社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

KUSK Object Dataset: 調理作業中の物体への接触履歴データセットの作成

橋本　敦史† 　飯山　将晃†† 　森　　信介†† 　美濃　導彦††

† 京都大学大学院教育学研究科
†† 京都大学学術情報メディアセンター
E-mail: †ahasimoto@mm.media.kyoto-u.ac.jp

あらまし　調理作業では多数の道具を使いながら，複数の食材を加工し一つの料理を作る．このような流れはワークフローとして記述できることが知られており，ワークフローと実際の調理作業との対応付けは重要な課題である．また，この対応付けを行う上で，調理者がどの物体に接触したかという履歴は重要な情報となる．そこで，本研究では先行研究で発表した Kyoto University Smart Kitchen Dataset (KUSK Dataset) に収録された映像を対象とし，調理台上に登場した物体と，その物体に対する調理者の接触履歴についてのアノテーションデータと，それらのアノテーションデータに基いた物体認識結果を備えた KUSK Object Dataset を作成する．調理台上の物体と調理者との接触を自動で検出する我々の先行研究の手法を適用することで効率的なアノテーションを行い，5947 個の物体領域画像を収集することができた．また，既存の物体認識手法により，70%程度の精度での認識が行えることを確認した．この認識結果は調理作業を対象とした物体認識問題におけるベースラインとしての利用だけでなく，自然言語処理と画像処理の融合研究における，仮想的な画像処理結果としての利用も想定している．
キーワード　食メディア, データセット, 物体認識, アノテーション

KUSK Object Dataset: Annotation of Access History to Objects on Cooking Counter

Atsushi HASHIMOTO†, Masaaki IIYAMA††, Shinsuke MORI††, and Michihiko MINOH††

† Graduate School of Education, Kyoto University, Yoshida Honmachi, Sakyo-ku, Kyoto
†† Academic Center for Computing and Media Studies, Kyoto University, Yoshida-Nihonmatsu-cho, Sakyo-ku, Kyoto
E-mail: †ahasimoto@mm.media.kyoto-u.ac.jp

1. はじめに

映像中の出来事を計算機に理解させ，自然言語により説明させることは，情報学分野の一つの大きな目標である．映像理解の問題における認識の対象には，単に物体認識や動作認識など従来研究において単体の認識問題として取り扱われて来た対象だけに留まらず，それらの要素によって構成される，より抽象的な概念も含まれる．近年の画像処理分野における Convolutional Neural Network(CNN) の成功により，物体認識の精度は飛躍的に改善し，ILMLS の 1000 カテゴリを対象とした物体認識のコンペティション型ワークショップ ImageNet Large-scale Visual Recognition Challenge (ILSVRC) では人間と大きく変わらない精度が達成されているとの報告もある [1]．このように映像を構成する重要な要素の一つである物体認識技術が成熟の時を迎えつつあり，物体認識結果に基づいた，より抽象的な対象の認識を研究する地盤が整いつつある．

他方，そのような抽象的な認識結果を社会で活用する上では，認識結果を人に伝える手段も問題になる．物体や動作間の関係性を人間に伝える際に，計算機上で用いられている多種多様なデータ構造は，その多くが計算機には理解しやすくとも人間にとってはわかりにくい．このため，得られた関係性を自然言語表現へ変換することは重要な課題である．自然言語は長年の人間の営みにより洗練された情報伝達手段であるというだけではなく，Web などを介して大量のコーパスが比較的容易に手に入

る．また，必要に応じてクラウドソーシングの手法を活用して容易にデータセットを作成することも可能である．このため，大規模データベースを前提とした機械学習手法との親和性が高く，近年，活発な研究が行われている [2]〜[8].

一般的に画像処理と自然言語処理は，数学的には共通する技術も多いものの，研究者コミュニティが異なっているという点で異なる専門分野といえる．異なる分野間の融合研究を進める上では，研究の基盤となるデータの存在が重要となる．例えば Pascal-sentence データセット [2] は画像処理の分野で良く用いられていた Pascal データセットのいくつかの画像に対して Amazon Mechanical Turk を利用して画像を説明する文が付加されたものであり，画像処理と自然言語処理の分野融合研究の初期の呼び水としていくつかの研究で利用された [3], [4].

我々の先行研究 [9] では，この Pascal-sentence をモデルケースとして，レシピテキスト解析の連携研究への利用を想定して 20 種類のレシピに対する複数の被験者による調理作業を記録した Kyoto University Smart Kitchen Dataset (KUSK Dataset) を作成した．調理作業は通常，レシピと呼ばれる指示文の集合により，表現される．このような指示文は言語の持つ性質上，全順序的に与えられるが，それが指示する手順自身はワークフローとして有向グラフで表現可能であることが指摘されている [10]．このワークフローと観測された調理作業の対応付けを取ることは調理作業を観測した映像の理解における重要な課題の一つである [11]〜[13]．ワークフローで記述可能な作業には，調理作業以外にも，組立作業や外科手術，手芸など幅広く存在するが，作業者の自由度や，自然言語表現の多様性という観点から見た場合，調理作業は軽微な失敗やアレンジを許容し，かつ，多様な言語表現を持つ挑戦的な課題である．

調理作業を理解する上では，調理器具や食材といった調理台上の物体への調理者の接触履歴が重要な役割を果たす．例えば，我々の先行研究 [13] では，ワークフローにより表現したレシピが既知であるという前提であれば，接触履歴のみからでも，行動予測としては高い 70%近い精度を達成できることがわかっている．そこで，本研究では，KUSK Dataset の映像 (図 1) に対して，各観測において実際に起きた物体への接触履歴のアノテーションを行うことで，多様なレシピに対して物体接触履歴を基にした映像理解手法の研究を行う基盤となる KUSK Object Dataset を作成し，提供を行う．

2. 食メディアに関連するデータセット

料理は映像や自然言語処理などの情報学の分野だけでなく，食品加工や家政学，栄養学など様々な分野を横断する研究対象である．研究の切り口も幅広く，国内外を問わず，幅広いデータセットが提供されている．以下では，特にキッチンにおける行動観測に対象を絞り，関連するデータセットを紹介した上で，我々の提供する KUSK Object Dataset の位置付けを明らかにする．

調理活動を観測した場合，得られる映像中の構成要素は主に身体動作と調理台上に現れる物体や調理設備の状況，という 2 種類に大別される．例えば，TUM Kitchen Data Set [14] と CMU Multi-Modal Activity Database [15] では，天井に取付たカメラや一人称視点映像により調理台上の状況を観測すると共に，Motion Capture によって身体の動きを記録している．一方，Actions for Cooking Eggs Dataset [16] や MPII Cooking Activities Dataset [17] では RGB-D カメラによって，調理活動を正面情報から観測することで，調理台上の状況と人物動作の両方を記録している．

調理作業を正面上方から撮影することが出来れば，比較的容易に調理台上と人間の身体動作の両方を観測することが可能であるが，観測条件としてアイランド型キッチンでの作業が前提となる．これに対して，より一般的な I 型や L 型の，壁面に備え付けられたキッチンを対象として，調理作業を横方向から観測した The Breakfast Actions Dataset [18] も公開されている．

一方，50 Salads dataset [19] や我々の先行研究である KUSK Dataset [9](注1) は調理台の状況の観測を重視し，調理台を真上から観測した映像を用いている．身体動作について，50 Salads dataset では映像に現れない人間の動きを補完する役割として三次元加速度センサーを調理者に装着させている．一方 KUSK Dataset では，調理台を横方向から撮影した映像も提供するとともに，水道やコンロなどの備え付け設備の利用については流量センサや電流センサによる計測を行っている．さらにカメラでは観測不可能な調理台にかかった力を観測するために，作業場所への荷重も計測している．

KUSK Dataset の重要な特徴として，自然言語処理によるレシピ解析のデータセットである Flow Graph Corpus [20](注2) との連携を行っている点が挙げられる．KUSK Dataset で対象とした調理作業には COOKPAD から選ばれた 20 種類の作業が含まれている．Flow Graph Corpus にも，同じ 20 種類の作業を記述したレシピテキストを解析して得られたレシピの Work Flow が公開されている．従って，Flow Graph Corpus は自然言語処理結果として KUSK Dataset と組み合わせて利用することが可能となっている．

上述のデータセットでは，基本的には映像を始めとする各種センサの時系列信号データと，各時刻における動作の種類に関する正解データなどが提供されてきた．一方で，調理作業中の抽象的な出来事を理解する上では，作業に関わる物体の種類や，その物体にいつ触れたか，という物体接触履歴は重要な役に立つ．本研究で提供する KUSK Object Dataset は，KUSK Dataset の映像を対象として，その作業映像の重要な構成要素である作業台上の物体に関する正解データを提供するとともに，自然言語処理研究者が仮想的な画像処理結果として利用ができるように，画像処理結果も提供するものとなっている．

3. 物体への接触履歴の半自動アノテーション

一般的に映像に対するアノテーションには，単純にその映像全体を見る以上の時間がかかってしまう．例えば前述の物体への接触についてだけでも，我々の先行研究 [21] における実験で

(注1)：http://kusk.mm.media.kyoto-u.ac.jp/
(注2)：http://plata.ar.media.kyoto-u.ac.jp/data/recipe/home.html

図 1　KUSK Dataset で公開されている可視光映像データの例
Fig. 1　Example of RGB Video Data provided in KUSK Dataset.

は，30 分程度の調理作業でもほぼ毎回の実験で 100 回近い接触が行われることが確認された．もちろんレシピに応じて物体への接触回数は増減するものの，物体への接触のみに限っても，完全に手動で調理作業へのアノテーションを行うには膨大なコストがかかってしまう．

そこで今回のアノテーションでは，先行研究 [21] で提案した背景差分に基づく手法を利用することで負担軽減を図った．背景差分は事前の物体モデルの学習なしで動作する物体検出手法であり，学習用データが存在していない状態でも利用が可能である．さらに，この手法では，単に背景差分により動物体を検出するだけでなく，物体が手に取られたり，置かれたりしたことを区別しながら，その物体領域を抽出することができる．

本研究では，手法 [21] の結果に対して「誤りの訂正」,「レシピ毎に現れる物体の語彙の確定」,「語彙に基づいた物体ラベル付与」という 3 つの手動でのタスクを行うことで，動画全てをチェックする必要のないアノテーション作業を実現した．特にレシピ毎の語彙の確定にあたっては，基本的に料理オントロジー [22] に存在する語彙から，レシピを実際に調理するにあたって使用する可能性のある調理器具，調味料，食材を選択してもらうことで，アノテータ毎の表記ゆれの問題に対処した．

ただし，調理作業特有の問題として，スポンジ，洗剤など直接調理に関係しないものや料理には投入されない食材の食べない部分，調味料ボトルのフタ，複数の食材が混ざったものなどの，料理オントロジーには含まれない物体も観測される．このうち，スポンジなど直接調理に関係しないものにはアノテータが名前を付与し，表記ゆれは手動で修正を行った．また，食材の食べない部分や調味料のフタ，包装容器などは，オントロジー中の各物体の子クラスとして「食べない部分」,「フタ」,「包装容器」を付与し，一律に扱った．

複数の食材が混ざったものは一律に「混合食材」というラベルを付与した．各データには，そのデータが観測されたレシピの ID も付与されている．従って「混合食材」というラベルが振られた物体は，そのレシピの ID に従って種類を区別した．この方法では，同一のレシピに複数の混合食材が現れてもそれらを区別することはできない．そのような混合食材同士の区別には，映像をチェックして混合食材の中に何が混ざっているかを記録する必要があるが，これは今後の課題である．

このようなアノテーション作業を通して得られたデータの例を図 2 に示す．それぞれの物体は種類の他に，手に取られた，置かれたと考えられる映像のフレーム番号が付与されている．また，検出結果とは別に，アノテータが改めて物体を囲った矩形の座標が記録されている．

調理作業という実問題に特有の難しさとしては，食材が取りうる状態の多様性と，ボウルやまな板などの調理器具には頻繁に別の種類の物体が上に置かれた状態で観測されるという点が挙げられる．例えば図 2 の中で卵は様々な形態のものが撮影されている．また，ボウルやザル，まな板の上に別の物体が置かれている状況も多く含まれている．

4. 物体認識手法の適用による画像処理結果の作成

このようにして得られたアノテーションデータを用いて，実際に物体認識器を構築し，それを手法 [21] の結果に適用した画像処理結果データを作成した．認識対象としては，手動で誤りが訂正されたデータに対する認識結果と，誤り訂正なしの結果にそのまま物体認識を適用した認識結果の 2 種類を用意した．これらのデータは単に画像処理研究者が認識結果のベースラインとして用いるだけでなく，自然言語処理研究者が仮想的な画像処理結果として用いることも想定している．

4.1　識別器の実装

調理台上に現れる調理器具などは同一キッチンであれば，毎回の調理で同一のものが多く用いられる．従って，多くの応用において，識別器はレシピやキッチンに応じたものを準備する状況が想定される．このような実用上の想定の下，本研究ではキッチン毎に柔軟に識別器の切り替えが可能な構成として，Convolutional Neural Network (CNN) による特徴抽出器と，Linear Support Vector Machine (L-SVM) の組み合わせを用いた．特に CNN のネットワーク構造には，現在最も一般的に用いられており，Caffe [23] で学習済みのネットワークが提供されている Alex Net を用いた．また，L-SVM には libSVM による実装を用いた．

4.2　CNN による特徴抽出のための画像領域の切り出し

ImageNet により学習済みのネットワークを利用する場合，入力形式は学習されたモデルに従う必要がある．Caffe による実装の場合には入力は 256x256pixels の RGB 画像となるため，アノテーションデータの中で与えられた矩形の長辺の長さに

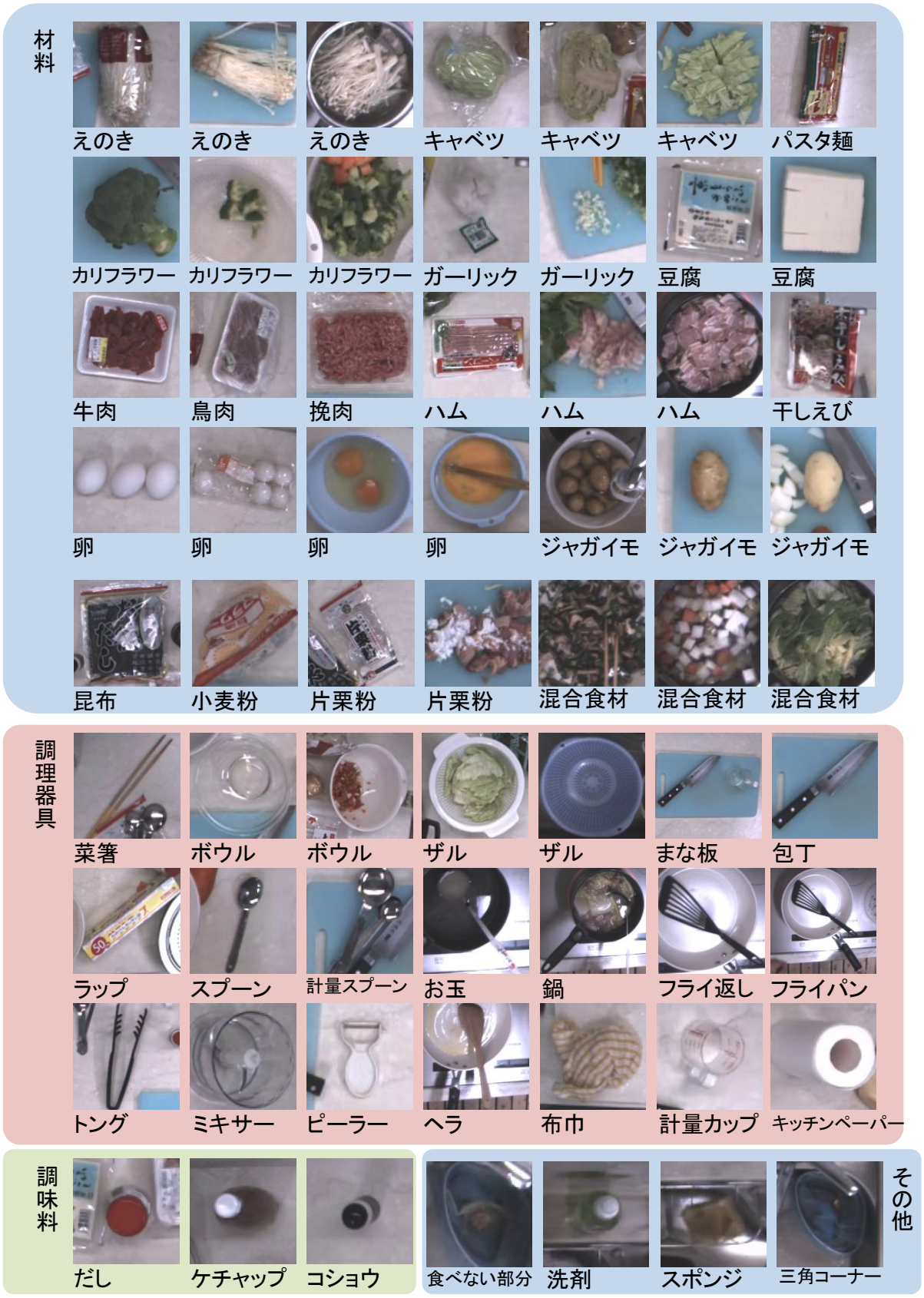

図 2　アノテーションによって得られた物体画像の例

Fig. 2　Examples of image data obtained through the annotation

よって正方形領域に画像を切り出し，256x256pixels へとリサイズする必要がある．ただし小さなボトルの調味料などは拡大による画像劣化の影響が予想されるため，切り出す正方形領域の辺の長さの最小値を設定して切り出しを行った．最小値を大きくすると，周辺にある他の物体が正方形領域に含まれてしまう恐れがある．このため，拡大による劣化が少なく，かつ，周

辺物体が含まれ肉い値として切り出す際の正方形領域の辺の最小値を 128pixels に設定した．

4.3 交差検定

前述の通り，調理作業における物体認識では，レシピ毎に語彙が異なる．また，作業者は調理中，何度も同じ食材の個体に接触する．このため，向きや姿勢が異なるものの，同一の調理映像から得られた食材の画像データは同一個体である場合が多い．そのため，本研究では同一の調理映像から得られた画像群を 1 つのグループとし，そのグループのための識別器を，他の調理映像から得られた画像群を用いて学習するという形式で交差検定を行った．学習にあたっては，ある調理映像でのみ使われた調理器具なども存在している．他の調理で用いられなかった調理器具等に関しては，学習データがそもそも存在しないため，認識の集計対象外として取り扱った．なお，今回の発表までにアノテーションの完了が間に合わなかった映像や，KUSK Dataset と同じキッチンで撮影された他の映像から収集されたサンプルも学習データには含まれている．

4.4 認識結果

上記の条件で実験を行った結果を表 1 に示す．なお置かれた物体は取られる直前まで同一の姿勢であることから，学習および評価の過程においては重複を避けるため物体が置かれた際のサンプルのみを用いている．これらの物体領域画像は合計 5943 個あり，KUSK Object Dataset の中で公開している．

材料，調味料，調理器具の分類は料理オントロジーに従った．オントロジーに存在しない物体のうち，混合食材はレシピ毎に別のクラスであるとしながら，材料として集計を行った．また，それ以外は全て背景クラスとして取り扱い，全種類の統計情報以外には含めていない．

KUSK Dataset では同一のキッチンで観測を行っているため，同一カテゴリの調理器具や調味料は数種類しか登場せず，個体識別問題に近い状況となっている．この結果，これらの物体は比較的高い精度での認識が達成されている．一方で，食材は前節で述べたように同じ食材でも様々な状態で現れるため，調味料や調理器具よりもかなり低い精度となっている．

作業の流れなどを利用して物体の認識精度向上を目指す場合には，正解となるカテゴリが認識結果の上位にいることが重要となる．そこで各レシピにおける累積照合特性曲線 (Cumulative Match Characteristic Curve: CMC Curve) を図 3 に示す．累積照合特性曲線の傾きを見ると，物体の種類によらず，上位 3, 4 個程度までは効率よく CMC のスコアが上昇している．上位 5 位までに正解が含まれる割合は材料，調味料，調理器具それぞれについて，0.865，0.875，0.958 となり，全種類の平均においては，0.912 となった．

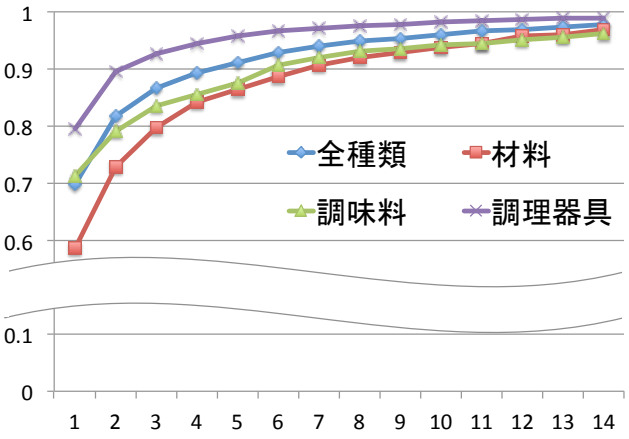

図 3 物体の種類ごとの累積照合特性曲線

Fig. 3 CMC Curve for each category group.

5. おわりに

本稿では，自然言語処理との連携を想定した KUSK Dataset における，調理台上の物体への接触履歴のアノテーションと，それらに対する CNN による画像認識結果のデータ提供について報告を行った．調理作業において，特に食材は様々な様態で観測されるため，CNN などの最新の物体認識手法を持ってしても，その精度は十分とはいえない．本研究で提供するデータセットは，実際の調理作業を観測したデータに加えて，作業中の物体への接触時刻と，その物体の種類の両方を広く公開することで，作業の文脈情報を用いた高度な物体認識や，レシピテキストの解析結果などを併用した分野融合的な研究を促進することを狙ったものである．

今後の課題としては，まだアノテーションが終わっていない映像へのアノテーションにより，各レシピ 2 本以上のアノテーション付きデータを作成することが挙げられる．また，人間の動きに注目した動作タグのアノテーションと同一個体の追跡結果のアノテーション，および追跡結果のアノテーションを利用した混合食材の細かな分類が挙げられる．

謝辞 本研究は JSPS 科研費 24240030 および 26280084 の助成を受けたものです．

文献

[1] 中山英樹，“深層畳み込みニューラルネットワークによる画像特徴抽出と転移学習，” 音声言語情報処理研究会 信学技報，pp.55–59, July 2015.

[2] C. Rashtchian, P. Young, M. Hodosh, and J. Hockenmaier, “Collecting image annotations using amazon’s mechanical turk,” Proc. of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk, pp.139–147, 2010.

[3] A. Farhadi, M. Hejrati, M.A. Sadeghi, P. Young, C. Rashtchian, J. Hockenmaier, and D. Forsyth, “Every picture tells a story: Generating sentences from images,” ECCV 2010, pp.15–29, Springer, 2010.

[4] Y. Yang, C.L. Teo, H. Daumé III, and Y. Aloimonos, “Corpus-guided sentence generation of natural images,” Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp.444–454, 2011.

[5] B.Z. Yao, X. Yang, L. Lin, M.W. Lee, and S.-C. Zhu, “I2t: Image parsing to text description,” Proceedings of the IEEE, vol.98, no.8, pp.1485–1508, 2010.

[6] C. Kong, D. Lin, M. Bansal, R. Urtasun, and S. Fidler, “What are you talking about? text-to-image coreference,” IEEE Conference on CVPRIEEE, pp.3558–3565 2014.

[7] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, “Show and tell: A neural image caption generator,” arXiv preprint arXiv:1411.4555, 2014.

[8] S. Venugopalan, H. Xu, J. Donahue, M. Rohrbach, R. Mooney, and K. Saenko, “Translating videos to natural lan-

表 1　レシピ毎の物体認識精度 (Cat.:カテゴリ数, Smpl.:サンプル数, Acc.:精度)

Table 1　Recognition Accuracy for each recipe (Cat.: # of Categories, Smpl.: # of Sampes, Acc.: Accuracy)

	全種類			材料			調味料			調理器具		
	Cat.	Smpls.	Acc.	Cat.	Smpls.	Acc.	Cat.	Smpls.	Acc.	Cat.	Smpls.	Acc.
2014RC01	24	363	70.25%	7	100	44.00%	5	54	75.93%	11	183	87.43%
2014RC02	24	172	68.02%	8	44	52.27%	4	18	83.33%	12	103	76.70%
2014RC03	25	331	69.79%	7	118	63.56%	5	42	69.05%	12	157	80.25%
2014RC04	21	96	63.54%	5	21	80.95%	6	8	75.00%	9	62	61.29%
2014RC05	21	141	75.76%	6	38	57.89%	3	25	64.00%	11	77	81.82%
2014RC06	24	139	52.52%	7	53	35.85%	5	16	56.25%	11	56	80.36%
2014RC07	17	181	66.30%	5	49	48.98%	3	30	86.67%	8	87	80.46%
2014RC08	15	30	56.67%	3	6	50.00%	4	0		7	16	87.50%
2014RC09	23	268	61.57%	6	66	46.97%	4	35	71.43%	12	145	75.17%
2014RC10	28	292	76.37%	7	107	66.36%	5	21	80.95%	15	159	83.02%
2014RC11	26	271	76.38%	7	66	65.15%	5	39	89.74%	13	159	81.13%
2014RC12	24	176	76.14%	7	56	60.71%	5	38	76.32%	11	79	88.61%
2014RC13	25	121	62.81%	6	21	66.67%	5	23	69.57%	13	72	63.89%
2014RC14	21	144	73.61%	5	51	62.75%	3	12	75.00%	12	81	80.25%
2014RC15	31	441	65.31%	8	119	54.62%	6	49	57.14%	16	264	73.11%
2014RC16	21	154	66.88%	4	43	60.47%	2	13	23.08%	14	94	78.72%
2014RC17	23	144	72.22%	7	49	48.98%	4	13	61.54%	11	82	87.80%
2014RC18	27	390	71.28%	6	141	70.21%	4	20	75.00%	16	215	76.28%
2014RC19	24	171	75.44%	5	29	55.17%	5	45	66.67%	13	94	87.23%
2014RC20	13	118	83.90%	4	25	88.00%	1	5	80.00%	7	84	86.90%
Total	84	4134	69.81%	38	1202	58.57%	23	506	71.34%	22	2269	79.51%

guage using deep recurrent neural networks," arXiv preprint arXiv:1412.4729, 2014.

[9] A. Hashimoto, S. Tetsuro, Y. Yamakata, S. Mori, and M. Minoh, "KUSK Dataset: Toward a direct understanding of recipe text and human cooking activity," Workshop on Smart Technology for Cooking and Eating Activities, pp.583–588, 2014.

[10] R. Hamada, I. Ide, S. Sakai, and H. Tanaka, "Structural analysis of cooking preparation steps," The Transactions of The Institute of Electronics, vol.85, no.1, pp.79–89, 2002.

[11] 山肩洋子，角所考，美濃導彦，"調理コンテンツの自動作成のためのレシピテキストと調理観測映像の対応付け," 電子情報通信学会論文誌 D，vol.90，no.10，pp.2817–2829，2007.

[12] A. Hashimoto, N. Mori, T. Funatomi, Y. Yamakata, K. Kakusho, and M. Minoh, "Smart Kitchen: A User Centric Cooking Support System," Proceedings of IPMU'08, pp.848–854, 2008.

[13] A. Hashimoto, J. Inoue, T. Funatomi, and M. Minoh, "How does user's access to object make hci smooth in recipe guidance?," Proc. of Human Computer Interaction International, pp.150–161, 2014.

[14] M. Tenorth, J. Bandouch, and M. Beetz, "The tum kitchen data set of everyday manipulation activities for motion tracking and action recognition," Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference onIEEE, pp.1089–1096 2009.

[15] F.D. laTorre, J. Hodgins, J. Montano, S. Valcarcel, R. Forcada, and J. Macey, "Guide to the carnegie mellon university multimodal activity (CMU-MMAC) database," Tech. report CMU-RI-TR-08-22, pp.1–17, Robotics Institute, Carnegie Mellon University, 2009.

[16] A. Shimada, K. Kondo, D. Deguchi, G. Morin, and H. Stern, "Kitchen scene context based gesture recognition: A contest in ICPR2012," Advances in Depth Image Analysis and Applications, pp.168–185, Springer, 2013.

[17] M. Rohrbach, S. Amin, M. Andriluka, and B. Schiele, "A database for fine grained activity detection of cooking activities," IEEE Conference on CVPRIEEE, pp.1194–1201 2012.

[18] H. Kuehne, A. Arslan, and T. Serre, "The language of actions: Recovering the syntax and semantics of goal-directed human activities," IEEE Conference on CVPRIEEE, pp.780–787 2014.

[19] S. Stein and S.J. McKenna, "Combining embedded accelerometers with computer vision for recognizing food preparation activities," Proc. of UbiComp 2013, Zurich, Switzerland, pp.729–738, ACM, Sept. 2013.

[20] S. Mori, H. Maeta, Y. Yamakata, and T. Sasada, "Flow graph corpus from recipe texts," Proceedings of the Nineth International Conference on Language Resources and Evaluation, pp.2370–2377, 2014.

[21] 橋本敦史，舩冨卓哉，中村和晃，美濃導彦，"机上物体検出を対象とした接触理由付けによる誤検出棄却," 電子情報通信学会論文誌 D，vol.95，no.12，pp.2113–2123，2012.

[22] 土居　他，"料理レシピと特許データベースからの料理オントロジーの構築," 電子情報通信学会技術研究報告, IMQ，vol.113，no.468，pp.37–42，2014.

[23] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," arXiv preprint arXiv:1408.5093, 2014.