

## 人が放置する物体の動的認識\*

渡辺 崇<sup>\*1</sup>, 前田 優樹<sup>\*2</sup>

## Dynamic Identification of Moved and/or Left Objects

Takashi WATANABE<sup>\*3</sup> and Yuki MAEDA<sup>\*3</sup> EcoTopia Science Institute, Nagoya University,  
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8603 Japan

We propose a video-based system which recognizes objects moved and/or left by human, and contribute for the realization of a ubiquitous environment which finds lost properties and suspicious objects in public areas. One method to identify an object moved by a human action is to trace the object and determine times at which the object is grasped and relieved. In case the sizes of the moved objects are at most similar to the sizes of human bodies and the human movements including those to handle objects occur intermittently, an another method is considered, which detects the starts and ends of the human actions and specifies the positions of objects before and after the movements. We adopt the latter approach. The starts and ends of the actions are detected by subtractions of frames captured from a video camera. A series of frame subtractions is used to distinguish between human movements and object movements. An effective background subtraction is developed to identify more than one moved object and objects going autonomously. The history of object movements is recorded and utilized to know where and when the object has appeared.

**Key Words:** Image Processing, Pattern Recognition, Sensor, Shape Measurement, Frame Subtraction, Background Subtraction, Object Detection, Ubiquitous System

## 1. 緒 論

ビデオカメラによるモニタリングシステムは、その利便性、省力性、安全性などのために、盛んに導入されている。本研究では、モニタリングシステムの一つの拡張として、人の行動に伴い移動し、放置される物体を認識するための方法について考える。

コンピュータビジョンを用いたイメージデータ処理の研究は、多くの分野でなされてきた。本研究で注目するような、移動物体の検出、追跡については、対象を人物や車両、交通標識などとして、それらが重なり合う場合も含めて、画像内で移動する物体の認識が行われている<sup>(1)-(4)</sup>。また、移動物体の表現に特定の特徴関数を仮定するより識別率を高める方法<sup>(5), (6)</sup>や、対象とは逆に背景を認識する方法<sup>(7)</sup>が提案されている。そして、これらの応用として、レーザレンジセンサによるロボット誘導も行われている<sup>(8)</sup>。

ビデオカメラでとらえた移動物体の検出、認識の

研究においては、そのほとんどが、物体が移動する過程をとらえようとしており、その機能を用いて、動的環境を対象とした各種のシステムが実現されている。例えば、大西ら<sup>(9)</sup>は、スポーツ中継や講義の撮影におけるカメラワークの省力化を目的に、撮影領域や複数カメラのスイッチングを、人の動きを追跡して自動的に決定している。また、連続した画像データから人の行動列をとらえ、そのパターンを学習することや、行動の意味を判定することで、人の作業環境の整備、改善が進められるであろうし<sup>(10)</sup>、連続した動きの中で、通常とは異なる行動を検出することにより、福祉や医療現場などでの実用も期待されている<sup>(11)</sup>。

防犯の場においては、移動や進入をする人物の発見を中心に研究が行われている。星野、益田<sup>(12)</sup>は、あらかじめ撮影されたパノラマ背景画像と、任意の時点の入力画像から、侵入物体、移動物体の位置を効率的に推定する監視機能を提案している。また、彼らは、設置が簡便で校正の必要がないことは、監視システムの普及に対して望まれることであると指摘している。

日常的な生活を考えると、人の動きのモニタリングに加えて、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、不

\* 原稿受付 2005年5月13日。

<sup>\*1</sup> 正員, 名古屋大学エコトピア科学研究所(☎ 464-8603 名古屋市千種区不老町)。<sup>\*2</sup> 名古屋大学大学院情報科学研究科 [現: キヤノン(株)](☎ 146-8501 東京都大田区下丸子 3-30-2)。

E-mail: watanabe@media.nagoya-u.ac.jp

審者の発見に加えて、ある物体が持ち去られるという置き引きや、不審物などが持ち込まれるという放置される物体の監視も、防犯の面で重要となる。このような監視を行うためには、人の動きと物体の動きを区別する必要があるとともに、ある物体の上に別の物体が重なる場合や、複数の物体が同時に動く場合の対処法が求められる。本論文では、このような問題を考慮した上で、コンピュータビジョンの応用として、人が動かしたり放置したりする物体を検出し、その位置の変化を認識する環境について提案する。

## 2. 人により動かされる物体の発見

本論文では、人が物体を動かし放置することを移置と呼び、移置される物体を移置物体と呼ぶことにする。人が移動させる物体を認識するためには、手、身体と、それらにより動かされる物体が、接したり、離れたりを適切に判断する必要がある。また、室内環境を想定しただけでも、刻々と変化する状況のなかで、服や靴、生活用品などの不特定な物体が認識対象となる。一方で、人や車などの特定では、あらかじめ獲得している2次元形状データや3次元形状データとの適合度を判定する方法がよく用いられており、これらは、本研究とは目的が異なる。

人が動かす物体を認識する方法として、人の操作対象である物体を撮影するための、尾関ら<sup>(1,3)</sup>やTsubukuら<sup>(1,4)</sup>の方法が利用できる可能性がある。これらは、人の作業状況を最適なシーンでとらえることを目的に、人の動きを追い、状況を推定して、人が持つ対象物を自動撮影するものである。これにならうのであれば、人による物体の移動を認識するために、人が物体を取り上げるタイミングを検出し、人とともに移動する物体の追跡し、そして、人が物体を放す時に、物体が置かれたことを知るというアプローチが考えられる。しかし、このためには、撮影領域におけるカメラを適切に配置した上で、人や物体の位置を逐次追跡するための工夫が必要である。事実、尾関らは、変化する人や物体の位置を知るために、磁気センサを併用している。

人が運動しているのであれば、連続した画像で、ある程度以上の変化が続く。また、人が物体を動かす場合にも、人は運動を行う。このため、人の運動が散発的に起こると仮定すれば、連続した画像の変化により人の運動の有無を判定し、人の運動の前後での画像間の差より、移動された物体を認識できると考えられる。このときには、物体は移動するが、それを移動中に追跡することはしない。その代わりに、人の運動の

開始と終了を推定し、その前後での画像の変化を調べることで、物体の移動を認識する。つまり、手で物体を持つことや放すことを陽に特定することなく、逐次撮影する画像間の変化の大きさから人の動きを判定し、移動前後における画像中の物体の位置を特定する、第2のアプローチである。

本研究では後者のアプローチをとる。そして、移動物体として文具や本、鞆などを想定したデスクワーク環境において、人が、物体を取り、移動させ、放置する時の、物体の位置の認識とその履歴の記録について考える。

以下では、人が物体を取り、移動させ、置くという動きは、人の比較的大きな動作により行われ、この動きは、ビデオカメラの画像では、比較的大きな変化として現れるとする。また、この画像の変化の始まりと終りを、それぞれ移動開始イベント、移動終了イベントと呼び、これらには含まれた時点を移動イベントとする。

## 3. 移置物体の認識

**3.1 概要** 移動開始、終了イベントの検出には、ビデオカメラが得るフレームの間に現れる差分領域に注目する。移動終了イベントが得られたならば、直前の移動開始イベントより前に取り込んだ画像を背景画像として、移動終了イベント後の画像との差分により、実際に移動した対象を取り出す。移動した対象には、画像中の人も含まれる可能性があるため、この人の領域を分離して、移置物体が存在する領域を特定する。物体の移動が観測された場合には、その時点の画像が、物体が移動した後の新しい状況をとらえているとして、背景画像を更新する。この過程を繰り返して、人が移動、放置する物体を認識する。

**3.2 移動イベントの検出** ビデオカメラから定期的にサンプリングされるフレームにおいて、連続するフレーム間で一定面積以上の差分領域が現れ始めた時を、移動開始イベントとし、その後の継続したフレームで差分領域の面積が小さくなった時を、移動終了イベントがおきたと判断する。これらは、人が物体を動かす前後のタイミングの候補となる。

移動イベント検出のためのフレーム間差分や、次節で述べる背景差分では、色相やRGBの色情報を用いた結果と、輝度情報を用いた結果に、大きな差はみられなかった。このため、処理の効率化も考慮して、カラー入力画像から算出した輝度値に基づき、差分処理を行う。フレームのサンプリング間隔を単位時間とした、時刻 $t$ における2次元画像座標 $(i, j)$ の輝度値を

$I(t, i, j)$  とする. そして, 隣接する直前の離散時刻  $t-1$  と  $t$  における輝度値を用いて,

$$g(t, i, j) = \begin{cases} 1 & |I(t, i, j) - I(t-1, i, j)| > \Delta I_g \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

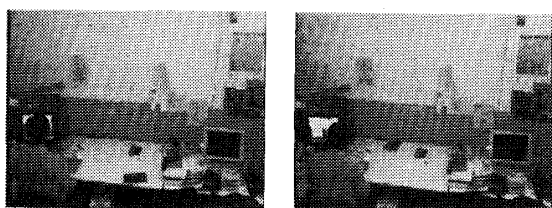
で与えられる 2 値画像を, 時刻  $t$  における 2 フレーム間の差分画像とし, 値が 1 の領域を差分領域とする. ここで,  $\Delta I_g$  は, 輝度値の差の有無を決定する閾値である.

ビデオカメラからの画像を用いると, 偶発的なノイズがフレーム間差分画像に表れることがある. このノイズの影響を抑えるために, 時刻  $t$  における移動イベントの検出には, 時刻  $t-2$  と  $t-1$  の画像から得られる  $g(t-1, i, j)$  と, 時刻  $t-1$  と  $t$  の画像から得られる  $g(t, i, j)$  の論理積として定義される, 時刻  $t$  における 3 フレーム間の差分画像を用いる. そして, 3 フレーム間の差分画像において, 値が 1 となり差が表れた領域の総ピクセル数が,  $S_{ms}$  以上になった時刻を, 移動開始イベント発生時とし, 移動開始イベント発生時以降に初めて  $S_{me}$  以下になった時刻を, 移動終了イベントが起きた時点とする.

**3.3 移置物体候補の検出** 移動開始イベント以前の画像を背景画像として, 移動終了イベント時の画像との背景差分を取ることで, 移動イベント前後の変化領域を得る. 画像座標  $(i, j)$  における背景画像の輝度値  $B_s(i, j)$  と移動終了イベント時  $t$  の輝度値  $I(t, i, j)$  より, 閾値  $\Delta I_h$  を用いた

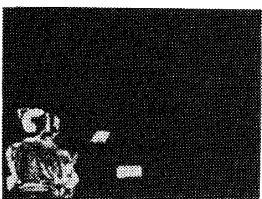
$$h(t, i, j) = \begin{cases} 1 & |B_s(i, j) - I(t, i, j)| > \Delta I_h \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

において, 値が 1 となる画素の 8 方向連結領域を, 背



(a) Background image.

(b) Input image.



(c) Background subtraction image.

Fig. 1 Background subtraction.

景差分領域として抽出する. ここで, 差分領域の面積が  $S_{min}$  未満の場合は, その領域はノイズであるとして除去する.

図 1 に背景差分の例を示す. 画像中の左端でパソコンの操作をしている人が, 図 1(a)では机上手前にある本を, 図 1(b)では奥方向に動かしている. この移動イベントによる背景差分の結果には, 移動前と移動後の本の領域が同時に現れており, これらを区別することはできない. また, 移置物体のほかにも, 左端で小刻みに動く人の領域が現れている. これらの点について, 次節以降で解決する.

**3.4 移動前後の物体位置の特定** 移動イベントの前後で検出される移置物体領域の, 時間的關係を判定するために, 移動前の領域の境界付近と, 移動後の領域の境界付近の画像情報の違いに注目する. つまり, 移動前の領域では, 物体が取り去られていることより, 領域境界の内側と外側において輝度値の差が小さいと想定される一方で, 移動後の領域は, 物体が新たに置かれたことにより, まわりとは異なる輝度値を持つであろうということを利用する.

ある背景差分領域に含まれる最も外側の画素を, 8 方向の連結成分としてつなげた画素列を内側境界とし, この内側境界に隣接する背景領域外側の画素列を, 外側境界とする. また, 内側境界に沿って測った境界の長さを  $L$ , 内側境界上の任意の位置  $k$  の画素と, この画素に隣接する外側境界上の画素の輝度値を, それぞれ  $E_{in}(k)$ ,  $E_{out}(k)$  とする. この時, フラグ  $e(k)$

$$e(k) = \begin{cases} 1 & |E_{in}(k) - E_{out}(k)| > \Delta I_e \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$(k = 0, \dots, L-1)$

の値が 1 となる割合が,  $k = 0, \dots, L-1$  に対して  $r$  以上になる場合は, その差分領域は移置物体が移された領域とし, その他の場合は, 物体が取り去られた領域と判定する.

**3.5 人の領域の分離** 背景差分領域が, 移置物体により現れるものであれば, 移動後すぐに物体を再び移動させない限り, この差分領域は変化しない. これに対して, 人は, たとえ物体を動かしてなくても少なからず動作をしており, 時間経過とともに, 人が占める差分領域は変化する. 移動終了イベントを検出したときの画像を, 取得画像と呼ぶことにする. 移置物体の領域と人の領域を区別するために, ある移動終了イベントにおける取得画像に加えて, その後の連続した複数回の移動終了イベント時での取得画像を取り

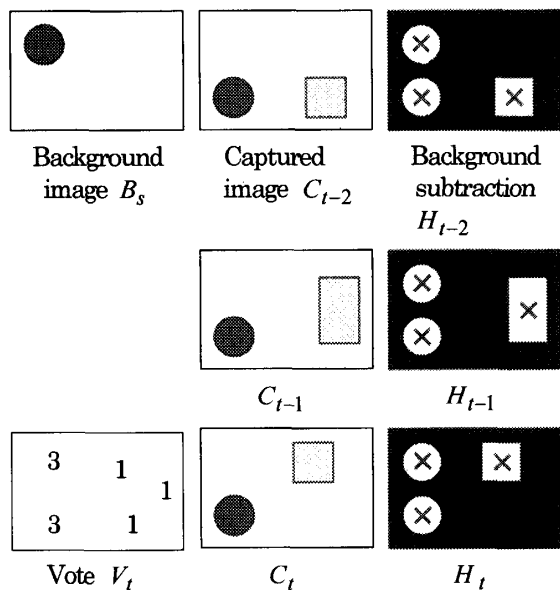
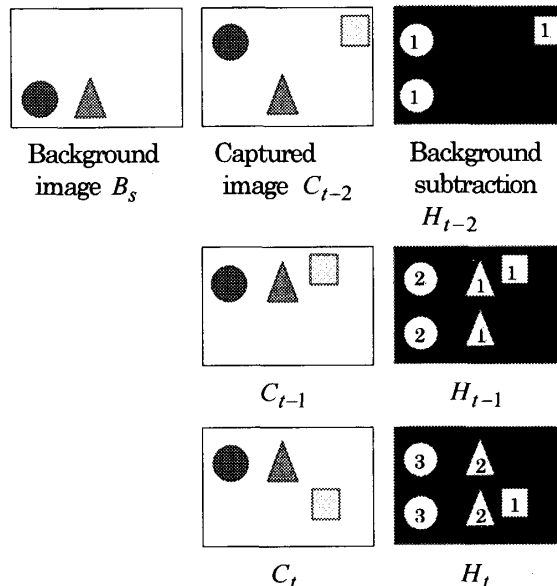


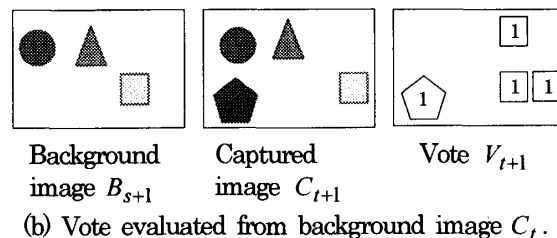
Fig. 2 Separation of human regions and object regions ( $N = 3$ ).

込み、これらの画像と背景画像との差分をとり、差分領域の位置の変化を調べ、人の領域を分離する。図2に、この処理の概要を示す。図中、丸が移置物体であり、四角形が常に動く人を表す。時刻  $t$  の移動終了イベントからさかのぼり、 $N$  回の連続した移動終了イベント時における、 $C_{t-N+1}$  から  $C_t$  の  $N$  枚の取得画像を保存する。ここで、 $C_t$  は、時刻  $t$  における最新の取得画像である。図2の例では  $N=3$  であり、取得画像は、3回の移動終了イベント時における  $C_{t-2}$  から  $C_t$  である。これらの取得画像と背景画像  $B_s$  との背景差分画像  $H_{t-2}$ 、 $H_{t-1}$ 、 $H_t$  を求め、差分領域中心の画像座標を決定する。一方、各画像座標にはカウンタを用意し、初期の全領域内あるいは認識直後の移置物体領域内のカウンタの値を0としておく。そして、差分領域中心の画像座標が決定されるたびに、そのカウンタ値を1つ増やす。この操作を投票とよぶことにする。図2では、投票される画像座標を×印で示す。四角形で表した人の領域には、取得画像ごとの動きがあり、領域の中心座標は常に異なる一方で、移置物体の領域では、連続して同じ座標に投票が行われる。投票状況  $V_t$  は、3枚の取得画像による投票の結果である。この投票状況を調べ、移動イベント終了後に、連続した投票数が  $N$  となる領域を、移置物体の領域と判定する。図2では、投票数が3に達した領域が、移置物体の領域となる。

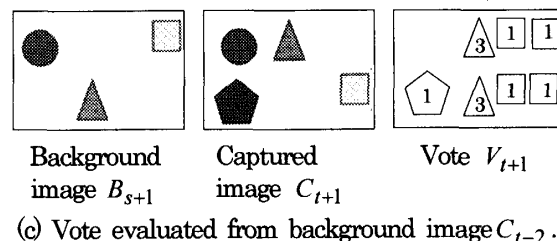
**3-6 背景画像の更新** 移動終了イベントを検出し、その移動イベントに伴う人の領域の分離、移置物



(a) Background subtraction of images including sequential movements of objects.



(b) Vote evaluated from background image  $C_t$ .



(c) Vote evaluated from background image  $C_{t-2}$ .

Fig. 3 Update of background image ( $N = 3$ ).

体領域の特定を行った後に、新しい移置物体を認識するために、背景画像を更新する。今、連続投票数が  $N$  になり、移置物体が特定されたときに、連続投票数が  $N$  に達していない移置物体の領域も存在しているとしよう。

$N$  が3の場合の、この状況の例を、図3に示す。図中、丸、三角形、五角形が物体の領域を、四角形が人の領域を表す。丸い物体を動かしたという移動イベントに続けて、三角形の物体を動かす移動イベントを発生させた。背景差分画像に記した数字は、差分領域の中心座標への投票数を示す。取得画像  $C_t$  の段階で、丸い物体の領域が規定の投票数  $N$  の3を獲得し、移置物体として認識され、また、三角形の物体が占める領域は、投票数2を得ている。ここで、移置物体検出

後の新たな背景画像  $B_{s+1}$  に取得画像  $C_t$  を割り付け、次の取得画像  $C_{t+1}$  には、五角形の物体が現れたとしよう。すると、連続した取得画像  $C_{t-1}$ ,  $C_t$ ,  $C_{t+1}$  に対して、図 3 (b) に示す投票結果が得られる。この結果では、取得画像  $C_t$  の段階まででは投票数 2 を得ていた三角形の領域が現れていない。このため、移置物体として認識された四角形の物体の移動イベントの後に移動を開始した三角形の物体が、認識から取りこぼされる。

この取りこぼし現象を排除するために、取得画像  $C_t$  において移置物体として認識された物体の領域が、最初に現れた取得画像を、新しい背景画像に設定する。つまり、図 3 では、四角形の物体の差分領域が現れた  $C_{t-2}$  を、背景画像  $B_{s+1}$  とする。取得画像  $C_{t-1}$ ,  $C_t$ ,  $C_{t+1}$  を用いた、このときの投票状況を、図 3 (c) に示す。三角形の物体は連続規定投票数  $N$  を獲得しており、移置物体として正しく認識される。

移動終了イベントが検出されない限り、取得画像は得られず、背景画像の更新も行われない。このような場合でも、画像外から物体が運び込まれるなどの、小さな物体の移動が存在するかもしれない。このため、移動イベントが検出されない場合にも、ビデオカメラが定期的にとらえるフレーム数が、直前の背景画像更新後から数えて閾値  $T_{nm}$  に達したときには、閾値に達した時点で移動終了イベントが検出されたとして、移動物体の特定を試みるとともに、その時点のフレームから取得画像を取り込み、この取得画像により背景画像を更新する。この定期的な更新は、小さな物体の移動の認識に加えて、照明条件が時間的に多少変化する環境においても、安定した認識をおこなうためにも有効であると考えられる。

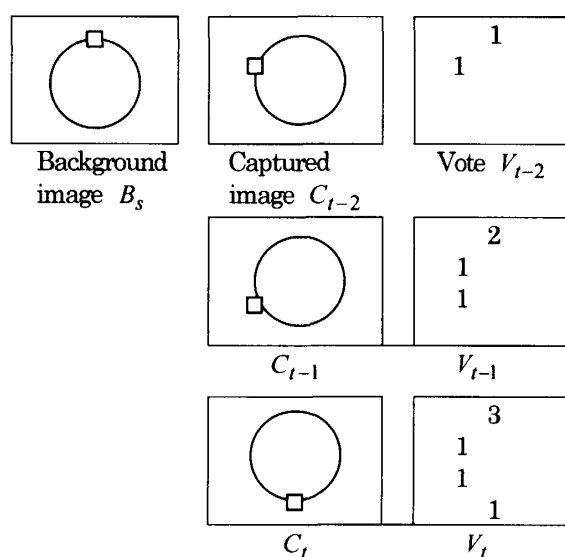
**3.7 連続的に移動する物体の特定** 本研究では、比較的大きな人の動きにより物体が移動するとしている。このため、人が動かすのではなく、自律的に移動し続ける小さな物体がある場合には、その動きは、3.2 節で述べた移動イベントとはされないが、他の移動イベントが検出されたときには、取得画像で背景差分領域を与える。

この状況を図 4 に示す。ここで、画像中には、人や移置物体も現れる可能性があるが、この図では、四角の物体が、反時計回りに円運動をしている部分のみを示している。また、移置物体検出のための投票数  $N$  は 3 とする。図 4 (a) は、背景画像  $B_s$ , 取得画像  $C_{t-2}$ ,  $C_{t-1}$ ,  $C_t$  における、物体領域への投票状況である。

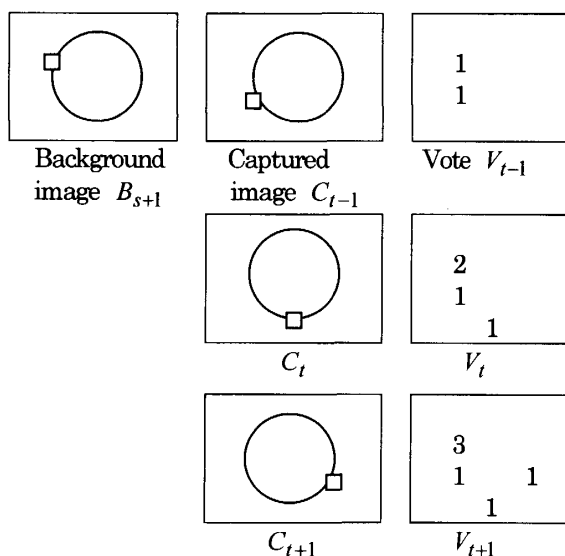
背景画像中で、円運動の頂点にある領域に対する投票数が規定数  $N$  に達し、この頂点に存在した物体が移

動したと判断され、取得画像  $C_{t-2}$  が新しい背景画像  $B_{s+1}$  となる。この背景画像の変更後は、図 4 (b) に示すように、新しい取得画像  $C_{t+1}$  を加えて、改めて投票が行われる。すると、更新した背景画像に現れている左上の領域が規定投票数に達し、この領域に対応する物体が、移動した物体と判定される。以降、物体が自律的に移動し続ける限り、背景画像に現れる物体の領域が、移動した物体の領域として認識され続ける。しかし、これらは、人が意図的に動かした物体ではないことより、人が動かした物体とは区別する。

図 3 において、人が移動させた後、移動後の場所に存在し続ける移置物体としての三角形の物体は、図 3 (c) では、背景画像を  $B_{s+1}$  のもとで規定投票数  $N$  を獲



(a) Recognition of an object starting to move.



(b) Recognition of a continuously moving object.

Fig. 4 Recognition of an autonomously moving object.

得しており、また、その一つ前の移動終了イベントの時、一つ前の背景画像  $B_s$  を用いる図 3 (a) では、 $N-1$  票を得ている。一方で、図 4 (b) で  $N$  票を獲得し、移動した物体のものであると判断された領域は、一つ前の移動終了イベント時の図 4 (a) では、1 票を獲得しているにすぎない。このため、一つ前の背景画像を用いた投票における得票数が  $N-1$  で、最新の背景画像を用いた得票数が  $N$  となる領域を持つ物体を移置物体とし、一つ前の背景画像を用いた得票数が  $N-1$  未満の場合には、移置物体とみなさないこととする。

ビデオカメラの起動時には、起動直後に取り込んだ最初のフレームから背景画像を得るが、この画像より時間的に前の背景画像は存在しない。このため、起動時には、一つ前の背景画像を用いた投票数は、全領域で 0 とする。これにより、起動時に既に動いている物体は、移置物体としては認識されない。また、起動後に移動し始め、動き続ける物体は、それが最初に存在した領域から取り去れた物体として認識される。

**3.8 移置物体の履歴の記録** 移置物体は、その固有番号、移動後に占める領域の中心の画像座標、色相情報分布、移動が認識された時刻を、移動後の領域を囲む矩形部分のカラー画像とともに記録される。ここで、色相情報は、画素における色情報であり、彩度が 15 を超える場合には、0 から  $2\pi$  の色相を  $\pi/3$  ごとに分けした 6 種類、彩度が 10 未満の場合には、輝度に応じた黒色、灰色、白色の 3 種類、そして、彩度が 10 から 15 の範囲では不定の、計 10 種類である。色相情報分布は、移動後の領域を囲む矩形部分の画素の色相情報に関するヒストグラムである。

本研究では、柔軟な衣服なども認識対象とするため、2 つの抽出矩形領域にある物体の同一性の判定には、変化しやすい形状でなく色相を用いる。つまり、両領域の色相情報のヒストグラムを 10 次元単位ベクトルに正規化し、それらの内積の値が 0.7 以上の場合、色相情報分布は類似し、物体は同一であるとする。

移置物体が認識された場合、移動後に占める矩形領域の色相情報分布を求める。そして、それが、既に記録されている移置物体の色相情報分布と同じであるならば、記録されている物体が動かされたものとして、固有番号継承し、新しい座標、色相情報分布、時刻を記録する。同じ色相情報分布をもつ物体が記録されていないならば、新たな物体が認識されたものとして、新規の固有番号を付けて記録する。これにより、移置物体の移動履歴を作る。作成された移動履歴は、後に、移置物体の位置検出などに利用できる。

## 4 実験

**4.1 実験環境** ビデオカメラは、高さ 2.7m の天井に設置し、斜め下方に向けて、机、床などを含む室内環境を観測する。使用機種はキヤノン(株)の VC-C4R であり、パンチルトを固定し、ズームを引いた状態で撮影した。処理用の計算機の CPU は Pentium 4, 2.4 GHz, メモリは 512 MB, OS は Fedora Core 2 である。画像キャプチャボードは、(株)メルコの CBP-AV (コントローラ, CONEXANT 社製, Bt878KHF) であり、RGB の各色の階調は 256, 画像サイズは  $640 \times 480$  画素に設定する。輝度値も 256 階調とする。

ビデオカメラからのフレームのサンプリングレートは 250 msec とした。予備実験において、このサンプリングレートでは、人の動きによる移動イベントが妥当に検出できることを確認している。移動イベントが検出されない場合に背景画像更新を行うフレーム数の閾値  $T_{nm}$  は 20 とした。これより、画像中に大きな変化がない場合にも、5 秒間に 1 回は背景画像が更新される。他の閾値は、次の 2 組から選ぶ。

閾値 1 :

$$\begin{aligned} \Delta I_g &= 20, S_{ms} = 1100 \text{ pixels}, S_{me} = 100 \text{ pixels}, \\ \Delta I_h &= 20, S_{min} = 20 \text{ pixels}, \\ r &= 0.3, \Delta I_e = 20, N = 4. \end{aligned}$$

閾値 2 :

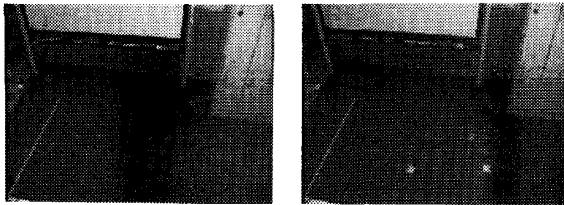
$$\begin{aligned} \Delta I_g &= 20, S_{ms} = 600 \text{ pixels}, S_{me} = 60 \text{ pixels}, \\ \Delta I_h &= 20, S_{min} = 20 \text{ pixels}, \\ r &= 0.3, \Delta I_e = 20, N = 5. \end{aligned}$$

予備実験結果より、ピクセル差数  $S_{ms}$  は、代表的な人の領域の 5% 程度とし、 $S_{me}$  は  $S_{ms}$  の 10% 程度とする。また、取得画像枚数  $N$  が大きいほど人領域が正しく分離できる傾向にあったことより、物体が長時間放置されるような場合は  $N$  の値を増す。以下、実験(1), (2), (3)では閾値 1, 実験(4)では閾値 2 を用いた。

**4.2 実験(1)** 散発的に起こる人の行動をとらえ、ほぼ同時に移動する移置物体が認識できることを確認するために、人が色の異なる 3 個の物体をカメラ視野内に持ち込み、その後、2 個の物体を移動する作業をモデル化する。人の動作を次に示す。

- (1) 青, 赤, 黄の物体を持ち込み, 床上に置く。
- (2) 以下を 3 回繰り返す。
  - (2-1) 青の物体と赤の物体を移動する。
  - (2-2) 赤の物体と黄の物体を移動する。
  - (2-3) 青の物体と黄の物体を移動する。

各動作の前に人は視野外から入り、動作後は視野外に去る。対象物体はフロップディスクである。各動作間隔を 10 秒, 20 秒, 30 秒として実験を行った。全動作間隔で、3 個の物体の持ち込みと移動が認識され、人の一部が移置物体と誤認識されることはなかった。動作間隔 30 秒の実験結果を図 5 に示す。図 5 (a), (b) は初期状態と物体の配置例である。



(a) Initial state.

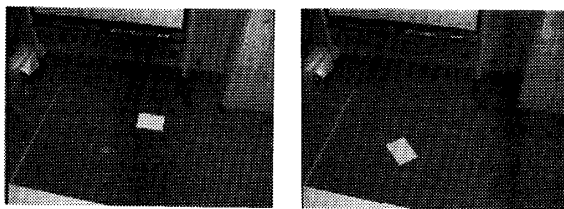
(b) Recognized three objects.

ID	x	y	Time
1	258	303	05:52
2	375	316	05:52
3	273	403	05:52
2	463	284	06:29
3	487	403	06:29
1	180	262	07:00
2	321	432	07:00
3	407	297	07:31
1	239	361	07:31
2	289	242	08:01

3	454	421	08:01
2	154	276	08:30
1	432	282	08:30
3	296	239	09:00
1	253	368	09:00
2	293	331	09:31
3	422	338	09:31
1	340	242	10:00
2	371	398	10:00
3	253	201	10:31
1	244	356	10:31

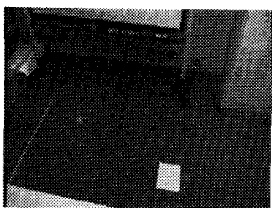
(c) History of object movements.

Fig. 5 Recognition of three objects carried into a camera view.



(a) First arrangement of two objects.

(b) Overlapping of objects.



(c) Rearrangement of objects.

ID	x	y	time
1	343	273	08:51
2	239	340	09:21
1	242	341	09:55
1	390	402	10:25
2	181	266	10:52

(d) History of object movements.

Fig. 6 Recognition of overlapping objects.

図 5 (c) は物体の認識履歴であり、認識物体の固有番号、物体が占める矩形領域の中心の画像座標と、認識時刻を示す。動作(1)で持ち込んだ青、赤、黄の物体には、それぞれ固有番号 3, 2, 1 が付けられた。その後、動作(2-1)では固有番号 2, 3 である赤と青の物体の移動などと、順次、移置物体が認識された。

4.3 実験(2)

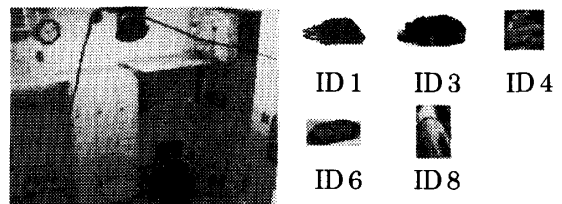
複数の物体が持ち込まれ、持ち込まれた物体が重なり合うように動かされる場合を考える。用いる物体は、物体 A と、その輪郭を覆い隠すような大きさを持つ物体 B の 2 個である。人の動作は次の通りである。

- (1) 物体 B を持ち込み、床上に置く。
- (2) 物体 A を持ち込み、床上に置く。
- (3) 物体 B を、物体 A を覆い隠すように移動する。
- (4) 物体 B を、物体 A とは別の場所に移動する。
- (5) 物体 A を、物体 B とは別の場所に移動する。

各動作には、ほぼ 30 秒の間隔をおき、各動作において、人はカメラ視野内に入り、動作を行い、その後、視野から去る。図 6 に、この実験状況を示す。図 6 (a), (b), (c) は、それぞれ動作(2)の後、(3)の後、(5)の後の画像である。図 6 (d) は物体の移動履歴を示す。フロップディスクである物体 A が固有番号 1, A4 ファイルである物体 B が固有番号 2 とされている。人の動きに対応して、移動履歴には、5 回の位置の変化が記録されている。

4.4 実験(3)

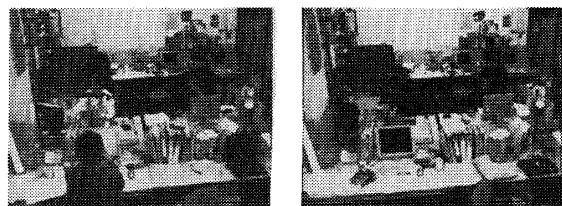
電車内などにおける、放置物体や



(a) Experimental environment.

(b) Recognized objects.

Fig. 7 Recognition of moved objects in a train-like situation.



(a) Initial state.

(b) Scene after 30 min.

Fig. 8 Recognition of moved objects in a room.

忘れ物を想定し、横に並べた肘掛け椅子とスチール棚で、ベンチシートおよび網棚を模して、実験を行う。一駅間の通過時間を1分とし、最初の三駅で、一人ずつ、各自の鞆を持ち込み、網棚の上に置く。その後、三駅で、一人ずつ下車するが、一人は鞆を置き忘れる。図7に、実験環境と、移置物体として認識された5つの物体を示す。図7(a)の上縁中央に、撮影用のビデオカメラがある。認識された移置物体のうち、固有番号が1, 3, 6の物体は、人が持ち込んだ鞆である。これより、持ち込まれた物体、および、持ち込まれた後で持ち去られる物体は、もれなく認識されていることになる。固有番号4, 5は、身体の一部の手である。駅間の通過中に、人の頭や胴体には多少の動きが現れるが、膝や座席の上にある手は、ほとんど動かない場合がある。このような身体の一部が、鞆と同様に、移置物体と認識されている。

**4.5 実験(4)** 実験環境として、デスクトップパソコンが置かれた研究室において、30分間にわたり、物体の移動認識を行った。図8に、実験開始時と実験終了時の画像を示す。この実験の間に、部屋への人の出入りがあり、また、ビデオカメラ近くを人が横切ることもあった。実験終了時に、13個の移置物体が認識された。そらは、コップ、カップめん、財布、2回移動したと認識された携帯電話、そして、手と頭部という身体の一部が8個であった。実験③と同様に、身体の一部が誤認識される可能性がある。実験後に、この実験環境に登場した人物に対して、意図的に動かしたものについてアンケートを行ったところ、携帯電話、コップ、カップめん、財布が具体的に挙げられた。これより、人が動かしたものは、ビデオカメラの視野内にとらえられている限り、移置物体として適切に認識できることが確認できる。

## 5. 結 論

ビデオカメラからの画像に基づき、人が移動させた物体を、人の動きの開始と終了を検出するフレーム間差分と、移動した物体を特定する背景差分を併用することにより認識する方法を提案した。人の行動による物体の移動と、常にわずかながらでも発生している人の動作、自律的に動く物体の運動を区別して、移動した物体を特定する。模擬的な環境および実環境で実験を行った。その結果、動きがほとんどない身体の一部が、動かされた物体として誤認識される場合があるが、人が実際に動かした物体は、妥当に認識できることを示した。

提案した方法は、逐次、高度な画像処理をすることなく、比較的簡単な処理で実現できるとともに、カメラの内部パラメータ、外部パラメータの校正も必要としない。これは、星野、益田<sup>(1)</sup>が述べている、監視システムで求められる要求とも合致している。今後の課題には、身体の一部に対する誤認識の排除や、複数台のカメラにより死角を抑えた観測環境の導入がある。

本研究の遂行にあたり協力頂いた、名古屋大学情報文化学部の元学生である北島利夫氏に、謝意を表す。

## 文 献

- (1) Shirai, Y., Miura, J., *IPSJ Transactions on Computer Vision and Image Media*, Vol. 43, No. SIG 4 (CVIM 4) (2002), pp. 33-41.
- (2) Hasegawa, O., Kanade, T., *Transactions of Information Processing Society of Japan*, Vol. 44, No. 7 (2003), pp. 1795-1807.
- (3) Satake, J., Shakunaga, T., *Transactions of the Institute of Electronics, Information and Communication Engineers D-II*, Vol. J86-D-II, No. 8 (2003), pp. 1212-1221.
- (4) Kastrinaki, V., Zervakis, M., Kalaitzakis, K., *Image and Vision Computing*, Vol. 21 (2003), pp. 359-381.
- (5) Liu, T.-L., Chen, H.T., *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 3 (2004), pp. 397-402.
- (6) Gelgon, M., et al., *Image and Vision Computing*, Vol. 23 (2005), pp. 19-31.
- (7) Sakaida, S., et al., *Transactions of the Institute of Electronics, Information and Communication Engineers D-II*, Vol. J84-D-II, No. 12 (2001), pp. 2541-2555.
- (8) Hashimoto, M., et al., *Transactions of the Japan Society of Mechanical Engineers, Series C*, 70-696 (2004), pp. 2344-2351.
- (9) Onishi, M., et al., *Transactions of the Institute of Electronics, Information and Communication Engineers D-II*, Vol. J85-D-II, No. 4 (2002), pp. 594-603.
- (10) Aoki, S., et al., *Transactions of the Institute of Electronics, Information and Communication Engineers D-II*, Vol. J87-D-II, No. 5 (2004), pp. 1083-1093.
- (11) Namri, T., Otsu, N., *IPSJ SIG Notes*, CVIM-145 (2004), pp. 179-186.
- (12) Hoshino, J., Masuda, I., *Transactions of Information Processing Society of Japan*, Vol. 40, No. 5 (1999), pp. 2367-2376.
- (13) Ozeki, M., et al., *Transactions of the Institute of Electronics, Information and Communication Engineers D-II*, Vol. J86-D-II, No. 11 (2003), pp. 1606-1617.
- (14) Tsubuku, Y., et al., *Lecture Notes in Computer Science* 3332 (2004), eds, Aizawa, K., et al., pp. 113-120.