SceneScan の FPGA リアルタイムステレオビジョン

Real-time Stereo Vision FPGAs with SceneScan

Nerian Vision社　CEO，Konstantin Schauwecker

概 要

本稿では、最大100フレーム／秒での処理、または最大3.4メガピクセルの画像処理を、わずか8Wの電力消費で行うフレキシブルなFPGAステレオ計測システムの実装手法を提示する。本手法では、既存の単純な手法よりも高精度なマッチング結果が得られるセミグローバルマッチング（SGM）アルゴリズムをベースとし、それを改変したアルゴリズムを使用する。そして、コストボリュームと視差マップを入力とする一連の後処理を行うことで、SGMによるステレオマッチング結果をさらに大幅に改善する。本手法に基づいて、ステレオ計測のための二つのシステム「SceneScan」と「SceneScan Pro」を開発した。これらは、現在のステレオ計測市場の成熟度を考慮して開発されたもので、現在Nerian Vision社(独)から提供されている。

［キーワード］ ステレオ計測、デプスセンシング、FPGA

1 はじめに

ステレオ画像計測は、コンピュータビジョンの分野の中では最も研究が進んでいる技術の一つであり、その研究が始まった1970年代以降、著しい科学的進歩を遂げてきた。密なデプスセンシングのための他の技術には、ToF（time-of-flight）カメラや構造化光（structured light）カメラなどがあるが、これらの技術と比較して、ステレオ計測にはパッシブな計測法であるという利点がある。これは、環境光以外のアクティブな光源を計測に必要としないことを意味する。

アクティブシステムによる計測は、明るい環境光の下にある屋外の状況には向かない。この状況では、アクティブ光源は識別可能な高コントラストなパターンをもはや提供できない。また、特別に強力な光源を必要とするロングレンジの計測も課題である。一方で、ステレオ計測は屋外やロングレンジに対して有望な技術であり、他の計測技術が利用できない環境で、デプスセンシングを利用した様々なアプリケーションを可能にする。

整備された屋内とは違う環境で活動する自律移動ロボットにとって、様々な光源環境での計測ロバスト性の欠如は重大な問題となる。高速移動する自動車の安全走行にはロングレンジの計測も可能でなければならない。ゆえに、移動ロボットこそリアルタイムステレオのようなロバストなセンサソリューションが必要といえる。

ロバストなデプスセンシングの必要性があるにもかかわらず、ステレオ計測が広まらない主な理由は計算コストの大きさにある。この10年の間に多くのステレオマッチングアルゴリズムが提案されてきたが、マッチング精度の良いものは、現代のハードウェアを利用したとしても依然として計算コストが非常に大きい。

汎用的なCPUでは、近年提案されたステレオマッチング手法が要求する膨大な計算をフレームレートで処理することは難しい。もちろん、例えば文献[1]でデモンストレーションされたように、ブロックマッチングのような単純なアルゴリズムであれば可能である。しかし、KITTIのステレオベンチマーク[2]でわかるように、そのような単純な処理は、マッチング精度やノイズ耐性の面で近年の手法よりも著しく劣る。より洗練されたステレオマッチング手法の最適な実装に成功した例[3,4]もあるが、トータルの計算コストは依然として大きく、使用可能な画像解像度と実現可能なフレームレートは比較的低いレベルに留まっている。

ステレオマッチング処理をスピードアップする方法として、大規模並列計算が可能なハードウェアの使用が考えられ、その一つのハードウェアプラットフォーム候補はGPUである。GPUの使用と適切なアルゴリズム実装によるリアルタイム計測をデモンストレーションした研究は数多く存在する[5,6]。これらの研究は、主に際立った演算性能を持つハイエンドなGPUに注目したものである。残念ながら、これらのハイエンドなGPUは電力消費量も際立っている。例えば、NVIDIAの GeForce GTX TITAN X は最大250Wの電力が必要となる[7]。そのような大量消費電力のハードウェアをバッテリ駆動のモバイルシステムに内蔵させることは非現実的であり、このことが、GPUベースのステレオ計測システムが移動ロボット分野で広がらない原因になっている。もう一つのハードウェアプラットフォーム候補はFPGA（field-programmable gate array）である。FPGAは汎用的な集積回路であり、特定のアプリケーションをプログラミングすることが可能である。FPGAでは回路レベルでプログラムが実行されるため、CPUやGPUの命令パイプライン（fetch-decode-executeサイクル）を強要されない。一方で、FPGAでは、解くべき問題をたくさんの小さな問題へと細分化する構造が内包されており、それらの小問題は最小電力で並列処理によって解かれる。

FPGAの欠点は、CPUやGPUを用いたケースよりも膨大なプログラミング努力を必要とする点にある。これにより、FPGAベースの画像処理は最近の画像処理システムにおいても未だ一般的ではない。Nerian Vision社は、FPGAを用いたステレオ計測システムを研究者やアプリケーション開発者が入手しやすくするために、SceneScanとSceneScan Proという二つのセンサシステムを開発した、SceneScan Proを図1に示す。FPGAを使用することにより、この小さなデバイスがステレオ計測処理をリアルタイム且つ高フレームレートで行う。これらのシステムの主な特徴は次のとおりである。

- 最大3.4メガピクセルの解像度を持つ入力ステレオ画像を処理
- 256ピクセルまでの視差をカバーし、視差のサブピクセル解像度は1/16ピクセル
- 処理速度は最大100fps
- 電力消費量は8W
- 算出された視差マップはギガビットイーサネットにてリアルタイム転送

SceneScanとSceneScan Proは、Nerian社のSP1システム[1]の後継機である。SP1と比較すると、SceneScan Proは3倍の処理パフォーマンスを持ち、計測精度が大幅に向上している。本稿ではその実現方法の概要を述べるとともに、得られたパフォーマンスを示す。

図1：SceneScan Proと接続されたKamin2ステレオカメラ

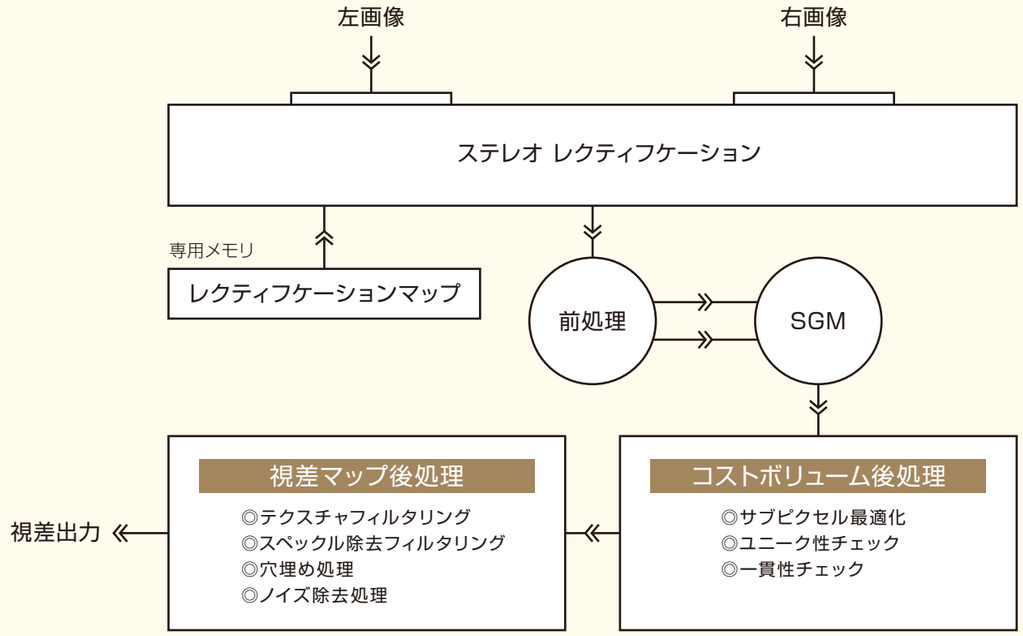


図2：処理構造のブロック線図

2 処理構造

ステレオ計測システムSceneScanとSceneScan Proは、Nerian社独自のステレオカメラKarmin2、もしくは標準的なUSB3Visionのカメラと接続する。システムは同期した左右のステレオ画像を取得し、内蔵FPGAによって処理を行う。算出されたステレオ視差マップ、すなわちインバースデプスはギガビットイーサ経由で接続された計算機に出力される。

SceneScanとSceneScan Proで実装されているステレオマッチング法は、セミグローバルマッチング（SGM）[9]をベースにしている。SGMは、高品質なマッチング結果と高い計算効率を持つため、現在では広く知られた手法である。SGMは、例えば Žbontarらの手法[10]のような、最高レベルのパフォーマンス（訳注：KITTIのステレオベンチマークでの高い成績を指す）を実現するステレオマッチング法のベースとなっている。しかし、実際には、SGMだけでこれらに比肩するマッチング精度が得られるわけではない。SGMを含み、かつ前処理および後処理をも含む、より広範な画像処理パイプラインが必要である。図2に示すように、SceneScanとSceneScan Proにはそのようなパイプラインが実装されている。

2.1 ステレオレクティフィケーション

入力画像に最初に適用される前処理ステップはステレオレクティフィケーションである。レクティフィケーションでは、事前校正にて算出されたレクティフィケーションマップを専用メモリから読み込む。このマップは、左右の画像の全画素について、方向と方向のサブピクセル量のオフセットを持つ。レクティフィケーション後の画像の画素値は、サブピクセル量のオフセット位置に対するバイリニア補間によって得られる。これらのオフセットは、単一のデータストリームを読めば両画像の全画素のオフセットベクトルが取得できるようにインターリーブされている。データ転送量低減のため、レクティフィケーションマップは圧縮されており、1画素のオフセットベクトルをエンコードするのに必要なメモリ量は平均で1バイトである。すなわち、レクティフィケーションマップの全体メモリ量は左右の入力画像の必要メモリ量とほぼ同じである。画像レクティフィケーションはウィンドウ走査によって処理しており、各画素の最大オフセット量は使用するウィンドウサイズに制限されている。SceneScanとSceneScan Proでは使用ウィンドウサイズは[pixels]であり、これにより、許されるオフセット量はから[pixels]となる。

2.2 画像の前処理

本システムでは、入力ステレオ画像に前処理を適用している。この前処理によって、2枚の画像間の明るさの違いやオクルージョンに対し、後段の処理がよりロバストになる。

2.3 ステレオマッチング

ステレオマッチングは、SGMアルゴリズムの改変版を適用することで実施される。小さな視差の変化と、大きな視差の変化に対するSGMのペナルティ値p_1とp_2は、実行時に設定することができる。ここでは、左画像の一つの画素に対し、いくつかの繰返し処理を必要とする。各繰返し処理では、左画像の注目画素値が右画像の画素グループと比較される。SceneScan Proでは右画素グループの画素数は$p=32$である（訳注：次段でpは"parallelization"と呼んでいることから、この比較は並列処理される模様）。左画像の画素ごとの繰返し回数n_iは、視差レンジを設定することによって実行時に設定できる。

ステレオマッチングにおける最小視差を示す視差オフセットo_dも実行時に設定できる。視差オフセットが$o_d>0$のとき、視差はo_dよりも小さくなることができないので、観測可能なデプスレンジは最大リミット値を持つ。視差オフセットo_d、繰返し回数n_i、および並列化（parallelization）pは、次式で最大視差d_{max}を定める。

$$d_{max} = o_d + n_i p - 1 \qquad (1.1)$$

2.4 コストボリューム後処理

SGMアルゴリズムは、左画像の画素と右画像の画素の有効な全組み合わせに対するコスト値を保持したコストボリュームを生成する。いくつかの後処理ステップがそのコストボリュームに直接適用される。

サブピクセル最適化

後処理として最初に適用されるステップは、サブピクセル最適化である。このステップは、各画素に対し、最小コストを持つ視差を取得した後に、その左右の視差コストを評価することによってデプス値の精度を向上させる。ここでは、左右の視差コスト値と最小コスト値を曲線にフィッティングし、その最小値に位置する視差をサブピクセル視差とする。次いで、その推定視差値は固定小数点値にエンコードされる。SceneScanとSceneScan Proではサブピクセル精度のために下位4bit固定小数点をサポートしており、これにより視差は1/16のサブピクセル解像度を持つ。

視差のユニーク性チェック

ここでは、視差のユニーク性拘束を導入することにより、不確実性が高いマッチング結果を排除する。ステレオマッチング結果がユニークとみなせるケースでは、最小マッチングコストc_{min}と、$q \in [C1, \infty)$との積算値が、二番目に小さいコスト値よりも小さくなければならない。この関係は次式で表すことができる。

$$c^* \cdot q < \{C \setminus \{c_{min}\}\} \quad (1.2)$$

ただし、Cは有効視差の全てのコストを表し、$c^* = c_{min}$は最小コストを示す。このユニーク性チェックによって排除されたマッチングを持つ画素は、無効な視差を持つ画素としてラベル付けされる。

視差の一貫性チェック

高い不確実性を持つマッチング結果をさらに排除するために、視差の一貫性がチェックされる。共通のアプローチは、反対方向のマッチングを実施する(本ケースでは、右画像から左画像へのマッチングを実施する)ことである。そして、次式を満たすマッチング結果だけを残す。

$$|d_l - d_r| \leqq t_c \quad (1.3)$$

ただし、d_lは先に行った左画像から右画像へのマッチング結果、d_rは右画像から左画像へのマッチング結果、t_cは一貫性チェック用の閾値である。FPGAのリソースを節約するために、ここでは反対方向のステレオマッチングを改めて実施することを避けている。代わりに、最初に行った左画像から右画像へのマッチングを行いながら、右から左へのステレオ結果の視差マップを、左から右へのマッチングコストから推測している。この一貫性チェックをパスしなかった画素も、無効な視差を持つ画素としてラベル付けされる。

2.5 視差マップ後処理

コストボリューム後処理の後に、コストボリュームは視差マップへと変換される。その視差マップに対して追加の後処理ステップが実施される。

テクスチャフィルタリング

わずかなテクスチャしかない画像領域、もしくは全くテクスチャのない画像領域に対するマッチングは、特に難しい問題である。特に、そのような画像領域が画像の境界付近に現れると、多くのミスマッチを引き起こす。この問題に対し、テクスチャフィルタリングを実施する。このフィルタリングでは、各画素に対し、その周辺領域のテクスチャ強度を表すテクスチャスコアを計算する。このスコアが、事前に定められる閾値よりも小さい場合は、無効な視差を持つ画素としてラベル付けされる。

スペックル除去フィルタリング

ここまでに述べてきた後処理手法を使っても、常に全ての誤対応を取り除けるわけではない。幸いなことに、残った誤対応は、類似視差を持つ画素どうしの小さな塊になって現れる傾向がある。これらのスペックルはスペックル除去フィルタによって排除される、このフィルタは、最小サイズよりも小さい連結成分を見つける。この最小サイズはスペックルフィルタのウィンドウサイズによって制御される。発見されたスペックルは、無効な視差を持つ画素としてラベル付けられる。

穴埋め処理

前述したすべての後処理ステップは、得られた視差マップから無効な視差を除去しているため、視差マップは穴があいた状態となる。もしその穴が小さい場合は、周辺の有効な視差値を使い、それらの視差値から補間することによって、その穴を埋めることができる。補間処理は、穴の水平方向の大きさl_hと垂直方向の大きさl_vが次式を満たす場合に実施される。

$$\{l_h, l_v\} \leqq l_{max} \quad (1.4)$$

ただし、l_{max}は穴の最大サイズを示す。この補間処理は、穴の周囲の視差値が類似した値を持たない場合にも省略される。

ノイズ除去処理

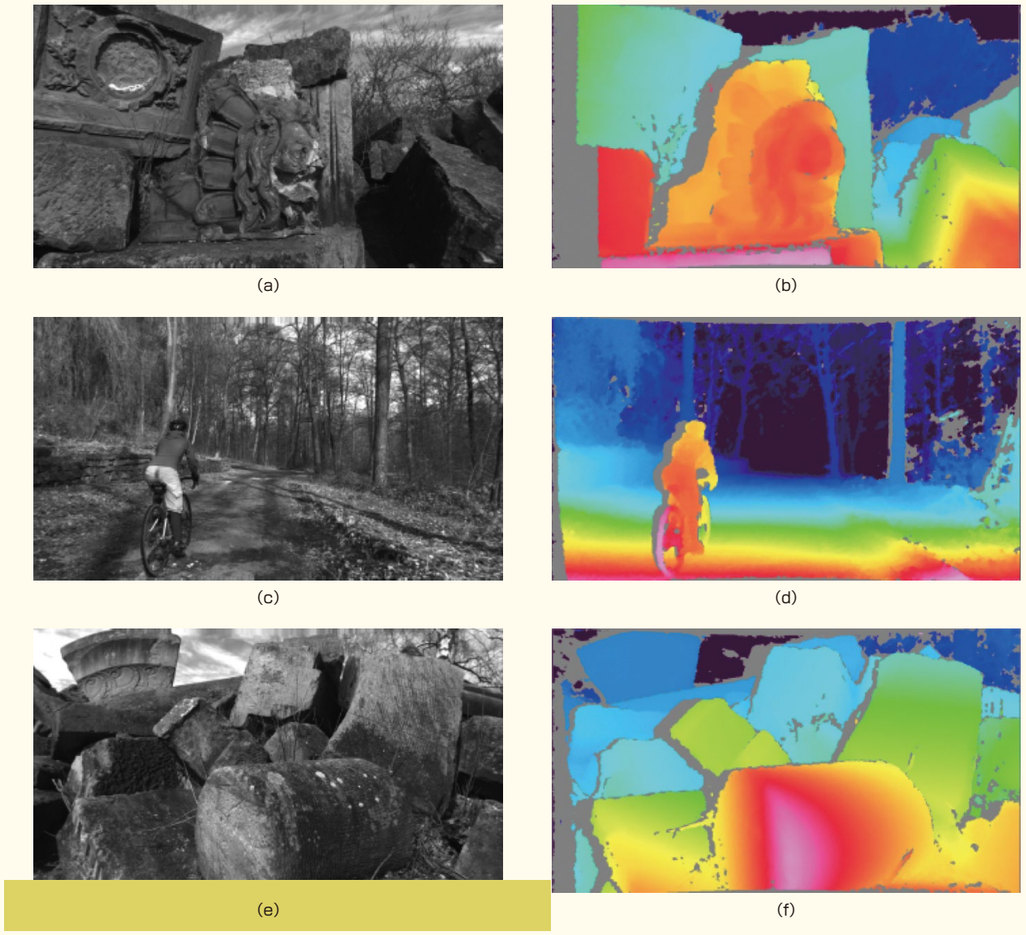
最後に、生成された視差マップにノイズ除去フィルタリングを施す。ここでは、視差マップの不連続性と無効な視差を認識した上で視差マップを平滑化する、処理結果は、SceneScanのイーサネット出力に直接転送される。

3 結果

表1は、異なる画像解像度と視差レンジにおいて、SceneScan Proの実現可能なフレームレートをリストしたものである。これらのフレームレート値は、接続された計算機が視差データを受け取った際の値であるので、カメラやネットワーク通信におけるオーバーヘッドが含まれたものである。すべての設定において、SceneScan Proは、カメラ電力を除き、8[W]の電力しか消費しない。最大のフレームレートは画像解像度が640×480で視差レンジが128[pixels]の時に得られ、SceneScan Proは最大100[fps]の高速処理ができる。画像解像度と視差レンジを増やすとフレームレートは下がってゆく、SceneScan Proの最大画像解像度は1856×1866、すなわち3.4メガピクセルである。最大の視差レンジを256[pixels]とすると、SceneScan Proは1秒あたり51億回の視差評価を実施することになり、それは1秒あたり2000万の視差出力に相当する。視差レンジを128[pixels]に減じた場合、処理パフォーマンスは1秒あたり3000万視差出力に増加する。このとき、各視差値は3次元座標に変換されるため、1秒あたり3000万個の3次元計測点を提供することになる。

視差レンジ	画像解像度			
	640 × 480	800 × 592	1280 × 960	1600 × 1200
128[pixel]	100[fps]	65[fps]	24[fps]	15[fps]
256[pixel]	70[fps]	45[fps]	15[fps]	10[fps]

表1:ScenScan Proの処理パフォーマンス



(a) (b) (c) (d) (e) (f)

図3:(a,c,e)左入力画像の例　(b,d,f)算出された視差マップ

図3は、入力画像とSceneScan Proによって算出された視差マップの例を示している。全ての画像は、晴天時に屋外で取得したものである。ステレオ計測は明るい環境光を苦にしないため、全ての例で密な視差マップを取得できている。
それぞれの画像は、カメラに近い物体と遠い物体を含んでいる。対象までの距離が計測密度や計測ロバスト性に影響を与えていないことがわかる。例えば図3(d)の中央部のように、相当遠い物体の画像領域でさえ、密なステレオ計測結果が得られていることがわかる。
総じて視差マップはとてもスムーズであり、わずかな誤対応しか見られない。このことは、コストボリュームと視差マップに対する広範な後処理によって信頼づけられている。図3では、前景物体の左側に明瞭なオクルージョンシャドウが現れていることから、オクルージョン領域も効果的にフィルタされていることがわかる。

4 結論

本稿では、計測ハードウェアFPGAを用いたステレオ計測のための二つのスタンドアロンシステムSceneScanとSceneScan Proを示した。FPGAを使うことにより、これらのシステムはとても高い電力効率を実現した。同時に、並外れた計算パフォーマンスを提供し、100[fps]でのステレオ画像処理が可能となった。特に、自律移動ロボットの分野では、様々なアプリケーションのための高速に取得できる高精度なデプスデータが不可欠である。ステレオ計測は、明るい環境やロングレンジの計測であってもそのようなデータの提供を約束する。SceneScanとSceneScan Proシステムは、消費電力が大きいGPUベースのシステムとは異なり、バッテリ駆動で消費エネルギーに制限のあるモバイルプラットフォームと簡単に統合できる。私たちは、これらのシステムにより、研究者や開発者にとってステレオ計測が今よりも入手可能になることを望むものである。さらに、ステレオ計測原理を利用する他のシステムの開発も促進できれば幸いである。

【参考文献】

1 . M. Humenberger, C. Zinner, M. Weber, W. Kubinger, and M. Vincze, "A fast stereo matching algorithm suitable for embedded real-time systems," Computer Vision and Image Understanding, vol. 114, no. 11, pp. 1180-1202, 2010.
2 . A. Geiger, P. Lenzm, C. Stiller, and R. Urtasun, "The KITTI vision benchmark suite - stereo evaluation," 2015, [http://www.cvlibs.net/datasets/kitti/eval_scene_flow.php?benchmark=stereo] 2018, accessed: 17.09.2018.
3 . S. K. Gehrig and C. Rabe, "Real-time semi-global matching on the CU," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, pp. 85-92.
4 . R. Spangenberg, T. Langner, S. Adfeldt, and R. Rojas, "Large scale semiglobal matching on the CPU," in IEEE Intelligent Vehicle Symposium (IV), 2014, pp. 195-201.
5 . I. Haller and S. Nedevschi, "GPU Optimization of the SGM Stereo Algorithm," in IEEE International Conference on Intelligent Computer Communication and Processing (ICCP), 2010, pp. 197-202.
6 . J. Kowalczuk, E. T. Psota, and L. C. Perez, "Real-time stereo matching on cuda using an iterative refinement method for adaptive support-weight correspondences," IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 1, pp. 94-104, 2013.
7 . NVIDIA, "GeForce GTX TITAN X specifications," [https://www.geforce.com/hardware/desktop-gpus/geforce-gtx-titan-x/specifications,] 2018, accessed: 19.09.2018.
8 . K. Schauwecker, "SP1: Stereo vision in real time," in MuSRobS@IROS, 2015, pp. 40-41.
9 . H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, 2005, pp. 807-814.
10. J. Zbontar and Y. LeCun, "Computing the stereo matching cost with a convolutional neural network," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
11 . S. K. Gehrig, F. Eberli, and T. Meyer, "A real-time low-power stereo vision engine using semi-global matching," Computer Vision Systems, pp. 134-143, 2009.
12 . C. Banz, S. Hesselbarth, H. Flatt, H. Blume, and P. Pirsch, "Real-time stereo vision systemusing semi-global matching disparity estimation: architecture and FPGA-implementation," in IEEE Int. Conf. on Embedded Comput. Syst. (SAMOS), 2010, pp. 93-101.

NANOXEED　nerian VISION TECHNOLOGIES