

Consensus based Object Localization

複数画像からの共通物体検出

坂田 惇[†] 和田 俊和[†][†] 和歌山大学システム工学部 〒 640-8510 和歌山県和歌山市栄谷 930 番地E-mail: [†]{j-sakata,twada}@vrl.sys.wakayama-u.ac.jp

あらまし 本論文では、インターネットの画像検索を用いて集められた画像群において、個々の画像上に写る対象の領域を推定する手法を提案する。提案手法では、複数の画像間で共通する特徴を合意 (consensus) として形成し、その consensus に最もマッチする領域を各画像の中から特定 (localize) する。そして各画像で localize された領域の特徴をもとにして再度 consensus を形成する、という一連の手続きを反復する。領域形状としては楕円を用い、同心楕円周上の輝度値に対するフーリエ変換によって特徴抽出を行う。また領域パラメータの探索には MCMC 法を用いる。実験により、比較的単純な画像群では対象の画像内領域をある程度推定することができ、提案手法の有効性が確認できた。

キーワード 複数画像, consensus, MCMC, object localization, 画像検索

1. はじめに

近年, Flickr などの画像アップロードサイトの普及に伴い, インターネット上には膨大な量の画像情報が蓄積されるようになってきている。これらの画像群には, 内容に関連するテキスト情報がけられているものも多く, テキスト情報を利用した画像検索はすでに一般的に行われている。例えば, Google 画像検索ではキーワードを入力するだけで, 特定の対象に関する画像を簡単に集めてくることができる。このようにして集めてきた大量の画像データを用いて, その対象の「見え」に関する情報を自動的に学習することが出来れば, 画像検索エンジンを用いて未知の対象を画像から検出, 認識するシステムが構築できるはずである。しかし, このようにして集めてきた画像集合はテキストで指定した対象を必ず含んでいるわけではなく, また画像上には指定した対象がどの程度の大きさでどこに写っているのかは予め分からず, 対象以外の被写体が一緒に写されていることもある。したがって, これら, ある対象に関する画像から, その対象の「見え」に関する情報を集約・学習し, その対象を自動的に検出, 認識するということは, 現在のところ実現できてはいない。本論文では, 複数枚の画像に共通に写されている対象が何であり, 個々の画像ではどこに写されているのかを解析する手法を提案する。この手法に類似する手法は従来から存在している [1][2] が, 画像信号や特徴ベクトルの一致に強く依存するボトムアップ的な手法がほとんどであるため, より実用的には, 対象に関する特徴ベクトルの統計なばらつきを許容する手法が必要であった。提案手法では, 複数の画像間で共通する特徴を合意 (consensus) として形成し, その consensus に最もマッチする領域を各画像の中から特定 (localize) する。そして各画像で localize された領域の特徴をもとに

して再度 consensus を形成する、という一連の手続きを反復する。この計算の枠組みを Consensus based Object Localization と呼ぶことにする。この計算を行うには, 領域からの特徴抽出と, その特徴に基づく領域の探索アルゴリズムが必要になる。本論文では, 領域形状として楕円を用い, 同心楕円周上の輝度値に対するフーリエ変換によって特徴抽出を行う。また領域パラメータの探索には, MCMC[3] を用いることにする。

2. 特徴ベクトルの記述

Consensus based Object Localization において, 画像上の localize された領域の特徴抽出が必要になる。本章では本論文でこの localize された領域の特徴抽出方法について述べる。

2.1 高速フーリエ変換

特徴抽出を行う際に用いる高速フーリエ変換 (Fast Fourier Transform, FFT) のアルゴリズムについて述べる。離散フーリエ変換 (Discrete Fourier Transform, DFT) の高速演算アルゴリズムとして, 高速フーリエ変換がデジタル信号処理における重要なツールの一つとして広く知られている。長さ N の DFT の直接計算には N^2 回の複素演算が必要なのに対し, FFT は, DFT を $N \log_2 N$ に比例する演算量で計算することができる。演算を飛躍的に高速化させる。整数 $m (> 0)$ に対して $Z_m := \{0, 1, 2, \dots, m-1\}$ とおき, データの長さを $N = 2^m$ とすると, 長さ N のデータ列 $[x(n)]_{n=0}^{(N-1)}$ に対する DFT は,

$$X(k) = \sum_{n=0}^{N-1} x(n) W_N^{nk}, W_N = e^{-j2\pi/N}, k \in Z_N$$

(1)

で定義される．ここで，長さ N を二つの因数の積 $N_1 N_2$ とおくことで，式 (1) のインデックス変換を

$$\begin{aligned} n &= N_1 n_2 + n_1, n_1 \in Z_{N_1}, n_2 \in Z_{N_2} \\ k &= N_2 k_1 + k_2, k_1 \in Z_{N_1}, k_2 \in Z_{N_2} \end{aligned} \quad (2)$$

とおくと，関係式

$$W_N^{nk} = W_{N_1}^{n_1 k_1} W_N^{n_1 k_2} W_{N_2}^{n_2 k_2} \quad (3)$$

が得られる．式 (1) は，式 (3) を用いると

$$X(N_2 k_1 + k_2) = \sum_{n_1=0}^{N_1-1} W_{N_1}^{n_1 k_1} W_N^{n_1 k_2} \quad (4)$$

$$\sum_{n_2=0}^{N_2-1} x(N_1 n_2 + n_1) W_{N_2}^{n_2 k_2} \quad (5)$$

と表される [8]．これは，長さ $N = N_1 N_2$ の DFT を長さ N_1 の DFT と長さ N_2 の DFT として演算することを意味し，必要とする複素乗算回数 M は

$$M = N_1 N_2 (N_1 + N_2 + 1) \quad (6)$$

となり，明らかに式 (1) を直接計算する場合の乗算回数 N^2 より少ない．整数である基数で割り切れる長さ N の DFT を，長さ $1/r$ の r 個の DFT に分解する基本式を設定し，基本式の分解構造にのっとって式 (1) の長さ N の DFT を $N = 4$ ，または 2 の DFT の組み合わせになるまで繰り返し分解するのが，Cooley-Tukey 型 FFT アルゴリズムである．

2.2 楕円特徴の生成

localize された楕円領域を特徴記述する手法を述べる．まず，画像上に楕円パラメータ x, y, a, b (楕円の中心，短径，長径を表す) によって表される楕円と，その楕円内部への同心楕円を合計 R 個生成する．そして各楕円周上で，画素を N 点サンプリングし (図 1) フーリエ変換を行う．ここで，サンプリングデータが画像信号であるので，高周波数成分はほとんどないことから N はあまり大きな値でなくてもよいと考えられる．フーリエ変換によって得られたフーリエ係数 $a_n^{(r)}, b_n^{(r)} (n = 1, \dots, N-1)$ の値は，サンプリングデータが画素値ゆえ実信号なので， $n = N/2$ を境に複素共役である．また，直流成分 a_0 はサンプリング楕円周上の平均画素値， $a_1, \dots, a_{N/2}, b_1, \dots, b_{N/2}$ はもと信号に含まれる各周波数の余弦波，正弦波の振幅を表す．これを R, G, B の各チャンネルに対して求め，1 列に並べて特徴ベクトル F とする (図 2)．

$$F = \{\mathbf{v}_{red}^T, \mathbf{v}_{green}^T, \mathbf{v}_{blue}^T\}^T \quad (7)$$

$$\begin{aligned} \mathbf{v}_{color} &= \{a_0^{(0)}, a_1^{(0)}, \dots, a_{N/2}^{(R-1)}, b_0^{(0)}, b_1^{(0)}, \dots, b_{N/2}^{(R-1)}\}^T \\ color &= red, green, blue \end{aligned} \quad (8)$$

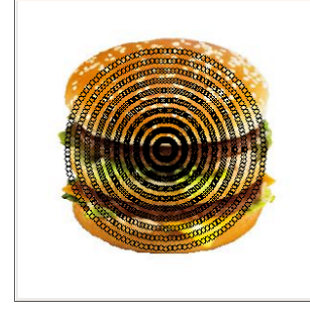


図 1 サンプリング

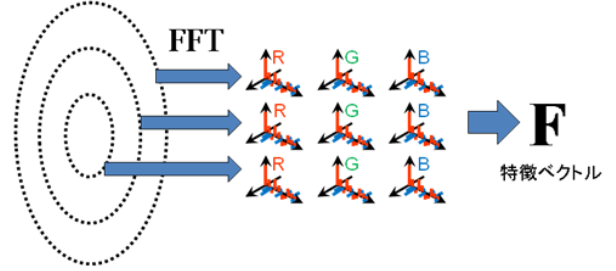


図 2 特徴ベクトル

3. MCMC のアルゴリズムと領域推定法

本章では MCMC のアルゴリズムと，MCMC を用いた領域推定方法について述べる．

3.1 MCMC のアルゴリズム

MCMC (Markov Chain Monte Carlo) は，統計物理学から統計学の文脈に輸入され目覚ましい成果を上げている統計手法の一つで，マルコフ連鎖サンプリングを用いてモンテカルロ積分を行うのがクリティカルなステップであることからマルコフ連鎖モンテカルロ法と総称される．追跡アルゴリズムとしてパーティクルフィルタ，ConDensation とも呼ばれている．この手法は追跡対象の位置・姿勢，形状などの状態の確率分布を，ランダムサンプリングによって時刻ごとに推定することで追跡を実現する手法である．本研究では物体の追跡ではなく，位置推定にこれを用いる．時刻 t における対象の状態変数を $\mathbf{X}_t$ と定義し，同時刻の観測を Z_t とし，時刻 t までに得られる観測系列を $Z_t = \{Z_1, \dots, Z_t\}$ とする．このとき， $\mathbf{X}_t$ の事後分布 $p(\mathbf{X}_t | Z_t)$ はベイズの定理から次式のようになる．

$$p(\mathbf{X}_t | Z_t) = k_t p(Z_t | \mathbf{X}_t) p(\mathbf{X}_t | Z_{t-1}) \quad (9)$$

ただし， k_t は正規化項である．また， $p(\mathbf{X}_{t-1} | Z_{t-1})$ は時刻 t における事前分布， $p(Z_t | \mathbf{X}_t)$ は，状態 $\mathbf{X}_t$ の尤度を表している．ここで，事前分布 $p(\mathbf{X}_t | Z_{t-1})$ は，前の時刻の事後分布 $p(\mathbf{X}_{t-1} | Z_{t-1})$ および，あらかじめ与えた $\mathbf{X}_t$ の推移確率 $p(\mathbf{X}_t | \mathbf{X}_{t-1})$ から次式のように求められる．

$$p(\mathbf{X}_t | Z_{t-1}) = \int_{\mathbf{X}_{t-1}} p(\mathbf{X}_t | \mathbf{X}_{t-1}) p(\mathbf{X}_{t-1} | Z_{t-1}) d\mathbf{X}_{t-1}$$

(10)

式(8)(9)より、各時刻における尤度 $p(Z_t|X_t)$ を求めることで、事後分布 $p(X_t|Z_t)$ の推定ができる。しかし(9)の積分範囲全体で事前分布を推定することは現実的でない。MCMCでは、事前分布 $p(X_t|Z_{t-1})$ に従ったランダムサンプリングによる仮説の生成と、その各仮説に対して尤度を求めることで、事後分布を離散的に近似し推定する。実際には、各仮説について尤度を求め、その比に従った重みづけを行うことで、事後分布を近似的に推定している。各仮説を $s_t^{(n)}$ とすると、その重み $\pi_t^{(n)}$ は次式により求める。

$$\pi_t^{(n)} = \frac{p(Z_t|X_t = s_t^{(n)})}{\sum_{i=1}^N p(Z_t|X_t = s_t^{(i)})} \quad (11)$$

MCMCでは、事前分布 $p(X_t|Z_{t-1})$ に従った仮説 $s_t^{(n)}$ の生成、各仮説の重み $\pi_t^{(n)}$ の計算を、「選択」、「予測」、「観測」の手順によって行う(図3)。アルゴリズムを以下に示す。

(1) 選択

時刻 t の仮説の集合 $s_{t-1} = \{s_{t-1}^{(1)}, \dots, s_{t-1}^{(N)}\}$ から、各仮説の重み $\pi_{t-1}^{(1)}, \dots, \pi_{t-1}^{(N)}$ の比に従い、仮説の選択を行う。ここで選択した仮説集合を $s'_{(t-1)}$ とする。

(2) 予測

選択した仮説集合 $s'_{(t-1)}$ の各仮説から推移確率 $p(X_t|X_{t-1})$ に従ったサンプリングを行う。

(3) 観測

s_t の各仮説 $s_t^{(n)}$ について尤度 $p(Z_t|X_t = s_t^{(n)})$ を推定し、各仮説 $s_t^{(n)}$ の重み $\pi_t^{(n)}$ を計算する。

これら「選択」、「予測」、「観測」の処理を繰り返すことで、各時刻の事後分布 $p(X_t|Z_t)$ が推定される。ある時刻 t における状態 X_t の期待値は次式のようになる。

$$E[X_t] = \sum_{n=1}^N \pi_t^{(n)} s_t^{(n)} \quad (12)$$

3.2 MCMC を用いた領域推定

複数の画像から共通性をもとに、各画像での対象の位置を推定する。ある程度の画像には中央に対象物があると仮定し、前処理として入力複数の画像それぞれの中央を任意の大きさの楕円で囲み、前章で述べた方法により特徴ベクトル F_i を生成する。そこから特徴ベクトルの平均 $F_{average(k=0)}$ を求め、これを初期の consensus とする。ここで、 i は各画像の添え字、 k は処理のステップ数を表す。次に、入力画像それぞれに MCMC を割り当てる。各画像に割り当てられた MCMC が、それぞれの画

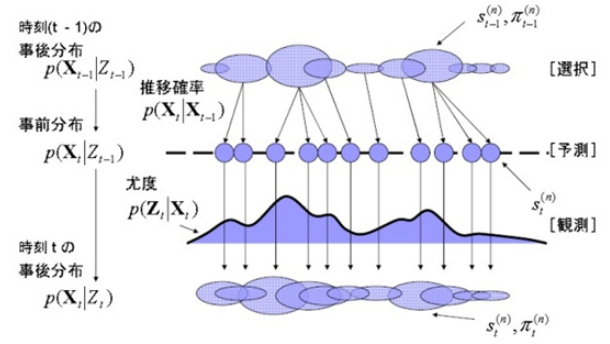


図3 MCMC

像内の対象領域を推定することになる。MCMCの仮説パラメータには、楕円パラメータの中心、短径、長径を示す x, y, a, b を設定する。したがって、仮説 $s_{i(k)}^{(n)}$ が与えられたとき、その楕円パラメータを用いて画像上に楕円で領域指定し、領域内の情報を用いることによって特徴ベクトル $F_{i(k)}^{(n)}$ を求める。初期ステップ ($k=0$) では画像上様々な位置、大きさの楕円を発生させるランダムサンプリングを行い、各仮説について尤度を評価する。初期ステップ以降 ($k>0$) は、前ステップの仮説の尤度による重みが大きくなるほど多くの仮説を生成することにより探索する領域を絞り込み、領域推定を行っていく。

3.3 尤度評価

consensus と localize した領域の類似度によって尤度を決定することにより、consensus によりマッチする領域の尤度を高くし、領域を絞り込みやすくする。具体的には、各仮説の尤度 $L(d)$ は consensus となる $F_{average(k)}$ と、仮説から求めた特徴ベクトル $F_{i(k)}^{(n)}$ との距離 d によって以下のように求める。

$$d_{i(k)}^{(n)} = |F_{i(k)}^{(n)} - F_{average(k)}| \quad (13)$$

$$L(d) = \exp\left(\frac{-d^2}{\sigma^2}\right) \quad (14)$$

3.4 consensus の更新

初期ステップでは consensus を各画像の中央へ楕円を配置することによって得られた特徴ベクトルの平均によって形成しているが、初期ステップ以降では次のように更新していく。まず、各画像において、各ステップでの状態ベクトルの期待値 $s_{i(k)}^{candidate}$ を、仮説の重みづけ平均によって以下のように求める。

$$s_{i(k)}^{candidate} = \sum_n \pi_{i(k)}^{(n)} s_{i(k)}^{(n)} \quad (15)$$

一般的に MCMC では、全ての仮説を用いて期待値を求めるが、本論文ではこの期待値を尤度上位 1% の仮説の重みづけ平均とした。尤度の低い仮説を無視することにより、良質な期待値を得ることができ、収束を早めることが期待できる。求められた仮説の期待値は、各ス

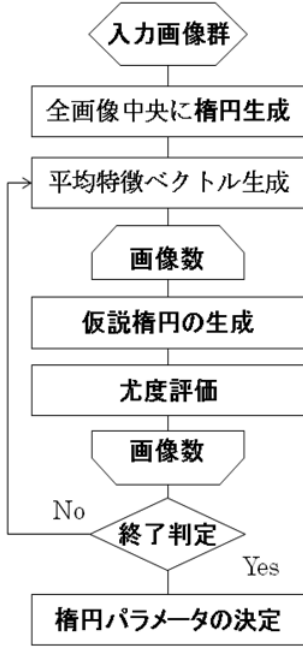


図 4 フローチャート

テップでの楕円パラメータの期待値といえる．この楕円パラメータの期待値 $s_{i(k)}^{candidate}$ を用いて，特徴ベクトルの期待値 $F_{i(k)}^{candidate}$ を求める．次に，平均特徴ベクトル $F_{average(k)}$ を，特徴ベクトルの期待値 $F_{i(k)}^{candidate}$ の重みづけ平均によって以下のように更新することによって，consensus を形成する．

$$F_{average(k)} = \sum \omega_i F_{i(k)}^{candidate} \quad (16)$$

$$\omega_i = \frac{L(|F_{i(k)}^{candidate} - F_{average(k)}|)}{\frac{1}{I} \sum_i L(|F_{i(k)}^{candidate} - F_{average(k)}|)} \quad (17)$$

3.5 終了判定

繰り返し計算の終了条件を，各画像での仮説パラメータの分布の広がりによって決定する．分布の広がりを計算するために，重みづけ分散を計算した．処理の段階ごとで分散の比率を求め，その比率が 1 に近づけば終了とする．このとき，このサンプルを $s_{i(k)}^{(n)}$ ，その平均を $\overline{s_{i(k)}}$ ，重みを $\pi_{i(k)}^{(n)}$ ，サンプル数を N とすると， $X_{i(k)}$ の分散 $D(X_{i(k)})$ は，

$$\overline{s_{i(k)}} = \frac{1}{N} \sum_n s_{i(k)}^{(n)} \quad (18)$$

$$D(X_{i(k)}) = \sum_n \pi_{i(k)}^{(n)} (s_{i(k)}^{(n)} - \overline{s_{i(k)}})^2 \quad (19)$$

となる．よって， D を閾値としたとき， $|1 - D(X_{i(k)})/D(X_{i(k-1)})| < D$ を満たせば終了させる．提案手法のフローチャートを図 4 に示す．

4. 実験

本章では，提案手法での実験結果を示す．提案手法が

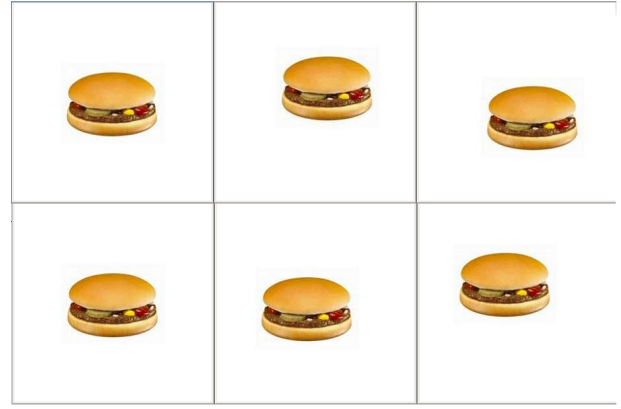


図 5 実験 1 の入力画像群

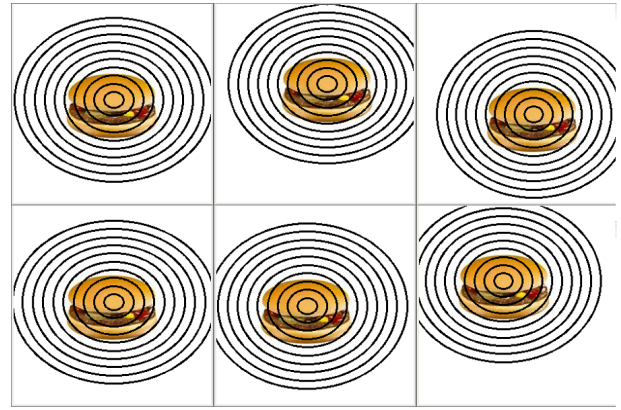


図 6 実験 1 の結果

どのような画像群に対して有効かを確認するために，単純な画像群から複雑な画像群まで順を追って実験を行った．実験内容は以下のように対象や背景を変えて実験を行った．

- (1) 同一対象と単純な背景の画像群
- (2) 同一対象と複雑な背景の画像群
- (3) 類似対象と単純な背景の画像群
- (4) 類似対象と複雑な背景の画像群
- (5) 一般的な画像群

実験条件は，画像サイズ 300×300 の 6 枚の画像群を入力とし，MCMC の仮説数を 1000，同心楕円数を 10，楕円周上サンプル画素数 256，終了判定閾値を $D=0.1$ とした．

4.1 同一対象と単純な背景の画像群での実験

無地の背景に，同一のハンバーガーを配置した画像群で実験を行った（図 5，6）．全ての画像で同じように対象領域を楕円で囲めている．楕円が大きいのはハンバーガーの背景の白色も，全画像共通であるためと考えられる．

4.2 同一対象と複雑な背景の画像群での実験

様々な背景に，同一のハンバーガーを配置した画像群



図 7 実験 2 の入力画像群

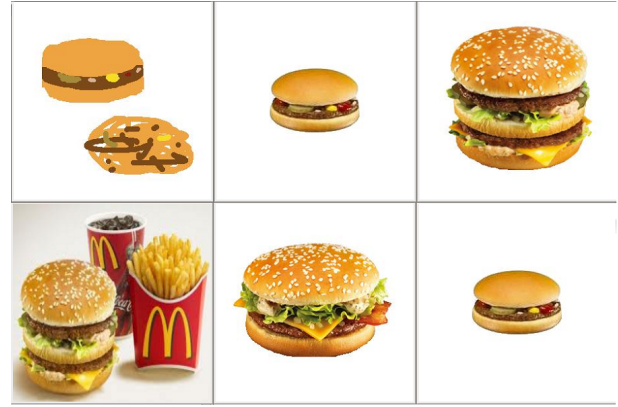


図 9 実験 3 の入力画像群



図 8 実験 2 の結果



図 10 実験 3 の結果

で実験を行った（図 7, 8）．左上の画像の楕円が小さくなった原因としては、他の画像の背景が暗いのにに対し、左上の画像の背景は明るい．このため、全画像の対象領域の平均特徴ベクトルによって算出している consensus に、ハンバーガーの外側は暗いという情報が付加されたため、楕円が大きくなると明るい背景を含んでしまう左上の画像では、背景を含まないように楕円が小さくなったと考えられる．

4.3 類似対象と単純な背景の画像群での実験

無地の背景に、類似したハンバーガーを配置した画像群で実験を行った（図 9, 10）．左上の画像において、結果では左上の手書きハンバーガーの領域を推定できたが、処理の途中段階では右下の同じ色を含む物体にも高めの尤度分布が得られ、楕円パラメータのばらつきが見受けられた．一般的な MCMC では、複数対象になると尤度分布にばらつきが生じ、状態ベクトルが定まらなくなるので、この問題を今後検討していきたい．また、この画像群でも背景の白色を共通部分とみなし、楕円が縦長になった．

4.4 類似対象と複雑な背景の画像群での実験

複雑な背景に、類似したハンバーガーを配置した画像群で実験を行った（図 11, 12）．一段のハンバーガーが

多いので、二段のハンバーガーでは一段目に楕円がかかっており、共通領域とみなされていることが分かる．

4.5 一般的な画像群

画像検索によって得られた、一般的なハンバーガーの画像群に対して実験を行った（図 13, 14）．いずれの画像においても楕円の中心はハンバーガー内にあるが、スケールをうまく捉える事ができていない．

5. ま と め

本論文では、対象を表すテキストをクエリとして検索した複数枚の画像集合の中から、共通に写された物体を抽出する Consensus based Object Localization を提案し、その実装と実験結果について述べた．実装では、キーポイント検出などのボトムアップ処理を伴わない画像領域の特徴記述方法と、得られた特徴から画像上の領域を決定するための Localization のアルゴリズムについて検討を行った．本論文では、画像上の対象領域を楕円で表現し、その楕円領域内の同心楕円上の画素値をフーリエ変換することによって領域の特徴記述を得る方法を採用し、Localization のアルゴリズムとしては MCMC を採用した．実験により、比較的単純な画像群では対象の画像内領域をある程度推定することができ、提案手法の有効性が確認できた．実験においては合成画像を用



図 11 実験 4 の入力画像群



図 13 実験 5 の入力画像群



図 12 実験 4 の結果

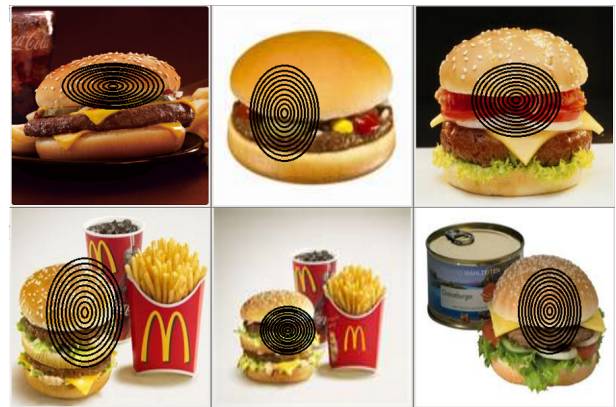


図 14 実験 5 の結果

い、どの画像にも対象が一つだけ含まれているという限定した条件において行ったので、今後様々な画像でも検討していく。また、領域推定の精度評価についても検討していく。今後の主な課題としては、次の2点を挙げることができる。1点目は、対象推定数を可変に拡張することである。画像内に複数の対象が含まれていると、それぞれの領域で尤度が高くなり期待値が定まらない。この問題については、対象の個数を推定することが必要になるため、Jump STATE など複数対象に対応させた MCMC[4,5,6,7] を利用する方法が考えられる。また、対象が含まれていない場合には、全仮説の尤度が全体的に低くなると考えられるので、全仮説の尤度が全体的に低い画像を処理の途中段階で除く方法が考えられる。2点目は、初期 consensus の生成方法である。一般的にの MCMC の反復処理では処理結果、収束時間が初期値に依存する。本論文では、初期 consensus を形成するために、各画像の中央に任意の大きさの楕円を配置する前処理を行っている。しかしこれは一般的な画像には対象が中央付近にあると仮定しているためであって、入力画像群の中に中央付近に対象が含まれていない、もしくはスケールが多様であると良い結果が期待できない。この問題を解決するためには、初期楕円の配置方法を検討する必要がある。

文 献

- [1] Rother, C., Kolmogorov, V., Blake, A: Grabcut: Interactive foreground extraction using iterated graph cuts. In: ACM SIGGRAPH, ACM(2004) 314
- [2] Wen-Sheng Chu, Chia-Ping Chen, and Chu Song Chen.: MOMI-Cosegmentation: Simultaneous Segmentation of Multiple Object among Multiple Images. In ACCV. (2010)
- [3] M.Isard and A.Blake: " CONDENSATION — Conditional Density Propagation for Visual Tracking ", International Journal of Computer Vision, pp .1-36(1998)
- [4] Isard, M. and MacCormick, J.: BraMBle: A bayesian multiple blob tracker, in International Conference on Computer Vision(ICCV2001), Vol. 2, pp. 34-41 (2001)
- [5] MacCormick, J. and Blake, A.: A probabilistic exclusion principle for tracking multiple objects, International Journal of Computer Vision, Vol. 39, No. 1, pp. 57-71(2000).
- [6] Tao, H., Sawhney, H. and Kumar, R.: A sampling algorithm for tracking multiple objects, in IEEE Workshop on Vision Algorithms, pp. 53-68 (1999).
- [7] Tweed, D. and Caway, A.: Tracking Many Objects Using Subordinated CONDENSATION, in Proceedings of the British Machine Vision Conference, pp. 283-292 (2002).
- [8] J.W. Cooley and J.W. Tukey, " An algorithm for machine calculation of complex Fourier series, " Math. Comput., vol19, no. 90, pp. 297-301, April 1965.