1P2-5in

多地点からの画像に基づくHAI用没入型仮想空間の構築

Constructing immersive virtual space for HAI with photos

森　槙悟　　大本　義正　　西田　豊明
Shingo Mori　Yoshimasa Ohmoto　Toyoaki Nishida

京都大学 情報学研究科
Graduate School of Informatics Kyoto University

This research presents system to construct immersive virtual spaces for HAI which added information for agents not to unnaturally behave by using photos. We need virtual space which looks like the real world to observe human agent interaction, because of expansion of research range. To make such space, we think that immersive space, photos and information of objects' position are needed. We use panorama images to show immersive space and depth map to describe objects' position. To make panorama images and depth map automatically from photos, photos' position and pose and a simple 3d model are need. Making good interpolated image need robust feature point and matching. To satisfy these requirements, this research use computer vision techniques, for example Structure from Motion and stereo method. Finally, this research succeeded in creating and implementing a system constructing virtual spaces for HAI in immersive environment.

1. 序論

バーチャルエージェントに関する研究は盛んに行われている。その中に HHI を仮想空間上での HAI を通して観察したいという欲求がある。例えば、観光案内を HAI で行えたとすると、海外などの遠い場所での案内を観察できるし、観測環境を整えやすいというメリットがある。

HAI を HHI の代わりに使おうとすると、HAI を行う環境はなるべく現実の空間に近い物でなければならない。現実の空間に近い仮想空間を実現するための重要な要素は、背景に写真を用いること、没入型環境で周囲を一目で把握できること、エージェントとオブジェクトの前後関係・位置関係が正しいことだと考えた。

本研究では観光案内というタスクを考慮に入れて写真は屋外の物を使用した。没入型環境はパノラマ画像を全方位ディスプレイに表示して解決し、位置関係は空間の 3 次元形状を復元し深度マップを作ることで解決した。また仮想空間でのユーザの移動を可能にするためにモーフィングで補間画像を作成した。

2. 関連研究

仮想空間を作成する主な方法としては Model Based Rendering と Image Based Rendering がある。Model Based Rendering は空間の 3 次元モデルを作成してしまう手法である。3 次元モデルを作ってしまうので任意視点の画像を作成するのが非常に楽というメリットがある。またどの点から見ても一貫した形状を保つことができる。Image Based Rendering は画像と画像の間をモーフィングにより補間し新しい視点の画像を作成する。この手法は木などの自然物に強く、得られる画像にホールが無いというメリットがある。しかしオクルージョンを表現するのが非常に難しいという欠点もある。

Furukawa ら [1] はマンハッタンワールド仮説を用いて 150 枚程度の写真から家の内部の空間・3D モデルを 1 日程度で復元している。しかし、屋外ではマンハッタンワールド仮説を用いることが難しいという問題がある。Gallup ら [2] は「現実世界は平面・非平面で構成されている」という仮定を用いて屋外の仮想空間を Model Based Rendering で復元している。しかし、この手法では木などの複雑な形状は復元することができず見た目の欠損が大きくなってしまう。Image Based Rendering を用いている伊吹 [3] の手法は撮影経路から 1,2m ほどしか離れることができないという問題点がある。

連絡先: 氏名:森　槙悟, 所属:京都大学情報学研究科, 住所:京都市左京区吉田本町 (工学部 10 号館 214 号室), 電話番号:075-753-5371, Fax 番号:075-753-4961, E-mail:mori@ii.ist.i.kyoto-u.ac.jp

図 1: 全方位ディスプレイ使用イメージ

3. HAI 用没入型仮想空間の構築

3.1 仮想空間の復元方法

本研究では没入型環境を実現するために全方位ディスプレイ (図 1) を用いて全方向に画面を表示することで解決することにした。このディスプレイは視点から近い位置に設置されており、またディスプレイ自体も大きいためドットが大きく見え、表示するための画像は高解像度・繊細なものでなければならない。Gallup ら [2] の手法では画像の質が足りず、Model Based Rendering の手法では満足な結果が得られそうにないことが分かった。実際の所、Google のストリートビュー程度の画質が必要で、Image Based Rendering を使用しなければ十分な画質を得られないことが判明した。

Image Based Rendering 自体には幾何情報は必須ではないのだが、エージェントを仮想空間上で動かすために 3 次元形状を復元する必要がある。Furukawa ら [1] や伊吹 [3] の手法の

図 2: 仮想空間として復元する広場

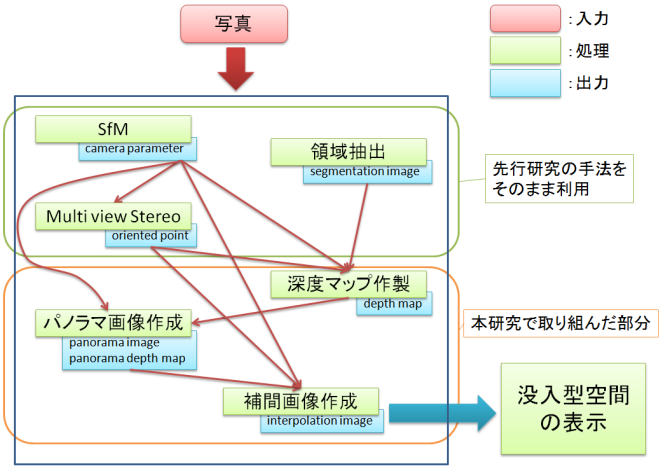

図 3: コンポーネント図

ように Structure from Motion とステレオ視を用い 3 次元形状と深度マップを作成すれば壁の推定を行うことができたり、オブジェクトとの前後関係を正しく保てると推測される。

本研究では大規模な空間の復元を行う前段階として図 2 のような 20m×20m 程度の広場の復元を試みることにした。この空間の復元を行うために、広場を 1-2m きざみのグリッド状におおよそに分割して、障害物の無いグリッドで撮影を行うことにした。各撮影地点では後で全方位ディスプレイに表示するためのパノラマ画像を作成できるように 18 枚ほどの写真で全周を撮影する。さらに撮影地点が近い所同士でモーフィングにより補間画像を作成して滑らかな移動が出きるようしなければならない。

3.2 システムの概略

3.1 節で述べたことを整理するとシステムの全体構造はコンポーネント図 (図 3) の様になる。各コンポーネントの簡単な解説を操作の順に述べると以下の様になる。

SfM Structure from Motion を用いてカメラパラメータの推定を行う。入力は写真のみである。出力は推定されたカメラの変換行列である。これにはオープンソースで公開されている Bundler[4] を用いた。

Multi view Stereo ステレオ視を用いて仮想空間の 3 次元形状を復元する。入力は写真とカメラの変換行列である。出力は 3 次元上の法線ベクトルと位置が分かっている点 (パッチ) である。これにはオープンソースで公開されている CMVS[5][6][7] を用いた。

領域抽出 深度マップの作成のために領域分割を行う。入力は写真のみである。これにはオープンソースで公開されている Felzenszwalb ら [8] の手法を用いた

深度マップ作成 現実世界は平面でできているという仮定を使用して写真の深度マップを作成する。入力は写真、カメラの変換行列、CMVS のパッチ、セグメンテーション画像である。出力は各写真の深度マップである。3.3 節で詳しく述べる。

パノラマ画像の作成 各カメラの位置関係を用いて写真をグループ化して全方位ディスプレイに表示するためのパノラマ画像を作成する。入力は写真、カメラの変換行列、深度マップである。出力はパノラマ画像とそれに対応する深度マップである。これにはフリーで扱える Microsoft ICE[9] を用いた。写真と深度マップ両方のパノラマ画像を作成した後に、向きの自動決定とずれの補正を行った。

補間画像作成 滑らかな移動に必要な補間画像を作成する。入力は写真、カメラの変換行列、CMVS のパッチ、パノラマ画像とそれに対応する深度マップである。出力は補間画像である。3.4 節で詳しく述べる。

没入型空間の表示 全方位ディスプレイに仮想空間を表示する。今回は作成したパノラマ画像、補間画像を連続で表示して移動の演出を行うのに止めた。

SfM、Multi view Stereo、領域分割とパノラマ画像の作成部分については先行研究の手法をそのまま使用した。本研究では主に深度マップを作る部分、パノラマ画像の補正部分、補間画像の作成の部分と各コンポーネントを統合して扱うことに力をいれた。

3.3 深度マップの作成

CMVS によって作成された 3 次元形状はホールや外れ値があるのでそれに対応しなければならない。本研究では現実の世界は平面で構成されるという仮定を用いて深度マップを作成する。植え込みや自転車群などは平面で構成されていないが HAI を行うのに十分な精度の深度マップが得られると予測される。

この仮定を用いるために写真を 60 個ほどの領域にセグメンテーションし、同じ領域に含まれるなら同じ平面に属するとみなすことにした。平面はその領域に含まれるパッチの位置と法線ベクトルの平均から求めた。領域があるパッチを含むかどうかは、各パッチを写真上に投影して NCC(正規化相互相関) を用いて十分相関があるかどうかで判定した。

3.4 補間画像の作成

まず、どのパノラマ画像の間で補間画像を作るかを決定しなければならない。今回は単純に撮影地点同士が近い場合に補間画像を作成することにした。撮影地点が近いかどうかの判定は Structure from Motion により得られたカメラの位置を用いれば簡単に行える。

本研究では補間画像の作成手法として 2 つの方法を試みた。1 つはパッチをパノラマ画像上に再投影して、その点を対応点とする手法。もう 1 つは深度マップを用いて視点が移動した際に各ピクセルがどの位置に移動するかを計算する手法である。後者よりも前者のほうがうまく動いたので前者について述

べる。　パッチをパノラマ画像上に再投影する際、パノラマ画像の位置と向きから計算するのではなく、一旦元の写真にパッチを投影してからパノラマ画像上に再投影を行う。写真上に投影されたパッチはさらにパノラマ画像上に再投影される。パノラマ画像上での写真の位置は Microsoft ICE のプロジェクトファイルを見れば分かるのでそれを使って計算する。

パノラマ画像を作った際に写真が重なっている部分で、どちらの画像を使用したかや、どの程度画像を歪めたかでパッチの投影先が少しずれることがある。ここでは写真の横方向の中央部分 2/3 にパッチが含まれていて、パッチの中央部分の色と投影先のピクセルの色が充分近く、NCC を計算し十分相関があるパッチだけを正しく投影できたパッチとみなした。　パッチが投影できたので 2 枚の画像間で対応点を求めなければならない。これは単純に 2 枚の画像間で同じパッチが投影された場所を対応点とした。このままでは対応点が多すぎるのでパノラマ画像上で 30×30 ピクセルに 2 つ以上の対応点が含まれないようにした。パノラマ画像間での移動方向は分かっているので、対応点が逆方向に進む場合も外れ値として除去した。　この様にして、得られた対応点を用いて補間画像の作成を行う。対応点を用いた補間手法で主流なのは、ドロネー三角形分割を用いた手法であるが、今回は画像上のエッジに対応点が乗ってない場合も多いので、ドロネー三角形分割を用いた手法を使うと不自然な補間になると考えられる。今回は画像の各ピクセルはもっとも近い対応点と同じ動きをするという仮定を用いて補間を行った。対応点自体の移動はパノラマ画像同士の距離の比率に応じて決定した。

この手法ではホールが出来てしまうので、ホールになった部分はもっとも近いピクセルと同じ色にした。補間を 2 枚のパノラマ画像両方に適用し、距離の比率でアルファブレンドを行えば最終的な補間画像が得られる。補間画像はパノラマ画像同士の間で 9 枚作成した。補間画像の深度マップも同様にして作成できる。

4.　評価

本研究の手法を用いて 3 箇所の仮想空間の復元を行った。撮影枚数、撮影地点数、各処理にかかった時間は表 1 のようになった。撮影にかかった時間はおよそ 1 時間ずつである。人の少ない早朝に撮影を行った。写真の解像度は 1024×768 であり、作成したパノラマ画像の解像度は 5376×768 である。計算に使用した PC の CPU は Core i7 975、メモリは 6GB である。

処理時間の大部分は Structure from Motion によるカメラのパラメータ推定である。広場の復元には 4.3 日程度かかっており、撮影枚数の二乗ほどのオーダで時間がかかっている。これは Agarwal ら [10] の手法を用いることにより大幅な改善が見込まれる。補間画像は 1 枚あたりに 5 秒程度で作成している。リアルタイムでの補間画像作成を行うには GPU を用いた高速な処理を行う必要がある。

4.1　画像・補間画像の評価

評価で生成された補完画像の典型例を図 4 中央に示す。この補間画像は 2 つの元画像のちょうど中間の位置の画像である。全方位ディスプレイに表示して見たところ撮影地点のパノラマ画像は、非常に高精度で注意深く見なければ画像上での不自然な点は見当たらない。撮影位置からある程度の距離がある場所やオクルージョンが起きない時は図 4 のベンチの様におおよそ正しい補間画像が生成できる。逆の状況、つまり撮影位置から近い場所やオクルージョンが起きると、1 つのオブジェクトが極端に分断されたり、ブレたりしていて違和感を与えてしまっている場合があった。図 5 はその典型例である。

表 1: 撮影枚数と処理にかかった時間。空間の広さは概算。

	建物前	プロムナード	広場
撮影枚数	427 枚	944 枚	1248 枚
撮影地点数	19 箇所	50 箇所	61 箇所
空間の広さ	10m×7m	45m×8m	20m×20m
パッチ総数	247325 点	629672 点	871859 点
SfM	375 分	1314 分	6194 分
CMVS	8 分	34 分	41 分
領域分割	7 分	16 分	21 分
深度マップ	19 分	40 分	47 分
パノラマ	64 分	170 分	210 分
補間画像	92 分	146 分	187 分

図 4: 元画像 (上下) と補間画像 (中央)

オブジェクトが分断される他の原因としては、一つのオブジェクトの中で対応点が複数取られてそれぞれが少しずれている場合と、オブジェクトの端で別のオブジェクトの対応点と同じ移動をしている場合もあった。特に前景と背景の境目付近でオブジェクトが分断されることが多かった。ブレに関しては視点に近い部分で多かった。これは視点が近いとパッチをパノラマ画像上に再投影した時に誤差が大きくなってしまい、正しく投影できない場合があることと、そもそも片方のパノラマ画像には映っていなくて対応点が取れないことが原因として考えられる。視点の移動による物体の見え方の変化・歪みに対応しきれておらず、また全方位ディスプレイで見ると、1 ピクセルあたりの面積が通常のディスプレイよりも大きいせいで汚い画像になってしまうという問題点もあった。これは各ピクセルが最も近い対応点と同じ動きをするという仮定を用いたことが原因であると考えられる。オクルージョンと視点に近すぎるパッ

図 5: 補間の失敗例

図 6: パノラマ化した元画像と深度マップ

チの問題を除くとマッチングは非常に綺麗に取れていて、各パノラマ画像の組に対して 500-1500 個ほどのマッチングが取れていた。

4.2 深度マップの評価

作成した深度マップをパノラマ化すると図 6 のようになる。写真に対して垂直な面はおおよそうまく復元できていたが、写真に対して平行な面、奥行きを間違って推定してしまうこともあった。平面が推定できなかった領域は主に遠景、木、地面である。遠景はインタラクションには関係ないと思われるので復元できなくても問題ない。木はセグメンテーションの段階で多くの領域に分割されるので、周りの領域が復元できていれば最も近いピクセルと同じ奥行きにしても問題ないと推測される。地面はエージェントが歩く際に使用すると思われるので別の手法で推定するのが望ましい。

4.3 総合評価

全体を評価すると、少なくともパノラマ画像の撮影地点で静止状態ならば十分な画質と深度マップが得られている。補間画像は遠くの物に対してはある程度うまく機能すると考えられるが、近くの物体や多層レイヤーがあるとブレや分断が生じてしまう。したがって、エージェントは自由に動き回り、ユーザは普段静止していてたまに動く程度のタスクならば行えると推測される。これは、目標のタスクを行うことができる空間を構築するシステムの開発・実装に成功したと言える。補間画像の改善、深度マップの高精度化、オブジェクトの配置をある程度推定すればより自然な HAI 用の環境を整えられると予測される。

5. 結論

本研究では、エージェントとインタラクションを行うための現実に近い仮想空間の構築を行うシステムの作成を目標とした。仮想空間を表示するために没入型の全方位ディスプレイを使用し、表示する画像には屋外の実際の写真を用いた。エージェントとの自然なインタラクションをするために写真の奥行きを復元し、移動も可能な空間の構築を行った。その過程において、現実世界は平面で構成されるという仮定を用いた深度マップの作成手法、対応点の検出手法と対応点を用いたパノラマ画像のモーフィング手法を提案した。その結果、現実に近い空間で HAI を行うのに必要な環境を整えることに成功した。

今後の課題としては、実際にこの空間上でエージェント動かした場合にどの程度自然に見えるかを調べる必要がある。本研究では最大 20m×20m の仮想空間しか構築していないが、もっと広大な仮想空間の構築をする必要が出てくる。そのような広大な環境に対応できる撮影手法を確立する必要がある。補間画像の質の向上を行いもっと違和感の無い仮想空間を構築しなければならない。最終的には、HAI の観察を行いそれを現実世界にフィードバックすることが望まれる。

参考文献

[1] Y. Furukawa, B. Curless, S.M. Seitz, and R. Szeliski. Reconstructing building interiors from images. In *Computer Vision, 2009 IEEE 12th International Conference on*, pp. 80–87. IEEE, 2010.

[2] D. Gallup, J.M. Frahm, and M. Pollefeys. Piecewise planar and non-planar stereo for urban scene reconstruction. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 1418–1425. IEEE, 2010.

[3] 伊吹拓也, 佐藤智和, 松下康之, 横矢直和. 視点位置に依存して変形する三次元メッシュモデルを利用した自由視点画像生成における違和感の低減. 2009.

[4] N. Snavely, S.M. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3D. In *ACM SIGGRAPH 2006 Papers*, pp. 835–846. ACM, 2006.

[5] Y. Furukawa, B. Curless, S.M. Seitz, and R. Szeliski. Towards internet-scale multi-view stereo. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 1434–1441. IEEE, 2010.

[6] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. *IEEE transactions on pattern analysis and machine intelligence*, 2009.

[7] Yasutaka Furukawa, Brian Curless, Steven M. Seitz, and Richard Szeliski. Clustering views for multi-view stereo. `http://grail.cs.washington.edu/software/cmvs`.

[8] P.F. Felzenszwalb and D.P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, Vol. 59, No. 2, pp. 167–181, 2004.

[9] MicrosoftCorporation. Microsoft image composite editor. `http://research.microsoft.com/en-us/um/redmond/groups/ivm/ice.html`.

[10] S. Agarwal, N. Snavely, I. Simon, S.M. Seitz, and R. Szeliski. Building rome in a day. In *Computer Vision, 2009 IEEE 12th International Conference on*, pp. 72–79. IEEE, 2010.