特別研究報告書

画像・振動音・荷重データを統合的に用いた食材認識

指導教員　　美濃 導彦　教授

京都大学工学部情報学科

井上 仁

平成24年2月2日

画像・振動音・荷重データを統合的に用いた食材認識

井上 仁

内容梗概

近年，一般の主婦等が自作のレシピを web 上に公開できるようなサービスが展開されている．このようなサービスは，自作のレシピを比較的簡単な手順で公開できるという点で，主婦をはじめとする多くの一般ユーザから支持を得ている．しかし，レシピを人手で作成することは手間がかかる．これに対して，調理観測データから自動でレシピを生成できれば，上記のようなレシピ公開の助けとなりうる．レシピとは加工対象食材（食材）と加工動作の対を単位作業として列挙したものであるので，レシピの自動生成処理の実現にあたっては，食材と加工動作を調理観測データから認識する技術が重要となる．本研究ではこのうちの食材認識に注目する．

食材認識を目的とした研究は従来からいくつか行われており，それらの従来研究では，調理観測データとして画像・振動音・荷重といったモダリティのデータを単独で利用することが検討されている．しかし，これらのモダリティには各々に問題点があり，単一のモダリティを用いるのみではその解決は難しい．一方で，これらのモダリティには，それぞれの問題点を相互に補い合うような関係もある．例えば，画像のみでは色の似た白ネギとダイコンの識別は難しく誤認識が発生するが，食材の切断時にかかる荷重であれば，この 2 つを容易に識別できる．逆に，荷重のみでは硬さの似たダイコンとニンジンを誤認識しやすいが，画像であれば容易に識別できると考えられる．そこで本研究では，画像・振動音・荷重という 3 つのモダリティを併用し，それぞれのモダリティに対応するデータを統合的に用いることにより，各々を単独で用いた場合に発生する誤認識が改善されるような食材認識手法の提案を目指す．

複数モダリティの時系列データを統合する際には，それぞれのデータから単純に時刻が一致している部分を取り出して統合しても認識に有効とはならない，という問題点がある．実際，振動音と荷重については切断に相当する区間のデータから特徴的な値を示す特徴ベクトルが抽出できる一方で，画像においては，切断時には調理者の手によって食材が遮蔽されるため，認識に有効な特徴ベクトルの抽出は難しい．そこで本研究では，各々のモダリティの性質を考慮して，最も認識に有効となるタイミングでそれぞれの特徴ベクトルを別々に抽出するこ

とを考える．

荷重と振動音は，切断時に特徴的な値を示すモダリティであるので，それぞれの時系列データから切断区間を検出することが必要となる．本研究では荷重データを用いてこの切断検出を行う．切断検出により得られる食材切断区間 T において，その区間に対応する荷重データから 10 次元の荷重特徴ベクトルを抽出する．振動音については，食材切断区間 T の中でも，包丁が食材に侵入してからまな板に衝突するまでの間において，特に有効な特徴ベクトルが抽出できると考えられる．そこで，区間 T の中から包丁とまな板が衝突する瞬間を探索し，その直前 0.2 秒分のデータから 16 次元の特徴ベクトルを抽出する．画像については，切断直前に食材がまな板の上に置かれた瞬間に注目し，その瞬間の画像における食材領域の色を 64 次元の特徴ベクトルとして抽出する．

以上のように各モダリティの特徴ベクトルを抽出し，各々を統合したときの食材認識精度を実験的に評価し，各モダリティを単独で用いた場合の精度と比較した．実験対象食材としては一般家庭でよく使われる被切断食材 23 種を用い，また，各特徴ベクトルの統合方法としては Ivanov らの手法を用いた．この結果，画像単独では同じ深緑色のキュウリとピーマンを識別できなかったが，振動音・荷重と統合することで，キュウリをピーマンと誤認識する割合，ピーマンをキュウリと誤認識する割合が，それぞれ 38.5%と 61.1%から 6.5%と 3.8%に改善された．他にも，荷重ではダイコンとニンジンの誤認識が発生しており，ダイコンをニンジンと誤認識する割合，ニンジンをダイコンと誤認識する割合はそれぞれ 12.9%と 9.0%であったが，統合により，それぞれ 0.3%と 0.0%に改善された．このように，各モダリティの問題点が他のモダリティにより補われることにより，誤認識が改善されていることが確認できた．

本研究では，食材一個体に対して行われる複数の切断一つ一つに対して，それぞれ別々に特徴ベクトルの抽出・統合および認識を行った．しかし，同一食材・同一個体であっても，切断の仕方によっては，特に振動音と荷重の特徴ベクトルに差が出る可能性がある．この問題への対処は今後の課題の一つである．例えば，同一個体に対する複数の切断から，特に認識に有効な特徴ベクトルが抽出できるような切断を選択する，などの対処法が考えられる．

Igredient Recognition using a Combination of Image, Vibration Sound and Load Data

INOUE Jin

Abstract

Nowadays, many homemakers use a web service in which they can publish their original recipes. The web service allows such personal users to publish their recipes easily. However, describing recipes requires great care. For helping the users to describe recipes, a system which can describe the recipes automatically through observing food preparation is useful. Since a recipe is a set of directions, each of which consists of culinary task and its target ingredient, it is important for that system to recognize the ingredients and the tasks through observing food preparation. In this paper we focus on the ingredient recognition.

Many of related works use *Image* data as an input of the ingredient recognition process, and some others use *Vibration Sound* and *Load* data at the cook's cutting actions. Although these modalities are used separately in the related works, each modality has its own drawback, and fails in the recognition. This can hardly be solved sololy, but would be solved with other modalities mutually. For example, *Image* is not suitable to distinguish cibols and radishes because they have similar colors, whereas *Load* suits because of their different hardness. By contraries, *Load* hardly discriminates radishes from carrots, whereas *Image* can. In this paper, we propose an ingredient recognition method by combinative use of these three modalities, which solves misrecognition caused by drawbacks from each modality.

Combining multi-modal time series data simply has the following problem: Each modality's data has effective parts and ineffective parts in its time series, which are different from those of the other modalities. For instance, the performance of *Vibration Sound* and *Load* will become extremely high when the cook cuts the ingredients, whereas that of *Image* will become low at that time because the ingredients are hidden by the cook's hands and their colors cannot be observed. Therefore, in this paper we separately extract the effective parts of each modality and combine them.

Since the performance of *Load* and *Vibration Sound* will become high when

the ingredients are cut, the cook's cutting actions should be detected first. In this paper, the detection is achieved by using *Load* data and an interval T corresponding to the cutting part is gotten, from which 10-dimensional vector of *Load* feature is extracted. The interval T is also used for extracting *Vibration Sound* features as follows: A moment t_c when a knife and a cutting board collide with each other is detected in the interval T, and 16-dimensional vector of *Vibration Sound* feature is extracted from the interval $T_S = [t_c - 0.2,\ t_c]$. As for *Image*, we focus on a moment when the cook puts the ingredient on the cutting board before cutting it, and extracted 64-dimensional vector of *Image* feature at the moment.

To examine the effectiveness of the proposed method, we conducted an experiment, in which we evaluated the recognition accuracy of the proposed method and compared the accuracy to those of uni-modal methods only using *Image*, *Vibration Sound* or *Load* data, respectively. 23 kinds of ingredients which often appear in ordinary home cooking were observed and the data obtained by the observation was processed by Ivanov's combination method. The results are as follows: By *Image*-based method, green bell peppers were often misrecognized as cucumbers and the misrecognition rate was 61.1%. However the rate was improved to 3.8% with the proposed method which combined *Vibration Sound* and *Load* with *Image*. In addition, radishes and carrots were sometimes misrecognized as each other by the *Load*-based method, and a rate of misregonizing radishes as carrots was 12.9%. With our method, the rate was improved to 0.3%. A rate of misrecognizing carrots as radishes was also improved from 9.0% to 0.0%. These results indicated that the misrecognition of each modality is successfully solved by the proposed multi-model method using the complementary relation between the three modalities.

The proposed method currently extracts two or more feature vectors from a single individual ingredient for each modality. However, some of them sometimes have largely different from each other due to a variety of cutting ways, especially for *Vibration Sound* and *Load* modalities. To solve this problem is one of the feature works. One solution will be to select the most effective features among the whole features extracted from each single ingredient.

画像・振動音・荷重データを統合的に用いた食材認識

目次

第1章　緒論

従来，料理のレシピは，主に専門家により作成され，本やテレビ番組を通して公開されてきた．しかし，近年，インターネットの普及により，一般の主婦等が自作のレシピを web 上に公開出来るようなサービスが展開されている．これらのサービスは，ユーザ側に本やテレビ番組を制作するような金銭的コストが掛からず，様々な人に自作のレシピを比較的簡単な手順で公開できるため，主婦をはじめとする多くの一般ユーザから支持を得ている．例えば，レシピコミュニティーサイトの最大手であるクックパッド [1] では，月間利用者数 1,005 万人，登録レシピ数 92 万件，月間レビュー数 4 億回を達成している [2]．しかし，こういったサービスを利用してレシピを公開する際には，ユーザは自ら調理の手順やポイントを文章化し，画像を付け加えるなどしてレシピを作成する必要がある．レシピの作成に必要となるこれらの工程は，レシピの公開までに必要な工程の大部分を占め，かつ手間のかかるものであるため，ユーザのレシピ公開を妨げる要因となり得る．これに対し，調理観測データからそのレシピを自動で生成することができれば，上述のレシピ作成作業の手間を省くことができ，一般ユーザによるレシピの公開が促進され，クックパッドのようなサービスがより活発になることが期待される．

調理観測データから自動的にレシピを生成するためには，その構成要素となる情報を調理観測データから取得する必要がある．山肩らによれば，レシピにはその料理中に行われるべき単位作業が列挙されており，各単位作業は，加工対象食材の名前と加工動作の名前の対により記述される [3]．これに従えば，レシピの構成要素となる情報の一部として，加工対象食材および加工動作の名前を挙げることができる．つまり，レシピを自動で生成できるようなシステムの実現には，そのための基盤技術の一つとして，加工対象食材と加工動作を調理観測データから認識することが必要となる．本研究では，このうち加工対象食材（以下，食材）の認識に注目する．

調理観測データからの食材認識を目的とした研究は従来からいくつか行われており [3, 4, 5, 6]，各研究ごとに様々なモダリティの利用が検討されている．食材を切断する前後の食材色を特徴量とした山肩らの研究 [3] や，料理番組などで食材がアップになったタイミングにおける食材色を特徴量とした柴田らの研究 [4] では，画像や映像といったモダリティが利用されている．他にも，食材切断

時に発生する振動音を利用した三功らの研究 [5]，切断時にまな板にかかる荷重を利用した土本らの研究 [6] がある．

上記の従来研究のように，何か一つのモダリティのみを用いた場合には，相異なる種類の食材から似たような特徴ベクトルが得られてしまう，ということが少なからず起こる．例えば，画像を利用して食材色を特徴量とする場合，ピーマンとキュウリは共に深緑色であり似たような特徴ベクトルが得られるため，その識別は難しい．このような食材組を識別することは，単一のモダリティを利用するだけでは原理的に難しく，誤認識は避けられない．実際，荷重を単独で利用した土本らの研究では，硬さの似た食材同士の識別が上手くできていないことが実験結果から分かる．このように，それぞれのモダリティには，それぞれ異なる問題点により，誤認識が発生しやすいような食材組が存在すると言える．上記の問題に対して，画像を利用した山肩ら・柴田らの研究では，それぞれレシピ情報・映像中のクローズドキャプションを食材色と併用することでこれに対処した．しかし，一般家庭で観測されたデータからレシピを生成することを目指す上では，これらの情報を利用することは適当ではない．

三功らが用いた振動音では，肉を切る場合など，あまり強い音が発生しないためその観測が難しい場合もある．このような問題も，振動音を単独で利用するだけでは対処することが困難である．

そこで本研究では，従来よりそれぞれ別々に用いられてきた画像・振動音・荷重の 3 つのモダリティを併用することで，個々のモダリティを単独で用いた場合に発生する誤認識が改善されるような食材認識の実現を目指す．

画像・振動音・荷重の 3 つを併用する際には，それぞれのモダリティに対応する時系列データのうち時刻の一致している部分同士を単純に併用しても認識に有効とはならない．例えば，振動音および荷重の時系列データでは食材切断時に相当する部分が特徴的な値を示すが，画像については，食材切断時には対象食材が調理者の手や包丁で遮蔽されるため，この時刻に相当する部分から食材色を観測することは難しい．また，荷重では食材切断時に相当する部分全体が特徴的であるが，振動音では，包丁が食材内を通過する瞬間のみが特徴的であり，その前後に相当する部分からは認識に有効な特徴ベクトルの抽出は難しい．この問題に対処するために，本研究では，各モダリティに対応する時系列データから最も認識に有効な部分をそれぞれ別個に切り出し，これらを併用することを考える．

本論文の構成は以下のとおりである．まず，2章では，食材認識に関する従来研究について概観する．次に3章では，画像・振動音・荷重に対応する時系列データからそれぞれどの部分を切り出すべきかについて議論し，その内容を踏まえ，各データの具体的な統合方法を提案する．4章では，実際の食材を観測することで画像・振動音・荷重の時系列データを実際に取得し，得られたデータを用いて，各モダリティ単独での認識手法で発生する誤認識が提案手法により改善されているかどうかを実験的に検証する．最後に5章で結論と今後の課題を述べる．

第2章　食材認識に関する従来研究

最も一般的な食材認識手法として，カメラで撮影した調理画像や映像から食材領域を抽出し，その領域から得られる特徴ベクトルを用いる手法がある．特徴ベクトルの具体的な例としては，食材領域の代表色やテクスチャ，形状などが挙げられる．山肩らは，食材に対して「切る」「剥く」等の加工動作が行われる際に観測される食材の外面色および内部色を特徴ベクトルの一つとして用い食材認識を試みている [3]．また，柴田らは，調理映像中における注目領域として食材がアップで映っているシーンにおける食材領域を求め，その領域の代表色を特徴ベクトルの一つとして用い食材認識を行っている [4]．

食材色を特徴ベクトルとして用いるこれらの手法には，色が似通っている食材同士を区別することが困難である，という根本的な問題点がある．この問題は，食材色のみを用いる手法では解決が難しい．そこで，山肩らは加工動作の情報を利用して認識の対象となる食材の候補を絞り込むことで認識精度の向上を図っている．すなわち，レシピ中で今対象としている加工動作と対になって現れる食材のみに認識結果の候補を絞り，その候補の中から食材色に基づいて単一の認識結果を返す，というものである．しかし，加工動作情報を利用するというこの対処法は，食材 - 加工動作の対としてあり得る組み合わせが明示的に記載されているレシピの存在を前提にしており，本研究のようにレシピ自動生成システムへの応用を想定した食材認識手法を目指す上では，現実的な対処法とはならない．一方，柴田らの手法では，調理映像に付与されているクローズドキャプションを食材色と統合して利用することにより認識精度の向上を図っているが，この手法は料理番組のように元々映像にクローズドキャプションが

付与されている状況でのみ有効な方法であり，主婦等の一般ユーザが自分で自分の調理風景を撮影するような状況を対象として想定している本研究にはそぐわない．

画像とは別種のモダリティを利用して食材認識を試みた従来手法もいくつか提案されている [5, 6]．一例として，三功らは，食材が切断される際にまな板付近で生じる振動音をコンタクトマイクで観測し，これを利用して食材認識を試みている [5]．また，土本らは，食材切断時にまな板にかかる荷重を計測するためのセンサボードを開発し，このボードから取得される荷重のデータを用いて食材認識を試みている [6]．三功ら，土本らはともに食材の切断時に着目しており，そのタイミングで特徴ベクトル取得および認識の処理を実行している．これは，食材を切断する際には，食材の内部構造に依存する切断音の発生や食材の硬さ・大きさ等に依存する圧力の発生，といったように，食材に関する様々な性質が様々な形で抽出できるためである．

食材切断時に食材内部を包丁が通過することで発生する振動音は，食材の内部構造を反映したモダリティであると考えられるため，層構造のタマネギと均一構造のジャガイモといったように，内部構造の明らかに異なる食材同士の識別に有効である．その一方で，トマトのように時間がたつと内部構造が大きく変化していく食材の認識に対しては，あまり有効とはならない．また，肉のように柔らかい食材を切断する場合には，あまり大きな振動音が発生しないため，認識のための特徴ベクトルが獲得できないこともある．硬さや切り難さといった食材の性質を反映したモダリティである荷重も，ダイコンと白ネギのように硬さが大きく異なる食材同士の識別には有効であるが，ニンジン，ダイコン，ショウガといった硬さが似通った食材同士の認識率は極端に下がってしまう結果となっている．

このように，三功ら・土本らの手法はある一定の成果を示しているが，振動音または荷重という単一モダリティのみを用いた認識手法であるため，各モダリティがもつ問題点により誤認識が発生している．このことは，それぞれの研究で報告されている実験結果からも確認できる．

第3章　画像・振動音・荷重データの統合利用

3.1　複数モダリティを用いた食材認識の検討

前述したように，従来手法で用いられてきた各モダリティにはそれぞれに問題点があり，その問題点が誤認識を発生させる要因の一つとなっている．これらの問題点を何らかの方法で解決できれば，より精度の高い認識手法が実現できると強く期待される．しかし，単一のモダリティを用いるのみではそのモダリティが持つ問題点を解決することは難しい．そこで本研究では，複数のモダリティを併用することで各々の問題点を相補的に解決することを考える．以下では，2章で挙げた画像・振動音・荷重の相補性について考察する．

画像から抽出される食材色のみを用いた場合，色が似通った食材同士の識別が困難となる．例えば，キュウリとピーマンは共に深緑色をしており，食材色のみでこの2つを識別することは難しい．しかし，この2つは．振動音を利用すればうまく識別できると期待される．つまり，食材の内部構造を反映していると考えられる振動音では，均一な内部構造をもつキュウリを切断したときと層状の内部構造をもつピーマンを切断したときとで，得られる特徴ベクトルに大きな違いがあり，この違いにより両者をうまく識別できると考えられる．同様に，ダイコンと白ネギの組も，共に白色をしていることから画像による両者の識別は困難である．しかし，この2つは硬さに大きな違いがあり，食材の硬さを反映していると考えられる荷重を用いれば容易に識別できると期待される．これらの事例は,「色が似通っている食材同士の識別が困難である」という画像の根本的問題が，食材切断時の振動音や荷重を用いることでうまく解決できる可能性がある，ということを示唆するものである．この一方で，例えばジャガイモとナスのような共に均一な内部構造をした食材の識別は，振動音では困難であるが，食材色を利用すれば容易である．同様に，ダイコンとニンジンのような硬さの似た食材同士は，荷重を用いるだけでは誤認識が十分に起こり得るが，画像を用いれば容易に識別できると考えられる．以上のことから，画像を用いる認識手法の問題点が振動音・荷重を利用することで解決できるというだけでなく，振動音・荷重では識別が難しい食材同士でも画像（食材色）の情報を利用することでその識別が可能になる，という可能性も示唆される．

振動音と荷重においては，この2つのモダリティは食材の性質に注目しているという点で，似た種類のモダリティと言える．しかし，包丁が食材を切断・破

壊する際に取得される振動音は食材の内部構造をよく反映している一方で，切断時にまな板にかかる荷重は食材の硬さをよく反映している，という違いがある．例えば，タマネギのような層状の内部構造をもつ食材を切断する場合，層の切れ目を包丁が通過する瞬間には，均一構造を持つ食材からは獲得できない何らかの特徴ベクトルが獲得できるはずであるが，このようなものの抽出には内部構造を反映したモダリティである振動音が向いている．一方で，硬さのみが異なるような食材組に対しては，微妙な硬さの違いを捉えることができる荷重が大きな役割を果たすと期待される．つまり，振動音と荷重は似たモダリティではあるが，各々に異なる特性があると考えられる．

以上のように，画像・振動音・荷重のモダリティにはそれぞれ異なる性質が備わっていることが分かる．各々のモダリティを単独で用いた場合には，その性質が問題点となって誤認識が発生してしまう．しかし，3つのモダリティに対応する画像・振動音・荷重の時系列データを統合利用すれば，各モダリティが相互に問題点を補い合うことにより，このような誤認識を改善できる可能性が高い．ただし，画像・振動音・荷重の3つのデータを統合利用するときに，それぞれの時系列データのうち単に時刻が一致している部分を抽出・併用しても，それらから得られる特徴ベクトルが全て認識に有効なものであることは期待できない．例えば，振動音・荷重では食材切断時に相当する部分が特徴的な値を示すが，画像から得られる食材色の特徴ベクトルについては，食材切断時には対象食材が調理者の手や包丁によって遮蔽されるため，この時刻において安定した特徴ベクトルを獲得することは難しい．また，荷重は食材切断時に相当する部分全体が特徴的であるが，振動音は，包丁が食材内部を通過している瞬間のみが特徴的であり，その前後から認識に有効な特徴ベクトルを抽出することは難しい．

この問題に対して，本研究では画像・振動音・荷重それぞれのモダリティの特性を考慮して，各々に対応するデータから最も認識に有効な部分をそれぞれ別個に切り出し，それらを併用する．3.2節では，この具体的な方法について述べる．

3.2　各データの性質に応じたタイミングでの特徴ベクトルの抽出

3.2.1　画像・振動音・荷重データの取得方法

画像・振動音・荷重の具体的な併用方法を述べる前に，まず，各々のモダリティに関する時系列データの取得方法について記載する．

荷重

荷重データの取得には，土本らが開発した荷重センサボードを用いる．荷重センサボードの上にまな板を置き，この上で食材の切断を行う．これにより，調理開始時から $t(\in \mathbb{R})$[1)]秒後の時点においてまな板にかかっている荷重 $F(t)$ を各時刻 t において取得する．食材切断時には，この荷重 $F(t)$ が特徴的な値を示す．

画像

調理を行っている様子を調理台の真上からでカメラで撮影する．これにより，調理開始時から $t(\in \mathbb{R})$ 秒後におけるまな板周辺の作業台の様子を映した静止画像 $I(t)$ を各時刻 t において取得する．この画像 $I(t)$ から，画像特徴ベクトルを抽出する．

振動音

振動音を用いた三功らの手法では，キッチンの作業台を振動音の伝搬しやすい硬化ガラス製にしてその裏に直接コンタクトマイクを貼り付け，この作業台の上に薄いまな板を置いて食材を切断することで振動音データを取得していた．しかし，本研究では，前述したように，荷重データ取得のために土本らの開発した荷重センサボードをまな板の下に置く必要がある．そのため，三功らのように硬化ガラスの下に取りつけたマイクで振動音を取得すると，荷重センサボードを伝搬した上で硬化ガラスに伝わった振動音が取得されてしまう．土本らは振動音を取得することを想定していないので，当然荷重センサボードは振動音が伝搬しやすい設計とはなっていない．このようにして取得された振動音が，食材認識に有効な情報を含んでいるとは考えにくい．

そこで本研究では，ガラス製のカッティングボード（Joseph Joseph 社 カッティングボード ミニモザイク 角型 30×30cm）をまな板として利用し，このまな板の裏に直接コンタクトマイクを貼り付けた．これにより，荷重ボー

[1)] 手法説明の便宜上，ここでは t が連続量であるとして議論を進める

ドを伝搬することなく，直接振動音を取得することが可能となり，三功らの実験環境にできる限り近い環境が構築できる（図 1 参照）．

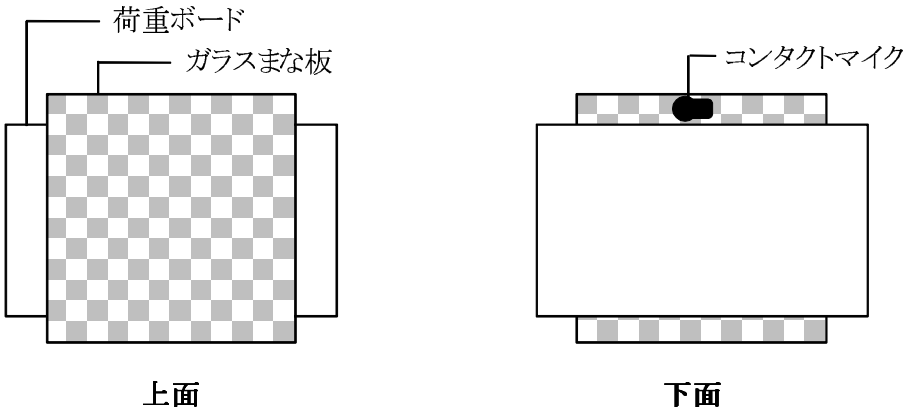


図 1: 荷重センサおよびガラス製まな板

コンタクトマイクについても，三功らが用いたものと同じ EDIROL 社製のマイクを使用する．以上のような環境の下，調理開始時から $t(\in \mathbb{R})$ 秒後の時点においてまな板上で発生している振動音の振幅値 $S(t)$ を各時刻 t において取得する．食材切断時には，この振幅値 $S(t)$ から特徴的な値が抽出できると考えられる．

3.2.2 特徴ベクトルの抽出タイミングの策定

3.1 節で述べたように，画像・振動音・荷重データを統合する際，全く同じ時刻 t において $I(t)$，$S(t)$，$F(t)$ のそれぞれから特徴ベクトルを抽出してそれらを併用しても，認識に有効とはならない．従って，各モダリティの性質を考慮した上で，適切なタイミングで特徴ベクトルを抽出することが必要となる．本研究で用いる 3 つのモダリティうち特に振動音と荷重は，食材切断時において特徴的な値が得られるモダリティであるため，それぞれのデータから食材切断時に相当する部分を検出することが必要となる．切断検出については，土本らの荷重を用いる手法により高い検出精度が実現されている [6]．本研究でもこの切断検出法を利用する．具体的な手順は以下の通りである．

食材切断時には，包丁が食材に押しあてられた段階（食材に侵入する少し前）で荷重 $F(t)$ が急激に大きくなり，切断が終わり食材から包丁が離れると $F(t)$ は急激に小さくなり切断前の値に戻る．そこでまず，$\frac{\mathrm{d}}{\mathrm{d}t}F(t) < -\theta_{\mathrm{c}}$ かつ $\frac{\mathrm{d}^2}{\mathrm{d}t^2}F(t) = 0$

が満たされる時刻（急激に荷重 $F(t)$ が立ち下がる時刻）t_{c} を検出する．この t_{c} に対して，$t \geq t_{\mathrm{c}}$ かつ $\frac{\mathrm{d}}{\mathrm{d}t}F(t) \geq -\theta_{\mathrm{e}}$ となる最初の時刻 t を切断終了時刻 t_{e} とする．t_{e} は荷重 $F(t)$ が切断前の値に戻る時刻に相当する．最後に，$t < t_{\mathrm{c}}$ かつ $|F(t_{\mathrm{e}}) - F(t)| < \delta$ かつ $\frac{\mathrm{d}}{\mathrm{d}t}F(t) > \theta_{\mathrm{s}}$ となる時刻のうち最も t_{e} に近いものを切断開始時刻 t_{s} とする．t_{s} は荷重 $F(t)$ が急激に立ち上がり始める時刻に相当する．正定数 $\theta_{\mathrm{c}}, \theta_{\mathrm{s}}, \theta_{\mathrm{e}}, \delta$ は経験的に決定された閾値である．

以上の手順により，荷重データを基に食材切断時に対応する時間区間 $T = [t_{\mathrm{s}}, t_{\mathrm{e}}]$ が自動抽出される．以下では，区間 T を基準として，各モダリティの性質を考慮した特徴ベクトルの抽出タイミングおよび抽出方法について述べる．

荷重

抽出された区間 T に対応する荷重 $F(t)(t_{\mathrm{s}} \leq t \leq t_{\mathrm{e}})$ から 10 次元の荷重特徴ベクトル L_T を計算する．この 10 次元特徴ベクトルは，$F(t)$ の平均，平均偏差，分散，標準偏差，歪度，尖度の 6 次元（付録参照），および土本らの提案した最大値・切断時間・力積・ピーク位置の 4 次元からなる．

画像

本研究では，複数のモダリティを併用した食材認識を行うため，食材を一回切断するごとに獲得できる振動音・荷重特徴ベクトルと対になる画像特徴ベクトルが必要になる．このような特徴ベクトルの最も単純な抽出タイミングとしては，切断が行われる瞬間の時刻 t_{m} が考えられる（ここで，t_{m} は区間 T に対して，$t_{\mathrm{s}} < t_{\mathrm{m}} < t_{\mathrm{e}}$ を満たす）．切断が行われる瞬間 t_{m} の画像 $I(t_{\mathrm{m}})$（の食材領域）から抽出された特徴ベクトルであれば，その抽出元の食材が振動音・荷重特徴ベクトルの抽出元食材と同一であることが保証される．しかし 3.1 節で述べたように，食材を切断する瞬間には調理者の手や包丁で食材が遮蔽されることが多く，食材色に関する特徴ベクトルの安定的な獲得が困難である．

そこで本研究では，切断の直前に食材がまな板上に置かれる瞬間に注目する．食材は切断が施される前に必ずまな板の上に置かれるので，切断の直前に置かれた食材は，切断対象の食材と一致する．つまり，この瞬間に抽出される画像の特徴ベクトルは，その直後に起きる切断から抽出される振動音・荷重の特徴ベクトルと対になる．また，食材をまな板に置いたときであれば，調理者の手などによって食材が遮蔽されることは少なく，安定した特徴ベクトルの抽出が期待できる．

以上のような考えの下，本研究では，橋本らのTexCut[7, 8]を用いて，まな板に物体が置かれた時刻$\tau_i(i = 1, 2, \cdots)$の集合$\{\tau_i\}$を特定・列挙するとともに，それらの時刻τ_iにおける画像$I(\tau_i)$から，置かれた物体の領域$I'(\tau_i) \subset I(\tau_i)$を抽出する．その上で，時刻集合$\{\tau_i\}$の中から切断区間$T$の直前に相当する時刻$\hat{\tau}$を次の方法で特定し，この$\hat{\tau}$に対応する食材領域画像$I'(\hat{\tau})$を対象として画像特徴ベクトルを抽出する．$\hat{\tau}$は$\hat{\tau} = \max\{\tau | \tau \in \{\tau_i\}, \tau < t_\mathrm{s}\}$として求める．

画像データから上記のようにして得られた食材領域画像$I'(\hat{\tau})$に対して，RGB値（それぞれ0〜255）をそれぞれ4ビン（0〜63，64〜127，128〜191，192〜255）に区切った食材領域画像$I''(\hat{\tau})$を新たに生成する．この食材領域画像$I''(\hat{\tau})$に対して$4 \times 4 \times 4 = 64$ビンの色ヒストグラムを作成し，これを64次元の食材色特徴ベクトル$\boldsymbol{C}_{\hat{\tau}}$とする．

振動音

振動音も，荷重と同様に，食材切断時に特徴的な値を示すモダリティである．よって，振動音データからも食材の切断時に相当する区間を検出することが必要となる．これに関して，振動音を用いた三功らの手法では，この切断検出を振動音のみから行おうとしているが[5]，物を置く音など切断時以外に発生する音も存在するため，振動音のみを用いて切断部分を正確に検出することは難しい．そこで本研究では，荷重を用いて検出された食材切断区間Tを振動音の場合にも用いる．

荷重により検出された区間Tは，切断に相当する区間全体を検出したものである．一方，振動音で最も食材の特徴を反映しているのは，食材の内部を包丁が通過している時に取得できる振動音であり，その前後の振動音は特段有効なものではない．よって，区間Tから更に振動音特徴ベクトルの抽出用区間T_Sを取り出す必要がある．そのために，「包丁が食材内部を通過し終えた直後にまな板と衝突することで非常に大きな振動音が発生する」という傾向を利用する．具体的には，以下のように区間T_Sを取り出し，その区間から，三功らの手法を基に振動音特徴ベクトルを抽出する．

まず，荷重により検出された切断区間$T = [t_\mathrm{s}, t_\mathrm{e}]$の範囲内から，振動音の振幅値$S(t)$が最大値を示す時刻$t_\mathrm{p}$を探索する．これは$t_\mathrm{p} = \arg\max_t |S(t)| (t_\mathrm{s} \leq t \leq t_\mathrm{e})$とすることで実現できる．この時刻$t_\mathrm{p}$からまな板と包丁の衝突が始まると考えて，これ以前の約0.2秒分（サンプリング周波数44.1kHzにお

いて 8832 サンプル分）の区間 $T_{\mathrm{S}} = [t_{\mathrm{p}} - 0.2, t_{\mathrm{p}}]$ を特徴ベクトル抽出用に取り出す．

区間 T_{S} に相当する部分の振動音波形に対して，ハミング窓を用い周波数解像度 128，オーバーラップ 50%でスペクトログラム分析を行う．これにより得られたスペクトログラムの低周波成分 8 次元の平均と分散をそれぞれ求め，計 16 次元のベクトルを得る．ここで，この分散値は平均値と比較して，非常に大きいので，平均値と分散値の両方に対して 10 を底とする対数をとる．こうして得た 16 次元のベクトルを振動音特徴ベクトル $\boldsymbol{V}_{T_{\mathrm{S}}}$ とする．

図 2 に，各モダリティの性質を考慮した特徴ベクトルの抽出区間を図示する．以上のようにして得られた $\boldsymbol{C}_{\hat{\tau}}, \boldsymbol{V}_{T_{\mathrm{S}}}, \boldsymbol{L}_T$ を併用して，食材認識を行う．次節では，実際の食材に対する食材認識実験に先立って，荷重に深くかかわる切断方法の制約について考察する．

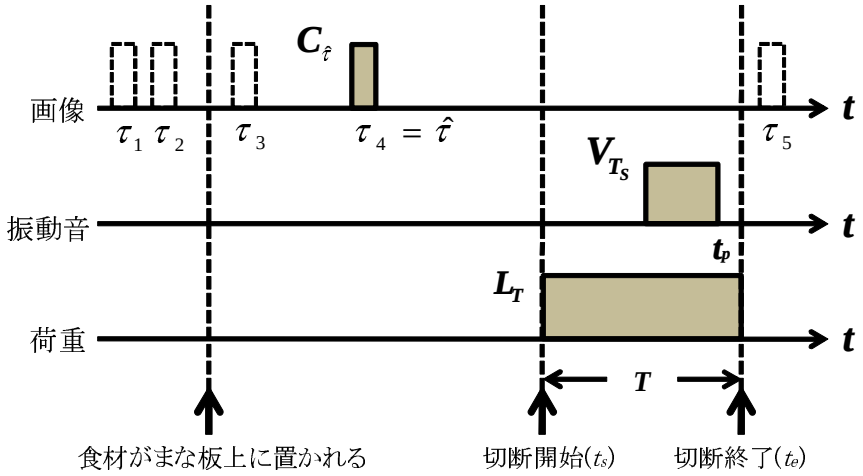


図 2: 各モダリティの特徴ベクトルの抽出区間

3.3 切断方法の違いが荷重データに与える影響の調査

土本らは，同一食材であっても，切断方法が異なれば，得られる荷重データにも大きな違いが現れる可能性を指摘している．実際，土本らは，切断方法に一定の制約を加え，これを統一した条件下で認識実験を行っている．しかし，実際の調理においてこのような制約を仮定することは不自然である．

切断方法の違いは，食材のどの部位（切断部位）にどのような角度（切断角度）で包丁を通すかによって生じると考えられる．また，切断部位の違いは切断面の面積の違いに直結すると考えられる．そこで本節では，切断面積および切断角度を変化させながら実際の食材を十〜数十回切断し，その際に得られた荷重データを用いて荷重単独による食材認識を試みることで，切断方法の違いが荷重データに与える影響の程度を検証する．

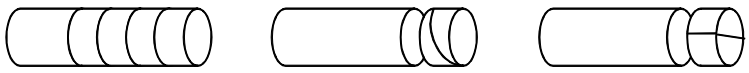
検証には，比較的硬さが似ており，また様々な面積・角度での切断が一般的に行われ得るような食材を用いる．硬さの違いは荷重に関する特徴ベクトルの違いとして現れると考えられるため，硬さが似ている食材群からなる食材集合を切断する際に観測される荷重特徴ベクトルは，食材の種類の違いによるばらつきが比較的小さいものとなっているはずである．このような，特徴ベクトルのばらつきが比較的小さくなるような食材集合を様々な面積・角度の下で切断し，その際に得られた特徴ベクトルを用いて認識処理を試みた場合でも，認識率が十分高ければ，切断方法の違いが荷重特徴ベクトルに与える影響は無視できるものと考えてよい．また，上記の検証の結果，認識率が低い値を示す食材組が得られたとしても，それらの間の誤認識が切断面積や角度の違いによらず一様に発生していることが確認できれば，それは切断方法の違いではなく，もともと誤認識が起こりやすい（似た特徴ベクトルが得られやすい）食材組であったことが原因と考えられる．従って，この場合も，切断方法が与える影響は無視できると言える．

このような検証を行うために，ジャガイモ・キュウリ・ダイコン・ニンジン・ゴボウの 5 種類を実験対象食材とする認識実験を試みた．各食材に対して行った切断方法を表 1 に示す．

表 1: 5 種類の食材に対する切断方法

	切断面積変化検証用	切断角度変化検証用
ジャガイモ	薄切り	—
キュウリ	斜め切り	縦・横・斜めに切る
ダイコン	薄切り	縦・横・斜めに切る
ニンジン	横切り	縦・横・斜めに切る
ゴボウ	斜め切り	縦・横・斜めに切る

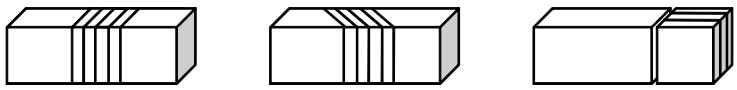
表 1 における「切断角度変化検証用」の切断方法に関して，キュウリは，それぞれの個体を約 2.5cm 間隔で横切りしていき（縦の切断），横切りされたキュウリ片の筋に対して 45 度（斜め）または 90 度（横）に包丁を入れることで，切断面積を一定に保ちつつ切断角度のみを様々に変化させた（図 3 参照）．ゴボウも，約 2cm 間隔で横切りしたあとに同様の切断を行うことで，切断角度のみを様々に変更した．



縦（0度）　斜め（45度）　横（90度）

図 3: キュウリ・ゴボウにおける切断角度の変化のさせ方（上から見た図）

ダイコンに関しては，事前に，(1) 底面が約 5cm × 5cm の直方体（スティック状），(2) 底面が約 4cm × 4cm の直方体（同），(3) 一辺が 5cm の立方体となるようにダイコンを加工し，(1) のダイコンの筋に対して 0 度（縦），(2) のダイコンの筋に対して 45 度（斜め），(3) のダイコンの筋に対して 90 度（横）に包丁を入れることで，切断面積を一定に保ちつつ切断角度のみを変化させた（図 4 参照）．ニンジンに対しても，事前に，底面が約 3cm × 3cm の直方体，底面が約 2.5cm × 2.5cm の直方体，一辺が約 3cm の立方体となるようにニンジンを加工し，その後ダイコンの場合と同様の切断を行うことで，切断角度のみを変化させた．



縦（0度）　斜め（45度）　横（90度）

図 4: ダイコン・ニンジンにおける切断角度の変化のさせ方（手前から見た図）

ジャガイモについては，他の 4 種類の食材とは異なり，その形状が球に近いため，食材に対して入れる包丁の角度が判断しにくい．そのため，切断角度に関する考察が困難と考え，切断面積の影響のみを検証した．

以上のような切断過程を観測することで獲得した荷重データを単独で用いて

食材認識を試みた．認識の際に用いた特徴ベクトルの種類およびその具体的な抽出方法については，前節で述べた内容に従った．この結果を表 2 に示す．

表 2: 切断面積・角度を変えながら荷重単独で食材認識を行った際の認識率 (%)

	ジャガイモ	キュウリ	ダイコン	ニンジン	ゴボウ
ジャガイモ	85.32	4.59	2.75	1.83	5.50
キュウリ	2.24	95.07	0.00	0.00	2.69
ダイコン	3.70	0.00	64.35	13.43	18.52
ニンジン	2.62	1.05	10.99	65.45	19.90
ゴボウ	5.58	1.72	6.87	11.16	74.68

表 2 のうち，ジャガイモとキュウリに関しては，面積変化・角度変化に関係なく，高い認識率を示した．一方で，ダイコン，ニンジン，ゴボウは比較的認識率が低くなっている．ダイコン・ニンジン・ゴボウについて，その誤認識の内訳を表 3 に示す．

表 3 において，まず切断面積の変化に注目する．ダイコン，ニンジン，ゴボウのような硬い食材では，切断面積が小さくなるにつれ，誤認識した食材名にキュウリなどの柔らかい食材が含まれるようになった．この結果は，荷重特徴ベクトルに対する切断面積の影響と食材の硬さ（これは食材の種類に依存）の影響との間にある程度の類似性が存在しているということを示唆している．従って，切断面積の違いが荷重特徴ベクトルに全く影響を与えないとまでは言い切れない．しかし，誤認識率自体はおおよそ 30%程度に抑えられているので，その影響の大きさは限定的であると言える．一部には，ダイコン，ニンジンの「面積小」のように，誤認識率が比較的高くなってしまっているものも存在するが，これらについても，誤認識した食材名を見れば，食材色が大きく異なる組み合わせであることが多いため，食材色を併用した認識を行うことにより，誤認識率を十分に低減できるものと考えられる．

次に，切断角度の変化に注目する．すじが多いごぼうは，0 度（縦）で切断する時にはすじを真っ二つにする必要があり，大きな力が必要となる．そのためダイコン・ニンジンという硬い食材との誤認識が多かった．ゴボウをすじに沿って横に切断した場合（90 度）は，小さな力で切断ができるため，キュウリ・ジャガイモといった柔らかい食材との誤認識が多かった．すじの少ないダイコ

表 3: 表 2 におけるダイコン・ゴボウ・ニンジンの誤認識の内訳

ダイコン	誤認識	切断総数	誤認識率	誤認識した食材名
面積小	38	54	70.37	ジャガイモ，ニンジン，ゴボウ
面積中	25	65	38.46	ニンジン，ゴボウ
面積大	7	28	25	ニンジン，ゴボウ
面積特大	0	14	0	—
0 度	2	19	10.53	ニンジン，ゴボウ
45 度	4	18	22.22	ニンジン
90 度	1	18	5.56	ゴボウ
ニンジン	誤認識	切断総数	誤認識率	誤認識した食材名
面積小	9	18	50	ジャガイモ，ダイコン，ゴボウ
面積中	7	23	30.43	ダイコン，ゴボウ
面積大	9	23	39.13	ダイコン，ゴボウ
0 度	20	52	38.46	キュウリ，ダイコン，ゴボウ
45 度	6	35	17.14	ジャガイモ，ダイコン，ゴボウ
90 度	15	39	38.46	ジャガイモ，ダイコン，ゴボウ
ゴボウ	誤認識	切断総数	誤認識率	誤認識した食材名
面積小	6	26	23.08	ジャガイモ，キュウリ
面積中	4	27	14.81	ダイコン，ニンジン
面積大	11	30	36.67	ダイコン，ニンジン
0 度	16	57	28.07	ジャガイモ，ダイコン，ニンジン
45 度	7	37	18.92	ジャガイモ，キュウリ，ニンジン
90 度	15	56	26.79	ジャガイモ，キュウリ，ダイコン，ニンジン

ン，ニンジンは，切断角度が変化しても，誤認識した食材名に変化は無く，切断角度と誤認識率の間にも明確な関連性は見られない．このことから，切断角度の変化は，基本的には誤認識率を大幅に変化させるほどの影響を荷重特徴ベクトルに与えることはないが，ゴボウにようにすじが多い食材に関しては，切断角度の違いが切断時の食材の硬さに変化を生じさせ，その結果荷重特徴ベクトルに多少の影響を与える，ということが分かる．しかし，面積変化に着目した場合と同様に，誤認識率自体は全体的にそこまで高くないことから，その影響の大きさは限定的であると考えられる．

以上の実験・考察から，荷重データは，多くの食材に関して切断面積の違いの影響を受けるのに加えて，すじの多い食材に限っては，切断角度の違いの影響も少なからず受けると言える．しかし，それらの影響は限定的であり，そこ

まで大きなものではない．また，切断面積や角度の違いの影響を大きく受ける一部のケースについても，画像から得られる食材色を併用することで十分に対処可能であることが期待できる．従って，複数のモダリティの併用を試みる本研究においては，土本らの従来研究 [6] で採用されていたような切断方法に関する制約を想定する必要はないと言える．

第 4 章　実際の食材を対象とした認識実験

4.1　実験対象とする食材

土本らは，妹尾が提案した食材相関図 [9] に基づいて抽出した頻出食材を認識実験に用いた．本研究でもこれを踏襲する．妹尾の食材相関図において，カテゴリ（和食，アジア食，洋食）を問わず頻出する食材に注目する．その中で，一般家庭での調理において切断が施される可能性が高いと考えられる食材（以下，被切断食材）を認識対象として用いた．具体的には，各カテゴリのレイヤー 1（頻出食材）・レイヤー 2（準頻出食材）における被切断食材（鶏肉・ホウレン草・しめじ・えのき茸・しいたけ・ゴボウ・ダイコン・ショウガ・白ネギ・ニラ・白菜・ニンニク・キャベツ・ピーマン・ナス・タマネギ・キュウリ・トマト・ブロッコリー・カボチャ），及び，カテゴリ分類はされていないが頻出食材である被切断食材（レモン，ニンジン，ジャガイモ）の計 23 種類の食材を用いた．

上記の食材集合に対して調理を行い，3.2 節で述べた方法に従って画像・振動音・荷重データを取得するとともに，認識に有効な特徴ベクトルを抽出した．以下の表 4 に食材ごとに得られたサンプル数をまとめた．

本研究では画像・振動音・荷重データを統合的に用いた食材認識を行っている．よって，これらのデータから得られた特徴ベクトルに何らかのパターン認識手法を適用し，最終的な認識結果を決定する必要がある．次節では，本実験で採用した Ivanov らのパターン認識手法 [10] について述べる．

4.2　実験に用いるパターン認識手法

個人識別や感情認識の分野では，複数のモダリティを統合することで認識精度の向上を図った研究が多数行われており [10, 11, 12, 13]，モダリティ統合のためのパターン認識手法も複数提案されている．これらのパターン認識手法は，主に feature-level fusion と decision-level fusion の 2 種類に大別される．

表 4: 本実験で用いた食材の種類と各食材から得られた特徴ベクトルのサンプル数

食材 ID	食材名	サンプル数	食材 ID	食材名	サンプル数
1	シイタケ	105	13	トマト	126
2	シメジ	116	14	ダイコン	286
3	エノキ	110	15	ジャガイモ	174
4	白ネギ	248	16	タマネギ	205
5	ショウガ	155	17	白菜	106
6	キュウリ	184	18	カボチャ	110
7	ホウレン草	113	19	ニンニク	506
8	レモン	128	20	ニンジン	158
9	ピーマン	237	21	ニラ	119
10	ナス	130	22	ブロッコリー	105
11	ゴボウ	159	23	鶏肉	251
12	キャベツ	112	—	—	—

feature-level fusion とは，各モダリティ $m \in \{\mathrm{m}_1, \mathrm{m}_2, \cdots\}$ に対応する特徴ベクトル $\boldsymbol{x}^m$ を連結することで連結ベクトル $\boldsymbol{x} = (\boldsymbol{x}^{\mathrm{m}_1\mathrm{T}}, \boldsymbol{x}^{\mathrm{m}_2\mathrm{T}}, \cdots)^{\mathrm{T}}$ を獲得し，この $\boldsymbol{x}$ を新たな特徴ベクトルとみなして認識を試みる方法である．これは，認識のための特徴空間を各々のモダリティに対応する特徴空間の直積空間として構成することに相当するため，認識のための特徴空間が高次元化する場合が多い．特徴空間が高次元の場合，安定した学習を行うために必要なサンプル数も大きくなるが，逆に十分な数のサンプル数があれば，良好な認識性能を示すことが多い．

一方，decision-level fusion とは，各々のモダリティ m に対応する特徴ベクトル $\boldsymbol{x}^m$ のみを用いてそれぞれの認識結果 C^m を求め，その後，各 C^m を基に最終的な認識結果 C を決定するという方法である．この方法は，モダリティ間の相関など一部の情報を無視するものであるため，feature-level fusion に比べ，若干認識性能が劣ることが多いが，比較的少数のサンプルでも安定した学習・認識が可能であるという利点がある．本実験では，表 4.1 に示したように，特徴ベクトルの次元数に対して十分な数のサンプルが得られなかった食材もあったことから，decision-level fusion に分類されるパターン認識手法を用いて認識実験を行う．

decision-level fusion の方法の一つとして，各モダリティの出力結果の尤度の和をとる統合方法 (sum) が提案されている．sum のような単純な方法には，どのような状況でも各々のモダリティを全て等価に扱ってしまうという問題がある．このため，認識結果の候補がある 2 つのクラス A，B に絞られているものの，その 2 つのクラスの何れであるかをあるモダリティ m_1 により判別することは難しい，といった状況でも，モダリティ m_1 に対応する認識結果 C^{m_1} を他のモダリティによる認識結果と同程度に重視してしまい，結果として認識精度が低下することがある [13]．このような decision-level fusion の問題点を回避する方法として，各モダリティを単独で用いて認識を行った際の誤認識率を基に各々のモダリティの重みを適応的に変更し（認識結果 C^m の値に応じて異なる重みを与え），その上で sum などにより統合する方法が Ivanov らによって提案されている [10]．本研究では，この Ivanov らの方法を用いて食材色・振動音・荷重の 3 つに対する decision-level fusion を行う．以下，その詳細について述べる．

モダリティ $m \in \{\text{image, vibration sound, load}\}$ に関する特徴ベクトルを $\boldsymbol{x}^m$ とおく．$\boldsymbol{x}^m$ が観測されたときの認識結果 f^* は一般に以下の式で計算される．

$$f^* = \arg\max_f P(f|\{\boldsymbol{x}^m\}) \tag{1}$$

ここで，$P(f|\{\boldsymbol{x}^m\})$ は，観測された特徴ベクトル群 $\{\boldsymbol{x}^m\}$ が食材 f から抽出されたものである確率であり，特徴ベクトル群 $\{\boldsymbol{x}^m\}$ に対する食材 f の尤度に相当する．

decision-level における単純な統合手法の 1 つである sum を用いる場合，上式の $P(f|\{\boldsymbol{x}^m\})$ を，単独の観測値 $\boldsymbol{x}^m$ に対する食材 f の尤度 $P(f|\boldsymbol{x}^m)$ の和で置きかえる．すなわち，認識結果 f^* を次式に従って計算する．

$$f^* = \arg\max_f P(f|\{\boldsymbol{x}^m\}) \approx \arg\max_f \sum_m P(f|\boldsymbol{x}^m) \tag{2}$$

上式は，尤度 $P(f|\boldsymbol{x}^m)$ を全ての m に対して等価に扱うことを意味しているが，実際の識別器では，尤度 $P(f|\boldsymbol{x}^m)$ のモデル化の為され方はモダリティの種類 m によって異なる．従って，上述のような decision-level fusion の問題が生じる．

ここで，Ivanov らの手法に従えば，モダリティ m による食材認識の結果 C^m を用いて食材 f を決定することを考え，C^m を確率変数として扱うことになる．

このとき，上記の $P(f|\boldsymbol{x}^m)$ は以下の式で計算される．

$$P(f|\boldsymbol{x}^m) = \sum_{j=1}^{K} P(f, C_j^m|\boldsymbol{x}^m) = \sum_{j=1}^{K} P(f|C_j^m, \boldsymbol{x}^m)P(C_j^m|\boldsymbol{x}^m) \quad (3)$$

$$\approx \sum_{j=1}^{K} P(f|C_j^m)P(C_j^m|\boldsymbol{x}^m) \quad (4)$$

ここで，(3) 式右辺における $P(f|C_j^m, \boldsymbol{x}^m)$ を実際に観測することは難しいため，これを $P(f|C_j^m)$ により近似したものが (4) 式である．なお，j は食材の種類を表すインデクスを，K は食材の種類の数 (クラス数) を表す．$P(C_j^m|\boldsymbol{x}^m)$ はモダリティ m から得られる各クラスへの尤度に相当するので，モダリティ m の特徴ベクトル $\boldsymbol{x}^m$ を識別器に投入することで獲得できる．$P(f|C_j^m)$ は，モダリティ m の特徴ベクトル $\boldsymbol{x}^m$ によって認識結果 C_j^m が定まった時に，実際には食材が f であった確率を示しており，以下のように変形できる．

$$P(f|C_j^m) = \frac{P(C_j^m|f)P(f)}{P(C_j^m)} = \frac{P(C_j^m|f)P(f)}{\sum_k P(C_j^m|f_k)P(f_k)} = \frac{P(C_j^m|f)}{\sum_k P(C_j^m|f_k)} \quad (5)$$

上式においては事前確率 $P(f)$ は任意の食材について一様であると仮定した．$P(C_j^m|f_k)$ および $P(C_j^m|f)$ は各モダリティ単独で食材認識を行った場合の食材組ごとの誤認識率に相当する．これは各モダリティを単独で用いた場合の Confusion Matrix を求めることで得ることができる．

以上のことから，

$$P(f|\boldsymbol{x}^m) \approx \sum_{j=1}^{K} P(f|C_j^m)P(C_j^m|\boldsymbol{x}^m) = \sum_{j=1}^{K} \frac{P(C_j^m|f)}{\sum_{k=1}^{K} P(C_j^m|f_k)} P(C_j^m|\boldsymbol{x}^m) \quad (6)$$

となり，最終的な認識結果は f^* は次式で与えられる．

$$f^* = \arg\max_f \sum_m P(f|\boldsymbol{x}^m) \approx \arg\max_f \sum_m \sum_{j=1}^{K} \frac{P(C_j^m|f)}{\sum_{k=1}^{K} P(C_j^m|f_k)} P(C_j^m|\boldsymbol{x}^m) \quad (7)$$

$P(C_j^m|\boldsymbol{x}^m)$ がどれか一つの食材 f_j に対してのみ 1 となり，それ以外の全ての食材に対して 0 となるような多数決（voting）型の統合を行う場合には，上式は次のように簡略化でき，これを計算することで最終的な認識結果を決定すればよい．

$$f^* = \arg\max_f \sum_m \frac{P(C_j^m|f)}{\sum_{k=1}^{K} P(C^m|f_k)} \quad (8)$$

4.3 実験手順

本研究では，食材認識に用いる識別器としてSVMを採用した．SVMは，ある特徴ベクトルが入力されたとき，その特徴ベクトルの発生元と考えられる単一のクラスのクラス名を返す．つまり，どれか一つのクラスに対してのみ尤度が1となり，それ以外の全クラスに対して0となるようなvotingとなる．よって，本実験では最終的な認識結果の決定式として，4.2節の式(8)を用いた．

Ivanovらの統合手法を用いる場合，各モダリティ単独による識別器を生成するための学習データ・その識別器を用いた場合のConfusion Matrixを獲得するための評価データ・最終的な認識精度を確認するためのテストデータの3つが必要になる．このとき，認識精度を不当に高く評価しないために，この3つのデータに重複があってはならない．しかし本実験では，この3つのデータを重複なく用意するだけの十分なサンプル数が獲得できなかった．そこで，獲得したサンプルに対してCross Varidationを適用した．その具体的な手順は以下の通りである．

まず，4.1節で得られた各食材に対する画像・振動音・荷重の特徴ベクトルを用いて各モダリティ単独での認識を行い，それぞれのConfusion Matrixおよび認識結果の組を獲得する．これは次の処理による．

1. サンプル集合を5つのグループに分ける．分割後の各グループに含まれる各クラスのサンプル数が，グループごとに出来るだけ均等になるようにする．
2. 5つのグループの内，4つのグループを用いて，以下の処理を行う．
   (a) 4つのグループの内，3つのグループを学習データとして識別器を生成する．残りの1グループを評価データとする．
   (b) (a)で生成した識別器に評価データを入力し，認識結果を獲得する．
   (c) 学習データと評価データを入れ替えながら(a)および(b)を4回繰り返すことで，認識結果が4組得られる．これらの正否を調べまとめることで，Confusion Matrixを獲得する．
3. 2の(a)〜(c)で用いた4つのグループ（学習データ＋評価データ）で識別器を学習し直す．その識別器に対して，残りの1グループ（テストデータ）を入力し，各モダリティ単独での認識結果を得る．
4. 4つのグループ（学習データ＋評価データ）と1つのグループ（テストデー

タ）を入れ替えながら 2〜3 を 5 回繰り返すことで，Confusion Matrix とモダリティ単独での認識結果の組を 5 組獲得する．

上記の処理を，すべてのモダリティに対して同様に行った．この結果得られた Confusion Matrix とそれに対応する各モダリティの出力を用い，4.2 節の式 (8) により，最終的な認識結果を決定した．

4.4 実験結果および考察

4.3 節において各モダリティごとに 5 つの Confusion Matrix が得られる．これをそれぞれのモダリティごとにまとめたもの（付録を参照）は，各モダリティ単独での認識結果に相当する．この結果から，3.1 節で述べたような各モダリティの性質が本実験でも表れていることが分かった．例えば画像では，本実験で用いた食材集合において特異な色を持つニンジンは，色が比較的近いトマトと多少誤認識する程度で，その他の食材とは誤認識することなく高い認識率を示した．逆に，食材集合に多数含まれる深緑色を持つキュウリ・ピーマン・ニラ・ブロッコリーなどは，これらの食材同士での誤認識が多数発生し，認識率は低くなった．また，本研究で用いた食材色特徴ベクトルの性質上，深緑色や紫色は黒色に近い値を示し，その結果シイタケ・ナスなどがキュウリ・ピーマンなどと誤認識される場合もあった．振動音では，均一構造をもつキュウリのような食材をホウレン草やニラのような葉物野菜と誤認識することは少なかった．また，鶏肉は切断時に振動音が発生しにくいと考えられていたにも関わらず，認識率が高くなっていた．これは他の食材と異なり，振動音が発生しにくいということが特徴的な値となって表れたからであると考えられる．しかし，ホウレン草・キャベツ・白菜・ニラのよう葉物野菜は切断時の振動音が似た音になりやすく，これらの食材同士での誤認識が避けられなかった．また，共にスジが多く内部構造が似通った食材であるショウガとゴボウ同士の誤認識も多かった．荷重では，硬い食材であるダイコンは，白ネギや白菜のような比較的軟らかい食材との誤認識は少ないが，タマネギ・カボチャ・ニンジンのような硬い食材との誤認識は多かった．各々のモダリティ単独での認識の平均精度は，画像が 61.0%，振動音が 35.3%，荷重が 39.2%であった．

表 5 に画像・振動音・荷重データを統合した結果を示す．表 5 における入出力の数字は，表 4 の食材 ID に対応する．個別の食材ごとの誤認識率の変化に注目

すると，大幅に改善された例が見られた．例えば，前述のように画像単独ではキュウリとピーマンの間で誤認識が多数発生しており，キュウリをピーマンと誤認識する割合，ピーマンをキュウリと誤認識する割合は，それぞれ 38.5%と 61.1%であった．しかし，振動音・荷重と統合することにより，これらの間の誤認識の割合は，それぞれ 6.5%と 3.8%まで改善された．他にも，荷重では大根と人参を識別することが難しく，大根を人参と誤認識する割合，人参を大根と誤認識する割合は，それぞれ 12.9%と 9.0%であったが，提案手法では，これらの間の誤認識の割合はそれぞれ 0.3%と 0.0%まで改善できている．このように，各モダリティの問題点が原因となり発生していた誤認識が，他のモダリティと統合することで改善されている例が多数確認できた．また，各食材の認識精度においても，大幅に改善された例が見られた．例えば振動音における生姜の認識精度は，ゴボウとの誤認識が原因となり，27.4%に止まっていたが，画像・荷重と統合された結果，提案手法では 75.5%まで向上している．提案手法の平均認識精度は 67.0%であり，各モダリティ単独での認識精度より高いものとなっていた．

一方，統合による認識を行った場合でも認識精度が高くならない食材もあった．例えば，ホウレン草の認識精度は統合による認識手法の場合でも 18.6%であり，ニラとの間で誤認識が多数発生していることが分かる．これは，ホウレン草とニラは共に深緑色をしており，共に葉物野菜であるため，抽出できる画像・振動音・荷重が全て似たものになってしまったためであると考えられる．その他にもキャベツ・タマネギ・白菜の間での誤認識も多く発生していた．この原因としては，これらの食材は色が似ており，いずれも層状の内部構造をしているため，得られる画像・振動音の特徴ベクトルが似たものになってしまったことが考えられる．また，本研究では切断方法に制約を設けなかったため，比較的硬いタマネギであっても切断部位によっては比較的軟らかい食材であるキャベツ・白菜と似た荷重特徴が得られてしまう場合もあり，この結果，3 種の食材間の識別ができなかったと考えられる．

以上のように，画像・振動音・荷重データを統合利用することで，それぞれのモダリティ単独での認識の際に発生していた誤認識が改善された食材組が複数確認できた．一方で，画像・振動音・荷重のどのモダリティを用いても識別が困難なために，3 つのモダリティを統合してもうまく識別できない食材組が存在することも分かった．

表 5: 画像・振動音・荷重データを統合した認識の結果（%）

入力＼出力	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	53.3	5.7	2.9	0.0	1.9	5.7	1.0	1.9	5.7	0.0	1.9	0.0	1.9	1.9	0.0	0.0	1.0	1.9	4.8	1.0	1.9	4.8	1.0
2	0.0	91.4	6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9
3	0.0	1.8	86.4	0.0	0.0	0.0	0.0	1.8	0.0	0.0	0.0	7.3	0.0	0.0	0.0	0.0	0.0	0.0	1.8	0.0	0.0	0.0	0.9
4	2.0	0.8	2.0	62.1	0.8	0.8	0.8	1.2	6.0	0.4	0.8	3.2	0.0	2.0	0.0	3.2	1.2	0.8	2.8	0.0	4.0	4.4	0.4
5	1.3	0.6	0.0	0.0	75.5	0.0	0.0	7.7	0.6	0.0	0.6	0.0	1.3	0.0	0.0	0.0	0.0	1.9	9.0	1.3	0.0	0.0	0.0
6	7.6	2.2	2.2	0.0	3.8	37.0	1.6	1.6	6.5	3.8	3.3	0.0	1.6	2.7	3.3	1.1	0.5	6.0	8.7	0.5	0.0	5.4	0.5
7	1.8	4.4	4.4	0.0	1.8	4.4	18.6	0.0	8.0	3.5	2.7	0.0	1.8	0.0	1.8	0.0	2.7	0.0	1.8	4.4	24.8	11.5	1.8
8	0.0	2.3	2.3	0.0	0.0	0.0	0.0	90.6	0.0	0.0	3.1	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8
9	5.5	2.1	0.4	0.0	3.0	3.8	3.4	3.4	47.7	6.3	2.5	0.0	1.7	1.7	3.0	2.5	0.4	1.3	7.2	0.0	0.8	3.0	0.4
10	5.4	7.7	3.8	0.8	1.5	1.5	0.8	4.6	6.2	37.7	2.3	0.0	6.2	0.0	0.8	3.1	6.2	3.8	0.0	0.0	0.8	6.2	0.8
11	1.3	8.8	1.3	0.0	17.0	0.0	0.0	0.0	0.6	0.0	59.7	0.0	0.0	1.3	0.0	0.0	0.0	0.0	0.0	0.6	1.3	5.7	2.5
12	0.0	0.9	1.8	1.8	0.0	0.0	0.0	1.8	0.9	0.0	0.0	43.8	0.0	6.3	0.9	21.4	16.1	0.0	0.0	0.9	0.9	0.0	2.7
13	0.8	0.0	0.0	0.0	5.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	84.1	0.0	0.0	0.0	0.0	8.7	0.0	0.0	0.0	0.0	0.8
14	0.3	0.7	1.4	6.3	0.0	0.0	0.0	0.7	0.0	0.3	2.8	3.5	0.0	69.6	0.0	1.7	7.0	0.0	0.7	0.3	3.8	0.0	0.7
15	0.0	0.0	0.0	5.7	0.0	0.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.0	92.5	0.0	0.0	0.0	1.1	0.0	0.0	0.0	0.0
16	0.0	0.0	4.9	0.0	1.0	0.0	0.0	0.5	0.0	0.0	0.0	12.2	0.5	4.4	0.0	66.8	2.4	5.9	0.5	0.0	0.0	0.0	1.0
17	0.0	0.0	0.9	2.8	0.0	0.0	0.0	0.9	0.0	0.0	0.0	15.1	0.0	8.5	0.9	7.5	58.5	0.0	0.9	1.9	0.0	0.0	1.9
18	0.0	0.0	0.0	0.0	5.5	0.0	0.0	0.9	0.0	0.0	0.0	0.0	9.1	0.0	0.0	0.0	0.0	80.9	0.0	0.0	0.0	0.0	3.6
19	0.0	6.9	15.4	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.6	0.0	1.8	0.0	4.9	0.0	2.0	68.2	0.0	0.0	0.0	0.0
20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.0	0.6	0.0	98.7	0.0	0.0	0.0
21	5.0	0.8	6.7	1.7	0.8	2.5	7.6	0.0	3.4	0.8	0.8	0.0	0.8	0.8	0.0	0.0	1.7	0.0	0.0	1.7	60.5	3.4	0.8
22	4.8	1.0	3.8	0.0	0.0	1.9	1.9	2.9	3.8	3.8	0.0	0.0	0.0	1.9	1.0	1.0	1.9	0.0	0.0	1.0	6.7	61.9	1.0
23	0.0	1.2	1.2	0.0	0.0	0.0	0.0	0.8	0.0	0.0	2.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	94.4

第5章　結論

本研究では，調理観測データからのレシピ自動生成に必要な基盤技術の一つである食材認識技術に着目した．食材認識を目的とした従来研究では，画像・振動音・荷重がそれぞれ単独で用いられてきたが，各々のモダリティにはそれぞれ問題点があり，その問題点により誤認識が発生していた．そこで本研究では，画像・振動音・荷重データを統合的に用いることで，各々のモダリティ単独での認識において発生する誤認識を改善するような食材認識手法を提案した．複数のモダリティを統合する際に，それぞれのデータにおいて単純に時刻が一致している部分から特徴ベクトルを抽出・併用しても認識に有効とはならない．この問題に対して，本研究では，各モダリティの性質を考慮したタイミングでの特徴ベクトルの抽出を行った．荷重については，土本らの手法により自動的に検出された切断区間から 10 次元の特徴ベクトルを算出した．画像については，食材をまな板に置く瞬間に注目し，その瞬間に相当する画像から獲得された食材領域に対して，64 次元の食材色特徴ベクトルを抽出した．振動音については，切断時にまな板と包丁が衝突する直前約 0.2 秒分から 16 次元の特徴ベクトルを算出した．23 種類の食材集合を対象に，上述の特徴ベクトルを用いて認識実験を行った結果，画像・振動音・荷重それぞれのモダリティを単独で用いた認識手法の際に発生していた誤認識が，他の 2 つのモダリティと統合することにより改善されていることが確認できた．

本研究で 3 つのモダリティを統合した認識手法における平均精度は 67%であった．これは，レシピ自動生成システムに用いることができるほど高い精度ではないと言える．この精度を向上させるためには，各モダリティから抽出する特徴ベクトルについて再検討することが必要である．例えば本研究では，一回の切断ごとに画像・振動音・荷重それぞれの特徴ベクトルを抽出し，認識を行った．しかし，本研究で用いたモダリティのうち特に振動音と荷重は，調理者の熟練度や食材のどの部位を切断するかによって，同じ食材・同じ個体であっても，得られる特徴ベクトルに差が出る可能性が考えられる．このような特徴ベクトルの差は，認識結果を下げる原因にもなりうる．これに対して，本研究のように得られる切断全てを対象にして特徴ベクトルを抽出し認識を行うのではなく，食材一個体に対して行われる複数回の切断のうち，認識に有効な特徴ベクトルが含まれるような切断を選択して，それから得られる特徴ベクトルによ

りその食材が何であるかを認識する，という解決法が考えられる．切断の選択方法としては例えば，一個体に対する複数の切断（切断系列）において初期の切断と後期の切断は切断部位が食材の端によりやすく，得られる特徴ベクトルも乱れやすい，という考えの下で切断系列の中ごろの切断を選択する，などが考えられる．また，各モダリティのデータに対し，より認識能力の高い特徴ベクトルを考察・抽出することも重要と考えられる．具体的には，本研究では画像から食材色を特徴ベクトルとして抽出したが，これ以外にもテクスチャや形状などを特徴ベクトルとして用いることで，画像というモダリティの認識能力は格段に向上すると期待される．

このような，統合に利用する特徴ベクトルの選択や各モダリティの認識能力の向上について実験と考察を行うことが今後の課題である．

謝辞

本研究を行うにあたり，多大な御指導を賜りました美濃導彦教授，椋木雅之准教授に深く感謝致します．また，日頃より研究について多くの助言を頂き，本論文の作成においても御指導を頂きました舩冨卓哉助教，中村和晃助手に厚く御礼申し上げます．最後に，本研究に対して多くの意見を下さった認識グループの皆様，並びに，美濃研究室の皆様に御礼を申し上げます．

参考文献

[1] URL: http://cookpad.com.

[2] クックパッド株式会社: 2011 年 4 月期第 3 四半期決算補足説明資料.

[3] 山肩洋子, 尾原秀登, 沢田篤史, 角所孝, 美濃導彦: 食材に視覚的特徴変化を生じさせる加工における食材と加工動作の同時認識, 電子情報通信学会論文誌, Vol. J90-D, No. 9, pp. 2550–2561 (2007).

[4] 柴田知秀, 加藤紀雄, 黒橋禎夫: 言語情報と映像情報の統合による物体のモデル学習と認識, 情報処理学会論文誌, Vol. 49, No. 3, pp. 1451–1464 (2008).

[5] 三功浩嗣, 山肩洋子, 角所孝, 美濃導彦: 調理加工に起因する振動音を用いた食材識別, 電子情報通信学会 2006 総合大会 (2006).

[6] 土本良樹, 橋本敦史, 船冨卓哉, 山肩洋子, 上田真由美, 美濃導彦: 調理における切断加工時の荷重特徴を用いた食材認識, 電子情報通信学会マルチメ

ディア・仮想環境基礎 (MVE) 研究会, Vol. 110, No. 456, pp. 55–60 (2011).

[7] 橋本敦史, 森直幸, 舩冨卓哉, 山肩洋子, 椋木雅之, 角所考, 美濃導彦: 把持の順序と外見の変化モデルを利用した調理作業における食材追跡, 信学論 A, Vol. J94-A, No. 7, pp. 509–518 (2011).

[8] 橋本敦史, 舩冨卓哉, 中村和晃, 椋木雅之, 美濃導彦: TexCut:GraphicCut を用いたテクスチャの比較による背景差分, 信学論 D, Vol. J94-D, No. 6, pp. 1007–1016 (2011).

[9] 妹尾紗恵: 食材相関図から見た料理構造解析一安定性と可変性にみる日本の家庭料理一, 日本家政学会誌, Vol. 59, No. 4, pp. 211–219 (2006).

[10] Ivanov, Y., Serre, T. and Bouvrie, J.: Error weighted classifier combination for multi-modal human identification, Technical report, MIT CSAI Laboratory Technical Report MIT-CSAIL-TR-2005-081 (2005).

[11] Gunes, H. and Piccardi, M.: Automatic Temporal Segment Detection and Affect Recognition From Face and Body Display, *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, Vol. 39, No. 1, pp. 64–84 (2009).

[12] Castellano, G., Kessous, L. and Caridakis, G.: Emotion Recognition through Multiple Modalities: Face, Body Gesture, Speech, *Affect and Emotion in Human-Computer Interaction*, Vol. 4868/2008, pp. 92–103 (2008).

[13] Tao, Q. and Veldhuis, R.: Threshold-optimized decision-level fusion and its application to biometrics, *Pattern Recognition*, Vol. 42, No. Issue.5, pp. 823–836 (2009).

付録

A.1　荷重特徴ベクトルで用いた 6 次元

本研究で用いた荷重特徴ベクトルのうちの 6 次元は，切断区間 $T = [t_s, t_e]$ における荷重 $F(t)$ に対して，以下の式により計算される．

平均

$$\bar{F} = \frac{1}{t_e - t_s} \sum_{t=t_s}^{t_e} F(t) \tag{A.1}$$

平均偏差

$$\mathrm{ADev} = \frac{1}{t_e - t_s} \sum_{t=t_s}^{t_e} |F(t) - \bar{F}| \tag{A.2}$$

分散

$$\mathrm{Var} = \frac{1}{t_e - t_s - 1} \sum_{t=t_s}^{t_e} (F(t) - \bar{F})^2 \tag{A.3}$$

標準偏差

$$\sigma = \sqrt{\mathrm{Var}} \tag{A.4}$$

歪度

$$\mathrm{Skew} = \frac{1}{t_e - t_s} \sum_{t=t_s}^{t_e} \left[\frac{F(t) - \bar{F}}{\sigma} \right]^3 \tag{A.5}$$

尖度

$$\mathrm{Kurt} = \left\{ \frac{1}{t_e - t_s} \sum_{t=t_s}^{t_e} \left[\frac{F(t) - \bar{F}}{\sigma} \right]^4 \right\} - 3 \tag{A.6}$$

A.2　各モダリティ単独による認識結果

画像・振動音・荷重をそれぞれ単独で用いた場合の認識結果を表 A.1, A.2, A.3 に示す．それぞれの表における「入力」「出力」の数字は，表 4 の食材 ID に対応する．

表 A.1: 画像を単独で用いた場合の認識結果

入力＼出力	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	19.0	0.0	0.0	0.0	3.4	48.6	0.0	0.0	28.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0
2	3.4	81.9	8.6	0.0	1.7	0.0	0.0	0.4	0.0	0.4	0.0	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.1
3	0.0	11.2	78.0	0.0	0.4	0.0	0.0	1.7	0.0	0.0	0.0	3.4	0.0	0.0	0.0	0.0	0.9	0.0	3.9	0.0	0.0	0.0	0.4
4	0.0	0.0	0.0	71.2	0.0	6.9	2.6	0.0	5.2	1.7	0.0	0.9	0.0	5.6	0.0	0.0	2.6	0.0	0.0	0.0	1.7	1.7	0.0
5	7.3	0.0	0.0	0.0	77.2	2.6	0.0	2.2	0.9	0.0	0.0	0.0	3.9	0.0	0.0	0.0	0.0	2.6	0.9	0.0	0.0	0.4	2.2
6	3.5	0.0	0.0	0.0	0.0	56.3	0.0	0.0	38.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.7	0.0
7	0.9	0.0	0.0	1.7	0.0	49.0	4.7	0.0	31.1	1.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.8	3.0	0.0
8	0.0	2.2	1.3	0.0	3.5	0.0	0.0	86.2	0.0	0.0	0.0	0.9	0.0	0.0	0.4	0.0	0.4	1.7	0.0	0.0	0.0	0.0	3.5
9	0.4	0.0	0.0	0.0	1.3	61.1	0.0	0.0	35.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.7	0.0
10	2.6	1.3	0.9	0.0	0.0	48.2	0.9	0.0	26.8	16.4	0.0	2.2	0.0	0.0	0.4	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0
11	1.7	0.4	0.4	0.4	12.9	0.0	0.4	0.9	0.0	0.4	72.9	0.0	0.4	0.0	0.0	1.3	0.0	0.0	0.0	0.0	2.6	0.0	5.2
12	0.0	1.7	6.5	1.7	0.0	0.0	0.9	0.0	0.0	0.0	1.3	40.5	0.0	3.9	0.9	16.8	21.6	0.0	2.6	0.0	0.9	0.9	0.0
13	1.3	0.0	0.0	0.0	7.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	91.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
14	0.0	1.7	0.0	6.0	0.0	0.0	0.4	0.0	0.0	1.3	3.0	4.3	0.0	73.7	0.0	1.3	7.8	0.0	0.0	0.0	0.4	0.0	0.0
15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	1.3	0.0	0.0	0.0	0.0	95.7	0.0	0.0	0.0	2.6	0.0	0.0	0.0	0.0
16	0.0	0.0	0.0	1.3	3.0	0.0	0.0	1.3	0.0	0.0	3.0	12.9	0.0	2.6	1.7	57.8	6.0	9.1	0.0	0.0	0.0	0.0	1.3
17	0.0	0.4	3.5	2.1	0.0	0.0	0.0	2.6	0.0	0.0	0.0	14.2	0.0	8.6	0.0	9.5	58.6	0.0	0.4	0.0	0.0	0.0	0.0
18	1.7	0.0	0.0	0.0	12.5	2.6	0.0	0.9	0.9	0.0	0.0	0.0	10.3	0.0	0.0	3.0	0.0	64.2	0.0	0.0	0.0	0.0	3.9
19	0.0	1.7	12.5	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	2.6	0.0	2.6	2.6	1.3	0.9	0.0	74.6	0.0	0.0	0.0	0.4
20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.2	0.0	0.0	0.0	0.0	0.0	0.0	94.8	0.0	0.0	0.0
21	5.6	0.0	0.0	0.0	0.0	31.4	9.9	0.0	16.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	32.7	3.5	0.0
22	0.0	0.0	0.0	0.0	0.0	25.4	4.3	0.0	16.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	54.3	0.0
23	0.0	2.6	0.0	0.0	13.8	0.0	0.0	3.9	0.0	0.0	3.9	0.0	9.5	0.9	0.0	0.4	0.0	0.4	0.0	0.0	0.0	0.0	64.7

表 A.2: 振動音を単独で用いた場合の認識結果

入力＼出力	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	35.2	5.2	3.3	9.0	1.7	3.1	1.7	1.9	5.5	0.5	3.1	1.0	3.1	1.9	1.2	1.7	2.1	3.1	5.2	1.0	1.0	5.7	2.9
2	3.1	29.3	8.6	1.2	3.8	3.6	5.7	10.7	0.7	6.0	1.0	3.1	1.9	1.9	1.9	0.5	2.6	1.0	0.5	3.3	3.8	3.8	2.1
3	1.2	7.6	40.0	1.9	2.6	1.2	4.0	2.4	1.0	3.8	1.7	6.4	0.7	1.2	1.0	1.9	4.3	0.2	0.5	2.1	8.1	1.9	4.3
4	9.0	2.9	4.3	15.2	3.8	3.3	8.1	1.9	6.7	3.3	3.6	1.0	1.4	3.8	1.9	2.6	5.5	2.9	8.6	2.6	2.9	1.7	3.1
5	1.0	4.3	3.6	5.5	27.4	3.3	2.9	5.5	1.4	2.9	8.6	1.0	4.3	5.2	4.5	4.3	2.6	5.5	3.6	2.6	0.0	0.2	0.0
6	4.5	2.4	1.9	6.0	4.0	36.9	2.6	1.7	2.4	3.3	3.1	1.9	2.6	0.5	2.6	1.7	1.9	6.7	9.0	1.2	0.7	1.7	0.7
7	2.4	2.4	6.0	6.7	2.9	1.7	27.4	0.0	5.7	1.4	4.3	3.3	2.6	0.0	1.4	0.7	4.0	1.7	2.9	3.3	16.2	1.7	1.4
8	1.4	13.8	0.5	1.4	2.9	1.0	0.0	48.8	2.6	7.4	2.6	2.1	0.5	3.1	2.6	2.4	1.9	1.7	0.0	1.4	0.0	1.7	0.2
9	4.5	1.4	1.2	5.7	3.8	2.9	2.1	1.4	43.1	3.1	3.1	5.0	2.9	0.7	3.1	3.1	1.2	5.5	4.3	0.0	0.0	1.4	0.5
10	0.7	8.8	5.5	1.9	3.8	2.6	2.1	7.9	3.6	22.6	4.0	6.9	5.7	1.0	2.1	7.4	6.7	3.1	0.0	0.0	0.0	3.3	0.2
11	3.1	0.7	4.5	3.6	10.2	4.0	3.8	4.8	2.9	1.7	28.6	3.3	1.2	5.5	4.8	5.2	2.9	3.1	1.2	1.4	1.2	1.0	1.4
12	4.0	5.2	6.7	1.9	3.8	0.5	7.4	2.4	6.7	7.9	3.8	16.2	2.6	2.9	3.1	2.6	10.5	2.1	0.5	1.4	3.1	2.6	2.1
13	3.1	2.6	5.0	1.4	2.9	3.6	6.0	1.9	1.9	6.2	1.9	9.5	25.7	2.1	3.8	3.6	3.8	2.1	3.8	0.2	3.8	3.6	1.4
14	1.4	4.5	1.7	5.2	5.0	1.0	2.1	1.4	0.0	3.1	4.8	1.7	3.3	37.1	3.8	5.2	6.2	2.4	1.0	1.2	1.0	5.0	1.9
15	2.9	3.1	0.0	2.4	7.6	3.1	1.0	1.4	2.9	3.1	0.7	1.7	3.8	2.1	44.3	5.2	0.7	6.4	4.3	1.4	0.0	1.7	0.2
16	1.0	1.0	3.8	5.2	9.5	2.9	1.4	2.1	3.3	5.0	5.7	2.6	4.5	5.5	5.0	25.0	4.0	8.6	1.7	1.0	0.2	1.0	0.0
17	2.9	3.3	5.2	6.9	1.4	4.0	6.0	4.8	1.9	6.2	4.3	7.1	4.8	4.0	0.2	1.4	22.6	1.0	1.9	1.7	2.6	4.0	1.7
18	5.5	2.9	1.0	4.3	3.1	4.0	0.7	5.5	4.5	5.2	4.3	0.7	1.9	2.1	4.0	6.7	2.4	33.8	3.1	2.4	0.0	1.2	0.7
19	8.1	0.0	0.2	6.7	2.1	5.2	1.7	0.2	1.9	0.2	1.9	1.0	1.9	0.0	0.7	0.7	1.0	6.0	60.5	0.0	0.0	0.0	0.0
20	2.1	5.0	6.0	2.6	3.3	1.2	5.5	2.9	1.7	0.5	2.6	1.9	0.5	3.6	1.0	2.6	1.7	1.9	1.0	46.2	2.1	1.4	2.9
21	1.0	1.9	7.9	4.3	0.7	1.0	14.0	0.2	2.4	0.2	1.7	3.1	2.1	2.4	0.0	0.5	2.9	0.0	0.5	1.7	48.1	2.6	1.0
22	5.2	4.3	4.8	4.5	1.0	1.7	7.1	2.6	4.3	4.8	1.0	3.1	3.3	6.2	1.0	3.8	7.4	0.5	1.0	2.6	7.4	21.7	1.0
23	4.0	0.5	0.0	1.2	0.2	0.0	2.1	1.9	1.0	0.0	1.9	3.1	2.4	1.4	0.2	0.2	0.0	0.2	0.0	1.7	0.7	0.7	76.4

表 A.3: 荷重を単独で用いた場合の認識結果

入力＼出力	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	36.7	4.0	0.0	0.2	1.2	9.3	2.9	0.0	0.0	0.0	0.0	0.7	1.4	0.0	12.6	0.2	0.0	1.0	11.4	0.0	16.4	1.9	0.0
2	3.6	36.2	6.0	4.3	4.5	2.6	5.2	0.7	1.0	7.1	0.5	2.6	1.0	0.2	9.8	3.3	3.8	0.0	1.0	3.8	1.4	1.2	0.2
3	0.0	10.2	24.5	3.3	7.4	0.2	1.0	1.0	0.0	17.1	6.7	4.3	2.6	2.4	0.0	12.4	1.9	0.0	0.7	3.1	0.0	0.5	0.7
4	3.3	5.2	2.4	40.2	9.3	0.5	17.6	0.0	3.1	1.4	2.9	2.9	0.0	1.0	0.0	1.9	1.7	0.0	0.0	1.9	4.5	0.0	0.2
5	0.5	14.5	10.0	5.5	20.7	2.1	2.6	0.0	5.7	4.5	1.9	5.2	1.9	0.2	0.0	10.7	6.7	2.1	0.0	2.6	0.5	1.2	0.7
6	5.0	0.0	0.0	0.0	1.9	46.2	2.1	0.0	0.2	0.0	0.0	1.0	0.0	0.0	4.8	0.0	0.0	0.0	28.8	0.0	7.9	1.4	0.7
7	5.2	3.3	1.2	14.5	4.3	3.3	50.5	0.0	5.7	0.5	1.7	1.2	0.5	0.0	1.4	0.2	2.6	1.0	0.0	0.0	2.9	0.0	0.0
8	0.0	0.0	0.7	0.0	0.5	0.0	0.0	49.0	0.0	0.0	6.4	0.0	1.7	5.0	0.0	3.1	2.4	16.4	0.0	11.7	0.0	1.4	1.7
9	0.0	1.9	3.3	5.5	7.1	1.4	7.1	0.0	45.7	0.7	8.6	4.3	3.3	0.2	1.4	0.5	4.3	1.4	0.0	0.5	0.7	1.9	0.0
10	0.0	6.0	12.1	1.4	4.5	0.5	0.5	0.0	0.0	42.4	0.2	1.2	11.7	3.1	2.9	6.4	0.7	0.0	1.0	3.8	0.0	1.7	0.0
11	0.0	1.7	6.2	8.1	0.5	0.0	0.0	6.0	8.1	0.0	44.8	5.5	0.0	5.0	0.0	2.6	2.9	1.7	0.0	7.1	0.0	0.0	0.0
12	1.7	11.2	13.6	10.5	7.9	1.2	7.4	0.2	5.2	4.0	3.3	6.9	8.1	1.9	1.9	3.1	5.5	0.0	1.0	4.3	1.2	0.0	0.0
13	0.5	0.5	1.7	1.4	0.7	1.0	9.0	1.4	1.0	18.1	3.1	2.6	47.6	2.4	3.6	1.0	0.0	0.0	0.7	0.5	0.0	1.7	1.7
14	0.0	3.8	8.8	0.5	0.2	0.0	0.0	1.9	0.0	5.7	4.8	2.4	5.7	18.1	1.2	12.9	3.6	14.3	0.0	12.9	0.0	1.2	2.1
15	5.0	4.8	0.7	0.0	1.0	0.2	0.0	0.0	0.0	1.7	0.0	0.5	2.6	0.2	59.5	1.9	1.4	0.2	14.0	1.4	0.2	4.5	0.0
16	0.0	4.3	14.8	1.0	4.8	1.0	0.0	2.6	0.0	10.5	2.6	1.9	1.9	8.1	0.5	33.8	1.0	3.3	0.0	7.1	0.0	0.2	0.7
17	4.5	4.8	3.6	6.0	10.2	0.2	13.3	3.6	7.9	1.0	3.3	5.0	1.7	3.1	1.0	3.1	19.3	3.8	0.7	0.7	1.4	1.0	1.0
18	0.0	0.0	0.5	0.0	0.0	0.0	0.0	17.1	0.7	0.0	6.2	0.5	0.0	3.3	0.0	2.9	0.5	60.2	0.0	8.1	0.0	0.0	0.0
19	6.9	0.0	0.0	0.0	0.2	12.1	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	10.5	0.0	0.0	1.0	63.8	0.0	4.0	1.0	0.0
20	0.0	2.1	2.4	0.2	1.7	0.0	0.0	7.1	0.5	4.8	5.5	0.0	0.2	9.0	11.4	10.0	0.7	3.3	0.2	37.1	0.0	2.4	1.2
21	16.7	5.5	0.0	2.1	0.5	8.6	2.6	0.0	0.2	0.0	0.0	0.7	0.0	0.0	4.8	0.0	1.9	0.0	5.7	0.0	50.7	0.0	0.0
22	9.8	4.3	1.0	0.0	2.9	2.9	0.2	2.1	0.0	4.0	0.0	0.0	5.0	1.9	21.0	0.7	0.2	1.9	19.0	1.4	4.5	16.0	1.2
23	1.0	1.4	1.0	0.0	0.2	1.9	0.0	2.4	0.5	1.2	0.0	0.0	6.4	1.9	1.9	1.7	0.7	14.5	3.6	3.8	0.5	3.3	52.1