動き特徴を考慮した領域分割による動物体領域抽出*

鈴木哲平** 高山将太** 磯部 翔† 増田 誠† 青木義満**

Moving Object Proposal by Grouping with Motion Feature

Teppei SUZUKI, Shota TAKAYAMA, Sho ISOBE, Makoto MASUDA and Yoshimitsu AOKI

Recently, object recognition is improved accuracy by Convolutional Neural Network(CNN) and growing object recognition demand for automatic driving system and security and so on. However, one of the problem of object recognition is to extract object region which have various size, scale and form. This paper propose moving object proposal for object recognition. Using region grouping and scoring with optical flow, we can propose object proposal for moving objects. The object proposal experiments show effective results compared with previous method on the UCF crowd dataset.

Key words: object proposal, dense crowd, optical flow, segmentation, object recognition

1. 序 論

近年、セキュリティシステムや自動運転システムなどの実用化に向けて、動画画における一般物体検出の需要が高まっている。監視カメラや車載カメラの映像において物体検出を行うことで危険物の検出や異常の発見が可能となる。そのため監視カメラや車載カメラにおいて物体検出を実現することは事件、事故の防止につながる。

高精度な物体検出を行うためには画像中における物体領域の抽出が重要となってくる。従来広く用いられている領域抽出方法としてスライディングウィンドウがあげられる。スライディングウィンドウでは、一定のウィンドウサイズで画像を走査し物体領域を抽出するため、多種多様な物体を対象とすると、形状やスケールが様々であり抽出矩形の数が膨大になってしまうため現実的ではない。また、混雑状況下や複雑背景下において遮蔽などによる見えの変化も個々の物体の切り出しの難しさの要因の一つとなる。これらの問題を解決するため、様々な領域抽出手法の研究が行われてきた。

物体の領域抽出方法の一つとして視覚的な顕著性を用いて画像中で注目度の高い領域を計算する顕著性マップを用いて領域を抽出するアプローチが数多く提案されている。Chen ら¹⁾の手法は顕著性マップを用いて高精度な物体検出を実現している。しかし、顕著性マップを用いた手法では多数の注目物体が存在するようなシーンや混雑状況下などの注目物体同士が重なり合うシーンにおいては検出精度の低下が起ってしまう。

また、高速な物体領域抽出手法として Binarized Normed Gradients(BING)²⁾があげられる。BING は抽出に用いる特徴量を 2 値化することで計算コストを減らし、高速な物体領域抽出を行う手法である。しかし、BING は高速ではあるが抽出される領域が余分な部分を含んでいたり物体の一部が欠けるといったような領域抽出精度の悪さが問題としてあげられる。

階層的なクラスタリングを用いた物体領域抽出手法に Selective Search³⁾がある。Selective Search は画像を小領域に分割し

たのち、小領域間の色などの類似度を用いて領域の結合を繰り返しその過程で領域を抽出する手法である。物体の形状やスケールに対し頑健な物体領域抽出が可能であり、高い再現率を実現していることが特徴である。しかし問題点として、抽出される領域の数が多く後段の処理に負担がかかることがあげられる。また、複数の色情報を持つ物体に対して抽出精度が落ちることも問題としてあげられる。

これらの物体領域抽出手法は画像中にあるすべての物体を対象とした抽出手法である。しかし、セキュリティシステムや自動運転技術といったシステムへの応用を考えた際重要となるのは動いている物体の検出である。急な飛び出しや不審な動きなどを検出することは事故や事件の防止において大変重要である。そのため、動物体に着目した領域抽出システムの実現が重要である。

階層的クラスタリングを用いた動物体領域抽出手法に Dan らの手法⁴⁾がある。superpixel を時系列に拡張した supervoxel を optical flow を用いて階層的にクラスタリングすることで動物体領域を抽出する手法である。しかし、静止した物体や一部分しか動いていない物体を抽出することができない問題がある。

近年様々な研究において成果を上げている畳み込みニューラルネットワーク (CNN) を用いた動物体領域のセグメンテーションによる動物体領域検出手法⁵⁾も提案されている。しかし、CNN は未知のデータに対応できないことや、多量の学習データを必要とするため限られたタスクにしか応用できないことが問題としてあげられる。

本稿では、Selective Search に optical flow による動き特徴を加えることで動物体に着目した領域抽出を行う。領域統合の際に optical flow をヒストグラムとして表現した特徴量を加えることで、従来手法では抽出を失敗していた複数の色を持つ物体の抽出を可能にする。さらに抽出の際のスコア計算に optical flow の強度を用いることで優先的に動物体の領域を抽出する。また、本手法を機械学習のための学習データの自動生成などに用いることも可能である。

最後に、本稿の構成を記す。第 1 章では、物体検出の研究背景、先行研究、本研究の位置付けについて述べた。第 2 章では、提案手法の処理について、第 3 章では本手法の領域抽出の再現率に関する実験と結果について、第 4 章では結論を述べる。

* 原稿受付 平成 28 年 5 月 8 日

掲載決定 平成 28 年 10 月 14 日

** 慶應義塾大学 (神奈川県横浜市港北区日吉 3-14-1)

† 沖電気工業株式会社 (埼玉県蕨市中央 1-16-8)

Fig.1 System flow

2. 提案手法

提案手法の流れを図1に示す。動画像を入力とし、Felzenszwalbら⁶⁾の手法を用いて領域分割を行う。その後、領域ごとに色、テクスチャ、optical flow、サイズを計算し、それらの値と隣接領域間の包含関係を用いて類似度を算出する。算出された類似度が最も大きくなる領域同士を統合し、統合された領域を動物体領域として抽出する。最終的に一つの領域になるまで統合と抽出を続け、抽出された全領域のスコアを算出する。算出されたスコアが高い領域を最終的な出力とする。

2.1 領域分割

領域分割の研究は幅広く行われているが、本手法ではFelzenszwalbら⁶⁾の手法を用いた。Felzenszwalbらの手法と同様に広く用いられるSLIC superpixel⁷⁾を用いた場合と比較してFelzenszwalbらの手法は領域分割の精度が高く、また処理もSLIC superpixelと同様に高速であることが利点としてあげられる。比較の結果は3章の実験で詳しく述べる。画像の各ピクセルを一つのノード v とし、各ピクセルの8近傍に対して隣り合うピクセルを結ぶエッジ e を考える。ある注目ノードに対し、そのノードとエッジで結ばれるノード間の輝度値の差をそのエッジの重み $w(e)$ とする。各領域 C に対して、領域内の相違度と領域間の相違度を(1)式、(2)式で定義する。

$$\text{Int}(C) = \max w(e) \quad (1)$$

$$\text{Dif}(C_1, C_2) = \min w((v_i, v_j)) \quad (2)$$

v_i, v_j は C_1, C_2 に含まれるノードを示す。これらの相違度算出した後(3)式にしたがって領域間の分離、結合を行っていく。

$$D(C_1, C_2) = \begin{cases} \text{true} & \text{if } \text{Dif}(C_1, C_2) > \text{MInt}(C_1, C_2) \\ \text{false} & \text{otherwise} \end{cases} \quad (3)$$

ただし、 $\text{MInt}(C_1, C_2)$ は(4)式で定義される。

$$\text{MInt}(C_1, C_2) = \min(\text{Int}(C_1) + \tau(C_1), \text{Int}(C_2) + \tau(C_2)) \quad (4)$$

また、 $\tau(C)$ は(5)式で定義される閾値関数である。

$$\tau(C) = k/|C| \quad (5)$$

$|C|$ は領域 C の総ピクセル数を表し、 k は定数である。 k の値が大きくなるほど大きなサイズの領域に分割される。以上の計算式を用いて、ガウシアンフィルタによってスムージングされた画像に対し相違度を算出し小領域に分割する。

Felzenszwalbらの領域分割手法の出力結果を図2に示す。

2.2 領域統合

領域分割の後に各領域間の類似度を計算して統合を行う。これによりあらゆるスケールの物体を抽出することができる。領域間の類似度を比較する際に色 S_{color} 、テクスチャ $S_{texture}$ 、optical flowの方向 S_{flow} 、領域サイズ S_{size} 、領域の包含関係 S_{fill} の5つの指標を用いる。

S_{color} はRGBの3チャンネルに対して0から255までの輝度値を25ビンに分割した、25ビン×3チャンネルの75次元で表現されたヒストグラムをL1ノルムで正規化した値で表現される。 $S_{texture}$ は勾配8方向×10ビン×3チャンネルの240次元で表現されたヒストグラムをL1ノルムで正規化した値で表される。

S_{flow} はFarneback法⁸⁾を用いて算出したoptical flowから生成したヒストグラムの値である。optical flowの計算には、テンプレートマッチングをベースとしたブロックマッチング法や勾配法などがあるが、Farneback法は連続する2フレームにおいて対象物の移動量が微小であることを前提としてoptical flowを求める勾配法から発展した手法である。ピラミッド画像を用いたoptical flowの算出を行っており、他手法と比較して精度と計算コスト両方において提案手法に適した手法であることからoptical flowの計算に採用した。Farneback法によって算出されたoptical flowの方向をいくつかに分割し、小領域内における各ピクセルのフローの方向と対応するビンに値を累積してヒストグラムとし、領域のサイズで正規化した値で表される。分割数が4の時のヒストグラム生成の例を図3に示す。

隣接領域 r_i, r_j の各ヒストグラムの類似度は(6)式で比較される。

$$S(r_i, r_j) = \sum_k^n \min(b_i^k, b_j^k) \quad (6)$$

Fig.2 Output of segmentation

Fig.4 Output of our approach

Fig.3 Optical flow histogram

b はヒストグラムの各ビンの値, n は次元数を示し, ヒストグラムが完全に一致した時最大値 1 となる.

S_{size} は (7) 式で定義され, サイズの小さい領域を優先的に統合するための値である.

$$S_{size}(r_i, r_j) = 1 - \frac{\text{size}(r_i) + \text{size}(r_j)}{\text{size}(img)} \quad (7)$$

size は領域の総ピクセル数を表し, img はフレームの総ピクセル数を表す. 隣接領域間の総ピクセル数が画像の総ピクセル数と一致するとき最小値 0 となる.

S_{fill} は (8) 式で定義され, 二つの領域が包含関係にある場合に優先的に統合するための値である.

$$S_{fill}(r_i, r_j) = 1 - \frac{\text{size}(BB_{ij}) - \text{size}(r_i) - \text{size}(r_j)}{\text{size}(img)} \quad (8)$$

BB_{ij} は二つの領域を完全に囲う矩形を表す. 二つの領域が完全な包含関係にある場合最も大きくなる.

以上の 5 つの指標を (9) 式のように重み $\{a_1, \dots, a_5\}$ を与えて最終的な類似度とし, 最も類似度の高い領域同士を統合する.

$$S(r_i, r_j) = a_1 S_{color}(r_i, r_j) + a_2 S_{texture}(r_i, r_j) + a_3 S_{flow}(r_i, r_j) + a_4 S_{size}(r_i, r_j) + a_5 S_{fill}(r_i, r_j) \quad (9)$$

2.3 領域抽出

領域の抽出は分割された各領域を完全に囲う矩形を対象の領域として行う. 領域抽出は領域分割時と領域統合の度に行われ, 全ての領域が統合され一つの大きな領域になるまで処理が繰り返される.

返される. 抽出の際に背景や静止物体を減らすために, 領域のサイズに対して閾値処理を加えた後 (10) 式を用いて optical flow の強度によるスコア計算を行い, スコアが上位の領域を動物体領域として抽出する.

$$\text{score}(r_i) = R \times \frac{1}{\text{size}(r_i)} \sum_k^p m_i^k \quad (10)$$

m は領域 r_i における各ピクセルの optical flow の強度, p はその領域における総ピクセル数を示す. また, R は 0 から 1 までの一様乱数を表す. 領域内の optical flow の強度が大きいくほどスコアが高くなりやすく, 静止している物体や optical flow がでない背景領域のスコアは低くなる. これにより, 動いている物体に着目した領域抽出が可能である.

optical flow の強度が強く出ている領域を優先的に抽出することでノイズ成分や静止物体の抽出領域数を減らす. また, 図 4 の赤い矩形で囲まれた領域の人物のように, 同様の物体に複数の矩形が表示される場合, 乱数を用いないとそれぞれの矩形のスコアが類似する. それにより, スコア上位の矩形を取り出す際に連続して同じ物体の矩形が抽出されてしまい, 抽出矩形数の増加の原因となってしまう. そのため, 乱数を掛け合わせスコアを散らすことで同一物体の抽出を減らし, 少ない抽出矩形数で高い再現率を実現することが可能である.

3. 実験

3.1 実験概要

本手法の有効性を示すために, 再現率の検証実験を行った. 実験データには UCF Crowd Data Set⁹⁾(解像度: 480×360 , フレームレート: 25fps, フレーム数: 50 フレーム, ground truth 総数: 1924 矩形)と TUD Stadtmitte¹⁰⁾(解像度: 640×480 , フレーム数: 179 フレーム, ground truth 総数: 871)を用い, UCF Crowd Data Set の正解データは手動により作成した. UCF Crowd Data Set は混雑状況を想定しており, 従来手法である Selective Search では抽出が難しい色情報が類似した物体同士が重なり合うシーンや複数の色を持つ物体などが含まれている. TUD Stadtmitte は実環境下での歩行者の検出を想定している. また, 本手法は結果が矩形として出力するため, 正解データと抽出矩形の比較には Intersection over Union(IoU)を用いた. IoU の計算は (11) 式を用い, IoU が 0.5 以上と 0.7 以上の 2 つの場合を正解として検証した. $\text{IoU} > 0.5$ は物体認識の研究において一般的に正解として用いられる値であり, $\text{IoU} > 0.7$ では高精度な認識が可能で

Fig.5 Division example of optical flow

あることからこれら二つの値を採用した。

$$IoU = \frac{\text{正解矩形} \cap \text{抽出矩形}}{\text{正解矩形} \cup \text{抽出矩形}} \quad (11)$$

また、再現率は (12) 式によって計算する。

$$\text{recall} = \frac{TP}{TP + FN} \quad (12)$$

TP は抽出に成功した物体の数, FN は抽出できなかった物体の数を表す。

実験は類似度計算における optical flow の方向の分割数による再現率の検証, 類似度の色と optical flow に対する重み a_1 , a_3 を変化した際の再現率の検証, SLIC superpixel⁷⁾ を用いた場合と Felzenszwalb ら⁶⁾ の手法を用いた際の再現率の比較, (10) 式によるスコア計算の際に乱数を用いた場合, 用いなかった場合による抽出矩形数に対する再現率の検証, 最後に従来手法である Selective Search³⁾ と BING²⁾ との再現率の比較を行った。従来手法は論文³⁾ を参考に追実装したものを用いて比較を行った。

3.2 結果・考察

3.2.1 optical flow の分割数による検証

optical flow の方向の分割数による再現率の検証結果を表 1 に示す。検証は UCF データセットにおいて, (10) 式のスコアが上位の 1000 矩形を用いて行い, (9) 式の類似度の重みは全て 1 として行なった。また, 分割 A は 0 度を起点として分割した場合, 分割 B は 4 分割, 6 分割, 8 分割を 45 度, 30 度, 22.5 度を起点に分割したものを表す。図 5 にそれぞれの例を示す。

結果から, $IoU > 0.5$ においてはどの分割数においても同程度の再現率となっていることがわかる。また, $IoU > 0.7$ においては分割 A, 分割 B どちらにおいても 4 分割における再現率が低くなっていることがわかる。分割数が少ないと移動方向に大きな違いがない物体同士が領域統合において結合しやすくなってしまう。今回評価に用いたデータセットには物体が同じ方向に動くような状況が多く存在したため 4 分割の再現率が低くなったと考えられる。しかし, 分割数を大きくしすぎると optical flow のノイズなどに影響を受け, 同一の物体領域の類似度が下がってしまうく結合できないことによる再現率の低下も考えられる。そのため, シーンに合わせて適切な分割数を選ぶことで再現率の向上が考えられるが, 表 1 に示したように再現率の変動は 2% 前後に収まっており影響はほとんどないと言えるため, optical flow の分割数のチューニングは必要ないと考えられる。

Fig.6 Comparison of recall vs. number of proposals(ss:selective search)

Fig.7 (a)Output of Selective Search (b)Output of our approach. (White bounding boxes are shown in output and red bounding boxes are shown in superior result of our approach)

3.2.2 重みによる検証

表 2 に $IoU > 0.5$, 表 3 に $IoU > 0.7$ における (9) 式の S_{color} に対する重み a_1 と S_{flow} に対する重み a_3 を変化した際の再現率の検証結果を示す。評価には UCF データセットを用いた。結果を見ると, S_{color} の重み a_1 を大きくすると再現率が下がる傾向にある。これは評価に用いたデータセットにビジネスマンが多く, 同様の色情報を持つ物体同士が重なること傾向が強いためと考えられる。同様の理由から S_{flow} の重み a_3 を大きくすると再現率が上がる傾向にある。また, 全体として再現率は両方の値を 1.0 にした場合と比較して多くが 2% 前後の差になっている。チューニングなしでの再現率との差が小さいことと, 今回用いたデータセットが色情報に偏りがあるという特殊な状況であることを踏まえると, 重みの値は 1.0 でも十分な再現率で領域抽出を行うことが可能であると考えられる。そのため, 重みのチューニングも必要ないと考えられる。

3.2.3 領域分割手法による検証

領域分割において代表的に用いられる SLIC superpixel⁷⁾ と本手法で用いた Felzenszwalb ら⁶⁾ の領域分割手法, それぞれを用いた場合の再現率の検証を行った。表 4 に結果を示す。抽出矩形数 1000 矩形, 類似度に対する重みは全て 1.0, optical flow の分割数は 6 分割 A を用いた。また, SLIC の分割数は 1000 とし UCF データセットで評価をおこなった。結果をみると, $IoU > 0.5$ においては SLIC と Felzenszwalb らの手法どちらも同じ値であるが, $IoU > 0.7$ においては Felzenszwalb らの手法の方が再現率が高くなっている。SLIC は入力画像を格子状に分割

Table1 Comparison of division of optical flow

division type		type A			type B		
number of division		4 division	6 division	8 division	4 division	6 division	8 division
recall	IoU > 0.5	69.2%	69.3%	68.5%	68.3%	68.9%	69.1%
	IoU > 0.7	24.4%	25.1%	25.7%	24.8%	26.4%	25.4%

Table2 Comparison of recall % vs. similarity weight (IoU>0.5)

		a_3						
		0.7	0.8	0.9	1.0	1.1	1.2	1.3
a_1	0.7	68.1	69.1	70.2	70.0	70.3	68.8	68.8
	0.8	67.7	68.3	69.4	69.2	69.3	69.0	68.9
	0.9	67.3	67.3	68.4	68.9	69.3	69.7	69.5
	1.0	66.4	67.3	67.8	68.9	68.4	69.2	69.1
	1.1	65.8	67.3	67.7	67.3	68.8	69.1	69.2
	1.2	65.2	66.0	66.9	67.3	68.7	67.6	68.3
	1.3	65.5	65.5	66.7	67.8	67.6	68.5	68.9

Table3 Comparison of recall % vs. similarity weight (IoU>0.7)

		a_3						
		0.7	0.8	0.9	1.0	1.1	1.2	1.3
a_1	0.7	26.1	26.2	27.7	27.9	27.5	26.9	26.7
	0.8	24.6	25.4	26.8	27.1	27.5	26.7	26.6
	0.9	23.7	25.0	26.4	26.8	26.9	27.3	26.7
	1.0	23.7	24.7	26.0	26.4	27.0	26.4	26.7
	1.1	23.6	24.1	25.4	24.8	26.5	26.5	26.4
	1.2	22.4	23.6	24.2	25.0	25.4	25.3	26.0
	1.3	22.1	23.3	23.8	24.3	24.8	25.2	25.7

Table4 Comparison of segmentation method

	SLIC ⁷⁾	Felzenszwalb ⁶⁾
IoU>0.5	68.9%	68.9%
IoU>0.7	22.0%	26.4%

した後位置と色の情報を用いて分割する手法である。格子の分割次第では異なる物体、背景を同様の領域として分割することがあり、抽出の際に余計な領域まで含んで抽出してしまうため、IoU>0.7において再現率が下がったと考えられる。

3.2.4 スコア計算による検証

selective search と (10) 式によるスコア計算の際に乱数を用いた場合と用いない場合の抽出矩形数と再現率の関係を図 6 に示す。比較には UCF Crowd Data Set を用いた。

スコア計算に乱数を用いた場合の方が用いない場合に対して同様の抽出矩形数で高い再現率を実現していることがわかる。乱数を用いない場合、スコア順に矩形を取り出す際に optical flow の強度が似通った同一物体の領域を囲う矩形が連続して抽出されてしまう。それに対し乱数を用いることで、同一物体の領域を囲う矩形を連続して抽出することが削減され、乱数を用いない場合と比較して同程度の再現率をより少ない抽出領域数で実現することができている。

また、抽出矩形数が 200 矩形を超えるあたりまで従来手法が本手法の再現率を上回っていることがわかる。(10) 式によるスコア計算では optical flow の強度が強くて出ているおかつ小さい領域のスコアが高くなる傾向にある。そのため、抽出矩形数が 200 矩形を過ぎたあたりまでは正解とならない小さな領域を抽出してしまい、従来手法の方が再現率が高くなっていると考え

Table5 Comparison of our approach to Selective Search(SS) and BING

			SS ³⁾	BING ²⁾	our approach
recall	UCF ⁹⁾	IoU > 0.5	46.9%	86.6%	68.9%
		IoU > 0.7	11.9%	14.0%	26.4%
	TUD ¹⁰⁾	IoU > 0.5	70.8%	90.6%	84.1%
		IoU > 0.7	30.1%	24.6%	45.9%

られる。しかし、抽出矩形数がそれ以上の数になると本手法の方が再現率が高く、また収束も早くなっている。これは、optical flow の強度が強く領域サイズが小さい領域の数が全体の抽出矩形数に対し少ないことや、それらの領域が乱数により確率的に小さなスコアをとること、動物体を抽出する上で optical flow の情報が効果的に働いたことなどによると考えられる。

3.2.5 従来手法との比較

従来手法²⁾³⁾と本手法の再現率比較結果を表 5 に示す。再現率の比較には抽出された全矩形を用い、また抽出矩形数 1000 矩形、提案手法の類似度の重みは全て 1、optical flow の分割数は 6 として行い、評価は UCF データセットと TUD データセットで行った。

Selective Search と比較して、IoU>0.5、IoU>0.7 において UCF crowd dataset ではおよそ 15% から 20% の再現率の向上をすることができた。出力結果を見ると図 7(b) の左赤枠の人物に示したような身体と頭部の色が大きく異なるような人物に対し、従来手法は身体と頭部をうまく結び付けられていないが本手法は頭部まで含めてうまく抽出することができていることがわかる。また同様に、図 7(b) の右赤枠に示したような類似した色情報を持つ物体同士が重なり合ったような状況においても抽出することができている。これにより再現率が向上したと考えられる。

また、BING は IoU>0.5 において高い再現率を実現しているが IoU>0.7 においては IoU>0.5 と比較して 70% 近く再現率が低下しており、これは BING が領域の抽出を正確に行えていないことを意味する。BING はバイナリ値で表された低次元特徴を用いた領域抽出を行うため人物の正確な位置が捉えにくいという欠点があり、IoU の閾値を高くすると再現率が大きく低下してしまう原因となる。しかし、UCF Crowd Data Set では人物が密集しているため、正確な位置を捉えられなくてもおよそその領域に矩形をあてれば正解となる可能性が高く、本手法と比較して再現率が高くなった要因と考えられる。より実環境に近い TUD stadtmittel においては本手法とおおよそ 6% 程度の差しかなく、また IoU>0.7 においては本手法の再現率が BING より大きく上回っている。また、TUD データセットは 1 フレームあたりの正解矩形が平均 5 矩形であることを踏まえると、6% の再現率の低下は手法として大きく劣る数値ではないため、本手法の有効性がうかがえる。

また、図 6 から Selective Search の全抽出領域に対する再現率を本手法は 600 から 700 領域程度で実現できている。Selective Search の全抽出領域数はアルゴリズムの関係上本手法と同数

になるため、抽出数を半分以下に抑えることができたと考えられる。

最後に従来手法と本手法の出力結果の一部を図 8 に示す。出力結果において、Selective Search と比較して未検出を示す緑の矩形が少なくなっている。また、IoU>0.7 において BING と比較すると本手法の検出成功数が大きく向上していることがわかる。以上の結果からも本手法の有効性を示せる。

4. 結 論

本稿では、監視カメラ映像において Felzenszwalb ら⁶⁾の領域分割手法と optical flow を考慮した領域統合、スコア計算による動物体領域抽出手法を提案した。これにより、従来手法で課題となっていた複数の色を持つ物体の抽出に成功した。

提案手法を実装し、混雑状況下の監視カメラ映像を用いた従来手法との比較実験や、パラメータを変えることによる再現率の変化、抽出矩形数による再現率の変化の検証などにより本手法の有効性を示し、従来手法である Selective Search と比較して上位 1000 矩形における再現率を UCF Crowd Data Set において約 20%、TUD Stadtmitte において約 15% 向上することができた。それと同時に従来手法と同様の再現率をより少ない抽出矩形数で実現することができた。

問題点として、類似した色情報を持つ動物体が重なり合った場合における領域分割の失敗があり、領域抽出の再現率の低下が起こる。この問題は領域分割に optical flow を用いることで解決ができると考えられる。しかし、optical flow の計算コストや分割領域数の増加による領域統合時の計算量の増加が予想されるため、処理速度の改善も同時に行っていく。また、optical flow の強度が強く出ている小さい領域のスコアが高くなる傾向にあるため、抽出枚数を極端に少なくすると従来手法より再現率が低くなる傾向にある。同時に動きがほとんど無い動物体の

スコアが低くなる傾向にあるため抽出枚数がある程度大きくしないとそのような物体の抽出が難しいことも問題としてあげられる。そのため、optical flow の強度以外の情報を用いたスコア計算の検討を行っていきたい。

参 考 文 献

- 1) Ming Cheng, Niloy J Mitra, Xumin Huang, Philip HS Torr, and Song Hu. Global contrast based salient region detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 37, No. 3, pp. 569–582, 2015.
- 2) Ming-Ming Cheng, Ziming Zhang, Wen-Yan Lin, and Philip Torr. Bing: Binarized normed gradients for objectness estimation at 300fps. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pp. 3286–3293. IEEE, 2014.
- 3) Jasper RR Uijlings, Koen EA van de Sande, Theo Gevers, and Arnold WM Smeulders. Selective search for object recognition. *International journal of computer vision*, Vol. 104, No. 2, pp. 154–171, 2013.
- 4) Dan Oneata, Jérôme Revaud, Jakob Verbeek, and Cordelia Schmid. Spatio-temporal object detection proposals. In *European Conference on Computer Vision*, pp. 737–752. Springer, 2014.
- 5) Katerina Fragkiadaki, Pablo Arbeláez, Panna Felsen, and Jitendra Malik. Learning to segment moving objects in videos. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4083–4090. IEEE, 2015.
- 6) Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, Vol. 59, No. 2, pp. 167–181, 2004.
- 7) Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 34, No. 11, pp. 2274–2282, 2012.
- 8) Gunnar Farneback. Two-frame motion estimation based on polynomial expansion. In *Image Analysis*, pp. 363–370. Springer, 2003.
- 9) Saad Ali and Mubarak Shah. A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pp. 1–6. IEEE, 2007.
- 10) Mykhaylo Andriluka, Stefan Roth, and Bernt Schiele. Monocular 3d pose estimation and tracking by detection. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 623–630. IEEE, 2010.

Fig.8 output result(upper:IoU>0.5, lower:IoU>0.7)Blue bounding boxes are shown in true positive, Ground truth bounding boxes are shown in Green and Red. Red bounding boxes were found and Green bounding boxes were not.(left:Selective Search center:BING right:our approach)