

ソーシャルネットワーキングサービスに適した料理レシピ動画の生成

祖父江 亮* 中澤 満† 益子 宗† 山下 隆義* 藤吉 弘巨*

概要. ソーシャルネットワーキングサービス (SNS) に投稿される調理工程を短時間に要約した動画の作成は、人的、時間的コストだけでなく、高度な編集技術が必要である。そこで本研究では、誰でも手軽に SNS に適した要約動画を作成できるように、調理動画を半自動で要約するシステムを提案する。動画を要約するにあたり本手法では、料理レシピサイトの手順画像が調理工程の流れを説明していることを活用し、撮影動画から手順画像と類似したシーンを抽出することで実現している。また、あらかじめ定めた SNS に適する要約動画の条件を満たすために、再生時間の調整やテロップの追加など必要機能を提案システムに組み込んだ。提案システムを用いた評価実験では、被験者から手軽に SNS に適した要約動画を作成できたと高い評価を得ることができた。

1 はじめに

ソーシャルネットワーキングサービス (SNS) に投稿される動画の中でも調理工程を要約した料理レシピ動画 (要約動画) は、工程や出来上がりを簡単・短時間に把握できることから、女性層を中心に人気を集めている。一方で、要約動画の作成には特別なスタジオやソフトウェアが必要であり、一般ユーザーが手軽に編集することは難しい。そこで本研究では、誰でも手軽に SNS に適した要約動画を作成できるようになることを目指し、調理工程を撮影した長時間映像から手軽に要約された動画を生成するシステムを提案する。専門家とのヒアリングより定めた SNS 映えする要素を満たしつつ、要約動画のシーン候補を複数提示してユーザーが確認および選択できるインターフェースを構築した。

2 SNS 映えの定義

SNS に適した要約動画の定義は、料理レシピのウェブサイトを運営する専門家にヒアリングを行い、次のように定めた。(1) 調理工程全体が見えるアングルの映像である、(2) 倍速再生で調理工程を把握できる、(3) 材料・工程を説明する短いテロップがある、(4) 完成を把握できる映像が最初と最後にある、(5) 30 秒から 60 秒程度の短時間にまとまっているこれらを満たす要約動画を作成するシステムを提案する。

3 提案システム

提案システムの利用の流れを図 1 に示す。Step1 では、ユーザーが料理レシピサイトから撮影したい料理のレシピを選択する。Step2 では、選択したレシピに従って調理を行い、全ての調理工程を撮影する。Step3 では、提案システムを用いて撮影動画からレ



図 1. 提案システムの利用の流れ

シピサイトの手順画像と類似したシーンを複数自動抽出する。ユーザーはそれらを確認し、最も自身が好むシーンを選択する。Step4 では、ユーザーが提案システムより要約動画を生成し、SNS に投稿する。

3.1 類似シーンの抽出

本研究では、レシピサイトの手順画像と類似する動画フレームを選択して要約動画のシーンとする。手順画像と動画の各フレームに対する特徴ベクトルを抽出する。特徴ベクトルの抽出には、まず InceptionV3 [1] をベースにした畳み込みニューラルネットワーク (CNN) を用い、手順画像と類似する動画フレーム間の特徴ベクトルが近くなるよう Triplet loss [2] を用いて学習した。CNN は 128 次元の特徴ベクトルを出力する。本研究では 199 種のレシピから 1014 枚の手順画像、17 本の調理動画を用いて学習を行った。特徴ベクトルの類似度はユークリッド距離で算出し、上位 5 フレームを調理工程のシーン候補とする。

3.2 要約動画作成システム

提案システムのユーザーインタフェースを図 2 に示す。各手順画像に対して、類似度の高いフレームを

Copyright is held by the author(s).

* 中部大学

† 楽天株式会社 楽天技術研究所



図 2. 提案システムのインタフェース

	前提条件	質問内容	回答人数	回答方式
Q1	提案システムを用いハンバーグの動画を要約させた	提案システムで要約動画は手軽に作れたか	9 人	5段階 (そう思う ややそう思う どちらとも言えない ややそう思わない そう思わない)
Q2	提案システムにより要約された動画の例を見せた	動画冒頭のタイトルの写真で料理の完成のイメージができたか	35 人	2段階 (はい、いいえ)
Q3		動画を見て調理工程が分かったか		
Q4		SNSに適する条件(5か条)を満たしたか		

表 1. アンケート質問項目

含む 30 秒間の範囲を動画クリップとする。そして類似度の高い上位 5 つの動画クリップを提示し、ユーザーに選択してもらう。各動画クリップの長さや手順の要約文章は、変更が可能である。また、レシピ動画の視聴者が、完成した料理をイメージしやすいように要約動画の冒頭には完成した料理の画像を挿入することが可能である。

4 評価実験

提案システムおよび要約した動画の有効性を示すために、アンケート評価を行なった。

4.1 実験概要

アンケート評価の内容を表 1 に示す。アンケートは本システムを利用して要約動画を作成した後に行うシステム評価、要約動画見た後に行う要約動画評価の 2 つのアンケート評価を行った。アンケートの回答者には調理をした本人や動画を撮影した人物は含まれていない。本実験ではあらかじめハンバーグのレシピ [3] に従って調理、撮影した 42 分間の映像を要約対象として用いた。

4.2 評価結果

アンケートの評価の結果を図 3 および図 4 に示す。Q2, Q3 では好意的な回答が否定的な回答の割合を上回ったが、Q1 では好意的な回答は半数にとどまっている。自由記述でのコメントでは、「どのフレームを選択したら良いか迷った」という意見が見られた。その理由として、似ているシーンが動画クリップ候補として提示されていたためであると考えられる。

また、SNS に適した動画の定義にもとづく 5 項目のアンケート評価の結果 (Q4) を図 5 に示す。全ての項目で 70% 以上が「はい」という回答であったことから、提案システムを利用することで SNS に適した要約動画を手軽に作成できることを示した。

Q1: 本システムで要約動画は手軽に作れましたか？

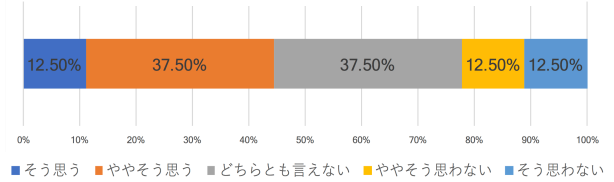
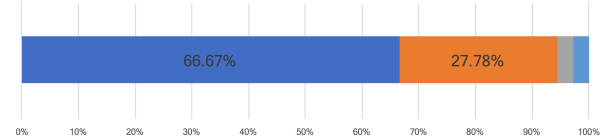


図 3. 提案システムに対するアンケート結果 (Q1).

Q2: 冒頭の写真で料理の完成のイメージができましたか？



Q3: 動画を見て、調理工程がわかりましたか？

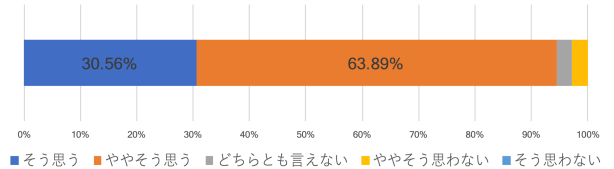


図 4. 要約動画に対するアンケート結果 (Q2~Q3).

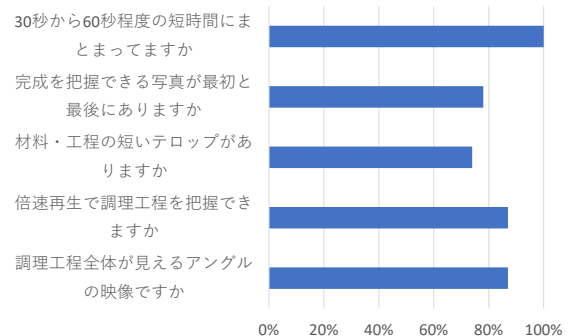


図 5. SNS に適した動画の定義にもとづく 5 項目のアンケート結果 (Q4)

5 まとめ

本研究では、SNS に適した要約動画の生成システムを提案した。今後は、類似シーンの抽出精度向上や提案システムのユーザインターフェース改善により、更なる手軽な動画要約を実現する。

参考文献

- [1] Christian Szegedy, et al. Rethinking the inception architecture for computer vision. *CoRR*, Vol. abs/1512.00567, , 2015.
- [2] Florian Schroff, et al. Facenet: A unified embedding for face recognition and clustering. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.
- [3] 楽天レシピ. 【その9】: シェフ直伝♪洋食屋さんのハンバーグ レシピ・作り方. <https://recipe.rakuten.co.jp/recipe/1860001109/> (accessed on June 1, 2018).