

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6903657号
(P6903657)

(45) 発行日 令和3年7月14日 (2021.7.14)

(24) 登録日 令和3年6月25日 (2021.6.25)

(51) Int. Cl.	F I
HO 4 N 13/261 (2018.01)	HO 4 N 13/261
HO 4 N 13/271 (2018.01)	HO 4 N 13/271
GO 6 T 7/50 (2017.01)	GO 6 T 7/50

請求項の数 15 (全 29 頁)

(21) 出願番号	特願2018-525821 (P2018-525821)	(73) 特許権者	518040390
(86) (22) 出願日	平成28年8月3日 (2016.8.3)		ヘフィーダ、ムハンマド エム
(65) 公表番号	特表2018-528733 (P2018-528733A)		カタール ドーハ ピーオー ボックス
(43) 公表日	平成30年9月27日 (2018.9.27)		5825 カタール ファンデーション
(86) 国際出願番号	PCT/GB2016/052388		フォー エデュケーション、サイエンス
(87) 国際公開番号	W02017/021731		アンド コミュニティー ディベロップメント
(87) 国際公開日	平成29年2月9日 (2017.2.9)	(73) 特許権者	518040404
審査請求日	令和1年7月30日 (2019.7.30)		カルガリ、キアナ アリ アスガル
(31) 優先権主張番号	62/200,311		カタール ドーハ ピーオー ボックス
(32) 優先日	平成27年8月3日 (2015.8.3)		5825 カタール ファンデーション
(33) 優先権主張国・地域又は機関	米国 (US)		フォー エデュケーション、サイエンス
			アンド コミュニティー ディベロップメント
			最終頁に続く

(54) 【発明の名称】 2Dから3Dへのビデオフレーム変換

(57) 【特許請求の範囲】

【請求項 1】

2Dビデオ画像を3D画像に変換するためにビデオストリームから抽出した2Dビデオ画像を処理する方法であって、

2D画像およびその2D画像に対応する深度情報とで構成される合成3D参照データベースを作成し、

入力ビデオフレームを合成3D参照データベースに投入し、

入力ビデオフレームの少なくとも一部を合成3D参照データベース内の2D画像の少なくとも一部とマッチングし、

その2D画像に対応する深度情報を選択し、前記深度情報は深度勾配であり、

選択した深度勾配をマッチングされた入力ビデオフレームの部分に適用して2D + 深度情報のフレーム部分を生成することから成る方法。

【請求項 2】

前記部分は $n \times n$ 個のピクセルで構成されるブロックであることを特徴とする請求項 1 に記載の方法。

【請求項 3】

入力ビデオフレームの別の部分を合成3D参照データベース内の前記2D画像または別の2D画像の部分とマッチングすることから成ることを特徴とする請求項 1 または 2 に記載の方法。

【請求項 4】

10

20

選択した深度勾配を前記マッチングされた入力ビデオフレームに適用することは、マッチングされた2D画像の部分の深度勾配を、対応するマッチングされた入力ビデオフレームの部分に適用することから成ることを特徴とする請求項1～3のいずれかに記載の方法。

【請求項5】

選択した深度勾配をマッチングされた入力ビデオフレームに適用することは、マッチングされた2D画像の部分の対応するピクセルを1個以上、マッチングされた入力ビデオフレームの部分の対応するピクセルにマッピングすることから成ることを特徴とする請求項4に記載の方法。

【請求項6】

入力ビデオフレームとマッチングする2D画像の候補をフレームのGISTおよびカラー情報を用いて特定することを特徴とする請求項1～5のいずれかに記載の方法。

10

【請求項7】

深度情報を以下の式に従いポアソン再構築を用いて再構築することからなる請求項1～6のいずれかに記載の方法。

【数1】

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) D = \nabla \cdot G,$$

20

ここで $G = (G_x, G_y)$ は深度勾配、 D は深度情報、 $\nabla \cdot G$ は G の発散である。

【数2】

$$\nabla \cdot G = \left(\frac{\partial G_x}{\partial x} + \frac{\partial G_y}{\partial y} \right).$$

30

【請求項8】

入力ビデオフレーム内のオブジェクトを特定し、

特定したオブジェクトのオブジェクトマスクを決定し、

ポアソン方程式を修正することでオブジェクト境界での深度不連続性を許容し、決定したオブジェクトマスクおよびマッチングされた入力ビデオフレームを用いて深度情報を推定することから成ることを特徴とする請求項7に記載の方法。

40

【請求項9】

オブジェクト境界のピクセルを有効な隣接ピクセルとして使用することを許容しないことにより、ピクセルをオブジェクト境界から切り離し、オブジェクト境界に隣接するピクセルは画像境界のピクセルと同様に処理されるために、境界ピクセルに隣接する各ピクセルに対して、 A の中の対応する連結を0に設定し、それに応じて G の値を更新することから成る $Ax = b$ と式化され、ここで $b = \nabla \cdot G$ 、 $x = D$ であり、 A にはポアソン方程式の係数が記憶されているポアソン再構築を用いて深度情報を推定することから成ることを特徴とする請求項7又は8に記載の方法。

【請求項10】

50

ポアソン再構築は一次および高次導関数から成ることを特徴とする請求項 7 ~ 9 のいずれかに記載の方法。

【請求項 1 1】

以下の式を用いて深度勾配を精細化させることを特徴とする請求項 7 ~ 1 0 のいずれかに記載の方法。

【数 3】

$$G_x = G_x \times \max\left(1 - e^{\left(1 - \frac{1}{\alpha |G_x|}\right)}, 0\right)$$

$$G_y = G_y \times \max\left(1 - e^{\left(1 - \frac{1}{\alpha |G_y|}\right)}, 0\right)$$

10

【請求項 1 2】

連続領域で次の設定を行い、高次深度微分係数を強制的にゼロにすることを特徴とする請求項 1 0 に記載の方法。

【数 4】

$$\left(\frac{\partial^4}{\partial x^4} + \frac{\partial^4}{\partial y^4}\right) D = 0.$$

20

【請求項 1 3】

2D + 深度情報を用いて左側ステレオ画像および右側ステレオ画像を生成することから成ることを特徴とする請求項 1 ~ 1 2 のいずれかに記載の方法。

30

【請求項 1 4】

2Dビデオ画像を3D画像に変換するためにビデオストリームから抽出した2Dビデオ画像を処理するシステムであって、

2D画像およびその2D画像に対応する深度情報とで構成された、合成3D参照データベースと、

入力ビデオフレームを合成3D参照データベースに投入するように構成された探索モジュールと、

入力ビデオフレームの少なくとも一部と合成3D参照データベース内の2D画像の少なくとも一部とのマッチングを行うように構成されるマッチングモジュールであって、そのシステムは入力ビデオフレームの少なくとも一部を合成3D参照データベース内の2D画像の少なくとも一部とマッチングし、前記2D画像に対応する深度情報、を選択するように構成されるマッチングモジュールと、前記深度情報が深度勾配であって、

40

選択した深度勾配をマッチングされた入力ビデオフレームの部分に適用して2D + 深度情報のフレームの部分の生成するように構成された生成モジュールとから成るシステム。

【請求項 1 5】

実行時に請求項 1 ~ 1 3 のいずれかに記載の方法を実行する命令でプログラミングされた、コンピュータで読み取り可能な媒体。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

50

本発明者が提案するのはデータ駆動型の2D-3Dビデオ変換技術である。本発明者の技術は高品質合成画像のデータベースから深度勾配を転送することに基づく。このような画像はビデオゲームから収集することができ、スポーツとそのサブカテゴリ、アクションとそのサブカテゴリ、通常のグラフィクス等の様々なジャンルで入手できることが多い。このような合成画像を2D+深度のレポジトリとして利用すれば、高額なステレオカメラを使用する必要はなくなる。それによりこの技術は最先端技術よりも拡張可能性の高いものになる。さらに、従来のデータ駆動型の技術とは異なり、本発明者の方法ではクエリー画像とデータベースとが全体的に類似する必要はなく、局所的なパッチ（部分）が類似してさえいればよい。それにより2D+深度のデータベースのサイズを大幅に縮小しながら類似深度の品質を維持することができる。その結果として、より一般的なデータへの拡張可能性が高まり、民生品への適用が容易になる。

10

【背景技術】

【0002】

従来技術

過去数年間、3D媒体のアプリケーションは映画の範囲をはるかに越えて、多くの研究者から大きな関心を寄せられている。Liuら [15] は、3D映画撮影術の原理とともに、映画以外の3Dコンテンツにその技術を適用する重要性を論じている。Wuら [23] は、3Dコンテンツの品質をリアルタイムの遠隔没入型アプリケーションに適合させている。Calagariら [9] は、多種多様なディスプレイ装置に応じて奥行をカスタマイズできる3Dストリーミングシステムを提案している。Yang [24] らは、クライアントの視角に基づいて遠隔没入型環境における3Dコンテンツストリーミングの優先順位を決めている。これらのシステムは利便性の高い3Dアプリケーションを提供するが、3Dコンテンツの量が限られていることは依然として3D技術にとって大きな障害となる。この問題に取り組むため、多くの研究者が2D-3D変換技術を研究している。しかしながら、従来の方法は、半自動的であるか [19、26]、複雑な動きに対応できないか [12、21、13、10、11] のいずれかである。本発明者の知る限りでは、多様なシーン構造を伴う複雑な動きを処理できるサッカー向けの2D-3D変換技術は存在していなかった。

20

【0003】

2D-3D変換では、画像または画像のシーケンスを、それらに対応する深度マップを用いて増強している。この情報を利用すればステレオ画像のペアの合成が可能である。深度マップは、動きから構造へ、またはデフォーカスから深度へ、などの従来のコンピュータビジョンの方法を使用して計算することができる。Rzeszutekら [19] は、動きに基づいて背景の深度を推定している。Zhangら [26] は、動きやデフォーカスなどの複数の深度表現に基づく半自動的な2D-3D変換システムを提案している。自動的な2D-3D変換技術および深度表現に関する調査は [25] で見ることができる。また、与えられたシーンの内部の深度分布に関しては強引な前提がなされることが多い。例えばKoら [12] は、映像をロングと非ロングに分類し、ロングの映像は視野が大きいことを前提とし、深度勾配を画像全体に割り当て、選手には一定の深度を割り当てている。同じくSchnyderら [21] も、選手を検出し、一定の深度を選手に割り当てている。しかし、この方法では、ステレオ映像の中で見るとオブジェクトに厚みが感じられない「書き割り効果」というよく知られた現象が生じる。

30

40

【0004】

データ駆動型方法は、深度マップとそれに対応するステレオ画像を合成するための代替的手段となる。Hoiemら [10] は、一つのシーンを平面領域にセグメント化し、各領域に方位を割り当てている。この方法では比較的粗い深度推定が可能である。Konradら [13] は、画像と深度マップのペアのデータベースに基づいて入力画像の深度を推定している。その研究は静止画像用に計画されたものであり、勾配ベースの特徴が類似している画像は深度も類似している傾向があるという前提に基づく。クエリー画像に対して最も類似した画像がデータベースで検索される。クエリー画像の深度は、検索された画像の深度の中央値として推定される。Karschら [11] はこの方法を画像シーケンスに拡張している。彼ら

50

も画像と深度マップのペアの大容量のデータベースを使用している。クエリーフレームに対して最も類似している画像がデータベースで検索され、検索された画像がクエリー画像にワープされる。そして最後に、ワープされた深度マップが組み合わされて最終的な深度が推定される。[11]の研究が本発明者の研究に最も近いと、本発明者はその研究を比較の対象とする。

【0005】

自動2D-3D変換ができるスタンドアロンボックス（例えば、JVC社のIF-2D3D1ステレオ画像プロセッサである3D Bee）またはソフトウェアパッケージ（例えば、DDD社のTriDef 3D）[25]製品がいくつか市販されている。しかしシステム詳細は広まっておらず、深度品質について課題が残されている。

10

【0006】

以下に示す従来技術は、ある面では本発明との関連性が認められるが、ある面では大きな相違がある。

【0007】

特許文献番号US 2013/0147911 A1、発明者：Karschら、発行日：2013年6月

US 2013/0147911の方法では、データベースからクエリー画像に最も類似した画像（候補）を選び出す。その候補をクエリー画像にワープし、深度と融合してクエリーの深度を推定する。この方法では、局所探索（ブロックマッチング）は行わず、深度勾配に依拠せず、ポアソン方程式を用いた深度勾配からの深度再構築も行っていない。本発明者の方法は以下の点でこの従来技術と区別される。

20

- ・ 合成3Dデータベース（2D画像と深度情報のペアのデータベース）を使用する。
- ・ 候補に対して局所探索（ブロックマッチング）を行う。すなわち、クエリー内のブロックごとに、候補画像内のすべてのブロック（部分）を探索して最もマッチするブロックを見つける。
- ・ 候補の絶対深度ではなく深度の空間勾配をクエリーにコピーする。
- ・ ポアソン方程式を用いてクエリーの深度勾配からクエリーの深度マップを再構築する。
- ・ オブジェクト境界を描画し、オブジェクト境界に対するポアソン方程式をカットすることにより深度不連続性を許容する。

30

【0008】

特許文献番号US 2015/0093017 A1、発明者：Hefeedaら、発行日：2015年4月

US 2015/0093017は、入力と出力が異なる、まったく別のシステムである。主な相違点は以下である。

- ・ 入力は3Dビデオであり（私たちが提案するシステムでは入力は2Dビデオ）であり、出力はそのビデオに対する固有のシグネチャである（そのビデオの3Dバージョンを出力とする本発明者のシステムとは異なる）。
- ・ このシステムでは画像の左側と右側のピクセルをマッチングするツールとしてSIFTを使用するが、本発明者はクエリー内の各ブロックと最もマッチするブロックをデータベースから検索し、検索したブロックの深度勾配をコピーするためのツールとしてSIFTを使用する。
- ・ このシステムは目的が異なるため深度マップの推定は行わず、したがって視覚探索、局所探索（ブロックマッチング）、勾配マッピング、境界カット、およびポアソン再構築の技術は使用していない。

40

【0009】

特許文献番号US 8,953,905 B2、発明者：Sandrewら、発行日：2015年2月

US 8,953,905 B2の方法は、本発明者の完全自動化された方法と比較すれば半自動的な方法である。この発明の諸側面は、「現在の多くの動画にはコンピュータで生成された要素が含まれ、それらの要素には例えばロボットや宇宙船のような現実には存在しないオブ

50

ジェクトや、動画に効果として加えられる埃、霧、雲などのオブジェクトがある」ことを前提としている。ここでいうオブジェクトとは、その深度が自動的に推定されるオブジェクトのことであり、それに関して従来技術では次のように記載されている。「本発明の実施形態は、コンピュータで生成された要素に関連付けられた任意の種類のデータファイルをインポートして、コンピュータで生成された要素に関連付けられた画像の部分に対して即時に深度値を与えるものである。」「コンピュータ生成オブジェクト以外のすべてのオブジェクトに対しては意図的に深度を調整する。」

この従来技術と本発明者の方法との主な相違点は以下である。

- ・ この従来技術は、本発明者の方法とは異なり、コンピュータで生成されていないオブジェクトの深度をデータベースから自動的に推定することができない。スポーツビデオのほとんどのシーンがコンピュータで生成されていないことを考慮すれば、この発明はスポーツビデオには適していない。

10

- ・ この従来技術ではコンピュータで生成されていないオブジェクトの他にキーフレームのオブジェクトマスクもインタフェースツールを使用して手作業で調整しているが、本発明者の方法ではオブジェクト境界は自動的に描画される。

- ・ この従来技術では視覚探索、局所探索（ブロックマッチング）、勾配マッピング、および深度再構築は使用されていないが、それらの技術は本発明者の方法の中心部分を構成している。

【 0 0 1 0 】

Calagari, Kianaら、「Anahita: A System for 3D Video Streaming with Depth Customization（深度をカスタマイズできる3Dビデオストリーミングのシステム）」、Proceedings of the ACM International Conference on Multimedia. ACM, 2014.

20

Calagariのシステムの目的および入力／出力は本発明者の方法とはまったく異なる。主な相違点は以下である。

- ・ このシステムの主要目的は3Dビデオの画質補正であるが、私たちが提案するシステムの目的は3Dビデオの生成である。このシステムの入力は3Dビデオであるが、本発明者の提案するシステムの入力は2Dビデオである。

- ・ 3Dデータベースが使用されない。

- ・ ビデオがすでに3Dであるため、このシステムでは深度の推定が行われない。したがって、視覚探索、局所探索（ブロックマッチング）、勾配マッピング、境界カット、およびポアソン再構築（深度勾配からの深度の再構築）の技術は使用されない。

30

【 0 0 1 1 】

Corrigan, Davidら、「A video database for the development of stereo-3D post-production algorithms（ステレオ3Dポストプロダクションアルゴリズムの開発のためのビデオデータベース）」、Visual Media Production (CVMP), 2010 Conference on. IEEE, 2010

Corriganの研究の目的は、通常の製作時に生成される映像の代表例となるステレオ3Dビデオのデータベースを提供して、色のアンバランス、ステレオペアの調整、深度の編集などの3Dポストプロダクションに関わる技術的難題への研究者の理解を深めることにある。本発明者の方法との主要な相違点は以下である。

40

- ・ この研究は3Dデータベースのみを提供し、変換方法は提供しない。

- ・ この3Dデータベースは3Dで撮影されたビデオの画質補正を行うことを目的とするが、本発明者の方法はデータベースを使用して、2Dで撮影されたビデオを3Dに変換することを目的とする。

- ・ 本発明者のデータベースとは異なり、このデータベースは合成ではないため、高品質の深度マップを利用できない。

【 0 0 1 2 】

Dominic, Jean MariaとJ. K. Arsha、「Automatic 2D-to-3D Image and Video Conversion by Learning Examples and Dual Edge-Confined Inpainting（実際例の学習および二重エッジ限定インペインティングによる画像およびビデオの自動2D-3D変換）」、Internat

50

tional Journal of Advanced Research in Computer Science and Software Engineering (2014)

Dominicの方法と本発明者の技術との主な相違点は以下である。

- ・ ここで使用されるデータベースは合成ではない。
- ・ この方法では、データベース画像の深度勾配ではなくデータベース画像の絶対深度を使用している。

- ・ この方法では、候補画像を検索した後、局所探索（ブロックマッチング）、勾配マッピング、境界カット、またはポアソン再構築を行わず、候補の深度マップの中央値をクエリーの推定深度として使用するだけである。

【 0 0 1 3 】

10

Kiana Calagari, 「2D to 3D Conversion Using 3D Database For Football Scenes (サッカーシーン向け3Dデータベースを使用した2D-3D変換)」、July 2013

Kianaの方法は前述のDominicの方法と類似している。Kianaが提示した技術と本発明者の技術との主な相違点は以下である。

- ・ この方法で使用されるデータベースは合成ではない。また、深度勾配のデータベースではなく2D+深度画像のデータベースである。

- ・ この方法では、データベース画像の深度勾配ではなくデータベース画像の絶対深度を使用している。今後の研究の一案として深度勾配を使用する必要性は提起されているが、その内容は説明されていない。

- ・ この方法では、候補画像を検索した後、局所探索（ブロックマッチング）、勾配マッピング、境界カット、またはポアソン再構築（深度勾配からの深度マップの再構築）は行わず、SIFTフローを使用して候補をワープし、ワープされた候補の深度マップの中央値をクエリーの推定深度として使用している。SIFTフローはSIFTを基本ツールとして使用するワープ方法だが、本発明者の方法とはやり方も目的も異なっていることに留意すべきである。SIFTフローでは、SIFTを使用して、SIFTで記述されたフローに基づいて各ピクセルを移動することによりRGB画像を別のRGB画像にワープしている。しかし、本発明者の方法では、SIFTを使用してクエリー内のブロックごとに最もマッチするブロックを探し、見つかったブロックの深度マップの勾配をクエリー内の対応するブロックにコピーしている。

20

【 0 0 1 4 】

30

Zhang, Chenxi, et al. "Personal photograph enhancement using internet photo collections. (インターネットの写真集を利用した私的写真の画像補正)" Visualization and Computer Graphics, IEEE Transactions on 20.2 (2014): 262-275.

Zhangの研究では、まったく同じ場所を写した非常に多くの写真をインターネットで入手できる、大都市や観光地の画像に特に注目している。彼らはこの膨大なインターネット写真集（IPC）を利用して多くの画像補正技術を実施している。これらの補正技術の一つが2D画像から3D画像への変換である。この研究と本発明者の方法との主な相違点は以下である。

- ・ この研究で、彼らは最初に前景／背景のセグメント化を行い、背景に深度を割り当てるだけのためにIPCデータベースを使用している。前景には手作業で深度を割り当てている。しかし、本発明者の方法では、前景と背景の両方の深度が自動的に推定される。

40

- ・ この方法は、本発明者の方法とは異なり、まったく同じ場所の写真が記憶されたデータベースを必要とするため、視覚的に類似しているだけの画像を使用して実施することはできない。

- ・ 背景の深度は、まったく同じ場所の写真を使用してその場所の3Dモデルを生成することにより推定される。彼らの2D-3D変換技術では、局所探索（ブロックマッチング）、勾配マッピング、境界カット、またはポアソン再構築（深度勾配からの深度マップの再構築）は行われず、ポアソン方程式は、画像の深度マップではなく画像自体を再構築する目的で彼らの他の測光補正技術（2D-3D変換ではない）で使用されている。

50

【発明の概要】**【発明が解決しようとする課題】****【0015】**

本発明とその実施態様は、従来技術が直面した問題を解決あるいは改善するとともに2D-3D変換の代替的な機構を提供しようとするものである。

【課題を解決するための手段】**【0016】**

本発明の一つの態様は、2Dビデオ画像を3D画像に変換するためにビデオストリームから抽出した2Dビデオ画像を処理する方法であって、

ビデオフレームの参照データベースを作成し、データベースの各エントリは2D画像およびその2D画像に対応する深度情報とで構成され、

入力ビデオフレームを前記参照データベースに投入し、

前記入力ビデオフレームを前記参照データベース内の2D画像とマッチングし、その2D画像に対応する深度情報を選択し、

前記選択した深度情報を前記マッチングされた入力ビデオフレームに適用して2D+深度情報のフレームを生成することから成る。

【0017】

本発明の別の態様では、

入力ビデオフレームを部分に分割し、前記入力ビデオフレームを前記参照データベース内の2D画像とマッチングすることは、

前記入力ビデオフレームの部分を前記参照データベース内の2D画像の部分とマッチングすることから成る。

【0018】

本発明のさらに別の態様では、前記部分は $n \times n$ 個のピクセルで構成されるブロックである。

【0019】

本発明のさらに別の態様は、前記入力ビデオフレームの別の部分を前記参照データベース内の別の2D画像の部分とマッチングして、前記入力ビデオフレームの複数の部分に対応する複数の2D画像の部分とマッチングするようにしたことから成る。

【0020】

本発明のさらに別の態様では、前記選択した深度情報を前記マッチングされた入力ビデオフレームに適用することは、前記マッチングされた2D画像の部分の深度情報を、対応する前記マッチングされた入力ビデオフレームの部分に適用することから成る。

【0021】

本発明のさらに別の態様では、前記選択した深度情報を前記マッチングされた入力ビデオフレームに適用することは、前記マッチングされた2D画像の部分の対応するピクセルを1個以上、前記マッチングされた入力ビデオフレームの部分の対応するピクセルにマッピングすることから成る。

【0022】

本発明のさらに別の態様では、前記入力ビデオフレームとマッチングする2D画像の候補を視覚的技術を用いて特定する。

【0023】

本発明のさらに別の態様では、視覚的技術は前記フレームのGISTおよびカラー情報を使用することから成る。

【0024】

本発明のさらに別の態様では、深度情報は深度勾配である。

【0025】

本発明のさらに別の態様は、

前記入力ビデオフレーム内のオブジェクトを特定し、

前記特定したオブジェクトのオブジェクトマスクを決定し、

前記決定したオブジェクトマスクおよび前記マッチングされた入力ビデオフレームを用いて深度情報を推定することから成る。

【0026】

本発明のさらに別の態様は、前記決定した深度情報をポアソン再構築を用いて推定することから成る。

【0027】

本発明のさらに別の態様では、前記ポアソン再構築は一次および高次導関数から成る。

【0028】

本発明のさらに別の態様は、2D + 深度情報フレームを用いて左側ステレオ画像および右側ステレオ画像を生成することから成る。

【0029】

本発明のさらに別の態様では、ソフトウェアで生成されたビデオフレームを用いて前記参照データベースを作成する。

【0030】

本発明のさらに別の態様では、前記ソフトウェアはビデオゲームである。

【0031】

本発明のさらに別の態様では、2Dビデオ画像を3D画像に変換するためにビデオストリームから抽出した2Dビデオ画像を処理するシステムであって、前記システムは、

各エントリが2D画像およびその2D画像の対応する深度情報とで構成される、ビデオフレームの参照データベースと、

入力ビデオフレームを参照データベースに投入するように作動する探索モジュールと、

入力ビデオフレームを参照データベース内の2D画像とマッチングし、その2D画像に対応する深度情報を選択するように作動するマッチングモジュールと、

選択した深度情報をマッチングされた入力ビデオフレームに適用して2D + 深度情報のフレームを生成するように作動する生成モジュールとから成る。

【0032】

本発明のさらに別の態様では、実行時にビデオストリームから抽出された2Dビデオ画像を3D画像に変換する命令でプログラミングされた、コンピュータで読み取り可能な媒体を備え、命令は、

ビデオフレームの参照データベースを作成し、データベースの各エントリは2D画像およびその2D画像に対応する深度情報とで構成され、

入力ビデオフレームを参照データベースに投入し、

入力ビデオフレームを参照データベース内の2D画像とマッチングし、その2D画像に対応する深度情報を選択し、

選択した深度情報を前記マッチングされた入力ビデオフレームに適用して2D + 深度情報のフレームを生成することから成る。

【0033】

本発明のさらに別の態様では、参照データベースを生成する方法は、ソフトウェアで生成されたビデオフレームを使用することから成る。

【発明の効果】

【0034】

本発明によれば、コンピュータで生成されたコンテンツを使用して、2D-3D変換のための高品質な3Dビデオ参照データベースを作成することができる。

【図面の簡単な説明】

【0035】

【図1】一つの実施形態の略図である。

【図2 - 1】深度推定を使用することによる効果として、(a) クエリー画像、(b) そのK候補のサブセット、(c) 生成されたマッチした画像、を示す。

【図2 - 2】深度推定を使用することによる効果として、(d) オブジェクト境界カット

10

20

30

40

50

、(e) ポアソン再構築を用いた深度推定、(f) 勾配精細化とポアソン再構築、(g) オブジェクト境界がカットされたオブジェクト、(h) 平滑化された最終的な深度推定、および (i) 黄色でマークされたブロック h の拡大および増幅されたバージョン、を示す。

【図 3】一つの実施形態の略図である。

【図 4】行列 A のポアソン方程式の略図であり、(a) はサンプルピクセル p とそれに隣接するピクセルを示す 4×4 ピクセルの画像の例、(b) はピクセル p に対する式 (3) の係数、(c) は行列 A におけるピクセル p に対応する行の非ゼロ値、を示す。

【図 5】上段は合成シーケンスのフレーム 3、下段は実写シーケンスのフレーム 24 を示す。ここに示す深度は、グラウンドトゥールス / ステレオ対応点探索 [8]、DT、DT+、および DGC を使用して抽出したものである。どちらのシーケンスでも本発明者の技術である DGC がグラウンドトゥールス / ステレオ対応点探索を最も良く再現している。

【図 6】合成サッカーシーケンスに対して、本発明者の方法である DGC、文献の中ではそれに最も近い方法である DT、および DT の拡張版である DT+ を適用した結果の客観的な比較である。

【図 7】本発明者の方法を使用して様々なサッカーシーケンスに対して行った深度推定を示す。本発明者の方法は、クローズアップ（例：上段左端）、ミディアムショット（例：下段左端）、鳥瞰図（下段右端）、およびロングショット（例：上段右端）を含めた多様な映像に対応する。

【図 8】様々な種類のサッカーシーンの奥行感覚および視覚的快適性の平均オピニオン評点である。

【図 9】サッカー以外の各種の野外スポーツシーンの奥行感覚および視覚的快適性の平均オピニオン評点である。

【図 10】（左から）DT、DT+、および本発明者の方法である DGC を使用して、様々なシーケンスに対して行った深度推定の結果である。DT では誤りのある推定が生成され、DT+ では測定結果にノイズがあって選手が検出されない。本発明者の方法は他の二つの方法よりも優れた結果を生成している。

【図 11】本発明者の方法で変換したシーケンスとオリジナル 3D との間の差分平均オピニオン評点（DMOS）である。DMOS がゼロであれば、本発明者の変換したシーケンスがオリジナル 3D と同じであることを意味する。

【図 12】本発明者の方法で変換したシーケンスと、深度転送法 DT+ で変換したシーケンスとの間の差分平均オピニオン評点（DMOS）である。正の DMOS は、本発明者の方法が DT+ よりも好ましいことを意味する。

【発明を実施するための形態】

【0036】

序論

立体 3D（S3D）動画は普及が進み、ほとんどの大手製作会社からこの形式の動画がリリースされている。しかし、ほとんどの動画が 2D で撮影されているため、手作業で深度マップをペイントし、対応するビューを描画（レンダリング）することにより動画を S3D にアップコンバートしているのが現状である。このプロセスでは非常に優れた結果が生み出されるが、膨大な費用と時間が必要である。ライブイベントの S3D を製作する場合はさらに多くの困難が伴う。手作業によるアップコンバートは不可能である。サッカーの試合などのライブイベントを直接にステレオ撮影するにはスタジアムに複数のステレオ装置を設置する必要があるが、それには大変な労力が要求されるため実際に試みられることは滅多にない。そのため、ライブイベントで自動的に高品質の 2D-3D 変換ができる方法が強く望まれている。現在の自動変換方法は性能が高くはない。ほとんどの方法は汎用であり、どの種類のビデオストリームにも適用できるようになっている。しかし、その出力はアーチファクトで損なわれているため多くの視聴者から受け入れられないか、または、アップコンバートの方法がきわめて旧式なため、変換後のビデオにはわずかな奥行感しか与えられない。

【0037】

本発明者が提案するのは、高品質な自動2D-3D変換を開発する方法である。この取り組みは、汎用的な方法を取らない、領域特化型のアップコンバートを開発することである。特に、S3Dのサッカービデオを生成する方法を提案する。本発明者の方法はデータ駆動型であり、S3Dビデオの参照データベースに依拠する。この方法は従来の研究 [13、11] に類似しているが、本発明者の鍵となる考え方は、コンピュータビジョン法を用いて計算した深度データや深度センサーで取得した深度データに依拠するのではなく、最新のコンピュータスポーツゲームからコンピュータで生成された深度データを抽出し、合成3Dデータベースを作成できることに着目している。最新のコンピュータゲームのビデオ品質は実写のそれに近づきつつあるため、本発明の方法には二つの利点がある。すなわち、1) 多様性のあるビデオフレームのデータベースを入手して、入力ビデオフレームとのマッチング度を高めることができる、そして、2) ビデオフレームごとに、完璧な深度不連続性を備えた正確な深度マップが得られる、ということである。クエリー画像が与えられれば、データベース内の類似画像およびそれらの深度マップに基づいて、その深度が推定される。本発明者が提案するのは、深度勾配（すなわち、x方向とy方向の深度値の変化率）を合成3Dデータベース内の類似画像からクエリー画像に転送する方法である。本発明の一つの態様は、クエリーをブロック（部分）に分割し、マッチングするブロック（部分）（合成3Dデータベース内の異なるフレームに属することもあり得る）から深度勾配を転送する。これは、フレーム全体に絶対深度を使用する従来のアプローチ [13、11] とはまったく異なる。本発明者の方法にはいくつもの利点がある。すなわち、(i) 小さい領域 / オブジェクト（例えば、選手）に対してより精細な深度の割り当てができる。(ii) データベースを大幅に縮小できる。これはフレーム全体（フレームにはきわめて多様な情報が含まれることがある）ではなく小さいパッチ（部分）に対してのみマッチングを行うからである。(iii) 参照として選ばれた類似画像の（不）正確さに対する高い堅牢性を持つ。これは、深度計算に個々のブロック（部分）だけが使用されるからである。深度勾配の転送後、ポアソン再構築を用いて深度の再生を

行う。ポアソン再構築とは、ポアソン方程式を解くことにより画像の深度勾配からその画像を再生する、従来から利用されている堅牢な技術である [18、7]。選手の境界は明瞭であることが望ましいが、本発明者の方法では、オブジェクトマスクを作成し、オブジェクト境界を検出することにより深度不連続性に対処している。本発明者の提案する技術により、カメラの様々な視野、オクルージョン、クローズアップ、散乱、複雑な動きが含まれた広範囲なサッカービデオショットを処理することが可能である。

【0038】

本発明者は、多種多様なビデオセグメントを使用して広範囲なユーザー調査を実施した。この調査を行うにあたってITU BT.2021の勧告 [6] に従っている。その結果は次のとおりである。(i) 本発明者の方法により、最初からステレオ撮影されたビデオとほとんど区別できない3Dビデオを生成できる。(ii) 本発明者の方法で生成したビデオは、調査対象者から奥行感覚の質および視覚的快適性の面で「非常に良い (Excellent)」と評価された。(iii) 本発明者の方法の性能は最先端の方法をかなり上回っている。

【0039】

本発明の理解を容易にするために、本発明の実施例を、添付図面を参照しながら以下で説明する。

【0040】

図3に、本発明者のサッカービデオ2D-3D変換システムの概要を示す。本発明者の技術では、合成して生成された深度のデータベースから深度を推定する。このデータベースはビデオゲームから収集されたものであり、高品質の深度マップを提供する。本発明者の方法では、データベースから深度勾配を転送し、ポアソン再構築を用いて深度の再構築を行う。オブジェクト境界の明瞭さと正確さを維持するために、オブジェクトマスクを生成し、オブジェクト境界に対するポアソン方程式を修正する。最後に、[11] のステレオワーブ技術を使用して、2Dフレームとそれらの推定深度を用いて左右のステレオペアをレンダリングする。この技術では、顕著な領域には変更を加えずに背景領域を広げて非オクルージ

ョン領域が埋められるように、推定深度に基づいて2Dフレームをワープする。このセクションでは、本発明者が使用する合成3Dデータベースおよびオブジェクトマスクの生成について論じる。セクション4では本発明者の深度推定技術について論じる。

【0041】

合成3Dデータベース：RGBD（赤、緑、青、および深度）[2, 1, 5] やビデオ [11, 3] のデータベースが数多く作成されている。深度チャンネルは、タイムオブフライト方式の画像処理 [20] またはアクティブステレオ（Microsoft Kinect等を使用）により取得されている。最新のRGBDデータベースでも、スポーツイベントの高画質2D-3D変換に使用できるものはない。スポーツイベントの場合、きわめて動きの多い環境において日光が当たっている状態で深度を取得する必要があるため、深度マップを取得することは難しい。

10

【0042】

この難題に対処するために、本発明者はビデオゲームから合成RGBD（SRGBD）データベースを生成することを提案している。ビデオゲームには非常に高画質の画像が含まれ、これらの画像から大量のコンテンツを容易に生成することができる。このようなデータベースはデータ駆動型の2D-3D変換に利用できる。本発明者は、合成データベース学習を経た推定技術で成果を上げているMicrosoft Kinectの姿勢推定からヒントを得ている [22]。本発明者の場合、ビデオゲームFIFA13から画像および深度情報を抽出することによりS-RGBDデータを収集している。Microsoft DirectXツールであるPIX [4] を使用した。PIXはアプリケーションから呼び出されるすべてのDirectXコマンドを記録する。これらのコマンドを再実行することで、PIXは記録された各フレームをレンダリングし、保存することができる。また、PIXではレンダリングされた各フレームの深度バッファにアクセスできるようになっている。抽出した各フレームの解像度は10 fpsで1916 x 1054である。本発明者は40の様々なシーケンスから16,500個の2D+深度フレームを抽出した。これらのシーケンスには、サッカーの試合でよく見られる、広範囲のカメラビュー、動きの複雑さ、およびカラーバリエーションを伴った多様な映像が含まれている。40シーケンス中の2つはハーフタイムがそれぞれ6~7分の長さで収められ、試合全体を通じてよく見られるシーンを映し出すよう設計されている。その他のシーケンスはそれよりも短く15~60秒の範囲だが、クローズアップ、ゴール裏から見たシーン、グラウンド上の場面のズームなど、あまり一般的ではない事象を映し出すことの方に重点が置かれている。本発明者のデータベースには、様々なチーム、スタジアム、シーズン、およびカメラアングルが含まれている。

20

30

【0043】

オブジェクトマスクの生成：本発明者の方法では、深度不連続性をより巧みに処理し、選手境界を明瞭にするために、オブジェクトマスクを生成することによりオブジェクト境界を描画している。オブジェクト境界を明確にしない場合、選手の深度がグラウンドと混合し、そのために深度の品質が劣化する。これらのマスクを生成するために、本発明者の方法では、動きと外観に基づいて各ビデオシーケンスを前処理することによりオブジェクトを自動的に検出している。スペースに限りがあるため、このステップについては簡単に説明する。本発明者は二種類のオブジェクト検出方法を提案している。一つはクローズアップであり、選手のサイズを大きくし、競技領域を小さくすることを特徴とする。もう一つは非クローズアップであり、視野を広くすることを特徴とする。非クローズアップのビデオセグメント化は、競技フィールドのカラーなどの大域的特徴に依拠している。これらの映像に対して、本発明者の方法ではカラーに基づくアプローチを用いて競技フィールドを検出している。競技フィールドから収集したサンプルにガウス混合モデル（GMM）で学習する。クローズアップでは、特徴点軌跡 [16] のような局所的な特徴に依拠するところが大きい。本発明者の方法では、特徴点軌跡のセグメント化で初期化を行うマット化に基づくアプローチ [14] を採用している。さらに、競技領域検出を用いて、競技フィールドの分類誤りがあれば修正している。

40

本発明者のシステムの中心部は深度勾配からの深度推定であり、入力2Dビデオに対して、本発明者の生成したS-RGBDデータベースに基づいて深度を推定する。図1はこの処理の概要を示している。被検査2Dフレームに対して、本発明者の生成したデータベース内のKが

50

最も近いフレームを検索する。被検査フレームに対してマッチングする画像を生成する。このマッチングする画像はブロックごとに生成するが、その場合、被検査フレーム内の各ブロックに対して、K候補画像内の最もマッチするブロックを検索する。次に、マッチしたブロック（部分）から被検査フレームに深度勾配をコピーする。最後に、ポアソン方程式を解くことにより、コピーされた深度勾配から深度を再構築する。本発明者の方法では、オブジェクトマスク（セクション3を参照）を使用してオブジェクト境界の周囲の深度不連続性を明瞭にしている。以下で各ステップをより詳細に説明する。

【0044】

被検査ビデオの各フレームに対して、本発明者のS-RGBDデータベースで視覚探索を行ってK（本発明者の研究では10）が最も近いフレームを特定する。視覚探索では2つの主要な特徴であるがGIST [17] とカラーが使用される。前者は類似構造全体とのマッチングに力点を置き、後者は類似カラー全体とのマッチングに力点を置く。カラーについては、色相値の正規化ヒストグラムを使用し、それに対して値0.1で閾値処理による2値化を行ってドミナントカラーのみを表現する。最終的な画像探索記述子はGISTおよびカラーヒストグラムを連結したものである。図2(b)に、図2(a)のフレームに対して生成されたK候補の4つのサンプルを示す。

【0045】

本発明者は、K候補画像を使用して被検査フレームに類似した画像を作成し、それをマッチした画像と呼んでいる。マッチした画像を使用することで候補と被検査フレームとの間のマッピングが可能になり、被検査フレーム内の各ピクセルが候補の対応するピクセルにマッピングされる。Karschら [11]は、このようなマッピングを行うために大域的なアプローチを用いている。彼らは候補をワープして被検査フレームに類似した画像を作成している。このアプローチは局所的な画像アーチファクトに対する堅牢性はあるが、被検査フレームとデータベースとの間の強い類似性が要求される。たとえば、被検査フレームに4人のプレーヤーが含まれている場合、データベースにはそれに類似の内容を持つ画像が含まれている必要がある。それに対して、本発明者は局所的なアプローチを用い、ブロック単位のマッチングを行うことにより類似画像を作成している。そうすることでマッチングの堅牢性を高めることができる。たとえば、撮影アングルや選手の人数や場所が異なっても、2つのフレーム間で良好なマッチングを行うことができる。図2の例はそのことを示している。この例では、図2(b)の画像を使用して高画質のマッチした画像（図2(c)）を作成したが、[11] の大域的なアプローチを使用していればそれと同等の画質の画像は作成できなかったかもしれない。本発明者の局所的なアプローチは大容量のデータベースを必要とせずに良好な深度推定を実現できるが、セクション3で論じたように正確なデータベースを作成することは難しく、このことは本発明者の方法にとってきわめて望ましい利点となる。

【0046】

マッチングする画像を作成するために、最初に被検査フレームを $n \times n$ 個のブロック（部分）に分割する。本発明者の実験では常に n を9に設定している。被検査フレームの各ブロックを、K候補画像内の可能なすべてのブロックと比較する。ユークリッド距離が最も小さいブロックを対応ブロックとして選択する。候補画像のサイズを被検査フレームのサイズに変更する。ブロック記述子として、そのブロックの平均RGB値で連結されたSIFTを使用する。SIFT記述子は、ブロックの中心に位置している、それよりもサイズが大きい $5n \times 5n$ のパッチに対して計算される。この計算によってより代表的なテクスチャをキャプチャできる。RGB値は0~1の範囲で正規化する。図2(c)に、本発明者のブロックマッチングのアプローチを用いて得られたマッチした画像を示す。垂直な広告板はすべて垂直なブロック（部分）にマッチし、水平な競技フィールドは水平な競技フィールドにマッチし、傾斜した観衆も観衆にマッチしていることに留意されたい。

【0047】

深度勾配の計算：入力フレームおよびS-RGBDでそれにマッチした画像を所与として、それに対応する深度勾配をコピーする。水平方向および垂直方向の一次空間微分係数 (G_x, G

10

20

30

40

50

y) をコピーする。画像マッチングと同様に、 $n \times n$ ピクセルで構成されるブロック（部分）の中の対応するブロック（部分）から深度勾配をコピーする。

【 0 0 4 8 】

ポアソン再構築：次のポアソン方程式を用いて、コピーされた深度勾配から深度値を再構築する。

【 0 0 4 9 】

【数 1】

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) D = \nabla \cdot G,$$

10

【 0 0 5 0 】

ここで、 $G = (G_x, G_y)$ はコピーされた深度勾配、 D は推定する深度である。 G は G の発散である。

【 0 0 5 1 】

【数 2】

$$\nabla \cdot G = \left(\frac{\partial G_x}{\partial x} + \frac{\partial G_y}{\partial y} \right).$$

20

【 0 0 5 2 】

離散領域では、式(1)および式(2)はそれぞれ式(3)および式(4)になる。

【数 3】

$$D(i, j+1) + D(i, j-1) - 4D(i, j) + D(i+1, j) + D(i-1, j) = \nabla \cdot G(i, j).$$

30

【数 4】

$$\nabla \cdot G(i, j) = G_x(i, j) - G_x(i, j-1) + G_y(i, j) - G_y(i-1, j).$$

40

【 0 0 5 3 】

解は $Ax = b$ の形式で示す。ここで、 $b = G$ 、 $x = D$ であり、 A にはポアソン方程式（式(3)）の係数が記憶されている。サイズ $H \times W$ の被検査画像に対して、 A はサイズ $HW \times HW$ の正方行列であり、その各行は被検査フレームのピクセルに対応している。この行の値は式(3)の係数に対応している。図4(a)は、小さいサンプル画像に対する A のセットアップを示している。1個以上の隣接ピクセルが存在していないため、画像境界のピクセルに特別な注意を要することに留意されたい。この場合、存在しないピクセルを参照している式(4)の項を削除することにより G の値を更新する。最後に、 $Ax = b$ を所与として x の値を求める。図2(e)は、再構築された深度(x)の一例を示す。

【 0 0 5 4 】

50

深度構造全体がキャプチャされているが、若干のアーチファクトが発生している（図2(e)の右下隅を参照）。

【0055】

このようなアーチファクトはSIFTマッチングが不正確であるために発生することが多い。たとえば、図4(c)では、いくつかのフィールドブロック（部分）が非フィールドブロックにマッチしている。深度の変化が滑らかであることが予想される領域から抽出されたクエリーブロックが、深度の急激な変化が含まれる参照ブロック（ゴールまたは選手境界など）に誤ってマッチした場合、参照ブロックから転送される急激な深度勾配は、その結果得られる深度に小さいアーチファクトを発生させることがある。この問題を解決するために、 x の値を求める前に、深度勾配を精細化することにより転送元の大きな深度勾配を最初に縮小しておき、代わりに本発明者が作成したオブジェクトマスクを使用して深度不連続性を適切な場所に置く。この2つのステップについて以下で説明する。

10

【0056】

勾配の精細化：不正確なブロックマッチングのために発生する誤りを減少させるために、次の式を用いて深度勾配を精細化する。

【0057】

【数5】

$$G_x = G_x \times \max\left(1 - e^{\left(1 - \frac{1}{\alpha |G_x|}\right)}, 0\right)$$

$$G_y = G_y \times \max\left(1 - e^{\left(1 - \frac{1}{\alpha |G_y|}\right)}, 0\right)$$

20

【0058】

これにより小さい勾配をそのまま維持しながら、誤って推定される可能性のある大きな勾配が指数関数的に縮小する。 α は精細化の強度を設定するパラメータである。 α を大きくすると正確な勾配が破損することがあり、 α を小さくするとアーチファクトを発生させることがある。本発明者実験では常に α を60に設定している。図2(f)に、図2(a)に対する深度推定において勾配の精細化を行った結果を示す。図2(e)と比較して、アーチファクトが除去され、深度の変化がより滑らかになっている。

30

【0059】

オブジェクト境界のカット：ポアソン再構築は、一つのピクセルをそれに隣接するすべてのピクセルに連結する。これによりほとんどのオブジェクト境界は色が薄れるため、急激な勾配が除去される勾配の精細化の後では特にその影響が顕著に現れる（図2(f)を参照）。本発明者はこの問題を解決するために、その場合のポアソン方程式を修正することにより、オブジェクト境界上の深度不連続性を許容している。オブジェクトマスクごとに、Canny法を用いてエッジを検出する（図2(d)を参照）。次に、オブジェクト境界のピクセルを有効な隣接ピクセルとして使用することを許容しないことにより、ピクセルをオブジェクト境界から切り離す。境界ピクセルに隣接する各ピクセルに対して、Aの中の対応する連結を0に設定し、それに応じて G の値を更新する。したがって、オブジェクト境界に隣接するピクセルは画像境界のピクセルと同様に処理される。

40

【0060】

ピクセルまたはピクセルのグループが画像の残り部分から完全に切り離された場合、ポアソン再構築の結果に誤りが生じることに留意する必要がある。それが原因で、切り離された領域が黒くなること、および/または画像全体の深度推定に影響することがある。したがって、オブジェクト境界のピクセルと画像の残り部分との連結を維持しつつ、しかし境界を挟んだ両側を切り離しておくことが重要となる。それをするために、本発明者は境界の各ピクセルを境界の最上部または最下部のピクセルに連結している。クエリー画像においてある境界ピクセルが最上部のピクセルの方により類似している場合は最上部のピク

50

セルに連結し、そうでない場合は最下部のピクセルに連結する。その結果、境界の各ピクセルは境界の上部または下部領域の一部となるが、それら二つの領域は互いに接近できなくなる。本発明者は、セグメント化の誤りに起因してオブジェクトマスクにしばしば穴が生じることに注目した。かかるマスクにエッジ検出を行うと、これらの穴が画像の残り部分から分離される。この問題を避けるために、エッジ検出の前にこの穴を埋めておく。ただし、オブジェクト自体にエッジ検出を行うと、オブジェクトが境界ピクセルに囲まれ、そのため背景から分離されることに注意すべきである。この問題を解決するために、本発明者は各オブジェクト境界を最下部（すなわち、選手の足）から開いている。それによりポアソンをグラウンドからオブジェクトに拡散させることができるため、自然な深度を生成しながら分離を回避することができる。図2(d)に、図2(a)に対して生成されたオブジェクト境界を示す。図2(g)に、ポアソン再構築時にオブジェクト境界がカットされたときの推定深度を示す。図2(f)と比較して、図2(g)では選手が視認しやすくなっている。

10

【 0 0 6 1 】

平滑化：本発明者は、高次深度微分係数を強制的にゼロにすることにより、ポアソン再構築に平滑化の制限を加えている。連続領域で次の設定を行う。

【 0 0 6 2 】

【 数 6 】

$$\left(\frac{\partial^4}{\partial x^4} + \frac{\partial^4}{\partial y^4} \right) D = 0.$$

20

【 0 0 6 3 】

離散領域では次のようになる。

【 0 0 6 4 】

【 数 7 】

$$12D(i, j) + D(i, j + 2) - 4D(i, j + 1) - 4D(i, j - 1) + D(i, j - 2) + D(i + 2, j) - 4D(i + 1, j) - 4D(i - 1, j) + D(i - 2, j) = 0.$$

30

【 0 0 6 5 】

Aの平滑化バージョンである A_s を生成する。 A_s に式(7)の新しい係数を格納する。オブジェクト境界の周囲の深度不連続性を保存するために、平滑化制限に境界カットを適用する。次に、Aを A_s と連結し、元の $Ax = b$ の代わりに次の式を解く。

【 0 0 6 6 】

【 数 8 】

$$\begin{bmatrix} A \\ \beta \cdot A_s \end{bmatrix} x = \begin{bmatrix} b \\ 0 \end{bmatrix},$$

40

【 0 0 6 7 】

は必要な平滑化の量を設定する。 を大きくすると平滑化が過剰になり、 を小さくすると平滑化が弱くなる。本発明者の実験では常に $\beta = 0.01$ と設定している。平滑化の結果は勾配の精細化の結果と異なることに留意すべきである。後者は鋭いアーチファクトを除去するが画像の残り部分を損なわないように設計されている。それに対して、平滑化

50

は深度テクスチャ全体に微妙なタッチを加える。鋭いアーチファクトを除去するために平滑化を使用すると、平滑化が過剰になることがある。また、過度に勾配細分化を行うと重要な勾配が損なわれることになる。

【0068】

最終出力の生成：推定された深度（式(8)の x ）は（0；255）の範囲で正規化され、クエリー画像と結合されて本発明者のクエリービデオの変換後の2D+深度を形成する。図2(f)に、図2(a)に対して平滑化に関わるすべてのステップを含めて推定した最終的な深度を示す。本発明者の方法で推定した深度は滑らかであり、フィールド、観客、およびプレイヤーが正確に再構築されている。本発明者の方法では、各プレイヤーに同じ深度が割り当てられる「書割り効果」が発生しないことにも留意されたい。このことを示すために、図2(h)のプレイヤーの一人から抽出した深度ブロックを拡大し、そのブロックの深度値を（0；255）の範囲に正規化することにより増幅する。図2(i)に、図2(h)の黄色でマークしたブロックを拡大および増幅したバージョンを示す。マークされたブロック内の選手は身体各部の深度値が異なっていることに留意されたい。この例は、勾配に基づいて小部分の深度を細部まで推定する本発明者の方法の効果を示している。

10

【0069】

本発明者は、図の中でDGC（Depth Gradient-based Conversionの略。深度勾配に基づく変換）と称される本発明の実施態様の評価を行った。合成シーケンスと実写シーケンスの両方を検討し、利用可能な場合はグラウンドトゥールースと比較した。また、文献[11]に示された最も類似するシステムであるDT（Depth Transferの略。深度転送法）と比較した。さらに、本発明者の技術を他の野外スポーツに適用する可能性を検討し、その結果からテニス、野球、アメリカンフットボール、およびフィールドホッケーの2D-3D変換にも有望であることが分かった。

20

【0070】

本発明者の方法には、すべてのシーケンスに対して今回に限って設定されたパラメータがいくつかあることに留意されたい。具体的には、 K （候補画像の数）が10、 n （ブロックサイズ）が9、（勾配精細化パラメータ）が60、（平滑化パラメータ）が0.01にそれぞれ設定されている。

【0071】

本発明者の2D-3D変換技術（DGC）を他のいくつかの技術と比較する。

30

【0072】

DT：独自のデータベースで学習した深度転送法（Depth Transfer）[11]。深度転送法はデータ駆動型2D-3D変換の最新技術である。そのデータベースであるMSR-V3DにはMicrosoft Kinectによりキャプチャされたビデオが収録され、オンラインで使用可能である。

【0073】

DT+：本発明者の合成データベース（参照データベース）であるS-RGBDで学習した深度転送方法。[11]で述べられているように、Kinectを用いた2D+深度のキャプチャは屋内環境に限定されている。それに加えて測定値の誤りの多さ、解像度の低さが大規模なサッカーデータベースを作成する能力を制限している。厳密な比較のために、本発明者の技術を、本発明者のサッカーデータベースで学習したデータ転送法と比較した。

40

【0074】

グラウンドトゥールース深度：グラウンドトゥールース深度マップは、セクション3で説明したようにPIX[4]を用いてビデオゲームFIFA13から抽出される。ただし、この方法は合成データにしか使用できない。

【0075】

オリジナル3D：3Dカメラでキャプチャされたオリジナルのサイド・バイ・サイド3Dビデオ。本発明者は結果の主観的な比較を行った。

【0076】

ステレオから抽出した深度：結果をオリジナル3Dの映像と客観的に比較するために、ステレオ対応点探索[8]を使用してグラウンドトゥールース深度を概算した。ステレオ対応

50

点探索の技術は必ずしも正確ではないことに留意すべきである。しかし、その技術は深度の全体構造をキャプチャするため、客観的分析には有効な場合もあることを本発明者の結果は示している。

【0077】

本発明の態様を8つの実写テストシーケンス(4つはサッカー、4つはサッカー以外)に適用した。さらに合成サッカーシーケンスのSynth(「合成」と呼ぶ)にも適用した。

【0078】

サッカー：本発明者が使用する実写サッカーシーケンスには、3D撮影のオリジナルビデオから抽出したクリップが含まれている。これらのシーケンスは、ロングショット、鳥瞰図、ミディアムショット、およびクローズアップの4つの主要カテゴリが含まれるように注意深く作成されている。ロングショットではカメラが高い位置に置かれ、フィールドのほぼ全体が視野に収められている(図7の下段右端)。ミディアムショットではそれよりも低い位置にカメラが置かれるため、それだけ視野が狭くなっている(図7の下段左端)。クローズアップでは狭い視野の中の一人ないし数人のプレーヤーにカメラをズームさせている(図7の上段左端)。

【0079】

サッカー以外：本発明者が使用するサッカー以外のシーケンスには、テニス、野球、アメリカンフットボール、およびフィールドホッケーの映像から抽出されたクリップが含まれている。これらのシーケンスを使用して、本発明者の方法を他の野外スポーツに適用する可能性を評価した。

【0080】

合成：本発明者は、S-RGBDの作成と類似した方法で120個の2D+深度合成フレームを抽出した。グラウンドトゥールース深度を所与として、この合成シーケンスを使用して本発明者の技術をDTおよびDT+と客観的に比較した。

【0081】

本発明者は、実写シーケンスと合成シーケンスの両方に対して本発明の態様を使用した客観的実験を作成し、本発明者の深度マップの品質を測定した。図5(上段)に、合成シーケンスのフレームとそのグラウンドトゥールース深度、さらに、DT、DT+、およびDGCを用いて推定したそのフレームの深度を示す。深度マップはすべて(0 - 255)の範囲で正規化されていることに留意されたい。MSR-V3Dはサッカーデータとの類似性がほとんどないため、DTは誤りの多い測定値を生成している。DT+は本発明者の作成したデータベースで学習しているため、DTよりもかなり良好な結果を生成している。とはいえ、ほとんどの選手が検出されていない。本発明者の技術であるDGCは、選手を検出し、滑らかな結果を生成し、グラウンドトゥールースに最も類似している。図6に、120個の合成フレームのグラウンドトゥールースに対する平均絶対誤差(MAE)を示す。この図から、本発明者の方法はDTおよびDT+と比較してMAEが大幅に低いことがわかる。

【0082】

グラウンドトゥールース深度がないため、実写シーケンスの客観的分析は困難である。[11]では、著者らはKinectで得られた深度をグラウンドトゥールース深度として使用している。しかし、Kinectは屋外環境では深度情報をキャプチャできないため、サッカーの試合のグラウンドトゥールースを推定することはできない。本発明者はそれとは違うアプローチをとった。すなわち、3D撮影されたサッカーシーケンスに対して、ステレオ対応点探索[8]を使用してグラウンド深度マップを概算した。次に、その深度マップを、2D-3D変換から推定された深度と比較する。図5(下段)に、最も難しいサッカーテストシーケンスの1つから抽出したフレームと、ステレオ対応点探索を用いてそこから抽出した深度を示す。完璧からはほど遠いが、深度構造全体が存在するため、これを利用して、変換された深度の品質を推測することができる。図5(下段)に、DT、DT+、およびDGCを用いて推定した深度を示す。本発明者の技術であるDGCは、グラウンドトゥールースを最も良好に再現している。これは約100個のフレームを比較して客観的に得られて評価であり、DGCはDTおよびDT

10

20

30

40

50

+と比較してMAEを最大でそれぞれ19%および86%まで減少させている。スペースに限りがあるため図は省略する。

【0083】

さらに本発明者は、合成データベース（参照データベース）のサイズの重要性を調査する実験を行った。最初に、サッカーの試合で見られる多様な映像の中から選び出した120個のフレームを使用して合成シーケンスを作成した。画像数1000、2000、4000、8000、13000、および16000の6種類のデータベースサイズについて調査を行った。その結果、画像数8000のサイズまでは、十分に大きなデータがないために性能に変動がありMAEが30前後になることが分かった。

【0084】

しかし、画像数13,000から性能が上がり、MAEは20前後に減少する。データベース内の画像数が16,000前後になると性能は安定する。そのため、本発明者が行った評価では画像数16,500のデータベースを使用した。

【0085】

本発明者は、いくつかの主観的な実験により3D視覚を評価した。本発明者の技術をDT+およびオリジナル3Dと比較した。

【0086】

セットアップ

本発明者は、3Dビデオの品質評価のための3つの主要な知覚的特質（画像品質、奥行品質、視覚的快適性（又は不快適性））を提言するITU BT.2021勧告に準拠して主観的な実験を実施した。画像品質は主に符号化または伝送により影響を受ける。奥行品質は知覚される奥行の量の尺度であり、視覚的快適性は3D知覚に起因するあらゆる形態の生理的不快感（疲労、眼精疲労、頭痛など）の尺度である。このような不快感は、3Dアーチファクト、奥行の変化、コンフォートゾーン侵害、またはクロストークに起因することが多い。本発明者の実験では奥行品質と視覚的快適性を評価した。圧縮または符号化のパラメータの変更やシーケンスの伝送は行っていないため、画像品質の評価は行っていない。

【0087】

各テストシーケンスの長さはITU勧告に従い10～15秒の範囲である。低照明条件でパッシブ偏光ガラス装備の55インチPhilips製テレビセットにシーケンスを表示した。視聴距離は、ITU勧告に従って解像度1920×1080のビデオは約2メートル、解像度1280×720のビデオは約3メートルとした。15人の被験者が主観的な実験に参加した。被験者は全員がコンピュータ科学の学生または研究者である。実験前に静的および動的ランダムドット・ステレオグラムを使用して被験者の立体視覚を検査した。実際の実験に入る前に被験者は安定化処置を受けた。予備試験では、被験者は最高から最低までの異なる3D品質を代表する4つのシーケンスを評価した。それら4つのシーケンスは実際の実験には使われていない。このステップにより被験者の予想を安定化し、評価のプロセスに慣れさせた。本発明者は被験者に、疑問点はすべてはっきりさせ、実験手順を十分に理解しておくように求めた。

【0088】

本発明者の技術の評価

本発明者が変換したシーケンスを観たときの被験者の平均的な満足度を測定することにより、本発明者の2D-3D変換の評価を行った。調査対象は4つのサッカーシーケンスと4つのサッカー以外のシーケンスである。ITU勧告の単一刺激（SS）法を使用して奥行品質および視覚的快適性を評価した。被験者はこれらのシーケンスをランダムな順序で見せられた。各シーケンスの長さは10～15秒だが、その前にシーケンスのコード名を表示したミッドグレーのフィールドが5秒間、その後に評価を行うよう被験者に求めるミッドグレーのフィールドが10秒間表示された。奥行品質および快適性の評価には標準的なITU連続尺度を使用した。連続尺度には、奥行品質のラベル（非常に良い（Excellent）、良い（Good）、普通（Fair）、悪い（Poor）、非常に悪い（Bad））および快適性のラベル（Very Comfortable（非常に満足）、Comfortable（満足）、Mildly Uncomfortable（やや不快）、Uncomfortable（不満）、Extremely Uncomfortable（非常に不満））が付けられている。

10

20

30

40

50

被験者はこれらの連続尺度上に評点のマークを付けるように求められた。次に、被験者の付けたマークを0～100の範囲の整数値に変換し、平均オピニオン評点（MOS）を算出した。

【0089】

図8にサッカーシーケンスに対するMOSを示す。図7に推定された深度画像をいくつか示す。様々なカメラビューやクラッターが含まれた多種多様なビデオに対して本発明者の方法がどの程度対処できているかに注目されたい。

【0090】

図9に、サッカー以外のシーケンスに対するMOSを示す。ホッケーは、サッカーに最も類似していることから最高の評点が付けられている。しかし、アメリカンフットボールの評点は最低になっている。奥行が非常に良いと評価した被験者もいたが、他の被験者はアメリカンフットボールは強度のオクルージョンとクラッターを伴う極めて動的な環境であるため奥行感覚に難があると評価した。それらの被験者は同じ理由で快適性を「やや不快」と評価した。サッカー以外の結果は本発明者の方法の可能性を示すためのものにすぎず、実際にはサッカーのデータベースを用いてサッカー以外のシーケンスを変換したことに留意することは重要である。将来的には、本発明者は様々なスポーツに対応可能な、より多様性のあるデータベースを作成する予定である。

【0091】

オリジナル3Dとの比較

本発明者が2D-3D変換した結果をステレオカメラで撮影したオリジナル3Dビデオと比較した。この実験にはITU勧告の二重刺激連続品質尺度法（DSCQS）を使用した。DSCQSに基づき、被験者は、両者の違いを適切に評価できるように、評点を付ける前に最低二回は各シーケンスのペア（本発明者が作成した3Dとオリジナル3D）を観た。視聴者には、どちらがオリジナルでどちらが変換結果であるかを知らせずにランダムな順序でシーケンスを見せた。被験者は標準的なITU連続尺度を使用して両方のシーケンスの奥行品質と快適性を評価するように求められた。次に、尺度上に付けたマークを0～100の範囲の整数に変換し、差分オピニオン評点（＝DGCに対する評点－オリジナル3Dに対する評点）を計算する。最後に、差分オピニオン評点の平均（DMOS）を計算する。

【0092】

DMOSがゼロであれば、本発明者が変換した3Dはオリジナル3Dと同じであると判定されたことを意味し、DMOSがマイナス値であれば、本発明者が変換した3Dはオリジナル3Dより奥行感覚／快適性が劣っていると判定されたことを意味する。図11に、各サッカーシーケンスの奥行品質および視覚的快適性のDMOSを示す。本発明者の変換結果は、とりわけサッカーゲームの中の約70%を占めているロングショットでは[9]、オリジナル3Dに匹敵している。被験者によっては本発明者の変換結果がオリジナル3Dよりも快適であると評価していることは注目に値する。それらの被験者は、オリジナル3Dのポップアウト効果は不快感を引き起こすことがあると報告している。

【0093】

最新技術との比較

本発明者の3D変換結果を深度転送法DT+ [11]と比較した。前述の実験と同様、DSCQS評価プロトコルを使用し、奥行品質と視覚的快適性のDMOSを計算した。最も難しいサッカーシーケンスのクローズアップとミディアムショットを調査した。多様なカメラアングル、複雑な動き、クラッター、およびオクルージョンが含まれているため、それらは2D-3D変換にとって最も難しいシーケンスになっている。図12に、DT+と比較した、クローズアップおよびミディアムショットのDMOSを示す。本発明者の技術は、DT+と比較して、ミディアムショットでは平均15ポイント、クローズアップでは平均12ポイント上回っている。また、15人の被験者全員が本発明者の技術にDT+よりも高い、または同等の評点を付けており、報告された差は統計的に有意であった（ p 値< 0.05）。図10に、DT、DT+、および本発明者の技術であるDGCで抽出された深度マップをいくつか示す。深度転送法の最初の実施形態はDTであり、これはDT+よりも品質が大幅に劣ることに留意されたい（図10を参照）

。また、DT+は主観的評価の評点が低いことに加えて、DT+による深度は非常に多くの雑音が発生する場合がある（図10および図5を参照）。そのため長時間の視聴で眼精疲労を引き起こす可能性がある。

【0094】

本発明者は、DGCおよびDT+でクローズアップを545フレーム、非クローズアップを1,726フレーム処理してフレームあたりの平均の実行時間を測定した。空間解像度は960×1080ピクセルである。DGCはクローズアップでは3.53分/フレーム、非クローズアップでは1.86分/フレームを要した。DT+の平均処理時間は15.2分/フレームで、クローズアップと非クローズアップのどちらの場合も本発明者の技術よりも遅かった。DGCは、より負担の大きいマスク作成のステップがあるため、クローズアップの方が所要時間が長かった。非クローズアップは1つのサッカーゲームの最大95%を占めているため[9]、非クローズアップの方が処理が速いことは利点である。とはいえ、クローズアップには豊富な深度情報が含まれていることが多いため、クローズアップを軽視することはできない。計算処理の複雑さを改善するための今後の取り組みとして、ビデオ処理の時空間多重解像度方式に重点を置くことが考えられる。すべての数値は、6個のIntel Xeon CPU E5-2650 0 (2.00 GHz、コア数8)、合計264 GBのRAM、86 GBのキャッシュを装備したサーバ上で実行された処理から報告された。

【0095】

本発明の態様は、2D-3Dビデオ変換方法を提供するものだが、本発明者は一例としてサッカーを採用し、コンピュータ生成の画像および深度情報が記憶された参照データベース（合成3Dデータベース）を用いて、リアルタイムの変換を行っている。従来の方法はこのサッカー試合のサンプルに見られるような多様なシーンや複雑な動きに対応することはできない。本発明者の方法は、合成データベース（参照データベース）から深度勾配を転送し、ポアソン再構築により深度を推定することに基づいている。本発明者は、実写シーケンスと合成シーケンスを使用して、提案した方法を実施し、その評価を行った。その結果、本発明者の方法は、たとえばサッカーゲームの中にある広範囲のビデオショット（様々なカメラビュー、動きの複雑さ、オクルージョン、クラッター、および様々なカラーが含まれている）に対応できることが分かった。本発明者が行った主観的調査の被験者は、本発明者の作成した3Dビデオはほとんど常に「非常に良い（Excellent）」と評価した。また実験結果は、実写シーケンスと合成シーケンスのどちらに対しても、本発明者の方法が客観的にも主観的にも最新技術の性能を上回っていることを示している。

【0096】

本発明の諸側面は、2D-3Dビデオ変換の分野に影響を与え、さらに3Dビデオ処理全般に影響を及ぼす可能性がある。第一に、領域特化型の変換は一般的な方法よりもはるかに優れた結果を生成できる。第二に、深度勾配をブロック単位で転送することで、滑らかで自然な奥行きが生成されるだけでなく、必要な参照データベースのサイズが縮小される。第三に、コンピュータで生成されたコンテンツで構成された合成データベース（参照データベース）により、様々な3Dビデオ処理アプリケーションが大量・多様・正確なテクスチャおよび深度のデータを容易に参照することができる。

【0097】

本発明の諸側面は多様な方面への拡張が可能である。たとえば、様々なスポーツのビデオを変換するには、もっとサイズの大きな合成データベース（参照データベース）が必要になるかもしれない。

【0098】

この明細書および請求項で使用される、「～から成る」という表現およびそのバリエーションは、特定の特徵、ステップ、または構成要素を含有することを意味する。この表現は、他の特徵、ステップ、または構成要素の存在を除外するという意味には解釈されない。

【0099】

前述の説明または後述の請求項または添付の図面で開示され、それらの具体的な形態で

、あるいは、開示された機能を実行するための手段もしくは開示された結果を達成するための方法またはプロセスの観点から適宜に表現された諸特徴は、単独に、または、かかる特徴のいずれかの組み合わせで、本発明を多様な形態で実現するために利用することができる。

【 0 1 0 0 】

- [1] Berkeley 3-D object dataset. (バークリー 3-Dオブジェクトデータベース) <http://kinectdata.com/>.
- [2] Make3D. <http://make3d.cs.cornell.edu/data.html>.
- [3] NYU depth dataset v2. (NYU深度データベースv2)
http://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html. 10
- [4] Performance Investigator for Xbox (PIX). (Xbox用パフォーマンスインベスティゲーター(PIX))
<https://msdn.microsoft.com/en-us/library/windows/desktop/ee663275%28v=vs.85%29.aspx>.
- [5] RGB-D object dataset. (RGB-Dオブジェクトデータベース)
<http://rgbd-dataset.cs.washington.edu/>.
- [6] ITU-R BT.2021, Subjective methods for the assessment of stereoscopic 3DTV systems. Geneva, Switzerland, November 2012. International Telecommunication Union. (ITU-R BT.2021. 立体3DTVシステムの評価のための主観的方法。スイス、ジュネーブ。2012年11月。国際電気通信連合) 20
- [7] P. Bhat, B. Curless, M. Cohen, and C. Zitnick. Fourier analysis of the 2D screened poisson equation for gradient domain problems. In Proc. of European Conference on Computer Vision (ECCV'08), pages 114{128. Marseille, France, October 2008. (P. Bhat、B. Curless、M. CohenおよびC. Zitnick。勾配ドメイン問題に対する2D表示ポアソン方程式のフーリエ解析。コンピュータビジョン欧州会議 (ECCV'08) 議事録 : p.114 ~ p.128。フランス、マルセイユ、2008年10月)
- [8] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In Proc. of European Conference on Computer Vision (ECCV'04), pages 25{36, Prague, Czech Republic, May 2004. (T. Brox、A. Bruhn、N. PapenbergおよびJ. Weickert。ワープの理論に基づく高精度光学的ow評価。コンピュータビジョン欧州会議 (ECCV'04) 議事録 : p.25 ~ p.36。チェコ共和国プラハ、2004年5月) 30
- [9] K. Calagari, K. Templin, T. Elgamal, K. Diab, P. Didyk, W. Matusik, and M. Hefeeda. Anahita: A System for 3D Video Streaming with Depth Customization. In Proc. of ACM Multimedia (MM'14), pages 337{346, Orlando, FL, November 2014. (K. Calagari、K. Templin、T. Elgamal、K. Diab、P. Didyk、W. MatusikおよびM. Hefeeda。Anahita。深度のカスタマイズ可能な3D動画ストリーミングのシステム。ACMマルチメディア会議 (MM'14) 議事録 : p.337 ~ p.346。フロリダ、オーランド。2014年11月)
- [10] D. Hoiem, A. A. Efros, and M. Hebert. Automatic photo pop-up. ACM Transactions on Graphics, 24(3):577{584, 2005. (D. Hoiem、A. A. EfrosおよびM. Hebert。自動フォトポップアップ。ACM Transactions on Graphics, 24(3) : p.577 ~ p.584、2005年) 40
- [11] K. Karsch, C. Liu, and S. B. Kang. Depth transfer: Depth extraction from video using non-parametric sampling. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(11):2144{2158, 2014. (K. Karsch、C. LiuおよびS. B. Kang。深度転送法 : 非パラメトリックサンプリングによる動画からの深度の抽出。EEE Transactions on Pattern Analysis and Machine Intelligence, 36(11) : p.2144 ~ p.2158、2014年)
- [12] J. Ko. 2D-to-3D Stereoscopic Conversion: Depth Estimation in 2D Im 50

ages and Soccer Videos. Master's thesis, Korea Advanced Institution of Science and Technology (KAIST), 2008. (J. Ko. 2D-3Dステレオ変換：2D画像およびサッカー動画の奥行評価。修士論文。韓国科学技術院 (KAIST)、2008年)

[13] J. Konrad, M. Wang, P. Ishwar, C. Wu, and D. Mukherjee. Learning-based, automatic 2D-to-3D image and video conversion. IEEE Transactions on Image Processing, 22(9):3485{3496, 2013. (J. Konrad、M. Wang、P. Ishwar、C. WuおよびD. Mukherjee. 学習型の自動的な2D-3D画像・動画変換。IEEE Transactions on Image Processing, 22(9) : p.3485 ~ p.3496、2013年)

[14] A. Levin, D. Lischinski, and Y. Weiss. A closed-form solution to natural image matting. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(2):228{242, 2008. (A. Levin、D. LischinskiおよびY. Weiss. 自然画像マッティングの閉形式解。IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(2) : p.228 ~ p.242、2008年)

[15] C.-W. Liu, T.-H. Huang, M.-H. Chang, K.-Y. Lee, C.-K. Liang, and Y.-Y. Chuang. 3D cinematography principles and their applications to stereoscopic media processing. In Proc. of ACM Multimedia Conference (MM'11), pages 253{262, Scottsdale, AZ, November 2011. (C.-W. Liu、T.-H. Huang、M.-H. Chang、K.-Y. Lee、C.-K. LiangおよびY.-Y. Chuang. 3D映画撮影術の原理および立体メディア処理への応用。ACMマルチメディア会議 (MM'11) 議事録 : p.253 ~ p.262、アリゾナ州スコッツデール、2011年11月)

[16] P. Ochs, J. Malik, and T. Brox. Segmentation of moving objects by long term video analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(6):1187{1200, 2014. (P. Ochs、J. MalikおよびT. Brox. 長時間動画分析による移動オブジェクトのセグメント化。IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(6) : p.1187 ~ p.1200、2014年)

[17] A. Oliva and A. Torralba. Modeling the shape of the scene: a holistic representation of the spatial envelope. International Journal of Computer Vision, 42(3):145{175, 2001. (A. OlivaおよびA. Torralba. シーンの形状のモデリング：空間エンベロープの全体的表現。International Journal of Computer Vision, 42(3) : p.145 ~ p.175、2001年)

[18] P. Perez, M. Gangnet, and A. Blake. Poisson image editing. ACM Transactions on Graphics, 22:313{318, 2003. (P. Perez、M. GangnetおよびA. Blake. ポアソン画像編集。ACM Transactions on Graphics, 22 : p.313 ~ p.318、2003年)

[19] R. Rzeszutek, R. Phan, and D. Androutsos. Depth estimation for semi-automatic 2D to 3D conversion. In Proc. of ACM Multimedia Conference (MM'12), pages 817{820, Nara, Japan, October 2012. (R. Rzeszutek、R. PhanおよびD. Androutsos. 半自動2D-3D変換の奥行評価。ACMマルチメディア会議 (MM'12) 議事録 : p.817 ~ p.820、日本、奈良、2012年10月)

[20] A. Saxena, S. H. Chung, and A. Y. Ng. Learning depth from single monocular images. In Proc. of Advances in Neural Information Processing Systems (NIPS'05), pages 1161{1168, Vancouver, Canada, December 2005. (A. Saxena、S. H. ChungおよびA. Y. Ng. 単一モノクロ画像からの深度の学習。神経情報処理システムの進歩 (NIPS'05) 議事録 : p.1161 ~ p.1168、カナダ、バンクーバー、2005年12月)

[21] L. Schnyder, O. Wang, and A. Smolic. 2D to 3D conversion of sports content using panoramas. In Proc. of IEEE Conference on Image Processing (ICIP'11), pages 1961{1964, Brussels, Belgium, September 2011. (L. Schnyder、O. WangおよびA. Smolic. パノラマを用いたスポーツコンテンツの2D-3D変換。IEEE画像処理会議 (ICIP'11) 議事録 : p.1961 ~ p.1964、ベルギー、ブリュッセル、2011年9月)

[22] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from sing

10

20

30

40

50

le depth images. In Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'11), pages 1297{1304, Providence, RI, June 2011. (J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. KipmanおよびA. Blake. 単一深度画像からの抽出部分に含まれる人間のポーズのリアルタイム認識。IEEEコンピュータビジョンおよびパターン認識会議 (CVPR'11) 議事録 : p.1297 ~ p.1304. ロードアイランド州プロビデンス、2011年6月)

[23] W. Wu, A. Are_n, G. Kurillo, P. Agarwal, K. Nahrstedt, and R. Bajcsy. Color-plus-depth level-of-detail in 3D tele-immersive video: A psychophysical approach. In Proc. of ACM Multimedia Conference (MM'11), pages 13{22, Scottsdale, Arizona, November 2011. (W. Wu, A. Are_n, G. Kurillo, P. Agarwal, K. NahrstedtおよびR. Bajcsy. 3D遠隔没入型動画におけるカラーおよび深度の詳細度: 心理生理学的アプローチ。ACMマルチメディア会議 (MM'11) 議事録 : p.13 ~ p.22、アリゾナ州スコッツデール、2011年11月)

10

[24] Z. Yang, W. Wu, K. Nahrstedt, G. Kurillo, and R. Bajcsy. Viewcast: View dissemination and management for multi-party 3D tele-immersive environments. In Proc. of ACM Multimedia Conference (MM'07), pages 882{891, Augsburg, Bavaria, Germany, September 2007. (Z. Yang, W. Wu, K. Nahrstedt, G. KurilloおよびR. Bajcsy. ビューキャスト: マルチパーティ3D遠隔没入型環境のためのビューの配布および管理。ACMマルチメディア会議 (MM'07) 議事録 : p.882 ~ p.891. ドイツ、バイエルン州アウクスブルグ、2007年9月)

20

[25] L. Zhang, C. Vazquez, and S. Knorr. 3D-TV content creation: automatic 2D-to-3D video conversion. IEEE Transactions on Broadcasting, 57(2):372{383, 2011. (L. Zhang, C. VazquezおよびS. Knorr. 3D-TVコンテンツの製作: 自動2D-3D動画変換。IEEE Transactions on Broadcasting, 57(2) : p.372 ~ p.383、2011年)

[26] Z. Zhang, C. Zhou, B. Xin, Y. Wang, and W. Gao. An interactive system of stereoscopic video conversion. In Proc. of ACM Multimedia Conference (MM'12), pages 149{158, Nara, Japan, October 2012. (Z. Zhang, C. Zhou, B. Xin, Y. WangおよびW. Gao. 対話型ステレオ動画変換システム。ACMマルチメディア会議 (MM'12) 議事録 : p.149 ~ p.158. 日本、奈良。2012年10月)

【産業上の利用可能性】

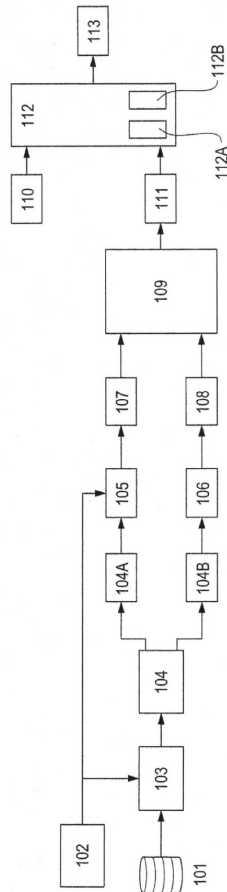
30

【0101】

本発明者の技術は高品質合成画像のデータベースから深度勾配を転送することに基づく。このような画像はビデオゲームから収集することができ、スポーツとそのサブカテゴリ、アクションとそのサブカテゴリ、通常のグラフィクス等の様々なジャンルで入手できることが多い。このような合成画像を2D+深度のレポジトリとして利用すれば、高額なステレオカメラを使用する必要はなくなる。それによりこの技術は最先端技術よりも拡張可能性の高いものになる。さらに、従来のデータ駆動型の技術とは異なり、本発明者の方法ではクエリー画像とデータベースとが全体的に類似する必要はなく、局所的なパッチ(部分)が類似してさえいればよい。それにより2D+深度のデータベースのサイズを大幅に縮小しながら類似深度の品質を維持することができる。その結果として、より一般的なデータへの拡張可能性が高まり、民生品への適応が容易になる。

40

【図 1】



110	オブジェクトマスク
111	推定深度の配
112	水アフィン相関演
112A	オブジェクト境界除去
112B	平滑化
113	2D + 深度

105	局所探索 (ブロックマッチング)
106	深度抽出
107	カメラパラメータ推定
108	差分係数 (Gr, Gb)
109	公配マッピングと精細化

101	合成データベース (2D + 深度)
102	カメラパラメータ
103	相関演算
104	2D + 深度 候補
104A	2D 候補
104B	深度候補

【図 2 - 1】



2(a)

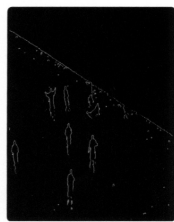


2(b)



2(c)

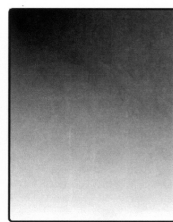
【図 2 - 2】



2(d)



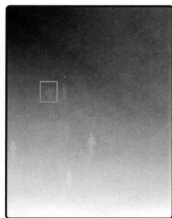
2(e)



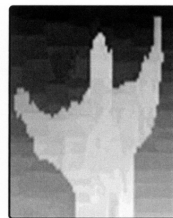
2(f)



2(g)

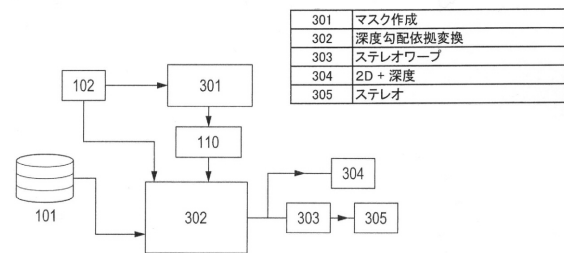


2(h)



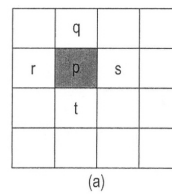
2(i)

【図 3】

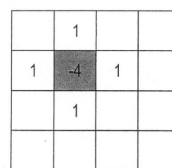


301	マスク作成
302	深度勾配依拠変換
303	ステレオワーブ
304	2D + 深度
305	ステレオ

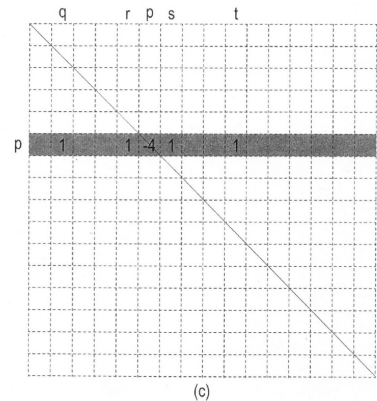
【図 4】



(a)

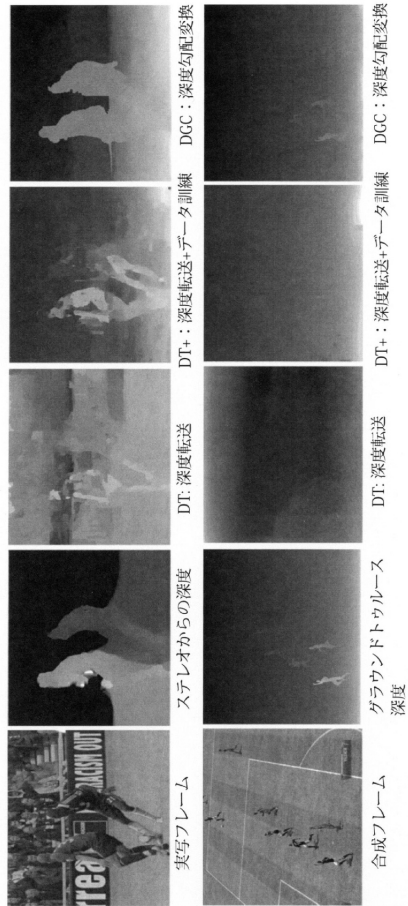


(b)

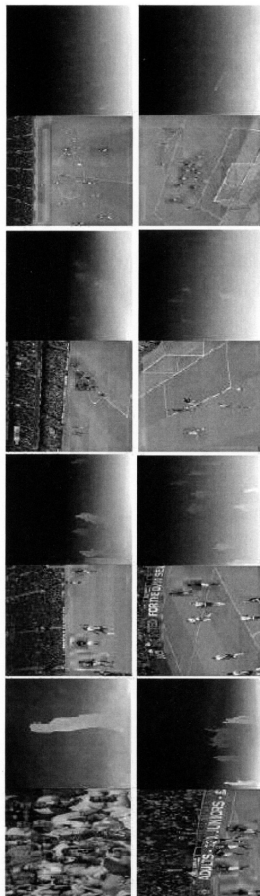


(c)

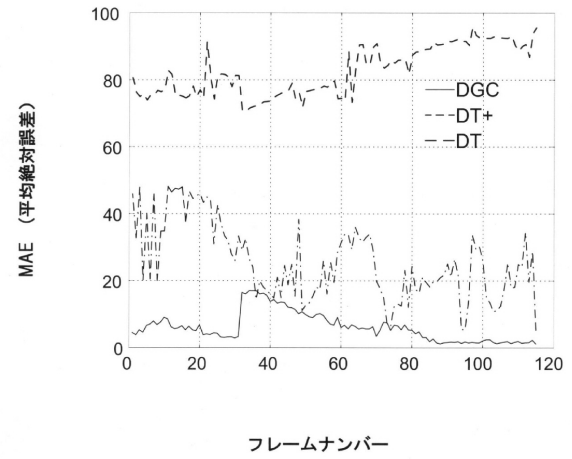
【図 5】



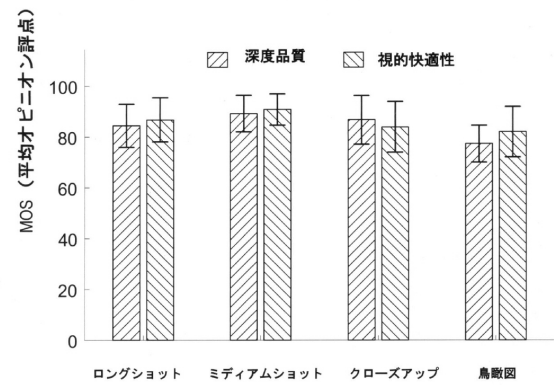
【図 7】



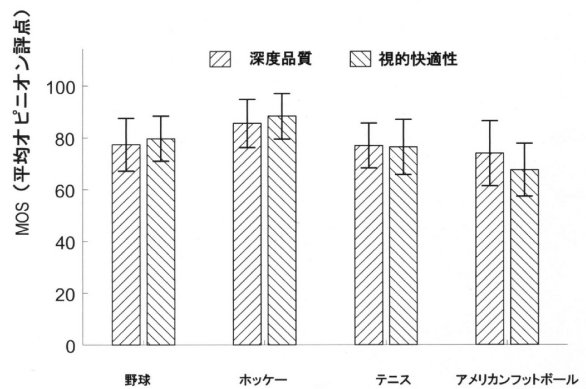
【図 6】



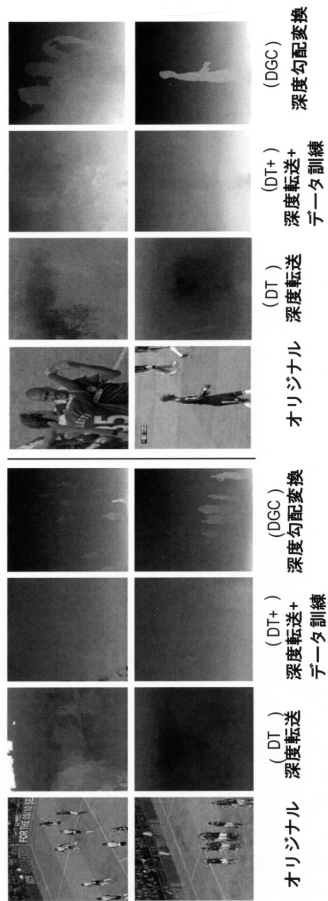
【図 8】



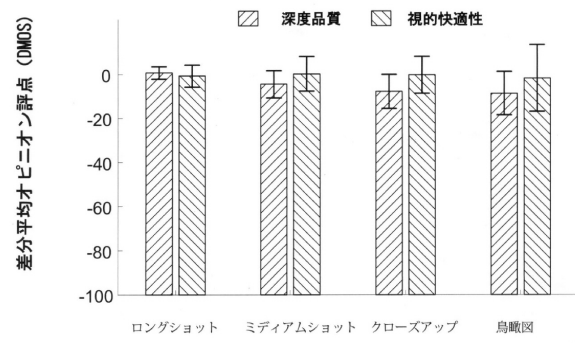
【図 9】



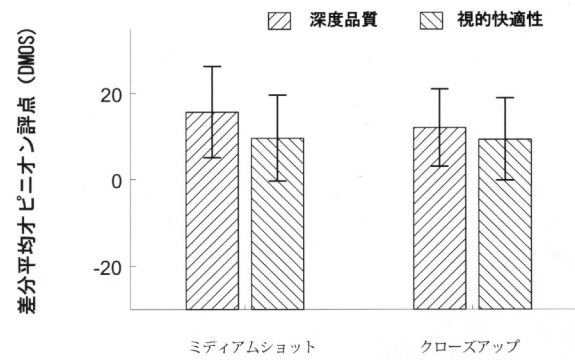
【図 10】



【図 11】



【図 12】



フロントページの続き

(73)特許権者 518040415

エルガリブ、ムハンマド アブデルアズィーズ エー ムハンマド
カタール ドーハ ピーオー ボックス 5 8 2 5 カタール ファンデーション フォー エデ
ュケーション、サイエンス アンド コミュニティー ディベロップメント

(73)特許権者 518040426

マチュシク、ヴォイチェフ
カタール ドーハ ピーオー ボックス 5 8 2 5 カタール ファンデーション フォー エデ
ュケーション、サイエンス アンド コミュニティー ディベロップメント

(73)特許権者 518040437

ディディック、ピオトル
カタール ドーハ ピーオー ボックス 5 8 2 5 カタール ファンデーション フォー エデ
ュケーション、サイエンス アンド コミュニティー ディベロップメント

(73)特許権者 518040448

カスパー、アレキサンダー
カタール ドーハ ピーオー ボックス 5 8 2 5 カタール ファンデーション フォー エデ
ュケーション、サイエンス アンド コミュニティー ディベロップメント

(74)代理人 100173901

弁理士 小越 一輝

(74)代理人 100093296

弁理士 小越 勇

(72)発明者 ヘフィーダ、ムハンマド エム

カタール ドーハ ピーオー ボックス 5 8 2 5 カタール ファンデーション フォー エデ
ュケーション、サイエンス アンド コミュニティー ディベロップメント

(72)発明者 カルガリ、キアナ アリ アスガル

カタール ドーハ ピーオー ボックス 5 8 2 5 カタール ファンデーション フォー エデ
ュケーション、サイエンス アンド コミュニティー ディベロップメント

(72)発明者 エルガリブ、ムハンマド アブデルアズィーズ エー ムハンマド

カタール ドーハ ピーオー ボックス 5 8 2 5 カタール ファンデーション フォー エデ
ュケーション、サイエンス アンド コミュニティー ディベロップメント

(72)発明者 マチュシク、ヴォイチェフ

カタール ドーハ ピーオー ボックス 5 8 2 5 カタール ファンデーション フォー エデ
ュケーション、サイエンス アンド コミュニティー ディベロップメント

(72)発明者 ディディック、ピオトル

カタール ドーハ ピーオー ボックス 5 8 2 5 カタール ファンデーション フォー エデ
ュケーション、サイエンス アンド コミュニティー ディベロップメント

(72)発明者 カスパー、アレキサンダー

カタール ドーハ ピーオー ボックス 5 8 2 5 カタール ファンデーション フォー エデ
ュケーション、サイエンス アンド コミュニティー ディベロップメント

審査官 益戸 宏

(56)参考文献 特表2014-532941(JP,A)

特開2012-103135(JP,A)

米国特許出願公開第2014/0210944(US,A1)

(58)調査した分野(Int.Cl., DB名)

H04N 13/00

G06T 7/50

G 0 6 T 1 / 0 0