

料理写真の魅力度推定器を使った写真撮影補助システムにおける撮影手法の改善 Improvement of Photography Method in Photography Assistance System Using An Attractiveness Estimator of Food Photos

大木 慧矢[†]
Seiya Ohki

マッキン ケネスジェームス[†]
Kenneth J Mackin

永井 保夫[†]
Yasuo Nagai

1. はじめに

近年スマートフォンの発展・普及により写真撮影が手軽となり、SNS などへの公開が容易になってきた。また、撮影上での色合いなどは最近の HDR (High Dynamic Range) 機能や写真の編集アプリでの調節が可能となるとともに、簡単に編集可能となる。そのため、写真に対する一層のクオリティの高さが求められている。しかしながら、露出や色合いは簡単に編集可能であるのに対して、最適な構図を決めて上手く撮影することは専門家でないとおこなえない。

我々はこの問題を解決するために、深層学習を使った魅力度推定器を利用し、魅力的な構図で撮影をするシステムを開発した[1]。本システムでは、佐藤ら[2]は、畳み込みニューラルネットワーク (CNN : Convolution Neural Network) である VGG16 が魅力度推定器に適していることを示した。ここではこの魅力度推定器を参考にスマートフォンなどから撮影した料理の動画から最適な構図の写真を出力する撮影補助システムを検討した。その結果、データセットにある料理の方がデータセットにない料理より良い構図の写真を出力できることがわかった。しかしながら、余計なものなどが映り込んだ画像や見切れている画像を選んでしまう場合もあり、写真撮影補助システムとして手振れなどの画像を選ばないような改良が必要となった。

そこで、本提案では、動画を画像化し入力する代わりに、カメラの連写機能を使うことで見切れている画像を少なくすることで魅力度の向上を行い、システムの精度を改善する提案を行う。

2. 提案手法

2.1 魅力度推定器

提案手法の魅力度推定器では、VGG16 に転移学習を行う。転移学習をするにあたって井手[2] らの作成したデータセット「NUFOOD 360×10」を用いる料理の写真を 244×244 にリサイズし入力し、画像に対する魅力度を教師データとして学習を行う。重みの学習を行う全結合層の活性化関数には ReLU を使用し、出力の全結合層にはシグモイド関数を使用し、最適化手法には Adam、評価関数 MAE (Mean Absolute Error) を用い、学習回数は 100 回とした。

2.2 入力写真の撮影方法

入力を行う際の動画の撮影方法として 2 つの手法を検討した。手法 1 では図 1 の A の様に対象の料理を俯角 0 度で 360 度撮影し、その後、図 1 の B の様に俯角をおおよそ 30 度

上げ、A の様と同じように撮影し、俯角 90 度まで繰り返し撮影を行った。

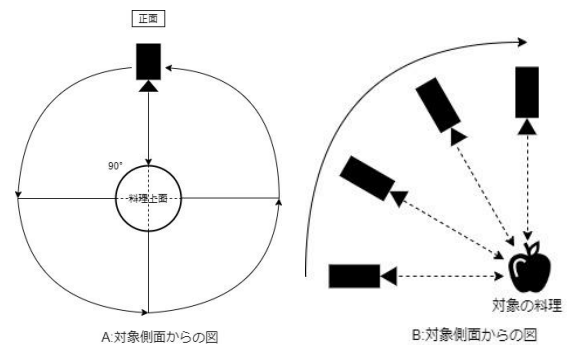


図 1 手法 1 のイメージ

また、手法 2 では、図 2 の様に対象の料理に対し正面から俯角 0 度で左右 90 度の 180 度撮影を行い、その後手法 1 の B と同じように俯角をおおよそ 30 度ずつ同じように撮影しながら俯角 90 度まで撮影を行った。

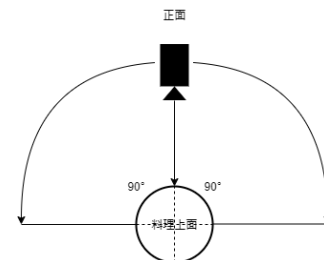


図 2 手法 2 のイメージ

手法 1 と手法 2 の撮影方法を使用し、カメラ (スマートフォン) の連写機能を使用し、撮影を行う。また、比較として前研究と同じ様に動画でも撮影を行った。

3. 評価実験

まず、VGG16 を使った魅力度推定器を作成する。次にデータセットに含まれている料理とデータセットに含まれていない料理を撮影、手法 1 と手法 2 に基づいて、動画と連写機能を使って撮影した写真を使い、撮影手法によっての見切れている画像の枚数を比較する

最後にそれぞれの動画と連写機能を使って撮影した写真をそれぞれ魅力度推定器に入力し比較を行い、評価を行った。

実験 1

手法 1 で撮影した動画を入力し実験 1 では VGG16 を使った魅力度推定器により評価を行った

[†] 東京情報大学 Tokyo University of Information Sciences

図 3 の A の画像は動画での魅力度最大, B の画像は連写での魅力度最大を表している。



図 3 手法 1 を使った出力

実験 2

実験 2 では, 手法 2 で撮影した動画を入力し VGG16 を使った魅力度推定器により評価を行った. 図 4 の A の画像は動画での魅力度最大, B の画像は連写での魅力度最大を表している。



図 4 手法 2 を使った動画での出力

それぞれの手法での出力された画像の魅力度は表 1 のようになった。

表 1 出力された魅力度

	魅力度最大
実験1-動画での出力	0.8086604
実験1-連写での出力	0.8642167
実験2-動画での出力	0.5976593
実験2-連写での出力	0.820611

4. 考察

今回の実験では, 表 2 のように動画より連写機能撮影のほうが見切れている画像の枚数を少なくできることが分かった。

表 2 入力データの見切れている画像の割合

	入力枚数	見切れている画像	割合
実験1-動画での出力	1570	496	0.315
実験1-連写での出力	763	219	0.287
実験2-動画での出力	264	30	0.113
実験2-連写での出力	164	35	0.213

また, 実験 1 における動画と連写機能での撮影では, 動画で撮った場合の方がデータセットの最大値の画像に近く, どちらも見切れている画像は抽出されなかった。

同様に, 実験 2 の動画と連写での撮影でも, 構図には変化がほとんどなく, データセットでの魅力度が最も高いチーズバーガーとは少し違う視点の画像が出力されたが, 実験 1 と同じように見切れている画像は出力されなかった。

以上のように, それぞれの手法での魅力度は動画で撮影するより, 連写で撮影することで魅力度を高くすることができた。また, 動画で撮影するよりも写真で撮影することにより, 入力データの見切れている画像の割合を少なくすることができ, 出力する際の魅力度を高めることができた。

しかしながら, 実験 1 の連写での出力画像は真上からの視点で, データセットでの魅力度が最も高いチーズバーガーの視点とは違うものになってしまった, これは違う料理の魅力度が高い画像に近い視点の画像を抽出してしまったためであると考えられる。

5. 結論

我々は, 最適な構図を決めた写真撮影が難しいことに対して, ディープラーニングによる料理写真の魅力度推定器を使い, 動画から最適な構図の写真を出力する写真撮影補助システムを検討してきた [1]。本研究では, さらに, 動画を画像化し入力する代わりに, カメラの連写機能を使うことで手振れや見切れている画像を少なくする撮影手法の改善策を提案し, より良い構図の写真を出力できるか実験を行った。

その結果, 連写で撮影を行うことで見切れている画像を減らすことができ, 出力される魅力度を前回報告より魅力度を高く出力することができた。

しかし, 出力された画像は対象の料理のデータセットとは違う視点になった画像が出ていたため, 今後は撮影している対象を判断する分類機を作成することで, 対象と違う料理の魅力度が高い画像に近い画像を選んでしまうのを防止するシステムを検討している。

参考文献

- [1] 大木 慧矢, ケネス マッキン, 永井保夫: 料理写真の魅力度推定器を使った写真撮影補助システムの検討, 第 82 回情報処理学会全国大会
- [2] 佐藤 陽昇, 道満 恵介, 平山 高嗣, 井手 一郎, 川西 康友, 出口 大輔, 村瀬 洋: 畳み込みニューラルネットワークを用いた料理写真の魅力度推定, 信学技報, vol. 117, no. 252, MVE2017-32, pp. 107-111, 2017 年.
- [3] Kazuma Takahashi, Keisuke Doman, Yasutomo Kawanishi, Takatsugu Hirayama, Ichiro Ide, Daisuke Deguchi, Hiroshi Murase: "Estimation of the Attractiveness of Food Photography Focusing on Main Ingredients", Proc. of 9th Workshop on Multimedia for Cooking and Eating Activities (CEA), pp. 1-6, Aug. 2017.