Vol.2012-CVIM-181 No.15
2012/3/16

特徴点選択とペア化による Naive-Bayes Nearest-Neighbor 手法の改良

秋山瑞樹[†1]　柳井啓司[†1]

本研究の目的は，特徴点マッチングベース手法である Naive-Bayes Nearest-Neighbor (NBNN) を改良することで画像認識の精度向上を目指すことである．

NBNN 手法では, 各クラスごとに局所特徴をデータベース化し, 未知画像の局所特徴をデータベース化した局所特徴に対して近似最近傍探索を行うことで, 未知画像とクラス間の距離を計算し, 画像対クラスの距離により分類を行う手法である．

本研究では, NBNN 手法に対して, (1)SaliencyMap によるデータベース特徴の重み付け, (2)Multiple Instance Learning によるデータベース特徴の重み付け, (3) 局所特徴のペア化という 3 点の改良を行った．データベースの重み付けでは, 画像の顕著性を考慮した SaliencyMap の利用とデータベース特徴の選別方法として使われる Multiple Instance Learning の導入を行った．データベース特徴への重み付けを行うことで, クラスを表すのに重要なデータベースの局所特徴が, 最近傍探索として選ばれた場合, 画像クラス間の距離が短くすることができる．局所特徴のペア化では, 隣接する局所特徴のペアに対して, 局所特徴を統合するすることで, より情報量の多い特徴を作成した．

Caltech101 のデータセットからランダムに選んだ 10 クラスでの認識実験を行った結果, ベースラインとなる NBNN 手法の分類率 0.633 を超える 0.673 の精度を得ることができた．

1. はじめに

1.1 背　景

実世界シーンの一般的な画像に含まれる物体に対して, 椅子, 自動車など一般的な物体カテゴリを言い当てる処理を計算機にさせる「一般物体認識」という研究がある．

こうした一般物体認識を行う手法として大きく分けて, 学習ベース分類とノンパラメトリックベース分類の 2 つに分けることができる．

学習ベース分類手法では, Support Vector Machine や Boosting といった機械学習を組み合わせた手法が代表される．認識に最適なパラメータを機械学習により推定することで, 高精度での認識が可能になっている．ノンパラメトリックベース分類手法では, Nearest-Neighbor 法が代表される．データベース画像と, 認識対象の未知画像との近傍画像を検索することで, 未知画像の認識を行う単純な手法であり, 学習手法よりも精度が劣るが, 学習にかかる時間が学習ベース手法に比べると, 短い時間ですむ．

一般物体認識手法では, 学習ベース手法が主流ではあるが, Naive Bayes Niearest Neighbor(NBNN) 手法の提案により, ノンパラメトリック手法でも, 学習ベース手法に匹敵する精度があげられることがわかった．本研究では, この NBNN 手法の改良を目指す．

1.2 目　的

NBNN 手法では, 各認識クラスごとに画像特徴をデータベース化し, 未知画像の特徴とデータベース化したクラスごとに近傍特徴との距離を計算することで, 未知画像, クラス間の距離を計算し, 各画像対クラスの距離により分類を行う手法である．

本研究では, NBNN 手法に対して, (1)SaliencyMap によるデータベース特徴の重み付け, (2)Multiple Instance Learning によるデータベース特徴の重み付け, (3) 局所特徴のペア化という 3 点の改良を行った．データベースの重み付けでは, 画像の顕著性を考慮した SaliencyMap の利用とデータベース特徴の選別方法として使われる Multiple Instance Learning の導入を行った．データベース特徴への重み付けを行うことで, クラスを表すのに重要なデータベースの局所特徴が, 最近傍探索として選ばれた場合, 画像クラス間の距離が短くすることができる．局所特徴のペア化では, 隣接する局所特徴のペアに対して, 局所特徴を統合するすることで, より情報量の多い特徴を作成した．

比較として, Caltech101 のデータセットからランダムに 10 クラスを選択し, ベースラインとなる NBNN 手法との比較実験を行った．

2. 関連研究

Naive-Bayes Nearest Neighhbor(NBNN)[1] は画像認識において, Niearest Neighbor 法を利用したノンパラメトリック手法にも関わらず, SVM や Boosting などの機械学習ベース手法の認識に匹敵する精度での認識が可能である．NBNN の特徴として 2 つの重要なポイントがある．

1 つ目の重要な点として特徴量を量子化せずにそのまま使うことである．量子化することでの精度低下は大きく, 学習ベース手法においては量子化をしたとしても, 学習ステップに

†1 電気通信大学院 情報理工学研究科 総合情報学専攻

よって情報欠落を補完できるが, NBNN に代表される Niearest Neighbor 手法においては量子化によって重要な特徴が消えてしまう危険性がある. 量子化をせずにデータベース化することで単純な手法でも精度を保つことができる.

2 つ目の重要な点として, 一般的な Niearest Neigbhor 手法で用いられる'Image to Image' 距離 (I2I) ではなく'Image to Class' 距離を使うことである.I2I ではクエリ画像とデータベース中の近傍画像がどのクラスに属するかで認識を行うが, I2C ではクエリ画像と各クラスまでの距離から認識を行う. クエリ画像のクラスを認識する式を表すと以下のようになる.

$$C_{test} = \arg\min_{c} \sum_{i=1}^{n} ||d_i - NN_c(d_i)||^2 \quad (1)$$

C_{test} はテスト画像のクラス, c はクラス, d_i は局所特徴, $NN_c(a)$ は c に含まれる a との最近傍特徴を表す.

NBNN の長所として, 特徴ベース手法の認識であるので, 同一物体や人工物などに対して高い精度を出せることや, 学習ベース手法に比べると, 学習に時間がかからないという点がある. しかし 短所としてデータベースの特徴点への依存が大きい点が問題となる. その問題を緩和するために色々な改良手法が提案されている.

Behmo らは[2] ,[1] のカーネル密度がクラス毎異なる点を指摘し, 各クラス適切な正規化を行った. 'Image to class' 距離にクラス毎に異なる 2 つのパラメータを加えることで, クロスバリデーションを行い, 最適なパラメータを推定し, 新たな I2C 距離を計算することで, 特徴の独立性仮定を緩和し, データベースへの依存を軽減した. またウィンドウ[3] ごとに評価を行うことで, 物体位置検出にも効果を示した. 2 つのパラメータを含めた I2C 距離は以下の式で表すことができる.α, β 2 つのパラメータを各クラスごとに最適化問題を解くことで推定し, 各クラス毎の異なる分布を正規化する距離計算を行うことが可能になり精度が向上した.

Wang らは[4] ,[5], 少ない特徴点でも高い精度が出るように, マハラノビス距離と空間ピラミッド[6] を導入することで, クラスごとに最適な距離空間で計算を行うことで NBNN を改良した. マハラノビス距離による I2C 距離は以下の式になる. 計量行列 M をクラスごとに推定することで, クラスに最適な距離空間での計算が可能になる.

また[5] では, データベースに対して, マージン最大化による問題を解くことで, データベース自体に重みをつけることを行なっている. ただし[5] での目的はマルチラベル認識となっており, 他研究と同様の分類率での評価ではないが, マルチラベルに対応した評価基準で,[1] を上回る精度が得られている.

Tuytelaars らは[7] , NBNN をカーネル化することで, 一般的な Support Vecroter Machine に組み込むことを可能とし, 学習ベースでの主要な画像全体特徴 Bag-of-Words[8] の認識手法と, NBNN における特徴ベースの認識手法を Multiple Kernel Learning (KML)[9] によって組み合わせを行った.

本研究では, 特徴自体に重み付けを行うことに関しては, Wang らの研究に近いが, 画像の顕著性に基づく Saliency Map と Multiple Instance Learning を利用し, 異なるアプローチでの重み付けを行った. また特徴自体をペアにすることで特徴量の情報量を上げる試みもし, これらのアプローチを組み合わせることで実験を行った. 各手法の簡単な比較は以下の表 1 になる.

表 1 各手法との比較

	パラメータ学習	距離空間の変化	重み付け	カーネル化	ペア化
Boiman et al.[1]	×	×	×	×	×
Behmo et al.[2]	○	×	×	×	×
Wang et al.[4]	○	○	×	×	×
Wang et al.[5]	○	×	○	×	×
Tuytelaars et al.[7]	×	×	×	○	×
提案手法	○	×	○	×	○

3. 提案手法概要

本研究での大まかな流れとしては以下のようになる. 本研究で新しく提案する点は, 特徴のペア化とデータベース特徴の重み付けになる.

手法概要

1 画像特徴量の抽出
2 特徴のペア化 (提案手法)
3 データベース化
4 データベース特徴の重み付け (提案手法)
5 クラス認識

Vol.2012-CVIM-181 No.15
2012/3/16

4. 手法詳細

4.1 局所特徴

SIFT(Scale Invariant Feature Transform)[10] は D.Lowe によって考案され，特徴点周りの局所画像パターンを 128 次元特徴ベクトルで表現し，回転・スケール変化・照明変化に対して耐性のある特徴である．本研究での, 特徴点検出には 8 ピクセルごとに 5 スケール毎に特徴を抽出し使用している．

4.2 データベース

データベースとして登録するクラスごとに, そのクラスに含まれる画像から得られた SIFT 特徴をまとめることでデータベースを作成する．クラスの数だけのデータベースができることになる．ただし, クラス間で登録した SIFT 特徴にばらつきがあると, 登録数が多いクラスに影響を受けてしまうので, ランダムサンプリングによって全てのクラスの特徴登録数を均一にした．

4.3 特徴のペア化

局所特徴として用いた SIFT 特徴のペア化を行う．

各画像から得られた SIFT 特徴に対して, 隣接する特徴同士を結合する．f を SIFT 特徴量とするとき, 隣接する特徴 a(x_a, y_a, f_a), 特徴 b(x_b, y_b, f_b) から得られるペア特徴 p は $((x_a + x_b)/2, (y_a + y_b)/2, f_a, f_b)$ となる．

本研究では, SIFT 特徴を, スケールを 5 種類に固定し, グリッドサンプリングによって抽出してるので, 隣接特徴を選ぶ場合, 同じスケール内で選ぶようにした (図 **??**).

4.4 データベースの重み付け

データベースの重み付けでは, 画像の顕著性を考慮した SaliencyMap の利用とデータベース特徴の選別方法として使われる Multiple Instance Learning の導入により, データベース特徴への重み付けを行った．

SaliencyMap では, 人間の視覚系に基づいて画像中の顕著な点を計算することができる．例えば入力画像として図 1 を与えると, 出力として図 2 が得られる．画像中の各ピクセルについて $[0, 1]$ の出力値が得られ, 0 に近いほど顕著度が低く, 1 に近いほど顕著度が高い結果となり, 図 2 では白い領域が顕著度の高い領域と検出された結果になる．SaliencyMap[11] を利用することで, クラス認識でより重要だと思われる物体領域からの特徴と背景特徴を区別することが期待できる．I2C 距離を計算する場合に, 近傍特徴として選ばれた特徴が, 物体領域から得られたデータベース特徴の場合は距離を短くし, 背景領域から得られたデータベース特徴の場合では距離が長くなるような重み付けが期待できる本研究では, 特徴点における SaliencyMap の出力値を重みとして利用した．

図 1 入力画像

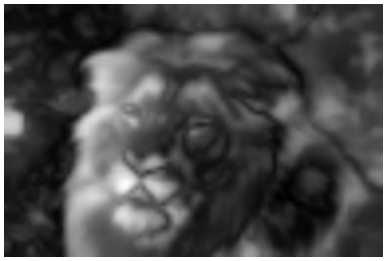

図 2 SaliencyMap

MIL(Multiple Instance Learning)[12] を用いることで, あるクラスのデータベースだけに現れる特徴だけを選別することができる．例えば飛行機クラスをポジティブクラスとし, ネガティブクラスをそれ以外のクラスとすると, ポジティブクラスだけに現れる独特な特徴である, 飛行機の羽やボディー, 雲などから得られた特徴と, ネガティブクラスでも現れる可能性が高い, 芝生や道路といった特徴を分けることが可能である．本研究では, MIL 手法として mi-SVM[13] を利用した. あるクラスのデータベース特徴をポジティブクラス, そのクラス以外から得たデータベース特徴をネガティブクラスとして MIL を計算することで, ポジティブクラスのデータベース特徴に正負の評価値を出力として得ることができる. その評価値を用いることで, I2C 距離を計算する場合に, 近傍特徴として選ばれた特徴が評価値が高い特徴の場合は距離を短くし, 低い特徴では距離が長くなるような重み付けが期待できる．本研究では, mi-SVM による出力値を特徴の重みとして利用した．

4.5 クラス認識

テスト画像からも同様に SIFT 特徴を抽出する．得られた SIFT 特徴すべてに対して, 各データベース毎に最近傍特徴を探索し, 最近傍特徴とのユークリッド距離をデータベースと対応したクラスの距離として足していくことで, クラスに対するテスト画像の'Image to Class' の距離 (I2C 距離) が得られる．すべてのクラスに対して I2C 距離を計算し, 距離が最小になるようなクラスを, テスト画像のクラスとして認識する．

式で表すと以下のようになる.

$$C_{test} = \arg\min_{c} \sum_{i=1}^{n} w(NN_c(d_i)) * ||d_i - NN_c(d_i)||^2 \quad (2)$$

C_{test} はテスト画像のクラス, c はクラス, d_i は局所特徴, $NN_c(a)$ は c に含まれる a との最近傍特徴, w は特徴の重みを表す. 本研究では, データベースの特徴に重み付けを行ったので, 重みを考慮した I2C の距離の式になっている.

5. 実　　験

5.1 データセット

データセットとして Caltech101[14)15)] を用いて提案手法の評価を行った. Caltech101 データセットは, 101 種類のカテゴリ, 各カテゴリ 31 枚から 800 枚で構成されている. 本研究では caltech101 データセットからランダムに 10 カテゴリを選択し, 10 クラス分類データセットとして実験を行った. 使用した 10 クラスは表 2 になる. 各カテゴリにはばらつきがあるため, 実験ではランダムに 15 枚を学習画像として利用し, 学習に使わない画像からランダムに 15 枚を実験画像として利用した.

表 2 10 クラスデータセット

barrel	bass	car_side	cougar_body	garfield
headphone	joshua_tree	laptop	saxphone	watch

5.2 実 験 内 容

10 クラス分類データセットを利用して, 各最適パラメータの推定を行う. 行った実験は以下になる.

(1) 特徴点座標特徴のパラメータ推定
(2) SaliecnyMap による重み付けのパラメータ推定
(3) mi-SVM による重み付けパラメータ推定
(4) 各パラメータを組み合わせた結果
(5) ペア化による結果

実験で利用する SIFT 特徴は 8 ピクセル毎, 5 スケールで抽出した特徴を利用した. また各クラスにおけるデータベースに登録する特徴数がクラスごとに異なると, 不利なクラスが出てきてしまうので, 特徴点をランダムにサンプリングすることで, 特徴点数を均一にした.10 クラスで利用した各クラスあたりの特徴点はペア化なしの場合 30,000 個, ペア化ありの場合 150,000 である. 比較のベースラインとして利用した手法は, NBNN[1)] 元論文を比較対象とした. またベースライン手法の実験も独自に行なった.

5.3 評 価 方 法

分類結果の評価に用いる基準として, 分類率を用いた. 分類率の式は以下で定義する.

$$\text{分類率} = \frac{\text{正しく分類されたテスト画像数}}{\text{テスト画像総数}} \quad (3)$$

実験では学習画像と実験画像をランダムで選出した画像で固定し, 1 回の分類精度を確認した.

5.4 結 果 結 果

各実験について細かく紹介する.

5.4.1 特徴点座標特徴のパラメータ推定

NBNN[1)] では, I2C 距離を計算する際に, SIFT 特徴のユークリッド距離だけでなく, 実験画像の SIFT 特徴とデータベースに登録されている最近傍特徴の座標の差を距離に加えている. これを式で表すと以下のようになる.

$$C_{test} = \arg\min_{c} \sum_{i=1}^{n} (||d_i - NN_c(d_i)||^2 + \alpha \cdot ||l_{d_i} - l_{NN_c(d_i)}||^2) \quad (4)$$

l_x は特徴点 x の座標情報を表す.[1)] では, このパラメータ α の値が明記されていないので, まずこのパラメータ α の推定を行った.

$0.12 \leq \alpha \leq 0.25$ の時, 良い結果となり, $\alpha = 0.14$ の時, 最高値 0.633 が得られた. よって以降の実験で座標情報を使う場合, 座標パラメータ α はこの範囲で実験を行う. また座標特徴を使わない場合の値は 0.587 となった. 以下の実験ではベースラインとして、α を利用するとき 0.633, 利用しないとき 0.587 として比較を行った.

5.4.2 SaliecnyMap による重み付けのパラメータ推定

データベース画像から特徴を抽出する際に, SaliencyMap を利用することで, その特徴点における顕著度を計算する. 顕著度から得られた特徴の重み $0 \leq w_{sm} \leq 1$ とすると, I2C 距離は以下の式で書き換えられる.

Vol.2012-CVIM-181 No.15
2012/3/16

$$C_{test} = \arg\min_{c} \sum_{i=1}^{n} exp(-w_{sm}/\beta) * ||d_i - NN_c(d_i)||^2 \quad (5)$$

w_{sm} は顕著度が高いほど 1 に近い値を示し, より重要性の高くなるが, I2C 距離においては距離が短くなるほど良い評価値となるので, w_{sm} の値を SaliecnyMap パラメータ β によって値が反転するようにした.

β の値の変化による精度は図 3 となった. ベースライン, 提案手法とも座標特徴を利用していない. 赤がベースライン, 青が SaliencyMap のみ使った場合の結果である.

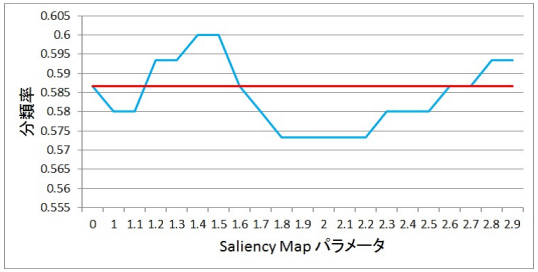

図 3 Saliency Map パラメータ推定.β の値を変化させて実験を行った.

ベースラインの値は 0.587 である. $0.12 \leq \beta \leq 0.16$ の時, 良い結果となり, $\beta = 1.4$ のとき最高値 0.6000 が得られた.

5.4.3 mi-SVM による重み付けパラメータ推定

mi-SVM における, SVM の精度を決める要素として, 学習データ数とコストパラメータが挙げられる. ここでは学習データ数 n とコストパラメータ C を変化させた時の, 重みの変化による精度の確認を行った. コストパラメータを大きくすればするほど, ハードマージンな SVM となる. 各データベース特徴には SVM の出力値が重み w_{SVM} として付けられるので, SaliecnyMap による重み付けと同じ式による認識となる. また SVM のカーネルは RBF カーネルを用いた.

パラメータを変化させた結果, mi-SVM で得られた特徴ごとの重みを w とすると, I2C 距離は以下の式で書き換えられる.

$$C_{test} = \arg\min_{c} \sum_{i=1}^{n} exp(-w_{SVM}/\gamma) * ||d_i - NN_c(d_i)||^2 \quad (6)$$

w_{SVM} は評価値が高いほど高い値を示し, より重要性の高くなるが, I2C 距離においては距離が短くなるほど良い評価値となるので, w_{SVM} の値を mi-SVM パラメータ γ によって値が反転するようにした.

n の変化による実験結果は図 4 となった. ベースライン, 重み付けどちらの実験でも座標情報を使わない. パラメータ γ を大きくすると, 重みは 1 に収束するので, パラメータが大きいほど重み情報を使わない結果に近づいていく.

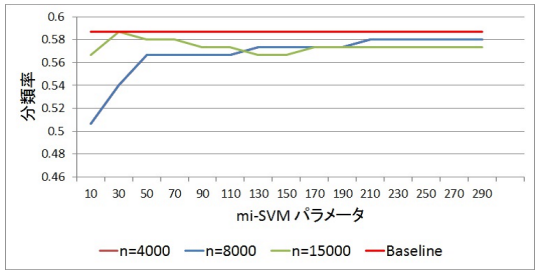

図 4 mi-SVM パラメータ推定における学習特徴数の変化.$n = 4000, 8000, 15000$ と γ の値を変化させて実験を行った.

$n = 4000$ の結果は, $n = 8000$ の時と同じ結果となった. どれもベースラインを下回る結果となってしまったが, 重みが重要になってくる mi-SVM パラメータが小さい場合では, 学習特徴数を増やしたほうが, 特徴数が少ない時よりも有効だと思われる. C の変化による実験結果は図 5 となった.

こちらの実験の場合も, ベースラインを下回る結果となってしまった. mi-SVM が小さい場合の比較結果を見てみると, コストパラメータを変化させた場合, $c = 100$ の結果が若干良く見えるが, mi-SVM の際の学習時間は c が大きいほど長くなってしまうので, 以下の組み合わせによる実験では $c = 1$ で実験を行った.

5.4.4 各パラメータを組み合わせ結果

以上の結果から, 座標パラメータ (α) を $0.12 \leq \alpha \leq 0.25$, SaliencyMap パラメータ

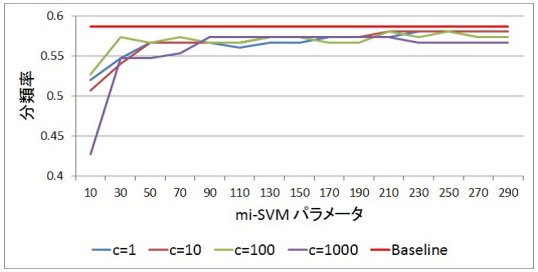

図 5 mi-SVM パラメータ推定におけるコストパラメータの変化.$c = 1, 10, 100, 1000$ と γ の値を変化させて実験を行った.

(β) を $1.2 \leq \beta \leq 1.6$, mi-SVM パラメータ (γ) を $10 \leq \gamma \leq 100$ とし各パラメータを組み合わせて実験を行った. 座標パラメータと SaliencyMap パラメータ, 座標パラメータと mi-SVM パラメータを組み合わせた結果は図 6 となった. ベースラインは実験 (a) での最高値 0.633($\alpha = 0.14$) の場合である. 座標と SaliencyMap を組み合わせた場合は $\alpha = 0.18, \beta = 1.3$ の時, 最高値 0.653 が得られ, 座標と mi-SVM を組み合わせた場合は $\alpha = 0.17, \gamma = 90$ の時, 最高値 0.633 が得られた. SaliencyMap では精度が向上したが, mi-SVM では精度が変わらなかった.

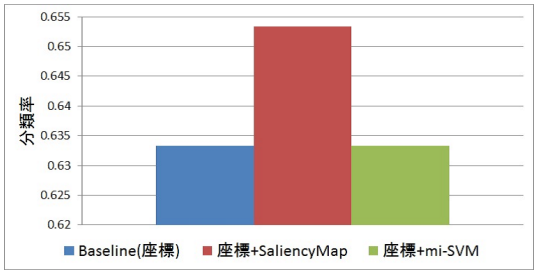

図 6 座標, 座標+SaliencyMap, 座標+mi-SVM の比較結果.

全てのパラメータを組み合わせた結果は図 7 となった. また全ての手法で座標情報を利用し, $0.12 \leq \alpha \leq 0.25$ の範囲で変化させた時の最高値の結果である. Baseline, SaliencyMap, mi-SVM は前の実験と同じであり, $n = 4000, 8000, 15000$ の場合の, それぞれパラメータ α, β, γ を変化させた結果である.

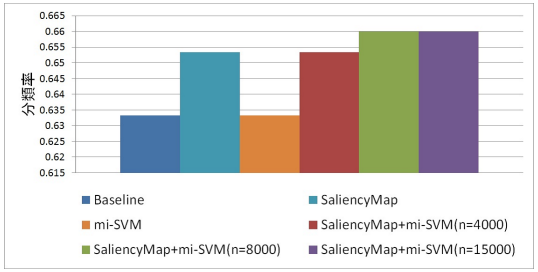

図 7 全てのパラメータを組み合わせた結果. 赤:($n = 4000, \alpha = 0.14, \beta = 1.2, \gamma = 90$), 緑:($n = 8000, \alpha = 0.16, \beta = 1.3, \gamma = 90$), 紫:($n = 15000, \alpha = 0.18, \beta = 1.2, \gamma = 70$)

最高精度は $n = 8000, 15000$ のとき最高値 0.660 が得られた. mi-SVM 単体では精度が下がっていたが, SaliencyMap と組み合わせることで, SaliencyMap だけを利用するよりもわずかに精度が向上した. ただし, SaliencyMap を加えた時の精度と比べると, あまり精度の変化がなかった. SaliencyMap に比べると, mi-SVM はあまりうまくいっていないことがわかった.

5.4.5 ペア化を組み合わせ結果

ペア化をしない実験と同様に, 座標パラメータの推定を行った. ただしペア化を行なっているので, ペアにした特徴同士についての座標情報, SaliencyMap の重みが得られることになる. f を SIFT 特徴量とするとき, 隣接する特徴 a(x_a, y_a, w_{sma}, f_a), 特徴 b(x_b, y_b, w_{smb}, f_b) から得られるペア特徴 p は $((x_a + x_b)/2, (y_a + y_b)/2, (w_{sma} + w_{smb})/2, f_a, f_b)$ となる. 本実験では, ペア化に用いた局所特徴点同士の座標の平均を, ペア特徴の座標情報として, SaliencyMap の重みの平均をペア特徴の SaliencyMap 情報として利用した. ペア化を行わない場合と同じように, α の変化による実験は図 8 となった.. 赤がベースラインとして座標情報を用いた NBNN, 青がペア化+座標情報の結果である.

ベースラインは実験 (a) での最高値 0.633($\alpha = 0.14$) の場合である. $0.34 \leq \alpha \leq 0.38$ の時, 良い結果となり, $\alpha = 0.38$ の時, 最高値 0.653 が得られた. ペア化をしない場合に比べて, α の値が高い時に良い結果が出たのは, SIFT 情報が 128 次元から 258 次元になるため, 全体としての I2C 距離が伸びたためであると考えられる. 結果としては, 隣接特徴同士を結

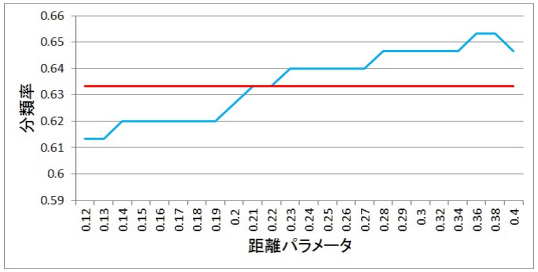

図 8 ペア化における座標パラメータ推定

合する, 単純なペア化だけでも精度が上がったことから, 登録するデータベースの数が重要だと考えられる.

実験 d と同様に, ペア化に対して, SaliencyMap パラメータ, mi-SVM パラメータを組み合わせた実験結果は図 9 となった.. 全ての手法で座標情報を利用し, $0.24 \leq \alpha \leq 0.40$ の範囲で変化させた時の最高値の結果である. また $n = 10000, c = 1$ として実験を行った. ベースラインは実験 (a) での最高値 0.633($\alpha = 0.14$) の場合である. ペア化は上の実験で得られた, $\alpha = 0.38$ の時, 最高値 0.653, ペア化と SaliencyMap を組み合わせた場合は $\alpha = 0.4, \beta = 1.4$ の時, 最高値 0.673, ペア化と mi-SVM を組み合わせた場合は $\alpha = 0.36, \gamma = 90$ の時, 最高値 0.653, 全てを組み合わせた場合は $\alpha = 0.40, \beta = 1.4, \gamma = 100$ の時, 最高値 0.647 が得られた.

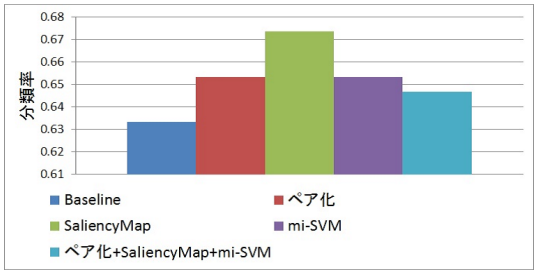

図 9 ベースライン, ペア化, ペア化＋ SaliencyMap, ペア化＋ mi-SVM, 全ての組み合わせの比較結果. $A : (n = 10000, \alpha = 0.40, \beta = 1.4, \gamma = 100)$

結果としては, ペア化を用いない場合と同じように SaliencyMap では精度が向上したが, mi-SVM では精度が下がってしまい, mi-SVM を組み合わせないほうが良い結果となってしまった.

6. 考　　察

6.1 SaliencyMap による重み付けに関する考察

SaliencyMap を用いた実験では, 適切にパラメータを設定することで, ベースラインよりも高い精度での認識が可能になった. この結果から, Caltech101 のような 1 枚の画像に対して, 1 つの物体が写っているようなデータセットに対しては, 背景領域の特徴よりも, 物体領域の特徴のほうが重要であると考えられる. 本研究で利用した SaliencyMap 手法では, ピクセル毎にガウシアンピラミッド画像の差分を利用して, 顕著度を検出しているので物体内の領域でも, 顕著度が画像によっては低くなってしまう場合があったり, そもそも顕著度がうまく取れていない場合があった. このような問題を解決するために, 例えば, 画像のエッジなどで物体の領域境界を考慮したり, 色に特化した別なアプローチの顕著度マップ[16] を組み合わせて利用することで, 物体領域の顕著度をより正確にできるのではないかと期待できる.

6.2 mi-SVM による重み付けに関する考察

mi-SVM を用いた実験では, mi-SVM を単体で使った場合, ベースラインよりも低い結果となってしまった. NBNN 手法に mi-SVM のような MIL を適用するとき, 学習に使うデータ数が重要となってくるため, 学習データ数を増やした実験を行う必要がある.

6.3 ペア化に関する考察

ペア化を用いた実験では, 隣接特徴同士をくっつけるという単純な方法にもかかわらず, ベースラインよりも良い結果が得られた.NBNN 手法では, データベースの特徴の質がとても重要なので, ペア化により登録する特徴の多様性に幅が生まれたことから精度が向上したと思われる. しかし特徴を増やすことと, 最近傍特徴の探索時間はトレードオフ関係になっており, ペアを増やせば増やすほど認識時間が増えてしまう. 本研究では, 最近傍探索手法として kd-Tree による木構造探索で行ったが, kd-Tree の改良[17] や, LSH(Locality Sensitive Hashing)[18),19)] のようなハッシュを使った近傍探索手法を利用することで, 計算時間を減らすこと可能ではないかと考えられる.

7. ま と め

本研究では, NBNN の改良として, SaliencyMap と mi-SVM によるデータベース特徴の重み付けと特徴点のペア化を行った. SaliencyMap では画像から顕著度マップを作成し, 物体領域から得られた画像特徴に対して良い重みを付け, mi-SVM ではデータベースごとに, クラス特有な特徴に対して良い重みを付けた. 特徴のペア化では, 隣接する特徴同士を結合することで, 特徴の情報量とデータベースの特徴数を増やすことを行った.

Caltech101 データセットを実験対象として, 10 クラス分類を行うときの認識結果として, NBNN[1] を上回る精度が得られた. しかし, SaliencyMap やペア化による精度向上は確認できたが, mi-SVM はあまり影響を受けてるいないことがわかった.

8. 今後の課題

精度向上の程度が低かった mi-SVM について見直す必要がある. 考察で述べたように, NBNN に対して, mi-SVM を適用する場合, 学習サンプル数と学習時間が問題となってくる. 学習サンプルを減らしてしまうと, 重要な特徴が抜けてしまい, 良い学習ができないと思われるが, 増やしてしまうと学習に非常に時間がかかってしまう. 今後は大規模かつ高速な MIL 手法を試す必要があると考えている. また今回の実験では, Caltech101 データセットに対しての精度評価しか行わなかったので, 他のデータセットでも評価を行い, 提案手法の有効性を確認する必要がある.

参 考 文 献

1) O.Boiman, E.Shechtman, and M.Irani. In defense of nearest-neighbor based image classification. *Proc. of IEEE Computer Vision and Pattern Recognition*, 2008.
2) R.Behmo, P.Marcombes, A.Dalayan, and V.Prinet. Towards optimal naive bayes nearest neighbor. *Proc. of IEEE International Conference on Computer Vision*, 2010.
3) C.H. Lampert, M.B. Blaschko, and T. Hofmann. Beyond sliding windows: Object localization by efficient subwindow search. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pp. 1–8. Ieee, 2008.
4) Z.Wang, Y.Hu, and L.T. Chi. Image-to-class distance metric learning for image classification. *Proc. of European Conference on Computer Vision*, 2010.
5) Z.Wang, Y.Hu, and L.T. Chia. Multi-label learning by image-to-class distance for scene classification and image annotation. *Proc. of ACM International Conference on Image and Video Retrieval*, 2010.
6) S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, Vol. 2, pp. 2169–2178. Ieee, 2006.
7) T.Tuytelaars, M.Frits, K.saenko, and T.Darrell. The nbnn kernel. *Proc. of IEEE International Conference on Computer Vision*, 2011.
8) G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV*, Vol.1, p.22, 2004.
9) G.R.G. Lanckriet, N.Cristianini, P.Bartlett, L.E. Ghaoui, and M.I. Jordan. Learning the kernel matrix with semidefinite programming. *The Journal of Machine Learning Research*, Vol.5, pp. 27–72, 2004.
10) D.G. Lowe. Local feature view clustering for 3d object recognition. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001.*, Vol.1, pp. I–682. IEEE, 2001.
11) L.Itti, C.Koch, and E.Niebur. A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol.20, No.11, pp. 1254–1259, 1998.
12) O.Maron and A.L. Ratan. Multiple-instance learning for natural scene classification. Vol.15, pp. 341–349, 1998.
13) S. Andrews, I. Tsochantaridis, and T. Hofmann. Support vector machines for multiple-instance learning. *In Advances in Neural Information Processing Systems*, 2003.
14) Caltech101. `http://www.vision.caltech.edu/Image_Datasets/Caltech101/`.
15) L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In *Computer Vision and Pattern Recognition Workshop, 2004 Conference on*, pp. 178–178, 2004.
16) M.M. Cheng, G.X. Zhang, N.Mitra, X.Huang, and S.M. Hu. Global contrast based salient region detection. In *cvpr*, pp. 409–416, 2011.
17) C.S. Anan and R.Hartley. Optimised kd-trees for fast image descriptor matching. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2008.
18) P.Indyk and R.Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pp. 604–613. ACM, 1998.
19) K.Min, L.Yang, J.Wright, L.Wu, X.S. Hua, and Y.Ma. Compact projection: Simple and efficient near neighbor search with practical memory requirements. In *Proc. of IEEE Computer Vision and Pattern Recognition*, 2010.