映像の蓄積と高度利用:
使える映像を集めることから始めよう

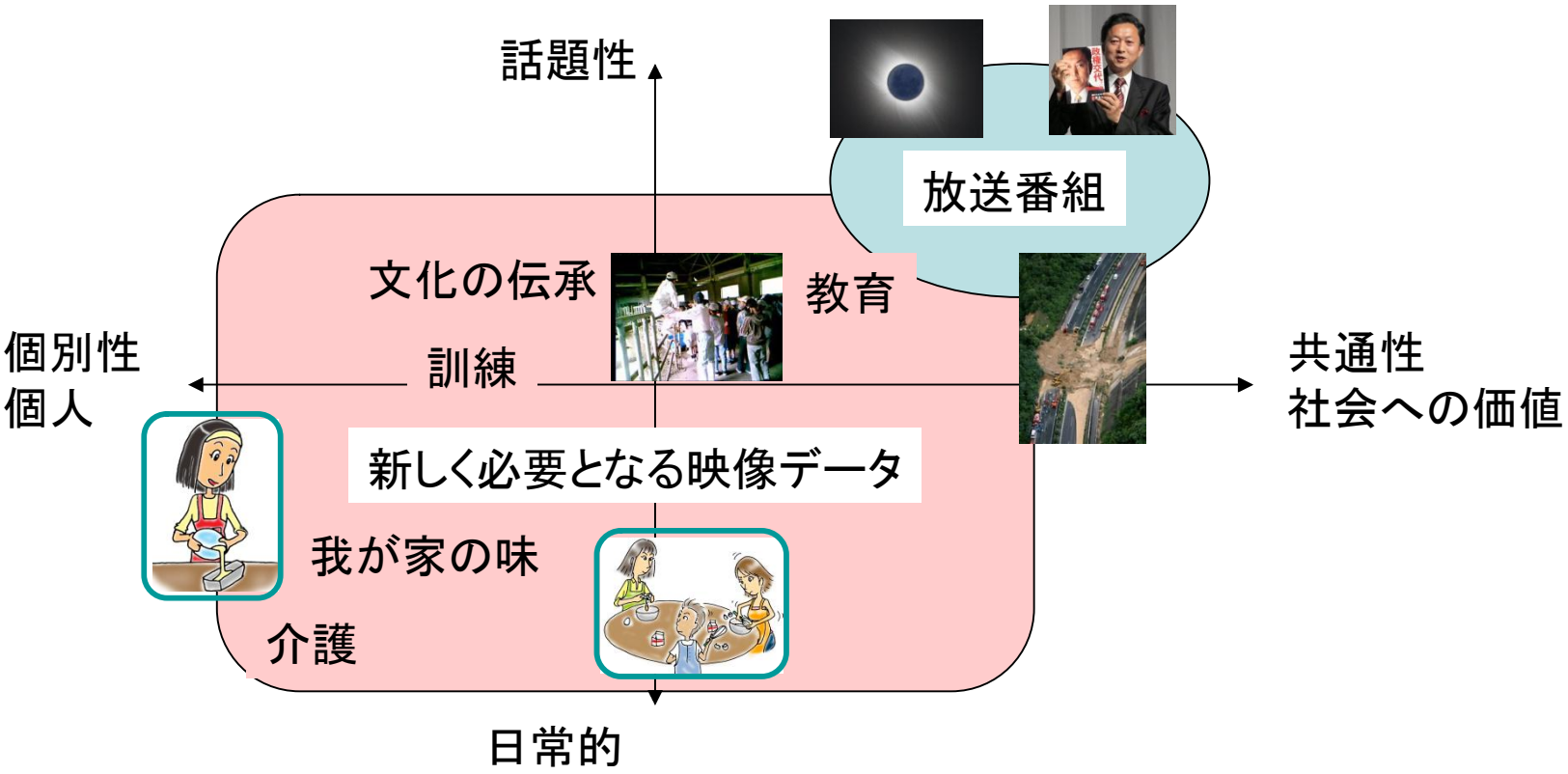
京都大学 学術情報メディアセンター
中村 裕一

概要

大量のデータを知識として蓄積し，適切なデータを適切なタイミングで得たり，大量のデータから新しい知識を自動的に得ることが行われ始めています．しかし，映像を大量に集めただけでは，そのような目的に簡単に使えるデータとはならないのが現状です．そのための仕組み，特にデータを収集する際の自動的な処理について考えます．具体的に映像を利用する場面や目的，そして，そのためにはどのような情報が必要か，さらに，そのような情報はどのようにして得られるか等，我々の研究室で行ってきた事例を紹介します．また，将来的にこのような処理を全自動化する展望について議論します．

ポイント（新しい映像利用を目指して）

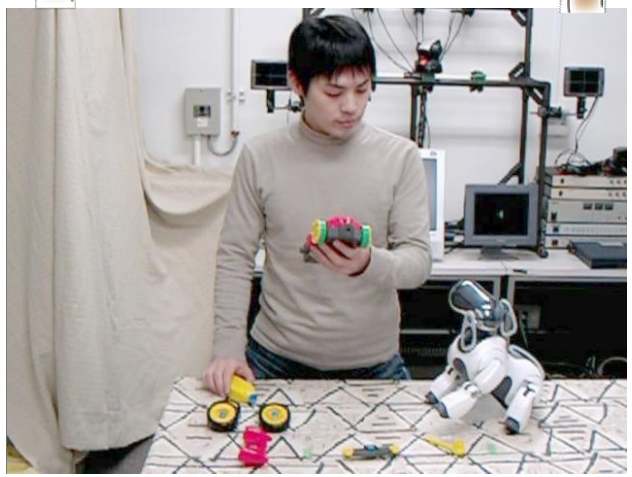
- アクティブな支援のためのデータとして
- 映像が扱う知識を個人レベルまで広げる



Before – After

Before
•常に適切な質問をするのは難しい
•作業の流れや思考が妨げられる

After
言語化されにくい知識，言語で検索しにくい知識をやさしく提供



質問に答えてくれる映像マニュアル
[伊津野・中村02, 03]

あらすじ

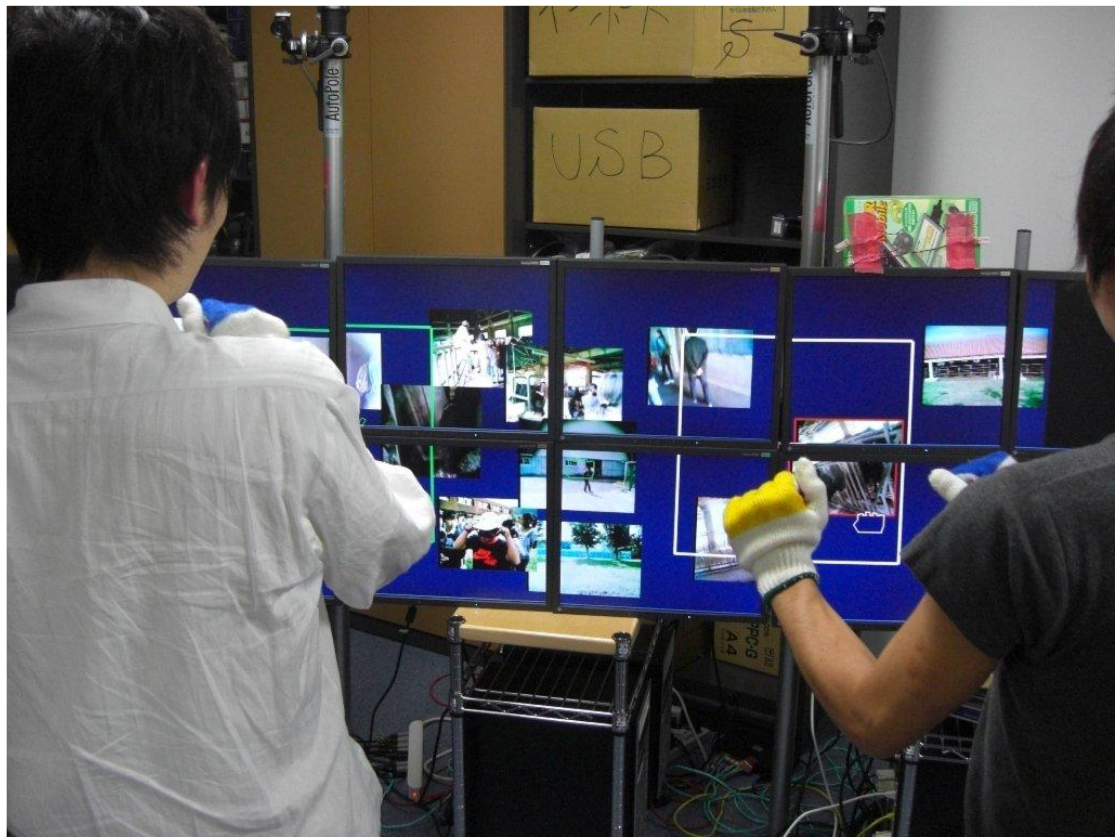
- 映像は使いにくい？(情報爆発，個別性)
- どんな使い方を目指すか
- 何が必要か(構造的，意味的な解析)
- これまでの研究(メタデータ付与，自動解析)と難しさ
- 映像取得時のメタデータの自動取得
- 作業映像取得の例
- 仮想アシスタントの例
- 会議支援システムの例
- ライフログの例
- 今後の展望(自動解析もしたい！)

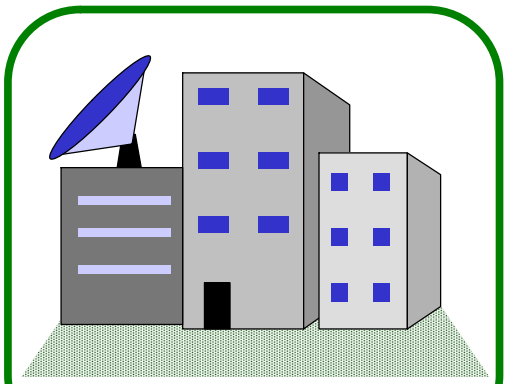
映像をためる
～量はたまってきたが，使い道は？～

- 国・公共機関
  - Digital Library (e.g., Informedia)
  - アニメの殿堂
- コンテンツホルダー
  - 放送局・映画会社
  - 商用DVDライブラリ
- 研究機関
  - 大学などの学術アーカイブ
  - NIIのアーカイブ
- コミュニティ・個人
  - YouTube, ニコニコ動画
  - ライフログ

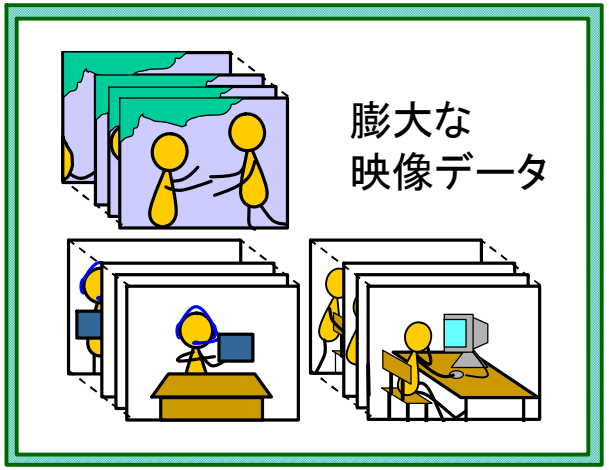
大量映像を使う
～どんな使い道があるか？～

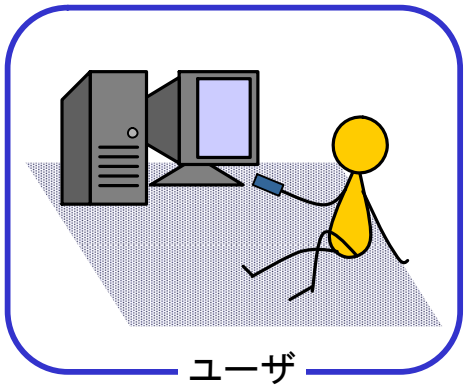
- 大量のデータはそれだけで存在価値がある
- 目的を持った検索
  - 検索・俯瞰を可能にする
  - 質問に答える
- 知識の発見
  - データマイニング
  - 機械のための知識の自動生成
- 人間の活動の支援（アクティブな利用）
  - 状況に応じた支援情報の提示
  - 記憶の代替


USB



放送局・映像制作会社



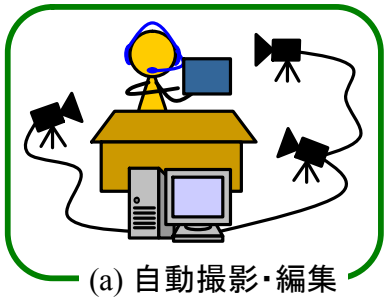


ユーザ

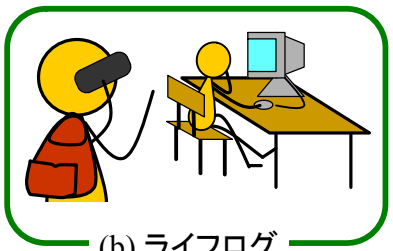
解析

提示

要求



(a) 自動撮影・編集

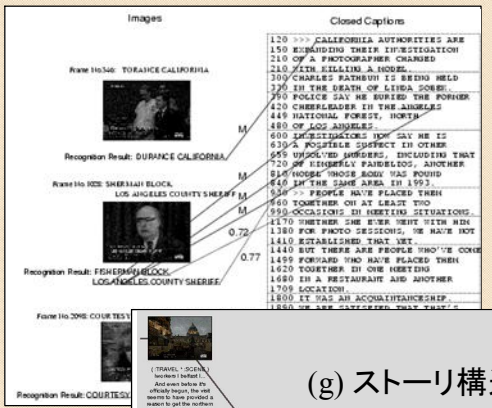


(b) ライフログ

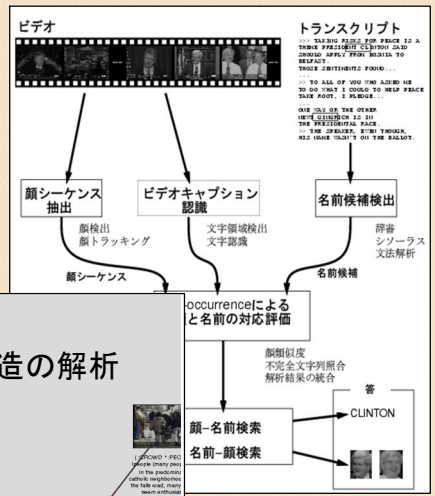
新しい映像撮影・製作手法

(c) 一覧表示

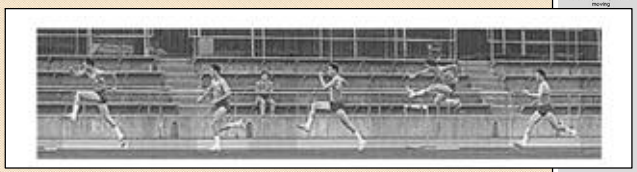
(e) 文字情報によるインデクシング



(f) クロスモーダルな
インデクシングと検索



(g) ストーリ構造の解析



(d) パノラマ表示

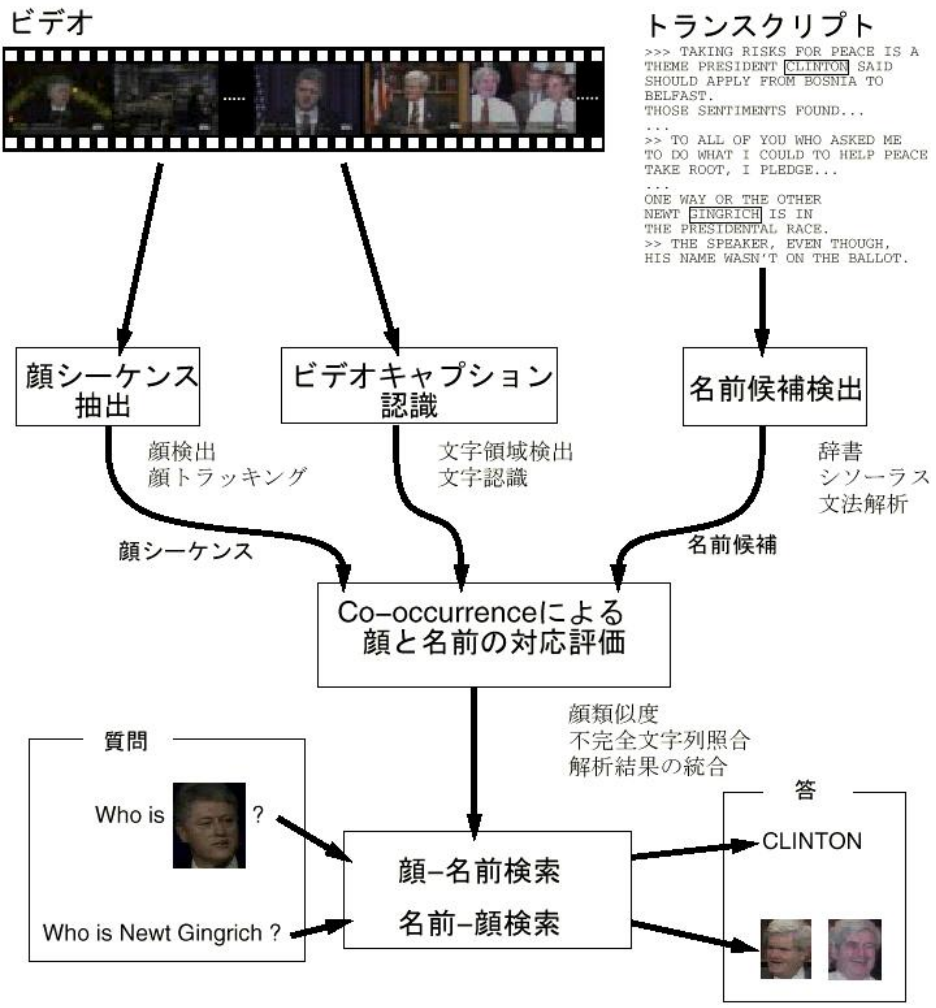
映像の解析技術
～構造の解析～

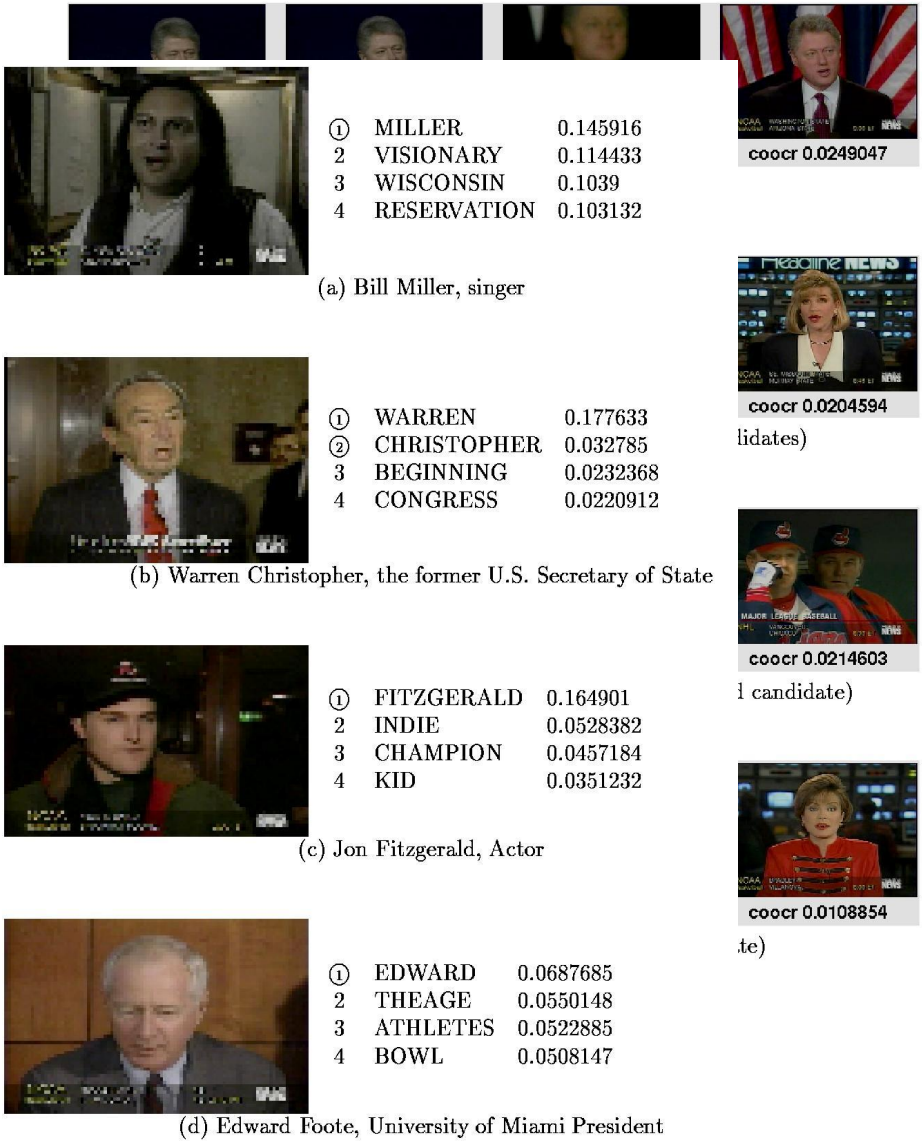
- 意味のある単位（セグメント）への分割
  - シーンチェンジの検出
  - カメラワークの検出
- セグメントの分類
  - 色，動き，人物等の特徴による分類
  - 発話や言語情報（トランスクリプト等）による分類
- セグメントの構造的役割
  - 映像構成の文法

情報の整理・新しい知識の発見

- 精度の高い内容認識
- 統計情報，データマイニング
- クロスモーダルな関係付け

あの人の名前/顔が知りたい: Name-Itシステム





映像の解析技術
～意味解析～

- セグメント内の特徴抽出
  - 画像からの情報抽出
    - 人物，物体の検出，特徴抽出，分類
    - シーン・状況の分類
  - 発話や言語情報（トランスクリプト等）からの情報抽出
    - 自然言語処理，IR技術
  - クロスモーダルな解析
- ストーリー構成の解析

QUEVICO

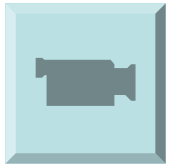
- マルチモーダルデータによる
  対話型メディア実現のための枠組み

想定する状況

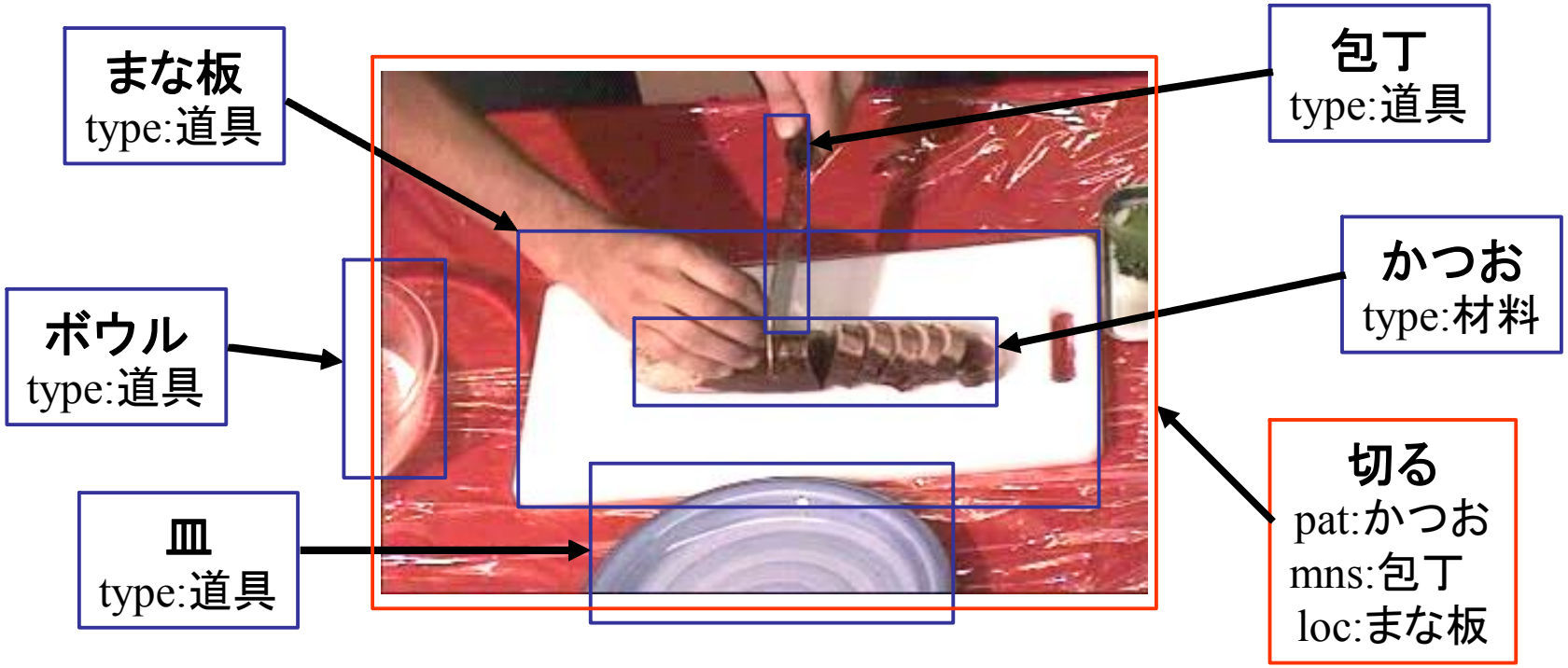
- 作業の実行中
- 自然言語による質問
- モニタでの答の提示
- 素材は、主に作業を
  撮影した多視点映像

実験例

- かつおのたたきの作り方に関する映像 (多視点，タグ付きデータ)
- 典型的な質問に対する応答
- いくつかの質問の複合 (タスクツリーによる状況推定の効果)



タグ付けによるデータの構造化



かつおを切る　盛りつける　ネギを振りかける

かつお

ネギ

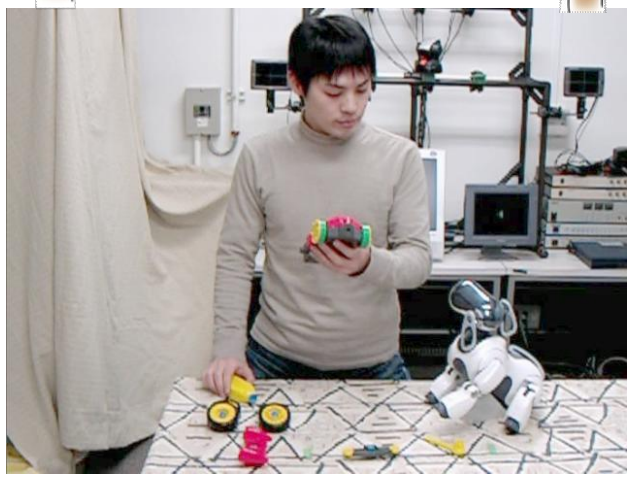
映像によるアクティブな支援

- 対話・質問応答
- 先生のように教えてくれる作業環境
  - ○○は何？ ○○はどうやって使うの？
  - 次はどうすればよいの？ これで正しいですか？
  - これについてどう思いますか？
- 個人やグループの体験を記録・再利用
  - ○○はどうしたっけ？
  - ○○君はどうしてたの？
  - 何か面白いこと見つけた？

Before – After

Before
•常に適切な質問をするのは難しい
•作業の流れや思考が妨げられる

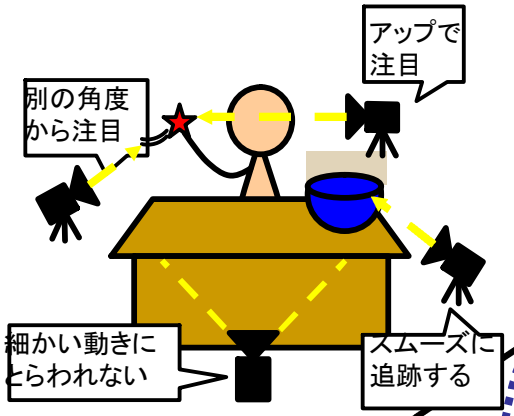
After
言語化されにくい知識，言語で検索しにくい知識をやさしく提供



質問に答えてくれる映像マニュアル
[伊津野・中村02, 03]

映像データの取得

- 自動撮影，必要となるデータを取得
- 認識技術によるインデックスの自動付与
- 人間を含む系としての再構築
- 過不足なく情報を引き出すしくみ



編集可能な
映像部品

目的別の映像

目的別の映像

目的

的に応じた
像提示

知的映像撮影・編集による
コンテンツ取得の枠組み

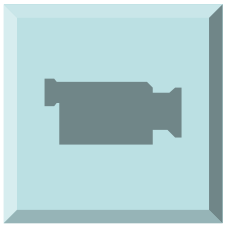
質問 ⟷ 説明

やり方: どういう風にやるの?
場所: どこに入れるの?
構成: どうなっているの?
形、色: どれ、どんな?
……..

こんな風に
細かく刻みます

煮る前に
レタスの中に入れます

自動撮影によるコンテンツ取得



ビデオ再生

左上: 提示される映像
左下: 映像切り替え
(スイッチャの状態)

右上: 動作認識の様子
右下: 手と全身を各々追跡する
可動カメラ

物体追跡

– テーブル上に複数の物体
– 背景に移動する人物

検出・追跡されている
把持物体を赤枠で表示

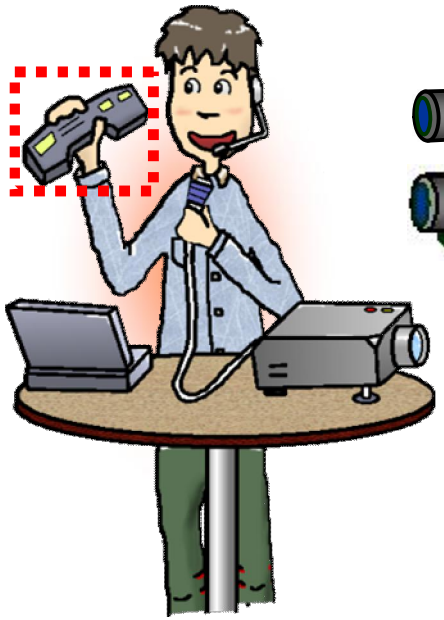
ムービー ファイル
(MPEG)

より高度な作業認識システム

天井設置

各カメラの特徴を利用、
物体や人体領域を検出

視体積交差法



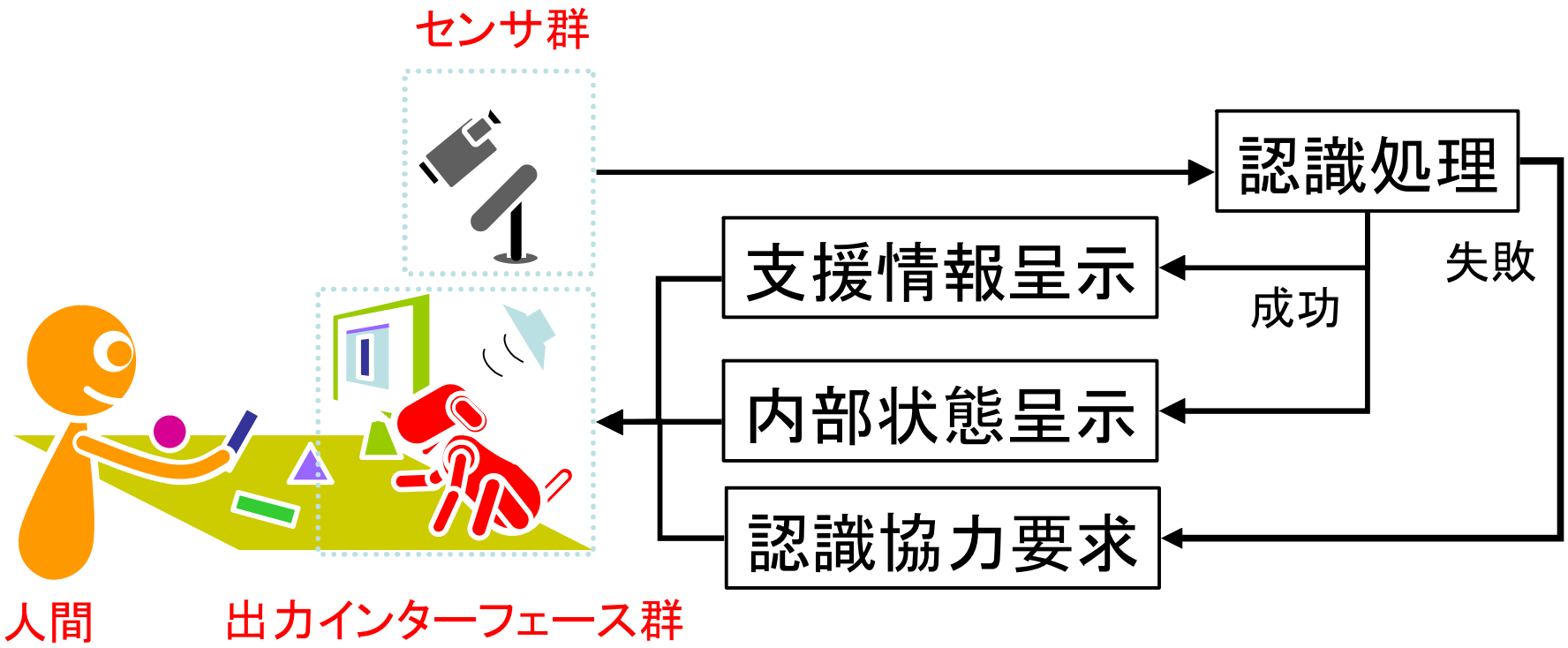
横設置

認識・追跡

人間を含んだ系としての再設計

- データの取得には人間が介在する
- 人間は優秀な認識器＋知能である
- 協力を仰ぐことにより，良いデータの取得が可能になる

認識協力要求による解決

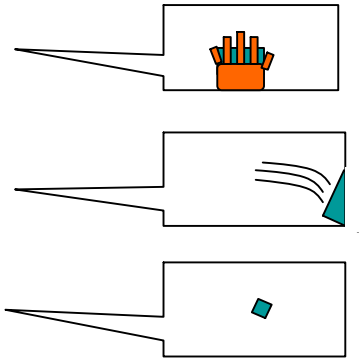


- どうすれば今の状況を解決できるかわからない

解決

具体例：認識が難しい状況

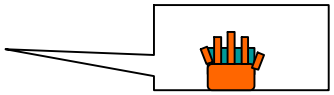
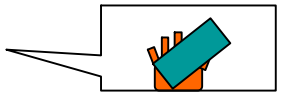
- 従来方法では未解決の問題
  - 物体が手に隠れている
  - 物体の移動が速い
  - 物体の見かけが小さい

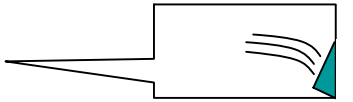
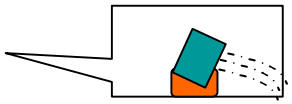


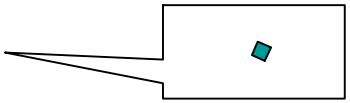
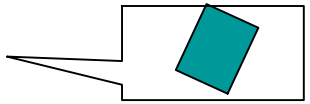
- 人間に位置センサ・物体にIDタグを付与
  - 人間への身体的制約，コスト面で利用できない場面が多々ある

認識が難しい状況を解決

- 人間に協力をしてもらう

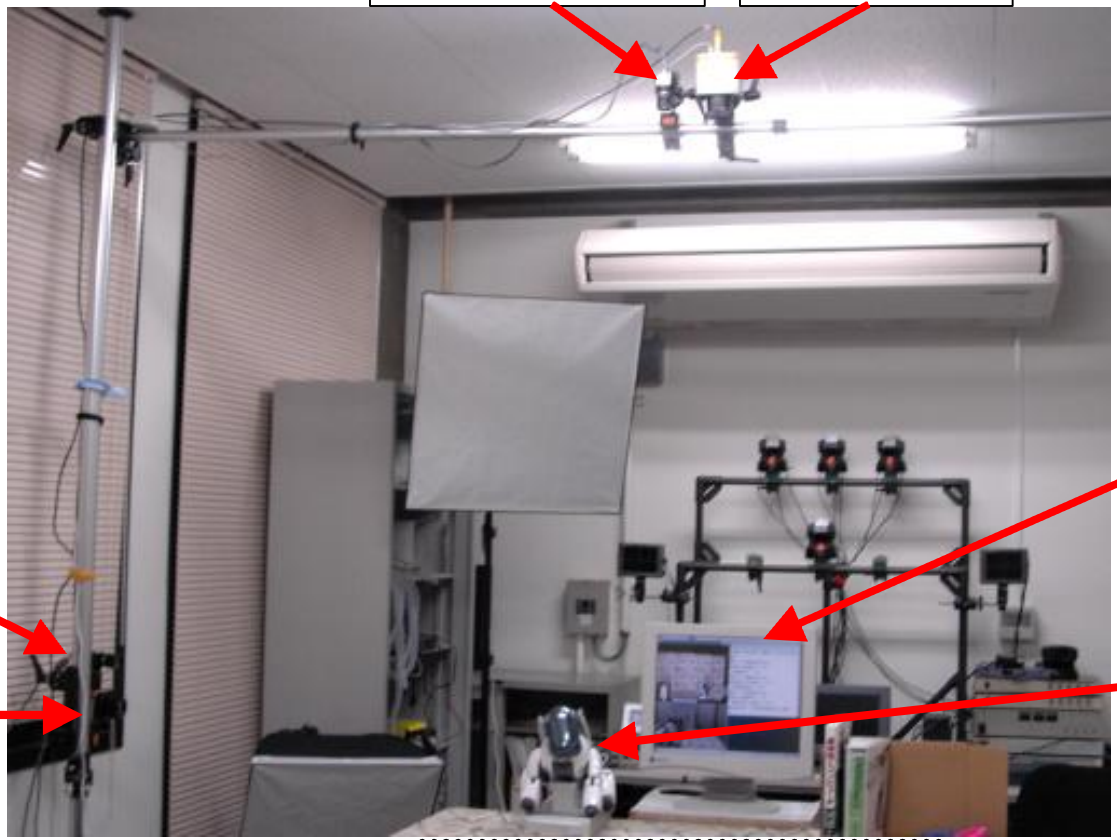

物体を手に乗せ見せる



物体を認識領域に戻す



カメラに近づける


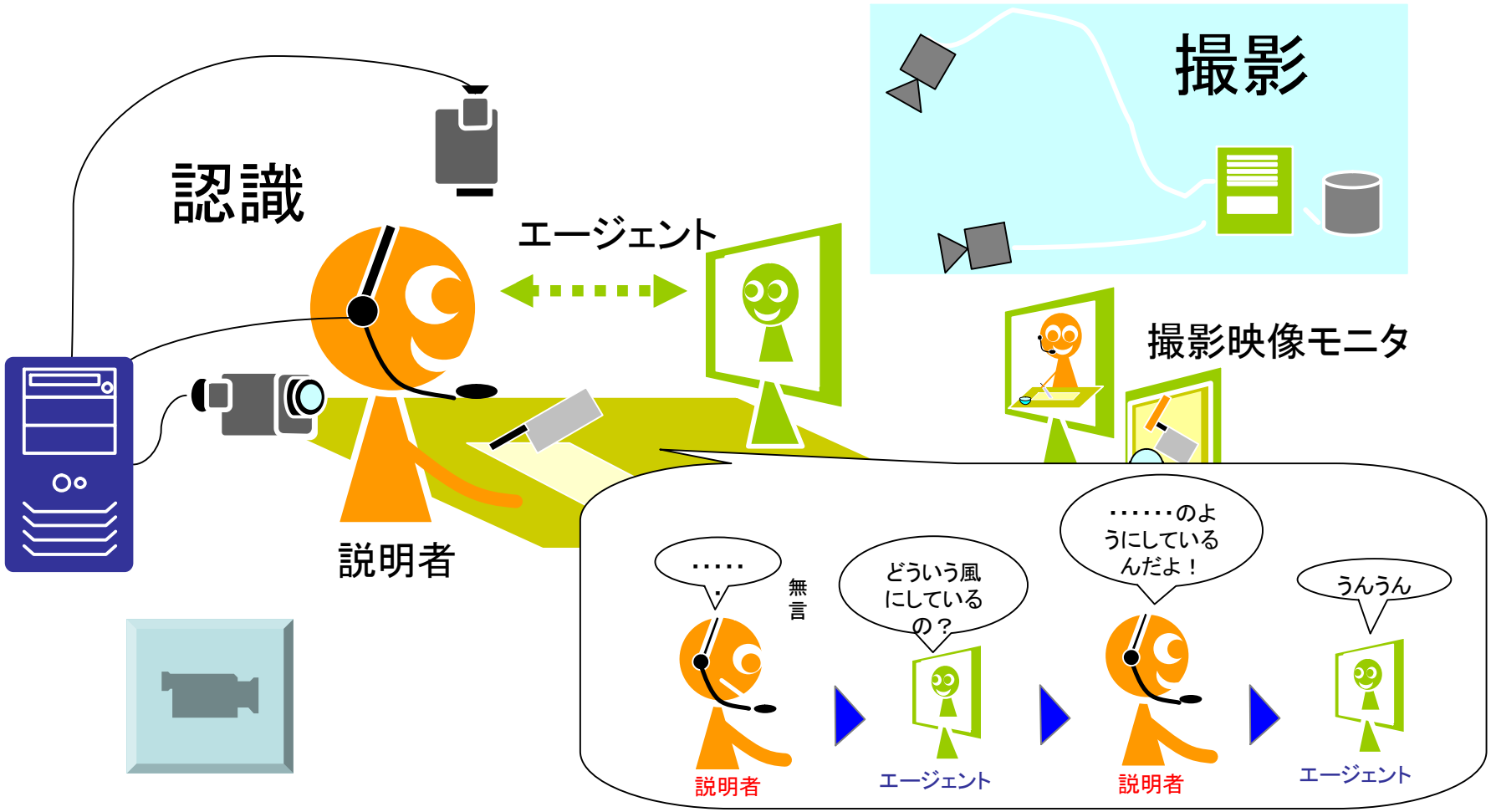
→　少しの協力で認識精度向上が見込まれる

作業支援システム


上RGBカメラ
上IRカメラ
情報呈示用
ディスプレイ
スピーカ
AIBO
横RGBカメラ
横IRカメラ
机上作業空間

エージェントによる情報記録支援 (仮想アシスタント)

映像コンテンツの取得，知識の伝承，体験の記録



説明者から引き出す情報

最低限必要な情報

器具名・食材名
手順・方法
説明者の状況
次の手順
作業の終了
分量
時間
加減

付加的な情報

食材の状態
代替品
代替方法
コツ・ノウハウ
理由

□仮想アシスタントが説明者の振る舞いや調理状況に応じてタイミング良く適切な反応をすることが期待される

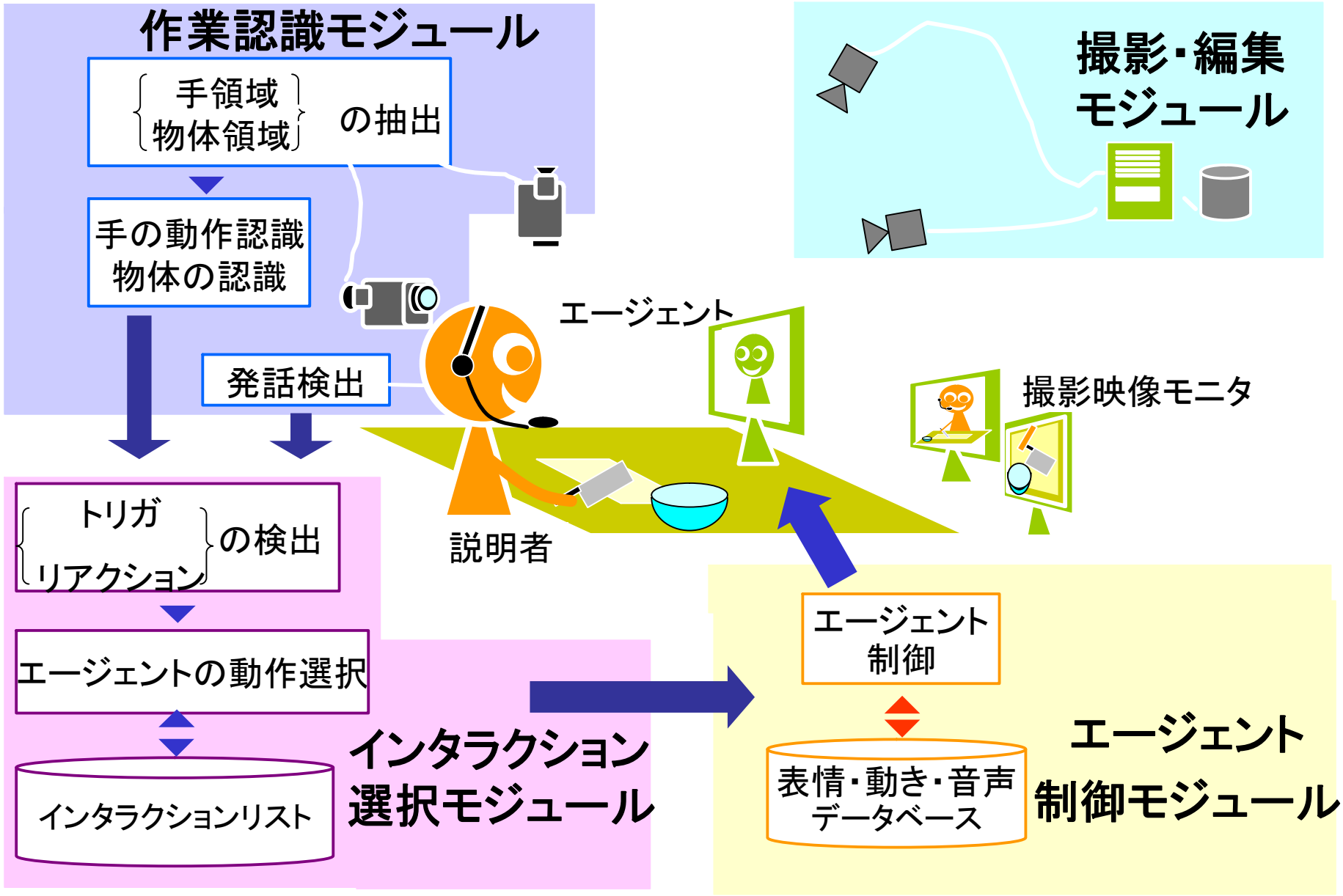
インタラクション

トリガ動作　　**アクション動作**　　**リアクション動作**　　**フォロー動作**

トリガ動作　：技術的な要素が含まれていそうな場面で発話をしていないなど

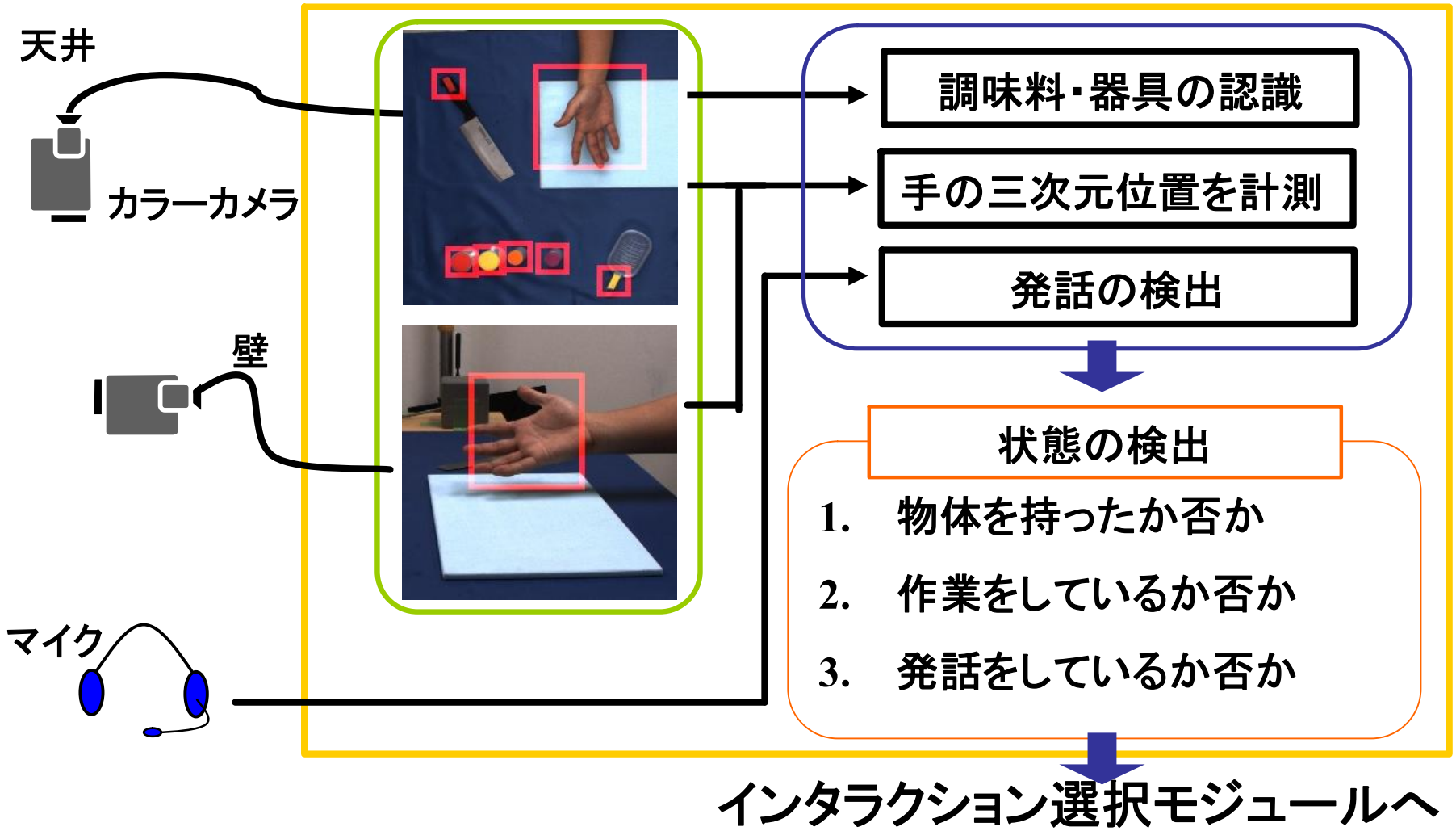
フォロー動作：説明者を不安にさせない、エージェントに協力を使用とする意思をなくさないという点で重要

システムの概要


作業認識モジュール
手領域
物体領域
の抽出
手の動作認識
物体の認識
発話検出
撮影・編集
モジュール
エージェント
撮影映像モニタ
説明者
トリガ
リアクション
の検出
エージェントの動作選択
インタラクションリスト
インタラクション
選択モジュール
エージェント
制御
表情・動き・音声
データベース
エージェント
制御モジュール

作業認識モジュール

□説明者の振る舞い・環境の状況を認識



インタラクション選択モジュール

- 作業認識モジュールから送られてくる情報に基づいてエージェントの動作を選択
  - トリガ⇒アクション
  - リアクション⇒フォロー
- トリガ動作は料理番組を参考にした

アクションの種類：9種類

フォローの種類：5種類

エージェントのアクション	説明者のトリガ動作
手順・方法を尋ねる	5秒以上物体を持ちながら作業をし、 且つ7秒以上発話をしていない
コツ・ノウハウを尋ねる	5秒以上作業をし、且つ7秒以上発話をしていない
:	:

インタラクションリスト例

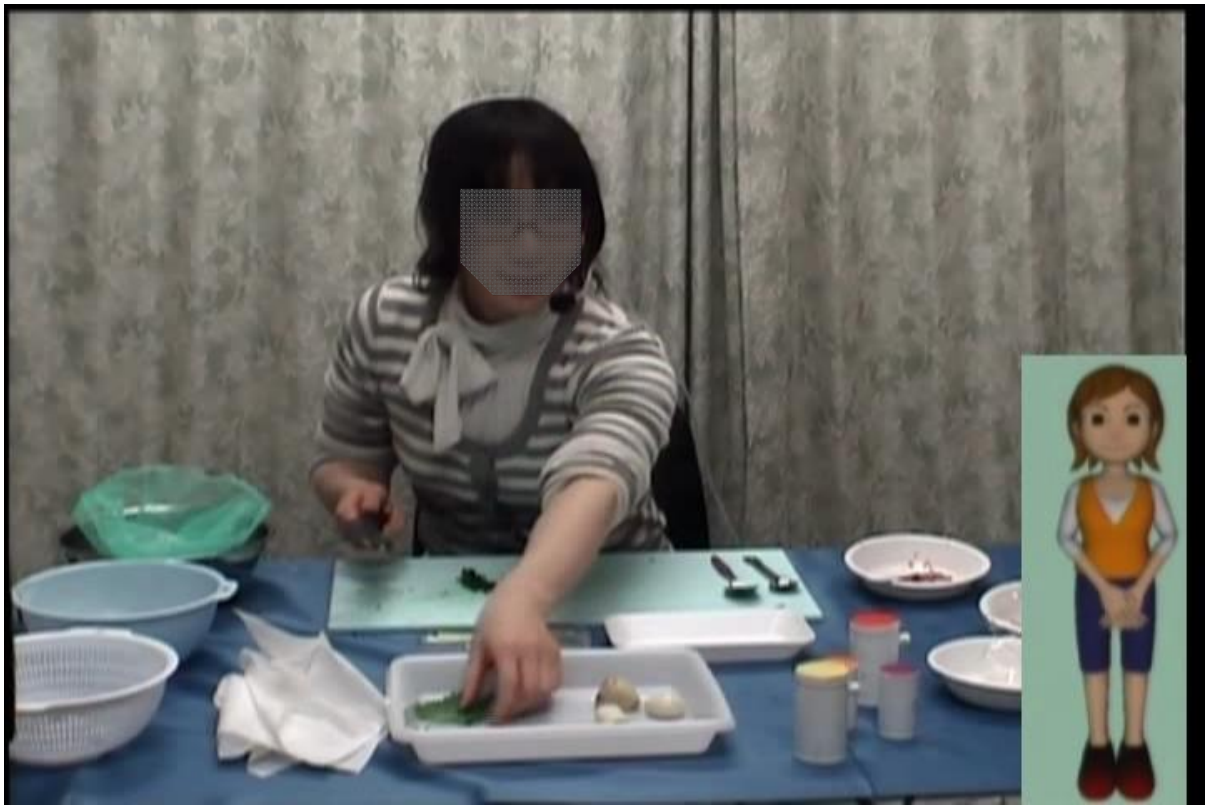
エージェント制御モジュール

- インタラクション選択モジュールからの命令に従ってエージェントを制御(表情＋動き＋発話)

アクション	表情	動き	発話
手順・方法を尋ねる	微笑み	首を傾けて手を差し出す	どういう風にやっているの？
コツ・ノウハウを尋ねる	眉を上げる	首を傾けて手を差し出す	コツはある？
:	:	:	:
フォロー	表情	動き	発話
相槌を打つ	微笑み	頷く	うんうん
:	:	:	:

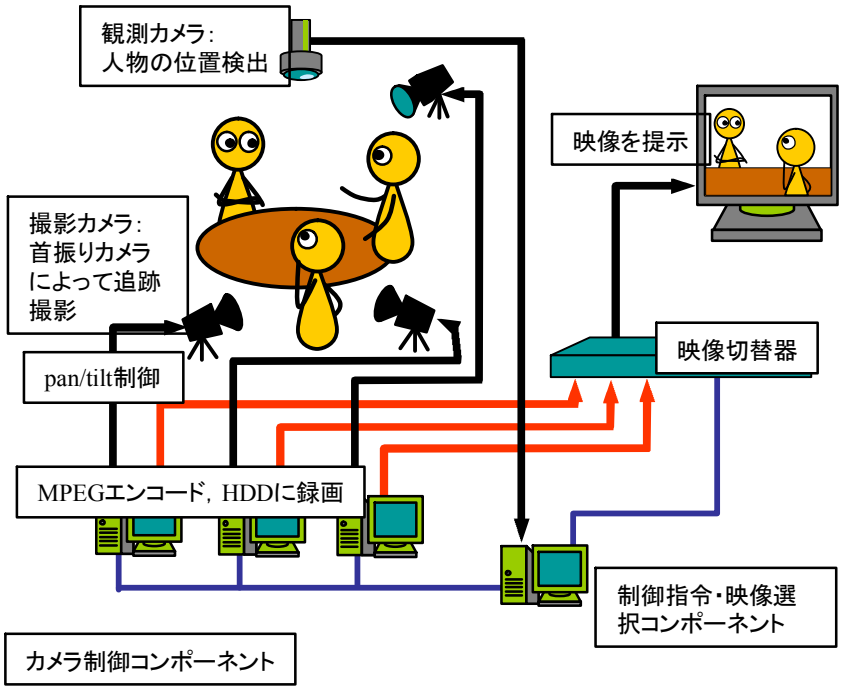
エージェントの動作リスト例

取得映像の例



会議・室内会話シーンの撮影

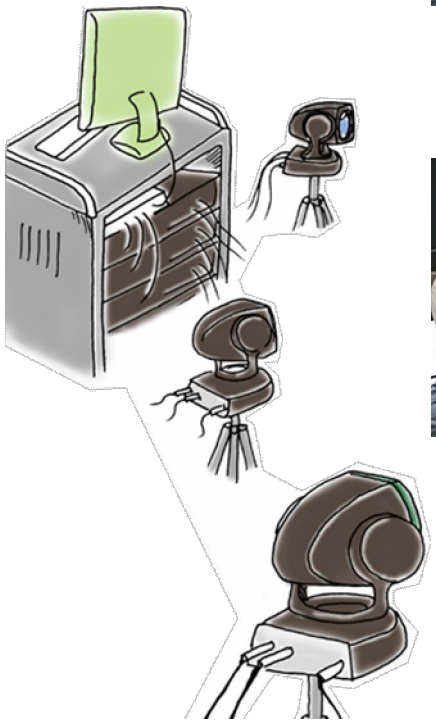
- 行動を制限したくない
  - 部屋に入ってから座るまで，その他の移動を追跡
- 座ってからも体(頭)の動きを補償したい
  - 構図を保った追跡撮影．種々の構図をあらかじめ設定



- 2種類のカメラ群
  - 環境カメラ
  - コンテンツ撮影カメラ
- 速い動きを追跡しながらの撮影は不可 (コンテンツとして見るに耐えない)

途中参加を支援するための会議ブラウジング

- **会議ブラウジング**
  - 会議の状況をわかりやすく提示
  - 準リアルタイムを想定
  - 議事録よりも生のデータに近い
  - 途中参加，発言を支援
- 研究概要
  - 自動撮影システム
  - 会議の特徴抽出
  - インタフェース

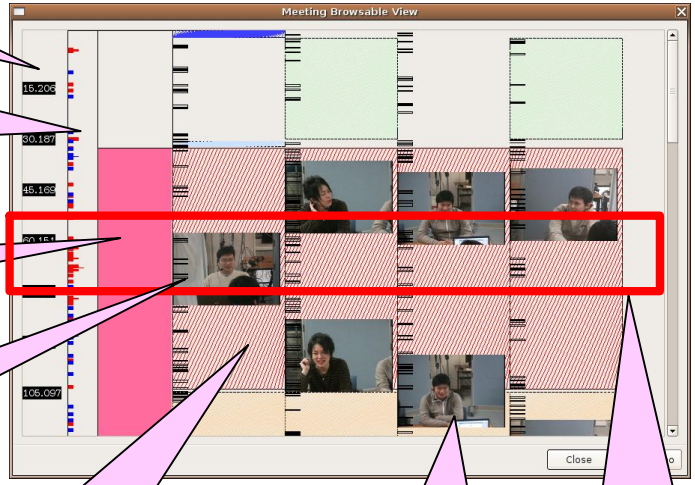


会議構成

盛り上がり点

合意点

次の話題へ



時刻

うなずき，相づち発生グラフ

話題バー

視線遷移区間，スナップショット

会話の盛り上がり（色）

話題区間

見たい区間をマウスで選択

実験

実際より短い時間で会議を把握する

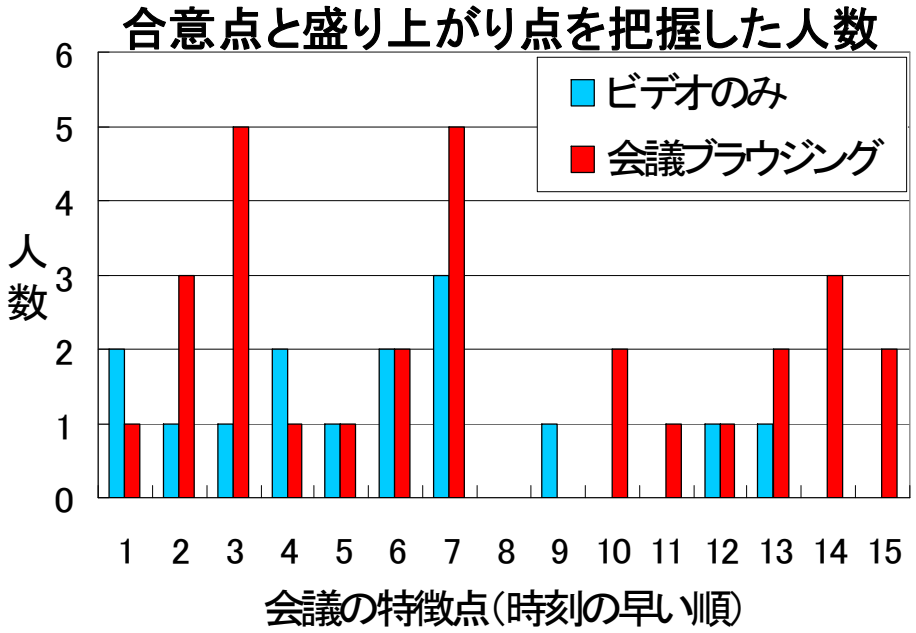
・ビデオのみ

・会議ブラウジング

合意点，盛り上がり点，話題の順番を把握するという課題に取り組む．

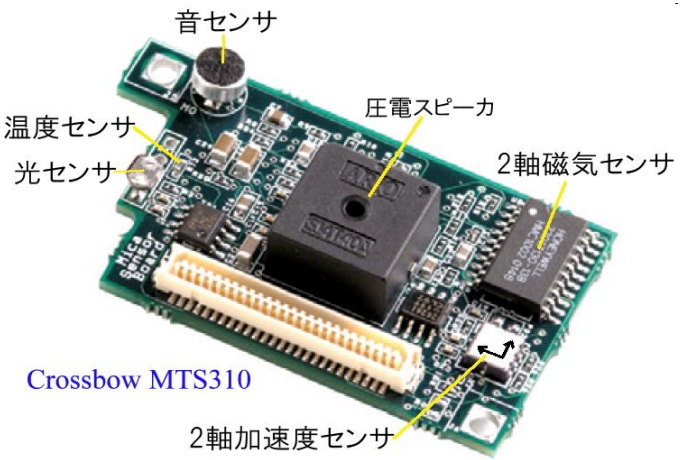
結果

会議ブラウジングのグループの方が全体を通して会議の特徴点を把握できることがわかった．



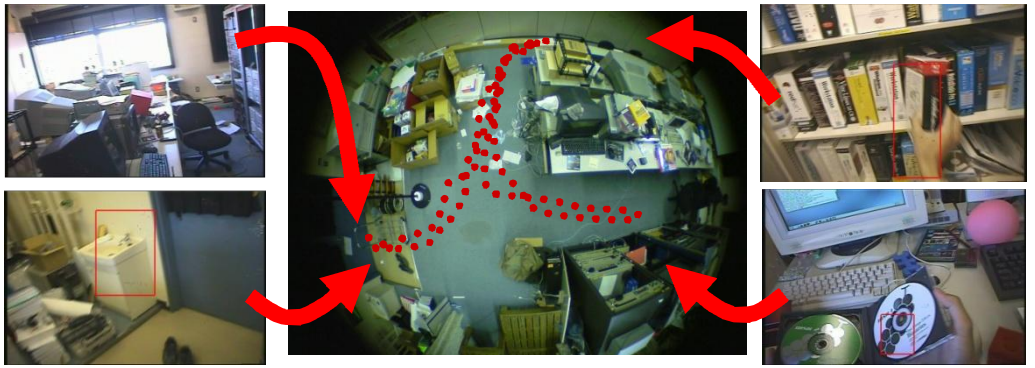
ライフログ（個人行動記録）

- 着るコンピュータ
  - いつでも，どこでも，支援してもらえる
  - いつでも，どこでも，情報を記録できる
- 個人の体験を記録する
  - 記憶の補助
  - 経験の共有



未来像？

部屋に入ってから

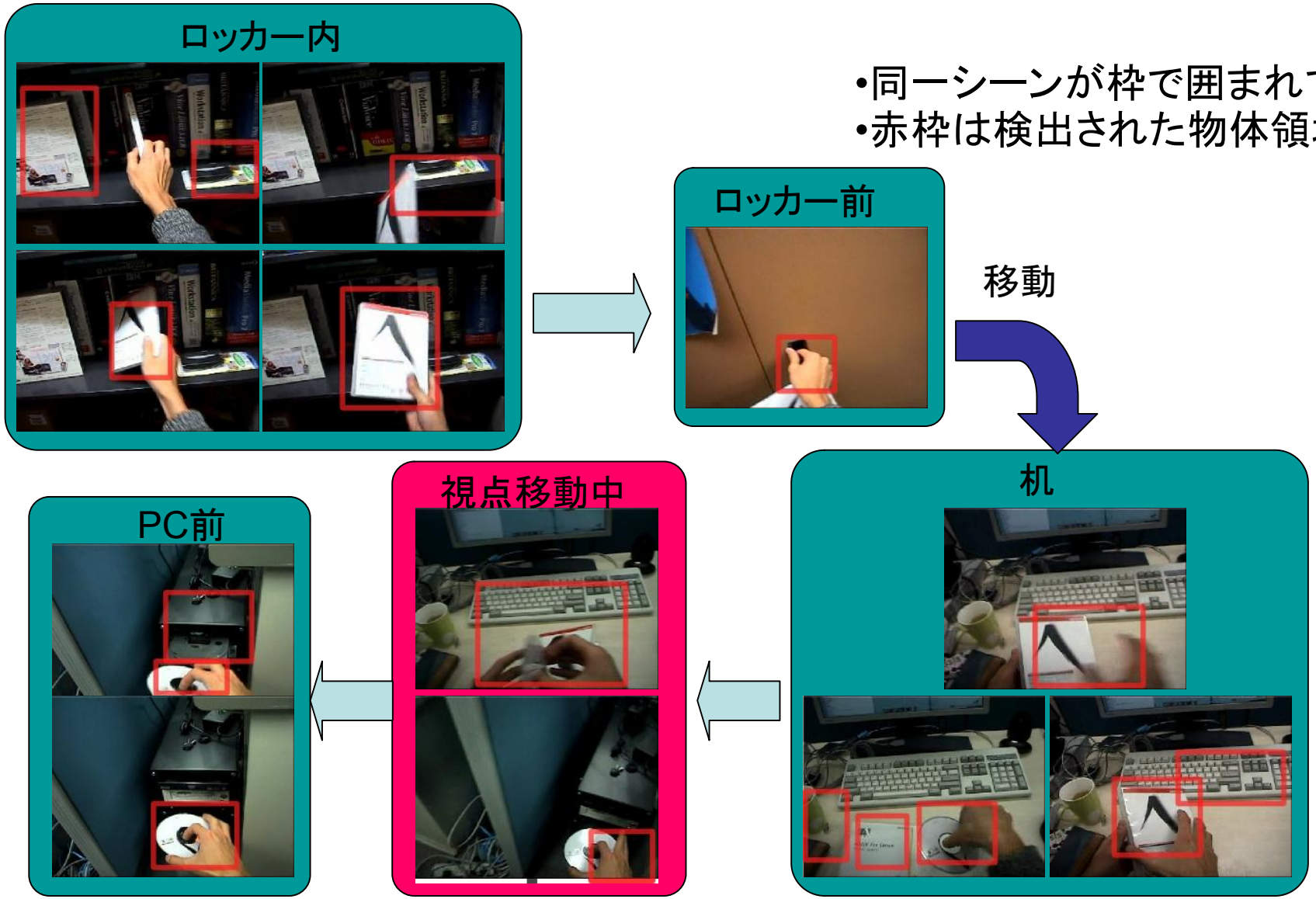


出て行くまで

ここから持っていく

ここで作業

関連付けの例



- 同一シーンが枠で囲まれている
- 赤枠は検出された物体領域

体験学習の記録

- 体験学習
  - 農場や動物園，博物館で様々な対象の観察や体験を行う
- 体験学習の記録
  - 参加者のメモ，写真など
    - 記録することに手間がかかる
    - 何かの体験中など記録が取れない場合がある

- ライフログ映像を用いた記録
  - ハンズフリーでの記録が可能
  - 記録のために観察を中断する必要がない
  - 観察対象に触れる，道具を使った体験の記録が可能

多人数の体験学習記録

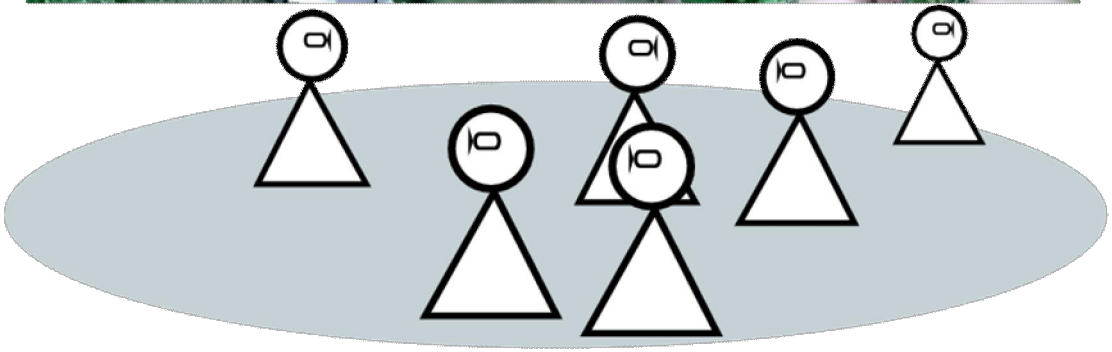
共通体験の発見
客観的な視点の追加
差異の発見
違う視点からの見え方
まったく異なる体験の発見

場の網羅的な記録

体験記録の比較

体験記録を通じた
コミュニケーションの発現

場の記録



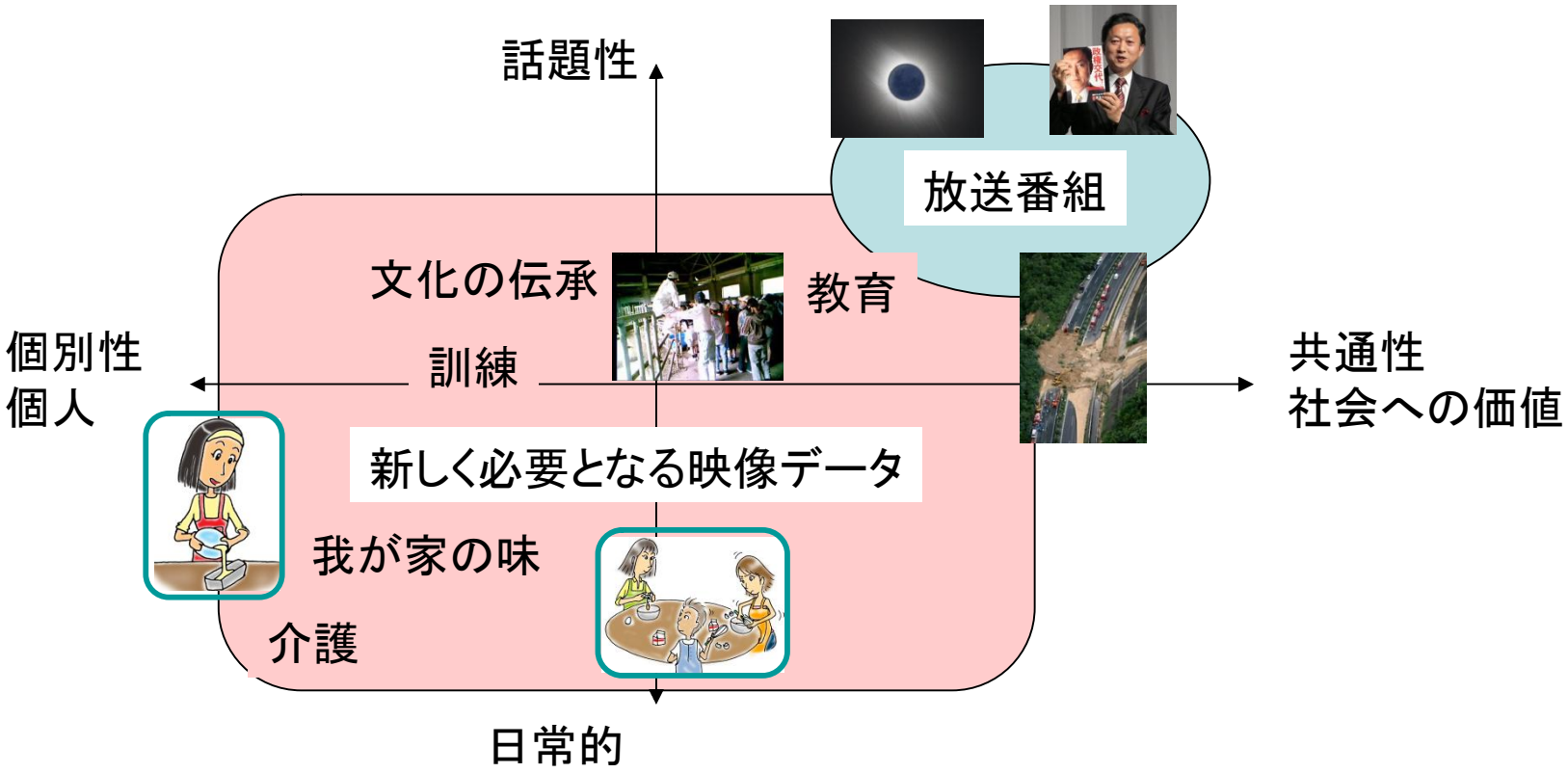
同一の場で別々のものを見聞きする様子

閲覧インターフェース


全員の体験学習記録映像
一人の体験学習記録映像のインデックス画像一覧
移動軌跡
北牛舎
広場
南牛舎
現在位置を表すアイコン
時刻を変化させるための
タイムスライダー

将来の展望（新しい映像解析を目指して）

- 放送映像（質が高く公共性の高い映像）と個人適応性の高い映像との対応
- 裾野の広い知識（一般性＋個別性）



将来の展望

- 自動解析と自動インデキシング
  - これまでのメディア処理技術の発展
  - 人間を含む系としての再構築
  - リアルタイム処理
- アプリケーションとしての映像利用
  - 個人適応と一般性の両立する支援
  - より広い状況での利用
  - リアルタイムインタラクション

- Y.Nakamura, T. Kanade: Semantic Analysis for Video Contents Extraction - Spotting by Association in News Video, Proc. ACM Multimedia, pp.393-401, 1997
- S.Sato, Y.Nakamura, T. Kanade: Name-It: Naming and Detecting Faces in News Videos, IEEE Multimedia, Vol.6, No.1, pp.22-35, 1999
- 佐藤真一, 中村裕一: 映像処理はもう古い？ そんなことはありません！？ ～映像処理の学術的意義と将来展望～,信学技報 PRMU-2003-56, pp.77-80, 2003
- H.Izuno, Y.Nakamura, Y.Ohta: QUEVICO QA Model for Video-based Interactive Media, Proc. Third International Workshop on Content-Based Multimedia Indexing, pp.413-420, 2003
- 尾関基行, 中村裕一, 大田友一: 机上作業シーンの自動撮影のためのカメラワーク, 信学論, Vol.DII-J86, No.11, pp.1606-1617, 2003
- 中村裕一: 会話ができる映像コンテンツを撮る・つくる ,電子情報通信学会ヒューマンコミュニケーショングループシンポジウム, 2004
- 尾形涼, 中村裕一, 大田友一: 制約充足と最適化による映像編集モデル, 電子情報通信学会論文誌, Vol.DII-J87, No.12, pp.2221-2230, 2004
- 尾関基行, 中村裕一, 大田友一: 注目喚起行動に基づいた机上作業映像の編集, 電子情報通信学会論文誌, Vol.DII-J88, No.5, pp.844--853, 2005
- 小泉敬寛, 亀田能成, 中村裕一: 個人行動記録の意味的構造を用いた効率的検索システム, 画像の認識理解シンポジウム(MIRU2005), pp.259--264, 2005
- M.Ozeki, Y.Nakamura, Y.Ohta: Automated Camerawork for Capturing Desktop Presentations, IEE Vision, Image & Signal Processing, Vol.152, No.4, pp.437--447, 2005
- M.Ozeki, Y.Nakamura: Evaluation of Self-Editing Based on Behaviors-for-Attention for Desktop Manipulation Videos,Proc. IEEE Int'l Conference on Multimedia and Expo, Vol.CD-ROM(MA2-L5.2), 2006
- 尾関基行, 宮田康志, 青山秀紀, 中村裕一:作業支援システムのための人工エージェントとのインタラクションを援用した物体認識, 画像の認識理解シンポジウム(MIRU), pp.81-86, 2007
- M.Ozeki, Y.Miyata, H.Aoyama, Y.Nakamura: Collaborative Object Recognition through Interactions with an Artificial Agent, Proc. International Workshop on Human-Centered Multimedia, pp.95-101, 2007
- Y.Nakamura:Video Content Acquisition and Editing for Conversation Scenes,"Conversational Informatics: An Engineering Approach", Chapter 12, Wiley, pp.363-393, 2008
- M.Ozeki, S.Maeda, K.Obata, Y.Nakamura: Virtural Assistant: Artificial Agent for Enhancing Contents Acquisition, Workshop on Semantic Ambient Media Experience (in conjunction with ACM Multimedia), pp.75-82, 2008
- 青山秀紀, 尾関基行, 中村裕一:インタラクション再生モデルによるさりげない学習支援,信学技報 MVE2008-60, pp.1-8, 2008