さりげなく作業支援を行う映像メディア
Video-Based Interactive Media for Gently Giving Instructions

小阪 拓也 † **中村 裕一** ‡ **大田 友一** † **亀田 能成** †
Takuya Kosaka Yuichi Nakamura Yuichi Ohta Yoshinari Kameda

1. はじめに

料理やプラモデルの組み立てのような作業を行っている場合，何をすべきかわからないといったことがしばしば起こる．そのような場合，もし熟練した人間の先生がいたならば，そのユーザの状況を判断して，状況に適したわかりやすい説明をしてくれるだろう．

本研究では，「さりげなく作業支援を行う映像メディア」を提案する．従来の「電子マニュアル」とは異なり，ユーザの状態に合わせた情報を提示してくれること，次に何をするかという主導権がユーザにあることが本研究の特徴である．現在は簡単な机上作業である，ブロックの組み付け作業を対象として研究を進めている．

2. さりげなく作業支援を行う映像メディア

本研究で必要な要素技術は以下の 4 つである．図 1 にその関係を示す．

1. 教示映像へのインデキシング
2. 物体とユーザの動作の認識
3. ユーザの現在の作業状況の推定
4. ユーザの作業状況に関連した情報の提示

このうち教示映像へのインデキシングについては既に我々は QEVICO[1] で提案している．物体と動作の認識は，現在のところ簡単な認識処理によって行っている段階であるが，物体追跡に関する研究 [2] の研究成果を利用する予定である．

一方，ユーザの現在の作業状況の推定部分については，検出した物体と，教示映像中のインデクスとを対応付けることで実現する．その際にユーザにできるだけ干渉しないように，可能な限りシステム内部で処理し，ユーザの状況が不明になった場合にのみ，ユーザに問い合わせる．これにより，ユーザの作業や思考を邪魔せずに作業支援することをめざす．以降，本稿では上記 2-4 の要素について述べる．

3. ユーザの作業状況の推定

3.1 定義と記述方法

まず，以下のような概念を定義する．

作業: 教示の目標である．タスクの集合により記述する．

タスク: 目的とした作業を達成するために行う動作．タスクは一連のアクションと関連する物体で構成される．本稿では，基本的な動作である「移動」「組み付け」の 2 種類を想定している．

アクション: 画像処理による認識に対応したプリミティブな動作を指す．本稿では，「持ち上げる」「置く」「(物体同士が) 接触する」を想定している．各アクションは，ID，動作名，関連する物体で記述する．

†筑波大学 大学院 システム情報学研究科
‡京都大学 学術情報メディアセンター

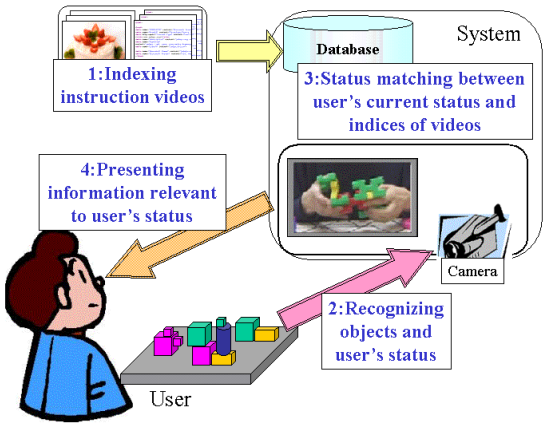

図 1: システムの概要

物体: 動作の対象となる具体的形状をもつもの．ID，画像特徴 (色，形等) で記述する．

3.2 ユーザの現在の作業状況の推定

ユーザの作業状況の推定方法の概要を示す．

(a) 教示映像のインデキシング情報を読み込み，作業のタスクと物体に関する情報を予め取得する．

(b) ユーザが何らかのアクションを行った場合，アクションと関連する物体の情報を時系列順に記録する．

(c) (a) と (b) を照合することで，ユーザの現在の作業状況を推定する．その際の照合には以下の類似度を用いる．

物体の類似度 $S(O_i, O_j)$**:** 色や形状等の画像特徴に基づく．本稿では，物体領域の色ヒストグラムを比較することで，類似度を求めている．

アクションの類似度 $T(A_i, A_j)$**:** 動作名の類似度と関係する物体の類似度の積で計算する．アクション $A_i(N_i, O_{i1}, O_{i2}, ..., O_{in})$ と，アクション $A_j(N_j, O_{j1}, O_{j2}, ..., O_{jn})$ との類似度 $T(A_i, A_j)$ は以下の式で求められる．ここで，N_i は動作名，O_{ik} は関連する物体を意味する．

$$T(A_i, A_j) = \delta(N_i, N_j)(\prod_{k=1}^{n} S(O_{ik}, O_{jk}))^{\frac{1}{n}} \quad (1)$$

$$\delta(X, Y) = \begin{cases} 1 & (X = Y) \\ 0 & (X \neq Y) \end{cases} \quad (2)$$

作業状況の推定は，以下で述べる部分探索と全探索の組合せで実施される．

部分探索: システムは，教示映像中のタスクと対応するユーザの一連のアクションを探す．その際，DP マッチングを適用することで，教示映像に記録されたとおりに行動していない場合でも，柔軟に対応す

る．この処理では，対応付けられる全ての可能性を求める．

全探索: タスクの一貫性を考慮に入れることで，起こりうる一連のタスクを推定する．本研究では，部分探索の結果を基に深さ優先探索を行うことで，起こりうるタスクの組み合わせを求める．

3.3 画像処理によるユーザ動作の認識

ユーザのアクションは物体に対して行われるため，把持物体の状態を追跡することで，ユーザのアクションを認識する．

フレーム間差分とテンプレートマッチングで求められた領域を動領域とし，動領域から肌色部分を除いた領域を物体領域とする．

各アクションの認識方法を以下に示す．

持ち上げる: 物体領域の中心座標が一定の高さよりも高くなった場合に認識する．

接触する: 二つの物体領域の中心座標が一定の高さよりも高く，さらにそれらの距離が接近している場合に認識する．

置く: 物体領域の中心座標が一定の高さよりも低くなった場合に認識する．

4. ユーザへの問い合わせ

ユーザの個々のアクションや物体の対応付けには常に曖昧性があるため，全探索の際，当てはまるタスクの組み合わせの数が爆発的に大きくなるという問題がある．

そこで本研究では，ユーザへの問い合わせにより曖昧性の解消を図る．まず，システムは最も曖昧な部分を調べる．その際，教示映像中の物体やタスクと対応付けられた数が曖昧性を示す一つの指標となる．そして，曖昧性が一定の閾値を超えた場合において，物体やタスクの名前を問い合わせる．ユーザからの回答を反映させることで，現場の物体/タスクと教示映像中の物体/タスクとを正しく対応付けることができ，曖昧性を大幅に削減することができる．

5. 実験

本システムの有効性を示すために，簡単な実験を行った．まず，アクションの認識精度，部分探索によるタスクの認識精度を調べた．組み立て作業の例として，ブロックを組み付けて「車」を作るという作業を対象とした．その作業は，50 個の物体と 30 のタスクで構成される．

画像認識の結果，ユーザが全ての作業を終えるまでの間に，70 のユーザのアクションを検出することができたが，50 パーセントのアクションは誤検出であった．

次に，画像認識の結果が正しく行われたものとして，作業状況の推定を行ったところ，図 2 に示されるような precision recall グラフを得た．このグラフは，部分探索の際の類似度の評価について，照合されたかどうかを判断するための閾値を変化させて得たものである．

また，ユーザとのインタラクションにより，解釈の曖昧性を減らす方法の有効性を調べた．図 3 は，検出された幾つかの物体とそれに対応付けられた物体の個数を示す．Num(1) は曖昧性を減らす以前の結果である．ID9 の物体は 19 個の解釈候補があり，最も曖昧性の高い物体である．ユーザに問い合わせることで，システムは物体同士の正しい対応付けを得て，その結果，図 3 の Num(2) に示される改善結果を得ることができた．

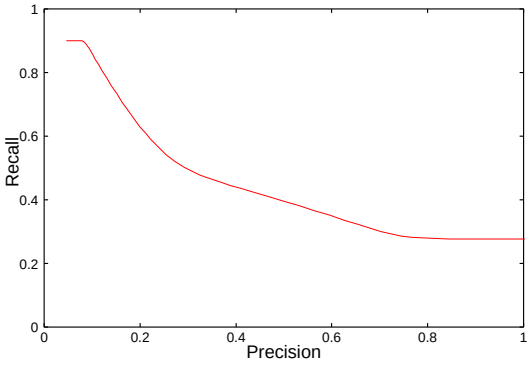

図 2: Precision recall グラフ

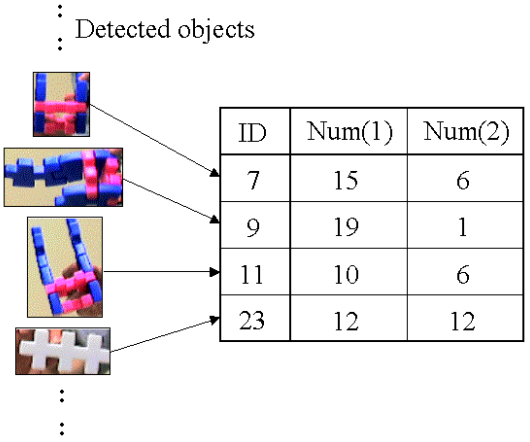

図 3: 対応付けられた物体の数

6. おわりに

本稿では「さりげなく作業支援を行う映像メディア」を提案した．ユーザの状態を認識するための枠組，またユーザに問い合わせることで曖昧性を減らす方法について述べた．今後の課題は，物体追跡システムと組み合わせることで物体認識の精度を高める，また実際にユーザがリアルタイムで利用できるシステムを構築することである．

参考文献

[1] H.IZUNO, et al: QUEVICO:A Framework for Video-Based Interactive Media. Proc. Int'l Workshop on Intelligent Media Technology for Communicative Reality (2002) pp.6-11.

[2] M.ITOH, et al: Simple and Robust Tracking of Hands and Objects for Video-based Multimedia Production. IEEE Conf. on MFI (2003) pp.252-257.