

実空間に対応した Master-Slaving による追尾カメラシステム

Visual Tracking System by Master-Slaving

小川雄三 † 藤吉弘亘 ‡

Yuzo Ogawa†, Hironobu Fujiyoshi‡

†TOA(株) ‡中部大学工学部情報工学科

†TOA, ‡Dept. of Computer Science, Chubu University

E-mail: ogawa_yuzo@toa.co.jp, hf@cs.chubu.ac.jp

Abstract

本報告では、キャリブレーション済みの広角カメラ (Master) 映像から移動体検出した結果を用いて、推定した実空間における物体位置を基に複数台の巡回型カメラ (Slave) を制御して追尾するシステムについて提案する。本システムの特長として、単眼カメラ (Master) のみによる検出した物体の実世界における位置と高さが取得可能、検出した人の高さ (身長) より複数の Slave カメラの注視点を個々に設定した自動追尾、複数台の巡回カメラにより多方向から撮影した対象物体の映像を取得可能が挙げられる。さらに、検出物体の移動軌跡を地図上に表示することで行動経路の把握が容易となる。実際に自動追尾システムの運用を考えた場合、物体検出の段階において、複数の移動体の重なりによるオクルージョンや誤検出が問題となる。そこで Master カメラの映像上でユーザがポインタを操作して Slave カメラを制御し、追尾する撮影システムについても提案する。

1 はじめに

近年、犯罪発生率の急増に伴い、ビデオカメラによるセキュリティシステムの重要性がますます高まっている [1]。特に防犯を目的とした場合、人物や行動を特定するために詳細な映像が必要である。特に監視エリアが広範囲になる場合、複数のカメラを効果的に配置し、さらにパン・チルト・ズーム機能を持つカメラによる追尾を行なうことで、対象物の詳細映像を取得することが考えられる。しかしながら、監視員は複数のモニタを同時に注視し、さらにカメラの巡回やズーム等の操作を行う必要があり、このような負担の増加が問題とされている。

これに対し、米国では DARPA の下、画像理解技術を

用いたビデオ監視システムの研究プロジェクト VSAM が行われた [2]。VSAM プロジェクトでは、屋外に 12 台のカメラを配置したテストシステムを構築し、侵入物体を複数のカメラの協調動作により追跡し、その状況をリアルタイムで監視員に提示することが可能となっている。国内においても、分散協調視覚に関する研究プロジェクト [3] において、複数カメラでの人物追跡による顔画像登録システム [4] や、分散カメラエージェントによる人物追跡システムが提案されている [5]。これらのシステムは、監視員の負担軽減と作業効率化に大きく貢献でき、次世代のビデオ監視システムとして期待されている。

本報告では、キャリブレーション済みの広角カメラ (Master) 映像から移動体検出した結果を用いて、推定した実空間における物体位置を基に複数台の巡回型カメラ (Slave) を制御する Master-Slaving による追尾カメラシステムを提案する。同様なシステムとして、親カメラと子カメラによる協調監視システムが提案されている [6]。このシステムでは親カメラの画像内における位置を用いて追尾しており、親カメラと子カメラは近い視点となるように配置しなければならないという制約がある。そのため、対象物の多方向からの映像を取得することが困難である。提案する本システムでは、単眼カメラ (Master) のみで検出した物体の実世界における位置と高さを推定する。検出した人の高さ (身長) を基に複数の Slave カメラの注視点を個々に設定した自動追尾を行なう。これにより、Slave カメラにより多方向から撮影した対象物体の顔や注目領域の詳細映像が取得可能となる。

しかし、実際に自動追尾システムを運用することを考えた場合、その検出段階において複数の移動体の重なりによるオクルージョンや誤検出が問題となるため、重要な局面においては監視員がカメラを操作して監視する場合も多い。そこで本研究では、Master カメラの映像上でユーザがポインタを操作するだけで複数の Slave

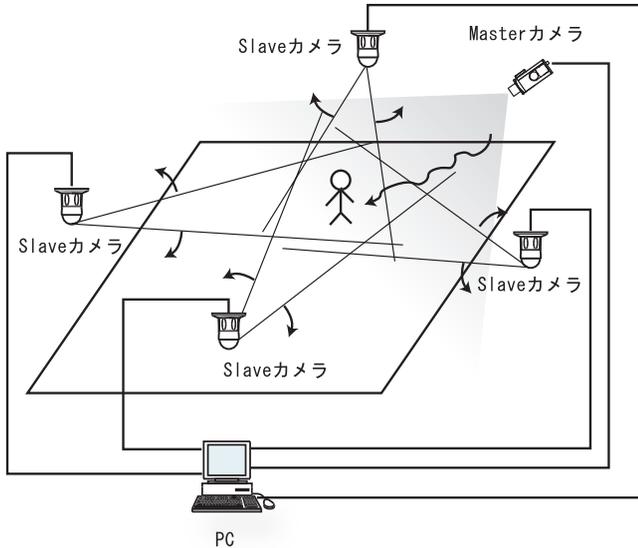


図1 追尾カメラシステムの構成例

カメラを制御し、対象物体の多方向から映像を獲得することのできる撮影システムについても提案する。

2 Master-Slavingシステム

Master-Slavingによる追尾カメラシステムの構成とその手法について述べる。

2.1 システム構成

本追尾カメラシステムは、対象領域全体を監視する1台の広角カメラと、パン・チルト・ズームが可能な複数台の回転型カメラから構成される。この広角カメラをMasterカメラと呼び、FOVが約90°のものを使用した。また、回転型カメラをSlaveカメラと呼ぶ。構築したMasterカメラ1台、Slaveカメラ4台からなるシステム構成を図1に示す。各カメラの設置位置、姿勢は予め取得しておくものとする。

本システムは、監視領域内に侵入した物体をMasterカメラより自動的に検出する。物体検出法には、対象物の領域をより正確に検出できるレイヤー型検出を用いる[7]。このとき得られた検出領域の下端部の中心点及び上端部の中心点を取得する。

次に、検出した画像座標に対応する実空間の位置を次節に述べる手法により推定する。推定した位置情報を基に複数のSlaveカメラのパン・チルト・ズームを制御して、多方向から対象物の映像を取得する。本システムでは、Slaveカメラにおいては画像処理を行わないため、1台のPCによる構成で、Master1対Slave*n*台のカメラシステムを容易に実現することができる。

2.2 床面上の位置推定

カメラ映像から得られる情報は2次元であるが、物体が存在する実空間は3次元である。物体の3次元位置情報を求めることができれば、カメラの制御を実空

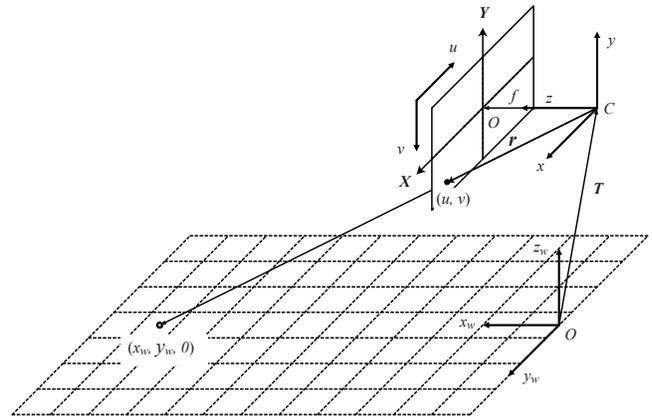


図2 光線情報による床面の位置推定

間に合わせて行うことが可能となるため、空間内の任意の位置に設置した複数のSlaveカメラから目標物体を撮影することができる。本システムでは、2次元カメラ映像から実世界の位置を推定する手法として、カメラキャリブレーション[8]により得られたカメラの内部・外部パラメータから、3次元空間の光線情報に基づく位置推定を行う。この手法では、図2に示すようなカメラ中心とカメラ画像上の座標を通る世界座標空間における直線を求め、その直線が床面に交差する点を対象物の位置として計算する。これにより、カメラ画像の座標 (u, v) に対応した床面上の世界座標 $(x_w, y_w, z_w = \text{床面})$ を求めることが可能となる。このとき、本システムにおける世界座標 (x_w, y_w, z_w) とカメラ座標 (x, y, z) の関係は次式で表される。 R は3行3列の回転行列、 T は平行移動ベクトル $[T_x, T_y, T_z]^T$ である。

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R \left(\begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} - T \right) \quad (1)$$

以下に画像座標 (u, v) より、世界座標 (x_w, y_w, z_w) を求める手法を示す。

Step1 画面平面 $X-Y$ 上の点 (X_d, Y_d) を次式から求める。 (u_c, v_c) は画像座標の中心である。

$$X_d = u_c - u, \quad Y_d = v_c - v \quad (2)$$

Step2 レンズ歪みを修正した点 (X_u, Y_u) を次式から求める。 r は画像中心からの距離であり、 S は画像座標の縦横比、 k_1, k_2 はレンズ歪み係数である。

$$\begin{aligned} X_u &= SX_d(1 + k_1 r^2 + k_2 r^4), \\ Y_u &= Y_d(1 + k_1 r^2 + k_2 r^4) \end{aligned} \quad (3)$$

Step3 図2において、カメラ座標の原点から画像上の任意点 p を通過する光線ベクトル $\mathbf{r} = [x_v, y_v, z_v]^T$ は次

式となる． f はカメラ座標における焦点距離とする．

$$\mathbf{r} = \begin{bmatrix} x_v \\ y_v \\ z_v \end{bmatrix} = R^T \cdot \begin{bmatrix} X_u \\ Y_u \\ f \end{bmatrix} \quad (4)$$

Step4 光線ベクトル \mathbf{r} を 3次元空間上の直線で表すと，次式となる． α は実数である．

$$\begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = \alpha \begin{bmatrix} x_v \\ y_v \\ z_v \end{bmatrix} + T \quad (5)$$

Step5 ここで床面が $x_w - y_w$ 平面に平行であり， $z_w = 0$ であると仮定すると，直線が平面と交わる時の α を決定できる．

$$\alpha = -\frac{T_z}{z_v} \quad (6)$$

Step6 α の値を式 (5) に代入することで，画像座標点 (u, v) に対応した世界座標点 $(x_w, y_w, 0)$ を求めることができる．

$$x_w = -\frac{T_z}{z_v}x_v + T_x, \quad y_w = -\frac{T_z}{z_v}y_v + T_y \quad (7)$$

2.3 実世界座標と高さの取得

次に，床面の位置情報を用いて対象物の高さを取得する手法を以下に示す (図 3 参照)．

Step1 検出領域の下端部中央点 (u_b, v_b) が，地面に接していると仮定し，2.2 で述べた手法により，対応する世界座標点 $(x_{wb}, y_{wb}, 0)$ を求める．

Step2 検出した対象物が地面から垂直に立っていると仮定し，世界座標点 $(x_{wb}, y_{wb}, 0)$ を通る $y_w - z_w$ 平面を考える．

Step3 人物の頭頂部にあたる点の画像座標点 (u_t, v_t) を通る光線を式 (5) より求める．ここでは，式 (5) における実数 α を β とする．その光線が Step2 の $x_w - z_w$ 平面と交差する β を以下より決定する．

$$\beta = \frac{(x_{wb} - T_x)}{x_v} \quad (8)$$

Step4 β の値を式 (5) に代入することで，画像座標点 (u_t, v_t) に対応した世界座標点 (x_{wb}, y_{wt}, z_{wt}) を求めることができる．

$$\begin{aligned} y_{wt} &= \frac{y_v(x_{wb} - T_x)}{x_v} + T_y, \\ z_{wt} &= \frac{z_v(x_{wb} - T_x)}{x_v} + T_z \end{aligned} \quad (9)$$

この z_{wt} は，検出した人物の実世界における床面からの高さであり，身長を示している．

Step5 人物の顔は頭頂部の鉛直下方向にあるので，頭頂部の世界座標点 (x_{wb}, y_{wt}, z_{wt}) から一定値 z_f 減算した世界座標点 $(x_{wt}, y_{wt}, z_{wt} - z_f)$ を，Slave カメラで狙うべき世界座標点とする．

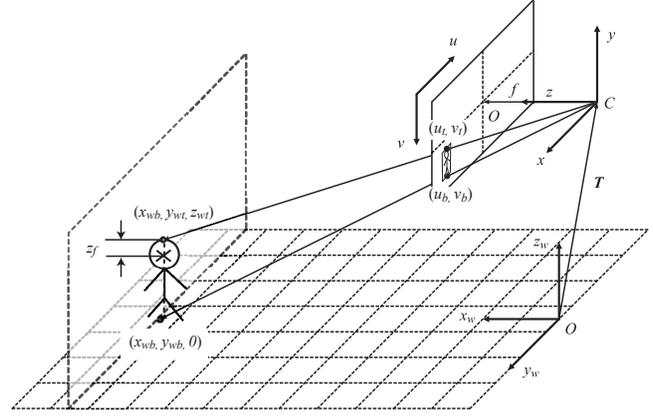


図 3 高さの取得

2.4 追尾動作

各 Slave カメラが求めた世界座標点となる目標を捉えるようにパン・チルト・ズームを制御する．パン・チルト角はカメラと対象物の位置関係から計算する．

Slave カメラと推定された物体の位置までの距離からズームパラメータを決定する．このとき，物体が一定の大きさに映るように，物体までの距離を基にズームパラメータを調整する．さらに本システムでは，推定位置毎の空間分解能を考慮し，対象物が空間分解能の低い領域にいるときは Slave カメラ視野から物体が外れないように，ズームを少し引くようにパラメータを調節する．推定位置毎の本システムの空間分解能については次章で述べる．

このように，Slave カメラのパン・チルト・ズーム量は世界座標点を基に計算するため，任意位置に設置された Slave カメラを，任意の世界座標点に向けて制御することができ，対象物を多方向から撮影することが可能となる．

2.5 Slave カメラ間の役割

Slave カメラの協調動作を行なうには，以下の 2 通りの制御が考えられる．

1) 全ての Slave カメラ上において，対象物が一定の大きさ映るように各カメラのズームパラメータを制御する方法

2) 物体の進行方向と各 Slave カメラ位置の関係より，対象物の注視点をカメラ毎に変える方法

2) において，ある Slave カメラの方向に人が移動している場合，その Slave カメラからは人の正面を捉えることが可能であると考え，頭部・顔領域をズームした状態となるように注視点を設定する．移動方向にない Slave カメラは，対象物全体を捉えるようにパン・チルト・ズームを制御する．このように，本システムでは複数の Slave カメラの役割をカメラの位置と対象物の進行方向より自動的に決定している．

表 1 $x_w - y_w$ 平面における位置推定誤差

平面の高さ (z_w)	推定誤差 [mm]	
	平均	標準偏差
0	46.57	28.46
500	43.01	31.50
1000	36.40	28.85
1500	48.13	36.52
2000	68.89	76.84

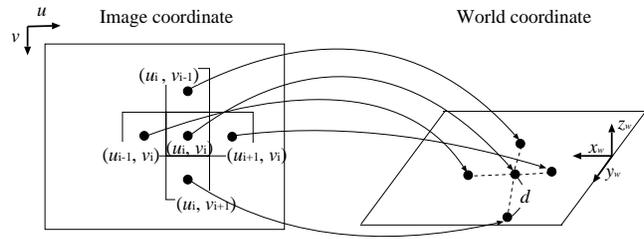


図 4 空間分解能の計算法

表 2 高さ方向の位置推定誤差

高さ	推定誤差 [mm]	
	平均	標準偏差
500	17.33	17.84
1000	11.61	10.61
1500	7.71	6.83
2000	11.47	6.42

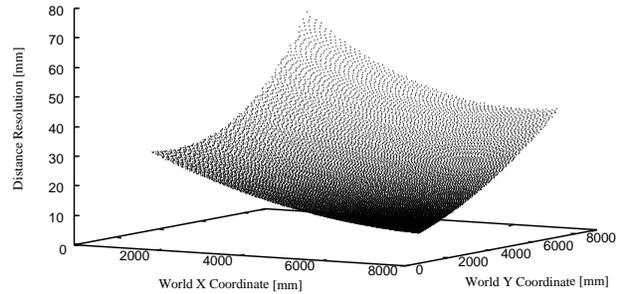


図 5 空間分解能の分布

3 実験と結果

評価実験には縦・横・高さが約 $9\text{m} \times 8\text{m} \times 3\text{m}$ の部屋を対象とし、Master カメラ 1 台、Slave カメラ 2 台を用いた。使用した Master カメラは通常の固定カメラ、Slave カメラは高速回転カメラ (回転速度: 最大 $360^\circ/\text{s}$ 、ズーム倍率: 光学 23 倍) である。

Master カメラのキャリブレーションには、高さを 0, 500, 1000, 1500, 2000[mm] とした平面上の、合計約 200 点を用いた。

3.1 位置推定精度の検証

得られたカメラパラメータを用いて、2.2 で述べた位置推定により求めた世界座標点とキャリブレーションに用いた点との $x_w - y_w$ 平面上における位置推定誤差を表 1 に示す。表 1 により、位置推定の平均誤差は 48.6[mm] であった。

次に、高さ方向の推定誤差として、各特徴点を $y_w - z_w$ 平面に対して位置推定し、推定した世界座標点の高さを真値と比較する。その位置推定結果を表 2 に示す。全体の平均推定誤差は 12.1[mm] であった。

このように、カメラキャリブレーション時に用いた特徴点に対しては位置推定誤差を求めることができるが、それ以外の点に対応する世界座標の真値が分からないため、誤差を求めることができない。そこで本研究では、図 4 のように、画像座標 (u, v) が 4 近傍に 1 画素ずれた際に対応する世界座標点 $(x_w, y_w, 0)$ 間のユークリッド距離の平均値を求め、これを空間分解能とし、推定誤差の指標とする。

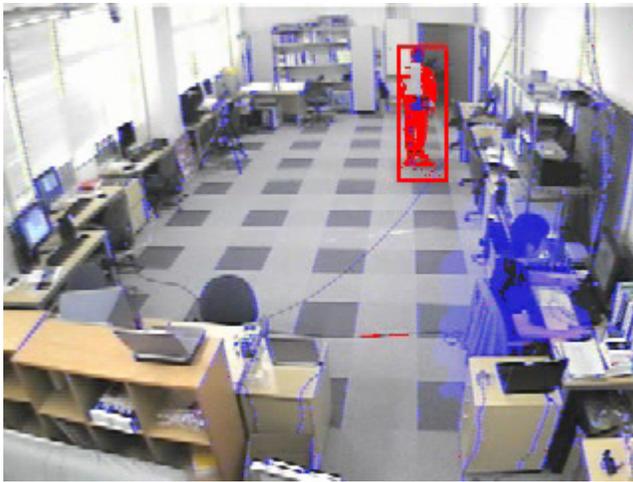
この空間分解能を、画像上で床面に相当する領域に

ついて算出した結果を図 5 に示す。図 5 より、カメラからの距離が離れるほど空間分解能が低くなり、最大で 1 画素のずれが約 80[mm] に対応していることがわかる。本システムでは、このように予め Master カメラの画像座標に対応した実空間の分解能を計算し、これを Slave カメラの制御の際に不確定要素の大きさとして扱う。

3.2 Master-Slaving による自動追尾

移動体の自動追尾における本システムの動作例を図 6 に示す。図 6(a) が Master カメラ映像で、長方形で囲まれた領域が検出領域を示している。検出領域から推定した位置を基に制御した Slave カメラ映像を (b) と (c) に示す。(d) は監視領域を真上から見下ろしたマップであり、検出した物体の軌跡、Slave カメラの位置とその視野等の状況をリアルタイムで提示する。この動作例における対象物は、Slave カメラ 1 の方向に移動しているため、Slave カメラ 1 は顔正面を捉えるように制御されている。一方、Slave カメラ 2 は、カメラと対象物の位置関係より対象物全体を捉えるように制御されている。このように、本システムは適応的に Slave カメラの注視点を位置関係から決定するため、常に正面顔領域をズームした状態の映像を取得することが可能である。

図 7 に高さ検出の例を示す。図 7 のグラフは、フレーム毎の検出した物体の高さを示している。また、その際の Slave カメラ映像を上を示す。図 7 より、しゃがんだ人物の高さを約 1[m] と検出していることがわかる。また、高さ情報を用いて Slave カメラを制御しているた



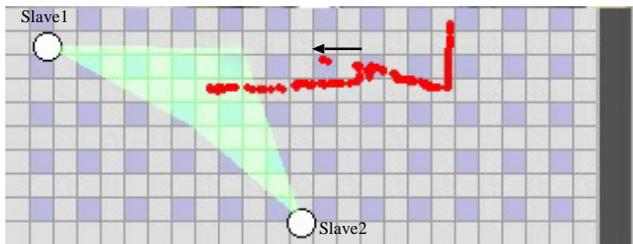
(a) Masterカメラ



(b) Slaveカメラ1



(c) Slaveカメラ2

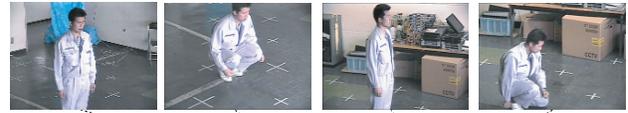


(d) マップ

図6 追尾カメラシステムの動作例

め、しゃがんだ人物の全体像を捉えることができている。このように、本システムでは検出物体の実世界における高さを推定することで、対象物の身長に合わせた追尾を実現することができる。

また、しゃがんでいない状態である200フレームまでの高さが周期的に約100[mm]の幅で変化していることがわかる。これは、物体がカメラに向かって歩いているため、進行方向に足を踏み出した際に、対象画像領域が大きく検出されるからである。このように、位置精度は物体検出の能力に大きく左右される。特に、複数物体が重なったときのオクルージョンや誤検出は、実際の運用上大きな問題となる。そこで、本システムではこのような状況の下において、監視員であるユーザがMasterカメラ映像上の複数のSlaveカメラを同時に操作することを考え、その操作ユーザーインターフェースを提案する。



Slave Camera's Image

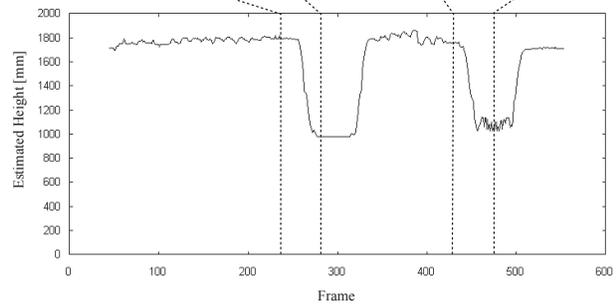


図7 高さの検出例

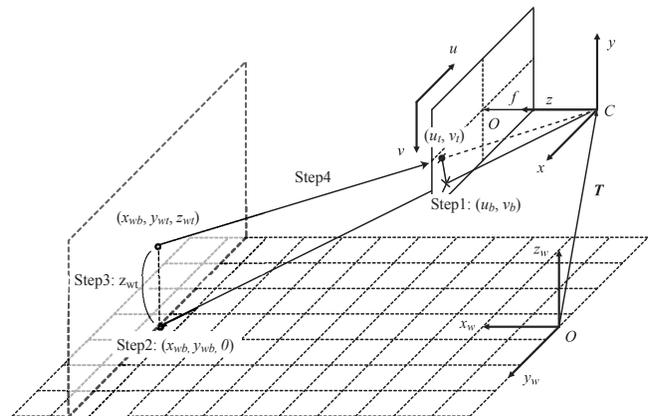


図8 操作ポイントによる処理の流れ

4 映像重畳型操作インターフェース

自動追尾における物体検出能力は万全ではないため、ユーザである監視員が手動でMasterカメラ映像上で指定した任意の点にSlaveカメラを向ける撮影システムの操作インターフェースについて述べる。

4.1 操作ポイントの重畳表示

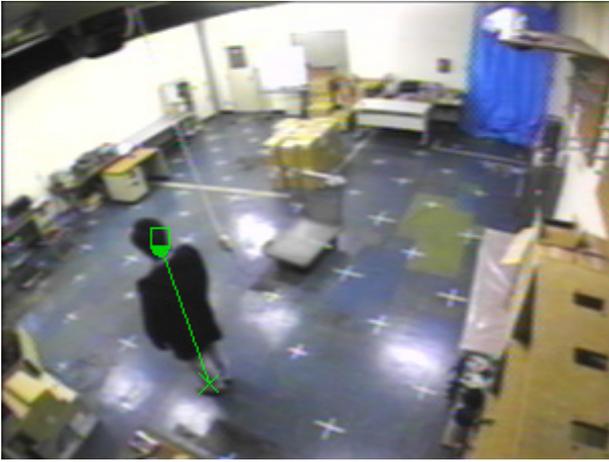
提案するインターフェースは、Masterカメラ映像上に、図9(a)のような実空間に対応した操作ポイントを重畳表示することで実現する。この処理の流れを以下に示す(図8参照)。

Step1 Masterカメラ映像上における対象物の下端点 (u_b, v_b) をマウスにより指定する。

Step2 画像座標点 (u_b, v_b) から対応する床面上の世界座標点 $(x_{wb}, y_{wb}, 0)$ の位置を推定する。

Step3 実空間における、対象の床面からの高さ(身長 z_{wt})を指定する。高さの指定はマウスのホイールでリアルタイムにユーザが可変することができる。

Step4 世界座標点 (x_{wb}, y_{wb}, z_{wt}) を画像平面に投影し、画像座標点 (u_t, v_t) を求める。その点にカメラの注視点を示すアイコンを表示する。その際には、画像座標点 (u_b, v_b) と (u_t, v_t) を結ぶ直線も表示する。



(a) Masterカメラと操作ポイント



(b) Slaveカメラ1



(c) Slaveカメラ2

図9 映像重畳型操作インターフェース

Step5 表示された操作ポイントがユーザの意図する位置を示している場合、マウスクリックによりその世界座標点 (x_{wb}, y_{wb}, z_{wt}) を取得し、その位置を基に Slave カメラを制御する。

4.2 操作例

本インターフェースの操作例を図9に示す。ユーザは人物が地面に接している点に、マウス操作で×印を合わせる。次にマウスホイールを操作することで、ポイント上部のアイコンを上下させて、顔領域に重ねる。このとき、ポイントは実空間における z_w 軸(垂直)方向に伸びているので、Master カメラからの距離に応じて伸縮する。また、この操作ポイントの直線は画面の端になるほどレンズ歪み考慮して傾くため、ユーザは直感的にレンズを通した際の3次元的な空間の広がりを把握できる。マウスの右クリックにより世界座標を取得し、Slave カメラを制御することで、図9(b),(c)のように複数の Slave カメラによる多視点映像を得ることが可能となる。

このように、本システムでは、ユーザはマウスだけの非常に簡単な操作で複数のカメラを同時に操作することができる。このため、対象物の多方向からの画像情報を素早く得ることができ、緊急事態発生時の対応能力向上が期待できる。また、マウスドラッグによる連続制御により、特定の物体をユーザが手動で追尾することも可能である。さらに、本システムは、スポー

ツを撮影した多視点カメラ映像の切り替えにより、あたかもカメラが動いているような効果の映像を表示するシステム[9]の撮影操作のインターフェースに応用することも考えられる。

5 まとめ

本研究では、単眼の Master カメラにより、物体の実世界座標と高さを取得し、その位置情報を基に複数の Slave カメラを制御する Master-Slaving システムを提案した。提案したシステムでは、物体の詳細映像を多方向から同時に取得することができるため、防犯上重要な情報を見逃しにくい撮影システムを構築することができた。また、映像重畳型インターフェースにより、物体検出が困難な場面ではユーザが簡単な操作で複数台のカメラを同時に制御することが可能となった。

参考文献

- [1] 東京・新宿広場に防犯カメラ設置計画, 毎日新聞 2002年3月6日(2002).
- [2] R.Collins, A.Lipton, H.Fujiyoshi and T.Kanade: "Algorithms for cooperative multisensor surveillance", Proceedings of the IEEE, Vol.89, No.10, pp.1456-1477(October, 2001).
- [3] 松山隆司: "分散協調視覚-視覚・行動・コミュニケーション機能の統合による知能開発", 画像の認識・理解シンポジウム(MIRU'98), pp.I-343-352(1998).
- [4] 加藤丈和, 向川康博, 尺長健: "安定な顔認識のための分散協調登録", 電子情報通信学会論文誌(D-II), Vol.J84-D-II, No.3, pp.500-508(2001).
- [5] 中澤篤志, 加藤博一, 日浦慎作, 井口征士: "分散視覚エージェントを用いた複数人物追跡システム", 情報処理学会論文誌, Vol.42, No.11, pp.2699-2710(2001).
- [6] 伊藤渡, 上田博唯: "広域監視用親カメラと追尾監視用子カメラによる協調監視システム", 画像の認識・理解シンポジウム(MIRU'00), pp.II-229-234(2000).
- [7] H. Fujiyoshi, and T. Kanade: "Layered Detection for Multiple Overlapping Objects", International Conference of Pattern Recognition Proc of ICPR2002, Vol.4, pp.156 - 161 (2002)
- [8] R. Y. Tsai: "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", IEEE Journal of Robotics and Automation, Vol. RA-3, No.4, pp.323-344 (August 1987).
- [9] <http://www.pvi-inc.com/eyevision/>