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

信学技報
TECHNICAL REPORT OF IEICE.

# 時間的冗長性の除去による調理履歴映像の要約

林　泰宏† 道満　恵介† 出口　大輔†† 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
†† 名古屋大学情報連携統括本部　〒 464–8601 愛知県名古屋市千種区不老町
E-mail: †{hayashiy,kdoman}@murase.m.is.nagoya-u.ac.jp, ††ddeguchi@nagoya-u.jp,
†††{ide,murase}@is.nagoya-u.ac.jp

**あらまし**　個人が調理過程を撮影した映像の要約手法について報告する．近年，個人の生活を記録するライフログが注目されている．しかし，ライフログは長時間に渡り記録されるものであり，データ量が膨大である本報告では，日常的な創作活動である料理に着目し，個人が調理過程を撮影した調理履歴映像の要約手法を提案する．調理履歴映像は，ログとしての利用以外に，インターネット上で公開するなどして，他人が調理する際の参考映像として提供することもできる．そのため，映像を要約することでより効率的に検索・閲覧できると考えられる．提案手法では，映像から時間的冗長な区間として静止区間と繰り返し区間を検出し，それらを除去することで調理履歴映像を要約する．実際に調理過程を撮影した映像を用いて区間検出実験を行い，静止区間は適合率 0.98，再現率 0.99，繰り返し区間は適合率 0.62，再現率 0.92 の精度が得られた．

**キーワード**　ライフログ, 映像要約, 調理履歴映像, 調理動作

# Cook-Log Video Summarization by Removing Temporal Redundancy

Yasuhiro HAYASHI†, Keisuke DOMAN†, Daisuke DEGUCHI††,

Ichiro IDE†, and Hiroshi MURASE†

† Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464–8601 Japan
†† Information and Communications Headquarters, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464–8601 Japan
E-mail: †{hayashiy,kdoman}@murase.m.is.nagoya-u.ac.jp, ††ddeguchi@nagoya-u.jp,
†††{ide,murase}@is.nagoya-u.ac.jp

**Abstract**　We report on a method for summarizing a video which recorded the process of cooking by an individual. In recent years, life-log which records the daily life of an individual has been attracting attention. However, since life-log is recorded over a long time, the amount of data is huge. In this report, we focus on cooking which is a creative activity in daily life, and we propose a method for summarizing a cook-log video which recorded the process of cooking by an individual. A cook-log video can not only be used as a kind of life-log, but it can also be able to be provided on the Internet as a reference for other people to cook. Thus, summarizing a cook-log video enables more efficient search and browsing. The proposed method detects the state sections and the repetitious sections as temporal redundant sections, and summarizes the cook-log video by removing them. We conducted a section detection experiment using an actual cook-log. A precision of 0.98 and a recall of 0.99 were obtained for detecting the state section, and a precision of 0.62 and a recall of 0.92 were obtained for detecting the repetitious section.

**Key words**　life-log, video summarization, cook-log video, cooking operation

## 1. はじめに

近年，個人の生活を記録するライフログが注目されている．日々の活動を画像や映像などで記録しておくことで，過去の行動を確認，分析して個人の生活スタイルに合ったサービスを提供できる．ただし，ライフログは長時間に渡り記録されるものであり，データ量が膨大となる．そのため，ライフログデータを効率的に検索・閲覧する技術が必要とされている．

畑田らは，ライフログデータの要約手法 [1] について提案している．ライフログデータは閲覧された回数が多いものほど重要なデータであると仮定し，データの閲覧された回数に基づいて要約している．一方，堀らは，ライフログデータの取得・検索・閲覧のためのライフログエージェントを提案している [2]．ライフログのために撮影された映像に，モーションセンサや GPS データなどの様々なデータから検索キーを設定することで映像を効率的に検索・閲覧することを可能にしている．このように，ライフログデータを検索・閲覧するための研究が行われている．

ライフログの効率的な検索・閲覧は，特に料理において必要とされる．料理は，豊富な知識や経験を必要とするものであり，熟練者の調理を記録した映像は，本人がログとして利用するだけでなく，映像を公開することで他人が調理の参考映像として利用することもできる．そこで，本研究ではライフログの中でも料理に関する映像の効率的な検索・閲覧に着目する．料理映像を効率的に検索・閲覧するための研究として，三浦らは，料理番組の要約手法 [3] を提案している．料理番組は教材映像であり，調理に関する教材として利用されている．しかし，料理番組には雑談などの冗長な部分も含まれている．したがって，閲覧にはある程度の時間が必要になる．また，大量に録り溜められている映像から必要とする映像を検索することにも時間を要する．そこで，この研究では料理番組映像の特徴を利用した映像要約を提案している．しかし，ライフログのような個人ユーザが調理を行う様子を撮影した映像を考えると，撮影条件の違いや編集の有無の違いから，調理履歴映像に料理番組の要約手法を適用するのは難しい．また，これまでのところ，個人が撮影した調理映像を対象とした要約手法は研究されていない．

そこで，本報告では個人が調理する様子を撮影した映像（以下，調理履歴映像）の要約手法を提案する．調理履歴映像の要約において，静止区間および繰り返し区間に注目する．料理番組の要約において，これらの区間は重要とされているが，編集が行われていない調理履歴映像では，その大部分は冗長な区間であり，すべてを要約映像に含める必要はないと考えられる．そこで，提案手法ではこれらを時間的に冗長な区間として検出する．静止区間は隣接フレームを比較することにより検出し，繰り返し区間は特徴量の類似性に着目して検出する．このとき，繰り返し区間検出で用いる特徴量として，調理動作解析 [4] などで用いられている CHLAC 特徴 [5] を利用する．最後に，これらの検出された区間を除去することで調理履歴映像を要約する．

以降，第 2 節で関連研究として，提案手法で利用する CHLAC 特徴について説明する．第 3 節では提案手法である冗長な区間の検出手法とその結果を利用した映像要約手法について述べる．

図 1　繰り返し動作の例

そして，第 4 節で区間検出手法の評価実験について述べ，その結果について考察する．最後に第 5 節で本報告をまとめる．

## 2. 関連研究

提案手法では，冗長な区間として繰り返し区間を検出する．その際，特徴量の類似性に着目するが，特徴量として動きの性質を背景や動作位置などに影響されずに表現できるものが望ましい．そこで本研究では，環境変化に頑健な特徴量である CHLAC 特徴 [5] を利用する．以降では，この CHLAC 特徴について詳しく説明する．

提案手法では，動作の繰り返しを検出するが，図 1 のように，同じ動作を行っていてもその見え方は変化する．そこで本研究では，動き特徴として見えの変化に頑健な CHLAC（Cubic Higher-order Local Auto Correlation）特徴を利用する．CHLAC 特徴は，1 枚の画像から特徴を抽出する HLAC（Higher-order Local Auto Correlation）特徴 [6] を時間軸方向を含めた 3 次元に拡張した特徴である．

CHLAC 特徴は，画像中での局所パターンの出現頻度に基づく特徴であり，画像中の位置に依存しない特徴量が得られる．また，差分画像から特徴抽出をすることにより，背景の影響を抑制することができる．このため，差分画像に対する CHLAC 特徴を用いることで，背景や動作位置などの違いに影響を受けず，動きのみに着目した特徴量を抽出することができる．

以下，CHLAC 特徴の基になる HLAC 特徴についてまず説明し，その後 CHLAC 特徴について説明する．

### 2.1 HLAC 特徴

画像データ $f$ の $N$ 次自己相関は以下のように定義される．

$$\int f(\boldsymbol{x})f(\boldsymbol{x}+\boldsymbol{\delta}_1)\cdots f(\boldsymbol{x}+\boldsymbol{\delta}_N)d\boldsymbol{x}, \tag{1}$$

ここで，$\boldsymbol{x}$ は画像中のある画素位置を表し，$\boldsymbol{\delta}_1,\cdots,\boldsymbol{\delta}_N$ は $\boldsymbol{x}$ からの変位を表す．HLAC 特徴は，注目画素とその近傍の $3\times3$ 画素の局所領域での相関を求めることで特徴量を抽出する．特徴量の算出は，局所領域の平行移動による冗長性を省いたすべての変位の組合せで行う．つまり，変位パターンは，$N=0$ の場合は 1 通り，$N=1$ の場合は 4 通り，$N=2$ の場合は，20 通り存在する．HLAC 特徴では $N=2$ までの相関を求めるため，相関の全パターンは 25 通りのパターンとなる．よって，画像から求まる HLAC 特徴の次元数は 25 次元となる．

### 2.2 CHLAC 特徴

CHLAC 特徴は，HLAC 特徴を時間軸方向に拡張したものであり，時間的に連続した複数の画像から抽出される．このとき，$3\times3\times3$ 画素の局所領域での相関を求めることで特徴量

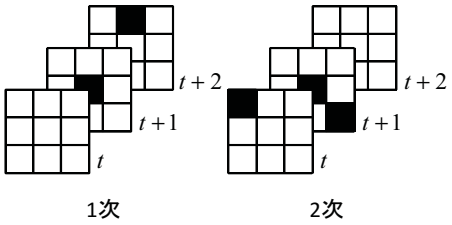


図 2　CHLAC 特徴の変位パターンの例

(a) 静止区間　　(b) 繰り返し区間

図 3　時間的に冗長な区間

を抽出する．そのため，変位パターンは図 2 のようになる．式 (1) により，CHLAC 特徴で $N = 2$ までの相関を求める場合，局所領域の平行移動による冗長性を省いたすべての変位の組合せは全部で 251 通りとなる．よって，CHLAC 特徴の次元数は 251 次元となる．提案手法では動き特徴として利用するため，フレーム間差分画像からこの CHLAC 特徴を抽出する．

## 3. 提 案 手 法

### 3.1 手 法 概 要

提案手法では，調理履歴映像における時間的冗長性に着目する．調理履歴映像における時間的冗長な区間として，以下の 2 種類が考えられる．

（1） 静止区間

焼く・煮るといった動作を行わない区間（図 3(a)）．

（2） 繰り返し区間

切る・混ぜるといった同じ動作を繰り返し行う区間（図 3(b)）．

提案手法では，これらの時間的冗長な区間を検出し，それらの冗長な区間を除去することで調理履歴映像を要約する．

以降の節では，まず本報告で対象とする調理履歴映像について説明し，その後，上記の区間検出方法とその結果を利用した要約手法について説明する．

### 3.2 調理履歴映像

調理履歴映像は,「調理台」や「コンロ」といった調理場所ごとにカメラを設置し，各調理場所の様子を撮影した映像を用いる．このとき，各カメラの撮影領域に重複はないものとする．これらのカメラで撮影された各映像から，各時刻において調理が行われている場所を検出し，それらを切り替えてつなぎ合わせることで 1 本の調理履歴映像を生成する．ただし，調理は同時に 2 か所以上の調理場所では行われないものとする．

### 3.3 区 間 検 出

#### 3.3.1 静止区間検出

静止区間は，フレーム間差分を利用し，画像的変化の小さいフレームを静止区間とみなすことで検出する．まず，映像中の

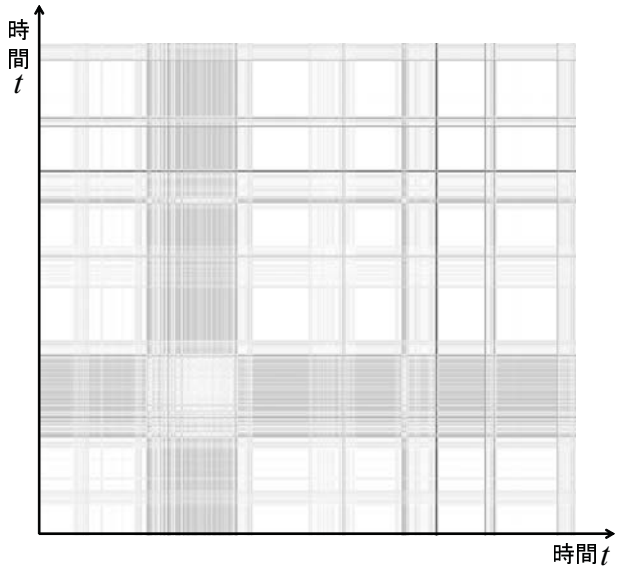


図 4　区間類似度の可視化画像．各画素は対応する横軸上の時刻での単位区間と縦軸上の時刻での単位区間の類似度を示す．黒い画素ほど低類似度であり白い画素ほど高類似度であることを意味する．

隣接するフレームの差分画像を生成し，それを 2 値化することで動きのあった画素と動きのなかった画素を求める．この画像の中で動きがあった画素の数が一定数以下のフレームを静止区間として検出する．これを映像のすべてのフレームに対して適用して，静止区間を検出し，区間長が一定フレーム以下の小区間を除去したものを最終的な静止区間とする．

#### 3.3.2 繰り返し区間検出

繰り返し区間は，類似した特徴を持つ区間を探索することで検出する．このとき，映像を 10 フレームごとに分割し，これを最小単位区間として繰り返し区間を検出する．まず，この単位区間ごとに CHLAC 特徴を算出する．隣接するフレームから差分画像を生成し，それを 2 値化する．その 2 値化した差分画像すべてを走査し，CHLAC 特徴の局所パターンを数え上げることで，単位区間から 1 つの CHLAC 特徴を抽出する．次に，すべての単位区間同士で特徴量間の類似度を計算する．これを可視化したものを図 4 に示す．このとき，特徴量が類似する連続した区間があれば，その区間内のどの単位区間同士の類似度も高くなるため，図 4 中に矩形として現れる．つまり，この類似度の画像から，対角線上に存在する類似度の高い矩形を見つけることで，繰り返し区間を検出できる．矩形検出は対角線上の矩形の内，領域内の平均類似度がしきい値 $\theta_1$ 以上，かつすべての類似度がしきい値 $\theta_2$ 以上となる矩形を検出することで行う．ただし，静止区間もこのような矩形を生じるが，繰り返し区間には含めない．これにより，最終的に検出された矩形領域に対応する映像区間を繰り返し区間とする．

### 3.4 映 像 要 約

前節で検出した静止区間・繰り返し区間の情報を利用し，調理履歴映像を要約する．

要約映像に用いる区間を図 5 に示す．静止区間は，時間的変化が重要である考えられるため，区間開始，区間終了，区間中

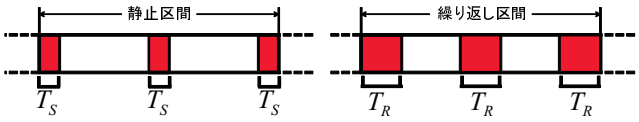


図 5　映像要約に用いる区間

表 1 「ハンバーグ」の調理手順(注1)

| 順番 | 手順 |
|---|---|
| 1 | 長ネギと椎茸をみじん切りにする. |
| 2 | フライパンを中火で温め，バターを溶かす. |
| 3 | 切った材料をフライパンに入れ，絡める．塩を入れ，水気を出す. |
| 4 | とろみが出てきたら，皿に移し，あら熱をとる. |
| 5 | 10 分ほどしたら，ボウルに挽肉・卵・炒めた材料を入れ，フォークでよく混ぜる．胡椒・醤油で味を調える. |
| 6 | 5〜10 分ほどなじませたら，フォークで半分に切り，パテを 2 つ作り，熱したフライパンで焼く. |
| 7 | パテの脇から肉汁が出てきたら返して裏面を焼く. |

間周辺の $T_S$ 秒をそれぞれ要約映像に用い，それ以外のフレームは省略する．繰り返し区間も静止区間と同様に，区間の開始，終了，中間の $T_R$ 秒をそれぞれ要約映像に用いるが，調理動作そのものも重要なので，$T_R > T_S$ とし，用いるフレームを静止区間の場合より長くする．よって，各区間から要約映像に用いられる映像の長さは，静止区間が $T_S \times 3$，繰り返し区間が $T_R \times 3$ となる.

これらの規則に基づいて要約することで，時間的に冗長な区間を省略し，短時間で調理動作の把握が可能な調理履歴映像を生成する.

## 4.　実験と考察

提案手法での要約において，静止区間と繰り返し区間を高精度に検出することが重要である．そこで，実験では実際に調理を行う様子を固定カメラで撮影した映像に本手法を適用し，静止区間・繰り返し区間の検出精度を評価した.

### 4.1　実 験 条 件

本実験では，表 1 に示すような手順でハンバーグの調理を行った調理履歴映像を使用した．実験での撮影環境を図 6 に示す．実験では,「調理台」と「コンロ」の 2 か所をそれぞれ固定カメラで撮影し，それらの映像から生成した調理履歴映像を使用した．実験で使用した調理履歴映像は，解像度が $1,920 \times 1,080$ pixels，フレームレートは 15 fps，映像長は 22 分 18 秒（20,085 フレーム）であった.

区間検出の精度評価に際して，各区間を人手で設定したものを真値区間とした．真値の静止区間は区間数が 5，総フレーム数が 7,496 フレーム，繰り返し区間は区間数が 9，総フレーム数が 6,845 フレームであった．精度評価は，以下の式により適合率・再現率・F 値を求めることで行った.

(注1)：http://cookpad.com/recipe/1452708


図 6　撮 影 環 境

表 2　区間検出結果の精度

| | 適合率 | 再現率 | F 値 |
|---|---|---|---|
| 静止区間 | 0.98 | 0.99 | 0.99 |
| 繰り返し区間 | 0.62 | 0.92 | 0.74 |

$$\text{適合率} = \frac{\text{検出に成功したフレーム数}}{\text{検出したフレーム数}} \tag{2}$$

$$\text{再現率} = \frac{\text{検出に成功したフレーム数}}{\text{真値区間の総フレーム数}} \tag{3}$$

繰り返し区間検出で用いるしきい値は，全類似度で最大のものを 1，最小のものを 0 と正規化したとき，$\theta_1 = 0.95$，$\theta_2 = 0.50$ とした．また，映像要約での各区間の長さに関するパラメータは，$T_S = 1$ 秒，$T_R = 4$ 秒とした.

### 4.2　実 験 結 果

検出結果の精度を表 2 に示す．静止区間については，真値として設定した区間を誤検出・未検出がなく検出できた．一方，繰り返し区間については適合率が 0.62 であり，38%の区間が誤検出された.

また，区間検出の結果を基に映像の要約した結果，映像長が 4 分 55 秒となり，約 4 分の 1 に短縮することができた．要約された映像の系列を図 7 に示す．要約映像では，冗長な調理動作を省略できていることが確認できた．真値区間を基に映像を要約した場合，映像長は 8 分 23 秒であった．実験結果では 4 分 55 秒であり，実験結果の映像は極端に短い映像となった.

### 4.3　考　　察

#### 4.3.1　区間検出の精度

各区間の検出精度を見ると，静止区間については真値の区間を過不足なく検出できており，有効性が確認できた．しかし，繰り返し区間については，真値より広い範囲を繰り返し区間として検出することや，繰り返し区間でない区間も繰り返し区間と誤検出する傾向があった.

この原因として，まず CHLAC 特徴が動きのある領域の大きさに依存していることが考えられる．提案手法では，CHLAC 特徴は動きのある領域を基に特徴抽出を行っており，動きの小さい区間では，異なる動作でも CHLAC 特徴に差が表れにくくなってしまう．そのため，あらかじめ動作領域を切り出すなど

表 3　繰り返し区間検出結果

| しきい値 | 適合率 | 再現率 | F 値 |
|---|---|---|---|
| $\theta_1 = 0.95, \theta_2 = 0.50$ | 0.62 | 0.92 | 0.74 |
| $\theta_1 = 0.96, \theta_2 = 0.50$ | 0.60 | 0.83 | 0.70 |
| $\theta_1 = 0.97, \theta_2 = 0.50$ | 0.59 | 0.64 | 0.62 |

により，動きの変化を抽出しやすくすることで改善できると考えられる．

また別の原因として，類似の条件が緩くなってしまっていたことが考えられる．繰り返し区間検出でしきい値を変化させたときの適合率・再現率・F 値を表 3 に示す．しきい値 $\theta_1$ を上げて，類似の条件を厳しくすると，適合率はあまり変化がないが，再現率が著しく低下する．これは，繰り返し区間よりも他に検出されやすい非繰り返し区間が存在することを意味している．つまり，現在の手法では，繰り返し区間を過剰に検出してしまう傾向があるといえる．そのため，今後は，大まかな動作位置や色などの他の情報を利用することで精度の向上を図る．

**4.3.2**　映像要約の精度

実験で要約した映像を見ると，冗長でない調理手順も省略されており，不自然な要約映像となってしまった．真値区間を基に生成した要約映像が 8 分 23 秒であるのに対し，実験で生成した要約映像が 4 分 55 秒と真値映像より極端に映像時間が短いことからも，適切でない省略があることが確認できる．省略された手順として，材料を加えて混ぜ合わせる手順（表 1 手順 5）や混ぜた材料からパテを作り焼く手順（表 1 手順 6）があった．前者では，卵を加える動作が要約に含まれておらず，調理手順を把握するための映像としては致命的といえる．また，後者では，パテを作る動作から焼く動作までをすべて同じ繰り返し区間としてしまっていた．しかし，提案手法では検出区間の一部を要約に含めるため，要約映像にパテを作る動作と焼く動作を部分的に含めることができていた．そのため，要約映像では，部分的に手順を含んでいれば調理手順の理解ができるため，要約において調理手順の欠落がないことが重要である．

区間検出結果からも分かるように，適切でない省略の原因は，繰り返し区間の過剰な検出が原因である．また提案手法では，検出された区間の開始・終了付近のフレームを要約映像に用いており，検出する区間の開始位置・終了位置がずれると生成される要約映像にも影響が出る．そのため，要約のためには区間の開始位置・終了位置を正確に検出することが特に重要といえる．今後は，区間検出の精度向上とともに，各調理動作の境界を検出することにも取り組んでいく．

一方，調理履歴映像は固定カメラで撮影するため，空間的にも冗長な映像となっており，要約しただけでは調理内容の把握しやすい映像は生成できない．そこで，調理履歴映像から必要な領域を切り出し，仮想的にカメラワークを生成することでより分かりやすい映像が生成できると考えられる．また，前述の特徴抽出においても，あらかじめ動作領域を切り出しておくことで，特徴抽出を改善できると考えられる．

## 5. むすび

本報告では，調理履歴映像の要約手法を提案した．提案手法では，調理映像の時間的冗長性に着目し，それらを除去することで映像を要約した．調理映像における時間的冗長な区間として静止区間，繰り返し区間を検出し，その精度を評価した．評価実験の結果，F 値が静止区間では 0.99，繰り返し区間では 0.74 であり，繰り返し区間検出の精度向上が必要であるといえる．

今後は，繰り返し区間の精度向上とともに，以下の点についても検討していく．

- 料理レシピの利用：料理レシピを利用し，料理レシピの各調理手順と映像を対応づけることで，調理手順も考慮したうえで冗長な区間を検出できる．例えば，複数の材料を切ることが手順から分かれば，すべての材料を切る様子を要約に含める必要はなく，一部の材料を切る様子は省略可能と考えられる．また，他の料理レシピで使われていない食材・調理方法ほど，その映像で重要な情報といえる．そのため，料理レシピの利用とともに，料理レシピに表れる食材や調理方法の統計情報を利用し，珍しい食材・調理方法を優先することで，より短時間で重要な情報を得られる映像が生成できる．
- 空間的冗長性の除去：調理映像は固定カメラで撮影されるため，不必要な領域まで撮影されることが多い．そのため，調理動作や動作領域に基づいて必要な領域を切り出すことで，仮想的にカメラワークを生成し，空間的冗長性を除去することで，より調理内容の把握しやすい映像が生成できる．また，動作領域をあらかじめ求めておくことは特徴抽出においても有効に働く．この調理履歴映像からのカメラワークの生成に関する研究は既に行われており [7]，今後は，この研究の手法を統合することでより分かりやすい調理履歴映像の要約を行っていく予定である．

## 文　献


[1] 畑田 晃希, 山崎 俊彦, 相澤 清晴, “ユーザの閲覧履歴を利用したライフログデータの要約,” 映像情報メディア学会誌, vol.64 no.2, pp.237–240, Feb. 2010.

[2] 堀 鉄郎, 相澤 清晴, “ライフログビデオのためのコンテキスト推定,” 信学技報, CS2003-152, Dec. 2003.

[3] 三浦 宏一, 浜田 玲子, 井手 一郎, 坂井 修一, 田中 英彦, “動きに基づく料理映像の自動要約,” 情処学 CVIM 研究会論, vol.44 no.SIG9, pp.21–29, Jul. 2003.

[4] 久原 卓, 出口 大輔, 高橋 友和, 井手 一郎, 村瀬 洋, “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,” 信学技報, MVE2010-144, Mar. 2011.

[5] T. Kobayashi and N. Otsu, “Action and Simultaneous Multiple-Person Identification Using Cubic Higher-Order Local Auto-Correlation,” Proc. 17th International Conference on Pattern Recognition, pp.741–744, Aug. 2004.

[6] N. Otsu and T. Kurita, “A New Scheme for Practical Flexible and Intelligent Vision System,” Proc. IAPR Workshop on Computer Vision, pp. 431–435, Oct. 1988.

[7] 兵庫 渉, 林 泰宏, 野田 雅文, 出口 大輔, 井手 一郎, 村瀬 洋, “調理手順に従った撮影対象領域の決定に基づく調理映像を対象としたディジタルカメラワーキング,” 信学技報, MVE2011-100, Mar. 2012.


図 7　要約結果．数字が○で囲まれたものは繰り返し区間として検出された区間，□で囲まれたものは静止区間として検出された区間を示す．