

ビデオシーケンスを利用した VR 環境マッピング用 テクスチャ及び空間の生成

吉田 幸代^{†1} 高田 雅美^{†1} 城 和 貴^{†1}

リアルな建物のモデルの作成はサイバシティアプリケーションの実装において、最も重要なタスクの 1 つである。本稿では見え方を重視し、且つユーザの位置を任意に変更できる臨場感ある屋内空間モデル作りについて考える。具体的には、ユーザの位置によって表示するテクスチャを動的に変更する手法を提案する。既存の見え方を重視した手法では、全方位カメラで撮影した写真をモザイクングし、1 枚の大きなパノラマ画像を用意し、半球にマッピングすることでサイバースペースを作成している。この既存の手法は、大抵街等の建物の外観を対象としており、広い空間作りに適している。廊下といった、屋内の狭い空間作りには不向きである。そこで本稿ではホモグラフィや SIFT を使用し、臨場感を維持しつつ、特に廊下といった狭い空間作りの手法を考えた。ユーザが空間内で移動したとき、ユーザの位置に合致したテクスチャに切り替える。臨場感を維持し、自由に空間内を移動できるようにもする。

Virtual View Generation in Hallway of Cybercity Buildings from Video Sequences

SACHIYO YOSHIDA,^{†1} MASAMI TAKATA ^{†1}
and JOE KADUKI^{†1}

Appearance of Three-dimensional (3D) building model is one of the most important components in a cyber city implementation and application. In this paper, we propose an approach for generating more freely and higher quality appearance of cyber city by using digital Video sequences than existing ones. This technique is updated depending on positions of the virtual view, and maps texture. In existing techniques, a large panoramic image is given by mosaicing all-directional camera. since the object of the techniques is large space such as city scenes, the techniques are unsuitable for small space such as indoor corridor. Therefore, the proposed a system aims to cyber city, which create automatically, for small space using homography, Scale Invariant Feature Transform (SIFT) and Video sequences. When users change their location in VR space, the system change textures newly. The system keeps realistic

sensation and free from location's limits.

1. はじめに

サイバースペースは、Google マップストリートビュー・医療における外科手術練習用アプリケーションなど様々な分野で利用されている。サイバースペースの必要構成要素は、体験可能な仮想空間の構築・5 感のいくつかに働き掛け得られる没入感・対象者の位置や動作に対する感覚へのフィードバック・ユーザが世界に働き掛けることができる対話性の 4 つである。これら構成要素の 1 つである没入感は、サイバースペースの臨場感の質を左右する最も大切な構成要素であり、中でも視覚的な感覚に働き掛けは最重要項目である。

従来のサイバースペースの作成方法として、モデルベースアプローチ¹⁾ とアピランスベースアプローチ²⁾ がある。モデルベースアプローチは、現実の建物に正確な形状のモデルを幾何学的に作成し、そのモデルにマテリアルや色を指定することで作成されている。より臨場感を出すため、実際に建物を撮影し、その写真をテクスチャとして 3D モデルにマッピングし作成される。ただし、実際の写真を使用する場合、テクスチャはサイバースペース内のどの位置から見ても一様であるため、ユーザの見る位置によっては環境光などと矛盾する可能性がある。そのため、違和感の原因になってしまい、臨場感を再現するには限界がある。一方、アピランスベースアプローチは、建物などの 3D モデルを一切使用せず見え方を重視した手法である。この手法は全方位カメラによって撮影した写真を一枚の 360 °パノラマ画像にモザイクングし、半球の内側にマッピングするというものである。全方位カメラは、屋内のような複雑な狭い空間の撮影には不向きである。また、サイバースペース内でユーザが移動できる場所は、実際にカメラ撮影した位置に限られる。そのため、自由にサイバースペース内を移動することができない。

そこで本稿では、狭い屋内空間に適した新しい手法のサイバースペース作りについて提案する。提案する手法は、単純な直方体に近似される廊下のような空間を対象とし、自由視点移動にあわせて動的に貼り付けるテクスチャを切り替えるものである。この際用いられるテクスチャは、全方位カメラのような高価なカメラを用いるのではなく、安価な一般ユーザ向

^{†1} 奈良女子大学 大学院 人間文化研究科
Graduate School of Humanities and Sciences, Nara Women's University

けカメラを用いる．これにより，任意のユーザが任意の場所をサイバースペース内に再現することを可能とする．

2章では既存の作成方法について説明する．3章では本稿で提案するシステムの手法について詳しく説明する．4章では提案手法を使用し，実際にサイバースペースを作成する．作成したサイバースペースは見る方向による違和感，ユーザの自由移動，一般のユーザにも扱うことができる容易なシステムであるかを確認する．

2. VR空間作成における既存の手法

2.1節ではモデルベースアプローチの作成方法・利点・問題点について説明する．2.2節ではGoogleマップストリートビューに代表されるようなアピランススペースアプローチについて述べる．

2.1 モデルベースアプローチ

モデルベースアプローチでは，最初に，3Dオブジェクトモデルを作成する．このモデルに対して，環境光とマテリアルが写真撮影されたテクスチャを設定し，サイバースペース内に配置する．この際，臨場感を出すためには，テクスチャを用いる方が好ましい．

モデルベースアプローチで作成したのサイバースペースの長所はユーザが自由に空間内を動き回り，視点を変更できることである．これはサイバースペースが3次元幾何に則って作成されているからである．一方で，問題点としてテクスチャ作成時の実際の環境光に強く影響を受けてしまうという点がある．固定されたテクスチャでは視点や位置を変更したとき，陰影のつき方が矛盾し，それが違和感になってしまい臨場感を損ねてしまう．モデルベースアプローチで臨場感を出すには限界がある．

2.2 アピランススペースアプローチ

アピランススペースアプローチの作成手法について説明する．アピランススペースアプローチでは実際の同一空間をある1点より満遍なく撮影し，それら撮影した複数の画像（タイル画像）をモザイクングし，1枚の大きなパノラマ画像に加工するように，半球モデルの内側にタイル画像をマッピングすることで実現している．タイル画像の撮影にはそれ専用のカメラである全方位カメラが使用されることが多い．視覚的にアピランススペースアプローチはモデルベースアプローチと比べリアリティ・臨場感があるのが特徴である．しかし，このコンテンツは撮影したカメラと同じ地点（位置）から，360°周囲を見渡すことはできるが，自由に空間的な移動ができない．アピランススペースアプローチは広い屋外空間の作成を前提としており，屋内の廊下といった狭い空間には不向きである．

3. システムの設計

3.1節では本稿の提案手法の概要について述べる．3.2節では入力データについて説明し，3.3節ではトラッキング，3.4節ではセグメンテーション，3.5節では位置情報の算出等の本提案手法で必要とされる技術について説明する．3.6節では本システムでのテクスチャの作成方法について，3.7節では動的なテクスチャの切り替え方法について，3.8節ではユーザによる初期値入力を支援するGUIについて説明する．

3.1 概要

本節では開発するシステムについての目的・処理の流れ・システムを利用することで得ることのできる効果について説明する．

2章で述べた既存技術は一長一短である．モデルベースアプローチは幾何学的に正確でありユーザはサイバースペース内を自由に移動できるが，デジタル的なマテリアルや，ある方向からしか撮影されていない写真画像を用いたテクスチャを使用しなければならないため，臨場感に欠ける．アピランススペースアプローチは，臨場感はあるが，空間的な自由な移動ができない．また屋内といった狭い空間を作成することは難しい．そこで，本稿では，これら2つのアプローチの長所を兼ね備えた新たなサイバースペースの作成手法を提案する．

提案手法では，形状モデルを使用し，動的にテクスチャを切り替える．具体的には，撮影しておいたビデオシーケンスよりテクスチャを同一平面に対し複数作成し，ユーザの指定する位置によって動的に適切なテクスチャを3次元モデルマッピングする．幾何学的に正確な3次元モデルを使用することによって，空間内の移動における自由度を確保できる．このモデルに一樣なテクスチャを張るのではなく，先にテクスチャデータベースを作成しておき，サイバースペース内でのユーザの位置に応じて適切なテクスチャを随時読み込みマッピングする．これにより，視覚的に矛盾がでる可能性を最低限に抑えることができる．つまり，モデルベースアプローチの自由度とアピランススペースの臨場感を兼ね備えた新しいサイバースペースを作り出すことが可能となる．

図1は，提案手法の流れ図である．この手法では，テクスチャデータベースの作成とサイバースペース内でユーザが移動した場合の描画のステージに分けてサイバースペースを開発する．データ・初期値入力においては，3.2節で述べる条件でビデオ撮影し，カメラレンズによる歪み補正を施し用意した複数の画像を読み込ませる．テクスチャを切り出す上で必要となる初期値もこの時点で入力する．セグメンテーションにおいては，グラフカットを使用し画像の領域を分割する．これは読み込んだデータを，特徴点の追跡の精度を上げるため

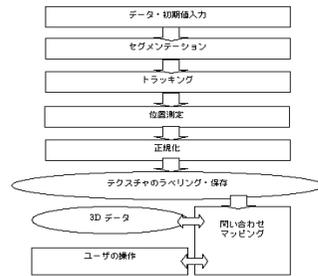


図 1 システムの流れ

に、廊下を構成する平面ごとに分類するためである。位置測定においては、ビデオシーケンスより位置情報を算出する。これによって、現実空間の位置とサイバースペース内での位置を関連付けることが可能となる。テクスチャのラベリング、保存においては、テクスチャの作成及びラベリングを行い、提案システムで必要となるテクスチャのデータベースを作成する。ユーザの操作においては、サイバースペース内でのユーザ操作・移動について設定する。このために必要となるのは、テクスチャの読み込みおよび3次元モデルへの貼り付けにおいて、テクスチャデータベースに問い合わせ、動的にテクスチャを変更するフィードバックを返すための機能である。この機能を実現することによって、動的にテクスチャが変更されるため、3次元モデルを使用している環境光に矛盾がなく臨場感を維持できる。また空間内を自由に移動できる。

3.2 入力データ

入力データセットについて説明する。

本稿では手持ち家庭用ビデオカメラで撮影した動画を入力データとする。廊下を撮影しデータを用意する際、各空間を往復して撮影する。これは全方位カメラと違い、一度に360度を撮影できないからである。

提案手法では、直方体に近い空間を対象とする。そのため、サイバースペース化できるのは出っ張りの少ない空間となる。極端に明度差のある空間についても本稿では対象外とする。極端に明度差のある空間では光がハレーションをおこし、感度の低いカメラでの撮影が困難となる。望ましい環境光条件は空間全体が均一な光量で満遍なく照らされていることである。

撮影したビデオシーケンスよりフレームを取り出し、カメラレンズによる歪みを補正す

る。これは特徴点追跡による移動量測定への影響をなくすためである。システムを動かすたびに歪み補正などの作業をするのは時間がかかるので、歪み補正をかけたフレームは画像データとして、サイズ720×480、Bmp(Microsoft Windows Bitmap Image)形式で保存しこれを入力データとする。

3.3 トラッキング

テクスチャはフレームより矩形に画像を切り出し、正規化することで作成できる。単純な直方体の内部に張るテクスチャを用意するためには、床・壁2面・天井・床・前方の合計5枚のテクスチャをフレームより切り出すこととなり、その境界にあたる8点指定する必要がある。この8点を全てのフレームに手動で与えることは実用的ではない。そこで最初のフレームのみ手動で点を指定し残りのフレームに対しては、自動的に点座標を求める手法を考える。この時与える最初の点座標を初期値と呼ぶこととする。

各フレームで面の境界が交差する8点を自動で定するために、初期値を、特徴点追跡を使って追跡する。特徴点の検出・追跡技術には、LK(Lucas & Kanade アルゴリズム)³⁾またはScale-Invariant Feature Transform(SIFT)⁴⁾などがある。LKとは動画像において、前の画像のある点が、次の画像のどこへ動いたのかを追跡するための理論とアルゴリズムである。コーナーを特徴点とし、その検出・追跡を行う。一方SIFTとは特徴点の検出と特徴量の記述を行うアルゴリズムである。検出した特徴点に対して、画像の回転・スケール変化・照明変化等に頑健な特徴量を記述するため、イメージモザイク等の画像のマッチングや物体認識・検出に用いられている。本研究対象である屋内空間は一様な壁が多く、コーナー情報が少ない。よって、本研究では特徴量による検出もできるSIFTを使用する。

SIFTにより検出・追跡した特徴点よりフレーム間のホモグラフィ行列を求める。ホモグラフィとは平面を2つの地点から撮影した場合にその画像間に成立する射影関係のことであり、式1のような関係式を満たす行列式である。

$$\begin{pmatrix} x' \\ y' \\ w' \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x \\ y \\ w \end{pmatrix} \quad (1)$$

(x, y, w) , (x', y', w') はそれぞれ変換前・変換後の座標である。また、 $h_{11} \sim h_{33}$ は画像の平面内の四頂点の座標から決まる定数である。

$$x'(h_{31}x + h_{32}y + h_{33}) = h_{11}x + h_{12}y + h_{13} \quad (2)$$

$$y'(h_{31}x + h_{32}y + h_{33}) = h_{21}x + h_{22}y + h_{23} \quad (3)$$

式 (2), 式 (3) より 4 組以上の対応点が分かれば, ホモグラフィ行列が一意に決定する.

本研究では 5 点以上の特徴点を使用し最小二乗法を用いてより正確なホモグラフィ行列を計算させる. ホモグラフィ行列を毎フレームごとに求めるだけでは, 微小な誤差が蓄積する可能性がある. そこで, よりロバストで正確な座標を求めるために毎フレームごとに求めたホモグラフィ行列と一定間隔ごとに求めたホモグラフィの 2 つを用意し, その 2 つの移動量を比べることとする. 以上により求めた特徴点の対応をもとに, ホモグラフィ行列を使用して最終的な移動量を決定する.

3.4 セグメンテーション

本節ではセグメンテーション, 領域分割について説明する.

ホモグラフィ行列は同一平面の射像関係あらず行列式である. 本システムでは 5 面のテクスチャを用意するためホモグラフィ行列は 5 個必要となる. ホモグラフィは同一平面上の特徴点より移動量を計算することができる. もし複数の平面上にある特徴点を使用し, ホモグラフィ行列を求めると, 適切な結果を得られない. 例えば, 床と天井は全く違う平面である. 床についてのホモグラフィ行列を求めるとき, 天井の特徴点を含めて計算をすると, その分ホモグラフィ行列の精度が落ちてしまう. 本システムではホモグラフィの精度を保つために, ビデオシーケンスより取り出した廊下の画像を, 床・壁・天井・などの平面に部類する. セグメンテーションによって平面を分類し, 単一な面の特徴点だけを取り出し, さらに RANSAC (RANDOM Sample Consensus)⁵⁾ にかけることでその平面上の窓といった微小な出っ張りについても例外処理を施し, 万全なものとする.

画像領域分割の手法の 1 つにグラフカットによるセグメンテーションがある. グラフカットは, ラベルに対するエネルギーを最小化することを問題として解く手法であり, これは Boykov らにより Interactive Graph Cuts として提案されている⁶⁾. ユーザーが seed と呼ばれる切り抜きたい画像のラベルを与え, 画像と併せてグラフを作成する. そのグラフに minimum/maximum flow algorithm を用いることで, エネルギー関数の最小化を行う.

図 2 は元画像であり, 図 3 は図 2 を実際にグラフカットによってセグメンテーションを施し各々の壁を切り出した画像である. 初めに seed と呼ばれるラベル物体と背景の 2 種類を最初のフレームのみ手動で指定する必要がある. seed はユーザが任意に図 3 のようにオブジェクトである部分, そうでない部分に適当に線を引くことで設定できる. この seed も初期値として設定する. 以降のフレームについてはビデオセグメンテーションを使用すれば自動的に分類することが可能である. 図 4 はグラフカットによって作成したマスク画像の例である.



図 2 オリジナル画像

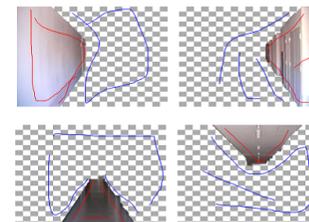


図 3 グラフカットを使用したセグメンテーションの例



図 4 作成するマスク画像の例

3.5 位置の計測

作成したテクスチャをテクスチャデータとして一旦保存し, 必要に応じて任意に呼び出し読み込めるようにする. そのためにはテクスチャに位置情報を自動的にラベリングし, 撮影している場所の特定ができるようにする必要がある. 本稿では SLAM (Simultaneous Localization And Mapping)⁷⁾ を使用することで, 位置情報を得る. SLAM とはコンピュータビジョンにおけるロボットの自己位置と地図の推定の技術である. SLAM を使用することで, 撮影したビデオシーケンスより, 撮影位置と地図情報を関連付けることができる. テクスチャにこの位置情報をラベリングし, ラベリングされた位置情報を参照することで, 随時サイバースペース内のユーザの位置に応じたテクスチャを読み込むことができる.

3.6 テクスチャの作成

3次元モデルの描画に OpenGL (Open Graphics Library) を用いる場合, 切り出した画像はテクスチャとしてふさわしい正方形に正規化する必要がある. OpenGL の特性として,



図 5 GUI の画面

この正方形のピクセル値は 2 のべき乗でなければならない。切り出した矩形のコーナー 4 点を正方形に正規化の際、ホモグラフィ行列を使用する。各コーナーの座標と正規化後のコーナーの座標を射像関係ととらえ、ホモグラフィ行列を求める。このホモグラフィ行列による射影変換を、切り出した画像に施し、正規化する。正規化し、できた画像はテクスチャとして保存する。保存するテクスチャの情報として、面の属性情報、3D 空間上の位置、往路・復路の判断の項目をファイル名とし保存する。

3.7 テクスチャのマッピングとユーザ操作

ユーザのサイバースペース内での現在位置と、テクスチャのデータベースのラベルに記載されている位置情報をリンクさせ、動的にマッピングする。なお本研究では、UI としてキーボードを用いているがこれ以外の UI を用いても構わない。

サイバースペース内でのユーザの位置が更新されたとき、すなわちキーボードから入力があった時、まず 3D データの表示位置の更新し、位置情報を用いてテクスチャデータベースに問い合わせを行う。保存されたテクスチャのラベルに記載されている情報を検索し、位置・進んでいる方向に応じた 5 つのテクスチャを読み込む。読み込んだテクスチャを 3D モデルに貼り付ける。最後に空間内を再描画し、ユーザの位置に見合った空間を表示させる。

3.8 開発補助ツールとしての GUI

3.3 節で述べたように、テクスチャを作成する上で、画像よりテクスチャに見合った範囲の画素を切り出す必要がある。そこで、初期値の入力を支援するため、firstframe での点を指定する工程を簡略化し、ある程度のガイドラインを示すための GUI を開発する。ユーザはこの 8 点を初期値として指定するものとする。このために必要となってくる機能として考えられるのは、切り抜く壁を矩形または平面の境界を直線として認識する、マウスなどの UI で簡単に選択できる機能である。そのために画像内の直線を検出するためにハフ変換を使用する。ハフ変換は、画像から直線や円を検出する技法として知られている。通常の直交座標上の画像を、直線検出の場合極座標の 2 次元空間に、円検出の場合 3 次元の空間等に変換し、そこで最も頻度の高い位置を求める。それを逆変換して、直線や円を検出する。

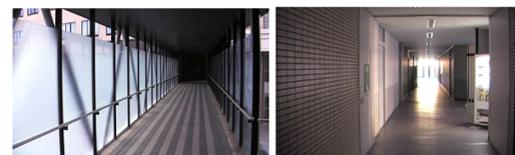


図 6 入力画像



図 7 作成した空間 (1)

一般的に、ハフ変換を用いる場合、非常に多くの直線が検出される。そこで、本稿では選択する直線の数を減らすために、確率的ハフ変換 (Probabilistic Hough Transform) を使用する。

図 5 は図 2 に確率的ハフ変換を施し、直線を検出し、4 本の直線を選択した状態の画面である。ユーザはまず切り出したい矩形の直線 4 本をマウスクリックで選ぶ。選んだ直線の両端に描画される 8 点はその直線上のみを動かすことができる。この点をマウスドラッグによってスライドさせ、その矩形のコーナー座標を指定する。このとき選択した点は緑色に描画される。

4. 提案手法を用いたサイバースペースの構築例

提案手法を用いて廊下のサイバースペースを作成する。ビデオカメラは victor の GZ-M505-B を使用する。開発環境は OS に windows XP サービスパック 3、コンパイラには visualstudio2005 を使用する。開発言語は C++ で、Intel 社がオープンソースで提供する画像処理ライブラリである OpenCV (Open Source Computer Vision Library) と OpenGL を利用する。OpenCV は画像処理に必要な画像フォーマットの解析・読み込み・表示・フィルタリングといった処理をするライブラリを提供している。

1 組のテクスチャを作成するための処理時間は RANSAC におけるモデル作成の繰り返し回数に依存する。本稿では 1000 回の繰り返しを行っており、1 組のテクスチャを作成するのに必要な処理時間は平均で 3.1458 秒かかる。



図 8 作成した空間 (2)

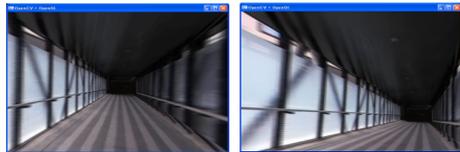


図 9 作成した空間 (3)

図 2, 図 6 は今回作成したシステムの入力データの例である。図 7, 図 8, 図 9 はそれぞれ今回作成したサイバースペースの画像であり, OpenGL で描画した直方体に本稿の提案手法によって作成したテクスチャを読み込み貼り付け, 元の現実空間を仮想的に再現している。

図 7 は異なった位置から見た画像であり, 図 7 右は図 7 左よりも奥にある空間である。テクスチャは図 2, 図 6 のような入力画像より各平面に切り抜かれ正規化することで作成される。正規化する前の切り抜いた画像は台形の形をしており, 長い底から短い方の底にかけて画像解像度としての情報は減ることになる。この情報量の少ない部分はサイバースペース内でも, 奥の方の部分に当たるため, 情報量が少なくても視覚的な違和感につながらないことがわかる。また, ユーザが奥に移動する際, 新たに適切なテクスチャに切り替えるため, 問題にならない。

図 8, 図 9 は同一空間内を違う視点からみた図である。テクスチャを作成するに当たり, サイバースペース内の位置とビデオを撮影した場所, すなわち現実空間との位置を関連付けている。そのため, 実際の空間と同じ視点の画像よりテクスチャを作成することができる。つまり, 動的にテクスチャを切り替えることによって, 臨場感を損なうことなく自由に位置や視点を変えることができるサイバースペースが作成できているものと考えられる。

5. まとめ

本稿では既存の手法であるモデルベースアプローチとアピランスペースアプローチの

長所を取り入れた, 新しいサイバースペースの作成手法について提案した。

初期値の入力簡略化のための GUI を開発した。直線の検出には一般的なハフ変換を用いる場合, 非常に多くの直線が検出されてしまうので, 本稿では確率的ハフ変換を使用し検出を行った。この GUI を使用することにより, 一般のユーザにも容易に初期値を入力することができ, サイバースペースを作成することができる。

テクスチャデータベースおよび GUI を用いることにより, 3 次元形状モデルを使用しユーザの位置によって動的にテクスチャを切り替えることで臨場感を維持し, 3 次元形状モデルを使用することで, 空間内の自由度を維持することが可能である。更に提案手法を用いることによって, アピランスペースアプローチでは対応できない廊下といった屋内空間をサイバースペースに再現することが可能である。そして, テクスチャデータを作成するために用いたカメラは, 一般的に市販されているものであるため, 任意のユーザが利用することが可能な手法である。

参考文献

- 1) Tsai, F., Chen, C.-H., Liu, J.-K. and Hsiao, K.-H.: Texture Generation and Mapping Using Video Sequences for 3D Building Models, *Springer Berlin Heidelberg*, pp.429–438 (2006).
- 2) Coorg, S. and Teller, S.: spherical mosaics with quaternions and dense correlation, *International Journal of Computer Vision*, Vol.37, pp.259–273 (2000).
- 3) Harris, C., Stephens, M. and Manor, R.: A combined corner and edge detector, *Proceedings of The Fourth Alvey Vision Conference*, pp.147–151 (1988).
- 4) 藤吉 弘巨: Gradient ベースの特徴点追跡-SIFT と HOG-, 電子情報通信学会技術研究報告 *CVIM 160*, pp.211–224 (2007).
- 5) Martin A. Fischler and Robert C. Bolles.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Commun. ACM*, Vol.24, number 6, pp.381–395 (1981).
- 6) Yuri Y. Boykov and Marie-Pierre Jolly.: interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images, *Eighth IEEE International Conference*, Vol.1, pp.105–112 (2001)
- 7) Howie Choset and K. Nagatani: Topological simultaneous localization and mapping (SLAM): toward exact localization without explicit localization *IEEE Transactions on Robotics and Automation*, Vol.17, pp.125–137 (2001)