屋内シーン変遷の対話的検索における検索の成否に基づく誤認識の発見と変遷系列の自動訂正

牧　和宏† 島田 伸敬† 白井 良明†

† 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: †maki@i.ci.ritsumei.ac.jp, ††{shimada,shirai}@ci.ritsumei.ac.jp

あらまし 自動画像認識によりシーン変化や物体の種類を誤りなく認識することは困難である．時系列にわたってシーン変遷を認識する場合においては，あるシーンの認識を誤るとそれ以降のシーンの認識も連鎖的に誤る可能性がある．誤認識を解消するためにユーザと対話的にやり取りを行う研究があるが，従来法では各認識の訂正を個別に行うため，訂正が複数必要な場合は効率が悪い．そこで，本論文ではシーン検索時のユーザとの対話をきっかけにしてシステムのシーンの誤認識を発見・訂正し，さらに訂正されたシーンの認識結果に依存した別シーンの検証，必要があれば訂正を行う手法を提案する．提案手法では，一度の訂正により複数の認識結果を訂正することができ，効率よくシステムのシーンの誤認識を解消することができる．最後に，提案手法を実装し，実験を行った．
キーワード 監視システム，対話的イベント検索，自動訂正，音声対話，シーン認識

Interactive Inquiry for Indoor Scene Transition with Awareness and Automaic Correction of Misunderstanding

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

† College of Information Science and Engineering,Ritsumeikan Univ. Nojihigashi 1–1–1, kusatsu-shi, Shiga, 525–8577 Japan
E-mail: †maki@i.ci.ritsumei.ac.jp, ††{shimada,shirai}@ci.ritsumei.ac.jp

Abstract It is difficult to recognize scenes and objects perfectly by an automatic manner of current image understanding techniques. In recognition of scene transition, one false scene recognition may cause other faults. We propose a method to be aware of the existence of the false scene recognition from user's suggestions via an interactive inquiry system. Furthermore, the system not only corrects the false recognition but also verifies other scene recognitions related to the corrected scene and corrects them if needed. Experiments on the system with the proposed method have shown that the system can find and correct false scene recognitions.
Key words surveillance system, interactive inquiry, automatic correction, speech dialog, scene recognition

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 Advanced Research Projects Agency) の下，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]．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 [2]，どこに何があるのか，誰が物体を「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 [3]．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 [4]．日常的な生活を考えると，これらの研究のように人が置き忘れた物体や，紛失した物体の捜査・検索を手助けする環境は有用である．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 [5]．その他にも，WWW(World Wide Web) 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 [6] もある．

そこで，本研究では，監視映像下から物体の「持ち込み／持ち去り」イベントを自動的に検知・整理しておき，ユーザが探しているシーンを対話的に検索できるシステムの開発を行う．ユーザがシステムと対話的に検索を行うシステムは開発されているが [7], [8]，これらのシステムは誤りのない完全な情報を対象とする．一方，本システムは自動画像認識によるシーン解析結果を対象とするため，誤ったデータを含む，必要なデータが存在しないなど，不完全なものとなる可能性がある．さらに，システムが誤認識を一つ起こすと他の認識にも影響を与えてしまうため，自動認識の誤りを発見・訂正し，それに関連した誤認識も訂正できる機能が必要である．画像認識の誤りを解消するために，ユーザと対話的にやり取りを行う研究があるが [9]- [12]，これらの研究では，各認識の訂正を個別に行うため，訂正が複数必要な場合は効率が悪い．そこで，本論文ではシーン検索時のユーザとの対話をきっかけにしてシステムのシーンの誤認識を発見・訂正し，さらに訂正されたシーンの認識結果に依存した別シーンの検証，必要があれば訂正を行う手法を提案する．提案手法では，ユーザは自分の知りたいシーンを探索するために必要な対話を行うだけで、その対話内容からシステム側が自ら誤認識を発見して正しい解釈を見つけ訂正する。また時空間的に関連する複数の認識結果を検証・修正することができ，効率よく自動シーン認識の誤りを解消することができる．

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 つから構成される．

イベント検知部は，天井に設置されたネットワークカメラからの入力画像を処理するための計算機によって構成される．常時，部屋の中の映像の処理を行い，映像中に人を検知すると”Human-DB”にそのシーンを保存する．さらに，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”Event-DB”に保存する．

イベント解釈部は，イベント検知部で”Event-DB”に保存されたシーンに「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,「持ち込み／持ち去り」イベントを判別するために検知された画像の処理が必要だが，イベント検知部と並列に処理させることにより，イベント検知部がイベント解釈部の処理を待つことなく入力画像を取得できる．

ユーザインタフェース部は，マイクとスピーカ，計算機及びネットワークカメラから構成される．イベント検知部と同一のネットワークカメラから画像を取得し，ユーザが現場で物体やかつて物体があった場所を指差しながら発話できるよう，ジェスチャ認識と音声発話の処理を行う．ユーザから問い合わせがあれば，保存されている画像データを検索し，物体を持ち込んだ・持ち去った人物の映っているシーンをユーザに提示する．また，ユーザーとの対話によって”Event-DB”に誤認識されたイベントがあったと判明した際には，”Human-DB”に保存されているイベントデータを用いて再解釈を行なう．なお音声対話は，音声認識エンジン Julius/Julian [18] を用いた対話システム [13] を拡張し実装した．

3. 屋内シーン変遷の認識

ユーザが探しているシーンをシステムが提示するためにシーン変遷を認識する方法について述べる．まず，シーン変化を自動的に検知する手法について述べ，検知したシーンがどのようなシーン (「持ち込み／持ち去り」イベント) なのかを解釈する手法について述べる．

3.1 イベント検知部

イベント検知部で行う屋内シーン変化の自動検知には，文献 [14] の手法を用いる．まず，急激な照明変化にロバストな手法 [15] で背景を推定し，入力画像との差分をとり，前景領域を検出する．次に，前景領域から影を取り除き [16]，人の持つ特徴 (大きさ，肌色，髪色) を用いて人を検出する．人を検知したシーンは”Human-DB”に保存する．人領域以外の領域を物体候補領域とし，ある一定区間同じ場所に観測された候補領域を物体領域とする．物体領域が見つかれば，そのフレームから過去にある 10 フレームまでイベント発生区間とし，その区間の画像列をまとめて”Event-DB”に保存する．図 2 に物体検知の様子を示す．図 2 における、黄色矩形，青色矩形，赤色矩形は，それぞれ人領域，物体領域候補，物体領域である．図 2-(a) は人が物体を置いた時の画像である．このフレームでは人と物体が離れていないので物が置かれたか検知できない．図 2-(b) と (c) では物体候補領域が観測された．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 2-(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，この時点で初めて当該候補領域が物体領域として検知される．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

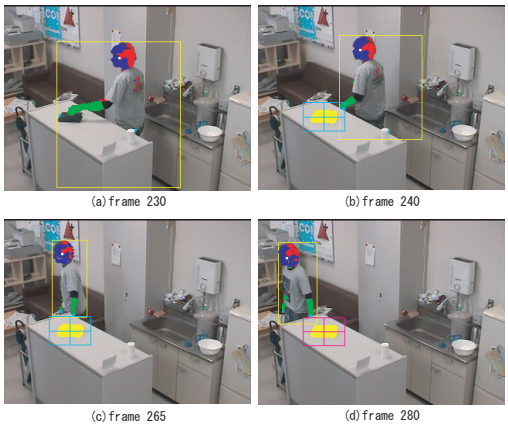


図 2　物体検知の様子

ができる．検知されたシーン変化がどのようなシーンであったかについては，イベント解釈部で判定され，情報を付与する．

3.2　イベント解釈部

イベント検知部で検知したシーン変化が,「物を持ち込んだ」／「物を持ち去った」のいずれのイベントなのか解釈を行なっているイベント解釈部について述べる．

3.2.1　シーン変化の事前予測によるイベントの解釈

文献 [14] では，物体が持ち込まれる時と持ち去られる時に検知される物体領域はほぼ同じ形状であるという特徴を用いて，時系列に物体領域をマッチングさせることによりイベントの解釈を行う．さらに，画面上での物体の重なりや配置をレイヤ構造で表現することにより，持ち込まれた各物体が持ち去られる時に検知される物体領域の形状を予想している．また，持ち込む前の背景画像のテクスチャを覚えておくことにより，持ち去られた後に見える背景画像のテクスチャとの比較を行い，持ち去られた物体の後ろに一部隠れていた物体領域を見つけ対処する．図 3 は，(a)，(b)，(c) の順にイベントが発生

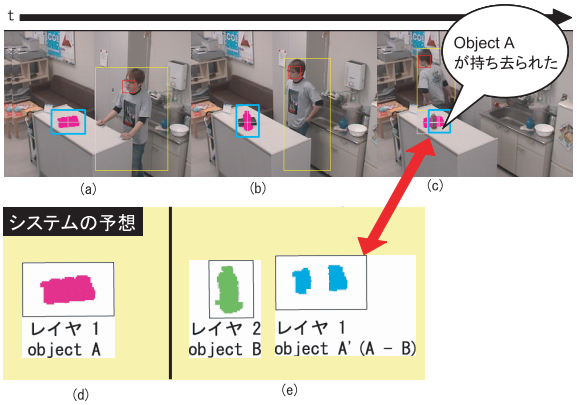


図 3　シーン変化の事前予測によるイベント解釈例

し，イベント解釈を行っている例である．ただし，黄色矩形，青色矩形はそれぞれ，人領域，物体領域である．まず，(a) というイベントが起こった時，物体は 1 つなので予想される持ち去り時の形状は，(d) のように (a) そのものである．次に，先ほどの (a) が隠されるような状態で (b) が置かれる．(b) の形状と，保存されている (d) の形状のマッチングを行なうと一致しないので，新たな持ち込み物体としてレイヤを追加し，その時の持ち去られる時の形状は (e) のようになる．(b) で検知された形状はレイヤ上一番上にあるので，持ち去り時の形状は持ち込まれた時と同じである．(a) の物体領域は，(b) の前に発生したことでレイヤ構造上，下のレイヤとなり (e) の右側のように形状が更新される．そのため，下の物が抜き取られるような動作の (c) のようなイベントが発生しても，システムの予想が (e) なので，右側とマッチして (a) が持ち去られたと解釈することができる．その後，持ち去られた (a) の物体レイヤを削除して (b) の物体形状のみがシステムの保持しているシーン記述による次回イベント時の変化予測となる．

3.2.2　イベント前後の物体と人の位置関係を考慮したイベント解釈の補強

シーン変化の事前予測によるイベント解釈では，物体のみに注目してイベント解釈を行っているが，物体の「持ち込み／持ち去り」は人によって起こされるので，人の情報も使うことによってイベント誤検知を訂正したり，物体情報だけでは誤解釈してしまうようなシーンに対しても解釈を改善することができる可能性がある．そこで，事前予測によるイベント解釈に，人の情報を用いたイベント解釈を追加することにより補強を行う．

事前予測による解釈が持ち込みと判定された場合，検知フレームから過去に遡って物体を追跡し，人と物体が連結していると判定されたのならば持ち込みと判断する．ここで，物体の追跡は CamShift [17] を用いて行う．また，ラベリング処理を用いて人領域と物体領域が連結しているか判定をする．図 4-(a) では，704 フレームで物体を検知し持ち込みと解釈されたので，過去に遡って物体追跡を行った例である．676 フレームで物体領域と人領域は連結しており，持ち込みと判定できる．また，持ち去りと判定された場合，検知した物体領域の差分が出る直前のフレームから現在方向に向かって物体を追跡し，人と物体が連結すれば持ち去りと判断する．図 4-(b) では，142 フレームで物体の変化を検知し、それが持ち去りと解釈されたので，物体の変化の差分を検知する前の 124 フレームから現在方向へ物体追跡させた例である．142 フレームで物体と人が連結しており，持ち去りと判定できる．

また，この解釈法を追加することにより，図 5 のようなシーン変化の誤検知にも対応できる．図 5-(a) は，人影を物体と誤検知した例だが，694 フレームから過去に遡って物体追跡を行っても人と連結することはないので，誤検知であることを認識できる．図 5-(b) は，照明変化による変化を物体と誤検知した例であるが，物体を検知した 1203 フレーム前後に人は存在しないので，誤検知であることを認識できる．

3.3　イベント検知・解釈の性能評価

イベントの自動検知・解釈の性能評価を検証するため

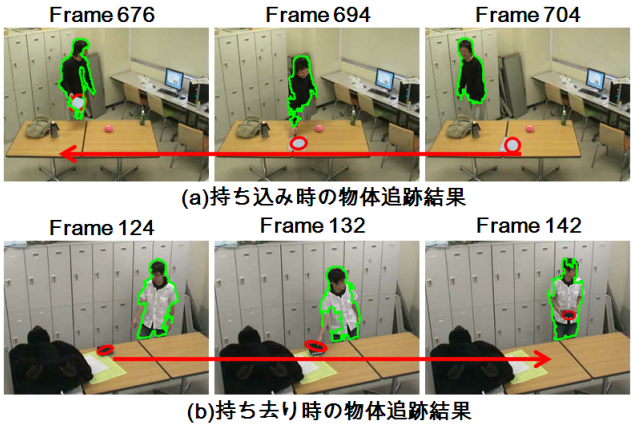


図 4　各イベント時の物体追跡結果例

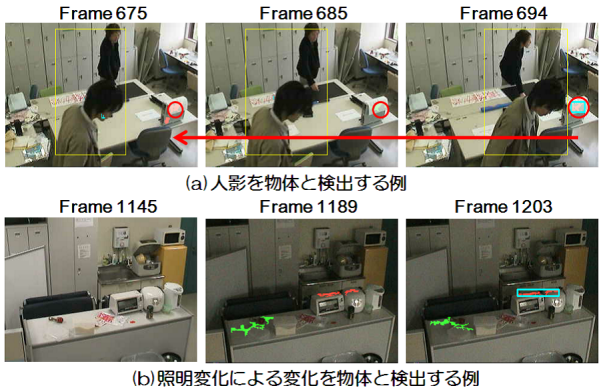


図 5　シーン変化の誤検知例

に行った実験結果を示す．実験は約 5 時間，32GB(画像数 23 万枚) のデータを使用して行い，人が頻繁に訪れる流し場や自然光の影響を受けやすい窓付近の休憩スペースなど複数の場所で行った．自動検知の性能評価には，正解検知率 (検知されるべき物体が検知された割合)，誤検知率 (誤検知数／総検知数) を用いて行う．表 1 に，イベント検知結果を示す．

表 1　イベント検知実験結果

正解検知率	93%
誤検知率	23%

未検知となった要因は，机と物体の色が似ているために前景領域抽出時に差分が出ない，物体が影と判定される，物体が小さいことが挙げられる．誤検知は，人の一部や影を物体と検知した．影の対処として，追跡を用いる処理を追加したが，人が通り過ぎることにより追跡対象が人に張り付いたシーンは検知された．次に，イベント検知実験の正解検知シーンを用いて，イベント解釈を行ったところ，正解解釈率は 87%となった．

4.　シーン検索時の対話を用いた誤認識の発見と変遷系列の自動訂正

物体領域をある程度忠実に検知することが出来れば高い精度で解釈することができる．しかし，依然未検出シーンや誤解釈シーンも存在し，ユーザがこのような誤認識したシーンを検索した場合，システムは正しく希望シーンを提示することができない．そこで，シーン検索時のユーザとの対話を用いて，誤認識シーンを発見・訂正し，さらに訂正されたシーンの認識結果に依存した別シーンの訂正も行う．

4.1　音声と指差しジェスチャを用いたシーン検索

図 6 に対話を用いたシーン検索の概念図を示す．ユーザは指差しと音声発話によりシステムに問い合わせる．ユーザが問い合わせると，システムは指差す場所付近で起こったイベントシーンを探し，検索結果の候補とする．さらに，音声認識により得られる情報（イベントの種類や時間情報）を用いて，検索結果を絞り，最有力の検索結果をユーザに提示し，その結果が正しいか確認する．もし，ユーザから検索結果が間違っていると指摘されれば，ユーザが希望するシーンが見つかるまで他の検索結果を提示し，確認を続ける．

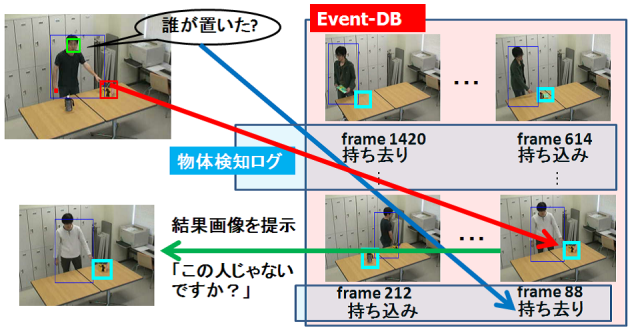


図 6　対話を用いたシーン検索

4.2　シーン検索時の成否に基づく誤認識の発見

ユーザが希望するシーンが見つかるまで検索候補を提示し続けたとしても，該当シーンが見つからない場合がある．人の認識能力がシステムよりも高いことを考えると，ユーザが提示シーンを見誤るよりもシステムがシーンの誤認識を起こしたと考えて，ユーザとの対話を利用してシーンを再検索する方が得策である．そこで，シーン検索に失敗した場合，システムは誤認識を起こしたことを発見し，システムが持つデータから候補となりそうなシーンを提示し，ユーザからの回答により誤認識シーンを特定する．

4.3　誤認識シーンの対話的特定

システムが起こす誤認識は，未検知と誤解釈の 2 つに分類される．システムが誤認識を発見した時点では，どちらの誤認識が起きたのかは判断できないので，両者の場合を仮定して誤認識シーンの検索を行い，ユーザに確認を取ることにより誤認識シーンと種類を決定する．

4.3.1　未検出シーンの特定

誤認識シーンが未検知だった場合，”Event-DB”内には該当するシーンがないので，”Human-DB”から該当シーンを検索する．

まず，システムが誤認識を起こしたことを発見する

と”Human-DB”の最新のデータから順（過去方向）に検索をバックグラウンドで開始しておく．それと同時に，ユーザから時間情報を教えてもらい，その時間付近のデータから現在の方向へ検索する．この時，

- 放置物検索の場合はまだ物体が置かれていなかったと分かっているできるだけ直近の時刻
- 紛失物検索の場合はすでに物体が置かれていたとわかっているできるだけ過去の時刻

を，それぞれ尋ねる．検索を並列でさせることにより，シーン候補の検知の高速化を図る．

シーン候補の検知には，3.1 節で述べたシーン変化検知処理を再度同条件で行っても検知できないので、以下の条件で検知を行う．

- 前景領域抽出時の背景差分の閾値を低くする
- 影判定を行わない
- 物体の大きさ定義範囲を広くする (より面積の小さい，大きい前景領域を物体と認識)

また，画像全体に対して処理するのではなく，ユーザが指差した領域付近のみを処理させることにより，短時間で処理することができる．

条件を緩めてシーン変化検知を行うので，既に検知され”Event-DB”に入っているシーンも再び検知される．そういったシーンは検索している未検知シーンではないので，候補から除外する．候補が見つかればそのシーンを動画で提示し，ユーザに確認を取る．希望シーンであれば，システムは見逃していた未検知シーンを発見し，”Event-DB”に追加する．もし希望シーンでない場合は，提示シーンに意中の物体が映っているかを尋ね，シーン検索すべき範囲を絞る．

4.3.2 誤解釈シーンの特定

誤認識シーンが誤解釈したシーンだった場合，”Event-DB”内に該当するシーンは存在する．そこで，”Event-DB”内のシーンをユーザに提示することにより誤認識シーンを特定する．

未検知シーン検索と同様，ユーザに時間に対する情報を音声入力させ，それ以降のシーンを提示する．ユーザが紛失物について尋ねた場合，システムはユーザ指定場所で起こった持ち去りシーンをすべて提示する．それでも該当シーンがない場合は，持ち去りを持ち込みであると解釈した可能性があると考え，持ち込みと解釈されたシーンをユーザに提示し確認する．ユーザが放置物について尋ねた場合，システムはユーザ指定場所で現在置かれていると認識されている物体の持ち込みシーンを全て提示する．それでも該当シーンがない場合は，持ち込みを持ち去りと解釈した可能性があると考え，持ち去りと判定されたイベントをユーザに提示する．また，持ち去りと誤解釈された物体は既に持ち去られたと考えており，ユーザにその物体が持ち込まれたシーンを提示しい．そこで，持ち去りシーンの提示で誤認識シーンを特定できない場合は，既に持ち去られた物体の持ち込みシーンをユーザに提示する．

4.3.3 誤認識シーンの特定機能を追加したシーン検索アルゴリズム

実際にユーザから問い合わせに対し，システムが誤認識を発見した段階では誤認識の種類は分からない．そこで，ユーザの問い合わせからシステムが誤認識を特定するまでの処理の流れを図 7 に示す． ユーザの問い合わ

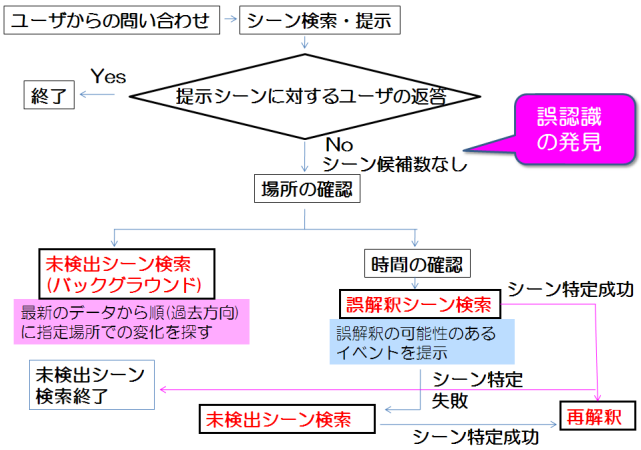


図 7　誤認識シーンの特定までの流れ

せに対し，シーンの候補を提示する．全ての候補がユーザの希望するシーンでなければ，ユーザが指定している場所が正しいか確認する．場所が正しかった場合に，システムは誤認識をしていることを発見し，誤認識シーンの特定作業を行う．まず，誤解釈シーンの検索をメインで行う．ユーザに物体が「置かれていた／置かれていなかった」時間を教えてもらい，その時間を用いて誤解釈シーン検索・提示を行う．これと同時に，”Human-DB”の最新データから順（過去方向）に指定場所での変化をバックグラウンドで検索させておく．候補が見つかっても未検出検索がメインの処理になるまでは提示を行わない．この作業を行っておくことで，後に誤認識シーンが未検出シーンだった場合により早く候補を提示することができる．メインの処理である誤解釈シーン検索の提示結果でユーザの希望するシーンが見つかれば，それで誤認識シーン（ユーザ希望シーン）は特定され，並列処理を行なっていた未検出シーン検索も終了する．希望シーンが見つからなかった場合はメインの処理を未検出シーンとし，時間情報付近のデータから現在方向に指定場所での変化を検索する．未検出シーン検索により誤認識シーンを特定できれば，検索を終了する．

4.4 誤認識の訂正と誤認識に関連したシーン変遷の再解釈

誤認識シーンを特定した後は，”Event-DB”にその情報を反映させる必要がある．まず，未検出だったシーンを特定した場合はその情報を”Event-DB”に追加し，誤解釈だった場合はそのシーンの記述を訂正する．追加・訂正後のシーン解釈が持ち込みになる場合は，そのシー

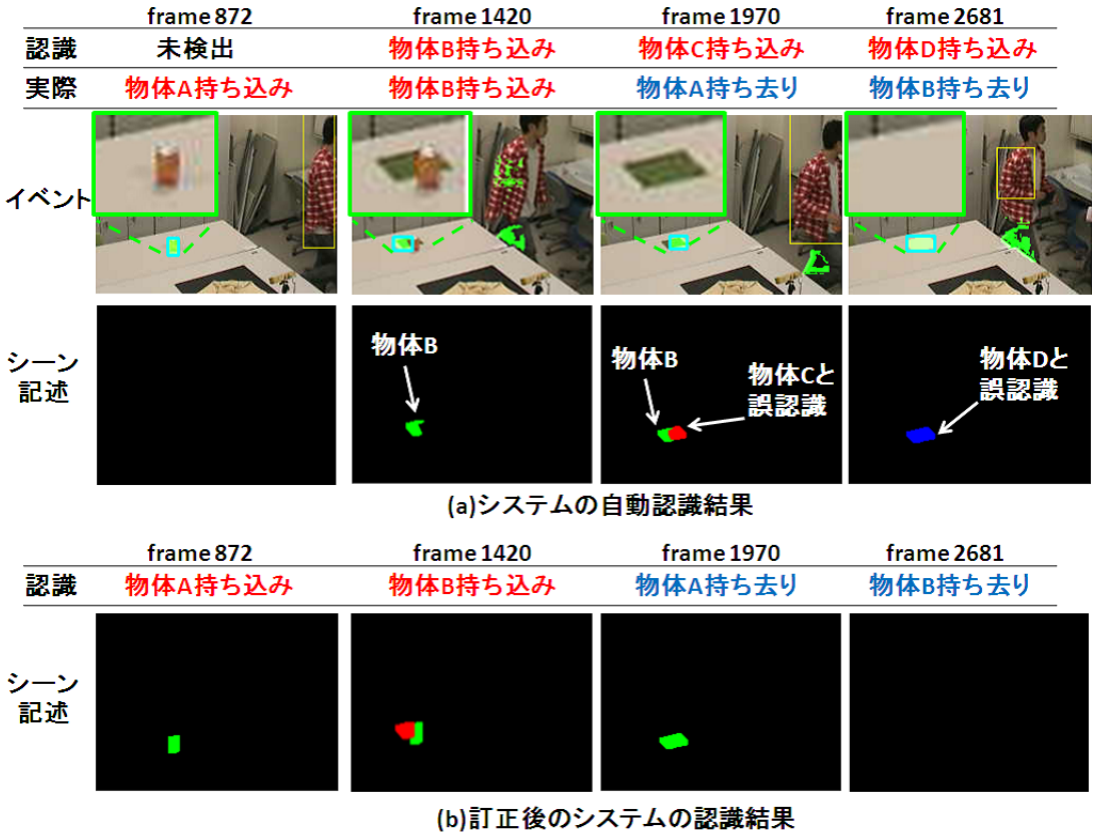


図 8　シーン変遷の自動認識結果と訂正後の認識結果

ン変化を追加・訂正し，それ以降に起こったイベントについても再解釈を行い，訂正がないか確認する．持ち去りだった場合は，そのシーン変化がどの物体の持ち去りによるものなのかを特定する．物体の特定はその時点で置かれたと判断された物体の中から最もシーン変化がマッチするものとする．物体が確定した場合は，その時点で特定物体が持ち去られたようにシーンの記述を訂正し，それ以降のイベントの再解釈を行い，訂正がないか確認する．

誤解釈の訂正によるシーン変遷の再解釈例

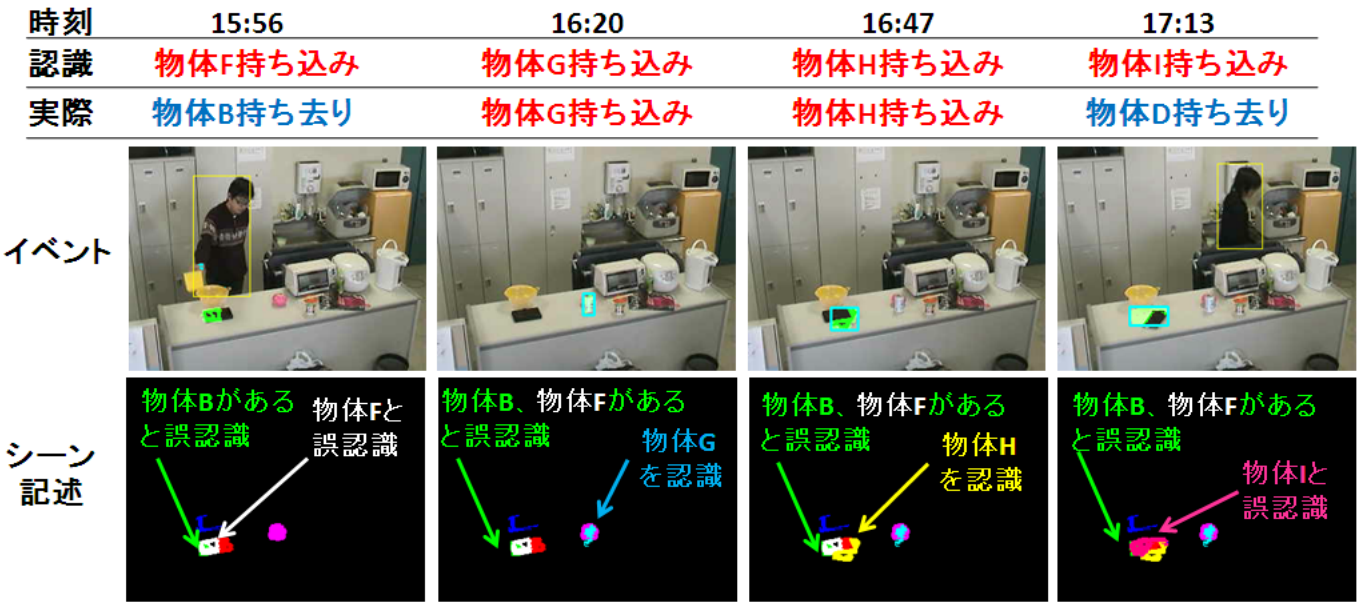
一つの誤認識を訂正することにより，その他のシーン変遷も訂正していく例を示す．まず，システムの自動認識の結果を図 8-(a) に示す．872 フレームで物体 A を持ち込んだシーンを見逃したため，1970 フレームでの物体 A の持ち去りを新しい物体 C を持ち込んだと誤認識した．さらに，この誤認識のために 1970 フレーム以降の各物体の持ち去り時の予想領域を間違い，2681 フレームで物体 B を持ち去ったシーンを新しい物体 D を持ち込んだと誤認識してしまった．

この状況で，ユーザが物体 A を持ち去ったシーンについて問い合わせた場合，システムが誤解釈検索により，1970 フレームのシーンが目的のシーン（物体 A を持ち去っている）であると特定できる．そこで，1970 フレームのシーンの記述を訂正するが，持ち去りと訂正するので，まず持ち去られた物体を特定する．この例では，以前検知したシーンに該当しそうなものがないので，”Human-DB”から見逃したシーンを探し，872 フレームで該当シーンを見つけることができる．それを”Event-DB”に追加し，それ以降のシーン解釈についても訂正の検討を行う．その結果，図 8-(b) のようにシーンの解釈・記述を訂正できる．1970 フレームでのシーンの記述は，物体 B の一部が物体 A の後ろに隠されていたことを認識したことにより記述を行った．これは，物体 A の持ち込みを検知したことにより，物体 A が置かれる前の背景のテクスチャを得ることができ，それと物体 A 持ち去り後の背景を比べることにより実現した．この結果、1970 フレームの誤認識のために持ち込みと判定していた 2681 フレームのシーンも正しく持ち去りと判定することができる．

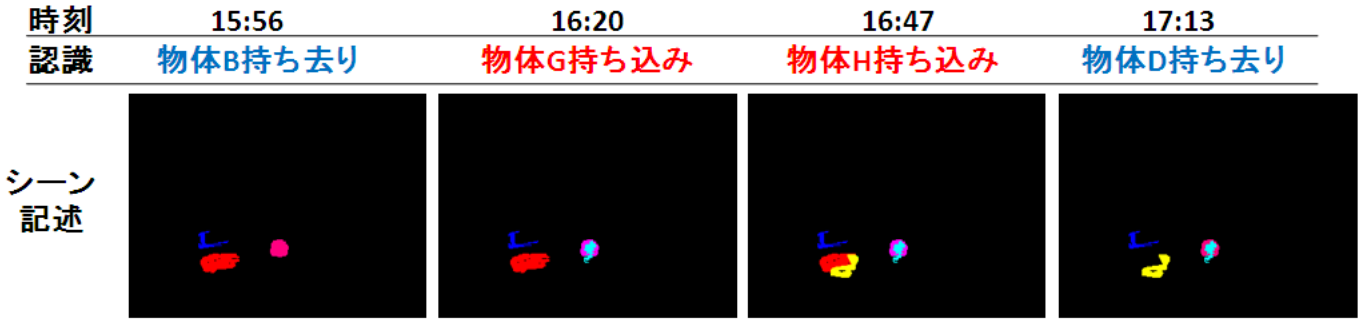
5.　誤認識シーン検索結果

最後に，誤認識シーンの検索を行った結果例を示す．まず，システムは図 9-(a) のような自動認識結果 (一部) を行った．時刻 15:56 に検知されたシーン (物体 B の持ち去り) の解釈を誤ったために，時刻 17:13 のシーンも誤解釈してしまった．ここで，ユーザが物体 B の持ち去りについて尋ねると (図 10)，候補となったシーンは物体 A の持ち去りシーンのみであった (図 11)．

そのため，ユーザの希望シーンが見つからないので誤認識検索を行う．まず，ユーザに指定物体が置かれていた時間を尋ね，15 時ぐらいには物体が置かれていたこ



(a)システムの自動認識結果(一部)



(b)訂正後のシステムの認識結果

(c)誤解釈シーン検索時の表示

図 9　シーン変遷の認識結果とユーザへの表示ウィンドウ

図 10　ユーザが尋ねている様子

図 11　候補シーン

とを知り，それ以降の”Event-DB”内の持ち込みシーンを表示する．この時の誤解釈検索のユーザへの表示ウィンドウは図 9-(c) のようになる．実際は各シーンが動画で表示される．これを見ると，ユーザは 3 番 (図 9-(c) 左下，時刻 15:56 に検知されたシーン) が物体 B の持ち去りシーンであると分かり，システムに伝える．ユーザは希望シーンを見ることができたので，ここで立ち去る．システムは時刻 15:56 に検知したシーンが持ち去りであるという情報から，訂正を行う．まず，どの物体が持ち去られた検討を行い，物体 B であると決定する．その結果，時刻 15:56 のシーンの記述は図 9-(b) のようになる．これは，物体 B が置かれる前の背景のテクスチャと物体 B 持ち去り後の背景を比べることにより，物体 D(シーンの記述上で赤で描画) の一部が物体 B の後ろに隠れていたことを認識したことにより実現した．さらに，物体 D を正しく認識できた結果，時刻 17:13 に検知したシーンが物体 D の持ち去りであると訂正できる．

その後，別のユーザが物体 D の持ち去り（先ほどと同じ場所に関する持ち去り情報）について尋ねると，図 12 の 3 つのシーンが候補となり，通常の検索のみで希望シーンを見つけることができた．

図 12　持ち去りシーン検索の問い合わせに対する候補シーン

6. ま と め

本論文では，物体と人の位置関係を考慮したイベント解釈を行うことにより，屋内シーン変化の誤検知・誤解釈の軽減を行った．また，ユーザとのシーン検索時の対話によって，システムが誤認識を起こしたことを発見し，その誤認識シーンを特定・訂正する手法を提案した．さらに，一つの誤認識シーンの訂正だけでなく，それに関連した他のシーン変遷の誤認識も訂正できる構造にした．今後の課題は，机上での物体の移動や向きの操作，さらに物体を鞄に入れる，人同士が物体を手渡すなど，より複雑なイベントを認識することが挙げられる．また，複数の人が関わるシーンや，複数物体が同時に動くシーンにも対応していく必要がある．

文　献

[1] R. Collins, et al.: ”A system for video surveillance and monitoring: VSAM final report”, Technicalreport CMU-RI-TR-00-12, Robotics Institute, CMU, May 2000.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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).

[7] 神田直之，駒谷和範，尾形哲也，奥乃博: データベース検索タスクにおける対話文脈を利用した音声言語理解，情報処理学会論文誌，Vol.47，No.6，pp.1802-1811，2006

[8] 駒谷和範，上野晋一，河原達也，奥乃博: 音声対話システムにおける適応的な応答生成を行うためのユーザモデル，電子情報通信学会論文誌，Vol.J87-D- ，No.10，pp.1921-1928，2004.

[9] 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.

[10] 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.

[11] 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.

[12] 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 (MIRU2007)，OS-B1-02，pp.81-86，2007.7.

[13] 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.

[14] 片山憲昭，牧和宏，島田伸敬，白井良明：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，画像の認識・理解シンポジウム (MIRU2007)，IS1-20，pp.588-593，2007.

[15] 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.

[16] Daniel Grest, Jan-Michael Frahm and Reinhard Koch: ”A Color Similarity Measure for Robust Shadow Removal in Real-Time”, VMV (Vision, Modeling and Visualization), 2003

[17] Gary.R.Bradski: ”Computer vision face tracking for use in a perceptual userinterface”, Intel Technology Journal, no.2nd Quarter, p.15, 1998

[18] 大語彙連続音声認識システム Julius: http://julius.sourceforge.jp/