グラスマン距離を用いた部分空間法による局所特徴量のモデリング

菊次　優太†　　ライチェフビセル††　　玉木　徹†††　　金田　和文††††

† 広島大学大学院工学研究科情報工学専攻
〒 739-8527　広島県東広島市鏡山 1-4-1
E-mail: †kikutsugi@eml.hiroshima-u.ac.jp, ††{bisser,tamaki,kin}@hiroshima-u.ac.jp

あらまし　本論文では画像から抽出された局所特徴量のモデリングを部分空間法で行う，Subspace of Features (SoF) を提案する．SoF において部分空間同士の類似度としてグラスマン距離を用いる，Grassmann distance Mutal Subspace Method (GD–MSM)，Grassmann Kernel Support Vector Machine (GK–SVM) を提案する．本論文ではスケール変化，ビュー変化，照明変化を受け雑多な背景を持つ 10 物体のデータベースで実験を行い，提案手法と Bag–of–Keypoints (BoK) の性能を比較した．認識性能においては BoK と SoF で同程度であること示し，計算時間は提案手法である SoF が大幅に早いことを示した．

キーワード　グラスマン多様体，グラスマンカーネル，正準角，正準相関，部分空間法，Bag–of–Keypoints，不変局所特徴量

1. はじめに

近年コンピュータビジョンやパターン認識の分野において SIFT [1] や SURF [2]，HOG [3] などの局所特徴量が広く使われている．これらは画像のレジストレーションや 3 次元復元，画像検索，ロボットナビゲーション，物体認識など様々な分野で使われている [4]．しかし局所特徴量，物体認識において，画像全体をベクトル化して，特徴量として用いるアピアランスベース手法 [5] のように直接的には認識に使えないという欠点がある．これは，物体の画像から抽出された特徴点の数が違うことに起因しており，識別器に適したベクトル型の特徴量を作る必要がある．近年，この問題に対する解決法の一つとして bag-of-keypoints (BoK) [6] が提案されている．BoK ではまずすべての学習画像から抽出した特徴量をクラスタリングする．そしてそれぞれの画像をクラスタの重心 (visual words [7]) を bin とするヒストグラムで表す．

本論文では，局所特徴量のモデリングを BoK とは異なる枠組みで行う手法を提案する．すなわち，画像から抽出された特徴量の集合を部分空間で表す Subspace of Features (SoF) を提案する．二つの部分空間を比較するためには，部分空間同士の最適な距離，もしくは類似度を決定する必要がある．部分空間同士の距離 [8], [9] は正準角で表現され，いくつかの手法が提案されている．相互部分空間法 (MSM) [10] では最小の正準角の余弦を用いて二つのクラスの類似度を決定する．

また近年，ユークリッド空間の k–次元の線形部分空間集合であるグラスマン多様体 [11] の幾何的構造を考慮した様々な距離および部分空間同士の正準角の表現が提案されている [12], [13]. [12] では projection metric と Binet Cauchy metric がグラスマン多様体における有用なメトリックであると示されており，対応するカーネル関数は projection kernel と binet-cauchy kernel である．これらのグラスマンカーネルは kernel LDA (Linear Discriminant Analysis) で用いられ，Grassman discriminant analysis (GDA) と呼ばれる．

本論文では，抽出された局所特徴量の集合を部分空間で表し，部分空間同士の類似度をグラスマン距離で決定する Grassmann Distance Mutual Subspace Method (GD-MSM) を提案する．さらにより識別能力の高い学習関数を得るために Support Vector Machine (SVM) [14] にグラスマンカーネルを適用する Grassmann Kernel Support Vector Machine(GK–SVM) も提案する．

本論文では GD–MSM および GK–SVM の性能を比較するため１０物体のデータベースを作成し，物体認識実験を行う．二つのパターンの識別辞書を用意し，その影響について調べる．一つ目は全ての画像を一つの識別辞書とする (各クラスの全ての画像から共通の部分空間を学習する)．二つ目の手法では複数の識別辞書を作成する (つまり各画像ごとに部分空間を作る)．

本論文の構成は次の通りである．2 節では提案手法と関連する用語について説明する．3 節では提案手法を用いた物体認識実験とその結果について述べる．

2. 提案手法

本節ではグラスマン距離及びグラスマンカーネルについて説明する．3.2 では BoK の代わりに SoF (Subspace of Features) を実現するために，GD–MSM，GDA，GK–SVM について説明する．

2.1 準　　備

二つの画像から抽出された局所特徴量の集合を $S_i =$

$\{\boldsymbol{x}_1^i, \boldsymbol{x}_2^i, \cdots, \boldsymbol{x}_n^i\}$ および $S_j = \{\boldsymbol{x}_1^j, \boldsymbol{x}_2^j, \cdots, \boldsymbol{x}_m^j\}$ とする．ここで $\boldsymbol{x}_n^i$ は i 番目の画像内の n 番目の特徴量である．それぞれの画像における特徴量の集合全体はユークリッド空間 $\mathcal{R}^D$ 中の対応する部分空間 $span(Y_i) = span\{\boldsymbol{u}_1, ..., \boldsymbol{u}_p\}$ および $span(Y_j) = span\{\boldsymbol{v}_1, ..., \boldsymbol{v}_p\}$ で表現することができる．ここで $span(Y_j)$ は $D \times p$ の行列 $\boldsymbol{Y}_i = [\boldsymbol{u}_1, ..., \boldsymbol{u}_p]$ の列ベクトルで張られる部分空間であり，$\{\boldsymbol{u}_1, ..., \boldsymbol{u}_p\}$ は正規直交である．

$\mathcal{R}^D$ の m 次元の線形部分空間の集合はグラスマン多様体 $G = (m, D)$ と呼ばれる．部分空間 $span(Y_i)$ および $span(Y_j)$ はグラスマン多様体 $G = (m, D)$ 上の 2 点と考えることができる．グラスマン多様体における様々な距離が [13] において定義されており，そのすべてが部分空間同士の正準角として表現することが出来る．

部分空間 $span(Y_1)$ と $span(Y_2)$ の正準角 $(0 \leqq \theta_1 \leqq \cdots \leqq \theta_m \leqq \pi/2)$ は以下で再帰的に定義される

$$\cos\theta_k = \max_{\boldsymbol{u}_k \in span(Y_1)} \max_{\boldsymbol{v}_k \in span(Y_2)} \boldsymbol{u}_k^t \boldsymbol{v}_k, \tag{1}$$

また以下の制約条件を持つ．

$$\begin{aligned} &\boldsymbol{u}_k^T \boldsymbol{u}_k = 1, \boldsymbol{v}_k^T \boldsymbol{v}_k = 1 \\ &\boldsymbol{u}_k^T \boldsymbol{u}_i = 0, \boldsymbol{v}_k^T \boldsymbol{v}_i = 0, (i = 1, ..., k-1) \end{aligned} \tag{2}$$

正準角は $\boldsymbol{Y}_1^T \boldsymbol{Y}_2$ の特異値分解により計算することが出来る．

$$\boldsymbol{Y}_1^T \boldsymbol{Y}_2 = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{V}^T, \Lambda = diag(\lambda_1, ..., \lambda_m) \tag{3}$$

ここで正規直交行列 $\boldsymbol{Y}_1$ および $\boldsymbol{Y}_2$ は $span(Y_1)$ および $span(Y_2)$ の行列表現である．そして $\lambda_i = cos\theta_i$ は正準角 θ_i の余弦であり，正準相関として知られている．

本論文では次のグラスマン距離について考える．これらは全て部分空間同士の正準角で表すことが出来る．

1. projection metric (正準角の正弦の 2 ノルム)

$$d_p(Y_1, Y_2) = \left(\sum_{i=1}^{m} \sin^2\theta_i \right)^{\frac{1}{2}} = \left(m - \sum_{i=1}^{m} \cos^2\theta_i \right)^{\frac{1}{2}} \tag{4}$$

2. Binet–Cauchy metric (正準角の余弦の積を用いる)

$$d_{BC}(Y_1, Y_2) = (1 - \Pi_i \cos^2\theta_i)^{\frac{1}{2}} \tag{5}$$

3. 最大相関 (最小の正準角のみを用いる，MSM [7] と同等)

$$d_{Max}(Y_1, Y_2) = (1 - \cos^2\theta_i)^{\frac{1}{2}} = \sin\theta_i \tag{6}$$

4. 最小相関 (最大の正準角の正弦のみを用いる)

$$d_{Min}(Y_1, Y_2) = (1 - \cos^2\theta_m)^{\frac{1}{2}} = \sin\theta_m \tag{7}$$

5. Procrustes (chordal) distance (二つの異なる部分空間 $\mathrm{span}(Y_1)$ と $\mathrm{span}(Y_2)$ の最小の距離，フロベニウスノルムを用いる)

$$\begin{aligned} d_{CF}(Y_1, Y_2) &= \min_{R_1, R_2 \in O(m)} ||Y_1 R_1 - Y_2 R_2||_F \\ &= 2\left(\sum_{i=1}^{m} \sin^2(\theta_i/2) \right)^{\frac{1}{2}} \end{aligned} \tag{8}$$

6. 行列の 2 ノルムを用いた Procrustes (chordal) distance

$$\begin{aligned} d_{C2}(Y_1, Y_2) &= \min_{R_1, R_2 \in O(m)} ||Y_1 R_1 - Y_2 R_2||_2 \\ &= 2\sin(\theta_m/2) \end{aligned} \tag{9}$$

7. geodesic distance (グラスマン多様体上を結ぶ最短測地線の長さ)

$$d_G(Y_1, Y_2) = \sum_{i=1}^{m} \theta_i^2 \tag{10}$$

8. 平均距離

$$d_{Mean}(Y_1, Y_2) = \frac{1}{m} \sum_{i=1}^{m} \sin^2\theta_i \tag{11}$$

第 4 節において物体認識実験を行い，上記の 8 つのグラスマン距離について比較する．

[12] では projection metric (4) および Binet–Cauchy metric (5) は次の正定値グラスマンカーネルとして用いられている．

1. Projection kernel

$$k_p(Y_1, Y_2) = ||\boldsymbol{Y}_1^T \boldsymbol{Y}_2||_F^2 \tag{12}$$

2. Binet–Cauchy kernel

$$\begin{aligned} k_{BC}(Y_1, Y_2) &= det(\boldsymbol{Y}_1^T \boldsymbol{Y}_2) \\ &= det(\boldsymbol{Y}_1^T \boldsymbol{Y}_2 \boldsymbol{Y}_2^T \boldsymbol{Y}_1) = \Pi_i \cos^2\theta_i \end{aligned} \tag{13}$$

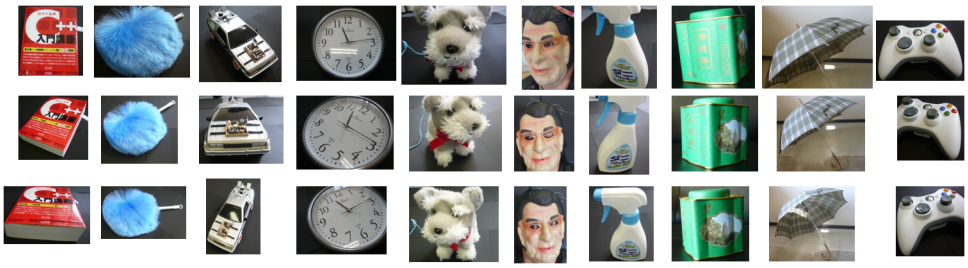
これらのカーネルは様々なカーネル手法 [18] に用いることができる．[12] では kernel LDA に用いられているが，次節では SVM [16] にこれらのカーネルを用いる．

2.2 Grassmann distance Mutal Subspace Method (GD–MSM)

本論文で提案する Grassmann Distance Mutal Subspace Method (GD–MSM) は前節で紹介したグラスマン距離を用いて MSM を拡張した手法である．GD–MSM では最小の正準角のみを用いる代わりに，部分空間同士の全ての正準角を計算に用いる．本論文では式 (4) から (11) で示した 8 つの距離を用いる．次節で示す実験結果においては mean distance が最も高い性能を示した．

2.3 Grassmann Kernel Support Vector Machine (GK–SVM)

本節で説明する Grassmann Kernel Support Vector Machine (GK–SVM) は SVM に Grassmann Kernel を用いた手法である．まず 2 クラス識別問題について考え



(a) 学習画像セット

(b) テスト画像セット

図 1　実験に用いたデータベース

る．学習セット $S = \{(\boldsymbol{Y}_1, y_1), ..., (\boldsymbol{Y}_N, y_N)\}$ が与えられたとする．ここで $\boldsymbol{Y}_1$ は $span(Y_i)$ の行列表現であり i 番目の画像に対応している．$y_i = \{-1, 1\}$ はクラスラベルである．このとき SVM は次の最適化問題を解く．

$$\underset{\boldsymbol{w},\boldsymbol{\xi},b}{min}\ \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{N}\xi_i \qquad (14)$$

$$subject\ to\ y_i(\boldsymbol{w}^T\phi_i + b) \geqq 1 - \xi_i, \xi_i \geqq 0$$

また双対表現は次のように与えられる．

$$\underset{\boldsymbol{\alpha}}{min}\ \frac{1}{2}\boldsymbol{\alpha}^T\boldsymbol{Q}\boldsymbol{a} - \boldsymbol{1}^T\boldsymbol{\alpha} \qquad (15)$$

$$subject\ to\ \boldsymbol{y}^T\boldsymbol{\alpha} = 0, 0 \leqq \alpha_i \leqq C$$

式 (15) および (16) において，$\boldsymbol{\xi}$ はマージンスラックベクトルで，ϕ は特徴空間の変換関数である．また C はスラック変数とマージンのトレードオフをコントロールするパラメータである．また $\boldsymbol{Q}_{ij} = y_i y_j \boldsymbol{K}_{ij}$ を満たす．決定関数は次式で与えられる．

$$sgn(\sum_{i=1}^{N} y_i\alpha_i\boldsymbol{K}(\boldsymbol{Y}_i, \boldsymbol{Y}_T) + b) \qquad (16)$$

ここで $\boldsymbol{Y}_T$ は $span(Y_T)$ の行列表現であり，テスト画像に対応する．本論文では多クラス問題に対応するため”one–against–one”アプローチを使用する．c をクラス数とすると，$c(c-1)/2$ 個の 2 クラス識別器を構成する．そして識別においては多数決により決定する．

2. 4　Grassmann Discriminant Analysis (GDA)

GDA は [12] において提案されており，式 (12)，(13) で Grassmann kernel k_p と k_{BC} を用いる．つまり GDA は kernel discriminant analysis に Grassmann kernel を用いたものである．GDA は Linear Discriminant Analysis において判別方向 $\boldsymbol{w}$ を求めるために用いるレイリー商 $L(\boldsymbol{w}) = \boldsymbol{w}^T\boldsymbol{S}_b\boldsymbol{w}/\boldsymbol{w}^T\boldsymbol{S}_w\boldsymbol{w}$ にカーネルトリックを適用する．ここで $\boldsymbol{S}_b$，$\boldsymbol{S}_w$ はクラス間分散行列およびクラス内分散行列である．ここで $\Phi = [\phi_i ... \phi_N]$ が学習サンプルの特徴行列のとき (各学習サンプルは部分空間である)，$\boldsymbol{w}$ は特徴ベクトルの線形結合 $\boldsymbol{w} = \boldsymbol{\Phi}\boldsymbol{\alpha}$ で表せる．またレイリー商は $\boldsymbol{\alpha}$ を用いて次のように表せる．

$$L(\boldsymbol{\alpha}) = \frac{\boldsymbol{\alpha}^T\boldsymbol{\Phi}^T\boldsymbol{S}_B\boldsymbol{\Phi}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^T\boldsymbol{\Phi}^T\boldsymbol{S}_W\boldsymbol{\Phi}\boldsymbol{\alpha}}$$

$$= \frac{\boldsymbol{\alpha}^T \boldsymbol{K}(\boldsymbol{V} - \mathbf{1}\mathbf{1}^T/N)\boldsymbol{K}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^T(\boldsymbol{K}(\boldsymbol{I} - \boldsymbol{V})\boldsymbol{K} + \sigma^2\boldsymbol{I})\boldsymbol{\alpha}} \quad (17)$$

ここで $\boldsymbol{K}$ は学習サンプルにグラスマンカーネルを適用して得られるカーネル行列であり，$\mathbf{1}$ は全ての要素が 1 の N 次元ベクトルであり，ブロック対角行列 $\boldsymbol{V}$ は c 番目のブロック (c 番目のクラスに対応) が $N_c \times N_c$ の全ての要素が 1 の行列を N_c で割ったものであり，$\sigma^2\boldsymbol{I}$ は regularizer である．GDA はまず式 (14) を最大化する $\boldsymbol{\alpha}$ を探す．次に $F_{train} = \boldsymbol{\alpha}^T\boldsymbol{K}$ と $F_{test} = \boldsymbol{\alpha}^T\boldsymbol{K}_{test}$ 間のユークリッド距離を用いて最近傍法による識別を行う．ここで $\boldsymbol{K}_{test}$ は学習サンプルとテストサンプルから得られるカーネル行列である．

3. 実　　験

本論文では提案手法の性能を調べるためにいくつかの実験で BoK との比較を行った．実験のため 1 0 物体のデータベースを作成した (図 1)．データベースはスケール変化，視点変化，照明変化があり，背景は一様ではない．このデータベースは学習画像セットおよびテスト画像セットを含んでいる．学習画像セット (図 1 (a)) は物体ごとに 3 枚の画像があり，背景は黒である．テスト画像セット (図 1 (b)) は物体ごとに 1 0 枚の画像があり，背景は一様ではない．

我々は Matlab で BoK，GD–MSM，GK–SVM および GDA を実装した．BoK および GD–MSM に入力する特徴量として、SIFT (t=128) および PCASIFT (t=10, 20, 36, 60, 100, 128) を用いた．SIFT については Lowe が実装したもの [17] を用い、PCASIFT は文献 [15] が提供しているものを用いた ([17] で公開されている)．BoK を実装した際，物体ごとに k–means クラスタリングを行うことでクラスタの重心を得る．ヒストグラムの比較の際はテスト画像と学習画像のヒストグラムのユークリッド距離を比較し，クラスの決定には最近傍法を用いる (SVM による比較も行ったがほとんど差がなかったため，最近傍法を採用した)．また GD–MSM，GK–SVM，GDA において，2 つの識別辞書の作り方を用意した．一つはクラスごとに一つずつ部分空間を作成する方法で，もう一つは画像ごとに部分空間を作成する方法である．

GD–MSM における 8 つのグラスマン距離と認識率を図 2，3 に示す．横軸は各グラスマン距離を表し，縦軸は認識率を示す．また横軸はそれぞれ，projection metric，Binet–Cauchy metric，maxcor (最大相関)，mincor (最小相関)，フロベニウス Procrustes distance (chordalF)，2 ノルム Procrustes distance (chordal2)，geodesic distance，平均距離 (mean) を表す．図より全ての入力特徴量において平均距離が安定して高い認識率を得ている．次いで geodesic distance，projection metric，Binet–Cauchy metric が高い認識率を得ているが，入力特徴量によって認識率に大幅な差があり，あまり安定していないことがわかる．従って以降の実験ではグラスマン距離

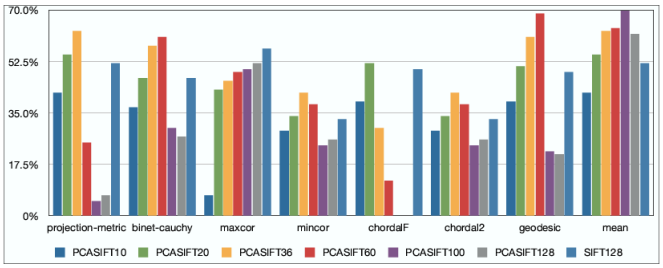


図 2　各グラスマン距離における認識率 (クラスごとに部分空間を作成)

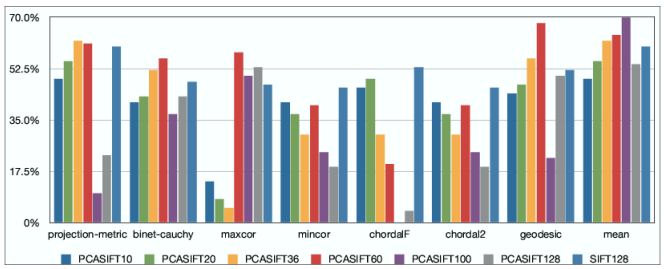


図 3　各グラスマン距離における認識率 (画像ごとに部分空間を作成)

として平均距離を採用する．また識別辞書による認識率の差はほとんど見られない．これはクラスごとに部分空間を作成する方法において，学習に用いた画像数が少なく，画像ごとに部分空間を作成する手法と比べてあまり差がなかったのではないかと考えられる．

BoK，GD–MSM，GK–SVM および GDA による認識実験の結果を図 4，5 に示す．図より各部分空間法における認識率の差はあまり見られず，特に PCASIFT60，PCASIFT100 で高い認識率を示している．また BoK では入力特徴量に依存していないが GD–MSM，GK–SVM および GDA の PCASIFT60，PCASIFT100 と同程度の認識率を得ている．ここでも識別辞書による認識率の差はほとんど見られなかった．

次に部分空間法の出力次元の影響を図 6 に示す．ここでは各部分空間同士であまり差がなかったので代表的なものとして入力特徴量が PCASIFT60 及び SIFT128 に GK–SVM を適用した場合を掲載している．図より PCASIFT60 では 10–20 次元で比較的高い認識性能を得ており，SIFT についてはほぼ全ての次元で安定した認識率を得ている．

最後に提案手法と BoK でそれぞれ最も高い認識率を示した時の計算時間について表 1 にまとめる．なお BoK のクラスタ数は 4000 である．表より GK–SVM は計算時間において BoK より大幅に高速であることが確認できる．

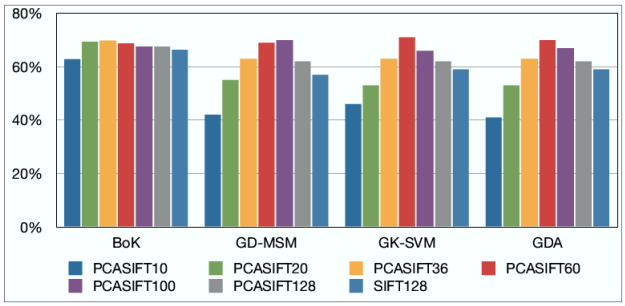


図 4　各手法における認識率 (クラスごとに部分空間を作成)

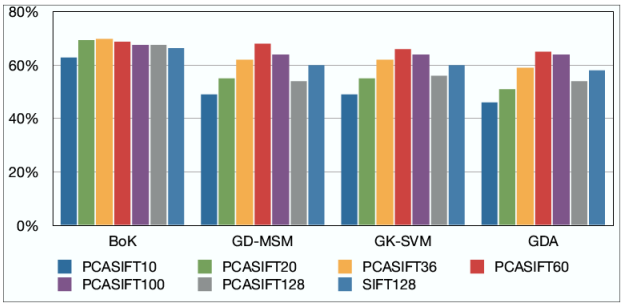


図 5　各手法における認識率 (画像ごとに部分空間を作成)

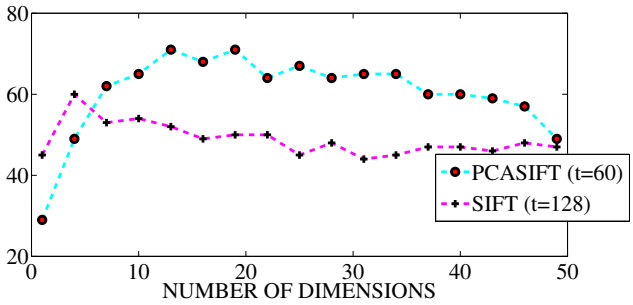


図 6　出力次元と認識率の関係

表 1　SoF および BoK の計算時間

手法	認識率	学習時間	テスト時間
BoK	69.8 [%]	7.4525[sec]	34.0828 [sec]
SoF (GK–SVM)	71 [%]	0.3067 [sec]	0.4779 [sec]

4. おわりに

本論文では BoK に代わる物体認識手法として，局所特徴量のモデリングを部分空間で行う Subspace of Features (SoF) を提案した．SoF の認識性能向上のために部分空間の類似度としてグラスマン距離，グラスマンカーネルを採用し，GD–MSM，GK–SVM を提案した．物体認識実験においては BoK と同程度の認識性能を得たが，BoK よりも計算コストの面で優れていることを示した．

本論文では SoF の有効性を示すため基本的な局所特徴量を用いた実験を行った．入力する特徴によって識別性能に違いがあることがわかったので，今後はより大きい規模のデータベースで様々な特徴量を入力とした時の認識性能の違いを調べることを予定している．

文　献

[1] Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 12(60), (2004) 91-110

[2] Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: SURF: Speeded Up Robust Features. Computer Vision and Image Understanding (CVIU) 110-3, (2008) 346-359

[3] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. Proc. IEEE Conference on Computer Vision and Pattern Recognition. (2005) 20-25

[4] Tuytelaars, T., Mikolajczyk, K.: Local Invariant Feature Detectors: A Survey. Foundations and Trends in Computer Graphics and Vision. 3-3 (2007) 177-280

[5] Murase, H., Nayar, S.: Visual Learning and Recognition of 3-D Objects from Appearance. International Journal of Computer Vision 14(1), (1995) 5-24

[6] Csurka, G., Bray, C., Dance, C., Fan, L.: Visual Categorization with bags of keypoints. Proc. ECCV Workshop on Statistical Learning in Computer Vision (2004) 1-22

[7] Sivic, J., Zisserman, A.: Video Google: A text retrieval approach to object matching in videos. Proceedings of the International Conference on Computer Vision. (2003) 1470-1477

[8] A. Bjorck and G.H. Golub, “Numerical Methods for Computing Angles between Linear Subspaces,” Math. Computation, vol. 27, no. 123, pp. 579-594, 1973.

[9] G. H. Golub and C. F. van Loan, Matrix Computations, Johns Hopkins University Press, 3rd edition.

[10] O. Yamaguchi, K. Fukui, and K. Maeda, “Face Recognition Using Temporal Image Sequence,” in Proc. Int. Conf. Automatic Face and Gesture Recognition, pp. 318-323, 1998.

[11] Y.C.Wong, ”Differential geometry of Grassmann manifolds,” in Proc. of the Nat. Acad. of Sci., Vol. 57, pp. 589–594, 1967.

[12] J. Hamm and D. D. Lee, ”Grassmann discriminant analysis: A unifying view on subspace–based learning,”in Proc. 25th Int. Conf. on Machine Learning, pp. 376–383, 2008

[13] A. Edelman, T.A. Arias and S.T. Smith, ”The geometry of algorithms with orthogonality constraints,” SIAM J. Matrix Anal. Appl., 20 (2), pp. 303–353, 1998.

[14] C. M. Bishop, Pattern recognition and Machine Learning, Springer-Verlag, 2006.

[15] Ke, Y., Sukthankar, R., ”PCA–SIFT: A More Distinctive Representation for Local Image Descriptors,” Proceedings IEEE Conference on Computer Vision and Pattern Recognition. 2 (2004) 506-513

[16] C. Cortes and V. Vapnik, “Support Vector Networks,” Machine Learning, 20, pp. 273-297, 1995.

[17] www.cs.cmu.edu/˜yke/pcasift/

[18] J. S. Taylor and N. Cristianini, ”Kernel Methods for Pattern Analysis”, Cambridge University Press, 2004.