

全方位画像センサを用いたネットワーク対応型遠隔監視システム

森田 真司^{†*} 山澤 一誠[†] 寺沢 征彦^{†**} 横矢 直和[†]

Networked Remote Surveillance System Using Omnidirectional Image Sensors

Shinji MORITA^{†*}, Kazumasa YAMAZAWA[†], Masahiko TERAZAWA^{†**}, and Naokazu YOKOYA[†]

あらまし カメラを使った遠隔監視においては、環境を常に広範囲に撮影し、かつ注目対象の実時間で検出・追跡が求められる。環境を常に広範囲に撮影するために回転カメラを用いる方法や全方位画像センサを用いる方法が提案されている。前者では見回しに機械的な時間遅延が生じる。これに対し後者では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が、取得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至っていない。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い。本論文ではネットワークを介して複数の全方位画像を送送する遠隔監視システムについて述べる。本システムは監視環境側をサーバ、監視者側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、サーバ側において移動物体の検出、クライアント側において移動物体の位置推定及び物体方向の画像提示による追跡を行う。また、ネットワークを利用した画像及び物体の検出情報の伝送を行うことで、クライアント側での一極集中型遠隔監視を実現する。

キーワード 全方位画像センサ HyperOmni Vision, ネットワーク, Digital Video Transport System, 遠隔監視システム, 移動物体追跡

1. まえがき

近年、銀行における防犯や交差点での交通量把握、ビル内の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が高まっている。監視システムの重要な機能として、環境を常に広範囲に撮影し、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。

監視システムにおいて、従来使用されている標準レンズを備えた CCD カメラは、画角が狭く、環境全体を監視できない。したがって環境を広範囲に動く複数の物体の位置すべてを知ることはできない。この問題の解決策として、監視者がカメラを制御し、そのカメラからの画像を得る方法がある。しかし、手動制御の場合、監視者が常に画像を注視している必要があり監視者の負担になる。一方、物体の動く方向を時系列差分やオプティカルフローなどにより予測し自動的にカメラを回転させることによって、一つの物体を監視す

ることができる [1] ~ [6]。しかし、一台のカメラでは環境全体の画像を一度に取得することはできないため、現在写っている物体の動く方向しか予測できず、複数の物体を同時に監視することができない。

また、複数の方向を制御できるカメラを使用することにより複数の物体を追跡する手法がある [7] ~ [9]。これらの手法では追跡する物体の数と同じかそれ以上の数のカメラを必要とし、物体の位置を推定するにはさらに多くのカメラを必要とする。また、カメラの方向を変化させるため常に環境全体を監視することは難しく、その時にカメラの死角があれば新たに監視範囲に入った物体を見逃すことがある。

もう一つの解決策として通常のカメラではなく、全方位画像センサを使用する方法が考えられ、これらの方法では、通常のカメラに比べ少ない台数で常に環境全体を監視することができる [10] ~ [12]。また回転カメラと違い、カメラの可動制御を行う必要がなく扱いやすい。これまでに複数の全方位画像センサを使用して、物体の位置を推定する手法が提案されている [13] ~ [15]。十河らは移動物体を人物と限定し、その位置を一定の大きさの円として推定する手法を提案している [14]。しかし円の大きさは経験的に求めた値で

[†] 奈良先端科学技術大学院大学
Nara Institute of Science and Technology

* 現在、株式会社日立製作所

** 現在、住友電工ネットワークス株式会社

あり、様々な大きさの物体の位置を推定するのは困難である。さらに円の中心は三角測量により求めるので、物体がベースライン付近に存在する場合、物体の方位角の導出誤差による推定位置の誤差が大きく、推定結果が不安定になる。さらに物体の方位角とベースライン方向が完全に一致すると、三角測量を行うことができない。また、障害物などによる部分的なオクルージョンや物体の向きなどによる物体の方位角の導出ずれにより推定位置がずれるという問題もある。このような問題に対して、寺沢らは、2次元的存在領域として物体の位置推定を行う手法を提案している[15]。この手法では、物体の存在領域を背景差分領域から見える物体の方位角の範囲を使って求めるので、センサで抽出される物体の角度情報の導出誤差をそのまま反映し、様々な大きさの物体の位置推定を行うことが可能である。しかし、これらの手法では、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、本論文では全方位画像センサとネットワークを利用した分散処理による実時間監視システムの開発を目指す。本システムは監視環境側をサーバ、監視者側をクライアントとしたサーバ/クライアントモデルである。クライアント側において安定した複数の移動物体の位置推定を行うために、1サイトにつき3台以上の全方位画像センサを用いて監視環境を撮影する。また、各サーバにおいて背景を逐次更新しながら背景差分を行い、屋内における照明条件の変化に対してもロバストに物体検出を行う。以上のようにして構築するシステムはセンサ数と監視者側の数に関して、容易に拡張が可能であり、ビル全体など多くのセンサを必要とする環境に適用が可能であり、ネットワークを介した実時間遠隔監視を行うことができる。

以下では、2.章において本システムの枠組を説明する。3.章ではサーバ側における移動物体検出と画像送信について述べ、4.章ではクライアント側における移動物体の存在領域推定と画像提示による追跡について述べる。5.章では構築したプロトタイプシステムを用いて行った実験により、開発したシステムが有効であることを示し、6.章でまとめる。

2. 全方位カメラを用いたネットワーク対応遠隔監視の枠組

本システムは監視環境の計算機をサーバ、監視者側

の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両者間のデータ通信による分散処理を行い、ネットワークを介した実時間遠隔監視を実現する。図1に本システムの処理の流れを示す。監視環境側にはサーバシステム(全方位画像センサと計算機)を3組以上、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。サーバ側では、3つ以上の全方位画像センサを用いて常に監視環境を広範囲に監視する。本システムでは全方位画像センサとして双曲面ミラーとビデオカメラから成るHyperOmni Vision[16]を用いる。また、サーバはデモンプロセスによりクライアントの要求によりコネクションを確立し、ネットワークを介して全方位動画画像を送送する。

各サーバにおいては屋内における照明条件の変化を考慮した背景更新を行い、注目物体を検出する。また、クライアントはネットワークを用いて、注目物体の検出情報を各サーバから受信することができる。クライアントはそれら複数のサーバから送信された検出情報を用いて、複数の注目物体の位置推定を行い、物体方向の画像を監視者に提示することで追跡を行う。HyperOmni Visionの入力画像は双曲面ミラーの内側の焦点を投影中心とした双曲面への一点中心透視投影画像であるため、入力画像(全方位画像)の一部を双曲面ミラーの内側の焦点から見た一般の平面透視投影画像に変換できるという特徴を持つ[16]。この特徴を利用して、物体方向の画像を監視者に提示する際には、全方位画像から物体方向の平面透視投影画像を生成する。

以下では本システムの処理手順をサーバ側の処理とクライアント側の処理、サーバとクライアント間のネットワークプロトコルに分けて概要を示す。ただし複数のサーバと複数のクライアントが存在し、各サーバは常時監視環境に存在する注目物体を検出し、1つのクライアントは常時3つのサーバと接続されていると仮定する。さらにクライアントは3つの全方位画像センサのうち1つの全方位動画画像のみを受信とする。

サーバ側の処理

- (1) 全方位動画画像を取得する。
- (2) 全方位動画画像から複数の注目物体を検出し、物体の方位情報を計算する。
- (3) 複数の注目物体の方位情報をクライアント側に伝送する。また、クライアントから要求がある場合は、全方位動画画像を送信する。

クライアント側の処理

- (1) 3つのサーバからの複数の注目物体の方位情報を受信する。
- (2) (1)で受信した各サーバからの方位情報を用いて、複数の注目物体の位置を推定する。
- (3) 1つのサーバに画像送信を要求し、全方位動画を受信する。
- (4) (1)の方位情報と(3)の全方位動画より、監視者に物体方向の画像を提示する。

サーバとクライアント間のネットワークプロトコル

- ・クライアント側からサーバ側への情報通信。
 - (1) 全方位動画の取得の要求。
 - (2) 上記(1)で受信した全方位動画の切断の要求。
 - (3) 注目物体の方位情報の取得の要求。
- ・サーバ側からクライアント側への情報通信。
 - (1) 全方位動画の取得の要求に対する確認情報と動画の送信。ただし、動画送信済みのサーバの場合、動画の送信は行わない。
 - (2) 全方位動画の切断の要求に対する確認情報の送信。
 - (3) 注目物体の方位情報の取得の要求に対する方位情報の送信。

以上のようなサーバ/クライアントモデルにおける処理構成により、複数のサーバと複数のクライアントが存在する場合への拡張が容易であると考える。

3. サーバ側の処理

本節ではサーバ側において注目物体をロバストに検出し、全方位画像を送信する手法について述べる。

サーバはまず注目物体の写った全方位画像から注目物体の存在を検出する。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。ここでは、ゲインアップによるごま塩ノイズなどのカメラにのみ依存する明度変化や、蛍光灯やCRTディスプレイなどのフリッカーなどによる明度変化を考慮して、ロバストに注目物体を検出することを目指す。

3.1 動的背景更新による物体検出

まず入力画像中の背景画素の輝度 I を以下の式でモデル化する。

$$I = \bar{I} + \sigma \sin(2\pi\omega t) + k\zeta \tag{1}$$

上式は入力画像中の各背景画素について成立し、 \bar{I} は

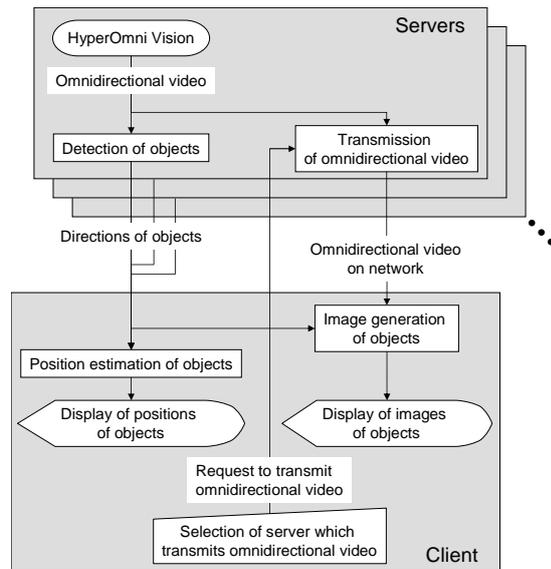


図1 遠隔監視処理の流れ
Fig.1 Flowchart of surveillance

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 $-1 \sim +1$ の値をとる係数、 ζ はカメラのみに依存した雑音の最大値を表す。ここで、 $\sigma \sin(2\pi\omega t)$ 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。

$$\bar{I} - \sigma - \zeta \leq I \leq \bar{I} + \sigma + \zeta \tag{2}$$

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、あらかじめ注目物体が存在しない状況でカメラの特性として 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 \bar{I} 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 は経験的に設定する値である。 \bar{I} 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 \bar{I} を更新する。ただし、 \bar{I}' は次フレームの \bar{I} とする。

$$\bar{I}' = (n-1)/n \times \bar{I} + 1/n \times I \tag{3}$$

n : 更新速度パラメータ

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) の範囲におさまり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 \bar{I} 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 $2 \times (I - \bar{I})^2$ とする。

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

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。

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

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。この時、 m を無限大にすることで静止物体は注目物体として検出され続け、背景画像に取り込まれない。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。

3.2 物体方位情報の計算と送信

次に毎フレームごとに物体とみなされた画素領域から注目物体の方位情報を求める。具体的には図2に示すように、 θ 軸方向に作成した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(図2においては $\Theta_\alpha(\theta_{\alpha 1}, \theta_{\alpha 2}), \Theta_\beta(\theta_{\beta 1}, \theta_{\beta 2})$)。次に各方位角範囲内について r 方向にヒストグラムを作成し0より大きい値を持つ部分を画素領域の仰角範囲として求める。ここで、方位情報とは全方位画像センサにおける方位角範囲と仰角範囲を表し、これらの方位情報をクライアント側に伝送する。

また、クライアントの要求に応じてサーバは全方位

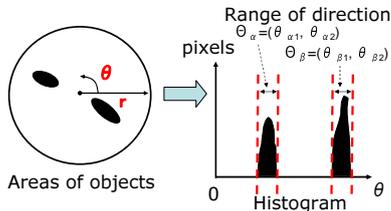


図2 極座標変換によるヒストグラムの作成
Fig.2 Generating a histogram using polar-coordinate conversion

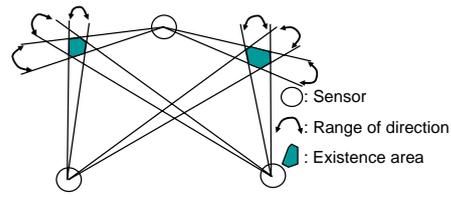


図3 注目物体の存在領域
Fig.3 Existence areas of focused objects

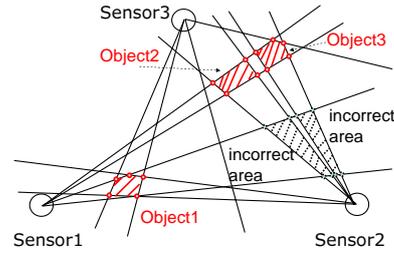


図4 複数物体の特殊な位置関係
Fig.4 Accidental positions of multiple objects

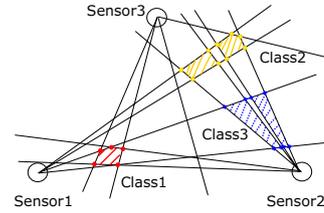


図5 存在領域候補のクラス分け
Fig.5 Grouping reliabilities of existence areas

画像をクライアントに送信する。本研究では、Digital Video Transport System(DVTS) [17] を用いて全方位画像を伝送する。DVTS はインターネットを介してIEEE1394 デジタルビデオストリーム (DV ストリーム) を伝送できるシステムであり、約 33Mbps の帯域幅を使用して、高品質動画像 (720 × 480 pixels, 30fps) を伝送することが可能である。

4. クライアント側の処理

本節では、クライアント側において、サーバ側から送られた複数の注目物体の方位情報を用い、物体の位置推定を行い、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。

4.1 多角形領域の辺導出による存在領域推定

サーバ側から送られた複数の注目物体の方位角範囲のみを用いて注目物体の位置推定を行う。本研究では、注目物体の位置推定手法に寺沢らの手法を採用した [15]。この手法は 2 次元的な存在領域として注目物体の位置を推定するものであり、図 3 のように各セン

サの位置からのばした方位角範囲が3つ同時に重なる多角形領域を注目物体の存在領域として位置を推定する。

以下、注目物体を表す多角形領域の辺導出による注目物体の存在領域推定を注目物体の存在領域候補の決定と信頼度に基づく存在領域候補のクラス分けに分けて述べる。

直線の媒介変数表現による交点座標の導出

2つの直線の交点の座標値は、世界座標系での2つのセンサの位置ベクトルをそれぞれ $\vec{S}_1(S_{x1}, S_{y1})$, $\vec{S}_2(S_{x2}, S_{y2})$, センサからのばした直線の単位ベクトルをそれぞれ $\vec{V}_1(V_{x1}, V_{y1})$, $\vec{V}_2(V_{x2}, V_{y2})$, 媒介変数をそれぞれ M_1, M_2 とすると、以下の式で表せる。

$$\begin{aligned} (\text{交点の } x \text{ 座標値}) &= S_{x1} + V_{x1} \cdot M_1 \\ &= S_{x2} + V_{x2} \cdot M_2 \\ (\text{交点の } y \text{ 座標値}) &= S_{y1} + V_{y1} \cdot M_1 \\ &= S_{y2} + V_{y2} \cdot M_2 \end{aligned} \quad (6)$$

以上の式より、各媒介変数は以下の式で求めることができる。

$$\begin{aligned} M_1 &= \frac{(S_{x2} - S_{x1}) \cdot V_{y2} - (S_{y2} - S_{y1}) \cdot V_{x2}}{V_{x1} \cdot V_{y2} - V_{y1} \cdot V_{x2}} \\ M_2 &= \frac{(S_{x1} - S_{x2}) \cdot V_{y1} - (S_{y1} - S_{y2}) \cdot V_{x1}}{V_{x2} \cdot V_{y1} - V_{y2} \cdot V_{x1}} \end{aligned} \quad (7)$$

異なるセンサからのばした全ての2直線について以上の処理を行い、各交点における媒介変数を求める。

次に求めた交点から注目物体の存在領域候補を構成する辺を導出し、その辺を連結することにより物体が存在するであろう領域を存在領域候補として決定する(詳細は参考文献[15]を参照)。

存在領域候補の信頼度に基づくクラス分け

複数の物体が特殊な位置関係で存在すると、物体数以上の存在領域候補が存在し、物体の存在しない領域が誤って検出される。図4のように、3つの物体の特殊な位置関係によって誤り領域が発生し、寺沢らの手法[15]ではこれらは削除している。しかし、実際に5つの物体がそれぞれの存在領域候補に存在した場合、誤り領域に存在する物体を検出することができない。

本研究では検出した注目物体の存在領域候補方向を示す各センサの方位角範囲を基に、それらが実際の存在領域としてどれだけ信頼できるかという信頼度ごとにクラス分けを行う。ここで、信頼度とは各センサの方位角範囲にある存在領域候補数の逆数であり、一つの存在領域候補は各センサからの3つの信頼度を持つと

考える。例えば、図4における物体1の領域はセンサ1の方位角範囲(1)に関しては1/3、センサ2の方位角範囲(2)に関しては1/1、センサ3の方位角範囲(3)に関しては1/1の信頼度を持つと考える。以下に存在領域候補の信頼度に基づく各クラスを示す。ただし、クラス1が最も物体の存在する可能性が高く、次いでクラス2、クラス3の順となる。

・クラス1

図5におけるClass1と示された領域。存在領域候補を持つ3つの信頼度の内、信頼度が1/1であるものが少なくとも一つ存在する領域。

・クラス2

図5におけるClass2と示された領域。存在領域候補を持つ3つの信頼度全てが1/2以下であり、かつ各方位角範囲にクラス1の存在領域候補が存在しない場合。

・クラス3

図5におけるClass3と示された領域。存在領域候補を持つ3つの信頼度全てが1/2以下であり、かついずれかの方位角範囲にクラス1の存在領域候補が存在する場合。

全ての存在領域候補を信頼度に基づいて、上記のようなクラスに分類し、監視環境内に存在する注目物体の存在領域とする。

4.2 注目物体方向の画像提示

最後に、受信する全方位動画から複数の物体方向の平面透視投影画像を生成し、監視者に提示する。

ここで、全方位動画の取得要求を送信するサーバは監視者の手動で選択されるものとする。クライアントは選択したサーバに全方位画像の伝送を要求し、DVTSにより伝送されてきたDVストリームをDVレコーダを用いて展開を行う。さらに、展開された全方位動画とそれに対応した注目物体の方位情報を用いて、物体方向の平面透視投影画像を生成する。平面透視投影画像への変換はイメージワーピングにより実時間で実行される(詳細は参考文献[18]を参照)。ただし、仰角範囲を基に生成する画像の焦点距離を変化させることで、仮想カメラのズーム機能を実現する。生成した画像を提示する際、フレーム間で最も方位情報が近いものを対応付けることにより対象の追跡を行い、提示画像の切り替えに利用する。

5. プロトタイプシステムの構築

本節では、提案手法を用いて構築したプロトタイプシステムについて述べ、次に、注目物体の検出及び画像提示による追跡実験について述べる。

5.1 システム構成

図6 システムの構成図を示す。構築したシステムでは、本学情報科学研究科 B 棟 3 階 B308 を監視環境とし、B211 を監視者側とした。また、ネットワークは有線 100Mbps の学内 LAN を、計算機は CPU Pentium4 2GHz、メモリ 512MB を使用した。各センサで取得した全方位画像をそれぞれに接続されたコンピュータに入力し、サーバ側で注目物体の検出を行い、クライアント側で物体の位置推定を行い、画像提示による追跡を行うものである。ただし、全てのサーバシステムにおいて、非同期で全方位画像の入力及び注目物体の検出が実行される。

5.2 人物の検出・追跡実験

構築したシステムを用いて、2 種類の注目物体の検出及び追跡の実験を行った。以下、それぞれの実験の設定、結果について述べる。

実験の設定

監視環境は $9.2m \times 9.2m$ の大きさの実験室であり、正方形の形をしているので、床面の角の 1 頂点を原点とし、この点を含む 2 辺を x 軸、 y 軸とする座標を世界座標とした。そして 3 台の全方位画像センサを $(0.0m, 0.0m)$, $(4.5m, 0.0m)$, $(3.0m, 4.5m)$ の位置に配置した。また床面から双曲面ミラーの内側の焦点までの高さを $1.7m$ とした。なお、サーバにおける注目物体の検出において、本実験では $0 \leq I \leq 255$ 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、ピルのロビーや会議室などの様々な屋内環境で適用できた。

・実験 1

1 人の人物の位置推定実験。人物は設定したルートを歩行することで、正しく注目物体の存在領域が検出されていることを確認する。なお、設定したルートは世界座標において $(1.5m, 0.0m) \rightarrow (3.0m, 0.0m) \rightarrow (3.0m, 2.0m) \rightarrow (1.5m, 2.0m) \rightarrow (1.5m, 0.0m)$ の順に直線移動するルートであり、人物は通

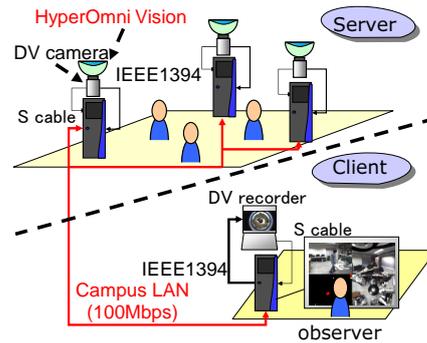


図6 遠隔監視システムの構成
Fig.6 Construction of surveillance system



図7 監視者への提示画像
Fig.7 Displayed image

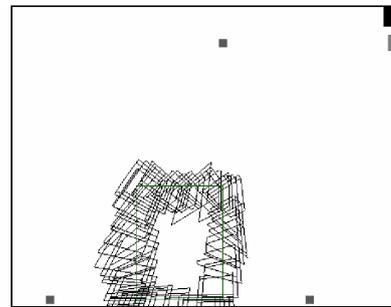


図8 推定した存在領域
Fig.8 Estimated existence areas

常の人間の歩行速度に近い約 $80cm/s$ で歩行する。

・実験 2

2 人の人物が監視環境を自由に歩行し、さらに 1 つの静止物体を監視環境に新たに置いた状態での注目物体の検出及び追跡。

クライアント側で監視者に提示される画面の例を図7

に示す。本実験では画面を4分割し、左下に位置推定結果を、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や全方位画像センサのミラーの内側の焦点から鉛直下向きの画像を提示するように実装した。なお、位置推定結果において、三角形に配置された小さな正方形がセンサの位置、画像全体が監視環境を表すようにし、4.1節において述べた各存在領域のクラスを黒の濃さで表した。右上の矩形がその指標で、それぞれクラス1,2,3に対応する。

実験結果

まず実験1において推定した全ての存在領域を描画したCGを図8の多角形領域(ワイヤフレーム)で示す。図8の中央下に矩形に書かれた線が人物が目標とした歩行ルートを示す。また、位置推定実験時の各サーバの入力全方位画像と注目物体の検出結果の内、選択したフレームを図9に、クライアントにおいて提示した平面透視投影画像の内、図9と同一のフレームを図10に示す。ここで図9は各センサの入力全方位画像にCGで描画した物体の検出結果を重畳表示したものであり、2本の直線で挟まれた領域が検出された注目物体を示す。図8より設定した移動ルートに沿って人物の存在領域が推定できていることが確認できる。また、図9より、各サーバにおいて、蛍光灯などのフリッカーは無視し、1人の人物が検出され、物体の方位情報が検出されていることが確認できる。また、図10より、各サーバから送信された方位情報を基に物体方向の透視投影画像を作成・提示できていることがわかる。ここで監視環境に存在する注目物体が1つであるため、図10(b)と(c)は物体方向の画像を提示しておらず、一定の速度で自動的に環境を見回す画像と決められた方向の画像を提示していることが確認できる。

また、実験2における各サーバの入力全方位画像と注目物体の検出結果の内、選択したフレームを図11に、クライアントにおいて提示した平面透視投影画像と位置推定結果の内、図11と同一のフレームを図12に示す。ただし実験1とは異なり、推定した注目物体の存在領域を各クラスの濃さで塗りつぶされた多角形領域で示す。図11と図12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物体は図11と図12のフレーム4で背景と判断され、注目物体として検出されていないことも確認できる。図12(d)から各サーバから送信され

た複数の注目物体の方位情報を基に複数の物体の存在領域を同時に推定できていることがわかる。人物が複数存在する場合、図12(d)のフレーム3の推定領域のCGからわかるように、2人の人物の特殊な位置関係により、存在確率が小さい濃さの薄い領域が生じている。

両実験において、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人物の方位情報は約0.2秒ごとに更新され、それに基づいた移動物体の位置推定を行うことができた。なお、人物の方位情報は約0.2秒ごとに更新されるものの、提示画像は約0.05秒ごとに更新できた。なお、遠隔地における全方位画像の取得から画像提示までの時間遅延は約1秒であり、クライアント側においてパケットロスなしに30fpsでデジタルビデオストリームを受信することができた。

5.3 考察

実験1の結果について考察する。クライアント側では各サーバにおいて非同期に検出された物体の方位情報を用いて、複数の物体の位置推定を行う。したがって、クライアント側での処理フレームごとに各センサの方位角範囲の時間的なずれが位置推定に影響を及ぼすことが考えられる。しかし、実験1の結果より通常の人物の歩行速度程度までならば移動する物体の存在領域の推定が可能であると考えられる。

実験2の結果について考察する。複数の物体を検出する際、特殊な位置関係により間違った存在領域も検出されるが、そのような領域は存在確率が小さい領域として推定可能である。さらに、その領域に物体が存在するかしないかは提示画像によりユーザが確認可能である。また、図12のフレーム1では(d)において注目物体が3つ検出されているが、検出された物体の画像は2つだけ表示されている。これは新たに置かれた静止物体と人物がセンサ1から見て重なっているからであり、その他のセンサから見た画像に切り替えれば3つの画像が表示されると考える。

以上のように、本研究で構築した遠隔監視システムは監視環境内に存在する複数の人物を同時に検出し、ネットワークを介して監視者が画像提示による追跡を行うことが可能である。さらにサーバ側における背景更新により、監視環境内に一定の時間存在した静止物体などは背景として判断され、注目物体として検出されない。照明条件などによる背景の変化がある場合でも、ロバストに注目物体の検出を行うことができる。監視環境内に存在する人数が増えた場合、存在領域の信

頼度が全般に低下し、クラス 2 及び 3 の領域が増加する。この場合でもクラス 1~3 の分類は信頼度を表し、最終的に提示される画像を監視者が見ることにより低い信頼度の領域を監視者が確認できる。現在のシステムでは画面を 4 分割しているため監視対象が 4 つ以上であれば分割数を増やすかウィンドウを対象の数だけ作る必要がある。また位置推定画面のみでは対象が人物か何か分からないが、画像を確認することにより監視者はその物体が何かを確認することができる。しかし、現状のインターフェースでは位置推定画面と対象方向の画像の対応が分かりにくいという問題がある。このため本システムを実用化するにはインターフェースの改良を必要とする。また、ネットワークを介してサーバ側とクライアント側に処理を分散させることで実時間での提示画像の更新を行うことができ、ビル全体などの多くのセンサを必要とする環境を監視する場合、有効なシステムであると考ええる。

6. む す び

本研究では、全方位画像センサによるネットワークを介した遠隔監視システムを開発した。開発したシステムは 3 対 1 のサーバ/クライアントシステムである。さらに、ネットワークを利用して多対多のシステムへの拡張も考えたネットワークプロトコルと分散処理を行うことでさらに大規模なシステムを考えた場合にも、実時間での提示画像の更新・監視を行うことができると考える。サーバ側においては、全方位画像において背景差分を利用し、かつ背景の更新を行うことで、蛍光灯のフリッカーなどによらずに複数の注目物体を同時にかつロバストに検出できた。またクライアント側においては、サーバ側から送られた注目物体の方位情報を用いて物体の存在領域を推定し、かつ物体方向の平面透視投影画像を提示することにより注目物体の確認が可能であった。また、受信したい全方位動画を監視者が手動で選択しているため、3 視点から自由に監視を行うことができる。したがって、ビルなどの多くのセンサを必要とする場合、3 台の全方位画像センサが同一直線上に並んでいなく、各センサから他の 2 センサが写っているという前提を満たしながら、3 台の全方位画像センサの組み合わせを変化させることで、複数の部屋の監視に適用可能であり、一極集中型の監視を行うことができる。

さらに有効な遠隔監視システムを構築するために、今後の課題としては、以下の 6 つが挙げられる。

- (1) 提示画像の解像度の向上。
- (2) 使用するネットワーク帯域幅の削減。
- (3) 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。
- (4) 多数の全方位画像センサ(サーバ)を用いた場合のシステムの拡張。
- (5) フレーム間での存在領域の対応付けによる同一人物の追跡。
- (6) 提示画面などインターフェースの改良。

(1)の問題に関しては、HyperOmni Vision を構成するビデオカメラに高精細なものを使用するか、その他の高精細な全方位画像センサを用いることにより解像度の向上が見込まれる。また、(2)の問題に関しては、30fps で全方位画像を伝送するための必要な帯域幅が大きく、ネットワークによっては使用する帯域幅を削減しなければならない。このような問題に対しては MPEG のようなさらに圧縮率の高い伝送方法を用いた帯域幅の削減が有効であると考ええる。(3)の問題に関しては、動画をマルチキャストすることによりシステムの拡張が可能であると考ええる。また、(4)の問題に関しては、今回のシステムでは一つのクライアントは 3 つのサーバにのみ接続することを想定しているが、監視範囲が広がるとクライアントシステムを拡張し 4 つ以上のサーバに同時にアクセスする必要が出てくる。その場合は監視範囲を分割し、それぞれの範囲で最適な 3 つのサーバの情報を選ぶことにより物体の位置を推定可能であると考ええる。本クライアントシステムにおける位置推定処理は画像処理を伴わないため処理が軽く、一台のクライアントで 4 つ以上のサーバの情報を処理することは困難でない。また、方位角情報は全方位画像と比較して必要とするネットワーク帯域ははるかに小さく、4 つ以上のサーバの情報を伝送する場合でも必要とする帯域の増大は少ない。(5)の問題に関しては、フレーム間で物体の存在領域が未対応であり、受信する全方位画像における検出された物体の画素数を基準に透視投影画像の提示位置を決定している。その結果、物体の画素数により表示位置が変更される場合が多く、監視者にとってわかりづらい。そのため、フレーム間で存在領域を対応付け、同じ表示位置に同一物体の画像を提示し続ける必要がある。(6)の問題に関しては提示画像を改良することにより 5.3 節で述べたインターフェースの問題を解決できる。

文 献

- [1] S. Rougeaux and Y. Kuniyoshi: "Velocity and disparity cue for robust real-time binocular tracking," Proc. of CVPR'97, pp.1-6, 1997.
- [2] 村瀬, 和田, 松山: "回転を伴うカメラによる移動物体の検出," 画像の認識・理解シンポジウム (MIRU'98), Vol.1, pp.425-430, 1998.
- [3] D. Murray and A. Basu: "Motion tracking with an active camera," Trans. Pattern Analysis and Machine Intell., Vol.16, No.5, pp.449-459, 1994.
- [4] J. Denzler and D. W. R. Paulus: "Active Motion Detection and Object Tracking," Proc. ICIP 1994, Vol.3, pp.635-639, 1994.
- [5] J. Batista, P. Peixoto, and H. Araujo: "Real-Time Active Visual Surveillance by Integrating Peripheral Motion Detection with Foveated Tracking," In Workshop on Visual Surveillance (in conjunction with ICCV98), pp.18-25, 1998.
- [6] A. J. Wavering and R. Lumia: "Predictive Visual Tracking," Proc. of SPIE Intelligent Robots and Computer Vision, Vol.2056, pp.86-97, 1993.
- [7] 中澤, 日浦, 加藤, 井口: "分散視覚エージェントを用いた複数人物追跡システム," 情処論文誌, Vol.42, No.11, pp.2699-2710, 2001.
- [8] H. Tsutsui, J. Miura, and Y. Shirai: "Optical Flow-Based Person Tracking by Multiple Cameras," Proc. 2001 Int. Conf. on Multisensor Fusion and Integration in Intelligent Systems, pp.91-96, 2001.
- [9] A. Utsumi, H. Mori, J. Ohya, and M. Yachida: "Multiple-view-based tracking of multiple humans," Proc. Int. Conf. on Pattern Recognition, pp. 597-601, 1998.
- [10] V. N. Peri, S. K. Nayar: "Generation of Perspective and Panoramic Video from Omnidirectional Video," Proc. of 1997 Image Understanding Workshop, pp.243-245, 1997.
- [11] A. Elgammal, R. Duraiswami, D. Harwood, L. S. Davis: "Background and Foreground Modeling Using Nonparametric Kernel Density for Visual Surveillance," Proc. of the IEEE, Vol.90, No.7, pp.1151-1163, 2002.
- [12] J. Illmann, B. Kluge, E. Prassler: "People Tracking Using Omnidirectional Vision And Range Information," Proc. of SPIE Sensor Fusion and Decentralized Control in Robotic Systems IV, Vol.4571, pp.128-138, 2001.
- [13] 三木, 山澤, 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," 信学技報, PRMU98-177, 1999.
- [14] 十河, 石黒, M.M. 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.
- [15] 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.
- [16] 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.
- [17] A.Ogawa, K.Kobayashi, O.Nakamura, and J.Murai: "Design and Implementation of DV Stream Over Internet", IWS Internet Workshop, 1999.
- [18] 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.399-406, 1998.

(平成 xx 年 xx 月 xx 日受付)



森田 真司

2001 年神戸大・工・電気電子工卒。2003 年奈良先端大・情報科学研究科・博士前期課程了。同年株式会社日立製作所入社。修士(工学)。



山澤 一誠 (正員)

1992 年阪大・基礎工・情報工卒。1996 年同大大学院博士後期課程中退。同年奈良先端大・情報科学研究科助手。2002 年奈良先端大・情報科学研究科助教授。コンピュータビジョン, 複合現実感の研究に従事。1997 年信学会論文賞受賞。2002 年信学会学術奨励賞受賞。博士(工学)。情報処理学会, 日本VR学会, IEEE 各会員。



寺沢 征彦

1999 年阪大・工卒。2001 年奈良先端大・情報科学研究科・博士前期課程了。同年住友電気工業株式会社入社。現在住友電工ネットワークス。修士(工学)。



横矢 直和 (正員)

1974 年阪大・基礎工・情報工卒。1979 年同大大学院博士後期課程了。同年電子技術総合研究所入所。以来, 画像処理ソフトウェア, 画像データベース, コンピュータビジョンの研究に従事。1986~87 年マツギル大・知能機械研究センター客員教授。1992 年奈良先端大・情報科学センター教授。現在, 同大情報科学研究科教授。1990 年情報処理学会論文賞受賞。工博。情報処理学会, 日本VR学会, 人工知能学会, 日本認知科学会, 映像情報メディア学会, IEEE 各会員。

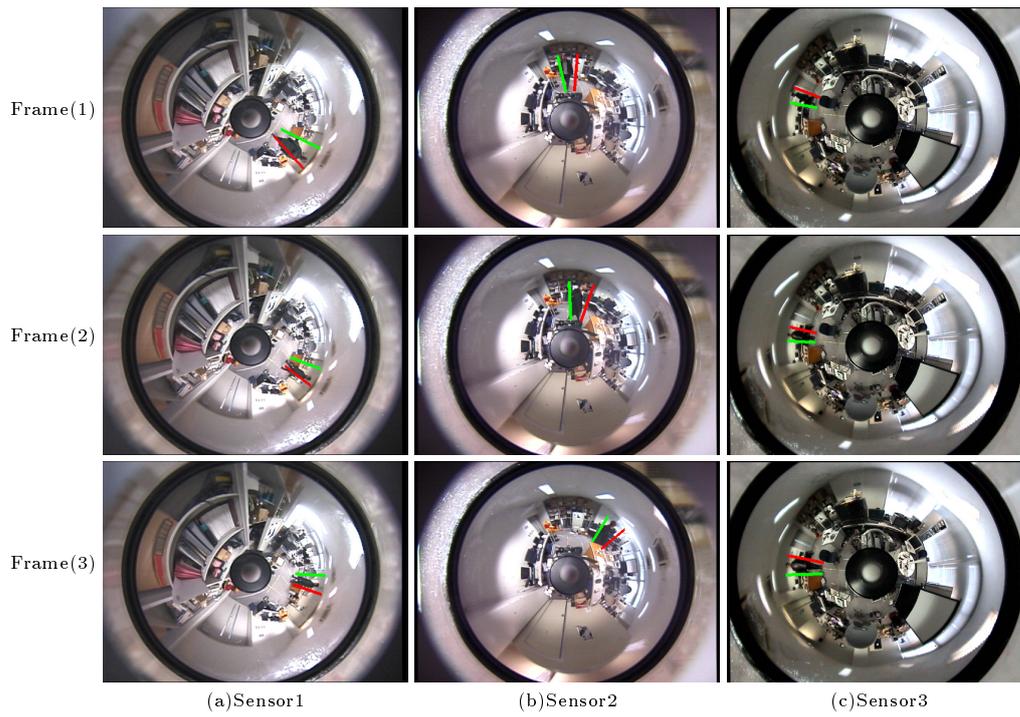


図 9 実験 1 でのサーバにおける入力画像と注目物体の検出
 Fig.9 Input images and detected objects in the experiment1



図 10 実験 1 でのクライアントにおける提示画像
 Fig.10 Displayed images in the experiment1

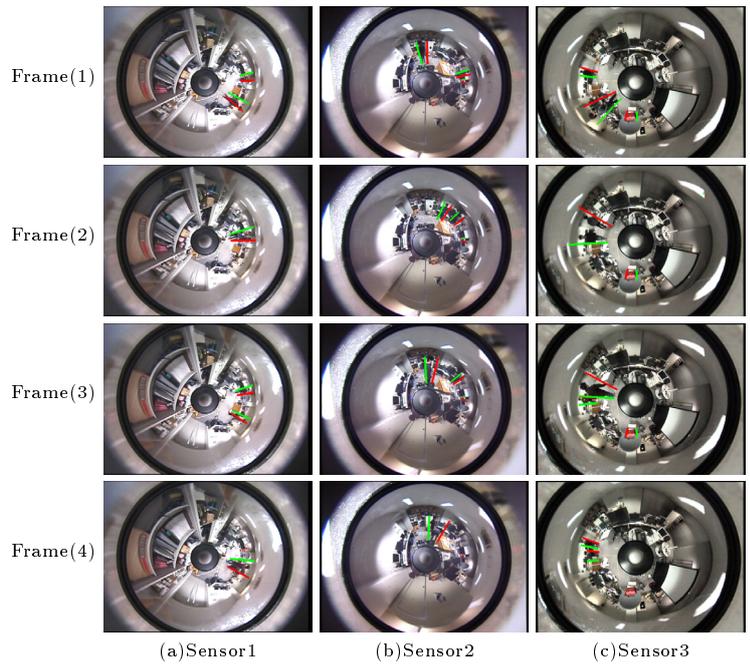


図 11 実験 2 でのサーバにおける入力画像と注目物体の検出
 Fig. 11 Input images and detected object in the experiment2



図 12 実験 2 でのクライアントにおける提示画像 (図 7 を分割)
 Fig. 12 Displayed image in the experiment2 (partitioned Fig.7)

Abstract Remote surveillance is required to always acquire wide environment, and to detect and track objects in the real time. In most conventional remote surveillance systems, active cameras and omnidirectional video cameras are generally used in order to acquire an image of the remote site. The former often suffers from the time delay of changing the user's viewing direction and updating the displayed image. The latter is not implemented over network so that it cannot be applied to remote surveillance using a large number of distributed sensors. This paper describes new remote surveillance system which transfers multiple omni-directional videos via network. The surveillance system is based on a server/client model: the server computers, each of which is connected to an omnidirectional video camera, are placed in the surveillance area and the client computer is placed in a user side. The servers detect moving objects. The client estimates object positions and presents object-centered perspective images to the user. Moreover, by transmitting detected information and video via network, the remote surveillance in the client side is realized.

Key words Omnidirectional video camera HyperOmni Vision , Network , Digital Video Transport System , Remote surveillance , Motion tracking