

全方位画像センサを用いた実時間テレプレゼンスと 画像表示遅延に関する実験的評価

尾上 良雄 山澤 一誠 竹村 治雄 横矢 直和

奈良先端科学技術大学院大学 情報科学研究科

〒630-01 奈良県生駒市高山町 8916-5

Tel: 0743-72-5296 Fax: 0743-72-5299

E-mail: {yoshio-o, yamazawa, yokoya, takemura}@is.aist-nara.ac.jp

あらまし: 移動ロボットの遠隔操作や多人数で行う遠隔通信会議などにおいて、遠隔地の情景を画像として提示するテレプレゼンス技術の必要性が高まっている。遠隔地の情景を、観察者の視線方向に追従して提示するには、回転台に取り付けたカメラを利用するのが一般的であるが、この方法では視線方向と提示される画像の方向に時間遅延が生じる。我々は全方位視覚センサで得られた全方位画像から利用者の視線方向の透視投影画像を生成・提示する実時間テレプレゼンスシステムを開発している。本稿では透視投影画像生成に必要となる処理時間を計測し、視線変化と画像提示の間の時間遅延について考察する。

キーワード: テレプレゼンス, 全方位画像センサ, 透視投影画像生成, HyperOmni Vision

Real-Time Telepresence System Using an Omnidirectional Image Sensor and Experimental Analysis of Time Delay

Yoshio Onoe Kazumasa Yamazawa Haruo Takemura Naokazu Yokoya
Graduate School of Information Science, Nara Institute of Science and Technology

8916-5, Takayama, Ikoma, Nara 630-01, Japan

Tel: [+81]743-72-5296 Fax: [+81]743-72-5299

E-mail: {yoshio-o, yamazawa, yokoya, takemura}@is.aist-nara.ac.jp

Abstract: Telepresence attracts much attention in the fields of tele-robotics and tele-conferencing, because it drastically improves the understanding of a real environment in the distance. Conventional telepresence systems that use a rotating camera suffer from the significant time delay. We have been developing a real-time telepresence system that is based upon video-rate omnidirectional imaging and view-dependent perspective image generation. In this report we measure the computation time to make a perspective image and consider the time delay to display an image.

Key words: telepresence, omnidirectional image sensor, perspective image generation, HyperOmni Vision

1 はじめに

近年、移動ロボットの遠隔操作や多人数が参加する遠隔通信会議などにおいて、遠隔地の様子をより臨場感豊かに伝送したいという要求が高まってきている。遠隔地の画像は一般的に固定されたビデオカメラで撮影されるために、視野が限定されている。より良いコミュニケーションをはかるためには、カメラの視野を限定せず周囲の様子を伝えることが必要不可欠である。

遠隔地の情景を提示し、その場にいるのと同様の感覚を与える仮想現実技術はテレプレゼンスと呼ばれる。我々は全方位視覚センサで得られた全方位画像から、利用者の視線方向の透視投影画像を生成・提示する実時間テレプレゼンスシステムを提案した[1]。このシステムでは全方位画像を一度に伝送しているため、観察者の視線変化からその方向の画像提示までの時間遅延が少ないという利点があげられる。本稿では、開発したプロトタイプシステムの概要について述べるとともに、透視投影画像生成に必要な処理時間を計測し、画像提示時間遅延について考察する。

1.1 従来手法とその問題点

遠隔地の画像を獲得するためには、視覚センサの使用が一般的である。しかし観測視野がレンズの画角により制限されてしまう。この解決策として、ビデオカメラを方向が制御できる回転台の上において、カメラからの画像を獲得する方法がある(図1)。この方法では利用者が指示を出してからビデオカメラを動かすため、機械的な時間遅延が生じてしまう。また、ビデオカメラと利用者が遠く離れている場合は、両者間の通信時間の遅延が問題になる。そのほかの問題点として、多人数でテレプレゼンスを行うときに、一台のビデオカメラでは複数人がそれぞれ異なる方向を見渡すことができないことが挙げられる。

一方、利用者の見たい方向を自由に提示する手法として、静止画を利用した QuickTime VR[2] が知られている。この手法はパノラマ静止画像を用いており、利用者が上下・左右任意の方向を見渡せる。しかし QuickTime VR では、パノラマ画像を回転カメラや何枚ものスチル写真のつなぎ合わせで作成していた。これらの方法では動画画像系列は作成でき

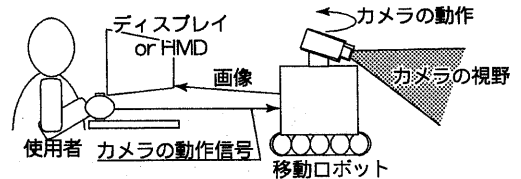


図 1: 可動カメラを用いたテレプレゼンス

ない。つまり、QuickTime VR は静止環境を対象とした蓄積再生型テレプレゼンスには利用できるが、実時間テレプレゼンスには不向きである。

実時間でテレプレゼンスを行なうために Virtual Dome[3] では、カメラを常時回転させて全方位の画像を撮影し、利用者が装着した HMD に利用者の視線方向の画像を提示している。この方法では見回しに関する時間遅延は少ないものの、画像の更新レートが遅いという問題点がある。

1.2 本研究の方針

テレプレゼンスにおいて観察者が遠隔地の情景を的確に把握するためには、広範囲の遠隔画像を見渡せることが望ましい。時間遅れなく見渡すためには、全方位の画像を一度に取得して、利用者の視線に追従した画像を実時間で生成する方法が必要である。

今までに、観測位置において周囲 360 度の画像を同時に取得できる様々な全方位視覚センサが提案されている [4] [5]。これらのセンサは何らかのミラーを通して周囲 360 度の情景を一つのカメラで撮像するものである。我々は、全方位視覚センサとして HyperOmni Vision[6] を用い、全方位画像から透視投影画像を実時間で生成するシステムを開発した。このシステムにより、時間遅れが少なく様々な方向を見渡せるようになった。本研究では透視投影画像生成に必要な処理時間を計測し、画像提示時間遅延について考察する。

2 全方位視覚センサ Hyper-Omni Vision

HyperOmni Vision は双曲面ミラーを用いた全方位視覚センサであり、センサの周囲 360 度の画像を一度に取得できる。また、他の全方位視覚センサとは異なり光学系の特性が透視投影であるため、得られる全方位画像から、ミラー焦点から見た通常の平面透視投影画像を生成できる。以下では、HyperOmni Vision の構成ならびに得られる全方位画像の特徴について述べる。

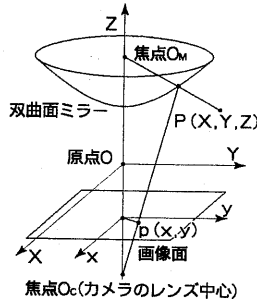
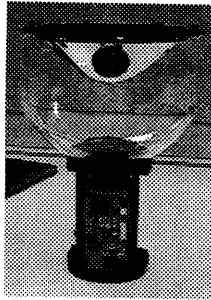


図 2: HyperOmni Vision の外観と光学系

2.1 HyperOmni Vision の特徴

図 2左に HyperOmni Vision の外観, 図 2右に光学系の構成を示す。HyperOmni Vision は, 下向きに設置した双曲面ミラーに写り込んだ鏡像を, 上向きに設置したビデオカメラで撮影するものである。その双曲面には回転 2 葉双曲面を用いており, 図 2 右のように 2 焦点 O_M, O_C を持っている。双曲面ミラーの外側にある焦点 O_C にカメラのレンズ中心が来るように, カメラを鉛直上向きに取り付ける。ここで図 2右に示すように Z 軸を鉛直軸とする 3 次元ワールド座標系を考えると, 双曲面ミラーの面, 双曲面ミラーの焦点 O_M , ならびにカメラのレンズ中心 O_C は次式で表せる (双曲面ミラーのパラメータ a, b, c は表 1 参照)。

$$\text{ミラー面} : \frac{X^2+Y^2}{a^2} - \frac{Z^2}{b^2} = -1 \quad (Z > 0)$$

$$\text{ミラー焦点 } O_M : (0, 0, +c) \quad (c = \sqrt{a^2 + b^2})$$

$$\text{カメラレンズ中心 } O_C : (0, 0, -c)$$

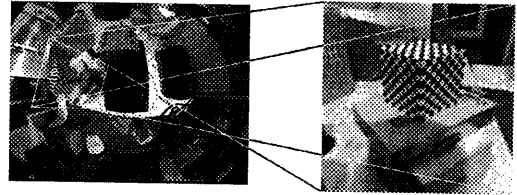
ただし, a, b, c : 双曲面ミラーのパラメータ (1)

内側の焦点 O_M に集まる像は双曲面ミラーを介して, 外側の焦点 O_C に集まる。したがって O_C にレンズ中心をおいたカメラで, 全方位の画像を撮影することができる。

双曲面ミラーの焦点 O_M とカメラのレンズ中心 O_C との位置関係ならびに双曲面の特性により, 3 次元環境中の点 $P(X, Y, Z)$ と, 全方位画像上の写

表 1: HyperOmni Vision パラメータ値

a	42.1 mm
b	42.7 mm
c	60.0 mm
焦点距離 f	14,500 pixel



(a) 全方位画像 (b) 透視投影画像
図 3: 全方位画像と透視投影画像の関係

像点 $p(x, y)$ の間には式 (2) の関係が成立する (図 2 右)。

$$x = X \times f \times \frac{(b^2 - c^2)}{(b^2 + c^2)Z - 2bc\sqrt{X^2 + Y^2 + Z^2}}$$

$$y = Y \times f \times \frac{(b^2 - c^2)}{(b^2 + c^2)Z - 2bc\sqrt{X^2 + Y^2 + Z^2}} \quad (2)$$

2.2 全方位画像と透視投影画像

HyperOmni Vision を用いて得られる全方位画像の例を図 3(a) に示す。同図からわかるように, HyperOmni Vision の双曲面ミラーの焦点を中心にして周囲 360 度の画像が取得できている。

この全方位画像の一部を変換して通常の平面透視投影画像を生成したものを図 3(b) に示す。この透視投影画像は, 同図 (a) の全方位画像上の四角形の枠と対応している。この変換を行なうためには, 通常, 透視投影画像上のすべての画素に対応する全方位画像上の対応点の位置を求める必要がある。しかしこの方法では, 計算量が膨大なものになり, 実時間で任意の方向の透視投影画像を得ることができない。このため本研究では次節に示すアルゴリズムを用いて, 簡易的な画像変換を行なう。

3 透視投影画像の生成

3.1 画像生成アルゴリズムの概要

HyperOmni Vision で撮影された全方位画像から通常人間が見ている透視投影画像に変換するためには, 式 (2) の変換式が必要である。したがって, 実時間での変換は難しい。

本研究では, 画像変換を実時間で行ない, かつ人間が不自然さを感じない程度に正確な透視投影画像を生成する必要がある。そのため, 生成画像上に図 4 のような格子をもうけ, 格子点ではソフトウェアにより正確な対応計算を行ない, 格子の中は使用した計算機のテクスチャマッピング機能を利用しハードウェアにより画像変形を行なう。このことにより変換を高速に行なえる。視線に追従した透視投影画像生成の概要は以下の通りである。

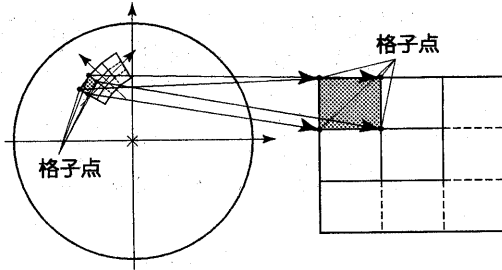


図 4: 入力画像 (左) と変換画像 (右) の関係

- (1) 磁気トラッカーにより視線を表す回転行列 K を得る。
- (2) 以下のステップ (i), (ii) により, 変換画像上の格子点 (図 4右) に対応する入力画像上の格子点 (図 4左) を計算する。
 - (i) 回転行列 K と格子点の変換画像上ローカル座標 (u, v) を次式 (3) 代入して, 格子点の 3 次元空間中の座標 (X, Y, Z) を得る。

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = K \begin{pmatrix} u \\ v \\ f_h \end{pmatrix} \quad (3)$$

ただし f_h : 視点と変換画像の間の焦点距離

- (ii) 3次元座標系での点 (X, Y, Z) を式 (2) に代入して, 全方位画像上の点 (x, y) を求める。
- (3) 入力画像を計算された格子で切り出す。
- (4) テクスチャマッピングにより, 変換画像上の各格子に対応する全方位画像の一部を張り付ける。これにより, 疑似平面透視投影画像を生成する。テクスチャマッピングには 2次元線形補間を用いる。

3.2 実時間テレプレゼンスシステム

前述の視線追従型画像生成提示手法を実装し, 実時間テレプレゼンスシステムのプロトタイプを構築した。

プロトタイプシステムの構成を図 5 に示す。また, 実際に使用した機器を表 2 に示す。全体の処理の流れは以下の通りである。

- (1) 全方位視覚センサ HyperOmni Vision で撮影した全方位画像を計算機に無線伝送する。
- (2) 利用者の装着する HMD に取り付けた 3D 磁気トラッカーで視線を計測する。
- (3) 計算機で視線に応じた透視投影画像を生成し, HMD に表示する。

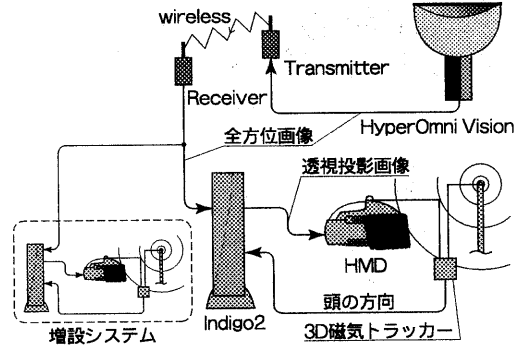


図 5: テレプレゼンスシステムプロトタイプの構成

4 実験と考察

前述のシステムを実装し, 近似変換による画像生成時間と画像ひずみの関係について調べた。ここで解像度は, 得られた全方位画像・生成された透視投影画像ともに 720 x 486 である。

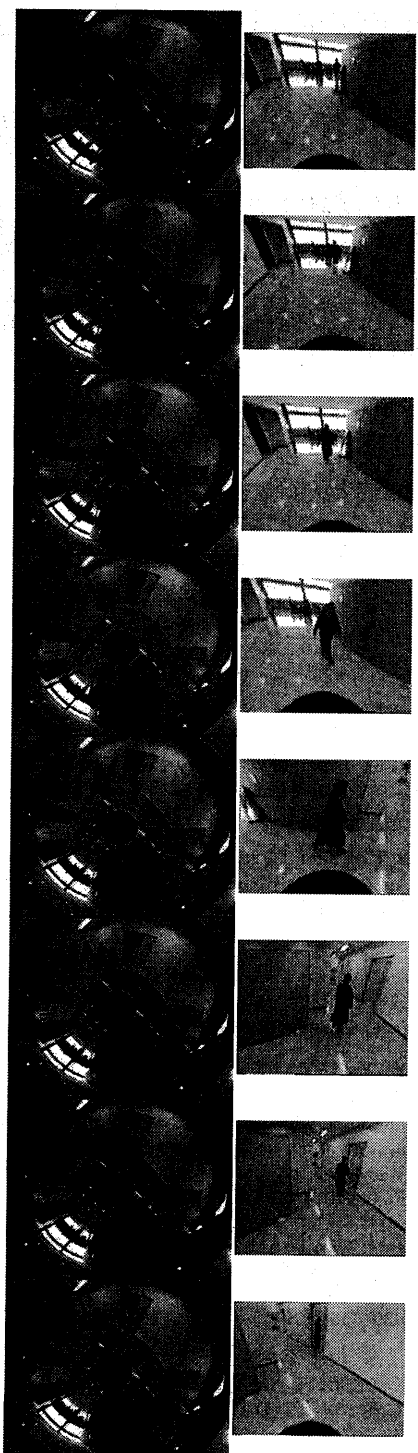
4.1 テレプレゼンスの例

図 6 に実験結果の一例を示す。左が全方位画像で右がそのときの利用者の視線に追従して生成された画像である (6.5 秒間のビデオストリームからのサンプル)。この場合, シーン内には 2 人の人物が存在し, 利用者の視線は 1 人の動きを追っているのが分かる。全方位の画像を一台のカメラで獲得して, その一部分を拡大して表示しているために解像度自体は粗くなっているが, 遠隔環境の状況把握には十分である。

従来の可動台の上においたカメラからの画像を用いる方法では, 多数の利用者がそれぞれ異なる方向を見渡すためにカメラを複数台用意しなければならない。しかし, 本手法の場合, 図 5 において点線の枠で囲まれた増設システムを複数用意することにより, 視覚センサは一台で複数人が同一視点から異なる方向を見渡すことができる。

表 2: 使用した機器

全方位視覚センサ	HyperOmni Vision ver.2
画像無線送受信器	Premier Wireless CS-200
計算機	SGI Indigo2 Impact + Impact Video (R4400, 250MHz)
磁気トラッカー	POLHEMUS 3SPACE FASTRAK
HMD	OLYMPUS Mediamask



(a) 全方位画像 (b) 透視投影画像
 図 8: テレプレゼンス実験における動画像系列

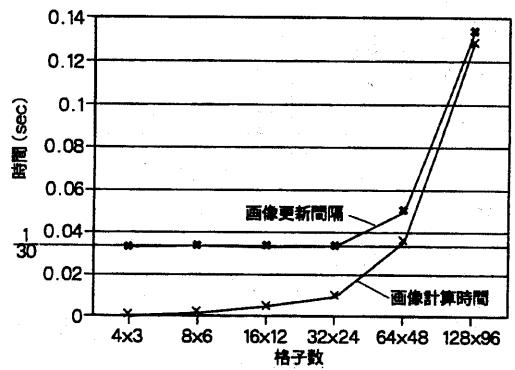


図 7: 格子数と時間

4.2 格子数と画像計算・提示時間

この節では格子数が画像計算・提示時間に及ぼす影響について述べる。図 7 に格子数と画像計算時間および画像更新間隔の関係を示す。画像計算時間¹は、格子数に比例して増えているのが解る。一方、画像更新間隔²は、1/30 秒を越えてからは画像計算時間と同様に増えているが、それまでは一定である。これは入力画像の取り込みが 1/30 秒ごとに行なわれており、この処理を待ってから画像が生成されるためである。このシステムでは、カメラと利用者の距離が離れている場合も、画像更新間隔や視線変化に対する画像表示時間の遅れが変わらない。

格子数が 16 x 12 の場合の画像表示の遅延は図 8 のようになる。視線取り込み・画像計算・画像表示はそれぞれ非同期で行なわれている。最もタイミングが良い場合 (図 8(a)), 画像計算開始時より 33.3ms 前の画像と 8.3ms 前の頭の方向が計算機に取り込まれる。そして、9.9ms かかって画像が生成された後、33.3ms までに HMD に表示される。最もタイミングが悪い場合 (図 8(b)) は、画像計算開始時より 33.3ms 前の画像と 16.6ms 前の頭の方向が計算機に取り込まれ、画像が 9.9ms かかって生成された後、66.6ms までに HMD に表示される。非同期のためタイミングが悪い時はタイミングが良い時と比べて、視線取り込みから表示まで 41.6ms 余分にかかる。したがって、同期について改善する余地がある。

¹ソフトウェアによる格子点の対応計算時間の実測値にハードウェアによる格子内のテクスチャマッピング処理時間の理論値を加えた値である。
²生成された画像が HMD に表示されてから次の画像が表示されるまでの間隔を意味する。

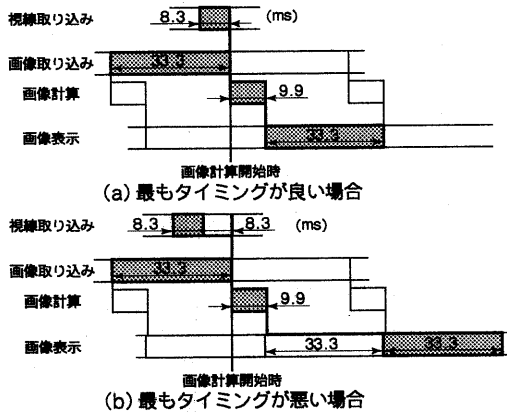


図 8: 画像表示の遅延

4.3 格子数と画像ひずみ

本研究では、実時間で全方位画像から任意の方向の透視投影画像を得るために、正確な平面透視投影画像変換を行っていない。実時間で変換するために格子をもうけ、格子点では正確な対応位置を計算し、格子の中はテクスチャマッピングで近似的に計算している。格子数が画像の幾何学的ひずみに及ぼす影響を調べるために主観評価実験を行なった。

図 9 にこの主観評価実験に用いた生成画像を示す。格子数が 4×3 の場合、変換画像が大きく湾曲している。格子数が 8×6 の場合、変換画像は 4×3 の場合と比べてかなり改善されたが、まだ違和感が残る。格子数が 16×12 の場合は、全方位視覚センサで撮影された直線が、変換画像ではほぼ直線に復元されて映っている。格子数が 32×24 の場合に交換された画像は、 16×12 の場合に交換された画像とほとんど変わらない。

4.4 考察

前節までの実験により、画像更新間隔が $1/30$ 秒でかつ変換された画像にゆがみを感じない格子の数は、 16×12 と 32×24 となる。したがって、これらの格子の数が今回の使用した機器での最適な値である。先に示した図 6 の実験例では 32×24 の格子を用いている。今後、使用する機器が変われば最適な格子の数も変化するものと思われる。

5 まとめ

これまでに我々は、HyperOmni Vision から得られる全方位画像を用いた視線追従型透視投影画像表示に基づく実時間テレプレゼンスシステムを提案した [1]。このシステムの特長として時間遅れが少ない

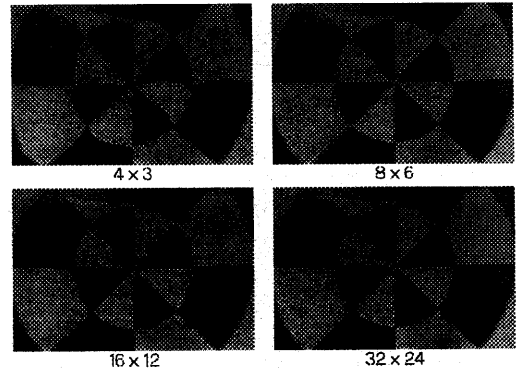


図 9: 格子数と画像の幾何学的ひずみ

く側方 360 度を見渡せることが挙げられる。本稿では、格子数と画像表示までの時間の関係や、格子数と画像ひずみの関係を調べ、最適な格子数を求めた。

今後の課題としては、移動物体を自動的に検出・追跡し、その物体の動きに追従した画像を提供する移動物体追跡型画像表示システムへの応用を考えている。

参考文献

- [1] 尾上, 山澤, 横矢, 竹村: “全方位画像からの視線追従型実時間画像生成によるテレプレゼンス”, 信学技報, PRMU97-20 (1997-05).
- [2] S. E. Chen: “QuickTime VR - An Image-Based Approach to Virtual Environment Navigation”, *SIGGRAPH 95 Conference Proceedings*, pp.29-38, 1995.
- [3] 広瀬, 佐藤: “人工現実感技術を用いた臨場感の合成とその伝達”, 9th Symposium on Human Interface, Oct.18-20, 1993.
- [4] Y. Yagi, Y. Nishizawa and M. Yachida: “Estimating location and avoiding collision against unknown obstacle for the mobile robot using omnidirectional image sensor COPIS”, *Proc. Int. Workshop on Intelligent Robots and Systems*, pp.909-914 (1991).
- [5] J. Hong, X. Tan, B. Pinette, R. Weiss and E. M. Riseman: “Image-based homing”, *Proc. IEEE Int. Conf. Robotics and Automation*, pp. 620-625 (1991).
- [6] 山澤, 八木, 谷内田: “移動ロボットのナビゲーションのための全方位視覚センサ HyperOmni Vision の提案”, 信学論 (D-II), Vol.J79-D-II, No. 5, pp.698-pp.707, 1996.
- [7] J. Neider, T. Davis, M. Woo: *OpenGL Programming Guide*, Addison-Wesley, 1993.