

2台の非同期カメラによる仮想高速ビジョンシステムの実現

Pseudo High Speed Vision System with Two Unsynchronized Cameras

清水彰一[†] 西 貴行[†] 藤吉弘巨[†] 長坂保典[†] 高橋友一^{††}

Shoichi Shimizu, Takayuki Nishi, Hironobu Fujiyoshi, Yasunori Nagasaka, and Tomoichi Takahashi

[†] 中部大学, ^{††} 名城大学

Chubu University, Meijo University

shiyou@vision.cs.chubu.ac.jp

Abstract

In order to control robots precisely and rapidly, an expensive vision system with 60 fps has been used. In this paper, we present a vision system with two normal unsynchronized cameras, that performs as fast as the expensive vision system. Frame grabbers for each camera are installed on a PC. Images from the camera are processed at 30 fps. Our two camera vision system processes robots positions as fast as the expensive vision system with 60 fps and gives better accuracy in position prediction than a single vision system.

1 はじめに

ロボットが行動するためには、視覚センサによる環境認識・理解技術による情報を基に行動プランニングを行いロボットを制御するビジュアルフィードバック機構が必要である。ロボットを迅速かつ正確に制御するためには、このビジュアルフィードバックのサイクルを高速に行う必要がある。特に、カメラから得られる画像から有用な情報を得るための高速なビジョンシステムが不可欠となる。その際には、ロボット等の対象とする物体の位置をより正確に検出する必要がある。

これまでにビジョンシステムの高速化として、1) 画像処理のハードウェア化による手法[Akita, 2000], 2) 高速な画像処理アルゴリズムの提案[Bruce, 2000], 3) 60fpsの高速カメラを用いる手法[D'Andrea, 2002, 加藤, 2003]が提案されている。1), 2)の手法は、カラー解析や領域検出等の画像処理アルゴリズムをハードウェア化や最適化することによる処理の高速化である。これらは、ビジュアルフィードバックにおけるビジョン処理にかかる時間を短くするものであり、30fpsのカメラを用いた場合それ以上のサイクルの高速化は期待できない。3)の手法は、60fps

のカメラを用いたダブルバッファリングによる高速処理を行うものである。この場合、カメラと専用のフレームグラバが高価であるという問題がある。

本報告では、通常のカメラ(30fps)を複数用いることにより、ビジュアルフィードバックのサイクルを高速化することを目的とする。このとき、複数のカメラを効果的に配置することで、オクルージョン領域の減少、画像の高解像度化等に対して大きな効果がある。このような複数カメラ群のシステムの取り組みとして分散協調視覚プロジェクトが挙げられる[松山, 1998]。中でも分散協調型対象追跡システムでは、多数の自律したエージェント間のコミュニケーションにより対象の追跡を実現している[中澤, 2001]。しかし、これらの多くは処理サイクルの高速化についてはまだ検討されていない。

我々は、分散協調型ロボットビジョンシステムの一アプローチとして、通常非同期カメラ(30fps)2台を用いた仮想高速ビジョンシステムについて提案する。2台のカメラは、1台のPCに接続されており、それぞれ30fpsで処理される。それぞれのカメラの処理結果を統合することで擬似的に30fps以上の結果を得ることが可能となる。

本報告では、2台の非同期カメラによる高速ビジョンシステムを提案し、物体の運動を予測する実験により、対象物の位置を高速カメラと同等のスピードで得ることができ、その位置の精度は1台のビジョンシステムよりも良いことを述べる。

2 2台の非同期カメラによるビジョンシステム

高速なビジュアルフィードバック機構を必要とするタスクとして、RoboCup小型リーグにおけるグローバルビジョンシステムが挙げられる[RoboCup, 2003]。グローバルビジョンシステムは、サッカーフィールド上の設置されたカメラ映像からロボットの位置、姿勢、IDとボールの位置を抽出し行動を決定するものである。本研究では、RoboCup

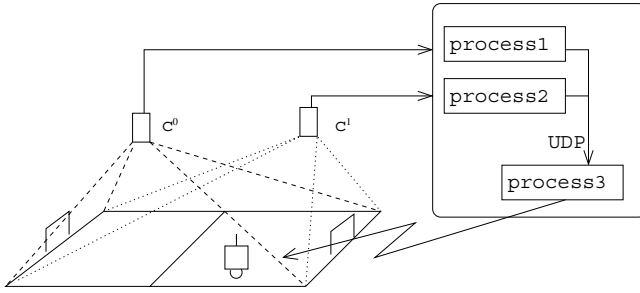
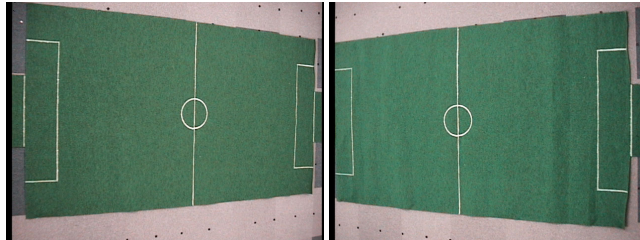


Figure 1: ビジョンシステムの概略



(a) c^0 からの映像

(b) c^1 からの映像

Figure 2: 2台のカメラ画像

小型リーグのグローバルビジョンシステムにおけるボールの運動復元を対象とする。

2台の非同期カメラ c^0 と c^1 を用いたビジョンシステムの概略を Figure 1 に示す。各々のカメラは、両ゴールの上部高さ 2,800[mm] にサッカーフィールド全体が視野に入るように固定されている。サッカーフィールドの大きさは、縦横約 2,000 × 3,000[mm] である。カメラからの映像信号は、1台の PC にインストールされた2枚のフレームグラバに入力され、それぞれの画像が取り込まれる。Figure 2 に各々のカメラ画像を示す。本ビジョンシステムの処理は、各カメラ画像内のオレンジ色のボール領域を検出し、その中心に対応するフィールド上の位置を求める。

2.1 位置推定

ロボットを制御するシステムにおいて、その制御領域は3次元である。画像処理の結果から得られたデータは2次元であり、その2次元情報からロボット等の、3次元情報を求める必要がある。これにより、行動計画を3次元で考えることができるため、制御システムを実空間に合わせて容易に作成することができる。

RoboCup 小型リーグでは、ロボットやボール位置を推定する手法に、カメラパラメータを用いる手法や、ホモグラフィを用いる手法[Ryad, 2001]が提案されている。カメラパラメータを用いる手法では、ピンホールカメラに基づく幾何学キャリブレーションを必要とする。単一のカメラに関する内部・外部パラメータを推定する手法として Tsai のキャリブレーション手法[Tsai, 1987]が多く用いられている。

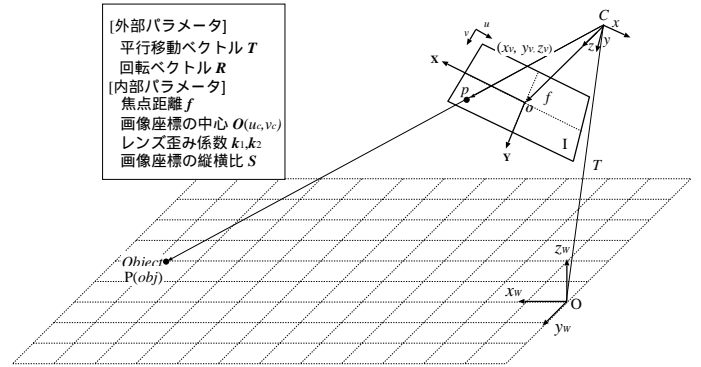


Figure 3: 視線交差によるフィールド上の位置推定

本システムでは、Tsai の手法で求めたカメラの内部・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。Figure 3 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、その直線がフィールドの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) に対応したフィールド上の世界座標 (X, Y, Z) を求めることが可能である。このとき我々のシステムでは、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) 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。

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

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 Figure 3 において、カメラ座標の原点から画像上の任意点 p を通過する視線ベクトル (x_v, y_v, z_v) は次式となる。 f はカメラ座標における焦点距離とする。

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

Step4 視線ベクトルを3次元空間上の直線式で表すと、

次式となる．

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

Step5 ここでフィールド面が $x_w - y_w$ 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ときの 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．

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

本手法では，Step5 における z_w の値として，ロボットやボールの高さを与えることにより，高さの異なる各対象物体の 3 次元位置を正確に求めることができる．

2.2 位置推定の誤差

前節の手法でカメラ校正時に用いたフィールド上の特徴点 ($c^0 = 98$ 点, $c^1 = 95$ 点) の位置を推定した結果，その位置推定誤差は平均 4.0[mm]，最大誤差は 17.8[mm] であった．フィールド上のオブジェクト obj の位置をカメラ c^i の画像を用いて算出したものを， $\mathbf{p}^i(obj) = (x^i(obj), y^i(obj))$ とする．ボールを片方のゴールからもう一方のゴールまで約 3.25[m/s] の速度で移動した際の軌跡を Figure 4 に示す．Figure 4 の横軸と縦軸はフィールドのサイドラインを示す．白丸はカメラ 0，黒丸はカメラ 1 より検出されたボールの位置を推定した結果である．カメラ校正時に用いたフィールド上の特徴点における位置推定誤差を求めることはできるが，それ以外の点に関しては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．本報告では，Figure 5 のように，画像座標 (u, v) が 4 近傍に 1 画素ずれた際の対応するフィールド上の世界座標点 (x_w, y_w) 間のユークリッド距離を推定誤差とする．

Figure 4 における三角形は，各カメラにおいてボールの位置を推定した際の誤差 $e^i(obj) = (e_x^i(obj), e_y^i(obj))$ を表している．位置推定誤差は，3.5[mm] から 6[mm] の範囲で各カメラから離れるほど増加していることが分かる．これにより，本システムでは，あらかじめ各カメラの画像座標における世界座標での誤差を計算できるため，位置に応じた誤差を行動プランニングの際に不確定要素の大きさとして扱うことが可能である．

2.3 2 台のカメラシステムの特徴

2 台のカメラを用いたビジョンシステムの特徴は，(a) オクルージョン領域の減少，(b) 各々カメラの位置推定結果を合成することによる誤差の減少，(c) ビジョンの高速化の三つが挙げられる．

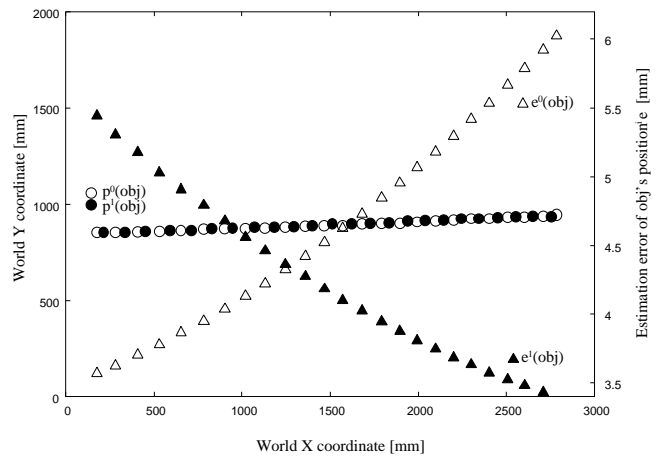


Figure 4: ボールの移動軌跡と位置推定誤差

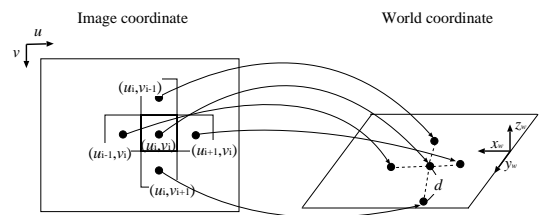


Figure 5: 推定誤差の計算法

最初の特徴は，同じ領域を異なる方向から複数のカメラで捉えることで，オクルージョン領域を減らすことが可能となる．

2 台のカメラを用いて位置推定した際の誤差がガウス分布であり，互いに相関がないと仮定すると，静止した対象物体 obj の位置 $\mathbf{p}(obj)$ とその誤差 e は次式により計算される．

$$\mathbf{p}(obj) = \frac{w_0}{w_0 + w_1} \mathbf{p}^0(obj) + \frac{w_1}{w_0 + w_1} \mathbf{p}^1(obj), \quad (8)$$

$$e^2 = \left(\frac{w_0}{w_0 + w_1}\right)^2 * e^{0^2} + \left(\frac{w_1}{w_0 + w_1}\right)^2 * e^{1^2}, \quad (9)$$

$$w_i = 1/e^{i^2}$$

合成した誤差 e は，それぞれのカメラの位置推定誤差 e^0 と e^1 より小さくなる．Figure 6 は，同時刻における各カメラのボール位置推定誤差を式 (9) により合成した結果を示す．

上記の二つの特徴は複数のカメラを用いた特徴としてよく知られているものである．本報告では，三つ目の特徴であるビジョン処理の高速化に関する手法について提案する．通常のカメラシステムでは 30fps で画像を取り込むため，ビジョン処理された結果は 30fps を超えることはない．

そこで，本システムでは 2 台のカメラの結果を統合することでビジョンの高速化を行う．本システムは，2 枚のフレームグラバを 1 台の PC 上に挿入したものでカメラ

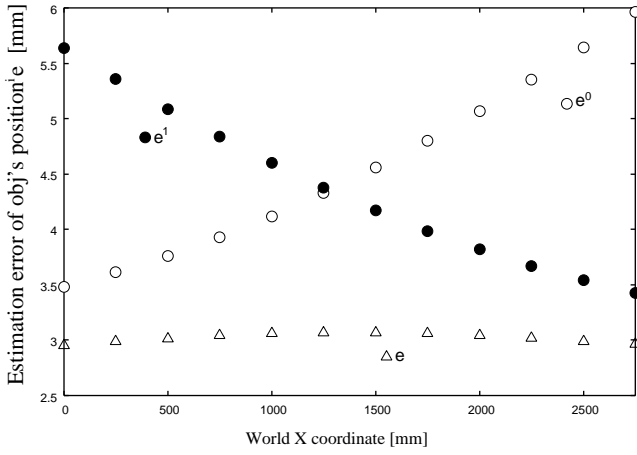


Figure 6: 合成誤差

間の同期機構を持たない．Figure 4 におけるカメラ 0 による結果 (白丸) とカメラ 1 の結果 (黒丸) は交互にプロットされており、これは 1 台のカメラシステムよりも早い周期でボールの等速直線運動を捉えていることを意味する．特別なカメラ同期機構を持たないでより高速な処理結果を得ることが、2 台のカメラを用いる三つ目の特徴であり、本報告で提案する仮想高速ビジョンシステムの特徴である．

3 高速な動き予測

2 台のカメラは外部同期を使用しない場合、それぞれのキャプチャタイミングは異なる．本報告では、この異なりにおける 2 台の非同期カメラによる仮想高速ビジョンシステムの特徴について検討する．

3.1 位置データの統合

Figure 1 における process-1 と process-2 はそれぞれのカメラからの画像を 1/30 秒毎に処理する．各プロセスは、カメラ c^i の処理結果である物体の位置 $p^i(obj)$ と、その処理した画像のキャプチャ終了時刻を UDP 通信により process-3 に送信する．process-1 と process-2 は、1 台の PC 上で実装されているため、それぞれのプロセスからキャプチャ終了時に参照する時刻のずれは生じない．process-3 では、process-1 と process-2 から送信された結果をキャプチャ終了時刻で整合し、その情報を基にロボットを制御する．

2 台のカメラのキャプチャタイミングが同じである場合、Figure 4 におけるカメラ 0 による結果 (白丸) とカメラ 1 の結果 (黒丸) は重なる．process-1 と process-2 の時間間隔が通常のビデオレートの半分 (1/60 秒) の場合、process-3 は 1/60 秒毎に位置データを交互に受け取ることができる．

Figure 7 に process-1 と process-2 のタイミングチャートを示す．ここで、 δ_{12} を process-1 から process-2 までの時間間隔、 δ_{21} を process-2 から次の process-1 までと

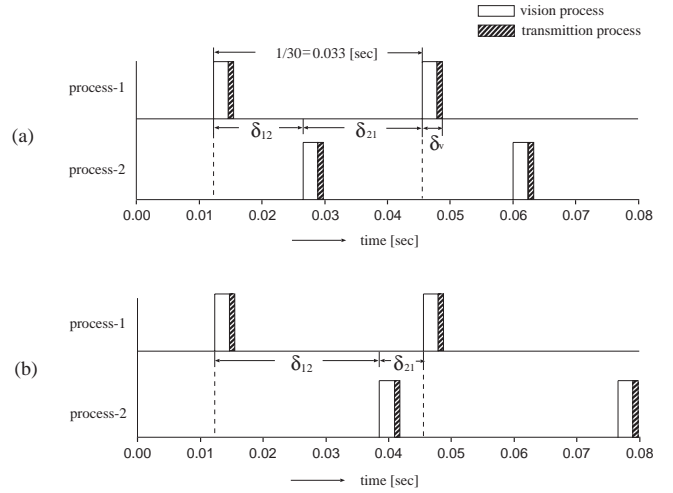


Figure 7: process-1process-2 のタイミングチャート

する．Figure 7 における (a) の場合は、2 台のカメラのキャプチャタイミングがほぼ等間隔 ($\delta_{12} \approx \delta_{21} \approx \frac{1}{60}$) に異なった例である．(b) の場合は、キャプチャタイミングが近い例 ($|\delta_{12} - \delta_{21}| \approx \frac{1}{30}$) である．本システムは、Dual processor Xeon の PC 上に実装され、ビジョンプロセスに 0.0026 秒、UDP 通信に 0.00018 秒を費やしている．

本システムでは、 δ_{12} と δ_{21} を等間隔 ($\delta_{12} = \delta_{21}$) にするためのカメラの外部同期機構を持たない．従って、タイミングはそれぞれのカメラのシャッタータイミングに依存している．本システムでは、画像のキャプチャとビジョンプロセス δ_v が δ_{12} と δ_{21} よりも短い時間で処理を行う必要がある．この処理時間が満足されないと、二つのカメラシャッターは同期した状態に近くなり、2.3 で述べた特徴 (b) の合成による位置精度の向上のみに使用されることになる．

3.2 線形予測による動き推定

ロボットを正確に制御するためには、対象物の動きを予測して制御する必要がある．得られた最新のボール位置から次の予測位置 $p_{t+1}(obj)$ を求める式を以下に示す．次式は、カメラ 0 の画像から計算された位置が最新の場合である．

$$\begin{aligned} p_{t+1}(obj) &= p_t^0(obj) + k(\delta_{12})v_t(obj), \\ v_t(obj) &= (p_t^0(obj) - p_{t-1}^1(obj))/\delta_{21} \end{aligned} \quad (10)$$

予測位置 $p_{t+1}(obj)$ は、現在と一つ前の結果より物体 obj の速度 $v_t(obj)$ を求め、次の位置を線形に予測するものである．カメラ 1 の結果が最新の場合は、上記の式の添字が入れ替わる．

3.3 実験

等速円運動における動き予測の実験として、Figure 8 に示すようにターンテーブルとボールを用いた．長さ 1000[mm]

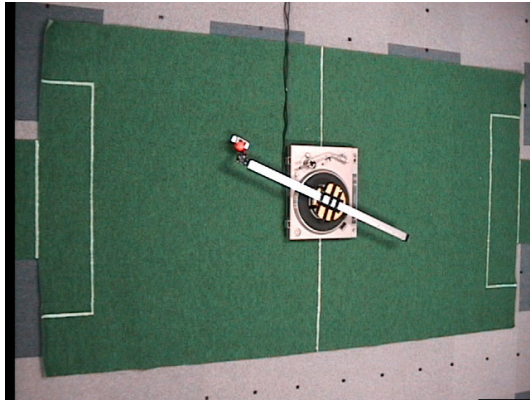


Figure 8: ターンテーブルを用いた実験の様子

の物差しの端にボールを固定し，ターンテーブルを回転させることで半径 500[mm] の等速円運動となる．ターンテーブルの回転数は 45[rpm] であるため，ボールの回転速度は $\frac{45}{60}2\pi = 0.478[\text{rad/s}]$ ，また，速度は 1023.7[mm/s] となる．

3.3.1 等速円運動の復元

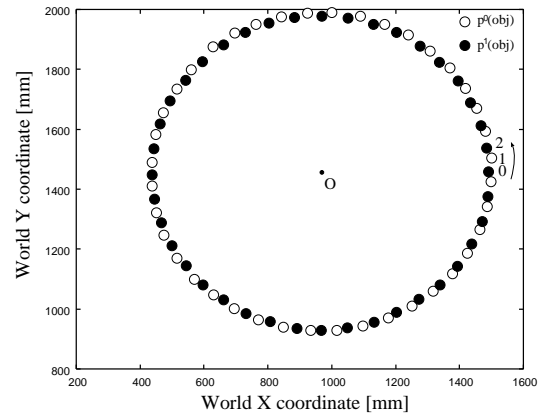
Figure 9 にボールの等速円運動における検出した結果の軌跡を示す．各プロット点は，時系列の順に 0, 1, ... とラベル付けされている．白丸 (奇数) はカメラ 0 のボールの位置推定結果，黒丸 (偶数) はカメラ 1 の結果を示す．

60fps の高速カメラを用いたビジョンシステムの場合，ボールが一周するには 1.33 秒かかるため， $1.33 \times 60 \approx 80$ 点のデータが得られることになる．Figure 9 は一周の実験結果であり，80 個のデータがプロットされている．これより，本ビジョンシステムのサイクルは高速カメラ (60fps) と同等のスピードを実現できていることを示している．

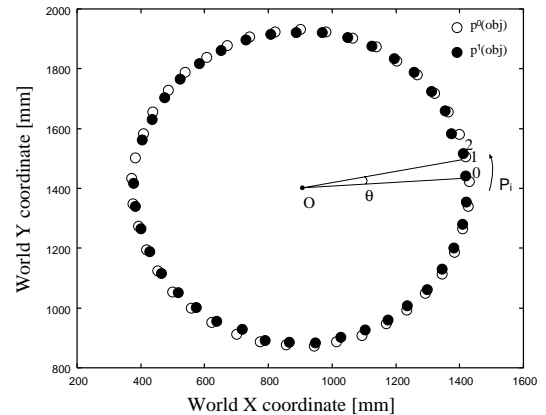
60fps で処理した場合，連続した 2 点 P_i と P_{i+1} の回転角度 $\angle P_i O P_{i+1}$ の理論値は 0.0785[rad] となる．Figure 10 は，回転角度 $\angle P_i O P_{i+1}$ の分布である．2 台のカメラの時間間隔が等間隔 ($\delta_{12} \simeq \delta_{21} \simeq \frac{1}{60}$) の場合，回転角度の平均は 0.0783[rad] であり 60fps で処理した際の理論値に近いことが分かる．角速度は， $\angle P_i O P_{i+1} / \delta$ で求められる．角速度の平均は， $\delta_{12} \simeq \delta_{21} \simeq \frac{1}{60}$ の場合 4.70[rad/s]， $|\delta_{12} - \delta_{21}| \simeq \frac{1}{30}$ の場合 4.50[rad/s] であった．その際の分散は前者が 1.06，後者が 1.52 であり，間隔が $\delta_{12} \simeq \delta_{21} \simeq \frac{1}{60}$ と等間隔である場合，より精度のよい円運動の復元が可能であるといえる．

3.3.2 動き予測

等速円運動するボールにおける次の予測位置を式 (10) より求めた．その際の予測誤差を Figure 11 に示す．予測誤差は，予測した次の位置と実際に観測された位置とのユークリッド距離である．グラフ中の + は，2 台のカメラシステムによる予測誤差，* は 1 台のカメラシステムによる予測誤差である．Figure 11 より，1 台のカメラシステ



(a) Equally separated



(b) Not Equally separated

Figure 9: 等速円運動の復元

ムよりも 2 台のカメラシステムの方が平均予測誤差が小さいことが分かる．Table 1 にカメラの時間間隔が等間隔である場合を A，そうでない場合を B としたときの予測誤差の平均と分散を示す．平均は時間間隔に関係なくほぼ同じであった．しかし，分散は等間隔の方が小さく，2 台のカメラシャッターが 1/60 秒ずれているとより精度のよい予測が可能であるといえる．

4 まとめ

本報告では，ロボットを迅速にかつ正確に制御するために，通常の非同期カメラ (30fps) 2 台を用いた仮想高速ビジョンシステムについて提案した．2 台のカメラは，1 台の PC に接続されており，それぞれ 30fps で処理される．

Table 1: 予測誤差 [mm]

		two cameras	single camera	
			c0	c1
A	Ave	9.4	23.8	22.8
	Var	23.2	17.0	10.4
B	Ave	10.2	22.2	21.8
	Var	42.3	10.3	36.4

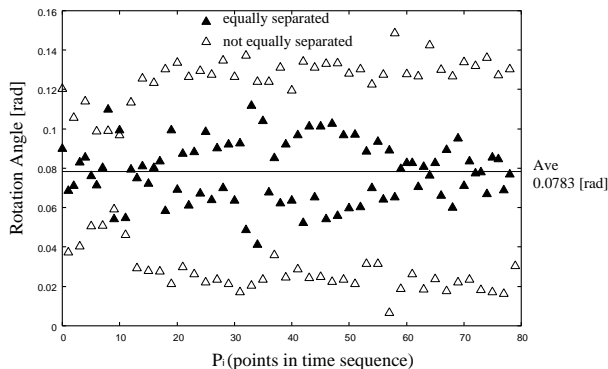


Figure 10: 回転角度の分布

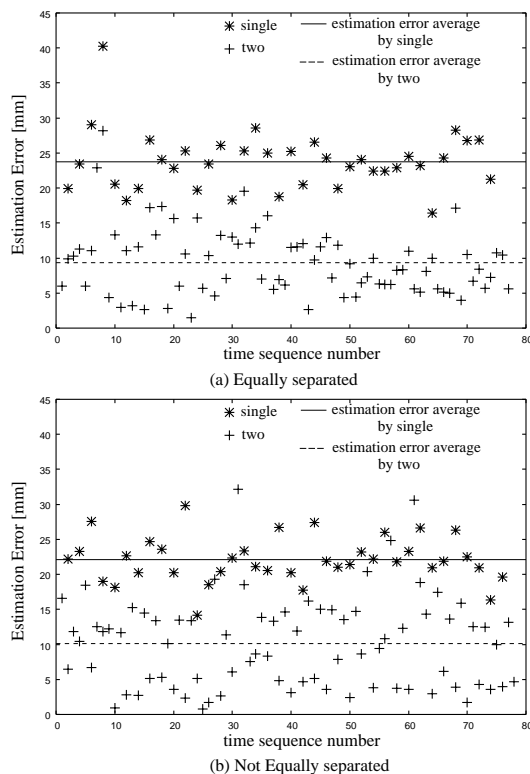


Figure 11: 予測誤差

それぞれの処理結果を融合することで擬似的に 30fps 以上の結果を得ることが可能となる。対象物の次の位置を予測する際に、2 台のカメラシャッターが約 1/60 秒ずれているとより精度のよい予測が可能であることが分かった。

参考文献

- [Akita, 2000] J. Akita.: Real-Time Color Detection System Using Custom LSI for High-Speed Machine Vision, in *RoboCup-99: Robot Soccer World Cup III*, Springer, pp128-135, 2000.
- [Bruce, 2000] J. Bruce, T. Balch, M. Veloso: Fast and Inexpensive Color Image Segmentation for Interac-

tive Robots, In Proceedings of IROS-2000, Japan, 2000.

[D'Andrea, 2002] R. D'Andrea, *et al.*: Detailed vision documentation, <http://robocup.mae.cornell.edu/>

[加藤, 2003] 加藤恭佑, 日比野晋也, 児玉幸司, 村上和人, 成瀬正: 情報処理学会研究報告, Vol. CVIM-136, No.16, pp115-122, 2003.

[Ohta, 2001] 太田紘高, 遠藤龍矢, 山本和彦, 加藤邦人, 神成敦司: 画像処理と戦略の協調による Robocup 小型システム, システムインテグレーション部門学術講演論文集, pp59-60, 2001.

[Tsai, 1987] 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, pp323-344, August 1987.

[Ryad, 2001] R. Benosman, J. Douret, J. Devars: A Simple and Accurate Camera Calibration for the F180 RoboCup League, in *RoboCup-2001: Robot Soccer World Cup V*, Springer, pp273-280, 2001.

[RoboCup, 2003] RoboCup: <http://www.robocup.org/>

[松山, 1998] 松山隆司: 分散協調視覚-視覚・行動・コミュニケーションの機能統合による知能の開発, 画像の認識・理解シンポジウム (MIRU '98), ppI-343-352, 1998.

[中澤, 2001] 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, pp2699-2710, 2001.