Snappy: 写真を用いたロボットへの物体配置の指示手法

Snappy: Snapshot-based Robot Interaction for Arranging Objects

橋本 直　　Andrei Ostanin　　稲見 昌彦　　五十嵐 健夫*

Summary. 写真は，実世界に存在する物体の状態を記述するための道具として有用である．人間は写真から瞬時に「どこに何が置かれているか」という情報を読み取ることができる．この行為は，言語情報を用いることなく視覚的な情報のみで行うことができるため，しばしば異なる言語を話す人種間でのコミュニケーションにおいて効果を発揮する．我々は，この性質が人間・ロボット間でのインタラクションに対しても有効であると考え，人間がロボットに対して物体配置の指示を行う際のフロントエンドとして写真を用いる手法「Snappy」を提案する．Snappy では，ユーザはカメラを使って対象物を撮影することにより，その物体の配置情報を登録することができる．そして，その写真をロボットに提示することによって，写真に写っているものと同じ配置でロボットに物体を並べさせることができる．本研究では，ロボットが食器の配膳を行うシステムを実験的に構築し，その中でロボットに対して配膳の指示を行う手法として提案手法を実装し，その有効性について検証を行った．

1　はじめに

人間・ロボット間のインタフェースをデザインする際，人間同士で交わされるコミュニケーションを模倣することは 1 つの有効な手段である．その代表例として，音声認識やジェスチャを用いた方法があり，これまでも数多く研究されてきた．しかし，音声認識やジェスチャによる方法には，指示に曖昧さが含まれるという問題や，指示内容から視覚的な情報を再現するのが難しいという問題がある．我々の研究における目標は，このような問題を解決することができる新しいヒューマン・ロボット・インタフェースを開拓することである．

本稿では，物体の配置というタスクに着目し，ユーザがロボットに対して物体の並べ方を指示するためのインタフェースを提案する．ここで我々が扱う問題は，「どのようにしてロボットに並べ方を教えるか」である．音声やジェスチャによる方法であれば，「物体 A と物体 B をここに移動」と発話する方法や，対象物と目的地を指やレーザポインタで指し示す方法が考えられる [1][2][3]．しかし，それを実現するためにはいくつかの問題がある．音声認識の場合，個々の物体を指定する際に複数の呼び方が存在し得るという問題がある．例えば，「お皿」と呼ぶこともあれば，「ボウル」，「食器」などと呼ぶ可能性もある．このバリエーションは無数に存在するため，すべての名前をロボットが記憶する方法は現実的ではない．

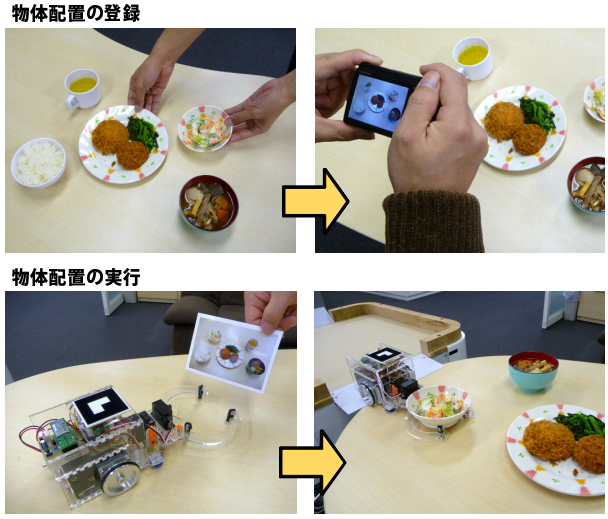

図 1. Snappy の利用イメージ

逆に，ロボットが認識可能な名前を人間が覚える方法では人間側の負担が大きくなってしまう．指によるジェスチャでは，指差し方向に複数の物体が存在した際に，どれが対象物かを特定する必要がある．また，これらの方法では，リファレンスとなる情報が残らないため，後に同じタスクを再度実行しようとしたときに，タスク内容を確認する手段がない．

そこで我々は，写真を物体配置タスクのリファレンスとして用いる手法「Snappy」を提案する．ユーザは写真を撮ることによって，ユーザが求めている結果を登録し，後に，その写真をロボットに提示することによって，ロボットに物体の並べ方を指示する．例えば，食卓に並んだカップや皿を撮影し，後日，その写真をロボットに見せれば，写真と同じ配置

* Sunao Hashimoto, JST ERATO 五十嵐デザインインタフェースプロジェクト, Andrei Ostanin, School of Computing, University of Utah, Masahiko Inami, 慶應義塾大学大学院メディアデザイン研究科, Takeo Igarashi, 東京大学大学院情報理工学系研究科

でロボットが配膳を行うことができる (図 1)．写真は，実世界に存在する物体の状態を記述するための道具として有用である．人間は写真から瞬時に「どこに何が置かれているか」という情報を読み取ることができる．この行為は，言語的な情報を付加することなく視覚的な情報のみで行うことができるため，しばしば異なる言語を話す人種間でのコミュニケーションにおいて効果を発揮する．本研究のアイデアは，この性質を人間とロボットのコミュニケーションに応用するものである．

この手法にはさまざまな利点がある．第 1 に，日常的に行われる写真を撮るという行為によって，自然かつ簡単にタスクの登録ができるという点が挙げられる．第 2 に，写真が物体配置タスクに必要な情報（どこに何を置くのか）を含んでいるという点が挙げられる．そして第 3 に，写真はタスクの内容を視覚的に表現しているメディアであるため，後にタスクを呼び出す際の良いリファレンスとなるという点が挙げられる．

本研究では，ロボットが食器の配膳を行うシステムを実験的に構築し，その中でロボットへの指示を行う手法として提案手法を実装した．本稿では提案手法の概要とプロトタイプシステムについて説明し，パイロットスタディの結果について報告する．

2 Snappy

2.1 シナリオ

Snappy を用いたシナリオの例を以下に示す．初日，ユーザは夕食のために食事の準備を行っていた．料理を終えたユーザは，食事を食器に盛り付け，食卓に並べた．その後，食卓に並んだ食事をカメラで撮影し，撮った写真を印刷してレシピブックの中に挟んだ．2 週間後，ユーザは同じ食事を作っていた．前回と同じレイアウトで食器を並べたいと考えたユーザは，レシピブックから写真を取り出した．キッチンにいる小型配膳ロボットに写真を見せると，ロボットは写真と同じレイアウトで食器を食卓に並べた．ロボットが配膳を行っている間，ユーザはキッチンの片付けを行うことができた．

2.2 ユーザインタラクション

Snappy には，登録フェーズと実行フェーズの 2 つのフェーズがある．それぞれについて説明する．

(1) 登録フェーズ

初回時，ユーザは実際に自分の手で物体を並べて所望のレイアウトを作成し，それをカメラを用いて撮影する．ユーザは，配置を覚えさせたい対象をカメラのフレームに収めることによって選択する．ユーザがシャッターボタンを押すと，写真とともに，フレーム内にある物体の ID，位置，姿勢が記録される．これらの情報は，無線でサーバに送信され，管理される．

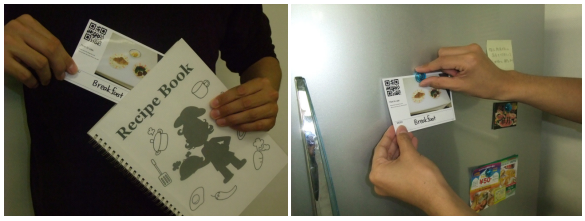

図 2. 写真をレシピブック（左）や冷蔵庫（右）に貼付する

(2) 実行フェーズ

登録フェーズで撮影した写真を使って，ロボットにタスクの実行を指示する．指示方法として，以下の 3 種類を提案する．(a) 印刷された写真を，テーブルや壁に設置された専用のリーダにかざす，(b) 印刷された写真をロボットの目に見せる，(c) 端末の GUI 上で写真を選択する．(a) と (b) では，写真の裏面に，物体配置に関する情報を記録した 2 次元コードを印刷する．装置がコードを読み取った際にタスクが実行される．(b) において，複数のロボットが存在した場合は，写真を読み取ったロボットが，ユーザからタスクの指示がきたことを他のロボットに伝え，協調して作業を行う．

印刷された写真を使う方法は，データが物理的な実体を持っているため，日常生活における利用においてさまざまな利点がある．第 1 に，複数の写真の中から素早く目的の写真を選び出すことができ，システムに「見せる」だけでタスクを実行できる点が挙げられる．この操作は，コンピュータ上のアプリケーションを起動する手間なしに行うことができる．第 2 に，写真に対してペンでメモ（例：「いつもの朝食」）を自由に記入できることが挙げられる．そして第 3 に，写真を冷蔵庫のドアやレシピブックの中など，日常生活のなかでアクセスする頻度が高い場所に貼付しておくことができるという利点がある (図 2)．

2.3 想定環境

ロボットの種類や大きさは特に限定せず，ヒューマノイドロボットや，小型の卓上ロボットなどさまざまなロボットを想定する．また，システムが，ロボットおよび運搬可能な物体の ID，位置，姿勢をトラッキングできているものとする．ユーザが写真を撮影する際に使用するカメラは，フレーム内にある物体が何であるかを認識できるものとする．これを実現する方法として，カメラ上での画像認識，またはカメラの視野範囲の情報と位置・姿勢のトラッキングによって算出する方法が挙げられる．

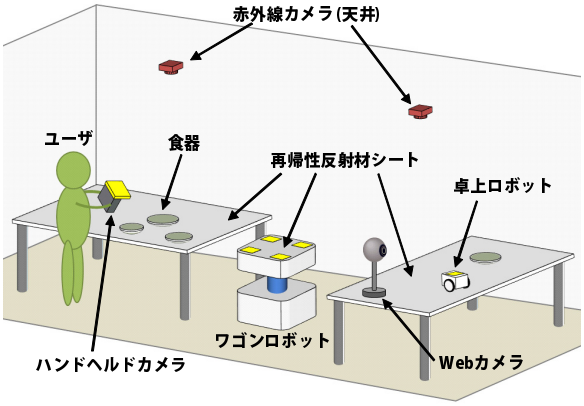

図 3. プロトタイプのシステム構成

3 プロトタイプ

提案手法を用いたプロトタイプとして，ワゴンロボットと小型の卓上ロボットを用いてキッチンからダイニングにあるテーブルまで食器を運搬し，ユーザが指示した場所に配置するシステムを構築した．システム構成を図 3 に示す．

3.1 ロボット

大きさ 130 × 100 × 70mm の小型の卓上ロボットを独自に開発した．皿やカップのような軽い物体（最大 300g）であれば押して運搬することができる．移動機構として，2 個の車輪とキャスターを備え，各車輪はステッピングモータによって駆動する．ロボットには Bluetooth による無線通信を行うマイコンを搭載しており，サーバからの遠隔操作によって制御される．物体（食器）の運搬アルゴリズムは [4] を用いた．

我々は，異なるテーブル間で卓上ロボットと物体の移動を行うために，ワゴンロボットも作成した．ワゴンロボットはリフタを備え，468〜930mm の高さで昇降させることができる．天板の大きさは 470 × 470mm で，最大積載量は 15kg である．ワゴンロボットは 4 個の車輪を持ち，全方位に移動することができる．ワゴンロボットは，サーバからの Wi-Fi 接続によって遠隔制御される．

3.2 物体のトラッキング

赤外線カメラ（NaturalPoint Optitrack FLEX: V100）を天井に 2 個設置した．ロボットとユーザが使用するカメラの位置計測にはビジュアルマーカ（ARToolKit[1]）を用いた．マーカは，下地は白色の再生反射材，パターン部分は白色の赤外線吸収素材で構成されている．このマーカは，人間の目にはただの白い板に見えるが，赤外線カメラからはクリアにパターンを認識することができる．食器の認識に

[1] http://www.hitl.washington.edu/artoolkit/

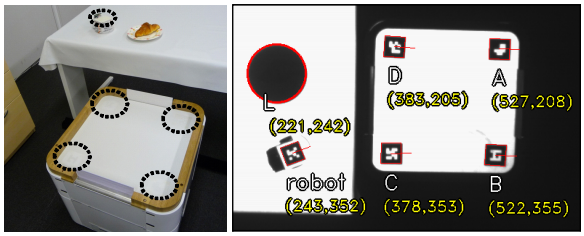

図 4. トラッキングシステムの外観（左）と検出結果（右）．点線内がマーカ．

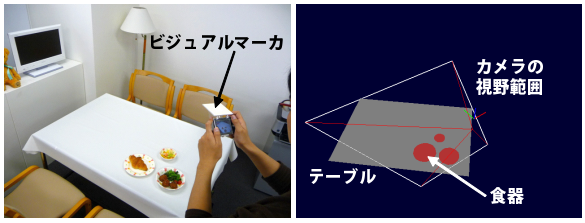

図 5. カメラフレーム内にある食器の認識

は円形ハフ変換を用いた．皿の大きさによって皿の種類を識別する．テーブルクロスに白色の再帰性反射材を用いることで，皿の検出性能を高めた．皿の上に食べ物を置いても検出には支障はなない．物体のトラッキングの履歴はサーバに記録される．トラッキングシステムを図 4 に示す．

3.3 ハンドヘルドカメラ

ユーザが食器の配置を登録するためのハンドヘルドカメラとしてディジタルカメラ（富士フイルム FinePix F100fd）を用いた．このカメラの位置と姿勢はビジュアルマーカを用いて計測される．ワイヤレス SD カード（Eye-Fi）を内蔵し，撮影された写真は Wi-Fi 通信によって即時サーバに送信される．サーバでは，写真のタイムスタンプの情報を用いて，位置計測の履歴の中から撮影時のカメラの位置姿勢を割り出し，カメラの視野範囲の情報から，カメラフレーム内にどの食器が写っていたかを推定する．カメラフレーム内にある食器の認識の様子を図 5 に示す．

3.4 サーバ

サーバは，物体の位置計測，ロボットの制御，および写真の管理を行う．ユーザが使用するハンドヘルドカメラから写真が送られてくると，カメラフレーム内にある食器を推定し，食器の絶対位置情報と写真を対応付ける．同時に，写真のシリアルナンバーを記述した 2 次元コードを生成する．サーバはこの 2 次元コードを付与した写真を印刷する．印刷された写真には，メモ欄も付与される．印刷された写真を図 6 に示す．本プロトタイプでは，写真に印刷された 2 次元コードを認識する方法として，Web カ

図 6. 印刷された写真

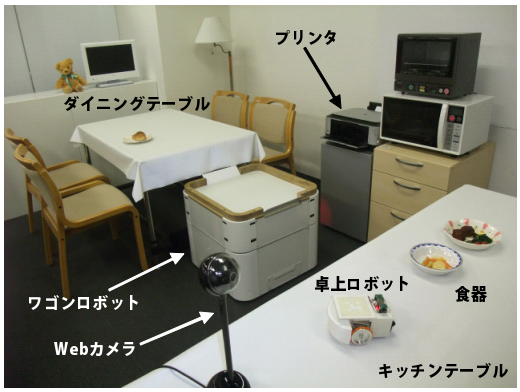

図 7. 実験環境

メラ（Logicool Qcam Orbit AF）を用いた．キッチンにいるユーザが使用するという想定で，Web カメラはキッチンにあるテーブル上に設置した．Web カメラが 2 次元コードを認識すると，その旨を知らせる音が鳴るようにした．

3.5 システムの仕様と制約

4 人がけのダイニングテーブルと，料理を載せた食器が置かれるキッチンテーブルが置かれた環境内にプロトタイプを構築した．実験環境を図 7 に示す．ダイニングテーブルの大きさは $1200 \times 800 \times 700$mm，キッチンテーブルの大きさは $700 \times 1600 \times 720$mm であり，テーブル間の距離は 1.6m である．ワゴンロボットはテーブル間を往復する．ワゴンロボットのリフタの昇降を行う位置は既知とした．配膳を行う際の食器の初期位置は，キッチンテーブルの中央付近とした．食器は，奥に置かれるものから順に運搬される．現在のプロトタイプでは，食器の数は大・中・小の 3 枚に限定している．

4 実験結果

開発したプロトタイプにおける実験結果を図 8 に示す．登録フェーズにおいて，ユーザは自分の手でダイニングテーブルに料理を載せた食器を並べ (a-1)，その後ディジタルカメラを用いて写真を撮影した (a-2)．撮影された写真は，サーバに送信され，自動的に印刷された (a-3)．実行フェーズにおいて，ユーザは調理した料理が載った 3 枚の皿をキッチンテーブルの中央に置いた (b-1)．その後，ユーザは登録フェーズで撮影した写真をキッチンテーブルに置かれている Web カメラにかざした (b-2)．写真に印刷されている 2 次元コードが認識されると，システムは食器配膳のタスクを開始した．食器配膳の流れは以下の通りである．まず，卓上ロボットが食器の 1 つをキッチンテーブルの端まで移動させる．次に，ワゴンロボットがキッチンテーブルに接近し，天板がテーブルと同じ高さになるようにリフトを上昇させる．天板上のタラップがテーブルに渡されると，卓上ロボットが食器をワゴンロボット上に運び入れる (b-3)．その後，タラップを上げ，リフトを下降させ，食器と卓上ロボットを載せた状態でダイニングテーブルに向かって移動する (b-4)．ワゴンロボットがダイニングテーブルに高さを合わせると，卓上ロボットは食器を押し出す (b-5)．卓上ロボットが目標位置に食器を運搬すると，ワゴンロボットを使って再びキッチンテーブルに戻る (b-6)．このシーケンスは各食器に対して行われる．3 枚の食器の配膳を行った結果を (b-7) に示す．3 枚の食器を指示通りの位置に運搬するのにおよそ 14 分を要した．

5 パイロットスタディ

プロトタイプのテストを 3 名のユーザに対して行った．システムの説明を行い，登録フェーズと実行フェーズの両方を体験させ，その感想を述べてもらった．

すべてのユーザは提案手法の使い方を理解し，正しく使うことができた．どのユーザも提案手法は簡単に使えると答え，食器の配膳を指示する方法として妥当だと答えた．

ユーザ全員から，印刷した写真をキッチンの食器棚やキッチンカウンターの近くで，カードホルダーに入れて収納したいという感想が得られた．ユーザの 1 人は「印刷した写真は選びやすいが，20 枚以上ある場合はディスプレイ上で選択したい」と答えた．また，ユーザ全員が撮影日時が写真にプリントされるべきだと答えた．

ユーザに対し，「もしマウスやタッチ操作で食器の並べ方を指示できるような GUI があるならば，GUI と提案手法のどちらを使いたいか？」と質問した結果，提案手法のほうがシンプルな操作で指示できることを理由に，全員が提案手法を選んだ．また，ユーザからは「実物の食器を持っているならば，実物を使って配置を指示したい」「物体の配置を教えるのにいちいちコンピュータを使うのは面倒」という意見が得られた．

提案手法の他のタスクへの応用可能性について質

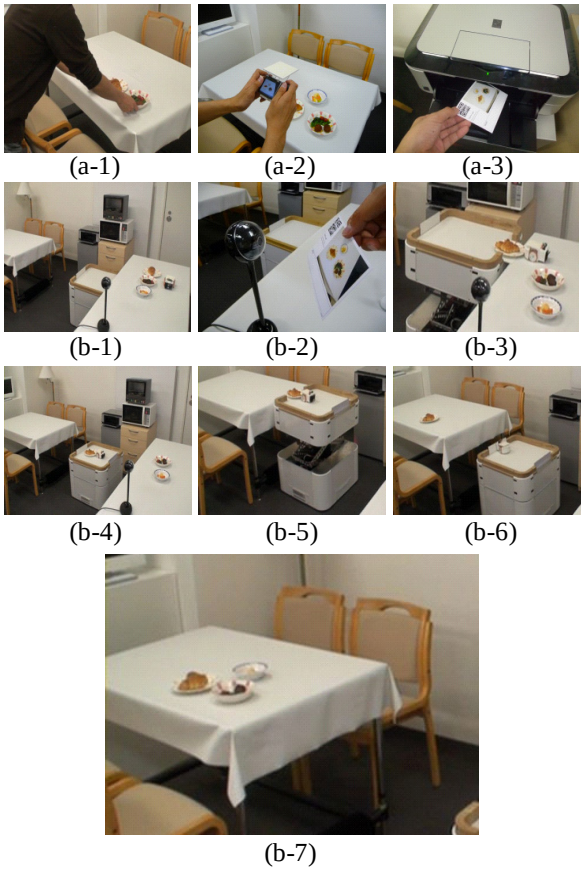

図 8. 実験結果

問した結果,「本棚，クローゼット，おもちゃ箱などの整理に使いたい」「会議室の机と椅子の並べ方を指示するのに使いたい」という意見が得られた．また，食器の配膳について,「大規模なパーティを実施する際に，同じ配置でたくさんの食器を並べるときに使いたい」という意見も得られた．

プロトタイプへの不満として，ロボットの遅さが指摘された．ユーザらは，1 人分の食器を並べるのに待てる時間は 1〜2 分だと答えた．2 人のユーザは，写真を読み取る装置として，Web カメラよりも，写真を挿入できるカードリーダのほうが良いと答えた．また，ロボットの目に写真を見せるという指示方法の可能性について尋ねた結果,「ロボットに見せるのであれば，写真を認識したとわかるように，ロボットの目が機械的に写真をトラッキングしてほしい」と答えた．

6 関連研究

ある時点の状態を記録し，後でそれを再現するための手法が提案されている．Time-Machine Computing[5] はユーザがコンピュータ上で行った操作の履歴を記録し，ユーザが指定した過去の状態を復元可能にするシステムである．Suzuki らは写真を使って，ネットワークに接続された家電を制御するインタフェースを提案した [6]．ユーザは，照明の明るさや，ビデオの再生状態などの家電の状態を，それらの写真を撮ることによって記録することができる．後に，ユーザは撮影した写真を PDA 上で選択することによって，家電の状態を再現することができる．我々の研究は，実世界の物体の配置を記録し，それを後からロボットを使って再現させるという点で異なる．

カメラを援用したインタフェースとして，代表的なものに NaviCam[7] がある．NaviCam では，実物体を撮影すると，その物体に関連する情報がオーバーレイ表示される．Hosoi らは，カメラを使ってロボットを直感的に操作するインタフェースを提案した [8]．ユーザは，ハンドヘルドデバイスに搭載されたカメラを用いてロボットを撮影し，移動させたい方向にカメラを向けることによってロボットを移動させることができる．また，綾塚らは，カメラを備えたコンピュータを用い，カメラで見ているものとの接続を確立する gaze-link メタファ [9] を提案している．我々の研究は，カメラを用いて実世界とインタラクションをするという点で，これらの研究と共通しているが，物体の物理的な状態を登録し，後からそれを再現するという点で異なる．

ペーパーメディアを用いたインタラクションはさまざまなものが提案されている [10][11][12]．Icon-Stickers[13] は，バーコードとアイコン画像が印刷された付箋紙を使って，実世界とコンピュータ内のデータを関連付けるシステムである．Sony 社のエンタテインメントロボット AIBO では，特殊なパターンが印刷されたカードをロボットに見せることによって，ロボットに特定の動作をさせることができる．Magic Cards[14] は，家庭用ロボットを操作するための紙ベースのインタフェースである．コマンドが書かれたカードを空間中に配置することによってロボットにタスクを指示することができる．これらの研究は，紙を用いたタンジブルなインタフェースが，GUI よりも，実世界におけるタスクの指示に適していることを示している．

7 まとめと今後の課題

本稿では，カメラと写真を使ってロボットに対して物体の並べ方を指示する方法を提案した．提案手法は，実世界の一部分の状態をワンボタンアクションで登録し，それを再度呼び出すことができる．このコンセプトを検証するために，ロボットが食器の配膳を行うシステムを試験的に実装した．

現在のシステムでは，登録時にユーザが意図した物体が登録できているかを確認するためのフィードバックや，実行時にロボットがタスクを遂行可能かどうかを確認するためのフィードバックがない．今後はこれらの要素について議論し，開発を進めていく．

参考文献

[1] R. A. Bolt. “Put-That-There”: Voice and Gesture at the Graphics Interface. In *Proceedings of SIGGRAPH'80*, pp. 262–270, 1980.

[2] C. C. Kemp, C. D. Anderson, H. Nguyen, A. J. Trevor, and Z. Xu. A point-and-click interface for the real world : Laser designation of objects for mobile manipulation. In *Proceedings of HRI'08*, pp. 241–248, 2008.

[3] K. Ishii, S. Zhao, M. Inami, T. Igarashi and M. Imai. Designing Laser Gesture Interface for Robot Control. In *Proceedings of INTERACT'09*, pp. 479–492, 2009.

[4] T. Igarashi, Y. Kamiyama, M. Inami. A Dipole Field for Object Delivery by Pushing on a Flat Surface. In *Proceedings of ICRA'10*, 2010.

[5] J. Rekimoto. Time-Machine Computing: A Time-Centric Approach for the Information Environment. In *Proceedings of UIST'99*, pp. 45–54, 1999.

[6] G. Suzuki, D. Maruyama, T. Koda, S. Aoki, I. Takeshi, K. Takashio and H. Tokuda. u-Photo Tools: Photo-based Application Framework for Controlling Networked Appliances and Sensors. In *Proceedings of UbiComp'04*, 2004.

[7] J. Rekimoto and K. Nagao. The World through the Computer: Computer Augmented Interaction with Real World Environments. In *Proceedings of UIST'95*, pp. 29–36, 1995.

[8] K. Hosoi and M. Sugimoto. A Mobile Interface for Robot Control from a User's Viewpoint. In *Proceedings of ROBIO'06*, pp. 908–913, 2006.

[9] 綾塚祐二, 松下伸行, 暦本純一, 実世界指向ユーザインタフェースにおける「見ているものに接続する」というメタファ, 情報処理学会論文誌, Vol. 42, No. 6, pp. 1330–1337, 2001.

[10] W. Johnson, H. Jellinek, L. Klotz, R. Rao and S. K. Card. Bridging the Paper and Electronic Worlds: The Paper User Interface. In *Proceedings of CHI'93*, pp. 507–512, 1993.

[11] P. Wellner. Interacting with paper on the DigitalDesk. Communications of the ACM 1993, 36, 7, pp. 87–96, 1993.

[12] C. Liao, F. Guimbretiere and K. Hinckley. PapierCraft: A Command System for Interactive Paper. In *Proceedings of UIST'05*, pp. 241–244, 2005.

[13] I. Siio and Y. Mima. IconStickers: Converting computer icons into real paper icons. In *Proceedings of HCI International'99*, pp. 271–275, 1999.

[14] S. Zhao, K. Nakamura, K. Ishii and T. Igarashi. Magic Cards: A Paper Tag Interface for Implicit Robot Control. In *Proceedings of CHI'09*, pp. 173–182, 2009.

未来ビジョン

カメラはワンボタンアクションで実世界の一部分をキャプチャすることができるツールである．また，写真は言葉やテキストでは伝えにくいことを表現できるツールである．我々はこれらの性質に着目し，カメラや写真を援用して，言葉やジェスチャでは表現しにくい複雑なタスクの指示をロボットに対して行えるようなインタフェースを実現する方法を模索している．

本稿では，静的な物体配置を指示する手法としてSnappyを提案した．今回実装した１人分の食器の配膳はあくまで一例であり，将来的には，１人分の食器の配置を記録した写真をベースに，同じ配置作業を多人数分実行するような「実世界コピー＆ペースト」を実現することを目標としている．また，本稿ではモノを移動させるためのロボットが存在する状況を想定していたが，我々はモノとロボットが融合し，モノ自身が提示された写真と同じ配置になるべく移動する，というデザインも視野に入れている．例えば，会議室の机や椅子を講演会用の配置に並べ替えたいというときに，ユーザが前回の講演会のときの写真をリーダに読ませると，アクチュエータが内蔵された机や椅子が自ら移動し，ユーザが会議室を利用する時間には整列が完了する．

我々は，連続した複数の写真を用いれば，シーケンシャルなタスクの指示も行えるだろうと考えている．例えばインスタントコーヒーを作るというタスクは，カップにコーヒーの粉を入れる，カップをポットの下に置く，ポットのお湯を注ぐ，などのいくつかのシーンによって表現することができる．実際にコーヒーを作りながら，ステップごとにカメラで撮影し，その写真群をタスクの手順書として保存しておけば，後日，ロボットに教えた通りの手順でコーヒーを作らせることができると考える．

将来的に，画像中の物体認識技術はより高精度かつ汎用的なものになると予想するが，コンピュータが写真からシーン理解を行うのは困難であると考える．タスクの実行に必要な情報の選択は人間が行う必要がある．また，物体の移動や破棄などのアクションについては写真という静的メディアでは表現しにくい．そこで，写真だけでは伝えきれない部分や，コンピュータが理解することができない部分についてはユーザが適宜アノテーションを付加することによって解決することを検討している．