

単眼カメラの動画像を用いたリアルタイムで緻密な3次元再構成技術

Real-Time Dense 3D Reconstruction Technology Using Video Images of Monocular Camera

関 晃仁 オリバー ウッドフォード

■ SEKI Akihito ■ Oliver WOODFORD

物体や周囲環境の3次元(3D)形状情報は、3Dコンテンツの作成をはじめ、自動車の運転支援装置、プラントの建設や保守、製品検査などに利用可能である。デジタルカメラなどの安価な手段で3D形状情報を取得できれば、コンシューマー向けから社会インフラ向けまで幅広く応用できる可能性があり、期待が高まっている。

東芝は、手持ちの単眼カメラの動画像からリアルタイムで3D形状を復元する技術を開発した。手持ちの単眼カメラでは、時々刻々と変化するカメラの位置や姿勢の推定と、画像間の対応関係を正確に求めることが必要である。今回、動画像を用いることで正確に3D形状情報を取得できるようにし、CPUとGPU(Graphics Processing Unit)を並列化して計算を行うことで、全体としてリアルタイムで3D形状の復元が可能になった。

Three-dimensional (3D) shape information of objects and their surroundings is used in various applications, including the creation of 3D contents, driving support systems, plant construction and maintenance, product inspection, and so on. Expectations are rising for the realization of a 3D reconstruction technology using inexpensive means such as a digital camera, in view of the strong possibility of applying it to a broad range of markets from the consumer market to the social infrastructure market.

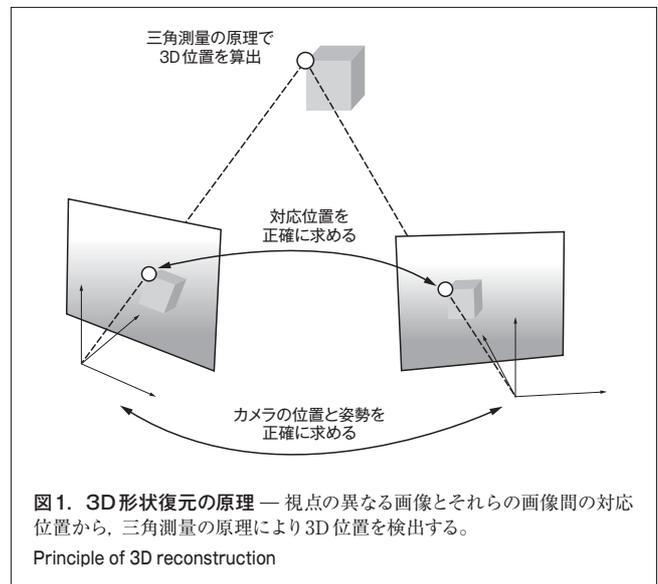
Toshiba has developed a real-time dense 3D reconstruction technology using a handheld monocular camera. In such types of cameras, it is crucial to achieve the precise estimation of camera positions/poses and point correspondences between images. This technology makes it possible to acquire 3D shape in real time by processing a sequence of motion videos through the simultaneous operation of both a central processing unit (CPU) and a graphics processing unit (GPU).

1 まえがき

プラントの建設や、保守、解体のほか、製品検査、ロボットの自律移動、自動車の運転支援装置などでは、周囲や現場の状況を把握するため、緻密で高速な3D再構成技術が必要である。また、近年の3Dテレビや3Dパソコン(PC)など、3D表示できる家電製品の登場により、3Dコンテンツはより身近なものとなってきた。更に普及するためには、3Dコンテンツを手軽に作成できる手段が必要になってきている。3D再構成技術は、カメラなどで撮影した静止画像や動画像から被写体の3D形状を復元する技術であり、デジタルカメラや携帯端末のカメラといった安価な手段で3D形状情報を取得できるため、コンシューマー向けから社会インフラ向けまで幅広く応用できる。また、近年のコンピュータの高速化やGPUを用いた並列計算により、高速な計算手段を安価で手に入れることができるようになってきている。

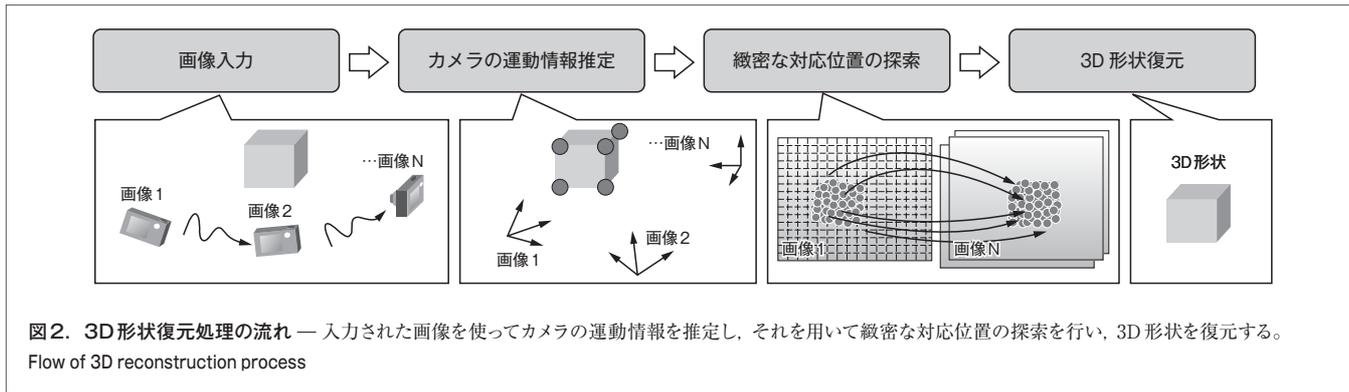
東芝は、3D形状の復元を少ない計算コストで行える方法を用い、かつそれぞれの処理に応じてCPUとGPUを使い分けることで、単眼カメラで撮影した動画像からリアルタイムで緻密な3D形状を復元する手法を開発した。

ここでは、単眼カメラを利用した3D再構成技術とその特長について述べる。



2 3D再構成技術の概要

カメラを使って物体の3D形状を計測するには、カメラの位置と姿勢(以下、カメラの運動情報と呼ぶ)が正確に求められた視点の異なる複数枚の画像を用い、同じものがどこに映っているか対応位置を正確に求める必要がある。これらの情報が



わかると、三角測量の原理によって3D位置を求めることができる(図1)。

単眼カメラでは、カメラの運動情報は時刻ごとに変わるため、それぞれの時刻ごとに運動情報を求めなければならない。また、物体の正確な3D形状情報を得るには、物体上の対応位置も緻密に求めなくてはならない。3D形状復元処理の流れを図2に示す。最初に、時刻の異なる複数の画像から各時刻におけるカメラの運動情報を推定する。続いて、この位置と姿勢を用いて膨大な数の対応位置を求めることで、三角測量の原理により、3D形状を緻密に復元する。

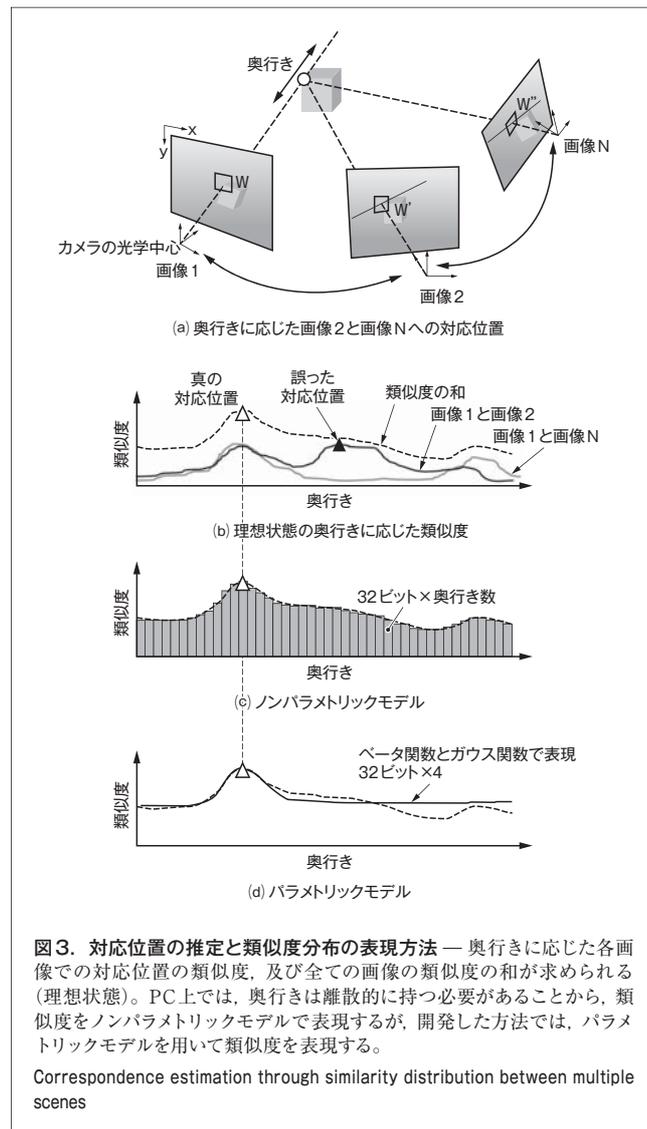
2.1 カメラの運動情報推定

3D形状の復元を周囲の環境に左右されずに行うために、マーカなど特殊な目印を必要としない方式とした。今回開発した方法では、画像が入力されると、画像からマーカの代わりとなる特徴点を自動的に抽出し、異なる時刻に撮影された画像間で追跡を行う。追跡された特徴点の位置を用いて、幾何学的な拘束条件により特徴点の3D位置とカメラの運動情報を求める。更に、オクルージョン(注1)の判定やバンドル調整(注2)により、安定して精度良く推定できるようにした。これらの処理は複雑であるため、CPU上で並列計算を行う。

2.2 緻密な3D形状の復元

求めたカメラの運動情報を用いて3D形状を緻密に推定する。そのためには、画像間で対応する位置を求め、その周囲の画像の類似度を求める(図3)。画像間の対応位置を求めるため、図3(a)に示すように、まず画像1の注目点Wと画像1のカメラの光学中心を通る直線上で奥行きを仮定する。次に、仮定した奥行きを使い、画像2に投影される点W'を求める。続いて、注目点Wと点W'の周囲のパターンの類似度を求める。パターンが似ているほど類似度が高い。異なる奥行きを仮定して、同様に類似度を求め、もっとも類似している奥行きを対応位置とする(図3(b)の“画像1と画像2”)。

画像1と画像2のように、画像が2枚だけであると、画像に



含まれるノイズや対象のパターンによって対応位置をまちがえる可能性が高いことが知られている。複数の画像が利用できる場合には、画像間で求められた類似度を加算(図3(b)の“画像1と画像2”及び“画像1と画像N”の加算により求めた“類似度の和”)することで、安定して対応位置を求められる⁽¹⁾。

(注1) 手前にある物体が、その背後にある物体を隠して見えなくすること。
(注2) 複数の画像で共通して観測される特徴点の位置を用いて、カメラの運動情報と特徴点の3D位置を精度良く求める手法。

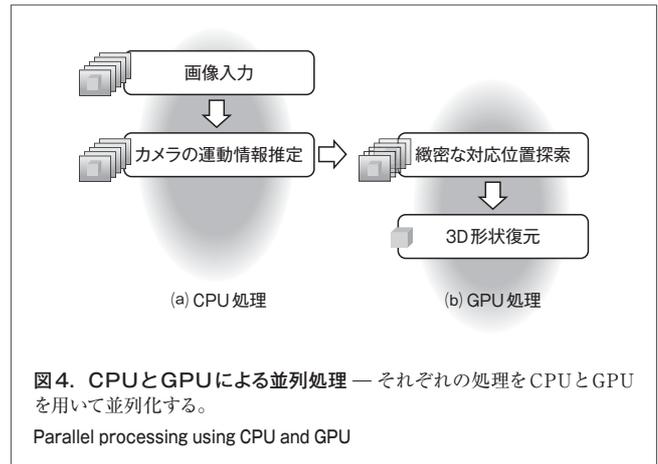
ここで、理想状態では奥行きを連続値としているが、実際の計算では、奥行きを離散化する必要がある(図3(c))。そして、離散化した各奥行きに対して類似度を保持しなくてはならない。離散化する間隔を広げるとPCのメモリ量と類似度の算出に掛かる計算コストを抑えることができるが、3D形状の推定精度が悪化するため、離散化する間隔を大きく広げることにはできない。また、3D形状を緻密に求めるには、画像中で注目点を多数設定する必要があるが、メモリ量と類似度算出に掛かる計算コストは注目点数に比例して増加してしまう。

これらの問題に対処して必要なメモリ量を削減するため、当社はパラメトリックモデル⁽²⁾を開発した。従来使われているノンパラメトリックモデルでは、奥行きに対する類似度を奥行きごとに持つため、メモリ量は奥行きの離散化の間隔に応じて増大していた。これに対しパラメトリックモデルでは、類似度をベータ関数とガウス関数の係数パラメータで表現することによって、必要なメモリ量を削減する。具体的には、各奥行きの類似度を32ビットの浮動小数点型で保持している場合に必要なメモリ量は、ノンパラメトリックモデルでは[(奥行き数) × 32ビット × (注目点数)]であるのに対し、パラメトリックモデルでは[関数パラメータ数 (= 4) × 32ビット × (注目点数)]となる(図3(d))。

パラメトリックモデルにおける対応位置は、ガウス関数の平均値として求めることができ、ガウス関数の分散が大きいほど対応位置の曖昧性が高いことを意味している。また、分散が大きい場所では、対応位置の候補から外れていると考えることもできるため、候補外の位置に対して類似度を計算する必要がなくなる。これにより、類似度計算に掛かる計算コストを削減できる。関数の係数パラメータは、画像が入力されるごとに、分散が一定値以内に収まっている奥行きの範囲で類似度を計算し、更新していく。この計算は、注目点ごとに並列計算でき、かつ処理が単純なため、GPU上で並列化して処理できる。

ノンパラメトリックモデルで対応位置を決定するには、類似度分布から最大となる類似度を探索する。実際には、類似度の最大値だけでは対応位置の誤りも含まれているため、最大値周りの類似度の急激さを評価したり、第2の最大値を評価したりして決定するが、計算コストがかさむ。パラメトリックモデルでは、類似度の最大値はガウス関数の平均値として求まるため、新たに探索処理は不要である。また、最大値周りの急激さは分散値そのものである。このようにパラメトリックモデルでは、対応位置を決定する際の処理も簡便に行うことができる。

複数の画像を用いる場合には、対応位置を決定するための画像枚数を判断しなくてはならない。今回の実装では、類似度分布がある一定の急激さを持った時点で対応位置を決定するため、対応位置を決定しやすい注目点は少ない画像枚数でよく、逆に、平たんな位置では、より多くの画像枚数を使って



安定に対応位置を決定できる。

2.3 CPUとGPUによる3D形状復元処理の分担

CPUとGPUによる3D形状復元処理の概要を図4に示す。画像入力とカメラの運動情報推定は、処理に掛かる時間に対して処理内容が複雑で、GPUでの並列処理に向かないため、CPU上で並列処理を行う。

カメラの運動情報を用いた緻密な対応位置探索は、注目点ごとに並列化して計算でき、処理内容もGPUに向いているため、GPU上で並列処理をする。前述のように、3D形状を求めるには、視点の異なる画像を用いる必要があるため、3D形状の復元に十分な動きがない画像に対しては、緻密な対応位置探索の必要はない。つまり、CPUでは入力画像を全て処理するが、GPUでは入力画像のうちの一部を処理すればよい。したがって、ある時刻にGPUに画像が入力されても、次の時刻に入力されるとはかぎらず、GPUによる処理を待っていると、CPUに待ち時間が発生してしまう。CPUによる待ち時間をなくすために、CPUとGPUの処理も並列化して実行する。

3 実験結果

GPUを搭載したノートPCを用いてデモンストレーションシステムを作成した(図5)。カメラには、解像度VGA(640×480画素)のUSB(Universal Serial Bus)カメラを用いた。またPCは、CPUがIntel® Core™(注3) i7 2630Q、メモリが20GB、及びGPUがNVIDIA® Quadro®(注4) 2000であり、C++とCUDA™(注5)を使って実装した。このシステムは、画像の入力から3D形状復元までの計算を37msで行い、毎秒58万点の対応点探索を行う。

(注3) Intel, Intel Coreは、米国又はその他の国におけるIntel Corporationの登録商標。

(注4) NVIDIA, Quadroは、米国又はその他の国におけるNVIDIA Corporationの登録商標又は商標。

(注5) NVIDIA Corporationが提供するGPU向けのC言語の統合開発環境。CUDAは、米国及びその他の国におけるNVIDIA Corporationの商標又は登録商標。

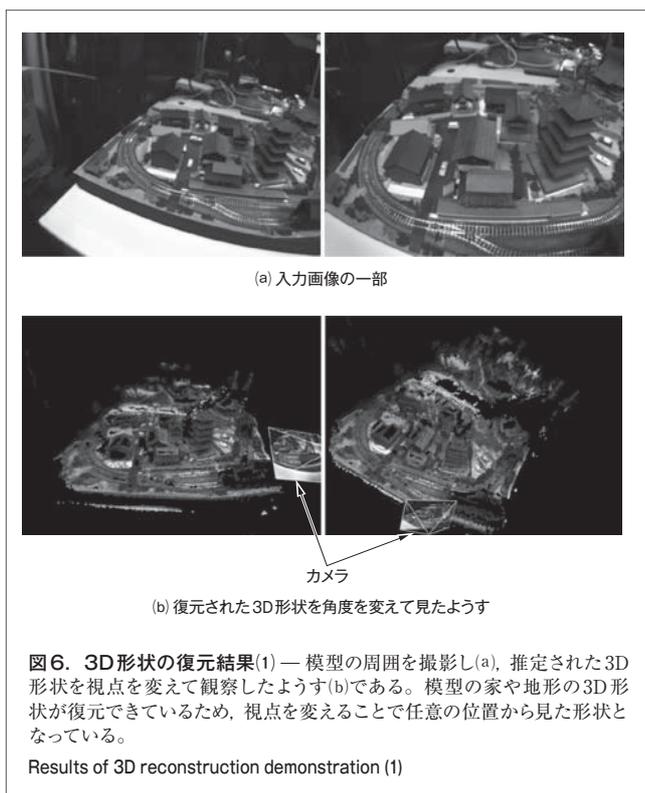
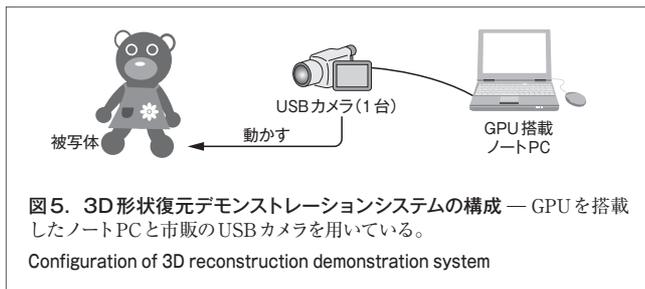


図6は、模型の左側から右側に向かってカメラを動かして撮影したシーンの画像の一部(a)と、推定された3D形状復元結果を角度を変えて観察したようす(b)を示している。建物などの3D形状が正しく復元されていることがわかる。また(b)には、推定されたカメラの運動情報を利用して、カメラの撮影場所とその撮影画像を書き込んでいる。

図7は、机上のノートPCと電話機の画像を復元した結果である。テクスチャのある部分は形状が正しく復元されている。

4 あとがき

手持ちの単眼カメラで撮影した動画像から、リアルタイムで緻密な3D形状を復元できる手法について述べた。

撮影環境にマーカなどの特殊なマークを必要としないため、様々な環境下で使用可能である。カメラの運動情報推定はCPUにより、また、運動情報から得られる拘束条件を使った



緻密な対応点探索はGPUによりそれぞれ並列処理を行うことで、リアルタイムな3D形状の復元を実現した。実験では、これらの処理を実装したノートPCを用いて、1フレーム当たり37 msでリアルタイムに処理できることを確認した。

文献

- (1) Okutomi, M.; Kanade, T. A multiple-baseline stereo. IEEE Trans. on PAMI. 15, 4, 1993, p.353 - 363.
- (2) Vogiatzis, G.; Hernandez, C. Video-based real-time multi-view stereo. Image and Vision Computing. 29, 7, 2011, p.434 - 441.



関 晃仁 SEKI Akihito, D.Eng.

研究開発センター インタラクティブメディアラボラトリー研究主務, 博士(工学)。3次元形状復元技術の研究・開発に従事。情報処理学会会員。
Interactive Media Lab.



オリバー ウッドフォード Oliver WOODFORD, D.Eng. 東芝欧州研究所 ケンブリッジ研究所 画像認識グループ, 博士(工学)。3次元形状復元技術の研究・開発に従事。
Toshiba Research Europe Ltd.