

作業空間への教師ビデオ重畳表示による AR 支援システム

後 藤 道 彦[†] 植 松 裕 子[†] 斎 藤 英 雄[†]
 仙 田 修 司^{††} 池 谷 彰 彦^{††}

[†] 慶應義塾大学

^{††} 日本電気株式会社

あらまし Augmented Reality(AR) 技術を利用した作業支援は、ユーザに対して直感的な支援が可能である。一般に、AR システムを構築するためには、重畳するコンテンツを用意する必要があり、従来はそのコンテンツを用途に合わせて CG など新たに作り出していた。本論文で提案する AR システムでは、そのコンテンツを新たに作るのではなく、既に多く存在しているビデオをお手本となる教師ビデオとして利用する。このとき、ユーザの視点に合わせて教師ビデオの視点を変換して重畳表示することで、ユーザが理解しやすいような情報提示を行う。また、ユーザの作業と教師ビデオ映像が同時に表示されることで、両者が視覚的に混同しやすいという問題を解決するために、教師ビデオに対して透明度変化や輪郭線の強調表示などの様々な加工を施す。さらに、作業内容に合わせて教師ビデオをいくつかの手順に分割し、ユーザ側の作業が完了してから次の手順へと進めるようなインタラクティブ性を持たせることで、ユーザが各自のペースで作業を行うことのできるシステムとなっている。本システムの有効性を検証するためのユーザ評価実験では、折り紙を折る作業とブロックを配置する作業の 2 つの状況を想定し、本システムでの教師ビデオの提示方法について評価を行った。その結果、本システムで提案するようにユーザの視点に合わせてビデオを提示することで、ユーザの視認性が向上するということが分かった。また、教師ビデオに施すエフェクトについての評価では、作業内容ごとに適したエフェクトを分類することができ、今後さらに他の作業へと適用する際の指標を得ることができた。

キーワード: Augmented Reality, ヒューマンコンピュータインタラクション, 組み立て支援

Summary Augmented Reality (AR) based work supporting system can help users intuitively. For developing AR systems, it is necessary to create contents to be overlaid. In most cases, such contents are newly-created with CG according to the application. In our system, we utilize already-existing videos as instruction videos which are transformed to the user's view point and overlaid onto the user's view. In order to solve the problem that the instruction video and the user's view may be visually confused, we add various visual effects to the instruction video such as transparency, enhancement of contours, and so on. By dividing the instruction video into some sections according to the work, moreover, the user interactively goes to the next step in the instruction video after the user's operation has finished. Therefore the user can carry on the work at his/her own pace. In the usability test, the users evaluated how to provide the instruction video in our system through two types of works; making origami and building block. As a result, we can find that user's visibility improves by transforming the instruction video to the user's view. As for evaluation of visual effects, we can classify those effects according to the work and obtain the guideline in applying the other work to our system.

Key words: Augmented Reality, human-computer interaction, assembly instruction

1. はじめに

現実世界の中に仮想的な情報を重ね合わせた映像を生成する「Augmented Reality(AR:拡張現実感)」と呼ばれる技術がある¹⁾²⁾³⁾。実世界と仮想情報を同時に重ねて見ることができるため、直感的な情報提示が可能となり、ナビゲーション、プラモデルの組み立て、楽器演奏支援などに応用されている。ナビゲーションでは、カメラで撮影された道路映像に対して、ルート案内情報を付加することで支援を行う⁴⁾⁵⁾。また、プラモデルの組み立てでは、次に必要なパーツを映像中のその部品の上に直接表示し、それを取り付けるべき場所も重畳表示することでスムーズな組み立てを支援する。

このような AR を実現する要素技術は、付加コンテンツの作成、カメラの位置姿勢推定、コンテンツとの位置合わせの3つに大きく分けられる。付加コンテンツの作成では、表示したい CG や画像、文字など仮想情報を生成する。位置姿勢推定では現実空間とカメラの画像座標との位置関係を特定する。位置合わせでは、位置姿勢推定によって求められた位置関係を用いて、現実世界に仮想情報を違和感なく合成する。AR は位置姿勢推定と位置合わせをリアルタイムで実現することで現実世界に違和感なく仮想情報を表示することを可能としている。

これまでに、AR を用いて直感的かつインタラクティブに作業の支援を行う研究が多く行われている。Flagg⁶⁾は、絵画をプロジェクターで白紙キャンパスに投影することで、直感的な絵画描画システムを実現している。また Robertson⁷⁾は、HMD をかけてレゴブロックを配置する指示を AR で表示するときに、指示を行う CG をどの位置に配置するのが作業を行う上で望ましいかを評価した。その結果、レゴブロックのように配置位置が重要な場合、位置合わせがなされていない表示より、位置合わせがなされている方が望ましいということが示されている。楽器演奏の支援システムでは、元川らによるギター演奏システムがある⁸⁾。このシステムでは、据え置き型のディスプレイに自分の姿を映し出すことで鏡を見ているような表示スタイルにしており、押さえるべき弦の位置の上に CG の手を重畳表示することによって、初心者でも直感的で分かりやすい支援を行っている。また Salonen⁹⁾は、AR を用いて積み木を組み立てる支援を行っており、Nilsson¹⁰⁾は、病院という限られた作業空間での AR インタラクションを行っている。さら

に Steven¹¹⁾は、HMD に装着された赤外線マーカを、環境内に 10 個設置されたカメラで読み取ることで、装着者の動きを検出し、適切な画像を表示する大規模な AR 作業支援システムを実現している。

このような従来研究では、支援情報である AR コンテンツを CG などで作っているため、支援内容が変わる度に、目的に合わせて適切な 3DCG コンテンツを作ることになる。そこで本論文では、新たに CG コンテンツを作るのではなく、同じ作業をしている様子をビデオで撮影しておき、そのビデオを作業支援の AR コンテンツとして利用するシステムを提案する。実際の工場などで部品を組み立てたりする際には、ビデオを目の前で流しながらその通りに作業を行い、習熟していくという事例が既に存在しており、またインターネット上に様々なインストラクションを目的としたビデオシーケンスが公開されている事例¹²⁾¹³⁾もあるため、そのようなインストラクションビデオを AR システム用の作業支援コンテンツとすることは十分に可能である。

しかし、インストラクションビデオを撮影するときには、撮影者によって様々な視点から作業対象を撮っているため、ユーザが見るときには、自分の手元とビデオの作業対象との向きを脳内で合わせる必要があり、作業が複雑になる恐れがある。実際に遠隔地での作業支援システムなどの研究でも、視点が合うと分かりやすいという事例が報告されており、木村ら¹⁴⁾の研究では、VR 環境下においてユーザの体と表示する 3D 教示像の体の軸に注目して位置条件を分類、検討しており、ユーザの視点に合わせて教示像を提示した方が、より細かな動作の模倣の理解度が向上することが示されている。そのため、ビデオを撮る際に作業視点でカメラを配置するということも考えられるが、作業環境にも依存するため必ずしも理想的な位置から撮影できるとは限らない。よって任意の視点から撮影済みのビデオをユーザが見る際に視点変換して提示することができれば、より分かりやすいシステムになると考えられる。

そこで本研究では、作業の様子が録画されているビデオを AR 付加情報のコンテンツとして適用した作業支援システムを提案する。ビデオは、ユーザの視点に合うように視点変換され、ユーザの作業空間へ重畳表示される。また、ビデオが重畳表示される際に、ユーザ自身の操作状況とビデオ中の作業情報が同時に見えることによって、両者の見分けが付きにくくなり、両者を混同してしまうという問題が生じるため、教師ビデオに対して様々な視覚効果を付加することで、問題解消を目指す。

“AR-Based Supporting System by Overlay Display of Instruction Video” by Michihiko GOTO, Yuko UEMATSU, Hideo SAITO, (Graduate School, University of Keio), Syuji SENDA, and Akihiko IKEYITANI, (NEC Corporation).

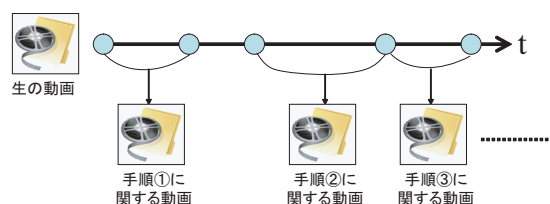


図 1 教師ビデオの分割

Fig. 1 Segmentation of instruction video

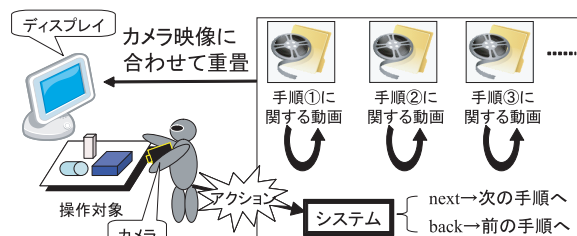


図 2 教師ビデオの提供方法

Fig. 2 How to provide instruction video

2. 提案システム

2.1 システムの概要

本論文において提案するシステムは、ユーザが作業する対象の前でカメラをかざすと、それを写した映像に対して教師ビデオが重畳表示されるシステムである。ここでは、二つの基本的なコンセプトがある。

一つ目は、表示する教師ビデオ映像をユーザの作業進捗状況に合わせるということである。それにより、ユーザはビデオ映像の作業に取り残されるという心配を感じることなく作業ができる。

まず、図 1 のようにあらかじめ手動で手順ごとに教師ビデオを分割しておく。その後システム側は、図 2 に示すように、ユーザからのアクションを受けるまで、その分割された教師ビデオの一つの手順を繰り返し流し続ける。各手順の最後では、教師ビデオ映像をしばらく静止させて完成形を見せる。ユーザは、自分が一つの手順を達成したと思えば、システムに合図を送り、次の手順に進むことができる。また、前の手順が気になれば、戻ることも可能であり、一時停止することもできる。こうしたシステムを設計することにより、ユーザはシステムに取り残される心配を感じることがなくなる。また納得のいくまで教師ビデオを見続けることができる。

二つ目は、カメラと操作対象の位置関係に合わせて教師ビデオを適宜変形させ、作業空間へと重畳表示することである。その結果、教師ビデオが撮影された視線とユーザの視線とを合わせて重畳することができ、現実世界にあたかも教師となる人が操作しているかのような映像を見ることができ、ユーザは何を操作するのか、またそれをどのように動かすのかを直感的に知ることがで

きる。

このような視点変換を行うためには、ユーザの作業空間中の操作対象とカメラとの位置姿勢関係、および教師ビデオ中の操作対象とカメラとの位置姿勢関係を求めることが必要になる。本論文で提案するシステムでは、テーブル上で行うような作業を対象としており、テーブル平面の平面性を利用した Homography による視点変換を行う。

このとき、対象物体が完全に平面近似できないような立体物である場合には、平面性が保たれずに歪んで変換されてしまう場合がある。一般に、このような平面近似できない 3 次元物体を視点変換する際には、対象物体の形状を把握した上で射影変換を行う必要があるが、対象物体の 3 次元形状をあらかじめ取得しておくことは、従来研究において 3D の CG コンテンツをその都度用意することと同じように、新たな AR コンテンツを作る上での障壁となる。本システムの目的は、事前に誰か 1 人でも同じ作業をしていて、その様子をビデオカメラで撮っておけば、それがそのままコンテンツとして使用できる、という点である。従って、本システムでは 2 次元のビデオをコンテンツとして用い、その視点変換に平面射影変換行列である Homography を用いている。Homography によって変換した場合でも、テーブル面に接している部分に関しては正しく変換できるので、対象物体全てが完全に一致していなくても、支援内容は伝わると考えている。

また、作業中に教師ビデオ中に写る作業者の手は、テーブル面から離れているため Homography によって完全に視点変換されることは難しいが、本システムでは操作対象同士を並べて表示することが目的であるため、ユーザの手と作業者の手を重ねて表示する必要はないと考えており、ここでは考慮しない。

さらに、教師ビデオが作業空間に重畳表示されると、ユーザは作業をより直感的に理解することができる反面、図 3 のように、ユーザは自分が操作している手や物体と、

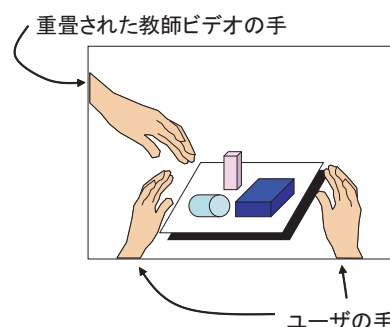


図 3 ユーザと教師ビデオとの見えの混同問題

Fig. 3 Appearance confusion problem of instruction video and user's view

3. 前 処 理

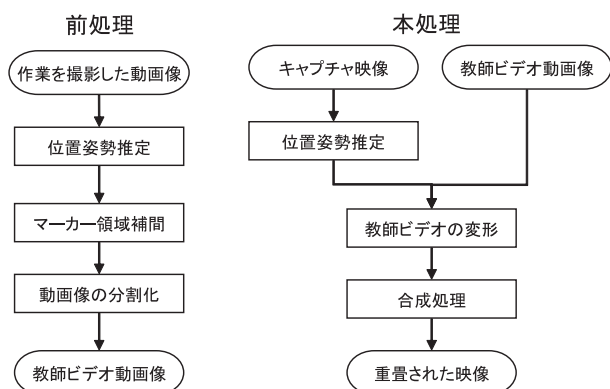


図 4 処理の流れ図

Fig. 4 Flow chart of proposed method

教師ビデオ側の手や物体とを混同してしまうといった問題が起こる．そのため、直感的に操作できるという利点を残しつつ、混同を如何にして防ぐかということについて取り組まなければならない．そこで本システムでは、教師ビデオに対して、透明化やエッジ強調などといった視覚効果を付加し、両者を並べて表示しつつも区別が可能になるようなコンテンツ作りを行う．

なお本システムでは、HMD などのウェアラブルディスプレイや、iPhone や iPad などの小型携帯ディスプレイ、据え置き型の大型ディスプレイなど、用途に応じて様々な表示デバイスを想定している．しかし、これらのディスプレイを用いる際には、それぞれのデバイスに依存した特性・問題が生じる．そのため本論文では、2次元のビデオを AR コンテンツとして用いたことに対する、コンテンツそのものの評価のみに焦点を当てる．

2.2 提案システムの流れ

本システムは、図 4 のように大きく二つの流れで構成されている．

- 教師ビデオの準備（前処理）
- ユーザによる作業（本処理）

前処理では、教師ビデオの作業空間とそれを写すカメラとの位置姿勢推定と、教師ビデオの加工を行う．この処理はオフラインで行われるため、リアルタイムで行う必要はない．

本処理では、作業空間とそれを写すカメラとの位置姿勢推定を行い、その幾何学的な位置関係を用いて教師ビデオの映像を変形し、教師ビデオの合成を行って AR 重畳表示をする．この処理はオンラインで行われるため、リアルタイムで処理を行う必要がある．

3.1 位置姿勢推定

ここでは、教師ビデオ中に映っている操作対象平面（主にテーブル平面）とユーザカメラの作業平面の位置姿勢関係を表す Homography を求める．対象平面とカメラの位置姿勢関係を求めるライブラリとしては ARToolkit¹⁵⁾ が広く知られている．これは平面形状の矩形マーカをカメラで撮影し、その矩形が画像上でどのような形に見えるかによって、そのマーカの属する平面とカメラの位置姿勢を推定することができる．しかし、マーカが画像上でフレームアウトしたり他の物体によるオクルージョンが発生すると、正しくマーカ領域を検出することができなくなり、位置姿勢推定がうまく行えないことがあるため、ある程度の範囲で視線を動かしながら作業を行う必要がある本システムにおいては、不向きであるといえる．

そこで本手法では、同様に位置姿勢関係を求めるマーカとして、ランダムな点群が紙面上に描かれたマーカ¹⁶⁾を用いる．これは、点群の局所的な配置状態を識別して、各点群をトラッキングすることにより、マーカが属する平面とカメラとの幾何学的な位置関係を推定をするものであり、多数ある点群の全てではなく一部が見えていれば検出ができるため、作業を行いながらも位置姿勢関係を求めることができる．

本システムでは、教師ビデオを撮影する際に、作業を行うテーブル上に、あらかじめ位置合わせマーカを設置しておくことを前提としている．この前提においては、現在既に存在しているビデオを用いることができなくなるが、今後教師ビデオを作成する際に、点群マーカを紙に印刷して置いておくだけで良いため、十分に許容できる条件であると考えられる．そして、教師ビデオは主に固定カメラで撮影されることが想定できるため、撮影され

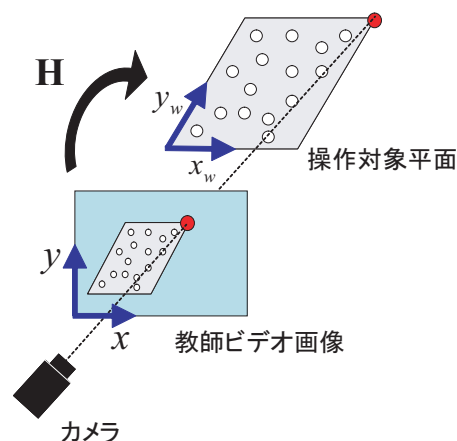


図 5 教師ビデオ画像と操作対象平面の関係

Fig. 5 Relationship of instruction video and real world



図 6 画像補完を行った教師ビデオ画像
Fig. 6 Inpainted image of instruction video

たビデオ中の 1 フレームについてマーカを検出し、マーカ平面と画像平面との Homography を推定する。この Homography は、作業支援を行う際に、ユーザのカメラと教師ビデオ映像とを位置合わせする際に用いられる。図 5 に教師ビデオ画像と操作対象であるテーブル（マーカ）平面の関係図を示す。

3.2 マーカ領域補完

ここでは、教師ビデオ中で撮影された点群の領域に対して、画像補完を行う。この処理は、本処理においてユーザ視点で教師ビデオを重畳するとき、マーカの点群が目障りになってしまうことが考えられるため、点群領域を補完して目立たないように加工する。作業支援中に置かれているマーカをリアルタイムで補完することは難しいため、ここでは教師ビデオ中の点群についての補完を行う。

画像補完とは、欠損領域の境界部分における色とテクスチャを用いて、それを欠損領域の内部へと広げて混ぜ合わせることによって補完する処理である。今回はマーカ点群部を欠損領域と見立てており、周囲の紙の色とテクスチャを用いて、図 6 に示すように、点群部分を補完した。これによって、教師ビデオを重畳したときにも、点群が目障りになる問題を軽減できる。

3.3 動画の分割化

最後に、教師ビデオの動画を手順ごとに分ける処理を行う。今回は、教師ビデオを提供する側が手動で分割することにしており、ビデオを見ながら各手順の開始フレームと終了フレームをそれぞれ指定することで行う。

4. 本 処 理

4.1 位置姿勢推定

ここでは、ユーザの視点に相当するカメラと作業対象であるテーブル平面との Homography を求める。本システムは教師ビデオと同じ環境で作業を行うこととしているため、あらかじめテーブル上にマーカを設置している。ここで推定した Homography は、前処理で求めた Homography と組み合わせることで、教師ビデオの変形

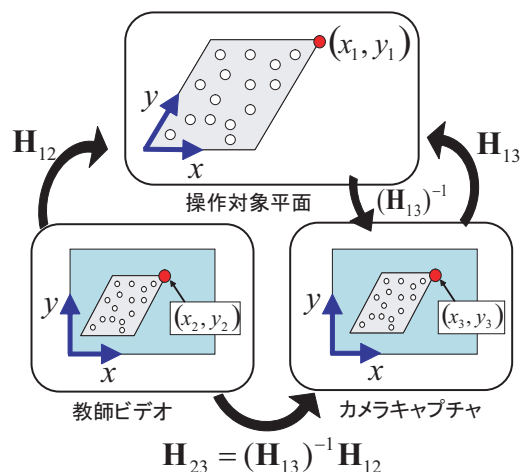


図 7 各座標間の対応関係
Fig. 7 Relationship of each coordinate system

と合成に用いられる。

4.2 位置合わせ

ここでは、教師ビデオの視線を、ユーザのカメラで撮影された映像の視線と合わせるために、教師ビデオの変形を行う。そして、変形された教師ビデオを、ユーザカメラの映像上に重畳表示する。これらの処理は、ユーザが作業している間リアルタイムに毎フレーム行われる。

図 7 に示すように、テーブル平面に設定した 2 次元の世界座標を $x_1(x_1, y_1)$ とする。また、教師ビデオで撮影された同じ点の画像座標を $x_2(x_2, y_2)$ とし、ユーザカメラが撮影した同じ点の画像座標を $x_3(x_3, y_3)$ とする。

このとき、教師ビデオ画像の座標と世界座標の関係は次のように表される。

$$x_1 \cong H_{12}x_2 \quad (1)$$

H_{12} は、前処理で得られる Homography であり、教師ビデオが固定カメラで撮影されている限り、変化しない。同様に、ユーザカメラで撮影したテーブル平面の座標と世界座標の関係は次のように表される。

$$x_1 \cong H_{13}x_3 \quad (2)$$

H_{13} は、本処理で得られる Homography であり、ユーザの視点に相当するカメラが絶えず動いているため、毎フレーム計算する必要がある。ここで、教師ビデオをユーザカメラの視線へと変換するために必要なのは、 x_2 から x_3 への Homography である。よって、式 (2) を変形して、

$$x_3 \cong H_{13}^{-1}x_1 \quad (3)$$

式 (3) に式 (1) を代入して、

$$x_3 \cong H_{13}^{-1}H_{12}x_2 \cong H_{23}x_2 \quad (4)$$

($H_{23} = H_{13}^{-1}H_{12}$ とする。)

このようにして、位置合わせのために必要な Homography である H_{23} を毎フレーム計算する。

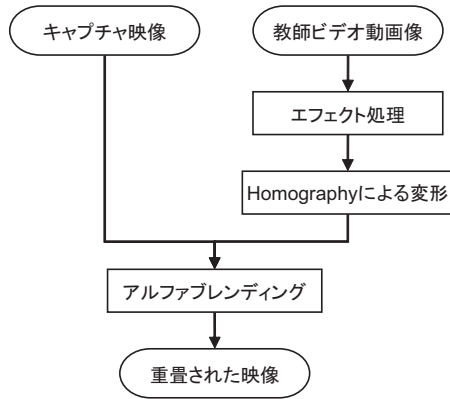


図 8 合成処理流れ図

Fig. 8 Flow of synthesis processing

4.3 合成処理

ここでは、教師ビデオの変形処理によって得られた Homography H_{23} を用いて、教師ビデオの画像をユーザカメラ画像へ合成する処理を行う。図 8 に処理の流れを示す。

エフェクト処理では、合成する教師ビデオの映像を、ユーザにとってより見やすくするような様々な処理を加える。これは、ユーザが必要に応じて加える処理である。処理の説明は次の章で詳述する。

Homography による変形では、ユーザの視点に合わせて教師ビデオ画像を適宜変形する。図 9 に示すように、教師ビデオ画像とユーザカメラ画像の幾何学的関係を示す Homography H_{23} (式 (4)) により、教師ビデオを現在のユーザ視点へと変換する。

アルファブレンディングでは、Homography によって変形された教師ビデオ画像とユーザカメラ画像が重なる部分に対して、明度値の割合を変化させて合成する。Homography によって変形された教師ビデオ画像中のある座標の明度値を $value_1(R_1, G_1, B_1)$ とし、同位置のユーザカメラ画像座標の明度値を $value_2(R_2, G_2, B_2)$ と

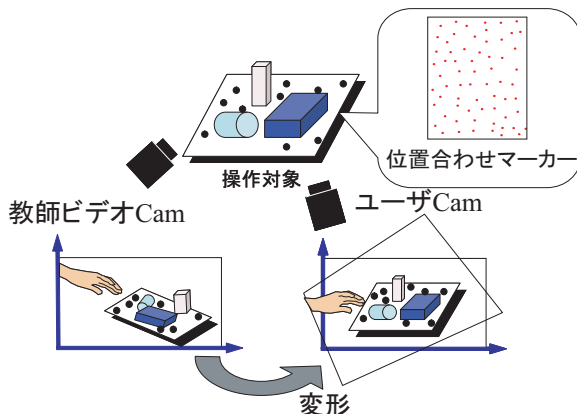
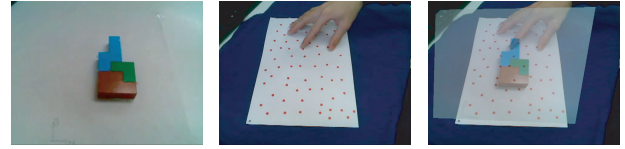


図 9 教師ビデオの重畳

Fig. 9 Overlaying instruction video onto user's view



(a) 教師ビデオ画像 (b) キャプチャ画像 (c) 重畳表示画像

図 10 画像補完した教師ビデオとの合成処理適用例

Fig. 10 overlay on image inpainted instruction video example

すると、合成後の $value(R, G, B)$ は以下ようになる。

$$value = \alpha \cdot value_1 + (1 - \alpha) \cdot value_2 \quad (5)$$

図 10, 11 に合成後の結果を示す。

5. エフェクト処理

ここでは、エフェクト処理について詳述する。教師ビデオ映像と実際に自分が見ている空間であるユーザカメラ映像の合成された映像は、ユーザにとって分かりやすくなる反面、どちらが自分の操作対象なのかの混同問題が生じる原因ともなってしまう。そこで、自分の操作対象がどちらであるかをより区別できるようにするため、様々なエフェクトを加えた。これは、ユーザが必要に応じて加える処理である。

5.1 透明度の変更

本システムでは、教師ビデオを重畳表示する際に、アルファブレンディングを行っており、それぞれの α 値の割合は式 (5) で表される。ユーザは、この α 値を変化させることで、両映像の優先度を決定することができ、教師ビデオ映像を優先して見たければ教師ビデオの透明度を下げ、反対に、自分の手元映像が見たい場合は教師ビ

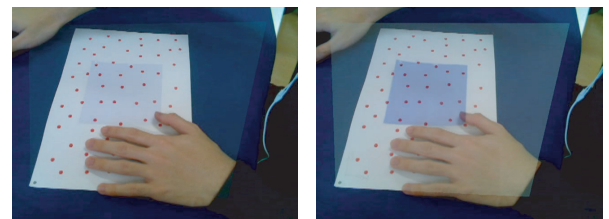
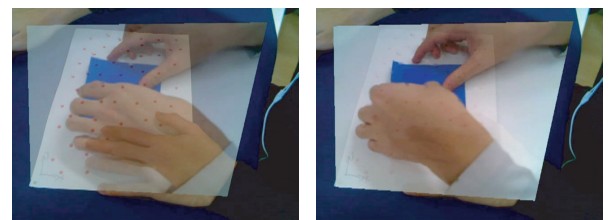
(a) $\alpha=0.1$ (ユーザ手元優先) (b) $\alpha=0.3$ (ユーザ手元優先)(c) $\alpha=0.7$ (教師ビデオ優先) (d) $\alpha=0.9$ (教師ビデオ優先)

図 11 各透明度での合成画像

Fig. 11 Synthesis images with each transparency

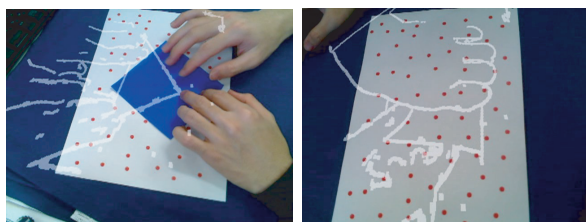


図 12 輪郭線を表示した合成結果

Fig. 12 Synthesis results with visualizing contour of instruction video

デオの透明度を上げることができる。これにより、ユーザは自由に見たい方の映像を見たいときに見ることができる。

5.2 透明度の点滅

ここでは教師ビデオの透明度を周期的に変動させる処理について説明する。式 (5) で表されるアルファブレンディングの α 値を時間ごとに周期的に変化 ($\alpha=0.0$ から $\alpha=1.0$ へ, $\alpha=1.0$ から $\alpha=0.0$ へ変化させることを繰り返す。) させると, 自分の操作対象と教師ビデオが交互に周期的に現れるようになるので, 区別が容易になる。このエフェクトは, 操作対象が動く (折り紙の折り方を示す教師ビデオなど) 場合は, その動きを見失ってしまうことが考えられるが, 逆に教師ビデオの操作対象が動かない (レゴブロックの配置を促す教師ビデオなど) 場合には効果的であると考えられるため, 必要に応じて用いることができる。

5.3 輪郭線の表示

ここでは, 教師ビデオをより区別させるために, 教師ビデオ中の輪郭線情報を表示することを試みた。まず, Canny のエッジ検出¹⁷⁾を用いて教師ビデオ中のエッジを抽出する。この検出法はノイズによる誤検出が少なく, 検出位置が正確であることが知られている。図 12 に, 教師ビデオの輪郭線を重畳表示し合成した画像例を示す。

また図 13 のように, 背景色によって輪郭線の色を変更することで, 様々な背景であっても対応可能である。

5.4 輪郭線の強調

前節で述べた輪郭線のための重畳表示は, 教師ビデオと自分の作業空間の混同を避けることはできても, 教師ビデオの詳細な情報が読み取れない場合が考えられる。そこで, 元の教師ビデオを表示しつつ輪郭線を表示するこ

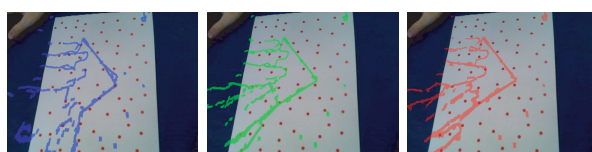


図 13 輪郭線の色を変更した合成結果

Fig. 13 Synthesis results with visualizing contour in various colors

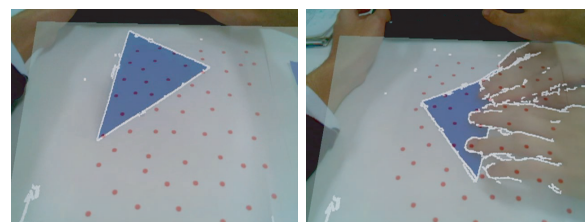


図 14 輪郭線を強調した合成結果

Fig. 14 Synthesis results with highlighted contour

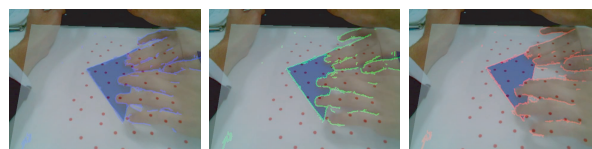


図 15 輪郭線の色を変更した合成結果

Fig. 15 Synthesis results with highlighted contour in various colors

とで, 元の情報を残しつつも教師ビデオと実空間の混同は避けられるのではないかと考えた。そこで, 同様にエッジを抽出し, エッジ (輪郭線) 領域以外は教師ビデオの映像をそのまま出力した図 14,15 に示す。

さらに, 輪郭線の太さも調整可能であり, その太さを周期的に変動させることで, 輪郭がまるで光っているかのような視覚効果をユーザに与えることができ, 教師ビデオとの識別がより容易になると考えられる。図 16 にその適用例を示す。

5.5 シフト AR 表示

作業支援を行う際に, その操作対象によっては, 教師ビデオを完全に自分の作業空間に重ねてしまうのではなく, 横にずらして重畳表示したい場合も存在する。例えば操作している対象自体が変形する教師ビデオ (折り紙の折り方などの) の場合, ユーザが操作している対象の上に教師ビデオが重畳されていると, ユーザは自分の操作状況が見づらくなるばかりでなく, 教師ビデオの操作も見づらくなってしまうことが考えられる。そこで, 図



図 16 輪郭線の太さを変更した場合の適用例

Fig. 16 Synthesis results with highlighted contour in various thickness

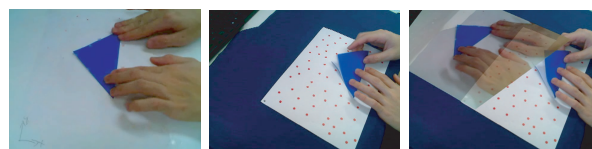


図 17 横にずらして重畳した合成結果

Fig. 17 Overlaying results with shift

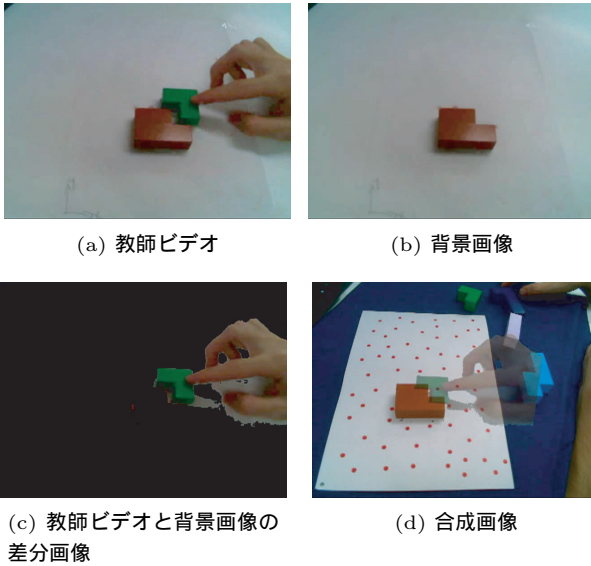


図 18 変化領域のみを重畳表示した合成結果

Fig. 18 Synthesis result with overlaying only changed area

17 のように横にずらして表示することによって、自分の隣で作業を教えてもらっているような支援ができ、教師ビデオの操作、ユーザ自身の操作の双方の動きが分かりやすくなる。

5.6 注目領域の表示

ここでは、教師ビデオの操作対象の中で、新たに変化が起きた部分のみを見たいときに効果的だと考えられる、注目領域の表示について説明する。

本システムで撮影された教師ビデオを変形してそのまま重ね合わせているだけであるため、背景などの操作対象以外の不必要な領域も重畳されてしまう。また、教師ビデオ中で全く変化が起こらない部分も常に表示されているため、新しく変化した箇所を優先的に表示した方が、より分かりやすくなると考えられる。そこで、教師ビデオの中から注目領域のみを抽出した映像を生成する。ここでは、変化の起きた領域を注目領域と考え、各手順ごとに教師ビデオ中から抽出する。

ここでは、変化領域検出法として広く用いられている背景差分法を用いる。背景差分法で用いた画像の表色系を、照明変化にもロバストな CIE $L^*a^*b^*$ 表色系で行った。以下にその結果、合成画像を示す。

この効果は主にブロックの配置など、パーツが増えていく指示を促す教師ビデオに適していると考えられる。

5.7 RGB 値のスワップ処理

ここでは、重畳する教師ビデオの色を極端に変化させることでユーザの操作と区別するための、RGB 値のスワップ処理について説明する。

教師ビデオの RGB 値をそれぞれ入れ替えることで、現実にはあり得ない映像になるため、混同を防ぐことが



(a) スワップなし (b) RGB → GBR (c) RGB → BRG

図 19 RGB 値のスワップ処理後に重畳した合成結果

Fig. 19 Synthesis results with RGB swapping

期待できる。RGB → GBR の入れ替え、RGB → BRG の入れ替えを行った処理例を図 19 に示す。

6. AR アノテーション

ここでは、AR アノテーションについて詳述する。AR アノテーションとは、AR 表示された注釈情報のことであるが、本システムでは、教師ビデオの操作などに用いるためのボタン、トラックバーを AR 表示する。以下にその詳細を記す。

6.1 AR ボタン

ここでは、AR ボタンについて説明する。このボタンの役割は 2.1 節で示した概念の通り、手順ごとに分割された教師ビデオの中で、前の手順や次の手順をユーザが見たくなったとき、システムへ合図を送るために用いる。次の手順の教師ビデオを要求するときは next ボタン、前の手順を要求するときは back ボタンが用いられる。ボタンの表示位置は、図 20 に示すようにマーカ座標を基準として決定している。

ボタンはマーカの左下を基準とした $x_w y_w$ 平面上で、 $(x_1 \leq x_w \leq x_2), (y_1 \leq y_w \leq y_2)$ に back ボタン、 $(x_3 \leq x_w \leq x_4), (y_3 \leq y_w \leq y_4)$ に next ボタンを表示している。ボタンが押されたかどうかは、その領域に手が入ったときに押した判定がなされる。まず、マーカ座標 $(x_1 \leq x_w \leq x_2), (y_1 \leq y_w \leq y_2)$ 領域を、式 (3) によってユーザ視点に変換する。その後、手の領域を認識するために肌色領域検出を用い、「ボタンが変換された領域」に対して肌色領域が 9 割満たされたときにボタンを押したと判断する。

肌色領域検出では、RGB 色空間で肌色領域判定を行

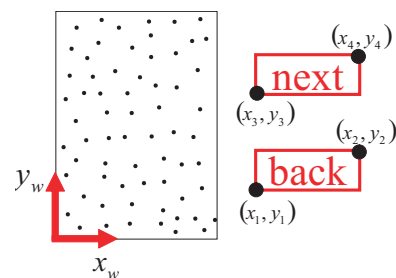


図 20 AR ボタンと座標系

Fig. 20 AR button and its coordinate system

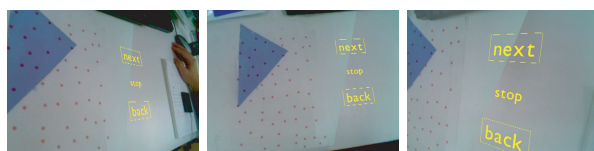


図 21 様々な視点から見た AR ボタン

Fig. 21 AR button looking from various view points

うと誤検出が多くなるため、HSV 色空間、YUV 色空間を組み合わせて行う。Sherrah ら¹⁸⁾により HSV 色空間で肌色領域は、照明変化によらず ($0 \leq H \leq 30$) に多く分布されるということが示されている。そこで、肌色を決めるしきい値としてこの値を用いた。また、YUV 色空間では、中川ら¹⁹⁾が肌色抽出の際に用いたしきい値 ($64 \leq Y \leq 240$), ($94 \leq U \leq 125$), ($131 \leq V \leq 255$) とした。

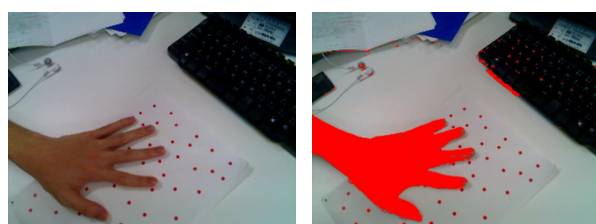
肌色領域検出結果を図 22 に示す。また、ボタン領域のみに対して肌色検出をおこなった例を示す。図 23(a) ではボタン領域に対して 9 割以下であり、図 23(b) は 9 割以上でボタンが反応している。

6.2 トラックバー

さらに本システムでは、教師ビデオがどこまで進んだかを示すため、図 24 のようなトラックバーを AR 表示する。教師ビデオ全体からの現在位置を青点で示している。これにより、ユーザは現在どこまで進んだのか、また、後どの位で終わるのが分かるようになっている。

7. ユーザ評価実験

提案システムの有効性を示すために、図 25 に示すような環境でユーザ評価実験を行った。今回使用した PC

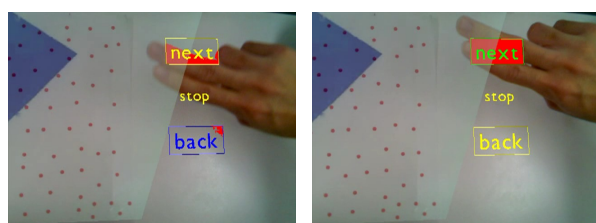


(a) カメラ画像

(b) 肌色検出画像 (赤)

図 22 肌色検出の結果

Fig. 22 Skin color detection



(a) 検出例:90 % 以下
(ボタン反応なし)

(b) 検出例:90 % 以上
(ボタン反応あり)

図 23 AR ボタン操作の検出の結果

Fig. 23 Detection of operating AR button

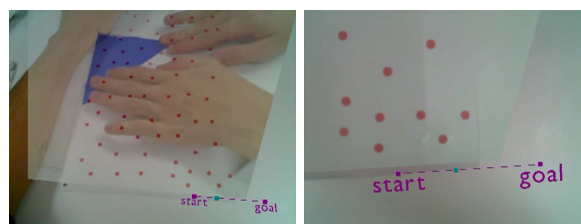


図 24 トラックバー

Fig. 24 Track bar

の仕様は以下の通りである。なお本実験では、本システムのコンテンツに関するの評価を行うため、教師ビデオに対してのエフェクト処理および視点変換の有効性に関するのユーザ評価を行う。

ここでは、据え置き型のディスプレイを表示デバイスとして用いた。これは、HMD や iPhone などの表示デバイスと比較して、ほとんどのユーザが最も慣れ親しんでいるディスプレイであるため、デバイス自身を利用することに付随する問題が生じにくいと考えたためである。また、ユーザのカメラはユーザの首元に装着した。首元に付けているため、ユーザの頭部が動いて前方のディスプレイを見た場合にも、カメラは作業空間である斜め下を向いている。また、ユーザが上半身を傾けると一緒に視点が移動するため、視線方向はユーザに追従することになる。

- カメラ：Logitech 社 web カメラ
(キャプチャ映像、教師ビデオ)
- カメラ解像度：640 × 480
(キャプチャ映像、教師ビデオ)
- PC：Dell Inspiron 530
- CPU：Intel Core 2 Duo 2.80GHz, 2.79GHz
- メモリ：3.25GB
- OS：Windows XP
- 開発環境：Visual Studio 2005 VC++ (Intel 社 画像処理ライブラリ (OpenCV) を使用して実装)

7.1 本システムの作業の視認性実験

本システムの目的である、AR 技術を利用した教師ビデオの重畳表示の有効性、また、重畳したときに生じる

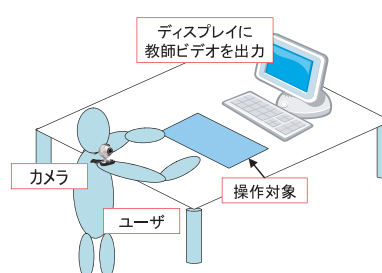


図 25 実験環境

Fig. 25 Experimental environment

視覚的混同問題を解決するためのエフェクト処理の有効性を示すために、2つの作業シーンにおいて作業の理解度を5段階で評価する主観評価実験を行った。

- 色付きパズルの配置支援
- 折り紙の折り方支援

パズルの配置支援では、色ごとに形の異なるパズルを順番に配置していくことで、ある形を作るという作業を教師ビデオとして用意した。折り紙の折り方支援では、折り紙で箱を作る作業を教師ビデオとして用意した。箱を作る折り紙は10手順に分けられ、細かい作業のない単純なものである。

<実験概要>

実験環境は図25に示す通りである。目の前に置かれたディスプレイで、ユーザ視点映像に教師ビデオがそれぞれ半透明で重畳表示された映像を見て作業を行うものである。被験者3人に対して本システムを実行し、その時、被験者は自由にエフェクト処理を選んでいき、その処理がこれらの作業シーンに向いているかどうかを5段階で評価した。

また、平面近似できない3次元物体を対象とした場合に、Homographyのみで視点変換を行った結果がどのように理解度に影響するのかを評価した。ここでは、パズルを12cmの高さに重ねる作業を1つの視点から撮影して教師ビデオとして用意しておき、被験者がカメラを対象物体の周囲で図26のように動かして視点を変えたときに、その視点に合わせて変形された教師ビデオを見て、平面拘束を用いていることによる歪みの影響があるかどうかを5段階で評価した。

<実験結果>

2つの作業シーンにおける教師ビデオの重畳表示の主観評価の結果と、それぞれのエフェクト処理の主観評価

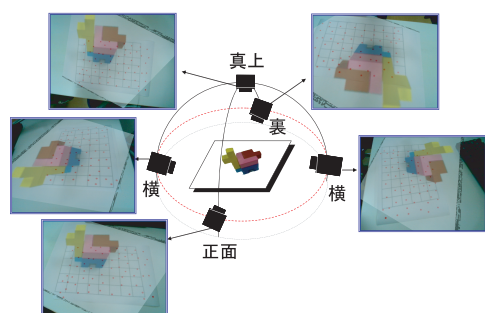


図26 平面射影変換による歪みの影響評価におけるユーザカメラ視点

Fig. 26 User's viewpoint for evaluating Homography distortion.

表1 エフェクト処理に対する主観評価結果

Table 1 Results of subjective evaluation of visual effects

エフェクト	パズル配置	折り紙
重畳表示	4.3	1.7
シフト AR 表示	2.7	4.0
透明度の変更	3.0	2.0
透明度の点滅	2.3	1.3
輪郭線の強調	3.3	2.3
注目領域の表示	4.0	1.0
RGB 値のスワップ	1.3	3.3

の結果をまとめたものを表1に示す。この作業に非常に向いていると思えば5を、この作業に全く向いていないと思えば1を選んでもらった。

重畳表示の評価結果では、パズル配置のように操作対象の位置が作業における重要な情報になる場合は、高い評価が得られている。しかし、折り紙の折り方のように、操作対象の形と作業者の手による操作そのものが重要な情報になる場合には、低い評価であり、教師ビデオの折り紙と自分が折る折り紙との見分けが困難であるという回答が得られた。一方で、ビデオを少しずらして表示させるシフト AR 表示においては、反対に折り紙で高い評価が得られている。これは前述の通り、隣で折り方を教えてもらっているような表示ができている効果であるといえる。よって、作業対象によって、完全に重畳するかシフトさせて表示するかを選択することで、操作対象に適した直感的な支援ができると考えられる。

また、パズル配置の場合では、注目領域の表示を行った場合において、高い評価が得られている。RGB 値のスワップで低い評価であるのは、今回用いたパズルは色付きであるために、色を変えてしまう RGB 値のスワップは本来持つ重要な情報を失うことになるためであると考えられる。折り紙の場合では、透明度の変更、透明度の点滅、輪郭線の強調、注目領域の表示を行った場合には、低い評価であり、単に重畳表示を行った場合とさほど変わらないことが分かる。これも、教師ビデオの折り紙と自分が折る折り紙との見分けが困難であるということが要因であるといえる。しかし、シフト AR 表示、RGB スワップにおいては、どちらかというやや向いている方に評価が得られている。これは、教師ビデオの折り紙と自分の折り紙との見分けがつくようになったからであるといえる。

このように、それぞれの操作対象に応じて適した表示方法およびエフェクト処理を選ぶことで、効果的な支援ができることが分かった。今後、他の作業を対象とする際には、ここで得られた指標を基にして、表示方法やエ

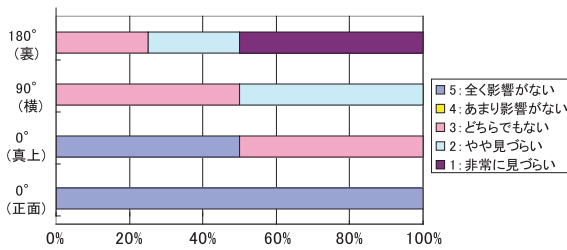


図 27 平面射影変換による歪みの影響に関する主観評価結果

Fig. 27 Result of subjective evaluation for Homography distortion.

フェクトを決定していくことが有効であると考えられる。

さらに、平面近似できない 3 次元物体を対象とした場合の評価結果を、図 27 に示す。この結果より、ユーザの視点によっては、高さのある物体を Homography によって視点変換しても、視認性にはほぼ影響を与えていないことが分かる。しかし、いくつかの視点においては歪みが影響する場合もあることが分かった。このような物体に対して、3 次元形状を持たずビデオのみで対処するためには、1 視点ではなく複数の視点で教師ビデオを撮影しておき、ユーザの視点に応じてビデオを選択するといったアプローチが想定できる。今回の結果から考えると、正面と裏側の最低 2 視点から撮影しておけば、歪みの影響を受けにくい視点変換ができると考えられる。

7.2 教師ビデオの向きと作業の視認性実験

本システムでは、教師ビデオをユーザの視点の向きに合わせて表示をすることで、作業を効率的に支援できると考えている。そこで、その有効性を確かめるための評価実験を行った。また、その教師ビデオがユーザ自身の視点からビデオの角度がどれだけずれると視認度に影響を与えるかを確かめる実験も行った。

< 実験概要 >

実験環境は前述の実験と同様であるが、今回は教師ビデオの向きのみに焦点を当てるため、ディスプレイにはユーザ視点の映像は表示せず、向きを変化させてある教師ビデオのみを表示して、それを見ながら作業を行った。ここでは、被験者 10 人に難易度が同程度の 2 通りの折り紙を折るビデオ A と B を提示し、それぞれビデオと同じ作業を行ってもらった。今回使用したビデオの中では、被験者の事前知識や経験に左右されないようにするために、特定のものを折るのではなく、無作為に数回折り続ける作業を行った。従って、被験者は自分が何を折っているのか分からず、次の手順も予測できないため、ビデオの操作についていかなと折ることができないよう

になっている。また、どちらも 7 回の手順で構成された作業となっている。

このとき、被験者を 2 グループに分け、一方のグループにはビデオ A をユーザ視点に合わせて表示し、ビデオ B を合わせずに表示した。またもう一方のグループには、ビデオ B をユーザ視点に合わせて表示し、ビデオ A を合わせずに表示した。そして、被験者には視点が合わせられたビデオとそうでないビデオで、どちらが視認しやすいと感じるのかについて 5 段階で評価する主観評価を行った。図 28 に結果を示す。

また、両方で作業速度に差が生じるかどうかを確かめるために、被験者 10 人の作業時間を計測した。なお、本システムではユーザが次の手順へと進める合図を行わない限り、ビデオが同じ手順を繰り返すようになっているため、ユーザが作業内容を理解できるまで、ビデオを繰り返すことになり、所要時間が増えていくことになる。図 29 に、それぞれのビデオでの作業時間を基準とし、その基準の時間と比べて視点合わせ有りと無しとのときで、どれだけ作業に対して平均時間から相対的な所要時間率（これは平均時間からの相対的な時間であるためこの値が 1 のときは平均的な時間ということになる。また平均時間は、被験者 10 人の作業時間を基にしている。）を比較した結果と、そのとき繰り返して見た回数をプロットした結果を示す。

さらに、ユーザ視点と教師ビデオの向きがどのように視認性に影響するのかを確認するために、教師ビデオの視点をユーザ視点から 10° ずつ回転させていき（180° 回転させた場合、教師ビデオの操作が逆向きの状態になる）、同じく 5 段階で評価する主観評価を行った。図 30 に結果を示す。

< 実験結果 >

図 28 の結果より、教師ビデオが、ユーザの視点に合わせている場合と、そうでない場合では、視点が合わせられている方がより視認しやすく感じるという結果

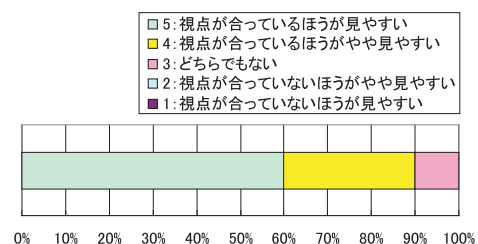


図 28 視点合わせに対する主観評価結果

Fig. 28 Result of subjective evaluation for view points of instruction video

表 2 視点合わせに対する作業経過時間率の平均と分散

Table 2 Average and Variance in relative hand working time rate of each subject for view points of instruction video

	作業経過 時間率の平均	作業経過 時間率の分散
位置合わせあり	0.54	0.02
位置合わせなし	1.46	0.94

が得られた。

表 2, 図 29 より, 10 人中 9 人は視点合わせありの方が作業を早く終えていることが分かる。また, 位置合わせありの場合には 10 人とも平均以下の時間で終えているのに対して, 位置合わせなしの場合は個人差が大きく, 人によっては何度も繰り返し見る必要があった。従って位置合わせありの方がより作業内容がユーザに理解されやすいことが分かる。

また図 30 より, 教師ビデオを回転させることによる視認性については, 自分の手元の向きに合わせられている状態から, $\pm 30^\circ$ 程度の範囲では視認性が高く, それを超えると視認性が落ちていることが分かる。自分の向きと正反対 (180°) のときには, 動きが鏡に映った自分のような状態になるため, 想像が付きやすいため分かりやすいと答えるケースや, 方向感覚が狂うので分かりにくいという答えなどがあったため, 被験者によって評価のばらつきが大きくなっている。一方で, ユーザの視点に近い 0° 付近でのばらつきは非常に小さく, 被験者に依らず評価が高い。また, 95 %信頼区間で (5: 分かりやすい) ~ (3: どちらでもない) が得られる角度は, 自分の手元の向きに合わせられている状態から $\pm 8^\circ$ である。従って, 教師ビデオをユーザの視点へ合わせて変換して表示する本システムの提示方法は, 多くのユーザに対して有効であるといえる。

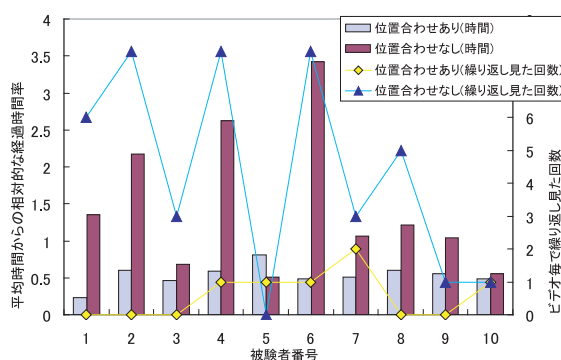


図 29 視点合わせに対する各被験者の平均時間からの相対的な作業経過時間率と被験者毎の繰り返し見た回数

Fig. 29 Relative hand working time rate of each subject for view points of instruction video

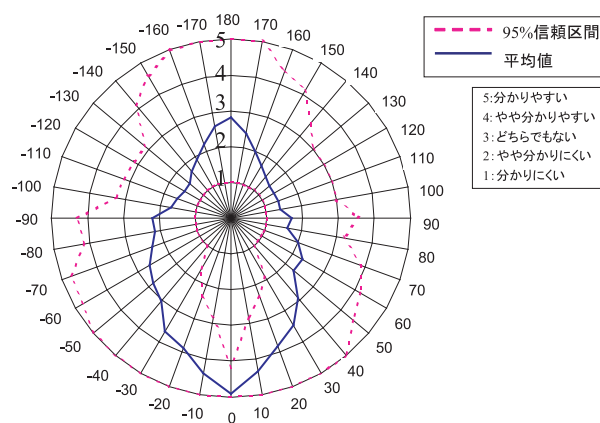


図 30 教師ビデオの回転による視認性評価結果

Fig. 30 Result of visibility evaluation for rotation of instruction video

8. む す び

本論文では, AR 表示コンテンツとして既存のビデオを利用可能な作業支援システムを開発した。また, AR 表示の際に起こる問題点である, ユーザ自身の操作とビデオ映像が視覚的に混同してしまう問題を解消するために, 教師ビデオに様々な視覚効果を与えることで, ユーザの視認性を向上させる提示方法を提案した。さらに, システム側がユーザの操作完了を待つような工夫を施し, ユーザがインタラクティブにビデオを操作できる AR ボタンや, 現在の進捗状況が知覚できるトラックバーなどを導入した。

本システムの有効性を確認するための評価実験では, 2 つの作業シーンを想定して評価を行った。折り紙の折り方支援では, 教師ビデオとユーザの作業をいかに区別して表示するかが重要であり, 操作対象の上に完全に重畳する必要はないが, 操作する向き (視点) が合わせられていないと分かりにくくなってしまうため, ややずらした場所に視線を合わせて重畳するのが効果的であることが分かった。またパズルの配置支援では, 教師ビデオとユーザの区別も必要であるが, 位置そのものが合っていることが直感的理解には重要であるため, 透明度やエッジ強調などのエフェクトを利用して区別を行うことが効果的であることが分かった。

このように, 支援する作業内容に応じてどのような見せ方が効果的であるかというコンテンツに関しての一定の指標を得ることができた一方で本システムでは, 作業内容に応じて, HMD などのウェアラブルディスプレイや, iPhone のような小型携帯ディスプレイ, 据え置き型の大型ディスプレイなど, 様々な表示デバイスを想定している。従って, コンテンツ内容と各種の表示デバイスとの組み合わせについての検討を今後行っていく予定で

ある．本論文では，表示デバイスに依らない評価を行ったが，今後の課題として，それぞれの表示デバイスを適用した場合に起こる影響を検討する．

参考文献

- 1) R.T. Azuma et al. A survey of augmented reality. *Presence-Teleoperators and Virtual Environments*, 6(4):355–385, 1997.
- 2) SK Ong, ML Yuan, and AYC Nee. Augmented reality applications in manufacturing: a survey. *International journal of production research*, 46(10):2707–2742, 2008.
- 3) R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, 21(6):34–47, 2001.
- 4) T. Hollerer, S. Feiner, T. Terauchi, G. Rashid, and D. Hallaway. Exploring MARS: developing indoor and outdoor user interfaces to a mobile augmented reality system. *COMPUT GRAPHICS(PERGAMON)*, 23(6):779–785, 1999.
- 5) W. Narzt, G. Pomberger, A. Ferscha, D. Kolb, R. M
"uller, J. Wiegardt, H. H
"ortner, and C. Lindinger. Augmented reality navigation systems. *Universal Access in the Information Society*, 4(3):177–187, 2006.
- 6) M. Flagg and J.M. Rehg. Projector-guided painting. In *Proceedings of the 19th annual ACM symposium on User interface software and technology (UIST)*, pages 235–243, 2006.
- 7) C.M. Robertson, B. MacIntyre, B.N. Walker, and GVV Center. An evaluation of graphical context when the graphics are outside of the task area. In *7th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 73–76, 2008.
- 8) 元川 洋一, 斎藤 英雄. ギター演奏支援のための構造特徴追跡を利用したマーカレス AR 表示. 日本バーチャルリアリティ学会論文, 13(2):267–278, 2008.
- 9) T. Salonen and J. Saaski. Dynamic and Visual Assembly Instruction for Configurable Products Using Augmented Reality Techniques. *Advanced Design and Manufacture to Gain a Competitive Edge*, page 23, 2008.
- 10) S. Nilsson and B. Johansson. Acceptance of augmented reality instructions in a real work setting. In *CHI'08 extended abstracts on Human factors in computing systems*, pages 2025–2032. ACM, 2008.
- 11) S.J. Henderson and S. Feiner. Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In *Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 135–144, 2009.
- 12) "How To Videos — Graspr.com"
. <http://www.graspr.com/>.
- 13) "Instructables - Make, How To, and DIY".
<http://www.instructables.com/>.
- 14) 木村篤信, 黒田知宏, 眞鍋佳嗣, 千原國宏. 動作学習支援システムにおける視覚情報提示方法の一検討. 日本教育工学会論文誌, 30(4):293–303, 2007.
- 15) H. Kato and M. Billinghurst. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *Proceedings of the Second International Workshop on Augmented Reality (IWAR)*, pages 85–94, 1999.
- 16) H. Uchiyama, H. Saito, V. Nivesse, M. Servieres, G. Moreau, and E.C.N.C. IRSTV. AR representation system for 3D GIS based on camera pose estimation using distribution of intersections. In *Proc. ICAT*, pages 218–225, 2008.
- 17) J. Canny. A computational approach to edge detection. In *IEEE Transactions Pattern Analysis and Machine Intelligence*, 1987.
- 18) J. Sherrah and S. Gong. Skin Colour Analysis. *University of Edinburgh*, 2001.
- 19) 中川弘隆. 高次局所自己相関特徴による高速画像認識モジュールの開発と自律走行型ロボットへの応用. 3 月 2002.

(2010 年 8 月日受付)