

屋外映像解析における物体間の関係を利用したイベントの記述

田代, 直之
九州大学システム情報科学研究院知能システム学部門

島田, 敬士
九州大学システム情報科学研究院知能システム学部門

有田, 大作
九州大学システム情報科学研究院知能システム学部門

谷口, 倫一郎
九州大学システム情報科学研究院知能システム学部門

<https://hdl.handle.net/2324/5958>

出版情報 : 火の国情報シンポジウム, 2007-03
バージョン :
権利関係 :

屋外映像解析における物体間の関係を利用したイベントの記述

田代直之† 島田敬士†, 有田大作‡, 谷口 倫一郎*

九州大学大学院システム情報科学府†

九州大学大学院システム情報科学府*

〒 819-0395 福岡県福岡市西区元岡 744 番地

財団法人九州システム情報技術研究所‡

〒 814-0001 福岡県福岡市早良区百道浜 2 丁目 1-22

E-mail: {tashiro,atsushi,rin}@limu.is.kyushu-u.ac.jp, arita@isit.or.jp

あらまし 屋外設置のカメラ台数が増え、セキュリティのための監視は人の目視による監視からシステムによる監視に移行し始めている。システムによる監視に重要なことは、映像解析によって映像内にどのような物体が出現し、どのようなイベントが発生したかを人に適切に提示することである。しかし、発生するイベントすべてをあらかじめ定義しておくことは困難である。本稿では、未知のイベントが繰り返し観測されたときに、新たな物体やイベントとして検出し、既知物体との物体間の関係を利用して、未知のイベントを記述する手法を提案する。これを繰り返すことにより、映像内で出現した物体や発生したイベントを定義することができる。

キーワード 屋外監視, 物体間関係, 未知イベント記述

Event Detection Using Interaction of Multiple Objects in the Outdoor Scene Analysis

Naoyuki Tashiro†, Atsushi Shimada†, Daisaku Arita‡, Rin-ichiro Taniguchi*

Department of Intelligent Systems, Kyushu University†*

744, Motooka, Nishi-ku, Fukuoka, 819-0395, Japan

Institute of System & Information Technologies/KYUSHU‡

2-1-22, Momochihama, Sawara-ku, Fukuoka, 814-0001, Japan

E-mail: {tashiro,atsushi,rin}@limu.is.kyushu-u.ac.jp, arita@isit.or.jp

Abstract Recently, the number of cameras which put in the outdoor have increased, and automatic surveillance by computerized systems becomes popular. It is the most important things to inform exactly what happened in the scene to the user of a surveillance system. However, it is very difficult to predict events happened in the scene in advance. In this paper, we propose an approach to describe unknown events using interaction of multiple objects, when unknown events happened frequently in the scene.

Key words Video Surveillance, Intraction of Multiple Objects, Description of Event

1. はじめに

屋外に設置した多数のカメラを用いて、撮影映像内で発生するイベントを監視する映像監視システムが多数導入されている。しかし、カメラの設置台数が増えれば、監視映像の数も増え、監視員にとって大きな負担となる。そこで、カメラから得られた映像を自動で解析し、発生したイベントがどのようなものであったかを人に知らせるようなシステムが必要である。しかし、発生するイベントすべてをあらかじめ分析し、そのイベントが発生する条件などを抽出しておくことは困難である。そこで、検出するイベントを選択し、そのイベントが発生する条件などをあらかじめ分析しておき、映像内に検出対象イベントが現れたときに通知するシステムがある。

Fengらは、物を置く動作を行ったときの人のシルエット形状を学習させ、観測されたシルエット形状の時系列情報を利用し、人の行動解析を行っている [1]。人が物を置くという行動を検出することで、不審物の検出を試みている。しかし、物体間の関係を静止物体と人と限定し、さらに物体間の距離しか考慮していないという問題がある。

藤吉らは、物体単体の状態を推定するだけでなく、物体間の距離を利用することで、物体間の関係を定義している [2]。抽出された人と車間の関係を推定することで、車の乗り降りや人の合流の検出を行っている。しかし、物体単体の状態として、出現、消失、移動、停止という基本的な動作に対して行っているため、複雑なイベントを検出することができないという問題がある。

これらの手法では、検出するイベントを増やすたびに、そのイベントの発生条件を人が定義しなければならない。そこで我々は、未知の物体やイベントが繰り返し観測されたときに、既知の物体との関係を利用して、未知のイベントを定義する手法を提案する。これを繰り返すことにより、映像内で出現した物体や発生したイベントを定義することができる。

2. システム構成

提案するシステムは、大きく分けて三段階の処理を行う。第一に、物体単体の属性、物体間の関係属性を抽出する処理を行う。第二に、抽出さ

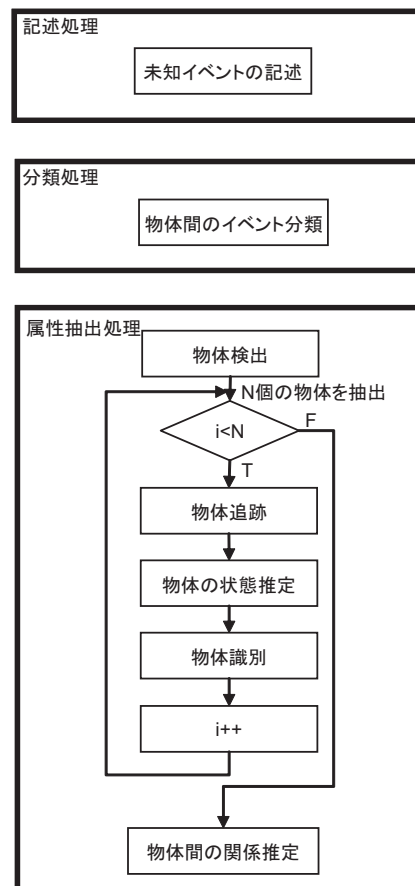


図 1: システムの構成

れた物体間の関係を利用して、未知物体を細分化する処理を行う。第三に、未知のイベントを記述する処理を行う。図 1 に提案するシステムの構成を示す。

2.1 属性のデータ構造

属性には、物体単体の属性と物体間の関係属性がある (図 2)。単体物体の属性には、物体の種類や出現・消失時刻、さらに物体の特徴量がある。物体の特徴量には、物体の位置、速度、形状があり、毎フレーム記録されている。

次に、物体間の関係属性には、関係の開始・終了時刻や関係の種類、関係があった物体へのポインタ、物体間の距離がある。ここで、物体間の距離は、毎フレーム記録されている。さらに、物体間の関係属性と物体単体の属性との間にリンクを張っておく。

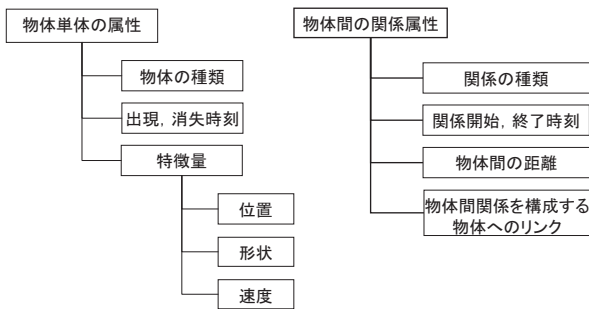


図 2: 属性のデータ構造

2.2 物体間の関係とイベントの定義

2.1 節で述べたように、物体単体は形状や速度といった特徴量を保持している。物体間の関係 R は、それらの物体がどのような係わり合いを持っているかを表し、ここでは、画像平面上の二つの物体の距離に基づいて定義する。一方、イベントは物体の識別結果と物体の関係によって定義される。例えば、「人が車から降りる」というイベントは、(人, 車, 分離) というように定義される。

3. 物体単体および物体間の属性抽出処理

2 節でも述べたように、属性抽出処理では、物体単体の属性、物体間の関係属性、イベントの抽出を行う。

3.1 物体検出および物体追跡処理

本システムでは、屋外環境下における物体の検出を想定しているため、天候による照明条件の変化などに対しても頑健な物体検出手法が必要となる。そこで、混合ガウス分布による動的な背景モデルを利用した背景差分を用いる [3]。この手法では、画素毎にガウス分布を持たせることで、背景モデルを構築している。また、背景の変化に応じて、画素毎にガウス分布の数を増減させることで柔軟かつ高速にモデルの更新を行うことが可能である。

次に、物体の状態を推定するため、抽出された物体の追跡処理を行う。本手法では、パーティクルフィルタを用いた物体追跡を利用している。パーティクルフィルタは状態の確率密度分布を多数の仮説の集合により表現する手法である。それぞれの仮説は固有の重みを持っているので、任

意の確率密度分布を近似することが可能である。

パーティクルフィルタを用いた追跡手法を用いることで、一時的な隠れなどに対しても安定した追跡が行える。パーティクルフィルタには、次状態を予測するための遷移モデルや仮説の重みを決定する評価基準が必要である。

本手法では物体の運動を等速直線運動と仮定し、物体の位置を予測している。また、仮説の重みを決定する評価基準には、追跡対象の色分布を利用している。

3.2 物体単体の状態推定

追跡処理によって得られた物体の推定位置の変化を用いて、物体単体の状態を以下のように定義する。

- 出現

物体検出により、時刻 t で物体が抽出され、その物体が追跡対象として未登録のとき、出現したと定義する。
- 移動と停止

物体を一定時間監視した結果、物体の速度が閾値より大きいときは移動、閾値未満のときは停止と定義する。
- 消失

時刻 $t-1$ まで追跡対象として登録されていた物体が時刻 t において観測されなくなった状態と定義する。

ここで、一定時間物体が停止の状態にあるとき、追跡対象が背景モデルとして登録され、前景として抽出されないという問題が生じる。そこで、追跡対象がある領域の画素に対しては、背景モデルの更新を行わないというフィードバックを行うことによって、追跡対象が停止しても物体が検出できるようにする。

3.3 物体識別

抽出された単体物体を既知の物体と未知の物体に分類する手法について述べる。ここで、既知の物体とは、物体の種類が分かっている物体を表し、未知の物体は、既知の物体として認識されない物体の総称を表す。本システムでは、屋外映像でよく観測される人と乗用車を既知の物体として選出した。

本節では、既知物体の形状などを学習させておいた識別器を用いて、追跡した物体を{人, 乗用車, 未知の物体}に分類する手法を述べる。識別器には、階層型ニューラルネットワークの一つである3層パーセプトロンを用いる。パーセプトロンには入力層, 中間層, 出力層があり, それぞれのニューロンはある入力に対して, 0から1の連続値を出力する。ここで, 入力層には入力する特徴ベクトルの次元数と同じ数のニューロンを, 出力層には分類するクラス数に応じたニューロン数を用意する。ニューラルネットワークの学習には, 教師あり学習のひとつである誤差逆伝播法を用いる。

物体識別は, 1フレームで観測された形状に対して処理を行うため, 一時的なオクルージョンなどの影響で形状を完全に抽出できなかったとき, 誤識別が起こるといった問題があった。そこで, 人と乗用車に対してそれぞれ信頼度 C_{human} と C_{car} を定義することで, 複数フレームを利用した物体識別を行う。

信頼度 C は, 式(1)によって算出される。ただし, k は ($0 < k < 1$) 定数とする。

$$C = k^d \quad (1)$$

ここで, i 番目の出力層のニューロンの出力値を $O[i]$, 教師信号を $T[i]$ とし, 出力差 d は式(2)のように定義する。

$$d = \sqrt{\sum_i (O[i] - T[i])^2} \quad (2)$$

$0.0 \leq d \leq 2.0$ の連続値を出力し, d が 0.0 に近いほど教師信号に近い特徴量を持った物体であることを意味する。

各フレームで人と乗用車に対する信頼度を算出し, 一定フレーム監視したときの平均の信頼度を利用して, 物体の分類を行う。人と乗用車に対する信頼度が低いものを未知の物体とする。

3.4 物体間の関係推定

物体間の関係推定では, 物体間の関係属性の関係の種類を推定する。関係の種類 R を {分離, 結合, その他} の3つに分類する。

本システムでは, 物体間の距離 D が閾値 D_{TH} 以下になったとき, それらの物体には関係があると判断して, 関係推定処理を行う。

ある時刻 t において, 物体検出により得られた前景領域の数を $B_n(t)$, 検出した前景領域を $B_{\text{rect}}^{(i)}(t)$, 追跡対象の数を $O_n(t)$, 追跡対象を $O^{(k)}(t)$, 状態を $O_s^{(k)}(t)$ と表記する。ここで, $0 \leq i, j \leq B_n(t)$, $0 \leq k, l \leq O_n(t)$ である。

物体間の関係推定ルールを以下に示す。

Rule1: $O^{(k)}(t-1) \in B_{\text{rect}}^{(i)}(t-1)$, $O^{(l)}(t-1) \in B_{\text{rect}}^{(j)}(t-1)$ が成り立ち, $O^{(k)}(t)$, $O^{(l)}(t) \in B_{\text{rect}}^{(i)}(t)$ のとき, $O^{(k)}(t)$ と $O^{(l)}(t)$ は結合と判断する。

Rule2: $O^{(k)}(t)$ と $O^{(l)}(t)$ の間の距離 D が十分に小さく, $O_s^{(k)}(t)$ が出現のとき, $O^{(l)}(t-1)$ から $O^{(k)}(t)$ が分離したと判断し, $O_s^{(l)}(t)$ が出現のとき, $O^{(k)}(t-1)$ から $O^{(l)}(t)$ が分離したと判断する。ただし, $O_s^{(k)}(t)$, $O_s^{(l)}(t)$ が共に出現のときは, 両方とも新規追跡対象として登録する。

Rule3: $O^{(k)}(t)$ と $O^{(l)}(t)$ 間の距離 D が閾値 D_{TH} 以下になると関係があると判断される。 $O^{(k)}(t)$ と $O^{(l)}(t)$ が結合も分離せずに, 関係が終了したときその他と判断する。

4. 未知物体の細分化処理

未知物体の中には多種類の物体が含まれていると考えられる。未知物体を分類する一つの手法として, 物体の形状や速度を利用したクラスタリングが考えられる。しかし, 形状や速度が酷似した物体は同じ物として扱われ, それ以上詳細な分類は困難である。一例として, 物体検出によって得られたバスとトラックの形状および速度は似ているため, 同じクラスに分類されてしまう。そこで, 物体単体の特徴だけでなく物体間の関係, 関係が発生した位置を利用してクラスタリングを行うことで, 未知物体の細分化を行う。

未知物体の細分化を行うための特徴量を以下に示す。

- 関係が発生した位置 x, y
分離, 結合関係が発生した画像中の位置を用いる。
- 関係をなす物体の形状 S_A, S_B
- 関係をなす物体の速度 V_A, V_B

ここで、結合と分離はそれぞれの異なる関係空間を用意し、細分化のための特徴量を各々の関係空間にマッピングする。つまり、結合と分離、それぞれの空間においてクラスタリングを行う。クラスタリングの結果、クラスを構成する要素数が多いもの、つまり発生頻度の高いイベントを提示する。

5. 未知イベントの記述処理

記述処理では物体間のイベントの内、未知のイベントに対して記述を行う。未知のイベントは、関係をなす物体に未知物体が含まれているために、記述を行うことができない。そこで、発生頻度の高い未知のイベント内の未知物体の情報を提示し、人が未知物体の種類を定義する。以下に、未知のイベント記述までの流れを示す。

Step1: 未知イベントから発生頻度の高いイベント $E_n (1 \leq n \leq S)$ を選択する。イベントをなす物体の種類を O_A, O_B とする。

Step2: (O_A, O_B のどちらか一方が未知のとき Step2-a へ
その他のとき Step3 へ。

Step2-a: 未知の物体 C_i の情報を提示し、その情報を基に人の手により未知物体の種類を定義する。

Step2-b: 物体間の関係属性の種類の種類と 2 物体の種類を基に、未知のイベントを記述し、人の手によって未知のイベントを定義する。

Step3: もし、未知イベントが存在するとき Step1 へ。

6. 実験と考察

解像度 640×480 で撮影した 20 分の屋外映像を用いて以下の実験を行った (図 3)。

6.1 既知物体の識別実験

学習サンプルには、人、乗用車、ノイズの 3 種類を与えた。既知物体単体を抽出できるか検証するため、人、乗用車が単体で抽出されているシーンに対して、識別実験を行った。



図 3: 実験に用いた映像

映像内で観測された人と乗用車の数は、それぞれ 33 人、14 台であった。観測されたすべての人および乗用車に対して、識別が成功した。

複数フレームを利用した既知物体の識別は、1 フレーム毎の識別に比べ、安定した識別結果を得ることができた。ただし、本実験では既知物体の識別精度の検証を行ったため、未知の物体は認識実験の対象としていない。

6.2 未知イベントの分類実験

検出した未知イベントを 4 節で述べた特徴量を用いて分離することができるか実験を行った。6.1 節で述べたように、既知の物体には人と乗用車を想定する。実験に用いた映像内には、未知の物体として、バス、トラック、バイク、複数の人の塊が出現していた。映像内で発生したイベントを分離、結合それぞれの空間に対して、Nearest Neighbor 法を用いてクラスタリングを行った。

実験映像内で発生したイベントは、22 種類あり、その中で発生頻度が高いものについて適合率、再現率を算出した。その結果を表 1 に示す。ここで、再現率と適合率を式 (3)、式 (4) で定義する。

$$\text{再現率 (\%)} = \frac{\text{正しく分類できたイベント数}}{\text{発生したイベントの回数}} \times 100 \quad (3)$$

$$\text{適合率 (\%)} = \frac{\text{正しく分類できたイベント数}}{\text{同クラス内の要素数}} \times 100 \quad (4)$$

Event	再現率 (%)	適合率 (%)
(人, バス, 分離)	48	72
(複数人, バス, 分離)	42	25
(車, バイク分離)	57	88
(車, バス, 結合)	55	47
(車, バイク, 結合)	58	63
(バス, バイク, 結合)	36	44

表 1: イベントの分類実験結果

さらに、表 1 中の太字は既知物体を表し、複数人とは、物体検出において複数の人間が重なって撮影されたために未知物体に分類されたものである。

6.3 考察

本システムでは、ユーザーにイベントを提示するときに、同じクラス内に異なる種類のイベントが含まれていると、間違った定義づけが行われるので、それを回避するために適合性を高める必要がある。そのため、本実験では、クラスタリングのための距離の閾値を小さくし、他のイベントの競合を回避しようとしたため、実際に分類したいイベントの種類よりも多くのクラスが生成されている。結果、全体的に再現率が低くなっていることがわかる。また、(バス, 複数人, 分離) の適合率が低い理由として、バスから降りた複数人が重なっているために、(人, バス, 分離) のクラスに属していると判定されているものが多かったためと考えられる。

7. おわりに

未知の物体やイベントが繰り返し観測されたときに、未知物体の細分化処理を行い、未知のイベントを記述する手法を提案した。今後の課題は以下のようなものが挙げられる。

・物体間関係項目の追加

物体間の関係に分離、結合のみを利用しては、より高精度の未知物体の細分化処理を行うためには、より詳細な関係を定義、抽出することが必要である。さらに、物体間の関係を二項関係だけでなく、 n 項関係まで考慮したときの物体間の関係を定義する

必要がある。ここで、 $(2 \leq n \leq N(t))$ で、 $N(t)$ は時刻 t における追跡対象物体の数を表す。

・未知物体の情報提示手法

監視員は未知のイベントを構成する未知物体の情報を利用して、未知物体を定義するため、提示する情報およびその提示方法を検討する必要がある。

- ・未知物体の細分化のオンライン化 現在のシステムでは、一定時間映像を撮影し、その間に観測された物体間関係を利用して、未知物体の細分化処理を行っている。また、人と乗用車という既知の物体をあらかじめ定義して行っている。しかし、実際に監視システムとして導入するためには、未知のイベントが頻繁に観測されるたび監視員に提示する必要がある。そのためには、未知物体のオンライン細分化手法が必要となる。

謝辞

本研究は、文部科学省の平成 18 年度科学技術振興調整費による「科学技術連携施策群の効果的・効率的な推進」の一環として実施した。

参考文献

- [1] S.Lu, J.Zhang and D.Feng. "A Knowledge-based Approach for Detecting Unattended Packages in Surveillance Video". IEEE International Conference on Advanced Video and Signal based Surveillance CDROM Proceedings, 2006.
- [2] N.Enomoto, T.Kanade and H.Fujiyoshi. "A method for monitoring activities of multiple objects by using stochastic model". IEICE Transactions on Information and Systems E84-D12, pp.1705-1712,2001.
- [3] 島田敬士, 有田大作, 谷口倫一郎. "混合ガウス分布による動的背景モデルの分布数の増減法". 画像の認識・理解シンポジウム (MIRU2006), pp.746-751, 2006.