

## Share-Z: 観察者視点の奥行き獲得のための

### クライアント・サーバ・センシング

大田友一, 菅谷保之, 五十嵐浩樹, 大槻俊和

筑波大学 機能工学系 (<http://www.image.esys.tsukuba.ac.jp>)

複合現実感を実現するために現実世界に仮想物体を違和感なく融合するためには, 現実世界の物体によって仮想物体が隠される現象や, 仮想物体によって現実世界に投影される影を表現することが重要である. そのためには, 観察者の視点から見た現実世界の奥行き情報を獲得することが必要となる. シースルー型のHMD(または, HWD)を装着した観察者は現実世界を動きながら見るのが前提であり, 観察者視点の奥行き情報を得るためには, HMDに奥行き獲得センサを装着する必要がある. しかし, HMDに装着できる程度の大きさや重さで高精度の奥行き情報を獲得できるセンサは, 能動的, 受動的のいずれ方式でも実現されていない.

この問題を解決するために, 本研究では, 観察者の近くに設置した高性能の奥行き情報センサをサーバとし, HMDに装着したセンサをクライアントとする, クライアント・サーバ型センシング方式(Share-Z)を提案する.

## Share-Z: Client-Server Depth Sensing for See-Through HMD

Yuichi OHTA, Yasuyuki SUGAYA, Hiroki IGARASHI, Toshikazu OHTSUKI

College of Engineering Systems, University of Tsukuba

(<http://www.image.esys.tsukuba.ac.jp>)

In the Mixed Reality, occlusions and shadows are important to realize a natural fusion between the real and virtual worlds. In order to do this, it is necessary to acquire the dense depth information of the real world from the observer's viewing position. In the MR, the depth sensor should be attached to the See-Through HMD of the observer because he/she is assumed to move around. It should be small and light enough to be attached to the HMD and can produce reliable dense depth map. Unfortunately, however, no such depth (or range) sensors are available. In this paper, we propose a client-server sensing scheme to solve this problem. In this scheme, called Share-Z, a server sensor (or sensors) located near the observer supplies the depth information to the client sensor mounted on the HMD.

### 1. はじめに

複合現実感において, 現実世界と仮想世界の違和感のない融合を実現するためには, 「幾何的整合性」, 「光学的整合性」, 「時間的整合性」を解決する必要がある. 幾何学的整合性には, 実世界と仮想物体の

位置や姿勢を正しく整合させる問題、及び、実物体と仮想物体の前後関係による隠れ（オクルージョン）や実世界に投影される仮想物体の影を表現する問題が含まれる。後者を正しく表現しないと図1（左）に示したように、位置や姿勢を正しく表現しても仮想物体の3次元位置の把握が著しく困難な融合結果になってしまう。実物体によって隠される仮想物体を表現するためには、図1（右）のように、隠される部分を消去して描画する必要があり、隠されるか否かを画素ごとに判断するために、観察者、すなわち、観察者が装着しているHMDの視点からの現実世界の密な奥行きマップが必要となる。

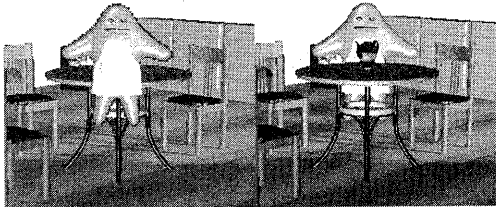


図1. 実物体と仮想物体の隠れの表現（左：隠れ処理なし、右：隠れ処理あり）

多視点ステレオ法を用いて観察者視点の奥行き情報をビデオレートで獲得し、実物体像と仮想物体像の合成を、隠れを考慮して実現した例として Kanade らの Z-key[1]がある。しかし、多視点ステレオ画像を撮影するための多眼カメラヘッドは、大きくて重い。超小型カメラを用いれば軽いカメラヘッドを製作することも可能かもしれないが、カメラヘッドの大きさを押さえる

ためにカメラ間隔を小さくしすぎると、得られる奥行き情報の精度が低下するというジレンマに陥る。複合現実感において、シースルー型HMDを装着した観察者は現実世界を移動することが前提であり、観察者視点の奥行きマップを得るためには、HMDに多眼ステレオヘッドを装着する必要がある。しかし、これは、大きさや重さの点で非現実的である。多視点ステレオ法だけでなく、レーザレンジファインディング法を見渡しても、HMDに装着可能な程度に小型軽量で、ビデオレートに近い速度で、密な奥行きマップを獲得可能なセンサを実現することは、現時点では不可能に近いと考えられる。

このようなジレンマを回避し、動き回る観察者が装着したHMDの視点からの現実世界の奥行き情報を獲得できる方法として、我々は、クライアント・サーバ方式による奥行きセンシング法を提案した[2]。これは、観察者の近傍にサーバセンサとして（大型の）奥行きセンサ（群）を固定して設置し、それによって得られる現実世界の奥行き情報を、観察者に装着されたセンサからえられる位置・姿勢情報に基づいて、観察者の視点からの奥行き情報として視点変換し供給するものである。なお、Share-Zのshareは、本方式によれば、1セットのサーバセンサの情報を、複数の観察者が共有できることに由来する。

## 2. Share-Zの基本概念

Share-Zシステムを構成するサーバセンサとクライアントセンサは、それぞれ、以下の役割を担う。

- サーバセンサ： 観察者が行動する空間に固定して設置され、その空間の3次元情報を、獲得・更新・管理し、クライアントセンサからの要求に対応して、クライアント側に3次元情報を供給する。複数の観察者が行動する空間を共有する場合には、1つのサーバセンサが複数のクライアントセンサに3次元情報を供給する。すなわち、サーバセンサが獲得した3次元情報を、複数の観察者が共有（share）する。
- クライアントセンサ： 観察者のHMDに装着され、観察者視点の奥行き情報を生成するため

に必要な観察者の位置と姿勢を獲得し、サーバセンサから供給される3次元情報を変換して観察者視点の密な奥行きマップを生成する。

空間に固定したサーバセンサと、HMDに装着したクライアントセンサに分離した構成をとることによって、以下の利点が得られる。

- 1) 既述のように、HMDに装着するセンサを小型軽量なものにできる。
- 2) サーバセンサを空間に固定することによって、サーバセンサから観測する現実空間は、大部分が静止世界と仮定できる。これによって、奥行き情報を安定に獲得できるとともに、比較的低速のセンサでも利用可能になる。これに対して、HMD視点から見た現実空間は、観察者の動きによって全画面が動くものとなり、安定な奥行き獲得が困難になるだけでなく、動きに追従可能な高速性が要求される。
- 3) サーバセンサとして、異なった特性を持つ複数のセンサを組み合わせる用いることが可能になる。例えば、現実空間中で動く物体の3次元情報を獲得するためには多視点ビデオレートステレオシステムの出力を、現実空間中で静止した部分についてはレーザレンジファインダの出力を用いて、空間の3次元情報を更新・管理する。これによって、ステレオ法では得にくい、テクスチャが少ない部分での3次元情報を、レーザレンジファインダからの情報で補完するといった効果も得られる。静止世界については、レンジファインダで計測するのではなく、あらかじめ、人手で3次元モデルを与えておくこともできる。サーバセンサが1視点の奥行きマップしか提供できない場合には、視点変換した場合に、奥行き情報が欠落した穴ができるが、複数のセンサを組み合わせることによって、これを軽減することも可能である。
- 4) 既述のように、1セットのサーバセンサを複数の観察者が共有できる。ビデオレートステレオシステムやレーザレンジファインダは高価なものであり、これらを、個々の観察者ごとに用意するよりは、複数の観察者で共有する方がコスト的に有利である。

### 3. Share-Z のバリエーション

Share-Z システムは、サーバやクライアントの実現法によって、技術的難易度や性能面で様々のレベルを考えることができる。

#### 3.1 サーバセンサの実現法

現実世界の中に動く物体が存在しない場合には、あらかじめ人手で3次元モデルを構築し与えておくことができる。これは、CGと実写の合成法として一般に用いられているものと同じである。また、3次元モデルを人手で構築するのではなく、レーザレンジファインダなどで計測したデータをもとにして作成することも考えられる。この場合にも、レンジデータを3次元モデルに加工する段階での、人手の介入は可能である。

動く物体が存在する場合には、ビデオレートステレオシステムのように、十分高速に実世界の3次元情報を獲得できるセンサが必要である。もっとも基本的な構成としては、1台のビデオレートステレオシステムで得られる奥行き情報のみを用いる方法が考えられる。ただし、この場合には、視点変換した奥行きマップに穴が空くことは避けられない。2台以上のステレオシステムを用いれば、この問題は軽

減できるが、テクスチャの少ない部分で奥行き情報の精度が低下する現象を避けることは困難である。

人手で構築した3次元モデルとビデオレートステレオシステムを組み合わせることによって、奥行きマップの穴の問題、及び、テクスチャが少ないことによる精度低下の問題は、軽減できる。複合現実感を実現しようとする一般の空間は、大部分が静止しており、その中で人間などの一部の物体が動き回るだけと仮定できる場合が多いと思われるので、静止している部分には、あらかじめ用意した高精度の3次元モデルを用いることは有効と考えられる。動いている部分と静止している部分で3次元情報を使い分ける処理は必要である。

現実世界には、静止物体と移動物体という単純な切り分けには、必ずしも馴染まない場合も多い。部屋の壁のように完全に固定されて決して動かない部分だけでなく、机の上の書類やコップ、椅子などのように、基本的には静止しているが、時には位置が移動する物体などが存在する。このような物体を準静止物体と呼ぶ。準静止物体については、物体の移動に応じて、その3次元モデルを移動させればよい。また、それが困難な場合には、変化のあった部分についてのみビデオレートステレオのデータで上書きするような対応も考えられよう。レーザレンジファインダを用いて定期的に静止部分と準静止部分の距離マップが得られる場合には、このデータを用いて準静止部分の検出と更新を行うのが精度面で有利である。

### 3.2 クライアントセンサの実現法

クライアントセンサが現実世界から得るべき情報は、観察者視点の位置と姿勢である。従って、クライアントセンサが画像センサを含むことは必須ではない。この位置・姿勢を用いて、クライアントセンサは、サーバセンサから供給される3次元情報を変換し、観察者視点の奥行きマップを生成する。

拡張現実感を実現するために開発されている観察者の位置・姿勢を求める手法は、基本的に、クライアントセンサの実現法として利用できる。例えば、赤外線を用いた位置・姿勢センサ、磁気を用いた位置・姿勢センサなどである。

しかし、観察者が見る現実世界像と正確に整合させた奥行きマップを得るためには、現実世界像そのものを用いて位置・姿勢を求める方が有利である。これには、磁気センサなどと画像中の特徴点追跡を併用し、磁気センサで求めた位置・姿勢を画像情報で修正する方法と、画像中の特徴点の配置のみから、観察者の位置・姿勢を求める方法がある。

HMDに装着された両眼2台のカメラの映像だけを用いて、観察者の位置・姿勢を求め、サーバセンサ供給される3次元情報を変換して奥行きマップを生成する方式が、研究課題としては興味深いが、動作の安定性という面からは、赤外線センサなどの利用が有利であろう。

HMDに装着された2台のカメラでステレオ処理を行い、サーバセンサから得られる情報が欠落している部分を補完すること、さらには、(複数の)観察者のHMDから得られるステレオ結果をサーバセンサに送りサーバセンサ側で統合管理して各クライアントに供給すること、なども、ストーリーとしては面白いが、現実的かどうかは検討を要すると思われる。

### 3.3 3次元情報の管理と視点変換処理

サーバセンサ側における3次元情報の管理は、以下の、一般には相反する要求を満たす必要がある。

- 3次元情報の表現形式は、サーバセンサ群から獲得される生のデータ形式に近い方が、更新が容易。
- 3次元情報の表現形式は、クライアント側計算機のグラフィックスエンジンが処理しやすいデ

ータ形式の方が、高速な視点変換処理が可能。

手動で与える静止世界の3次元モデルについては、自動更新の必要がないため、比較的少数のポリゴンで表現したグラフィックスエンジンでの処理に適した表現形式をとることができる。一方、ビデオレートステレオシステムで得られる奥行きマップは、秒間30フレームでデータが流れるため、ポリゴン化などの処理は殆ど行わずに、生に近い状態で視点変換処理をせざるを得ない。現実世界の動物体の3次元情報を獲得してから、観察者に供給するまでのレイテンシーも最少にする必要がある。

レーザレンジファインダによって獲得されるデータの表現形式は、上記2つの中間的な性質を持つ必要がある。長期的に静止している部分については、出来るだけ効率のよいポリゴン化を行うべきであるが、現実世界で準静止物体が移動した場合の変化などは、出来るだけ迅速に観察者視点に反映する必要がある、生に近い形式で用いることが考えられる。

複数のセンサから得られるデータをどのように統合するかも重要な課題であろう。

#### 4. 現在までの実現例

提案した方式の有効性を検証するために試作したシステムの概念図を図2に示す。サーバセンサとしては、多視点ステレオ法によるステレオシステムが1台あることを想定して、それによって得られる離散的な奥行き値を持つ奥行きマップを入力として与えている。クライアントセンサが位置と姿勢を獲得する方式としては、HMDに装着されたカメラで撮影された映像上で特徴点を追跡することによって、画像情報だけで位置・姿勢を推定する方法を採用した。

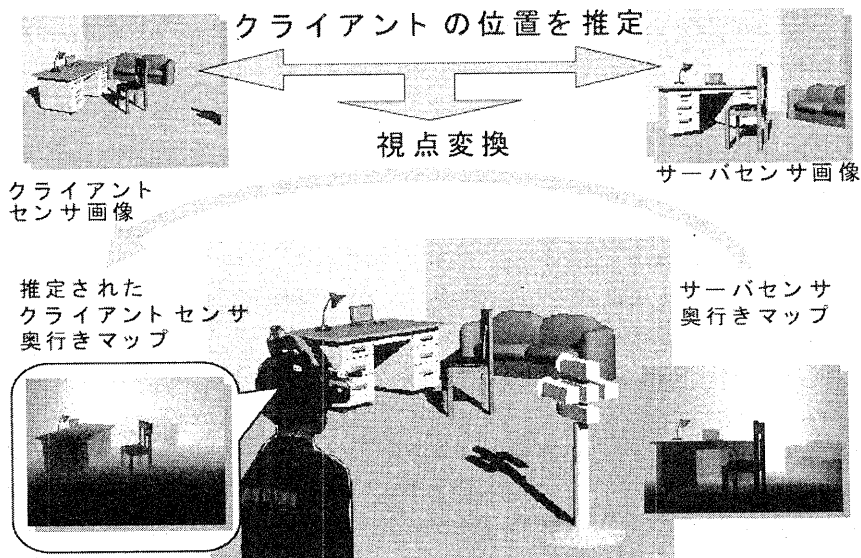


図2 システム概念図

実装は、SGI の Onyx2 InfiniteReality 上で OpenGL を用いて行っている。特徴点追跡には富士通のトラッキングビジョンを用いた。カメラの映像は Onyx2 およびトラッキングビジョンの両方に供給される。トラッキングビジョンを装備した PC と Onyx2 は 100Base-TX で接続されており、カメラ映像から探し出した特徴点の画像座標が Onyx2 に送信され、Onyx2 で位置・姿勢を推定、奥行きマップの視点変換処理をした上で、隠れを考慮して仮想物体像を合成し、その映像を観察者に提示する。

ステレオシステムで得られる奥行きマップを観察者視点に変換する場合の変換方法として、以下の 3 つの方法を試みた。

- 1) 同じ視差をもつ画素全体を一枚のテクスチャとして表現し視点変換する。視差の数だけテクスチャプレーンを生成する。
- 2) 方法 1) では生じてしまう奥行き情報が空白となる領域を、ポリゴンパッチで埋める。
- 3) 方法 2) で一枚のテクスチャとして表現していた領域を、画素単位の 3 次元ポリゴンパッチで構成する。画素数分のポリゴンを生成する。

視点変換の処理速度を表 1 に示す。これにはポリゴンを生成するのに必要な時間も含まれている。

観察者の位置・姿勢の推定に用いる特徴点は図 3 で示された 7 点である。最初の画面におけるこれらの特徴点の位置は、ワールド座標との対応も含めて手入力を与える。観察者の視点の変化にともなう

表 1 視点変換処理の速度 (msec/frame)

視差画像のサイズ	256 × 256	512 × 512
方法 (1)	81	230
方法 (2)	140	470
方法 (3)	89	360

特徴点の追跡はトラッキングビジョンにより実時間で行う。図 3 中の実物のテレビラック上に仮想のデスクランプ、後方に仮想の椅子を配置した。奥行き情報を用いない合成例は図 4 に示す。椅子の足がテレビラックの上に描画されるためテレビラックの前に浮いているかのような印象をうける。観察者視点の奥行きマップを生成した結果が図 5 である。黒い部分は、奥行き情報が欠落している部分である。これを HMD カメラからの映像の depth buffer に入力し、隠れ処理をして仮想物体を合成した結果が図 6 である。別の視点からの結果を図 7 に示す。前述の方法 3) を用いた場合毎秒 6 フレームの更新が可能であった。

## 今後の展開

サーバセンサとして、Kanade らのビデオレートステレオマシンをベースに SONY が開発した 5 眼ビデオレートステレオシステム (図 8) を用いた実験システムの構築を現在進めている。また、静止世界に対して 3 次元ポリゴンモデルを手動で与える方法と、レーザレンジファインダ (図 9) を併用する方法について並行して研究を進めている。レンジファインダで撮影した intensity map と depth map を図 10、図 11 に示す。

Share-Z の方式は、HMD に装着可能な程度に小型軽量で、かつ、高精度の奥行きマップが獲得可能なセンサを求めて考案したものであるが、現実世界とサーバセンサとの相対位置が変化しないという利点から、比較的低速で高精度のセンサを利用する道が開かれた。この利点を利用すれば、スタジオカメラのように、必ずしも、小型軽量を要求されない状況で奥行きマップを獲得する方式へも発展可能ではないかと思われる。

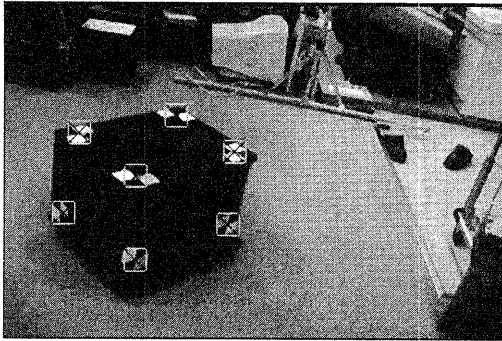


図 3. 観察者視点映像と特徴点

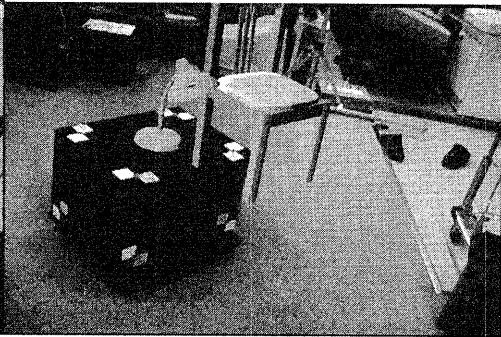


図 4. 奥行き情報を用いない合成例

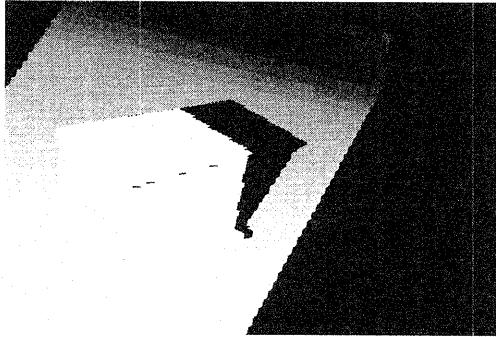


図 5. 観察者視点に変換した奥行き情報



図 6. 奥行き情報を用いた合成例 1

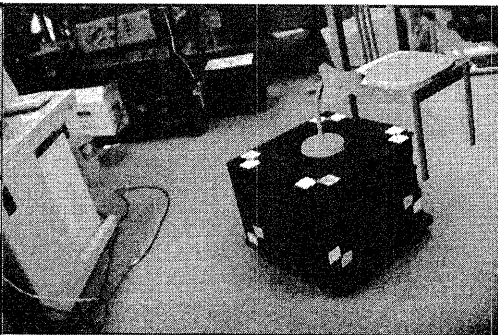


図 7. 奥行き情報を用いた合成例 2

謝辞

ビデオレートステレオマシン，レーザレンジファインダは，共同研究先のMRシステム研究所のものを利用してゐる．関係者各位に謝意を表す．

参考文献

[1] Kanade, et.al. : A stereo machine for video-rate dense depth mapping and its new applications, Proc. IEEE CVPR'96, June 1996.

[2] 五十嵐, 大槻, 大田: シースルーHMDのためのクライアント・サーバ方式による観察者視点の奥行き情報推定, 信学技報 PRMU98-247, Mar. 1999.

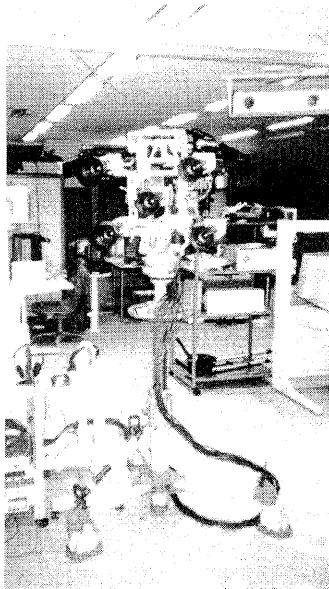


図 8.五眼ビデオレートステレオシステム

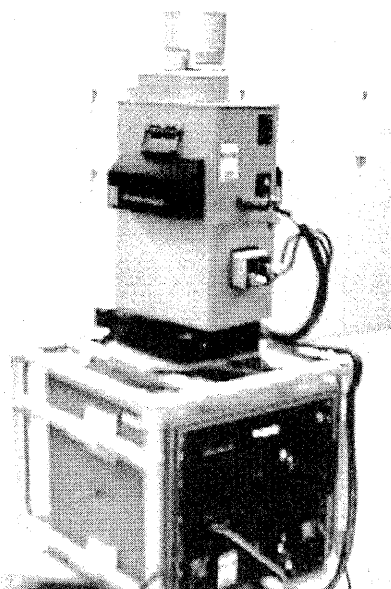


図 9.レーザーレンジファインダー

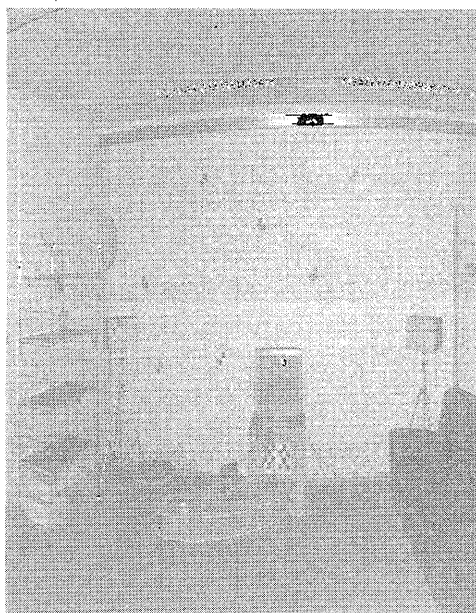


図 10. レーザーレンジファインダーによる  
intensity map

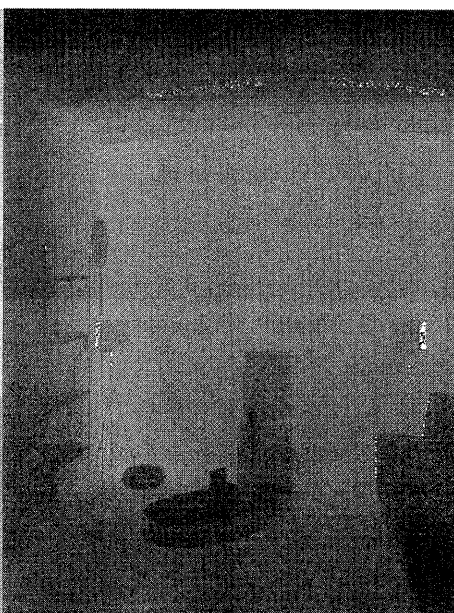


図 11. レーザーレンジファインダーによる  
depth map