

ソーシャルメディア上での動画コンテンツ提示を目的とした料理のハイライト映像自動生成

著者	北清 拓実
出版者	法政大学大学院情報科学研究科
雑誌名	法政大学大学院紀要. 情報科学研究科編
巻	13
ページ	1-6
発行年	2017-03-31
URL	http://doi.org/10.15002/00021522

ソーシャルメディア上での動画コンテンツ提示を目的とした 料理のハイライト映像自動生成 Automatic Creation of Cooking Highlight Videos for Contents in Social Media

北清 拓実*

Takumi Kitase

法政大学大学院 情報科学研究科 情報科学専攻

Email: takumi.kitase.4j@stu.hosei.ac.jp

Abstract—We present a method for creating a highlight video of cooking as social media. If we can easily prepare cooking videos, it is expected to be used for communication. However, it is difficult to extract entertaining scenes to make highlight video of your cooking because cooking videos also contain so many scenes which are not contain factors which attract viewers. We use two videos which captured different angles as input. At first, our system extract entertaining scenes using image similarity and two evaluate functions defined by us: frame score, which is defined as the standard distribution of the optical flow in each frame divided by the mean of the optical flow, and scene score which defined by several video frames. We focus on the difference of scene scores and image similarity between two continuous scenes to extract scenes for highlight. Next, our system applies color correction by applying gamma correction. Finally, our system combines two extracted scenes of each input video. Experimental results show our method can create cooking highlight videos which can provide positive impressions compared to previous method.

1. はじめに

料理は食事の為だけではなく他者とのコミュニケーションを行う為にも利用される。代表的な例としてソーシャルメディアに投稿する活動があげられる。投稿を行う際は、自身が調理した料理の画像の投稿、調理し完成までの様子を編集した動画の投稿などが一般的である。このような活動の主要な目的は、ソーシャルメディアにおいて自身の活動を投稿することで多くの人にコメントなどの反応を得ることである。

ソーシャルメディアにおいて画像や動画として投稿することによる大きな利点は、視覚的な情報を付随することが可能という点である。文章のみの投稿と比較して閲覧者により強く印象付けられることが期待できる。個人だけではなく企業なども、BuzzFeed社のTasty[1]のように、料理が作られていく様子や完成した料理などを短い動画コンテンツとして提示している。このように、料理がエンターテイメントとしての役割を担っている事例はソーシャルメディアにおいて顕著である。

著者らは画像特徴量を用いて、ソーシャルメディアへの利用を想定した料理のハイライト映像の自動生成の手法を提案した[2][3]。画像による投稿の場合と比較して、動画で投稿する際には多大な手間がかかってしまう。動画をソーシャルメディアに投稿する場合、撮影した映像を簡潔かつ魅力的なものに編集をする必要があ

る。ソーシャルメディア上では短時間に莫大な量の情報が発信されている為、短い時間で閲覧を終了できるメディアコンテンツを好む傾向が強い。撮影した映像を魅力的なものに編集するとは、色調の補正や早送りなど、シーンの抜粋以外にも編集の行程として行う要素が存在する事を表す。

しかし、これらの行程を実行することはユーザにとって大きな負担となる。料理の場合、完成までの間に多数の調理行程が存在する為、料理活動を撮影した動画は長時間の映像になりやすく、そのような映像から使用する場面を選ぶことは困難である。また、色調補正などをかける場合でも、補正をかけ忘れてしまう等のミスが発生する可能性もある。ハイライト映像制作の作業量の問題から、動画によるソーシャルメディアへの投稿は画像と比較して盛んではない。

本研究では、著者らの提案した手法[2][3]をもとに、ソーシャルメディアコンテンツとしての料理のハイライト映像の自動生成を行う手法を提案する。本研究の手法によって生成されるハイライトは、連続した映像フレームによって構成された映像中の一場面(以下、シーンセグメント)を複数組み合わせる構成されている。本手法におけるハイライト生成には次の処理手順が存在する。まず、調理場を真上から撮影した映像に対してハイライトに使用するシーンセグメントを選出する。シーンセグメントの選出方法は、色ヒストグラムを用いた連続するシーンセグメント間の類似度推定と、オプティカルフロー推定を使用したシーンセグメントの評価スコアを利用して選出を行う。具体的には、色ヒストグラムを用いた類似度が小さい、または評価スコアの差が一定値以上の連続したシーンセグメントを選出していく。その後、調理場を横方向からやや俯瞰で撮影した映像も同様の処理を行う。調理場を真上から撮影した映像からは主に調理時の動作を含むシーンセグメントの選出、調理場を横方向からやや俯瞰で撮影した映像からは主に盛り付け、完成の場面を含むシーンセグメントの選出を行う。それぞれの映像から選出されたシーンセグメントを組み合わせるハイライト映像を生成する。生成された映像に対してアンケート評価を行い、今回の手法で生成されるハイライト映像がどのような印象を持たれる傾向のあるコンテンツであるかを考察する。

2. 関連研究

2.1. ソーシャルメディアに関連する調査

総務省が定義した「ソーシャルメディア」と呼ばれるサービス[4]の平成28年における利用率は71.2%であり、平成24年度の利用率から約30%増加している。ま

* Supervisor: Prof. Takafumi Koike

た、Leonardiらはソーシャルメディアがユーザにどう認識されているのか調査を行った[5]。同研究ではソーシャルメディアを主要なコミュニケーションツールとして利用しているかという質問を行っており、62.5%が“Yes”と回答した。以上のように、ソーシャルメディアは多くのユーザが利用しており、ユーザの多くがコミュニケーションツールとして利用していることがうかがえる。

Pittmanらは、InstagramやSnapchat等、画像を投稿し共有するソーシャルメディアを利用しているユーザは、テキストベースのソーシャルメディアであるTwitter等を利用しているユーザと比較して、孤独感を感じにくく、生活の充実感をより得られている傾向があるという事を示唆した[6]。その理由として、視覚的な情報が閲覧者にコンテンツが示す状況を想起させやすいということが考察されている。この結果から、ユーザがソーシャルメディア上で動画を閲覧した際、ユーザに静止画よりもさらに強くコンテンツの状況を想起させられると考えられる。また、閲覧者や投稿者のコミュニケーションがポジティブなものになり積極的なコミュニケーションが発生する事が期待できる。

2.2. 料理をテーマとした映像要約技術

これまでに多くの映像要約の手法が提案されており、いくつかの手法は料理映像を取り扱っている。Hayashiらは料理レシピを利用した料理映像の自動要約システムを提案している[7]。料理レシピの参照により、調理行程が行われていないなどの不必要な場面を除外している。Domanらは料理レシピ中に記載された調理動作と対応する動画を結び付け、調理動作の理解を支援するシステムを提案した[8]。これらの手法では、料理レシピに記載された調理行程を視覚的な情報として必要十分に動画で提示する事を目的としている。

一方、ソーシャルメディアにおけるコンテンツの役割は、閲覧者により印象的な場面を提示することが重要である。Twitter社が提供するTwitterサービス[9]を利用するアプリでは、撮影した画像を投稿する際にカラー補正をかけられる機能を提供している。このように、ソーシャルメディアのコンテンツ提示を目的とする場合、印象的な場面を選出したり、色調などに補正をかけることが重要になる。

以上の理由から、料理映像の自動要約を行う手法[7][8]では、料理の様子をハイライト映像としてソーシャルメディアに投稿する際の支援システムとして利用するのは困難である。

3. 提案手法

図1に提案手法の処理の流れを示す。入力となる映像は異なる場所から撮影した2種類の映像を使用する。この提案手法内では、シーンセグメントを2倍速で1秒ぶんの連続した映像フレームと定義する。色ヒストグラムによって計算された画像類似度と、オプティカルフローを利用した評価関数によるスコアの、2種類の画像特徴量を利用して、ハイライトに使用する映像を選出していく。画像類似度またはオプティカルフローを利用した評価関数に設定された条件を満たすシーンセグメントを、ハイライト映像に使用するものとして選出する。その後、選出したシーンセグメントこの処理を2種類の映像それぞれで行い、選出が終わった後、一方の映像から選出されたシーンセグメントをもう一方の前後に

それぞれ挿入する。この時に挿入される映像は同一の映像である。

3.1. 使用する映像

本手法で使用する映像は2つある。1つは調理場をほぼ真上から撮影した映像(以下、映像A)と、映像Aのカメラに対して右横から俯瞰で撮影した映像(以下、映像B)の2種類の映像(図2)を使用する。映像Aでシーンセグメントの選出に使用する範囲は、最後の2分間以外のすべて、映像Bでシーンセグメントの選出に使用する範囲は最後の1分間である。映像Aでは調理時の行程の提示、映像Bでは完成した料理の提示を行うシーンセグメントをハイライト映像に使用する事を目的とする。その為、映像Aでは最後に位置する料理周辺の飾り付けと完成した料理を映す部分以外を選出範囲とし、映像Bでは最後の完成部分を選出範囲とする。選出範囲の時間は実際に図2の環境下で撮影した映像群から、該当する部分の大よその所要時間を推定し決定した。

3.2. 映像の取り扱い

ハイライト映像として生成する映像は、映す領域が入力動画の中心を中心点とする縦横比1:1の動画である。動画の一边の長さは入力映像の短辺の長さと同じである。以降で述べる画像特徴量の取得も、正方形の領域内のみで行う。理由としては、左右両端の領域から取得できる画像特徴量の情報が乏しいことが挙げられる。左右両端で行われる行為は、生ごみの処理や使用済み食器などの片付けなど料理には直接関わらない動作が多くなる。さらに、それらの動作が料理中に占める割合も低い。左右両端で行われる動作が少ない為、左右両端の領域の画素が定常成分になる可能性が高く、画像類似度の変動が小さくなる可能性が高い。その為、正方形の領域内のみ画像特徴量から取得できる特徴のみを利用する。

ハイライト映像を生成する際、入力映像は撮影時の2倍速のものとして扱う。具体的には入力映像の奇数番目の映像フレーム取得を行わない。短い時間でテンポよく場面を展開していく都合、等倍速では動作を十分に提示するのが困難になりやすい事が理由として挙げられる。

3.3. 色ヒストグラムによる画像類似度の計算

ここでは、色ヒストグラムの作成手順と2画像間の類似度の計算方法、ハイライト映像のシーンセグメント選出への利用方法を述べる。映像中に映っているものが変わった場合、そこで新しい行程に進んだ可能性が高いと考えられる。行程の変化をハイライト映像に取り込むことで映像にストーリー性を導入でき、閲覧者に強くハイライト映像を印象付けることができる可能性がある。

3.3.1. 色ヒストグラムの作成. 映像中の映像フレームを1秒ごとに1フレームサンプリングしてその映像フレーム内の色ヒストグラムを求め、ベクトル表現を行う3. サンプリングするフレームは各シーンセグメントの最初のフレームである。色ヒストグラムを求める際の映像フレームの色空間はRGB色空間である。色ヒストグラムを作成する際、サンプリングした映像フレームの色数を $R \times G \times B = 256 \times 256 \times 256$ 色から $R \times G \times B = 8 \times 8 \times 8$

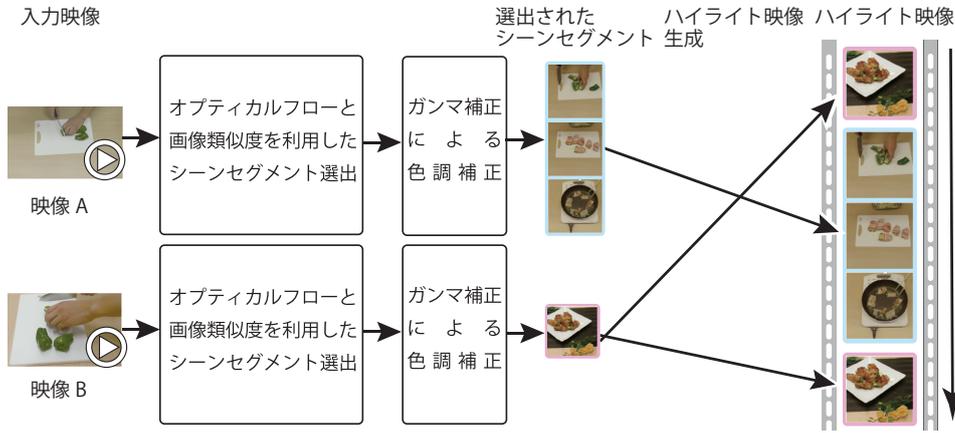


図 1. 提案手法のシステム概要 .

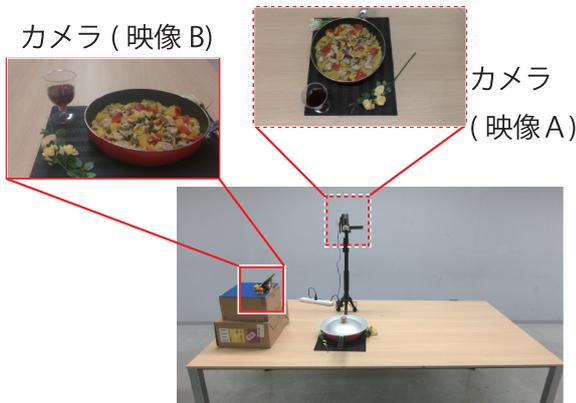


図 2. 撮影位置と各映像の構図 . 映像 B はスマートフォンを傾けた状態で台の上部に設置 .

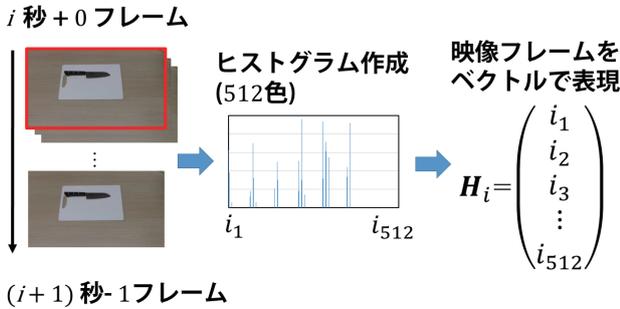


図 3. 色ヒストグラム取得方法 .

色に減色する . サンプルした映像フレームから生成される色ヒストグラムは , 各色の画素が映像フレーム中に存在する個数を度数とする . 2 つの画像のそれぞれの色ヒストグラムの情報は , それぞれ配列 H_1, H_2 として格納される . 以降 , $H_1[i], H_2[i]$ はそれぞれ 1 つ目の画像の色番号 i の度数 , 2 つ目の画像の色番号 i の度数を表す .

3.3.2. 画像類似度の計算. 3.3.1 3.3.1 で求められた 2 つの画像の色ヒストグラム情報が格納された配列 H_1, H_2 を利用して , 2 画像間の画像類似度 S は以下の式を通して計算される .

$$S(H_1, H_2) = \frac{\sum_{i=1}^c \min(H_1[i], H_2[i])}{\sum_{i=1}^c H_1[i]} . \quad (1)$$

ここで , c は画像の色数を示すので , 3.3.1 より , $c = 512$ である . $0 \leq i \leq c$ の任意の整数 i において , $0 \leq \min(H_1[i], H_2[i]) \leq H_1[i]$ である為 , 式 (2) の値の範囲は $0.0 \sim 1.0$ であり , 1.0 に近いほど 2 つの画像が近い特徴であると考えられる .

3.3.3. 画像類似度のシーンセグメント選出への利用. 画像類似度の計算は現在のシーンセグメント N の最初の映像フレーム N_1 と一つ前のシーンセグメント P の最初の映像フレーム P_1 で行われる . この時 , それぞれの映像フレームの色ヒストグラム情報が格納された配列 H_{N_1}, H_{P_1} から計算される画像類似度を $S(H_{N_1}, H_{P_2})$ とする . 次の条件

$$S(H_{N_1}, H_{P_1}) < \theta_S \quad (2)$$

を満たす場合 , ハイライト映像に出力するシーンセグメントとして N と P を選出する . P がすでにハイライト映像に出力するシーンセグメントとして選出されていた場合は , N のみを新規のシーンセグメントとして選出する . θ_S は画像類似度の閾値である .

3.4. LK 法によるオプティカルフローの利用

ここではオプティカルフロー推定で求める数値 , 求めた数値を利用した動きの評価関数によるシーンセグメントのスコア評価 , ハイライト映像選出における動き評価関数のスコア利用について述べる . 著者の過去の手法 [2][3] では , Lucas-Kanade 法 [10] によるオプティカルフローの推定を利用した評価関数が定義されている . 調理動作を行っているなど , 映像中に動きがある場面は何かしらの行程を行っている可能性が高いと考えられることが理由である .

3.4.1. オプティカルフロー推定と評価関数に使用する値. オプティカルフロー推定によって映像フレーム間の動きの大きさを求める . オプティカルフロー評価関数に使用する値は , オプティカルフローのノルムである . 映像フレーム内にあるオプティカルフローは 16 ピクセル間隔で評価関数に使用した 4 . 以降 , 映像フレーム内の (x, y) の位置にあるピクセルのオプティカルフローのノルムを $s|(x, y)|$ と表現する .

3.4.2. 動きの評価関数. シーンセグメントのスコアを評価するために著者の過去の手法 [3] で定義された評価関数を使用する . この手法では , 固定されたカメラで撮影された映像を取り扱うことを想定している . そのため , 固定されていないカメラで撮影された映像を想定した手

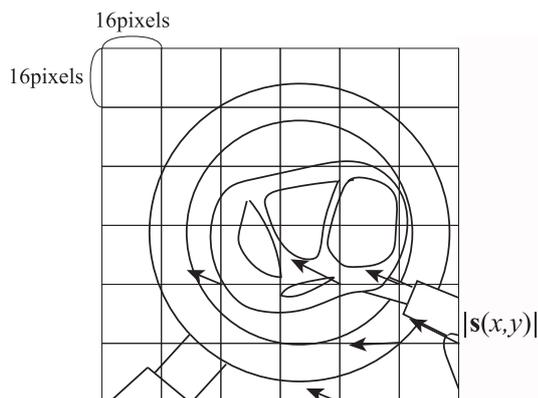


図 4. オプティカルフローの推定 .

法 [2] の評価関数を改変したものを使用している . シーンセグメントのスコアは , 次のように計算される .

まず , シーンセグメント内の k 番目の映像フレームのスコア F_k を次のように定めている .

$$F_k = \frac{\sigma(s)}{M}. \quad (3)$$

$\sigma(s)$ は映像フレーム内の $s|(x, y)|$ の標準偏差を表す . M は映像フレーム内の $s|(x, y)|$ の表すため , 次のように定義される .

$$M = \sum_{m=1}^X \sum_{l=1}^Y \frac{1}{s(d \times m, d \times l)}. \quad (4)$$

d は映像フレーム内の使用するオプティカルフローのピクセル間隔を表す . 3.4.1 での記述より , $(d \times m, d \times l)$ で表現される座標は必ず 16 の倍数の座標である . そのため , $d = 1/16$, X は計算を行う範囲の $1/16$, Y は計算を行う範囲の $1/16$ である . 映像内で物体が動いた場合 , 物体周辺の $s|(x, y)|$ は大きな値を取る . この時 , $\sigma(s)$ は M よりも大きな値を取る傾向が強い . したがって , F_k は k 番目の映像フレームのオプティカルフローが大きいほど大きな値を取る .

i 番目のシーンセグメントのスコア R_i を次のように定める .

$$R_i = \sum_{k=2}^f \overline{F_k}. \quad (5)$$

(5) において , シーンセグメントのフレームレートは f で表現される . (5) では k の初期値は 2 となっている . これは , シーンセグメント単位でスコアを計算するためである . シーンセグメント内の最初のフレームから計算を行う場合 , オプティカルフローの計算のために 1 つ前のシーンセグメントの映像フレームが必要となる . 今回は , シーンセグメントの独立したスコアを計算するように位置付けた .

3.4.3. 動きの評価関数によるシーンセグメント選出 .

3.4.2 で述べた評価関数を用いて , ハイライト映像に使用するシーンセグメントの選出を行う . 現在のシーンセグメント N と一つ前のシーンセグメント P の評価関数によるスコアをそれぞれ R_N, R_P . とする . このとき , 次の条件

$$R_N - R_P \geq \theta_R \quad (6)$$

を満たす場合 , ハイライト映像に出力するシーンセグメントとして N と P を選出する . P がすでにハイライト映像に出力するシーンセグメントとして選出されてい

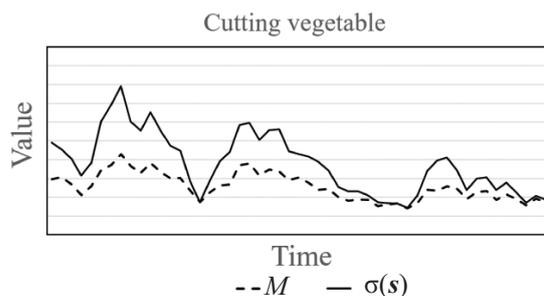


図 5. M と $\sigma(s)$ のデータ (上) と対応する場面のサンプル (下) .

た場合は , N のみを新規のシーンセグメントとして選出する . θ_R は評価スコアの差の閾値である . 式 (6) を設定した理由として , 取得するシーンセグメントの構図的な偏り回避が挙げられる . 著者の過去の手法 [2][3] ではスコアの高いシーンセグメントを取得していたが , 同じような構図のシーンセグメントの取得が問題点となっていた . 動作の開始地点周辺のみを取得することで , 似たような構図のシーンセグメントを過剰に取得をする事を防ぐとともに , 料理の行程を簡潔に分かりやすくする事をねらう .

3.5. ガンマ補正による色調調整

食べ物の色合いは , 暖色系色彩のものは嗜好度が高く , 寒色系色彩や明度の低い色調は嗜好度が低い傾向がある [11] . 生成するハイライト映像をより魅力的に見せるために , RGB3 色のカラーチャンネルに対しガンマ補正を適用する . ガンマ補正の式は以下の通りである .

$$Y = 255 \times \left(\frac{Y}{255} \right)^{(1.0/\gamma)}. \quad (7)$$

式 (7) 中の Y は , $0 \sim 255$ の値をとる画素のチャンネルの数値を表し , γ はガンマ値を表す . ガンマ補正では , γ の値が 1.0 より大きいときにそのチャンネルの強さを増幅し , 1.0 より小さいときにそのチャンネルの強さを減衰させる .

式 (7) を選出されたシーンセグメント中にある映像フレームの各カラーチャンネルに適用する . このとき , カラーチャンネルは RGB 色空間にする . RGB 色空間の各カラーチャンネルのガンマ値 $\gamma_R, \gamma_G, \gamma_B$ の大小関係は , $\gamma_R > \gamma_G > \gamma_B \geq 1.0$ である .

3.6. ハイライト映像の生成

映像 A と映像 B のシーンセグメント選出が終了した後 , ハイライト映像の生成を行う . ハイライト映像の生成方法は , 映像 A の選出されたシーンセグメント群 C_1 の前後に映像 B で選出されたシーンセグメント群 C_2 を挿入

しハイライト映像を生成する。この時、 C_1, C_2 内のシーンセグメントは時系列順になっている。生成されたハイライト映像のシーンセグメント再生順は $C_2 \rightarrow C_1 \rightarrow C_2$ となる。

4. 評価実験

提案手法によって生成された映像は、閲覧者に対してどのような印象を与えるのかを調査する実験を行った。従来の手法と比較して、今回の手法はどのような印象を与える傾向が強いかを考察する。

4.1. 実験に使用したハイライト映像

評価に使用する為、2 の環境で 2 種類の料理映像を撮影した。料理はそれぞれ kurashiru[12] に掲載されている「チーズたっぷり！ピーマンの肉巻き」(以下、このレシピの料理を「料理 A」と呼ぶ)「揚げずに簡単 フライドポテト風」(以下、このレシピの料理を「料理 B」と呼ぶ)というタイトルのレシピをもとにしたものである。1 つの料理映像につき今回の手法で生成したハイライト映像と従来手法 [3] で生成したハイライト映像が存在する。今回の実験では、従来手法の入力映像は映像 A を使用した。また、従来手法のハイライト映像のデュレーションは可変であるため、従来手法で生成したハイライト映像のデュレーションは、今回の手法で生成されたハイライト映像とほぼ同じになるようにした。生成されたハイライト映像の例として、図 6 に料理 B の映像をもとにして生成された本研究手法のハイライト映像と従来手法 [3] のハイライト映像の映像フレームのサムネイルを示す。

4.2. 評価方法

被験者に 4.1 で生成したハイライト映像を閲覧してもらい、閲覧後の動画から受ける印象について評価してもらい。実験の際は、4.1 で生成した動画を順番に閲覧し、1 つの動画の閲覧が終わるごとに閲覧した動画の評価を行ってもらい。また、各動画に対する自由記述も任意で記入してもらい。閲覧の順番は以下のとおりである。

- 1) 従来手法で生成された料理 A のハイライト映像 (A-1)
- 2) 本研究の手法で生成された料理 A のハイライト映像 (A-2)
- 3) 従来手法で生成された料理 B のハイライト映像 (B-1)
- 4) 本研究の手法で生成された料理 B のハイライト映像 (B-2)

評価手法には 7 段階評価の SD 法 (Semantic Differential Scale Method) を用いた。被験者は、20 ~ 24 歳の男女 10 名 (男性 6 名、女性 4 名) である。従来手法で生成されたハイライト映像と、それに対応する本研究の手法で生成されたハイライト映像の各評価項目の評価値の平均値間に有意差が存在するかどうかを、評価項目ごとに t 検定で求める。有意水準を 0.05 として t 検定を行い、どのような評価項目で有意差があるかを調べた。

4.3. 実験結果

t 検定を行い、有意差があると認められた項目を表 1 に示す。料理 A、料理 B とともに有意差があると認めら

れた項目は 3 項目、料理 A のみに有意差が認められた項目は 3 項目、料理 B のみに有意差が認められた項目は 14 項目だった。なお、評価項目のうち、「個性的」と「平凡な」の形容詞は対になる形容詞である為、異なる形容詞として評価した。

表 1. 有意差があると認められた項目

有意差の有無	項目	p 値 (料理 A)	p 値 (料理 B)
料理 A, B で有意差あり	明るい	4.40×10^{-3}	3.11×10^{-3}
	スピード感がある	1.07×10^{-3}	8.19×10^{-3}
	速い	6.33×10^{-3}	1.90×10^{-3}
料理 A で有意差あり	なめらか	9.99×10^{-3}	2.73×10^{-1}
	鮮明な	1.28×10^{-2}	1.11×10^{-1}
	個性的	2.20×10^{-2}	本文参照
料理 B で有意差あり	短い	8.68×10^{-2}	2.50×10^{-3}
	心地よい	8.76×10^{-1}	6.74×10^{-3}
	迫力がある	8.11×10^{-2}	9.53×10^{-3}
	きれい	2.39×10^{-1}	5.12×10^{-3}
	分かりやすい	6.33×10^{-1}	5.12×10^{-3}
	カッコいい	7.70×10^{-1}	4.40×10^{-3}
	食べたい	1.93×10^{-1}	8.63×10^{-3}
	印象的な	5.43×10^{-2}	3.11×10^{-3}
	濃厚な	7.26×10^{-1}	1.33×10^{-2}
	おいしい	2.69×10^{-1}	1.15×10^{-2}
	見たい	3.09×10^{-1}	4.14×10^{-2}
	面白い	8.86×10^{-2}	4.65×10^{-3}
	平凡な	本文参照	1.50×10^{-2}
	他人に知らせたい	2.44×10^{-1}	1.07×10^{-2}

4.4. 考察

料理 A, B とともに有意差があると認められた項目は「明るい」「スピード感がある」「速い」の 3 項目であった。従来手法と比較して「スピード感がある」「速い」という項目に有意差が生じたのは、主に 2 倍速で映像を扱ったことが要因として挙げられる。「明るい」という項目に有意差が生じた要因は、ガンマ補正による効果が考えられるが、二つの手法で使用する映像の数が異なる為、映像 A、映像 B のカメラの性能差や異なるカメラアングルであったことも影響していると考えられる。

料理 A と比較して、料理 B がより多数の項目で有意差が生じた理由は、B-1 が A-1 よりも選出されたシーンセグメントの似通った場面の偏りが強かったためであると思われる。B-1 のアンケートにおいて「同じような場面が多い」という内容の意見が A-1 よりも多かった。図 6 に示した通り、B-1 のハイライト映像の大部分が細切りにしたジャガイモを炒めている場面を映したシーンセグメントで構成されている。そのため、B-1 の印象が比較的否定的なものであったことが理由であると思われる。同様に、「短い」という項目は、場面の偏りが要因となったと考えられる。B-1、B-2 とともに動画のデュレーションは同じであるが、場面が偏っているため、同じような場面が長時間続いていることが強く印象づけられたと考えられる。「分かりやすい」という項目についても、全行程中の 1 つの場面のみを映している印象から生じるものだと思う。

自由記述においては、本研究の手法を適用した A-2、B-2 のほうが分かりやすかったという意見が多かった。また、「最初に完成したのを見せると印象が良くなる」という内容の意見もあった。

A-2、B-2 において、形容詞対である「個性的な」「平凡な」でそれぞれ有意差が発生してしまった要因として、評価実験の過程で先に A-2 を閲覧した事が仮説として挙げられる。A-2 を閲覧したことで、本手法で生成

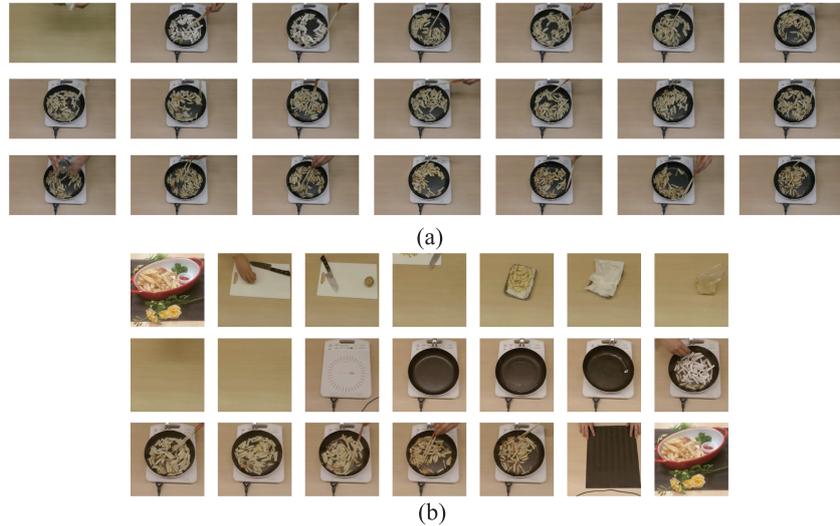


図 6. 料理 B の映像から生成されたハイライト映像フレームを 3 秒おきに示したサムネイル。(a) 従来の手法 [3] (b) 本研究の手法

されるハイライト映像のイメージが多少形成された可能性があると思われる。

本研究の手法で生成された料理のハイライト映像に対する被験者の印象傾向や、それに関連する被験者のコメントを踏まえ、本手法によって従来の手法よりも肯定的な印象を提示する動画を自動生成できると言える。

5. 今後の展望

今後の展望としては「評価する料理映像の種類を増加」「選出アルゴリズムの改善」「視聴順の影響の調査」「閲覧者の能力と評価の相関関係の調査」の 4 つが挙げられる。

まず、「評価する料理映像の種類を増加」について述べる。今回の実験では、2 種類の料理映像を使用してシステムを提示したが、調理方法は多種多様に存在するため、様々な調理方法の料理を撮影した映像を使用して評価を行いたい。次に「選出アルゴリズムの改善」について述べる。撮影した料理映像の中には映像 B からシーンセグメントを 1 つも選出することが出来なかったものも存在した。その為、選出するシーンセグメントの下限、または上限を決定するアルゴリズムの考案も行いたい。3 つ目に、「評価実験手順の改良」について述べる。1 で述べた通り、動画コンテンツは閲覧者に強く印象が残るコンテンツである。その為、被験者によって動画を閲覧順を変更するなどの措置をとり、視聴順の影響の有無を調査したい。最後に、「閲覧者の能力と評価の相関関係の調査」について述べる。ソーシャルメディアの多様化により、提供されるサービスによって、サービスが重点を置くメディアコンテンツ、ユーザのメディア形態の嗜好が異なる。また、同じサービス内においても、メディアコンテンツの制作能力にはユーザ毎に大きな差が存在する。その為、同じ動画コンテンツでも、ユーザによって印象が異なる可能性がある。今後の展望として、これらの要因の相関関係を調査し、システムに反映できるように研究を行いたい。

6. 結論

本研究では、ソーシャルメディアコンテンツとしての料理のハイライト映像の自動生成を行う手法を提案した。色ヒストグラムによる画像類似度の計算とオプティカルフローを利用した動きの評価関数、色調補正により、撮影された 2 つの料理映像から料理のハイライト

映像を生成した。生成されたハイライト映像が閲覧者にどのような印象を与えるかを調査するため、SD 法を用いたアンケート調査を行った。実験結果を分析した結果、従来の手法よりも肯定的な印象を与えることが出来る可能性があるとして示唆された。

参考文献

- [1] BuzzFeed, “Tasty,” <http://bit.ly/tastyYT>, (2018 年 1 月 30 日に閲覧).
- [2] 北清拓実, 小池崇文, “画像特徴量を利用した料理のハイライト映像の自動生成,” 情報処理学会 第 78 回全国大会, 2016.
- [3] T. Kitase and T. Koike, “Automatic creation of cooking highlight videos by sound loudness and optical flows,” in *Proceedings of the 5th IIAE International Conference on Intelligent Systems and Image Processing*, 2017.
- [4] 総務省, “平成 28 年情報通信メディアの利用時間と情報行動に関する調査報告書,” 2017. http://www.soumu.go.jp/menu_news/news/01iicp01_02000064.html (2018 年 1 月 30 日入手).
- [5] P. M. Leonardi, M. Huysman, and C. Steinfield, “Enterprise social media: Definition, history, and prospects for the study of social technologies in organizations,” *Journal of Computer-Mediated Communication*, vol. 19, no. 1, pp. 1–19, 2013.
- [6] M. Pittman and B. Reich, “Social media and loneliness: Why an instagram picture may be worth more than a thousand twitter words,” *Computers in Human Behavior*, vol. 62, pp. 155–167, 2016.
- [7] Y. Hayashi, K. Doman, I. Ide, D. Deguchi, and H. Murase, “Automatic authoring of a domestic cooking video based on the description of cooking instructions,” in *Proceedings of the 5th international workshop on Multimedia for cooking & eating activities*. ACM, 2013, pp. 21–26.
- [8] K. Doman, C. Y. Kuai, T. Takahashi, I. Ide, and H. Murase, “Video cooking: Towards the synthesis of multimedia cooking recipes,” in *International Conference on Multimedia Modeling*. Springer, 2011, pp. 135–145.
- [9] Twitter, Inc., “Twitter,” <https://twitter.com/>, (2018 年 1 月 30 日に閲覧).
- [10] B. D. Lucas, T. Kanade *et al.*, “An iterative image registration technique with an application to stereo vision,” 1981.
- [11] 豊満美峰子, 松本仲子, “食物・食器・食卓の配色が嗜好に及ぼす影響,” *日本調理科学会誌*, vol. 38, no. 2, pp. 181–185, 2005.
- [12] dely, Inc., “kurashiru,” <https://www.kurashiru.com/>, (2018 年 2 月 1 日に閲覧).