高速プロジェクタとウェアラブル型トラッキングに基づく低遅延な空間型拡張現実の検討

天野 光[1,a)]　渡辺 義浩[1]　石川 正俊[1]

概要：空間型拡張現実 (Spatial Augmented Reality，SAR) は，日常的な空間へ環境に応じた情報を自在に投影できる技術として注目されている．しかし，既存システムでは，プロジェクタの映像投影やセンシング時における処理の遅延により，人間の身体動作と遅延なく協調した情報提示を行うことは困難であった．身体動作に対する情報提示の遅延が観測者に知覚されると，観測者が提示された映像に対し違和感を感じ，映像酔いなどを生じる原因となる．したがって，人間の身体動作に対して，遅延なく情報提示を行うことは重要である．そこで，本稿では，高速プロジェクタと高速な画像センシングを組み込むことで，スループット 1000 fps，遅延 ms オーダの SAR システムを構築できる可能性を示す．さらに，このような応用では，人間の視点や動作を広い範囲で高精度かつ高速に捉える必要がある．そこで，本稿では SAR における視点把握のために，ウェアラブルな高速ビジョンと高速プロジェクタを組み合わせる構成を提案する．

Low-latency Spatial Augmented Reality Based on High-speed Projector and Wearable Type Tracking

Hikaru Amano[1,a)]　Yoshihiro Watanabe[1]　Masatoshi Ishikawa[1]

Abstract: Spatial augmented reality is useful technique due to project information corresponding to the environment. However, with existing systems, it was difficult to realize a system that interacts with people because the system has delay of projection and sensing technologys. When the observer perceives the delay of the image, the observer feels a sense of incompatibility and motion sickness. Therefore, it is important to present images without delay. In this research, by incorporating a high-speed projector and high-speed image sensing, we show the possibility to construct the system that throughput is 1000 fps and latency is millisec order. Additionally, in such applications, we need to caputre the human viewpoint and motion with high accuracy and high speed in a wide space. Therefore, in this paper, we propose a configuration combining a wearable high-speed vision and a high-speed projector due to capture the observer viewpoint.

1. はじめに

プロジェクタを用いた投影型のディスプレイによって，実空間を没入型の仮想環境に変換する応用や，日常空間の実物体に応じた情報を提示する空間的拡張現実 (Spatial Augmented Reality，SAR) [1] のような応用は，広い分野でニーズが高い．前者には，CAVE [2] がある．このシステムは，特殊なスクリーンやセンサなどの専用の装置を用いることによって，没入感の高い仮想環境を構築している．また，後者の SAR の枠組みには，様々な研究がある [3], [4]．これらのシステムは，環境側に空間全体をセンシングするデプスセンサを設置することで，日常的な空間を没入型の空間に変換し，人間とのインタラクションを可能としている．このように，技術の進展とともに，業務的な専用システムから，日常的な空間に適用しようとする流れが生み出されており，今後ますますの発展が期待できる．

一方で，このような投影型の仮想情報提示技術におけるシステム遅延には課題が残されている．例えば，Ng らは，人間の投影遅延の知覚実験を行い，被験者平均で 6.04 ms

[1] 東京大学 情報理工学系研究科
Graduate School of Information Science and Technology, University of Tokyo
[a)] hikaru_amano@ipc.i.u-tokyo.ac.jp

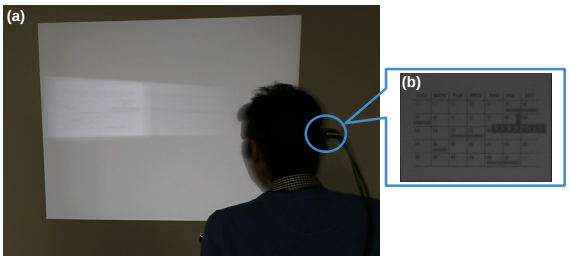

図 1 観測者の視界のある位置に対し，静止して見えるようにスケジュールを表示している．(a) は動いている観測者に対して，映像も静止して見えるように動いている様子である．(b) はウェアラブルビジョンで投影映像を撮像したときの様子である．

Fig. 1 We displayed the schedule seen in static state in sight. (a) The observer and the image are moving together for making the image seen in a static state. (b) The image captured with wearable camera

以上の投影遅延が生じると人間に遅延が知覚されるという結果を報告している [5]．これはタッチパネルの入力動作に限定された議論ではあるが，ビデオシースルー型の Head Mounted Display (HMD) [7] や，プロジェクションマッピングの遅延解消 [8] においても目標性能として参考にされている．このような流れの下，投影型の仮想情報提示技術のシステムにおいても同様の性能が有効であることが期待できる．しかし，SAR の枠組みの 1 つの RoomAlive [3] で使用しているプロジェクタは最大 144 fps であるから，最低でも 7 ms 程度の投影遅延が発生する．また、投影遅延とともに，センシングも課題を抱えていると考えられる．このタイプの技術では，広い範囲で対象の動作を捉えられるとともに，遅延解消の観点から高速なセンシング技術が必要となる．上述のシステムでは，いずれも環境側に設置されたセンシング装置から対象の運動を捉えるタイプのものであり，範囲制約とともに，推定精度を上げるために処理量が増大し，高速性を満たすことが難しかった．

以上の背景に基づき，本稿では，人間の身体運動の中でも，視点の変化に焦点を当て，高速な視点把握のために，高速プロジェクタとウェアラブルビジョンに基づいた構成を提案する．高速プロジェクタは，従来のプロジェクタの投影遅延の問題を解消し，また，映像とマーカの高速投影を行うことで，観測者に不可視なマーカを提示映像に埋め込む．さらに，このマーカを人間の頭部に設置された高速ビジョンにより撮像することで，広い空間において，高速高精度なセンシングを達成できなかったという問題を解消する．実験の結果，システムの処理におけるスループットとしておよそ 530 fps，レイテンシとしておよそ 5.43 ms の性能での実現の可能性を示す．また，図 1 のように，観測者に対して静止して見えるような情報提示を可能とするアプリケーションを実装する．

2. 関連研究

2.1 Spatial Augumented Reality

SAR は，プロジェクタを用いることで，実世界の空間に存在する物体に情報提示を行い，現実を拡張する．この技術により，日常的な空間で仮想環境を実現するシステムや，それに加えてインタラクションすることが可能なシステムが開発されている．例えば，Illumiroom [4] や Invoked computing [6] がある．これらのシステムは，日常空間にある実世界の物体に対し，適切な映像投影を行うことで仮想的に空間を拡張し新たな体験を可能とするシステムとなっている．また，日常的な空間を没入型の仮想環境に変えて，インタラクションするシステムとして，RoomAlive [3] がある．このシステムは，環境側にセンサを設置し人間の運動を推定することで，日常的な空間でインタラクションすることを可能としている．しかし，想定されているプロジェクタによる投影遅延の問題や，環境全体をセンシングする必要があるために，システム全体の遅延で ms オーダでの高速性を満たすのが困難であるという問題がある．

2.2 ウェアラブルビジョンを用いた人間の身体動作のセンシング

環境側にセンサを設置すると，推定精度を上げるためには環境全体をセンシングする必要があり，高速性を保つことが困難である．この問題を解決する 1 つの方法として，人間の身体のセンシングしたい部位にビジョンを装着する手法がある．このようなウェアラブルビジョンを用いることで人間の歩行動作の推定や人間の身体動作をセンシングした研究がある [9], [10], [11]．しかし，これらの手法には問題がある．例えば，歩行動作のセンシングのみでは，映像とインタラクションするような今回のシステムにおいては十分な情報が得られないことや，最適化問題の処理コストが大きく動作速度が遅いという問題がある．また，これらのような自然特徴点を用いた手法では，投影された映像により自然特徴点の検出が困難になったり，投影対象が壁のような単色な物体であると自然特徴点が検出できないことが考えられる．

2.3 不可視マーカを用いたセンシング

自然特徴点を利用しない手法としてマーカを用いた手法がある．この手法では，予めセンシング対象に既知のパターンを設置し，そのパターンをセンシングする．しかし，設置したマーカが観測者に知覚されるという問題がある．そこで，様々なマーカ不可視化手法が提案されてきた．対象物体に不可視マーカを設置しそれをビジョンによりセンシングする研究がある [12]. しかし，このような設置型マーカは，センシングする対象物体にマーカを設置しなくては

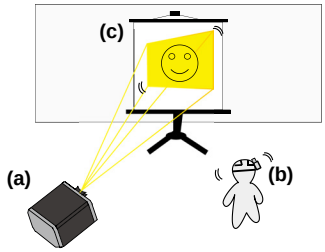

図 2　システムの全体図．(a) は高速プロジェクタ，(b) はウェアラブルビジョンを頭部に設置した観測者，(c) はスクリーンを示す．

Fig. 2　The whole system. (a) a high-speed projector and (b) a wearable vision and (c) a screen

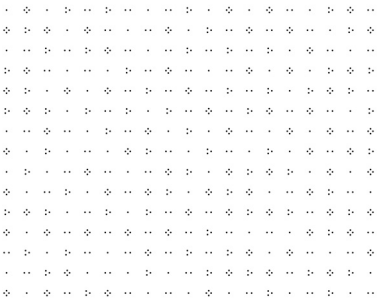

図 3　今回使用した 19×15 の DDCM

Fig. 3　DDCM(19×15) currently used.

ならず，日常的な空間に対しては壁や床などにマーカを設置する必要があり，対象物体を加工しなくてはならないという問題がある．

このような設置型のマーカに対して，設置する必要のない投影型の不可視マーカを用いた手法がある [13], [14]．しかし，一方は，埋め込まれたマーカのデコード時に遅延が発生する問題があり，また，もう一方はマーカの変形に弱く，日常的な空間に適応するのは難しいという問題がある．

3. 低遅延な空間型拡張現実

3.1 システムの概要

本節では，高速プロジェクタと高速な画像センシングを組み込むことで，スループット 1000 fps，遅延 ms オーダの SAR システムを構築できる可能性を示す．今回は，人間の身体動作の中でも視点変化に着目し，SAR における視点把握のためのウェアラブルな高速ビジョンと高速プロジェクタを組み合わせる構成を提案する．

提案システムは，図 2 のように，高速プロジェクタとウェアラブルビジョンからなる．先に記述した遅延 ms オーダを達成する可能性のある高速プロジェクタとして，DynaFlash [15] と呼ばれる，最小 3 ms のレイテンシで 8 bit 階調の映像を 1000 fps で投影可能な高速プロジェクタを使用する．高速プロジェクタは環境側に固定し，観測者に提示する映像と，頭部センシングのためのマーカを投影する．このとき，マーカが観測者に知覚されないように，人間の臨界融合周波数以上で映像を高速投影することで，投影映像に観測者に不可視なマーカを埋め込む．投影するマーカとそのマーカの不可視化手法については 3.2 節で詳細を述べる．

また，ウェアラブルビジョンは観測者の頭部に，観測者の顔と同じ方向を向くように設置し，投影されたマーカを撮像する．このとき，マーカをビジョンにより正確に撮像するためには，プロジェクタとビジョンを同期させる必要がある．しかし，非同期の場合でもマーカを撮像できる手法が近年提案されている [14], [16]．これらの手法を利用すれば非同期でも撮像できることが期待できるが，今回は有線により同期信号を送る方法を採用する．

ウェアラブルビジョンの画像上の点と世界座標系の点の対応関係がわかれば，世界座標系に対するウェアラブルビジョンの位置姿勢がわかる．ここで，投影面の形状が既知であると仮定すると，プロジェクタは環境側に固定されているので，投影面に投影されたマーカの世界座標系における座標がわかる．したがって，プロジェクタで投影するマーカ画像とウェアラブルビジョンにより撮像された，ビジョン画像上の対応点の関係がわかれば，ウェアラブルビジョンの位置姿勢を取得できる．今回は，投影面の形状を平面と仮定することで，プロジェクタとビジョンの座標関係をより簡略な形で記述した．詳細な手法については 3.3 節で述べる．

3.2 不可視マーカを用いたウェアラブルビジョンによる頭部のセンシング

本節では，自然特徴点が存在しないような場所でも高速高精度に頭部の位置姿勢を取得する手法について述べる．システム全体の遅延が ms オーダとなるために，頭部のセンシングの処理が 1 ms 以下を達成でき，また，周辺環境に左右されないために，自然特徴点を利用しない手法として投影型のマーカを用いた手法を採用する．しかし，投影型のマーカは投影環境によりマーカが変形してしまう可能がある．したがって，変形に頑健なマーカである必要がある．他のマーカの要件として，プロジェクタによる投影範囲に対して，ビジョンの画角が狭くマーカ全体を撮像できない場合も想定される．このような状況も考慮し，マーカの一部のみからビジョンがマーカのどの部分を撮像したのかがわかる必要があると考えられる．これらの要件を満たすマーカとして，Deformable Dot Cluster Marker (DDCM) [8] が報告される．したがって，本稿では，提案システムのセンシングの要件を満たす可能性がある DDCM を用いる．

DDCM は，図 3 のように，ドットクラスタと呼ばれる変形の影響を受けづらい円形のドットの集合から成る．ま

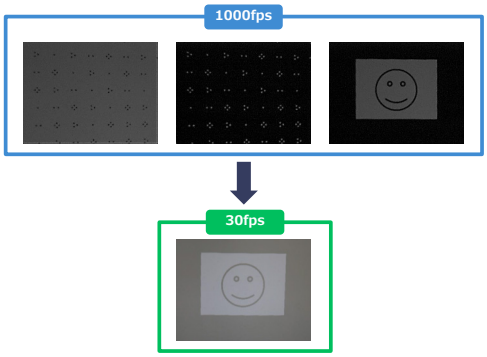

図 4　1000 fps のビジョンで撮像した映像と 30 fps のビジョンで撮像した映像．30 fps のときは，投影している 3 枚の画像が積分されたような映像が撮像される．

Fig. 4　Projection image taken with 30 fps vision and 1000fps vision.

た，処理として，detection と frame-by-frame tracking の 2 つがある．detection 処理では，それぞれのドットクラスタの ID を同定する．frame-by-frame tracking では，前フレームで検出されたドットクラスタの情報をもとに処理することで，計算量を減らし高速なトラッキングを可能としている．しかし，このまま使用すると観測者にマーカが知覚されるので，人間の知覚特性を利用して不可視化を行う．

人間は点滅する光を見るとちらつきを感じるが，点滅周波数がある値を超えると，ちらつきを感じなくなる．この周波数は臨界融合周波数と呼ばれる．しかし，臨界融合周波数は点滅している光の空間周波数や視線移動などにより大きく変動する．そこで，このような視線移動も考慮して，それぞれの空間周波数に対してどのような時間周波数なら光の点滅が知覚されなくなるかを調査した研究がある [13]．この研究によれば，どのような空間周波数においても 700 fps 以上で点滅させれば 75%以上の確率で人間にはちらつきが知覚されないことが報告されている．そこで，本稿では，高速プロジェクタを用いて 1000 fps で，マーカパターン，輝度反転したマーカパターン，観測者に提示する画像の 3 枚の画像を高速で切り替えることで，マーカの不可視化を試みる．このとき，実際に投影した映像をプロジェクタと同期した 1000 fps で撮像可能なビジョンで 3 フレーム撮像したときの画像が図 4 の上の 3 枚である．また，人間の視覚に近い，30 fps のビジョンで撮像したときの画像が図 4 の下の 1 枚である．

3.3　観測者に対して静止して見える映像の提示

本節では，アプリケーションの 1 つの観測者に対して静止して見える映像を提示する方法について述べる．同次座標系において，世界座標系 $\mathbf{X} = (x, y, z, 1)^{\mathrm{T}}$ とビジョン上の画像座標 $\mathbf{U} = (u, v, 1)^{\mathrm{T}}$ の対応関係は，透視投影行列 $\mathbf{P}$ を用いて以下のようにかけることが知られている．

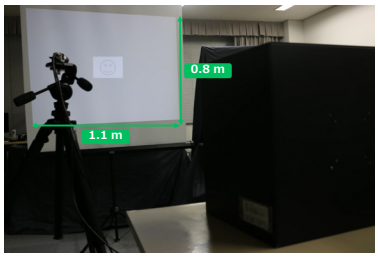

図 5　実験時のシステムの様子

Fig. 5　The whole system at the time of experiment

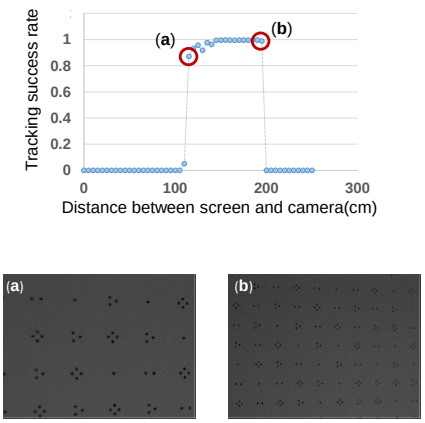

図 6　上図はビジョンのスクリーンからの距離とトラッキング成功率を示す．下図 は上図の (a)，(b) の位置での実際のビジョンから撮像されるマーカの様子．

Fig. 6　The upper figure is that system tracking accuracy at varying distances between vision and screen. (a) and (b) are images of the marker taken by the visions at each point.

$$\mathbf{U} \sim \mathbf{PX}$$

ただし，$\sim$ は定数倍の不定性を除いて等しいことを表す．また，この透視投影行列を推定することで，ビジョンの位置姿勢を求めることができる．ビジョンの内部パラメタが既知の場合，透視投影行列を求めるには，画像上の 2 次元点と世界座標系の 3 次元点の対応関係を取得できれば良い．2 次元点の座標は DDCM から推定され，3 次元点の座標は投影形状が既知であれば予め求めることが可能となる．

ここで，観測者に静止した映像を提示するのみならば，投影環境が平面であると仮定することでより単純なホモグラフィ行列を用いることが可能となる．ホモグラフィ行列は未知数が 8 つなので，DDCM から画像座標と投影平面上の座標の対応する点が少なくとも 4 点求まればよい．

また，簡単のために，本稿では観測者が単眼であることを仮定すると，頭部に装着したビジョンに対して，投影映像を静止させれば，観測者に対しても映像が静止して見えることになる．したがって，予めビジョン画像上のどこに映像を静止させるか指定することで，求めたホモグラフィ

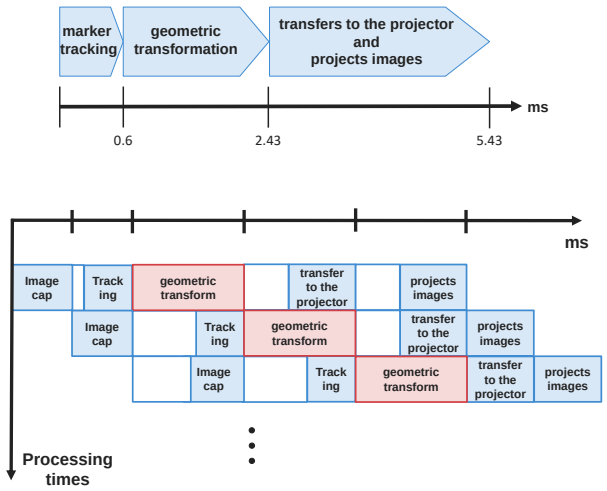

図 7　レイテンシとスループット
Fig. 7　Latency and throughput

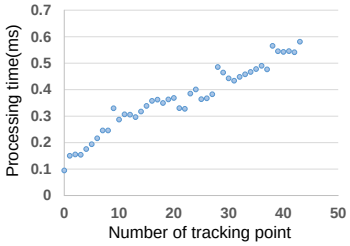

図 8　トラッキング成功点とトラッキング処理時間
Fig. 8　Tracking processing times at varying tracking success points.

行列からどこに投影すれば良いかがわかる．

4. システムの動作可能範囲と高速性の評価

本節では，システムの性能を示す．システムの性能の評価として，図 5 のように環境をセットアップし，スクリーンからの距離に対するトラッキング処理の動作可能範囲を調査した．また，システムのレイテンシとスループットの概算を行った．今回，スクリーン上に，およそ 1.1×0.8 m の映像が投影されるようにプロジェクタを設置し，プロジェクタは 1024×768 pixel の画像を 1000 fps でマーカ，輝度反転したマーカ，観測者に提示する画像の 3 枚を投影した．また，ビジョンは basler の acA-750uc を使用し，320×240 pixel の画像を 1000 fps で撮像した．ただし，ビジョンには焦点距離が 6 mm のレンズを使用した．また，マーカのドットクラスタの個数としては 19×15 個描画されたマーカを使用した (図 3)．本稿では，GPU は使用せずに CPU のみを使用した．CPU として，Intel Xeon CPU E5-2687W v4 @ 3.00GHz を 2 つ搭載している計算機を使用した．

図 6 はスクリーンからの最短距離に対して，ビジョンをスクリーンから 5 cm 刻みで離していき，それぞれの距離で 10000 フレーム処理を行ったときのホモグラフィ行列の推定成功率を示す．ただし，今回は，投影されたマーカの座標と画像座標上のドットクラスタの座標の対応関係が 4 点以上検出できたとき，ホモグラフィ行列の推定が成功したと仮定して，成功率を算出した．スクリーンからの距離が 120 cm から 180 cm のとき安定してトラッキングできていることがわかる．ただし，本システムで採用した DDCM は，トラッキング可能な距離がマーカの密度，ドットのサイズに影響されるので注意が必要である．例えば今回の場合，ビジョンから見て図 6 の下図のようにマーカが撮像されたときが，トラッキングの限界となる．図 6 の (a) は，それぞれのドット間の距離が撮像された画像上で離れすぎてしまい，ドットクラスタを正しく生成できないことが原因だと考えられる．また，図 6(b) は，撮像された画像上でドットのサイズが小さく，ドットとして検出されないことが原因であると考えられる．

また，処理の流れは，図 7 の上図のようになっている．今回は，それぞれのプロセスの処理時間を別々に計測し足し合わせることで，画像撮像後のシステム全体の処理のレイテンシを概算する．1 つ目の処理として，トラッキングしているマーカ数に対する処理時間を図 8 に示す．このグラフから処理時間は，トラッキングが成功したドットクラスタの数に支配的であることがわかる．また，今回セットアップした環境でトラッキング可能な範囲での最大の処理時間として，およそ 0.6 ms かかっていることがわかる．2 つ目の処理として，ホモグラフィ行列の推定も含めた，観測者に提示する映像を生成するときの処理時間を計測する．1000 回計測し，その平均値を処理にかかった時間とした．結果，この処理は 1.83 ms であった．今回は CPU のみで実装を行っており，他の研究 [8] では，GPU を利用して 1 ms 以下の映像生成を達成しているので，本稿においても GPU を使用することで，1 ms 以下の映像生成を達成できることが期待できる．最後にプロジェクターの投影遅延として，3.0 ms と報告されているので 3.0 ms で計算を行う [15]．以上より，およそ 5.43 ms のレイテンシを達成できることが概算できる．スループットは，図 7 の下図のように処理が独立であるのでパイプライン化でき，また本システムは，プロジェクタ内にバッファを設けているので，投影画像の転送と投影の処理を分離して考えられる．したがって，ボトルネックは幾何変換の 1.83 ms であり，スループットはおよそ 530 fps である．

5. アプリケーション

本節では，アプリケーションについて述べる．3.3 節で述べた処理を用いて観測者に対して静止して見える映像を提示した．頭部にビジョンを装着した観測者が，高速プロジェクタにより投影されたマーカを見たときに投影される映像を図 3.3 に示す．(a) は左右，(b) は上下，(c) は前後，

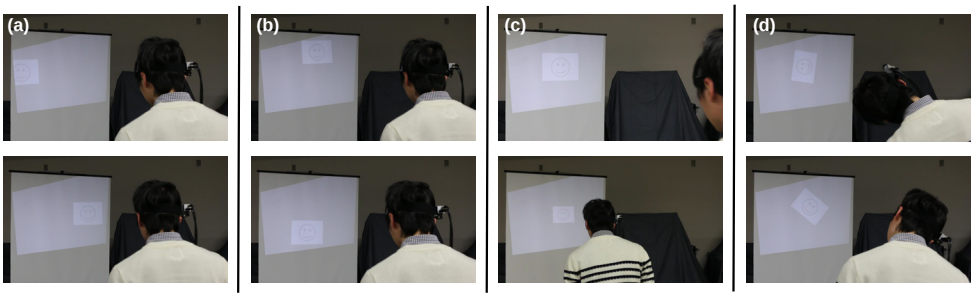

図 9 人の動きに対する幾何変換の様子．(a) は左右に，(b) は上下に，(c) は前後に，(d) は左右に頭部を回転したときの様子．

Fig. 9 A state of geometric transformation when the observer moves. (a) Moving left and right. (b) Moving up and down. (c) Moving back and forward. (d) Rotate to left and right.

(d) は左右に回転したときのの投影映像の様子である．

今回は，性能評価時と同様に 1.1×0.8 m の映像が投影されるようにプロジェクタとスクリーンをセットしたが，プロジェクタからスクリーンをさらに離し画角をより広くすることや，複数台のプロジェクタを使用することで，常に視界内の同じ場所に同じ形で情報を提示することができる．例えば，いつでもどこでも，持続的にメールやスケジュールなどの情報にアクセスしたり，自動車における Head-Up Display (HUD) のように移動方向の指示をするなどの応用が考えられる．また，観測者が運動中や，車内などの振動する環境にいる場合でも，観測者の視界に対し静止して見えるように情報を提示することで，振動をキャンセルすることが可能となり，映像が見えやすくなるほかに映像酔いなどを回避できることが期待できる．

6. おわりに

本稿では，処理のレイテンシがおよそ 5.43 ms の低遅延な SAR の実現可能性を示した．本システムの性能の限界として，高速プロジェクタによる画角の狭さが挙げられる．これは，複数台の高速プロジェクタを利用することで解決可能であると期待できる．また，DDCM のドットサイズを用いたアルゴリズムに起因するトラッキング可能範囲の限界も挙げられる．この限界は，観測者のスクリーンからの距離に応じて，マーカサイズを動的に変えることで，処理的な付加なく解決可能だと考えられる．

参考文献

[1] Bimber, Oliver, and Ramesh Raskar. Spatial augmented reality: merging real and virtual worlds. CRC press, 2005.
[2] Cruz-Neira, Carolina, et al. “The CAVE: audio visual experience automatic virtual environment.” Communications of the ACM 35.6 1992: 64-73.
[3] Jones, Brett, et al. “RoomAlive: magical experiences enabled by scalable, adaptive projector-camera units.” Proceedings of the 27th annual ACM symposium on User interface software and technology. ACM, 2014.
[4] Jones, Brett, et al. “IllumiRoom: peripheral projected illusions for interactive experiences.” Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2013.
[5] Ng, Albert, et al. “Designing for low-latency direct-touch input.” Proceedings of the 25th annual ACM symposium on User interface software and technology. ACM, 2012.
[6] Zerroug, Alexis, et al. “Invoked computing: Spatial audio and video AR invoked through miming” Proceedings of Virtual Reality International Conference. 2011.
[7] Lincoln, Peter, et al. “From Motion to Photons in 80 Microseconds: Towards Minimal Latency for Virtual and Augmented Reality.” IEEE transactions on visualization and computer graphics 22.4 2016: 1367-1376.
[8] Narita, Gaku, et al. “Dynamic Projection Mapping onto Deforming Non-rigid Surface using Deformable Dot Cluster Marker.” IEEE Transactions on Visualization and Computer Graphics 2016.
[9] 渡辺義浩，ほか “単一のウェアラブルカメラを用いた人間の歩行動作推定 (〈特集〉 実世界イメージング).” 日本バーチャルリアリティ学会論文誌 17.3 2012: 219-229.
[10] Shiratori, Takaaki, et al. “Motion capture from body-mounted cameras.” ACM Transactions on Graphics (TOG). Vol. 30. No. 4. ACM, 2011.
[11] Rhodin, Helge, et al. “EgoCap: egocentric marker-less motion capture with two fisheye cameras.” ACM Transactions on Graphics (TOG). Vol. 35. No. 6. ACM, 2016.
[12] Asayama, Hirotaka, et al. “Diminishable visual markers on fabricated projection object for dynamic spatial augmented reality.” SIGGRAPH Asia 2015 Emerging Technologies. ACM, 2015.
[13] 後藤正太郎，ほか “可視光通信プロジェクタ映像鑑賞時の時空間周波数特性の計測 (知覚・触覚, 人工現実感).” 電子情報通信学会技術研究報告. MVE, マルチメディア・仮想環境基礎 113.109 2013: 71-76.
[14] Niidome, Hidetaka, et al. “Camera synchronization to imperceptible frames embedded in a displayed video sequence.” SIGGRAPH Asia 2014 Posters. ACM, 2014.
[15] Watanabe, Yoshihiro, et al. “High-speed 8-bit image projector at 1,000 fps with 3 ms delay.” Proceedings of the International Display Workshops. 2015.
[16] Goto, Akifumi, et al. “Display tracking using blended images with unknown mixing ratio as a template.” SIGGRAPH ASIA 2016 Technical Briefs. ACM, 2016.