

令和元年6月21日現在

機関番号：11201

研究種目：基盤研究(C)（一般）

研究期間：2016～2018

課題番号：16K00225

研究課題名（和文）多視点ビデオ映像を利用した実世界映像のインタラクティブ編集技術の開発

研究課題名（英文）Development of interactive editing technique of real world video using multi-viewpoint video

研究代表者

藤本 忠博（Fujimoto, Tadahiro）

岩手大学・理工学部・教授

研究者番号：00312512

交付決定額（研究期間全体）：（直接経費） 3,500,000円

研究成果の概要（和文）：複数カメラによる多視点ビデオ映像から生成した実世界の自由視点映像上の目的物体をインタラクティブ編集し、仮想世界の自由視点映像をリアルタイム生成する技術の開発を目指した。先行技術の2次元幾何変換を3次元に拡張するため、プレーンスイープ法の導入により他物体に遮蔽された凹凸物体を鮮明に抽出する技術を開発した。多様な要素を考慮した基礎実験に向けて実環境を模擬する仮想環境を構築し、クラスタリング処理を高速化するシェーダプログラムの改良も試みた。応用の可能性に向け、複数のRGB-Dカメラによる自由視点映像生成技術、複数の実世界画像から仮想世界画像を合成するイメージモンタージュ技術の開発も行った。

研究成果の学術的意義や社会的意義

研究代表者の知る限り、自由視点のビデオ映像を撮影と同時にリアルタイムでインタラクティブに編集する技術はなく、本研究は学術的に非常に独創的であると言える。応用分野として、映像コンテンツ制作の他に、遠隔地の実世界を仮想的に操作可能とする遠隔コミュニケーション（遠隔会議、遠隔展示など）やナビゲーションなどが挙げられ、さらには、本技術を実用化する映像スタジオの構築、制作者と視聴者や視聴者同士のリアルタイムインタラクションを可能とするライブ映像を配信する放送システムの実現なども考えられる。我々の生活を豊かにする大きな可能性を持つという観点から、社会的意義のある研究であると言える。

研究成果の概要（英文）：The goal of this research is to develop a technique to enable a user to interactively edit an existing object displayed in a free-viewpoint video of a real world, which is generated by multiple cameras, and create a free-viewpoint video of a virtual world in real time. In order to extend 2D geometric transformations to 3D ones, we developed a method to purely extract an object with a bumpy surface occluded by other objects using plane sweep. We constructed a virtual environment in a computer simulating a real environment for fundamental experiments with various conditions. We tried to improve a shader program to make a clustering process more efficient. In addition, we developed a free-viewpoint video technique using multiple RGB-D cameras, and image montage techniques to generate a virtual world image from real world images.

研究分野：メディア工学、コンピュータグラフィックス、画像処理、コンピュータビジョン

キーワード：イメージベースCG ビデオベースCG 知的映像編集 多視点ビデオ映像 自由視点映像 実世界 仮想世界 カメラアレイ

## 様式 C-19、F-19-1、Z-19、CK-19（共通）

### 1. 研究開始当初の背景

実世界映像上に実際には存在しない物体を付加することで仮想世界映像を生成する「拡張現実 (Augmented Reality)」の研究が盛んであり、また、最近では、実世界映像上の不要な物体を除去した仮想世界映像を生成する「減損現実 (Diminished Reality)」の研究も行われている。これらは、映像コンテンツ制作、遠隔コミュニケーション、ナビゲーションなど、我々の生活の豊かさと利便性を向上させる大きな役割が期待される。それらに関連する技術として、実世界の撮影対象（物体や風景など）を複数のカメラで異なる位置から撮影したビデオ映像（多視点ビデオ映像）を利用して、撮影対象を自由な視点位置から見た映像（自由視点映像）を生成する技術は、コンピュータグラフィックスとコンピュータビジョンの分野において、これまで、国内外で活発に研究されてきた。また、関連して、静止画像やビデオ映像の知的編集技術の研究も国内外で活発である。

本研究の研究代表者は、これまでに、自由視点映像生成、ならびに、知的映像編集技術の研究を行ってきた。特に、それらを融合し、拡張現実や減損現実を実現する研究として、複数のカメラで構成されたカメラアレイによる多視点ビデオ映像から遮蔽を伴う目的物体を（遮蔽物体によってぼやけることなく）鮮明に抽出し、インタラクティブな2次元幾何変換（拡大縮小、回転、平行移動）を与えた仮想世界の自由視点映像を撮影と同時にリアルタイムで生成する技術を開発している。本研究では、この先行技術における幾何変換を3次元に拡張する。

### 2. 研究の目的

本研究では、複数のカメラで撮影した多視点ビデオ映像から生成した実世界の自由視点映像上の目的物体にインタラクティブな編集を行い、仮想世界の自由視点映像を生成する技術を開発する。ユーザが自由視点映像上の任意の目的物体を選択して抽出し、回転や拡大縮小などの3次元幾何変換を施し、他の位置に3次元移動して合成することで、実世界とは異なる仮想世界の映像をリアルタイムで生成することを目指す。自由視点位置からは実際には見えない側の目的物体の表面や他物体の背後に隠れた表面の完全な抽出を可能とし、他物体の背後への移動や不要物体の除去も可能とし、実世界映像と同品質の違和感のない仮想世界映像を生成する。

### 3. 研究の方法

先行技術における2次元幾何変換を3次元に拡張するにあたり、3次元回転などでは、自由視点から見えない側の目的物体の映像も要し、自由視点とは異なる編集視点から見た目的物体の表面を抽出する必要がある。そこで、編集視点に対する目的物体と自由視点に対する前方・後方物体を抽出する要素技術のため、他物体に部分的に遮蔽された物体も抽出可能とするように先行技術を改良する。特に、焦平面による先行技術では表面に凹凸を持つ物体の鮮明な抽出が困難であったため、その解決も図る。具体的には、プレーンスweep法の導入により凹凸物体の鮮明な抽出や動物体の連続的な抽出も可能とする技術の開発を行う。また、先行技術では、その核となるクラスタリング処理の効率化が課題であったため、その処理の高速化を試みる。また、その他にも、適宜、3次元への拡張のために必要な技術を検討し、実現を試みる。

### 4. 研究成果

上記のプレーンスweep法の導入により凹凸物体の鮮明な抽出や動物体の連続的な抽出を行う技術の開発を行った。また、多様な要素を考慮した試行錯誤的な実験を容易に行うため、実環境を模擬する仮想環境をコンピュータ内に構築した。そして、上記のクラスタリング処理の高速化にあたり、GPU上で高速に動作するようにシェーダプログラムの改良を試みた。

また、本研究への応用の可能性から、複数のRGB-Dカメラを利用した自由視点映像生成技術、複数の実世界画像から仮想世界画像を合成するイメージモンタージュ技術の開発も行った。

最終的には当初の目標の十分な実現までには至らなかったが、本研究の今後の継続に向けた多くの知見や解決すべき問題点が明らかとなり、有意義な結果が得られた。以下、本研究の実施期間中に得られた成果について述べる。

#### (1) プレーンスweep法の導入による遮蔽された凹凸物体の鮮明な抽出

##### ① 背景

物体を異なる位置からカメラで撮影した複数の画像（多視点画像）を用いて、ユーザが指定する自由な視点位置から物体をレンダリングして自由視点画像を得る様々な手法が提案されており、イメージベースレンダリングと呼ばれている。イメージベースレンダリングのうち、ライトフィールドレンダリングと呼ばれる手法では、多視点画像を膨大なレイの集合とみなし、自由視点画像を構成する適切なレイを利用して自由視点画像を合成する。さらに、合成開口法 (Synthetic Aperture Imaging) と呼ばれる手法では、実際のカメラにおけるレンズと焦点面によるフォーカスの原理に基づき、カメラアレイで撮影された多視点画像を用いて、撮影後のピント調整や、前方の他の物体（遮蔽物体）に部分的に遮蔽されている後方の注目物体の全体像を得ることなどを可能とする。先行技術では、多視点画像を用いて、合成開口法の原理を利用することで、他の物体に部分的に遮蔽されている注目物体の全体像を背景から実時間で鮮明に抽出する。しかし、カメラアレイに対して奥行き方向に大きな凹凸のある注目物体は鮮明な

抽出が困難であった。そこで、本研究では、この先行技術を改良し、大きな凹凸のある注目物体を鮮明に抽出する技術を提案する。本技術では、プレーンスweep法を利用し、複数の焦点面を用いることで、自由視点画像の画素ごとに注目物体の表面が存在する位置を判定することで、表面の凹凸が大きい物体に対しても鮮明な抽出が可能となる。

## ② 先行技術の概要

先行技術では、合成開口法を利用して、遮蔽物体を除去して注目物体を鮮明に抽出し、平行移動、回転、拡大縮小などのエディット処理を適用した後、背景と合成した仮想世界映像を実時間で生成する。合成開口法では、通常、注目物体の表面が平面であると仮定し、その表面が存在するものと仮定した位置に焦点面を配置する。そして、自由視点画像の各画素に対応する焦点面上の3次元点を映す全カメラの画素値を平均化して、自由視点画像の画素値とする。一方、先行技術では、注目物体を鮮明に抽出するため、各カメラの画素値の類似性から、その焦点面上の3次元点が注目物体の表面上にあるかを判定する。各カメラの画素値が互いに類似しているかどうかの判定にはクラスタリングによる方法を用いる。先行技術では、クラスタリングの手法として、階層的クラスタリングの中でも単純な手法である重心法を用いている。R、G、Bの3要素からなる画像値を3次元空間上のサンプル点とし、以下の処理を行う。

- ・ クラスタリングの初期化として、N台のカメラに対して、それぞれ、各カメラの画像値だけを持つN個のクラスタを構築する。
- ・ 3次元空間上で、クラスタの重心間のユークリッド距離が小さな順に、2つのクラスタを併合する処理を反復する。
- ・ 重心間のユークリッド距離の最小値が閾値を超えた時点で、クラスタリングを終了する。

クラスタリングの終了後、画素値の個数が最大のクラスタについて、その個数が閾値以上の場合、その画素値が注目物体の表面の色であるとみなし、その表面が焦点面上にあると判定する。そして、その画素値の平均を自由視点画像に与えることで、遮蔽物の色を除去した注目物体の鮮明な抽出がなされる。

## ③ 本研究で提案する技術

先行技術では、通常、合成開口法を利用し、注目物体の表面が平面であると仮定するため、カメラアレイから見て奥行き方向に大きな凹凸のある注目物体は鮮明な抽出が困難であるという問題がある。この問題を解決するため、本研究ではプレーンスweep法を用いる。プレーンスweep法では、3次元空間上で自由視点画像の画像面に平行な平面（プレーン）を奥行き方向に移動（スweep）しながら、自由視点画像の画素ごとに各カメラの画素値の類似性からプレーン位置に物体の表面があるかを判定する。実際には、離散的なプレーン位置で判定を行う。先行技術では、焦点面を一つの位置に固定した。一方、本研究では、焦点面をプレーンとしてスweepし、複数の位置に焦点面を置く。そして、クラスタリングによる各カメラの画素値の類似性の判定とプレーンスweep法によるプレーン位置ごとの画素値の類似性の判定を組み合わせることで、注目物体の表面が存在する適切な位置を求め、かつ、遮蔽物体の色を除去して注目物体の色だけを鮮明に得る。

スweepしたそれぞれの焦点面の位置において、先行技術と同様の方法で各カメラの画素値のクラスタリングを行う。そして、以下の2つの方法のいずれかにより、注目物体の表面の位置と色を決定する。

方法1：焦点面の位置ごとに、最大クラスタの画素値の分散値が閾値以下の場合、候補クラスタとする。全ての候補クラスタのうち、画素値の分散値が最小となる焦点面の位置に注目物体の表面があると判定する。

方法2：焦点面の位置ごとに、最大クラスタの画素値の個数が閾値以上の場合、候補クラスタとする。全ての候補クラスタのうち、画素値の個数が最大となる焦点面の位置に注目物体の表面があると判定する。

注目物体の表面があると判定された候補クラスタの画素値の平均値によって、注目物体を鮮明に抽出した自由視点画像が得られる。

## ④ 実験結果

カメラ 20 台によるカメラアレイを用い、全てのカメラを1台のPC（インタフェースカード（4ポート）×5個を装着）にUSB接続して実験を行った。PCのスペックは以下である。

CPU : Core i7-4790 3.6-4.0GHZ      GPU : GeForce GTX 980 4GB GDDR5

メモリ : DDR3-1600 8GB×2 (16GB)      ハードディスク : 1TB

実験結果を図1に示す。(a)では、2つの注目物体をカメラアレイに対して平行にし、かつ、カメラアレイから等距離に配置した。(b)では、右側の注目物体をカメラアレイのほうに近づけた。さらに、(c)では、右側の注目物体をカメラアレイに対して斜めに配置した。2つの注目物体の奥行きが異なる(b)と(c)の場合は大きな凹凸のある注目物体を模擬しており、先行技術に比べて本研究の技術のほうが良好な結果が得られたことから、その有効性が確認された。



図 1：実世界の自由視点画像上の物体の抽出

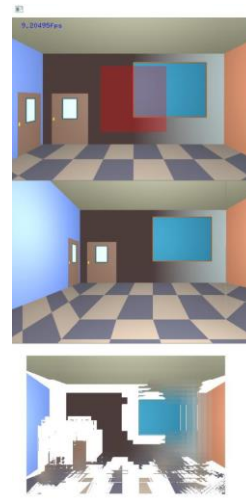


図 2：仮想環境の構築

## (2) 実環境を模擬する仮想環境の構築

様々な撮影条件のもとでロバストかつ高精度に動作するシステムの開発にあたり、多様な要素を考慮した試行錯誤的な実験を容易に行うため、本研究を実施する実環境を模擬する仮想環境をコンピュータ内に構築した。実験結果を図 2 に示す。

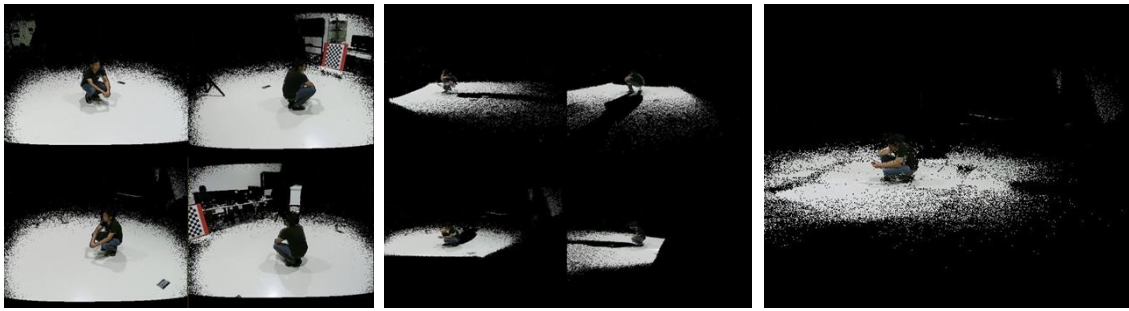
## (3) クラスタリング処理の高速化のためのシェーダプログラムの改良

(1) で述べたように、物体の鮮明な抽出のためには、カメラアレイ上の各カメラの画素値の類似性から、焦点面上の 3 次元点が注目物体の表面上にあるかを判定する必要がある。そして、各カメラの画素値が互いに類似しているかどうかの判定にはクラスタリングによる方法を用いる。先行技術で用いた重心法では、クラスタの重心間のユークリッド距離が小さな順に 2 つのクラスタを併合する処理を反復するが、反復ごとに全てのクラスタの組み合わせで重心間距離を計算するシンプルなアルゴリズムで実装を行っていた。しかし、前回の反復で変化のないクラスタどうしの距離は不変のため、毎回の反復で全ての組み合わせで計算することには無駄が多い。そこで、変化のないクラスタどうしの距離は再計算せず、前回の反復で併合して新たに生成されたクラスタに関わる距離だけを計算するようにアルゴリズムを改良した。

新たなアルゴリズムの有効性を検証するため、まずは C 言語で実装して CPU 上で実行することにより、既存のシンプルなアルゴリズムと計算時間を比較した。その結果、新たなアルゴリズムのほうが高速であることが分かった。続いて、同じアルゴリズムをシェーダ言語 GLSL により実装し、GPU 上で実行する比較実験を行った。その結果、予想に反して、新たなアルゴリズムのほうが多くの計算時間を要することとなった。この原因は現在も調査中であるが、GLSL による実装の仕方の問題がある可能性が考えられる。GLSL による実装は、本来、グラフィック処理を GPU 上で高速に実行することが目的であるため、汎用的な利用のためには制約がある。今後、その点を考慮した新たなアルゴリズムを設計する予定である。

## (4) 複数の RGB-D カメラを利用した自由視点映像の生成

自由視点映像は 3 次元空間上の自由な視点位置から対象とする物体やシーンを見ることができ、以前の代表的な方法では、通常の RGB (カラー) 映像を取得する複数の RGB カメラによって物体を周囲から撮影した多視点 RGB 映像を用いて、RGB 情報から物体の 3 次元形状を推定することで自由視点映像を生成する。しかし、この方法は、一般に、2 次元情報から 3 次元情報を推定する精度と処理時間に問題があった。一方、最近では、画素ごとにカメラから物体表面までの距離を持つ Depth (デプス) 映像を RGB 映像に加えた RGB-D 映像を取得できる RGB-D カメラの利用により、2 次元情報から推定することなく、直接、距離情報から高精度の 3 次元形状を高速に得ることが可能となった。そして、自由視点映像の生成のため、物体を囲む複数の RGB-D カメラによる多視点 RGB-D 映像を用いる方法も提案されてきている。しかし、通常、各 RGB-D カメラを個別の PC に接続してネットワークを構築するため、ネットワークを介して PC 間で RGB-D 映像のデータを転送する必要があり、その膨大なデータ量が自由視点映像のリアルタイム生成のボトルネックとなる。そこで、本研究では、フレーム時刻ごとの自由視点の位置に応じて、自由視点映像の生成にとって必要な軽量化された RGB-D データを生成して転送することで、効率的に自由視点映像を生成する手法を提案する。提案手法では、まず、各 RGB-D カメラが取得した個々のローカル RGB-D データのみを用いて並列に自由視点に向けた「視点依存プリレンダリング」を行い、軽量化された自由視点 RGB-D データを生成する。そして、そのデータをネットワーク転送して統合することで自由視点映像をリアルタイムで生成する。データをネットワーク転送する際には Run Length Encoding (RLE) による符号化を行うことで、さらなる軽量化を図る。



各カメラが取得したローカル RGB-D データ      ローカル RGB-D データから生成された自由視点 RGB-D データ      全ての自由視点 RGB-D データを統合して生成された自由視点映像

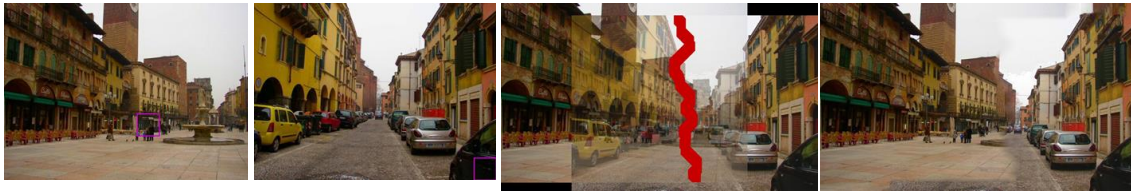
図 3：複数の RGB-D カメラを利用した自由視点映像の生成

4 台のカメラを用いた実験結果を図 3 に示す。視点依存プリレンダリングによりデータの軽量化がなされ、視点の位置に応じた適切な自由視点映像が生成されている。

#### (5) 複数の実世界画像から仮想世界画像を合成するイメージモンタージュ技術

##### ① 2 画像の相対配置の最適化によるイメージモンタージュ

本研究では、異なるシーンを撮影した 2 枚の入力画像の最適な相対配置を自動的に決定して自然な出力画像を合成するイメージモンタージュ法を提案する。本手法では、2 枚の入力画像が部分的に重なるように相対的に配置し、その重なり領域において入力画像間の色の分布が最も類似する縦横方向の画素列を最類似パスとして求める。そして、効率的な多重解像度処理により、全ての相対配置の中から類似度が最も高い最類似パスを持つ最適な相対配置を決定する。最類似パスの探索は、イメージキルティングテクスチャ合成法で用いられる画素ベースの探索法をパッチベースに発展させた「パッチベースイメージキルティング探索法」によって行う。このパッチベース探索法では、入力画像間で重なり合うパッチどうしの類似度を考え、パスに沿った連続するパッチ群の類似度が最も高いものを最類似パスとして得る。パッチ間の類似度の評価にはカラーヒストグラムを用いる。最適な相対配置の決定後は、2 枚の入力画像をつなぐ「境界領域」と出力画像上に複製する各入力画像上の「複製領域」を縦横方向の最類似パスから決定する。そして、境界領域に対するフェザリングとテクスチャ合成、あるいは、ポアソン画像合成によって自然な出力画像を得る。なお、出力画像上に残したい各入力画像上の「保持領域」の指定も可能である。実験結果を図 4 に示す。異なるシーンを撮影した 2 枚の入力画像から自然な出力画像が合成されている。

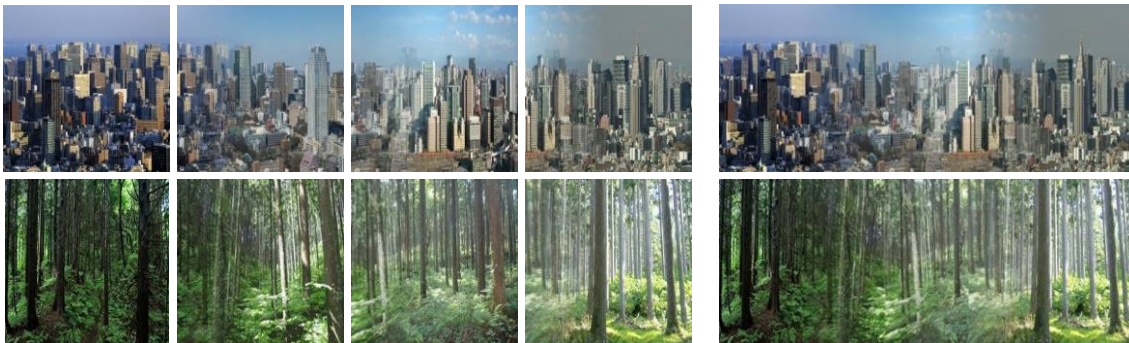


入力画像 1      入力画像 2      最適な相対配置と境界領域      出力画像

図 4：2 画像の相対配置の最適化によるイメージモンタージュ

##### ② 異なるシーンを融合して仮想世界パノラマ画像を生成するイメージモンタージュ

本研究では、異なるシーンを撮影した大量の入力画像を持つデータベースから適切な画像群を自動的に選択して合成することで、実在する一つのシーンのように見える仮想世界パノラマ画像を生成するイメージモンタージュ法を提案する。提案手法では、各入力画像が同じシーン



異なるシーンの入力画像群

仮想世界パノラマ画像

図 5：異なるシーンを融合して仮想世界パノラマ画像を生成するイメージモンタージュ

内で共通の視点から周囲の異なる水平な視線方向に向けて撮影した画像であるとみなす。まず、視線方向の回転によってシーン中の共通部分が左右にずれて写り込んでいるかのように見える、左右部分が互いに類似する2枚の画像のペアを求める。そして、複数の画像ペアから、視線方向の回転に合わせて並べたときに左右の隣接する画像と互いに類似部分を持つような連続する画像群を求め、類似部分を重ね合わせて合成することでパノラマ画像を生成する。類似部分を得るための画像間の重なり領域の類似性の評価には GIST 特徴量とカラーヒストグラムを用いる。複数の画像ペアからパノラマ画像を生成する画像群を求める最適化問題は、グラフ上の制約付き閉路探索問題に帰着させて解く。実験結果を図5に示す。左右部分が互いに類似している隣接する画像群からパノラマ画像が合成されている。

## 5. 主な発表論文等

〔雑誌論文〕(計1件)

- ① 鈴木康太, 藤本忠博, パッチベースイメーজキルティング探索法を用いた2画像の相対配置の最適化によるイメーজモンタージュ法, 芸術科学会論文誌, 16巻, 3号, pp.56-74, 2017年, 査読有  
<https://www.art-science.org/journal/v16n3/>

〔学会発表〕(計11件)

- ① 溝江史也, 藤本忠博, ビデオ映像上の遮蔽物体除去を伴う注目物体抽出法の高速化, 平成30年度第3回芸術科学会東北支部研究会, 2018年
- ② 徐迅, 藤本忠博, プレーンスイープ法を用いた部分遮蔽された凹凸のある物体の実時間抽出手法, 平成30年度第1回芸術科学会東北支部研究会, 2018年
- ③ Yusuke Sasaki, Tadahiro Fujimoto, Efficient Virtual View Rendering by Merging Pre-rendered RGB-D Data from Multiple Cameras, IWAIT2018 (国際学会), 2018年
- ④ Mizuki Tachibana, Tadahiro Fujimoto, Image Montage for Constructing Photorealistic Virtual World from Different Real Scene Images, IWAIT2018 (国際学会), 2018年
- ⑤ 佐々木悠輔, 藤本忠博, 複数の RGB-D カメラを用いた視点依存プリレンダリングによる効率的な自由視点映像生成法, NICOGRAPH2017, 2017年
- ⑥ 立花瑞樹, 今野尚輝, 藤本忠博, 異なるシーンを融合して仮想世界パノラマ画像を生成するイメージモンタージュ法, NICOGRAPH2017, 2017年
- ⑦ 向井尚人, 藤本忠博, 焦平面スイープを用いた合成開口法による注目物体抽出の改善, 平成28年度芸術科学会東北支部大会, 2017年
- ⑧ 立花瑞樹, 今野尚輝, 藤本忠博, シーン画像連結のための GIST 特徴とカラーヒストグラムの効果について, 平成28年度芸術科学会東北支部大会, 2017年
- ⑨ Kota Suzuki, Tadahiro Fujimoto, Montage of Multiple Images using Optimal Relative Arrangement by Patch-based Image Quilting Search, IWAIT 2017 (国際学会), 2017年
- ⑩ 鈴木康太, 藤本忠博, パッチベースイメージキルティング探索法を用いた2画像の最適な相対配置によるイメージモンタージュ法, NICOGRAPH2016, 2016年
- ⑪ 立花瑞樹, 今野尚輝, 藤本忠博, GIST 特徴を利用したシーン画像の連結, 平成28年度第2回芸術科学会東北支部研究会, 2016年

〔図書〕(計0件)

〔産業財産権〕

○出願状況 (計0件)

○取得状況 (計0件)

〔その他〕

なし

## 6. 研究組織

(1) 研究分担者

なし

(2) 研究協力者

なし

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。