

全方位画像センサによるネットワークを介した テレプレゼンスシステムと遠隔監視システム

森田真司[†] 山澤一誠[†] 横矢直和[†]

[†] 奈良先端科学技術大学院大学 情報科学研究科 〒630-0101 奈良県生駒市高山町 8916-5

E-mail: {shinji-m, yamazawa, yokoya}@is.aist-nara.ac.jp

あらまし 本稿では、全方位画像センサによるネットワークを介したテレプレゼンスシステムと遠隔監視システムについて報告する。テレプレゼンスシステムは、遠隔地の全方位画像を獲得・伝送し、その画像より利用者の視線方向の画像を生成し、提示する。この手法は利用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ほとんどなく、ネットワークを介して遠隔地を広範囲に見回すことができる。また、遠隔監視システムはサーバ/クライアントモデルであり、サーバ側で移動物体を検出し、センサからの物体の方位を推定する。クライアント側では、ネットワークを介してサーバから受信した全方位画像と物体の方位情報を用いて、時系列での物体の同定を行い、画像提示による対象の自動追跡を行う。また、テレプレゼンスシステムと同様に、監視環境を利用者が自由に見回すこともできる。

キーワード ネットワークテレプレゼンス, Digital Video Transport System, 遠隔監視システム, 全方位画像センサ, 移動物体追跡

Networked Telepresence and Remote Surveillance Systems Using an Omnidirectional Image Sensor

Shinji MORITA[†] Kazumasa YAMAZAWA[†] and Naokazu YOKOYA[†]

[†] Graduate School of Information Science, Nara Institute of Science and Technology

8916-5 Takayama, Ikoma, Nara, 630-0101 Japan

E-mail: {shinji-m, yamazawa, yokoya}@is.aist-nara.ac.jp

Abstract This paper describes a networked telepresence system and a remote surveillance system using an omnidirectional image sensor. The networked telepresence system transports an acquired omnidirectional video stream of a remote site to an observation site, and then generates a view-dependent perspective video stream at the observation site. The method enables us to instantly look around a visualized space of a dynamic real world through the network. The remote surveillance system is based on a server/client model. The server PC of the surveillance system detects intruders and estimates their directions from a sensor. The client PC determines the correspondence among intruders in the current frame and those in the previous frame from their directions received from the server PC, and presents their perspective images to the user like the telepresence system. The user can also look around a visual space of a surveillance area.

Keyword Networked Telepresence, Digital Video Transport System, Networked Remote Surveillance System, Omnidirectional Imaging, Motion Tracking

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、バーチャルツアーなどにおいて、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感技術はテレプレゼンスあるいは没入型テレプレゼンスと呼ばれる[1]。

テレプレゼンスシステムとしてまず考えられる方法は、カメラを方向が制御できる回転台の上に置いて、

カメラからの画像を獲得する方法である。しかし、この方法では利用者が指示を出してから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、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も無視できなくなる。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。以上のような問題に対して我々は全方位画像センサ[2]を利用したテレプレゼ

ンスを提案した[3]．これは全方位画像センサにより遠隔地の周囲 360 度の全方位画像を獲得・伝送し，その画像より利用者に視線方向の画像を提示する．この手法では利用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く，広範囲の遠隔画像を見渡すことができる．しかし，従来のシステム構成では，全方位画像センサで撮影した全方位画像を直接，計算機に伝送しており，ネットワークを介した実装には到っていなかった．

以上の問題に対して本研究では，ネットワークを介して全方位画像を伝送するテレプレゼンスシステムを開発した．我々のシステムはネットワークを使って全方位画像を送受信し，かつ観測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．したがってネットワークを介して広範囲の遠隔画像を実時間で見渡すことができる．またひとつの全方位画像センサだけで複数人が同時に異なる方向を見渡すことができるという特徴をもっている．

さらに，テレプレゼンス技術の応用として，銀行や自宅における防犯や交差点での交通量把握，ビルのように複数の部屋や複数のビルを一括監視する場合などを目的としたカメラによる遠隔地の監視に対する要求も高まっている．監視システムの重要な機能として，環境を常に広範囲を撮影し，かつ注目すべき対象を実時間で検出・追跡することが求められる．

環境を常に広範囲を撮影するために，全方位画像センサを用いた方法がある[4, 5]．全方位画像センサを用いることにより，センサ 1 台または少数で環境全体を監視することができる．しかし従来の全方位画像センサを用いた手法は 1 台の計算機ですべてのセンサを管理しているものがほとんどで，ビル全体など多くのセンサを必要とする環境には適用できなかった．

以上の問題に対して，ネットワークテレプレゼンスを応用した実時間監視システムを開発した．本システムはサーバ/クライアントモデルであり，まずサーバ側において監視したい環境を全方位画像センサ HyperOmni Vision によって撮影し，その画像より移動物体を検出し，センサからの移動物体の方位を推定する．次にその全方位画像と方位情報を，ネットワークを利用してクライアント側に伝送する．クライアント側では時系列の方位情報を用いて移動物体の同定を行い，物体方向の画像提示による追跡を行う．これにより本システムではネットワークを介して実時間監視を行うことを可能とする．

以下では，2 節において本研究で採用した全方位画像センサ HyperOmni Vision の概要について説明し，3 節ではネットワークテレプレゼンスシステムの処理概要とシステム構成及び実験について述べる．4 節では

遠隔監視システムの処理概要，移動物体の検出・追跡，及びシステム構成と実験について述べる．最後に 5 節でまとめる．

2. 全方位画像センサ HyperOmni Vision

本研究で利用する全方位画像センサ HyperOmni Vision は，図 1 に示すように鉛直下向きに設置した双曲面ミラーとその下に鉛直上向きに設置したカメラから構成される．これにより図 2 のようにセンサの周囲 360 度を一度にカメラで撮影することができる．また，入力画像は双曲面ミラーの内側焦点を投影中心とした双曲面への中心投影画像であり，入力画像を内側焦点からみた一般の平面透視投影画像（図 3）やパノラマ画像に変換できる．

本研究ではネットワークテレプレゼンスシステムと遠隔監視システムにおいて送信（サーバ）側で撮影した全方位動画を，ネットワークを通じて受信（クライアント）側に送り，実時間で物体方向の平面透視投影画像に変換し利用者（監視者）に提示する．



図 1 HyperOmni Vision の概要

図 2 全方位画像

図 3 平面透視投影画像

3. ネットワークテレプレゼンスシステム

本節では，ネットワークテレプレゼンスシステムの全体の処理概要とシステム構成及び実験について述べる．

3.1. システムの処理概要

本システムは次の 3 つのステップから実現される．

- (1) ビデオレートでの全方位画像の取得
- (2) ネットワークを利用した全方位画像の伝送
- (3) 視線追従型実時間透視投影画像生成

なお，本システムは，センサ側に設置した計算機を送信側，観測者側に設置した計算機を受信側とした peer to peer 型システムであり，受信側において複数の利用者に遠隔地の画像を提示するものである．

3.2. 全方位画像の取得

前節で述べた全方位画像センサを遠隔地に設置し，全方位動画を取得する．例えばビルのロビーや会議室などにおいては，ロビー全体や会議室全体が映るような位置と高さに全方位画像センサを固定し，全方位動画を取得する．次に得られた全方位動画を DV ストリームとして送信側の計算機に入力する．

3.3. 全方位画像の伝送

本システムでは Digital Video Transport System (DVTS)[6]を用いて、前節で取り込んだ全方位動画画像を、ネットワークを介して受信側に伝送する。DVTS はネットワークを介してデジタルビデオ画像 (DV データ) を伝送できるシステムであり、約 33Mbps の帯域幅を使用して高品質動画画像 (640×480 pixels, 30fps) を伝送することが可能である。

3.4. 透視投影画像の生成

受信側では、送信側から伝送された DV ストリームを計算機で受信し、DV カムコーダなどを用いてこれを展開する。このようにして得られた全方位動画画像より、観測者の視線に追従した平面透視投影画像を生成する。透視投影画像生成の概要は以下の通りである。

(1) 磁気トラッカー、マウスなどのインタラクティブデバイスにより観測者の視線 (頭部の動き) を計測する。

(2) (1)で計測された視線情報と全方位動画画像から観測者の視線に応じた透視投影画像を生成する。この時、全方位動画画像から平面透視投影画像へは参考文献[3]の手法によりハードウェアのテクスチャマッピング機能を用いて実時間で変換する。

観測者が複数の場合は上記(1)と(2)の手順を各観測者に対して適用する。

3.5. システム構成と実験

前節で述べたネットワークテレプレゼンスシステムを構築した。図4と表1にシステムの構成図と使用した機器の仕様を示す。このテレプレゼンスシステムにおける解像度は DV カメラから得られる全方位画像と送信される全方位画像、入力される全方位画像ともに 640×480pixels, 出力される平面透視投影画像が NTSC の 720×486pixels である。

このシステム構成で、本学情報科学研究科棟のロビー(1階)を送信側として HyperOmni Vision を設置し、研究室(3階)を受信側として、学内 LAN を用いて動作実験を行った。実験では、送信側の環境中を複数の人物が歩行する状態での見回しを行った。なお、使用したネットワークは有線 100Mbps である。

得られた入力画像と 4 人の利用者の内、2 人の利用者の視線に追従して生成した出力画像の系列を図5に示す。実験より、全方位画像センサ 1 台で複数人が同時に同一視点から異なる方向を自由に見回すことができることを確認した。送信側で画像を撮影してから受信側の使用者に透視投影画像が提示されるまでの時間遅延は約 1 秒程度であった。また、30fps で全方位画像を受信することができ、かつ複数の平面透視投影画像を生成することができた。しかし、本システムは peer

to peer 型のシステムであり、現状では複数の受信側に全方位動画画像を伝送することはできない。また、全方位動画画像を伝送するために 33Mbps の帯域幅を使用しているため、実用的なアプリケーションを考えた場合には使用する帯域幅の削減が望まれる。

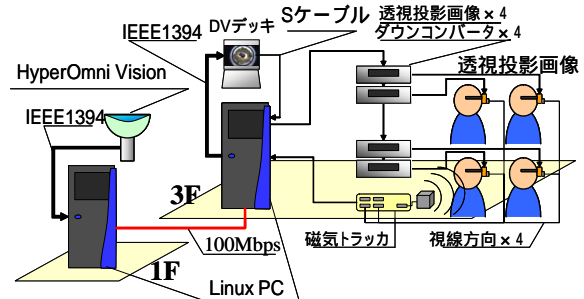
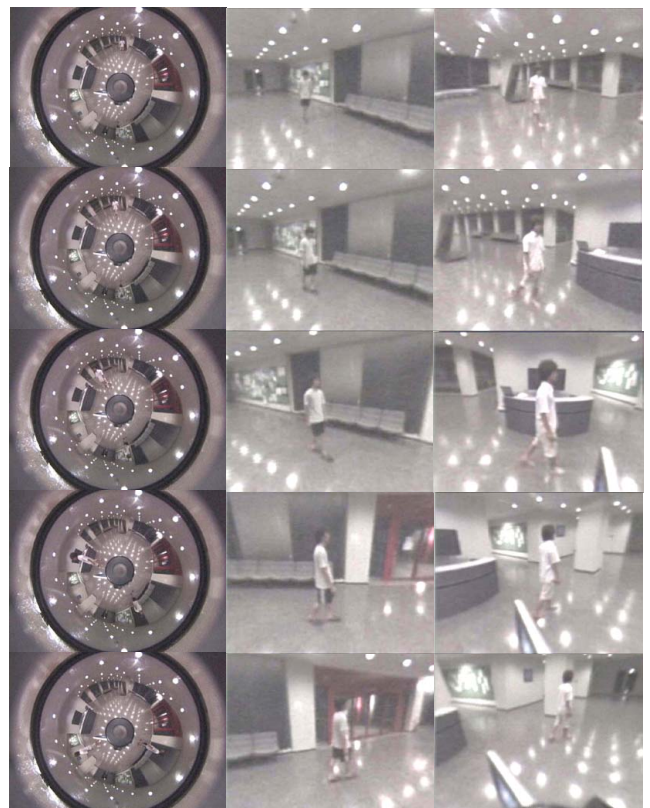


図 4 テレプレゼンスシステムの構成

表 1 テレプレゼンスシステムの構成機器

| 送信側・受信側 計算機 | OS:Linux CPU:Pentium4 2GHz Memory: 512MB |
|----------------|---|
| DV 伝送システム | Digital Video Transport System (dvts-0.9a22) |
| FMD | OLYMPUS EyeTrek FMD-700 |
| 磁気トラッカー | POLHEMUS 3SPACE FASTRAK |



(a)全方位画像 (b)透視投影画像 1 (c)透視投影画像 2

図 5 ネットワークテレプレゼンス

実験における動画像系列

4. 遠隔監視システム

本節では、ネットワークテレプレゼンスシステムを応用した遠隔監視システムについて述べる。

4.1. システムの処理手順

本システムは監視環境に設置した計算機をサーバ、監視者側に設置した計算機をクライアントとしたサーバ/クライアントモデルであり、ネットワークを用いて二者間のデータ通信を行い、ネットワークを介した遠隔監視を実現する。ネットワークテレプレゼンスの送信・受信側と同様に、サーバ側には全方位画像センサを1台と計算機を1台設置し、クライアント側には計算機を1台とDVカムコーダなどを1台設置し、画像提示により監視を行う。サーバ計算機はデーモンプロセスにより、クライアントの要求に基づいてコネクションを確立し、DVTSを用いてネットワークで全方位動画像を送信する。以下に本システムの処理手順を示す。

サーバ側の処理

(1) 全方位画像センサを監視する環境に設置し、全方位動画像を取得する。

(2) (1)で取得した全方位動画像から複数の移動物体を検出し、物体の方位情報を計算する。

(3) (1)と(2)で取得した全方位動画像と複数の移動物体の方位情報をクライアント側に送信する。

クライアント側の処理

(4) 全方位動画像と複数の移動物体の方位情報を受信する。

(5) (4)で取得した全方位動画像と方位情報より、時系列での複数の物体の同定を行い、監視者に物体方向の画像を提示する。

なお複数のサーバ、つまり複数の全方位画像センサを考えた場合でも、上記のようにサーバ側とクライアント側に処理を分散することで実時間の監視を行うことができる。

4.2. 移動物体の検出

本節では注目する移動物体をロバストに検出する手法について述べる。ここで、注目物体を監視環境中に新たに現れた物体、もしくは監視環境の中で移動する物体とする。本研究では背景差分に基づいた注目物体の検出手法を採用した。まず入力画像中の背景画素の輝度 I を以下の式のように仮定する。

$$I = I_{ave} + \sigma \sin(2\pi\omega t) + k\zeta \quad (1)$$

上式は入力画像中の各背景画素について成立し、 I_{ave} は輝度値の時間平均、 σ は輝度の振幅、 ω は輝度の周波数、 t は時間、 k は $-1, 0, +1$ のいずれかの値、 ζ はカメラのみに依存した雑音を表す。また、 $\sigma \sin(2\pi\omega t)$ の項は蛍光灯や CRT ディスプレイなどの

フリッカー、窓の外の樹木の揺らぎなどの影響を表し、 $k\zeta$ の項はゲインアップによるごま塩ノイズなどカメラにのみ依存する雑音を表す。このとき背景画素の輝度値 I は以下の範囲をとり得る。

$$I_{ave} - \sigma - \zeta \leq I \leq I_{ave} + \sigma + \zeta \quad (2)$$

本手法では、入力画素の輝度 I が式(2)の範囲におさまる場合はその画素は背景の画素、そうでない場合は物体の画素とする。

まず、あらかじめ移動物体が存在しない状況で ζ を測定しておく。この ζ はカメラのみに依存する値であるので、環境によって変更する必要はない。また、注目物体をロバストに検出するために、最初の j フレームにおいては I_{ave} のみを計算し、式(2)の判定を行わない。すなわち、注目物体の検出を行わない。ただし、 j はユーザの設定する値である。 I_{ave} の初期値を1フレーム目の I とし、最初の j フレームにおいて、以下の式を用いて毎フレームごとに I_{ave} を更新する。ただし、 I_{ave} は次フレームの I_{ave} とする。

$$\begin{cases} I'_{ave} = (n-1)/n \times I_{ave} + 1/n \times I \\ n: \text{更新速度パラメータ} \end{cases} \quad (3)$$

次に、背景のゆっくりとした変化を考慮し、各画素の輝度 I が式(2)の範囲におさまらず背景であると判断した場合、式(4)を用いて毎フレームごとに I_{ave} と σ を更新する。ただし、 σ' は次フレームの σ とし、 σ の初期値を $j+1$ フレーム目の $2 \times (I - I_{ave})^2$ とする。

$$\begin{cases} I'_{ave} = (n-1)/n \times I_{ave} + 1/n \times I \\ \sigma' = (n-1)/n \times \sigma + 1/n \times 2 \times (I - I_{ave})^2 \\ n: \text{更新速度パラメータ} \end{cases} \quad (4)$$

また、各画素が物体の部分であると判断された場合でも、急激な背景の変化や新たに監視環境中に置かれた静止物体に対応するため、式(5)を用いて σ を更新する。

$$\begin{cases} I'_{ave} = I_{ave} \\ \sigma' = (m-1)/n \times \sigma + 1/m \times 2 \times (I - I_{ave})^2 \\ m: \text{更新速度パラメータ} \quad (m \geq n) \end{cases} \quad (5)$$

これによって、新たな物体などがその場に静止し続けた場合には σ が増加し、しばらくするとその物体が背景と判断されるようになる。なお、本手法では、式(2)を用いて I が背景であるか、物体であるかの判定を行うため、 ω を設定する必要はない。

次に注目物体の方位情報を求める。具体的には毎フレームごとに物体とみなされた画素領域から図6に示すような方向の画素数のヒストグラムを作成する。このヒストグラムから連続した0より大きい値を持つ部分を画素領域の方位角範囲とする。この方位角範囲

の平均から、画素領域の中心角を求め、注目物体の方位角情報 θ_i ($i=0, 1, 2, \dots, l$) とする。図 6 においては全方位画像中に 2 つの物体が存在し、それらが黒色の領域 K_1 と K_2 で表されている。 K_1 と K_2 に含まれる画素からヒストグラムを生成し、 K_1 と K_2 の方位角範囲 (θ_1, θ_2) と (θ_3, θ_4) を求め、中点 $\theta_1 = (\theta_1 + \theta_2)/2$ と $\theta_2 = (\theta_3 + \theta_4)/2$ をそれぞれの方位角とする。さらに、各方位角範囲で検出された物体を囲むように仰角の範囲も求め、方位角と同様に、仰角情報 ϕ_i とする。これら θ_i と ϕ_i を注目物体の方位情報とし、クライアント側に送信する。ただし、 l は注目物体数を表す。

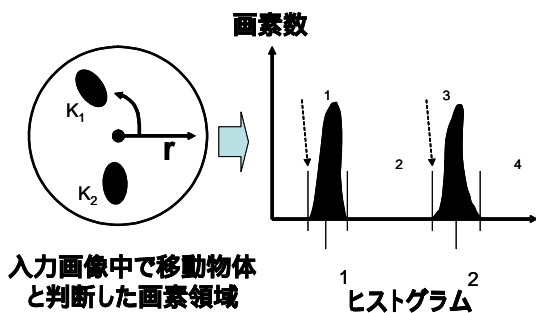


図 6 極座標変換によるヒストグラムの作成

4.3. 注目物体の追跡

本節では、クライアントにおいて時系列での注目物体の同定を行い、監視者に物体方向の画像を提示する手法について述べる。ただし、本節では、クライアント側においてサーバから伝送された方位情報を受信したフレームについて考える。

まず、サーバ側から更新された複数の注目物体の方位情報 θ_i と ϕ_i から式(6)より各注目物体の方位を 3 次元の単位ベクトル $G_{i(t)}$ で表現する。

$$G_{i(t)} = (G_{i(t)x}, G_{i(t)y}, G_{i(t)z}) \\ = (\cos \phi_i \cos \theta_i, \cos \phi_i \sin \theta_i, -\sin \phi_i) \quad (6)$$

次に前フレームと現フレームでの各単位ベクトル $G_{j(t-1)}$ と $G_{i(t)}$ の全ての組み合わせに対して、式(7)を用いて各組のなす角 ψ_{ij} を求める。

$$\psi_{ij} = \cos^{-1} \left(\frac{G_{i(t)} \cdot G_{j(t-1)}}{|G_{i(t)}| \times |G_{j(t-1)}|} \right) \quad (7)$$

各 $G_{i(t)}$ に対して、なす角 ψ_{ij} の値が最も小さい $G_{i(t)}$ と $G_{j(t-1)}$ とを対応付け、注目物体の同定を行う。最後に受信した全方位動画画像から対応付けされた複数の物体方向の平面透視投影画像を生成して監視者に提示する。全方位動画画像から平面透視投影画像へはネットワークテレプレゼンスと同様に参考文献[3]の手法によりハードウェアのテクスチャマッピング機能を用いて実時間で変換する。

4.4. システム構成と実験

図 7 と表 2 に遠隔監視システムの構成図と使用した機器の仕様を示す。このようなシステム構成で、3.5 の実験と同様に、本学情報科学研究科棟のロビー(1 階)を監視するようにサーバシステムを置き、研究室(3 階)を監視者側としてクライアントシステムを置いて、学内 LAN を用いて動作実験を行った。監視環境中に複数の人物が歩行し、さらに 1 つの物体を環境中に新たに置いた状態での監視を試みた。なお、使用したネットワークは有線 100Mbps である。本実験では、 $0 \leq l \leq 255$ に対して ζ は 4 とした。また、 j は 360、 n は最初の 360 フレームまではフレーム番号、361 フレーム以降は 360 とし、 m は 1080 とした。ここで n と m はいくつかの実験から決定されたヒューリスティックな値であるが、ビルのロビーや会議室などの一般的な環境に適用できると考える。

クライアント側で監視者に提示された画面の例を図 8 に示す。本実験では画面を 4 分割し、左下に全方位画像を、残り 3 つに検出された物体を入力画像中で面積の大きい方から提示した。ただし、注目物体が 3 つ未満の場合は、あらかじめ決められた速度で自動的に環境を見回す画像を提示するように実装した。また、提示画像を分割した提示画像の系列を図 9 に示す。図 9 において 2 人の人物と新たに置かれた物体が検出できていることがわかる。また、一定の時間存在した黄色の物体は背景と判断されていることも確認できた。

サーバ側における物体検出により方位情報は約 0.2 秒間隔で更新できた。またクライアント側においては平面透視投影画像の方位は約 0.2 秒ごとに更新されるものの、監視者に提示される画像は 22fps で更新できた。また蛍光灯などのフリッカーなどは無視し、ロバーストに移動物体のみを検出できることを確認した。しかし、本システムはネットワークテレプレゼンスと同様に、全方位動画を伝送するために 33Mbps の帯域幅を使用しているため、更なる帯域幅の削減が望まれる。また、サーバ側とクライアント側での方位情報と提示画像の更新速度が異なるために、クライアント側において方位情報が更新されないフレームが存在する。そのため、提示画像においてなめらかに注目物体を追従することができない。また、複数の注目物体が歩行中に交差した場合などでは、それらを 1 つの物体と認識してしまい、間違った対応付けが生じるなどの問題点が挙げられる。

5. まとめ

本報告では、全方位画像センサによるネットワークを介したテレプレゼンスシステムと遠隔監視システムについて述べた。本テレプレゼンスシステムにおける実験より、観測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く、ネットワークを介して、広範囲の遠隔画像を見渡すことができた。提案システムはさまざまな分野で応用が期待できる。例えば、遠隔操作支援、遠隔観察・監視、遠隔通信会議・授業支援、仮想観光、水族館テレプレゼンスなどでの実利用が考えられる。また遠隔監視システムは1対1のサーバ/クライアントシステムであり、ネットワークを利用して多対1のシステムへの拡張も考えた分散処理を行った。背景差分を利用し、かつ背景の更新を行うことで、ロバストに複数の移動物体を検出し、クライアント側において平面透視投影画像を提示することにより移動物体の確認が可能であった。

今後の課題としては、以下の3つが挙げられる。

・両システムの共通課題

(1) 複数の受信(クライアント)側に全方位動画を伝送するマルチキャスト型システムへの拡張と使用する帯域幅の削減を行うことより、実用的なアプリケーションとする必要がある。

・遠隔監視システムの課題

(2) さらに精度のよい物体検出を行うために、複数の全方位画像センサを用いることを考える。これにより死角領域が少なく、かつ複数のセンサを含む世界座標での物体の位置検出が可能である。

(3) クライアント側において、カルマンフィルタ等を用いることにより移動物体の方位情報を推定し、提示画像において滑らかな仮想カメラワークを可能とする。また、サーバ側から更新される方位情報との関係を考えることにより、フレーム間での間違っただけの対応付けを減少させることを考える。

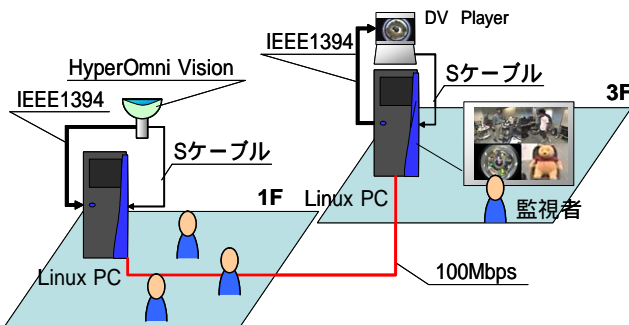


図 7 遠隔監視システムの構成

表 2 遠隔監視システムの構成機器

| | |
|----------------------|--|
| 計算機 (サーバ, クライアント) | OS:Linux CPU: Pentium4 2.0GHz x 2 Memory:512MB |
| DV 伝送システム | Digital Video Transport System (dvts-0.9a22) |



図 8 提示画像

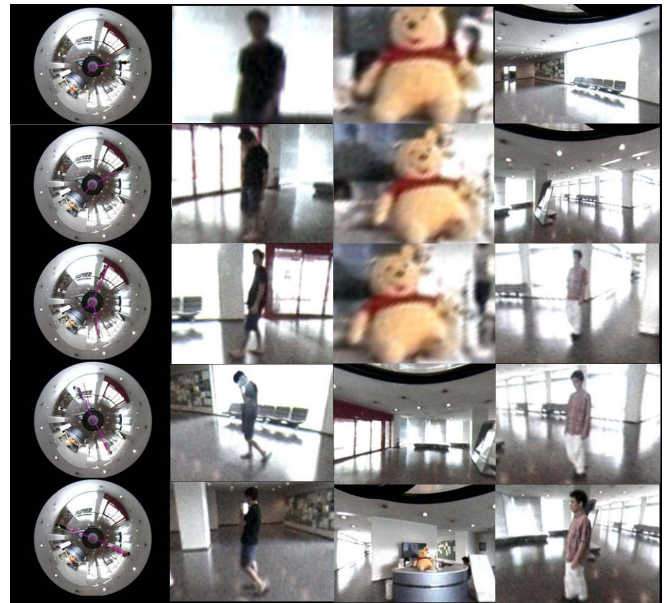


図 9 遠隔監視システム実験における動画像系列

文 献

- [1] "Special issue on immersive telepresence," IEEE Multi Media, vol.4, no.1, pp.17-56, 1997.
- [2] 山澤, 八木, 谷内田: "移動ロボットのための全方位視覚センサ HyperOmni Vision の提案," 信学論, Vol.J79-D-II, No.5, pp.698-707, 1996.
- [3] Y. Onoe, K. Yamazawa, H. Takemura, and N. Yokoya: "Telepresence by real-time view-dependent image generation from omnidirectional video streams," Computer Vision and Image Understanding, Vol.71, No.2, pp.154-165, 1998.
- [4] 十河, 石黒, Trivedi: "複数の全方位画像センサによる実時間人間追跡システム," 信学論, Vol.J83-D-II, No.12, pp.2567-2577, 2000.
- [5] 寺沢, 山澤, 竹村, 横矢: "複数の全方位画像センサを用いた遠隔監視システムにおける複数移動物体の存在領域推定," 信学技報, PRMU2000-195, 2001.
- [6] A. Ogawa, K. Kobayashi, O. Nakamura, and J. Murai: "Design and implementation of DV stream over internet," Proc. IWS Internet Workshop, No.99EX385, 1999.