

# 写真の中の仮想世界へ

## 1枚の景観画像からのインタラクティブな3Dシーン生成システム

1枚の景観画像から3Dモデルを容易に生成できるシステムを開発した。生成される3Dモデルは背景モデルと前景物モデルから構成され、ユーザが地面領域と立体物との「境界線」をインタラクティブに指定することで、各モデルの3次元座標が算出される。さらに本システムでは、前景物モデリングやテクスチャ生成を簡単に行うことができ、簡素ながら十分な3D効果の得られる3Dシーンを効率よく作成することができる。

### ○ はじめに

近年、写実的なレンダリングを効率よく行うため、イメージベースによるレンダリング手法が多く研究されている。この手法は画像をテクスチャとして用いることで複雑な物体形状を正確にモデル化しなくてもリアリティのある3D空間が作成でき、高速なレンダリングが可能となる。作成された空間の視点を変更することにより、ユーザはその空間を実際に歩いているような体験をすることができ、バーチャルツアーなど特定の景観を3次元的に表現するときによく用いられる。しかし、一般的にそのような3Dシーンを作成するには多くの入力画像<sup>1)</sup>や専門的な知識が必要である上、非常に手間のかかる作業である。

これらの問題に対して、近年では自動または手動によって1枚の画像のみから3Dシーンを生成する手法が研究されている。自動生成の場合、領域分割や機械学習を用いて画像中の各領域の幾何ラベルを推定する手法が提案されている<sup>4)12)</sup>。これらの手法は手作業を必要としないが、奥行き推定や領域推定に失敗する 경우가少なくなく、適用可能な画像に限られる。また、木や人などの前景物があるシーンではうまくいかない場合が多い。これに対し、ユーザが画像の領域情報や奥行き情報を与えることで対話的に3Dシーンを生成する手法が提案されている<sup>5)6)7)9)</sup>。これらの手法は自動生成に比べ、より多くの画像を3D

モデル化できるが、領域指定やテクスチャ編集などに手間がかかる。また、ユーザインタラクションもあまり直感的ではない。

そこで我々は1枚の画像を入力として、簡単で直感的な操作で3Dシーンを作成できるシステムを開発した。本システムではまずユーザが地面と立体物の境界線をインタラクティブに指定する。この境界線によってモデルの3次元座標が算出され、3Dシーンモデルが生成される(図1)。また、入力画像の領域分割とグラフィック<sup>2)</sup>ベースの最適化を組み合わせることにより、簡単に前景物の抽出とモデリングを行うことができる。これにより、



(a) 入力画像



(b) 境界線と前景物

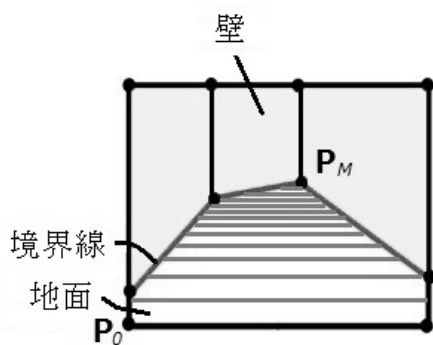


(c) 3Dモデル

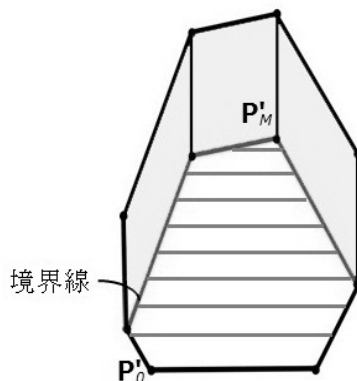


(d) 視点移動

第1図 提案システムによる3Dシーン生成の流れ



(a) 2D 画像



(b) 3D モデル

## 第2図 3D モデル生成

前景物を含む景観画像でも 3D シーンモデルが容易に作成できる。生成されるモデルは単純なモデルであり、入力画像の 3 次元構造を正確に再現するものではないが、ユーザに十分な 3D 効果を与えることができる。

### ○ 提案システム

本システムによって生成される 3D シーンは背景モデルと前景物モデルから成る。これらのモデルは少数の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。このシステムにおいて、3D シーン作成のためにユーザは 2 つの簡単な作業を行う。すなわち入力画像の境界線と前景物の指定である。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。

### 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 2)。この各頂点に適切な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。また、3D 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。ここで同次座標  $(x, y, z, w)$  は  $w \in [0, 1]$  が小さいほど 3 次元空間の奥に透視投影変換される。よって入力画像のスクリーン座標の原点  $P_0$  を  $(x_0, y_0)$ 、境界線上で最も  $y$  座標が大きな頂点  $P_M$  を  $(x_M, y_M)$  とし、ワールド座標系における  $P_0, P_M$  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  $P'_0, P'_M$  は以下のように表される：

$$P'_0: (x'_0, y'_0, f, 1)$$

$$P'_M: (x'_M, y'_M, f, w_{min})$$

ここで  $w_{min}$  とは小さな正の値であり、本システムでは 0.1 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：

$$P'_i: (x'_i, y'_i, f, w_i)$$



(a) 入力画像と前景領域 (破線)

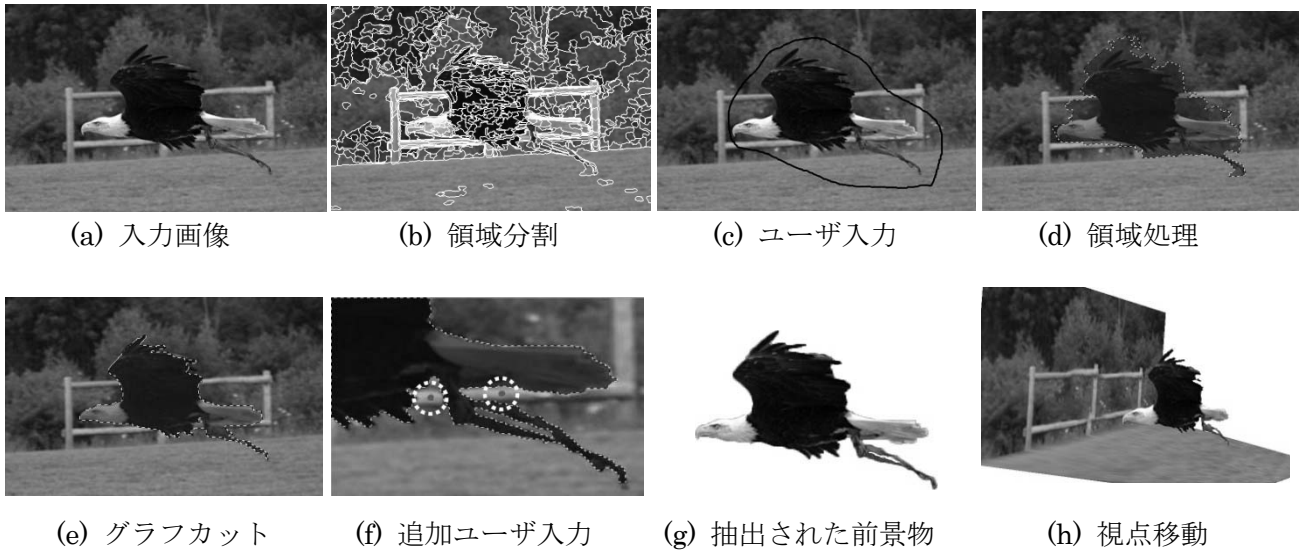


(b) 背景推定 (制約なし)



(c) 背景生成 (制約あり)

## 第3図 背景テクスチャの生成



第4図 前景物の抽出

ただし、

$$w_i = \frac{y_i - y_0}{y_M - y_0} w_{min} + 1$$

$$- \frac{y_i - y_0}{y_M - y_0}$$

... (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3D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<sup>6)</sup>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座標も算出できる。

**背景テクスチャ** 生成されたポリゴンモデルに背景テクスチャを貼り付けることで背景モデルが完成する。本稿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。本システムでは補完のためにBarnesらのPatchMatchと呼ばれる手法<sup>7)</sup>を用いる。彼らの手法により、計算コストの高い類似領域探索を高速化し、画像を高速に自動補完することができる。本システムではこの手法を実装し、指定された前景物が抜き出されると同時にその領域が自動的に補完される。よって、すべての前景物の抽出が完了すると同時に完全な背景テクスチャが生成される。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ふ(図3(b))。

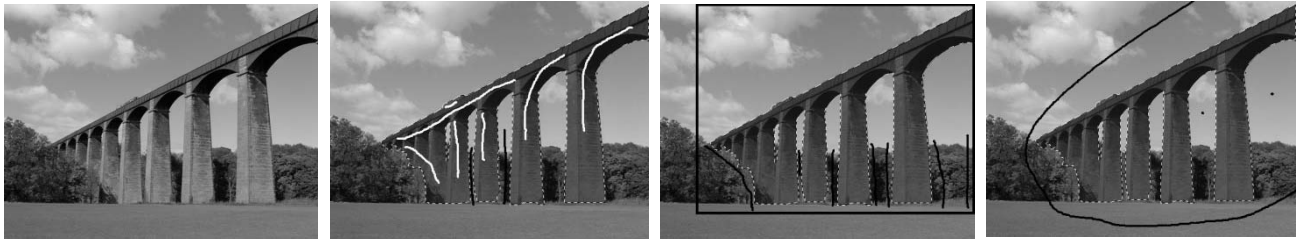
Barnesらは、画像補完の際にガイド線上の類似領域探索をその線上に拘束することで画像構造を補完に反映させる手法を提案している。本システムではこの理論にもとづき、境界線を補完の拘束条件として用いることで前景物の補完精度を向上させている(図3(c))。つまり、境界線は3次元座標の算出と背景テク

スチャ生成の精度向上という2つの役割を果たしている。

#### 前景物モデル

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<sup>10)</sup>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<sup>8)14)</sup>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<sup>2)</sup>を組み合わせることで前景最適化を行う。

本システムでは入力画像は読み込まれると同時に領域分割が適用される(図4(b))。この領域分割のために、我々はComaniciuらの手法<sup>3)</sup>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



(a) 入力画像

(b) Photoshop

(c) *GrabCut*<sup>11)</sup>

(d) 提案手法

第 5 図 提案手法と既存手法との比較

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。

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 4 (c))。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 4 (d))。これを初期状態とし、さらにグラフカットをベースとする *GrabCut* と呼ばれる手法<sup>11)</sup>にもとづいて領域の最適化を行う (図 4 (e))。

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処理を追加し、ユーザによる前景/背景指定、指定範囲を領域単位に拡張、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。

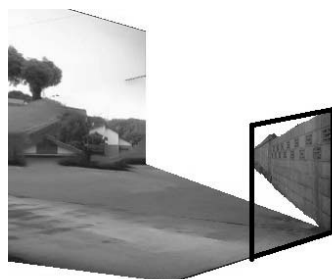
本システムによる前景物抽出結果を図 5 に示す。入力画像のサイズは 800×600 ピクセルである。提案手法のユーザ入力に対する最適化処理時間は約 0.41 秒であった。また、ユーザの作業時間も含めて前景物抽出にかかった時間は、強力な画像編集ツールである Adobe Photoshop CS5 の *Quick Selection* ツールでは 107 秒、我々の実装による Rother らの *GrabCut* では 115

秒、提案手法では 18 秒であった。よって図 5 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。

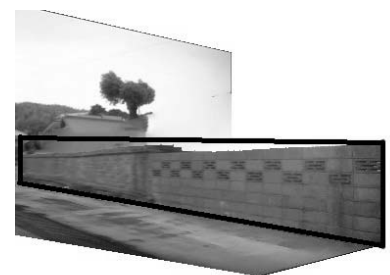
このようにして抽出された前景物画像を 1 枚の四角形ポリゴンに前景物テクスチャとしてマッピングすることで前景物モデルが生成され、背景モデルの地面領域に垂直に配置される。この前景物モデルの各頂点の 3 次元座標は、前景物領域の中で最も小さい y 座標にもとづき式(1)と垂直条件から計算できる。



(a) 前景と接地制約 (白線)



(b) 接地制約なし



(c) 接地制約あり

第 6 図 前景物の接地制約



(a) 入力画像と境界線 (白線)



(b) 3D モデル



(c) 視点移動

第 7 図 曲線形状の境界線をもつ室内画像



(a) 入力画像とユーザ入力



(b) 視点移動



第 8 図 提案システムによる 3D シーン生成

**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。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 (図 6)。

### ○ 結果と考察

本稿において、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。図 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 7 は曲線と直線を含む境界線をもつ屋内の写真である。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 3D シーンを生成することができる。また、図 8 のように複数の前景物や夜のシーンなどでも、本システムを用いて簡単に 3D モデルを生成することができる。

これらの 3D 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 7 の場合、3D シーンを作成するのににかかった時間が 14 秒であったのに対し、複数の前景物をもつ図 8 下段では作業時間は 110 秒であった。このように 3D 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高速化は 3D シーンを作成する上で非常に重要な役割を果たしていることがわかる。

## ○ 結論

本稿では、我々が開発した 1 枚の景観画像から容易に 3D シーンを生成できるシステムについて述べた。生成されるシーンモデルは背景モデルと前景物モデルから構成され、ユーザがインタラクティブに指定した地面と壁の境界線からその 3 次元座標が算出される。この境界線を用いたモデリングにより、様々な景観画像の 3D シーンを作成することが可能となる。さらに提案システムでは前景物の簡易抽出やモデリング、背景テクスチャの自動生成を行うことができる。これによりユーザの負担が少ない効率的なシーンモデリングが可能となる。

今後の研究では、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、それを反映した立体的な前景物モデルが作成できる機能を実現

したいと考えている。また、透視投影によって引き伸ばされるテクスチャを高解像度化することでよりシーンのリアリティを向上させたい。

## 参考文献

- 1) BARNES, C., SHECHTMAN, E., FINKELSTEIN, A., AND GOLDMAN, D. B. 2009. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Trans. Graph.* 28, 3, 24:1–11.
- 2) BOYKOV, Y., AND JOLLY, M. P. 2001. Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. In *Proceedings of ICCV*, 105–112.
- 3) COMANICIU, D. AND MEER, P. 2002. A robust approach toward feature space analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24, 5, 603–619.
- 4) HOIEM, D., EFROS, A. A., AND HEBERT, M. 2005. Automatic photo pop-up. *ACM Trans. Graph.* 24, 3, 577–584.
- 5) HORRY, Y., ANJYO, K.-I., AND ARAI, K. 1997. Tour into the picture: using a spidery mesh interface to make animation from a single image. In *Proceedings of ACM SIGGRAPH* 97, 225–232.
- 6) KANG, H., PYO, S., ANJYO, K., AND SHIN, S. 2001. Tour into the picture using a vanishing line and its extension to panoramic images. In *Proceedings of Eurographics*, 132–141.
- 7) LIEBOWITZ, D., CRIMINISI, A., AND ZISSERMAN, A. 1999. Creating Architectural Models from Images. In *Proceedings of Eurographics*, 39–50.
- 8) MORTENSEN, E. N., AND BARRETT, W. A. 1995.

Intelligent scissors for image composition. In *Proceedings of ACM SIGGRAPH*.

- 9) OH, B. M., CHEN, M., DORSEY, J., AND DURAND, F. 2001. Image-based modeling and photo editing. In *Proceedings of ACM SIGGRAPH*, 433–442.
- 10) OLSEN, JR., D. R., AND HARRIS, M. K. 2008. Edge-respecting brushes. In *UIST*, 171–180.
- 11) ROTHER, C., BLAKE, A., AND KOLMOGOROV, V. 2004. Grabcut - interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.* 24, 3, 309–314.
- 12) SAXENA, A., SUN, M., AND NG, A.Y. 2009. Make3D: Learning 3D Scene Structure from a Single Still Image. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 31, 5, 824–840.
- 13) SNAVELY, N., SEITZ, S. M., AND SZELISKI, R. 2006. Photo tourism: Exploring photo collections in 3d. In *SIGGRAPH Conference Proceedings*, ACM Press, New York, NY, USA, 835–846.
- 14) WANG, J., AGRAWALA, M., AND COHEN, M. F. 2007. Soft scissors: an interactive tool for realtime high quality matting. *ACM Trans. Graph.* 27, 3, 9.

## 【筆者紹介】 飯塚里志

筑波大学大学院

システム情報工学研究科 コンピュータ

サイエンス専攻 博士前期課程 2 年

〒305-8573 茨城県つくば市天王台

1-1-1

TEL : 029-853-5388

FAX : 029-853-5388

Email : iizuka@npal.cs.tsukuba.ac.jp

## 金森由博

筑波大学大学院

システム情報工学研究科 コンピュータ

サイエンス専攻 助教

〒305-8573 茨城県つくば市天王台

1-1-1

TEL : 029-853-5388

FAX : 029-853-5388

Email : kanamori@cs.tsukuba.ac.jp

## 三谷純

筑波大学大学院

システム情報工学研究科 コンピュータ

サイエンス専攻 准教授

〒305-8573 茨城県つくば市天王台

1-1-1

TEL : 029-853-5388

FAX : 029-853-5388

Email : mitani@cs.tsukuba.ac.jp

## 福井幸男

筑波大学大学院

システム情報工学研究科 コンピュータ

サイエンス専攻 教授

〒305-8573 茨城県つくば市天王台

1-1-1

TEL : 029-853-5524

FAX : 029-853-5524

Email : fukui@cs.tsukuba.ac.jp