

追跡対象の動きを利用した多数カメラの位置・姿勢推定

山添 大丈^{†‡} 内海 章[†] 鉄谷 信二[†] 谷内田 正彦[‡]

[†]ATR メディア情報科学研究所

〒 619-0288 京都府相楽郡精華町光台 2-2-2

{utsumi, tetsutani}@atr.co.jp

[‡]大阪大学大学院基礎工学研究科

〒 560-8531 大阪府豊中市待兼山町 1-3

yamazoe@yachi-lab.sys.es.osaka-u.ac.jp

yachida@sys.es.osaka-u.ac.jp

あらまし: 本報告では追跡対象の動きを利用して、多視点システムのカメラの位置・姿勢を推定する手法について述べる。多視点人物追跡システムにおいて、カメラ数の増加によって、手作業でのカメラ校正は困難になる。また、カメラ数が多いと長期にわたってカメラの位置・姿勢を維持・管理していくためのコストも大きくなる。これらの問題を解決するため、カメラ間の整合性をとりながら多数のカメラ位置・姿勢を推定する手法を提案する。本手法ではカメラ位置・姿勢に関する情報(相関・共分散行列)を各カメラで分散的に持つことで、計算量の増大を避ける。実験により本手法の有効性を示した。

Position and Pose Estimation for Multiple Cameras in Distributed Human Tracking System

Hirotake Yamazoe^{†‡} Akira Utsumi[†] Nobuji Tetsutani[†] Masahiko Yachida[‡]

ATR Media Information Science Laboratories

2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288 Japan

{utsumi, tetsutani}@atr.co.jp

Graduate School of Engineering Science, Osaka University

1-3 Machikaneyama-cho, Toyonaka-shi, Osaka 560-8531, Japan

yamazoe@yachi-lab.sys.es.osaka-u.ac.jp

yachida@sys.es.osaka-u.ac.jp

Abstract: We describe a camera calibration method to determine the positions and poses of all installed cameras in a system by using target human motions. In multiple-camera human tracking systems, increasing the number of cameras causes serious problems for camera calibration. Manual calibration becomes a heavy (boring) task for humans. In addition, maintaining the geometrical consistency among multiple cameras in a time sequence becomes difficult when the number of cameras is huge. To reduce these problems, we propose an automatic camera calibration method that enables both initialization and maintenance of all camera positions and poses. To avoid explosion of computation, our system maintains the calibration information (correlation and covariance matrices) in a distributed way. Each camera maintains a minimum amount of information. Experimental results show the effectiveness of our method.

1 はじめに

我々は、計算機との非接触インタフェースを目指し、人間の動きを画像処理によって検出する手法の検討を進めている [1, 2]。人間の動き情報として、ここでは各人物の同定に必要な顔画像、背丈、服の色等の抽出、追跡に必要な位置・移動方向の検出、さらに着座動作を含むいくつかの動作の検出について検討している。これらの動作情報を検出することにより、仮想環境でのインタラクション、監視システムといった応用が可能になる。

従来より画像を用いた人物追跡については多くの提案があったがその多くは単眼または2眼画像によるものであり [3, 4, 5, 6]、オクルージョンに対応できない、検出エリアが狭い、といった問題があった。これらの問題を解決するため、近年多視点画像を利用した人物追跡システムが盛んに研究されている [7, 8, 9]。多視点画像を利用することでオクルージョンの発生を低減し、より安定した検出が可能になると考えられる。

しかし、このようなシステムで広範囲の人間の動きを追跡するには検出エリアに合わせて多くのカメラが必要となり、そのために生じる問題も多い。例えば、多くのビジョンシステムは3次元計測のために各カメラが同時刻に観測を行うことを前提としており、そのための同期機構の導入によりシステムが複雑化する。さらに、複数の観測を同時に行うことは観測間の冗長性を増大させ、システムの処理効率を低下させる。また、多数のカメラ全てをあらかじめキャリブレーションしておくことは、視点数(カメラ数)が増えるにしたがって困難になる。これらの問題は利用する視点数が増加するに伴ってより顕著になると考えられる。我々は多視点画像による追跡システムにおいては、これら規模の拡大によって生じる問題が本質的であると考えている。

本報告では、多視点システムにおけるカメラの位置・姿勢推定について考える。多視点システムでは事前のカメラ校正に加えて、運用中の障害によるカメラ位置・姿勢の変化への対応など維持・管理のための手法の確立も重要となる。我々は事前のカメラ校正だけでなく、運用中の障害の発見・修復も視野に入れながら手法の検討を進めている。

多視点追跡システムにおけるカメラ校正については既にいくつかの手法が提案されている。

斎藤らは2台の基準カメラと他のカメラ間の基礎行列を両者間で共有する観測を用いて求めることで大規模な3次元映像生成システムの構築を容易にした [10]。Leeらは、平面上を運動する対象物体を利用して基準カメラと各カメラとで共有する観測により、基準カメラからの相対位置・姿勢を求める手法を提案した [11]。しかしながら、これらの手法では校正対象のカメラと基準カメラとが観測を共有しない場合には適用できないという問題がある。これに対して、我々は追跡中の人物の3次元運動を利用しカメラの位置・姿勢を推定する手法を提案した [12]。しかしながら、この手法では人物の運動を求めるために複数の校正済カメラを要する。

本報告では任意の2カメラ間で共有される観測情報によってカメラの位置・姿勢の推定値を逐次修正し、最終的に多数のカメラの位置・姿勢を推定する手法を提案する。本手法では全てのカメラが基準カメラと観測を共有する必要はなく、基準カメラと共有されない局所的な観測も推定に利用することができる。

次節では我々の人物追跡システムの構成と処理の流れを簡単に述べる。3節では校正に用いる画像特徴について、4節では校正アルゴリズムについて述べる。5節でシミュレーション実験とその結果を示し、6節で本報告をまとめる。

2 人物追跡システムの構成

ここでは我々が本手法の適用を考えている人物追跡システムの概要と処理の流れを述べる。全体の構成と処理の流れを図2に示す。まずカメラ毎に独立して得られた入力画像に対し特徴抽出処理を行う(観測ノード)。ここで得られた特徴量(代表点位置と頭頂点、服領域の色)は、追跡ノードから送られてくる予測観測値をもとに追跡モデルと対応づけられた後、観測時刻の情報とともに追跡ノードに送られる。対応がとれなかった特徴量は発見ノードに送られる。前述の通り各観測ノードは独立に動作する。

発見ノードでは送られてきた未対応点情報を用いて、シーン中に新たに現れた人物の検出を行う。新規人物の検出結果は追跡ノードに送信され、追跡が開始される。

追跡ノードにおいて新規人物情報・観測情報は、それぞれ初期値・入力値となり、カルマンフィルタを用いることによって人物状態(位置・方向角・背丈等)が更新される。また位置・方向角推定結果を利用して顔領域の検出を行う。本システムによる追跡結果の一例を図2に示す。

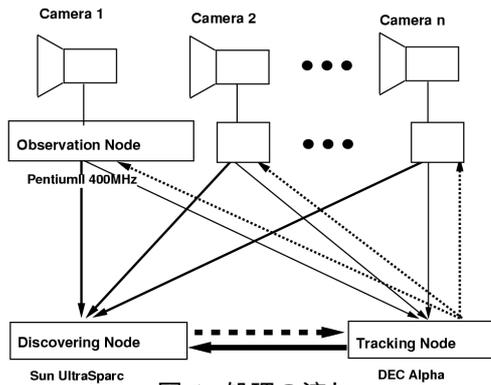


図1: 処理の流れ

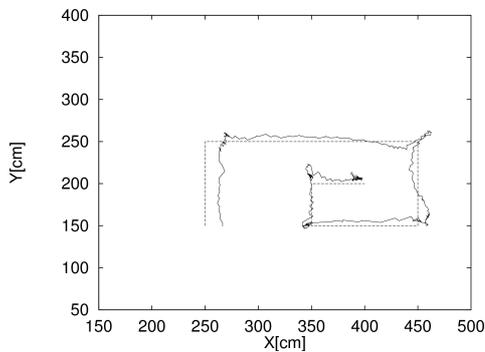


図2: 追跡結果の例

ここに述べたように、本システムでは各カメラによる観測を独立した観測ノードによって処理しており、カメラの追加・削除が容易に行えるという利点を持つ。このような分散型の追跡システムではカメラの位置・姿勢情報についても各カメラが独立に維持・管理することが処理効率・対障害性の確保から望ましいといえる。次節以降では我々の提案する分散型のカメラ位置・姿勢推定アルゴリズムについて述べる。

3 カメラ位置・姿勢の校正

本節では、提案するカメラ位置・姿勢推定のアルゴリズムについて説明する。多数のカメラを同時に推定するためには、現在の推定値、共分散行列等の情報を保持する必要があるが、カメラ数の増加とともに保持すべき情報の量も増加し、大規模なシステムで全てのカメラの位置・姿勢推定に関する情報を一元的に保持することは困難となる。そのため、提案手法ではカメラの位置・姿勢推定に関する情報を各カメラ毎に分散して保持することを考える。

図3にカメラ位置・姿勢推定の流れを示す。各カメラは新たな観測が得られるたびに周囲のカメラと共有する観測情報と周囲のカメラのその時点での位置・姿勢推定値を用いて自身の位置・姿勢を修正する。その際、基準カメラの位置・姿勢にはより高い重みが与えられる。なお、本手法では世界座標を決定するために、最低限カメラ1台の姿勢とカメラ2台の位置を基準として与える必要がある。

次節では、今回考える観測情報について述べる。

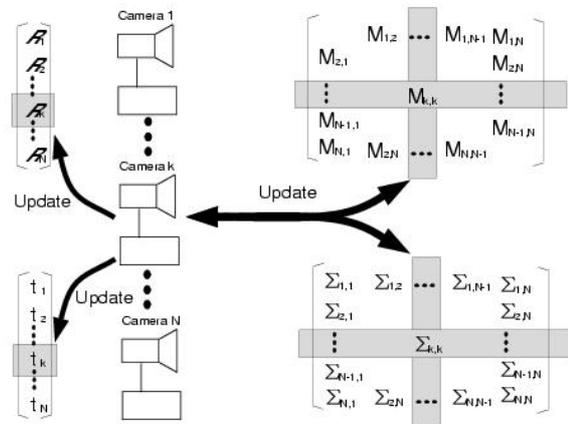


図3: カメラ校正

3.1 位置・姿勢推定に用いる観測情報

3.1.1 特徴抽出

ここでは、位置・姿勢推定に用いるために各観測ノードで得られる画像特徴について述べる。

まず入力画像を人物領域と背景領域に分割し[13]、得られた人物領域に対して距離変換を行う。

距離変換された画像における領域内の各画素は、最近傍の領域境界までの距離を画素値として持つ (図 4 右)。領域内で画素値が最大となる点をその領域の代表点として選択する。さらに、画像中から頭頂点と足先点の位置、胴体部分の色情報を抽出する (図 5)。色情報は異なるカメラ間での画像特徴の対応づけに用いられる。頭頂点と足先点の画像上での距離を特徴点 (代表点) の大きさとして利用する。

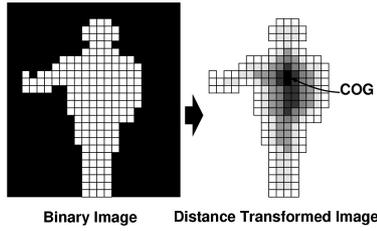


図 4: 特徴抽出



図 5: 特徴抽出の例

3.1.2 観測情報

本節では、次節以降で述べる姿勢・位置推定値の更新に利用する観測情報について説明する。

ここではカメラ C_k について、カメラ C_l と共有する複数回の観測からカメラ C_k とカメラ C_l の間の相対姿勢に関する行列 m_{kl} とカメラ C_k のカメラ座標系でのカメラ C_l の方向 y_{kl} が観測情報として得られると考える。

ここで、 m_{kl} は以下のように定義される。

カメラ C_k, C_l と同じ姿勢を持つカメラ C'_k, C'_l がともに世界座標原点にあるとする。両者の姿勢を R_k, C_l とすると C'_k のカメラ座標系上で方向 $X_{k,i}$ に観測される観測点 i は C_l 上では次の方向 $X_{l,i}$ に観測される。

$$X_{l,i} = R_l R_k^{-1} X_{k,i} + e \quad (1)$$

(e_k は観測誤差)。

m_{kl} は以下のように計算する。

$$m_{kl} = \sum_{j=0}^s X_{k,j} X'_{l,j}. \quad (2)$$

カメラ C_k, C_l 間の相対姿勢 $R_l R_k^{-1}$ は (2) 式から特異値分解を利用して求めることができる [14]。

なお、前節で述べた画像特徴から m_{kl}, y_{kl} を得る方法については付録 A に示す。

カメラ C_k と各カメラについて求めた相関行列・相対位置から、カメラ C_k の位置・姿勢を推定する。

次節以降ではカメラ C_k から見たカメラ C_k, C_l 間の相対姿勢・相対方向を用いてカメラ C_k の位置・姿勢を更新することを考える (図 6)。

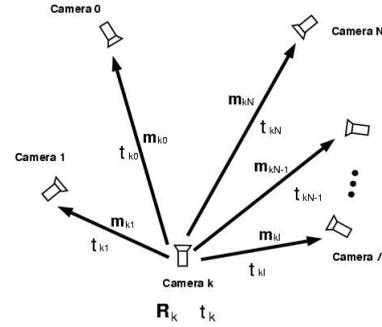


図 6: 位置・姿勢の推定

3.2 初期状態

本節では各カメラの位置 t_i 、位置に関する共分散行列 Σ_{ij} 、姿勢 R_i 、姿勢推定に関する行列 M_{ij} の初期状態について述べる。

位置の基準となるカメラ C_{t_1}, \dots, C_{t_n} ($n \geq 2$) については世界座標上での位置 t_{t_1}, \dots, t_{t_n} を与える。また、

$$\Sigma_{ii} = \begin{cases} p_1 I & (i \in [t_1, \dots, t_n]) \\ p_2 I & (i \in [1, \dots, t_n]) \end{cases} \quad (3)$$

とする ($p_1 \ll p_2$)。

姿勢の基準となるカメラ C_{r_1}, \dots, C_{r_m} ($m \geq 1$) については世界座標上での姿勢 R_{r_1}, \dots, R_{r_m} を

与える。また、

$$M_{ij} = \begin{cases} q_1 \mathbf{R}_{r_i} & (i = j, i \in [r_1, \dots, r_n]) \\ q_2 \mathbf{I} & (\text{otherwise}) \end{cases} \quad (4)$$

とする ($q_1 \gg q_2$)。

3.3 位置・姿勢推定値の更新

本手法では、新たに観測が得られる度に各カメラでの推定値を局所的に最適化することにより、最終的に全カメラの位置・姿勢推定値を得る。

カメラ C_k では自己の周囲にある全てのカメラ $C_i (i \neq k)$ の位置推定値 t_i 、姿勢推定値 R_i 、位置推定値に関する共分散行列 Σ_{ii} 、自己の姿勢に関する行列 M_{ii} 、さらに自己とカメラ C_i の間の相対姿勢に関する行列 M_{ki} を保持する (図 6)。

以下では、カメラ C_k において新たな観測 y_{kl} 、 m_{kl} が得られたときの姿勢推定値、位置推定値の更新方法について説明する。

なお、今回はあるカメラで新たな観測が得られる度にそのカメラでのみ更新を行うこととしたが、推定値の更新間隔についてはさらに検討を進める必要がある。

3.3.1 姿勢推定値の更新

本節では、カメラ C_k の姿勢情報の更新について説明する。

カメラ C_k では、周囲のカメラの姿勢推定値 R_i とカメラ C_k と各カメラ C_i の相対姿勢に関する行列 M_{ki} をもつ。

ここで、観測により得られた m_{kl} により、行列 M_{kl} を更新する。

$$M_{kl}^t = \frac{1}{u} M_{kl, t-1}^{t-1} + m_{kl} \quad (5)$$

(u は忘却係数)。さらに自己姿勢に関する行列 M_{kk} を更新する。

$$M_{kk}^t = \sum_{j=0}^N \mathbf{R}_i^{-1} M_{ki}^t \quad (i \neq k) \quad (6)$$

新たな姿勢推定値 R_k は、 M_{kk} を特異値分解することで得られる。

3.3.2 位置推定値の更新

本節では、さらにカメラの位置推定値の更新について述べる。カメラ C_k では、各カメラの位置に関して現時点での位置推定値 t_i と各カメラとの共分散行列 Σ_{ii} を保持する。

この時、観測 y_{kl} による新たな位置推定値 t_k は以下のように求められる。

$$t_k^t = t_k^{(t-1)} + K(Ht_l^{(t-1)} - Ht_k^{(t-1)}) \quad (7)$$

ここで、

$$H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{R}_k^{-1} \mathbf{R}_{kl}^{-1} \quad (8)$$

R_{kl, t_n} は $R_{kl, t_n} \frac{y_{kl, t_n}}{\|y_{kl, t_n}\|} = [0 \ 0 \ 1]^T$ を満たす回転行列。

$$\Sigma_{kk}^t = \Sigma_{kk}^{(t-1)} - K^t H^t \Sigma_{kk}^{(t-1)} \quad (9)$$

ただし、

$$K = (H^{(t-1)})' (H \Sigma_{kk}^{(t-1)} H' + G)^{-1} \quad (10)$$

なお、(7)~(11) はカルマンフィルタの更新式である。

4 実験

本手法の有効性を確認するため、次のシミュレーション実験を行った。

5台のカメラ(カメラ1-5)を用い、世界座標系を決めるためにカメラ1の位置・姿勢とカメラ2の位置はあらかじめ与えた。

シーン中の $150 \times 150 \times 150$ [cm] の領域から10000点をランダムに選択し、各点について、5台のカメラからランダムに2台を選択した。その2台による2次元観測位置と2次元大きさを計算し、さらにガウス誤差を加えたものを、各カメラでの観測とした。

各カメラに保存されている位置・姿勢に関する校正情報は、各観測によって更新される。

図7に全カメラパラメータの推定結果を示す(横軸: データ数)。ここに見られるように、校正の初

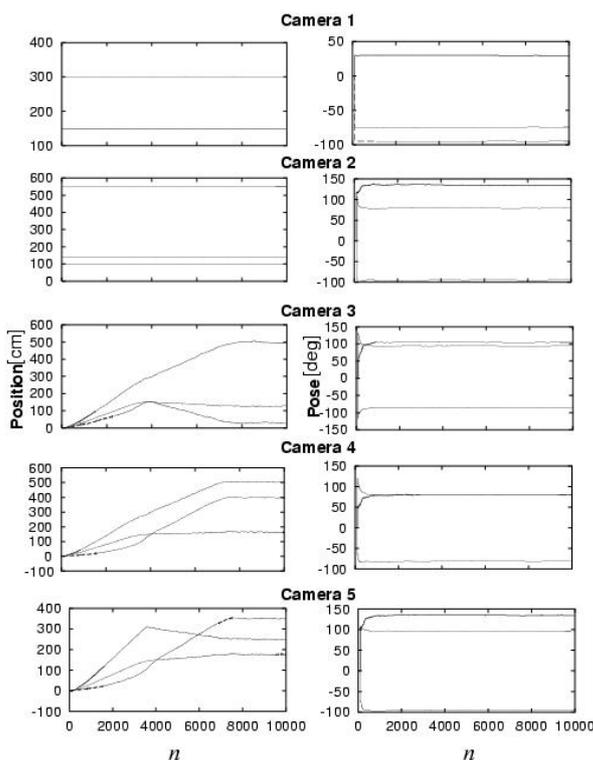


図 7: カメラパラメータ推定結果

期段階では位置・姿勢ともに大きな誤差を含んでいる。しかし、データ数の増加に伴い推定誤差が減少しており、最終的には位置で平均 7.35 [cm]、姿勢で平均 0.85 [deg] の高い推定精度が得られていることがわかる。以上の結果から本手法によって多数カメラの位置・姿勢推定が行えることが示された。

5 まとめ

本報告では、多数のカメラからなる人物追跡システムにおいて、シーン中の人の動きを用いて、カメラの位置・姿勢を推定する手法について述べた。提案手法では各カメラが周囲のカメラとの相対的な位置・姿勢の観測結果をもとに自己の位置・姿勢を逐次更新する。このため各カメラは全てのカメラの位置・姿勢推定に関する情報を保持する必要がなく、本手法は分散型の追跡システムに適しているといえる。一方で、各カメラの位置・姿勢の推定誤差の評価が不十分であるため、推定の収束に時間がかかるという問題がある。

今後は、推定誤差評価の改善によって収束性・口

バスト性の向上を計るとともに推定に利用する観測情報についてもさらに検討を進めていきたい。また、本手法を実システムに実装し、障害からの自律回復を含めた応用を目指していきたい。

参考文献

- [1] Hiroki Mori, Akira Utsumi, Jun Ohya, and Masahiko Yachida. Human tracking system using adaptive camera selection. In *Proc. of RO-MAN '98*, pp. 494–499, 1998.
- [2] Howard Yang, 内海章, 大谷淳. 非同期多視点画像を用いた複数人物追跡の安定化. 信学技報, PRMU99-150, pp. 1–7, 1999.
- [3] D. M. Gavrila and L. S. Davis. 3-d model-based tracking of humans in action: a multi-view approach. In *Proc. of CVPR*, pp. 73–80, 1996.
- [4] Ali Azarbayejani and Alex Pentland. Real-time self-calibrating stereo person tracking using 3-d shape estimation from blob features. In *13th International Conference on Pattern Recognition*, pp. 627–632, 1996.
- [5] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfunder: Real-time tracking of the human body. In *SPIE proceeding vol. 2615*, pp. 89–98, 1996.
- [6] M. Patrick Johnson, P. Maes, and T. Darrell. Evolving visual routines. In *Proc. of Artificial Life IV*, pp. 198–209, 1994.
- [7] Jakub Segen and Sarma Pingali. A camera-based system for tracking people in real time. In *Proc. of 13th International Conference on Pattern Recognition*, pp. 63–67, 1996.
- [8] Q. Cai, A. Mitiche, and J. K. Aggarwal. Tracking human motion in an indoor environment. In *Proc. of 2nd International Conference on Image Processing*, pp. 215–218, 1995.

- [9] Q. Cai and J. K. Aggarwal. Tracking human motion using multiple cameras. In *Proc. of 13th International Conference on Pattern Recognition*, pp. 68–72, 1996.
- [10] Hideo Saito and Takeo Kanade. Shape reconstruction in projective grid space from large number of images. In *Proc. of CVPR*, pp. 49–54, 1999.
- [11] L.Lee, R.Romano, and G.Stein. Monitoring activities from multiple video streams: Establishing a common coordinate frame. *IEEE Pattern Anal. Machine Intell.*, Vol. 22, No. 8, pp. 758–767, 2000.
- [12] Hirotake Yamazoe, Akira Utsumi, Nobuji Tetsutani, and Masahiko Yachida. Automatic camera calibration method for multiple camera based human tracking system. In *Proc. of IWAIT 2001*, pp. 77–82, 2001.
- [13] 内海章, 大谷淳. 画素値分布の推定による時系列画像からの動物体領域抽出. *信学論*, Vol. J81-D-II, No. 8, pp. 1768–1775, 1998.
- [14] F.Landis Markley. Attitude determination using vector observations and the singular value decomposition. *the Journal of the Astronautical Sciences*, Vol. 36, No. 3, pp. 245–258, 1988.

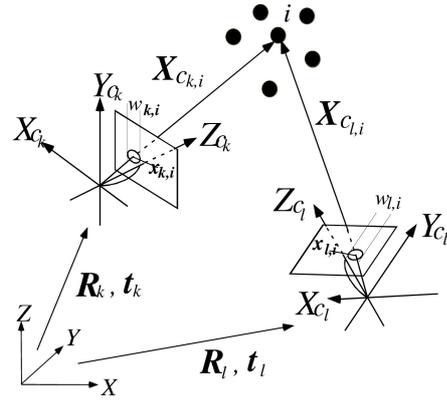


図 8: 座標系と特徴点の投影

カメラ座標系での人物の 3 次元位置を X_{c_k} とすると、 X_{c_k} と X には次の関係が成り立つ。

$$X_{c_k} = \begin{bmatrix} R_k & t_k \end{bmatrix} \begin{bmatrix} X \\ 1 \end{bmatrix}. \quad (12)$$

$a_3 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$ より $a_{m3} \begin{bmatrix} x \\ 1 \end{bmatrix}$ は一定なので、 X_{c_k} は以下のように表せる。

$$X_{c_k} \propto \frac{1}{w_k} A_k^{-1} \begin{bmatrix} x_k \\ 1 \end{bmatrix}, \quad (13)$$

ここで、 w_k は特徴点 x_k の画像上での大きさを示す。

カメラ C_l 座標系での人物の 3 次元位置 X_{c_l} も同様に得られ、

$$X_{c_l} \propto \frac{1}{w_l} A_l^{-1} \begin{bmatrix} x_l \\ 1 \end{bmatrix}. \quad (14)$$

R_l, t_l をカメラ C_l の姿勢・位置とし、 s 回の観測について考えると、次式が得られる。

$$\begin{aligned} X_{c_k,1} &= R_k R_l^{-1} X_{c_l,1} + R_k (t_l - t_k), \\ &\vdots \\ X_{c_k,s} &= R_k R_l^{-1} X_{c_l,s} + R_k (t_l - t_k). \end{aligned} \quad (15)$$

よって、次式を得る。

$$\bar{X}_{c_k} = R_k R_l^{-1} \bar{X}_{c_l} + R_k (\bar{t}_l - \bar{t}_k), \quad (16)$$

(ここで、 $\bar{X}_{c_k} = \frac{1}{s} \sum_{i=1}^s X_{c_k,i}$ 、 $\bar{X}_{c_l} = \frac{1}{s} \sum_{i=1}^s X_{c_l,i}$)

(15) 式と (16) より、

付録: m_{kl} 、 y_{kl} の観測

3 次元位置 X_k の点のカメラ C_k の画像上の観測位置を x_k とする。

$$\begin{aligned} s \begin{bmatrix} x_k \\ 1 \end{bmatrix} &= A_k \begin{bmatrix} R_k & t_k \end{bmatrix} \begin{bmatrix} X \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \begin{bmatrix} R_k & t_k \end{bmatrix} \begin{bmatrix} X \\ 1 \end{bmatrix}. \end{aligned} \quad (11)$$

ここで、 A_k は既知のカメラ C_k の内部パラメータ行列とし、 R_k と t_k が世界座標に対するカメラの姿勢を表す推定する回転行列とカメラ位置を表すベクトルである。なお、 $a_3 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$ とする。

$$\begin{aligned}\hat{\mathbf{X}}_{c_k,1} &\propto \mathbf{R}_k \mathbf{R}_l^{-1} \hat{\mathbf{X}}_{c_l,1}, \\ &\vdots \\ \hat{\mathbf{X}}_{c_k,s} &\propto \mathbf{R}_k \mathbf{R}_l^{-1} \hat{\mathbf{X}}_{c_l,s},\end{aligned}\quad (17)$$

ここで、 $\hat{\mathbf{X}}_{c_k,i} = \frac{(X_{c_k,i} - \bar{X}_{c_k})}{\|X_{c_k,i} - \bar{X}_{c_k}\|}$, $\hat{\mathbf{X}}_{c_l,i} = \frac{X_{c_l,i} - \bar{X}_{c_l}}{\|X_{c_l,i} - \bar{X}_{c_l}\|}$.

(17) 式を満たす回転行列は、次式を最小にする $\mathbf{R}_k \mathbf{R}_l^{-1}$ である。

$$\sum_{i=1}^s \left\| \hat{\mathbf{X}}_{c_k,i} - (\mathbf{R}_k \mathbf{R}_l^{-1}) \hat{\mathbf{X}}_{c_l,i} \right\|^2 \rightarrow \min \quad (18)$$

この解は、以下の相関行列 m_{kl} を特異値分解することで求められる。

$$m_{kl} = \sum_{i=0}^s \hat{\mathbf{X}}_{c_k,i} \hat{\mathbf{X}}_{c_l,i}^t. \quad (19)$$

一方、カメラ C_k のカメラ座標系から見たカメラ C_l の方向ベクトル $\mathbf{y}_{kl} = \mathbf{R}_k(\mathbf{t}_l - \mathbf{t}_k)$ は得られた $\mathbf{R}_k \mathbf{R}_l^{-1}$ を (16) 式に代入して得られる。