点から平面的にしか見ることができず，その 3 次元構造は個々の想像で補っている．しかし，近年の 3D 映画や 3 次元 CG などの普及からもわかるように，我々は映像を現実世界のように立体的に見たいという心理的欲求がある．このため，写真の 3 次元シーンモデルの作成は，コンピュータグラフィクスやコンピュータビジョンの分野で盛んに研究されてきた．3 次元シーンモデルを作成することで，ユーザはそのシーンを立体的に見ることができ，より直感的かつ印象的な映像効果を得ることができる．最近では，この技術は一般的なサービスにも利用されている．例えば Google maps のストリートビュー [2] などの地図閲覧サービスでは，大量の写真で構成された 3 次元空間をユーザが仮想的に歩けるようにすることで，ユーザがそのシーンをより立体的に感じられるようになっている．このように 3 次元映像の提示はその有用性から多く研究され実用化されてきている．

3 次元モデルの生成にはその用途によっていくつかのアプローチがある．工業分野で精密なモデルを作成するためには一般的に CAD に代表されるような計算機を用いて製品の形状情報などを精密に解析・処理する方法が用いられる．コンピュータゲームや CG 映画などでは 3 次元 CG を駆使して任意のシーンの 3 次元形状をモデリングする方法がとられる．これらはそれぞれ精密なモデリングや任意のシーンのリアリスティックな表現が可能となるが，専門的な知識を多く必要とし制作に非常に手間がかかる．この制作コストの高さは 3 次元コンテンツを提供する上で大きなハードルになっている．

これらの手法に対し，写実的な 3 次元シーンを効率よく生成するため，画像ベースによるモデリングおよびレンダリング手法が提案されている．この手法は画像からシーン構造を推定し，画像をテクスチャとして 3 次元シーンモデルを構築する．これにより，複雑な幾何形状を正確にモデリングしなくてもリアリティのある 3 次元空間が作成できる上，高速なレンダリングが可能となる．さらに生成されるシーンは参照画像をテクスチャとして用いるため，高品質な画像を使うだけで高品質なシーン生成が可能になる．しかし，そのような 3 次元シーンを自動で構築するためには視差を計算できるように撮影した大量の写真が必要となる．また，入力画像が 1 枚の場合はモデル作成に専門家による多くの試行錯誤が必要となり手間がかかる．単視点の画像ベースでも写実的な 3 次元シーンモデルを容易に得られるようになれ

ば，3 次元コンテンツをさらに多くの場面で活用できると考えられる．

そこで，単視点の画像を入力としてその 3 次元シーンを単純なユーザ操作のみで生成する手法について研究を行うこととした．次節ではまず 3 次元シーンを活用できると考えられる分野について述べる．

1.2 画像ベースの 3 次元シーンの活用

画像をベースにした 3 次元シーンは幅広い分野で活用されている．本節では画像ベースでモデリング・レンダリングした 3 次元シーンまたは推定したデプス情報を活用できると考えられる分野について，それぞれの事例を交えて紹介する．また，画像の奥行き情報は多くの画像編集ツールでも有用と考えられ，その応用例について考察する．

空間ナビゲーション

対象となるシーンの 3 次元構造が推定できれば，ユーザが指定した任意のカメラ方向から見たシーンを描画することで仮想的にシーンを動き回ることができる．これは高臨場感の映像をユーザに提供することができるため，エンターテイメントだけでなく広告やナビゲーションシステム，教育現場など幅広い分野で応用が期待できる．例えば，サッカー場の中を複数視点のビデオカメラで撮影し，その情報をもとにサッカー選手の 3 次元位置を推定することで，任意の視点から見たサッカー場の様子をリアルタイムで表示するシステムが提案されている [44]．また，一般に公開されている商用ソフトの Microsoft Photosynth[3] ではユーザが携帯電話のカメラで自分の周りを撮影していくことで，自分の立っている位置を中心としたパノラマ写真を作成できる．これは通常の単視点画像と異なり，空間を 3 次元的に把握することができるため，物件の紹介や観光地の様子を効果的に伝えられるウェブ広告などにも用いられる．また，すでに述べた地図閲覧サービスの一つである Google Maps のストリートビュー機能では，自由視点画像を提示することで従来のように 2 次元で経路を表示するよりも直感的に対象空間を理解できるようなっている（図 1.1）．これらの技術は一般的に大量の写真をつなぎ合わせることで 3 次元空間を形成している．また，特殊な撮影装置が必要となる場合もある．

3 次元映像

画像の奥行きに合わせて視差が生成されるように左目用画像と右目用画像を作成することで，3 次元映像を作ることができる．近年の 3D テレビや立体視ディスプレイなどの 3D 対応デバイスの進化によって，3 次元映像は映画やテレビ放送などのエンターテイメント分野を中心に急速に普及してきている（図 1.2）．これら 3 次元映像は従来の 2 次元映像に比べシーンを立体的に感じることができるため，あたかもその場にいるかのような高臨場感をユーザに与えることができる．日本政府の技術戦略指針でも 3 次元映像技術の研究開発の推進が提言

図 1.1: Google Maps のストリートビュー機能 [2].

されており，将来的には通信，広告，遠隔医療など様々な分野での実用化が期待されている．しかし，これら 3 次元映像コンテンツの制作には専用の高価な撮影機材が必要であり，利用できる場面が限られる．また，既存の 2 次元映像を 3 次元映像化するには高度な映像制作技術を有するクリエータによる多くの試行錯誤が必要となる．例えば 2006 年公開の映画「スーパーマンリターンズ」では，20 分間の 2 次元映像を 3 次元映像に変換するのに 10,000,000 ドルもの費用がかかったことが報告されている．このように制作コストが大きいことが消費者に十分な 3 次元映像コンテンツを提供できていない一因であり，3 次元映像の普及の妨げになっている．また，3 次元映像の表示・通信に関連する研究開発においても評価用の 3 次元映像素材が不足しているという問題がある．このように，3 次元デバイスの進化に 3 次元コンテンツ制作が追いついていないのが現状であり，2 次画像から容易に 3 次元映像コンテンツを生成する技術が確立できれば，3 次元映像コンテンツの発展と普及に貢献できると考えられる．このためには，画像から容易に奥行きを推定する手法を考える必要がある．

コンテンツ制作

画像を用いたコンテンツ制作においてもシーンの奥行き情報は幅広い形で利用できる．例えば，対象画像に別の画像から切り抜いたオブジェクトを合成する際に，シーン構造を考慮することで遮蔽の再現やオブジェクトサイズの調整などが可能となり，自然な合成結果が得られる．また，画像中に 3 次元オブジェクトを挿入し，画像のシーン構造を合わせてアニメーションさせることも可能となる．この他にも，画像処理によるカメラの撮影技法や絵画の空気遠近法の再現などにも利用することができる．背景をぼかすことで対象物を強調する *depth-of-field* と呼ばれる撮影技法は，本来カメラの絞りなどを調整することによって可能となるが，これは初心

図 1.2: 3 次元映像 [46] と 3D テレビ.
（出典：http://blog.livedoor.jp/zzcj/lite/article/51803092/image/1264005）

者には難しく高価なカメラが必要になる場合もある．これに対し，写真の奥行きが推定できればそれに合わせて背景を画像処理でぼかすことで，普通に撮影した写真からでも *depth-of-field* を再現できる．また，遠くにあるものほど色彩が大気の影響を受けて色が薄く均一になっていく現象を利用した空気遠近法も，推定された奥行き情報を用いて霞を合成することで簡単に再現することできる．さらに，画像の 3 次元構造がわかれば画像の簡易な照明調整（リライティング）を行うことも可能になる．このように画像の 3 次元構造の推定は幅広い分野で有用である．

1.3 本研究の目的

本研究は，単視点の画像から複雑なユーザ操作なしに 3 次元シーンモデルを効率よく生成できるようにすることで，人々にとって映像コンテンツをさらに有用なものにすることを目的とする．これにより，今まで専門家が多大な労力をかけて作成していた 3 次元シーンモデルを専門知識がないユーザでも容易に利用できるようになり，3 次元コンテンツのさらなる普及に貢献できると考えている．

単視点画像ベースの 3 次元モデリングに関連する手法として，画像中の特定のオブジェクトを対象としたものや入力画像のデプス推定のみを目標としたものなどがある．本研究では画像全体のシーンモデリングを目標とする．このためには，入力画像の奥行き推定や前景物体領域の抽出，前景物体の背後領域の合成などを効率的に行うワークフローの提案が必要となる．また，本研究では精密な幾何形状を復元するのではなく，上述したイメージベースドレンダリングの利点を生かした写実的な 3 次元シーンモデルを，できるだけ単純で少ないユーザ入力のみで生成することを目指す．

デジタル画像には，広い屋外を撮影した風景写真や物体を近くから撮影した写真など様々なシーンが記録されている．図 1.3(a) のように平坦な地面をもつ景観画像では，消失点や消失線の位置が推定できれば，地面領域の簡易な奥行きを算出することができる．また，地面上

(a) 平坦な地面をもつ景観画像

(b) 物体を近距離から撮影した画像や平坦な地面をもたない景観画像

図 1.3: 本研究で対象とする画像.

に置かれた物体（前景物）については，あらかじめ個別に抽出しておくことで地面領域からその奥行きを算出できる．よって，このような景観画像の 3 次元シーンを生成するためには，消失線の位置推定や前景物の領域抽出が必要となる．これに対し，図 1.3(b) のように，物体を近距離から撮影した画像や平坦な地面をもたない景観画像については消失線を定義出来ない場合が多く，奥行き推定には別のアプローチを考える必要がある．以上のように多様な画像の 3 次元シーン生成に対応するため，本研究では対象とする画像を限定した 2 通りの 3 次元シーンモデル生成手法を提案する．

- 地面の境界線にもとづく少数のポリゴンから構成される 3 次元シーンの構築
- 少数のデプス入力による滑らかな表面形状をもつ 3 次元モデルの生成

まずはじめに，入力画像を平坦な地面をもつ景観画像に限定し，地面領域と残りの領域の境界線を利用して簡易な 3 次元シーンモデルを構築する手法を提案する．ここでは 3 次元モデル生成のための全体のワークフローの提案を主とし，シーンの 3 次元座標の計算や地面に置かれた物体の対話的な抽出とモデルリング，背後領域のテクスチャの生成を効率よく行えるようにする．生成される 3 次元シーンモデルは少数のポリゴンから構成される簡易なモデルであるが，十分に立体感のある 3 次元ウォークスルーが可能になる．

その後，上記の手法で対象外となるシーンに対応できるように，ユーザが画像上でまばらに指定したデプスから 3 次元シーンモデルを生成する手法を提案する．この手法は前述の手法に比べユーザ入力が増えるが，地面が写っていない写真や物体を近くから撮影した写真など幅広いシーンに適用できる．この手法では，3 次元シーンモデルを Layered Depth Image[70] として表現する．本研究の Layered Depth Image は前景レイヤと背景レイヤから構成され，それぞれのレイヤは対応するテクスチャとデプスマップを保持している．このモデル表現は視点移動したときに前景領域の背後が「穴」になってしまうことを防ぐことができる．このようなモデルを生成するには，効率的なデプスマップの生成，デプスの不連続箇所の抽出，遮蔽領域の抽出とデプスとテクスチャの生成が必要になる．

さらに，すでに述べたように画像のシーン構造推定は幅広い画像コンテンツ制作に利用できる．本研究ではシーンの奥行きを考慮したコンテンツ制作の例として，まず画像の構図編集に焦点を当て，シーンの遠近を考慮して物体の配置編集が行えるアプリケーションについて提案する．これは 3 次元シーン構造推定を利用し，前景物体の抽出や影の抽出・合成を行うことで実現できる．さらに，推定した奥行き情報を利用した 3 次元映像の生成や画像の焦点調整などのアプリケーションを示すことで，単視点画像の 3 次元構造を推定することの有用性について議論する．

本研究の成果および今後の発展によって，写真は単に 2 次元的に「見る」ものではなく，立体的に「感じる」ことができるようになることが考えられる．これは，2 次元静止画像として記録されたシーンの理解をより深め，効果的な学習コンテンツやプレゼンテーションなどに利用できる．また，上述した 3 次元映像制作のコスト削減や画像を利用したコンテンツ制作の現場など幅広い分野での利用も期待できる．

1.4 本論文の構成

本論文は本章を含めて全 6 章から構成される．

第 2 章では，画像からの 3 次元モデルの生成について，入力が複数の場合や単視点の場合などに分類して紹介する．特に，本研究の対象としている単視点画像からの 3 次元モデル生成について，対象とするモデル形状やユーザ操作について分けて説明する．その後，単視点画像からデプスマップを生成する手法についてもまとめる．

第 3 章では，少数の平面ポリゴンから構成される 3 次元シーンを地面の境界線にもとづき対話的に構築する手法について述べる．まず，対象とする画像および生成される 3 次元モデル形状について述べ，3 次元座標の推定方法や効率的な前景物モデリング，背後領域の補完手法について述べる．提案手法によって生成された 3 次元シーンの例を示し，既存研究と比較することによってその有効性について検証する．

第 4 章では，スパースなデプス入力から 3 次元モデルを生成する手法について述べる．ここで対象とするモデルは 3 章で示したものとは異なり，滑らかな表面形状をもつ 2 つのレイヤで構成される Layered Depth Image として表現される．まず入力したデプスから画像領域を考慮したデプスマップを生成する手法について述べ，レイヤ構造を構築するためのデプスの不連続箇所の抽出や遮蔽領域の推定手法について説明する．実際に計算機上にアプリケーションを実装した例を紹介し，提案手法によって生成された 3 次元シーンモデルおよび既存研究の比較を示すことで，提案手法について議論する．

第 5 章では，前章までに述べてきた単視点画像の 3 次元構造推定を画像コンテンツ制作に応用した例を示す．まず，シーンの遠近を考慮して画像中の物体の配置編集を行うシステムについて詳しく説明する．次に，画像の奥行き情報を利用する，3 次元映像制作や焦点調整などのアプリケーションについて説明する．これらのアプリケーションについて，実際に提案手法によって生成された結果を示し，提案手法の有用性を示す．

最後に，第 6 章で本研究の成果によって得られた結論および今後の展望についてまとめる．

第2章　画像を用いた3次元シーンの生成に関する研究

本章では，画像から 3 次元シーンモデルを生成する手法（イメージベースドモデリング）についてこれまで提案されてきた手法について紹介する．ここでは表 2.1 のように，イメージベースドモデリングの特徴を大まかに分類して説明する．イメージベースドモデリングは複数の画像または動画を入力として用いる手法と，1 枚の画像のみを用いる手法に大別できる．また，モデリング対象をシーン全体かオブジェクトのみに限定するか，全自動で生成するかユーザ入力にもとづき生成するかでアプローチが異なる．まず，複数視点の画像から 3 次元シーンを復元するアプローチと単視点の画像のみから 3 次元シーンを生成するアプローチに対してそれぞれの特徴で分類して紹介する．その後，本研究で対象とする，単視点画像から 3 次元モデルを生成する手法について詳しく述べる．

表 2.1: イメージベースドモデリングの分類

使用する画像	複数（動画）　又は　単一
対象	シーン　又は　オブジェクト
ユーザ入力	自動　又は　半自動

2.1 複数視点の画像からの3次元シーンの復元

画像の3次元構造の復元はコンピュータビジョンにおいて最も知られた問題の1つである．Debevec ら [25] は写真の中の建築物の輪郭から基本立体形を当てはめることで建築物の3次元モデルを対話的に作成する手法を提案している．近年の研究では，ステレオ視や Structure from motion (SFM) と呼ばれる手法に基づき3次元モデルを生成する手法が多く提案されている [69, 54, 13, 72, 29]．これらの手法は2枚または複数の画像で特徴点を抽出し，それぞれの対応する点の位置関係からエピポーラ幾何によって3次元座標を計算している（図 2.1）．また，パノラマ写真のように複数の写真をつなぎ合わせて自由視点画像を作成するようなアプローチ [16] や，複数視点のビデオカメラの映像から任意の視点から見た空間の様子をリアルタイムで表示するシステムも提案されている [44]．これらの手法は広範囲にわたる良好な3次元シーンを生成できるが，用意する入力画像に制約がある．SFM ではエピポーラ幾何によってカメラパラメータを計算するため対応する特徴点を計算する必要があり，視点が大きく移動したり動的な物体が画像中に存在したりするとうまく3次元座標を計算できない．パノラマ写真の場合，写真同士をつなぎ合わせられるように大部分の領域を共有するような写真を撮影する必要がある．

上記のアプローチに対し，単視点の画像のみからその3次元シーンを復元するための手法が研究されている．これらの手法は1枚の画像を入力として，シーンの構造に制約を与えることで自動または半自動で生成している．このような研究はコンピュータグラフィクスやコンピュータビジョンにおいて非常に多く存在するため，以下では特に本研究と関連すると思わる研究のみに焦点を当てて紹介していく．

図 2.1: 複数の画像から3次元構造の復元．
出典：Photo Tourism[72]

2.2 単視点の画像を用いた3次元シーンの生成

1枚の入力画像のみからその3次元構造を計算機を用いて復元するのは非常に困難な問題である．実際，単視点の画像から正確に3次元座標を復元するのは幾何学的には不可能であることが知られている [40]．また，写真には前景物の背後領域など遮蔽される領域が存在することが多く，この領域の構造を正確に再現することはできない．しかし，実際には人間はこのような3次元構造を経験にもとづき容易に想像することができる．このため，単視点画像から3次元構造を復元しようとする場合は人間が自身の経験にもとづき手作業で3次元シーンをモデリングしていくことが一般的である．しかしこの作業は非常に手間がかかる．本節では単視点画像から3次元シーンを自動生成する手法をはじめに紹介し，その後簡単なユーザ入力を用いて3次元モデルを半自動生成する手法について述べる．

2.2.1 自動生成

前節で述べたように，単視点の画像からその3次元構造を正確に復元することは困難である．しかし，現実世界の3次元構造には制約があり，その制約を考慮することで視覚的に良好な3次元シーンモデルを自動生成することが可能となる．

Shape from Shadingと呼ばれる手法 [86, 69] では，モデリング対象を画像中のオブジェクトに限定し，その表面の陰影から3次元形状を復元している．また，オブジェクトの表面の模様の歪みから3次元形状を推定する，shape from textureと呼ばれる手法 [49, 55] も提案されている．しかし，これらの手法はオブジェクトの表面が均質な色や模様でないと適用できない．また，これらの手法は景観画像などのシーン全体に適用することは難しい．

Hoiemら [31] は画像が平らな地面，空，物体領域の3つから構成されると仮定し，それぞれの3次元構造をあらかじめ制約することで簡易な3次元シーンを自動で構築する手法を提案している．この手法では，画像を上記3つの領域に分割し，地面領域の奥行きを消失点から推定し，物体領域は地面に対して垂直に置かれていると仮定し，空領域を削除することで簡易な3次元シーンを生成している．この3つの大きな領域を推定するために，領域分割 [27] によって得られた小さな均質領域 (スーパーピクセル) を機械学習したパラメータにもとづき統合していくことアプローチを提案している．この手法はその後の研究で，3次元構造の大まかなラベリング [32, 34] や遮蔽された物体境界の復元 [35] などに応用されている．また，Liuら [50] は入力画像を空や道路，木，建物など，より多種類の領域に分割し，この領域情報と消失線の位置から画像の奥行きを推定する手法を提案している．Saxenaら [67, 68] は3次元シーンがスーパーピクセルの集合で表現できると仮定し，画像のテクスチャの色や勾配などの特徴量から機械学習によってスーパーピクセルの3次元位置を推定し，自動で3次元シーンモデルを生成する手法を提案している．これらの手法は，パラメータの調整は別として，完全に自動で3次元シーンを生成できるため，誰でも簡単に利用できる．しかし，これらの手法は対象とするシーンに厳しい制約を用いており，3次元座標の計算や領域推定に失敗する場合が少なくない．また，前景物が存在する画像ではその領域を地面領域と分離できず，背後領域なども推定できないため，良好なシーンモデルが得られない．また，少数のポリゴンま

たは小領域の集合でモデルを構成しているため，滑らかに変化するような表面形状を生成することはできない．

Karsch ら [42] は主に 3 次元映像制作を目標とし，単視点映像からでも画像のデプスマップが計算できる手法を提案している．この手法ではあらかじめ画像とそのデプスマップが対応付けされたデータベースを用意し，はじめに入力画像と似たシーンをもつ画像を複数抽出する．次に抽出された画像群をピクセル単位で入力画像と対応付け，対応付けされたピクセルと一致するデプスを重み付きで混合することで，入力画像のデプスマップを自動で生成している．この手法は動画への拡張も提案されており，動的な物体を含む映像などステレオアルゴリズムなどが適用できないシーンにも利用できることが示されている．しかし，3 次元推定の精度は使用するデータベースに依存し，生成されるデプスマップも物体の輪郭がぼやけた粗いものである．

このように全自動による 3 次元シーンの生成は専門知識をもたないユーザなどでも簡単に利用できるが，入力画像をうまく領域分割できない画像や前景物が地面に置かれている画像，遮蔽領域がある画像などではうまくシーンを推定できない．これに対し，人間の奥行き知覚能力を生かし，ユーザが 3 次元構造を推定する上での手がかりを入力することで対話的に 3 次元シーンを生成する研究がある．次節からはこの半自動生成のアプローチについて議論する．

(a) 入力画像　　(b) 3 次元シーンモデル

図 2.2: Hoiem らの手法 [31] による単視点画像からの 3 次元シーンの自動生成．

2.2.2　半自動生成

前節で述べた自動生成手法に対し，いくつかのシーン情報をユーザが入力することで 3 次元モデルを生成する手法が提案されている．これらは特に画像中のオブジェクトの 3 次元モデル生成を目的としたものと，シーン全体の 3 次元モデルを生成する手法で分類できる．本

節では，はじめにイメージベースでオブジェクトの 3 次元モデリングを行う手法をいくつか紹介し，その後本研究で対象とするシーンの 3 次元モデリング手法について関連する研究を述べる．

オブジェクトの 3 次元モデリング

画像中に写る箱や自動車などのオブジェクトを対象して 3 次元モデリングを行う技術は多く研究されている．これらの研究の多くは，最初に対象とするオブジェクト形状を定義し，その形状を復元するのに最適なユーザ入力と形状処理アルゴリズムを提案している．Jiang ら [39] は対称性をもつ建築物のモデリングを目標とし，ユーザがその輪郭をなぞっていくことで精密な 3 次元形状をもつ建築物のモデルを生成できる手法を提案している．また，ユーザが指定したオブジェクトの輪郭線に基づきその閉領域を膨らませることで滑らかな曲面形状をもつ 3 次元モデルを作成する手法 [59, 74] が述べられている（図 2.3）．また，あらかじめ想定した円筒形状をオブジェクトの輪郭線にもとづき変形して当てはめる手法 [17] も提案されている．これらの手法はそれぞれの研究が対象とするオブジェクト形状を良好に復元することができるが，背景領域のモデリングや遮蔽領域の合成などは行わないため，画像全体のシーンモデリングには利用できない．

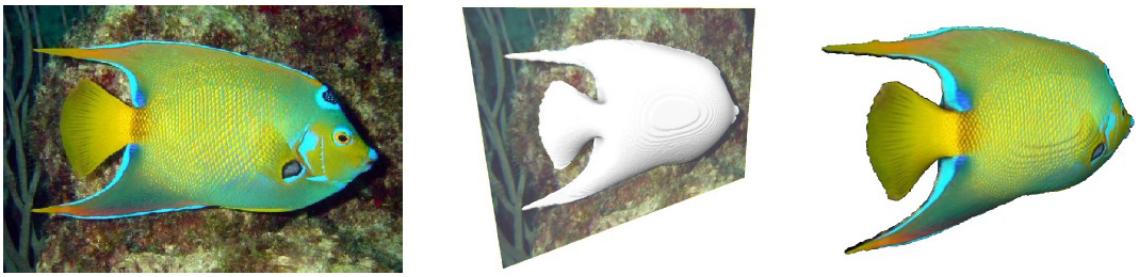


図 2.3: 画像中のオブジェクトの 3 次元モデリング.
出典：Fast and Globally Optimal Single View Reconstruction of Curved Objects[59]

シーンの 3 次元モデリング

図 2.4 は画像のシーン全体の 3 次元モデルを生成する手法について，ユーザ入力の多さと生成されるモデルの品質によって分類したものである．Criminisi ら [23] は画像の物体輪郭の中の平行線や消失点をユーザがしていすることで画像の 3 次元座標を計算する手法を紹介している．Oh ら [57] は 1 枚の画像を入力として，画像中の木や建物を階層的に並べることで 3 次元シーンをインタラクティブに生成するシステムを提案している．このシステムでは画像の照明なども編集でき，非常に良好な結果を得られるが，領域分割やその奥行き割り当てなどを手動で行う必要があり，1 つの 3 次元シーンを作成するのに数時間かかったことが報告されている．また，Zhang らは自由形状モデルを 1 枚の画像から対話的に生成する手法を提案

している．この手法では，ユーザが画像上で法線や領域の不連続箇所，平面形状や曲面形状などの制約を指定していき，この情報をもとにモデルを変形させることで曲面形状をもつ複雑な 3 次元モデルを作成することができる．Assa ら [8] は 3 次元モデルを直接生成するのではなく，立体感をユーザに与えられるジオラマモデルを生成する手法を提案している．この手法は，大気散乱や焦点ぼけの度合いなどの奥行きの手がかりから画像中の奥行きが大きく変化している箇所を抽出し，その部分を膨張させて描画することで，わずかな視点移動で写真が立体的に見えるようにしている．この手法は焦点ぼけや大気散乱などが起こる広大な屋外画像のみを対象としており，デプスの計算にも時間がかかるため対話的な編集は行えない．

上記のように複雑なシーン形状を構築する手法は多くのユーザ入力を必要とし，目標とする 3 次元モデルを作成するために試行錯誤を行うため手間がかかる．これに対し，簡単なユーザ入力のみで十分な 3 次元効果を生成できる簡易な 3 次元シーンモデルを構築する手法が提案されている．この手法は Tour Into the Picture と呼ばれ，本論文の 3 章で述べる手法はこの研究の着眼点にもとづいている．

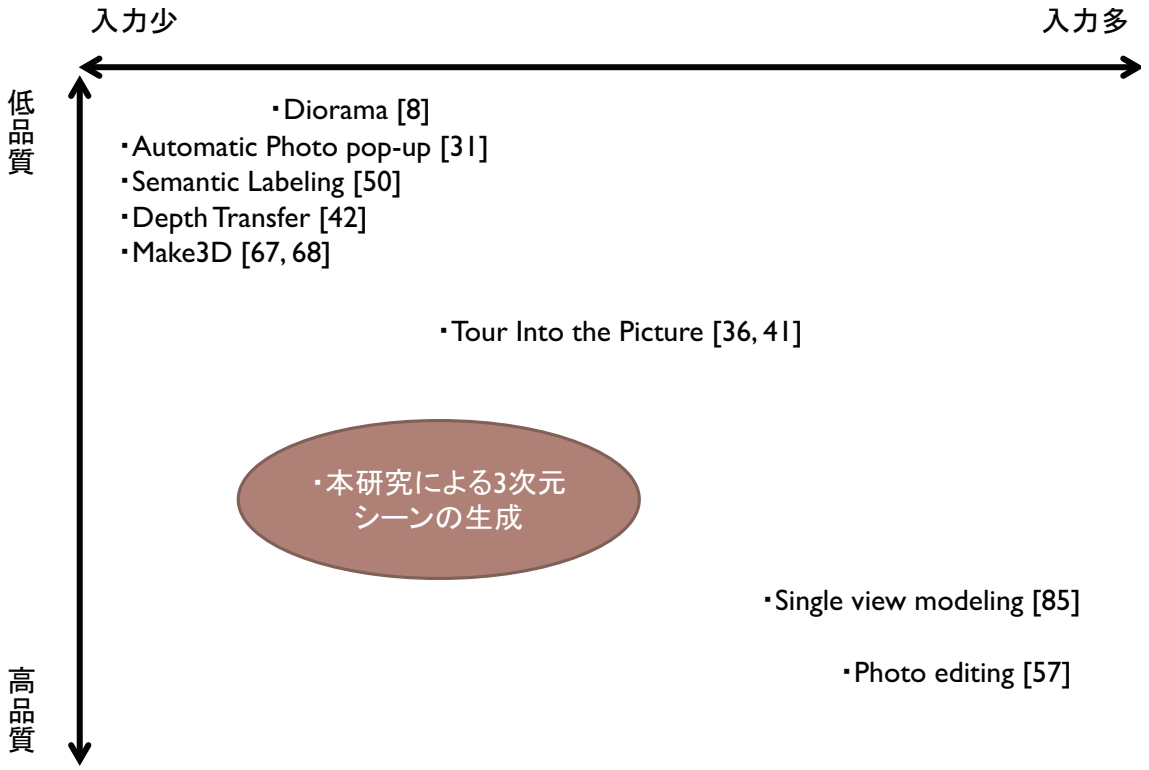
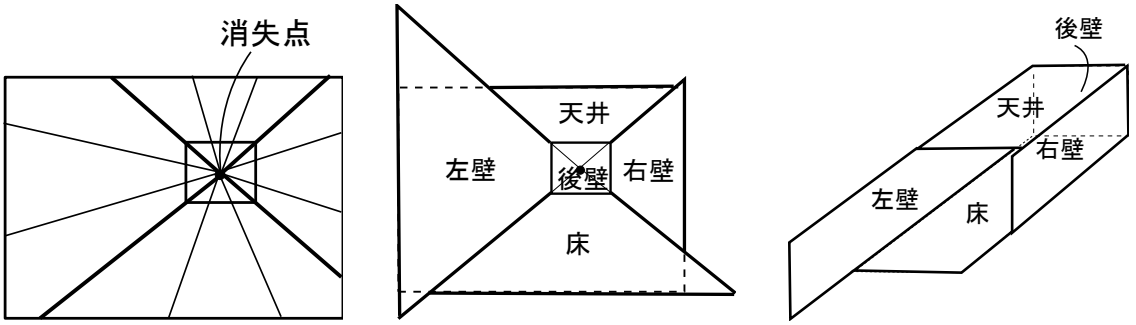


図 2.4: 1 枚の画像を入力とした 3 次元シーンの生成に関する手法の分布．

Tour Into the Picture

Horry らが提案した Tour Into the Picture と呼ばれる手法 [36] は，1 枚の画像から単純な 3 次元シーンを作成することでウォークスルーアニメーションの簡単な作成を可能にした．この手法で生成されるシーンモデルは背景モデルと複数の前景物モデルからなる．まず，ユーザはスパイダリーメッシュと呼ばれるユーザは消失点とその消失点から放射状に伸びる線分を指定する (図 2.2.2)．これにもとづき入力画像は「床」「右壁」「左壁」「後壁」「天井」の 5 つの領域に分割される (図 2.2.2)．この 5 つの領域から図 2.2.2 のような背景ポリゴンモデルが生成される．



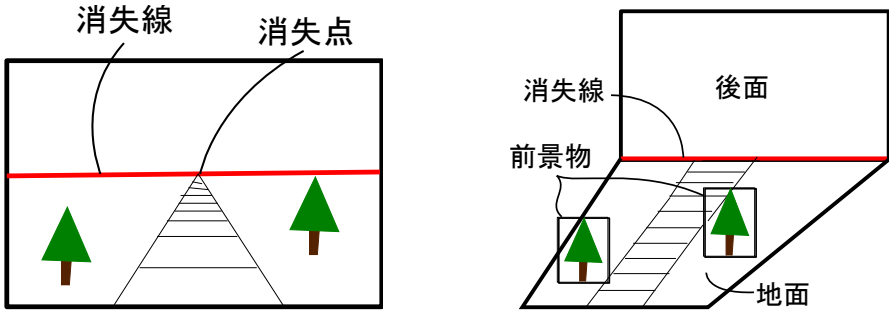
(a) スパイダリーメッシュ　(b) 領域分割　(c) 背景ポリゴンモデル

図 2.5: Horry らの手法 [36]. (a) 入力画像上で消失点とスパイダリーメッシュの位置を指定し, (b)「床」「右壁」「左壁」「後壁」「天井」の 5 つの領域に分割して, (c)3D モデルを生成する．

背景モデルが生成された後，前景物も画像上でインタラクティブに指定される．この前景物は複数のポリゴンの集合でモデリングされ，背景モデルの床に設置される．このようなポリゴンモデルに，ユーザが一般的なペイントツールを用いて入力画像から作成した背景画像と前景画像をテクスチャとして参照することで，出力画像がレンダリングされる．この手法は入力画像を 1 点透視図と仮定しているため，画像中に消失点が存在しない場合や消失点が複数存在する画像には適用が困難である．

Kang らは Horry らの手法を発展させ，より単純で効果的な 3 次元モデリング手法 [41] を提案した．この手法では入力画像を消失線によって 2 つの領域に分割し，3 次元モデルを生成する．ここで述べる消失線とは消失点を含む「地平線」である．消失線より下に位置する領域は前景物などが設置される地面であり，消失線より上の領域は画像の後景に相当する後面である．図 2.6 は，Kang らの手法による 3 次元モデルである．この手法では，まず入力画像から背景画像と前景物画像を作成する．背景画像は一般的なペイントツールを使用して入力画像から前景物領域を消すことで作られる．前景物画像は入力画像から前景物をそれぞれ四角形で抜き出し，四角形の中で前景物領域以外は透過させることで作成される．次に背景画像上で消失線をユーザが指定することで，3 次元背景モデルが生成される．最後に背景モデル

に前景物画像を貼り付けた 1 枚の四角形ポリゴンが設置され，3 次元シーンモデルが完成する．この手法は消失点が存在しない画像や消失点が複数存在する画像に適用することができる．しかしこの手法は地平線をもつ広い景観画像を対象としているため，適用できる画像が限られる．また，背景画像や前景物画像を作成するのにも手間がかかる．



(a) 入力画像　　(b) 3 次元モデル

図 2.6: Kang らの手法 [41]. (a) 消失線によって画像を 2 つの領域に分割する. (b) 消失線より下を地面, 上を後面とし, 前景物画像をそれぞれ板状ポリゴンに貼り付けて設置することで 3 次元モデルを生成している.

スクリブルを用いたデプスマップの生成

単純で対話的なデプスマップの生成手法として，Wang ら [75] はスクリブルでデプスを指定していくことで，1 枚の画像からステレオ視画像を生成する手法を提案している（図 2.7）．ここでスクリブルとはブラシストロークを指す．この手法では，画像上でスクリブルによって指定したデプスをエネルギー最適化によって画像全体に伝播させることでデプスマップを生成している．このようなスクリブルによるデプスマップの生成は Ribera ら [65] によって動画を入力とした 3 次元モデルの生成にも利用されている．Yucer ら [83] は *Transfusive Image Manipulation* と呼ばれる手法 [82] を発展させ，不等式制約をエネルギー関数に取り入れることで Wang らの手法よりもエッジを保持したデプスマップを計算する手法を提案している．これらの手法は粗くデプスを指定する単純な操作のみで画像全体のデプスを計算できるが，入力画像に対して 1 枚のデプスマップしか生成できないため，前景物体による遮蔽領域などは考慮できない．また，エネルギー最適化のみによるデプスの伝播は計算コストが高く，多くのデプス入力が必要になる場合もある．Lopez ら [53] は勾配が小さい領域は似たデプスをもつという仮定を用いてエネルギー関数を定義し，パースや相対性などの制約を与えることで画像全体のデプスマップを計算する手法を提案している．

このように，スクリブルベースのデプスマップの計算は単純なユーザ入力のみしか必要としないため直感的な編集が可能である．しかし，上記の手法はデプスを入力した位置から離

れたピクセルにはうまくデプスが伝播しないため，多くのデプス入力が必要となる．また，ピクセルベースの最適化は反復計算に時間がかかるため，対話的な編集が難しい．

(a) ユーザ入力　　(b) デプスマップ　　(c) アナグリフ

図 2.7: デプススクリブルを用いたデプスマップの生成 [75]．(a) ユーザは入力画像上でスパースにデプスを指定すると，(b) そのデプスを画像全体に伝播させることでデプスマップが生成される．(c) このデプスマップを用いてアナグリフ画像が生成される．
出典：StereoBrush[75]

第3章　地面の境界線を利用した3次元シーンモデルの生成

本章では，1 枚の景観画像を入力とし，地面の境界線にもとづき 3 次元シーンを生成する手法を述べる．ここで対象とするモデルは少数のポリゴンから構成される単純なものであるが，ユーザに十分な 3 次元効果を与えることができる．また，ユーザ入力は画像上で線やブラシストロークを引くという簡単な操作であり，直感的に編集作業を行うことができる．まず，本手法で生成する 3 次元シーンモデルの特徴やそのモデリング手法について概説する．その後，それぞれの手法の詳細について述べ，前景物抽出については既存手法との比較を示す．最後に提案手法によって得られた 3 次元モデルの例を示し，既存手法と比較しながら考察を行う．

3.1 概要

提案システムでは，地面領域をもつ 1 枚の景観画像を入力とし，その 3 次元シーンは 1 つの背景モデルと複数の前景物モデルで構成される．これらのモデルは少数の平面ポリゴンに入力画像の各領域がテクスチャマッピングされたものであり，背景モデルの地面領域に前景物モデルが垂直に設置される．本手法の主な貢献はそのような 3 次元シーンモデルを効率よく生成するための全体的なフレームワークの提案である．

提案システムにおいて，3 次元シーン作成のためにユーザは下記の 2 つの簡単な作業を行う．

1. 地面の境界を折れ線で指定
2. 前景物を粗く囲む

これらの作業は入力画像上で行われ，3 次元的な編集を必要としない．

提案手法の流れを図 3.1 に示す．ユーザが境界線と前景物を指定すると（図 3.1(b)），入力情報にもとづき 3 次元ワイヤーフレームモデルと背景・前景画像が生成される（図 3.1(c)）．それぞれ対応するポリゴンに背景画像と前景画像をテクスチャマッピングすることで 3 次元モデルが生成される（図 3.1(d)）．カメラ視点を移動させることでユーザは画像の 3 次元ウォークスルーを体験することができる（図 3.1(e)）．以下では本手法における境界線と前景物に定義について述べる．

3.1.1 境界線

入力画像の 3 次元座標は境界線により決定される．ここで述べる境界線とは「地面」領域と「壁」領域の境目を指し，ユーザによって折れ線で指定される．例えば 図 3.1 (a) のような写真では地面と建物の境目が指定される．ここでは建物や空が「壁」領域となる．また，図 3.11(a) のように境界に曲線が現れるような画像に対しても，折れ線の頂点を増やした近似的な曲線で境界線を指定することができる．この境界線の各頂点座標にもとづいて背景モデルの 3 次元構造が決定され，前景物モデルの 3 次元位置を算出することができる．さらに境界線は背景テクスチャ生成で必要な前景物領域の穴埋めの際に拘束条件としても用いられ（3.2.2 節で後述），背景テクスチャ生成の精度を向上させる役割もつ．

3.1.2 前景物

入力画像の前景物はユーザによって指定される．本論文における前景物とは境界線によって分割された地面領域に置かれている立体物のことであり，基本的に 1 枚の平面ポリゴンでモデル化される．しかし前景物領域を正確に抽出する作業は手間がかかる．そこで本手法では画像の領域分割 [22] を用いた選択領域の拡張と GrabCut [66] をあわせて前景物抽出を行う．この手法ではユーザははじめに前景物を粗く囲み，この情報をもとにシステムが前景物領域を抽出する．もし前景物が正確に抽出できていない場合はその領域をユーザが指定し，それ

(a) 入力画像

(b) ユーザ入力

(c) テクスチャとワイヤーフレームモデル

(d) 3 次元モデル

(e) 視点移動

図 3.1: 提案システムの流れ.

をシステムが学習して再度抽出処理を行う．この手法により少ないユーザ入力で簡単に前景物を抽出することが可能となる．さらに，前景物領域の中で地面に接している部分をユーザが明示的に指定することで遠近を考慮した前景物をモデル化できることを示す．

3.2 シーンモデリング

この節ではシステムの基盤となるアルゴリズムについて述べる．最初に，境界線にもとづく背景モデルの生成アルゴリズムとそのテクスチャ生成について述べる．次に簡易ユーザ入力による前景物の抽出手法とそのモデリングについて説明する．さらに，前景物モデルに適用されるビルボード変換や接地制約について述べ，これによりシーンのリアリティが向上することを示す．

3.2.1 背景モデル

入力画像はユーザによって指定された折れ線にもとづいて地面ポリゴンと複数の壁ポリゴンに分割される（図 3.2）．この各頂点に適切な 3 次元座標を割り当てることで背景モデルが構築される．スクリーン座標系において左下を原点とし右方向を $+x$，上方向を $+y$ とする．ここで，3 次元モデルにおいて原点はカメラ位置と一致し，視線方向を $+z$，カメラの焦点距離を f とする．

入力画像のスクリーン座標の原点 $\mathbf{P}_0$ を (x_0, y_0)，境界線上で最も y 座標が大きな頂点 $\mathbf{P}_M$ を (x_M, y_M) とし，ワールド座標系における $\mathbf{P}_0, \mathbf{P}_M$ をそれぞれ $(x'_0, y'_0, f), (x'_M, y'_M, f)$ とすると，その同次座標 $\mathbf{P'}_0, \mathbf{P'}_M$ は以下のように表される．

$$\mathbf{P'}_0 : (x'_0, y'_0, f, 1) \qquad \mathbf{P'}_M : (x'_M, y'_M, f, w_{min})$$

ここで w_{min} とは十分に小さな正の値であり，本システムでは 0.2 としている．この 2 点を基準として入力画像の地面領域の各頂点 $P_i(x_i, y_i)$ が以下のように算出される．

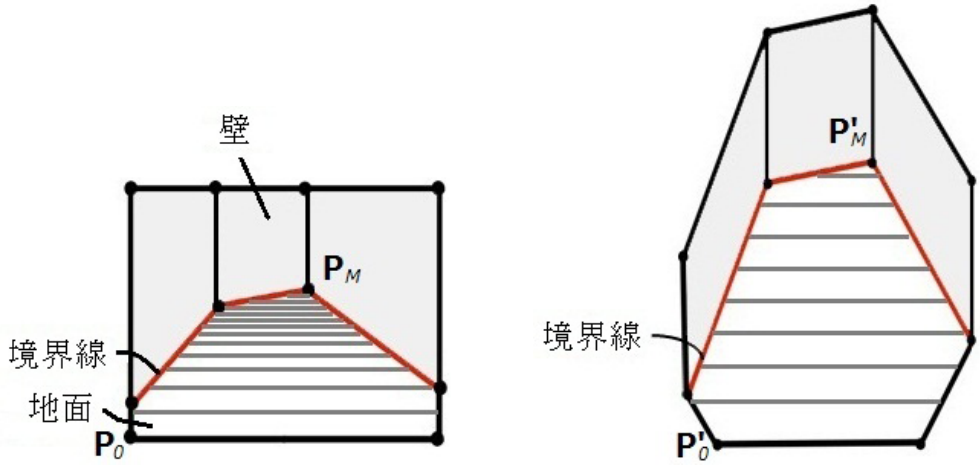
$$\mathbf{P'}_i : (x'_i, y'_i, f, w_i)$$

ただし，

$$w_i = \frac{(y_i - y_0)}{(y_M - y_0)} w_{min} + 1 - \frac{(y_i - y_0)}{(y_M - y_0)} \tag{3.1}$$

また，残りの壁領域の各頂点は壁と地面は垂直であるという制約条件のもとに算出される．これにより視線方向に奥行きをもつ 3 次元モデルが生成される．生成されるモデルは Kang らの消失線の理論 [41] にもとづいているが，本手法は各頂点の y 値によって座標を決定している

ため，Kang らが示した算出手法よりも単純に計算でき，さらにこの方程式を用いて前景物モデルの 3 次元座標も算出できる（節 3.2.3）．



(a) 2 次元画像　　(b) 3 次元モデル

図 3.2: (a)2 次元入力画像と (b) 生成される 3 次元モデル．

3.2.2　背景テクスチャの生成

生成されたポリゴンモデルに背景テクスチャをマッピングすることで背景モデルが完成する．本論文における背景とは入力画像の中で前景物が含まれない領域である．しかし前景物領域を取り除いた入力画像を背景テクスチャとして用いると，前景物の部分が「穴」になってしまう．これを避けるため，前景物領域はその他の背景領域で補完する必要がある．このため，本研究では Wexler ら [77] や Simakov ら [71] のように，画像の前景物領域を未知領域とし，この領域をその他の画像領域から類似パッチを探索し合成していくことで未知領域を補完する．これは以下の式を最小化することで表される．

$$d(U,O) = \frac{1}{N_o} \sum_{o \subset O} \min_{u \subset U} D(o,u) \tag{3.2}$$

ここで，U は未知領域，O は画像中の U 以外の領域，o, u はそれぞれ O と U に含まれる画像パッチ，N_o は O のパッチ数を表している．また，D はパッチ間の Lab 空間における SSD (sum of squared differences) であり，$N \times M$ サイズのパッチでは下記の式で表される．

$$D(i,j) = \sum_{j=0}^{N-1} \sum_{i=0}^{M-1} ||\mathbf{C}_i - \mathbf{C}_j||^2 \tag{3.3}$$

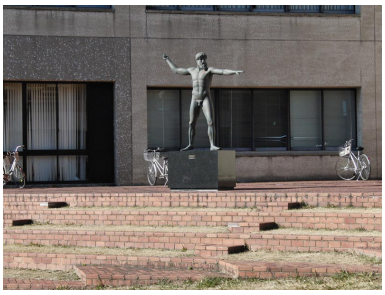
ここで，$\mathbf{C}_i, \mathbf{C}_j$ はピクセルの Lab 空間における色を表す 3 次元ベクトルである．

d は次のように最小化される．まず，未知領域を境界の色で塗りつぶして初期化し，未知領域の全ピクセルにおいて SSD が最小となる類似パッチを計算する．次に，算出された類似パッチの中心座標の色を未知領域に割り当てていくことで未知領域を埋める．この 2 つのステップを繰り返すことで，自然な補完結果を得ることができる．計算結果が局所解に陥らないように，これらの処理はガウシアンピラミッドを用いたマルチスケールで行われる．

しかし，未知領域の類似パッチを全画像領域から探索するのは非常に時間がかかり，対話的な編集が行えない．これに対し，Barnes らは無作為探索を利用することで類似パッチ探索を数十倍から数百倍に高速化する手法を提案している [10, 11]．提案システムではこの手法を用いることで，オブジェクト領域の補完を対話的な速度で行えるようにしている．この手法の詳細については付録 A を参照されたい．

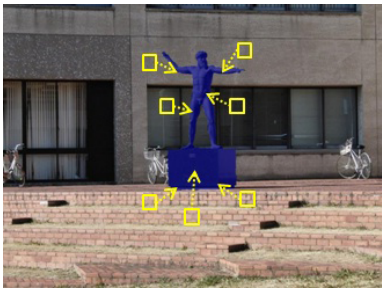
しかし画像補完は対象領域が大きくなるほど精度が低くなってしまう（図 3.3(b)-(d)）．Barnes らは，画像補完の際にユーザが指定したガイド線上に探索領域を拘束することで，画像構造を補完に反映させる手法を提案している．本手法ではこの手法にもとづき，境界線を補完の拘束条件として用いることでオブジェクトの補完精度を向上させる (図 3.3(e)-(g)). つまり，入力画像の境界線上のオブジェクト領域では背景領域の境界線上のみから類似パッチを探索するようにすることで，より自然な補完結果を得られるようにする．

図 3.3 は本手法における背景画像の生成結果である．選択された前景領域（青）をそのまま補完すると，本来建物の壁であるはずの領域に階段の一部が合成されてしまう (図 3.3(b)-(d)). これに対し，奥行き推定のために指定した境界線（赤）を類似パッチ探索の制約として用いることで壁と階段が区切られた良好な背景画像が生成されている（図 3.3(e)-(g)）．このように境界線は 3 次元座標の算出の他にもオブジェクト領域の補完精度を向上させる役割も果たしている．



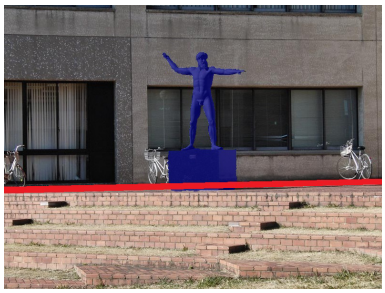
(a) 入力画像

(b) オブジェクト領域

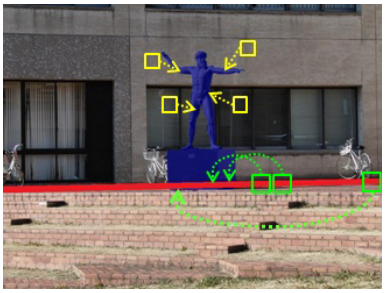


(c) 制約なしの類似パッチ探索

(d) 補完結果



(e) オブジェクト領域と境界線



(f) 境界線による探索範囲の制約

(g) 補完結果

図 3.3: 背景画像の生成．(a)(b) 入力画像のオブジェクト領域（青色）に対して，(c) 制約なしの類似パッチ探索による画像補完では (d) 不自然な結果になってしまっている．これに対し，(e) 境界線（赤）によって (f) パッチ探索範囲を制約することで，(g) 良好な背後領域の生成を行うことができる．

3.2.3 前景物モデル

本節では提案手法における前景物モデリングについて述べる．まず入力画像から前景物領域を抽出する手法について述べ，その後前景物のモデリング手法について説明する．

前景物領域の抽出

前景物の抽出は画像編集の際に最も手間のかかる作業の 1 つである．既存手法では前景をブラシで直接塗りつぶしていくペイントベースの抽出手法 [15][48][51][58] や背景と前景の境界をなぞっていく境界線ベースの抽出手法 [43][56] が示されている．しかし，これらは対象となる前景物の形状によっては細かく正確な作業が必要となる．提案システムではユーザ入力をなるべく少なくし，かつ細かく正確な作業を必要としない前景物抽出を目指し，領域分割とグラフカットベースの手法を組み合わせて前景領域の抽出を行う．

まず，本システムでは入力画像は読み込まれると同時に色特徴が均質な領域（スーパーピクセル）に分割される（図 3.4(b)）．この領域分割のために，我々は Comaniciu らの手法 [22] を用いる．この手法では入力画像の色を表す $L^*u^*v^*$ と位置を表す x, y の 5 属性を特徴として Mean Shift を行い，近接領域同士の色空間におけるユークリッド距離が閾値以下のものを統合することで領域分割を行う．この手法は多くの画像で高い精度を実現することが示されている．

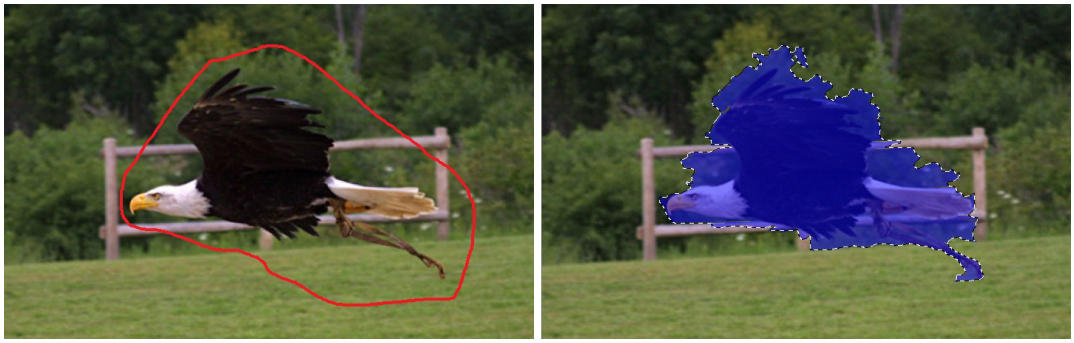
提案システムによるオブジェクト抽出の手順は以下の通りである：

1. オブジェクト領域をユーザが粗く囲む
2. ユーザが指定した線をスーパーピセルベースに拡張
3. GrabCut によりオブジェクト領域を抽出
4. 正確に抽出できなかった領域をユーザが指定
5. 対象オブジェクトを抽出できるまで 2 から 4 を繰り返す

前景物抽出のため，まずユーザはなげなわツールのようなインタフェースを用いて前景物領域を粗く囲む（図 3.4(a)）．囲まれた領域の外側は背景領域とし，さらにガイド線が含まれる各領域は背景であると推測できるため，その領域もすべて背景とみなす（図 3.4(d)）．これを初期状態とし，さらに Rother らの GrabCut と呼ばれる手法 [66] にもとづいて領域の最適化を行う（図 3.4(e)）．通常，GrabCut はグラフカットによる分割結果から前景と背景の色分布を再学習し，反復的に前景分割を行うことでその精度を向上させている．しかし，反復処理は収束まで時間がかかりインタラクティブな編集には向かないことが多い．本システムでは，先に領域分割を用いた領域の最適化を行っておくことで繰り返し処理を行わずに 1 度だけの GrabCut による最適化で十分な前景物分割が可能となる．

しかし正確な前景物抽出は困難な課題であり，上記の処理だけではうまく抽出されない場合がある．GrabCut ではユーザが明示的に前景や背景領域を指定し，この情報をもとに再度分

(a) 入力画像　　(b) 領域分割



(c) 物体領域の指定　　(d) スーパーピクセルによる抽出

(e) GrabCut による抽出　　(f) 抽出されたオブジェクト

図 3.4: オブジェクトの抽出．入力画像 (a) は前処理としてスーパーピクセルに分割される (b). (c) ユーザがオブジェクトを粗く囲むと，(d) そのガイド線の外側とガイド線が含まれている領域が背景として処理される．(e) この情報をもとに GrabCut[66] を適用することで，(f) 前景領域が抽出される．

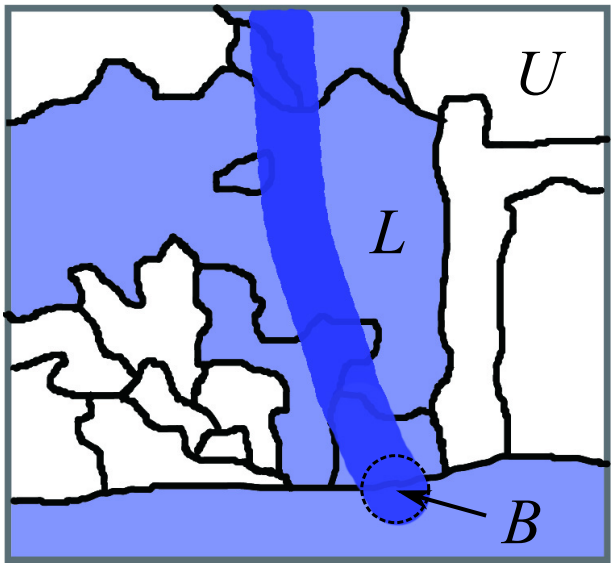
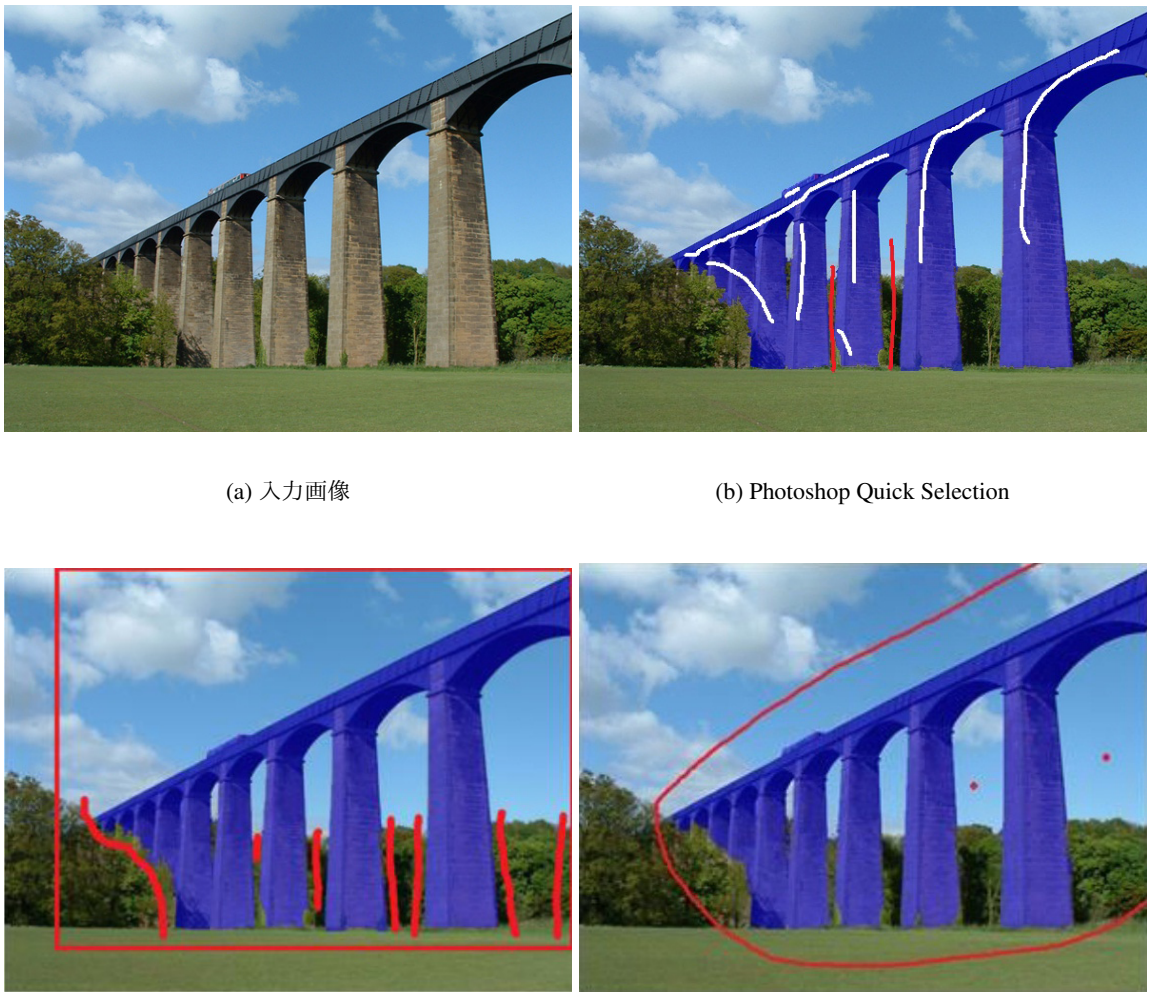


図 3.5: スーパーピクセルを用いたラベリング．B はユーザ入力，L は選択済みのスーパーピクセル，U は未知のスーパーピクセルを表している．

割を行うことでより正確な前景抽出を行う編集機能が示されている．本システムではこれに上記の領域分割を用いた最適化を追加し，ユーザによる前景/背景指定，領域分割を用いた指定領域の拡張（図 3.5），グラフカットによる最適化の 3 段階を繰り返すことでより粗く少ないユーザ入力で正確な前景物抽出を行うことができる．

本システムによる前景物抽出結果を図 3.6 に示す．入力画像のサイズは 800 × 600 ピクセルである．提案手法のユーザ入力に対する最適化処理時間は約 0.41 秒であった．また，ユーザの作業時間も含めて前景物抽出にかかった時間は，商用の画像編集ツールである Adobe Photoshop CS5 [1] の Quick Selection ツールでは 107 秒，我々の実装による Rother らの GrabCut では 115 秒，提案手法では 18 秒であった．よって図 3.6 に示す例では，提案手法は他の手法と比べ，前景抽出にかかる時間が約 85% 低減されている．ただし，本手法の前景抽出精度は GrabCut による最適化に依存しており，既存手法よりも精度が向上するわけではない．しかし他の手法に比べてより粗く少ないユーザ入力で前景物を抜き出すことができ，ユーザの負担を軽減することが可能となる．

提案手法における前景物抽出は，前処理で行われる領域分割において背景と前景物が同じ領域として分割された場合にはうまく最適化が行えない．しかし本システムでは領域分割を用いた最適化を行わず，ブラシによる直接塗りつぶしと GrabCut による最適化だけを用いた前景抽出や，単純にブラシによる塗りつぶしだけで前景物を抽出することもできるため，ほぼすべての前景物に対応できる．



(a) 入力画像　(b) Photoshop Quick Selection

(c) GrabCut　(d) 提案手法

図 3.6: オブジェクトの抽出結果．赤線が背景のガイド線，白線が前景のガイド線，青領域が抽出されたオブジェクト領域を表している．(a) 入力画像が与えられたとき，(b)(c) Photoshop CS5 の Quick Selection や Rother らの GrabCut に比べ，(d) 提案手法は粗く少ないユーザ入力でオブジェクトを抽出できている．

3.2.4 前景物のモデリング

このようにして抽出された前景物画像を 1 枚の四角形ポリゴンに前景物テクスチャとしてマッピングすることで前景物モデルが生成され，背景モデルの地面領域に垂直に配置される．3 次元モデルにおけるポリゴンのそれぞれの頂点の座標は，背景モデルの最も奥に位置する頂点座標から式 3.1 と同じ要領で求められる．

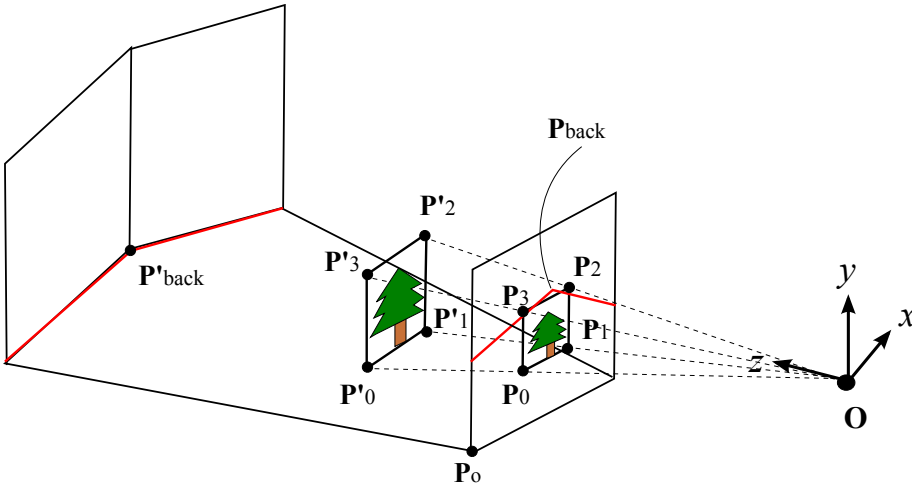


図 3.7: 前景物モデルの頂点座標. 入力画像の前景物の座標から 3 次元シーンにおける前景物の座標を算出する．

図 3.7 は本手法による前景物モデリングである．ここでは 2 次元画像の中で抽出された前景物領域が四角形で囲まれており，その頂点は $\mathbf{P}_0, \mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3$ で表されるとする．この指定された前景物は $\mathbf{P'}_0, \mathbf{P'}_1, \mathbf{P'}_2, \mathbf{P'}_3$ を頂点とするポリゴンとしてモデル化され，背景モデルに設置される．背景モデルにおいて最も奥に位置する頂点を $\mathbf{P}_{back}(x_{back}, y_{back}, f)$ とし，カメラからの視線方向を $+z$，上方向を $+y$，焦点距離を f とすると，それぞれの頂点の座標は以下のように与えられる．

$$\mathbf{P}_0 : (x_0, y_0, f) \qquad \mathbf{P}_1 : (x_1, y_1, f)$$
$$\mathbf{P}_2 : (x_2, y_2, f) \qquad \mathbf{P}_3 : (x_3, y_3, f)$$

ここで，画像の左下 $\mathbf{P}_o$の同次座標として $(x_o, y_o, f, 1)$，$\mathbf{P}_{back}$ の 3 次元モデル上の頂点 $\mathbf{P'}_{back}$ の同次座標として $(x_{back}, y_{back}, f, w_{min})$ が与えられると，式 3.1 からそれぞれの頂点は以下のように定まる．

$$\mathbf{P'}_0 : (x_0, y_0, f, w_0) \qquad \mathbf{P'}_1 : (x_1, y_1, f, w_1)$$
$$\mathbf{P'}_2 : (x_2, y_2, f, w_2) \qquad \mathbf{P'}_3 : (x_3, y_3, d, w_3)$$

このような頂点座標の算出をすべての前景物について行い，算出された座標に基づいて前景物を背景モデルに設置することで，シーンモデルが完成する．

ビルボード変換

前景物モデルは板状ポリゴンであるため，視点を横へ移動すると前景物の立体感が失われてしまう．これを解決するため，本システムではビルボード変換を前景物モデルに適用する．ビルボード変換とは対象のポリゴンが常に視点方向を向くように座標変換を施す手法であり，本システムでは木や円筒状の前景物に適用される．このビルボード変換行列 Π_b は以下のように表される．

$$\Pi = T^{-1}RT \tag{3.4}$$

ただし，

$$T = \begin{pmatrix} E & \mathbf{V} \\ \mathbf{0}^t & 1 \end{pmatrix}, R = \begin{pmatrix} R_y & \mathbf{0} \\ \mathbf{0}^t & 1 \end{pmatrix} \tag{3.5}$$

この式において E は単位行列，$\mathbf{V}$ は視点座標と対象ポリゴンの回転軸座標の差を表すベクトル，R_y は y 軸周りの回転行列である．これにより，対象の前景物ポリゴンが常に視点方向を向くようになり，前景物の立体感を表現することができる．

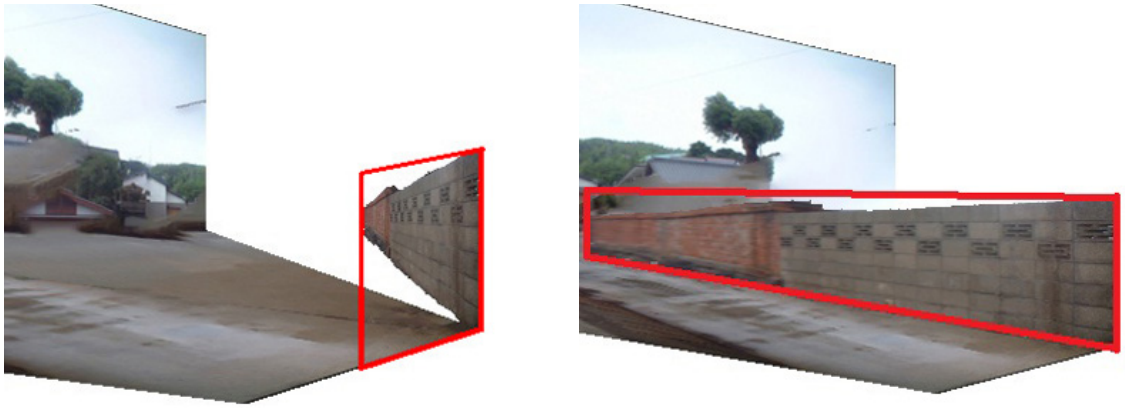
接地制約

四角形の板状ポリゴンである前景物モデルの各頂点には同じ奥行きが与えられている．しかし前景物が奥に向かって地面領域に置かれている場合には不自然なモデルが生成されてしまう（図 3.8）．そこで本システムでは前景物が地面領域に接している部分を直線で指定することで，前景物モデルの 3 次元座標を修正できる機能を実装している．この機能を用いることで前景物に立体形状を与えることも可能となる．このような前景物モデルは回転すると不自然なため，ビルボード変換は適用しない．

並列処理

上記で述べた画像補完や前景物最適化処理の高速化はインタラクティブな編集には重要な問題である．Barnes らは画像をタイル状に分割して並列に処理することでより高速に類似領域探索が行えることを示した．本システムではこの並列処理手法を用いて，画像の補完を複数の CPU コアによって並列処理している．これによりコア数に比例して類似領域探索処理が高速化され，2 コアで約 40% の処理速度の向上がみられた．

(a) 入力画像　　(b) 前景物と設置制約



(c) 制約なし　　(d) 制約あり

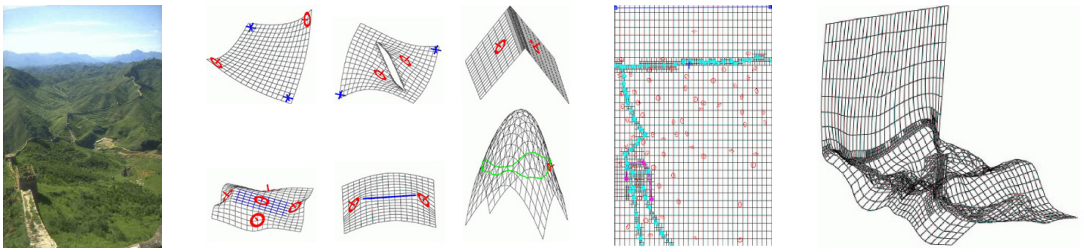
図 3.8: 前景物の設置制約.

3.3 考察

本節では提案手法の 3 次元シーン生成と従来手法について例を示しながら考察する．

従来手法の中でも，単視点画像から 3 次元シーンを自動生成する手法 [31, 67, 68] は木や車などの前景物があると 3 次元シーン生成に失敗する．これは，前景物を自動で認識し背景と区別してモデリングすることが非常に困難であることが主な原因である．意味のあるオブジェクト領域を自動認識する技術はコンピュータビジョンにおいて重要な研究分野であるが，これは未だに困難な問題である．これに対し，人間はたとえ子供であってもいとも簡単に画像中の前景物領域を区別することができる．本システムのキーとなるアイデアは，人間の認識能力と計算機の正確な処理能力を最大限に生かして 3 次元シーンモデルの構築を行うことである．すなわち，前景物や境界線の大まかな指定は人間が行い，前景物の正確な領域抽出や背後領域の合成などは計算機に行わせることで，幅広い画像の 3 次元シーンを効率よくモデリングすることを可能にしている．

はじめに述べた通り，本システムの目的は専門知識をもたないユーザでも 1 枚の画像の 3 次元ウォークスルーができるようなモデルを生成できるようにすることである．従来の対話的なシーンモデリング手法 [57, 85, 53] は，法線の指定や照明効果の計算によって正確なシーン形状をモデリングできる．しかし，これらの手法は手動によるレイヤ構造の生成やデプスの指定，法線や不連続箇所の指定など，非常に多くの複雑なユーザ入力が必要となる（図 3.9）．これはユーザの大きな負担となるため，本研究目的には適さない．



(a) 入力画像　(b) 制約条件　(c) 制約の指定　(d) メッシュモデル

図 3.9: Zhang らの手法による 3 次元シーンモデルの生成 [85].
出典：Single View Modeling of Free-Form Scenes[85]

Tour Into the Picture と呼ばれる手法 [36, 41] は本研究と同じ目標にもとづいた 3 次元モデリング手法である．すなわち，なるべくユーザ入力を単純なものに限定し，少数のポリゴンから構成される簡易な 3 次元シーンモデルによって十分な 3 次元効果を得られる自由視点映像を実現することができる．しかし，この手法は前景物体の正確な抽出や背後領域の合成などはユーザが個別にペイントツールなどを用いて行う必要があるため，全体的なシーンモデリングには手間がかかる．また，生成される 3 次元シーンモデルは消失点をもつスパイダリーメッシュや消失線にもとづいているため，それが当てはまらないような画像ではうまくいか

ない．これに対し，提案手法は境界線と効率的な前景物抽出，および背後領域の自動合成によって幅広いシーンを効率よく 3 次元モデリングすることができる．

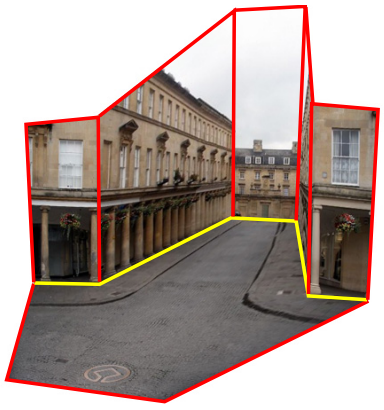
図 3.10, 3.11 は提案手法と既存手法との比較を示している．既存手法は折れ曲がった地面境界や曲線状の地面境界をもつような画像ではうまく 3 次元モデルを当てはめることができないが，提案手法では折れ線を調整するだけで良好な 3 次元シーンモデルを生成することができる．

(a) 入力画像

(b) Hoiem らの手法 [31]

(c) Horry らの手法 [36]

(d) Kang らの手法 [41]



(e) 提案手法

図 3.10: 提案手法と既存手法の比較．黄線がユーザ入力，赤線が生成されたモデルのワイヤフレームである．(a) 入力画像に対し，(b)-(c) 既存手法は地面領域と建物をうまく分離できていないが，(e) 提案手法は地面と壁が分割された良好なモデルが生成されている．

(a) 入力画像

(b) Hoiem らの手法 [31]

(c) Horry らの手法 [36]

(d) Kang らの手法 [41]

(e) 提案手法

図 3.11: 提案手法と既存手法の比較．黄線がユーザ入力，赤線が生成されたモデルのワイヤフレームである．このように提案手法は地面の境界が曲線状になっていても近似的にモデル化することができる．

3.4 結果

3.4.1 実装・実行環境

本システムはライブラリとして OpenGL，GLUT，GLUI，OpenMP を用いて C++言語で実装し，Intel Core i7 620M (2.67GHz, 4.00GB RAM) と NVIDIA Quadro NVS 3100M グラフィックカードが搭載された PC 上で実行した．使用した画像のサイズは全て 0.5 から 1.0 メガピクセルの範囲内である．

3.4.2 結果の考察

図 3.1 では 2 基の街灯が前景物として指定され，背景の境界が 5 個の頂点をもつ折れ線で指定されている．この街灯はビルボード変換が適用され，横からの視点に対しても立体感のある自然な前景物が生成される．街灯のような円柱状の物体や一般的な木などの前景物は，回転してもあまり形状が変わらない場合が多く，ビルボード変換で十分に対応することができる．しかし，本手法では地面や壁のそれぞれの領域は 1 つの平面としてモデル化されているため，視点によっては階段が後ろの建物に張り付いているような違和感をユーザに与えることがある．

図 3.10（e）は前景物がない画像を本手法によって 3 次元モデリングした結果である．このように前景物がない画像は境界線を指定するだけで 3 次元シーンを生成することができる．また，図 3.12 の 1 段目の例では建物に折れ線で接地制約を与えることで，建物が 2 枚のポリゴンで立体的にモデル化されている．このように接地制約を折れ線で指定することにより，通常 1 枚の板状ポリゴンモデルである前景物を，複数のポリゴンから構成される立体的なモデルにすることができる．さらに図 3.12 の 2 段目のように，入力が絵画のように複数視点の情報が得られないようなシーンであっても，提案システムによりその 3 次元シーンを作成することができる．

これらの 3 次元シーンを作成するのにかかった時間はすべて 3 分以下であった．この作業時間において大きな割合を占めるのは，入力画像の前景物抽出にかかる時間である．例えば前景物がない図 3.10（a）の場合，3 次元シーンを作成するのにかかった時間が 10 秒程度であったのに対し，2 つの前景物をもつ図 3.12 の 3 段目では 1 分程度の作業時間がかかった．このように 3 次元シーンを作成するための作業時間は，前景物の数に応じて増加していく．よって，提案システムにおける前景物抽出の効率化は 3 次元シーンを作成する上で重要な役割を果たしていることがわかる．

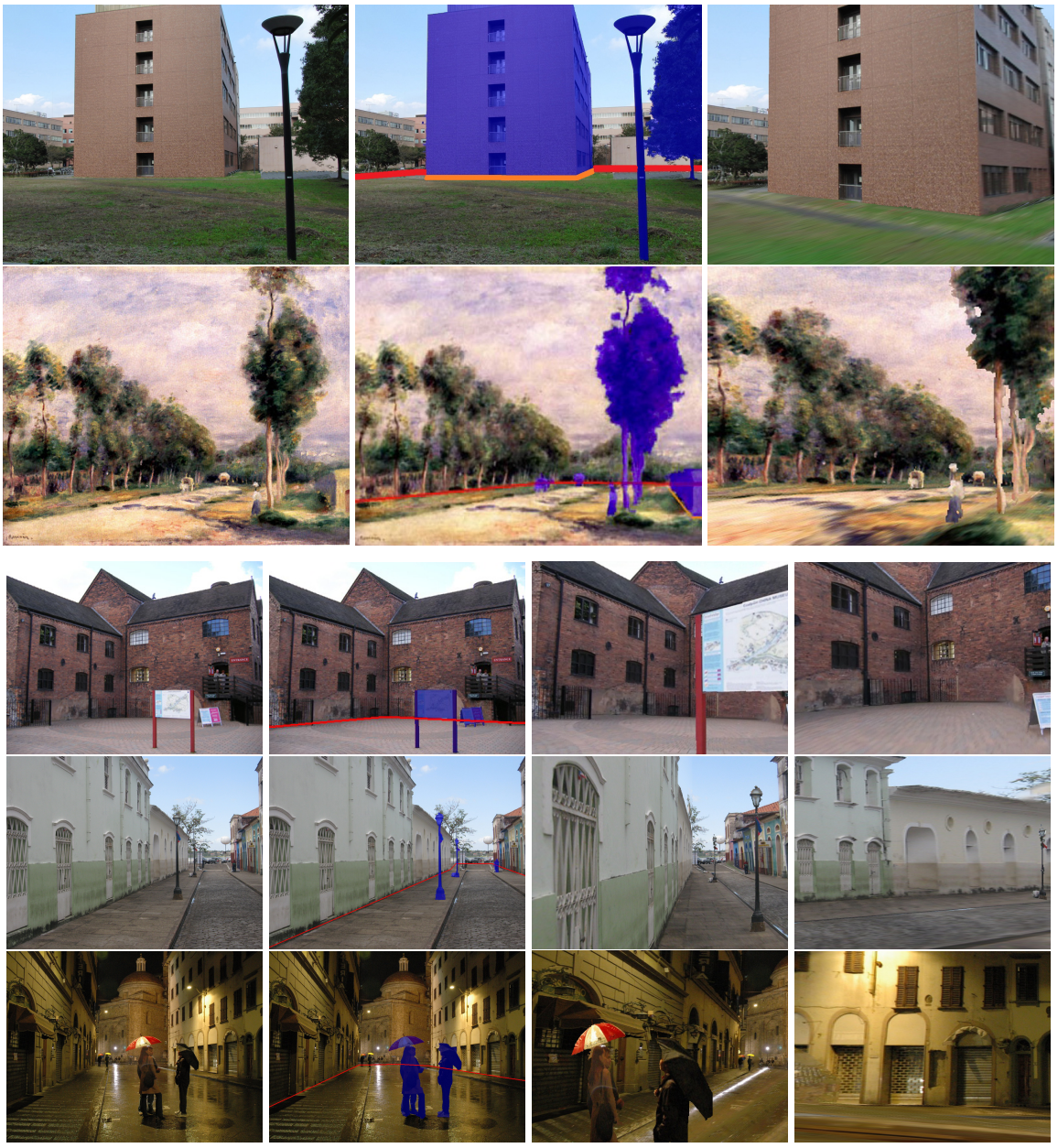


図 3.12: 提案手法による 3 次元シーンの生成．左から入力画像，ユーザ入力，生成された 3 次元シーンで視点移動した結果である．

第4章　少数のデプス入力にもとづく3次元モデルの生成

前章で述べた手法は，平坦な地面をもつ景観画像において良好な 3 次元シーンを生成できる．しかし，地面が写っていない写真やオブジェクトを近くから撮影した写真など，異なる環境のシーンに対応するには別の方法で 3 次元情報を推定する必要がある．そこで本章では，少数のデプス入力によって滑らかな表面形状をもつ 3 次元モデルを 1 枚の画像から生成する手法について述べる．はじめに従来手法について再度簡単に紹介した後，本手法で目標とする 3 次元シーンモデルの特徴と手法の概要を述べる．続いて，提案手法を構成するそれぞれの処理工程について，既存手法との考察も含めて詳しく述べる．その後，提案手法によって生成される 3 次元モデルの例を示しその有効性について述べる．ここでは既存手法と提案手法の比較も示すことで，より詳しい議論を行う．

4.1 概要

2 章で議論したように，単視点静止画像のみから 3 次元モデルを生成することは，シーンの奥行きの曖昧性や前景物体による遮蔽のために困難な問題である．そこで前章では景観画像の地面の境界線にもとづき，数枚の平面ポリゴンで構成される 3 次元シーンモデルを対話的に生成する手法について提案した．この手法では，平坦な地面をもつ広い景観画像から 3 次元シーンを効率よく作成できる．本章ではより多様なシーンに対応できるように，ユーザが画像上でスパースに入力したデプスから滑らかな表面形状をもつ 3 次元シーンモデルを生成する手法を提案する．この手法は前章の手法に比べユーザ入力が増えるが，地面が写っていない写真やオブジェクトを近くから撮影した写真など幅広いシーンに適用できる．入力画像のシーンによって前章と本章の手法を適宜使い分けることで，用途に合った幅広い 3 次元シーン構築を実現できる．

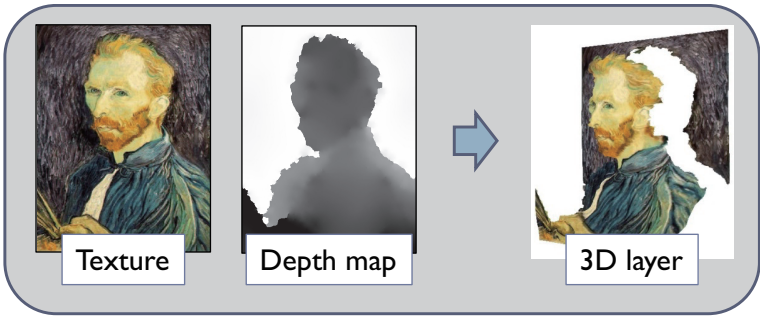
本手法では幅広い単視点画像の 3 次元シーンを表現するため，その 3 次元モデルを Layered Depth Image(LDI)[70] として表現する．本手法における LDI は前景レイヤと背景レイヤから構成され，それぞれのレイヤはテクスチャとデプスマップを保持している（図 4.1）．LDI は滑らかな表面形状をもつ 3 次元シーンを表現できるだけでなく，背景レイヤによって前景物の背後領域が「穴」になることを防ぐことができる．このように LDI はより立体的な 3 次元シーンを表現できるが，これを作成するには効率的なデプスマップの生成や前景と背景の切れ目（不連続箇所）の抽出，遮蔽領域のテクスチャとデプスの推定と合成などの課題がある．

そのような LDI を効率よく構築するため，提案手法は主に以下の 3 つの要件を満たすように設計されている．

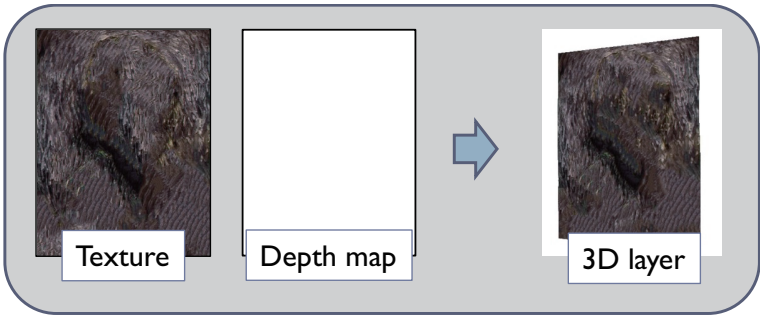
- できるだけ少数のユーザ入力
  提案手法では人間の奥行き知覚能力を生かしユーザによるデプス入力を用いてシーンの奥行きを推定するが，この入力をなるべく少なくすることでユーザの負担を減らす必要がある．そこで提案手法では少数のデプスストロークを画像全体に伝播させることでデプスマップを生成するようにする．このようなデプス伝播ベースの手法は Wang らなどによって提案されているが [75, 65]，これらの手法はデプスを入力した位置から離れた領域にはうまくデプスが伝播せず計算に時間がかかるため，対話的な編集に向かない．

- 不連続箇所を考慮した滑らかなデプスマップ
  生成されたデプスマップから 3 次元モデルが生成される．このときデプスマップのわずかなノイズも 3 次元モデル上では強調され，粗い表面形状が生成されてしまう．このため，デプスマップはデプスが不連続になっている箇所を除き十分に滑らかである必要がある．

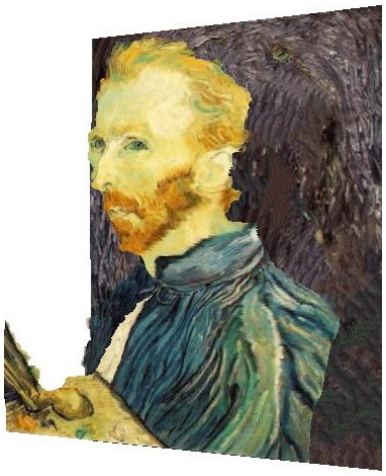
- 実時間のフィードバック
  ユーザ入力が必要なシステムにおいてインタラクティブ性は効率的な編集のために重要である．このためには上記のデプスマップの生成を高速に計算し，ユーザ入力に合わせて計算結果を表示することで直感的な編集が可能となる．



(a) 前景レイヤ



(b) 背景レイヤ



(c) Layered Depth Image

図 4.1: 本手法における Layered Depth Image.

これらを踏まえ，本研究ではスパースなデプス入力のみで高速に LDI を生成できる手法を提案する．提案手法では，ユーザが入力したデプスを均質な小領域（スーパーピクセル）ベースの重み付き測地距離とエネルギー最適化によって画像全体に伝播させることでデプスマップを計算する．提案手法によるデプス伝播は実時間で計算されるため，ユーザは対話的に LDI を作成することができる．さらに，提案手法は奥行きが大きく変化する不連続箇所から遮蔽領域を推定し，遮蔽領域のテクスチャとデプスを含む背景レイヤを自動生成する．また，ユーザがデプスを直感的に指定できるようにするため，3 次元モデルに直接ストロークを引き，3 次元モデルの変化を確認しながらストロークのデプスを変化させていくインタフェースを採用している．

本研究の主な貢献は以下のとおりである：

- LDI を効率的に生成するためのワークフローの提案
- スーパーピクセルベースの重み付き測地距離とエネルギー最適化による効果的なデプスの伝播
- 直感的なインタフェースの設計

提案手法によって単純な操作のみで十分な 3 次元効果を得られる 3 次元モデルが生成できることを示す．

4.2 提案手法の流れ

提案手法の流れを図 4.2 に示す．提案手法では，まずユーザはスパースにデプスを入力する．これをもとに以下の情報が自動で算出される：

- 画像全体のデプスマップ
- 奥行きが大きく変化する不連続箇所
- 遮蔽領域を含む背景テクスチャ
- 遮蔽領域を含む背景デプスマップ

はじめに，入力したデプスにもとづき画像全体のデプスマップを計算する．次に算出したデプスマップからデプスが大きく変化する不連続箇所を抽出する．この不連続箇所で前景レイヤと背景レイヤが分割される．最後に，不連続箇所から遮蔽領域を抽出し，遮蔽領域のテクスチャとデプスを計算することで遮蔽領域を含む背景レイヤを生成する．以下ではこれらの処理の詳細について述べる．

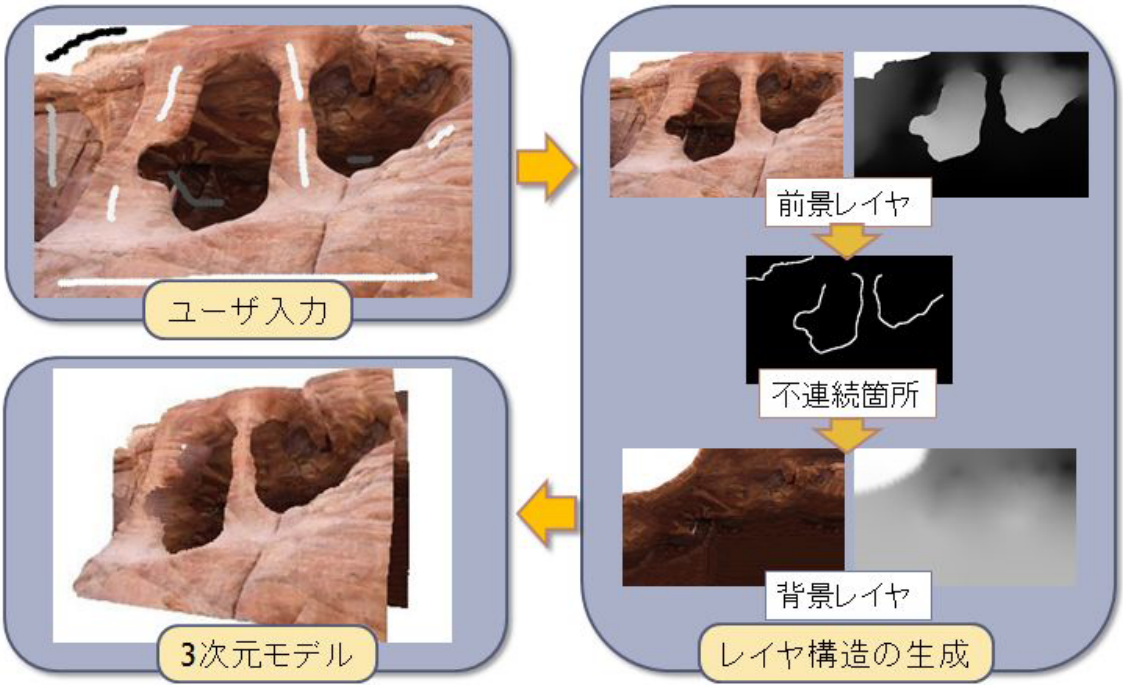


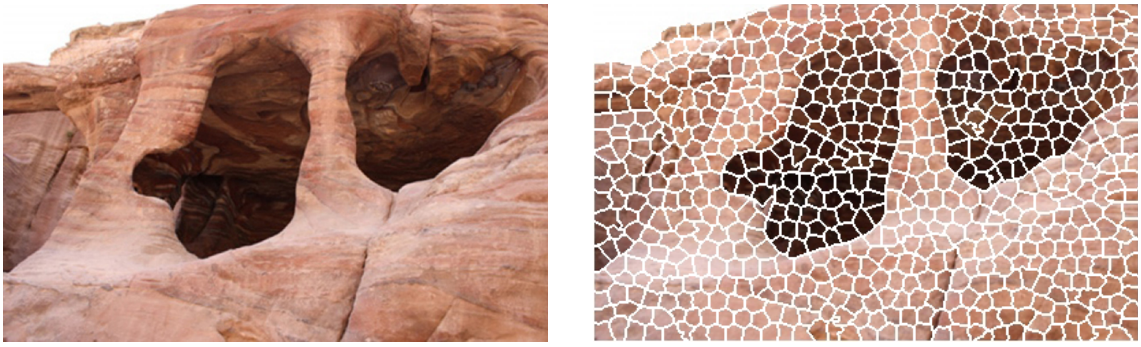
図 4.2: 提案システムの概要． ユーザがデプスを入力すると，システムがデプスマップを計算し, 前景物レイヤが生成される．次にこのデプスマップから不連続箇所が計算される．不連続箇所とデプスマップから遮蔽領域のテクスチャとデプスを含む背景レイヤが生成される．この前景レイヤと背景レイヤを重ねることで 3 次元モデルが生成される．

4.3 デプスマップの計算

提案手法ではスパースなデプス入力を画像全体に伝播させることでデプスマップを計算する．既存のエネルギー最適化ベースのデプス伝播 [75, 65] は，デプスを入力した位置から離れた領域にはうまくデプスが伝播せず計算にも時間がかかる．これに対し，提案手法ではスーパーピクセルベースの重み付き測地距離にもとづきデプスを伝播させる．これにより，デプス入力から離れた位置にも効果的にデプスが伝播する．測地距離は線形時間で計算できるため，実時間でのデプスマップ生成が可能になる．

スーパーピクセルの利用

デプスマップを計算するため，本手法では最初に入力画像をスーパーピクセルに分割する．ピクセルの代わりにスーパーピクセルを用いることで，計算時間を大幅に削減できるだけでなく画像のノイズによるデプスマップのノイズを避けることができる．提案手法をピクセルベースで計算した結果とスーパーピクセルベースで計算した結果を比較した結果は図 4.14 で示されている．



(a) 入力画像　　(b) スーパーピクセル

図 4.3: スーパーピクセルの生成．

スーパーピクセルの計算には MeanShift による領域分割 [22] やグラフベースの領域分割 [27] が広く用いられている．近年，Simple Linear Iterative Clustering(SLIC) と呼ばれるスーパーピクセル生成手法が提案されている [6]．この手法では，画像を格子状に分割し，格子を画像の勾配を考慮して変形していくことで領域分割を行っている．これにより，従来手法よりも比較的均一な形状をもつスーパーピクセルを高速に生成できる．本手法では，SLIC アルゴリズムによるスーパーピクセルを計算に利用する．本論文で示す結果では，スーパーピクセルの数は全ピクセル数の 10% に設定している．

4.3.1 重み付き測地距離

スーパーピクセルの初期デプスを計算するため，本手法ではまず全スーパーピクセルからユーザが入力したデプスストロークを含むスーパーピクセルまでの測地距離を計算する．その後，算出した測地距離によって各ストロークのデプス値を混合することで，画像のエッジを考慮したデプスマップを計算できる．測地距離は画像の色付け [81] や前景物抽出 [24, 9] など様々な画像編集技術で用いられている．

Chaurasia ら [14] は，マルチビューステレオアルゴリズムで復元できなかったデプスを合成するためにスーパーピクセルベースの測地距離を利用する手法を提案している．この手法では，対象のスーパーピクセルが属しているオブジェクトに含まれるスーパーピクセルを探索するために測地距離を用いている．探索したスーパーピクセルのデプスをユークリッド距離と確率密度関数によって混合することでデプスマップを生成している．この手法と提案手法の大きな違いは，提案手法は測地距離を各デプス値を混合するための重みとして用いる点である．これにより，画像のエッジを考慮して滑らかに変化するデプスマップを生成することができる．提案手法と Chaurasia らの手法の比較は図 4.12 で示す．

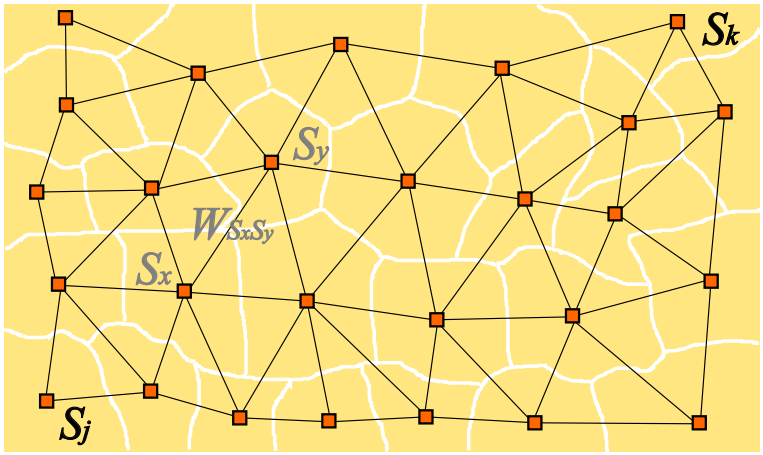


図 4.4: スーパーピクセルベースの測地距離の計算．

ユーザが異なるデプス値をもつ L 本のストロークを入力したとして，ストローク $l(=1,2,...,L)$ がもつデプスを $d_l \in [0,1]$ とする．ここでスクリブル l に含まれるスーパーピクセルの集合を Ω_l とし，Ω_l 内のスーパーピクセルはすべて同じデプス値 d_l をもつものとする．ストローク l からスーパーピクセル S_k までの測地距離 D_l は以下の式で計算される．

$$D_l(S_k) := \min_{S_j \in \Omega_l} dist(S_j, S_k) \tag{4.1}$$

$$dist(S_j, S_k) := \min_{C(S_j,S_k)} \sum_{S_x,S_y} W_{S_x S_y} \tag{4.2}$$

ここで $C(S_j, S_k)$ はスーパーピクセル S_j と S_k をつないだ経路，S_x と S_y は経路上の隣接したスーパーピクセルである．重み $W_{S_xS_y}$ はスーパーピクセル S_x と S_y からランダムにサンプリングした Lab 色空間のピクセル値の 2 乗距離の和で定義される．

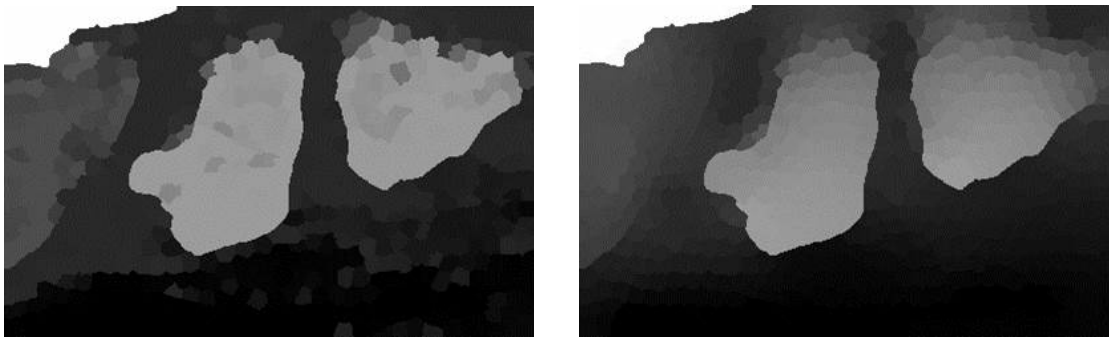
$$W_{S_xS_y} = \sum_{s \in S_x, t \in S_y} \|\mathbf{c}_s - \mathbf{c}_t\|^2 \tag{4.3}$$

ここで s と t はそれぞれスーパーピクセル S_x と S_y からサンプリングされたピクセルであり，$\mathbf{c}_s$ と $\mathbf{c}_t$ はピクセルの色を表している．この測地距離は改良ダイクストラアルゴリズム [80] により線形時間で計算できる．

算出した測地距離に応じて各ストロークのデプス値をブレンドすることで，スーパーピクセル S_k のデプス値 $G(S_k)$ が計算される．

$$G(S_k) = \frac{\sum_l D_l(S_k)^{-b} d_l}{\sum_l D_l(S_k)^{-b}} \tag{4.4}$$

ここで b は測地距離の影響力を定める定数であり，提案手法では $b = 2$ としている．この値が大きいほど測地距離が短いスクリブルのデプス値が強調されるようになる．図 4.5(a) はスーパーピクセルベースの測地距離を用いてデプスマップを計算した結果を示している．



(a) 測地距離ベース　　(b) 最適化

図 4.5: スーパーピクセルベースのデプスマップの計算．入力画像とユーザ入力は図 4.2 と同一である．(a) 最初にスーパーピクセルベースの測地距離にもとづきデプスマップが計算され，(b) その後エネルギー最適化によってデプスマップが平滑化される．

エネルギー最適化によるデプスマップの平滑化

測地距離を用いることでユーザ入力から離れた領域にまで効果的にデプスを伝播させることができる．しかし，生成されるデプスマップ G は局所的に粗くなりやすい（図 4.5(a)）．提案手法ではエネルギー最適化によってデプスマップのエッジを保持した平滑化を行う．まず目的となるデプスマップ U のコスト関数 $E(U)$ をデータ項 $E_{data}(U)$ と平滑化項 $E_{smooth}(U)$ で定義する．データ項 $E_{data}(U)$ によって測地距離ベースのデプスマップ G のデプスを保持させつつ，平滑化項 $E_{smooth}(U)$ によって対象のスーパーピクセル S_i のデプスを近傍のスーパーピクセル $N(S_i)$ のデプスの加重平均によって均一化する．これは以下の式で表される．

$$E(U) = E_{data}(U) + \gamma E_{smooth}(U) \tag{4.5}$$

$$E_{data}(U) = \sum_{S_i}(U(S_i) - G(S_i))^2 \tag{4.6}$$

$$E_{smooth}(U) = \sum_{S_i}\left(U(S_i) - \sum_{S_j \in N(S_i)} w_{S_i S_j} U(S_j)\right)^2 \tag{4.7}$$

$w_{S_i S_j}$ は以下の式で表される．

$$w_{S_i S_j} = \exp(-W_{S_i S_j}/2\sigma^2) \tag{4.8}$$

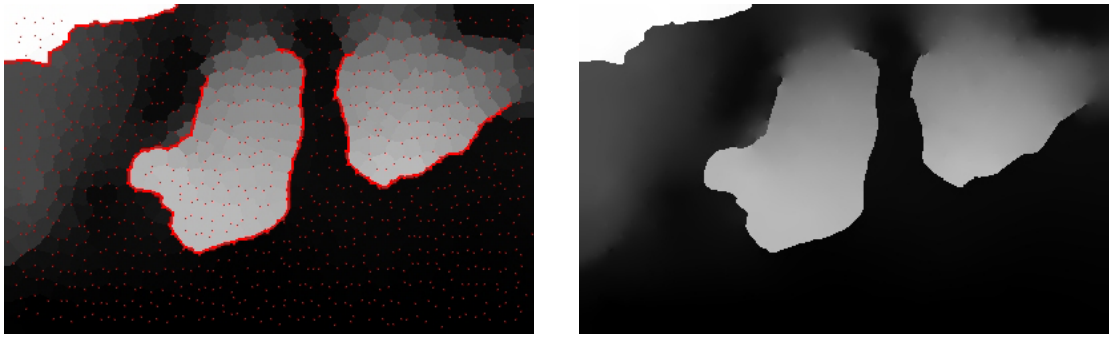
ここで，σ は定数，γ は平滑化の度合いを調整する定数である．デプスマップ U を G で初期化することで，Gauss-Seidel 法による最適化計算はすぐに収束する．この最適化処理により，滑らかなデプスマップが生成される（図 4.5(b)）．

4.3.2 ピクセル単位のデプス割り当て

上記の処理によって算出されたデプスマップはスーパーピクセル単位でデプスが割り当てられているため，スーパーピクセル間でデプスの段差が生じてしまう．滑らかなデプスマップを生成するにはこの段差を解消してピクセル単位のデプスマップを生成する必要がある．しかし，バイラテラルフィルタのような一般的なエッジを保持する平滑化フィルタは，デプスの不連続箇所周辺のデプスなど特定のデプスを保持することができないため適当でない．そのため，提案手法では適切な境界条件を設定してラプラス方程式を解くことでピクセル単位にデプスを割り当てる．ピクセル p のデプスを d_p，境界条件を設定するピクセル以外の領域を Ω とすると，これは以下の式で表される．

$$\Delta d_p = 0 \quad over \quad \Omega, \quad with \quad d_p|_{\partial\Omega} = d_p^0, \tag{4.9}$$

ここで d_p^0 は境界条件として固定されたデプス値である．まず 3 つ以上の異なるデプス値をも



(a) デプスの固定　　(b) ピクセル単位のデプスマップ

図 4.6: ピクセル単位のデプスマップの計算．(a) 3 つ以上の異なるデプス値をもつスーパーピクセルの境界の交点に位置するデプス値を固定し，(b) ラプラス方程式を解くことでピクセル単位のデプスマップが生成される．

つスーパーピクセルの交点を探索し，その位置のピクセルに隣接するスーパーピクセルのデプス値の平均値を割り当て境界条件とする（図 4.6(a) 赤）．また，次節で述べるデプスの不連続箇所周辺のデプスも固定して境界条件とする．この境界条件を用いてラプラス方程式を解くことでピクセル単位で計算された滑らかなデプスマップが得られる（図 4.6(b)）．

4.4 Layered Depth Image の生成

本節では LDI を構築するための残りの工程である不連続箇所の抽出，遮蔽領域の抽出，遮蔽領域のテクスチャとデプスの計算について述べる．

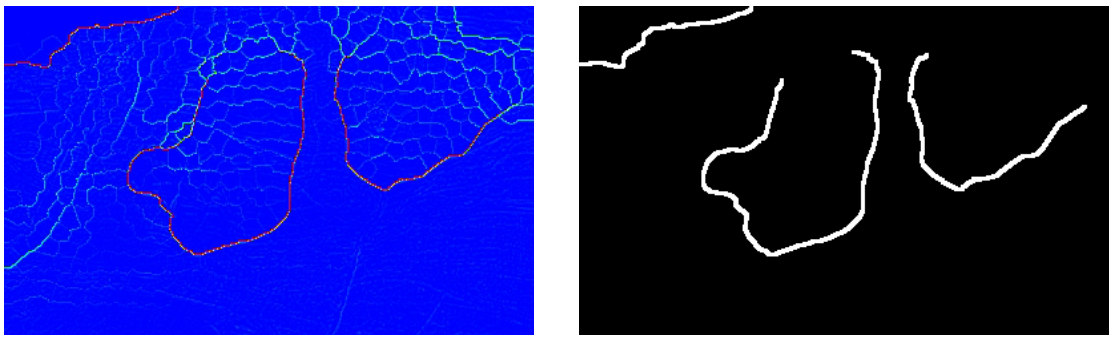
4.4.1 不連続箇所の抽出

まず，算出したスーパーピクセルベースのデプスマップにおいて，スーパーピクセル間でデプスが大きく変化する箇所は領域が不連続になっていると考えられる．この不連続箇所を抽出することで，前景と背景レイヤを分割でき，また遮蔽領域の検出のための手がかりとして用いることができる．

本手法では，まず前節で計算したデプスマップのエッジ E_d と画像のエッジ E_i をラプラシアンフィルタで検出し，これらを重み付きで結合してエッジ強度 E_s を定義する．これは以下の式で表される．

$$E_s = w_d E_d + w_i E_i \tag{4.10}$$

ここで w_d, w_i はそれぞれデプスエッジと画像エッジの重みであり，本手法では $w_i = 0.2, w_d = 0.8$ としている．計算されたエッジ強度を閾値処理することで不連続箇所を抽出している（図 4.7）．



(a) エッジ強度　　(b) 不連続箇所

図 4.7: 不連続箇所の抽出．(a) エッジ強度をデプスマップのエッジと画像のエッジから計算し，(b) 閾値処理を行うことで不連続箇所を抽出する．

4.4.2　遮蔽領域の抽出

次に抽出した不連続箇所から遮蔽領域を推定する．不連続箇所周辺ではデプスが大きく変化しているため，不連続箇所周辺のピクセルはそのデプス値によって前景と背景に容易に分類できると考えられる．提案手法では以下の手順で遮蔽領域を抽出する．

1. 不連続箇所を膨張処理
2. それぞれの不連続箇所のデプスヒストグラムを生成
3. ヒストグラムの閾値を計算
4. 閾値処理により Trimap を生成
5. バイナリラベリングにより遮蔽領域を抽出

まず前節で抽出した不連続箇所をモルフォロジー演算により膨張させる（図 4.8(a)）．ここでは膨張幅は 5 ピクセルとしている．次にそれぞれの不連続箇所のピクセルのデプスからヒストグラムを構築する．このヒストグラムから判別分析法 [60] で閾値を計算する．この閾値より大きなデプス値をもつピクセルを「背景」，小さいデプス値を「遮蔽」，それ以外のピクセルを「未知」として分類することで Trimap を生成する（図 4.8(b)）．この Trimap をもとに測地距離ベースのバイナリラベリング [9] によって遮蔽領域が抽出される（図 4.8(c)）．

(a) 不連続箇所の膨張処理　(b) Trimap　(c) 遮蔽領域

図 4.8: 遮蔽領域の抽出．(a) まず不連続箇所を膨張させ，(b) その中で背景（黒）と遮蔽（白）を分類することで Trimap を生成する．(c) Trimap を用いて，測地距離ベースのバイナリラベリング [9] により遮蔽領域（白）を抽出する．

4.4.3　遮蔽領域のテクスチャとデプスの推定

最後に，検出した遮蔽領域のテクスチャとデプスを推定する．まずはじめに遮蔽領域のテクスチャ生成を行う．このテクスチャ合成には，3.2.2 節で示した高速な類似パッチ探索を用いた手法 [10] を用いる．これにより，遮蔽領域を含む背景レイヤのテクスチャが生成される（図 4.9(a)）．

次に背景レイヤの遮蔽領域を含むデプスマップを計算する．まず背景デプスマップを入力画像のデプスマップの遮蔽領域以外のデプスで初期化する．これをもとに残りの領域にデプスを式（4.5）を用いて伝播させることで背景レイヤのデプスマップが算出される（図 4.9(b)）．なお背景レイヤのデプスマップの計算には背景テクスチャを用いる．

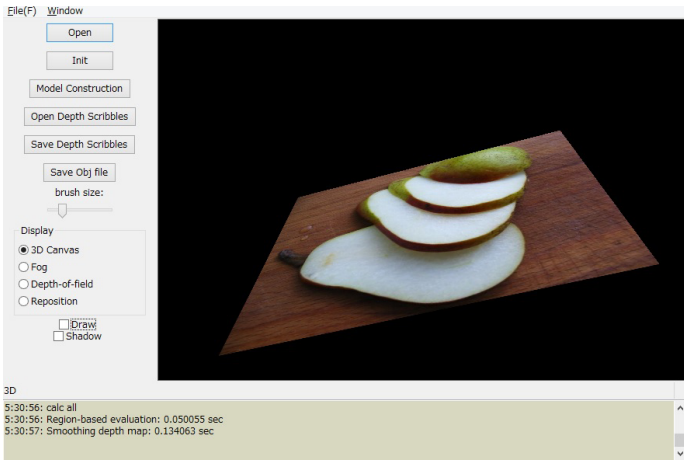
(a) 背景テクスチャ　(b) 背景デプスマップ

図 4.9: 遮蔽領域のテクスチャとデプスを含む背景レイヤのテクスチャとデブスマップ．

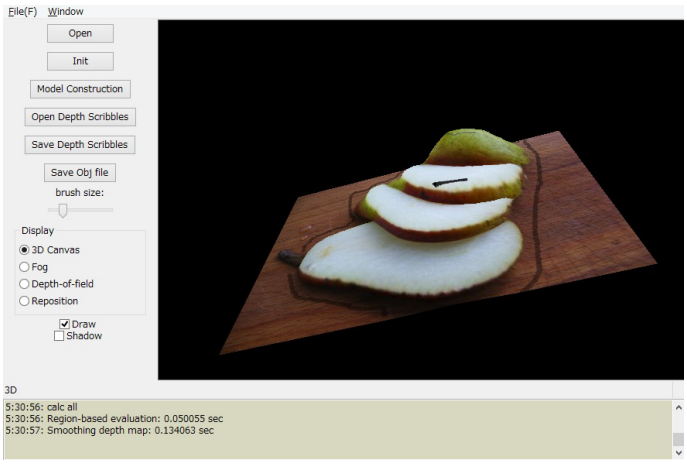
4.5　ユーザインタフェース

本節では，提案手法によって直感的に LDI を作成するためのインタフェースについて述べる．

図 4.10 は提案システムのスクリーンショットである．提案システムでは，ユーザは 3 次元空間上に配置された入力画像にブラシストロークでデプスを直接ペイントしていく．このブラシストロークのデプスをマウスホイールによって変化させることで，再度デプスを指定し直すこともできる．指定されたデプスはただちに画像全体に伝播し 3 次元モデルが生成されるため，ユーザは 3 次元モデルを確認しながら対話的に編集することができる．



(a) 初期状態



(b) デプスの編集

図 4.10: 提案手法のインタフェース．

4.6 結果

4.6.1 実装・実行環境

本システムは C++言語で実装し，3.4GHz の CPU と 8GB のメモリおよび NVIDIA Quadro NVS 5200M グラフィックカードが搭載された PC で実行した．

入力として使用した画像のサイズは 0.5 メガピクセルから 1 メガピクセルである．デプスマップを計算するためのそれぞれの工程にかかった平均時間は，測地距離ベースのデプス計算に 0.01 秒，エネルギー最適化に 0.02 秒，ピクセル単位のデプス計算に 0.2 秒，不連続箇所の抽出に 0.02 秒，合計で約 0.25 秒であった．また，背景レイヤの生成に約 5 秒，ユーザ入力まで含めて全体でかかった時間は 2 分以内であった．

4.6.2 3 次元シーンモデルの生成

図 4.11 は提案手法によって生成された 3 次元シーンモデルである．提案手法は風景写真や自画像など幅広いシーンの 3 次元モデルを生成することができる．生成されるモデルは滑らかに変化する表面形状を表現することが可能である．また，LDI によって前景物体の背後領域が「穴」になることなく立体感のある 3 次元シーンを表現できていることがわかる．本手法を様々な画像に試したところ，境界がはっきりとしない物体境界の周辺をデプスストロークで指定することで良好な結果が得られることがわかった．

4.6.3 比較

既存手法との比較と考察

図 4.12 は，同じデプス入力をした場合の既存手法 [18, 47, 75, 14] と提案手法のデプス伝播の比較である．Chen らの手法と Li らの手法によって生成されるデプスマップは非常に粗くなっている．これらの手法はデプスを指定したピクセルと似た色をもつピクセルにそのデプスを伝播させるため，デプスマップの局所的な滑らかさは考慮されず粗くなってしまう．これは全ピクセルとの色と位置によって入力した値を画像全体に伝播される手法 [7, 79] でも，同様の理由でデプスの伝播はうまくいかない．このデプスマップのノイズは 3 次元モデルでさらに強調され，非常に粗い表面形状をもつ 3 次元モデルが生成されてしまう．Wang らの手法 [75] は局所的な滑らかさを考慮してデプスマップを生成している．そのため，生成されるモデルは他の従来手法に比べ滑らかになっている．しかし，この手法ではユーザ入力から離れた領域にはうまくデプスが伝播せず，エッジがぼやけた不自然なデプスマップが生成されてしまう．Chaurasia らの手法 [14] はスーパーピクセル単位でデプスが割り当てられる傾向にあり，デプスマップは平坦で局所的に粗くなっている．これに対し，提案手法はデプス入力がエッジを考慮して滑らかに伝播していることがわかる．

図 4.13 は，似たデプスマップを生成しようとした場合の既存手法と提案手法のデプス入力の比較である．既存手法 [75, 83] に比べ，提案手法は少ない入力で目的のデプスマップを生成

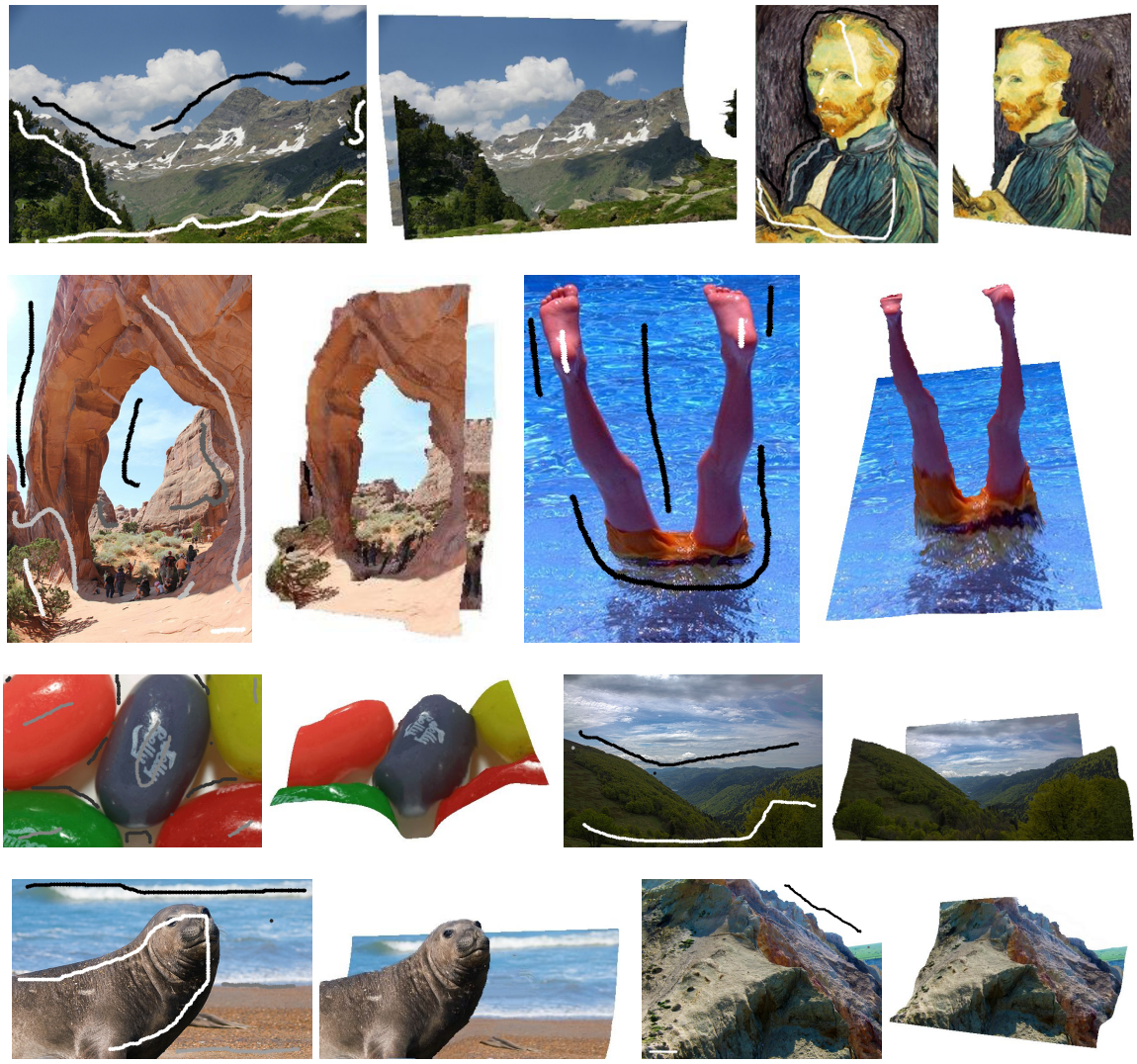


図 4.11: 3 次元シーンモデルの生成．それぞれのペアにおいて左の画像がユーザ入力，右が 3 次元モデルである．

できる．さらに，生成されるデプスマップをデプスの不連続箇所を考慮した滑らかな変化を実現している．

既存手法との計算時間の比較

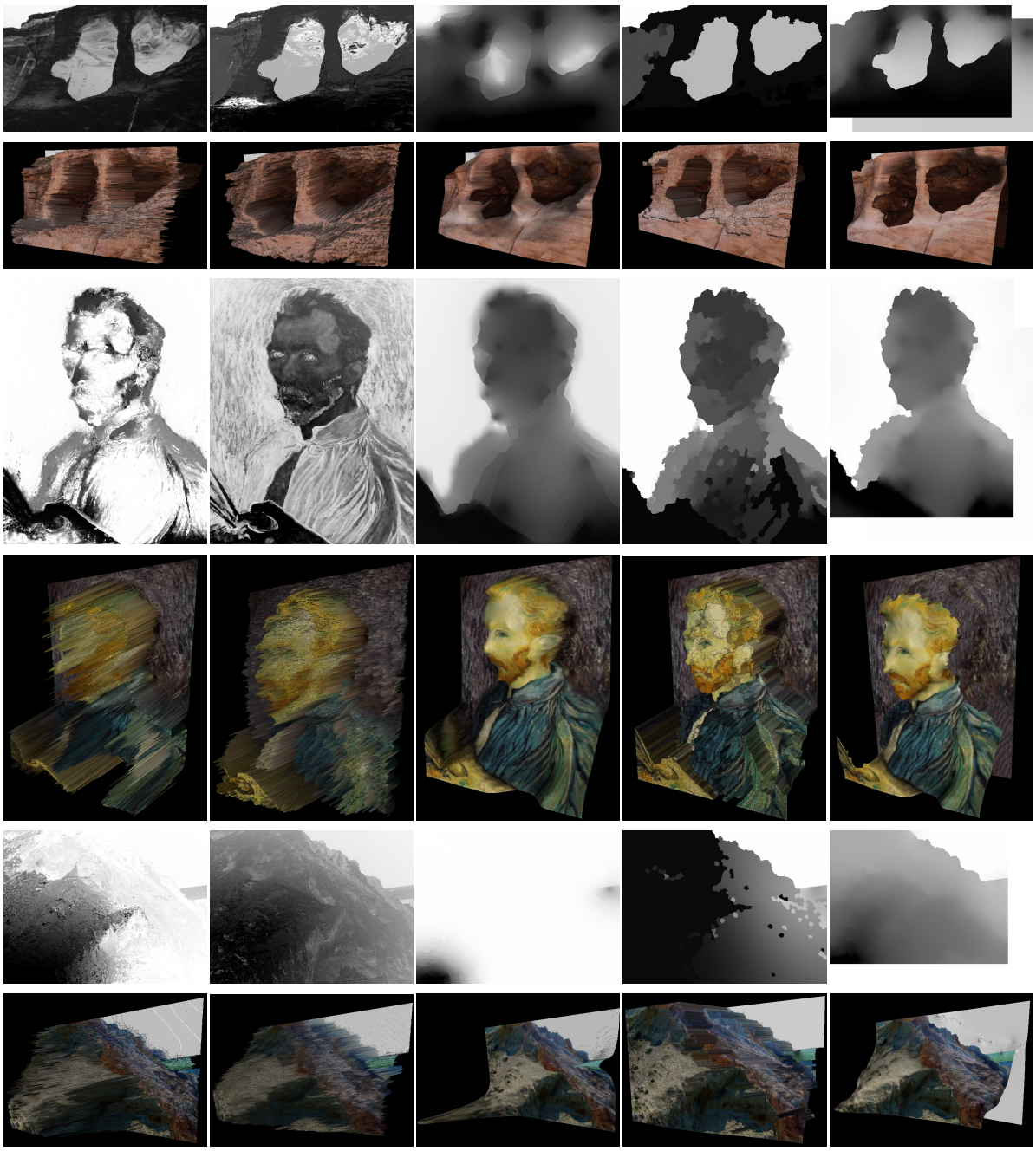
約 1 メガピクセルの画像に対して，提案手法は 0.2 秒程度でデプスマップを生成できる．これに対し，Chen らの手法と Li らの手法はそれぞれ 1 分と 2.5 秒程度の時間がかかった．エネルギー最適化ベース手法である Wang らの手法は Matlab を用いた実装で約 10 秒の時間がかかった．Chaurasia らの手法の計算時間は，デプス入力を含むスーパーピクセルの数に依存するが約 5 秒程度であった．Yücer らの手法は 250×250 の画像で $5 - 15$ 秒の計算時間が必要であることが文献 [83] に示されている．このように，既存手法に比べ提案手法は十分に対話的な編集が可能な速度でデプスマップを生成することができる．

ピクセルベースとスーパーピクセルベースの比較

図 4.14 は提案手法をピクセルベースで計算した場合とスーパーピクセルベースで計算した場合の比較である．すなわち，ピクセルベースで測地距離を計算してデプスマップを生成し最適化によって平滑化した場合と，最初にスーパーピクセルベースでデプスを計算した後ピクセル単位のデプスに変換した場合の比較である．ピクセルベースで計算するとデプスマップは荒くなり，デプスの入力位置から離れた領域にはうまく伝播しない．これに対し，提案手法のスーパーピクセルベースの手法は滑らかなデプスマップを生成できている．また，スーパーピクセルベースの手法はピクセルベースの手法に比べ約 10 倍の処理速度を実現している．

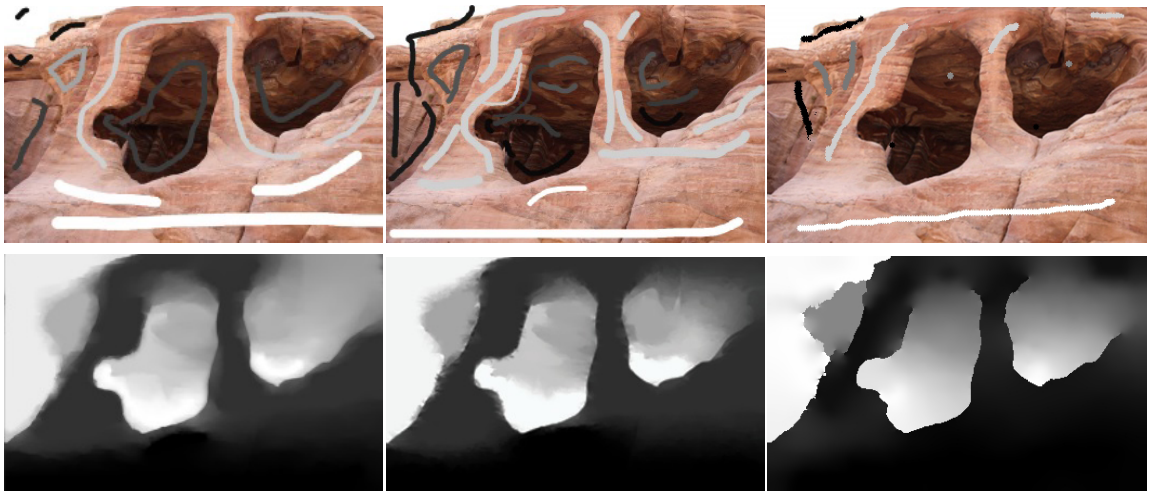
4.6.4 制約

提案手法にはいくつかの制約がある．まず，球体や円錐体のような特定の形状を復元することはできない．図 4.15 はその典型的な例である．ユーザ入力を増やすことで 3 次元モデルを目的の形状に近付けることはできるが，正確な球面などを再現することはできない．また，複雑なシーン構造をもつ入力画像に対しては必要なユーザ入力が増えてしまう．例えば，多数の人や車などのオブジェクトが存在するような画像では，遮蔽領域の抽出やテクスチャの推定に失敗する可能性が高い．この場合，正確な前景物体抽出や遮蔽領域のテクスチャ生成のためのさらに多くの入力が必要となる．



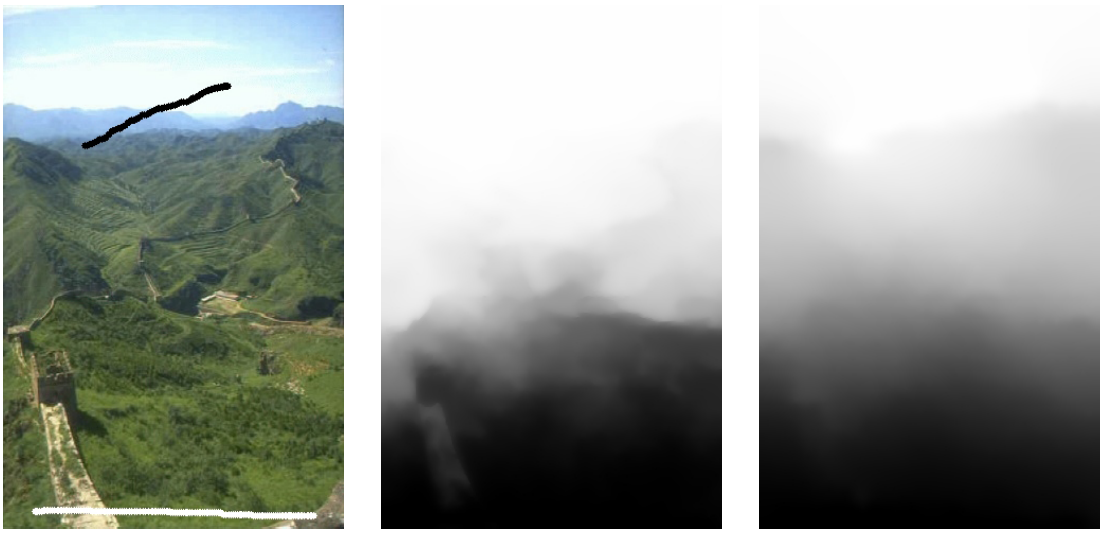
(a) [18] (b) [47] (c) [75] (d) [14] (e) Ours

図 4.12: 提案手法と既存手法のおけるデプスマップと 3 次元モデルの比較．ユーザ入力はそれぞれ図 4.2, 4.11 のものである．(a)-(d) 従来手法ではデプスがうまく伝播せず粗いデプスマップが生成されてしまい，3 次元モデルの生成がうまくいかない．(e) これに対し，提案手法では物体形状を考慮した滑らかデプスマップが生成されており，3 次元モデルも表面形状が滑らかになっている．



(a) [75] (b) [83] (c) 提案手法

図 4.13: 似たデプスマップ（下段）を生成するために必要なユーザ入力（上段）の比較．提案手法は既存手法に比べ少ないユーザ入力で目的のデプスマップを生成できる．さらに，提案手法によって生成されるデプスマップは不連続箇所でデプスのエッジを保持しつつ滑らかな変化を実現している．



(a) ユーザ入力 (b) ピクセルベース (c) 提案手法

図 4.14: 提案手法によるデプス伝播をピクセルベースで計算した結果との比較．(a) デプス入力に対し，(b) ピクセルベースで計算するとデプスマップは荒くなり，デプスの入力位置から離れた領域にはうまく伝播しない．これに対し，(c) スーパーピクセルベースで計算したあとにピクセルベースでデプスを計算する提案手法は，滑らかなデプスマップを生成できている．

(a) ユーザ入力

(b) 3 次元モデル

(c) ユーザ入力

(d) 3 次元モデル

図 4.15: 提案手法の制約．(a) ユーザ入力に対して，(b) 球体形状が生成できない．(c) ユーザ入力を増やすと，(d) 3 次元形状を半球体に近付けることはできるが正確な形状は復元できない．

第5章　3次元情報を利用した画像コンテンツ制作

本章では，これまで述べてきた単視点画像の 3 次元シーン推定をコンテンツ制作に応用する例として，画像の構図編集を行うシステムを中心に解説する．画像の構図編集はコンテンツ制作の現場で多く利用される技術であり，コンピュータグラフィクスの分野でも多くの研究が存在する．本手法では特に画像のオブジェクトの配置編集に着目し，一般的なユーザがこの高度な画像編集をシーンの遠近を考慮して直感的に行えるようにすることを目的とする．まずはじめに画像の構図編集に関連する既存研究について紹介する．続いて，本手法の概要ついて述べ，それぞれの工程について詳しく説明する．その後，本手法による物体の配置編集結果を示し，その有用性について検証する．最後に，3 次元映像や空気遠近の生成など，奥行き情報を利用したその他のアプリケーションについて述べ，提案手法が様々なコンテンツ制作に応用できることを示す．

5.1 オブジェクトの配置編集に関する研究

デジタル画像の普及により，画像をユーザの思い通りに編集・加工する技術はますます重要になっており，幅広く研究が行われている．そのような画像編集技術の一つに画像中のオブジェクトの配置編集がある．これは画像中の任意のオブジェクトを好きな位置に移動させて自然に合成することで，ユーザが画像の構図を好きなように編集する技術であり，コンテンツ制作の現場で重要技術である．以下ではこのオブジェクト再配置に関する研究を述べる．

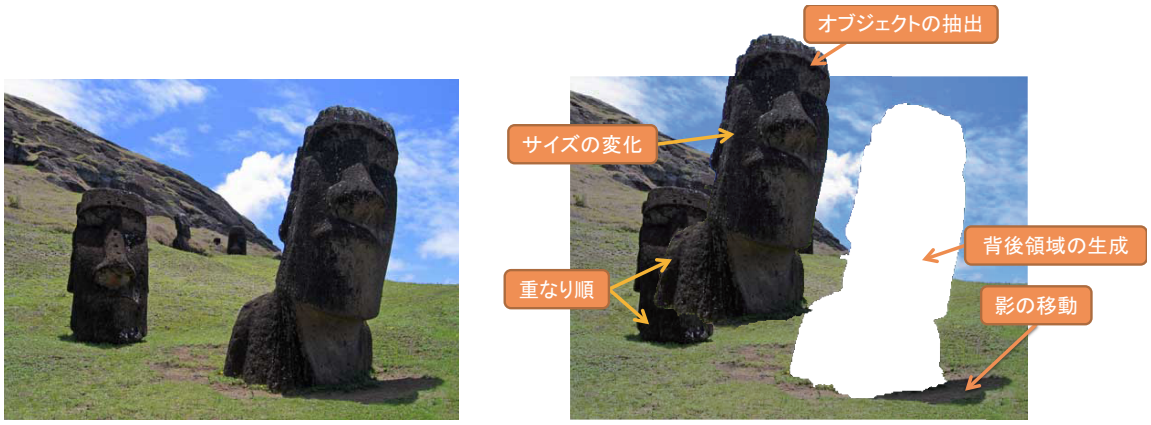
5.1.1 オブジェクトの配置編集

Cho ら [21] は入力画像を細かいパッチに分解し，任意の領域を移動先に自然に合うようにパッチを組み換える手法を提案している．これは，分解されたパッチをマルコフ確率場とみなし，パッチ組み換えの最適化問題を確率伝搬法を用いて解くことで実現している．また，Simakov ら [71] や Barnes ら [10] は，大域的なエラー関数を定義し，これを最小化するようにパッチを合成していくことで自然な再配置を行える手法を提案している．これらパッチベースの手法は単に指定領域を移動させるだけでは歪みが発生する場合が多く，ユーザが直線箇所などを明示的に指定する必要がある．

これらパッチベースの手法に対し，Cheng ら [19] は形状が酷似したオブジェクトが多く写る画像を対象に，簡単なユーザ入力で似た形状をもつオブジェクトをすべて抽出し編集できるシステムを提案している．この手法では，ユーザが背景領域と 1 つのオブジェクトを粗くブラシストロークで指定すると，そのオブジェクトと境界形状が類似したオブジェクトをすべて抽出し，変形や再配置などを行える．この手法はあらかじめ抽出したオブジェクトを移動させているため，パッチベースの手法のようにパッチを組み替えるときに起こる歪みが発生しない．しかし，この手法は類似したオブジェクトが多く映るような画像でないと適用できない．また，シーンの遠近に伴うオブジェクトの変化は考慮されておらす，オブジェクトの影の移動も扱えない．

5.1.2 オブジェクトの挿入

オブジェクトの配置編集の重要な関連研究として，オブジェクト挿入がある．この技術は他の画像からオブジェクトを切り抜いて対象画像に挿入し，自然に合成するものである [62, 38]．これは 2 次元画像だけでなく 3 次元映像にオブジェクトを挿入する研究も行われている [52]．特にシーンの遠近を考慮した画像編集ツールとして，Lalonde らが提案した Photo Clip Art[45] がある．このシステムでは，ユーザはあらかじめ用意されたデータベースから人や車などのオブジェクトを選択し，入力画像に挿入することができる．この際，オブジェクトの元の画像と挿入先の画像においてユーザが指定した消失線の位置関係から，シーンに合うようにオブジェクトサイズを調整している．しかし，このシステムでは入力画像にはじめから存在するオブジェクトを直接編集することはできない．また，それぞれのオブジェクトの位置関係



(a) 入力画像　　(b) オブジェクトの移動

図 5.1: オブジェクトの配置編集の課題.

を考慮することができないため，入力画像のオブジェクトの後ろに新たにオブジェクトを配置するような操作はできない.

5.2　提案システムによるオブジェクトの配置編集

本節ではまずオブジェクトの配置編集を行う際の課題について述べる．その後，提案システムの目標を述べ，その概要と要素技術について説明する.

5.2.1　1 枚の画像オブジェクト配置編集の課題

1 枚の画像のオブジェクトの配置編集は一般的に高度な画像編集として知られている．これは図 5.1 に示すような以下の課題が挙げられるためである.

- 効率的なオブジェクトの抽出
- オブジェクトの背後領域の生成
- 影の移動
- シーンの遠近によるオブジェクトのサイズの推定
- オブジェクト同士の重なり順の変化の考慮

まず移動したいオブジェクト領域を画像から正確に分離する必要がある．しかし，正確なオブジェクト抽出は手間のかかる作業である．また，抽出したオブジェクト領域をそのまま

移動させると背後領域が「穴」になってしまう．これを避けるため，何らかの方法でこの背後領域を補完する必要がある．オブジェクトに影がある場合，影をオブジェクトと同様にそのまま移動すると，元の地面と移動先の地面の色が異なる場合に不自然な合成結果が得られてしまう．また，画像のシーンには奥行きが存在し，オブジェクトを移動させたときシーンの遠近に合わせてオブジェクトのサイズや重なり順を変化させなければ自然な合成結果は得られない．このためにはあらかじめシーンの奥行きを推定する必要がある．

5.2.2 システム概要

提案システムは，画像中のオブジェクトをユーザが指定した位置に遠近を考慮して簡単に移動できるようにすることを目標とする．

本システムの流れを図 5.2 に示す．提案システムでは，入力は 1 枚の景観画像とし，再配置するオブジェクトは地面に対して垂直に配置されているものとする．この仮定にもとづき，ユーザは始めに地面領域を入力画像から折れ線で分割し，オブジェクトの移動はこの地面領域上で行う．

ユーザインタフェース． 本システムではユーザは 2 つまたは 3 つの単純な操作を行う．すなわち，(1) 奥行き推定のため地面領域の境界を折れ線で指定，(2) バウンディングボックスを対象オブジェクトを囲むように配置，(3) オブジェクトに影がある場合は影をスクリブル（ブラシストローク）で指定という 3 つの操作である．この操作の後，ユーザはオブジェクトをドラッグして好きな位置に移動することができる．移動の際，オブジェクトのサイズや重なり順は奥行きに合わせて自動で調整される．これにより，ユーザは直感的に物体の配置編集を行うことができる．

本システムでは前処理として入力画像はスーパーピクセルに分割される．その後，提案システムはユーザが入力した情報をもとに下記の 3 つの処理を行う．

1. 入力画像をレイヤ構造へ変換
2. オブジェクトの影マットを抽出
3. シーンの奥行きを推定

まず，入力画像は複数のオブジェクト画像と背景画像で構成されるレイヤ構造に変換される．オブジェクト画像は人間の目を引きやすい領域を計算した顕著性マップにもとづき抽出される．その後，抽出されたオブジェクトの背後領域は画像パッチベースの補完手法によって自動で生成される．もしオブジェクトに影がある場合，少しのスクリブルで指定するだけで影マットをシステムが抽出し，オブジェクトの移動先の地面に合成する．これにより，自然な影の移動を実現できる．最後にシステムがユーザが指定した境界線から奥行きを推定することで遠近に合わせたオブジェクトサイズや重なり順を自動で決定することが可能になる．次節からはこれらの工程の詳細について述べる．

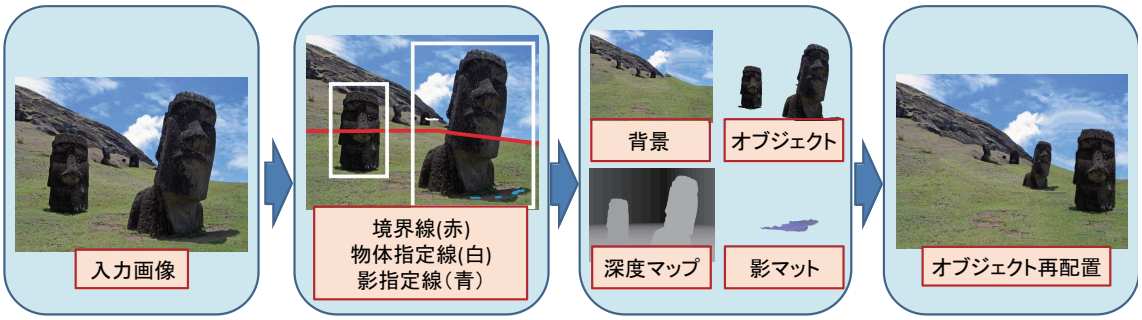


図 5.2: システム概要．1 枚の景観画像の境界線（赤）とオブジェクト（白）と影（青）をユーザが指定すると，影マットを含むオブジェクトと背景からなるレイヤ構造が構築され，デプスマップに基づき遠近に合わせたオブジェクトの再配置が可能となる．

5.3 レイヤ構造の生成

本節では，1 枚の画像からレイヤ構造を生成する手法について述べる．ここで述べるレイヤ構造は，複数のオブジェクト画像とオブジェクトの「穴」を埋めた背景画像を重ねたものである．

5.3.1 オブジェクト画像の生成

対話的なオブジェクト抽出の流れ

オブジェクト画像を生成するため，オブジェクト領域を入力画像から分離する必要があるが，これをユーザが手動で正確に行うのは手間がかかる．このため，3 章では領域分割 [22] と GrabCut[66] による効率的なオブジェクト抽出の手法について述べた．ここでは，さらに精度よく少ないユーザ入力でオブジェクトを抽出するため，ユーザ入力をバウンディングボックスとし，顕著性マップを利用してオブジェクト抽出を行う手法を提案する．ここで顕著性マップとは人の注目を引きやすい顕著領域を計算したグレイスケール画像である．本システムでは Mean Shift による領域分割 [22] とユーザが入力したバウンディングボックスから顕著性マップを計算し，この情報を GrabCut のフレームワークに取り入れることで，単純かつ少ないユーザ入力でオブジェクトを抽出できるようにする．オブジェクト抽出における本手法の主な貢献は,「オブジェクトらしさ」の手がかりとして用いることができる顕著性をバウンディングボックスベースで計算する手法の提案である．オブジェクト抽出のため，本手法は下記エネルギー関数を最適化する．

$$E(I) = \sum_{p} R(I_p) + \lambda \sum_{(p,q)\in C} [I_p \neq I_q]\, B(I_p, I_q) \tag{5.1}$$

ここで I は入力画像，I_p はピクセル p におけるピクセル値，C は隣接ピクセルのペア p,q の

集合，R はデータ項，B は平滑化項，λ は平滑化項の影響を決める定数，および $[\cdot]$ は指標関数である．本手法では平滑化項は Boykov らの手法 [12] のように下記式で表される．

$$B(I_p, I_q) = \exp(-\frac{\|I_p - I_q\|^2}{2\sigma^2}) \cdot dist(p,q)^{-1} \tag{5.2}$$

ここで，σ は定数，$dist(p,q)$ はピクセル p と q の空間的な距離である．データ項 R を定義するため，提案手法では Gaussian Mixture Models(GMM) と顕著値を用いる．次節ではこの顕著値を計算する手法について詳しく述べる．

顕著性マップの生成

一般的に画像中の顕著性が高い領域はオブジェクト領域である可能性が高いため，顕著性マップはオブジェクト認識などに広く利用される．既存の顕著性マップ作成手法 [5, 20, 37, 84] の多くは，人間の目の神経細胞の受容野が画像のコントラストに強く反応する理論 [64] にもとづき，画像全体で大域的に顕著性を計算している．しかし，景観画像においてこれらの手法による顕著性マップは，オブジェクト以外の領域の顕著性が高く算出されてしまう場合が多い．そこで本システムでは，粗くオブジェクトを囲む矩形のガイド線を利用した顕著性マップ生成手法を提案する．本論文で提案する顕著性マップは以下の仮定に基づいている．

- ガイド線と交差するスーパーピクセルは背景領域に分類できるため，顕著値は小さくなる
- オブジェクト領域は背景領域との色空間における距離がコントラストが高い
- 空間的に近いスーパーピクセル同士のコントラストほど重要である
- 景観画像において背景領域は連続している可能性が高いため，ガイド線上のスーパーピクセルはピクセル数が多いほど背景らしい
- ガイド線に近い領域ほど背景である可能性が高い

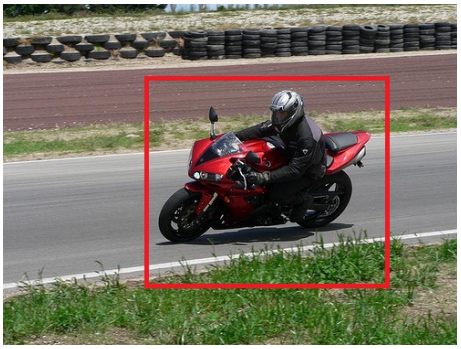
この考察に基づき，本システムでは画像全体で大域的にコントラストを計算するのではなく，オブジェクトを囲むガイド線と交差するスーパーピクセル（図 5.4(a) 黒領域）とその内部のスーパーピクセル（図 5.4(a) 灰領域）のコントラストによって顕著性マップを計算する．この際，それ以外のスーパーピクセル（図 5.4(a) 白領域）は計算に使わないようにする．

ガイド線の内部のスーパーピクセルを r_i，ガイド線上のスーパーピクセルを r_j とすると，スーパーピクセル r_i の顕著値 $S(r_i)$ は下記の式で表される．

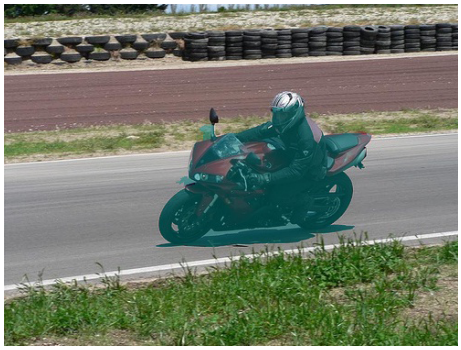
$$S(r_i) = \sum_{r_i \neq r_j} e^{-\frac{d_s(r_i,r_j)}{\sigma_1^2}} (1 - e^{-\frac{d_s(r_i,b)}{\sigma_2^2}}) f(r_j) d_c(r_i, r_j) \tag{5.3}$$

(a) 入力画像

(b) スーパーピクセル



(c) ユーザ入力

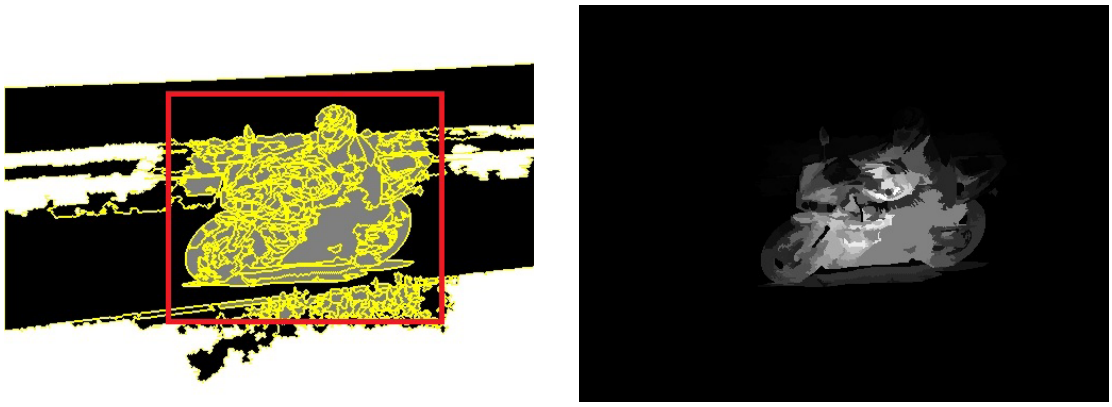


(d) 物体抽出

(e) 追加入力

(f) 抽出結果

図 5.3: オブジェクト抽出の流れ．(a)(b) 前処理として入力画像はスーパーピクセルのに分割される．(c) ユーザがオブジェクトを粗く囲むと，(d) これをもとにオブジェクト領域 (緑) が分離される．(e) うまく分離できていない箇所はユーザが指定（赤線）することで，(f) 正確にオブジェクトを抽出できる．



(a) 使用するスーパーピクセル　　(b) 顕著性マップ

図 5.4: 顕著性マップの計算. (a) 顕著性はバウンディングボックスと交差するスーパーピクセル（黒）とその中のスーパーピクセル（灰色）から (b) 計算される．

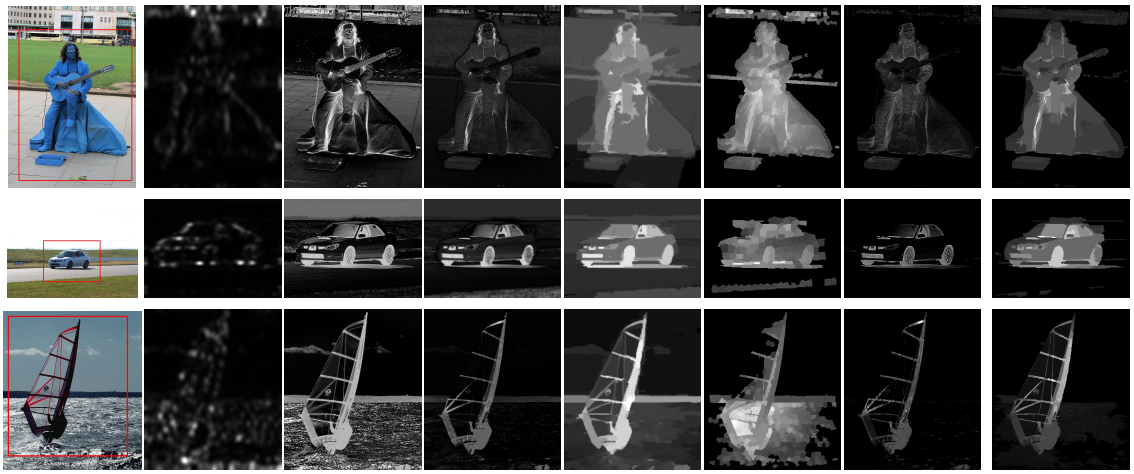
ここで $d_s(r_i, r_j)$ は r_i と r_j の重心距離であり，$d_s(r_i, b)$ はガイド線とスーパーピクセル r_i との最短距離，σ_1, σ_2 はこの重みを調整する定数であり，本システムでは $\sigma_1 = 0.4, \sigma_2 = 0.7$ としている，この項により，距離の近いスーパーピクセル同士のコントラストが強調され，ガイド線に近いスーパーピクセルの顕著値は小さくなる．また，$f(r_j)$ はガイド線上のスーパーピクセルの面積の総和に対する r_j の面積の割合であり，ガイド線上のサイズの大きなスーパーピクセルとのコントラストが強調される．$d_c(ri, rj)$ は r_i と r_j における平均色の Lab 空間の距離である．

算出した顕著値 $S(p)$ を用いて式 5.1 のデータ項を次のように定義する．

$$R_p(`obj') = -\log\left(Pr(I_p|obj)S(I_p)\right) \tag{5.4}$$

$$R_p(`back') = -\log\left(Pr(I_p|back)(1 - S(I_p))\right) \tag{5.5}$$

ここで，$Pr(I_p|\cdot)$ はカラー GMM を用いた尤度 [66] である．グラフカットによってこのエネルギー関数を最小化することでオブジェクト領域が抽出される．もしオブジェクト領域が正確に抽出できていない場合，その部分をスクリブルで指定することで簡単に修正することができる．抽出したオブジェクト領域の境界にポワソンマッティングをよばれる手法 [73] を適用する．これにより境界部分のアルファ値が調整され，オブジェクトを移動させたときに背景とうまくなじむようになる．



(a) Input (b) [37] (c) [84] (d) [5] (e) [20] (f) [76] (g) [61] (h) 提案手法

図 5.5: 提案手法と既存手法による顕著性マップの比較. (a) ユーザが指定したバウンディングボックスは赤で示されている. 既存手法に比べ，提案手法はオブジェクト領域の顕著性が高く背景領域の顕著性は低くなっている.

顕著性マップとオブジェクト抽出の考察

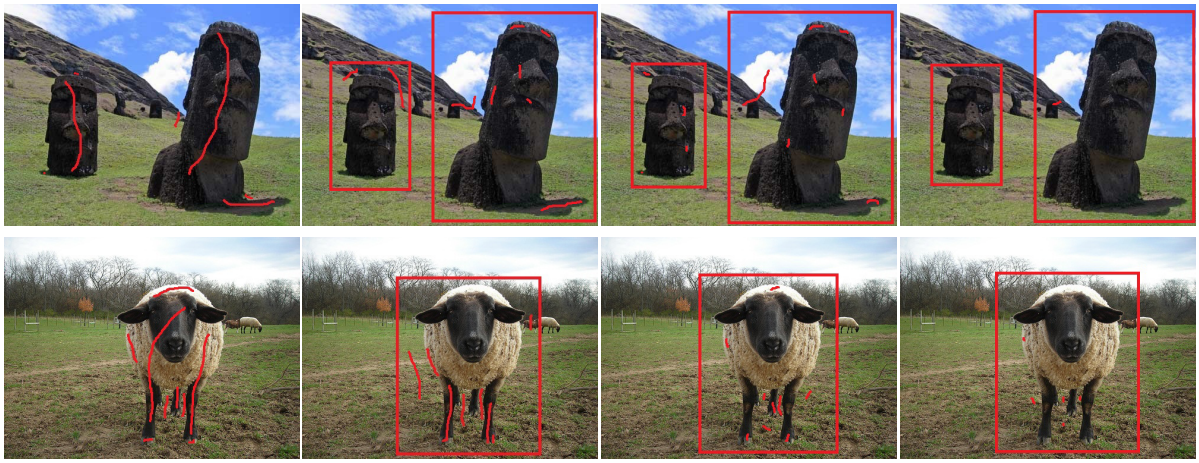
図 5.5 は提案手法と既存手法 [37, 84, 5, 20, 76, 61] の顕著性マップの比較である．より公平に比較するため，既存手法の顕著性マップをユーザが指定した矩形領域内で計算することでそれぞれの手法の精度を向上させている．既存手法ではオブジェクト以外の領域も顕著性が高くなってしまっているのに対し，提案手法ではオブジェクト領域の顕著性が高くなっていることがわかる．これにより，提案手法は既存手法よりも正確にオブジェクト領域を推定することができる．

図 5.6 は商用の画像編集ツールである Adobe Photoshop CS5 の Quick selection ツールと Rother らの GrabCut[66]，3 章で提案した領域分割と GrabCut を合わせた手法と提案手法との比較である．図 5.6 において，提案手法は Photoshop や GrabCut に比べ，単純で少ないユーザ入力しか必要としない．また 3 章で示した手法と比べても，オブジェクト領域を分離するためのストローク数が半分に減り，より少ないユーザ入力でオブジェクトを抽出できていることがわかる．

ただし，入力画像によってはオブジェクト以外の領域の顕著性が高くなり，誤ってオブジェクト領域に分類される場合がある．このような場合でも，スーパーピクセルを利用した追加ユーザ入力によって簡単にオブジェクト領域を修正することができる．

5.3.2 背景画像の生成

1 枚の画像のオブジェクトを再配置するためには，移動するオブジェクトの背後領域を推定し合成する必要がある．本システムでは，3.2.2 節で示した手法と同様のアプローチで背景画



(a) Photoshop (b) GrabCut (c) 領域分割と GrabCut (d) 提案手法

図 5.6: オブジェクト抽出の比較. ユーザ入力は赤で示されている. Photoshop Quick Selection や GrabCut [66], および 3 章で示した領域ベースのオブジェクト抽出に比べ, 提案手法は粗く少ないユーザ入力でオブジェクトを抽出できている.

像を生成する. すなわちオブジェクト領域を「穴」領域とし, 残りの領域から類似画像パッチを探索して「穴」領域に合成していく. これにより背景画像が生成され, これにオブジェクト画像を重ねて動かすことで自然なオブジェクトの移動が可能となる.

5.4 影の抽出と合成

オブジェクトに影がある場合, 影をそのまま移動すると移動先の地面の色相と合わない (図 5.7(a)). そこで本システムでは Wu らの手法 [78] を改良し, 影マスクを指定するだけで影のマットを生成できるようにする. この影マットを移動先の地面に合成することで自然な影を再現できる (図 5.7(b)).

5.4.1 Natural Shadow Matting

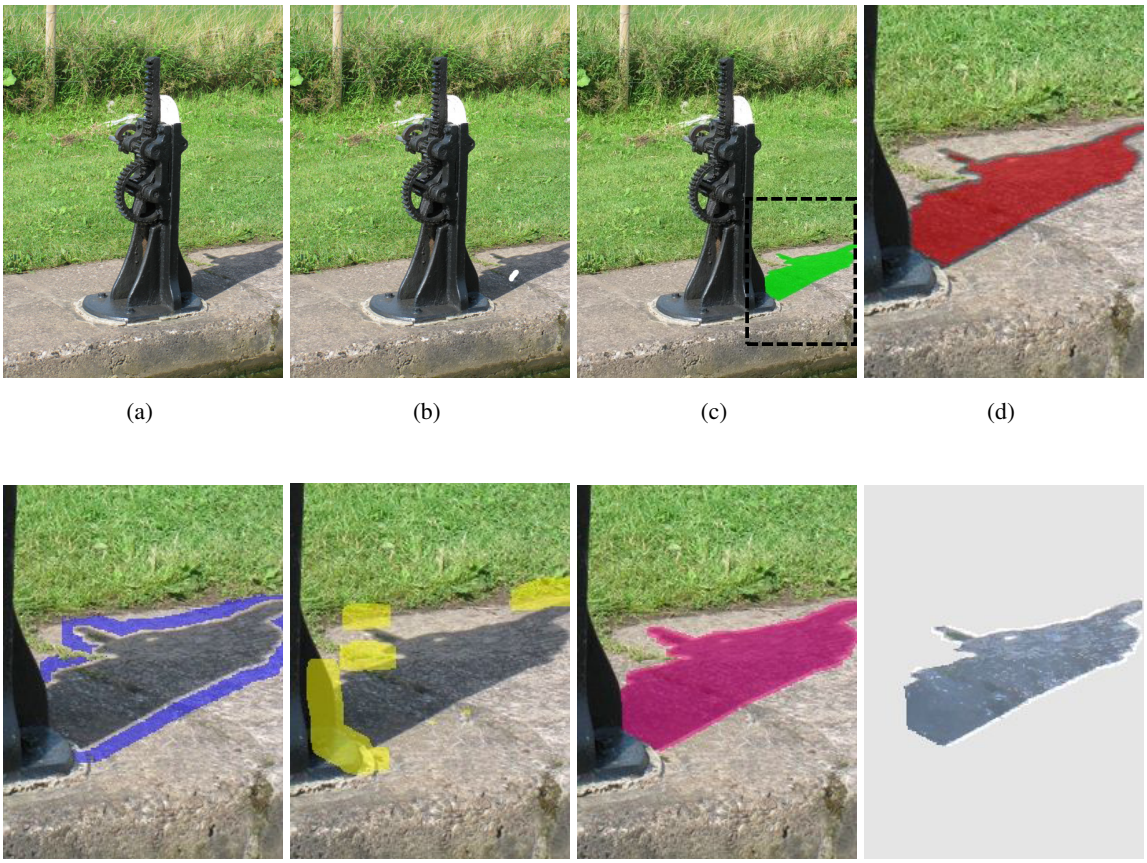
Wu らが提案した Natural Shadow Matting[78] では, 目に入る光は反射率と照明光の積で決まるという Retinex 理論に基づき, シーン画像 I を影マット β と非影画像 F の積 $I = \beta F$ として定義し, これを解くことで影を除去した画像と影マットを生成している. この手法では, まずユーザが「影領域」「非影領域」「未知領域」「除外領域」から構成される quad map を生成する. この各領域の色分布から粗い非影画像 $\hat{F}$ を Color Transfer[63] によって計算し, これを用いてエネルギー関数を定義し, 最小化することで影マットと非影画像を計算している. この手法は多くの画像において, 良好な結果が得られることが示されている.

(a) 影をそのまま移動　　(b) 提案手法

図 5.7: 影の抽出と合成．(a) 影をそのまま移動すると，移動先の地面の色と影の色が合わない．これに対し，(b) 本システムではあらかじめ影マットを抽出しておき，移動先の地面に合成することで，自然な影の移動を実現している．

5.4.2　Gaussian mixture model を利用した quad map の生成

Wu らの手法によって良好な影マットが抽出できるが，オブジェクトごとに「影領域」「非影領域」「未知領域」「除外領域」を指定するのは手間がかかる．そこで本システムでは影マスクを指定するだけでマットを地面から抽出できるようにする．まず影マスクとして指定した領域を縮小させ「影領域」とする．次に，指定した領域を 2 回に分けてモルフォロジー演算により膨張させ，1 回目と 2 回目の差分領域に対し Gaussian mixture model を用いてクラスタリングを行い，2 つのクラスタを生成する．本手法では 1 回目と 2 回目の膨張幅はそれぞれ 5 ピクセル，17 ピクセルとしている．このクラスタの内，大きなクラスタに含まれる領域を「非影領域」として分類する．また，小さいクラスタに含まれる領域を膨張させて「除外領域」とする．最後にユーザが指定した領域を膨張させ「未知領域」とすることで quad map が生成され，影マットの抽出が可能となる．本システムでは，影領域の指定をスーパーピクセルを利用して行えるようにすることで，少ないユーザ入力で指定できるようにしている（図 5.8 白）．



(a) (b) (c) (d)

(e) (f) (g) (h)

図 5.8: quadmap の生成．入力画像 (a) に対して，(b) ユーザは影マスクをスクリブル（白）で指定すると，(c) 影マスク（緑）がスーパーピクセルにもとづき抽出される．(d)-(h) は (c) の画像中の黒の四角形領域を拡大したものである．影マスクを用いて 4 つの領域が自動で抽出される：(d)「影領域」(赤），(e)「非影領域」（青），(f)「除外領域」（黄），(g)「未知領域」（ピンク）である．この quad map を用いて, (h) 影マットが計算される．

5.5 シーンの奥行き推定

本システムでは，3 章で示した手法と同じようにユーザが折れ線によって分離した地面領域からシーンの奥行きを推定する．奥行き推定には 4 章のスパースなデプス入力にもとづく手法を用いてもよいが，ここではオブジェクトが地面に設置されていると仮定しているため，その大まかな 3 次元位置は 3 章の手法で容易に推定できる．そのため，本システムでは境界線を指定するというより単純なユーザ入力である 3 章の手法を採用した．本研究の目的はシーンの遠近に合わせて物体の配置編集を行うことであり，大まかな奥行き推定で十分に良好な結果を得ることが可能である．5.7 節で示すその他のアプリケーションについては，入力画像の種類に応じて 3 章と 4 章の手法を使い分けることで，効果的に画像の奥行きを推定するようにしている．

まず，オブジェクト画像の底辺の座標から計算した z 座標にもとづきオブジェクト同士の重なり順が決定される．また，オブジェクト i の画像上の高さ h_i' は Hoiem らの手法 [33] にもとづき以下のように計算される．まずオブジェクト i のワールド座標での高さ y_i は $y_i = \frac{y_c h_i}{v_0 - v_i}$ で表される．ここで h_i は画像での元の高さであり，y_c はカメラの高さ，v_0 は消失線の y 座標，v_i はオブジェクトの底辺の y 座標である．また，v_i' を再配置されたオブジェクトの底辺の y 座標とすると，$y_i = \frac{y_c h_i'}{v_0 - v_i'}$ と表すことができる．上記 2 式から y_c を消すことで次式を得ることができる．

$$h_i' = \frac{v_0 - v_i'}{v_0 - v_i} h_i \tag{5.6}$$

この式によって計算される座標は厳密なものではないが，シーンの遠近に合わせたオブジェクト再配置を十分に良好に行うことができる．

5.6 結果

本システムは C++言語で実装し，Intel Core i7 (2.67GHz, 4.00GB RAM) が搭載された PC 上で実行した．使用した画像のサイズは全て 0.3 から 1.0 メガピクセルの範囲内である．結果に示した画像において，顕著性マップの計算には約 0.5 秒，影マット抽出にかかった時間は約 0.5 秒，デプスマップを含むレイヤ構造に変換するのにかかった時間はユーザ入力も含めて 2 分以内であった．

提案システムにより，入力画像は影マットを含むオブジェクト画像と背景画像から成るレイヤ構造に変換される．移動に伴うオブジェクトサイズや位置関係の変化はデプスマップにもとづき自動で調整されるため，ユーザは移動させたいオブジェクトをドラッグして動かすだけで直感的にオブジェクトの配置編集を行うことができる．図 5.2,5.9,5.10 は提案システムによるオブジェクトの再配置結果である．オブジェクトが遠近に合った自然なサイズになっていることがわかる．また，図 5.9 中段では，片方の牛がもう片方の牛に隠れるように配置されている．このようなオブジェクト同士の位置関係の変化も本システムでは自動で再現できる．

図 5.9: 提案システムによるオブジェクトの配置編集．左の列がユーザ入力を示している．

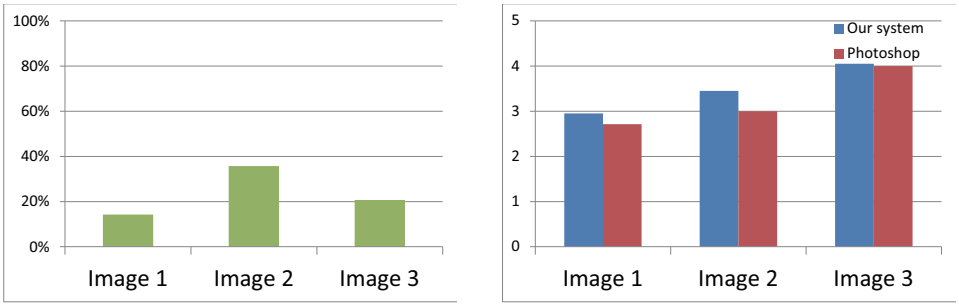
提案システムの目的は，入力画像のシーンに合った自然なオブジェクト移動を可能にすることである．提案システムでは，カメラは地面に対して並行であると仮定しているため，奥行き情報の推定は正確ではない．しかし，オブジェクトの再配置に伴うサイズや重なり順の変化を考慮するという目的においては，多くの景観画像で十分良好な結果を得ることができる．

5.6.1 ユーザテスト

提案手法の有効性を確かめるため，ユーザテストを実施した．ユーザテストの内容は，画像編集の経験がほとんどない 5 人のユーザに提案システムを用いてオブジェクトの配置編集を行ってもらい，一方で画像編集に長けたユーザに商用の画像編集ツール Adobe Photoshop CS6 を用いてオブジェクトの配置編集を行ってもらい，結果を比較した．なお，テストに用いた画像は図 5.2, 5.9 上段と中段の画像である．具体的な編集内容は，画像中の手前にあるオブジェクトを奥に配置し，逆に奥にあるオブジェクトは手前に配置するというものであり，その際オブジェクトのサイズや重なり順などが遠近に合わせて自然になるように調整してもらうようにした．それらの編集時間を記録し，編集結果の自然さは別の 4 人の協力者に 5 段階で評

図 5.10: 提案システムによるオブジェクトの配置編集の例.

価してもらい，結果を集計した．その結果，編集時間は提案システムを用いたユーザの作業時間は約 1-2 分であり，商用の画像編集ツールを用いたユーザよりも 3 倍から 6 倍の速さで目的の編集結果を得ることができた．また，編集結果の自然さは同程度の評価であった．この結果から，画像編集の知識を持たないユーザでも提案システムを用いることで，画像編集に長けたユーザの構図編集に匹敵する結果を短時間で得られることを確認した．



(a) 編集時間　　(b) 結果の自然さ

図 5.11: 提案システムと商用ツールを用いたオブジェクト配置編集の時間と結果の自然さの比較．(a) 提案システムを用いたときの編集時間は Photoshop(100%) を用いたときよりも 3-6 倍短かったことがわかる．(b) また，提案システムによる編集結果は Photoshop の編集結果と同等の自然さを実現している．

5.6.2　制約

現在の提案システムにはいくつかの制約が存在する．まず図 5.12 のように前景物同士が重なり合った複雑なシーンでは，遮蔽領域の推定は困難であるため背景画像の生成に失敗する場合が多い．このような問題には単視点画像の情報だけでは対処が困難であるため，データ駆動型のアプローチ [30] が有効であると考えられる．また，提案手法の奥行き推定は屋内画像のような狭いシーンではうまくいかない場合が多い．これはユーザが指定した境界線の最も上に位置する頂点は，カメラから十分に遠い位置にあると仮定して奥行き推定を行っているためである．また，提案システムはオブジェクトの配置編集に伴う照明環境の変化は考慮できない．例えば，日なたに置かれているオブジェクトを日陰に移動させた場合，オブジェクトの陰影の変化は考慮できない．このようなリライティングを行うためには，オブジェクト形状やシーンの照明環境を推定する必要がある．

(a) 入力画像　　(b) 物体の配置編集

図 5.12: 提案システムの失敗例．(a) 左の馬と騎手を移動すると， (b) 背後領域の補完に失敗してしまう．

5.7　奥行き情報を利用するその他の画像コンテンツ制作

本節では，これまで述べてきた画像の構図編集以外の奥行き情報を利用するコンテンツ制作について述べる．

5.7.1　3 次元映像の生成

近年の立体視ディスプレイなど 3D 対応デバイスの進化によって，3 次元映像の需要はますます高まっている．このような 3 次元映像は特殊な撮影装置を用いて撮影することによって得られるが，そのような撮影装置を利用できる場面は限られる．これに対し，既存の 2 次元映像を 3 次元映像に変換することで 3 次元映像を制作することもできる．これはシーンの奥行きを何らかの方法で取得し，その奥行きに合わせて画像をワーピングして左目用と右目用

の画像を作成することで得られる．しかし，一般的に画像の奥行きを推定することは困難であるため，この 2 次元映像から 3 次元映像を作成するには非常に手間がかかる．これに対し，提案手法によってスパースなデプス入力のみから入力画像のデプスマップを生成できるため，このデプスマップを用いて Wang らの手法 [75] によりアナグリフ画像を自動生成することができる．図 5.13 は提案手法によって計算されたデプスマップを用いて生成したアナグリフ画像である．このように提案手法は 3 次元映像制作の大幅なコスト削減にも貢献できる．

(a) ユーザ入力　(b) デプスマップ　(c) アナグリフ

図 5.13: 提案手法によるアナグリフ画像の生成．

5.7.2 空気遠近の生成

提案手法は推定したデプスに合わせて霞を合成していくことで，空気遠近による奥行き感の強調を行うことできる．ここで空気遠近とは遠くにあるものほど色がぼやけていく現象を指し，遠近法の一つとして広く用いられる．提案システムでは下式のような一般的な霧モデル [26] を利用し，霧を合成することができる．

$$\mathbf{I}_{out} = t(\mathbf{x})\mathbf{I}_{in} + (1 - t(\mathbf{x}))\mathbf{A} \tag{5.7}$$

$$t(\mathbf{x}) = e^{-\frac{b}{z}} \tag{5.8}$$

ここで，$\mathbf{A}$ は霧の色，b は定数，z は座標 $\mathbf{x}$ におけるデプス値を表している．提案システムでは，$\mathbf{A} = (0.9, 0.9, 0.9)$ かつ $b = 0.07$ としている．図 5.14 は霞を合成することで空気遠近を生成した結果である．シーンの奥行きに合わせて視覚的に良好な結果が得られている．

図 5.14: 霞の合成による空気遠近の再現．左が入力画像，右が合成結果である．

5.7.3 Depth-of-field

Depth-of-field は，対象物体を強調するために背景領域を故意にぼかして撮影するのはカメラの撮影技法のひとつである．しかし，そのような写真を初心者が撮影するのは技術的に難しく，また高価なカメラが必要となる場合もある．これに対し提案システムでは，算出したデプスマップに合わせて背景領域に平滑化フィルタをかけることで，この技術を簡易的にではあるが再現することができる．

ぼかしの再現のため，提案システムではガウシアンフィルタを用いる．ガウシアンフィルタはガウス分布の関数を利用した平滑化フィルタであり，以下の式で表される．

$$f(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \tag{5.9}$$

ここで，x, y はピクセルの座標，σ は標準偏差である．ガウシアンフィルタにおいて，σ が大きいほど平滑化の度合いが強くなり，逆に σ が小さければ平滑化の度合いも弱くなる．この性質を利用し，提案システムではパラメータ σ を奥行きに合わせて変化させていくことで背景をぼかし，オブジェクトを強調させている．つまり，焦点を合わせる物体の 3 次元位置を基準として，そこから離れるほど σ を大きくすることで奥行きに合わせたぼかしを再現している．これは次の式で表される．

$$\sigma = \frac{\sigma_{max} - \sigma_{min}}{d_{max} - d_{min}}(d - d_{min}) + \sigma_{min} \tag{5.10}$$

ここで d_{min} と d_{max} はそれぞれデプスの最大値と最小値であり，σ_{min} と σ_{max} はぼけの強さ調整する定数である．

図 5.15 は画像中のオブジェクトに焦点を合わせた depth-of-field 効果を再現した結果である．オブジェクトに焦点が合った良好な結果が得られていることがわかる．

図 5.15: Depth-of-field によってオブジェクトに焦点を合わせた結果．下段の例では，提案手法によってオブジェクトの位置を移動し，移動したオブジェクトに焦点を合わせている．

第6章　結論と今後の展望

6.1　結論

本研究では，1 枚の画像からの複雑なユーザ操作を必要としない多様な 3 次元シーン生成を目的とし，地面の境界と前景物の指定による 3 次元シーンの生成手法と，スパースなデプス入力による滑らかな表面形状をもつ 3 次元シーンモデルの生成手法を提案した．これにより，従来は熟練者が多くの手間をかけて作成していた，イメージベースの 3 次元シーンモデルを容易に作成できるようにした．また，3 次元映像生成や depth-of-field の再現など推定した 3 次元情報を利用した複数のアプリケーションについて示した．その中でも画像中のオブジェクトの配置を遠近を考慮して編集できるシステムについて提案することで，手法の有用性および拡張性を確認した．さらに，既存手法と比較を行うことでそれぞれの手法の有効性を示した．

テーマ毎については，以下のような結論を得た．

6.1.1　境界線を用いた 3 次元シーンモデルの生成

単視点の景観画像を対象とし，地面と壁の境界線にもとづく 3 次元シーンモデリングおよび前景物の効率的な抽出，前景物の背後領域の自動合成を一連の流れとして効率よく行えるフレームワークを提案した．境界線にもとづく 3 次元シーンモデリングは，既存手法よりも幅広いシーンに対応できることを既存手法との比較により確認した．また，境界線は前景物の背後領域を補完する際の制約条件として利用することで，補完精度を向上させられることを示した．前景物が存在するような入力画像においても，提案手法によって前景物を効率よく抽出できることを既存手法との比較によって確認し，接地条件やビルボード変換によって立体感のある 3 次元シーンを構築できることを示した．提案手法によって生成される 3 次元シーンの例を複数示し，考察を行った．

6.1.2　スパースなデプス入力による 3 次元モデルの生成

3 章で提案した 3 次元モデルよりも幅広いシーンに対応するため，スパースなデプスのみを入力として前景レイヤと背景レイヤから構成される Layered Depth Image（LDI）を容易に生成できる手法を提案した．LDI の対話的な編集を可能とするため，デプス入力をスーパーピクセルベースの重み付き測地距離とエネルギー最適化によって画像全体に伝播させることで，

既存手法より良好なデプスマップを高速に計算できることを示した．さらに，奥行きが大きく変化する不連続箇所を自動抽出し，これをもとに遮蔽領域のテクスチャとデプスを含む背景レイヤを自動で生成する手法を提案した．これにより，スパースなデプス入力のみで十分に立体感のある LDI を容易に生成できることを確認した．デプス入力の伝播について既存手法と比較を行い，精度や処理速度の観点からその有効性について確認した．また，提案手法によって生成されるデプスマップは 3 次元映像生成にも利用できることを示し，制作コストの高い 3 次元映像の制作に提案手法が有効であることを示した．最後に，特定の形状への適用などを含めた今後の課題についてまとめた．

6.1.3 3 次元情報を利用した画像コンテンツ制作への応用

提案システムによる画像の 3 次元シーンの推定は幅広い応用が可能であることを示すため，画像の構図編集を行うシステムや 3 次元映像制作などを解説した．まず画像の構図編集に焦点を当て，シーンの遠近にもとづきオブジェクトの配置編集を行えるシステムを提案した．提案システムにより，境界線の指定やバウンディングボックスの指定など単純なユーザ入力のみで，遠近を考慮したオブジェクトの配置編集が行えることを確認した．また，バウンディングボックスを利用した新しい顕著性マップの計算と前景物抽出を提案することで，より少ないユーザ入力でオブジェクトを抽出できることを既存手法と比較することで確認した．また，上記 2 つの手法によって推定した奥行き情報を用いることで，良好な 3 次元映像や depth-of-field，霞の合成による空気遠近の生成など，提案システムが多様な画像コンテンツ制作への応用も可能であることを示した．

6.2 展望

前節でまとめたように，本研究によって単視点画像から少数の単純なユーザ入力のみで十分に立体感のある 3 次元シーンの生成を実現した．

境界線を用いた 3 次元シーンの生成について，今後はより多様なシーンに対応できるようにする必要があると思われる．例えば，ユーザが前景物に対して立体のプリミティブを当てはめたり，立体形状を表す稜線をひいたりすることで，立体的な前景物モデルが作成できると考えられる．近年では，立方体を当てはめる手法 [87] や円筒形状を当てはめる手法 [17] が提案されており，写実的な 3 次元オブジェクトの生成に成功している．しかし，より詳細な 3 次元モデルを生成しようとするとユーザ入力が増えてしまうため，どこまでユーザ入力を許容できるかは目標とするシーンにあわせて検証する必要がある．

スパースなデプス入力による 3 次元モデル生成について，まず複数のレイヤからなる Layered Depth Image の構築が考えられる．現在のモデルは前景レイヤと背景レイヤで構成されており，複数の前景物体が複雑に重なり合ったシーンなどはうまくモデル化できない．そのため，複数レイヤを効率よく生成する手法は今後の大きな課題である．また，本手法は 3 次元オブ

ジェクトの挿入やリライティングなど様々な応用が可能と考えられる．これらについても今後検証していく必要がある．

また，3 次元シーン構築に関わる研究としてデータベースを利用したアプローチも今後重要になると考えられる．近年，Microsoft Kinect のように深度センサーによってデプスを取得できるデバイスが安価で手に入るようになり，高精度のデプスデータセットがインターネットを通じて簡単に手に入るようになっている [4]．すでにこのようなデプスデータセットを用いて単視点画像の粗いデプスを推定する手法 [42] も提案されており，このようなデータセットの利用法などは今後興味深い課題である．

以上で本研究で扱った単視点画像からの 3 次元シーン生成ついての展望を述べたが，これらの技術をどのように社会に役立てることができるかを検討することも重要と考えられる．単視点画像の 3 次元モデリングは非常に興味深い分野ではあるが，商用の技術として実用化するには精度などで課題が多い．3D デバイスや通信技術の進化によって，今後はさらに 3 次元映像コンテンツ制作技術は重要になる考えられ，用途に応じてデータセットの利用など柔軟な技術研究が望まれると考えられる．

発表論文

投稿論文他（査読付）

1. Satoshi Iizuka, Yuki Endo, Jun Mitani, Yoshihiro Kanamori and Yukio Fukui,“An Interactive Design System for Pop-Up Cards with a Physical Simulation”, The Visual Computer (Proc. of Computer Graphics International 2011), 27, 6, 605-612, 2011.

2. Satoshi Iizuka, Yoshihiro Kanamori, Jun Mitani and Yukio Fukui, “ Efficiently Modeling 3D Scenes from a Single Image”, IEEE Computer Graphics and Applications, 32, 6, 18-25, 2012.

3. Satoshi Iizuka, Yuki Endo, Hirose Masaki, Yoshihiro Kanamori, Jun Mitani and Yukio Fukui, ”Object Repositioning Based on the Perspective in a Single Image”, Computer Graphic Forum, Volume 33, Issue 8, pages 157-166, 2014.

4. Satoshi Iizuka, Yuki Endo, Yoshihiro Kanamori, Jun Mitani, Yukio Fukui: ”Efficient Depth Propagation for Constructing a Layered Depth Image from a Single Image”, Computer Graphics Forum (Proc. of Pacific Graphics 2014), Volume 33, Issue 7, pages 279-288, 2014.

5. 飯塚里志，金森由博，三谷純，福井幸男.「境界線とモーフィングを用いた景観画像からの 3D シーン生成」，NICOGRAPH SPRING, 2010-3.

6. 飯塚里志，遠藤結城，三谷純，金森由博，福井幸男.「物理シミュレーションを用いたポップアップカード設計支援システム」，画像電子学会・情報処理学会 Visual Computing/グラフィクスと CAD 合同シンポジウム, 2011-6.

7. 飯塚里志, 遠藤結城, 廣瀬真輝, 金森由博, 三谷純, 福井幸男. 「シーンの奥行きを考慮した景観画像における対話的なオブジェクト再配置」, Visual Computing/グラフィクスと CAD 合同シンポジウム 2012, 2012-6.

8. 飯塚里志, 遠藤結城, 金森由博, 三谷純, 福井幸男. 「スクリブルを用いた 1 枚の画像からの対話的なレイヤ状 3 次元モデルの生成」, Visual Computing/グラフィクスと CAD 合同シンポジウム 2013. 2013-6.

講演論文他

1. 飯塚里志，金森由博，三谷純，福井幸男.「簡易ユーザ入力による 1 枚の画像からの 3D シーン生成システム」，第 142 回グラフィクスと CAD 研究会，2011-2.

2. 飯塚里志，遠藤結城.「バネマスモデルを用いたポップアップカード設計支援システムの開発」，ソリューション型研究開発プロジェクト 2010 年度研究成果報告, 2011-2.

3. 飯塚里志，遠藤結城.「3 次元構造を考慮した対話的な景観画像編集システム」，ソリューション型研究開発プロジェクト 2012 年度研究成果報告, 2012-2

4. 鶴田直也, 飯塚里志, 胡健雄, 山田裕貴.「合同な正三角形で構成される立体の生成手法」, ソリューション型研究開発プロジェクト 2012 年度研究成果報告, 2013-2.

付録 A

PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing

この手法の核となる部分は，類似パッチを高速に算出するアルゴリズムである．これにより，類似パッチ探索を用いる画像補完や構図変換，画像再構成などの画像編集を対話的な速度で行うことができるようになる．ここでいう類似パッチ探索とは，各ピクセルをその類似度が最も高いパッチ座標 Nearest Neighbor (NN) に変換する関数として定義される Nearest Neighbor Field (NNF) を計算することである．これは図 6.1 のように 3 つの手順で算出される：

1. 初期化
2. 伝搬
3. ランダム探索

初期化を行った後，伝搬とランダム探索を繰り返し行うことで NNF が算出される．以下ではそれぞれの詳細について述べる．

初期化：まず，NNF には初期値としてピクセル座標がランダムで割り当てられる．ただし，1 度ランダム割り当てを行うだけだと，局所解に陥ってしまう場合がある．これを避けるため，ランダム割り当てを数回繰り返し，もっとも類似度が高い座標を初期値として割り当てるようにする．

伝搬：自然画像では，似た領域が続く場合が多い．つまり，あるピクセル座標 (x_1, y_1) の NN が $f(x_1, y_1)$ とすると，その隣接するピクセル $(x_1 + 1, y_2)$ の NN は $f(x_1 + 1, y_2)$ である場合が多い．この性質を利用して，既に計算された NN である $f(x - 1, y)$ と $f(x, y - 1)$ を用いて $f(x, y)$ を更新する．つまり，(x, y) における現在のオフセット $f(x, y)$ を $f(x - 1, y) + (1, 0)$ と $f(x, y - 1) + (0, 1)$ のパッチと類似度を比較して，より類似しているパッチ座標を (x, y) に割り当てる．これは画像の左上から順に水平走査で行われ，それが終わると次に右下から順に水平走査で行われる．

ランダム探索：$\mathbf{v}_0 = f(x, y)$ とすると，この $f(x, y)$ を更新するため，$\mathbf{v}_0$ を中心としてその周りから徐々に近づくようにランダムでパッチの類似度を調べ更新する．これは次の式で表される．

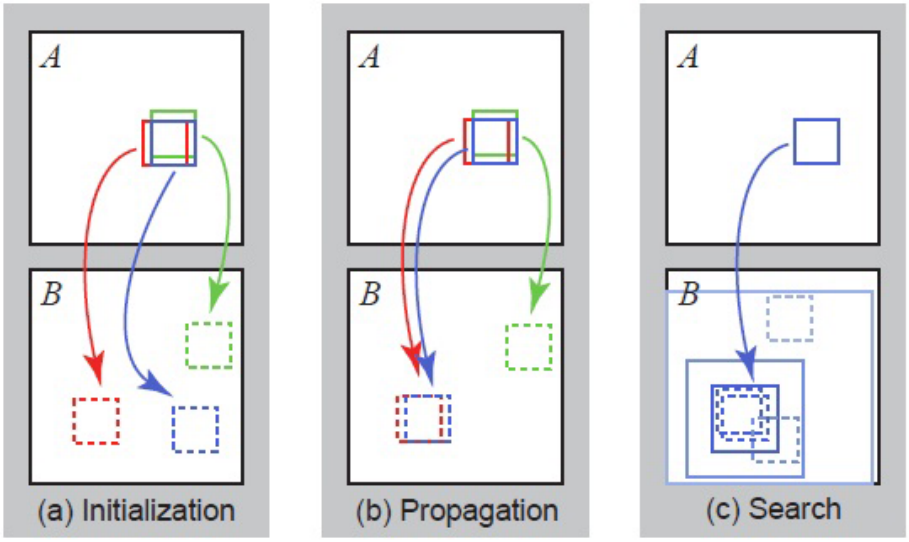


図 6.1: PatchMatch による類似パッチ探索．(a) まずランダムでパッチが割り当てられ，(b) 青パッチの上/緑パッチと左/赤パッチの類似パッチを調べ，現在のものより類似度が高ければ更新する．(c) そのパッチの周りをランダムに探索し，より類似度の高いもので更新する（図は文献 [10] より引用）．

$$\mathbf{u}_i = \mathbf{v}_0 + w\alpha^i \mathbf{R}_i \tag{6.1}$$

ここで，w とは最大探索半径，α は探索範囲の固定比率，$\mathbf{R}_i$ は $[-1,1] \times [-1,1]$ のランダム値である．

表 6.1 は PatchMatch アルゴリズムと kd-tree を用いた approximate nearest neighbor matching を比較したものである．kd-tree に対し，PatchMatch では計算速度が約 20 倍から 100 倍向上し，メモリー使用量は 1/20 以下に抑えられていることがわかる．

表 6.1: PatchMatch と kd-tree による approximate nearest neighbor matching の比較．

Megapixels	Time [s]		Memory [MB]	
	PatchMatch	kd-tree	PatchMatch	kd-tree
0.1	0.68	15.2	1.7	33.9
0.2	1.57	37.2	3.4	68.9
0.35	0.35	87.7	5.6	118.3

謝辞

本研究を進めるにあたり，本学大学院システム情報工学研究科の金森由博助教をはじめとし，三谷純准教授，福井幸男教授には多くのご指導・ご協力をいただきました．ここに深く感謝の意を表します．

遠藤結城氏には，プログラムの提供や研究に関する議論など，本研究に対する多くのご協力をいただきました．ここに感謝いたします．

また，本学システム情報工学研究科西原清一名誉教授や拓殖大学の水野一徳講師，および非数値処理アルゴリズム研究室の皆様には多くの有益なご意見をいただきました．ここに感謝いたします．

本学の大田友一教授，福井和広教授，酒井宏教授，高橋伸准教授には，本論文の審査をしていただき，貴重なコメントとアドバイスをいただきました．ここに感謝の意を表します．

本研究は多くの方々のご指導，ご協力によりなされたものであり，改めて心より御礼を申し上げます．

最後に，学生生活を支援してくれた両親に心から感謝いたします．

2015 年 3 月飯塚里志

参考文献

[1] Adobe photoshop. URL https://www.adobe.com/products/photoshop.html

[2] Google street view. URL https://www.google.com/maps/views/streetview

[3] Microsoft photosynth. URL https://photosynth.net

[4] A Category-Level 3-D Object Dataset: Putting the Kinect to Work (2011)

[5] Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned Salient Region Detection. In: Proceedings of CVPR 2009, pp. 1597 – 1604 (2009)

[6] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Su, S.: Slic superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2274–2282 (2012)

[7] An, X., Pellacini, F.: AppProp: All-pairs appearance-space edit propagation. ACM Trans. Graph. **27**(3), 40:1–40:9 (2008)

[8] Assa, J., Wolf, L.: Diorama construction from a single image. Computer Graphics Forum **26**(3), 599–608 (2007)

[9] Bai, X., Sapiro, G.: Geodesic matting: A framework for fast interactive image and video segmentation and matting. Int. J. Comput. Vision **82**(2), 113–132 (2009)

[10] Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B.: PatchMatch: A randomized correspondence algorithm for structural image editing. ACM Trans. Graph. (Proc. of SIGGRAPH) **28**(3), 24:1–24:11 (2009)

[11] Barnes, C., Shechtman, E., Goldman, D.B., Finkelstein, A.: The generalized patchmatch correspondence algorithm. In: Proceedings of ECCV 2010, pp. 29–43. Springer-Verlag, Berlin, Heidelberg (2010)

[12] Boykov, Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. In: Proceedings of ICCV, vol. 1, pp. 105–112 (2001)

[13] Brown, M., Lowe, D.: Unsupervised 3D object recognition and reconstruction in unordered datasets. In: Proceedings of the 5th International Conference on 3D Imaging and Modelling (3DIM05). Ottawa, Canada (2005)

[14] Chaurasia, G., Duchene, S., Sorkine-Hornung, O., Drettakis, G.: Depth synthesis and local warps for plausible image-based navigation. ACM Trans. Graph. **32**(3), 30:1–30:12 (2013)

[15] Chen, J., Paris, S., Durand, F.: Real-time edge-aware image processing with the bilateral grid. ACM Trans. Graph. **26**(3) (2007)

[16] Chen, S.E.: Quicktime vr: An image-based approach to virtual environment navigation. In: Proceedings of the 22Nd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1995, pp. 29–38. ACM, New York, NY, USA (1995)

[17] Chen, T., Zhu, Z., Shamir, A., Hu, S.M., Cohen-Or, D.: 3-sweep: Extracting editable objects from a single photo. ACM Trans. Graph. (Proc. of SIGGRAPH Asia 2013) **32**(6), Article 195 (2013)

[18] Chen, X., Zou, D., Zhao, Q., Tan, P.: Manifold preserving edit propagation. ACM Trans. Graph. **31**(6), 132:1–132:7 (2012)

[19] Cheng, M.M., Zhang, F.L., Mitra, N.J., Huang, X., Hu, S.M.: RepFinder: Finding approximately repeated scene elements for image editing. ACM Trans. Graph. **29**(4), 83:1–8 (2010)

[20] Cheng, M.M., Zhang, G.X., Mitra, N.J., Huang, X., Hu, S.M.: Global contrast based salient region detection. In: Proceedings of CVPR 2011, pp. 409–416 (2011)

[21] Cho, T.S., Avidan, S., Freeman, W.T.: The patch transform. IEEE Trans. Pattern Anal. Mach. Intell. **32**(8), 1489–1501 (2010)

[22] Comaniciu, D., Meer, P., Member, S.: Mean shift: A robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. **24**, 603–619 (2002)

[23] Criminisi, A., Reid, I., Zisserman, A.: Single view metrology. Int. J. Comput. Vision **40**(2), 123–148 (2000)

[24] Criminisi, A., Sharp, T., Blake, A.: Geos: Geodesic image segmentation. In: Proceedings of ECCV 2008, pp. 99–112. Springer-Verlag (2008)

[25] Debevec, P.E., Taylor, C.J., Malik, J.: Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. In: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, SIGGRAPH '96, pp. 11–20. ACM, New York, NY, USA (1996)

[26] Fattal, R.: Single image dehazing. ACM Trans. Graph. **27**(3), 72:1–72:9 (2008)

[27] Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. Int. J. Comput. Vision **59**(2), 167–181 (2004)

[28] flickr: URL https://www.flickr.com

[29] Furukawa, Y., Ponce, J.: Accurate, dense, and robust multiview stereopsis. IEEE Trans. Pattern Anal. Mach. Intell. **32**(8), 1362–1376 (2010)

[30] Goldberg, C., Chen, T., Zhang, F.L., Shamir, A., Hu, S.M.: Data-driven object manipulation in images. Computer Graphics Forum **31**(2pt1), 265–274 (2012)

[31] Hoiem, D., Efros, A.A., Hebert, M.: Automatic photo pop-up. ACM Trans. Graph. **24**(3), 577–584 (2005)

[32] Hoiem, D., Efros, A.A., Hebert, M.: Geometric context from a single image. In: Proceedings of ICCV 2005, pp. 654–661 (2005)

[33] Hoiem, D., Efros, A.A., Hebert, M.: Putting objects in perspective. In: Proceedings of CVPR 2006, pp. 2137–2144. IEEE Computer Society, Washington, DC, USA (2006)

[34] Hoiem, D., Efros, A.A., Hebert, M.: Recovering surface layout from an image. Int. J. Comput. Vision **75**(1), 151–172 (2007)

[35] Hoiem, D., Efros, A.A., Hebert, M.: Recovering occlusion boundaries from an image. Int. J. Comput. Vision **91**(3), 328–346 (2011)

[36] Horry, Y., Anjyo, K.I., Arai, K.: Tour into the picture: using a spidery mesh interface to make animation from a single image. In: Proceedings of the 24th annual conference on Computer graphics and interactive techniques, SIGGRAPH '97, pp. 225–232. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA (1997)

[37] Hou, X., Zhang, L.: Saliency detection: A spectral residual approach. In: Proceedings of CVPR 2007, pp. 1–8 (2007)

[38] Jia, J., Sun, J., Tang, C.K., Shum, H.Y.: Drag-and-drop pasting. ACM Trans. Graph. **25**(3), 631–637 (2006)

[39] Jiang, N., Tan, P., Cheong, L.F.: Symmetric architecture modeling with a single image. ACM Trans. Graph. **28**(5), 113:1–113:8 (2009)

[40] Kanade, T.: Recovery of the three-dimentional shape of an object from a single view pp. 409–460 (1981)

[41] Kang, H.W., Pyo, S.H., Anjyo, K., Shin, S.Y.: Tour into the picture using a vanishing line and its extension to panoramic images. Computer Graphics Forum **20**(3), 132–141 (2001)

[42] Karsch, K., Liu, C., Kang, S.B.: Depth extraction from video using non-parametric sampling. In: Proceedings of ECCV 2012, pp. 775–788 (2012)

[43] Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. Int. J. Comput. Vision **1**(4), 321–331 (1988)

[44] Koyama, T., Kitahara, I., Ohta, Y.: Live mixed-reality 3d video in soccer stadium. The Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR2003), pp. 178–187 (2002)

[45] Lalonde, J.F., Hoiem, D., Efros, A.A., Rother, C., Winn, J., Criminisi, A.: Photo clip art. ACM Trans. Graph. (Proc. of SIGGRAPH 2007) **26**(3), 3 (2007)

[46] Lang, M., Hornung, A., Wang, O., Poulakos, S., Smolic, A., Gross, M.: Nonlinear disparity mapping for stereoscopic 3d. ACM Trans. Graph. **29**(3), 10 (2010)

[47] Li, Y., Ju, T., Hu, S.M.: Instant propagation of sparse edits on images and videos. Computer Graphics Forum **29**(7), 2049–2054 (2010)

[48] Li, Y., Sun, J., Tang, C.K., Shum, H.Y.: Lazy snapping. ACM Trans. Graph. **23**(3), 303–308 (2004)

[49] Lindeberg, T., G arding, J.: Shape from texture from a multi-scale perspective. In: Proceedings of ICCV 1993, pp. 683–691 (1993)

[50] Liu, B., Gould, S., Koller, D.: Single image depth estimation from predicted semantic labels. In: Proceedings of CVPR 2010, pp. 1063–1260 (2010)

[51] Liu, J., Sun, J., Shum, H.Y.: Paint selection. ACM Trans. Graph. **28**(3), 69:1–69:7 (2009)

[52] Lo, W.Y., van Baar, J., Knaus, C., Zwicker, M., Gross, M.: Stereoscopic 3D copy & paste. ACM Trans. Graph. **29**(6), 147:1–147:10 (2010)

[53] Lopez, A., Garces, E., Gutierrez, D.: Depth from a single image through user interaction. In: Proceedings of CEIG 2014, pp. 1–10 (2014)

[54] Lu, Y., Zhang, J., Wu, Q., Li, Z.N.: A survey of motion-parallax-based 3-d reconstruction algorithms. IEEE Trans. Systems, Man, and Cybernetics **34**, 532–548 (2004)

[55] Malik, J., Rosenholtz, R.: Computing local surface orientation and shape from texture for curved surfaces. Int. J. Comput. Vision **23**, 149–168 (1997)

[56] Mortensen, E.N., Barrett, W.A.: Intelligent scissors for image composition. In: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques, SIGGRAPH '95, pp. 191–198. ACM (1995)

[57] Oh, B.M., Chen, M., Dorsey, J., Durand, F.: Image-based modeling and photo editing. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, SIGGRAPH '01, pp. 433–442. ACM, New York, NY, USA (2001)

[58] Olsen Jr., D.R., Harris, M.K.: Edge-respecting brushes. In: Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology, UIST '08, pp. 171–180. ACM, New York, NY, USA (2008)

[59] Oswald, M.R., Toeppe, E., Cremers, D.: Fast and globally optimal single view reconstruction of curved objects. In: Proceedings of CVPR 2012, pp. 534–541. Providence, Rhode Island (2012)

[60] Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man and Cybernetics **9**, 62–66 (1979)

[61] Perazzi, F., Krähenbühl, P., Pritch, Y., Hornung, A.: Saliency filters: Contrast based filtering for salient region detection. In: Proceedings of CVPR 2012, pp. 733–740 (2012)

[62] Pérez, P., Gangnet, M., Blake, A.: Poisson image editing. ACM Trans. Graph. **22**(3), 313–318 (2003)

[63] Reinhard, E., Ashikhmin, M., Gooch, B., Shirley, P.: Color transfer between images. IEEE Computer Graphics and Applications **21**(5), 34–41 (2001)

[64] Reynolds, J.H., Desimone, R.: Interacting roles of attention and visual salience in v4. Neuron **37**(5), 853–863 (2003)

[65] Ribera, R.B.i., Choi, S., Kim, Y., Lee, J., Noh, J.: Video panorama for 2D to 3D conversion. Computer Graphics Forum **31**(7pt2), 2213–2222 (2012)

[66] Rother, C., Kolmogorov, V., Blake, A.: GrabCut: interactive foreground extraction using iterated graph cuts. ACM Trans. Graph. **23**(3), 309–314 (2004)

[67] Saxena, A., Chung, S.H., Ng, A.Y.: 3D depth reconstruction from a single still image. Int. J. Comput. Vision **76**(1), 53–69 (2008)

[68] Saxena, A., Sun, M., Ng, A.Y.: Make3d: Learning 3d scene structure from a single still image. IEEE Trans. Pattern Anal. Mach. Intell. **31**(5), 824–840 (2009)

[69] Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comput. Vision **47**(1-3), 7–42 (2002)

[70] Shade, J., Gortler, S., He, L.w., Szeliski, R.: Layered depth images. In: Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '98, pp. 231–242. ACM (1998)

[71] Simakov, D., Caspi, Y., Shechtman, E., Irani, M.: Summarizing visual data using bidirectional similarity. In: Proceedings of CVPR 2008, pp. 1–8 (2008)

[72] Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: Exploring photo collections in 3d. ACM Trans. Graph. **25**(3), 835–846 (2006)

[73] Sun, J., Jia, J., Tang, C.K., Shum, H.Y.: Poisson matting. ACM Trans. Graph. **23**(3), 315–321 (2004)

[74] Töppe, E., Nieuwenhuis, C., Cremers, D.: Relative volume constraints for single view 3D reconstruction. In: Proceedings of CVPR 2013, pp. 177–184 (2013)

[75] Wang, O., Lang, M., Frei, M., Hornung, A., Smolic, A., Gross, M.: StereoBrush: interactive 2D to 3D conversion using discontinuous warps. In: Proceedings of the Eighth Eurographics Symposium on Sketch-Based Interfaces and Modeling, SBIM '11, pp. 47–54. ACM (2011)

[76] Wei, Y., Wen, F., Zhu, W., Sun, J.: Geodesic saliency using background priors. In: Proceedings of ECCV 2012, pp. 29–42. Springer-Verlag (2012)

[77] Wexler, Y., Shechtman, E., Irani, M.: Space-time completion of video. IEEE Trans. Pattern Anal. Mach. Intell. **29**(3), 463–476 (2007)

[78] Wu, T.P., Tang, C.K., Brown, M.S., Shum, H.Y.: Natural shadow matting. ACM Trans. Graph. **26**(2) (2007)

[79] Xu, L., Yan, Q., Jia, J.: A sparse control model for image and video editing. ACM Trans. Graph. **32**(6), 197:1–197:10 (2013)

[80] Yatziv, L., Bartesaghi, A., Sapiro, G.: O(n) implementation of the fast marching algorithm. Journal of Computational Physics **212**, 393–399 (2005)

[81] Yatziv, L., Yatziv, L., Sapiro, G., Sapiro, G.: Fast image and video colorization using chrominance blending. IEEE Transaction on Image Processing **15**, 2006 (2004)

[82] Yücer, K., Jacobson, A., Hornung, A., Sorkine, O.: Transfusive image manipulation. ACM Trans. Graph. (Proc. of SIGGRAPH Asia 2012) **31**(6), 176:1–176:9 (2012)

[83] Yücer, K., Sorkine-Hornung, A., Sorkine-Hornung, O.: Transfusive weights for content-aware image manipulation. In: Proceedings of the Vision, Modeling and Visualization Workshop (VMV), pp. 57–64. Eurographics Association (2013)

[84] Zhai, Y., Shah, M.: Visual attention detection in video sequences using spatiotemporal cues. In: Proceedings of ACM Multimedia 2006, pp. 815–824 (2006)

[85] Zhang, L., Dugas-Phocion, G., Samson, J.S., Seitz, S.M.: Single view modeling of free-form scenes. In: Proceedings of CVPR 2002, pp. 990–997 (2002)

[86] Zhang, R., Tsai, P.S., Cryer, J.E., Shah, M.: Shape from shading: A survey. IEEE Trans. Pattern Anal. Mach. Intell. **21**(8), 690–706 (1999)

[87] Zheng, Y., Chen, X., Cheng, M.M., Zhou, K., Hu, S.M., Mitra, N.J.: Interactive images: Cuboid proxies for smart image manipulation. ACM Trans. Graph. **31**(4), 99:1–99:11 (2012)