

地面位置の推定に基づく2次元画像からの擬似3次元復元

石橋 薫[†] 有木 康雄[‡] 滝口 哲也[‡]

[†] 神戸大学大学院工学研究科 〒657-8501 神戸市灘区六甲台町 1-1

[‡] 神戸大学自然科学系先端融合研究環 〒657-8501 兵庫県神戸市灘区六甲台町1-1

E-mail: [†] ishibashi@me.cs.scitec.kobe-u.ac.jp [‡] { ariki, takigu } @kobe-u.ac.jp

あらまし 動画画像からの3次元復元では様々な研究がなされているが、静止画からの3次元復元の研究は少なく、地面や壁面などの簡素な復元にとどまっている。この問題に対して、本研究では、地面の輪郭線や人の存在を考慮した静止画擬似3次元復元を行う。まず画像を地面、壁、空の3領域に分割し、構図の情報に従って3次元に復元する。人を検出した場合には、既に設定した地面に配置することでその場所に立っているように見せる事が可能である。

キーワード 3次元, 画像ラベリング, 構図情報, 人物認識

Pseudo Three-Dimensional Reconstruction from Two Dimensional Image Based on Presumption of Ground Position

Kaoru ISHIBASHI[†] Yasuo ARIKI[‡] and Tetsuya TAKIGUCHI[‡]

[†] Graduate School of Engineering, Kobe University 1-1Rokkodaicho, Nada-ku, Kobe,657-8501 Japan

[‡] Organization of Advanced Science and Technology, Kobe University 1-1, Rokkodai, Nada, Kobe, 657-8501 Japan

E-mail: [†] ishibashi@me.cs.scitec.kobe-u.ac.jp [‡] { ariki, takigu } @kobe-u.ac.jp

Abstract There are various researches for three dimensional reconstruction from the dynamic scene, but few for the 3D reconstruction from the still image. Moreover, it is limited to the simple reconstruction such as ground and walls. In order to improve the 3D reconstruction quality, this paper processes a pseudo-3D reconstruction that takes into consideration of the ground line and the person existence. The image was first labeled into ground, walls and sky, and then 3D reconstruction is carried out according to the information of their composition. When the persons are detected as standing, they are placed on the detected ground.

Keyword three dimensions, image labeling, composition information, person recognition

1. はじめに

今日、エンターテインメントでは、3次元のコンピュータグラフィクスが一般的に用いられている。また、実写画像、動画の3次元化も始まっており、3次元写真を視差画像から簡単に作成できるデジカメや、動画を3次元として出力するテレビなどが存在する。このような状況から、ユーザーは3次元のモデルにより親しむようになり、今後、ユーザー自身が3次元のモデルを作り出したいという欲求も生まれてくると考えられる。しかし、簡単に3次元のモデルを作ることが出来るようになったとはいえ、それは簡単な形状を作る場合の話である。例えば写真に写っている光景をそのまま3次元化するようなことは、専門知識のない人々にとっては困難である。

1枚のデジタル画像を3次元に置き換える研究には、3

次元レーザースキャナから得た色・テクスチャ情報を学習し、そこから推論を行う Make3D[3]の方法や、画像に水平面、垂直面(4種)、空というラベリングを施し、それを元に3次元構造を推定する Automatic Photo PopUp[4]の方法がある。しかし、前者はデータの収集に高価な機材が必要であり、しかもその普遍性ゆえに安定した結果が得られないという問題がある。後者は構図が簡易すぎて、しばしば不都合な結果を作り出している。

本研究では、[4]の手法をベースに、各段階の手法を改善することにより、1枚の写真から簡易な擬似3次元復元を可能とする方法を提案する。

本研究の手法は、画像中の水平面(≒地面)の領域に着目し、ラベリング結果とカメラ姿勢の推定を基に、画像中の地面位置から3次元位置を計算するものである。人

が検出された場合は、地面位置を考慮し結果を統合する。

2. 提案手法

特定の建物や人物ではなく、画像に写っている全体を3次元復元する。

2.1 問題設定

本実験の復元対象は図1に示すように、1枚のデジタル画像である。対象画像には地面(基点となる水平な領域)が写っている必要がある。人物については水平面、垂直面、空という3種類のラベルとは別に、人領域を検出する。その理由は、人は地面を含む画像領域に写っている可能性が高いこと、先行研究により人の検出が比較的容易であることが挙げられる。



図1 復元対象の画像例(左:復元可能 右:復元不可能)

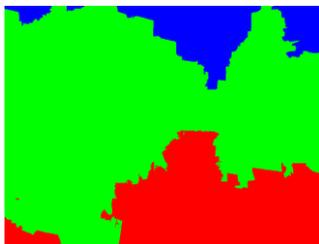
2.1 提案手法の流れ

まず、図2のように、2次元静止画像を地面、壁、空の3種にラベル付けする。この時、人物が写っていればラベル画像とは別に人検出結果の画像を用意する。



(a)入力画像

(b)人検出結果画像



(c)ラベル画像

(青:空 緑:垂直面 赤:水平面(地面))

図2 入力画像とラベル画像

ラベリングには[1]の手法を、人検出にはOpenCVに実装された[12]の手法を用いた。

次に、カメラの姿勢を決定する。今回の実験では、地面からカメラまでの距離を手動で設定した定数とし、水平線を[6]の手法により推定した。具体的には、図3に示すように画像中最も上方にある地面領域(図3 赤色)のピクセル P_{ground} と、画像中最も下方にある空領域(図3 青色)のピクセル P_{sky} のY座標を用いて、画像中の水平線の高さ Y_{hor} を次式で求める。

$$Y_{hor} = \frac{(P_{sky} + P_{ground})}{2} \quad (1)$$

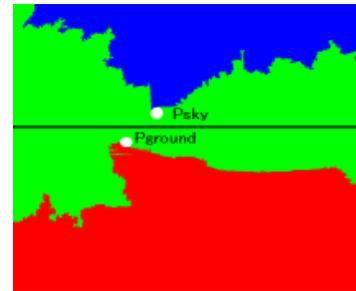


図3 水平線位置の推定例

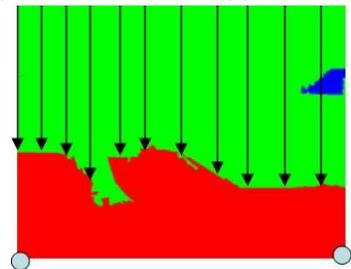
画像の横方向が水平線と平行になっていると仮定し、カメラの仰角を決定する。カメラの地面からの距離と仰角によりカメラの姿勢を決定する。

次に、図4に示すように地面領域の輪郭を取得する。画像を1列ずつ画面上部から下方向に走査し、最初に検出した地面領域を、復元に使用する地面の輪郭として登録する。



(a)入力画像

(b)ラベル画像



(c)地面領域の輪郭

図4 地面領域の輪郭検出

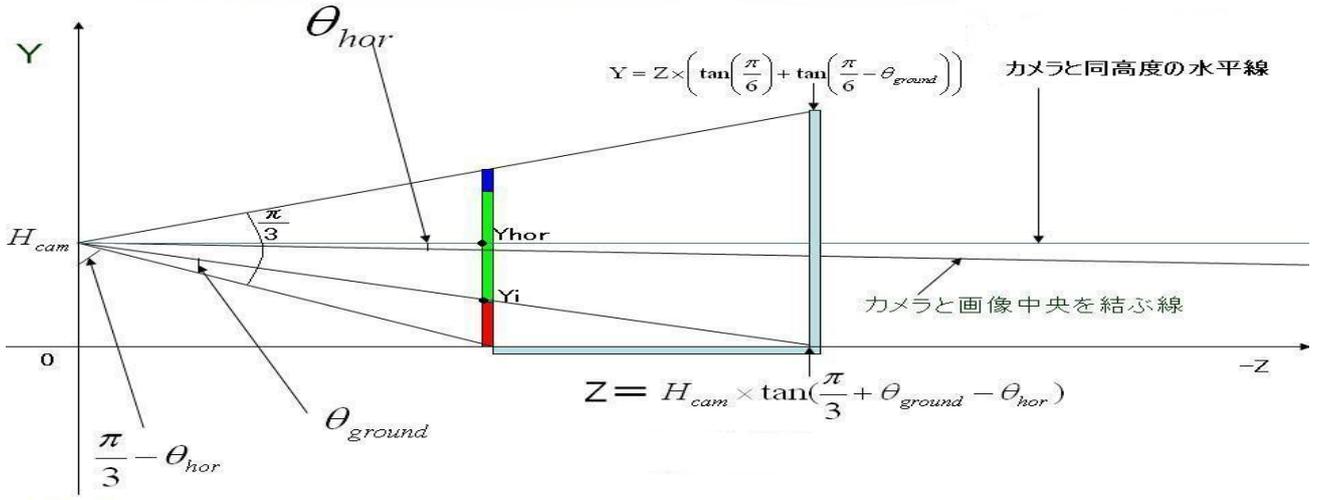


図5 地面と垂直面の取得

また、視野角はX方向Y方向ともに60度にする。

登録した地面の輪郭は画像中の座標であるので、これを3次元座標に変換する。図5に示すように、まず、

$$\theta_{hor} = \frac{\left(\frac{Y_{img} - Y_{hor}}{2}\right)}{Y_{img}} \times \frac{\pi}{3} \quad (2)$$

$$\theta_{ground} = \frac{(Y_{img} - Y_i)}{Y_{img}} \times \frac{\pi}{3} \quad (3)$$

を計算する。 Y_{img} は入力画像の縦のピクセル数、 Y_i は登録した地面座標の画像上のY座標値であり、 Y_{hor} はカメラと同じ高さの画像上のY座標値である。画像中の原点は、画像の上側左にある。 θ_{hor} はカメラと同じ高さの水平線が画像中央からどの程度ズレているかを表しており ($-30^\circ \sim +30^\circ$)、 θ_{ground} は画像中の地面座標の高さに対応する角度を表している ($0^\circ \sim 60^\circ$)。地面は「高さ0の水平な領域」と定義する。つまり、登録した地面座標のX座標を X_i 、入力画像の横のピクセル数を X_{img} とすると、画像上の点に対応する3次元空間上の地面の位置は、 $Y=0$ であり、 Z 、 X は次式で求められる。

$$Z = H_{cam} \times \tan\left(\frac{\pi}{3} + \theta_{ground} - \theta_{hor}\right) \quad (4)$$

$$X = Z \times \tan\left(\frac{\left(\frac{X_i - X_{img}}{2}\right)}{X_{img}} \times \frac{\pi}{3}\right) \quad (5)$$

次に、壁を指定する。地面座標点1つにつき壁座標点1つが対応するように設定する。即ち、 Z と X が式(4)、(5)と等しく、

$$Y = Z \times \left(\tan\left(\frac{\pi}{6}\right) + \tan\left(\frac{\pi}{6} - \theta_{ground}\right)\right) \quad (6)$$

である点を壁座標点として設定する。隣り合う壁座標点と地面座標点の2組4点で囲まれた範囲を順次壁面として設定する。

2. 2人物の描画

入力画像に対して人物検出を行い、その結果を2.1節で再構成した3次元環境に追加して出力する。

図6に示すように、人領域として検出された中で、画像中最下部にある点を接地点とし、 Z 座標を求める。この Z 座標を基準点と呼ぶ。この Z 座標は式(4)で求められる。基準点における人領域のY座標は、画像中の人領域のY座標を Y_i 、一番下のY座標を Y_{low} とすると、次式のように表される。

$$Y = Z \times \left(\tan\left(\frac{\pi}{6}\right) + \tan\left(\frac{\pi}{6} - \theta_{ground}\right)\right) \times \left(\frac{|Y_{low} - Y_i|}{Y_{low}}\right) \quad (7)$$

これは式(6)に示した壁座標点の縮小になっている。

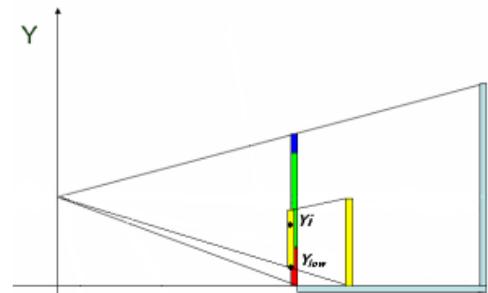


図6 人領域の3次元変換

3. 実験

複数枚の画像を用いて3次元復元した結果を示す。実験に使用する画像は、3次元カメラで撮影した左目画像1枚である。

また、3次元カメラから得られる左目、右目画像から算出した奥行き画像と、3次元復元によって得られた奥行き画像を比較して示す。

実験に使用した画像の解像度は878×684画素である。



(a)入力画像

(b)視点を前方へ移動した画像

(c)上方から見下ろした画像



(a)入力画像

(b)視点を前方へ移動した画像

(c)上方から見下ろした画像



(a)入力画像

(b)視点を前方へ移動した画像

(c)上方から見下ろした画像

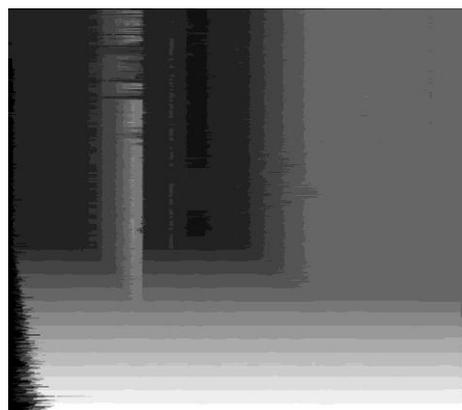
図7 復元結果



(a1)入力画像



(b1) 3次元カメラから算出した奥行き画像



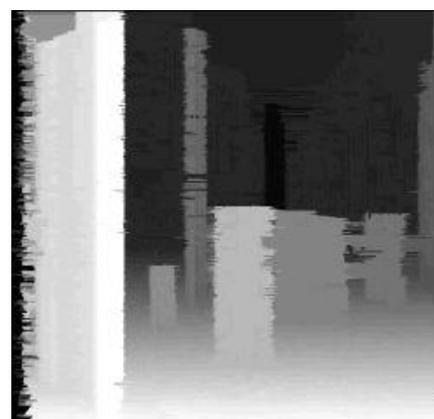
(c1)復元した奥行き画像



(a2)入力画像



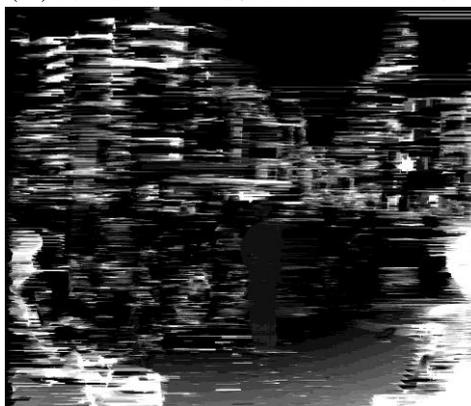
(b2) 3次元カメラから算出した奥行き画像



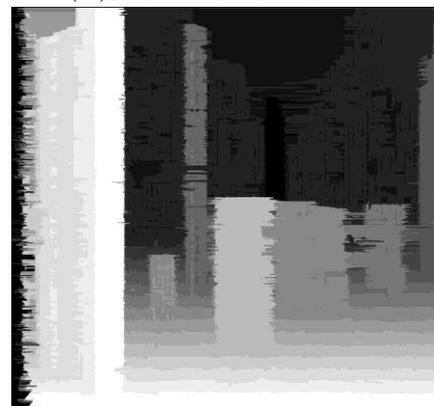
(c2)復元した奥行き画像



(a3)入力画像



(b3) 3次元カメラから算出した奥行き画像



(c3)復元した奥行き画像

図8 3次元カメラから算出した奥行き画像と復元した奥行き画像

4. 結果

図7の結果の考察をまとめる。

上段のように水平な地面の上に建物が建っているような画像に対しては、復元できている。しかし、木に対して処理を施していないので画面左については不自然になっている。また、手前の地面と奥の水面は高低差があるが、それが表現できていない。

中段のように遮蔽物が無く、平面のみの画像に対しては有効に処理できている。ポリゴンモデルの表示限界を超えて、遠くの距離を表現することが出来ている。画面左の歩道と柵は表現できていない。

下段のように人の写った画像に対しては、前述したアルゴリズムにより、写真に写った位置に問題なく立たせることに成功している。ただ、背景に適用するテ

クスタチャとして元画像を使用しているため、人の影が地面に写ってしまっている。

5. まとめと今後の展望

本論文では、画像を地面、壁、空という3つのラベルに分け、地面位置の情報を使うことで3次元復元する方法を提案した。

特長を挙げると、ラベル画像から計算して3次元形状を決定するので、ラベリングを終えた後であれば、何度でも高速に3次元復元が可能であることと、地面があれば遠さを損なわず再現できること、ラベル画像が3色の単純な画像なので、ユーザーが自作で用意することが比較的容易ということである。

人を検出して立たせることにより、3次元復元の可能性が確認できたので、今後、人検出の精度を向上させるとともに、木や、柵などの遮蔽物、他の動物などについて個別に対応することで復元精度が向上していくものと考えられる。実際、木などオブジェクトのラベルを用意するという提案が[11]でなされている。

本研究の根本的な問題として、地面領域が存在しない画像に対しては処理することが出来ないといった点が挙げられる。カメラが上下に傾きすぎている場合には、カメラの姿勢推定をより詳細にし、カメラ位置が高く地面が写っていない場合は、画像中のカメラ焦点の位置から奥行きに補正をかけ、その後、地面領域を仮定するという対応が考えられる。

地面を水平な1枚の平面としてしか再現できない問題については、ラベリングの時点で地面領域に傾きや段差を表すサブクラスを用意できれば、(4)式の結果に手を加えることで傾きや段差を表現できる。本論文の手法では、水平線は空と地面の間にあると定義されるが、もし水平線が地面領域と交わるならば、そこから地面の傾きを求めることが可能であると思われる。

今後、以上述べた点を基に、3次元復元の精度向上を目指す予定である。

文 献

- [1] 奥村健志, "大域的特徴として BoF を導入した CRF による一般物体認識" *MIRU* 2009. pp95-102
- [2] Ashutosh Saxena, Min Sun, Andrew Y. Ng, "Make3D: Learning 3-D Scene Structure from a Single Still Image "
- [3] Derek Hoiem, Alexei A. Efros, Martial Hebert "Recovering Surface Layout from an Image"
- [4] Jean-François Lalonde, Alexei A. Efros, and Srinivasa G. Narasimhan "Estimating Natural Illumination from a Single Outdoor Image" 2009 12th ICCV pp183-190
- [5] Derek Hoiem, Alexei A. Efros, Martial Hebert "Putting Objects in Perspective" 2008 IJCV
- [6] Jean-François Lalonde Derek Hoiem, Alexei A. Efros, Carsten Rother John Winn Antonio Criminisi "Photo Clip Art" *ACM SIGGRAPH* 2007,
- [7] Jiang Ze-tao , Zheng Bi-na , Wu Min, Chen zhong-xiang "A 3D Reconstruction Method Based on Images Dense Stereo Matching" 978-0-7695-3899-0/09 2009 IEEE
- [8] D. Hoiem, A. A. Efros, and M. Hebert. "Automatic photopop-up". In *SIGGRAPH*, 2005.
- [9] D. Hoiem, A. A. Efros, and M. Hebert. "Geometric context from a single image". In *ICCV*, 2005.
- [10] D. Hoiem, A. A. Efros, and M. Hebert. "Closing the loop on scene interpretation". *CVPR*, 2008.
- [11] Beyang Liu Stephen Gould Daphne Koller "Single Image Depth Estimation From Predicted Semantic Labels" In *CVPR*, 2010
- [12] Navneet Dalal, Bill Triggs and Cordelia Schmid. "Human detection using oriented histograms of flow and appearance" *European Conference on Computer Vision (ECCV)*, vol. II, pages 428-441, May 2006.