

建物が疎に並ぶ通りの多視点パノラマ画像生成

柳澤 惇^{†1} 手塚 大基^{†1} 櫻田 健^{†1}
岡谷 貴之^{†1} 出口 光一郎^{†1}

本稿では、市街地の通りのような長いシーン、特に通り沿いに建物がまばらにしか立っていないような場合を対象に、多視点パノラマ画像を生成する方法を述べる。過去の研究で知られているように、そのように奥行きの変化幅が大きいシーンを対象に、ひずみのないパノラマ画像を生成することは原理的に不可能である。この困難さを克服するため、提案方法は、通りに面する建物の表面および地面のみを鮮明に描画する。それ以外の物体はぼかして提示することで、それらの存在感は保ったまま幾何学的な歪みをなるべく目立たなくする。これらを実行するため、最初に対象とするシーンの3次元構造を SfM (Structure from Motion) によって推定し、シーンの各点が建物表面、地面とそれ以外のどれに属するかを、MRF (Markov Random Field) モデルに基づく最適化により同定し、これを元に、これら3種類の物体の画像をモザイクングすることでパノラマ画像を合成する。画像のぼけは、ライトフィールドフォトグラフィーで使われるのと同様の方法で生成する。2011年3月11日の東日本大震災に伴う津波被災地を対象に、提案方法を用いて生成したいいくつかのパノラマ画像を示す。

Creation of Multi-Viewpoint Panoramas for Sparse Streets with a Small Number of Buildings

JUN YANAGISAWA,^{†1} DAIKI TETSUKA,^{†1} KEN SAKURADA,^{†1}
TAKAYUKI OKATANI^{†1} and KOICHIRO DEGUCHI^{†1}

This paper presents a method for creating multi-viewpoint panoramas that is particularly targeted at streets with sparsely located buildings. As is known in the literature, it is impossible to create panoramas of such scenes having a wide range of depths in a distortion-free manner. To overcome this difficulty, our method renders sharp images only for the facades of buildings and the ground surface (e.g., vacant lands and sidewalks) along the target streets; it renders blurry images for other objects in the scene to make their geometric distortion less noticeable while maintaining their presence. To perform these, our method first estimates the three-dimensional structures of the target scenes using the results obtained by SfM (structure from motion), identifies to which category (i.e., the facade surface, the ground surface, or other objects) each scene point belongs based on MRF (Markov Random

Field) optimization, and creates panoramic images of the scene by mosaicing the images of the three categories. The blurry images of objects are generated by a similar technique to digital refocus of the light field photography. We present several panoramic images created by our method for streets in the tsunami-devasted areas in the north-eastern Japan coastline because of the Great East Japan Earthquake of March 11, 2011.

1. はじめに

本稿では、市街地のストリートのように長いシーンを対象に、その多視点パノラマ画像を生成する方法を述べる。提案方法は、特に通り沿いに建物がまばらにしか立っていないような場合を対象に、通りから見た建物の見えおよび空き地や歩道の状況を、単一画像上で詳細に提示したパノラマ画像を生成する。

この研究は、2011年3月11日の震災による津波被害を対象とした、われわれの一連の研究の一つである。岩手・宮城・福島各県沿岸部の市街地および住宅地において、多数の建造物が甚大な被害を受けた。それら一連の研究の目的は、これら被災状況とその後の同地域の復旧・復興の過程を可視化する方法を研究開発することにある。

われわれは東京大学池内研究室と合同チームを作り、被災後1か月後ごとに被災地に入り、それ以降車載全方位カメラによる被災状況およびその後の復旧過程の記録を行ってきた。開始後約一か月の撮影期間で、3県にまたがる被災地の撮影を完了し、約10テラバイトの映像アーカイブを取得した。画像の一部を図1に示す。同様の活動はほかのグループも行っている。例えばグーグルは、われわれの3か月程度後から同様の記録を開始し、その映像記録は現在、同社のサイト(ストリートビュー)で一般公開されている*1。

全方位画像はシーンへの没入感を生成できる一方で、単一視点の画像を与えているに過ぎない。グーグル社のストリートビューのように、より広範囲の空間を把握するには視点を変える必要があるが、視点を変化させるとその分閲覧時間がかかる。上記の津波被災では、一つの街全体が壊滅の状況にあることも多く、広範囲の空間の損害状況を一瞥して把握できるような画像表示方法が、復旧・復興にともなう都市計画や教育の観点から必要とされている。なお、航空写真は一つの選択肢ではあるが、地上視点の映像がもつ情報を与えるものではなく、代替とはならない。

^{†1} 東北大学
Tohoku University

*1 われわれはプライバシーへの配慮から現時点では一般公開は行っていない。



図 1 東北地方沿岸部の津波被災地を記録した画像の一部．1 視点からの円柱パノラマ画像として表示してある．

以上に基づき、本稿では、通り沿いのような長いシーンを、図 12 にあるような、一枚の横長のパノラマ画像として表示する方法を考える。このような、長いシーンのパノラマ画像生成の研究は長い歴史があり^{2),6),8)-13)}、われわれの方法はそれらの研究の成果の上に成り立つものであるが、対象とするシーンの性質が違う点で既存研究と異なる。われわれが対象とするシーンは、通り沿いに建物がまばらにしか存在せず、空き地やがれきが散乱したような通り沿いの情景である。上述の津波被災エリアでは、津波による直接被害で、さらにはその後復旧過程で取り壊されるなどにより、建物がまばらにしか残っていないところが多い。このことはパノラマ画像生成を本質的に難しくする要因となる。

本稿の構成は次の通りである。次節で、関連研究および長いシーンを対象としたパノラマ生成における本質的な制約を議論し、その制約に対するわれわれのアプローチについて述べる。3 節では、対象シーンの 3 次元構造のモデル化の方法および、SfM(structure from

motion) によって得られる復元結果を元にこれを推定する方法を述べる。4 節では、推定したシーンの構造をもとにパノラマ画像を生成する方法を述べる。5 節で実験結果を示し、6 節で本稿をまとめる。

2. 関連研究と提案方法のアプローチ

長いシーンのパノラマ画像には、奥行き異なる物体をひずみなく（正しいアスペクト比で）表示することができない本質的制約がある。手前の物体ほど狭く、遠いほど幅広くなる。このアスペクト比のひずみは、奥行きがほぼ一定のシーンでは目立たないが、奥行きの変化幅が大きなシーンでは顕著となる。

この問題を完全に解決するのは原理的に不可能なので、既存研究では、ひずみがなるべく目立たないパノラマ画像をいかにして作るかに主眼が置かれてきた。例えば、cross slit¹³⁾ に基づく半自動の方法が文献²⁾ にある。対象とするシーンの密な奥行きを陽に推定し、その上でアスペクト比のひずみを最小化するようなパノラマ画像を合成する方法が文献⁸⁾ にある。Microsoft 社で実用化されている Street Slide⁶⁾ を含め、これらの方法では、シーンのうち奥行きの変化が大きい部分は局所的に限定されることを前提としている。例えば、市街地の通りのシーンでは、通り沿いに建物が密集しており、奥行きが大きく変化するのは交差点などに限定されることが多い。

このような性質は、本研究が対象とするシーンに対しては成り立たない。通り沿いにまばらにしか建物が存在せず、これらの建物は遠方の物体（他の建物や山、空など）と混ざり合う。重要なのは、しばしばシーンの大部分がそのような遠方の物体で埋め尽くされることである。これは、ひずみの少ないパノラマ画像を生成するのが最も困難なシーンであると考えられる。

本稿では、この困難さを次のようなやり方で克服することを考える。

- 対象とする通り沿いの建物の表面と地面（空き地や歩道など）を選び、これらのみ詳細な画像を提示する。この選択は、これら 2 つが応用上重要であることによる。そのため、SfM で復元したカメラ運動軌跡と 3 次元点群を用いて、空間中の建物表面および地面の位置を推定する。
- これら 2 つ以外の物体は、画像上でぼかして表示する。その詳細が表示されないことで、その存在感は保ったままひずみを目立たなくできる。画像のボケはライトフィールドフォトグラフィー^{1),7)} で一般的なリフォーカスの技術と同様の方法によって生成される。

- カメラの運動軌跡を近似する連続曲線上に投影中心を有するプッシュブルーム方式によって画像を生成する。建物表面は一定の奥行きにあるのでそのテクスチャにはアスペクト比のひずみは生じない。地面は、奥行きが空間的に変化するが、地面は視線方向に平行で、視線方向のアスペクト比をもともと定義できないので、生成される画像は違和感を生じない。
 - パノラマ画像の各画素を、特定の奥行きにある建物面、地面およびそれ以外のどれに分類するかを認識し、その結果に基づいて画像を合成する。画素ごとの特定の奥行き面、地面、それ以外の3分類は、密なステレオ復元で一般的なMRF(Markov Random Field)モデルを用いた最適化手法により行う。具体的には、各画素のラベルを変数とし、視点移動にともなう視差の発生有無を、画像間類似度によってデータ項、隣接画像間のラベルの類似性を平滑化項とする最小化問題として定式化し、グラフカットによって解く。
- 以上の方法の背後にある狙いは、シーンの重要構造物を幾何学的に正確に提示することと、単一画像でシーンへの没入感を生成することの2つを両立させることである。

3. シーンのモデル化

対象となるシーンの3次元構造を、3つの曲面、すなわち(i)通りに面した建物表面を近似する曲面、(ii)地面を近似する曲面、および(iii)その他のシーン中のオブジェクトで表現する。最も単純な場合、(i)および(ii)の2つの曲面は平面としてモデル化できるが、通りに並ぶ建物の表面は、必ずしも単一平面で近似するのが適当でないことが多い。本稿では急激なカーブを持つ通りは除外して考えるが、緩やかなカーブを持つ通りは扱いたい。そこで、(i)および(ii)の曲面は、空間中を線分をスイープして得られる曲面(線織面)でモデル化する。

3.1 カメラの運動軌跡のモデル化

建物表面および地面を表すこれらの曲面は、SfMによって推定されたカメラの運動軌跡に基づいて決定する。SfMを行う画像は車載の全方位カメラで撮影した。画像の例を図1に示す。カメラは画像は通りに沿って約2メートル間隔で撮影する。撮影時には、システム上GPSやオドメトリなどの情報を利用しているが、SfMによる3次元復元の行程においては、それらの情報は用いていない。

カメラの運動軌跡は、画像が撮影されたカメラ姿勢の離散集合として与えられ、さらに自動車の偶発的な姿勢変化のせいで滑らかではない。そこでパラメトリックな曲線をこれにあてはめ、カメラ軌跡の連続かつ滑らかな近似を得る。後述のように、あてはめられた曲線は

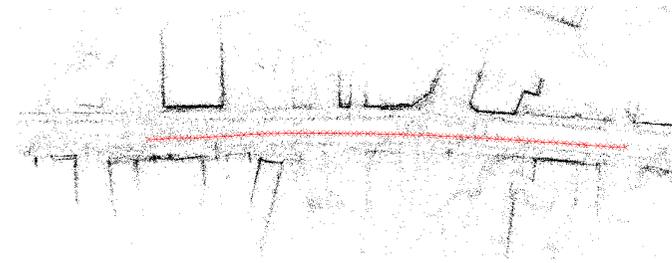


図2 SfM(Structure from Motion)によって数百枚の画像をもとに得た点群とカメラの運動軌跡(約70視点)の例。

生成するプッシュブルーム画像の投影中心の役割を果たす。

シーンに固定されたワールド座標系と、カメラに固定されカメラとともに動くカメラ座標系を考える。 \mathbf{X} と \mathbf{X}' をそれぞれ、空間における同一点に対するワールド座標とカメラ座標とする。これらは

$$\mathbf{X}' = \mathbf{T}\mathbf{X} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}. \quad (1)$$

によって関係づけられる。ただし \mathbf{R} は回転行列、 \mathbf{t} は並進ベクトルである。 $k=1,2,\dots$ を、各カメラ姿勢と画像を指すインデックスとし、空間の点のカメラ座標を \mathbf{X}'_k 、姿勢パラメータを \mathbf{T}_k あるいは $\mathbf{R}_k, \mathbf{t}_k$ のように表記する。

画像を K 枚の用いたとき、カメラの運動軌跡は $\{\mathbf{T}_1, \dots, \mathbf{T}_K\}$ のように与えられる。 k 番目のカメラの姿勢 \mathbf{T}_k は6次元ベクトル $\mathbf{c}_k = [\mathbf{t}_k^T, \omega_k^T]$ によって表される。ここで ω_k は回転 \mathbf{R}_k の回転軸・角度表現で、Rodriguesの公式で与えられるである。このカメラの運動軌跡を、次のような多項式を用いて近似する：

$$\mathbf{c}(s) = \sum_{l=0}^L \mathbf{a}_l s^l. \quad (2)$$

ただし $s \in [0:1]$ である。実験では $L=4$ とした。係数 \mathbf{a}_l は次のように最小二乗あてはめにより決める：

$$\min_{\mathbf{a}_0, \dots, \mathbf{a}_L} \sum_{k=1}^K \|\mathbf{c}(s_k) - \mathbf{c}_k\|^2. \quad (3)$$

具体的には、まず曲線のパラメータを $s_k = (k-1)/(m-1)$ とし、この最小二乗あてはめを行った後、曲線のパラメータが得られた曲線の空間における弧長に比例するように、 $s \leftarrow \tilde{s}(s)$ と取り直す。

3.2 建物表面および地面

建物表面を近似した曲面は次のように生成する。まず初めに、カメラ座標系においてパノラマを作成する方向 \mathbf{v}' を定める。例えば車の進行方向に対して右側のシーンのパノラマを作成したい場合、 \mathbf{v}' はカメラ（車）の進行方向に対して90度右にセットする。同じくカメラ座標系での天頂方向 \mathbf{z}' も定める。また、 \mathbf{v}' に対し直交、かつ \mathbf{v}' と \mathbf{z}' が張る平面上のベクトルを $\mathbf{n}' (\propto (\mathbf{I} - \mathbf{v}'\mathbf{v}'^T)\mathbf{z}')$ と書く。これらのベクトルはすべてカメラ座標系上で定められることに注意する。

図3(a)に示すように、建物表面の近似曲面を生成する線分を、 \mathbf{v}' にそって距離 d に位置し、 \mathbf{n}' に平行になるようにとる。 d はカメラから建物表面までの距離であり、SfMによって得た空間の点群を使って各 s に対して定める。これを行うため、空間の点群から、建物表面に属する（とみなされる）点を単純な閾値処理で取り出す。各カメラ姿勢 k について、 z 軸方向が \mathbf{v}' と、 y 軸方向が \mathbf{n}' とそれぞれ一致するような局所的な xyz 座標系をとり、SfMの点群を $\{(x_j, y_j, z_j)\}$ のように記述する。そして次の条件を満たす点を抽出する：

$$\begin{cases} |x_j| < \alpha \\ \beta_1 < y_j < \beta_2 \\ |z_j - d_0| < \gamma d_0 \end{cases} \quad (4)$$

ここで $\alpha, \beta_1, \beta_2, \gamma$ は閾値であり、 d_0 は手動で選んだ d_k のラフな見積もりで、全 k に対して一定とする。抽出された点の z 座標 z_j の平均値を、奥行き推定値 d_k とする。残りの閾値は次のように定める。 α は隣接カメラ間の距離に基づいて定め、 β_1, β_2 はカメラの地面からの高さ、最終的に生成するパノラマ画像の鉛直方向の視野角の大きさから定める。 γ は0.2とした。カメラの運動軌跡 $\mathbf{c}(s)$ の場合と同様に、多項式を d_1, \dots, d_K にあてはめ、奥行き連続表現 $d(s)$ を得る。

2つの関数 $\mathbf{c}(s)$ および $d(s)$ は、各 s ごとに一本の空間直線を与える。図3(a)に示すように、この空間直線上に範囲を定め、線分を取り出す。具体的には、線分の中点が $\mathbf{c}(s)$ から

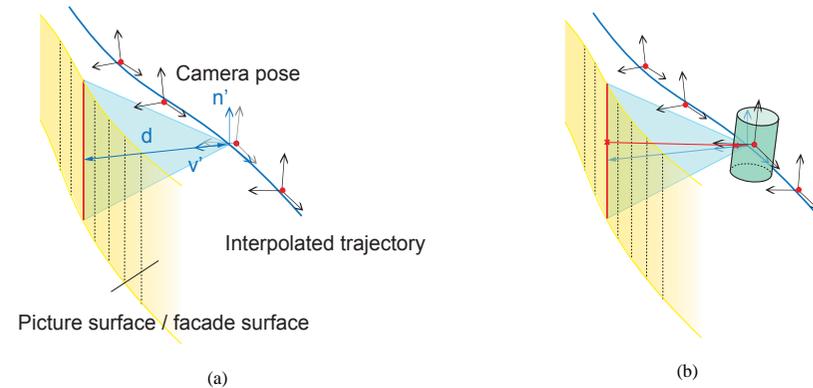


図3 カメラの運動軌跡とパノラマ画像の画像面の関係。(a) 補間されたカメラの運動軌跡と建物表面の近似曲面 (= パノラマ画像の画像面)。(b) パノラマ画像生成における画像面と撮影画像の関係。

出た \mathbf{v}' に平行な光線上に位置し、かつ線分の長さが $2d(s) \tan(\phi/2)$ (ϕ はパノラマ画像の垂直視野角) となるようにこの範囲は定める。こうして決まる線分上の点を、パラメータ $t \in [0:1]$ によって表し、 $t=0$ が線分の上端を、 $t=1$ が下端を表すように、また t は線分の長さに比例するように決める。図3(a)のように、この線分をカメラの運動軌跡 $\mathbf{c}(s)$ に沿って動かすことで、建物表面の曲面を生成する。この曲面は2つのパラメータ (s, t) によって表現できるから、 $\mathbf{f}(s, t)$ と表すことにする。

地面の近似曲面も同様に決める。今度は \mathbf{n}' に垂直で、カメラの直下、距離 $h(s)$ の位置にある線分を考える。 $d(s)$ の場合同様、 $h(s)$ は、まず SfM の点群から地面に属する点を Eq.(4) と同様の閾値処理によって選別し、これらの点を対象に各視点 k ごとに距離 h_k を計算し、最後に多項式を h_1, \dots, h_K にあてはめる。こうして得た線分を $\mathbf{c}(s)$ に沿って動かすことで、地面を近似する曲面 $\mathbf{g}(s, t)$ を生成する。地面の場合、パラメータ t は、 $\mathbf{g}(s, t)$ の点が、 $\mathbf{f}(s, t)$ の点と $\mathbf{c}(s)$ を結ぶ光線上にあるように決めておく。

4. パノラマ画像の合成

4.1 画像生成の幾何

前節の方法で生成した3次元空間のモデルを対象に、仮想的なプッシュブルームカメラによって画像を生成することで、目的のパノラマ画像を合成する。プッシュブルームカメラの投影中心は、カメラの運動軌跡 $\mathbf{c}(s)$ とする。そして、建物方面の近似曲面 $\mathbf{f}(s, t)$ をそのま

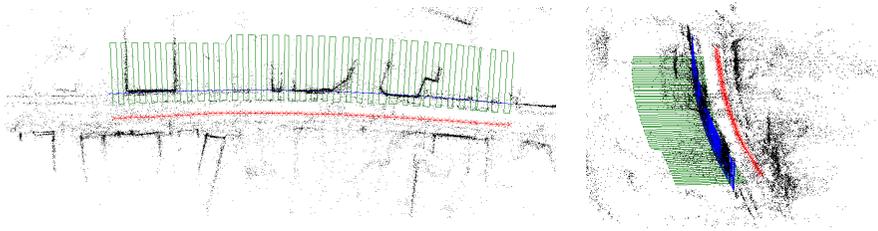


図 4 Fig.2 の 3 次元復元に対して推定した各曲面．青い線は建物表面の近似曲面を，赤い線は地面の近似曲面をそれぞれ表す．

まパノラマ画像の画像面とし，シーンの各点からの光がそこに投影されるとする．パノラマ画像の縦の列がこの曲面を生成する線分に相当する．パノラマ画像のサイズを $W \times H$ ピクセルとすると，画像の点 $(u, v) \in [0: W - 1] \times [0: H - 1]$ は，建物表面の近似曲面上の 1 点 (s, t) に， $s = u / (W - 1)$ および $t = v / (H - 1)$ によって変換されることとなる．パノラマ画像の幅 W は自由に決め，建物表面上のテクスチャのアスペクト比が 1:1 となるように H を定める．以上の画像生成の幾何は，建物表面の近似曲面以外のシーンの点が，画像面のどこに投影されるかを定める．

4.2 成分画像の合成

次に，シーンの各点のカラーをどうやって定めるかを述べる．実際のカメラ（撮影時のカメラの位置姿勢）は，あてはめた曲線 $c(s)$ の近くにまばらに分布するが，図 5(b) に示すように，指定したシーンの点に対し，各カメラ位置から見たその点の方向はただちに定まり，そこから，その点の（そのカメラに関する）カラーが決定できる．

SfM の復元では疎な点群しか存在せず，シーンの密な形状はわからないため，スパノラマ画像上の各点に対応する空間の点の位置はわからない．代わりに，（パノラマ画像に投影される）空間中の全点が建物方面の近似曲面上にあると仮定すると，各カメラ k のパノラマ画像を，図 reffig:facadeimg に示すように合成できる．同様に，地面についても図 6 に示すように合成できる．建物表面および地面について， k 番目のカメラを用いてこのように合成した画像のカメラ位置に最も近い点 (u, v) を，そのカメラ位置の投影位置とする．そして，画像の 1 点 (u, v) に対し，最も投影位置に近いカメラを $k = \tilde{k}(u, v)$ のように記述する．これを使って，パノラマ画像の各点のカラーを，その点に最も近い投影位置を持つカメラの合成画像の同じ点のカラーと一致させた次のような画像



図 5 各視点の画像を用いて合成した建物表面の（パノラマ画像面への）投影像．全 70 枚のうちの 3 枚．

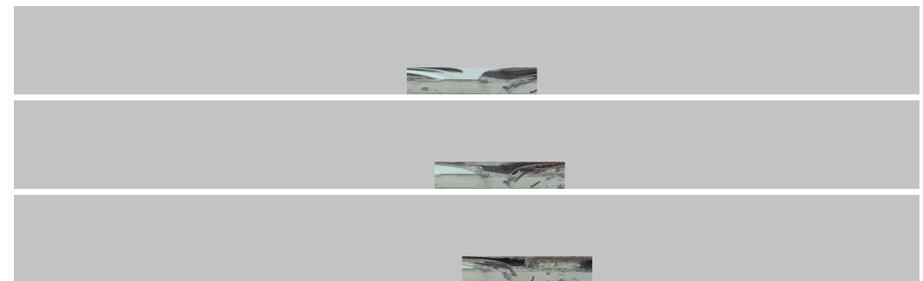


図 6 Fig.5 と同じ視点での地面の投影像．

$$I_f(u, v) = I_f^{\tilde{k}(u, v)}(u, v), \quad (5)$$

$$I_g(u, v) = I_g^{\tilde{k}(u, v)}(u, v), \quad (6)$$

を考える．図 7 に示すように，異なるカメラ画像から合成した建物表面および地面の投影像から，縦方向に長い短冊状の領域を切り出し，順番につなぎ合わせたようなものとなる．この画像のことを短冊画像と呼ぶことにする．図 7 から，建物表面の短冊画像では，建物表面のテクスチャだけが滑らかに張り合わさっていて，地面の短冊画像では地面のテクスチャだけが滑らかに張り合わされているのが分かる．

建物表面と地面のいずれにも近くないシーンの物体は，ぼかした画像によって表現する．画像のボケは，全視点 $k = 1, \dots, K$ にわたる建物表面の投影像の平均を計算することで生成する．



図7 建物表面の短冊画像(上)と地面の短冊画像(下)。



図8 建物表面の投影像の全視点にわたる平均画像。

$$I_{\bar{f}}(u, v) = \frac{1}{K} \sum_{k=1}^K I_f^{(k)}(u, v). \quad (7)$$

なお、各建物表面の投影像の非投影領域は平均の計算からは除外する。

図8に平均画像の例を示す。対象の奥行き(つまり建物の表面)に近いシーンの点のみがシャープなテクスチャを持ち、それ以外のシーンの点はボケたテクスチャしか持たないことが分かる。これは、異なる視点から同一シーンの複数の画像を撮影し、任意の奥行きにフォーカスを合わせた画像を再構成する、マルチカメラアレイを用いたデジタルリフォーカスやライトフィールドフォトグラフィーと呼ばれる技術と同じである。われわれの場合、カメラ(もしくは投影中心)は1次元状にしか分布しないので、画像は同じ1次元方向(つまり水平方向)にしかばげない。にもかかわらず得られるボケの効果は、われわれの目的には十分である。すなわち、対象とする奥行きから外れた位置にある物体が、パノラマ画像上で正しくないアスペクト比を持つことを目立たなくする。

このように合成される物体の画像上のボケは、そのボケの強さが物体と建物表面の距離に応じて自動的に調節されることに注意されたい。より遠方の物体のボケはより強くなり、その逆も真である。これは、今のボケの合成の仕組みが、画像のボケ生成の物理的過程を(荒くではあるが)再現していることによる。また、各建物表面画像の水平方向の視野角を変化させることで、全体的なボケの強さを制御できることにも注意する。視野角が大きいほど、画像のボケがより強くなる。

4.3 成分画像のモザイクキング

最終的なパノラマは上述の3成分、つまり建物表面と地面の短冊画像 $I_f(u, v)$ および $I_g(u, v)$ 、建物表面の平均画像 $I_{\bar{f}}(u, v)$ をモザイクキングすること等で合成する。パノラマ画像の各ピクセル i (画像座標 (u_i, v_i)) に対して、変数 l_i を導入し、 $l_i = 0$ はこの画像の点に対応するシーンの点が建物表面にあることを指し、 $l_i = 1$ は地面、 $l_i = 2$ はそれ以外の物体上にあるとする。そして、最終的なモザイク画像を次のように合成する：

$$I(u_i, v_i) = \begin{cases} I_f(u_i, v_i) & \text{if } l_i = 0, \\ I_g(u_i, v_i) & \text{if } l_i = 1, \\ I_{\bar{f}}(u_i, v_i) & \text{if } l_i = 2. \end{cases} \quad (8)$$

各ピクセル i のラベル l_i もまた、これらの画像から推定し定める。その推定は、ステレオマッチングで広く使われている MRF 最適化によって行う。

各視点での建物表面の画像(図5)に示されるように、ある点が建物表面に属する場合、その点は異なる視点の画像間でその位置を変えない。地面の画像でも(図6)同様である。したがって、各画素のラベルすなわち、どのカテゴリーのシーンの点に対応するかは、これらの画像上での移動の有無を調べることで推定できる。これは、画像間の濃淡の類似度を調べることで可能である。ある画素 (u_i, v_i) に最も近い画像面への投影位置を持つカメラを $\tilde{k}_i = \tilde{k}(u_i, v_i)$ と書く。この類似度の評価に SSD を用いることとし、 (u_i, v_i) での SSD の値を、基準となる \tilde{k}_i の画像と、その近傍の視点 $\tilde{k}_i + k$ での何枚かの画像 ($k = \dots, -2, 1, 2, \dots$) との間で計算する。

$$E_f(u_i, v_i) = \sum_{k \neq 0} \sum_{(u, v) \in \mathcal{W}(u_i, v_i)} |I_f^{(\tilde{k}_i)}(u, v) - I_f^{(\tilde{k}_i + k)}(u, v)|^2, \quad (9)$$

$$E_g(u_i, v_i) = \sum_{k \neq 0} \sum_{(u, v) \in \mathcal{W}(u_i, v_i)} |I_g^{(\tilde{k}_i)}(u, v) - I_g^{(\tilde{k}_i + k)}(u, v)|^2, \quad (10)$$

ここで $\mathcal{W}(u_i, v_i)$ は (u_i, v_i) を中心とする 5×5 ピクセルの正方領域である。

E_f は (u_i, v_i) に対応するシーンの点が建物表面上の点であるときに小さく、 E_g は地面の点であるときに小さくなる。それ以外の場合には一般に大きな値をとる。したがって各ピクセルに対するデータ項を



図9 グラフカットによる最適化の例．上：推定された各ピクセルのラベル．下：推定されたラベルをもとに合成したパノラマ画像．

$$E_{data}(u_i, v_i, l_i) = \begin{cases} E_f(u_i, v_i) & \text{if } l_i = 0 \\ E_g(u_i, v_i) & \text{if } l_i = 1 \\ E_a & \text{if } l_i = 2 \end{cases}, \quad (11)$$

と定める． E_a は定数で，当該画素が最初の2カテゴリに属さない場合のコストである．

MRF グラフの全コストは次のように定義できる：

$$E(l_1, \dots, l_N) = \sum_{i=1}^N E_{data}(u_i, v_i, l_i) + \beta \sum_{(i,j) \in \mathcal{E}} \delta(l_i, l_j). \quad (12)$$

ここで N は画像のピクセル数であり， \mathcal{E} は画像内で隣接するピクセルのペアの集合である． $\delta(l, l')$ は平滑化項で， $l \neq l'$ ならば1， $l = l'$ ならば0とする．このコストの最小化は，グラフカットのアルファ拡張を用いる．実験では，GCOptimization (バージョン 3.0)³⁾⁻⁵⁾ を使用した．

5. 実験結果

Fig.9 に，これまでと同じデータに対し前節で述べた最適化を適用した結果を示す．推定されたラベルの分布をみると，建物表面，地面，およびそれより遠方あるいは近くにあるその他の物体はすべて，かなり正確に認識されているのが分かる．

図10は，図7に示した短冊画像と合成したパノラマ画像の同じ部分を拡大表示したものである．同図に見られるように，短冊画像は，ターゲットとした建物表面から離れた奥行きにある建物がおかしな具合で表示されており，見る人に誤った印象を与える危険がある．一方でわれわれの方法によるパノラマ画像は，これらの物体をぼかして表示しており，そのような誤った印象を与えるのを回避できている．



図10 画像の一部を拡大したもの．左：短冊画像．右：提案方法により合成されたパノラマ画像．短冊画像は，遠方の建物について不正確な印象を与える．提案方法のパノラマ画像では，この建物をぼかすことでこれを回避している．



図11 短冊画像(左)と提案方法のパノラマ画像(右)の拡大図．別の例．

図11は，別の画像での同様の比較を示す．短冊画像では，遠方の建物が誤って表示されているだけでなく，この通りと交差する道路も誤って表示してしまっている．実際にはこの通りには1本の道路が交差しているだけだが，あたかも複数の道路が交差しているかのように表示されている．提案方法のパノラマ画像では，交差する道路の幾何を正しく伝えている．

最後に図12に，いくつかのシーンに対して合成したパノラマをいくつか示す．同図から，シーンの構造を幾何学的に正しく表示し，同時に1枚の画像で没入感を生成するという目標をほぼ達成できていることが分かる．

6. まとめ

本稿では，建物がまばらに点在する通りの多視点パノラマ画像を合成する方法を述べた．奥行き範囲が大きいシーンに対してパノラマを合成する困難さを緩和するために，提案方法では，建物表面と地面のみシャープな画像で提示し，それ以外の物体をぼかして提示する．これにより，シーンの重要な構造物を幾何学的に正確に再現しつつ，単一画像によって没入感を生成するという相反する目標を達成した．

参 考 文 献

- 1) T. Adelson and J. Y. A. Wang. Single lens stereo with a plenoptic camera. *IEEE PAMI*, Vol.14, No.2, pp. 99–106, 1992.
- 2) A. Agarwala, M. Agrawala, M. Cohen, D. Salesin, and R. Szeliski. Photographing long scenes with multi-viewpoint panoramas. *ACM Transactions on Graphics*, Vol.25, No.3, pp. 853–861, 2006.
- 3) Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE PAMI*, Vol.26, No.9, pp. 1124–1137, 2004.
- 4) Y. Boykov, O. Veksler, and R. Zabih. Efficient approximate energy minimization via graph cuts. *IEEE PAMI*, Vol.20, No.12, pp. 1222–1239, 2001.
- 5) V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE PAMI*, Vol.26, No.2, pp. 147–159, 2004.
- 6) Johannes Kopf, Billy Chen, Richard Szeliski, and Michael Cohen. Street slide: Browsing street level imagery. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2010)*, Vol.29, No.4, pp. 96:1 – 96:8, 2010.
- 7) R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. Technical report, Stanford University Computer Science, 2005.
- 8) A. Rav-Acha, G. Engel, and S. Peleg. Minimal aspect distortion (MAD) mosaicing of long scenes. *International Journal of Computer Vision*, Vol.78, No. 2-3, pp. 187–206, 2008.
- 9) A. Román, G. Garg, and M. Levoy. Interactive design of multi-perspective images for visualizing urban landscapes. *IEEE Visualization*, pp. 537–544, 2004.
- 10) A. Román and H. P. A. Lensch. Automatic multiperspective images. In *Eurographics symposium on rendering*, 2006.
- 11) S. M. Seitz and J. Kim. Multiperspective imaging. *IEEE Computer Graphics and Applications*, Vol.23, No.6, pp. 16–19, 2003.
- 12) J. Y. Zheng. Digital route panoramas. *IEEE Multimedia*, Vol.10, No.3, pp. 57–67, 2003.
- 13) A. Zomet, D. Feldman, S. Peleg, and D. Weinshall. Mosaicing new views: The closed-slits projection. *IEEE PAMI*, Vol.25, No.6, pp. 741–754, 2003.



(a)



(b)



(c)



(d)

図 12 奥行き不連続の少ないシーンの結果。ただし建物内部に続く地面領域が正確に再現できている。(b) 建物が 1 棟だけのストリート。建物を取り壊した後の空き地のテクスチャがよく再現されている。(c) 別の建物がまばらなストリートの例。なかほどにある 3 本の電柱とビル 1 棟は、対象とする奥行きからわずかに外れているため、存在がうかがえる程度にボカされた。(d) 屋上まで津波が到達した陸前高田市役所のパノラマ画像。