部屋内シーン変遷の自動管理システム

Dialog-based Inquiry System of Missing or Left-behind Objects in Office Scene

片山憲昭, 島田伸敬, 白井良明
Noriaki Katayama , Nobutaka Shimada , Yoshiaki Shirai

立命館大学大学院 理工学研究科
katayama@i.ci.ritsumei.ac.jp, {shimada, shirai}@ci.ritsumei.ac.jp

概要 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．また自動認識の誤りがおこった場合，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望ましい．そこで本研究では，監視映像下に起こる「物体の持ち込み・持ち去り」のイベントを自動的に検知・整理する手法について述べる．そしてGUIで直観的な検索操作インターフェースの実装を試みていたのでそれについても報告する．

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．多数の人物が出入りする室内環境で，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 [1, 2]，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 [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]．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]．シーンの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 [6, 7]，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，ユーザー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ながら，ものをと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はじめている [8]-[11]．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ような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て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．そこで本研究では，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,「これをもってきたのは誰？」といった直観的な検索操作インター

フェースの実装を試みていたので，実際の実装によるイベント検知およびシーン検知の結果について報告する．

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．システムの機能は，大きく分けて監視部とインタラクション部からなる．

監視部は天井に設置されたカメラと PC で構成される．カメラは部屋の画像をキャプチャし画像処理を行うことで様々な情報を常時獲得している．そして物に注目し人が物を「持ち込んだ・持ち去った」瞬間を自動的に検知する．検知した画像データは，それぞれシーンごとに保存される．

インタラクション部はユーザとやりとりをする部分で，入力である指差しなどのジェスチャと，発話を処理する．構成はカメラとスピーカとマイクそしてそれらの情報を処理する PC で構成される．カメラは監視部と同一のもので，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．音声対話部は，フリーの音声認識エンジン Julius/Julian[12] を用いて実装したシステムを筆者ら [13] がすでに報告しているので，このシステムを拡張し使用する．

3　人と物の検知

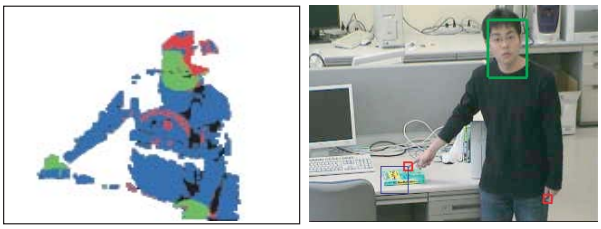
物体検知はこれまでに様々な手法 [15] が提案されている．本研究では背景差分法を用いるが，研究室のような室内環境では，

- 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 [14] 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．背景差分領域をラベリング処理した後，そのラベル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[16] 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



(a)　(b)

図 2: 人検知の様子

の顔と指差しを検知している様子を示している．顔がを検知していれば，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．

1. 「物体候補領域」を発見したフレームから過去一定フレーム「物体候補領域」同士の論理積 (AND) をとる
2. 一定フレーム AND をとって真であった「物体候補領域」を「物体領域」とする
3. 「物体領域」を検知した一定フレームを画像列にして保存
4. 「物体領域」となった領域のみ背景に更新する

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

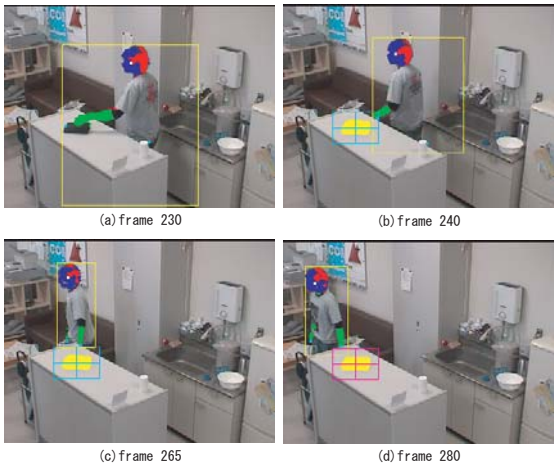


図 3: 物体検知の様子

像である．こ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．

4　シーンの解釈

前章では物体検知法について述べた．物体領域は物の持ち込み・持ち去りが起こった時に検知される．しかし，物の重なりなどによって物の領域が上手く抽出できない場合もある．そこでシステムは，何が起こった時に検知した物体領域か解釈する必要がある．以下の節では，物体領域を抽出した時のシステムの解釈についてその仕組みについて述べる．

4.1　持ち込み・持ち去りの判定

図 4: 持ち込み・持ち去りの例

図 4 は (a) で物が置かれ，(b) で物体領域を獲得，(c) で持ち去りが発生し，(d) でまた物体領域を獲得するという時系列を示している．持ち込み・持ち去りの判定は，過去に保存されたデータと物体の形状を用いる．物体検知した時，過去に保存したデータと物体領域同士の論理積 (AND) をとり一致した画素数がある一定以上の場合は同じ形の物が無くなったとし「持ち去り」と解釈する．一致しなければ新たな物の「持ち込み」があったと解釈しデータを保存

する．図 4 の場合は (b) と (d) の物体領域の AND がほぼ一致したので持ち込みの後，持ち去りが起こったと解釈できる．

4.2 物の重なりが起こった時

物の重なりを考慮した物体検出方法はレイヤー法 [17] などがある．本研究では，検出された物体画像を時系列に階層化して考える事で物同士の重なりを判別する．例えば，図 5 のように，物体を下から抜

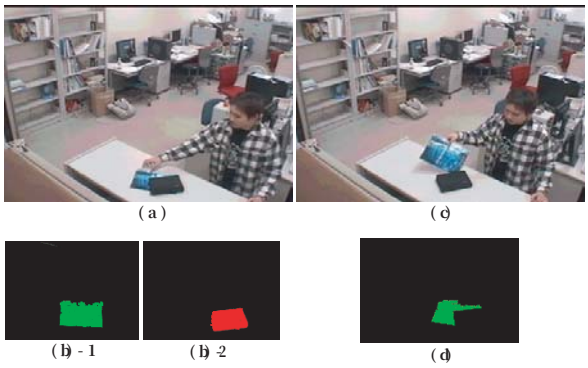


図 5: 重なりが起きたシーンの例

き取られた場合 (例：(a) から (c) の状態が起こった時)，図 5-(d) の様な物体領域が検知される．そのため，システムは (b)-1 の上に (b)-2 の物体が置かれていることを階層的に覚えておく．こうした階層関係を使うことで，事前に次状態で起こりうる差分の形状を予想することが出来る．もし予想していた差分の形状を検知した時は「持ち去り」とし，予想外の差分の形状を検知すれば新たな物が「持ち込み」とする．

4.3 物体が移動した時

物が監視映像内を移動した判定するにはその物の領域を画像処理で追跡していれば可能である．しかし移動する事で物の見え方が変わった場合自動で識別するのは困難である．しかし，人間は多少物体の形が変わろうが，色が変化しようが物が移動する変遷を全て見ていれば移動したかどうか簡単に判断できる．このため，物が移動したかどうかはユーザとの対話を用いて判別する．図 6 は対話をすることで物の移動を解釈し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 6(a) が現在シーンだ

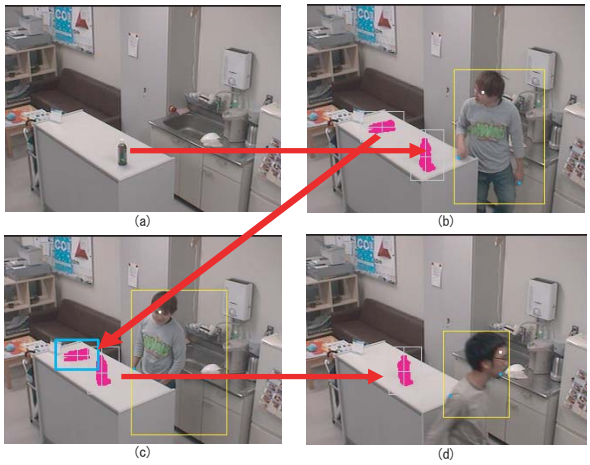


図 6: 対話でできたリンク

として，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．すると (b) のような結果が返ってくる (結果は動画であるのでその中の一枚)．人間はその一部始終を見れば (b) に映っている人がテーブルの上で持ち上げてすぐ横に置いた瞬間に出た，無くなった時と現れた時の 2 つの同じ物体領域だと理解できる．さらに (b) の状態で持ち上げた時に出た上にある物体領域について問い合わせると (c) の画像が返ってくる．しかし，ここでもテーブル上を移動させたということが分かったので (c) の下の持ち去り時に出た物体領域について問い合わせると (d) の画像が結果としてでてくる．(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) に映っている人物が本当に (a) にあった物を持ちこんだ人物になる．ユーザとやりとりを行うことで，(a)-(b) にしか張られていなかったリンクは，過去を遡ることによってやりとりの履歴がシーン同士のリンクになり「誰が持ってきたの？」という本来の問い合わせに対する結果に繋がる．

5 実験

5.1 実験環境

実験に使用した場所は以下の図 7 のような場所である．この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

図 7: 実験に使用した場所

像について物体検知を行った.

5.2 実験結果・考察

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 8 に示す. 図 8-(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.

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い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一度，冷蔵庫にとりに行くような人と物が離れる時間があるからだと考えられる. 物体検知されるべきものでないものが検知された割合（誤検知数／総検知数）は約 50%ほどであった．多くの誤検知は，図 8-(e) のような人影を物体としてしまったものと，(f) のように人が画面内に長時間居た事で徐々に背景に更新され，物体領域ほどの大きさになり保存されてしまったものがあった.

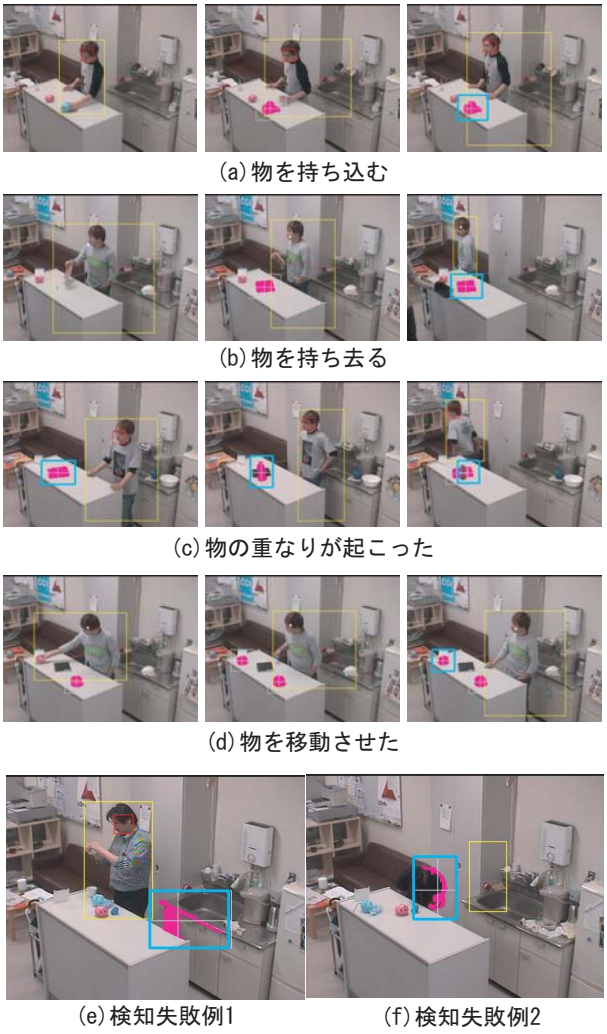


図 8: 検知したシーンの代表例

6 実装した検索アプリケーション

現在システムを，図 9 のような GUI で実装している．この GUI では，ユーザは物体をマウスで指定し、キーボードから文字を入力し，問い合わすことができる．そして，指定された物について保存された画像をポップアップしてユーザに提示する．システム

図 9: 検索アプリケーションの概観

は物体検知した時，物体の重心座標とその画像データのパスを対応させて保存する．そしてマウス入力があった時，そのマウスの座標をキーにして保存されているファイルを表示する．このシステムでは一回の問い合わせで求めた結果が出なくても，図 6 で説明したように，次々に物体を指定し過去に遡って問い合わせる事も可能である．

7 まとめ

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た．そして，マウスを使って直接物体に問い合わせる可能な検索システムを GUI アプリケーションで実装した．今後はユーザがインタラクティブに問い合わせしたり，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機能を作っていく必要がある．音声対話部はまだシステムと繋がっていないので，ユーザが監視映像下の現場に来てジェスチャと発話によってシーンに問い合わせが何度も出来るようなインタフェースに実装する予定である．

参考文献

[1] 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.

[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] 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

[7] 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 .

[8] 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.

[9] 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.

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

[13] 小倉英樹，島田伸敬，白井良明，片山憲昭: 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] 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.

[15] 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

[16] OpenCV: http://www.intel.com/technology /computing/opencv/

[17] 藤吉弘亘, 金出武雄複数物体の重なりを理解するレイヤー型検出法, 第 7 回画像センシングシンポジウム,2001.

片山憲昭：平成 17 年立命館大学理工学部情報学科卒．現在，同大学院理工学研究科博士前期課程在学中．コンピュータビジョンの研究に従事．

島田伸敬：平成 4 年阪大・工・電子制御機械卒，平成 7 年同大大学院博士後期課程了．博士 (工学)．同年同専攻助手．平成 13 年同研究科研究連携推進室情報ネットワーク部門講師，同助教授を経て，平成 16 年より立命館大学情報理工学部知能情報学科助教授，現在に至る．コンピュータビジョン，ジェスチャ認識，ヒューマンインターフェース，インターネットソリューションの研究に従事．電子情報通信学会，人工知能学会，IEEE 各会員．

白井良明：昭和 39 年名古屋大学工学部機械工学科卒業．昭和 44 年東京大学大学院工学系機械工学専攻博士課程修了．同年電子技術総合研究所研究官．昭和 60 年 5 月同制御部部長．昭和 63 年 4 月大阪大学工学部（後に大学院工学研究科）教授．平成 8 年 4 月〜平成 11 年 3 月東京大学大学院工学研究科教授併任．平成 14 年 8 月現在情報学研究所客員教授．平成 17 年 4 月立命大学情報理工教授．知能ロボット、画像処理、ヒューマンインターフェイスの研究に従事．IEEE、電子情報通信学会、人工知能学会、日本ロボット学会などの会員．