画像に基づく部屋内シーン変遷の自動検知と対話的イベント検索システム

片山憲昭† 牧和宏†† 島田伸敬†† 白井良明††

† Yahoo Japan 〒 106–6182 東京都港区六本木 6–10–1
†† 立命館大学大学院 理工学研究科 〒 525–8577 滋賀県草津市野路東 1–1–1
E-mail: †{katayama,maki}@i.ci.ritsumei.ac.jp, ††{shimada,shirai}@ci.ritsumei.ac.jp

あらまし 本論文では，音声対話とジェスチャ認識を用いて屋内シーンで起こったイベント情報に対話的に問い合わせが可能なシステムの提案を行う．監視映像の検索性を向上させるには、自動画像処理によって重要な場面を検知して索引付けする技術が不可欠であるが，シーンの変化や物体を全自動で完全に識別することは困難な問題である．一方，人間の識別能力は正確ではあるが，監視カメラに保存された長時間の記録映像から目的のシーンを人が探し出す作業は多大な労力を要し，見逃しの問題も抱えている．このような問題から，監視映像下に起こったイベントシーンをある程度自動的に検知・整理しておき，ユーザがシステムと対話的に検索操作を進めていくことによって，正確で素早いシーンイベント検索システムの実装を試みた．

キーワード 物体検知，イベント検索システム，音声対話，ジェスチャ認識，ヒューマンインタフェース

Image-Based Automatic Detection of Indoor Scene Events and Interactive Inquiry

Noriaki KATAYAMA†, Kazuhiro MAKI††, Nobutaka SHIMADA††, and Yoshiaki SHIRAI††

† Yahoo Japan Roppongi 6–10–1, Minato-ku, Tokyo, 106–6182 Japan
†† College of Information Science and Engineering,Ritsumeikan Univ. Nojihigashi 1–1–1, kusatsu-shi, Shiga, 525–8577 Japan
E-mail: †{katayama,maki}@i.ci.ritsumei.ac.jp, ††{shimada,shirai}@ci.ritsumei.ac.jp

Abstract This paper proposes a system for automatic detection of indoor scene events with interactive inquiry based on speech dialog and gesture recognition. The system detects the events that various objects are brought in or taken out by image understanding. The user of the system inquires the stored events in the past by directly pointing the objects or spaces in the real environment. Since automatic event detection may fail in complicated indoor scenes, the system can use interactive inquiry to correct such failures.

Key words object detection, event inquiry，speech dialog，gesture recognition，human interface

1. はじめに

近年，犯罪数の増加と共に監視カメラや映像レコーダといった，セキュリティ用映像装置の需要が高まり急速に普及してきている．米国では DARPA(Defence Ad-vanced Research Projects Agency) の基，複数のカメラが協調して動作するビデオ監視システムの研究プロジェクト VSAM(Video Surveil-lance and Monitoring) が行われている [1], [2]．これらの屋外の監視映像を対象とするシステムでは，天候等による環境変動の影響を受けずに侵入物体を検出することが重要となる．

一方，室内環境を監視するシステムは，多数の人物の出入りなどで観測する画像が複雑になり移動物体の抽出は困難になってくる．加えて画像から人間の行動を理解する研究も盛んに行われ，室内の物体を操作したり椅子に座ったりしたことを検知するインテリジェントルームの研究や [3]〜[5]，どこに何があるのか，誰が物を「持ち込んだ/持ち去った」のかを認識・管理するシステムの研究開発が行われている [6]．身に着けた画像センサの情報から，ユーザ自身の作業内容や物体の置き忘れなどを記憶・報告してくれるシステムの研究もある [7]．また部屋内のユーザの動作を認識することによって家電機器を操作するジェスチャインターフェースが研究されている [8]．その他にも WWW を用いて，ユーザがブラウザから見たいシーンの表

示・検索ができるシステム [9] もある．シーン変化や室内の物体の種類を識別する技術は，画像認識の分野で研究されているものの [10], [11]，多様なシーンに対して全自動で誤りなく認識することは現状でまだ困難である．そういった避けがたい自動認識の誤りを，ユーザがシステムとインタラクティブにやり取りを行うことで誤りを訂正し，物を取ってきたり情報収集を行うサービスロボットが研究されている [12]- [15]．これらの研究では，多量のデータ収集や情報処理を機械にまかせ，暫定的に出力された認識結果（冷蔵庫の中のどこにどんなものがあるか）をユーザが確認して，誤りがあったり何かの影に隠れて見つからない場合には,「○○の後ろ」などのアドバイスを行うことで認識誤りを回復し，作業を完遂する．このように，機械が不得意な場面での人間の識別能力は確かに正確ではあるが，監視カメラに保存された長時間の記録映像から目的のシーンを人手で探し出す作業は多大な労力を要し，見逃しの問題も抱えているため，安易にユーザが介在すれば解決するわけではない．

そこで，このような大量の映像データをシステムが自動整理して効率よく保存しておき，ユーザは必要最低限の指示を行うだけで対話的に検索することができれば便利である．監視映像の検索性を向上させるには、自動画像処理によって重要な場面を検知し索引付けする技術が不可欠である．また自動認識の誤りがおこった場合に，ユーザが簡単に適切な指示を行って再検索できるインターフェースがあることが望ましい．

本研究では，監視映像下に起こる「物体の持ち込み/持ち去り」のイベントをある程度自動的に検知・整理しておき，あとから放置された物体や，かつて物体があった場所をユーザがGUI で指し示すことで，それを持ち込んだ・持ち去ったシーンを特定できるシステムの開発を試みた．また，GUI 越しにだけではなく，実シーン中で直接ジェスチャによって物体や空間を指し示してシーン検索を行い,「これを持ってきたのは誰？」といった直観的な問い合わせが可能な検索操作インターフェースの実装を試みた．

2. システムの概要

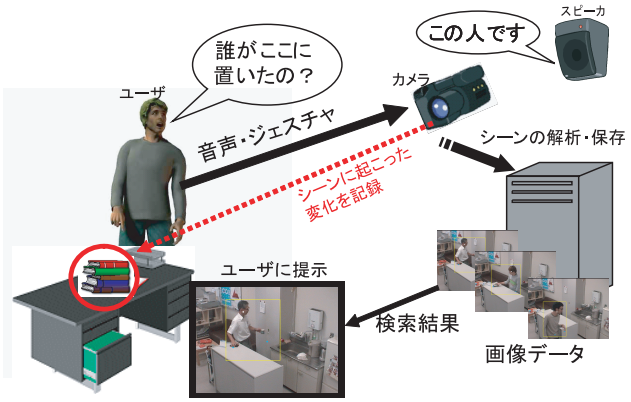


図 1　システムの概念図

図 1 にシステムの概念図を示す．システムの機能は，

- イベント検知部
- イベント解釈部
- ユーザインタフェース部

の 3 つからなる．

イベント検知部は，天井に設置されたカメラと，カメラからの入力画像を処理するための PC によって構成される．PC は常時，部屋の中の映像の処理を行っていて様々な情報を獲得している．人間の動作に注目していて，人が何か物を「持ち込んだ/持ち去った」シーンを自動的に検知し保存する．

イベント解釈部は，イベント検知部で保存されたシーンに「いつ？ どこで？ だれが？ 何をしたか? 」索引付けし，その情報をデータベースで管理する．イベント解釈部では，検知された画像の処理を行う PC が必要だが，イベント検知部と同様の PC 内のプログラム上で別のプロセスとして行う．

ユーザインタフェース部は，ユーザの入力となる指差しなどのジェスチャ認識と，音声発話の処理を行い，その問い合わせに応じた結果を提示しやり取りをする仕組みになっている．構成はカメラとスピーカとマイク，そしてそれら処理する PC となる．カメラはイベント検知部と同一のもので，ユーザが監視映像下の現場に行ってその物を指しながら発話できるよう，ジェスチャを検知する処理を行っている．システムはユーザの問い合わせに対し，保存されている画像データを検索し物を持ち込んだ・持ち去った人物の映っている画像をユーザに提示する．なお音声対話は，フリーの音声認識エンジン Julius/Julian [21] を用いたシステム [16] を拡張し実装した．

3. 部屋内イベントシーンの自動検知

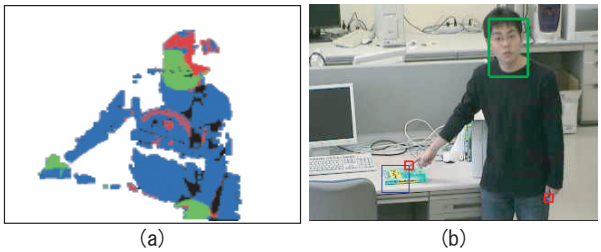
物体検知はこれまでに様々な手法 [18] が提案されている．本研究では背景差分法を用いるが，研究室のような室内環境では，

- 多数の人の往来
- 照明の ON/OFF などによる急激な照明変化
- 日光による穏やかな照明変化

といった現象が背景画像に影響を与え，移動物体を含まない背景画像を生成するのは困難になる．そこで，照明変動などの背景画像への急激な変化にロバストで，背景の時間的変化の追従性を高めた統計的手法 [17] で背景画像を推定し，それとの差分をとり移動物体を検出した．背景差分領域は,「人」と「物」の両方の領域を含んでいるので，２つを区別する．背景差分領域の連結領域を調べ，その連結領域が人の持つ特徴であれば人領域，それ以外は物体候補領域と定義した．

3.1 人間の動作検知

人領域は面積が大きく，顔は肌色領域の上に髪色領域があり，そこから離れた場所にも肌色領域 (手) を持っていると定義する．図 2-(a) はある大きな面積のラベルを肌色，髪色，それ以外の 3 値画像で示したものである．この様に，肌色領域の上に髪色領域を持つ大きなラベルは人領域の可能性があるのでこの領域の中から顔を検出する．顔検出には画像処理ライブラリの OpenCV [22] を用いた．もし顔が見つかればこの領域は人領域となる．図 2-(b) は，ユーザの顔と指差しを検知している様子



(a)　(b)

図 2　人検知の様子

を示している．顔を検知していれば，顔領域の重心座標から，その人領域の中で一番距離の離れた肌色領域を「指」とする．指の座標を獲得することができれば「これ，だれが置いたの？」という問い合わせの「これ」とユーザが指している領域を決定することができる．

3.2 物体の検知

物体検知には，背景差分領域の時系列データを用いる．物体検知アルゴリズムは以下のものを用いた．物体は自ら動かないという特性を持っているとし，ある一定区間の時系列データを観測する検知ウィンドを設け，それを画像シーケンス上をスライドさせながら，物体領域のみを抽出する．物体検知のアルゴリズムは以下のものを用いる．あるフレームでの人領域以外の領域を物体候補領域と定義する．

Step1 「物体候補領域」を発見したフレームから，検知ウィンドに入ってるフレーム全ての「物体候補領域」の論理積をとる
Step2 Step1 で過半数以上が真であった「物体候補領域」を「物体領域」とする
Step3 「物体領域」を検知した検知ウィンドの全ての画像列をまとめて保存する
Step4 「物体領域」となった領域を背景に更新する

物体を検知した瞬間「物体領域」を背景画像に急激に更新する事で次フレーム以降に何度もその物体が同じ場所で検知されないように工夫した．今回は 6fps のフレームレートで画像をキャプチャし，一定フレームを 10 フレームとした．10 フレーム中 8 フレーム以上の物体候補領域を観測することができればイベントが起きたシーンとして保存した．以下の図 3 に物体検知の様子を示す．図 3-(a) は人が物を置いた時の画像である．こ

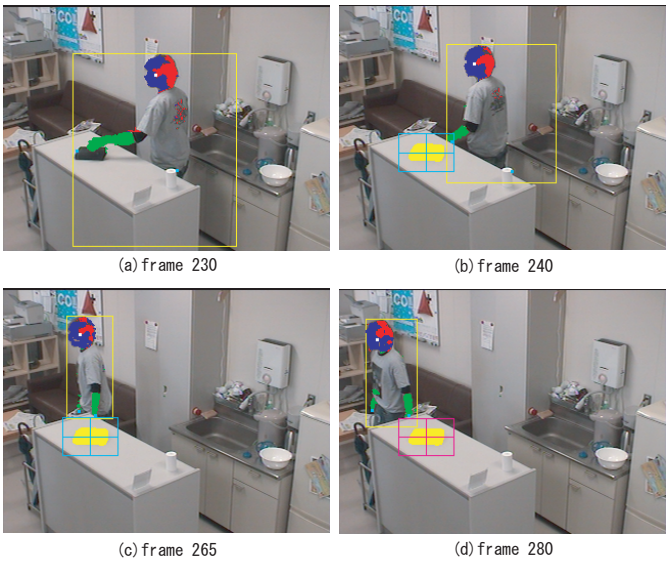


図 3 物体検知の様子

のフレームでは人と物が離れていないので物が置かれたか検知できない．図 3-(b) と (c) では物体候補領域が観測されている．この時点ではまだ連続観測されたフレーム数が一定値に達していないので，物体領域候補として観測を続ける．図 3-(d) のフレームにおいて一定フレーム，物体候補領域を観測したので，初めて物体領域を検知したことを示している．一定フレーム数連続して観測された物体領域候補を物体領域とすることで，一瞬物体の前を横切る人がいても安定して検知することができる．

しかし，この手法では図 4-(a) のように，人影を物体領域としてイベント検知される場合が多く存在した．また，図 4-(b) のように，光が局所的にあたりイベント検知される場合もあった．そこで，誤検知の多かった影を除去するために，影領域と背景画像の正規化相関係数の値を求め，相関が低いもののみを物体領域とみなす処理を加えた．

図 4 イベント検知の失敗例

イベント検知で抽出した画像列はデータベースに保存する．検知されたイベントがどういったシーンであったかは，イベント解釈部で判定され，情報を付与する．

3.3 イベント検知実験・考察

実験に使用した場所は以下の図 5 のような場所である．この

図 5 実験に使用した場所

場所はシンクや冷蔵庫やコーヒーメーカなどがあり人が頻繁に訪れる場である．実験には監視カメラ (SONY EVI-D100), PC(Pentium4 2.66GHz) を用いた．期間は約 1 週間で約 100 万枚 (220GB) の画像について物体検知を行った．

検知されたシーンを目視で分類し、代表的なシーンについて図 6 に示す．図 6-(a),(b) のように 1 つの物の場合，持ち込みでも持ち去りでも物体領域を得ることができたことが分かる.(c) は左から時系列になっていて物が置かれ，上にさらに物が重なり，そして下の物を持ち去った時の物体領域を示している．これについても下の物体の持ち去った時の領域のみを検知していることが分かる．(d) は人手で見て判断したが，物を移動させた時は移動元と移動先の物体領域が 2 つ同時に出ることが分かる．

物体検知されるべき物が検知された割合は，約 80%，物体検知されるべき物が検知されなかった割合は，約 20%ほどであった．一番多く検知したシーンは，コップに冷蔵庫からお茶を取り出して注ぎ持ち去る動作であった．このシーンはほぼ 100%に近い形で検知できた．その理由は，コップを置いて冷蔵庫を開ける時人と物が離れるからだと考えられる．物体検知されるべきではない物が検知された割合（誤検知数／総検知数）は約

図 6　イベント検知の成功例

50%ほどであった．この割合は影を考慮せずに出したもので，影除去の判定を追加した場合での割合は約 30%に改善された．

4.　レイヤ構造を用いたイベントの解釈

イベント検知部で保存された画像は「もち込み/持ち去り」の両方の画像を含んでいる．イベント解釈部ではイベント検知が起こった時，どの物体が「持ち込み/持ち去り」だったのか判定を行なう．図 7 にイベント解釈までの流れを示す．

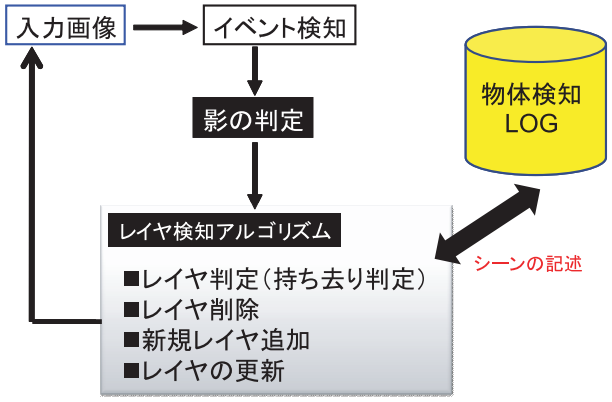


図 7　イベント解釈の流れ

前章で述べた手法で入力画像の中からイベントを検知する．検知した物体領域に関してレイヤ操作を行なう．レイヤ検知アルゴリズムの基で，物体が持ち込まれた/持ち去られたのかを判定する．イベント検知が発生するたびに，シーンの記述を逐次データベースに書き込んでいく．シーンの記述は，検知されたシーンに「いつ？どこで？だれが？何をしたか？」を索引付けするもので，イベント検知部で保存された画像列を代表する 1 画像をキーフレームとして，物体領域の面積，物体の重心座標，検知した時刻，シーン解釈結果を記述したものである．システムはシーン解釈部で誰がイベントを起こしたか分からないが，イベント検索時の対話の中で，結果画像に映っているが誰かをユーザに尋ねることにより，シーンの記述に「だれが？」のインデックスを付与することができる．

また，シーンの記述以外に検知された物体領域の画像パターンを覚えておき，これを用いて次に説明するレイヤ構造アルゴリズムを実現する．

4.1　レイヤ構造を理解するアルゴリズム

図 8　持ち込み・持ち去りの例

「もち込み/持ち去り」の判定は図 8 に示すように，ある物体 O_i が (a) で持ち込まれ，その後 (b) で持ち去られたシーンを見ると，検知される物体領域はほぼ同じと考えられる．そのため，イベント検知した時の物体領域と，それより過去に時系列に形状のマッチングを行う事で持ち込み/持ち去りを判定する．形状のマッチングは現在のマスクが過去と一致するか論理積 (AND) をとって調べる．論理積をとって 80% 以上の面積が真であれば一致とし「持ち去り」と記述する．図 8 の場合は (a) と (b) の物体領域の形状がほぼ一致したので，物体 O_i が持ち込まれその後，持ち去られたと記述できる．

日常生活のシーンでは机の上が散らかるなど，物体同士のオクルージョンは頻繁に発生し，物の形状が完璧に抽出できない場合がある．そこで物体同士のオクルージョン問題を解決するために，検知された物体領域を時系列のレイヤ構造として考える．移動物体の重なりを考慮した物体検出方法は，レイヤー型検知法 [19] があるが，この手法では物体のオクルージョン下で起こったことに対する処理，例えば，物体を下から抜き取とる，物体の隠れている奥に物が置かれるといったシーンは対応できない．レイヤ構造を理解するアルゴリズムは以下のように定義した．

Step1　ある物体領域と，全てのレイヤごとの予想した次状態の物体領域との前後関係の判定

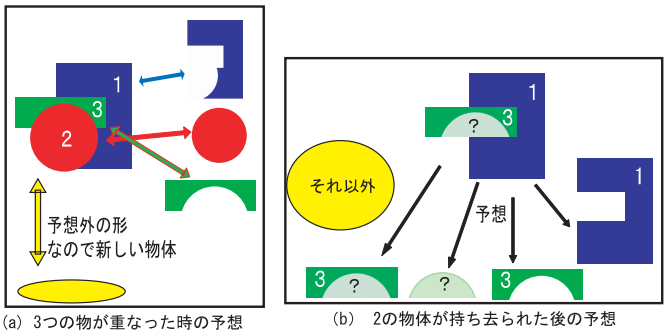
Step2　同じ形状があれば，そのレイヤの物体領域を「持ち去り」領域としそのレイヤを削除する

Step3　Step2 でレイヤを削除したら，全てのレイヤの次状態の形状の予想を行い Step1 に戻る

Step4　次状態の形状とマッチしなければ「持ち込み」領域とし，新規レイヤとして重なっているレイヤの最大値より 1 つ上のレイヤとして追加する

Step5　レイヤ追加後に「持ち去り」によって検知される差分領域を予想し形状を生成する

次状態の物体領域の予想とは，持ち込まれた時の物の形状を物の重なりを考慮した上で持ち去られる時の形状を生成することである．積み重なったレイヤの上から順に，画像中に見えている部分の形状で次状態の「持ち去り」予想画像を生成する．



(a) 3つの物が重なった時の予想　　(b) 2の物体が持ち去られた後の予想

図 9　レイヤ検知の概念

図 9-(a) は番号の順に物体を検知したときの予想を示す．この時，予想されるのは 3 つの物体の重なりを考慮した形状か，それ以外の持ち込み時の物体形状のみとなる．図 9-(b) は (a) の状態から 2 の物体が持ち去られた時の状態を示す．予想で 2 の下のレイヤにある領域をシステムは，見たことが無いので「？」の領域としておく．この領域と同じ階層のレイヤをつなぎ予想することでどの物体が持ち去られても対応することができる．

4.2　イベント解釈例

図 10 は，物が奥に隠れた形で持ち込まれた場合に，物体領域を予想しイベント解釈を行なう例である．frame92 で物体を検知し，何も物が置かれていないのでレイヤを追加する．frame221 では検知した領域は frame92 の領域とマッチしないので，物が持ち込まれ 2 つの物体があることを示している．frame507 で検知した領域は，frame221 で予想した形状とマッチするのでレイヤを削除する．frame606 では，過去に置かれた物の奥に物が持ち込まれる様子を示している．frame606 では，物体 A(frame221 の物体) は物体 B(frame606 の物体) の１つ下のレイヤとシステムは解釈している．図 11 に示すように実際は frame606 のレイヤ構造は物体 B の 1 つ上のレイヤに物体 A がある．frame733 で物体 A が持ち去られる事で現れる物体 B の領域は，システムは見たことが無い領域なので「？」の領域が現れる．「？」の領域は，ある物の奥に物が置かれた場合の例外措置として，同レイヤ近辺の frame606 で検知した領域と結合した形状を予想 (図 11) し，その 1 つとして図 10 の frame733 の右のような画像を生成することができる．実際は frame840 で検知された物体領域は，frame733 で予想した形状とマッチしたので「持ち去り」と解釈しレイヤを削除することができた．このことからオクルージョンの問題にも対応できたことがわかる．

5.　情報検索・提示のためのユーザインタフェース

5.1　対話インタフェースの構造

ユーザとのやり取りには，イベント解釈によって記述された物体検知ログの情報を用いる．物体検知ログは，ユーザの指さした付近の座標や時刻をキーにして検索できるよう提示する画像がセットになっている．タスク決定には，音声認識によって得られるユーザの発話に含まれるキーワードを用いる．タスクは「誰が持ってきたのか？」と「誰が持っていったのか？」という 2 種類に限定し，数種類の認識用辞書をそれぞれ用意した．システムは，ユーザの問い合わせの座標の値から候補となった

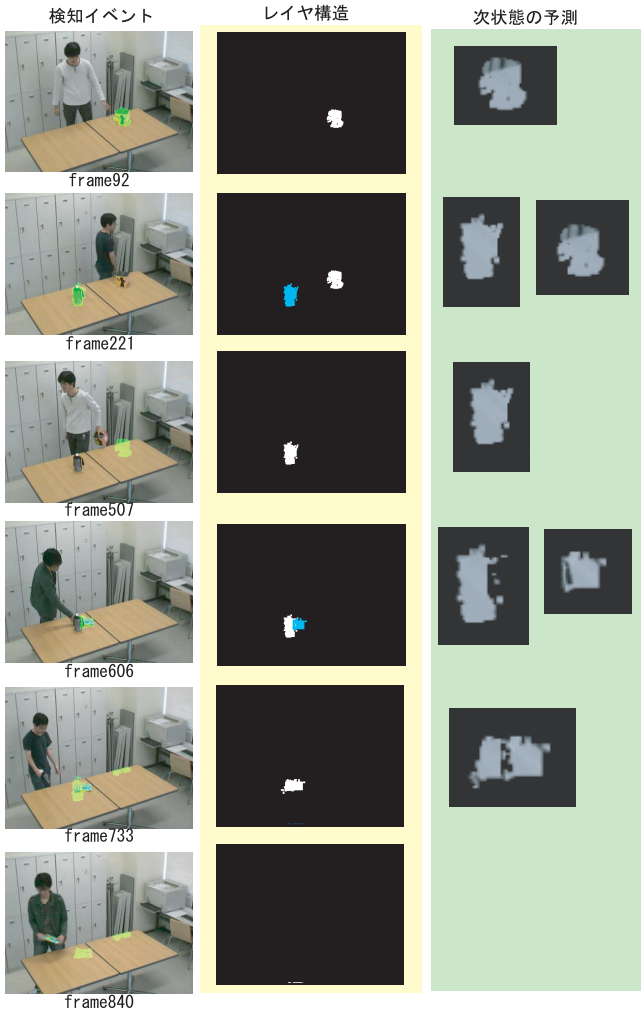


図 10　イベント解釈の成功例

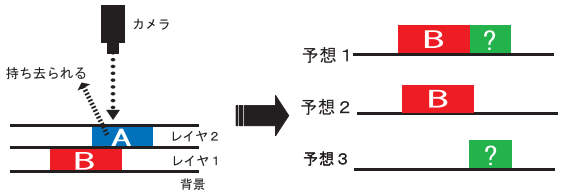


図 11　frame606 のレイヤ構造と次状態の予想

結果画像は，全てユーザに確認を取りながら聞いていく構造になっている．画像処理だけでユーザを判断するのではなく，音声入力をトリガにして問い合わせを開始していくことで，通り過ぎていくだけの人をユーザと誤認識しないようにした．

5.2　対話的な問い合わせの実現

物が監視映像内を移動したと解釈するには，その物体の領域を画像処理で追跡すれば可能である．移動した事で物の見え方が変わった場合，自動的に識別するのは困難になる．人間は物体の色や形が変わっても，物が移動していくシーンを見れば，移動したかどうかは簡単に判断できる．図 12 はユーザが追従して問い合わせをすることでシーンの変遷のリンクが張られていく様子を示している．まず図 12-(a) は現在シーンで，テーブルの上の物について「誰が持って来たか？」問い合わせると，(b) のような結果が返ってくる (結果は動画なのでその中の一枚)．ユーザは (b) の一連の画像を見て映っている人がテーブルの上で物を移動させたシーンだと理解できる．そこでさらに

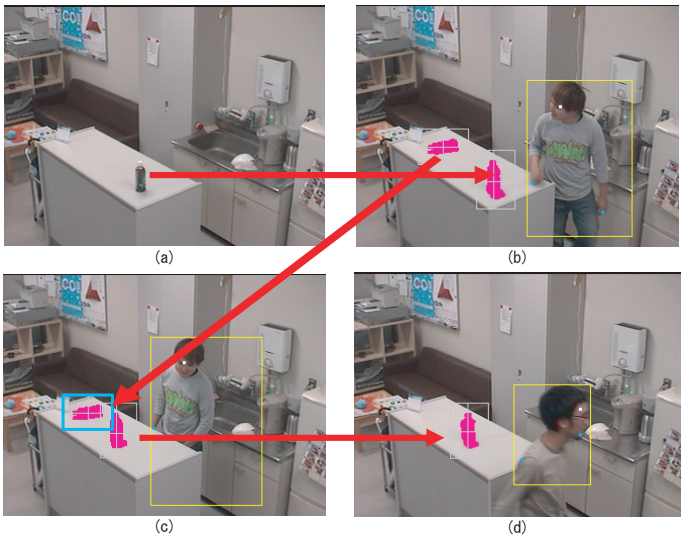


図 12　対話的な問い合わせでリンクが張られる様子

(b) に対して，問い合わせると (c) の画像が返ってくる．ここでもテーブル上を移動させただけということが見て解ったので (c) に対して問い合わせると (d) の画像が提示される．(d) より過去には移動させたシーンは発見できなかったので (d) に映っている人が本当に (a) のシーンの物を持ち込んだ人とユーザは判断する．ユーザとやりとりを行い，過去を遡って問い合わせをする事で本来のユーザの問い合わせに対する結果を提示することが可能になる．

5.3　イベント検索アプリケーション

これまで述べてきたシステムを，

- マウスポインティングを用いた検索アプリケーション
- 指差しジェスチャを用いた検索アプリケーション

の 2 つの GUI アプリケーションで実装している．システムの動作画面を以下の図 13 に示す．マウスを用いた検索アプリは，

図 13　イベント検索アプリケーションの動作画面

ディスプレイを見ながらマウスで大まかな領域を指定し，キーボードからテキスト入力によって問い合わせを行う．システムはイベントデータベース中の現在物体の存在する領域や，かつて物体があった領域をマウスポインティングの領域と照合することで，ユーザの問い合わせたい物体について候補を提示し，ユーザはそれを選択することで簡単に指示を出せる．図 13-(a) では，テーブルの上の物体にカーソルを合わせた状態で「ここに誰が置いたの？」という問い合わせに，ポップアップして結果画像を提示している様子を示している．ジェスチャ認識や音声認識を組みあわせることで，GUI ではなく，ユーザが監視シーンの現場で問い合わせを実施することも可能である．図 13-(b) では，ユーザの指先の座標と音声の入力を基にした，検知イベントの画像をユーザに提示している様子を示している．

6.　おわりに

シーンに起こったイベントを自動検知し，解釈・保存するシステムを提案し，後からユーザが希望するイベントシーンをマウスやジェスチャを使って簡単に問い合わせることができる検索システムを実装した．シーン変化の検出に誤りが多いと対話回数が増えユーザの負担が増すので，物体領域の画像パターンの変化をより詳細に解析しイベント自動検知の性能を向上させる必要がある．ユーザが対話的に問い合わせをしたり，システム側から検索の手助けとなる情報を提示してユーザを補助するためには，マウスやジェスチャによる指示と，音声やテキストによる指示をどう組み合わせるか，またどのような情報を提示すればユーザの支援になるか検討することが今後の課題である．

文　献

[1] VSAM:“ Section I, video surveillance and monitoring ”, Proc. of the 1998 DARPA Image Understanding Workshop, Vol.1, pp. 1.400 (Nov. 1998).
[2] R. Collins, A. Lipton, T. Kanade, H. Fu-jiyoshi, D. Duggins, Y. Tsin, D. Tolliver,N. Enomoto, and O. Hasegawa: “ A sys-tem for video surveillance and monitor-ing: VSAM nal report ”, In Technicalreport CMU-RI-TR-00-12, Robotics Insti-tute, CMU, May 2000.
[3] 森武俊, 野口博史, 佐藤知正: センシングルーム —部屋型日常行動計測蓄積環境第 2 世代ロボティックルーム—, 日本ロボット学会誌,Vol.23 No.06,pp.25-29,2005.
[4] 橋本秀紀, 新妻実保子, 佐々木毅: 空間知能化—インテリジェント・スペース—, 日本ロボット学会誌,Vol.23 No.06,pp.34-37,2005.
[5] 若村直弘，鈴木健一郎，入江耕太，梅田和昇：”インテリジェントルームの構築 - 直感的なジェスチャを用いた家電製品の操作-”，画像の認識・理解シンポジウム (MIRU2005)，IS3-108，pp.1074-1081，2005.7.
[6] 市川 徹，山澤 一誠，竹村 治雄，横矢 直和: 高解像度全方位ビデオカメラを用いた遠隔監視システムにおけるイベント検出，電子情報通信学会 技術研究報告,PRMU2000-213,pp.87-94,2001.
[7] 上岡隆宏, 河村竜幸, 河野恭之, 木戸出正継: I’m Here!：物探しを効率化するウェアラブルシステム，ヒューマンインタフェース学会論文誌,Vol.6,No.3,pp.19-30,2004.
[8] 鈴木健一郎，和田正樹，梅田和昇：インテリジェントルームにおける家電機器操作の高度化，日本機械学会ロボティクス・メカトロニクス講演会’06，2P1-E21，2006.5.
[9] 藤吉弘亘, 榎本暢芳, 長谷川修, 金出武雄: “ アクティビティモニタリング. 屋外監視映像の要約と WWW 上表示・検索システム. ”, 第 7 回画像センシングシンポジウム論文集, pp. 423.428(2001-6).
[10] 松井康作，浜田玲子，井手一郎，坂井修一: 監視映像におけるオブジェクト移動履歴検索，第 67 回情報処理学会全国大会講演論文集，Vol.3，pp.79-80,2005.
[11] 槇原靖，白井良明，島田伸敬: 対話を用いた物体認識のための照明変化への適応，電子情報通信学会論文誌 D-II，Vol.J87-D-II，No.2，pp.629-638，2004.
[12] 滝澤正夫，槙原靖，白井良明，島田伸敬，三浦純: サービスロボットのための対話システム，システム制御情報学会論文誌，Vol.16，No.4，pp.24-32，2003.
[13] 井本浩靖，白井良明，島田伸敬，三浦純: インタラクティブビジョンにおいてユーザから有用な助言を得るための手法，電子情報通信学会 PRMU 研究会,2005.9.
[14] 宮本圭，上野敦志，武田英明：オフィス環境における文字情報の検出と利用に関する研究，人工知能学会知識ベース研究会，2000.
[15] 佐治禎基，上野敦志，武田英明：移動のある物体の認識・管理を行うオフィスロボットの構築，電子情報通信学会人工知能と知識処理研究会，知能ソフトウエア工学研究会,2000.
[16] 小倉英樹，片山憲昭，島田伸敬，白井良明: 複数の認識エンジンを併用したビデオ操作支援システム，電子情報通信学会総合大会 A-19-5，2006.3.
[17] 島井博行，栗田多喜夫，梅山伸二，田中勝，三島健捻: ロバスト統計に基づいた適応的な背景推定法，電子情報通信学会論文誌，Vol.J86-D-II，No.6，pp.796-806，2003.6.
[18] 鷲見和彦，関真規人，波部斉：物体検出 - 背景と検出対象のモデリング -，情報処理学会研究報告 (CVIM)，Vol.2005，No.88，pp.79-98,2005.9
[19] Hironobu FUJIYOSHI,Takeo KANADE: ”Layered Detection for Multiple Overlapping Objects”, IEICE TRANSACTIONS on Information and Systems Vol.E87-D No.12 pp.2821-2827,2004.12
[20] 滝澤正夫, 白井良明, 三浦純, 島田伸敬, 先山卓朗: ”状況，文脈，発音，類似度，を考慮して発話を解釈する対話システム”, 大阪大学修士論文，2003.
[21] 大語彙連続音声認識システム Julius: http://julius.sourceforge.jp/
[22] OpenCV: http://www.intel.com/technology/computing/opencv