

全方位カメラを用いた監視追跡システム

佐藤 洋介[†] 米田 裕也[‡] 橋本 浩二[†] 柴田 義孝[†]
岩手県立大学ソフトウェア情報学研究科[†] 株式会社アイソニック[‡]

人間の建物への侵入や室内での行動を監視するシステムにおいて、主に人物を実時間で追跡する事や、広範囲で鮮明な映像を撮影する事などが求められている。また、広範囲且つ人物の拡大顔画像を撮影することで、一般に利用されている監視カメラの台数や記録メディアの削減につながると考えられる。これまで、カメラの画像から人間の顔を検出し検出範囲を拡大するシステムや、Pan/Tilt/Zoom カメラ(PTZ カメラ)を用いて追跡するシステムの研究がなされてきた。一方、全方位映像カメラを用いた自律ロボット制御や、複数の移動物体を検出するシステム等の研究が多くなされている。そこで、全方位映像カメラを用いて広範囲に人間の顔を検出し、その検出位置情報に基づき、PTZ カメラを制御して顔を適切な位置・大きさで撮影することで、広範囲に人物の鮮明な拡大顔画像を撮影することが可能となる。本稿では人物の顔を特定して抽出し、自動追跡可能な新しい監視カメラシステムについて提案を行う。

Face Image Tracking System for Surveillance using an Omni-directional Camera

Yousuke Sato[†], Yuya Maita[‡], Koji Hashimoto[†], Yoshitaka Shibata[†]

[†]Graduate School of Software and Information Science, Iwate Prefectural
University

[‡]AI Sonic Corporation.

In the surveillance system which observes the behavior of human invasion to building or indoor, it is required not only to capture the high quality and wide area images, but also to automatically track to the specific suspicious person in realtime to reduce the number of the required surveillance cameras. So far, there are many researches which have investigated the facial motion detection and extraction and the area expansion, and automatic motion tracking using Pan/Tilt/Zoom camera (PTZ camera). On the other hand, the other systems which perform autonomous robotics control and multiple moving objects using omni-directional cameras and PTZ cameras has been also investigated. Using the suggested system, the location of the target human face in wide area with 360 degree surround can be detected and the moving face can be tracked and captured his high quality images in realtime.

1.はじめに

最近、国内外において、殺人、強盗、ひき逃げなどの凶悪犯罪が急増し、住民の生活が脅かされており、安心や安全を確保するための監視システムが重要になってきた。特に、人間の建物への侵入や屋内での行動をリアルタイムに追跡し、広範囲で高品質な映像の記録や、行動の分析を行うことが求められている。一般に利用されている監視カメラを使用し、広範囲の監視をするためには、カメラの台数や記録メディアの数が多く必要となり、監視システムが大規模で複雑なものとなる。これは、カメラの撮影範囲が狭いことや、複数のカメラ映像から人物の特定や、

追従することが困難なためである。これらの問題を解決すべく、撮影画像中の人間の顔を検出し、その検出範囲の顔画像を切り出し、拡大表示を可能とするシステム[10]や、Pan/Tilt/Zoom 可能な制御カメラ(以下 PTZ カメラ)を用いて人物を追跡するシステム[4]の研究がなされてきたが、一般的な単一方位型カメラを利用した場合、一台で検出可能な範囲が狭く、広い範囲を網羅することができない。

一方、全方位映像カメラを用いて広い範囲の画像を取得することにより、ロボットを自律的に制御する研究や、複数の移動物体を同時に検出するシステムなど、全方位映像カメラをセンサとして用いる研究が多くなされている。全方位映像カメラは1つ



図 1 Telegnosis System クライアント画面

の CCD で 360 度の広い範囲を撮影することが可能なため、周囲の認識や監視を行うためのセンサとして非常に有効である。しかしながら、撮影される全方位映像は環状映像として撮影されるため、有効画素数が CCD の最大画素数の 30%~40%程度になる。そのため、撮影された画像をそのまま監視に利用することは十分な解像度を確保できないという点で問題がある。

筆者らはこれまで、全方位映像カメラと PTZ カメラを組み合わせ利用し、PTZ カメラを制御するシステム(Telegnosis System)[3]の開発を行ってきた(図 1)。このシステムでは、全方位映像カメラと PTZ カメラを利用し、サーバでキャプチャされた映像を IP ベースで Video Stream として送信する。これをクライアントで受信し、ディスプレイに表示する。全方位映像は、サーバもしくはクライアントで展開処理を行い、全方位パノラマ映像上の任意の位置を指定することで、遠隔から PTZ カメラの Pan/Tilt/Zoom 操作を行うことを容易にした。このシステムを、ユーザの要求通りの視野を提供可能な遠隔テレビ会議や遠隔ヘルスケア実習の支援システムとして使用し、機能追加と実用性の検証を行ってきた。

そこで本稿では、前述のシステムの機能を拡張した顔追跡を行う新しい監視システムを提案する。このシステムでは全方位映像カメラをイメージベースのセンサとして用いることで、広範囲に人間の顔の位置を検出する。検出された位置情報に基づき、PTZ カメラを制御することにより、顔を適切な位置や大きさで撮影し、ユーザに広範囲かつ高精細な顔の拡大画像を提供する。また、顔が検出された際に発生する検出イベントと Video Stream の録画を同期させ、変化のない場面の記録を削除することで、記録時間を減少させ、低圧縮率の高品位映像の記録を図る。これらの機能を実現するために、前述の Telegnosis System に人物追跡アルゴリズムを実装した Face Tracking Module を機能を拡張する。本稿では、システムの構成とアーキテクチャ、想定している顔検出手法や各種機能について述べる。

2. 関連研究

人物追跡を行う監視カメラシステムには、PTZ カメラを利用したもの[4][5]や、固定カメラ、全方位映像カメラ[6]、または、それらを複数個利用したもの[7]がある。PTZ カメラを利用した[5]のシステムでは、1 台の PTZ カメラを入口が撮影できるよう部屋の壁沿いに 1 台設置する。このカメラによって取得した画像からフレーム間差分法と画素濃度分布法によって、人物の場所を検出し、人物の場所の推定を行い、室内での移動軌跡を自動的に算出している。また、[4]のシステムでは、固定した単一方位型 CCD カメラを利用して人物の顔の位置と大きさを検出し、その領域に PTZ カメラをパン/チルト/ズームし顔の拡大画像を効果的に撮影する。このシステムでは登録された背景画像と、新しく撮影された画像の各画素の平均をとし、背景画像を更新する。この手法によって、フレーム間差分によって起こる「静止している人物のフェードアウト」を防ぐことが可能となる。

一方、全方位映像カメラを利用したシステムとしては[6]のシステムのように、1 台のカメラで複数の移動物体を検出することが可能なシステムがある。このシステムでは撮影した画像から登録した背景画像を除去し、抽出された人物の領域のピクセルカラーとその位置情報に ID を付加して識別し、カメラの視界に映っている間、位置と色情報を記録する。また、[7]のシステムでは、複数個の全方位映像カメラを利用し、これらのサーバで検出した人物の方向情報と映像を送信する。クライアントでは受信した方向情報を基に、全方位映像からその範囲を切り出して表示し、特定の人物の多方向からの映像を監視することが可能である。

3. システム構成

前述のとおり監視映像システムの多くは、単方位カメラや PTZ カメラ、全方位映像カメラをそれぞれ単独で用いたものである。

一方、本システムでは図 2 に示すように、複数のサーバと、全方位映像カメラと PTZ カメラの組み合わせによるカメラユニット、監視を行うクライアントから構成される。全方位映像カメラは PAL レンズを装着した DV / HDV または C-MOUNT USB カメラを IEEE1394 または USB インタフェースでサーバに接続する。また、PTZ カメラは NTSC 出力されるアナログ信号を A/D コンバータで IEEE1394 に変換後にサーバに接続する。サーバでは、全方位映像カメラから取得される画像中の人物の顔の位置とサイズを検出し、検出した位置とサイズに応じて PTZ カメラに対し Ethernet 経由で制御コマンドを送信し Pan/Tilt/Zoom 操作を行う。PTZ カメラで撮影された

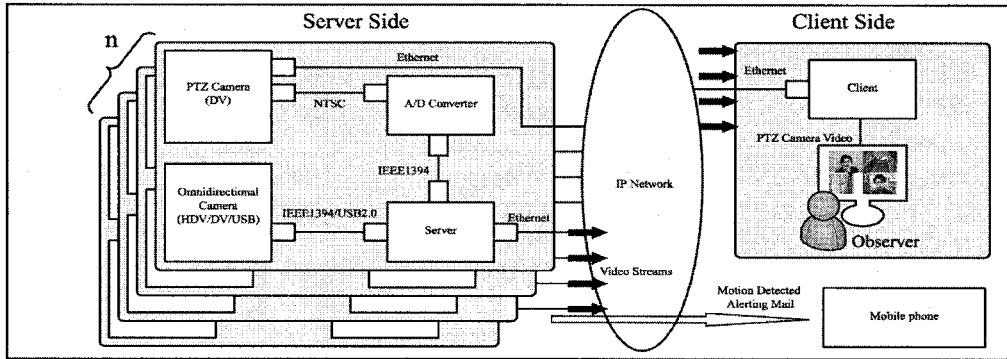


図 2 監視カメラシステム構成図

映像は **Video Stream** としてネットワークを利用してクライアントに送信される。また、人物の検出イベントの発生に同期して録画を行う。クライアントでは、受信した **Video Stream** を再生表示する。また、サーバに対して、蓄積されたビデオを送信するための **Video-File Stream** を遠隔操作で生成させ、受信表示を行う事(**Replay**)も可能である。このとき、サーバとカメラユニットを n 台設置することで、クライアント側では多地点からの視点映像を一度に監視することが可能となる。再生表示はすべての映像を分割ウィンドウに表示することや、検出イベントが発生した映像だけを自動的に切り替えて表示することも可能である(**Motion Detected Cam Switching**)。また、全方位映像環状画像を送信してクライアント側でパノラマ映像に展開して監視を行うことも可能である。これにより、たとえ検出に失敗した場合でも、手動で制御することが可能となる。その他に、検出イベントの発生に応じて通知メッセージをクライアントや携帯端末へ通知する機能も実現可能となる。

提案システムにおいて、これまで筆者らが開発してきた次節で述べる 2 つのミドルウェアをライブラリとして利用する。まず、全方位映像や PTZ カメラ映像の IP 転送やファイルへの保存(録画)、ストリーム生成の遠隔操作などの機能のために **MidField System**[2]を、全方位映像のパノラマ展開処理のために全方位ミドルウェア[1]をそれぞれライブラリとして使用する。

4. アーキテクチャ

4.1. Telegnosis System

本システムは図 3 に示す通り、**Telegnosis System** 層としてアプリケーション層に 2 階層・3 プレーンで構成される。**Interface** 層では、ユーザの操作情報(**Event Process plane**) やマルチメディアストリームのディスプレイ出力(**Stream Control plane**)、システムステータ

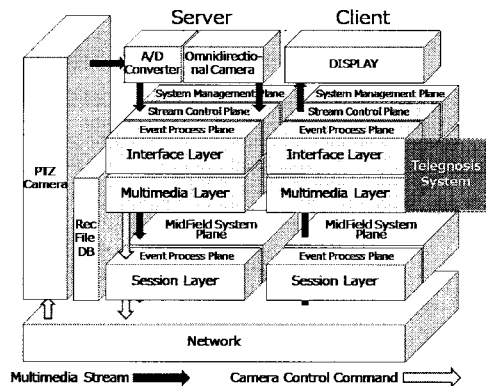


図 3 システムアーキテクチャ

スの表示 (**System Management plane**) を行う。**Multimedia** 層では全方位映像のパノラマ展開処理、人物検出、ストリームの生成と制御(**Stream Control plane**)、PTZ カメラのカメラワークの計算と制御コマンド、検出通知の生成、**Motion Detected Cam Switching** のための画面表示切り替え(**Event Process plane**)、システムステータスチェック、記録ビデオの管理 (**System Management plane**) を行う。**Session** 層では **Event Process Plane** にて PTZ カメラ制御コマンドをソニー考案の **VISCA** プロトコルに基づき **VISCA** コマンドに変換し、**HTTP** リクエストメッセージを PTZ カメラに送信する。**MidField System** を **MidField System Plane** に位置づける。

図中の **Server** ブロックでは、全方位映像を入力とし人物追跡を行い、顔の位置(角度)と距離(大きさ)に基づきカメラ制御コマンドを送信する。PTZ カメラの映像は A/D Converter 経由で取得し、Client へ送信する。

4.2. MidField System[2]

MidField System は、コンピュータネットワークを

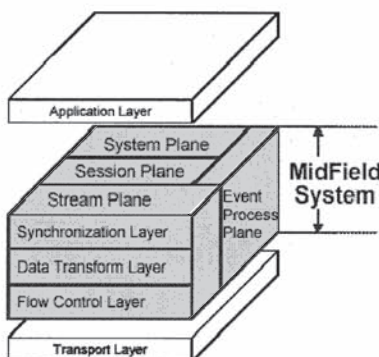


図 4 MidField System アーキテクチャ

利用した多地点相互通信に必要となる各種機能をアプリケーションへ提供するマルチメディア相互通信ミドルウェアであり、図 4 に示すようにトランスポート層とアプリケーション層の間に 3 階層・4 プレーンで構成されている。

MidField System は、利用者のネットワーク帯域などの通信環境に適したオーディオ・ビデオフォーマットを用いて多地点相互通信を実現するために、ネットワーク上の適切なコンピュータ(トランスコーダー)上でフォーマット変換機能(トランスコーディング機能)を動作させる仕組みを実現している。WMV(Windows Media Video)フォーマットを利用して、数 100kbps～10 数 mbps の範囲におけるトランスコーディング処理に対応する一方、DV (Digital Video) / HDV (High-definition Digital Video)フォーマットを用いた映像伝送にも対応している。また、リアルタイムオーディオ・ビデオ通信に加えて、ビデオファイルとして録画や再生、再生内容を送信することも可能である。

また、最新バージョンの MidField System1.30 は遠隔操作機能を有しており、操作コマンドを送信することで Multimedia Stream の生成や一時停止といった制御が遠隔から実行可能である。[11]

4.3. 全方位映像ミドルウェア[1]

本システムでは Telegnosis System 層の機能を実現するために、全方位映像ミドルウェア[1]を利用し、全方位環状映像のパノラマ展開処理を行う。全方位映像カメラは高精細化・小型化しており、カメラの種類を利用目的に応じて選択でき、全方位映像ミドルウェアもそれに応じた動作が可能となっている。しかしながら、全方位環状映像のパノラマ展開処理には各フレームに処理を実行するため、サーバもしくはクライアントの CPU 負荷が非常に大きくなるという問題点がある。通常の全方位環状映像ストリームをパノラマ展開処理し送信した場合は、フレーム

レートの低下を招く。特に HDV フォーマットの場合、MPEG-2 に伴うビデオフォーマットの符号化のための処理量の増加のため、遅延の要因となる。このため、全方位映像ミドルウェアでは、CPU 能力に応じたフレームレート制御などの QoS 制御方式を検討している。

5. 自動追跡監視手法

5.1. 顔抽出手法

動画像から動物体を抽出する手法として、背景差分法やフレーム間差分法、画素濃度分布を利用した手法[11]があげられる。背景差分法は、予め設定した背景画像と取得画像の差分によって動物体の領域を抽出する。この方法は変化の少ない環境であれば正確な抽出が可能であるが、日光による明るさの変化などの時間的変化がある場合は、背景画像と取得画像の差が大きくなり動物体の抽出が難しくなる。長時間かつ変化の激しい環境下では背景モデルの更新が必要となる。

また、画像濃度分布法は過去の数フレームから各画素における濃度値分布を計算し、各画素の 255 段階の濃度値を蓄積する。動物体が存在しない画素は分布の最頻値と平均値の差が小さく、単峰性になる。逆に動物体が存在する画素は両者の差が大きいため多峰性になることを利用した抽出手法である。この手法は、分布を計算することにより背景差分法に比べ高い精度の領域抽出が可能であるが、画素毎に全時系列の濃度値分布を考慮するためにリアルタイム処理には適さないとされている。

一方、フレーム間差分法は、時刻(t)の入力フレームと時刻(t-n)の入力画像との差分をとることにより、動物体を抽出する抽出する手法である。フレーム間差分法は対象に動きがない場合、差分が表れないなど不安定な部分もあるが、常に背景画像の更新を行っているため、環境の変化に強いといえる。

本システムでの人物抽出は背景画像更新を行う背景差分法の利用を検討している。抽出を行う全方位映像は位置推定の容易さや、現在のシステムの PTZ カメラと全方位パノラマ画像のマッピングを利用するため環状画像から展開処理を行い、その後展開画像を利用して抽出を行う。ただし、全方位展開処理と人物抽出処理を同時に行うことは非常に多くの処理量を必要とするので、環状画像のまま抽出・追跡を行うことも検討している。

5.2. Face Tracking

図 5 に示すように、2 値化された抽出画像には人物がシルエットとして映る。このシルエットから顔の

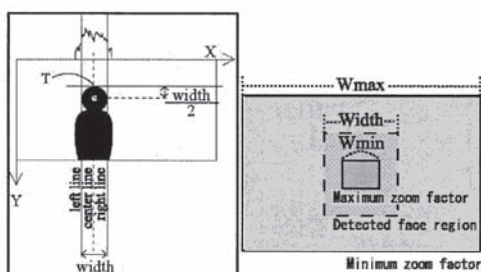


図 5 左:顔領域推定 右:ズーム倍率計算

位置を以下の手順で推定する。まず、縦列の画素の重みが高い範囲を顔の幅とし、左右端線を決定する。次に左右端線の二等分線を中心線とし、この中心線は顔の中心を通るものとする。次にシルエットのうち中心線上で最も上に位置する画素の座標を(Tx, Ty)とし、そこから顔の幅の半分下の位置を顔の中心(Tx, Ty+width/2)とする。顔の中心は全方位パノラマ映像上に決定される。この座標を以下の計算式によって PTZ カメラの座標にマッピングする。

$$PTZx = \frac{2\pi(Px - Ex)}{W} \times \frac{360}{2\pi}$$

$$(0 \leq Ex \leq W/2)$$

$$PTZy = \frac{H - Ey - Py}{H} \times \text{omniELRange}$$

$$\text{omniELRange} = \text{omniTop} + |\text{omniBottom}|$$

$$Ey = \frac{H \times |\text{omniBottom}|}{\text{omniELRange}}$$

ここで、W, H を全方位パノラマ画像の幅と高さ、(Tx, Ty+width/2) を(Px, Py)とする。このとき、(Px, Py)に対応する PTZ カメラの視角度度(PTZx, PTZy)を求める。また、Ex は全方位カメラと PTZ カメラの設置状況に応じた縦軸に対する水平回転のずれであり、設置後に調整される。Ey は全方位映像の仰角 0 度の線と、パノラマイメージの下端の間の長さを示す。omniTop, omniBottom は、全方位カメラの可視仰角の上端と下端を表し、可視仰角の範囲 omniELRange を求める。たとえば、本システムにおいては、H = 200, W = 800, omniTop = 40°, omniBottom = -20°, Ey = 200/3 ≈ 66.667 となる。

次に、ズーム倍率 Z を以下の式で求める。最もズーム倍率が低い際に PTZ カメラで撮影される範囲に対応する全方位パノラマ画像上の範囲の横幅のピクセル数を Wmax。反対に最もズーム倍率が高い際の横幅のピクセル数を Wmin として、あらかじめ測定し定義する。求められた顔の幅のピクセル数を Width とする。Zmax, Zmin はそれぞれ PTZ カメラのズーム

倍率の最大値と最小値を示す。

$$Z = \left(1 - \frac{\text{Width}}{W_{\max} - W_{\min}}\right) \times (Z_{\max} - Z_{\min})$$

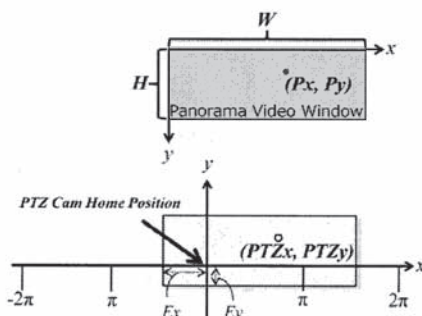


図 6 パノラマ映像と PTZ カメラ座標の対応

5.3. Motion Detected Cam Switching

Motion Detected Cam Switching は、移動物体が検出されたカメラの映像のみを切り替えてディスプレイに表示する機能である。これは、検出イベントが発生したサーバから、検出イベント通知を受信したクライアントが、Window Manager Module によって自動的にウィンドウの表示と非表示を制御する。このウィンドウ制御に連携して、サーバに対して遠隔操作コマンドを送信し、非表示になった Video Stream の送信を一時停止する。これにより、不必要なネットワーク転送量と CPU 処理量を削減できる。

この機能は、一定間隔でカメラユニットが設置された通路を移動する人間を、シームレスに監視する際に用いることで、効果的な監視を可能とする。

6.実装

Omni-Cam	SONY HDR-HC1 / COREFRONT IUC-130CN(USB)
PTZ Cam	SONY SNC-RZ30N Pan:+170° ~-170° Tint:-25° ~+90° Swing Speed:1.5 sec
Media Converter	CANOPUS ADVC-100
Computer Server/Client	CPU:Pentium4 3.4GHz Memory:1.5GB OS:WindowsXPProSP2
Network	100Base-T
Image Transport System	MidField System Ver.1.30

表 1 利用機器表

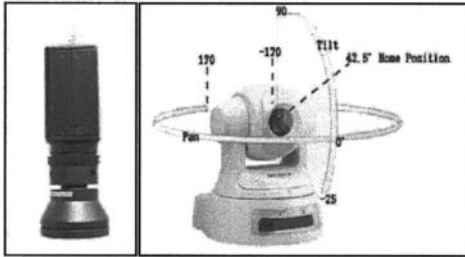


図 7 左:C-Mount 全方位カメラ 右:SNC-RZ30N

本システムで用いる機器を表[1]に示す。監視カメラシステムはサーバ側の **Telegnosis System** に **Face Tracking Module** を **Directshow** フィルターとして組み込むことで、検出した顔の範囲と位置情報を得る。**Face Tracking Module** は全方位画像から一定間隔で顔の座標と幅を算出する。使用言語は **Visual C++**, **Microsoft Windows Vista** 上での動作を想定する。

7. まとめ

本稿では、全方位映像カメラと PTZ カメラを利用した人物追跡遠隔監視システムについて述べた。本システムはネットワークを利用したサーバ/クライアントシステムである。サーバ側では全方位映像カメラにより取得した画像からリアルタイムに人物の顔の位置を検出し PTZ カメラで追跡を行い、PTZ カメラによって撮影された映像をクライアントに送信する。また、設定次第で全方位パノラマ展開映像をクライアントに送信することで、全方位パノラマ映像を頼りにユーザが PTZ カメラを制御することも可能である。また、サーバとカメラを複数設置することによって多地点を監視することも可能となる。

今後の展開として、システムの設置状況や解像度、PTZ カメラの動作速度を考慮の上、提案するシステムに最適な人物追跡手法の検討と、機能の実装、性能評価を行う。特に、本稿の 5.2 で述べた顔検出手法は不完全な手法であるため、より正確でロバストな顔検出手法の検討と実現を行う。

また、現在検討中の監視カメラシステムでは全方位映像カメラと PTZ カメラが 1 対 1 で組み合わせられているため、複数地点に設置する場合は全方位映像カメラと PTZ カメラが同じ台数必要となる。将来的に全方位映像カメラと PTZ カメラの台数に偏りがある場合でも利用可能とするために、複数の全方位映像カメラを設置し、それらの映像をつなぎ合わせることでシームレスな映像を生成する技術や、1 台の全方位映像カメラと複数の PTZ カメラを利用し、あらゆる角度からの画像を撮影するといった技術の導入も期待できる。

参考文献

- [1] 米田祐也, 橋本浩二, 柴田義孝: 全方位映像通信のためのミドルウェアの研究, 情報処理学会第 67 回全国大会, pp595-596, (2005)
- [2] 橋本浩二, 柴田義孝: 利用者環境を考慮した相互通信のためのミドルウェア, 情報処理学会論文誌, Vol.46, No.2, pp.403-417, (2005)
- [3] 佐藤洋介, 米田裕也, 橋本浩二, 柴田義孝, “全方位カメラと制御カメラを利用した遠隔カメラワークシステム” 情報処理学会研究報告マルチメディア通信と分散処理 (DPS 研究会), No. 130, pp. 429-434, 2007 年 3 月
- [4] T. Funahashi, M. Tominaga, T. Fujiwara, and H. Koshimizu, "Hierarchical Face Tracking by Using PTZ Camera", Proc. The 6th International Conference on Automatic Face and Gesture Recognition (FGR 2004), pp.427-432, Seoul, Korea, (May.2004)
- [5] 大島直也, 齊藤剛史, 小西亮介: 回転可動カメラによるリアルタイム人物位置推定システム, IEEJ Trans. EIS, Vol.126, No.8 (2006)
- [6] H. Liu, W. Pi, H. Zha, "Motion Detection for Multiple Moving Targets by Using an Omnidirectional Camera", Proc. IEEE Int. Conf. on Robotics, Intelligent Systems and Signal Processing, Changsha, China, Sep., 2003
- [7] Shinji Morita, Kazumasa Yamazawa, Naokazu Yokoya: Networked Video Surveillance Using Multiple Omnidirectional Cameras, Proc. IEEE Int. Symposium on Computational Intelligence in Robotics and Automation, Kobe, Japan, July., 2003
- [8] Yi Yao, Besma Abidi, Mongi Abidi: Fusion of Omnidirectional and PTZ Cameras for Accurate Cooperative Tracking, Proc. IEEE Int. Conf. on Video and Signal Based Surveillance(AVSS'06), 2006
- [9] Tang yipping, Ye Youngjie, Jin Hairmin: Development of Multi-object Tracking System Based on Omni-Directional, Proc. Of the 16th International Conference on Artificial Reality and Telexistence Workshops(ICAT'06), 2006
- [10] 馬場 功淳, 大橋 健, 乃万 司, 松尾 英明, 江島 俊朗: HeadFinder: フレーム間差分をベースにした人物追跡, 画像センシングシンポジウム 2000, pp329-334, 2000
- [11] MidField System Ver. 1.30 : <http://www.sb.soft.iwate-pu.ac.jp/~hashi/MidField-System/>, 橋本浩二(岩手県立大学 ソフトウェア情報学部) (2007/3)