

パラメトリック固有空間法による物体認識のための 複数カメラ配置の決定手法

高橋 友和^{1,2}, Lina¹, 松ヶ野 修功¹, 井手 一郎¹, 目加田 慶人³, 村瀬 洋¹

¹名古屋大学大学院情報科学研究科 〒464-8603 名古屋市千種区不老町

²日本学術振興会

³中京大学生命システム工学部

E-mail: ¹ttakahashi@murase.m.is.nagoya-u.ac.jp

あらまし 複数のカメラを用いて物体認識を行う場合、認識率はカメラの配置に大きく影響を受ける。本稿では、複数のカメラを用いて精度良く物体認識を行うためのカメラ配置の決定手法を提案する。物体認識のフレームワークとしてはパラメトリック固有空間法を用いる。パラメトリック固有空間法では、学習段階において固有空間上に各物体に対応する多様体を形成する。提案手法は、これらの多様体の空間的な分布の形状特徴に基づいた評価関数を用いてカメラ配置を決定するものである。多様体間の距離に基づく評価関数を定義し、カメラを2台用いる場合、3台用いる場合についてカメラ配置決定に関する実験を行った。実験の結果から提案手法が認識性能の向上に有効であることを示す。

キーワード 物体認識, 複数カメラ, カメラ配置, パラメトリック固有空間法

Planning of Multiple Camera Positions for Object Recognition by Parametric Eigenspace Method

Tomokazu TAKAHASHI^{1,2}, Lina¹, Osanori MATSUGANO¹, Yoshito MEKADA³ and Hiroshi MURASE¹

¹Graduate School of Information Science, Nagoya University Furou-cho, Chikusa-ku, Nagoya, 464-8603, Japan

²Japan Society for the Promotion of Science

³Department of Life System Science and Technology, Chukyo University, Japan

E-mail: ¹ttakahashi@murase.m.is.nagoya-u.ac.jp

Abstract When objects are recognized by using multiple cameras, recognition rates strongly depend on the camera arrangement. In this paper, we propose a new method for planning a multiple camera arrangement for accurate recognition. We use a parametric eigenspace method for the recognition framework in which objects are represented as manifolds in an eigenspace. The proposed method evaluates the adequacy of camera arrangement according to the relations between the manifolds in the eigenspace. In the experiments, we defined a function that measures relations by the distances between manifolds. The experimental results show the effectiveness of the proposed method.

Keyword Object recognition, Multiple cameras, Camera arrangement, Parametric eigenspace method

1. はじめに

3次元物体の認識は監視システムやヒューマンインタフェース、産業応用の分野で重要な技術である。物体認識を実現する手法では様々なアプローチが用いられるが、我々はノイズに対しロバストなアピアランススペースの物体認識に着目する。パラメトリック固有空間法[1]はその中で広く用いられる手法の一つであり、物体の認識とその物体の姿勢、光源方向などのパラメータの推定を同時に行う手法である。この手法に関して様々な関連研究が報告されている[2, 3, 5]

アピアランススペースの物体認識に関して、認識性能を低下させる主な要因として以下の2つが考えられる。一つは撮

影状況に起因する入力画像の品質低下(図1)であり、もう一つは入力画像が撮影されたカメラの位置が物体の識別に適さない場合に生じる(図2)。認識を行う際に複数の入力画像を用いることにより、これらの問題の解決を図る試みが報告されている。山口らは顔認識の入力画像として、ビデオシーケンス中の複数のフレーム画像を用いている[4]。アクティブリコグニション[5, 6]のフレームワークでは、認識の過程において適応的にカメラ位置を移動させながら複数の入力画像を得ることにより、認識性能の向上を図っている。

一方、複数の入力画像を得る際により一般的に用いられる方法は、あらかじめ配置された複数のカメラから入力画像

を得るものである。Selinger らは複雑な背景を含むシーン中の物体認識の問題に複数のカメラを用いている[7]。Shakhnarovich らは複数のカメラを用いて人物の顔と歩行姿勢の認識を行っている[8]。実際の応用を考えた場合、認識性能はカメラの配置と認識対象となる物体の形状特徴やテクスチャ特徴に依存すると考えられるため、あらかじめ設置するカメラの配置は慎重に検討しなければならない。認識に適したカメラ配置は、各カメラの位置を変化させながら認識実験を繰り返し行うことによって決定することも考えられるが、カメラ位置の組み合わせ方はカメラの台数乗で増加し、これらすべてに対して認識実験を行うのは現実的ではない。

そこで、我々は複数のカメラを用いて精度良く物体認識を行うためのカメラ配置の決定手法を提案する。認識にはパラメトリック固有空間法を用いる。提案手法は、学習段階において固有空間上に形成される多様体の情報のみを用いて、与えられたカメラ配置を評価し、実際の認識実験を行うことなく適切なカメラ配置を決定するものである。以下、2.で複数カメラを用いる場合のパラメトリック固有空間法について述べ、3.で提案手法の詳細を述べる。4.で実験の方法を述べ、5.で結果の考察を行い、6.で本稿をまとめる。

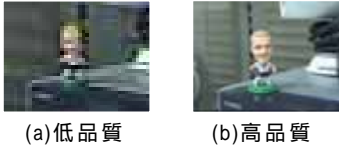


図 1. 撮影状況に依存する画像の品質の違い



図 2. カメラ位置の違いによる見え方の違い

2. 複数カメラを用いたパラメトリック固有空間法

図 3 に本稿で用いるカメラの座標系を示す。各カメラの位置は ϕ_m^h と ϕ_m^v の 2 つのパラメータで表される。これらはそれぞれ水平方向、垂直方向のカメラの角度を表す。以下、複数カメラを用いたパラメトリック固有空間法について述べる。

2.1. 学習段階

各学習画像に対し、画素値を要素とした特徴ベクトル \hat{x} を $x = \hat{x} / \|\hat{x}\|$ によって正規化したものから行列 X を以下のように作成する。

$$X = [x_{1,1}^{(1)} - c, \dots, x_{h,v}^{(p)} - c, \dots, x_{H,V}^{(P)} - c] \quad (1)$$

ここで、 h, v はそれぞれ物体 $p (=1, 2, \dots, P)$ の水平方向、垂直方向の姿勢を表す識別子である。 c はすべての h, v, p に関する x の平均ベクトルである。ユニバーサル固有空間を XX^T の固有値の上位 k 個に対応する固有ベクトル e_i ($i=1, 2, \dots, k$) から形成する。式(2)を用いて各特徴ベクトル $x_{h,v}^{(p)}$ を固有空間上の点 $g_{h,v}^{(p)}$ に投影する。

$$g_{h,v}^{(p)} = [e_1, e_2, \dots, e_k]^T (x_{h,v}^{(p)} - c) \quad (2)$$

$g_{h,v}^{(p)}$ から 3 次スプライン補間を用いて、各物体に対応する多様体 $G^{(p)}(\theta)$ を形成する。ここで、 $\theta = (\theta^h, \theta^v)$ は物体の姿勢パラメータベクトル、あるいは物体撮影時のカメラ位置を表す。

同様に各物体に対して、その物体に関する学習画像のみを用いてオブジェクト固有空間を形成し、オブジェクト固有空間上に多様体 $G_O^{(p)}(\theta)$ を形成する。これは、物体の姿勢パラメータ推定に用いる。

2.2. 複数カメラを用いる認識段階

カメラ位置を既知とする M 台のカメラを用いた物体認識について説明する。入力される物体の姿勢パラメータベクトル $\tilde{\theta}$ が未知である場合を想定し、 $m (=1, 2, \dots, M)$ 台目のカメラ位置 $\phi_m = (\phi_m^h, \phi_m^v)$ を 1 台目のカメラからの相対位置 $\alpha_m = \phi_m - \phi_1$ で表す。学習段階と同様に、 m 台目のカメラから入力される画像の特徴ベクトル y_m をユニバーサル固有空間上の点 z_m に以下の式を用いて投影する。

$$z_m = [e_1, e_2, \dots, e_k]^T (y_m - c) \quad (3)$$

z_m と $G^{(p)}(\theta)$ の距離を以下のように定義する。

$$d^{(p)}(\theta) = \sum_{m=1}^M \left\| z_m - G^{(p)}(\theta + \alpha_m) \right\| \quad (4)$$

認識結果として、 \tilde{p} を以下の式で求める。

$$\tilde{p} = \arg \min_p \left(\min_{\theta} d^{(p)}(\theta) \right) \quad (5)$$

姿勢パラメータの推定に関して、1 台目のカメラ位置 ϕ_1 に

対する入力された物体の姿勢パラメータベクトル $\tilde{\theta}$ は以下の式を用いて推定する。

$$\tilde{\theta} = \arg \min_{\theta} d_O^{(\tilde{p})}(\theta) \quad (6)$$

ここで、 $d_O^{(\tilde{p})}(\theta)$ は物体 \tilde{p} に関するオブジェクト固有空間上の投影点と多様体の間の距離を表すもので、式(4)の $G^{(p)}(\theta)$ をオブジェクト固有空間上の多様体 $G_O^{(p)}(\theta)$ に置き換えたものである。

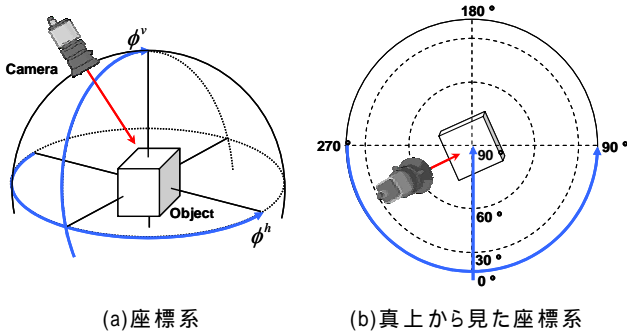


図 3. カメラ位置の座標系

3. 物体認識に適切なカメラ配置の決定

3.1. パラメトリック固有空間上でのカメラ配置の決定

複数カメラを用いた物体認識を精度良く行うためのカメラ配置の決定手法を提案する。提案手法は、学習段階において形成されたユニバーサル固有空間上の多様体の情報のみから、与えられたカメラ配置の認識に対する適切さを評価し、カメラ配置を決定するものである。

求めたいカメラ配置をカメラ位置の集合 $\tilde{\alpha} = \{\tilde{\alpha}_m\}$ で表すとき、提案手法は $\tilde{\alpha}$ を以下の式を用いて決定する。

$$\tilde{\alpha} = \arg \max_{\alpha} F(G, \alpha) \quad (7)$$

$G = \{G^{(p)}(\theta)\}$ はユニバーサル固有空間上の多様体の集合を表し、 $\alpha = \{\alpha_m\}$ はすべての可能なカメラ配置を表す。そして、 $F(G, \alpha)$ はカメラ配置の適切さを測る評価関数である。

3.2. 多様体間距離に基づいたカメラ配置の評価関数

ユニバーサル固有空間上には、複数の多様体同士が互いに交差、あるいは極端に接近している領域が存在する。認識の際、入力された特徴ベクトルがこれらの領域に投影された場合、一般に認識は困難となる場合が多い。このこと

から、カメラ配置に関する評価関数の 1 つとして、多様体間距離に基づく評価関数を定義する。

M 台のカメラによる任意の配置 α が与えられたとき、2 つの物体 p, q に対応する多様体間の距離を以下の式で定義する。

$$D^{(p,q)}(\alpha) = \min_{\theta_p, \theta_q} \sum_{m=1}^M \left\| G^{(p)}(\theta_p + \alpha_m) - G^{(q)}(\theta_q + \alpha_m) \right\| \quad (8)$$

この式から、 P 個の多様体が与えられたとき、各多様体間の距離の最小値を用いて、カメラ配置 α に対するそれらの多様体同士の分散の度合いを以下の式で定義する。ただし、 $p = (1, \dots, P), q = (1, \dots, P), p \neq q$ とする。

$$D(\alpha) = \min_{p,q} D^{(p,q)}(\alpha) \quad (9)$$

式(9)は与えられたカメラ配置 α に対する物体の認識の容易さを表すと考えられるため、カメラ配置に関する評価関数の一つとして考えることができる。式(7)と式(9)より、 $F(G, \alpha) = D(\alpha)$ として、求めたいカメラ配置 $\tilde{\alpha}$ を以下のように決定する。

$$\tilde{\alpha} = \arg \max_{\alpha} D(\alpha) \quad (10)$$

式(8), (9), (10)をまとめ、以下の式を得る。

$$\tilde{\alpha} = \arg \max_{\alpha} \min_{p,q} \min_{\theta_p, \theta_q} \sum_{m=1}^M \left\| G^{(p)}(\theta_p + \alpha_m) - G^{(q)}(\theta_q + \alpha_m) \right\| \quad (11)$$

4. 実験方法

カメラが 2 台の場合と 3 台の場合について、提案手法によって決定されたカメラ配置を用いて認識実験を行うことにより、提案手法の有効性を調査した。カメラ配置の評価関数として式(9)を用いた。72 種類の物体を図 4 に示す 4 つの物体セットに分割して用いた。物体セットの内訳を以下に示す。

- **FP**(Football Players): 27 種のサッカー選手人形
- **CC**(Cartoon Characters): 15 種の漫画キャラクター人形
- **WB**(Wooden Blocks): 9 種の積み木
- **FC**(Faces): 21 人の顔画像

学習段階では、各物体について、カメラ位置を水平方向 0 ~ 350 度を 10 間隔、垂直方向 0 ~ 90 度を 15 度間隔で変化させながら、合計 252 枚の画像を得た。各画像から物体が外接する最小の正方形領域を切り出し、画像サイズをそれぞれ 64 x 64 に伸縮させたものを学習画像とした。各物体セットに対してユニバーサル固有空間を形成し、各物体に対応する多様体を物体が属する物体セットの固有空間上

に形成した。このとき、固有空間の次元数を 4 とした。本実験では、パラメトリック固有空間法による物体の認識のみを扱い、物体の姿勢パラメータ推定は扱わなかったため、物体毎のオブジェクト固有空間は形成しなかった。

カメラ配置の決定の際には、各カメラの位置を水平方向 12 段階 (0 ~ 330 度, 30 度間隔), 垂直方向 4 段階 (0 ~ 90 度, 30 度間隔) で変化させ、それらの組み合わせ数分のカメラ配置を用意した。カメラを 2 台用いる場合と 3 台用いる場合について、用意したすべてのカメラ配置の中から提案手法を用いて適切なカメラ配置を決定した。多様体間距離を算出する際、任意の異なる多様体上の点同士の距離を算出する必要があるが、本実験では 3 次スプライン曲面で表された各多様体上に 360 (水平方向) × 90 (垂直方向) の点群をサンプリングし、それらの点同士の距離を近似的に用いた。

認識段階では、学習画像を撮影したカメラ位置を水平方向に 5 度ずらした位置から撮影した画像に対し、学習画像と同様に物体領域を正方形で切り出した。学習時と認識時の撮影環境の違いを想定して、それらの画像をサイズ 32 × 32, 16 × 16, 8 × 8 の 3 段階に伸縮させたものを入力画像とした (図 5)。入力画像を固有空間に投影する際には、線形補間を用いて画像のサイズを 64 × 64 に伸長したのから特徴ベクトルを形成し、それを投影した。各物体セット、各画像サイズに対して、提案手法によって決定されたカメラ配置と、それ以外のカメラ配置を用いた場合について、入力する物体の種類、ならびにその物体の姿勢パラメータを変化させながら認識を繰り返し行い、実際の認識率を算出した。

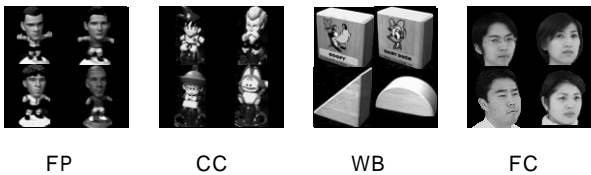


図 4. 実験で用いた物体の例



図 5. 実験で用いた 3 段階の入力画像サイズ

5. 結果と考察

表 1 は、物体セット FC, 画像サイズ 16 × 16 に対して、カメラを 3 台用いて行った実験に関して、そこで用いられたカメラ配置の例とそのときの認識率を示している。カメラ配置 (c) は提案手法によって決定されたものであり、他のカメラ配置 (b), (c) を用いた場合よりも高い認識率が得られた。図 6 は、物体セット FC, 入力画像サイズ 16 × 16 に対して、カメラを 3 台用いた実験において用いられたすべてのカメラ配置に関して、式 (9) で算出されたカメラ配置の評価値 (Camera arrangement adequacy) とそのカメラ配置を用いて実際に調べた認識率 (Recognition rate) の関係を示している。図 6 の (a), (b), (c) はそれぞれ表 1 のカメラ配置に対応している。この図から、評価値の高いカメラ配置からは高い認識率が得られ、認識率の低いカメラ配置の評価値は低いという関係が見られる。この関係は、物体セットや入力画像サイズに依らず、本実験のすべての場合において確認された。このことから、提案手法を用いて評価値からカメラ配置を決定することは、認識性能の向上に有効であるといえる。

表 1. 実験で用いた 3 台のカメラ配置の例 (物体セット FC, 入力画像サイズ 16 × 16): (c) は提案手法によって決定されたカメラ配置

カメラ配置	(a)	(b)	(c)
認識率 [%]	59.0	83.7	100.0

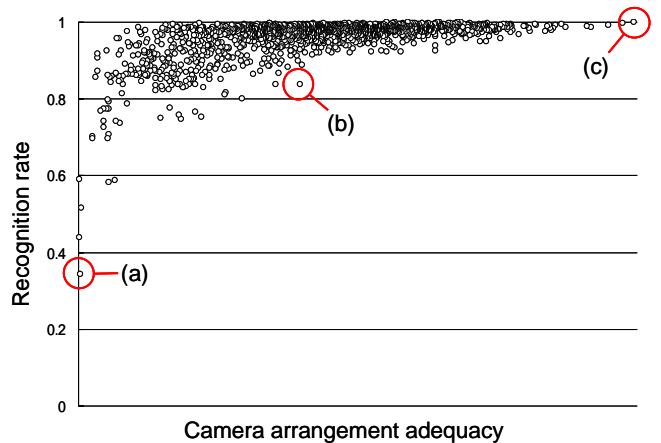


図 6. 3 台のカメラを使用した場合の認識率とカメラ配置の評価値の関係 (物体セット FC, 画像サイズ 16 × 16): (a), (b), (c) は表 1 に示したカメラ配置にそれぞれ対応している

表 2, 表 3 はそれぞれカメラ 2 台を用いた場合と 3 台を用いた場合について, 提案手法によって決定されたカメラ配置(最大の評価値を持つカメラ配置)と不適切なカメラ配置(最小の評価値を持つカメラ配置)を用いたときの認識率を示している. 適切でないカメラ配置を用いた場合と比較して, 提案手法によって決定されたカメラ配置を用いることにより, 入力画像サイズが小さく認識が困難な場合でも精度良く認識が行われていることが確認できる.

表 2. 提案手法により決定されたカメラ配置を用いることによる認識率の向上(カメラ 2 台)

入力画像サイズ	物体セット	認識率 [%]	
		不適切なカメラ配置	適切なカメラ配置 (提案手法)
32 × 32	FP	71.5	99.4
	CC	74.6	97.0
	WB	86.7	96.9
	FC	90.4	99.9
16 × 16	FP	36.8	95.7
	CC	45.7	93.0
	WB	65.8	84.4
	FC	59.0	94.3
8 × 8	FP	20.3	61.4
	CC	18.5	60.7
	WB	49.4	75.0
	FC	28.0	48.8

表 3. 提案手法により決定されたカメラ配置を用いることによる認識率の向上(カメラ 3 台)

入力画像サイズ	物体セット	認識率 [%]	
		不適切なカメラ配置	適切なカメラ配置 (提案手法)
32 × 32	FP	71.5	99.4
	CC	74.6	98.1
	WB	86.7	97.5
	FC	90.4	100.0
16 × 16	FP	36.8	99.4
	CC	45.7	95.7
	WB	65.8	90.0
	FC	59.0	100.0
8 × 8	FP	20.3	89.8
	CC	18.5	76.1
	WB	49.4	64.2
	FC	28.0	59.3

6. まとめ

複数カメラを用いた物体認識に関して, 精度良く認識を行うためのカメラ配置の決定方法を提案した. 提案手法は, 適切なカメラ配置を評価関数によって決定するものであり, そのような評価関数の一例として, 多様体間距離に基づくシンプルな評価関数を定義した. その評価関数を用いて, カメラが 2 台の場合と 3 台の場合に対するカメラ配置の決定に関する実験を行い, 実験の結果から, 提案手法が認識性能の向上に有効であることを示した.

今後は, カメラ配置決定のためのより正確で有用な評価関数の設計, 学習サンプル数が変化した場合の影響の調査を検討している.

謝 辞

日頃より熱心に御討論頂く名古屋大学村瀬研究室諸氏に感謝する. 本研究の一部は日本学術振興会科学研究費補助金, 21 世紀 COE プログラム「社会情報基盤のための音声・映像の知的統合」による.

本研究では, 画像処理にソフトウェアライブラリ MIST(<http://mist.suenaga.m.is.nagoya-u.ac.jp/>)を使用している.

文 献

- [1] H. Murase and S. K. Nayar, "Visual Learning and Recognition of 3-D Object Recognition from Appearance," *Internat. J. Comput. Vision*, vol. 14, pp. 5-24, 1995.
- [2] H. Murase and S. K. Nayar, "Illumination Planning for Object Recognition using Parametric Eigenspaces," *IEEE Trans. PAMI*, vol. 16, no. 12, pp.1218-1227, 1994
- [3] H. Tanaka, I. Kitahara, H. Saito, H. Murase, K. Kogure, and N. Hagita, "Dynamic Visual Learning for People Identification with Sparsely Distributed Multiple Surveillance Cameras," *Proc. SCIA2005*, pp. 130-140, 2005.
- [4] O. Yamaguchi and K. Fukui, "Smartface" – A Robust Face Recognition System under Varying Facial Pose and Expression," *IEICE Trans. Inf. Sys.*, vol. E86-D, no.1, pp. 37-44, 2003
- [5] H. Borotschnig, L. Paletta, M. Prantl, and A. Pinz, "Active Object Recognition in Parametric Eigenspace," *Proc. BMVC'98*, vol.2, pp. 629-638, 1998.
- [6] F. G. Callari and F. P. Ferrie, "Autonomous Recognition: Driven by Ambiguity," *Proc. CVPR'96*, pp.701-707, 1996.
- [7] A. Selinger and R. C. Nelson, "Appearance-based Object Recognition using Multiple Views," *Proc. CVPR2001*, vol. 1, pp. 905-911, 2001.
- [8] G. Shakhnarovich, L. Lee, and T. Darrell, "Integrated Face and Gait Recognition from Multiple Views," *Proc. CVPR2001*, vol. 1, pp. 39-46, 2001.