社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
IE2010-156,MVE2010-144(2011-03)

CHLAC 特徴の周期性解析による料理映像中の繰り返し調理動作区間の抽出と識別

久原 卓† 出口 大輔† 高橋 友和†† 井手 一郎† 村瀬 洋†

† 名古屋大学 大学院情報科学研究科 〒464-8601 愛知県名古屋市千種区不老町
†† 岐阜聖徳学園大学 経済情報学部 〒500-8288 岐阜県岐阜市中鶉 1-38

E-mail: † tkuhara@murase.m.is.nagoya-u.ac.jp, {ddeguchi,ide,murase}@is.nagoya-u.ac.jp
†† ttakahashi@gifu.shotoku.ac.jp

あらまし 本報告では，料理映像から「切る」や「混ぜる」といった繰り返し動作が行われている映像区間を抽出し，その区間の調理動作を識別する手法を提案する．繰り返し動作区間の抽出には，映像フレーム中の動作の位置に依存しない特徴量と依存する特徴量という 2 種類の特徴量を用いる．これにより，繰り返し動作の振動中心が移動する動作および一定である動作の両方に対して，高い抽出精度を維持する．そして，これら特徴量の周期性をフーリエ変換により解析し，区間抽出を行う．一方，調理動作の識別では，映像フレーム中で調理動作が行われる位置が多様であることを考慮し，動作の位置に依存しない特徴量を用いる．繰り返し動作区間抽出の実験では 0.78，調理動作識別の実験では 0.77 の精度が得られ，このことから本手法の有効性を確認した．

キーワード 料理映像，動作解析，CHLAC 特徴，フーリエ解析

Extraction and Recognition of Repetitive Cooking Motion Segments in Cooking Video by Periodicity Analysis of CHLAC Feature

Taku KUHARA†, Daisuke DEGUCHI†, Tomokazu TAKAHASHI††, Ichiro IDE†, and Hiroshi MURASE†

† Nagoya University, Graduate School of Information Science Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan
†† Gifu Shotoku Gakuen University, Faculty of Economics and Information 1-38, Nakauzura, Gifu-shi, Gifu, 500-8288 Japan

E-mail: † tkuhara@murase.m.is.nagoya-u.ac.jp, {ddeguchi,ide,murase}@is.nagoya-u.ac.jp
†† ttakahashi@gifu.shotoku.ac.jp

Abstract This paper proposes a method for extracting segments that contain repetitive cooking motions such as “cutting” and “mixing”, from cooking videos, and for recognizing the cooking motions in the segments. The proposed method extracts repetitive cooking motion segments by two types of features; One is a feature that depends on the location of the motion within video frames and the other is a feature invariant to the location. As a result, high identification accuracy is expected to be maintained on both a repetitive motion whose oscillation center moves along time and a repetitive motion with a constant oscillation center. Next, the proposed method analyses the periodicity of these feature values by Fourier transform and extracts the segments. On the other hand, in cooking motion classification, considering that the location of the cooking motion within video frames is various, the proposed method classifies the repetitive cooking motion segments by using a feature value invariant to the location of the motion within video frames. In an experiment for extracting segments that contain repetitive cooking motions, the method obtained an accuracy of 0.78, and for cooking motion classification, an accuracy of 0.77 was obtained. From these results, the effectiveness of our method was shown.

Keyword Cooking video, motion analysis, CHLAC feature, Fourier analysis

1. はじめに

近年，マルチメディア環境の普及により，社会や家庭において様々な映像技術が日常生活支援に導入されつつある．その 1 つとして，料理という日々の生活に必要不可欠な行為に対する支援が挙げられる．例えば，我々の周りの料理情報源として，これまでは出版物である料理教材本および放送料理番組が主流であった．

しかし，最近では Web 上の料理レシピや動画投稿サイトの料理映像を参考にして料理を行うことが流行となっており，それらの情報量は既存メディアよりも遙かに速い勢いで増大している．たとえば，大手料理レシピサイト COOKPAD[1]は，2010 年 10 月時点で月間ユニークユーザ数 1069 万人を有し，掲載されている料理レシピはユーザによる新たな投稿により日々増加している．しかし，それらレシピをコンテンツという観点から見ると，短いテキストと数枚の画像のみで構成されている場合がほとんどである．一方，任天堂（株）が携帯ゲーム機向けに開発した「しゃべる！DS お料理ナビ」[2]というソフトウェアには，食材や調理過程の詳細な説明や映像による解説など，初心者でも基礎から理解できるように工夫がされている．このソフトウェアの商業的成功から，利便性と楽しみを感じるような料理情報源を求める需要が高まっていると考えられる．

そのような料理情報源の 1 つの形態として，レシピ中の各調理過程に対応する調理映像が付加された料理レシピ，が考えられる．任意の料理レシピの調理過程に対して，この映像付加を実現するには，調理食材と調理動作に関して索引付けされた大規模な調理映像群が必要となる．そのため，この 2 点に基づく調理映像の自動索引付け手法が必要となる．調理食材の索引付け手法として，柴田[1]らは，映像に現われる食材の物体モデルを自動的に学習し，それを用いて素材認識を行う手法を提案した．一方，調理動作に関する従来手法として，浜田ら[2] は重要な調理動作が含まれている可能性が高い，図 1 のような繰り返し動作の映像区間を局所的周波数解析により抽出する手法を提案した．ただし，この手法では，図 2 のような繰り返し動作の振動の中心が時間経過につれて移動する動作に対して適用困難である．一方，カイら[3]は画面全体の大域的な特徴の時間的変化を解析することにより，繰り返し動作区間を抽出する手法を提案した．この手法では，動作速度が遅い場合やズームアウトされている場合のように，動作領域が小さい映像に対し適用困難である．そこで，本報告ではこれらの欠点を持たない手法により繰り返し動作区間抽出を行い，その区間の調理動作を識別することを目的とする．

2. 提案手法

提案手法の処理は繰り返し動作区間の抽出および調理動作の識別の 2 段階に分けられる．まず 2.1 節で本手法において主要な役割を果たす特徴量を概説した後，2.2 節で抽出，2.3 節で識別の手法を述べる．

[1] http://cookpad.com
[2] http://www.nintendo.co.jp/ds/a4vj

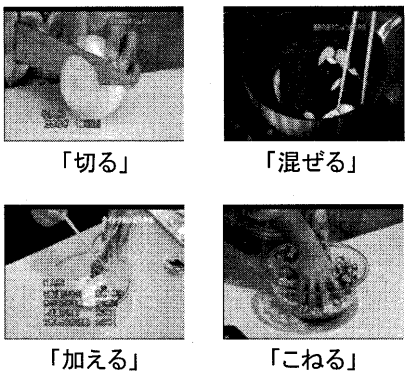

図 1：繰り返し動作の調理例

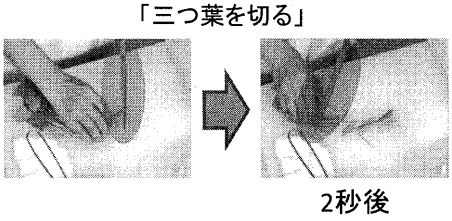

図 2：繰り返し動作の振動中心が移動する例

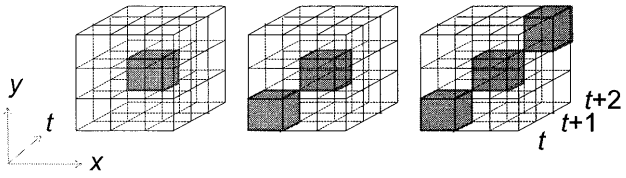

図 3：CHLAC 特徴の 3 次元マスクの例

2.1. CHLAC 特徴

本節では，提案手法の抽出と識別の両段階で用いる CHLAC（Cubic Higher-order Local Auto-Correlation；立体高次局所自己相関）特徴[4]の性質および計算方法について概説する．CHLAC 特徴は画面全体に現われる時空間変動を表現する画像特徴量の 1 つである．具体的には，画面中の動作が起こっている領域の「形状」と「動き」を表現する 251 次元の特徴量である．例えば、映像中で手が右へ移動する動作および左へ移動する動作の映像それぞれから求めた CHLAC 特徴は対称性を持っており，この性質は繰り返し動作を検出する上で必要な性質である．また等速な平行移動では CHLAC 特徴は一定の値を保ち，さらに対象物体の動作の位置に依存しないという極めて重要な性質を持つ．この性質のため，図 2 のような短時間の時間変化の間に繰り返し動作の振動中心が移動する場合でも，その周期性を解析することが可能になると考えられる．

以下では，CHLAC 特徴の計算方法について述べる．

計算のために，251 種類の 3 次元マスクを映像に対して 1 画素単位でずらしながら走査する．図 3 にマスクの 3 つの例を示す．各マスクは，水色の画素の値の積を計算し，映像走査の間，これらの値の累積和をとる．最終的に，各マスクの値が CHLAC 特徴の各次元の値に対応するため，CHLAC 特徴は 251 次元の特徴量となる．続いて，CHLAC 特徴の 2 種類の利用方法を述べる．図 4 はマスクを x, y 方向に動かす場合である．このとき，3 フレームごとに 1 つの CHLAC ベクトルが得られるため，走査完了後には，CHLAC ベクトルの時系列データが得られる．一方，図 5 ではマスクを x, y, t 方向に動かす場合である．このとき，映像 1 本から 1 つの CHLAC ベクトルが得られる．提案手法では，これら 2 通りの特徴ベクトルを用いる．

2.2. 繰り返し動作区間の抽出

本節では，料理映像から繰り返し動作が起こっている映像区間を抽出する手法について述べる．本手法では，短時間の解析窓を映像上で走査することにより，繰り返し動作区間を抽出する．解析窓は，映像中の動作の位置に依存しない特徴と依存する特徴の 2 種類の特徴の時間的周期性を解析し，繰り返し動作区間か否かを判別する．

2.2.1. 解析窓

映像区間を入力とする解析窓の処理の流れを図 6 に示す.始めに，入力映像区間に対してフレーム間差分としきい値処理により 2 値化を行い，2 値画像列を得る．次に，この画像列から CHLAC と 2 値ベクトルの時系列データを取り出す．2 値ベクトルとは，2 値画像に格納されている数値をラスタスキャン方式で取り出し、ベクトルとみなしたものである．そのため，2 値ベクトルは動作の位置に依存するという性質を持つ．これ以降，CHLAC と 2 値ベクトルの両方に同じ処理を施す．まず，PCA により次元削減を行う．図 7 は，繰り返し動作の振動中心が時間経過と共に移動するような調理映像から求めた上記 2 種類の特徴の第 1 主成分の時間変化の例である．続いて，FFT により上位の主成分に対応するパワースペクトルを求める．これらパワースペクトルは，パワーの総和が 1 になるように正規化を行う．これにより，映像中の動作の大きさに依存しない性質を持つ繰り返し動作区間抽出手法を構築する．次に，各スペクトルから文献[2]で示されている周期性の度合いを表す 5 つの特徴量を取り出す．そのためにまず，図 8 に示すように，人の繰り返し動作の速さを考慮した周波数帯 $f_0 \leqq f \leqq f_1$ を設定する．そして，この範囲内のパワーのピークに対応する周波数 f_p を求める．入力映像が繰り返し動作区間である場合，そのスペクトルには以下の 2 つの性質がある．

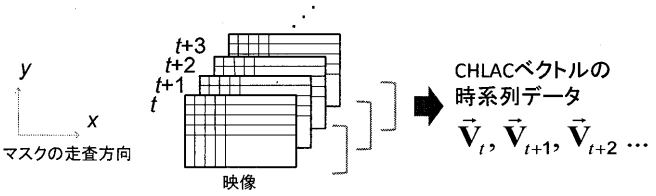

図 4：CHLAC ベクトルの時系列データの計算

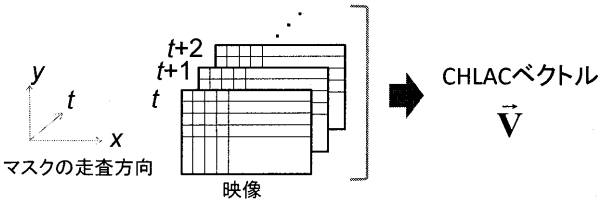

図 5：CHLAC ベクトルの計算

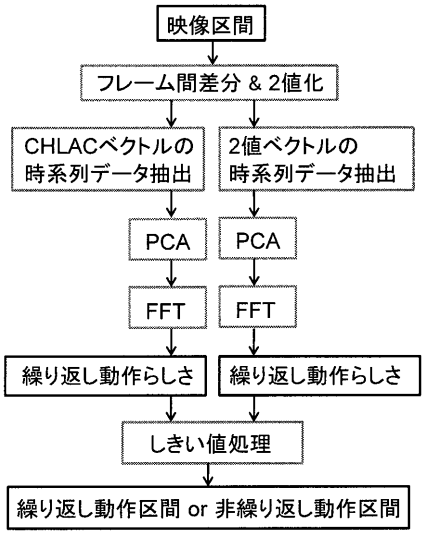

図 6：繰り返し動作区間抽出における解析窓の処理

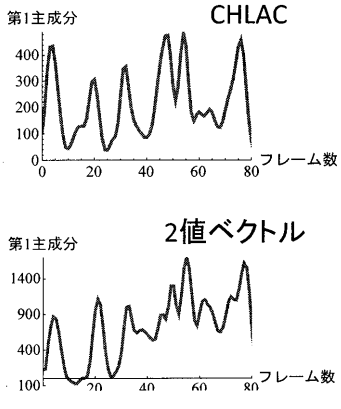

図 7：各特徴の第 1 主成分の時間的変化の例

・P(f_p) の値がスペクトルにおいて突出して大きい
・低周波数（$f \leqq f_0$）のパワーが小さい

このような性質を考慮して，5 つの特徴量として以下を計算する．

・周波数帯 $f_0 \leqq f \leqq f_1$ における P(f)の総和
・P(f_p)の局所的鋭敏さ
・P(f_p)の大域的鋭敏さ
・P(30/80)に対する P(f_p)の比
・P(60/80)に対する P(f_p)の比

そして，各スペクトルの各特徴量に関して正の重み付け和を取り，「繰り返し動作らしさ」と定義する．最後に，CHLAC 特徴と 2 値ベクトル，つまり動作の位置に依存しない特徴と依存する特徴から抽出した繰り返し動作らしさをしきい値処理することにより，入力映像区間が繰り返し動作区間か否かを判別する．

2.2.2. 解析窓の走査

本節では，前節で述べた解析窓を料理映像に対して走査させ，繰り返し動作区間を抽出する手法について述べる．走査の流れを図 9 に示す．映像に対して，窓幅 80 フレームの解析窓を窓ステップ 10 フレーム単位でずらして走査する．そして，複数の解析結果の多数決により，10 フレーム単位で繰り返し動作区間か否かを判別する．ここで，抽出結果として得られた 40 フレーム（約 1.3 秒）以下の繰り返し動作区間および非繰り返し動作区間は，その識別結果を反転させる．一般に，繰り返し動作が行われる最中でも，ごく短時間手を止めたり，または特異な動きを 1 回挟むというような場合は珍しくない．そのため索引付けを行う上で，このような短時間の孤立区間を消去する．

2.3. 調理動作の識別

提案手法は図 10 で示すように学習段階と識別段階に分かれる．学習段階として，映像から 1 つの CHLAC ベクトルを取り出す.そして，フレームの長さに応じて正規化を行う．次に，PCA に基づく固有空間法により次元削減を行い，学習識別器 SVM により学習を行う．識別段階では，同様に正規化された CHLAC 特徴を求め，固有空間に投影することで次元削減を行う．最後に，SVM により識別を行う．

3. 実験

3.1. 繰り返し動作区間の抽出実験

実験対象の料理映像として，NHK 地上波放送の料理番組「きょうの料理」10 本から手元ショットとして 323 個を人手で取り出した．手元ショットとは，図 11 のように手元が拡大されて撮影された映像であり，調理上重要な調理過程であることが多い．2.2 節の提案手法をこれら映像に適用し，実験を行った．

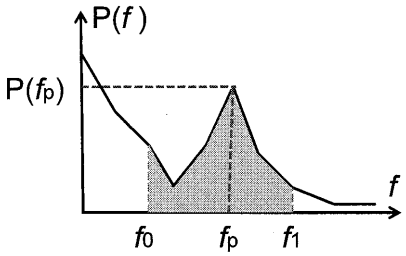

図 8：パワースペクトル（横軸:周波数，縦軸:パワー）

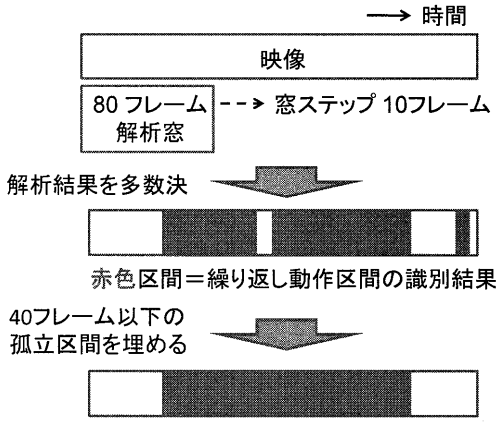

図 9：解析窓の走査

図 10：調理動作の識別の流れ

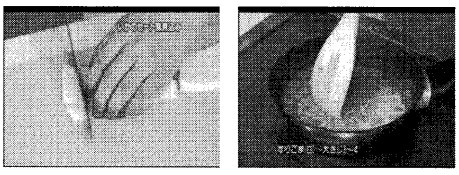

図 11：手元ショットの例

3.1.1. 評価方法および比較手法

提案手法の繰り返し動作区間の抽出精度を評価するにあたり，30 フレームの誤差許容範囲を設定した．具体的には，図 12 のように，識別結果と真値それぞれの開始時間および終了時間の誤差が 30 フレーム以下のとき正検出とする．また，非繰り返し動作区間を繰り返し動作区間として抽出した場合は誤検出、それ以外の場合は検出漏れとする．また比較手法として，CHLAC 特徴および 2 値ベクトルを単独で用いる手法，および従来手法[2][3]を用いて実験を行う．

3.1.2. 結果・考察

提案手法を前述の手元ショットに適用した際の実験結果を，表 1 に示す．提案手法同士および従来手法との比較においても，CHLAC 特徴と 2 値ベクトルの両方を用いた手法が最も良い結果となった．このことから，提案手法の有効性を確認した．また，CHLAC 特徴および 2 値ベクトルを単独で用いた手法ではほぼ差がなかった．しかし，動作の振動中心の位置が移動する距離が長いほど，CHLAC 特徴を単独で用いた手法が良い抽出結果をもたらした．一方，位置がほぼ一定の場合は，逆に 2 値ベクトルを単独で用いた手法の方が僅かに良い結果であった．そのため，繰り返し動作の振動中心の位置が移動するのか，または一定なのかという情報を元に，CHLAC 特徴および 2 値ベクトルの繰り返し動作らしさにそれぞれ重み付けを行うことで，より高い精度で抽出が可能となると考えられる．続いて，提案手法と従来手法の抽出性能の比較として，以下の 5 種類の映像に関して述べる．

・カメラワークが発生する映像
・湯気，水面の沸騰が発生する映像
・動作速度が遅い繰り返し動作の映像
・繰り返し動作の領域が小さい映像
・繰り返し動作の振動中心が移動する映像

提案手法および従来手法の誤検出の一番の原因はカメラワークである．図 13 のようなカメラワークが発生する映像に対して，提案手法でフレーム間差分と 2 値化を行った際求まる 2 値画像列における前景領域は，極めてランダムで偶然的に高い周期性を持つことがある．しかし，「切る」という動作を追尾するようなゆっくりとしたカメラワークではそれほど影響がないことから，大きなカメラワークを検出し，その区間を非繰り返し動作区間と判別するような処理を加えることで，各手法の抽出精度が高まると考えられる．

2 つ目の湯気と沸騰は，各手法で誤検出が多かったが，それでも提案手法が最も良い結果であった．それは，図 14 のような映像に対し，提案手法ではフレーム間差分により，ノイズのような輝度変化を無視することができる一方，従来手法では輝度値を直接解析するため大きな影響を受けるためである．

3 つ目は，動作速度が遅い繰り返し動作の映像である．各手法の性能に共通している点は，動作速度が速い繰り返し動作ほど正検出されやすいという性質である．これは各手法で用いる解析窓のフレーム長の範囲で，繰り返し動作の回数が多いほど周期性が高いと判別されやすいからである．しかし，動作速度に強く依存する大域解析[3]よりは，提案手法の方が良い結果であった．

4 つ目は，繰り返し動作の領域が小さい映像である．

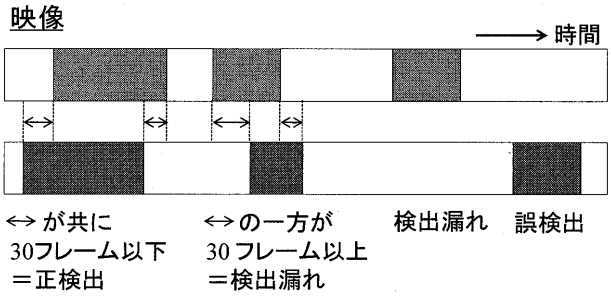

図 12：繰り返し動作区間抽出の評価方法

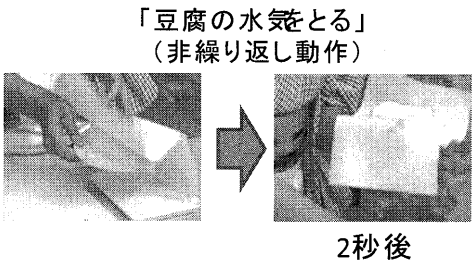

図 13：カメラワークが発生する映像

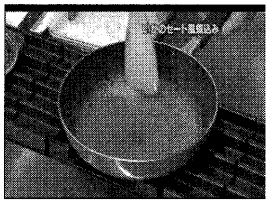

図 14：水面の沸騰が発生する映像

表 1：繰り返し動作区間の抽出の実験結果

		F 値
提案手法	CHLAC 特徴 & 2 値ベクトル	0.78
	CHLAC 特徴	0.68
	2 値ベクトル	0.65
従来手法	局所解析 [2]	0.47
	大域解析 [3]	0.56

従来手法[2][3]は動作が発生している領域の大きさに強く依存するため，検出漏れが多かった．一方，提案手法は動作領域の大きさに依存しないため，特徴量が捉えきれないほど小さい場合を除いて，正検出が可能であった．しかし，極めて稀ではあるが，動作領域が小さな非繰り返し動作を誤検出する場合があったため，あまりに小さな場合は前処理としてその区間を除去することが必要であると考えられる．また，繰り返し動作区間であっても，動作領域が小さな映像は，索引付けされた調理映像群としての価値が低いので，これらも除去しても良いと考えられる．

5 つ目の繰り返し動作の振動中心が移動する映像は，提案手法が圧倒的に良い結果であった．次いで，大域解析[3]が良く，局所解析[2]では検出漏れが非常に多かった．

3.2. 調理動作の識別実験

学習用映像として，映像中で常に繰り返し動作が行われている手元ショットの映像区間 60 個を用いた．一方，識別用映像として，抽出実験で用いた 323 個の手元ショットから 199 個の繰り返し動作区間を人手で抽出した．調理動作の内訳は，切る 25 個，混ぜる 97 個，加える 30 個，こねる 15 個，その他 30 個である．2.3 節の提案手法を，繰り返し動作により構成されるこれらの映像区間に適用した結果を表 2 に示す．全体の識別率は 0.78（154/199）となった．その他の調理動作には，学習用映像に存在しなかった調理動作が含まれていたため，識別率は 0.47（14/30）と低かった．大規模な調理映像群を構築するにあたり，学習用映像として様々な調理映像を用意しなければならない．しかし，学習のために必要な調理映像の本数は，CHLAC 特徴の位置不変性のために比較的少なくて済むというメリットがあると考えられる．さらに，リアルタイムな調理動作識別として利用できる程度の高速性を持っている．

表 2：調理動作の識別の実験結果

識別結果＼真値	切る	混ぜる	加える	こねる	その他	合計
切る	**22**	1	1	0	3	27
混ぜる	0	**85**	1	2	2	90
加える	1	4	**23**	1	8	37
こねる	0	2	1	**10**	3	16
その他	4	5	4	2	**14**	29
合計	27	97	30	15	30	**199**

4. むすび

本報告では，料理映像から繰り返し動作区間を抽出し，それらの区間の調理動作を識別する手法を提案した．提案手法は，位置に依存しない特徴（CHLAC 特徴）及び依存する特徴（2 値ベクトル）の周期性を解析することにより，多様な繰り返し動作の調理映像に適用可能になるものであった．また，調理動作の識別では，比較的少量の学習用映像から CHLAC 特徴を抽出および学習し，低次元な特徴であるため高速に識別可能であった．繰り返し動作区間抽出実験では F 値 0.78 が，調理動作識別実験では識別率 0.77 が得られ，提案手法の有効性を確認した.今後は，繰り返し動作の動きの性質に応じて 2 種類の繰り返し動作らしさに重み付けを行い，さらに抽出精度を向上させる手法の検討を行う．

謝辞　本研究の一部は文部科学省科学研究費補助金による

文　献

[1] 柴田知秀，加藤紀雄，黒橋禎夫，“言語情報と映像情報の統合による物体のモデル学習と認識，” 情報処理学会論文誌，Vol.49, No.3, pp.1451--1464, Mar. 2008.

[2] R. Hamada, S. Satoh, S. Sakai, and H. Tanaka, “Detection of Important Segments in Cooking Videos,” Proc. IEEE Workshops on CBAIVL, pp.118--123, Dec. 2001.

[3] カイ承穎，高橋友和，井手一郎，村瀬洋，“画像特徴の時間変化に基づく料理映像の分類，” 電子情報通信学会 2009 年総合大会，A-16-2, Mar. 2009.

[4] N. Otsu, “Towards Flexible and Intelligent Vision Systems -- From Thresholding to CHLAC --,” Proc. 9th IAPR Conf. on Machine Vision Application, pp.430--439, May. 2005.