

RGB 画像と音声を用いた冷蔵庫内食品管理システム

Refrigerator food management system using RGB images and speech interface

謝 二ノ†
Xie Ninuo下ノ村 和弘‡
Shimonomura Kazuhiro

1. はじめに

近年, IoT や AI 情報技術を駆使より, 住宅の「スマート化」が世界で急速な拡大が見込まれている. 一方で現状, パナソニック株式会社のアンケート結果[1]によると, 主婦にとって, 家事の効率化・時間短縮化を図りたいが, 料理に関しては手を抜きたくない意識が明らかである. そのために, 食材管理又はレシピ提案などの機能をもたらすスマート冷蔵庫では, 家事の効率化が実現できるし, 食材ロスも減らせる.

冷蔵庫の食材管理に関するシステムの一例として, ボッシュの「Smart food management」システム[2]があげられる. システムの概要としては, 冷蔵庫の食材を自動認識し, その食材だけを使ったレシピを提案, スマートフォンアプリ上でレシピを選択すると, その調理メニューがオープンへ送られて, 調理を行なえるというものである.

先行研究によると, スマート冷蔵庫システムが既に食材の自動認識やレシピ提案などの機能をもたらせる. しかし, 賞味期限や食材の使用量や残量を把握し管理することができないという問題が挙げられる. 以上の問題点を解決するため, 本研究では RGB 画像と音声を用いた冷蔵庫内食品の種類を識別する能力を備え, 賞味期限や食材の量を把握する食品管理を行う方法について述べる.

2. 提案手法

本論文は, RGB 画像を用いた食材の種類を識別する能力を備え, 音声対話による賞味期限の提示するシステムを提案する.

システムには, 二つの段階に分けられる.

段階 1, システムが初めてデータを入力する初期化状態である. ここでは, 「ミルク」に関するデータが新たに入力されたものと仮定する. ロボットアームが冷蔵庫のドアを開け, 画像識別による冷蔵庫内の食品識別を行う. 「ミルク」というラベルが識別できたら, 音声対話によるミルクの賞味期限を入力し, 経過時間のカウントを開始する. 賞味期限のデータに従って, 期限の一日前になるとシステムが音声でユーザーに通知する. 最後に, ロボットアームがドアを閉めて, 段階 1 は終了する. ただし, 冷蔵庫のドアが閉めても経過時間のカウントを続ける.

段階 2, システムが既に段階 1 で食品を識別し, 賞味期限を取得した状態である. ロボットアームが冷蔵庫のドアを開けて, 画像識別による冷蔵庫内の食品識別を行い, ミルクのラベルがあるかどうかを確認する. ラベルが検出できる場合, 経過時間のカウントを続ける. そうでなければ, ミルクが飲み終わった状況と見なされ, カウントを中止,

ミルクのデータをリセットする. ミルクの賞味期限切れる場合, カウントも中止する. 最後に, ロボットアームがドアを閉めて, 段階 2 は終了する.

新たな食材が検出されると, 段階 1 の初期化から始まる. ただし, 冷蔵庫内複数の食材が検出される場合, 食材ごとに個別の計算が実行される.

3. 実験と結果

本章では, 食材検出と賞味期限管理の両面からシステムの実験を行う.

3.1 実験対象と目的

本実験では, 「ミルク」を実験対象として選定する.

リアルタイムでデブスカメラが冷蔵庫内のミルクを画像認識による検知し, 検知後に音声で賞味期限を記録する予定である.

本実験の目的は, 冷蔵庫内のミルクを正確に検知し, 入力された賞味期限の 1 分前に音声で提示することである.

ただし, ここでは実験よりシステムの正しさを証明するために賞味期限の 1 分前としており, 通常の提示時間は有効期限の 1 日前である.

3.2 画像認識

本研究提案する食品管理システムは, 画像識別による冷蔵庫内の食材の種類を識別する能力を備えたいため, リアルタイムオブジェクト検出アルゴリズムが必要である. このために, デブスカメラで取得した画像を YOLOv4[1]で食材検出を行う. 本研究では, 教師データの収集として, 同じ角度と高さで冷蔵庫内の 171 枚の画像を撮影した. 続いて, 畳み込みニューラルネットワークと YOLO の誤差関数を定義し, トレーニングを開始する. トレーニングが終了後, 学習済みモデルを取得した.

3.2.1 画像認識の実験過程と結果

画像認識のプログラムを実行した後, ビデオの fps は 13 から 14 の間で推移し, 「ミルク」ラベルの信頼度は最小 97%, 最大 100%となった.

また, ミルクの座標に関する出力の抜粋を以下に掲載する.

b'milk 0.99' 213 670 331 779

b'milk 1.00' 212 671 333 779

b'milk 1.00' 212 670 334 779

ここで, [b'milk]はラベルがミルクであることを示す. 最初の数値は, 信頼度を示す. 次の 4 つの数値は, それぞれ画像中のアンカーボックスの座標を表す.

上記のデータは, リアルタイム画像認識の結果の正確さを示している

† 横浜国立大学 Yokohama National University

‡ 立命館大学 Ritsumeikan University

3.3 賞味期限

これまでの食材検出では画像認識が用いられたが、本研究で用いたカメラはロボットに設置された外部カメラである。冷蔵庫内の二次元コードを正確に取得することは困難であり、食材が包装されていない場合にも二次元コードを取得することができない。そのため、賞味期限を画像認識で取得することは難しいと想定し、音声入力での賞味期限を取得することにした。

具体的には、システムとユーザーの音声対話による情報入力を採用した。

システム音声には、google が提供している Google Text-to-Speech という API を用いて、テキスト入力から音声ファイルを生成する。

ユーザーの音声入力には、Speech Recognition という API を用いて、リアルタイムで音声データをテキストに変換する。生成したテキストがデータとして保存する。

画像認識により、冷蔵庫内の食材を検出した後、食材の賞味期限をユーザーに確認、ユーザーが回答した後、賞味期限をデータとして記録される。

3.3.1 賞味期限の実験過程と結果

ミルクが検出されると、同時に音声対話プログラムの実行が開始される。出力欄の結果を以下に掲載する。

ミルクの賞味期限は
翻訳中
記録が完了しました。

Milk time is 120s

ただし、[ミルクの賞味期限は]と[記録しました]はシステム音声、[2分です]はユーザーの音声応答を変換したテキストである。[翻訳中]はユーザーの音声をテキストに変換していることを示す。

ユーザーの音声が賞味期限と関係ない場合、あるいは音声翻訳に誤りがある場合、システムはユーザーに新しい音声の入力を求める。ユーザーの回答は、時間単位さえ入っていれば、正確な入力と判定される。例えば、[2分]、[賞味期限は2分です]は正しく入力できる。このプログラムでは、文字列から数値を取り出し、int 関数に変換する。単位が「分」の場合は、入力された数値に 60 を掛けて「秒」で格納し、結果を出力する。

この時点から賞味期限までのカウントダウンが開始される。

残り時間が 1 分の場合の出力欄の抜粋を以下に掲載する。

Milk Time is 60.054603576660156s

Milk Time is 59.996758222579956s

ミルクは賞味期限まであと 1 分

Milk Time is 56.55343770980835s

賞味期限切れの場合の出力欄を以下に掲載する。

Milk Time is 0.989347219467131s

ミルクは期限切れ

Milk Time is -0.9775378704071045s

テキスト出力には、対応するシステム音声で提示する。出力欄の秒数は、有効期限まであと何秒あるかを表している。ミルクが検出された時刻を開始時刻とし、開始時刻から有効期限を引いて加算し、有効期限までの時間を算出する。

これより実験終了、食材の種類検出と賞味期限を管理するシステムの有効性が実証された。

3.4 考察

本実験では、精度を上げるために、冷蔵庫内のターゲットが 1 つしかない状態で行われた。

ターゲットが複数ある場合のコード設計はされたが、ターゲット数が増えると処理が遅くなり、精度もかなり落ちる。そのため、本論文で提案するシステムでは、複数のターゲットを扱う方法について改善の余地があると考えられる。

また、本実験では冷蔵庫のドアを開け続けた状態で行われ、冷蔵庫のドアが閉まってから開くという検知は行われておらず、日常生活に即していないため、改善が必要である。ただし、画像認識用のプログラムでは、初めて食材を検出した時点で座標が記録されることになる。冷蔵庫のドアを閉めた後に再び開けても、座標が同じであれば同じ食材と判断してタイミングを継続するという効果を狙っている。これは今後の課題としたいと考える。

4. おわりに

本論文が提案した食品管理システムは、実験により、YOLOv4 を用いた画像認識による食材の種類、個数及び座標を検出すること及びシステムとユーザーの音声対話による賞味期限を管理することが可能と実証された。

本実験では、YOLOv4 を用いた画像認識の結果がより良好で、様々な食材の配置を想定してデータセットを多くの変数で調整したため、予測時に基本的に 95% 以上の結果が得られた。

一方、音声入力の精度はそれほど高くなく、正しく入力できるまでに何度も繰り返す必要があることが多く、達成感はあるが効率は悪い。この部分の改善と見直しが必要であると考えられる。

また、本実験では 4 つの食材しか入力されていないため、冷蔵庫に入力されていない食材があった場合、宛も入力されたように誤って記録される可能性がある。本実験では、ラベルの信頼度がほぼ 95% 以上であることから、信頼度の閾値などのパラメータを調整することで、精度を向上させる。

ただし、肉類など画像で識別しにくい食材は、システムで識別することが困難である。そこで、信頼度が 80% 未満の項目を検出したときに、システム音声で「これは新しい食材ですか」と尋ねて、ユーザーが「はい」と答えたら新しいデータとして記録し、学習させることはできないかと考えている。

システムがユーザーと音声で対話することで、画像認識の可能性はより広がると思っている。

参考文献

- [1] 主婦のキッチンや料理に関する意識・実態調査、<https://news.panasonic.com/jp/press/data/2013/04/jn130405-3/jn130405-3.htm>, 2013 年 4 月 5 日, 2022-05-31 閲覧
- [2] smart-food-management, <https://www.bosch-home.com.sg/highlights/home-connect/smart-food-management> 2022-05-31 閲覧
- [3] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." arXiv preprint arXiv:2004.10934 (2020).