

幾何学的 AIC によるカメラモデル選択

木下 敬介[†] Michael Lindenbaum^{‡‡}

[†] ATR 人間情報通信研究所

^{‡‡} The Technion, Israel

kino@hip.atr.co.jp, mic@cs.Technion.AC.IL

3次元空間がカメラによって画像に投影される場合、それらの間の関係は、透視カメラモデルや、アフィンカメラモデルなどによってモデル化される。どちらのモデルを採用するかは、3次元物体の形状や、物体とカメラの間の距離、そしてもちろん、カメラの内部パラメータ（焦点距離など）に依存するが、従来は、システム的设计者の直感によって決められてきた。本研究では、透視変換カメラモデルをより一般化した、射影カメラモデルと、アフィンカメラモデルの間で、どちらのカメラモデルがより適切なのかを、3次元物体形状や、画像データから定量的に判断する。判断基準は、金谷によって提案されている幾何学的 AIC を用いた。これは、データをモデルに当てはめた時の残差と、モデルの複雑さを総合的に判断するものである。単眼カメラの場合と未校正ステレオカメラの場合について解析・実験を行った。

Camera Model Selection based on Geometric AIC

Keisuke Kinoshita[†] Michael Lindenbaum^{‡‡}

[†] ATR Human Information Processing Research Labs.

^{‡‡} The Technion, Israel

When a camera projects 3D objects onto an image plane, the perspective camera model or affine camera model are often used to describe the relation between them. However, there is no quantitative criterion for a decision: which camera model should be used? Obviously, it depends on the object shape, the distance between the camera and the object, and the intrinsic camera parameters(i.e. focal length). In this paper, a new criterion for selecting camera models between projective(which is a more general model than perspective) and affine camera models from object shape and/or image data is proposed. We use *geometric AIC* proposed by Kanatani, which evaluates the residual of the data fitting to the model as well as the complexity of the model. Both monocular and uncalibrated stereo settings are analysed.

1 はじめに

実世界の対象を何らかの形で抽象し、それを一定の記述形式のもとで表現することは、モデリングと呼ばれる [1]。実世界では、ありとあらゆることが起こり得るが、その全てを抽象し、記述することは不可能である。モデルとは、その実世界の現象の中から、重要な、あるいは、興味のある要素間の関係を簡潔に記述したものである。実世界を完全にはモデル化できない以上、モデルはあくまでも実世界の近似に過ぎない。

コンピュータビジョンの世界に目を向ける。実世界が、カメラを通して画像に投影される。実世界では様々な物体が存在し、本来ならその全てを記述したいのだが、数学的な取り扱いやすさから、点、線、平面、二次曲面といった幾何学的なモデルが好んで使用されている。

実世界から画像への投影に関しても、本来なら様々な要素を考慮しなくてはならない。レンズの歪み、絞りの影響、CCD 素子の特性、AD 変換器の精度などである。それらの中で、まだ扱いやすいのは、遠くのもの小さく、近くのもの大きく写るといふ、幾何学的な投影関係である。この枠組みの中では、投影関係以外の要素、(レンズの歪みなど)は、ノイズとして扱われる。

この投影関係をモデル化したものがカメラモデルである。幾何学的な投影関係は、射影幾何学、あるいは、エビポーラ幾何学という数学的な枠組みで簡潔に記述できる [2][3][4]。その中でも、射影カメラモデルとアフィンカメラモデルは、代表的なカメラモデルである。射影カメラモデルや、透視カメラモデルは、3次元空間と画像の間が、非線形な関係で結ばれているのに対して、アフィンカメラモデルでは、線形な関係で結ばれているという性質を持つ。

では、3次元復元や画像生成などのアプリケーションにおいて、いったい、どちらのカメラモデルを使って、設定した問題を解けば良いのであろうか。一般には射影カメラモデル、あるいは、透視カメラモデルが使われている。しかし、対象の厚みがあまりなく、対象とカメラの間の距離が長い場合は、アフィンカメラモデルを用いる方が良いともいわれている。実際、現実の多くの問題に、アフィンカメラモデルが用いられ、よい結果を産み出している。しかし、アフィンカメラモデルを使うことの正当性はどこにあるのだろうか。あるいは、どのような場合に、射影カメラモデルを使い、どのような状況でアフィンカメラモデルを使えばよいのであろうか。

本研究では、金谷の幾何学的 AIC [5][6][7] を適用することで、カメラモデル選択の定量的な判断基準を提供する。この判断基準を使うことで、状況に応じて、適切なカメラモデルを、動的に切り替えるようなシステムを実現することができる。

幾何学的 AIC は、統計的な推定方法と密接な関係がある。コンピュータビジョンにおける統計的な推定は、金谷によって導入され、[8] に詳しく、そして厳密に解説されている。さらに、最近、統計的な判断基準を、画像からの運動推定に適用した Torr らの研究 [9][10] も大きく注目されている。

以下、まず、カメラモデルについて述べた後、金谷の幾何学的 AIC に関して簡単に説明する。次に、射影カメラモデルとアフィンカメラモデルのどちらを採用すればよいのかについて、(単眼カメラ+対象の3次元情報既知)の場合と、(未校正ステレオカメラ+対象の3次元情報未知)の場合について、それぞれ解析する。

2 カメラモデル

3次元空間が画像にどのように投影されるかを記述しているのが、カメラモデルである。カメラモデルを含め、幾何学的な投影関係は、射影幾何学、あるいはエビポーラ幾何学と呼ばれる枠組みで簡潔に記述できる [2][3][4]。いま、3次元空間の点を X 、その3次元射影座標を $\bar{X} = (X, Y, Z, W)^T$ とする。また、その画像への投影点 x の2次元射影座標を $\bar{x} = (x, y, w)^T$ とする。3次元空間から画像への投影の関係は、 3×4 の行列 P を使って、

$$\bar{x} \simeq P\bar{X} \quad (1)$$

と書き表せる。ただし、 \simeq は、定数倍の不定性を除いて等しいことを表す。この 3×4 の行列 P は投影行列、あるいは、カメラ行列と呼ばれる。3次元空間から画像への投影の情報は全てこの行列に集約されている。 P の各要素にどのような関係があるかによって、様々なカメラモデルを表現できる。

まず、最も一般的なカメラモデルとして知られているのが、射影カメラモデルである。射影カメラ行列 P_p の各要素は、どのような値をとってもよく、

$$P_p = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix}$$

という形をもつ。カメラ行列には定数倍の不定性があるので、実質的な自由度は 11 である。

射影カメラモデルでは、3 次元空間と画像の間は、齊次座標で表せば、線形な関係を持っているかのようにみえるが、実際の 3 次元空間座標と画像座標の間は、非線形な関係にある。そのため、扱いが難しく、画像に乗るノイズの影響を大きく受け、3 次元復元などへの応用では、必ずしも安定な解が得られない。しかしながら、このカメラモデルは、射影幾何学の枠組みの中では、現実の投影関係を正確に記述している。

この射影カメラモデルの特殊な場合として、アフィンカメラモデルがある。アフィンカメラ行列は、

$$P_a = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ 0 & 0 & 0 & p_{34} \end{pmatrix}$$

のような形を持つ。アフィンカメラ行列の場合も、定数倍の不定性があるので、実質的な自由度は 8 である。射影カメラ行列と比較すると、第 3 行 1 列目から 3 列目までが 0 であるという特徴を持つ。この特徴により、実際の 3 次元空間の座標と画像の間は線形な関係を持つ。線形であるので、画像に乗るノイズの影響は、非線形な場合に比べて小さくなる。

撮影している対象の奥行き（厚み）が、カメラから物体までの距離に対して十分小さい場合、アフィンカメラによる近似が有効であると言われている（図 1）。それは、 P_p の P_{31}, P_{32}, P_{33} の値が他の要素と比べて十分小さくなっているということである。そのような状況では、射影カメラ行列ではなく、アフィンカメラ行列を採用することで、すべてが線形な関係となり、3 次元復元や、画像生成などのアルゴリズムが簡潔になるというメリットがある。ところが、アフィンカメラモデルによる近似が有効かどうかは、カメラの性質（焦点距離など）、物体までの距離、物体形状などに依存するのだが、いまだ、定量的な判断基準は提案されていない。もし、アフィンカメラモデルによる近似が有効でない状況でアフィンカメラモデルを採用したならば、モデルの誤差が顕著に結果に現れてきて、たとえば、数値計算上、安定に 3 次元情報を復元できたとしても、もはや、意味のある情報とは言えなくなる。

表 1 に、射影カメラモデルとアフィンカメラモデルの性質を簡単にまとめ、それぞれのカメラモデルの一種である、透視カメラモデルと弱透視カメラモデルの投影図を図 2 に示した。両者の中間的な性質を持つカメラモデルを新たに作る試み [11] も報告されているが、本研究では、どちらか一方を選択することを考え

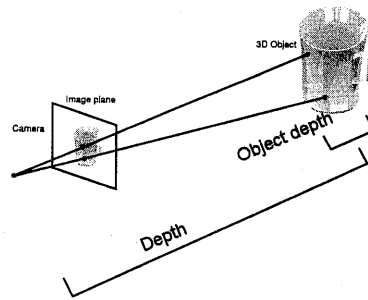


図 1. 対象物体の厚みが、カメラまでの距離に比べて十分小さい場合は、アフィンカメラモデルを採用するのが望ましい。

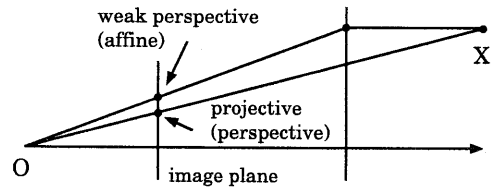


図 2. 3 次元空間の点が画像面に投影される。射影カメラの一種である透視カメラモデルと、アフィンカメラモデルの一種である弱透視カメラモデルの比較。

る。どのような場合に、より簡単なアフィンカメラモデルを採用すべきか、また、どのような場合は、複雑な射影カメラモデルを採用しなくてはならないかについて解析していく。

3 幾何学的 AIC

射影カメラモデルとアフィンカメラモデルの選択基準を説明するために、コンピュータビジョンにおけるパラメータ推定法と、幾何学的 AIC について簡単に

表 1. 射影カメラモデルとアフィンカメラモデルの性質の比較

	射影カメラモデル	アフィンカメラモデル
モデルの記述	非線形	線形
自由度	11	8
モデルの精度	正確	近似
数値的な安定性	不安定	安定

説明する。

3.1 パラメータの最適な推定法

金谷によって提案された、最適なパラメータ推定方法 [8] を紹介する。観測データに乗るノイズの分布形状を与えるだけで、最適なパラメータを推定できるのはもちろん、ノイズの大きさ（ノイズレベル）までも得られるのが特徴である。

対象としているのは、 N 個の観測データ $a_1, \dots, a_\alpha, \dots, a_N$ が与えたときに、(2) を満たすような、パラメータ u を推定するという問題である。(2) は L 個の連立同次方程式¹ で構成されている。コンピュータビジョンの多くの問題は、この形式に書き直すことができる。

$$\begin{cases} (a_\alpha^{(1)}, u) = 0 \\ \vdots \\ (a_\alpha^{(L)}, u) = 0 \end{cases} \quad (2)$$

$V[a_\alpha^{(k)}, a_\alpha^{(l)}]$ を観測データ a の共分散行列とする。パラメータ u を最適に推定するには、

$$W_\alpha^{(k,l)} = (u, V[a_\alpha^{(k)}, a_\alpha^{(l)}]u)^-, \\ M = \frac{1}{N} \sum_{\alpha=1}^N \sum_{k,l=1}^L W_\alpha^{(k,l)} a_\alpha^{(k)} a_\alpha^{(l)T}$$

とにおいて、

$$J[u] = (u, Mu) \rightarrow \min \quad (3)$$

となるような u を求めればよい。

ここでは、 $V[a_\alpha^{(k)}, a_\alpha^{(l)}] = \epsilon^2 V_0 [a_\alpha^{(k)}, a_\alpha^{(l)}]$ のように、 V を、観測データに乗るノイズの大きさ（ノイズレベル） ϵ と、ノイズの正規化された分布（正規化共分散行列） V_0 に分けて考える。一般に、 V は事前には知ることはできないが、その分布形状 V_0 は既知であるとみなしてよい場合が多い。例えば、画像に乗るノイズは、どのくらいの大きさのノイズが乗っているかは分からないが、 u, v 方向で無相関であると仮定してもおかしくはない。この場合、 $V_0 = I_2$ となる。以降、 V_0 は既知であるとする。

(3) を計算する過程で、 V の代わりに、 V_0 を使い、そのとき、得られる最小値（残差）を J_0 とする。すると、残差 J_0 から、逆に、ノイズレベル ϵ を

$$\epsilon^2 = \frac{J_0}{rN - n'} \quad (4)$$

¹(\bullet, \bullet) はベクトルどうしの内積を表す。

と推定することができる (r, N, n' については後述)。つまり、ノイズの分布形状 V_0 さえ分かっているならば、最適なパラメータ u が求まるのはもちろん、未知であったノイズレベル ϵ までも推定することが可能になる。

3.2 幾何学的 AIC

コンピュータビジョンの問題において、どのようなモデルを採用すべきなのかを決定する基準が、金谷によって提案されている [6][8][12]。この基準は幾何学的 AIC と呼ばれ、赤池の情報量基準 (Akaike Information Criterion) [13] を元に、金谷が拡張したものである。[7] では、幾何学的 AIC を用いて、3次元空間中の点群に対して 3次元平面、直線のどのモデルを当てはめるべきかについて検討している。また、[5] では、画像からの 3次元運動解析にも適用している。

モデル S の幾何学的 AIC は、

$$AIC(S) = J[S] + 2(dN + n') \quad (5)$$

で定義される。ただし、

- S : モデル
- $J[S]$: 観測データ a_1, \dots, a_N を、モデル S に最適に当てはめた時の残差
- d : S の次元
- n' : モデル S の自由度
- m' : 観測データ a の自由度
- N : 観測データ数
- r : 余次元。 $r = m' - d$ である。(2) の L 個の方程式の中で、独立なもの数。

である。この AIC が小さいほど、そのモデルは、観測データを良く説明しているといえる。

本来なら $J[S]$ を得るためには、 a の共分散行列 V を知っている必要がある。しかし、 V の代わりに V_0 を使って得られた残差 $J_0[S]$ と、 $J[S]$ の間には、 $J = \frac{1}{\epsilon^2} J_0$ という関係があるので、 V_0 が既知の場合には、(5) を、

$$AIC_0(S) = J_0[S] + 2(dN + n')\epsilon^2 \quad (6)$$

で代用することができる。

もし、二つのモデル S_1 と S_2 の間に、包含関係 ($S_2 \subset S_1$) があるとする (例えば、 S_2 : 直線、 S_1 : 平面)。両者

の AIC_0 は、それぞれ、

$$AIC_0(S_1) = J_0[S_1] + 2(d_1N + n'_1)\epsilon^2, \quad (7)$$

$$AIC_0(S_2) = J_0[S_2] + 2(d_2N + n'_2)\epsilon^2 \quad (8)$$

である。ここで、観測データに含まれるノイズの大きさ ϵ は、両者で共通である。この ϵ には、より一般的で、近似のないモデル S_1 の残差 $J_0[S_1]$ から推定されるノイズレベル ϵ を用いる。この ϵ を (7),(8) に代入して $AIC_0(S_1)$ と $AIC_0(S_2)$ の大きさを比較する。もし、 $AIC_0(S_2) < AIC_0(S_1)$ なら、モデル S_2 を選択することが合理的である。あるいは、 ϵ を (7),(8) から消去し、

$$\frac{J_0[S_2]}{J_0[S_1]} < 1 + \frac{2(d_1 - d_2)N + 2(n'_1 - n'_2)}{r_1N - n'_1} \quad (9)$$

が成立するならば、モデル S_2 を選択するべきである。

ここで、問題としているのは、より一般的なモデル S_1 か、それと包含関係にある、より特殊なモデル S_2 か、である。 S_1 はすでに、観測データを説明するモデルとして正しいと仮定していることに注意する。 S_1 はもちろん、更に S_2 をモデルとして適用しても良いかどうかを判断しているのである。

4 カメラモデル選択

4.1 単眼、対象の3次元情報既知の場合

3次元空間中の点 X と、その画像への投影点 x が与えられたときに、それらの関係を、射影カメラモデル S_p でモデル化した方がいいのか、アフィンカメラモデル S_a でモデル化した方がいいのかを解析する (図3)。

射影カメラ行列 P_p を

$$P_p = \begin{pmatrix} p_1^T \\ -p_2^T \\ -p_3^T \end{pmatrix}$$

と書く。3次元空間中の点 $X_\alpha = (X_\alpha, Y_\alpha, Z_\alpha, 1)^T$ を、このカメラで投影すると、

$$\bar{x}_\alpha \simeq P_p \bar{X}_\alpha$$

という関係より、画像上 $(u_\alpha, v_\alpha)^T$ に投影される。ただし、

$$u_\alpha = \frac{(p_1, X_\alpha)}{(p_3, X_\alpha)}, \quad v_\alpha = \frac{(p_2, X_\alpha)}{(p_3, X_\alpha)}$$

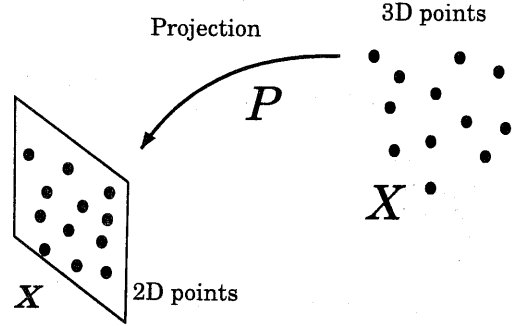


図3. 3次元空間の点が画像に投影されているときに、どのようなカメラモデルでモデル化するのがよいのか。

である。整理すると、

$$\begin{pmatrix} X_\alpha^T & 0^T & -u_\alpha X_\alpha^T \\ 0^T & X_\alpha^T & -v_\alpha X_\alpha^T \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = 0 \quad (10)$$

が得られる。(2)に対応させるならば、

$$a_\alpha^{(1)} = \begin{pmatrix} X_\alpha^T & 0_4^T & -u_\alpha X_\alpha^T \end{pmatrix},$$

$$a_\alpha^{(2)} = \begin{pmatrix} 0_4^T & X_\alpha^T & -v_\alpha X_\alpha^T \end{pmatrix},$$

$$u = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$$

である。

一方、アフィンカメラ行列を、

$$P_a = \begin{pmatrix} \cdots & p_1^T & \cdots \\ \cdots & p_2^T & \cdots \\ 0 & 0 & 0 & p_{34} \end{pmatrix}$$

と表す。変数は、 p_1, p_2, p_{34} の計9個であるが、実質的な自由度は8である。射影カメラモデルの場合と同様に、(2)に対応する書き方をすると、

$$\begin{pmatrix} X_\alpha^T & 0^T & -u_\alpha \\ 0^T & X_\alpha^T & -v_\alpha \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_{34} \end{pmatrix} = 0 \quad (11)$$

が得られる。

観測データの自由度は、画像の u, v 座標の二つであり、 $m'_p = m'_a = 2$ である。点 X の3次元座標は与えられるもので、観測データではないので、 m' の数には入らない。 $\bar{x} \simeq P\bar{X}$ は、二つの拘束式を与えるの

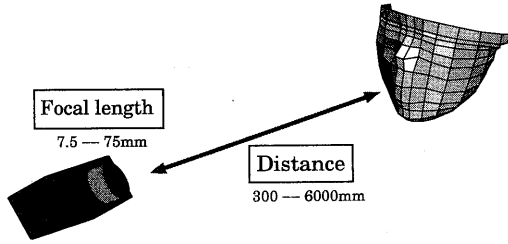


図 4. 3次元物体（49点で構成される3次元顔データ）を、物体までの距離を変えながら撮影する。この物体が、ほぼ画像いっぱいになるように焦点距離を調整している。

で、 $r_p = r_a = 2$ である。よって、 d は、 $d = m' - r$ より、 $d_p = d_a = 0$ となる。各カメラモデルを表現するのに必要なパラメータは、 $n_p = 11$, $n_a = 8$ である。ゆえに、これらの数を(5)に代入し、各カメラモデルのAICは、

$$\text{AIC}_p = J[S_p] + 22, \quad (12)$$

$$\text{AIC}_a = J[S_a] + 16 \quad (13)$$

として得られる。これらと比較し、小さい方のモデルを採用する。もちろん、 ϵ を計算し、(7)と(8)を比較しても良い。あるいは、ノイズレベル ϵ を消去して、

$$\frac{J_0[S_p]}{J_0[S_a]} < 1 + \frac{6}{2N - 11} \quad (14)$$

が成立すれば、 S_a を選択する。ここで注目したいのは、このAICが、データの個数 N に依存しないということである。

計算機によるシミュレーションにより、このカメラモデル選択法の妥当性を検証した。

対象物体としては、顔の3次元データを用いた。このデータは、49個の点から構成されている。カメラと物体との距離を8段階に変化させて撮影した。この時、レンズの焦点距離も切りの良い8段階で調整し、この物体がほぼ画面いっぱいになるようにしている(図4)。また、画像上の点に対しては、標準偏差1ピクセルのノイズを加えている。

得られたAICを、図5に示す。縦軸は対数になっている。顔のデータを撮影した場合は、どの場合も射影カメラモデルを採用する方がよいという結果になった。しかし、もう少し遠方から撮影したならば、アフィンカメラモデルを採用した方がよくなると予想される。

次に、同じ設定で、ほぼ平面であるような物体についても調べた(図6)。この場合は、射影カメラモデル

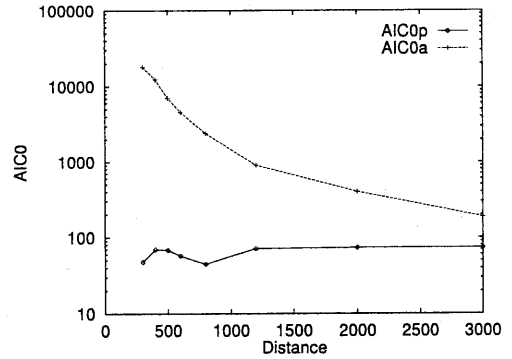


図 5. 3次元物体を撮影し、撮影距離を変化させたときのAICの変化。

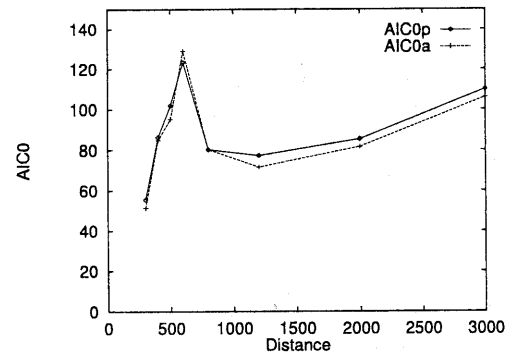


図 6. 平面的な物体を撮影し、撮影距離を変化させたときのAICの変化。

でも、アフィンカメラモデルでも、ほとんど変わりはないが、やや、アフィンカメラモデルを採用した方がよいようである。

ここでは紹介しないが、画像に加えるノイズを大きくしていくと、アフィンカメラモデルを選択する割合が大きくなっていくという現象も確認した。

4.2 未校正ステレオカメラの場合

対象物体の3次元形状は与えられず、未校正の二枚の画像だけが与えられた時(図7)に、それを撮影したカメラが、どのようなカメラモデルでモデル化するのが適切であるかを解析した。

二枚の画像の対応点 x_1^T と x_2^T の間には、エビポー

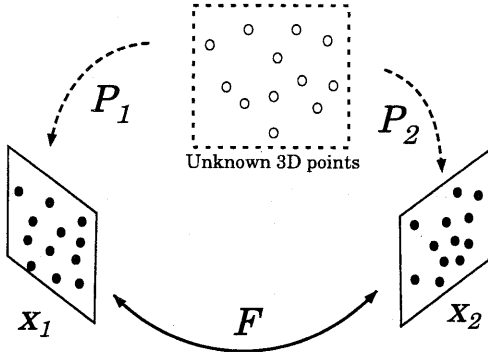


図 7. 3次元物体を、2台のカメラで撮影すると、画像間にはエピポーラ拘束が存在する。それぞれのカメラ行列 P_1, P_2 は一意に決定できないが、基礎行列 F は一意に決定できる。

ラ拘束が存在する。

$$x_2^T F x_1 = 0 \quad (15)$$

F は基礎行列 (fundamental matrix) と呼ばれ、 3×3 行列、ランクが 2 という性質を持つ。

ここで、それぞれの画像を撮影したカメラが、同じカメラモデルでモデル化できると仮定し、射影カメラモデルとアフィンカメラモデルの間で、適切なものを選択する。

二枚の未校正画像からは、それぞれのカメラ行列 P_1, P_2 は、射影的不定性があるため、一意には決まらない。しかし、基礎行列は、一意に決定でき、射影カメラモデルとアフィンカメラモデルで、その形に違いがある。射影カメラモデルの場合、基礎行列の要素には、先に挙げた性質以外には、特に拘束条件は見いだせない。基礎行列の要素の数は 9 であるが、定数倍の不定性を持つことと、ランクが 2 であることを差し引くと、自由度は $m'_p = 7$ であることがわかる。

一方、アフィンカメラモデルの場合は、基礎行列は、

$$F_a = \begin{pmatrix} 0 & 0 & f_{13} \\ 0 & 0 & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix}$$

という形をとる。左上の 2 行 2 列が零行列となるのである。変数の数は 5 に減っている。この場合、ランクが 2 であることは既に満たされているので、定数倍の不定性分だけを差し引いて、実質的な自由度は $m'_a = 4$ になる。

エピポーラ拘束式 (15) は、拘束条件の一つを与えるので、 $r_p = r_a = 1$ である。また、観測データの自由

度は、 x_1, x_2 の u, v 座標の $m'_p = m'_a = 4$ である。ゆえに、 $d_p = d_a = m' - r = 3$ となる。

射影カメラモデルとアフィンカメラモデルの AIC はそれぞれ、

$$AIC_p = J[S_p] + 2(3N + 7), \quad (16)$$

$$AIC_a = J[S_a] + 2(3N + 4) \quad (17)$$

となる。ノイズレベル ϵ を消去したならば、

$$\frac{J_0[S_a]}{J_0[S_p]} < 1 + \frac{6}{N-7} \quad (18)$$

を判定基準とすればよい。(18) が成立すれば、アフィンカメラモデルを採用する。

基礎行列、 J_0, ϵ 等は、射影カメラモデルの場合は [8]² に基づいて計算した。アフィンカメラモデルの場合は

$$(u_{2\alpha}, v_{2\alpha}, u_{1\alpha}, v_{1\alpha}, 1) \begin{pmatrix} f_{13} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{pmatrix} = 0$$

を、一次のくりこみ法を使って計算した。

実画像による実験を示す。立方体形状の対象物体を、カメラの前で運動させ、画像を二枚撮影する。二枚の画像間で人手により、対応点を決定した (図 8)。これら $N=32$ 点の対応点を使って、基礎行列、ノイズレベル ϵ を計算し、 AIC_{0a}, AIC_{0p} を求めた。得られた結果を表 2 に示す。図 8(a) のように、対象物体の奥行きがあり、動きが大きい場合は、射影カメラモデルを採用するべきであるが、図 8(c) のように、奥行きがあまりなく、動きが小さい場合はアフィンカメラモデルで十分であることがわかる。

表 2. 無校正ステレオ画像から得られたノイズレベル ϵ 、AIC、そして、採用すべきカメラモデル

	ϵ_p	ϵ_a	AIC_{0p}	AIC_{0a}	Model
(a)	0.89	5.13	183.4	896.6	PROJ
(b)	0.38	0.59	33.89	39.27	PROJ
(c)	0.61	0.63	85.53	85.08	AFF

²[8], pp.332-337 に、essential matrix の求め方が記されているが、 $\|FF^T\| = \sqrt{2}$ という拘束条件をはずすことで、fundamental matrix を計算することができる。

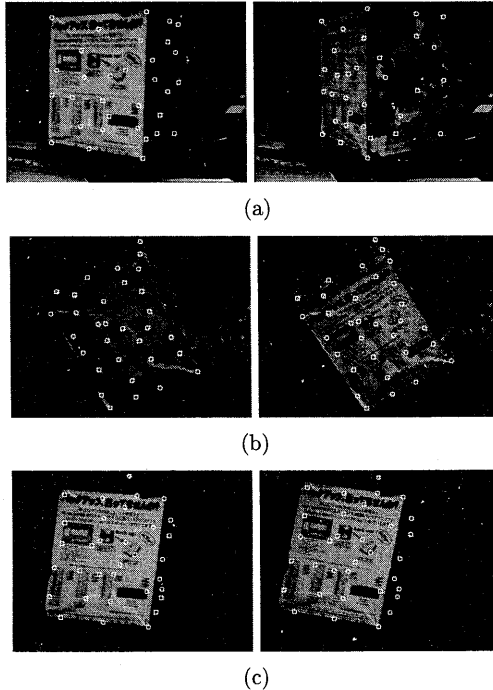


図 8. 立方体形状の物体を運動させて、その画像を二枚撮影する。画像間の対応点 32 点は人手でとった。これをカメラから物体までの距離を変化させて撮影した。物体が画像上でほぼ同じ大きさになるように、焦点距離も変化させている。

5 おわりに

金谷の幾何学的 AIC を使用したカメラモデル選択方法について述べた。本手法は、カメラに関する知識をいっさい必要としないにも関わらず、適切なカメラモデルを決定することができる。さらに、ステレオカメラの場合は、対象に関する知識をも必要とせず、画像だけから、適切なカメラモデルを決定できる。適切なカメラモデルを採用することで、3次元復元や、新規画像生成などのアプリケーションにおいて、状況に適したアルゴリズムを選ぶことができ、「牛刀を以て鶏を割く」といった事態を避けることができる。

現状では、ステレオカメラの場合には、基礎行列を解析して、射影カメラモデルか、アフィンカメラモデルかを判断している。しかし、特別な対象、特別なカメラ配置の場合には、それぞれのカメでは、射影カメラモデルでモデル化した方がいい場合も起こり得る。

このような特異な状況を、今後、詳しく解析していくつもりである。

謝辞

ATR 人間情報通信研究所 Martin Tonko 研究員には実画像データを、磯野勝宣研究員には 3次元顔データを提供していただきました。各種カメラモデルの定義に関して、有益な議論、アドバイスをいただいた、名古屋工業大学 佐藤淳助教授に心から感謝します。

参考文献

- [1] 木村英紀. モデルとは何か. 数理学, Vol. 36, No. 9, pp. 5-10, 1998.
- [2] 佐藤淳. 視覚の幾何学. コロナ社, 1999.
- [3] Olivier Faugeras. Stratification of three-dimensional vision: projective, affine, and metric representations. *Journal of the Optical Society of America A*, Vol. 12, No. 3, pp. 465-484, 1995.
- [4] Olivier Faugeras. From geometry to variational calculus: theory and applications of three-dimensional vision. In *Computer Vision for Virtual Reality Based Human Communications*, pp. 52-70, 1998.
- [5] 金谷健一. 情報量基準による幾何学的推論. 情報処理学会研究報告 AI-101-1, pp. 1-6, 1995.
- [6] 金谷健一. 幾何学的モデル選択の情報量基準. 信学技報 IT-95-16-7, pp. 19-24, 1995.
- [7] 金谷健一. 幾何学的モデルの選択基準について. 情報処理学会研究報告 MSP-2-5, pp. 27-32, 1995.
- [8] Kenichi Kanatani. *Statistical Optimization for Geometric Computation: Theory and Practice*. Elsevier Science, 1996.
- [9] P.H.S. Torr. Geometric motion segmentation and model selection. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, Vol. 356, No. 1740, pp. 1321-1340, 1998.
- [10] Phil Torr, Andrew W. Fitzgibbon, and Andrew Zisserman. Maintaining multiple motion model hypotheses over many views to recover matching and structure. In *International Conference on Computer Vision*, pp. 485-491, 1998.
- [11] 木下敬介, 佐藤淳. 疑似透視カメラモデルの提案. 画像の認識・理解シンポジウム (MIRU), pp. I-249-254, 1998.
- [12] Kenichi Kanatani. Statistical optimization and geometric inference in computer vision. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, Vol. 356, No. 1740, pp. 1303-1320, 1998.
- [13] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Automatic Control*, Vol. 19, No. 6, pp. 716-723, 1974.