Vol.2012-CG-149 No.5
Vol.2012-CVIM-184 No.5
2012/12/3

# 背景特徴量と物体の出現頻度に基づくイベント検出

勝手 美紗[1,a)]　内海 ゆづ子[1,b)]　黄瀬 浩一[1,c)]

**概要**：本稿では，インターネット上の大規模な動画像を自動で分類することを目的とし，動画像中のイベントを検出する．インターネットで配信されているホームビデオなどの動画の多くは視点が固定されずに撮影されているため，動画像から人物や物体の動きの情報を取得することは困難であり，行動認識などの技術を用いてイベントを検出することは困難となる．また，イベントは行われる環境や環境を構成する物体に大きく依存する．そこで，動画像中の背景や動画像中に登場する物体に着目し，イベントの検出を行う．背景からは，Opponent SIFT 特徴量を Bag-of-Feature で表現したものを特徴量として抽出する．物体特徴量には，動画像から物体検出器により検出した物体の頻度と識別器の信頼度の値を用いる．それぞれ特徴量を用いて最近傍探索を行い，結果を統合することでイベントの認識を行った．評価は TRECVID2012 Multimedia Event Detection タスクのデータセットを用いて行った．その結果，特定の環境でのみ行われるイベントと動画中の物体を高い精度で検出できたイベントを検出できた．

## 1. はじめに

近年，You Tube [1] などの動画共有サイトが増加し，大規模な映像コンテンツが蓄積されつつある．大量に蓄積された動画から，ユーザが必要とする動画像を検索するには，一つ一つの動画像に対して関連するテキストデータなどのメタデータを付与する必要がある．各動画に適したメタデータの付与は，人手に頼ることとなるが，大量な動画像を人手で管理することは困難である．そこで，動画像に関連するメタデータを自動で抽出し，大量の動画像を分類できるシステムがあれば有益である．その際に，管理者の要望に合った分類に動画像を分けることが望ましいと考えられる．こういった管理者の要望を満たした大量の動画像を分類するための手法の一つとして，イベント検出という技術がある．イベントとは，人により定義された動画像中の事象のことで，動画像中の人物や物体などの動作や音声により特徴づけられる出来事を指す．動画像中からイベントを検出することができれば，動画像を分類することができる．本稿では，大規模なインターネット上の動画像を自動で分類し管理することを目的とし，動画像中のイベントを検出する手法を提案する．

動画像中のイベントを検出する手法として，動画像上の人物や物体などの動きの時系列の状態を確率モデルで表現したものがある [2], [3]．しかし，web 上の動画像のイベントは多種多様であり，同じイベントであっても，動きの時系列がいつも同じになることはなく，時系列のモデルが有効でない場合もある．

他に，動きの情報そのものを特徴量として得る手法 [4], [5] がある．これらは，固定カメラで撮影された画像から動きの特徴量を抽出している．web 上にある多くの動画像は視点が固定されていないため，動画像から得られた動きの特徴量はカメラの動きと物体や人物の動きの両方を含んでいる．そのため，固定したカメラを想定した手法の場合，web 上の動画像からカメラの動きと物体の動きを分解して利用する必要がある．しかし，カメラの動きを推定して，カメラの動きと物体や人の動きを分離することは容易でなく，固定したカメラを想定した手法は，web 上の動画像のイベント検出は困難である．そこで本研究では，人物や物体などの動き情報を用いずにイベントを検出する．

また，イベントには，特定の環境でのみ行われるものが多数存在するため，動画像中の背景とイベントには相関があると考えらえる．例えば，ロッククライミングのイベントの動画像は，森の中や崖などで撮影されたものが多く存在する．よって，動画像中の背景を認識することで，イベントを検出することが出来ると考えられる．また，人物の行動には物体と共に行われるものもある．洗浄のイベントでは，洗浄機器が使われるし，自転車に乗るイベントでは，自転車が使われる．このことを利用して，イベントに関係している物体の情報をイベント検出に用いたり，[6], [7] 背景情報を用いてシーンの検出をする手法 [8] が提案されて

[1] 大阪府立大学
OPU, Sakai, Osaka 599–8531, Japan
[a)] katte@m.cs.osakafu-u.ac.jp
[b)] yuzuko@cs.osakafu-u.ac.jp
[c)] kise@cs.osakafu-u.ac.jp

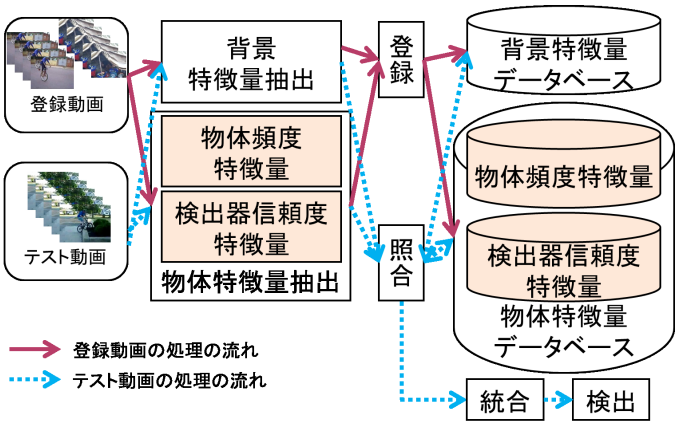



**図 1** 処理の流れ

きた．これらの手法はより精度の高い認識精度を示している．そこで，本研究では，イベント検出に，イベントに関連する物体の情報とイベントが行われている背景の情報を用いる手法を提案する．

よって，動画中に登場する物体を検出することができれば，イベントを推定することができると考えられる．本稿では，これら二つの情報を用いてイベント検出することを提案する．背景情報は，色情報を含んだ Opponent SIFT 特徴量により表した．物体情報は，物体検出器を使用して，動画中の物体を検出した結果により表した．イベントの検出は，それぞれの特徴量で最近傍探索を行い，それらの結果を統合することで行った．

TRECVID2012 Multimedia Event Detection タスクのデータセットを用いて提案手法の評価を行った．その結果，Rock climbing や，Grooming an animal といった限られた場所で行われるイベントや Attempting a bike trick や Getting a vehicle unstuck などの，自転車や車などの特定の物体が登場するイベントを検出できた．

以下，2 章でシステムの流れについて説明し，3 章で背景特徴量と物体特徴量の抽出手法について，4 章で検出手法について説明する．そして，5 章で実験について述べ，6 章で本稿をまとめる．

## 2. イベント検出システムの概要

本章では，提案手法の概要について述べる．前章で述べたように，本稿では動画像上の背景情報と登場する物体情報を用いてイベントを検出する．背景情報により，動画像中のイベントがどこで行われているものかを表現し，物体情報により，イベントにはどのような物体が使われているかを表現することを目的とする．これら二つの情報を用いてイベントを検出する．図 1 に，イベント検出までの処理の流れを示す．まず，動画像からキャプチャした画像から，背景特徴量と物体特徴量を抽出する．物体特徴量は物体検出器 [9] を用いて二つの特徴量を抽出する．一つは，検出物体の数を示す物体頻度特徴量とし，もう一つは，検出器の出力値（検出物体の信頼度）を使用した検出器信頼度特徴量とする．それぞれの特徴量で動画像が各イベントとなる確率を求め，結果を統合し動画像中のイベントを検出する．次章から各特徴量抽出手法と検出手法について述べる．

## 3. 特徴抽出手法

本章ではイベント検出に使用した特徴抽出手法ついて述べる．はじめに背景特徴量について説明したあと，物体特徴量について説明する．

### 3.1 背景特徴量

背景特徴量として，Opponent SIFT 特徴量 [10] を用いる．この特徴量は，Koen らの色情報を用いた特徴量による背景認識実験で高い精度を示している [11]．Opponent SIFT 特徴量は，特徴点検出と特徴量記述の二段階を経て抽出される．特徴点検出には，Harris Laplace detector [12] を用いる．これは輝度値の変動に元づく特徴点の検出方法である Harris detector にスケールを変化させながら生成した Laplacian of Gaussian (LoG) を組み込んだ検出器である．この特徴点検出器を用いることでスケール変化に頑健な特徴点を得ることができる．Harris Laplace detector により得られた特徴点から色情報を含む Opponent SIFT 特徴量を抽出する．まず，画像を RGB 色空間から以下の式によって Opponent 色空間 [13] に変換する．

$$\begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} = \begin{pmatrix} \frac{R-G}{\sqrt{2}} \\ \frac{R+G-2B}{\sqrt{6}} \\ \frac{R+G+B}{\sqrt{3}} \end{pmatrix} \quad (1)$$

ここで得られたチャンネル $O_3$ は HSV 色空間の明度に等しく，$O_1$ と $O_2$ はそれぞれ赤と緑，黄色と青の反対色の組の情報を保持している．そして，$O_1$ から $O_3$ のチャンネルごとに SIFT 記述子を用いて特徴抽出をすることで 128 次元 × 3 チャンネル = 384 次元の色情報を持つ特徴量を得る．

動画像より数千個抽出された Opponent SIFT 特徴量を Bag-of-Feature により表現し，動画像毎で背景情報を一つの特徴ベクトルで表現する．動画像の背景の類似度はベクトル同士のユークリッド距離により測定する．また，抽出される Opponent SIFT 特徴量の数は画像によるため，特徴ベクトルの要素の総和が 1 となるように正規化する．こうして得られた特徴量を，PCA により次元圧縮し，背景特徴量として用いる．

### 3.2 物体特徴量

動画像中の物体特徴量は，Felzenszwalb ら [9] の物体検出器を用いて抽出する．この物体検出器は，大局的なフィルタ global root filter と局所的な 6 つのフィルタ part fileter を用いて，二段階で物体を検出することにより画像中から

高精度に物体を検出することができる．この物体検出手法では，sliding window approach により画像を走査することで物体を検出する．フィルタは HOG 特徴量により構成される．root filter は粗い領域から HOG 特徴量を抽出したもので，物体全体の輪郭を表現している．part filter は細かい領域から HOG 特徴量を抽出したもので，物体の詳細の輪郭を表現している．物体検出をする際には，まず画像全体から粗く HOG 特徴量を抽出し，画像全体に対して root filter を走査させ，各領域で物体が存在する信頼度を算出することにより，物体領域となる候補の領域を検出する．次に，細かく HOG 特徴量を抽出し，part fileter を走査させ，再度領域内に物体が存在する信頼度を算出することにより，領域に物体があるかを決定する．

物体特徴量として，動画像からキャプチャした画像に対して物体検出器を用いることで，二つの特徴量を抽出する．一つは，動画中から検出した物体の頻度ヒストグラム（物体頻度特徴量）である．これは，イベントによってどのような物体がどのくらいの数だけ登場するかを表すことを目的としている．また，物体検出器による検出結果には，誤検出も多く含まれている．それらの影響を軽減することを目的として，もう一つの特徴量には，物体検出時の信頼度の値を使用する（検出器信頼度特徴量）．使用する信頼度の値は，動画像全体から物体を検出した中で，最も高かったものとし，特徴量は，各物体検出器の検出結果の中で最も高い信頼度であった値を並べたものとする．二つの特徴量の次元は，使用する物体検出器の数と一致する．

## 4. 検出手法

背景特徴量，物体頻度特徴量，検出器信頼度特徴量の三つの特徴量を用いて動画像が各イベントとなる確率を求める．各特徴量で次元数などの性質が異なるため，それぞれで K 近傍法によりクエリ周辺の特徴量の数を求め動画像が各イベントとなる確率を求める．そして，その結果を統合する．動画像 $v$ のイベント $E$ となる確率 $P(E|v)$ を次式に示す．

$$P(E|v) = \frac{\#KNN(E)}{\#DataClips(E)} \tag{2}$$

$\#KNN(E)$ はクエリの K 近傍内にあるイベント $E$ の特徴量の数である．$\#DataClips(E)$ はデータベースに登録したイベント $E$ の動画像数である．この割合を三つの特徴量毎に算出し，それらの平均値を統合した結果とした．そして，統合して導いた値が，閾値以上であるならば，該当するイベントが動画上にあると判別する．

## 5. 実験

提案手法がイベント検出に有効であるか評価するために実験を行った．

### 5.1 実験条件

実験は，TRECVID2012 Multimedia Event Detection タスクのデータセット [14] に対して行った．データセットはインターネット上から収集した動画から構成されている．このデータセット中の 20 イベントに対して提案手法の評価実験を行った．データセットは，各イベントに対して positive 動画と related 動画から構成されている．positive 動画は，一つ以上のイベントが含まれている動画のことを指し，related 動画は，イベントは含まれてはいないが，イベントに登場する人物や物体が登場する動画のことを指す．例えば，イベント Attempting a bike trick の related 動画には，自転車に乗って芸をするというイベントは含まれていないが，自転車が多く登場する動画などがある．表 1 に，20 種類のイベント名とそれぞれの学習動画数を示す．データベースとして，各イベントで 8 割の positive 動画と全ての related 動画を登録した．残りの positive 動画のうち 1 割を最近傍探索における最適な K を調査する時に使用し，残りの 1 割を評価用のクエリとした．このようにして 5 回データベースとクエリの選択を行った．

背景特徴量の抽出は，動画像から 2 秒間に 1 枚キャプチャした画像に対して行った．Bag-of-Features を構成するときの, visual word の数は 3969 個とし，PCA により特徴ベクトルを 3204 次元に圧縮した．

物体検出器は PascalVOC2009 database [15] と INRIA Person Dataset [16] により学習した．これら二つのデータベースには我々の身近にあるような多様な物体の多くの画像データが含まれており，それらの画像を用いて 21 物体の検出器（飛行機，自転車，鳥，ボート，ボトル，バス，車，猫，椅子，牛，ダイニングテーブル，犬，馬，人，バイク，植木，羊，ソファ，電車，モニタ）を作成した．物体特徴量は，動画像からランダムにキャプチャした 3 枚の画像に対して物体を検出することにより抽出した．

評価方法は，TRECVID MED タスクと同様のものを用いた．TRECVID MED タスクでは， Missed Detection probabilities($P_{MD}$) と False Alarm probabilities($P_{FA}$) の二つの指標を用いる．二つの値はそれぞれ次式により定義されている．

$$P_{MD}(E, DT) = \frac{\#MD(E, DT)}{\#Targets(E)} \tag{3}$$

$$P_{FA}(E, DT) = \frac{\#FA(E, DT)}{\#TotalClips - \#Targets(E)} \tag{4}$$

$E$ はイベント，$DT$ はイベントを検出する時の確率の閾値を表す．$\#Targets(E)$ は検索対象となるイベント $E$ の動画数を示し，$\#MD(E, DT)$ はシステムが検出できなかったイベントの動画数，$\#FA(E, DT)$ はシステムが誤検出したイベントの動画数を示す．また，$\#TotalClips$ は全クエリ動画の数である．TRECVID MED タスクでは，$P_{MA} : P_{FA} = 12.5 : 1$ となるように，イベント検出時の確

**表 1** Events for MED 12

| Event name （学習動画数） | Event name （学習動画） |
|---|---|
| Birthday party (173) | Attempting a bike trick (200) |
| Changing a vehicle tire (111) | Cleaning an appliance (200) |
| Flash mob gathering (173) | Dog show (200) |
| Getting a vehicle unstuck (132) | Giving directions to a lodation (200) |
| Grooming an animal (138) | Marriage proposal (200) |
| Making a sandwich (126) | Renovating a home (200) |
| Parade (138) | Rock climbing (200) |
| Parkour (112) | Town hill meeting (200) |
| Repairing an appliance (123) | Winning a race without a vehicle (200) |
| Working on a sewing project (120) | Working on a metal crafts project (200) |

**表 2** 各データベースにおける特徴量毎の最適な K の値

| データベース | 背景特徴量 | 物体頻度特徴量 | 検出器信頼度特徴量 |
|---|---|---|---|
| 1 | 20 | 140 | 240 |
| 2 | 50 | 80 | 110 |
| 3 | 100 | 250 | 250 |
| 4 | 350 | 40 | 200 |
| 5 | 350 | 50 | 350 |

**図 2** 車の検出例

率の閾値を調整することが望ましいとされており，本稿でも同様の値をとるような閾値でのイベント検出結果を評価した．

K 近傍法で用いる K の値は, 各データベースで実験により求めたものを使用した．各特徴の 5 通りのデータベースで，K の値を変化させ検出の精度を求め, $P_{MA} : P_{FA} = 12.5 : 1$ となる点で最も良い精度となる K の値を調査した．表 2 に結果を示す．この K の値を使用して認識精度を評価した．

### 5.2 結果と考察

図 9 と図 10 に，各イベントの検出結果を示す．グラフは横軸を $P_{FD}$ とし，縦軸を $P_{MA}$ とし，検出時の閾値を変化させた時のグラフであり，原点に近いカーブほど精度が高いといえる．グラフの opponent は背景特徴量のみを用いた精度を示し，maxConf は検出器信頼度特徴量のみを用いた精度を示し，objNum は物体頻度特徴量のみを用いた精度を示す．fusion はこれら三つの特徴量を統合した時の精度である．

図 9 (e), 図 10 (b), (d) の oppnent のグラフから, 背景特徴量により Grooming an animal, Rock climbing, Winning a race without a vehicle のイベントが検出可能であることが明らかとなった．イベント検出できた動画像をみると Grooming an animal では，白を基調とした浴槽で洗浄しているイベントが多く検出できていた．一方で，イベント検出できなかった動画像を見ると，屋外で行われているイベントは殆ど検出できていなかった．Rock climbing も動画の殆どが森や岩壁，ロッククライミングウォールで構成されているものが多く検出できていた．一方でピンク色のロッククライミングウォールや緑のシートなどが動画像の多くを占めている動画は検出できていなかった．Winning a race without a vehicle では，競技場やプールで撮影された動画像が多くあり，こういった環境の中でのイベントは検出することができた．このように限られた場所で行われるようなイベントは，Opponent SIFT 特徴量を使用することが有効であると考えられる．

図 9 (d)，(k) の maxConf のグラフから，検出器信頼度特徴量により，Getting a vehicle unstuck と Attempting a bike trick のイベントが検出可能であることが明らかとなった．Getting a vehicle unstuck のイベントでは車が多く登場している．Attempting a bike trick のイベントでは自転車やバイクが多く登場している．図 2 と図 3 に車，自転車の検出例の一部を示す．これらのイベントでは，物体検出器により高精度に検出できたため，高い精度でイベントを検出できたものと考えられる．一方で，図 9 (e) , (m) より，犬や猫が多く登場する Grooming an animal や Dog show のイベントの検出率は低くなった．これは犬や猫などの物体検出器が車や自転車の物体検出器と比べ精度が低いためであると考えらえる [9]．図 4，図 5 に示すように，物体の検出結果を見ると，犬や猫を牛や馬として検出している例が多くあった．物体の特徴量は動画像からランダムにキャプチャした三枚の画像から抽出したが，画像中にイベントに関連した物体を含まない画像も多く存在していた．イベントに関連した物体の取りこぼしを避けるために，物体特徴量を抽出する画像の枚数を増やす必要があると考えられる．

図 9 , 図 10 の objNum のグラフより，物体の頻度特徴量のみを用いた場合，イベント検出の精度が低くなった．

**図 3** 自転車の検出例

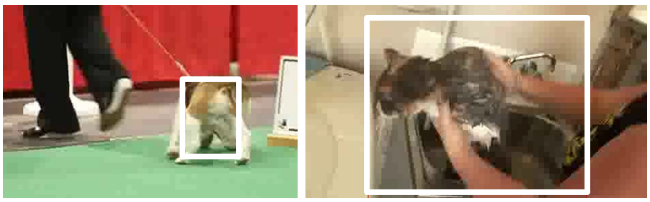


**図 4** 牛と検出された犬や猫の例

**図 5** 馬と検出された犬の例

**図 6** 自転車の誤検出例

図 6 に，自転車の誤検出例を示す．この他の物体検出器も同様に誤検出が多数あるため，イベント検出の精度が低かったものと考えられる．

図 9 (j)，(l)，(n)，(m) の fusion のグラフから，Working on a sewing project や，Cleaning an appliance，Giving directions to a location，Dog show のイベントでは，全ての特徴量を統合することで，検出の精度が向上することが明らかとなった．背景情報では，室内や路上などの特定の環境下で行われるイベントを検出することができ，物体特徴量では，物体検出器により車や自転車などの高精度に検出できる物体を含んだイベントを検出することができた．このように各特徴量で性質が異なるため，各特徴量で検出できるイベントは異なる．そのため，統合することにより精度が向上したものと考えられる．図 7 にイベント Dog show の動画像から物体を検出するために，キャプチャした画像を示す．Dog show では，このように芝生の上や緑のマットの上でイベントが行われているものが多く存在する．図 7 の画像中には物体が登場しておらず，物体が検出

**図 7** Dog show の画像例

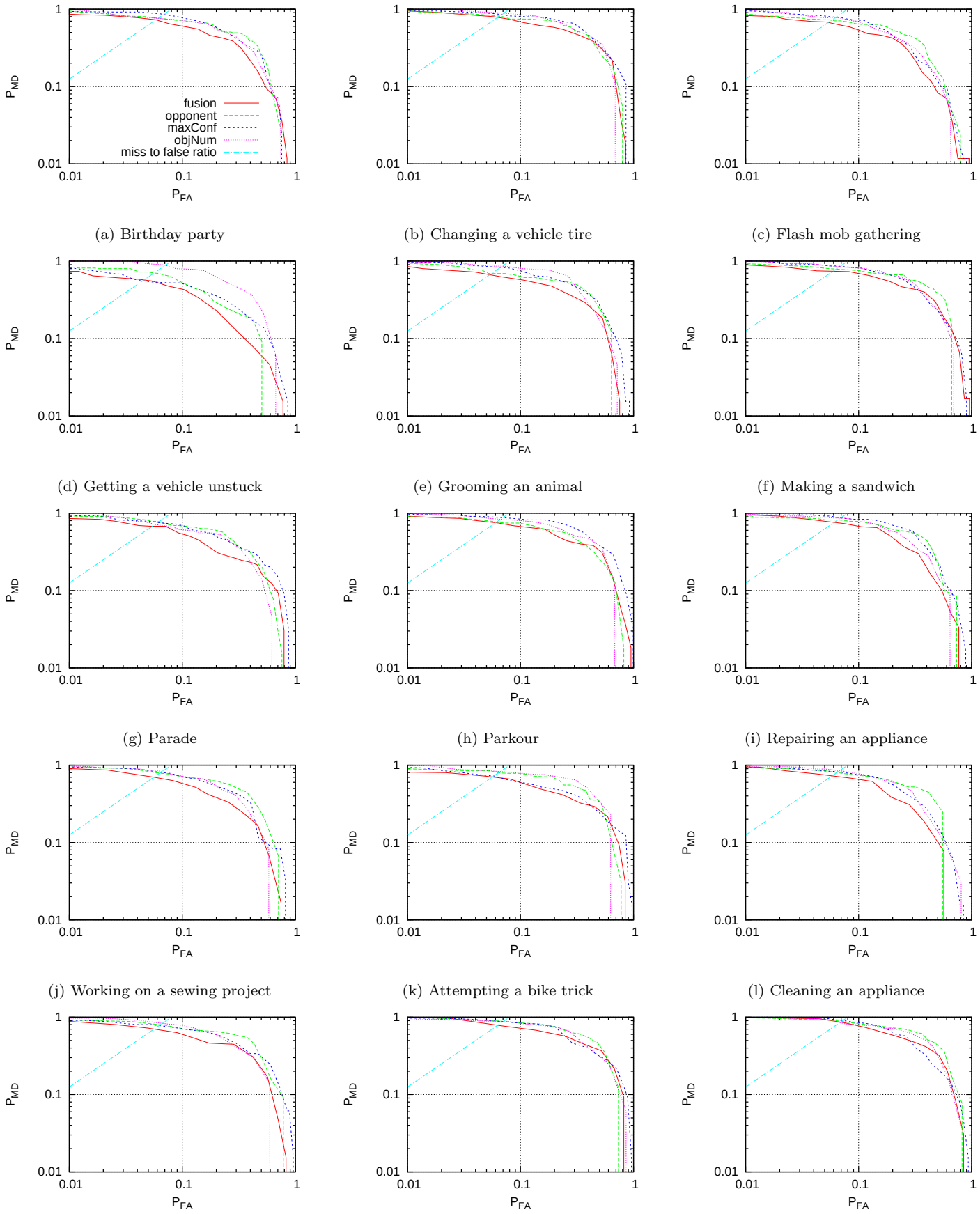
**図 8** Marriage proposal の画像例

できなかったため，物体特徴量では，イベントを検出できなかった．一方で，背景特徴量では，背景の情報を得ることができたため，イベントを検出することができた．そのため，物体特徴量のみでは検出できなかったイベントも，特徴量を合わせることでイベント検出することができた．逆に，背景情報のみではイベント検出を失敗していた場合でも，車や自転車などの物体を高精度に検出できた場合は，物体特徴量によりイベント検出できており，結果を統合することで検出することができると考えられる．

図 9 (b)，(o)，(p)，図 10 (e) より，殆ど検出できないイベントがあった．これらのイベントは，同じイベントであっても異なる環境下で行われていたり，登場する物体の物体検出器が，実験で使用した検出器の中にないことが検出に失敗した原因だと考えらる．また図 8 に示すように，Marriage proposal のイベントでは，イベントが庭や屋内，山の中など様々な環境下で行われている．さらに，登場する物体検出の対象が人のみであるため，イベントを特定するような物体情報が乏しく，イベント検出することが困難である．Marriage proposal にはキスやハグなどの特定の行動が多く登場する．これらを表現できるような，人物の姿勢情報や行動情報などの新たな特徴量を加える必要があると考えらえる．

## 6. まとめ

本稿では，動画像の背景情報と，登場する物体情報を用いてイベントを検出した．背景特徴量として，色情報を含んだ Opponent SIFT 特徴量を使用した．物体特徴量として，物体検出器を使用して，検出できた物体の頻度ヒストグラムと，検出した物体の信頼度の値を用いた．背景特徴量では，浴槽や競技場など限られた場所で行われるイベントを検出することができた．物体の頻度ヒストグラムを使用した特徴量では，低い検出結果となった．物体の信頼度を使用した特徴量では，物体検出器が精度の高い検出を出来た場合に，イベントを検出することができた．それぞれの特徴量がイベントを検出できる動画像は異なるため，これらの特徴量を統合することで，検出精度を向上することができた．今後の課題として，物体検出器の種類を増やすことと精度の向上が挙げられる．インターネット上にある動画像を対象としているため，物体検出器の学習画像も動画像から収集することが考えらる．また，検出に有効な新




(a) Birthday party (b) Changing a vehicle tire (c) Flash mob gathering

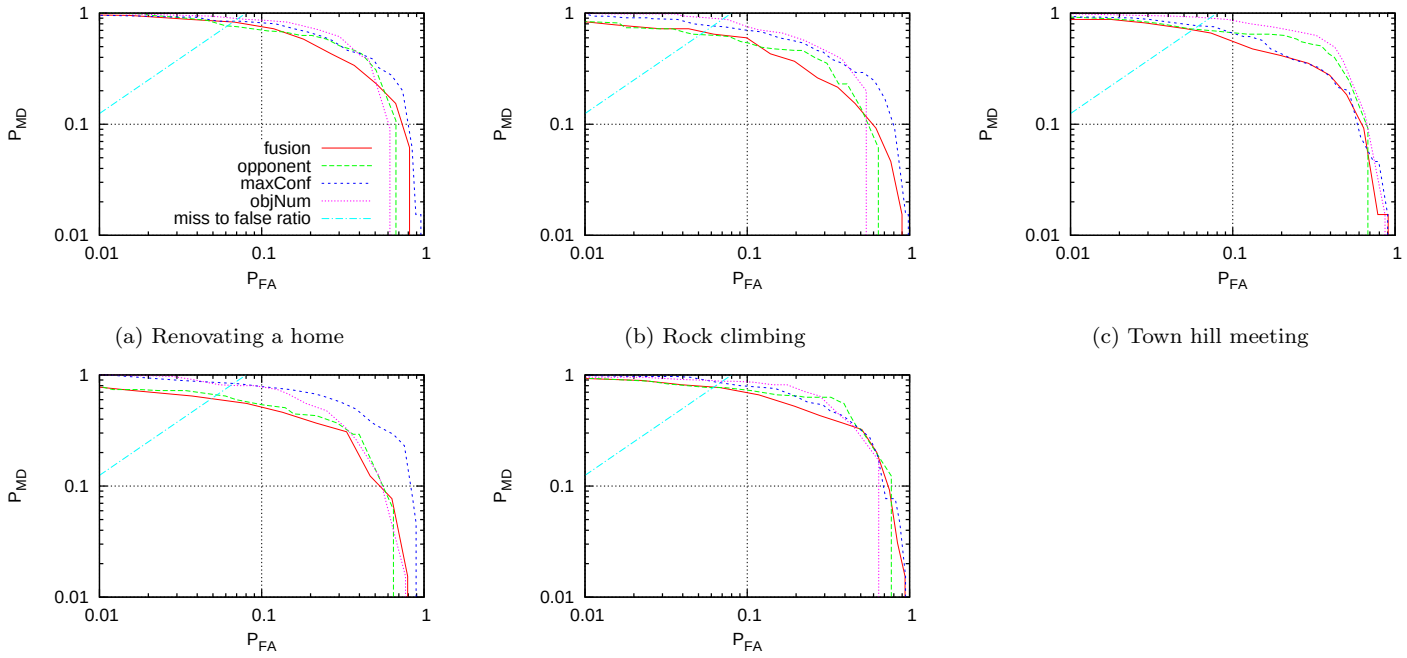
(d) Getting a vehicle unstuck (e) Grooming an animal (f) Making a sandwich

(g) Parade (h) Parkour (i) Repairing an appliance

(j) Working on a sewing project (k) Attempting a bike trick (l) Cleaning an appliance

(m) Dog show (n) Giving directions to a location (o) Marriage proposal

**図 9** 各イベントにおける認識結果 (1)




(a) Renovating a home (b) Rock climbing (c) Town hill meeting

(d) Winning a race without a vehicle (e) Working on a metal crafts project

**図 10** 各イベントにおける認識結果 (2)

たな特徴量を調査することも今後の課題である.

## 参考文献

[1] http://www.youtube.com/.
[2] K. Tang, L. Fei-Fei, and D. Koller, "Learning latent temporal structure for complex event detection," Proceedings of 2012 IEEE Conference onComputer Vision and Pattern Recognition, pp.1250–1257, 2012.
[3] M. Albanese, R. Chellappa, V. Moscato, A. Picariello, V.S. Subrahmanian, P. Turaga, and O. Udrea, "A constrained probabilistic petri net framework for human activity detection in video," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol.10, no.6, pp.982–996, 2008.
[4] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol.30, no.3, pp.555–560, 2008.
[5] Y. Ke, R. Sukthankar, and M. Hebert, "Efficient visual event detection using volumetric features," Proceedings of Tenth IEEE International Conference on Computer Vision, vol.1, pp.166–173, 2005.
[6] A. Gupta, A. Kembhavi, and L.S. Davis, "Observing human-object interactions: Using spatial and functional compatibility for recognition," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol.31, no.10, pp.1775–1786, 2009.
[7] F. Wang, Y.-G. Jiang, and C.-W. Ngo, "Video event detection using motion relativity and visual relatedness," Proceedings of the 16th ACM international conference on Multimedia, pp.239–248, 2008.
[8] L.-J. Li and L. Fei-Fei, "What, where and who? classifying events by scene and object recognition," IEEE 11th International Conference on Computer Vision, pp.1–8, 2007.
[9] P.F. Felzenszwalb, R.B. Girshick, and D. McAllester, "Discriminatively trained deformable part models, release 4," http://people.cs.uchicago.edu/~pff/latent-release4/.
[10] K.E.A. van deSande, T. Gevers, and C.G.M. Snoek, "Color descriptors for object category recognition," European Conference on Color in Graphics, Imaging and Vision, pp.378–381, 2008.
[11] K.E.A. van deSande, T. Gevers, and C.G.M. Snoek, "Evaluating color descriptors for object and scene recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.32, no.9, pp.1582–1596, 2010.
[12] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," Int. J. Comput. Vision, vol.60, pp.63–86, Oct. 2004.
[13] J. van deWeijer and Th. Gevers, "Boosting saliency in color image features," Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01, pp.365–372, Washington, DC, USA,
[14] The National Institute of Standards and Technology(NIST), "TREC video retrieval evaluation". http://www-nlpir.nist.gov/projects/trecvid/
[15] M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2009 (VOC2009) Results," http://www.pascal-network.org/challenges/VOC/voc2009/workshop/index.html.
[16] "INRIA Person Dataset". http://pascal.inrialpes.fr/data/human/