2009－CVIM－166 (13)
2009／3／13

Multiple Kernel Learningを用いた食べ物画像の分類

上東太一[†1] 柳井啓司[†1]

近年，食事に関する健康管理が注目され，より簡単に食事内容が記録できるシステムが望まれている．そこで，本研究では，画像認識技術を用いて食事内容を記録するシステムを提案する．画像認識手法として，最新の機械学習の手法である Multiple Kernel Learning(MKL) を用いて，局所特徴，色特徴，テクスチャ特徴などの複数種類の画像特徴を統合して，高精度な認識を実現することを提案する．MKL を用いることにより，カテゴリ毎に認識に有効な画像特徴を自動的に推定し，各特徴に対して最適な重みを学習することが可能となる．それに加え，本研究では，提案した食事画像認識手法を組み込んだ食べ物画像認識システムのプロトタイプを実装した．

実験では，50 種類の食べ物画像データセットを構築し，提案手法の評価を行ない，平均分類率 61.34%を達成した．50 種類もの大規模な食事画像の分類は，実用的な精度で実現することが困難であったため報告例がないが，本研究では MKL による特徴統合を行なう提案手法によって，初めて大規模食事画像分類において高い認識精度を達成することができた．

Classification of Food Images with Multiple Kernel Learning

Taichi Joutou[†1] and Keiji Yanai[†1]

Since health care on foods is drawing people's attention recently, a system that can record everyday meals easily is being awaited. In this paper, we propose an automatic food image recognition system for recording people's eating habits. In the proposed system, we use the Multiple Kernel Learning (MKL) method to integrate several kinds of image features such as color, texture and SIFT adaptively. MKL enables us to estimate optimal weights of image features for each category. In addition, we implemented a prototype system to recognize food images taken by cellular-phone cameras. In the experiment, we have achieved the 61.34% classification rate for 50 kinds of foods. To the best of our knowledge, this is the first report of a food image classification system which can be applied for practical use.

1. はじめに

近年，健康管理への関心が高まってきている．特に「食事」に関する健康管理が注目され，より簡単に食事内容が記録できるシステムが望まれている．一般的に，「食事」は健康の変化に対して最も影響を及ぼす一因として認識されている．そのため，農林水産省と厚生労働省は共同で，健康づくりのために食生活指針を具体的な行動に結び付けるものとして，「食事バランスガイド」を策定した．医療機関等が行っている食生活アドバイスサービスでは，この「食事バランスガイド」をベースに，日頃の食事情報を把握し，食生活の改善やより良い健康づくりを支援するケースが多い．その場合，現状では管理栄養士が写真を見て手入力により食事情報の管理をしている．しかし，これはリアルタイム性に欠け，写真の枚数が多くなるほど即座に食事を分析することが困難になる．

食事の管理に関する問題を解決するために，画像認識技術を用いて，食事写真から画像中に含まれる料理の名前が自動で選択されるシステムの作成が望まれる．本研究の目的は，このようなシステムの認識エンジンを作成することである．認識エンジンで用いる画像認識手法として，Multiple Kernel Learning(MKL) を用いた方法を提案し，高精度な認識を実現する．それに加え，本研究では，提案手法を用いた認識エンジンを組み込んだ食べ物画像認識システムのプロトタイプを実装する．これは，利用者が撮影した食事写真の画像データを携帯電話から認識システムに送ることにより，写っている食べ物名を認識し，データベースに食事記録を自動生成するシステムである．

2. 関連研究

本研究では，用いる特徴量として，局所特徴による bag-of-keypoints 表現，色特徴，ガボール特徴を用いている．その特徴を Support Vector Machine(SVM) を用い 50 種類の食べ物クラスに分類する．また，各特徴を結合するために機械学習の手法である Multiple Kernel Learning(MKL) を使う．これは各特徴のカーネルを重み付き線形結合することで各特徴を結合する

†1 電気通信大学大学院 電気通信学研究科 情報工学専攻
Department of Computer Science, The University of Electro-Communications

図 1 50 種類の画像のサンプル

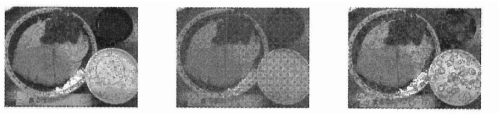

図 2 異なるオペレータによる特徴点の検出の例

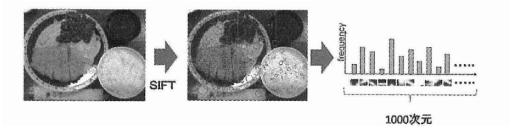

図 3 画像から抽出された局所特徴の出現頻度によるヒストグラム

出した局所特徴をクラスタリングすることで求める．この量子化された特徴 (visual words) を記述子として用いて，画像の局所特徴をコードブック中の一番近い visual word に割り当てていき，画像を 1 つの特徴ベクトルとして出現回数のヒストグラムで表す (**図 3**)．その後，その画像の局所特徴の総数でヒストグラムの各 bin を割ることによってヒストグラムを正規化する．

3.1.2 色 特 徴

色特徴とは，画像の色の分布を表現した特徴量である．色の表現方法は，RGB，HSV，Lu*v*など様々な色空間で表現されるが，本実験では，一般的によく使われる光の 3 原色の RGB 色空間のカラーヒストグラムを用いた．画像を色の特徴ベクトルに変換する際，RBG 値をそのまま特徴ベクトルにすると $256^3 = 16,777,216$ 次元になってしまうので扱いにくい．したがって，R，G，B それぞれの要素を 4 つに分割し，色空間を $64(4 \times 4 \times 4)$ 色に減色する．つまり，画像を 64 次元のヒストグラムとして表現する．基本的に，色特徴は画像のピクセルの RGB 値から生成されるヒストグラムなので，位置情報や隣接関係の情報は無視される．そこで本研究では，**図 4** のように，画像を 2×2 分割し，位置情報を考慮した色の分布で表現する．分割したそれぞれの部分画像からヒストグラムを作成し，それらを統合することにより画像を 1 つの特徴ベクトルで表現する．

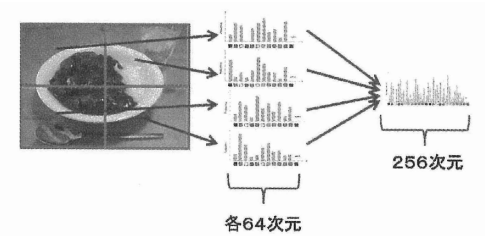

図 4 位置情報を考慮した色特徴のヒストグラム

3.1.3 ガボール特徴

ガボール特徴は，画像から局所的な濃淡情報の周期と方向を表した特徴量である．カーネルの形を固定し，それを周期を変えて伸び縮みさせたり，回転させて方向を変えたりして，様々な周期や方向のカーネルフィルタカーネルを作成する．解像度 m，方向 n のガボールフィルタは次式で表される．

$$g_{m,n}(x,y) = \frac{k_m^2}{\sigma^2} \exp\left\{-\frac{k_m^2(x^2+y^2)}{2\sigma^2}\right\} \times \left[\exp\left\{jk_m(x\cos\theta_n + y\sin\theta_n)\right\} - \exp\left(-\frac{\sigma^2}{2}\right)\right] \tag{1}$$

ここで，式 1 の k_m および θ_n は，以下のように表される．

$$\begin{aligned} k_m &= a^m \quad (0 \le m \le S-1) \\ \theta_n &= \tfrac{n\pi}{K} \quad (0 \le n \le K-1) \end{aligned} \tag{2}$$

K は方向の数，S は解像度の数，a は拡大率を表す．式 1 で表されるフィルタを用いて，それぞれに対応した空間周期の特徴を抽出 (パターン強度を数値化) する．ガボールフィルタは，特定の向きのエッジと特定の幅のエッジを抽出する．あるサンプリング点においてガボール特徴が得られる様子を**図 5** に示す．最後に，各フィルタ毎に強度の平均を求め，それをヒストグラムとする．例えば，6 方向，4 周期の場合，24 次元のベクトルとなる．ガボール特徴は局所的な情報を見るので，画像の照明変動の影響を受けにくいという利点

手法である．このセクションでは，食べ物画像を対象とした画像認識の研究と，本研究で用いる手法の既存研究について述べる．

2.1 食べ物画像の認識

食べ物画像を対象にした画像認識 (物体認識) の研究は以前から行なわれている．D. Pishva ら[1] は色特徴を用いてパンの種類を分類する手法を提案している．彼らの分類実験では，73 種類の手作りのパンの画像を用いて，95%の精度を達成している．しかし，彼らの実験で用いている画像データセットは，背景がないパンのみが写っている画像で，写っている位置も固定されているため，クラス数は豊富だが本実験で用いるデータセットと比較すると分類は容易である．

2.2 Bag-of-keypoints 表現

近年の一般物体認識の研究は，位置情報を無視して多数の局所特徴の集合によって認識物体を表現する bag-of-keypoints[2] と呼ばれる手法によってさらに発展した．Bag-of-keypoints は，統計的言語処理における bag-of-words のアナロジであり，bag-of-words は文章中に出現する単語のコードブックを基に，語順を無視して，文章を単語の集合と考える．同様にして，bag-of-keypoints では，位置を無視して画像を局所特徴の集合として考える．Csurka らは局所特徴に SIFT を使用し，画像を Bag-of-keypoits で表現し，分類器に Naive Bayes と SVM を用いることにより，7 クラスのデータベースで分類実験を行なっている[2]．この研究は SIFT と bag-of-keypoints モデルを用い，従来の手法と比べて高い性能を示している．Bag-of-keypoints 表現では，特徴点の表現方法においては，SIFT 記述子を用いることが一般的であるが，特徴点の抽出方法においては，特徴点オペレータや，グリッド，ランダムなどの様々な方法が提案されている．Nowak ら[3] は，様々な特徴点の抽出方法を試し，分類の結果を比較した結果，ランダムによる特徴点抽出方法が平均的に最も優れていると報告している．本実験では，Csurka ら[2] の提案手法である bag-of-keypoints を用いる．また，Nowak ら[3] の実験のように，局所特徴の抽出方法を，SIFT の標準的な特徴点オペレータである DoG(Difference of Gaussian) 検出だけでなく，グリッド点とランダム点で試し，比較する．

2.3 Multiple Kernel Learning (MKL)

Multiple Kernel Learning (MKL) は，SVM などのカーネルを用いた識別器を複数用いたときに，それぞれのカーネルに対し最適な重みを学習する手法である．MKL を画像認識の研究に取り入れた例は，まだ数少ないが，Varma ら[4] は，分類クラス毎に対して，最適な特徴を MKL を用いて学習する研究をしている．彼らは MKL をカーネルの選択のために用いるのではなく，複数の特徴の最適な重みを計算するために用いている．つまり，特徴を統合する方法として MKL が利用されている．これにより，クラスの分類タスク毎に対して，最適な特徴で認識することができる．彼らは，MKL で学習された重みを用いて，Caltech 101/256 などの大規模なデータセットにおいて，現在で最も良い結果を達成している．Nilsback ら[5] は，Varma らの方法を用いて，103 クラスの花画像データセットで分類実験を行なって，MKL による特徴結合の有効性を示している．また，Lampert らは,[4][5] の方法とは異なり，単に特徴を統合するのではなく，MKL を物体間の関連性の学習に適用している[6]．これにより，他の物体の存在 (context) を考慮した認識ができる．このように，MKL は画像認識の研究において様々な用途で利用できる可能性がある．本研究では，それぞれのクラスの各特徴の重みを Varma ら[4] の方法のように MKL を用いて最適な重みを求る．

3. 画像認識方法

本研究では，提案手法によって，画像データをその画像中に含まれる料理クラスに分類する．

画像認識手法の流れとして，まず，学習データ用の画像を用意し，全画像から画像特徴を抽出する．次に，学習データの特徴を用いて 1-vs-rest SVM 分類器を学習させる．最後に，学習した分類器を使って，テスト画像の料理クラスを決定する．

本実験での画像の表現方法として，局所特徴，色特徴，ガボール特徴を用いている．実験では，5 fold cross validation を用いて分類手法の性能を検証する．画像中の食事部分のバウンディングボックス (bounding box) の情報が予め与えられている 50 種類の画像のデータセット (**図 1**) を用いてマルチクラス分類を行い，平均分類率をその分類手法の性能とした．

本セクションでは，最初に，本実験で用いた特徴量の説明をする．次に，分類方法について述べる．最後に，これらの特徴を統合する方法について説明する．

3.1 特徴抽出と表現

3.1.1 局所特徴

局所特徴とは，**図 2** 左のように特徴点オペレータにより画像中の濃淡変化が大きい特徴点を多数検出し，その特徴点周りの領域を画素値や微分値等により特徴ベクトルにしたものである．画像は局所特徴の組み合わせで認識する．本研究で用いる局所特徴として，SIFT を選択した．また，3) の実験のように，局所特徴の検出方法に，SIFT の標準的な特徴点オペレータである DoG(Difference of Gaussian) 検出だけでなく，グリッド点 (図 2 中) とランダム点 (図 2 右) の 3 つのタイプを使い，画像から局所特徴を検出する．本研究では，bag-of-keypoints 表現[2] を用いて，画像を局所特徴の出現頻度として一つの特徴ベクトルに表現して認識を行なう．

Bag-of-keypoints 手法とは，画像を局所特徴の集合と捉えた画像の表現方法である．bag-of-keypoints による画像の特徴ベクトルは，visual words と呼ばれるコードブックに基づいた局所特徴の出現頻度のヒストグラムである．visual words は全学習データから抽

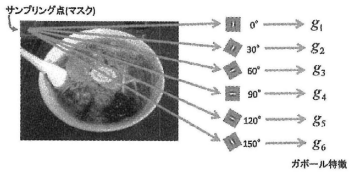

図 5 ガボールフィルタで特定のエッジを抽出した例 (K=6, S=1)

がある．ガボール特徴は局所的なパターンの特徴を表現するため SIFT と類似しているが，回転に対して不変性がないため，SIFT と異なる表現ができる．

3.2 分類・識別

本研究では，50 種類のマルチクラスに分類するために，1-vs-rest 分類法を用いて分類を行なう．あるクラスを正とし，それ以外のクラスを負としてシングルクラスの分類器を学習する．これを全てのクラスについて行なう．つまり，N クラスに分類する場合は，N 個の分類器を学習する．テストデータの分類は，学習したすべてのシングルクラス分類器にテストデータを入力し，出力値の最も大きなクラスに分類する．1-vs-rest 分類では，正解クラス以外がすべてネガティブクラスとなるので，ネガティブの学習データが多すぎてしまうという問題があるが，本研究では，そのまますべてネガティブデータとして学習することにする．その分類に用いる分類器として，SVM を用いる．

3.3 MKL による特徴統合

本研究では，特徴を統合して画像を認識するために，複数の特徴量のカーネルを線形結合することにより統合カーネルを作成し，それをサポートベクターマシン (SVM) に適用して特徴統合による画像認識を実現する．最適なカーネル (カーネルを重みつきで線形結合したカーネル) のサブカーネルに対する重み β_j を学習する．これは Multiple Kernel Learning (KML) 問題[7] と呼ばれ，統合カーネルは以下の式のように表される．

$$K_{combined}(\mathbf{x},\mathbf{x}') = \sum_{j=1}^{K} \beta_j k_j(\mathbf{x},\mathbf{x}')$$
$$\text{with } \beta_j \geq 0, \quad \sum_{j=1}^{K} \beta_j = 1. \quad (3)$$

各サブカーネルをそれぞれの特徴と対応させることによって，MKL は特徴選択や特徴統合に用いることができる．

この MKL 問題は，すべての β_j の組合わせを cross-validation により解くことできるが，カーネルの数 (特徴の数でもある)K が大きくなるにつれ，$\beta_j(j = 1,..,K)$ の組み合わせが爆発的に増大する．また，計算量との兼ね合いで刻み幅も粗くなってしまい真の最適な値が求まらない可能性が高い．そこで，最近の研究では，この MKL 問題を凸面最適化問題として効果的に解く方法が提案されている[8]．その研究の一つとして，Sonnenburg ら[8] は，単一カーネルの SVM 学習を反復することによって，最適なカーネル重み β_j を SVM の学習パラメータと同時に求める方法を提案している．MKL を画像認識における特徴統合に初めて適用した Varma ら[4] も基本的には 8) の方法を用いて最適な重みを求めている．Varma らは MKL のサブカーネルを画像の特徴と対応づけて，カーネルの線形結合を特徴の統合とすることで，MKL を画像認識に適用している．本研究でも，同様に MKL による特徴統合を試みる．

サブカーネルは，4)–6)，9) を参考にして，χ^2 カーネルを使うことにする．Zhang ら[9] は，画像分類において χ^2 カーネルは，最も性能が良かったカーネルの一つであると報告している．

カーネル法は SVM のみを前提としたものではないが，Multiple Kernel Learning は SVM のフレームワークで解く方法が一般的で，MKL-SVM と呼ばれることもある．2 クラス分類に対する MKL 問題において，N 個のデータ点 $(\mathbf{x}_i, y_i)(y_i \in \{\pm 1\})$ が与えられたとする．MKL において解くべき最適化問題の主問題は以下のように示される．

$$\begin{aligned} \min \quad & \frac{1}{2}\left(\sum_{k=1}^{K} ||\mathbf{w}_k||^2\right)^2 + C\sum_{i=1}^{N}\xi_i \quad (4) \\ \text{w.r.t.} \quad & \mathbf{w}_k \in \mathbb{R}^{D_k}, \xi \in \mathbb{R}^N, b \in \mathbb{R}, \\ \text{s.t.} \quad & \xi_i \geq 0 \text{ and} \\ & y_i\left(\sum_{k=1}^{K}\langle \mathbf{w}_k, \mathbf{\Phi}_k(\mathbf{x}_k)\rangle + b\right) \geq 1 - \xi_i, \\ & \forall i = 1,\ldots,N \end{aligned}$$

ここで，$\mathbf{w}_k = \beta_k \mathbf{w}'_k (\beta_k \geq 0, \forall k = 1,\ldots,K)$, $\sum_{k=1}^{K}\beta_k = 1$ である．

Bach ら[10] は式 4 に対して双対問題を導いた．以下に，MKL の双対問題を示す．

$$\begin{aligned} \min \quad & \gamma \quad (5) \\ \text{w.r.t.} \quad & \gamma \in \mathbb{R}, \alpha \in \mathbb{R}^N \\ \text{s.t.} \quad & \mathbf{0} \leq \alpha_i \leq C, \quad \sum_{i=1}^{N}\alpha_i y_i = 0 \\ & S_k(\alpha) = \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j \mathbf{k}_k(\mathbf{x}_i,\mathbf{x}_j) - \sum_{i=1}^{N}\alpha_i \leq \gamma, \\ & \forall k = 1,\ldots,K \end{aligned}$$

ここで，$\mathbf{k}_k(\mathbf{x}_i,\mathbf{x}_j) = \langle \mathbf{\Phi}_k(\mathbf{x}_i), \mathbf{\Phi}_k(\mathbf{x}_j)\rangle$，$N$ は学習データの個数である．単一カーネルの双対問題との違いは，カーネル毎に $S_k(\alpha) \leq \gamma$ という拘束条件があって，$\sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j \mathbf{k}(\mathbf{x}_i,\mathbf{x}_j)$ を最大化する代わりに，全カーネルで共通の上限値の γ を (符合が逆であるために) 最小化する点である．$K = 1$ の場合は，上記の問題は標準の SVM の双対問題と等価になる．

この双対問題を解くために Sonnenburg ら[8] は，次

のような単一カーネルの SVM 学習の反復による解法を提案している．(1) 最初に β_j を均等重みとする．(2)β_j を固定して，統合カーネルを単一のカーネルとみなし，通常の SVM 学習を行い，$\alpha_i(i=1..N), b$ を求める．(3) 求めた α_i を固定して，$\sum_{k=1}^{K}\beta_k S_k(\alpha)$ が増加するように β_j を変化させる．(4) 終了条件に達するまで (2),(3) を繰り返す．

4. 実　　験

実験では，提案手法をバウンディングボックス付き食事画像データベースに適用して，その有効性を検証する．MKL の重み β_f を求める方法として MKL-SVM を用いて，1-vs-rest の SVM により，クラス毎に最適な特徴の重みを求め，分類する．求められた重みは，食べ物毎の分類に必要な視覚的な特性を表していると言える．

4.1 実験データセット

実験で用いる画像データセットは，50 種類の食べ物の名前をキーワードとするテキスト検索によって Web から画像を収集してきたものを使用する．実際に実験で使用する画像は，収集してきた画像から人手で各種類 100 枚ずつ選択したものを用いる．選択される画像の基準は “ready to eat” とし，調理済の料理で，しかもすぐに食べられる状態になっているものだけを選ぶ．例えば「ラーメン」の場合，調理中の麺を茹でているところの画像や，インスタントラーメンのパッケージしか写っていないものなどは正解画像として採用していない．また，認識対象の学習にはできるだけノイズを含めないようにすることが望ましい．しかし，Web から集めた画像データには多くのノイズが含まれている．したがって，データベース中の画像を認識の対象物体が含まれている領域とそうでない背景領域を分離することが必要である．実験で用いるデータセットには，bounding box を用いて領域指定する．これらの画像データは，学習データ，評価データとして利用する．50 種類の食べ物画像データセットは図 1 で示す．

4.1.1 実験で用いる特徴

ここでは，実験で使う特徴である色特徴，局所特徴，ガボール特徴について説明する．色特徴は，色空間を 64 次元に量子化して画像を 2×2 に分割しているので 256 次元のベクトルで表現する．局所特徴とガボール特徴に関しては，複数の種類の表現がある．局所特徴は，異なる特徴点の抽出方法 (3 タイプ: DoG, グリッド, ランダム) とベクトルの次元数 (2 タイプ:1000, 2000) があり，合計 6 種類の bag-of-keypoints のベクトルで表現する．グリッド点は bounding box 内から半径 4, 8, 12, 16 の局所領域を 10 ピクセル間隔で検出する．ランダム点は bounding box 内からランダムに 3000 個検出する (半径は 0.8 から 10.0 の間)．ガボール特徴は，4 スケール，6 方向の合計 24 個のガボールフィルタを使って bounding box 内のすべての領域に対して特徴を抽出する．色特徴と同じように画像を分割するが，3×3 と 4×4 の 2 タイプあるので，結果として 216 次元と 384 次元のベクトルで表現している．最終的に，合計 9 種類の方法で画像を表現する．

4.2 パラメータ設定

この実験で用いる分類器は multiple kernel を使った SVM である．本実験では，カーネル関数は χ^2 カーネルを用いているため，ある 2 つのデータ点 i,j に対し，最終的な結合カーネルは以下の形式をとる．

$$K_{comb}(i,j)=\sum_{f=1}^{9}\beta_f k_f(i,j)$$
$$=\sum_{f=1}^{9}\beta_f\exp\Big(-\gamma_f\chi_f^2(\mathbf{x}_f(i),\mathbf{x}_f(j))\Big)$$
$$\text{where } \chi^2(\mathbf{x},\mathbf{y})=\sum\frac{(x_i-y_i)^2}{x_i+y_i}$$

ここで，$\mathbf{x}_f$ は特徴 f の特徴ベクトル (色や BoK のヒストグラム) で，β_f は特徴 f に対する重みである．パラメータ γ_f は次のように設定する．

4.2.1 cross-validation によるパラメータ設定

求めるパラメータは SVM のコストパラメータ C とカーネルパラメータ γ_f である．これらのパラメータをデータセットの cross-validation により推定する．各特徴毎に $C\in\{1,10,100,1000,10000,10000\}$ と $\gamma_f\in\{0.0001,0.0005,0.001,0.005,0.01,0.05,0.1,0.5,1.5,10,50,10\}$ の組み合わせで，最も良い結果を出したパラメータで MKL を行なう．今後このパラメータ設定方法の MKL のことを MKL(cross-validation) と書く．

4.2.2 距離の平均によるパラメータ設定

Zhang らは[9] は，χ^2 カーネルのパラメータ γ_f に，全ての学習画像のベクトルの組み合わせの χ^2 距離の平均の逆数を設定することによって，良い結果を報告している．距離の平均で割ることは，特徴空間の正規化を意味し，異なる特徴空間でも同じ距離尺度で比較することができる．cross-validation で求めたパラメータの他にこの値でも分類実験を行なってみる．しかし，SVM のコストパラメータ C に関しては，あらかじめ設定しておかなくてはならないため，cross-validation によるパラメータ設定と同様に，$C\in\{1,10,100,1000,10000,10000\}$ で実験し，最も結果が良かった C の値を選択する．結果として，$C=1$ 以外性能に違いがみられないので，代表して，$C=1000$ の結果を示す．今後，このパラメータ設定方法の MKL のことを MKL(mean-χ^2distance) と書く．

このようにして求めたカーネルパラメータを用いて，1-vs-rest 分類器として，それぞれのクラスに対して重みパラメータ β_f は学習される．

4.3 評 価 方 法

分類実験の手法は，実験データセットの 5-fold cross validation によって評価する．各種類の画像セットを 5 つのグループに分割し，ある 1 つのグループを評価

データとして利用するとき，残りの4グループは学習データとして分割器を訓練させる．これを5通りそれぞれのグループが評価データとして分類実験を行い，その5回の結果の平均を分類結果とする．

分類結果の評価に用いる基準として，正確性の観点からみた**適合率** ((正しく分類されたデータ集合)/(分類されたデータ集合)) と，完全性の観点からみた**再現率** ((正しく分類されたデータ集合)/(分類されるべきデータ集合))，さらに，与えられた画像がどのクラスに分類されたか知るために混合行列 (confusion matrix) も作成する．

混合行列 (confusion matrix): マルチクラス分類システムを評価する基準として混合行列 M_{ij} をこのように定義する．

$$M_{ij} = \frac{|\{I_k \in C_j : h(I_k) = i\}|}{|C_j|}$$

ここで，$i, j \in \{1, \ldots, N_c\}$($N_c$ はクラスの数)，C_j はクラス j のテストデータ集合，$h(I_k)$ は，画像 I_k に対して分類器から出力された値の最大値を得たクラスとする．混合行列の対角成分の平均値 ($\sum_{i=1}^{N_c} M_{ii}/N_c$) を**平均分類率**として，50 クラス分類全体の評価に用いる．

4.4 実験結果

MKL を用いて特徴を統合した結果と各特徴を単独で用いた結果を**表 1** で示す．各特徴単独の結果は，χ^2 カーネルのパラメータ γ_f を χ^2 距離の平均の逆数に設定したときのものである．この結果をみると，MKL を用いて特徴を統合することによって大幅に分類精度が向上したことが分かる．また，カーネルパラメータ推定に χ^2 距離の平均を用いることによって，cross-validation による最適なカーネルパラメータ選択の結果よりも良くなることが分かった．**図 7** は分類率が高かったクラスの認識結果である．SVM の出力値の高いものはほとんど正解している．MKL(mean-χ^2distance) の食べ物の種類別の分類精度は**図 6** の混合行列で示す．コロッケ vs ロースカツ vs エビフライやチャーハン vs ピラフなど，視覚的に類似しているものは部分的に他クラスと混合しているが，全体的には正確に分類できている．**図 8** は，結果が良かった MKL(mean-χ^2distance) の食べ物の各種類 (各 1-vs-rest 分類器) に対して学習した重みを示したものである．一番，重要視されていた特徴は BoK(DoG2000) の特徴であった．色は平均の重みが約 0.1 で全体としてはあまり重要ではない特徴という結果になったが，クラスによっては色が重要な特徴になることもある．色の重みが大きかった 「オムライス」や「エビチリ」は**図 9** にある結果を見ても，色に特徴がある種類であることが分かる．「オムライス」は背景の色が様々あるように見えるが，対象領域の bounding box 内では玉子の黄色とケチャップの赤色の占める割合が多い．また図 8 の平均の重みをみると，同じ特性のある特徴 (BoK(DoG1000) と BoK(DoG2000) のペアや gabor3 × 3 と gabor4 × 4 のペアなど) 同士で比較してみると，次元数の多い方が重みが大きくなっている．これは表 1 からみても，次元数の多い方が分類精度が高いことから理解できる．

表 1 特徴統合と特徴単独で分類した結果

特徴	平均分類率
color	38.18%
BoK(dog1000)	26.52%
BoK(dog2000)	27.48%
BoK(grid1000)	26.10%
BoK(grid2000)	27.68%
BoK(random1000)	28.42%
BoK(random2000)	29.70%
gabor3 × 3	31.28%
gabor4 × 4	34.64%
MKL(cross-validation)	53.16%
MKL(mean-χ^2distance)	61.34%

表 2 MKL(mean-χ^2distance) 手法の分類率 (再現率) の上位 5 位と下位 5 位

TOP5	クラス名	分類率	WORST5	クラス名	分類率
1	味噌汁	0.97	1	角煮	0.18
2	ざるそば	0.94	2	生姜焼き	0.28
2	うな重	0.94	3	トースト	0.31
4	ポタージュ	0.91	4	ピラフ	0.39
5	オムライス	0.87	4	春巻き	0.39

表 3 MKL(cross-validaion) 手法の分類率 (再現率) の上位 5 位と下位 5 位

TOP5	クラス名	分類率	WORST5	クラス名	分類率
1	味噌汁	0.96	1	角煮	0.03
1	うな重	0.96	2	たこ焼	0.06
2	ざるそば	0.95	3	トースト	0.08
4	ポタージュ	0.90	4	春巻き	0.16
5	オムライス	0.87	5	生姜焼き	0.18

2 つの MKL の上位 5 位と下位 5 位の分類結果 (再現率) を**表 2** と**表 3** で示す．再現率を比較してみると，上位 5 位にあるクラスについては 2 つの MKL 手法ともほとんど同じである．違いがみられたのは下位 5 位の部分である．「角煮」，「生姜焼き」，「トースト」，「春巻き」など下位 5 位に入っているクラスは両方とも同じであるが，再現率に大きな違いがみられる．結果が悪かった「角煮」の適合率をみてみると，再現率 3% (3 枚正解) であるのに対し，適合率が 50%もあった．つまり，角煮に分類された画像の枚数が 6 枚しかない．他にも，「トースト」，「たこ焼」は 13 枚しかそれらのクラスに分類されていなかった．一方，再現率が上位の「うな重」と「ざるそば」を見てみると，うな重として分類された画像の枚数は 173 枚，ざるそばとして分類された画像は 167 枚もあった．このように MKL(cross-validation) は偏った分類がされてしまい，結果が悪くなったのだと考えられる．

最後に，2 つの MKL の分類の許容クラス数を変化させたときの平均分類率の変化を表したものはは**図 10** である．ランキング 3 位までを許容するすと MKL(mean-χ^2distance) の分類率は 80%を越えていることがわかる．

5. 食べ物認識システムの作成

本セクションでは，カメラ付き携帯電話で撮影した画像を認識するプロトタイプシステムを紹介する．

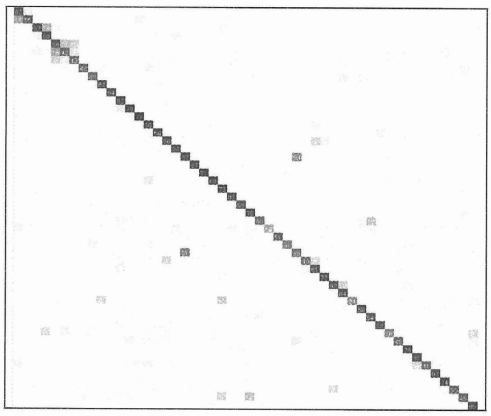

図 6 MKL (mean - χ^2 distance) の混合行列

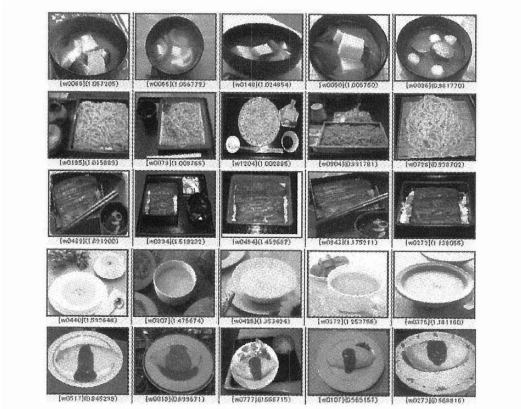

図 7 分類率が高かった種類の結果

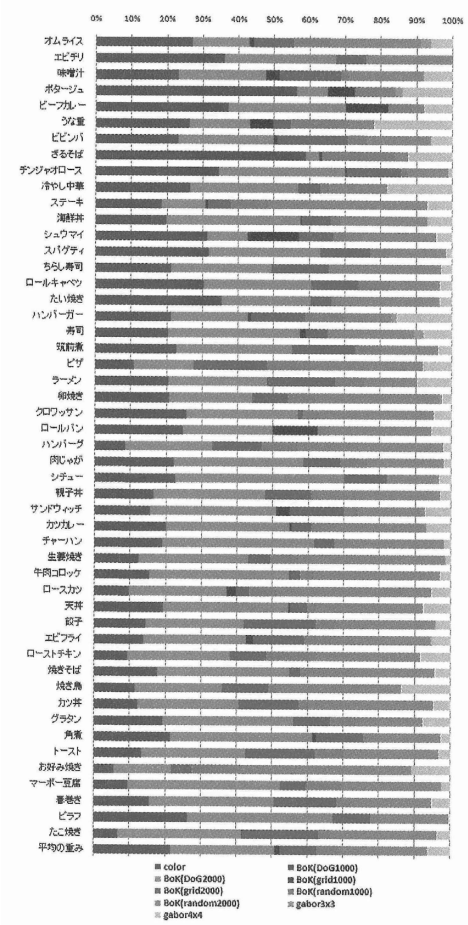

図 8 MKL(mean-χ^2distance) の各 1-vs-rest 分類器に対して学習した重みの割合

携帯版システムの構成は**図 11** のようになる．まず始めに，ユーザは画像をメールでアップロードする．アップロードされた画像はサーバーの画像データベースに登録される．その時，アップロードした時間やユーザのアドレスも画像 ID と一緒に記録される．これにより，いつ，誰が，何をアップロードしたのかが把握できる．

次に，アップロードされた画像は，認識システム部によってその画像に含まれる物体が何の料理なのか認識される．認識システムは，データベースにある各料理に対しての類似度を計算する．出力値が高いものから順に料理のランキングを生成し，その結果を記録する．その後，ユーザのもとにもランキングを返す．

最後に，ユーザはそのランキングをもとに正しい記録を選択することにより正解情報をフィードバックし，サーバー側はその情報を記録する．

このシステムの特徴は，

- 画像をアップロードするとその画像の対象を認識して，類似度の高い順にデータベースにある料理名を出力する．

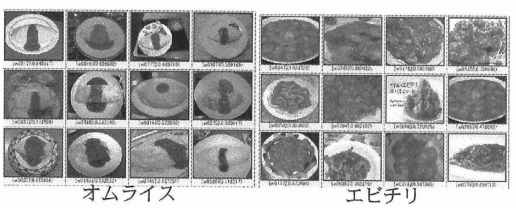

図 9 色の重みが大きかったオムライスとエビチリ

- アップロード時に画像の他に，時間，アドレスの情報も記録する．
- 正しい結果を知るために結果に対してフィードバック機能も持つ．(今後その画像が学習画像にも使用できる)

である．認識には本研究で提案した手法を用いる．認識対象となるデータベースには，本実験で用いた 50 種類データベースをもとにしている．

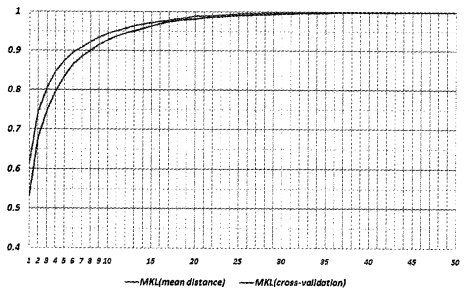

図 10 2 つの MKL の分類率の変化

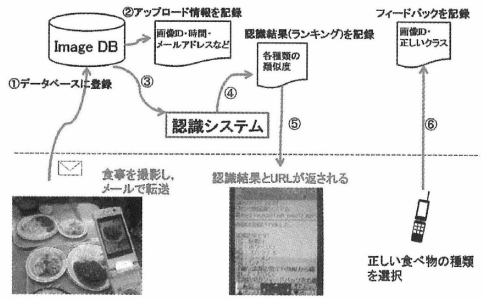

図 11 携帯版システム構成

本システムは，10ヶ月間試験的に運用した．その結果，10ヶ月間に 166 枚がアップロードされ，166 枚中 62 枚が正解で分類率 37.35%，3 位以内を許容すると分類率 55.42%，5 位以内では分類率 57.23%であった．

提案手法の評価実験では，バウンディングボックス付の画像から学習して，画像のバウンディングボックスの領域のみを分類したが，本システムでは実際に携帯電話に内蔵されているカメラで撮影した食べ物画像を対象にしているため，認識精度は 61.34%に比べると低いものに留まっている．一因としては，食事画像の撮影方法をユーザに特に指定しなかったため，斜めから撮影した画像や，暗くてコントラストが低い画像も含まれていることが挙げられる．そこで，写真の撮影の仕方をシステムが認識しやすいように，例えば，真上から画面いっぱいに対象が写るようにするなどと，ユーザに指定することによって，手法の評価実験の認識精度に近づけられると考えられる．

6. おわりに

本研究は，食べ物画像から料理を認識する手法を提案し，その提案手法を用いて，携帯電話などで撮影した写真を認識するシステムを作成した．最新の機械学習の手法である Multiple Kernel Learing を適用し，各カテゴリ毎に複数の特徴の最適な重みを求めて統合する方法を提案した．本実験では，学習画像データセットで提案手法の精度を検証し，50 種類で平均分類率 61.34%を達成した．また，分類の許容クラス数を変化させることにより，上位 3 位までの分類を許容すると提案手法の分類率は 80%を越えることができた．しかし，実際に，携帯電話で写した画像に対する認識精度は 37.35%で，5 位まで考慮すると 57.23%であり，写し方により認識精度が低下する．

さらに認識精度をあげる方法としては，本実験で用いた特徴とは異なる識別能力の新しい特徴を用いて，MKL で特徴を統合することが考えられる．また，学習画像データベースの拡張が考えられる．これは，システムの運用にあたり，現在のデータベースの種類では対象となる食べ物がまだ少ない．そこで，現在では 85 種類の食べ物画像のデータベースを構築中である．

本研究の最終的なねらいは，提案手法の認識エンジンを取り入れた自動カロリー計算システムを構築することである．このシステムが実現すれば，食事写真を毎食ごとに撮影してシステムに入力していくだけで，利用者の食生活の傾向が見えてくる．最終的には，食事以外にも睡眠や運動の情報も合わせ，利用者の生活パターンの解析まで発展する．

参 考 文 献

1) Pishva, D., Kawai, A., Hirakawa, K., Yamamori, K. and Shiino, T.: Bread Recognition Using Color Distribution Analysis, *IEICE Trans. on Information and Systems*, Vol.84, No.12, pp.1651–1659 (2001).
2) Csurka, G., Bray, C., Dance, C. and Fan, L.: Visual categorization with bags of keypoints, *Proc.of ECCV Workshop on Statistical Learning in Computer Vision*, pp.1–22 (2004).
3) Nowak, E., Jurie, F. and Triggs, B.: Sampling Strategies for Bag-of-Features Image Classification, *Proc.of European Conference on Computer Vision* (2006).
4) Varma, M. and Ray, D.: Learning The Discriminative Power-Invariance Trade-Off, *Proc.of IEEE International Conference on Computer Vision* (2007).
5) Nilsback, M. and Zisserman, A.: Automated flower classification over a large number of classes, *Proc. of Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing* (2008).
6) Lampert, C.H. and Blaschko, M.B.: A Multiple Kernel Learning Approach to Joint Multi-Class Object Detection, *Proc. of the German Association for Pattern Recognition Conference* (2008).
7) Lanckriet, G. R. G., Cristianini, N., Bartlett, P., Ghaoui, L. E. and Jordan, M. I.: Learning the Kernel Matrix with Semidefinite Programming, *Journal of Machine Learning Research*, Vol.5, pp.27–72 (2004).
8) Sonnenburg, S., Rätsch, G., Schäfer, C. and Schölkopf, B.: Large Scale Multiple Kernel Learning, *Journal of Machine Learning Research*, Vol. 7, pp. 1531–1565 (2006).
9) Zhang, J., Marszalek, M., Lazebnik, S. and Schmid, C.: Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study, *International Journal of Computer Vision*, Vol.73, No.2, pp.213–238 (2007).
10) Bach, F. R., Lanckriet, G. R. G. and Jordan, M. I.: Multiple kernel learning, conic duality, and the SMO algorithm, *Proc. of International Conference on Machine Learning* (2004).