広域監視のための両眼監視カメラシステム

A Binocular Camera System for Wide Area Surveillance

谷 宇章[†], 正会員 佐藤 誠[†], 張 曉林[†]
Yuzhang Gu[†], Makoto Sato[†] and Xiaolin Zhang[†]

Abstract An advanced binocular active camera system for wide area surveillance is proposed. The system is equipped with high-resolution actuators, wide angle lens cameras, and telescope lens cameras, which enables it to smoothly track a pedestrian from a far distance and to continuously shoot a stable video in detail. Depth information is available because it is a stereo vision system. Differences in depth between the target and the background can be used to distinguish the one from the other. To achieve robust tracking when the background is complicated, a 3D tracking image processing algorithm based on binocular disparity is proposed to separate the target being tracked from the complicated background, even though the target has changed its appearance according to its rotation. Moreover, to avoid the difficulty of looking for corresponding points between a pair of stereo images, the baseline distance between the two eyes of the proposed system is kept very short. However, this leads to poor vergence control, a problem that is rectified by applying a mathematics model based on the neural pathways of human binocular eye movement, thus enabling stable binocular tracking.

キーワード：広域，3次元追跡，両眼，監視カメラ，SIFT

1. ま え が き

近年，安心・安全な社会の構築への関心が高まっている。そのための手段の一つとして，公共の安全を確保するための監視カメラシステムの設置が有効であると認識されており，町中に設置された監視カメラの数は日々増えている。しかし，現在実用化されている広域監視カメラシステムの多くは，画角の広い固定式カメラを用いており，監視範囲が広い反面，監視対象である人物や事件の詳しい映像情報を得ることができないという問題点がある。

そこで，その問題点を克服すべく，PTZ（パン・チルト・ズーム）カメラを利用する監視カメラシステムが開発されている。PTZカメラ監視システムでは，監視領域中の詳しく見たい対象物にカメラの視線を向け，ズームインして詳細な映像を得ることが可能である。しかしながら現状のPTZカメラ監視システムでは，前提条件として対象物がほぼ静止していることが必要である。なぜならば，ズームインすることでカメラの画角が非常に狭くなるため，対象物の動きによって激しい映像のブレが発生する，また対象物がカメラ視野の外に移動してしまうという恐れもある。そのた

め，PTZカメラ監視システムにおいては，対象物の動きに合わせてPTZカメラの向きを常に調整する必要がある。

先行研究として，いくつかの自動追跡PTZカメラ監視システムが研究されているが^{1)～3)}，それらは監視領域を屋内に限定しており，屋外での広域監視には不向きなシステムである。屋外での監視においては，背景や遮蔽物の問題・追跡対象物自身の移動や回転など動きに対する頑健性などが必要となり，対象を追跡監視するための画像処理が非常に困難になるという問題がある。これらの問題点のため，著者らの知る限りでは広域監視の研究において完成された自動追跡PTZカメラシステムは見当たらない。

本論文では，広域監視における両眼監視カメラシステムを提案する。提案するシステムにおいて二つのカメラを人間の両眼のように隣接して配置する。このシステムでは両眼ステレオ視により距離計測ができるため，距離の違いから追跡対象物と背景，遮蔽物とを区別することができるという特徴がある。また，両カメラを隣接配置することにより，両眼での取得画像における画像間の違いが微小になり，両眼ステレオ画像間の対応付けが非常に容易になるため，3次元情報が容易に計測できるという特徴がある。しかし，このようなカメラ配置では，カメラ間の基線距離が短くなり，奥行の計測精度が落ちるという欠点もある。そこで十分な精度を維持するために，視野角の狭い高倍率なカメラレンズを使用することが不可欠である。そのため，視野角

2009年1月21日受付，2009年7月1日再受付，2009年9月18日採録

[†] 東京工業大学 精密工学研究所

(〒 226-8503 横浜市緑区長津田町4259, TEL 045-924-5069)

[†] Precision and Intelligence Laboratory, Tokyo Institute of Technology
(4259 Nagatsuta, Midori-ku, Yokohama, 226-8503, Japan)

の狭い左右のカメラでの共通視野を維持するために、左右のカメラにおける輻輳角を協調的に制御する必要がある。

われわれは、屋外での使用を考慮し、監視距離を3m～70mと定め、屋外での使用が可能な両眼監視カメラシステムを開発した。本論文において、その性能と効果を検証した。

2. システムの構成と動作の仕組み

図1は提案する両眼監視カメラシステムの構成図である。

人間の眼球運動において、両眼の眼球は共役運動（左右の連動した動き）と輻輳運動（寄り眼方向の動き）を行い、常に同一の対象を注視することができる。この仕組みを監視カメラに応用するために図1のようなカメラシステムを作成した。本研究に利用した眼球運動に関しては4章において詳しく解説する。

まず、駆動系に関して解説する。Motor1とMotor2によって、左右カメラにおけるパン方向の共役運動および輻輳運動ができる。また、チルト方向の動きは共役運動のみ考慮すればよいという特徴により、左右のカメラは連動して動くため、Motor3だけで左右の両方のカメラを制御できる。Motor4は人間に例えると頸に相当するパン方向の回転を制御する。

次に、画像取得系に関して解説する。Camera1とCamera3は望遠レンズカメラであり、Camera2とCamera4は広角レンズカメラである。本システムでは、望遠レンズカメラと広角レンズカメラを一つのセットとして制御している。この望遠レンズカメラと広角レンズカメラを組合せることにより、近距離から遠距離までの幅広い監視距離が実現できる。距離が遠い対象を監視する場合は、望遠レンズカメラからの画像を利用し、距離が近い対象の場合は、広角レンズカメラからの画像を利用して監視を行う。Camera5は、基盤部分に固定し動かない視野角180°の超広角レンズカメラであり、領域全体を監視することができる。

図2は輻輳平面における座標系である。本システムにおける基線距離 l は0.2mである。監視距離は3m～70mを想定しているので、カメラを輻輳させて監視対象を撮影する場合、左右カメラの輻輳角の変化範囲は約0.16°～3.82°であり、画像の回転変換が少ないため、ステレオ画像を構築するための対応点の対応付けが非常に容易である。

本システムにおけるカメラの解像度とレンズの視野角を表1に示す。人間の頭部を直径0.25mの球体と想定すると、本システムにおいて70m先の人間の頭部は望遠カメラの画像上の直径57ピクセルの円形に相当する。これは顔認識において十分な画素数である。

図2における φ_l (φ_r) は視標とカメラのレンズ中心の作る直線とゼロ線 (図中のZero Line) のなす角である (以降、視標回転角と呼ぶ)。この視標回転角は、カメラの水平回転角度 E_l (E_r) と水平方向における視標の視軸から離れる角度 β_l (β_r) の和でもある。図2において時計回り方向を視標回転角の正回転方向と定める。また α を輻輳角と呼び、

図1 両眼監視カメラ機構の構成図

Picture of the binocular surveillance camera system.

図2 輻輳平面における座標系

Coordinate system in vergence plane.

表1 本システムにおけるカメラの仕様
The specification of cameras.

	Camera1,3	Camea2,4	Camera5
解像度[pixel]	1024×768	640×480	1600×1200
水平視野角[deg]	3.67	21.73	180
垂直視野角[deg]	2.75	16.38	180

$\alpha = \varphi_l - \varphi_r$ が成り立つ。

視標の奥行距離 z_T の計算式は以下の式 (1) で求めることができる。

$$z_T = \frac{l}{\tan \varphi_l - \tan \varphi_r} \quad (1)$$

本システムでは $l=0.2\text{m}$ 、 $z_T \in (3\text{m}, 70\text{m})$ となり、奥行距離 z_T は基線距離 l より遥かに長ので、 φ_l と φ_r は十分に小さく、

図3 視標距離と奥行距離計測誤差の関係
Target distance VS. Distance measurement error.

式 (1) は次の式 (2) に近似することができる。

$$z_T = \frac{l}{\varphi_l - \varphi_r} \quad (2)$$

ここで、奥行距離の計測誤差について考える。

奥行距離 z_T の微分は式 (3) で求めることができる。

$$\begin{aligned} dz_T &= \frac{\partial z_T}{\partial l} dl + \frac{\partial z_T}{\partial \varphi_l} d\varphi_l + \frac{\partial z_T}{\partial \varphi_r} d\varphi_r \\ &= \frac{dl}{\varphi_l - \varphi_r} - \frac{l d\varphi_l}{(\varphi_l - \varphi_r)^2} + \frac{l d\varphi_r}{(\varphi_l - \varphi_r)^2} \\ &= \frac{z_T}{l} dl - \frac{z_T^2}{l} d\varphi_l + \frac{z_T^2}{l} d\varphi_r \end{aligned} \quad (3)$$

上記の式 (3) からわかるように、奥行距離の計測誤差は視標の奥行距離に大きく影響される。特に視標回転角の誤差 $d\varphi_l$, $d\varphi_r$ に対し、視標の奥行距離の2乗倍の計測誤差が発生するといえる。本システムにおけるモータ回転の空間分解能は $7.5e-4^\circ$ であり、望遠レンズカメラのピクセル空間分解能は $3.6e-3^\circ$ である。また、サブピクセル精度位置推定によって、カメラの空間分解能をさらに向上させることができる。そのため、本システムにおいて用いる位置推定精度を10倍に高めることが可能である^{4) 5)}。したがって、望遠レンズカメラの空間分解能を $3.6e-4^\circ$ とみなすことができる。

論理的に計算される本システムでの視標の奥行距離とその計測精度の関係を図3に示す。このグラフの示すとおり、本システムの構成において、理論的には70m先の視標に対して約0.95mの奥行距離の分解能が限界であるといえる。また、基線距離を適度に長くすることやカメラの解像度を大きくすること、望遠レンズの視野角を狭くすることなどによって、システムの奥行距離の分解能をさらに向上させることが可能である。

本システムの動きは三段階の動きによって構成される。

まず、第一段階では超広角レンズカメラCamera5により監視範囲内の移動物体を検出する。移動物体が存在する場合、各モータを駆動しその移動物体にサッカード運動(跳躍性眼球回転運動)によって大まかに両眼カメラの視線を移動物体に向ける。

次に、第二段階として両眼カメラCamera1~Camera4を移動物体に向けた後に、両眼カメラから取り込んだ映像から顔検出を行う。顔を検出する場合、その顔を視標とし、両眼カメラはその顔に向けて再びサッカードを行う。

最後に、第三段階として両眼カメラにて視標をスムーズに追跡しながら撮影することで、ブレの少ない鮮明な顔映像を撮るとともに、視標の3次元位置を計測する。

本研究では、システムの追跡性能を評価することを目的としている。そのため、視線の複数監視対象間の切替えのタイミング、いわゆるマルチターゲット監視ストラテジーを考慮しない。前述第一段階と第二段階において複数の対象を検出した場合、カメラ視軸に最も近い対象を追跡目標とした。

本システムにおける画像処理とモータ制御は、1台のPC (Intel Xeon 3.16GHz, 4 Core×2) によって実現した。このシステムの画像処理速度は約30FPSである。

3. 画像処理アルゴリズム

図4は、本システムにおける処理の全体的な流れを示すフローチャートである。

本システムでは、図4に示すような流れで処理を行っている。本システムには、サッカードモードとスムーズ追跡モードの2種類の動作モードがある。以下、この2種類の動作モードについて詳しく解説する。

図5に示したサッカードモードのフローチャートに沿って説明する。

- (1) 超広角レンズカメラCamera5の画像から移動物体を検出する。移動物体検出アルゴリズムは、森田らが提案した動的更新背景差分法を利用している⁶⁾。複数の移動物体を検出した場合、画像中心に最も近い移動物体を視標として選択する。
- (2) 移動物体を検出した後、左右のカメラの視軸をサッカード運動によってその移動物体に向け、移動物体を広角レンズカメラCamera2とCamera4の視野中に収める。
- (3) Camera2とCamera4の画像からブーストされたカスケード分類器^{7) 8)}を用いて顔検出を行う。
- (4) Camera2とCamera4の画像から顔を検出できない場合、視標がシステムから遠く離れた位置に存在し、広角カメラに写った顔部分の画像上のピクセルサイズが小さすぎる可能性が高いので、望遠レンズカメラCamera1とCamera3の画像を用いて顔検出を行う。
- (5) Camera1とCamera3では画角が非常に狭いため、一度のサッカードでは、Camera5が検出した移動物体がCamera1とCamera3の撮像範囲に入っていない可能性がある。そのため、Camera2とCamera4の画像に対して、動的更新背景差分法を用いて移動物体を検出し、左右のカメラは再度サッカード制御を行うことで、移動物体がCamera1とCamera3の画像中に

図4 システム処理の流れ
Flowchart of system processing.

写るようになる。

- (6) 最後に、左右のカメラの画像から検出した顔は同一人物の顔であることを確認できるのであれば、その顔を追跡対象とし、システムのスムーズ追跡モードに入る。

続いて、スムーズ追跡モードについて解説する。このモードは、追跡対象の顔がサッカーモードによって照準された後の滑らかな追跡制御を行うモードである。追跡アルゴリズムについて、従来手法であるオプティカルフローとブロックマッチングを用いて実験を行った結果、背景環境の複雑さ・追跡対象自体の回転・他の物体による部分遮蔽の影響により、追跡の失敗率は高いという問題があった。この問題を解決するために、われわれは両眼視差とSIFT (Scale Invariant Feature Transform) を基にした動的テンプレート更新アルゴリズムを提案した。図6にスムーズ追跡モードのフローチャートを示す。

図6のフローチャートにおける両眼協調制御モデルについては4章で詳しく解説する。

図5 サッカーモードフローチャート
Flowchart of saccade mode.

両眼視差とは左目と右目の視点の違いによる両眼像のずれのことである。図7に注視点と視差の関係を示す。左右のカメラにおけるレンズ光学中心と注視点の3点を通る円はホロプタ円と呼ばれており、ホロプタ円上に存在する物体の両眼視差は論理上ないといえる。つまり、ホロプタ円に近い物体の両眼視差は小さく、ホロプタ円からの距離が遠くなるほど、物体の両眼視差は大きくなっていく特徴がある。

SIFTは、特徴点の検出と特徴量の記述を行うアルゴリズムである。検出した特徴点に対して、画像の回転・スケール変化・照明変化等に頑健な特徴量を記述することができ

図6 スムーズ追跡モードフローチャート
Flowchart of smooth pursuit mode.

図7 ホロプタ円と両眼視差
Horopter circle and binocular disparity.

るため、画像のマッチングや物体認識・検出に広く用いられてきている^{9)~11)}。

図5と図6のフローチャートに示すように、本システムでは、まず追跡対象となる顔を決めた後、その顔のSIFT特徴点と特徴量を抽出し、テンプレートとして登録する。追跡する時は、フレーム毎にSIFT特徴点と特徴量の抽出を行い、あらかじめ登録しておいたテンプレートとマッチングして⁹⁾追跡対象の位置を検出するという追跡方法である。しかしながら、追跡対象自体の回転や部分遮蔽などによってマッチングに失敗してしまう場合がある。そこでテンプレートを更新または追加登録を行う必要がある。その際には、背景や遮蔽物などの追跡対象に属しないSIFT特徴点を除外して登録しなければならない。

追跡対象以外のSIFT特徴点の除去方法について解説する。本システムにおいて、左右のカメラの視線は常に追跡対象を注視するように制御するので、テンプレート更新が必要な場合においても追跡対象は必ず注視点の近くに存在するといえる。そのため、追跡対象と背景や遮蔽物のシステムからの距離の差を考慮すると、追跡対象の両眼視差は背景や遮蔽物の両眼視差より小さいといえる。この特徴により、SIFT特徴点における両眼視差の大きさによって、追跡対象に属する特徴点か否かの判別ができる。

図8は両眼視差を利用して追跡対象を検出した例である。図中の上の画像が左画像であり、下の画像が右画像である。緑の点は両眼視差が2ピクセル以内でマッチングしたSIFT特徴点であり、追跡対象に属する特徴点である。赤い点は両眼視差が2ピクセル以上でマッチングしたSIFT特徴点であり、追跡対象には属さない特徴点である。この実験により、両眼視差を用いて追跡対象と背景のSIFT特徴点の判別ができることを確認した。

4. 両眼カメラの協調制御

本システムでは左右のカメラ間の基線距離が短いため、輻輳角の誤差が視標距離計測値の精度に大きな影響を与えてしまう。左右のカメラがそれぞれ独立なPID制御ループで制御されている場合は、輻輳角を安定させるためにそれぞれのモータ回転運動における伝達関数が大きな時定数を持つ必要がある。しかし、時定数を大きくするとカメラの視軸を横切る方向へ移動する物体の追跡が難しくなる。そこで本システムでは、人間の両眼眼球運動の神経経路に基づいた両眼協調制御モデル¹²⁾を用いて、輻輳角の安定化と優れた追跡性能両方を同時に実現した。

図9は人間の水平両眼眼球運動神経経路の数学モデルである。 σ , v , κ , η , ρ_1 , ρ_2 は眼球運動モデルのゲインである。

この眼球運動モデルを解析すると、以下の式(4)が得られる。

$$\begin{bmatrix} E_l(s) \\ E_r(s) \end{bmatrix} = \frac{1}{2} \frac{\sigma + \kappa + (v + \eta)s}{\sigma + \kappa + \left(\frac{1}{\rho_1 + \rho_2} + v + \eta\right)s} \begin{bmatrix} (\varphi_l(s) + \varphi_r(s)) \\ (\varphi_l(s) + \varphi_r(s)) \end{bmatrix} + \frac{1}{2} \frac{\sigma - \kappa + (v - \eta)s}{\sigma - \kappa + \left(\frac{1}{\rho_1 - \rho_2} + v - \eta\right)s} \begin{bmatrix} (\varphi_l(s) - \varphi_r(s)) \\ (\varphi_r(s) - \varphi_l(s)) \end{bmatrix} \quad (4)$$

図9 人間の水平両眼運動神経経路の数学モデル

Mathematical model of human horizontal binocular motor neural pathways.

上記の式 (4) からわかるように、左右視軸の回転角 E_l と E_r は左右それぞれの眼における視標回転角 ϕ_l と ϕ_r の和と差の結合である。つまり、モデルには共役運動項と輻輳運動項二つの異なる1次系の伝達関数が存在し、共役運動時定数 T_c は式 (5)、輻輳運動時定数 T_v は式 (6) で表れる。

$$T_c = \frac{1 + (\rho_1 + \rho_2)(v + \eta)}{(\rho_1 + \rho_2)(\sigma + \kappa)} \quad (5)$$

$$T_v = \frac{1 + (\rho_1 - \rho_2)(v - \eta)}{(\rho_1 - \rho_2)(\sigma - \kappa)} \quad (6)$$

システムが安定な系になるためには、 $\sigma > \kappa$, $v > \eta$, $\rho_1 > \rho_2$ を満足する必要がある¹²⁾。本研究では次のようにゲインの設定を簡単化する。図9からわかるように、両眼を互いに影響し合う主なゲインは κ , η , ρ_2 であるが、 ρ_2 だけが機能した場合においても両眼協調制御が成り立つため、ゲイン設定の簡単化として κ , η を0にすることが可能となる。また、正規化するために $\rho_1 + \rho_2 = 1$ とする。式からわかるように、 ρ_1 と ρ_2 の差によって両眼輻輳運動時定数 T_v が変わるという関係がある。つまり、 $\rho_1 - \rho_2$ が小さくなるほど T_v が大きくなるといえる。また、両眼が常に同一の対象を追跡するので、輻輳角の変化は常に緩やかであり、大きい T_v が望ましい。したがって、 $\rho_1 - \rho_2$ が小さくなるように ρ_1 と ρ_2 を設定する。また、共役運動時定数 T_c が小さいほど追跡性能が向上する。これらの関係を元に ρ_1 と ρ_2 を決めた後、共役運動時定数 T_c が小さくなるように、かつ輻輳運動時定数 T_v が大きくなるように σ と v を決める。

本システムでは、図9で示したモデルのゲインを $\sigma=50$, $v=0.1$, $\kappa=0$, $\eta=0$, $\rho_1=0.51$, $\rho_2=0.49$ に設定した。したがって、共役運動時定数 T_c は0.022sとなり、輻輳運動時定数 T_v は1sと定めた。

両眼協調制御モデルの制御効果を確認するため、検証実験を行った。静止視標に対する注視点保持実験の結果を図10に示す。この実験では、システムの前方約5.5mの位置に配置した椅子に、動かないように人間を座らせて被験者とした。システムが被験者の顔を注視するように制御し、システムに両眼協調制御モデルを適用した場合と適用しない場

(a) 視標の奥行距離

(b) 視標の奥行速度

図10 両眼モデルの輻輳安定効果

Effect of binocular model for vergence stability.

合の比較検証を行った。モデルを適用しない場合は、両眼はそれぞれ独立なPIDフィードバック制御ループによって制御される。PID制御の各パラメータは限界感度法を用いて設定した。図10の横軸は時間である。図10(a)での縦軸は計測した視標の奥行距離であり、輻輳角の値と関連する。図10(b)における縦軸は視標の奥行速度である。図10での点線は、両眼協調制御モデルを適用しない時のデータであり、実線はモデルを適用する時のデータである。比較した結果、両眼協調制御モデルを適用することで、距離(輻輳角)の計測ノイズは大幅に低減した。

両眼協調制御モデルは、本システムにおいて二つの重要な役割を果たす。距離計測ノイズを大幅に低減することができるという役割と輻輳角が安定的に制御できるという役割である。安定した輻輳角制御により、前章で述べた両眼視差によるテンプレート更新が可能になる。なぜならば、不安定な輻輳角制御により、システムの注視点の位置が前後大きく揺れることになり、追跡対象がホログラムから離れやすくなる。その結果、前章で述べた手法でのテンプレートの更新はできなくなる確率が高く、結果的に視標追跡性能が低下することになる。

5. 評価実験

従来法における遠距離追跡での問題点として、追跡対象の回転が問題になってきた。従来の手法では追跡対象の回

図11 頭部の回転と検出された頭部に属するSIFT特徴点数の関係
Head rotation angle VS. number of SIFT key points belong to the head detected.

表2 頭部検出手法の比較
Comparison of several head detection methods.

手法	検出できる頭部の回転範囲	特徴
提案手法	-90°～+90°	背景と追跡対象の間にある程度の距離が必要である
従来の SIFT 手法 ⁹⁾	-20°～+20°	なし
CAMShift ¹³⁾	無制限	皮膚が見えること・背景の色が皮膚と異なることが必要
Multi-level Affine Image Registration ¹⁴⁾	約 -70° ～ 約 +70°	ローカル特徴を利用するため、ノイズに弱い
Shape Based Tracking ¹⁵⁾	無制限	楕円形頭の輪郭がはっきり見える、複雑な背景に弱い

転により、追跡対象を追跡できなくなるという問題があった。そこで本章では、最初に追跡対象の回転に対する追跡のロバスト性に対する評価実験を行った。

まず、両眼視差を利用して回転する追跡対象の頭部を検出する実験を行った。約3.3m離れた位置にある回転椅子に座る被験者を本システムで注視した。広角レンズカメラ Camera2とCamera4を利用すると、被験者の顔の画像サイズは約100×100ピクセルである。被験者は顔の正面(0°)から10°ずつ90°まで回転した。3章にて説明した手法により検出した被験者の頭部に属するSIFT特徴点数の数を記録した。図11はこの実験により得られた、頭部の回転角度に対する検出されたSIFT特徴点の中で頭部に属する特徴点数の数を示したグラフである。

図11の実験結果から、提案手法により被験者の頭部が90°回転しても、検出された頭部に属するSIFT特徴点数の数が、一定の数以上取得できていることが確認できた。つまり、追跡対象の向きが正面から±90°回転しても、本システムは安定に追跡対象を検出することができるといえる。

表2は頭部の検出手法に関する表である。この表は提案システムの頭部検出手法と既存のいくつかの頭部検出手法の比較を示している。われわれの提案手法では、頭部が±90°回転しても検出できる。ただし頭部は背景物から少し離れる必要がある。従来のSIFT手法では、頭部が正面から±20°

図13 両眼カメラシステムの座標系
Coordinate system for the binocular cameras system.

以上回転するとテンプレートとのマッチングができなくなり、頭部の検出ができなくなる。また、CAMShift法、Multi-level Affine Image Registration法およびShape Based Tracking法では、背景ノイズの影響を受けやすく、追跡対象検出の頑健性に欠ける。

次に、提案システムにおける追跡のロバスト性に対する評価実験を行った。追跡対象は本システムの前方をパン方向に折り返して歩く。本システムはこの実験において30FPSのフレームレートで合計1,000組のステレオ画像を記録した。図12は実験の様子を示す。

図12 (a) にはフレームシーケンス中の3組の画像を例として示した。上の段は左画像、下の段は右画像である。この実験において、追跡対象の向きが追跡中に大きく変化するような経路で被験者を移動させた。また、背景には特徴点を多く検出できるようなポスタを配置した。図12 (b) では更新されたテンプレートを示した。本システムでは、まず1フレーム目の正面顔をテンプレートとして登録し、続いてのフレームの中からSIFTマッチングを利用して顔の位置を検出していく。第224フレームにおいて、顔の向き変化による画像の変化のために顔テンプレートとの照合が失敗した。その時、両眼視差の小さいSIFT特徴点(図中の緑色の点、黄色枠線の中に包まれる)を新しいテンプレートとして登録し、以降の追跡に利用する。同様に、第271, 352, 366, 386, 415, 651, 800フレームにおいてもテンプレートとの照合に失敗したため、両眼視差の小さいSIFT特徴点を新たなテンプレートとして追加登録し、以降の追跡に利用した。

ここまでの評価実験により、本システムが回転する追跡対象を安定的に背景から切出すことができること、さらに長時間にわたるロバストな追跡ができることが確認できた。

次に、提案システムの3次元追跡性能の評価実験を行った。図13は本システムにおける座標系である。提案システムに対して、z軸上に無限遠の視標を注視するようなカメラの向きをカメラの初期位置とし、空間上の点Tの座標を(x_T, y_T, z_T)、カメラ間の基線距離をlとする。座標原点は左右のカメラ回転中心の真ん中に定めた。二つのカメラ回転中心のなす直線方向をx軸、システムの頸のモータの回転

軸を y 軸，システムの奥行方向を z 軸と定める． E_l (E_r) は左 (右) カメラの回転初期位置から点 T までのパン方向の回転角である． θ は初期位置から点 T までのチルト方向の回転角である． ω は頸の回転角度である．

点 T の座標は下記式 (7) ～式 (9) から求められる．

$$x_T = \frac{l \tan E_l + \tan E_r}{2 \tan E_l - \tan E_r} \cos \omega + \frac{l \sin \omega \cos \theta}{\tan E_l - \tan E_r} \quad (7)$$

$$y_T = -\frac{l \sin \theta}{\tan E_l - \tan E_r} \quad (8)$$

$$z_T = -\frac{l \tan E_l + \tan E_r}{2 \tan E_l - \tan E_r} \sin \omega + \frac{l \cos \omega \cos \theta}{\tan E_l - \tan E_r} \quad (9)$$

上記の式 (7) ～式 (9) からわかるように，視標の距離が遠くなるほど， E_l と E_r の誤差の及ぶ3次元計測誤差は大きくなっていく．本提案システムにおいて， E_l と E_r の誤差にはギア減速機の角度伝達誤差もあるので，遠距離の3次元追跡精度は充分とはいえない．そこで，3次元追跡の精度評価実験は近距離 (3m～5m) のみで行った (図14)．また，カメラのエンコーダの分解能は $7.5e-4^\circ$ と非常に高精度なので，両眼協調制御モデルによって遠距離においても滑らかな追跡ができ，追跡対象の両眼視差を小さく抑えるように制御することが可能である．この機能を検証すべく，遠

距離 (約70m) における追跡実験を行った (図15)．

3次元追跡の精度評価実験は地面の平らかな研究室内で行った．カメラのチルト方向回転軸は地面から約1.23mの位置に設置してある．被験者があらかじめ決めた矩形ルー

図14 測定軌跡と正解軌跡

Preset trajectory and measured trajectory.

(a)実験の概観

(b)水平方向におけると視軸のずれ

(c)水平方向における両眼視差

図15 遠距離追跡と両眼視差

Long distance pursuit and binocular disparity.

表3 計測位置と被験者に歩かせた軌跡との距離の統計結果
Statistical result of distance from measured position to preset trajectory.

平均値	47.0mm
標準偏差	35.3mm
最大誤差	112.9mm

トに沿って歩き、本システムによって計測した3次元位置の軌跡と被験者に歩かせた軌跡の比較を行った。

図14は被験者に歩かせた軌跡と本システムにより測定された軌跡を表すグラフである。細い線は被験者に歩かせた軌跡、太い線は測定された軌跡である。測定された軌跡が被験者に歩かせた軌跡と概ね重なっていることが確認できる。実験結果において、被験者に歩かせた位置と計測した被験者の位置との時間的な同期を取ることが不可能であるため、数値評価として本システムが計測した位置と被験者に歩かせた軌跡との直交距離の平均値、標準偏差、最大誤差を求めた。表3にその結果を示す。

遠距離での追跡実験を行った。実験条件と結果を図15に示す。まず図15 (a) に本システムの設置状況の概観を示す。本システムを地面から約26mの高さの建物に設置し、建物から約70m離れた通路を約1.2m/sで歩く歩行者を追跡させた。結果として得られた水平方向における視標の左右望遠カメラ画像中心からのずれを図15 (b) にグラフで示し、左右の望遠カメラによる両眼視差を図15 (c) にグラフで示した。その結果、400組の追跡画像の両眼視差の平均値は-1ピクセル、標準偏差は1.08ピクセルとなった。400組のステレオ画像の内、両眼視差が ± 2 ピクセル以内の数は392組であった。したがって、本実験において提案システムが遠距離追跡において両眼視差を ± 2 ピクセル以内に抑えるように制御できる確率は98%であるといえる。本システムを用いた遠距離追跡実験を繰り返し行い、良好な追跡結果を得られた。

6. む す び

本論文では、両眼が隣接に配置する監視カメラシステムを開発した。このシステムの特徴として以下の三つがあげられる。

- (1) 左右カメラを隣接させて配置し制御することによって、両眼が同一視標に注視している場合、視標に対する両眼の視差が小さくなり、両眼ステレオ画像の対応付けが容易になる。これにより、視標の3次元空間位置情報の取得は簡単になった。
- (2) 追跡対象の距離と背景の距離が区別できるため、その距離情報を利用して追跡対象を背景と分けることができ、従来の監視システムにおける回転する対象が追従できないという問題点を克服することができた。

- (3) 人間の両眼運動制御モデルを利用することで、両眼の共役運動と輻輳運動を区別して制御することができた。つまり、素早い共役運動による対象の追跡が実現でき、さらに安定した輻輳運動も可能になった。その結果、短い基線距離による奥行計測での誤差を大幅に減らすことができると同時に、安定な両眼追跡が可能となった。

今後は本システムを中心にネットワーク型の監視カメラシステムへの拡張を行い、広域監視に適した高度な監視システムへと発展させていきたい。

本研究は、独立行政法人科学技術振興機構 (JST) 大学発ベンチャー創出推進プロジェクトの援助金を受けたものである (課題番号1907)。

〔文 献〕

- 1) C. Chen, Y. Yao, D. Page, B. Abidi, A. Koschan, and M. Abidi: "Heterogeneous Fusion of Omnidirectional and PTZ Cameras for Multiple Object Tracking", IEEE Trans. On Circuits and System for Video Technology, 18, 8, pp.1052-1063 (Aug. 2008)
- 2) S. Kang, J. Paik, A. Koschan, B. Abidi, and M. Abidi: "Real-time Video Tracking Using PTZ Cameras", in Proc. of SPIE 6th International Conf. On Quality Control by Artificial Vision, 5132, pp.103-111 (May 2003)
- 3) R. Collins, O. Amidi, and T. Kanade: "An Active Camera System for Acquiring Multi-view Video", in Proc. of the 2002 Intl. Conf. On Image Processing, pp.517-520 (Sep. 2002)
- 4) D. Lowe, and J. Little: "Mobile Robot Localization and Mapping with Uncertainty using Scale-Invariant Visual Landmarks", the Intl. Journal of Robotics Research, 21, 8, pp.735-758 (2002)
- 5) S. Se, and P. Jasiobedzki: "Instant Scene Modeler for Crime Scene Reconstruction", IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops 2005, pp.123 (2005)
- 6) 森田真司, 山澤一誠, 寺沢征彦, 横矢直和: "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", 信学論, J88-D-II, 5, pp.864-875 (2005)
- 7) P. Viola, and M.J. Jones: "Rapid Object Detection Using a Boosted Cascade of Simple Features", in Proc. of IEEE Computer Society Conf. On Computer Vision and Pattern Recognition, pp.511-518 (2001)
- 8) R. Lienhart, and J. Maydt: "An Extended Set of Haar-like Features for Rapid Object Detection", in Proc. of Int. Conf. On Image Processing, 1, pp.900-903 (Sep. 2002)
- 9) D.G. Lowe: "Object Recognition from Local Scale Invariant Features", in Proc. of IEEE Int. Conf. On Computer Vision, pp.1150-1157 (1999)
- 10) Y. Ke, and R. Sukthankar: "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors", In Proc. of IEEE Conf. On Computer Vision and Pattern Recognition, pp.511-517 (2004)
- 11) A. Stein, and M. Herbert: "Incorporating Background Invariance into Feature-based Object Recognition", Proc. of IEEE Workshop on Applications of Computer Vision), pp.37-44 (Jan. 2005)
- 12) X. Zhang, and H. Wakamatsu: "Mathematical Model for Binocular Movements Mechanism and Construction of Eye Axes Control System", Journal of the Robotics Society of Japan, 20, pp.89-97 (2002)
- 13) G. R. Bradski: "Computer Vision Face Tracking as a Component of a Perceptual User Interface", in Workshop on Applications of Computer Vision, pp.214-219 (1998)
- 14) T. Bausch, P. Bayerl, and H. Neumann: "Multi-level Face Tracking for Estimating Human Head Orientation in Video Sequences", Perception and Interactive Technologies, 4021, pp.179-182 (2006)
- 15) S. Birchfield: "An Elliptical Head Tracker", 31st Asilomar Conf. On Signals, Systems, and Computers, pp.1710-1714 (Nov.1997)

図8 両眼視差で追跡対象を検出する例
An example of detect visual target by binocular disparity.

(a) フレームシーケンス

(b) 追加テンプレート

図12 追跡ロバスト性実験

Demonstration of tracking robustness.

谷 宇 1997年，浙江大学工学部光电科学仪器工程学科卒業。2000年，同大学大学院修士課程修了。同年，(株)オプトラン入社。2004年，東京工業大学電子機能システム専攻入学。2009年，同大学大学院博士課程修了。現在，同大学産学官連携特別研究員。視覚情報処理，ロボットビジョンの研究に従事。

佐藤 誠 1973年，東京工業大学工学部電子物理工学科卒業。1978年，同大学大学院博士課程修了。同年，同大情報工学科助手。1986年，東京工業大学精密工学研究所助教授を経て，現在，同大学精密工学研究所教授。コンピュータビジョン，パターン認識，ヒューマンインタフェース，VRの研究に従事。工学博士。正会員。

張 曉林 1995年，横浜国立大学工学研究科博士修了。同年，東京歯科大学医学部保健衛生学科助手。2003年より，東京工業大学大学院助教授。2007年より，同大学大学院准教授。制御工学，ロボット工学，視覚情報処理，ロボットビジョン，神経細胞の等価回路の研究に従事。博士(工学)。