Vol.2011-MPS-86 No.17
Vol.2011-BIO-27 No.17
2011/12/1

MKLによる食事画像認識の追試

内村　麻里奈[†1]　　高田　雅美[†2]　　城　和貴[†2]

本研究では食事画像認識をする．食事のカロリーなどの情報を必要としているユーザがいる．しかしながら，食事の種類はとても多いため，一般ユーザにとって，正しい食事の情報をえることは困難である．この問題を解決するために，食事画像を使って専門家がアドバイスするシステムがある．ただし，このシステムでは，処理能力に限界がある．そこで，食事画像に対して，画像認識技術を適用することによって，認識の自動化を行う．分類機の 1 つとして，SVM（Support Vector Machine）がある．この SVM は，入力された画像の特徴を用いて，正しく分類することができる．食事の種類は多いため，複数の特徴で表す必要がある．そのため，これらの特徴を SVM で利用するために，統合しなければならない．このための手法として，MKL（Multiple Kernel Learning）を用いる．この MKL を用いた SVM の性能を調べるために，85 種類の食事画像を用いて，実験を行う．

Additional Test of Food Image Recognition by MKL

Marina Uchimura,[†1] Masami Takata[†2]
and Kazuki Joe[†2]

In this paper, we recognize food image. Users need information about food calorie so on. However, it is too difficult to have correct information about food because of variety foods. To solve this problem, the system which specialists advise with food image has been developed. However, this system has a limitation in capacity. Therefore, food images should be recognized automatically by using image-recognition technique. SVM (Support Vector Machine) is one of supervised learning. The SVM can classify correctly with given image features. A number of features are needed to treat so many food. These features have to be combined by using SVM. Consequently, MKL (Multiple Kernel Learning) should be adopted. We experiment with 85 categories of food images to find out quality of SVM using MKL.

1. はじめに

食事を摂取する際に，カロリーや塩分などの栄養に関する情報を必要とするユーザがいる．特に，糖尿病，腎疾患，肝疾患，冠動脈疾患，コレストロール高値などのユーザにはそれぞれ特別な食事が必要なものであるため，外食の際などに注意が必要となる．この際，ユーザが食事の内容を判断するのではなく，食事画像を携帯電話などで送ることで，専門の栄養士が判断し，アドバイスを返すシステム[1)]がある．しかし，このシステムでは栄養士が目視で認識しているため，リアルタイム性に欠け，人手がかかり，効率的ではない．

そこで，画像認識技術[2)]を使い，携帯電話などのカメラで送られてきた食事画像を自動で認識し，料理名，含まれている栄養素，カロリーなどその料理に関する情報を表示するシステムの開発が望まれている．すでに，MKL（Multiple Kernel Learning）[3)]を用いて SVM（Support Vector Machine）[4)]で食事画像を認識する研究がある[5)]．この研究では，特徴量として，色特徴の他に，局所特徴量である SIFT 特徴[6)]やテクスチャ特徴量であるガボール特徴量[7)]を用いている．これにより，分類数が多い食事画像を認識することが可能となると考えられている．この食事認識では，複数の特徴量を 1 つに統合するために，MKL を使い，それぞれの特徴量に最適な重みを学習させることによって，統合カーネルを作成する．この総合カーネルを用いて，SVM で分類する．

SVM において，教師データとして与える画像が，認識性能に大きな影響を与える．また食事画像を対象としているため，分類数が多くなり，複数の特徴量を用いて分類するべきである．そこで，本論文では，MKL を用いた SVM の性能を調べるために，85 種類の食事画像を各 100 枚用意し，検証を行う．

第 2 章では SVM で用いる特徴量として SIFT 特徴，ガボール特徴，色特徴について紹介する．第 3 章では MKL について説明する．第 4 章では食事を認識するための既存研究ついて述べる．第 5 章では実験結果を説明する．

†1 奈良女子大学理学部情報科学科
Departmant of Information and Computer Sciences, Faculty of Science, Nara Wumen's University
†2 奈良女子大学大学院人間文化研究科複合現象科学専攻
Departmant of Advanced Information and Computer Sciences, Graduate School of Humanity and Sciences, Nara Wumen's University

Vol.2011-MPS-86 No.17
Vol.2011-BIO-27 No.17
2011/12/1

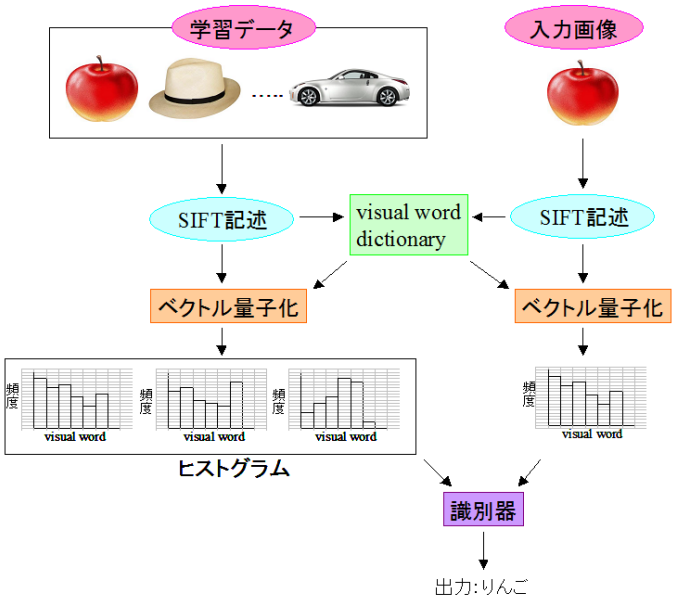

図 1 bag-of-keypoints の流れ

2. 特 徴 量

SVM で分類するためには画像情報から特徴量を得ることが必要である．また，MKL は複数の特徴量を 1 つに統合することが可能であるため，本論文では，3 種類の特徴量を用いる．

2.1 節では SIFT 特徴，2.2 ガボール特徴，2.3 色特徴について説明する．

2.1 SIFT 特徴

SIFT(Scale Invariant Feature Transform) 特徴を用いることによって，スケール変化，回転変化に不変な特徴量を記述することができる．ゆえに，画像認識の際に必要となる特徴量として有用である．SIFT のアルゴリズムは次の 4 つの手順 で行われる．

手順 1 スケールとキーポイントの検出
手順 2 キーポイントのローカライズ
手順 3 オリエンテーションの算出
手順 4 特徴量の記述

手順 1 のスケールとキーポイントの検出では，DoG （Difference of Gaussian）処理[7] やグリッド点，ランダム点などでキーポイントの候補を検出する．DoG 処理とは，入力画像とスケールの異なるガウス関数 $G(x, y, \sigma)$ の畳み込みによって求めた平滑化画像 $L(u, v, \sigma)$ の差分（DoG）から求める．式で表すと，

$$D(u, v, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(u, v) \quad (1)$$
$$= L(u, v, k\sigma) - L(u, v, \sigma) \quad (2)$$

となる．手順 2 のキーポイントのローカライズでは，特徴点としてふさわしくない点を削除する．手順 3 のオリエンテーションの算出では，特徴点ごとに方向を正規化することで回転に不変な特徴量を求める．手順 4 の特徴量の記述では，SIFT descriptor により 128 次元の特徴量を記述する．

SFIT は異なる画像間で抽出された各キーポイントの SIFT 特徴量を比較することで，画像間の対応点の検索が可能である．そのため，特定の物体が同定であるか判断するには有効な手段だが，同じ種類ではあるが，異なる画像に対して，SIFT による対応点を求めることはできない．よって，同じ分類か判断するにはそのまま使うことは難しい．そこで，Bag-og-Features 手法[8] を使う．Bag-of-Features とは画像を局所特徴量の集合とみなし，位置情報を無視して画像認識を行うことである．**図 1** では，SIFT 特徴量を用いて Bag-of-Features で画像の分類を行う流れである．全学習データの SIFT 特徴ベクトルを N_{SIFT} 個のクラスタにクラスタリングする．N_{SIFT} 個のクラスタの各セントロイド（中心となるベクトル）を visual word とする．それぞれ特徴ベクトルから一番近い visual word を検索し画像中に visual word がぞれぞれいくつあったかでヒストグラム化する．そして学習画像群から識別器を作成，判定させる．

2.2 ガボール特徴

ガボール特徴と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．様々な方向と周期が設定できるため，高精度な認識ができ，画像処理では虹彩認識や指紋認証にも利用されている．

解像度 r，方向 d のガボールフィルタは次式で表される．σ はガウス関数である．

Vol.2011-MPS-86 No.17
Vol.2011-BIO-27 No.17
2011/12/1

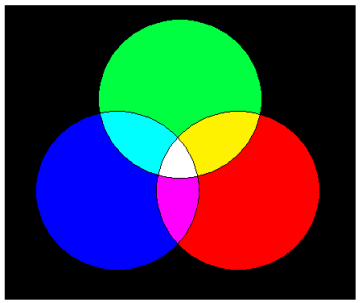

図 2 加法混色

$$g_{r,d}(x,y) = \frac{k_r^2}{\sigma^2}\exp\left\{-\frac{k_r^2(x^2+y^2)}{2\sigma^2}\right\} \times \left[\exp\{jk_r(x\cos\theta_g(d)+y\sin\theta_g(d))\} - \exp\left(-\frac{\sigma^2}{2}\right)\right] \quad (3)$$

ここで，式の k_r および $\theta_g(d)$ は，以下のように表される．

$$k_r = m^r(0 \le r \le N_r - 1) \quad (4)$$

$$\theta_g(d) = \frac{d\pi}{N_d}(0 \le d \le N_d - 1) \quad (5)$$

ここで，N_d は方向の数，N_r は解像度の数，m は拡大率を表す．式 (3) で表したフィルタを画像の各ピクセルに対して重なるように合成する．画像中のエッジがフィルタの向き，周期と同じであった場合，フィルタの山や谷に重なる部分の値だけ増幅される．また，周辺全体の値に変化がなければ，山と谷のそれぞれで増幅された値同士が打ち消されるため，全体の和はゼロになる．よって，ガボール特徴は特定の向きのエッジと特定の幅のエッジを抽出することができる．ゆえに，局所的な情報をみるため，画像の照明変動の影響を受けにくいという利点がある．

2.3 色

色の表現方法として赤（Red），緑（Green），青（Blue）の原色を用いる RGB 法[9] がある．RGB 法を計算機で扱う場合，各原色に 8 ビットを割り当て，0 から 255 の整数で表し，それらの数値の違いで RGB の割合を変えさせることによって，色を表現する．そのため，計算機が認識できる色の種類は，$256 * 256 * 256 = 16777216$ となる．RGB では，さらに輝度に関する値も保持している．この輝度は，各ピクセルごとに保持されており，ピクセル間は，ガンマ補正をかけることで表現できる．そのため，赤，青，緑の各要素にどれだけ含まれているかで幅広い色が表現できる．各要素は輝度最小から輝度最大まで範囲があり，すべての要素が最小であれば黒，すべての要素が最大であれば白となる，加法混色である．**図 2** は加法混色の例である．

特徴量として用いる場合は，画像の各ピクセルに含まれている RGB 値を求め，それをヒストグラム化する必要がある．この際，色の分布などの情報は取得されない．つまり，1 つの画像に対して，1 つの特徴量しか得られないことに注意が必要である．

3. Multiple Kernel Learning

MKL とは，SVM などのカーネルを用いた識別器を複数用いる際に，それぞれのカーネルに対して最適な重みを学習する手法である．論文 10) において，Varma らは各クラスに対して，適切な重み学習して，各特徴量を総合し，その総合カーネルをサポートベクターマシン（SVM）に適応することで画像認識させている．

カーネルに重みをつけて統合したカーネルは以下の式で表現される．

$$K_{combined}(t,t') = \sum_{j=1}^{N_T} w_{MKL}(j)k_j(t,t')$$
$$with\ w_{MKL}(j) \ge 0, \sum_{j=1}^{N_T} w_{MKL}(j) = 1 \quad (6)$$

ここで，各サブカーネルを k_j，重みを $w_{MKL}(j)$，N_T をサブカーネル数とする．それぞれの特徴量を各サブカーネル k_j に対応させることで，それぞれの特徴量に適切な重み $w_{MKL}(j)$ をつけ統合する．6 を解く方法として，すべての重み w_{MKL} の組み合わせを cross-validationni で解くことができるが，カーネルの数（特徴量の数）N_T 大きくなるにつれて w_{MKL} の組み合わせが膨大になる．そこで，凸面最適化問題として効果的に解く研究が行われている．[5]. その 1 つとして単一カーネルでの SVM 学習を反復することによって，最適な重み w_{MKL} を求める方法がある．この方法では，大規模なデータに対してよい結果を出している．サブカーネルを画像の各特徴量と対応させ，1 つのカーネルをつくり，画像の特徴量とする．分類するクラスが多い場合，1 つの特徴量では分類しきれない場合があると考えられる．よって，複数の特徴量を用いて，より画像の特徴を表現する．また，MKL でどの特徴量を重要とするか学習するため，より精度が高い認識ができると考えられる．Varma ら[10] も基本的

Vol.2011-MPS-86 No.17
Vol.2011-BIO-27 No.17
2011/12/1

には同じような手法を用いて最適な重みを求めている[5]．

2 クラス分類に対する MKL 問題において，N 個のデータ点 $(x_i, y_i)(y_i \in \pm 1)$ が与えられたとすると，MKL において解くべき最適化問題の主問題は，以下の式で表される．

$$\min \frac{1}{2}\left(\sum_{j}^{N_T} \parallel w_{MKL}(j) \parallel^2\right)^2 + C\sum_{i=1}^{N}\xi_i \tag{7}$$

ただし，$\xi_i \geq 0$ and

$$y_i\left(\sum_{N_T}^{j=1}\langle w_{MKL}(j), \Phi_j(x_j)\rangle + b\right) \geq 1 - \xi_j, \forall = 1, \ldots, N_T \tag{8}$$

ここで，

$$w_{MKL} \in R^{D_j}, \xi \in R^{N_T}, b \in R \tag{9}$$

であり，

$$w_{MKL}(j) = \beta_j w_j'(\beta_j \geq 0, \forall_j = 1, \ldots, N_T), \sum_{j=1}^{N_T}\beta_j = 1 \tag{10}$$

である．$\Phi_j(x_j)$ はカーネルマップである．Bash ら[11] は式 7 に対して双対問題を導いている．この双対問題は以下で表せる．

$$\min\{\ \gamma - \sum_{i=1}^{N_T}\alpha_i\} \tag{11}$$

ここで，

$$\leq \alpha_i \leq C, \sum_{i=1}^{N_T}\alpha_i y_i = 0,$$

$$S_j(\alpha) = \frac{1}{2}\sum_{i,l=1}^{N_T}\alpha_i\alpha_l y_i y_l k_j(x_i, x_l) - \sum_{i=1}^{N_T}\alpha_i \leq \gamma,$$

$$\forall j = 1, \ldots, N_T \tag{12}$$

となる．各変数は，$\gamma \in R$，$\alpha \in R^N$，$k_j(x_i, x_l) = \langle \Phi_j(x_i), \Phi_j(x_l)\rangle$ を意味する．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り，

$$\sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,l=1}^{N}\alpha_i\alpha_l y_i y_l k(x_i, x_l) \tag{13}$$

を最大化する代わりに，全カーネルで共通の上限値の γ を符号が逆であるため最小化する点である．もし，$N_T = 1$ のとき，この式は通常の SVM の双対問題と等価になる．この双対問題を解くために，以下のような，単一カーネルの SVM での学習の反復を使った方法が提案されている．

(1) 最初に β_l を均等重みとする．

(2) β_l を固定し，総合カーネルを単一カーネルとみなし，通常の SVM 学習を行い，$\alpha_i(i = 1 \ldots N), b$ を求める．

(3) 求めた α_i を固定して，

$$\sum_{j=1}^{N_T}\beta_j S_\alpha$$

が増加するように β_l を変化させる．

(4) 終了条件に達するまで 1，2 を繰り返す．

4. Multiple Kernel Learning を用いた Support Vector Machine

食事画像を認識させるための手段として SVM がある．SVM とは，教師ありデータを用いる機械学習の 1 つであり，計算量が比較的少なく，単純な原理にも関わらず，未知なデータに対して識別性能が優れている．データの集合 P を分離する超平面は

$$\langle w_{SVM}(p_i), p\rangle + b = 0$$

で定義でき，このときの超平面を $(w_{SVM}(p_i), p)$ と表す．データ p が超平面のどちら側にあるかによってクラス分類を行う．w_{SVM} は重みベクトル，b は閾値である．式で表すと，

$$f(p) = \langle w_{SVM}(p_i), p\rangle + b$$
$$= \sum_{i\in P}\langle w_{SVM}(p_i \cdot p_i) + b \tag{14}$$

である．判別式は，

$$sgn(f(p)) = \begin{cases} 1 & (f(p) > 0) \\ -1 & otherwise \end{cases} \tag{15}$$

Vol.2011-MPS-86 No.17
Vol.2011-BIO-27 No.17
2011/12/1

1,-1 はそれぞれデータが属するクラスのラベルである．

MKL を用いて食事画像を認識を行った研究がある．色特徴，ガボール特徴，SIFT 特徴など複数の特徴量を用い，MKL で各特徴のカーネルに重みをつけ総合カーネルをつくり，SVM に適応する．Varma ら[10] は，MKL を使って複数の特徴の最適な重みを計算し，Caltech 101/256 などのデータセットにおいて，最も良い結果を出している．

局所特徴量として SIFT 特徴量を用いている．特徴点を求め，全学習データから局所特徴量をクラスタリングし，visual words を求める．そしてそれを基に画像中の各特徴に visual words を割りふっていき，それぞれの回数をヒストグラムで表す．画像の特徴数は異なるので，総特徴数でヒストグラムの要素を割ることで正規化をしている．特徴点は DoG 処理，グリッド点，ランダム点の 3 通りからそれぞれ求めており，次元は 1000, 2000 の 2 通り求め，合計 $3*2=6$ 通りの bag-of-keypoints のベクトルで表現する．グリッド点は画像中から半径 4，8，12，16 の局所領域を 10 ピクセル間隔で検出する．ランダム点は画像中から半径は 0.8 から 10.0 の間でランダムに 3000 個検出する．

ガボール特徴は式 3 で表されたフィルタを使い，それぞれに対応した空間周期の特徴を抽出し，各フィルタごとに強度の平均を求め，それをヒストグラムとする．4 スケール，6 方向の 24 個のフィルタを使って特徴量を抽出するので，24 次元のベクトルができる．色特徴と同様に画像を分割して $3*3$ と $4*4$ の 2 通りで求めるため，実際には，216 次元と 384 次元になる．

色特徴は，各ピクセルの RGB 値をヒストグラムにしたものである．それぞれの要素は 256 通りで表されており，$256*256*256$ 通りなのでそのまま特徴量としてヒストグラムで表すと次元数が多くなってしまう．よって，各要素を 4 通りに減色することで $4*4*4$ 通りとして 64 次元のヒストグラムで表す．ただし，この方法では，画像全体に含まれる色の出現頻度の分布はヒストグラムで表されるが，色の出現情報は保持されない．そこで，画像を $2*2$ の 4 分割にし，各部分ごとに色特徴を求め，$64*64*64$ 次元のヒストグラムを作ることで，位置情報を考慮することができる特徴量を計算する．

これら 9 種類の特徴量で画像を表現し，MKL を用いて総合カーネルを作成し，SVM を用いて画像認識させる．SVM のカーネル関数は χ^2 カーネルを使うため，総合カーネルは

$$K_{combined}(i,l) = \sum_{f=1}^{9} \beta_f k_f(i,l)$$
$$= \sum_{f=1}^{9} \beta_f \exp\left(-\gamma_f \chi_f^2(x_f(i), x_f(l))\right) \quad (16)$$

ここで，

$$\chi^2(x,y) = \sum \frac{(x_i - y_i)^2}{x_i + y_i}$$

x_f は特徴 f の特徴ベクトルであり，β_f は特徴 f に対する重みである．

5. 追試実験

本研究では MKL を用いて複数の特徴量を統合し，SVM で食事画像を認識する．

用いる特徴量はガボール特徴，SIFT 特徴，色特徴である．本研究では，その中で MKL においてガボール特徴，SIFT 特徴はそれぞれ重要視されていたパラメータの方を用いる．SIFT 特徴の場合は DoG 処理でキーポイントを検出し，次元数を 2000 で表した特徴量であり，ガボール特徴は画像を $4*4$ に分割したものを特徴量とする．

食事画像は 85 種類を集め，1 種類につき 100 枚用意する．**図 3** は，85 種類の食事の名前とそのサンプル画像を列挙したものである．画像はすぐに食べられる状態のものをインターネット上から無作為に収集する．また，画像中に食事以外の背景は食事を認識するためには不必要な情報なので，削除する．

また，SVM および MKL の実行には，SHOGUN toolbox[12] を使う．SHOGUN toolbox はカーネル法に関するツールボックスである．機械学習を実装するために多数のアルゴリズムを提供しており，様々な実装を統一的なインターフェースで利用できる．

6. まとめ

本研究では，食事画像を自動的に分類するための手法について紹介した．この手法では，自動的に分類するために，SVM を適応している．SVM で画像を扱うためには，画像から特徴量を得る必要がある．また，食事画像は，種類が多いため，1 つの特徴量で全ての特徴量を表すことは困難であると考えられる．そこで，特徴量として，SIFT 特徴，ガボール特徴，色特徴の 3 種類を用いる．これらを SVM に適応するためには，複数の特徴量を 1 つに

Vol.2011-MPS-86 No.17
Vol.2011-BIO-27 No.17
2011/12/1

統合する必要がある．そこで，MKL を用いて，総合カーネルを作成している．この MKL を用いた SVM の食事画像認識の検証を行うために，85 種類の食事画像を 100 枚用意した．SHOGUN を用いて実行したところ，2011 年 11 月 3 日現在実験中であり，結果は発表時に行う．

参 考 文 献

1) 旭化成ライフサポート株式会社：げんき！食卓コンシェルジュ（オンライン），入手先⟨http://shoku365.com/⟩（参照 2011-11-03）．
2) 奈良先端科学技術大学院大学 OpenCV プログラミングブック制作チーム：OpenCV プログラミングブック, 株式会社毎日コミュニケーションズ（2007）
3) Sonnenburg, S. , Rätsch, G., Schäfer, C. and Schölkopf, B.: Large Scale Multiple Kernel Learning, *Proce.Intl.Conf .Computer Vision*, pp.1150–1157 (1999).
4) Nello, C. and Jhon, S.T.：*An Introduction to Support Vector Machines and other kernel-based learning methods*, Cambridge University Press, (2000).（大北剛訳：サポートベクターマシン入門, 共立出版（2005））．
5) 上東太一，甫足創，柳井啓司：Multiple Kernel Learning による 50 種類の食事画像の認識，電子情報通信学会論文誌 D， Vol.J93-D, No.8, pp.1397–1406 (2010).
6) David, G. Lowe.: Object Recognition from Local Scale-Invariant Features, *Proce. Intl.Conf.Computer Vision*,pp.1150–1157 (1999).
7) Manjunath, B.S.: Texture features for browsing and retrieval of image data, Vol.18,pp.837–842 (1996).
8) Eric, Nowak. , Frëdëric, Jurie. , Bill, Triggs.：Sampling Strategies for Bag-of-Features Image Classification, Vol.60, pp.91–110 (2004).
9) 大田登：色彩工学（第 2 版）， 東京電機大学出版局（2001）．
10) Varma, M. and Ray, D.: Learning The Discriminative Power-Invariance trade-Off, *Proce.Intl.Conf.Computer Vision*, pp.1–8 (2007).
11) Bash, F.R., Lanckriet, G. R. G., Jordan, m.l.: Multiple kernel learning, conic duality, and the SMO algorithm, *Proce.Intl.Conf.Machine learning*,(2004).
12) Shogun: ,available from ⟨http//www.shogun-toolbox.org/⟩（参照 2011-11-01）．

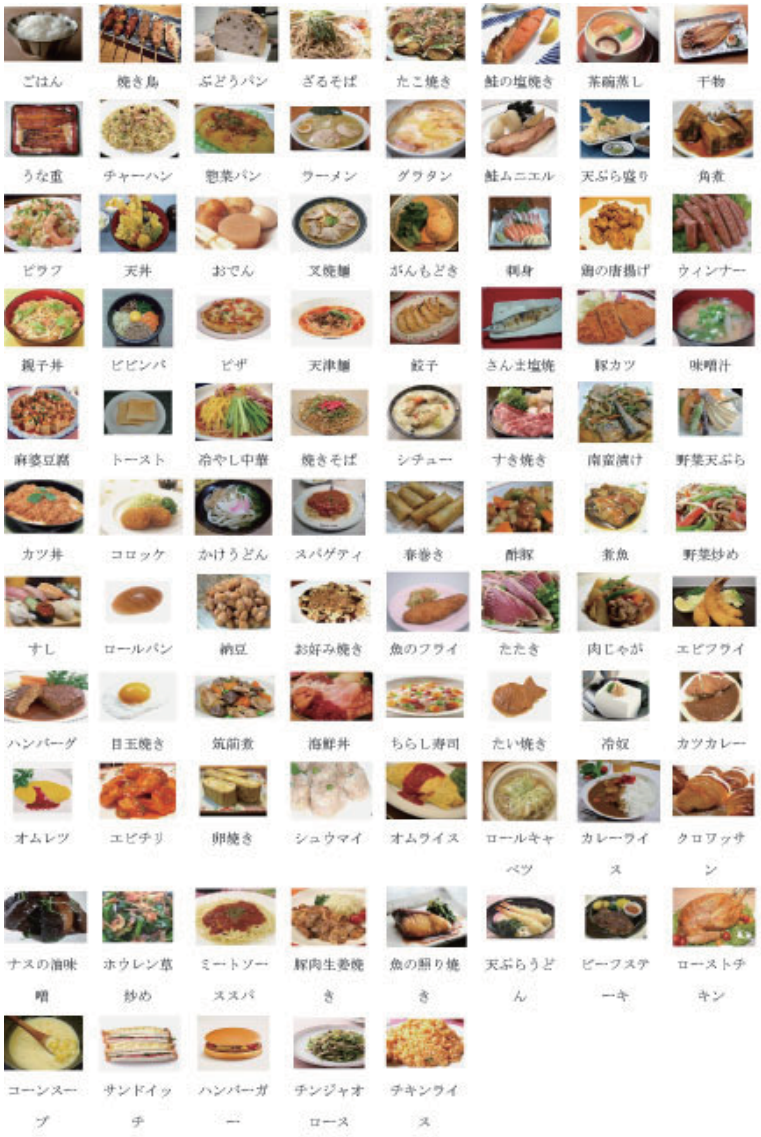

図 3 食事画像 85 種類