

# 動画像からの対話的な情報抽出

石山佳史\*

戸崎哲也\*\*

## Interactive Extraction of Significant Information from Fixed Observation Camera

Keishi ISHIYAMA\*

Tetsuya TOZAKI\*\*

### ABSTRACT

The movies taken by the fixed video camera have been used to recognize to some significant informations. In this study, we propose the method that can extract the significant informations such as colors or moving directions from the movie data taken by fixed observation camera. Firstly, we extract only moving objects based on wavelet decomposition. Wavelet decomposition is a method to analyze time and frequency at once. Secondly, we analyze the feature of moving objects. The contents of analysis are colors, the movements and the moving directions. For the color determinant we use the color circle. We calculate the movement and the moving direction are calculated by presuming optical flow. We extract the target whose characteristics are given interactively. Finally we show the effectiveness of our method by the experimental results.

*Keywords* : wavelet decomposition, optical flow, labeling

### 1. はじめに

ビデオカメラにより撮影された映像は、一般的には旅行や催しの思い出の記録、あるいは動物や植物などの生態や自然風景をとらえておくなど、色々な場面で幅広く利用されている。これら映像には、人物や物体の動きや風景が記録として保存されるので、そこから何らかの情報を取り出すことによって様々な方面に利用することが可能であると考えられる。例として、防犯カメラの映像から犯人を特定したり、未然に犯罪を防止する事例、高速道路上に設置したカメラから特定の車を認識する事例等が挙げられる。このような動画像情報は、常時物体を映し出しているため多くの情報を有し、この中から必要な情報のみ抜き出して利用することが効果的であるとされる。しかし、情報量が莫大であるため、要素を特定するために必要な情報を限定的に取り出すには相当な労力を必要とする。そこで本研究では、ユーザが様々な特徴を対話的に与え、動画像からそれに合致する情報を自動的に抽出する手法の開発を目的とする<sup>(1)</sup>。本稿では、ユーザが提示する情報として色、対象物体の移動量／移動方向について

検討を行ったのでその報告を行う。

映像情報から特定情報を抽出する手法は、以下のとおりである。まず、動画像から移動物体のみを取り出す。ここでは、動画像を時間方向に積み上げた3次元データとしてみなし、時間軸方向に画像列を周波数解析し高周波成分に相当する部分を移動物体と考える。周波数解析には、ウェーブレット分解を用い、これにより得られた高周波成分を取り出すことで移動物体の検出を実現する。次に、移動物体の特徴解析を行う。考慮する特徴は色、移動量・移動方向である。色の情報に関しては色相環と呼ばれるものを用いる。また、移動量・移動方向に関してはオプティカルフローを推定して導出する。最後に関心領域を限定して観察できるように関心領域の切り出しを行う。関心領域ごとに矩形領域を取り、この領域内にもみ2値化を施す。

本論文の構成は次のとおりである。2章では移動物体の検出や色の特定、および移動量・移動方向の特定に関する手法について述べ、3章では2種類の動画像に対して適応した結果を述べ、4章ではその考察を行い、5章で本報告のまとめを行う。

\* 専攻科 電気電子工学専攻

\*\* 電子工学科 准教授

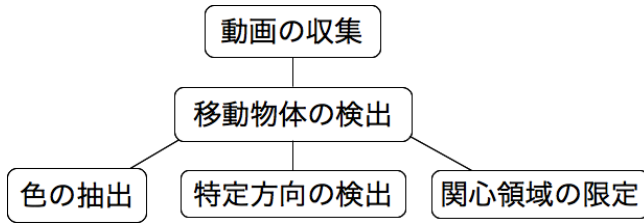


図1 システムの流れ

2. 手法

2.1 手順 図1に本研究で提案するシステムの流れを示す。撮影された動画像に対してウェーブレット分解を用いて移動物体を検出する。これにより、背景を除去して移動物体のみを検出することが可能となる。さらに、検出された移動物体から色の抽出、移動方向の検出を行い、それらを解析する。また、移動物体には個別にラベルを付与し、与えられたラベルごとに矩形領域を設け、関心領域を限定して重点的に観察できる機能を設ける。

2.2 移動物体の検出 移動物体の検出には、ウェーブレット分解を用いる。動画像を時間方向の画像列とみなし、時間軸方向に画像列を追跡すると、画素値に関する時系列信号が得られる。この時系列信号の画素値は、移動物体の境界付近で急激に変化する時点がある。この変化をウェーブレット分解の高周波成分としてとらえることができ、これを各座標に対して行うと、移動物体の輪郭が得ることができる。つまり時系列信号を分解して高周波成分を取り出すと、移動物体の検出ができる。これを基に移動物体検出フィルタを作る。また、今回は2次の Daubechies ウェーブレットを用いる。このウェーブレットを用いる理由は、連続関数であり、必要なフレームが少ないからである。

座標(k,l)における時系列信号 $\{C_{k,l}^1(t)\}_{0 \leq t \leq T}$ をウェーブレット分解すると、低周波成分と高周波成分はそれぞれ式(2.1)、式(2.2)で表される。

$$C_{k,l}^0(t) = \sum_{\tau} \alpha_{\tau-2t} C_{k,l}^1(\tau) \quad (2.1)$$

$$D_{k,l}^0(t) = \sum_{\tau} \beta_{\tau-2t} C_{k,l}^1(\tau) \quad (2.2)$$

このとき $\tau$ の動く範囲は $0 \leq \tau - 2t \leq 3$ となり、 $\tau$ のとりうる値は  $2t, 2t+1, 2t+2, 2t+3$  となる。これは低周波画像  $C^0(t)$ と高周波画像  $D^0(t)$ を得るのに4つの画像フレーム  $C^1(2t), C^1(2t+1), C^1(2t+2), C^1(2t+3)$ が必要であることを意味する。次の低周波画像  $C^0(t+1)$ と高周波画像  $D^0(t+1)$ は、因子2でサンプリングされるので、4つの画像フレーム  $C^1(2t+2), C^1(2t+3), C^1(2t+4), C^1(2t+5)$ から計算する。また、ここでの $\alpha$ はスケール関数、 $\beta$ はウェーブレット関数と呼ばれる。 $\alpha_n, \beta_n$ は文献(2)より表1の通りとなることが示されている。

表1 N=2のときの $\alpha_n, \beta_n$

n	$\alpha_n$	$\beta_n$
0	0.482962913145	-0.129409522551
1	0.836516303738	-0.224143868042
2	0.224143868042	0.836516303738
3	-0.129409522551	-0.482962913145

ある時間帯において、ある座標における画素値が一定の値を保っていれば、その時間帯のその場所は背景領域、逆に画素値が変動していると、その時間帯のその場所は動領域であると考えられる。上述の方法で画像列をウェーブレット分解することで、高周波画像列が得られる。1つの高周波画像は4つの画像フレームにわたる画素値の変化をとらえることを意味している。また、この分解では因子2のサンプリングがされてきたので、原画像フレームが時間方向に1ずれると、元と異なる高周波成分が得られる。偶数番目のフレームから出発して分解すると式(2.3)、奇数番目のフレームから出発して分解すると式(2.4)で表される。

$$\bar{D}_{k,l}^0(t) = \sum_{\tau} \beta_{\tau-2t} C_{k,l}^1(\tau) \quad (2.3)$$

$$\bar{D}_{k,l}^0(t) = \sum_{\tau} \beta_{\tau-2t} C_{k,l}^1(\tau+1) \quad (2.4)$$

移動物体を検出するためのフィルタを次の手順によって作成する。最初に式(2.5)を用いて高周波成分 $\bar{D}_{k,l}^0(t)$ の二値化を行う。

$$F_{k,l}^{(4)}(2t) = \begin{cases} 1, & |\bar{D}_{k,l}^0(t)| > \theta \\ 0, & \text{その他} \end{cases} \quad (2.5)$$

ここで $\theta$ は閾値である。 $F_{k,l}^{(4)}(2t)$ は高周波成分 $\bar{D}_{k,l}^0(t)$ が $C^1(2t)$ から始まる4つの原画像フレーム  $C^1(2t), C^1(2t+1), C^1(2t+2), C^1(2t+3)$ から計算されることを意味する。同様にして、 $F_{k,l}^{(4)}(2t)$ の時間tを1だけ進めた $F_{k,l}^{(4)}(2t+2)$ を用意する。式(2.6)に示すように、ここまでで得られたフィルタ $F_{k,l}^{(4)}(2t)$ と $F_{k,l}^{(4)}(2t+2)$ の論理積を取る。

$$F_{k,l}^{(2)}(2t+2) = F_{k,l}^{(4)}(2t) \wedge F_{k,l}^{(4)}(2t+2) \quad (2.6)$$

図2は $F_{k,l}^{(4)}(2t+2)$ の構成をまとめたものである。次に、式(2.7)を用いてもう一つの高周波成分 $\bar{D}_{k,l}^0(t)$ の二値化を行う。

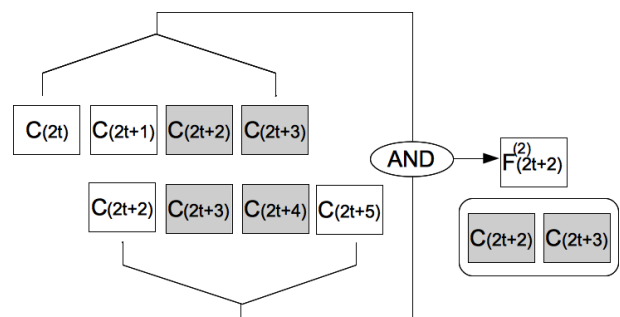


図2 フィルタ $F_{k,l}^{(2)}(2t+2)$ の構成

$$F_{k,l}^{(4)}(2t+1) = \begin{cases} 1, |\hat{D}_{k,l}^0(t)| > \theta \\ 0, \text{その他} \end{cases} \quad (2.7)$$

先程と同様に  $F_{k,l}^{(4)}(2t+1)$  は高周波成分  $\hat{D}_{k,l}^0(t)$  が  $C^1(2t+1)$  から始まる 4 つの原画像フレーム  $C^1(2t+1), C^1(2t+2), C^1(2t+3), C^1(2t+4)$  から計算されることを意味する。またここでも時間  $t$  を 1 進めたフィルタ  $F_{k,l}^{(4)}(2t+3)$  を用意し、式(2.8)に示すように  $F_{k,l}^{(4)}(2t+1)$  と  $F_{k,l}^{(4)}(2t+3)$  の論理積を取りフィルタを構成する。

$$F_{k,l}^{(2)}(2t+3) = F_{k,l}^{(4)}(2t+1) \wedge F_{k,l}^{(4)}(2t+3) \quad (2.8)$$

次に式(2.6)と式(2.8)で得られたフィルタ  $F_{k,l}^{(2)}(2t+2)$  と  $F_{k,l}^{(2)}(2t+3)$  の論理積を取り、新たにフィルタを式(2.9)で構成する。

$$F_{k,l}^{(1)}(2t+3) = F_{k,l}^{(2)}(2t+2) \wedge F_{k,l}^{(2)}(2t+3) \quad (2.9)$$

こうして 1 フレーム  $C^1(2t+3)$  において動領域を取り出すことができる。図 3 にフィルタ  $F_{k,l}^{(1)}(2t+3)$  の構成をまとめる。これを繰り返すことで動領域の検出を行う。

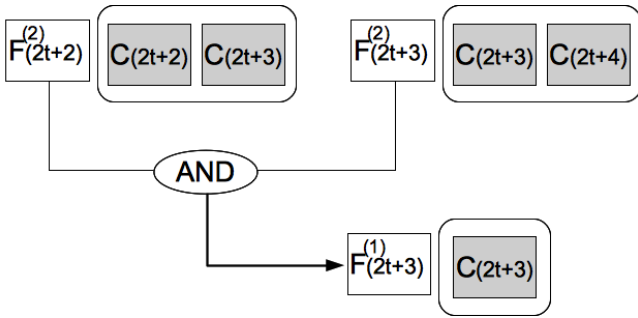


図 3 フィルタ  $F_{k,l}^{(1)}(2t+3)$  の構成

**2.3 色の抽出** 色は、HSV 色空間（色相・彩度・明度）からなる。色相とは色味の違いのことで赤・緑・青・黄・紫などがある。彩度は色の鮮やかさ、つまりカラーのことを表す。彩度が上がると色がカラフルになり、反対に彩度が下がると白黒に近づく。明度は色の明るさのことで明度が上がるとだんだんと白っぽく、反対に明度が下がると暗くなる。これら 3 つの値を各々で変化させることで色に違いが出る。



図 4 色相環

色の抽出は色相を用いることにした。図 4 のように色相を円状に並べ、各色の間にそれぞれに相当する中間色を加えたものが色相環<sup>(3)</sup>と呼ばれており、1 周で元の色に戻るのので色の分割をしやすいからである。

画像を読み込む際、色の情報は初めに赤緑青の RGB (3 原色) で認識される。RGB のままでは色相の利用ができないために、事前に RGB から HSV へ変換を行う。明度と彩度はここでは変化をさせないために、そのままにする。色相環は 360 度で 1 周して元の色に戻るのので、360 度の間を等間隔で分割し 12 色の中から色を選択できるようにする。例えば、0 度が赤とされているのだが、その「赤」を取り出したい場合は 0 度の前後 15 度ずつを含めた 30 度の範囲を「赤」として抽出する。他の色も同様に 30 度ずつに範囲としている (12 色: 赤, 赤橙, 黄橙, 黄, 黄緑, 緑, 青緑, 緑青, 青, 青紫, 紫, 赤紫)。

**2.4 移動方向の検出** オプティカルフローを推定することで、物体の移動方向の検出を行う。オプティカルフローとは、時間が連続する画像列を用いて、物体の動きをベクトル化するものである<sup>(4)</sup>。時刻  $t$  において点  $(x, y)$  の輝度を  $\rho(x, y, t)$  とし、微小時間  $\delta t$  後の移動点を  $(x + \delta x, y + \delta y)$  とする。物体上の輝度は不変であるとする (2.10) が成り立つ。

$$\rho(x, y, t) = \rho(x + \delta x, y + \delta y, t + \delta t) \quad (2.10)$$

(2.10) の右辺をテイラー展開すると (2.11) の式が得られる。

$$\rho = \rho + \frac{\partial \rho}{\partial x} \delta x + \frac{\partial \rho}{\partial y} \delta y + \frac{\partial \rho}{\partial t} \delta t + O(\delta x, \delta y, \delta t) \quad (2.11)$$

ここで  $O(\delta x, \delta y, \delta t)$  は  $\delta x, \delta y, \delta t$  に関する 2 次以上の高次項であり、微小であるとして無視する。両辺を  $\delta t$  で割ると (2.12) が導かれる。

$$\frac{\partial \rho}{\partial x} \frac{\delta x}{\delta t} + \frac{\partial \rho}{\partial y} \frac{\delta y}{\delta t} + \frac{\partial \rho}{\partial t} = 0 \quad (2.12)$$

$\delta t$  の極限として  $\delta t \rightarrow 0$  とすると (2.12) は (2.13) に置き換えられる。

$$\frac{\partial \rho}{\partial x} \frac{dx}{dt} + \frac{\partial \rho}{\partial y} \frac{dy}{dt} + \frac{\partial \rho}{\partial t} = 0 \quad (2.13)$$

ここでオプティカルフローの速度成分  $(u, v)$ 、空間的な輝度勾配  $\rho_x, \rho_y$ 、時間的な輝度勾配  $\rho_t$  を (2.14) のようにおくと (2.13) は (2.15) のように表される。

$$u = \frac{dx}{dt}, v = \frac{dy}{dt}, \rho_x = \frac{\partial \rho}{\partial x}, \rho_y = \frac{\partial \rho}{\partial y}, \rho_t = \frac{\partial \rho}{\partial t} \quad (2.14)$$

$$\rho_x u + \rho_y v + \rho_t = 0 \quad (2.15)$$

しかし (2.15) だけではオプティカルフローを一意に決定することができず、もう 1 つ拘束条件が必要である。そこで、同一物体の濃度パターン上の局所領域においては、オプティカルフローは一定であるとする。すなわち (2.16) を仮定する。

$$\frac{\partial(u,v)}{\partial x} = \frac{\partial(u,v)}{\partial y} = 0 \quad (2.16)$$

(2.15)と(2.16)の2つの拘束式から誤差Eは(2.17)で表される。

$$E = \sum_x \sum_y \left( \rho_x(x,y,t)u + \rho_y(x,y,t)v + \rho_t(x,y,t) \right)^2 \quad (2.17)$$

(2.17)の評価関数が最大となるようにオプティカルフローの速度成分  $u,v$  を決定する。このオプティカルフローを特徴点の移動ベクトルとする。

**2.5 関心領域の限定** 注目する領域を1物体に限定することで、より重点的に観察することができる。得られた領域から関心領域を選択し、その領域のみに2値化処理等の動画処理を施す。こうすることで領域を限定して個体を抽出することができる。

### 3. 実験結果

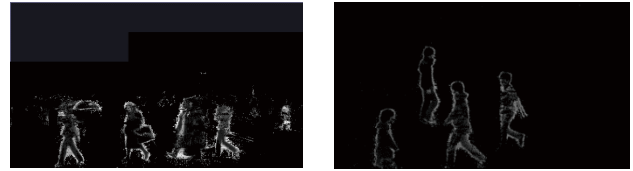
動画を使用し、移動物体の検出、色復元、物体の個数判別、色の抽出・移動方向の検出を行った。なお、カメラによるブレは考慮しない。

**3.1 使用動画** 特徴抽出のために2つの動画を使用した。図5はそれぞれの動画の1フレームを抜き出したものである。(a)は4人の人物に異なる色の服を着てもらい左右方向へ移動する様子を意図的に撮影した動画、(b)は撮影が意図的でないNHKクリエイティブ・ライブラリー(<http://www1.nhk.or.jp/creative>)でダウンロードしたミラノの横断歩道の動画の2種類である。(a)は10秒295フレーム、(b)は15秒463フレームである。



(a)意図的に撮影した (b)撮影が意図的でない  
図5 動画の1フレーム

**3.2 移動物体の検出** ウェーブレット分解を用いた移動物体の検出を行った。その様子を図6に示す。図から背景が取り除かれ、移動物体のみを検出出来ることが確認できる。どちらの画像も移動している人物、車以外でも背景の一部が細かい部分で残ってしまったり、頭や足元などの暗い部分などで動いているのに検出できていない部分がみられたりしてしまった。また、(b)の実際の映像では奥で人が動いているが、物体が小さすぎるためにうまく検出はできていないように見て取られる。これは小さい画素がフィルタリングされてしまうからである。ノイズが残っているため特徴抽出の際の障壁となってしまう可能性がある為、1回の収縮と3回の膨張による処理を行った。



動画(a) 動画(b)  
図6 移動物体のみを抜き出した画像

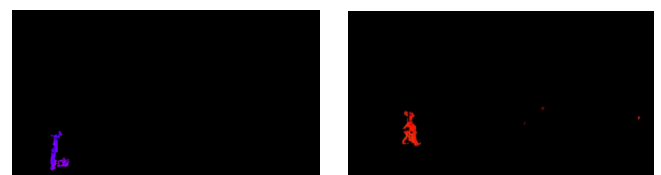
**3.3 移動物体の色復元** 先ほどの画像に元々付いていた色をつけた。撮影した動画のフレーム（元画像）と移動物体を検出した画像（処理後画像）の画素値の論理積を取ることで、処理後画像のある座標が白色であれば元画像の色が塗られ、背景と認識された黒色の座標ならばそのまま黒色とした。図7はそれぞれの動物体に対して実際に色をつけたものである。この画像から元の色が復元されていることが確認できる。ただしノイズの影響が大きいいためか(b)の復元画像に粗さが残る。



動画(a) 動画(b)  
図7 色を復元させた画像

**3.4 物体の個数判別** 物体の個数を判別するために、ラベリング処理<sup>(5)</sup>を行った。図7の画像をそれぞれラベリングした所、ラベル数は(a)が4、(b)が6となった。画像と比較したところ、実際の人数と合致しており正確に計算されていると考えることができる。ノイズによって両画像ともこれ以上のラベル数が計数されていたのだが、それを考慮して今回は画素数400を閾値とし、閾値以下のものはラベルとして考慮しなかったために、ラベル数がこの結果となった。また、(b)の動画は(a)と比べて正確にラベル数が計数できていないフレームが多く見られた。

**3.5 色の抽出** 移動物体を検出した画像から指定した色を取り出した。2.3章の12色から(a)の画像では紫色、(b)の画像では赤色を選択した結果が図8である。

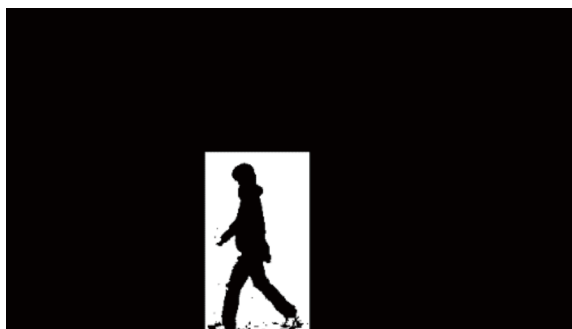
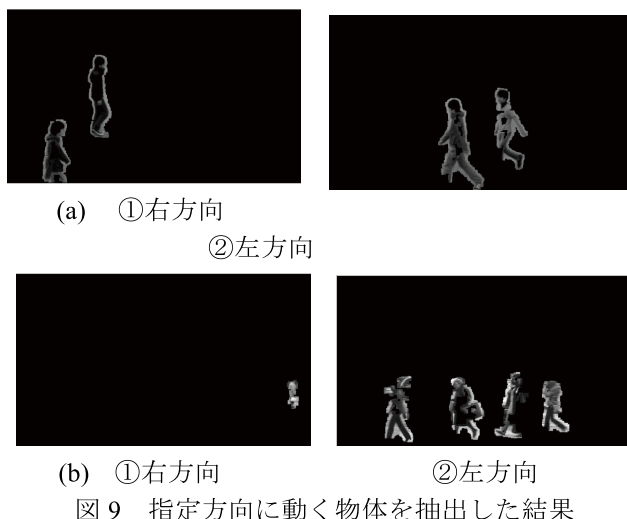


(a)の画像から紫を抽出 (b)の画像から赤を抽出  
図8 特定色の抽出結果



結果画像から、(a)では左端の紫色の服の人物の紫色が主に抽出、(b)では左端の赤い服の人物の赤色が主に抽出され、色が十分取り出されているといえる。特定色が抽出出来たため、ここでいう紫色の服の人物、赤い服の人物の特定につなげることができる。ただし、色相による色抽出では(a)の中央の人物のような白や黒、灰色というような色味のない色の抽出はできないことがわかった。

**3.6 移動方向の検出** 3次元データ化した動画から2つのフレームを用いてオプティカルフローを導出する。ここでは空間最適化法によって求められるオプティカルフロー全てから、右方向のベクトルを正方向として、ラベルごとに方向の平均をとった。平均が正の場合は右方向、負の場合は左方向にそれぞれ物体が移動しているという判定をここで用いて、右方向へ動く物体、左方向に動く物体それぞれを抽出した。その結果が図9である。どちらの画像も、元動画を比較した所、移動物体の方向に間違いはなく、指定した方向の物体が正しく抽出出来ていることが確認できる。



**3.7 関心領域の限定** 注目領域の最小座標、最大座標から矩形領域をとった。その後この領域のみに二値化処理を行った。その結果が図10である。図より特定の関心領域が良好に抽出処理できていることが確認できる。

#### 4. 考察

まず、動画から得られたフレームごとの画像をウェーブレット分解することによって、移動している物体だけを画像に残すことができることがわかった。しかしこの状態ではノイズが残されてしまうことから二値化や拡大縮小といった画像処理が必要となった。処理の結果、意図的に撮影した図3.1(a)の動画は比較的良好に移動物体が取り出されていたのに対し、(b)の動画ではノイズが多く残った。小さい領域はフィルタリングされるため、これはカメラ奥で移動しているような小さい物体の移動が主な原因と考えられる。

ラベリング処理によって、対象である物体を計数することができることがわかった。しかしラベリングするにおいて二値化の処理を通さなければならないため、物体が暗すぎる場合などはラベルの数のぶれがあり、正確に対象である物体数を計数できないフレームも見つけられた。また、ある程度の画素数以下を無視することで、ノイズの計数を回避することができた。しかしその分小さな物体には対応できないため、現行のプログラムでは比較的手前の物体以外に対応しづらい仕様であると考えられる。また、動画(a)でのラベリングはどのフレームでも比較的良好であったが、(b)の動画は(a)と比較すると正確なラベル数の取れないフレームも多く見られた。背景に映る人物が微小な動作をしたためと考えられる。

(a)(b)どちらの場合も検出した画像に元の色を正確に復元することができることを確認でき、色相環をもとに指定した12色から服の色を抜き出すことができた。暗い色のような無彩色は色相を用いた特定色の抽出では抜き出すことが不可能であることも分かった。これは、明度の度合いにフィルタをかけることで白や黒、灰色を表現できると考えられる。また、取得したラベルからフィルタリングをして、対象色が多く含まれないノイズとなる部分について除去をすることで、物体個々の色抽出も可能であることから、色抽出が利用できることも分かった。

オプティカルフローの推定と、与えられたラベルを用いることによって、指定した方向へ移動する物体の抽出ができることを(a)(b)の動画両方の結果から確認できた。また、オプティカルフロー推定や、ウェーブレット分解において個々の処理に長い時間を要してしまっていることが問題点として挙げられる。

## 5. まとめ

意図的に撮影した動画，Web上で公開されている任意の動画の2つをそれぞれフレームごとに分割し，ウェーブレット分解によって移動物体のみを取り出した。そこから二値化・ラベリングといった処理をすることでノイズを除去した。また与えられたラベリングによって物体数の計数を行った。ここまでで得られた情報から色の抽出，移動方向の検出を行ったところそれぞれの特徴を抽出することが出来た。最後に関心領域を限定してその領域を重点的に観察できる機能を設けたところ，与えられたラベルによって注目領域を指定して2値化することでその領域に限定した処理ができるようになった。

考えられる課題として，色の抽出においては，色相環を用いた12色では色の表現に限界が有ることがわかっており，黒や白といった彩度の低い色の抽出ができるように進めている。次に物体の計数である。今回はラベリングによって物体を計数したが，物体が重なった場合に1つの物体としてラベルが振られてしまうために，物体分別のためには他の手法を考える必要がある。また，(b)の動画においては背部のノイズが多かったこともあり，物体が2つに分けられたり，ノイズ対策のために画素数の小さいものは画像から除去したことによって小さく映る物体は特徴抽出に利用できなかつたりと，ラベリングをするまでに問題点が多かった。撮影条件を限定することで解決を進める。もう1つの特徴抽出として移動方向の検出を行ったが，こちらは両動画ともに良好な結果が得られたと言える。ウェーブレット分解で移動物体の抽出，オプティカルフローの推定では指定方向へ移動する物体の抽出を実現できたが，この2つの処理をするためには多くの計算時間を要する。そのため，並列処理等のアルゴリズムの導入が有効と考えられる。

### 参考文献

- (1) 石山佳史・戸崎哲也：「動画像からの対話的な情報抽出法の検討」，平成24年電気関係学会関西連合大会講演論文集，pp.459，2012.
- (2) 新島耕一：「ウェーブレット画像解析」，科学技術出版，1999.
- (3) コニカミノルタ：「マンセル表色系」，  
<http://www.konicaminolta.jp/instruments/knowledge/color/part1/09.html>  
(閲覧2013/09/27)
- (4) 三池秀敏・古賀和利：「パソコンによる動画像処理」，森北出版，1993.
- (5) 酒井幸市：「デジタル画像処理入門」，コロナ社，1997.