

実世界の取り込みとモデル構築のための被写体追尾システム

秋山 拓也[†] 関谷 隆行 有澤 博[‡]

[†] 横浜国立大学大学院環境情報学府 〒240-8501 横浜市保土ヶ谷区常盤台 79-7

[‡] 横浜国立大学環境情報研究院 〒240-8501 横浜市保土ヶ谷区常盤台 79-7

E-mail: [†] takuya@arislabs.ynu.ac.jp, [‡] arisawa@ynu.ac.jp

あらまし 近年、実世界のさまざまなデータをコンピュータに取り込む技術が盛んに研究されている。特にカメラ画像においては取得できるデータの種類が多い。しかし、カメラは撮影できる範囲が限られているため、広範囲を取得するには、カメラの数を増やす、もしくは、カメラを操作して適宜必要な範囲を撮影するなどの方法を用いなければならない。ここで人手を介さず、コンピュータ自身が環境の変化に応じて、カメラ装置を最適に制御できれば非常に有用である。そこで本論文では、被写体として最も重要な人間を対象にした追尾システムを構築した。

キーワード 画像処理、マルチメディア処理、コンピュータビジョン、最適制御

Active camera system for taking and model construction of real world

Takuya AKIYAMA[†] Takayuki SEKIYA and Hiroshi ARISAWA[‡]

[†] Graduate School of Environment and Information Sciences, Yokohama National University

[‡] Faculty of Environment and Information Sciences, Yokohama National University

79-7 Tokiwadai, Hodogaya-ku, Yokohama, 240-8501 Japan

E-mail: [†] takuya@arislabs.ynu.ac.jp, [‡] arisawa@ynu.ac.jp

Abstract The real world into the computer is actively researched. Especially, there are a lot of kinds of the data that can be acquired in the camera image. However, it is necessary to use the method such as operating the camera of increasing the number of cameras to acquire the large range because the range that can be taken a picture is limited as for the camera. If the computer can control the input device best according to the change of scene, it is very useful. Then, the pursuit system intended for man was constructed in this thesis.

Keyword Image processing, Multimedia processing, Computer vision, Optimal control

1. はじめに

1.1. 実世界データの取り込みとモデル構築

実世界の3次元時空間情報、たとえば、人体や機械の位置情報、姿勢・形状の変化情報、音声や各種のセンサ情報などを計器によって正確に計算機中に取り込み、これをデータベース化することによって、3次元空間内でのイベントの発見、運動評価、姿勢や動作の最適化などに役立つシステムを構築することは非常に広い応用範囲を持つと考えられる。我々はこの概念をリアルワールドデータベース(Real World Database)と称し、その構成要素技術の研究開発や、人体の幅広い負荷評価を目指したシステム構築などを行ってきた。^[1]

高精細ビデオカメラによる画像を実世界の情報を取り込む手段として考えると、データ量が膨大であるという問題がある反面、適切な解析を行うことで、非常にリッチな情報が引き出し得るという点で、きわめて有用である。カメラ画像を計算機に取り込む仕組み

は、テレビ会議、監視カメラなど多くの実用的な応用システムで実現されており、複数台のカメラ画像を同期して取り込むことにより、被写体の位置解析を行うモーションキャプチャシステムも実用化している。しかし、これら多くの応用分野において常に問題となるのは、一般的なカメラの視野は狭く、撮影範囲が極めて限られてしまうことである。視野を広角にすることは同時に画像の精度を落とすことでもあり、適切な解決手段にはなり得ない。一方、対象世界の状況を判断し、その時々に必要な場所だけを自動的に写しこむことができる装置があれば、その利用価値は非常に大きい。

被写体を自動追尾するシステムは既にさまざまなものが提案されているが、これらの手法は、差分処理で得られた被写体領域のみを頼りに追尾を行う、また被写体にセンサを取り付けるなどの方法がとられている。この方法では、被写体に負担をかけ、カメラの制御に制約があるなど問題が多い。

そこで我々は、撮影対象となっている世界をすばやく把握し、世界モデルを計算機内に構築すると同時に、その世界で注目すべき被写体を発見して、カメラのパン、チルト、ズーム、フォーカスを制御し、自動的に特定の被写体を追尾する装置を考えることとした。

被写体追尾システムを構築するためには、次の2つの要素技術が必要である。

(1) 対象世界の把握技術

撮影対象となる3次元空間、また被写体の情報をリアルタイムに取得する技術

(2) カメラの制御技術

対象世界のデータを元に、最適な位置にカメラを向けるためのデータをリアルタイムに算出し、実際にカメラを制御する技術

これらの問題を解決するために、我々は、対象世界を把握するための装置を導入することを考えた。さらに、追尾する被写体の情報を効率よく算出するアルゴリズムを開発し、得られた被写体情報を元にリアルタイムにカメラを制御する方法を提案する。

以降2章では自動追尾の構成要素と既存手法について述べ、3章では提案システムの概要、4章では被写体情報抽出手法について述べ、5章ではカメラ制御の方法、6章で実験結果と考察、7章でまとめを述べる。

2. 自動追尾の構成要素

2.1. 自動追尾に必要な技術

被写体追尾システムを構築するためには次の問題を克服する必要がある。

(1) 対象世界の把握

撮影対象となる3次元空間がおおよそどのような形をしていて、被写体の動く空間はどこかを把握する。また、自動追尾システムを構築するのに、最も重要な技術は被写体情報の把握方法である。ここで、被写体情報とは、被写体の現在位置（カメラの向きを決めるために）、被写体の大きさ（画角（ズーム）を決定するために）である。どのように被写体情報を把握するかは、自動追尾システムを構築する上で、追尾精度を左右する非常に重要な問題である。また、情報把握のリアルタイム性も重要な項目である。

(2) カメラの制御

得られた対象世界のデータを元に、カメラを自動制御する必要があるが、その際、カメラの位置姿勢を把握し適切なコントロールデータをリアルタイムに算出、送信しなければならない。

上記の問題、特に対象世界の把握方法について既存手法の考察を次節で行う。

2.2. 自動追尾の既存手法

自動追尾に関する既存手法にはさまざまなものがあるが、大きく分けると以下に示すものがある。

(a)動作中は処理を行わず、停止中にフレーム間差分をとる方法

(b)パン、チルトした量を計算して、前フレームをその分移動させて差分をとる方法[2][3]

(c)カメラ移動範囲の背景を先にすべて取得し、パン、チルト量からその時点の背景領域を切り出し、その画像で追尾カメラの画像と差分処理を行う方法[4][5]

(d)人間の肌色領域を抽出する方法[6]

これらの4つの方法に共通しているのは、追尾カメラから得られる情報だけで、被写体を追うことを中心に考えているということである。基本的な処理方法として、差分処理が用いられている。これは、処理が簡単であり、比較的短時間で被写体抽出が行えるため、リアルタイム性を重視する自動追尾システムに適しているからだと考えられる。しかし、差分処理から得られる情報は、画像上における被写体領域の重心の位置や、領域の大きさといったものに限られる。実際、上に挙げた方法のほとんどが、この2つの情報をもとにカメラを制御している。しかし、被写体領域には偏りがあるため、領域から重心を求めたとしても、実際の画面上で被写体を中心にくるとは限らない。また、画面から被写体が外れたら追いつけられなくなるといった本質的な問題があるため、それを回避するため、ズームに制限を加えるなど、制約が多くなってしまう。

(e)被写体にセンサを付け、それから得られた3次元座標を利用する方法[7][8]

これは、追尾カメラからの情報ではなく、それから独立したセンサの情報を利用する方法である。この方法では、画像から被写体が外れたとき追いつけられない問題は解決されている。しかし、被写体にセンサを付けるため、被写体に負担がかかる。大きさは得られないため、ズームの制御はあらかじめ決められた被写体の大きさから行わねばならず、正確性に欠ける。被写体にセンサを付けたとしても、3次元位置からだけでは、カメラを向けることはできても、正確なズームができる条件が狭い。

上記の問題をできるだけ解決した方法の提案を次章で行う。

3. 被写体追尾システムの提案

3.1. 提案手法

前章で述べた問題を解決するために、われわれが提案する基本的なアイデアは次のようなものである。

(1) 対象世界の把握のための装置の提案

カメラを操作するときには重要なのは、被写体がどこにいて、画角を決めるために被写体の大きさがどのくらいかを把握することである。実際のカメラマンの撮影方法を考えても、カメラのビューファインダを覗くだけでなく、視線を外に向けることでの被写体情報把握も頻繁に行っている[9]。ビューファインダ内の限られた情報だけでは被写体を追うことは非常に難しく、撮影範囲の全体的な情報を利用することで被写体情報を安定して正確に把握することができる。つまり、追尾カメラの画像とは別の装置から情報を得ることができれば、情報把握量は格段に増え、精度よく追尾できると考える。以上の考察から我々は、固定カメラ1台とレーザレンジファインダ（以下 LRF）を用いてシステムを構築することとした。

LRF は、照射したレーザーが、物体に当たって反射してくる時間を測定することにより、LRF と物体との距離を計測することができる。測定範囲は、LRF の中心から半円を描く水平面である。得られるデータは、距離と角度の2次元情報であり、角度 1° 程度の間隔で測定できる。これを用いることで被写体にマーカ、センサを付けることなく2次元上での位置がわかる。このため、被写体情報をリアルタイムに把握するのに適していると考えた。

2. の固定カメラは、被写体の大きさ、被写体の中心を計測するのに用いる。被写体の大きさ、中心とも実際の長さを求めることができるため、その後のカメラの制御が正確に行えるはずである。

(2) リアルタイムに被写体を効率よく抽出し、3次元位置を計算するアルゴリズムの提案

上記2つの装置から得られるデータは、まったく違う種類のものである。それらから、必要なデータを抽出し、被写体情報を計算するアルゴリズムを提案する。

(3) 被写体情報を元に、カメラの制御を行う方法の提案

得られた被写体情報からカメラを最適に制御するメカニズムを提案する。

3.2. 全体の流れ

全体の処理の流れは以下ようになる。

- ・ 準備

システムを動作させる前の準備として、追尾カメラ、固定カメラ、LRF を設置する。設置時の注意として、LRF のスキャン面が床と平行になるようにすること、カメラ雲台の水平を調整することがあげられる。その後、追尾カメラ、固定カメラのキャリブレーション、LRF のキャリブレーションを行う。被写体のいない状態で、固定カメラの背景画像、LRF の背景距離データ

を取得しておく。それぞれのキャリブレーション方法の詳細については第3章で述べる。

次にシステムの動作の流れを順に説明する。

1. データ取得

固定カメラからの画像と、LRF からの距離データを取得

2. 被写体抽出処理

LRF と固定カメラから被写体を抽出

3. 被写体情報計算

抽出したデータから、被写体の大きさ、位置を算出

4. カメラ制御データ作成、送信

追尾カメラのキャリブレーションデータと計算した被写体の中心からパン、チルトの値を算出し、被写体の大きさから、ズームの値、カメラ中心から被写体中心までの距離でフォーカスの値を算出し、制御データを作成、送信する。

以上4つの段階をループさせることで、各時点のカメラ制御を行う。

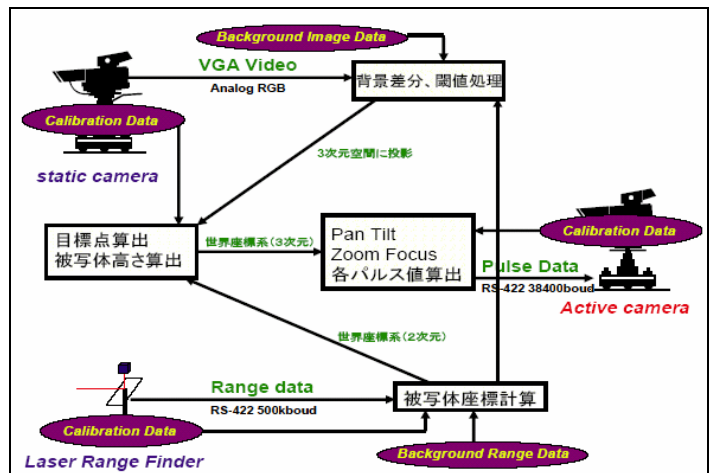


図 1 システムの流れ

3.3. 使用するハードウェアについて

3.3.1. 追尾カメラ

自動追尾を行うにはPC制御できる電動雲台が必要である。さらに、問題となるのはPC制御できることだけでなく、パン、チルト、ズーム、フォーカスの絶対位置制御ができることが挙げられる。本研究では、独自に開発した電動雲台を使用し、パン、チルト、ズーム、フォーカスについてそれぞれの速度と位置（パルス値）が指定できる。具体的には、パン、チルトは絶対位置、ズーム、フォーカスはリングの回転位置を制御することができる。

3.3.2. レーザレンジファインダ

LRF の原理は、レーザー光を照射し、それが物体に

当たって反射してくる間の時間を計測することで、その物体までの距離が取得できるというものである。本研究で用いる LRF は、内部の回転ミラーによってレーザー光を扇形に照射し、2次元空間のデータ（角度と距離のデータ）が取得できるものである。

4. 被写体情報抽出手法

前述した外部センサである、固定カメラ、LRF から得られるデータは、それぞれ画像データ、距離角度データと種類の違うデータである。この章ではこの2つのデータから被写体情報を抽出する方法について述べる。

4.1. レーザレンジファインダから被写体の抽出

LRF の被写体抽出処理は、差分処理で行う。あらかじめ背景距離データを取得し、そこで得られた値は、すべて壁であると考え、それより遠い位置に被写体は存在し得ないとする。また、背景距離データで囲まれた領域には、被写体以外は物体が侵入してこないこととする。よって、同じ角度上で、背景距離データの距離情報より小さい値が得られたら、それは被写体であると考えられる。この手法で、得られる値は複数あると考えられるので、角度、距離それぞれ平均をとる。そこで得られた角度、距離データを次の段階に送る。そのデータを角度 ϕ 、距離 r とする。

4.2. 固定カメラから被写体の抽出

固定カメラの被写体抽出処理は、基本的に背景差分処理で行う。理由は、簡単な処理で被写体領域が比較的正確に抽出できるからである。しかし、背景差分処理でもノイズが発生することがあり、確実に被写体領域を検出することは難しい。その解決策として、閾値処理をして差が激しい部分だけを抽出する、領域の大きさを計測しある大きさよりも小さいものは省くなどの方法が考えられる。しかし、閾値処理をして、差が激しい部分だけを抽出しても、今度は被写体領域が削られてしまう。また、大きさを計測する方法も、計算時間がかかってしまい、リアルタイム性を要求される自動追尾システムには適さない。そこで、LRF のデータ用い、固定カメラの画像中で被写体が映っている領域を計算し、差分する領域を狭め、ノイズの影響をできるだけ受けないようにした。これによって、画面全体を背景差分する方法よりも、被写体領域をできるだけ正確に取得できるようになった。

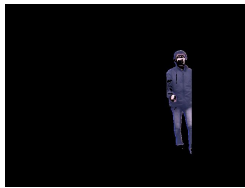


図 2 差分画像

4.3. 被写体情報の算出

前述した外部センサである、固定カメラ、LRF から得られるデータは、それぞれ画像データ、距離角度データと種類の違うデータである。我々は、2種のデータを、統一的な座標系である世界座標系での値に変換することで統合しようと考えた。このとき必要とされるのは、それぞれのデータを世界座標系での値に変換するための対応関係や諸パラメータの算出（キャリブレーション）である。はじめにそれぞれの機器のキャリブレーションについて述べる。

4.3.1. カメラのキャリブレーション

2次元のカメラ画像から実世界の3次元情報を得るには、実世界とカメラ画像が論理的、幾何学的にどのような関係にあるかがわからなければならない。本稿では、3次元復元や運動解析などでよく使用されているピンホールカメラモデルを用いる。

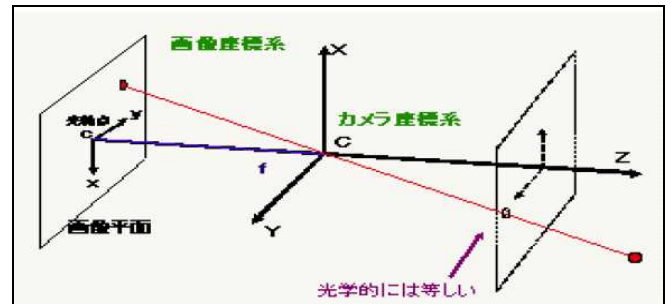


図 3 ピンホールカメラモデル

(x, y) を画像座標、 (X, Y, Z) を世界座標、 f を焦点距離とすると次式が成り立つ。

$$x = f \frac{X}{Z} \quad (1)$$

$$y = f \frac{Y}{Z} \quad (2)$$

また、カメラ座標系と世界座標系との関係は

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} = R \begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} + T \quad (3)$$

と表すことができる。ここで、カメラ座標系 (X_c, Y_c, Z_c) 世界座標系 (X_w, Y_w, Z_w) 回転行列 R 平行移動 T とする。レンズの半径方向の歪みを考慮した座標 (X_d, Y_d) は、

$$(1+k_1r^2+k_2r^4+\dots)x_d = x, \quad (1+k_1r^2+k_2r^4+\dots)y_d = y$$

ただし $r = \sqrt{x_d^2 + y_d^2}$ である。

画像座標系からカメラ座標系への変換は、

$$\mathbf{u} = S_x \cdot \mathbf{x}_d / d'_x + C_x \quad (4)$$

$$\mathbf{v} = y_d / d_y + C_y \quad (5)$$

ただし、 $d'_x = d_x \cdot x_d / N_x$

S_x …スケールファクタ (CCD 素子の縦、横の比)

$d_x d_y$ …受光素子の中心間距離 (mm)

N_x …1 走査線のサンプリング数

$C_x C_y$ …画像の中心の x,y (mm)

以上をまとめると

$$d'_x / S_x \cdot (\mathbf{u} - C_x) \cdot (1 + k_1 R^2 + \dots) = f \frac{r_1 X_w + r_2 Y_w + r_3 Z_w + T_x}{r_7 X_w + r_8 Y_w + r_9 Z_w + T_z} \quad (6)$$

$$d_y \cdot (\mathbf{v} - C_y) \cdot (1 + k_1 R^2 + \dots) = f \frac{r_4 X_w + r_5 Y_w + r_6 Z_w + T_y}{r_7 X_w + r_8 Y_w + r_9 Z_w + T_z} \quad (7)$$

この式には未知変数が含まれている。それを求めるために、今回は Tsai のキャリブレーション手法を用いた。[10]

4.3.2. レーザレンジファインダのキャリブレーション

LRF から得られるデータは、角度、距離のデータであるため、言い換えるとそれ自身が 2 次元座標系を持っているということである。この座標系と世界座標系との対応関係を求めるわけだが、世界座標系は 3 次元である。ここで以下の仮定を行う。

- LRF は、世界座標系の 1 平面と、LRF 自身を持つ 2 次元座標系の座標平面が平行になるように、設置されている。

こうすることにより、平行に置かれた面上の 2 次元座標同士の対応関係という問題に帰着できる。実際には、世界座標系を、床が XY 平面となるようにとり、としたとき、その XY 平面と、LRF の座標系の座標平面が平行になるように、すなわち、LRF の底面が床と平行になるように水準器などを用いて設置する。こうすることで、対応関係がわかれば、LRF から得られたデータは、世界座標系の XY 平面上の点に変換できる。ここで、2 次元座標同士の対応関係がどのように表されるかについて述べる。

$$Mw = R(\theta)M1 + T \quad (8)$$

$$\text{回転行列} \quad R(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

$$\text{平行移動} \quad T = \begin{bmatrix} T_x \\ T_y \end{bmatrix} \quad \text{世界座標系} \quad Mw = \begin{bmatrix} X_w \\ Y_w \end{bmatrix}$$

$$\text{LRF の座標系} \quad M1 = \begin{bmatrix} x_l \\ y_l \end{bmatrix}$$

上記の式を利用することで LRF の座標系から世界座標系の変換ができる。この式を利用するために、未知変数を求める (LRF のキャリブレーション)。未知変数は、回転角度 θ と平行移動 T である。カメラのキャリブレーションと同様に、目標物体を計測することにより、その計測値と世界座標系の座標のいくつかの対応組から算出する方法を使用する。具体的方法としては、目標物体を用い、異なる 3 点を計測することによって、世界座標系と測定値の対応組 3 つから算出する。ここで 3 組とした理由は、誤差を少なくすることと、測定にかかる作業量とを勘案して、提案するものである。

まず、回転角度 θ を求める。誤差を少なくするために、3 つの対応組すべての組み合わせで、測定点間の距離を算出し、最も距離の長い 2 組を利用する。最も距離の長かった 2 組を i, j とする。

$$\text{世界座標系での点を} \quad \begin{pmatrix} X[i] \\ Y[i] \end{pmatrix}, \begin{pmatrix} X[j] \\ Y[j] \end{pmatrix}$$

$$\text{LRF のデータを} \quad \begin{pmatrix} x[i] \\ y[i] \end{pmatrix}, \begin{pmatrix} x[j] \\ y[j] \end{pmatrix}$$

とする。

回転角度 θ は以下の式で求まる。

$$\theta = \arctan \left(\frac{Y[i] - Y[j]}{X[i] - X[j]} \right) - \arctan \left(\frac{y[i] - y[j]}{x[i] - x[j]} \right) \quad (9)$$

次に、平行移動 T を算出する。式(4.12)を変形して

$$T = Mw - R(\theta)M1 \quad (10)$$

ここで Mw、M1 に測定したデータを代入するのであるが、代入するデータは、誤差を少なくするために、LRF に最も近い測定点の組の値とする。この組の値と、先ほど求めた θ を代入することにより、T を求める。

4.3.3. 被写体情報算出

まず、4.1 で得た LRF の情報の意味について考える。得られたデータは、2 次元座標である。これを式 (10) に代入することにより、世界座標系の 1 平面上の点 (床を XY 平面とすると、XY 平面上の点) として変換することができる。これは、被写体 (正確には被写体のほぼ中心) が、その点からその点が存在する平面に垂直な方向に伸ばした直線上のどこかに存在している、ということの意味している。この直線を直線 1 とする。一方、固定カメラの画像データは、式 (6) (7) により世界座標系に変換できる。このとき 2 次元の画像上で 1 点が決まると、3 次元の世界座標系に 1 本の直線が決定する。この直線を直線 m とする。固定カメラの画像すべてのピクセルについて、世界座標系に投影し、直線 1 と直線 m の最小となる距離を算出す

る。被写体のほぼ中心が、直線 1 上のどこかに存在することがわかっているから、算出した距離がある一定の範囲に入っているものだけを背景差分処理する。ここで差分のあるピクセルを被写体であると認識する。こうすることによって、背景差分する領域を狭め、ノイズの影響を受けにくくする。また、差分のあったピクセルについては、最小となる距離を構成する直線 1、直線 m 上の 2 点の 3 次元座標を算出する。被写体領域とされるすべてのピクセルについてこれを求め、直線 1 上の点の Z 座標の最小値、最大値を算出する。(最大値-最小値)を被写体の大きさとする。また、LRF から変換した世界座標系での 2 次元座標 (X Y 座標) と (最大値-最小値) ÷ 2 + 最小値 を加えた 3 次元座標を被写体の位置とした。

また被写体までの距離は、世界座標系における追尾カメラのカメラ座標の原点座標と上記の点間の距離で求めた。

5. カメラ制御手法

ここでは、得られた被写体情報から、追尾カメラの制御データを算出する手法を述べる。追尾カメラは 4.3.1 節の方法でキャリブレーションされている。

5.1. パン、チルト

追尾カメラは、4.3.1 節の方法でキャリブレーションされており、世界座標系とカメラ座標系との対応関係は既知である。そのとき、世界座標系で 1 点 A が決まると (この点を求める方法は後述) 図 4 のように投影される。このとき、空間上の直線 (直線 1 とする) と、光軸であるカメラ座標系の Z 軸が重なることで、画像の中心に点 A を映しこむことができる。

ここで、以下の仮定をする。

- ・ パンはカメラ座標系の X Y 平面に沿って動く。
- ・ チルトはカメラ座標系の Y Z 平面に沿って動く。
- ・ 雲台の回転中心は、カメラ座標系の原点と等しい。
- ・ パン、チルトは 0.0072deg/pulse であり、誤差はない。

上の仮定をすることにより、直線 1 とカメラ座標系の X Y 平面との角度からパンの移動すべき角度、直線 1 とカメラ座標系の Y Z 平面との角度からチルトの移動すべき角度が求まる。ここで得られた角度を 1 パルスあたりの移動角 0.0072deg で割ることにより、パン、チルトそれぞれのパルス値が求まる。

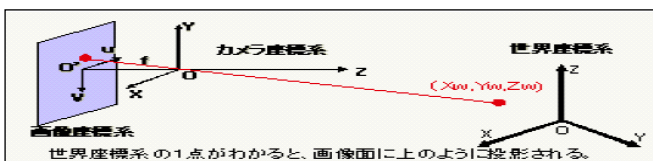


図 4 世界座標系とカメラの対応関係

$$\begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} \quad \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix}$$

世界座標系

カメラ座標系

$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} = R \begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} + T \quad (11)$$

パンパルス値 pan_pulse チルトパルス値 tilt_pulse
追尾カメラのキャリブレーションはパン 0 pulse チルト 0 pulse の初期位置で行われたとする。

世界座標系での点を追尾カメラのカメラ座標系に直す

$$\text{pan_pulse} = \arctan\left(\frac{X_c}{Z_c}\right)/0.0072 \quad (12)$$

$$\text{tilt_pulse} = \arctan\left(\frac{Y_c}{Z_c}\right)/0.0072 \quad \arctan \text{の単位は deg とする} \quad (13)$$

5.2. ズーム

ズームを制御するには、まず、レンズの構造把握が必要である。しかし、我々が実際に使用するレンズは、複数のレンズから構成されていて、光学的に複雑な構造をしている。これを、そのままモデル化するのは非常に困難である。よって 4.3.1 節で述べたピンホールカメラの構造に近似して考える。

図 5 で示したように、レンズを介してカメラ内部 (焦点から画素) と実世界 (焦点から被写体) に相似な三角形ができる。この三角形の相似関係を利用し次式で焦点距離を求め、画角を決定することができる。

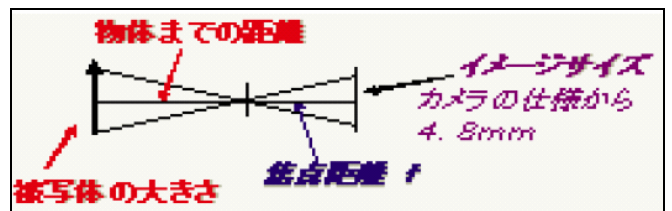


図 5 ズーム制御

$$f[\text{mm}] = \frac{(\text{レンズから物体までの距離}[\text{mm}]) \times (\text{イメージサイズ}[\text{mm}])}{\text{被写体の高さ (幅)}[\text{mm}]} \quad (14)$$

ここで必要なのは、

- ・ レンズから物体までの距離
- ・ 被写体の高さ (幅)
- ・ イメージサイズ

である。イメージサイズは、カメラの仕様から決定した。

ズームを制御するためには、適切なズームパルス値を算出しなければならない。そのため、焦点距離とズ

ームパルスの対応付けを行い、焦点距離からズームパルスを算出できるようにした。フォーカス、ズームを一定間隔で移動させて、それぞれのパルス値で Tsai のカメラキャリブレーションを行い、焦点距離の算出対応付けを行った。

これで得られた値から、今回は、フォーカスパルスを1つ決め、ズームパルスと焦点距離の対応付けは近似曲線を用いることでおこなった。

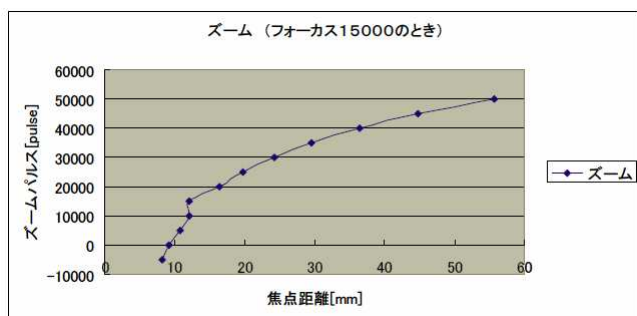


図 6 ズームパルスの対応付け

5.3. フォーカス

フォーカスの動きは、レンズの仕様により、レンズから被写体の距離に関係している。このため、距離とフォーカスパルスの対応付けができていれば、被写体の位置からカメラ座標系の原点までの距離を算出することにより、フォーカスパルスを導き出すことができる。この際、ズームと同様に、フォーカスパルスと距離の対応付けも行わなければならない。本研究では、目標物を実際に映し、最もズームした状態で、フォーカスを目視で合わせ、そのときの目標物とカメラとの距離を測定し、対応付けを行った。

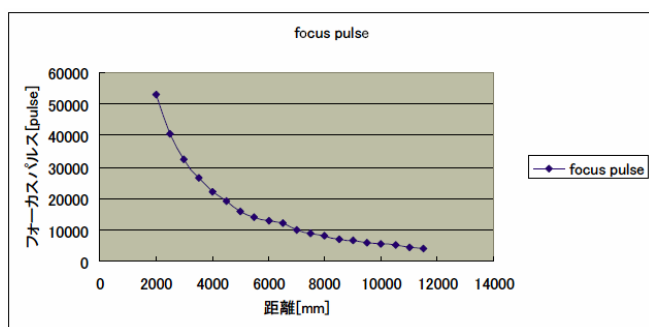


図 7 距離とフォーカスパルスの対応付け

6. 実験結果

提案したシステムを実際に PC に実装し動作の検証を行った。使用した PC 環境は次のとおりである。

CPU : pentium4 3.2GHz

メモリ : 1GB

OS : WindowsXP Professional Edition

開発環境 : Microsoft Visual Studio .net 2003

実験した空間は 1.2m × 6m の部屋で行った。得られた結果画像の一部を示す。

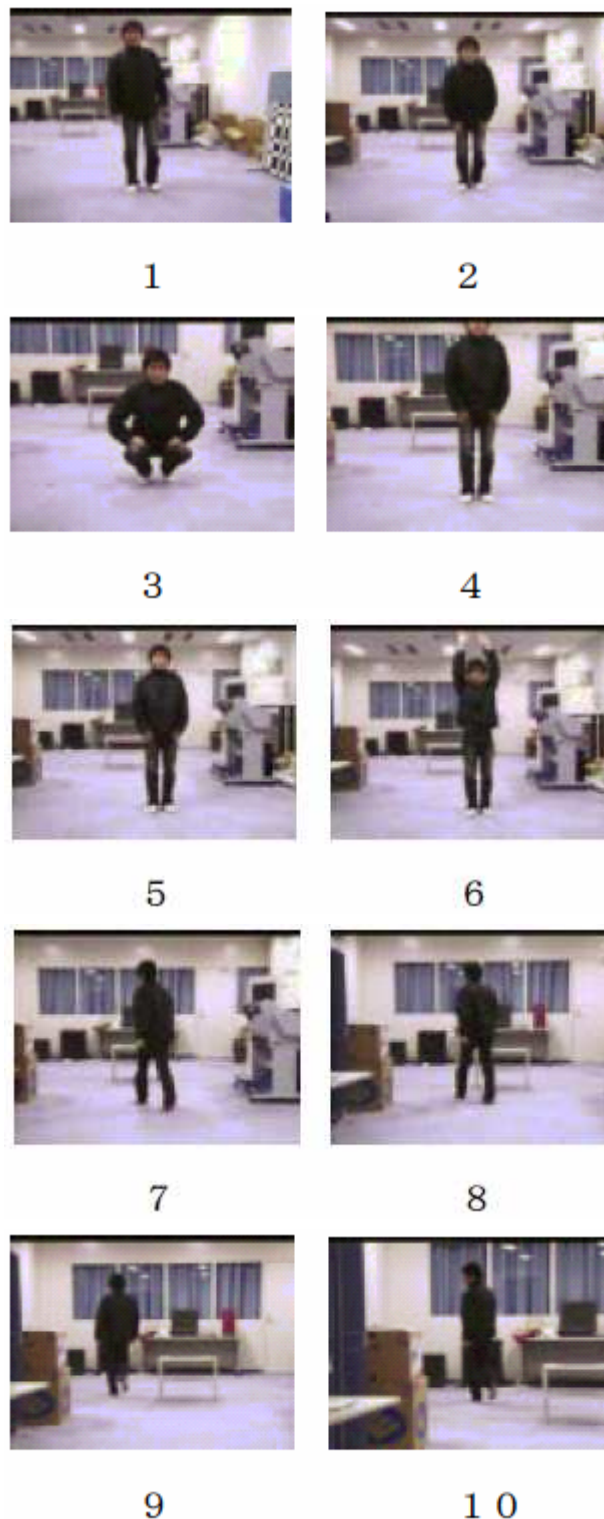


図 8 結果画像

図 8 からほぼ被写体を追えていることがわかる。し

かし、図8の4番目の画像では画面から外れてしまった部分があることがわかる。また、ここでは示せないがカメラの動きが滑らかでなかった。これについての考察は次章で行う。

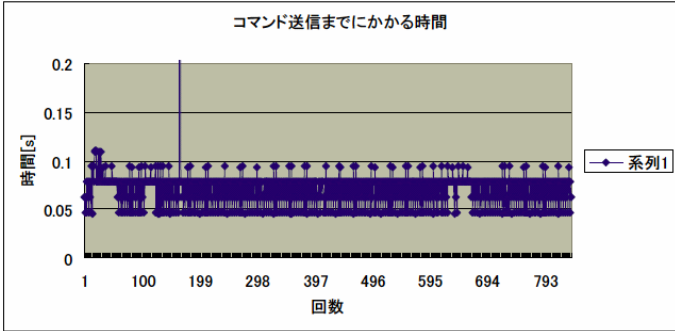


図 9 計測時間

計測時間については、ほぼ0.1秒以下で1回のループが終了している、これは、1秒間に10回コマンドを送信できることになる。

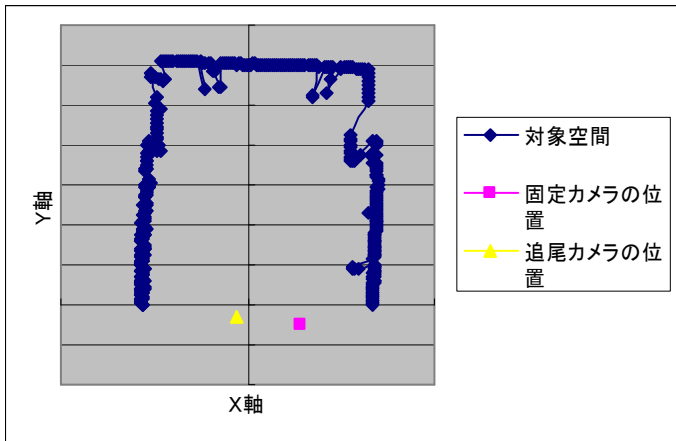


図 10 対象世界のモデル化

図10はLRFのデータから計算機内に作成された対象世界のモデルを可視化したものである。前述の方法でカメラのキャリブレーションデータと統合することにより、対象世界の把握ができる。

7. まとめ

撮影対象となっている世界をすばやく把握し、世界モデルを計算機内に構築すると同時に、その世界で注目すべき被写体を発見して、カメラのパン、チルト、ズーム、フォーカスを制御し、自動的に特定の被写体を追尾する装置を提案、実験、考察を行った。

現段階での問題点を整理する。

- ・ 動きが滑らかでない
- ・ 早い動きをした場合に画面から外れてしまうことがある
- ・ 被写体の動作範囲は、固定カメラ、LRF 両方で観測できる部分に限られる

- ・ 被写体1人に限られている点
- ・ 被写体の大きさは床に垂直な方向のみから計測している点

上の2つの原因は、コマンド送信までの計算時間と、ハードウェアの応答速度が挙げられると考えられる。3つ目の問題は、カメラの台数を増やす、LRFを増やすなどの解決策が考えられるが、機材の増加という問題との兼ね合いで考えねばならない。4つ目の問題は、被写体の認識の問題がある。固定カメラの色情報などを使い区別する方法が考えられる。5つ目の問題は、画角を決定するための情報の取得問題である。直立、歩行、しゃがむなどの姿勢、動作には、前述のような精度で制御できた。しかし、床に寝転ぶなどの姿勢、動作には対応できない。現在のデータ統合部分の方法の拡張などが求められる。以上が今後の課題である。

文 献

[1] H.Arisawa “Real World Database System - Design and Implementation “ADVANCED TECHNOLOGY AND SCIENCE, Proceedings of THE JAPAN-KOREA JOINT Symposium, pp.35-41, 1996.10

[2] 堀口昌彦、竹内義則、大西昇 “滑らかなパンチルトによる移動物体追跡”、電子情報通信学会技術研究報告 IE、画像工学 Vol.99 Num.610 pp.43-48 (2000.02)

[3] 森田俊彦 “局所相関演算による動きの検知と追跡”、電子情報通信学会技術研究報告 PRMU,パターン認識・メディア理解 Vol.100 Num.44 pp.55-62 (2000.05)

[4] 波田尚哉、三宅哲夫 “アクティブビジョンシステムによる遮へいを伴う移動物体の追跡”、電子情報通信学会論文誌 D-II Vol.J84-D-II No.1 pp.93-101 (2001.1)

[5] 松山隆司、和田俊和、物部祐亮 “視点固定型パン・チルト・ズームカメラを用いた実時間対象検出・追跡 (<特集>画像の認識・理解)”、情報処理学会論文誌 Vol.40 No.8 pp.3169-3178 (1999.08)

[6] 三塚和幸、山村恵一、山中徳唯 “インテリジェンスロボットカメラの開発”、テレビジョン学会技術報告 Vol.17 No.51 pp.33-37 (1993.9)

[7] 東和秀、亀田能成、美濃導彦 “パン・チルト角速度及びズーム制御による動物体の追跡撮影手法”、第6回知能情報メディアシンポジウム 予稿論文集 pp101-106

[8] 畠山祐里、加藤大一郎、津田貴生、井上誠喜 “知的ロボットカメラの被写体検出に関する一検討”、電子情報通信学会ソサイエティ大会講演論文集, Vol.2001年.情報・システム Num. pp.210 (2001.08)

[9] 加藤大一郎、山田光穂、阿部一雄、石川秋男、尾幡昌芳 “スポーツ中継でのTVカメラの動特性と放送カメラマンの視線の動き”、電子情報通信学会技術研究報告 IE、画像工学, Vol.96 Num.116 pp.123-128 (1996.06)

[10] 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, RA-3(4) pp.323-344 (1987.08)