

グラスレス3Dレグザ (REGZA) の多視差変換アルゴリズム

Multiparallax Conversion Algorithm for Glasses-Free 3D REGZA

三田 雄志 菊池 義浩 山内 日美生

■ MITA Takeshi ■ KIKUCHI Yoshihiro ■ YAMAUCHI Himio

グラスレス3Dレグザでは、9方向の異なる視点から見た視差画像^(注1)を表示することで、専用の眼鏡を使わず(グラスレス)に自然で見やすい3D(立体視)映像を視聴できる。多視差変換アルゴリズムは、一般のテレビ放送のような単眼映像やBlu-ray 3DTM^(注2)のようなステレオ映像から複数の視差画像を作り出す技術であり、デプス(奥行き情報)生成処理と多視差画像生成処理の二つから成る。東芝は、動き検出、人物検出、及び構図識別を組み合わせた高度なデプス生成処理技術を開発し、様々な映像をグラスレスで3D視聴できるようにした。更に、多視差画像生成処理では、生成されたデプスから視聴者の好みに応じた視差量を求めているため、画面から飛び出して見える度合いを調節することが可能であり、快適な視聴を実現できた。

The glasses-free 3D REGZA series enable users to watch high-quality three-dimensional (3D) images by generating nine parallax images from regular two-dimensional (2D) TV programs and stereoscopic 3D contents such as Blu-ray 3DTM movies without the need for dedicated glasses.

As a key technology to realize these products, Toshiba has developed a multiparallax conversion algorithm consisting of two processing technologies: a depth generation technology that makes it possible to convert any contents into 3D images with a combination of motion estimation, face detection, and scene classification; and a parallax image creation processing technology that allows users to adjust the amount of disparity according to personal preference in order to comfortably watch 3D contents.

1 まえがき

グラスレス3Dレグザは、九つの異なる角度から観測した画像を表示することにより、専用の眼鏡を使わず(グラスレス)に、自然な3D(立体)映像を視聴することができる。

ここでは、通常のテレビ放送やホームビデオなど1台のカメラで撮影された単眼映像や、近年増加している3D映画や3D放送など2台のカメラで撮影されたステレオ映像から9視差^(注1)の映像を作り出す、東芝独自の“多視差変換アルゴリズム”について述べる。

単眼映像やステレオ映像をグラスレス3Dレグザにそのまま表示しても、3D映像を視聴できるわけではない。当社は、多視差変換アルゴリズムを開発したことで、世の中に流通している多くの映像コンテンツを3Dで視聴できる製品を実現した。

2 多視差変換アルゴリズムの概要

多視差変換アルゴリズムの概要を図1に示す。

まず、入力された単眼映像又はステレオ映像の画像を基に、シーンの奥行きを表現するデプスを求める。図1では、被写体

(注1) 異なる複数の視点から観測した画像内で、同じ物体が異なった位置に映るときは位置ずれを視差と言う。また、視差のついた画像を視差画像と言う。
 (注2) Blu-ray DiscTM、Blu-rayTM、Blu-ray 3DTMは、Blu-ray Disc Associationの商標。

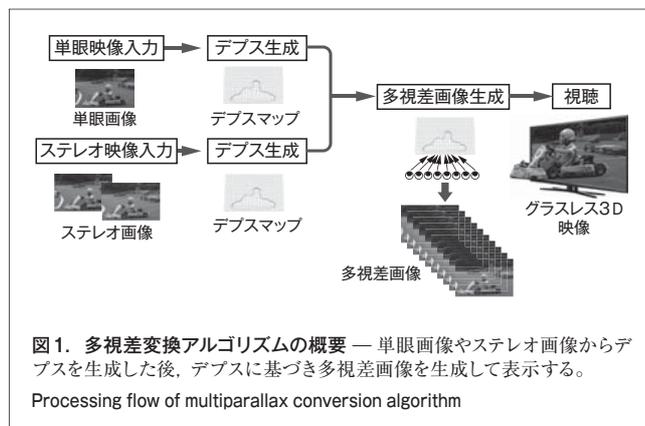


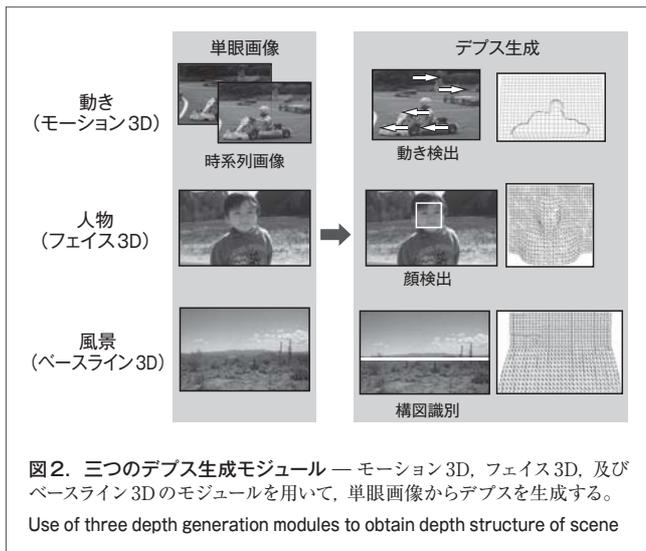
図1. 多視差変換アルゴリズムの概要 — 単眼画像やステレオ画像からデプスを生成した後、デプスに基づき多視差画像を生成して表示する。
 Processing flow of multiparallax conversion algorithm

となっている車が手前にあり、背景が奥にあることを示すデプスが得られている。

次に、このデプスに基づいて、仮想的に設置した九つのカメラからそのシーンを観測するとどのような画像が見えるかを再現し、多視差画像を生成して表示する。デプスの生成方法は、単眼画像が入力された場合とステレオ画像が入力された場合で異なるが、これらについて、以下で順を追って述べる。

3 単眼画像からのデプス生成法

単眼画像からデプスを生成する仕組みを図2に示す。ホームビデオや写真に加えて、24時間分の放送映像を調査



した結果, 多くの映像には, カメラワークなど動きを伴うシーン, 人物が映っているシーン, 及び水平線や地平線など典型的な構図のある風景シーンの3種類が含まれることが判明した。そこで, それぞれのシーンに対応した, “モーション3D”, “フェイス3D”, 及び“ベースライン3D”の三つのデプス生成モジュールを用いて画像からデプスを生成し, デプスマップを算出するようにした。

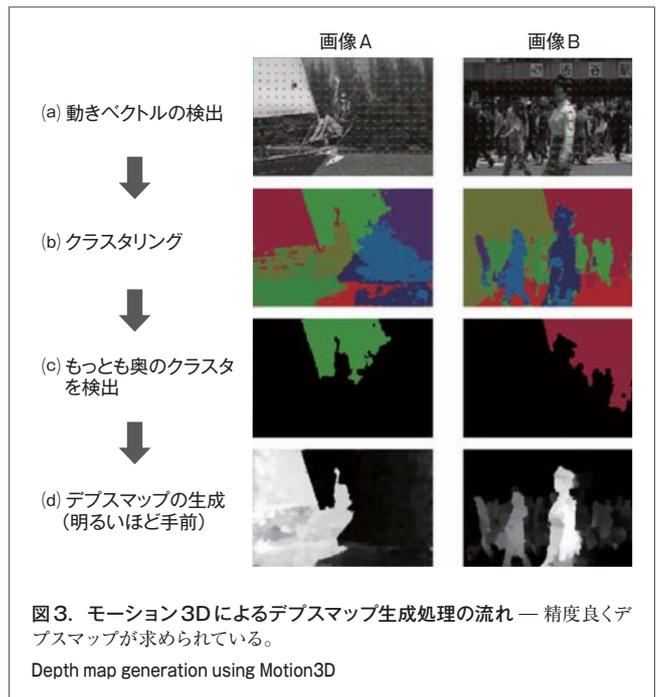
これら三つのデプス生成モジュールを組み合わせることで, 単一のモジュールを用いるのに比べて, そのシーンに適応したデプスを生成することができる。また, 人物のように見慣れた対象は違和感が生じやすいので, 注意深くデプスを作る必要があるが, 顔検出により人物が映っているかどうかを把握したうえで人物の形状モデルを当てはめることにより, 自然な立体感が得られるよう工夫した。

3.1 動き検出に基づくモーション3D

カメラの動きからデプスを推定する代表的な方法はモーションステレオ法である。これは, 移動カメラによって撮影された映像から時間的に異なる2枚のフレーム画像を取得し, その画像間で対応付けを行うことによりデプスを推定する。カメラから遠い物体ほど2枚の画像間での移動量が小さくなる, という基本原理に基づいている。静止しているシーンに対してカメラだけが移動している理想状態であれば正確なデプスを取得できるが, 移動物体を含むシーンでは基本原理が成り立たなくなるため, デプスの正確性は劣化する。また, Factorization⁽¹⁾に代表される Shape from Motionの手法も, 基本的に前記の理想的な状態を前提としている。

モーション3Dは, 前記の問題を完全にではないが大幅に解決することが可能である。モーション3Dによるデプスマップ生成処理の流れを図3に示す。

まず, 時間的に前後する2枚のフレーム画像からブロックごとに移動を検出し動きベクトルを算出する。対応するブロック



の探索には, 4-D Recursive 探索⁽²⁾と呼ぶ独自の方法を用いる(図3(a))。この方法は, 少ない計算量で精度の高い動きベクトルが得られるという特長がある。

次に, 長さや方向が類似した動きベクトルをいっしょにして複数グループに分ける, クラスタリングという処理を行う。各グループをクラスタと呼ぶが, これらのクラスタ間で互いの領域を遮蔽する関係を調べることにより, いずれのクラスタがカメラから遠い領域なのかを調べる。カメラに近い物体(前景)に対応するクラスタは, その周辺の背景クラスタを覆い隠す動きをする。クラスタの境界領域の動きベクトルを調べ, 他を隠す動きをしているベクトルを前景と判定する。このようなクラスタ間の遮蔽関係に基づいて, カメラからもっとも遠い, すなわちもっとも奥のクラスタを求める(図3(b), (c))。

最後に, もっとも奥のクラスタの平均動きベクトルと各画素の動きベクトルの2乗誤差を求めることで, もっとも奥の動きに対する相対的な動きを全ての画素について算出する。この補償された動きをデプスに変換する。

図3の画像Aは, 航行するヨットに乗っている人物をカメラで追いつながり撮影したものである。カメラの移動によって背景領域は動いて見える一方, ヨットと人物は画像上ではほぼ静止して見えるような映像である。図3(c)を見てもっとも奥のクラスタとして背景の森が正しく選択されており, 図3(d)のデプスマップもヨットと人物に手前のデプスを与えている。

画像Bは, 静止したカメラからスクランブル交差点を歩く人々を撮影したものであり, デプス生成が非常に難しい映像である。しかし, クラスタ間の遮蔽関係を評価しているため, 人々の前後関係がほぼ矛盾なく求められ, 良好なデプスマップ

プが得られている。

モーション3Dの利点は、もっとも奥の動きに対する相対的な動きを扱うという工夫により、様々な方向の動きを統一的に扱え、高精度なデプスを生成できることである。もう一つの利点は、モーションステレオ法における理想的な状態、すなわち静止したシーンに対してカメラのみが移動する状態だけでなく、移動物体を含むシーンも扱える点である。複数の移動物体が重なり合って移動するような場合には、前記の画像Bのように前後関係を判定できる。

3.2 顔検出に基づくフェイス3D

フェイス3Dによる人物領域のデプス生成処理の流れを図4に示す。

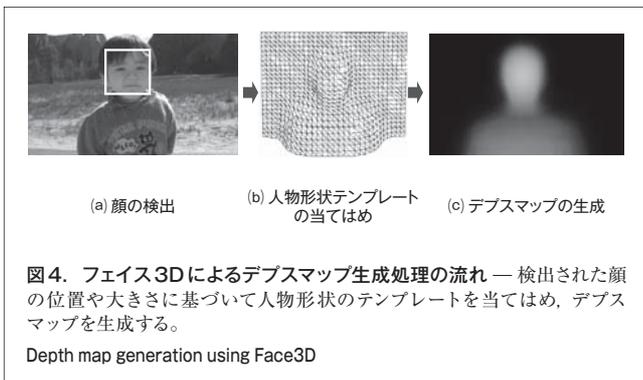
まず、入力画像から顔を検出する。検出された顔の位置と大きさに応じて人物形状のテンプレートを当てはめ、デプスマップを生成する。これには、高精度な顔検出アルゴリズムが必要となる。特に、顔でない背景領域を誤って顔として検出してしまふ過検出を避けなければならない。過検出した場合、人物が存在しない領域が不自然にふくらんでしまうからである。

そこで、当社が提案している顔検出方法⁽³⁾を用いた。この方法は、高い検出精度が得られ、過検出を少なくできる。人物形状テンプレートは、約200名からデータを取得し作成した。これにより、検出された人物に適合する自然なデプスを生成できている。

一方、他の研究では、入力画像のRGB(赤、緑、青)値のうちR成分の大きさによって、デプスを制御している⁽⁴⁾。R成分は肌色に多く含まれるため、肌が露出している領域はある程度うまく動作するが、髪の毛や衣服などR成分が少ない領域はうまく扱えない。

別の研究では、二人の人物が対話している典型的なシーンを簡便な手法で検出し、二つの円体をデプスとして当てはめることを提案している⁽⁵⁾。しかし、当てはめるモデルは人物形状に沿ったものではなく、十分な立体感とは得られにくい。

フェイス3Dは、本格的な人物検出を導入した初めてのデプス生成方法であり、人物領域において、従来に比べ自然な立体感が得られる方法である。



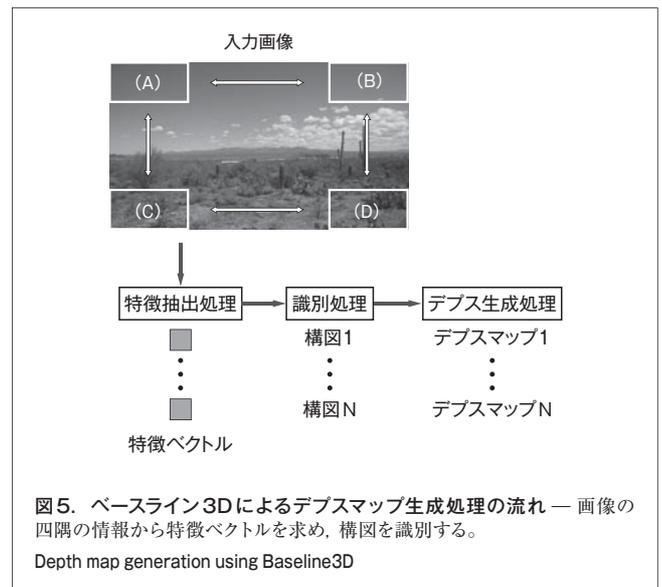
3.3 構図識別に基づくベースライン3D

ベースライン3Dは、典型的な奥行き構造を持つ風景シーンのデプスを生成するモジュールである。構図の識別によって、入力された風景が複数あるデプスパターンのいずれかを判断し、もっとも当てはまりの良いデプスを選択する。構図識別の具体的な手順は、三つのステップから成る。

第1のステップは、識別のための特徴抽出である。図5に示すように、入力画像の四隅の方形領域から色ヒストグラムを算出し、矢印で示す領域間でヒストグラムの類似度を求める。この類似度を特徴量として識別を行う。図のように地平線を含むシーンの多くは、領域(A)と(B)、領域(C)と(D)間のそれぞれの類似度は高く、逆に領域(A)と(C)、領域(B)と(D)間のそれぞれの類似度は低くなる傾向がある。色そのものには構図を決定付ける情報はあまりないが、領域間の相対的な関係を調べることで有益な情報を引き出すことができる。

第2のステップは、前述の特徴量に基づく識別処理である。構図ごとに手作業で分類した学習用の画像から特徴を抽出し、構図を識別するための関数を学習しておく。識別にはマルチクラス サポート ベクターマシン^(注3)を用いており、単純なしきい値処理による識別方法⁽⁵⁾に比べて高い識別精度を期待できる。

第3のステップはデプス生成である。構図ごとに事前に定めておいたルールに従ってデプス値を決定する。例えば、地平線の構図については、水平方向のエッジを検出するオペレータを走査して地平線位置を検出し、地平線の位置で折れ曲がる2枚の平面を当てはめることでデプスマップを得る。一連の処理は極めて簡便な計算で実現でき、リアルタイム処理やハードウェア化に適している。

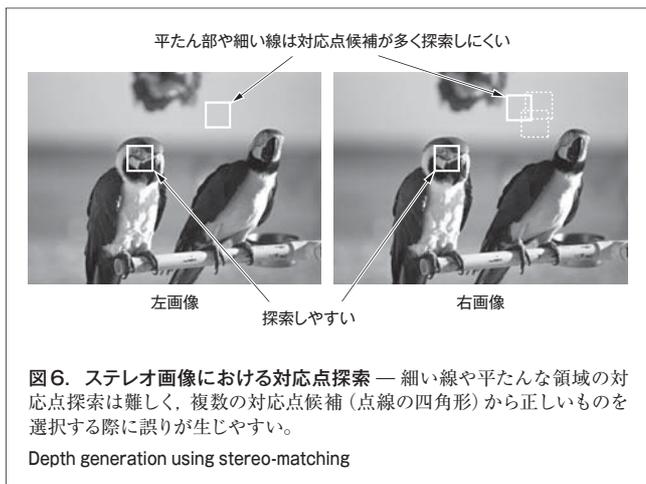


(注3) 複数のクラス(ここでは“構図”と言う)を正確に識別する方法。

4 ステレオ画像からのデプス生成法“ステレオ3D”

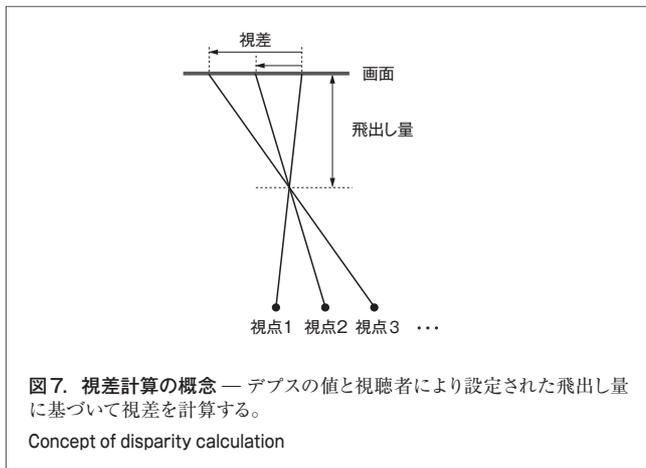
3D放送やBlu-ray 3D™などの2枚の視差画像（ステレオ画像）が入力された場合は、3章とは別の方法を用いてデプスを生成する。具体的には、入力された左右2枚の画像から対応点を探索し、それらの水平方向のずれ量に応じてデプスを求める。

ところが、図6に示す空のような平たんな領域や細い線は、おうむの顔のように複雑な模様のあるテクスチャ領域に比べて、対応点の候補が多く探索が難しいという問題がある。ステレオ3Dは、平たんな領域の探索に適したマッチング手法とテクスチャ領域のマッチングに適した手法の二つを組み合わせることで、探索の誤りを大幅に減らしている。



5 多視差映像の生成と表示

前章までの処理により、入力画像の各画素の奥行きを表すデプス値が得られる。このデプス値に基づいて各画素を水平方向にシフトさせることで、9枚の視差画像を作ることができる。シフト量すなわち視差は、デプス値、事前に定められた九つの仮想的な視点の位置、更に視聴者が設定する好みの奥



行き強度に応じて算出される。

視差計算のための概念を図7に示す。好みの飛出し量を得るには、図に示した幾何学的な関係に基づいて視差を求め、入力画像をシフトさせればよい。九つの仮想的な視点は、より広い視域で高画質な立体映像を視聴できるように、この特集の別の論文で紹介する“視域最適化技術”(p.18-20参照)やパネルの表示能力を考慮して設計されている。また、視聴者は、好みに応じて視差を大きくしたり小さくしたりすることで、画面から飛び出して見える度合いを調整できるようにしている。

6 あとがき

グラスレス3Dレグザで自然な3D映像を映し出すために、当社が開発した多視差変換アルゴリズムについて述べた。このアルゴリズムによって、通常のテレビ放送やホームビデオなどの単眼映像や、近年増加している3D映画などのステレオ映像から9視差の画像を生成し、専用眼鏡なしで自然で見やすい3D映像を視聴できるようになった。

今後は、更に高画質で迫力ある3D映像の生成を目指して、多視差変換アルゴリズムの改良を進めていく。また、グラスレス3Dレグザで培った画像処理技術を更に深耕し、パソコンなどテレビ以外の当社製品に展開していく。

文 献

- (1) Tomasi, C.; Kanade, T. Shape and motion from image streams under orthography : a factorization method. IJCV. 9, 2, 1992, p.137-154.
- (2) 山本琢磨 他. "4-D Recursive探索を用いた高精度動き推定". 映像情報メディア学会年次大会. 松山市, 2010-09, 映像情報メディア学会. 2010.
- (3) Mita, T. et al. Discriminative feature co-occurrence selection for object detection. IEEE Trans. on PAMI. 30, 7, 2008, p.1257-1269.
- (4) Yamada, K.; Suzuki, Y. "Real-time 2D-to-3D conversion at full HD 1080P resolution". Proc. of IEEE ISCE. Kyoto, 2009-05, IEEE. 2009.
- (5) Burazerovic, D. et al. "Automatic depth profiling of 2D cinema and photographic images". Proc. of ICIP. Cairo, 2009-11, IEEE. 2009.



三田 雄志 MITA Takeshi

研究開発センター マルチメディアラボラトリー研究主務。
3D映像処理技術の開発に従事。電子情報通信学会、IEEE Computer Society会員。
Multimedia Lab.



菊池 義浩 KIKUCHI Yoshihiro

デジタルプロダクツ&サービス社 コアテクノロジーセンター
エンベディッドシステム技術開発部グループ長。映像・音響
処理技術の開発に従事。
Core Technology Center



山内 日美生 YAMAUCHI Himio

デジタルプロダクツ&サービス社 コアテクノロジーセンター
AV技術開発部グループ長。高画質化技術の開発に従事。
映像情報メディア学会会員。
Core Technology Center