

平成 27 年度卒業論文

画像変換を用いた多数の候補領域抽出と
検証による物体検出

宮崎大学 工学部 情報システム工学科

赤松 龍太

指導教員 古谷博史, 椋木雅之

目次

1	はじめに.....	2
2	複雑背景における物体検出.....	3
2.1	複雑背景とは.....	3
2.2	従来の物体検出研究.....	4
2.3	複雑背景での物体検出.....	5
3	多数の候補領域抽出と検証による物体検出.....	6
3.1	画像変換による多数の候補領域抽出.....	6
3.2	物体検出の流れ.....	7
3.3	候補領域の抽出.....	8
3.3.1	検出器の学習.....	8
3.3.2	検出器を用いた候補.....	11
3.4	候補領域の検証.....	11
3.4.1	識別器の学習.....	12
3.4.2	識別器を用いた検証.....	13
4	物体検出実験.....	15
4.1	検出対象の選定.....	15
4.2	実験手順.....	15
4.3	実験結果.....	20
4.4	一般背景での比較実験.....	22
4.4.1	比較実験の手順.....	22
4.4.2	実験結果.....	23
4.4.3	考察.....	24
4.5	画像変換の有効性評価.....	25
5	おわりに.....	30

1. はじめに

近年，レスキューロボットによる被災者検出や，運転中の衝突回避システムなど物体検出分野の研究が多くされてきている．しかし，一般背景に対する物体検出は多いが，複雑背景に対して物体検出を行う研究はまだ少ない．ここで複雑背景とは，検出したい物体に対してまぎらわしい背景であり，一方，一般背景とは，検出したい物体と区別が付きやすい背景のことを指す．実生活において複雑背景における物体検出の研究が求められている．例えば海外からの輸入品に虫やごみがついていないか検品する場面や，森の中で人物を検出する場面などは複雑背景における物体検出となる．複雑背景での物体検出は一般背景に比べて困難と考えられる．本研究では，多数の候補領域抽出と検証処理を組み合わせることで，複雑背景にも適用できる物体検出を目指す．

2. 複雑背景における物体検出

2.1 複雑背景とは

本研究で扱う複雑背景とは、検出したい物体と類似した物体が多く含まれるような背景である。例えば、人物検出の場面での木の幹が多く含まれる背景や、擬態している昆虫検出の場面での、昆虫によく似た樹皮や葉などの背景である。物体が多数存在し、そこから目的の物体を検出するという意味での複雑背景からの物体検出ではなく、検出したい物体とまぎらわしい背景が、本研究の複雑背景の定義である（図1）。



図1 複雑背景の例

2.2 従来の物体検出研究

従来から、物体検出の研究が多く行われてきている。例えば[1]では人物の顔検出、[2]では、自動車検出が行われている。物体検出においては、検出対象となる物体が画像内で回転や拡大・縮小、明るさが変化して映されていると、一般に検出が困難となる。[3]の研究は、画像の明るさによる物体検出の困難性に着目した研究である。明るさによる未検出に対して **RRF(Radial Reach Filter)**と呼ばれる手法を提案し対処している。[4]の研究は、画像の回転・拡大縮小による物体検出の困難性に着目した研究である。**HOG(Histograms of Oriented Gradients)**と呼ばれる特徴量と、**Log-polar 変換**と呼ばれる手法を合わせることにより、認識対象の回転や大きさの変化に対応した手法を提案している。このように従来から物体検出の研究が多く行われているが、それらは一般背景を想定しており、複雑背景において手法を評価、検討していたものはなかった。

2.3 複雑背景での物体検出

従来の手法を複雑背景に適用すると、まぎらわしい背景からの誤検出が多くなる。また、検出対象の大きさの違いや回転など様々な要因により検出漏れが生じる。

これに対処するために、本研究では、まず回転や拡大・縮小といった様々な要因を考慮して画像変換を行い、多数の候補領域を抽出する。これにより検出漏れを少なくする。一方で、誤検出は非常に多くなる。そこで画像内には必ず対象物体が1つのみ存在するという仮定をおき、これらの誤検出を減らすために多数の候補領域の中から、最も確からしいもののみを選択する。これにより、複雑背景下での物体検出の性能向上を目指す（図2）。

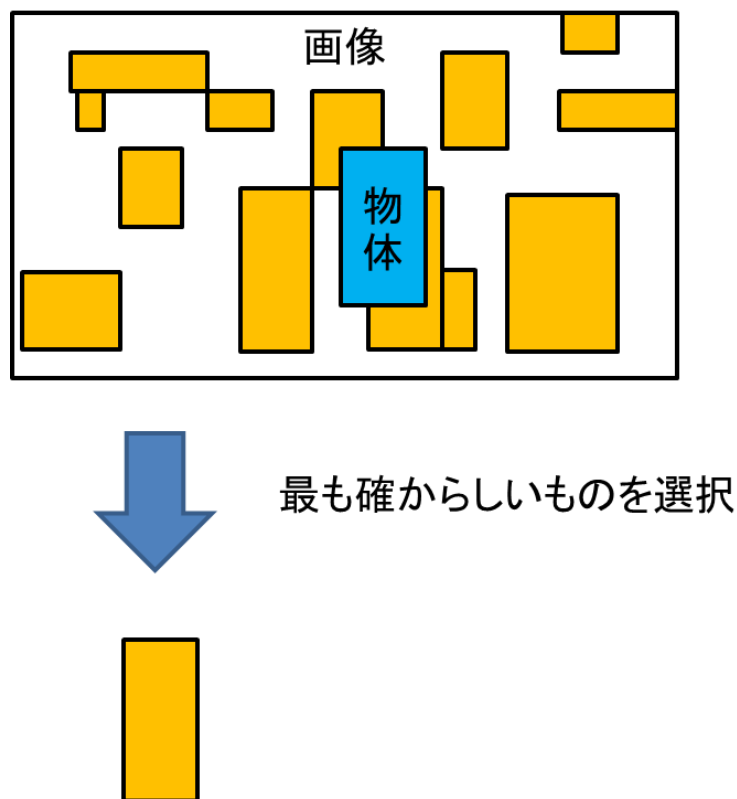


図2 複雑背景での物体検出

3. 多数候補抽出と検証による物体検出

3.1 画像変換による多数候補抽出

複雑背景での物体検出を困難にしている要因は，検出したい物体の周りにそれと類似した物体もしくは背景があること，回転や拡大縮小，明るさの変動など様々である．これに対処するために，検出を行いたい入力画像に画像変換を施したものからも物体検出を行う．要因毎に対応する画像変換を行うことで，その要因に対処しやすくする．例えば，入力画像を回転させた画像から物体検出を行えば，画像中で回転していた対象物体の検出が行いやすくなる．どのような要因が複雑背景での物体検出を困難にするかは，自明ではない．本研究では，候補抽出数の増加に効果的であると考えられる画像変換を複数用意し，それらすべてを適用することで，対象物体の検出漏れを少なくする．

3.2 物体検出の流れ

処理の流れは、大きく2段階に分かれる(図3)。1段階目は、検出したい物体の候補領域の抽出である。ここでは、複雑背景における物体検出という問題に対して、求めている物体を漏れなく抽出することを優先し、他の誤った部分が抽出されても可とする。

2段階目は候補領域の検証である。1段階目において抽出した候補領域に対して、それが検出対象の物体であるかの検証を行う。ここでは、検出対象の物体が候補の中に必ず1つ含まれていると仮定し、最も確からしいもののみを選択する。これらの2段階の処理を行うことにより目的物体の検出を行う。

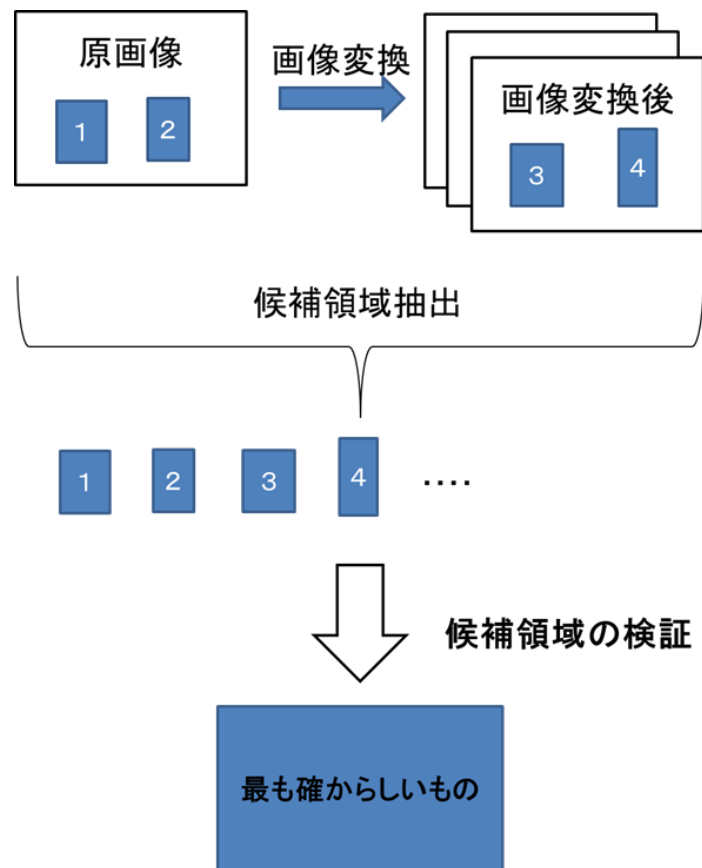


図3 物体検出の流れ

3.3 候補領域の抽出

候補領域を抽出するために，物体の検出器を作成する．その際に，検出したい物体の事例を計算機に与え自動で学習させることにより物体検出器を作成する．

3.3.1 検出器の学習

正事例（検出したい物体の画像）と負事例（検出したい物体が写っていない画像）を用いて物体検出器の学習を行う．学習には[6]と同様の LBP 特徴量と AdaBoost を組み合わせた手法を用いる．

LBP(local binary patterns)特徴量とは，明るさの変化の影響を受けにくい局所特徴量である（図 4）．画像内の 3×3 領域の画素の中で，中心画素とその周辺画素（縦横斜めの 8 か所）を用いて計算される．中心画素より大きい点は 1，中心画素以下の点は 0 とし，左上から時計回りに 1 画素について 8 ビットで表現される．

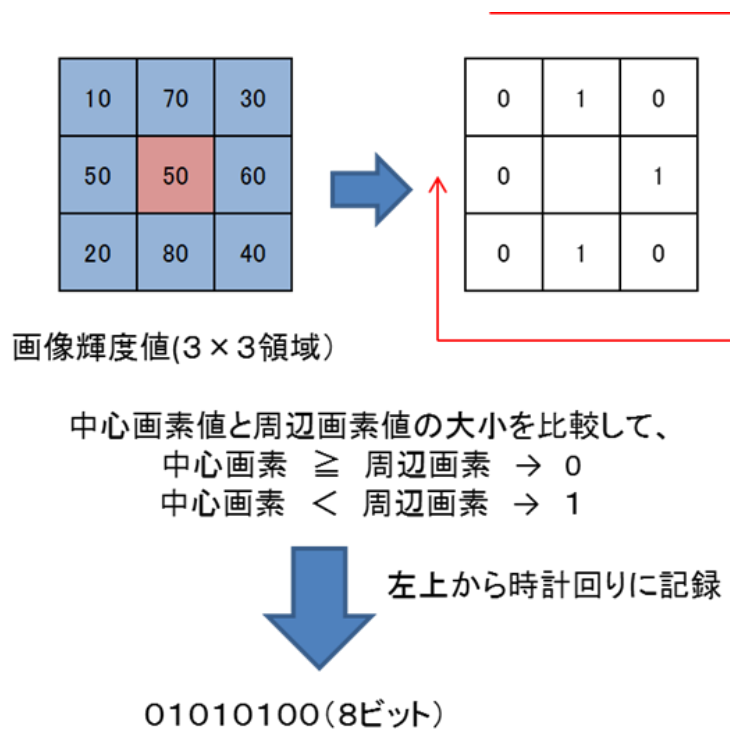


図4 LBP 特徴量の計算

まず各画素に対して LBP 特徴量を計算する。次に、画像を M 個のブロックに分け、それぞれのブロックで LBP 特徴量のヒストグラムを計算する。それにより $256(8\text{bit}) \times M$ 個の弱検出器ができる。それらの精度の悪い多数の弱検出器を AdaBoost で統合することにより強検出器である物体検出器を作成する。AdaBoost の流れを以下に示す。実装には、openCV の `opencv_traincascade` を用いた。

AdaBoost の流れ

N 個の学習サンプルがありそれぞれには正解 1, 非正解-1 のラベル付け $y_i (i = 1 \dots N)$ が与えられているとする.

step1 各サンプルの重みを $D_1(i) = \frac{1}{N}$ で初期化する.

step2 For $t = 1, \dots, T$:

(A) サンプル重み D_t において, 弱検出器 $h_t(x)$ を学習する. 学習サンプルに対する重み付き誤り率

$$\varepsilon_t = \sum_{i: y_i \neq h_t(x_i)} D_t(i)$$

が最小となる $h_t(x)$ を選択する.

(B) 誤り率から信頼度 α_t を計算する.

$$\alpha_t = \frac{1}{2} \log \left(\frac{1 - \varepsilon_t}{\varepsilon_t} \right)$$

(C) サンプルの重みの更新を行う. 弱検出器で正しくラベル通りに識別できたサンプルの重みは小さく, 間違えたサンプルの重みは大きくする. またこのとき $\varepsilon_t = 0.5$ の場合は $\alpha_t = 0$ となり重みの更新がされないため学習をする意味がなくなる.

(D) サンプルの重みの和が 1 になるように正規化を行う.

step3 最終的な強検出器 $H(x)$ は, 得られた弱検出器を信頼度で重み付けることで作成する.

$$H(x) = \text{sign} \left[\sum_{t=1}^T \alpha_t h_t(x) \right]$$

3.3.2 検出器を用いた抽出

学習により得られた検出器を用いて，物体の候補領域の抽出を行う．画像中から領域を切り出して，学習時と同様に LBP 特徴量を計算し，ブロック毎にヒストグラムを作成する．この特徴量を作成した検出器 $H(x)$ に与える．

$H(x) = 1$ のとき対象物体， $H(x) = -1$ のとき非対象物体と判定する．

検出器は，学習に用いた画像と同じ大きさの物体しか抽出できない．物体の拡大・縮小に対処するために，検出最小サイズを定めて，画像を一定の割合ずつ小さくしていき，学習させた検出器により候補領域を探索し，候補領域の出力を行う．実装には，openCV の `detectMultiScale` を用いた．

検出漏れを少なくするために，原画像と画像変換後の画像に検出器を適用し，出てきたものすべてを候補領域とする．

3.4 候補領域の検証

候補領域の検証を行うために，識別器を作成する．識別器は 1 段階目と同じように計算機にあらかじめ検出したい物体の事例を与え，学習により作成する．

3.4.1 識別器の学習

正事例と負事例を用いて、物体識別器を作成する。画像全体からその画像の特徴量を計算する。特徴量としては、Caffe[5]により算出した4096次元の特徴量を用いる。識別器の学習には、SVM(Support Vector Machine)を用いる。

Caffeとは、人の脳を模した深層学習(Deep Learning)を応用した機械学習のライブラリである。入力層にデータを与え、隠れ層で様々な角度から学習し、出力層でそのデータが何であるかを判別する(図5)。Caffeを用いて出力層の1つ前の隠れ層のデータである特徴量をとることができ、それをCaffe特徴量という。

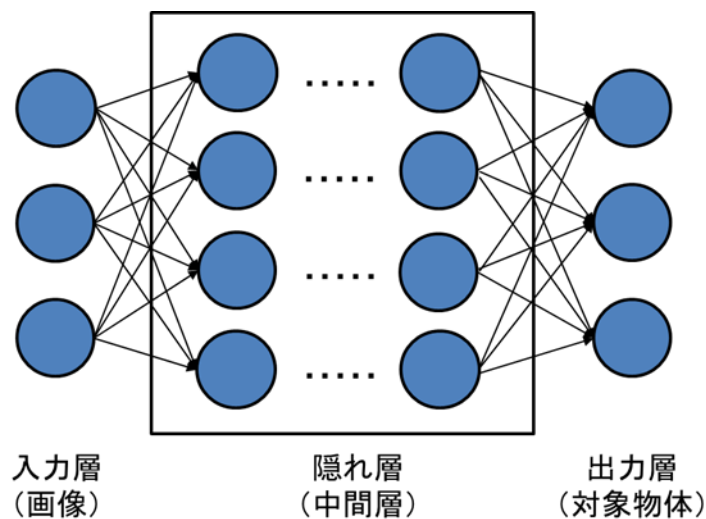


図5 深層学習

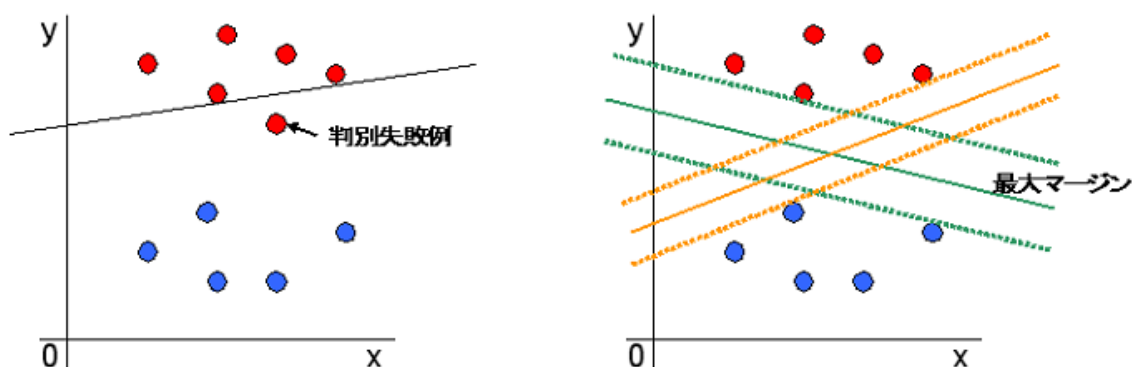


図 6 SVM の最大マージンについて([7]より引用)

対象物体の識別を行うために SVM を用いる。SVM とは教師あり機械学習の手法であり、高いパターン識別能力をもっている。対象物体と対象物体以外の境界線を引く際に、図 6 の左のように境界線を引くとうまく識別が行う事ができない。SVM を用いると、可能な境界線の中から対象物体とそれ以外との距離（マージン）が最も大きくなる線を引くことができる（図 6 右）。また、カーネルトリックと呼ばれる手法により、非線形な境界線を求めることができる。

3.4.2 識別器を用いた検証

まず、抽出した候補領域それぞれに対して Caffe[5]により特徴量を計算する。そして、その特徴量に対して学習済みの SVM を用いて候補領域の対象物らしさを算出する。抽出したすべての候補領域の中で、最も値が大きい候補領域を検出結果とする（図 7）。実装には LIBSVM を用いた。LIBSVM では、識別結果として、対象物らしさの確率も得られる。

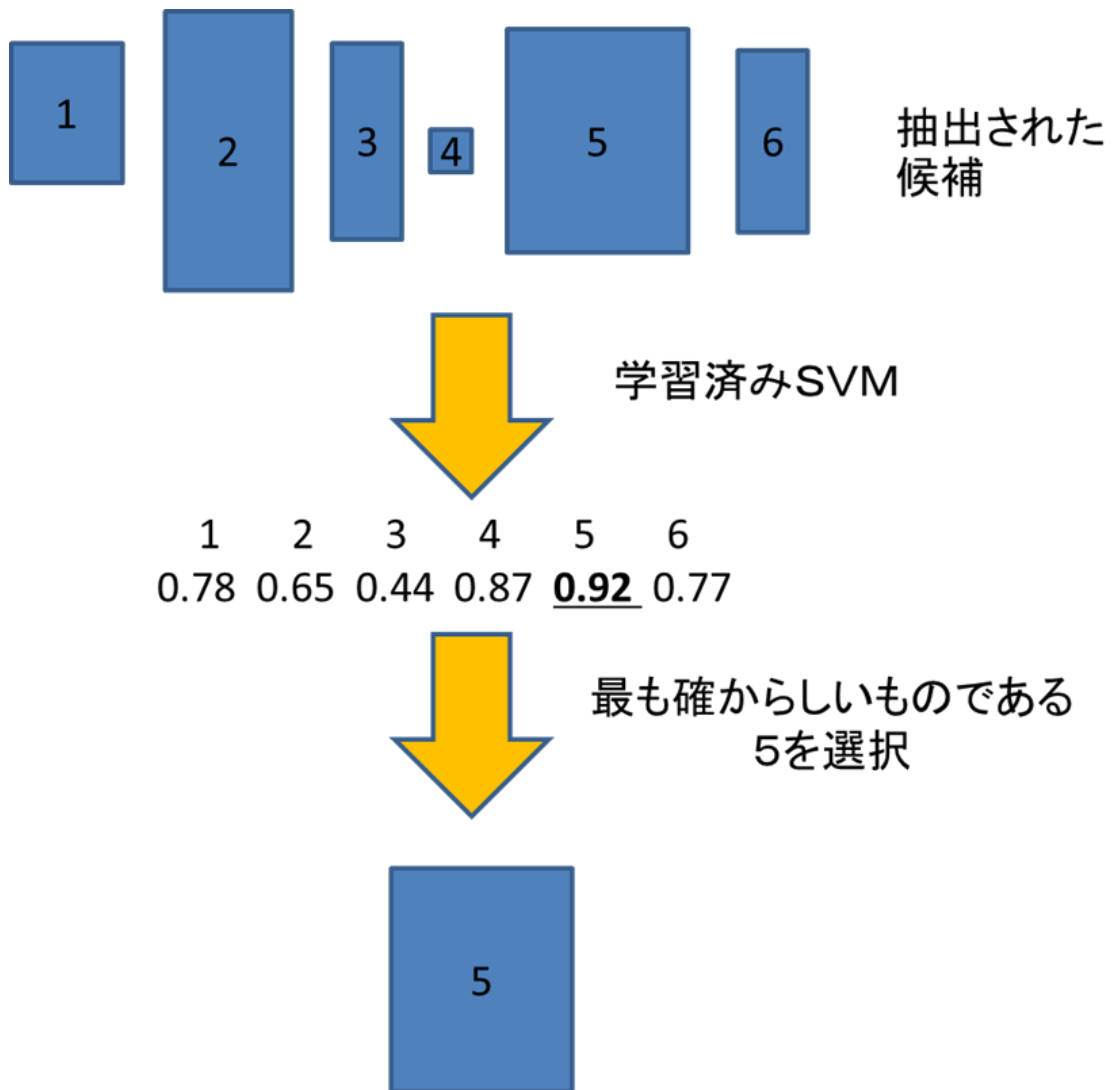


図7 識別器を用いた検証

4. 物体検出実験

4.1 検出対象の選定

今回選んだ検出対象の物体はアブラゼミである。アブラゼミは木に擬態しており、木に止まっているセミは複雑背景にある物体とみなせるため、本研究での検出したい対象として適している。また他の擬態昆虫に比べて写真を撮る際に写る向きに変化が少ないため、学習に必要な事例数が比較的少なくすむ。

4.2 実験手順

1段階目の検出器の学習では、Web上からダウンロードしたセミの画像100枚（図8）からセミの写っている部分を切り出した100領域を正事例とした（図9a）。また、同じ100枚の画像中で、セミが写っていない木の幹の部分を人手で切り出した50領域を負事例とした（図9b）。2段階目の識別器の学習では、1段階目と同じセミの領域100個を正事例、Web上からダウンロードした風景画像100枚を負事例として用いた。識別器であるSVMのカーネルには、ガウシアンカーネルを使用した。



図 8 学習用の原画像



a) 正事例



b) 負事例

図 9 学習画像

評価には、手持ちのカメラにより屋外で撮影したセミの画像 226 枚を用いた。これらの評価用画像に対してセミをちょうど含む矩形を正解領域として人手で与えた。

候補領域の検出では、セミの検出漏れを少なくするために評価用画像に対して、19 種類の画像変換を行った。各処理について以下に示す。

- **モノクロ**

画像をモノクロ化させる処理.

- **補色変換**

画像内の色をその色と補色関係にある色と置換させる処理.

- **コントラスト強調**

画像の明暗の輝度の差を強調する処理.

- **エンボス処理**

画像の濃淡差を用いて輪郭部分を立体化させる処理.

- **均等化 (ヒストグラム平坦化)**

各画素をヒストグラム化したものに対して、画素が集中している部分を平坦化させることにより見やすい画像にする処理.

- **スムージング (別名平均化フィルタ)**

中心画素と一定範囲の周辺画素の輝度値の平均をとり中心画素値をその平均とする処理. 画像のノイズなどを除去するために用いる. 周辺画素の範囲は 3x3,5x5,7x7 の 3 種類を用いる.

- **回転**

画像を左右に 90 度ずつ傾ける処理.

・階調変換

256階調で表現されている画像の階調数を減らす処理。128,64,32,16階調の4種類を用いる。

・メディアン

中心画素と一定範囲の周辺画素の輝度値を昇順に並べその中央値を中心画素値とする処理。周辺画素の範囲は3x3,5x5,7x7の3種類を用いる。

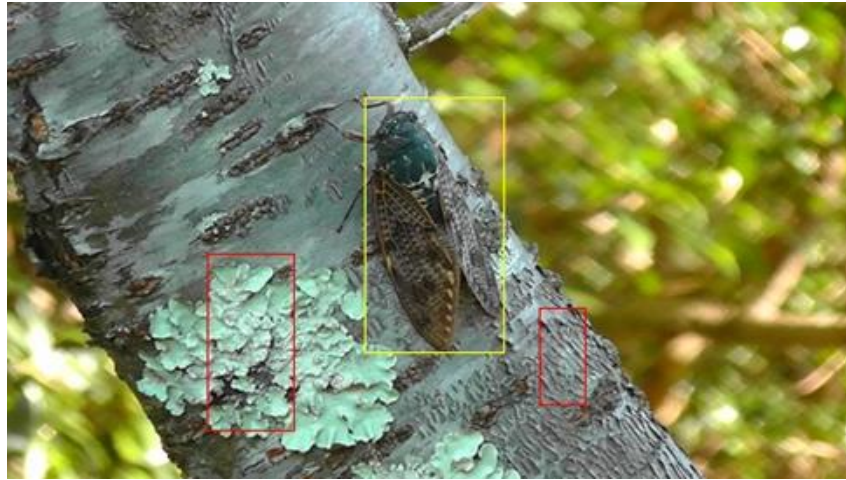
・拡大縮小

拡大は、画像のサイズを2倍にする処理。縮小は、画像のサイズを半分にする処理。

上記の19種類の画像変換を行い、セミ検出器を用いてセミの候補領域抽出を行った。

得られた候補領域に対して、識別器を用いて、最も確からしいものを検出結果とした。人手で与えた正解領域をA、検出された候補領域をB、AとBの重なり領域をCとし、式(1)が50%以上の場合検出できたとした。

$$\frac{C}{A+B-C} \quad (1)$$



a) 原画像



b) 画像変換の一例(階調変換)

図 10 候補領域の抽出例

表 1 複雑背景での実験結果

候補抽出	原画像のみ	画像変換後
候補領域数	3564	68911
正解検出数 (総数226枚)	117	216
検証		
最終検出数(総数226枚)	103	164
最終検出率	45.6%	72.6%

4.3 実験結果

原画像から、候補抽出した結果の一例を図 10 a), 原画像に 19 種類の 1 つである階調変換処理を施して検出した結果を図 10 b)に示す. 黄色の枠は人手で与えた正解領域で, 赤色の枠がセミ検出器による物体候補領域である. 図 10 a)の例では, 原画像に対して検出をするだけではセミを検出できなかったが, 階調変換処理を施すことによって検出できた.

226 枚×(19+原画像)に対してセミ検出器をかけた結果を表 1 の上段に示す. 画像変換処理を施すことでセミの検出数が約 2 倍に上がった.

次に, ここで得られた候補領域に対して, 2 段階目の検証を行った結果を表 1 の下段に示す. 画像変換を行った結果, 検出率は原画像のみに比べ, 27.0 ポイント向上した. 複雑背景において, 画像変換を行うことにより検出性能が向上することが分かった.

図 11 は SVM によるセミであるかの確率を表したものである. オレンジ色の点が正解で, 青色の点が非正解である. 一定確率以上の正解数, 非正解数を表 2 に示す. また, 図 12 は, 適合率と再現率を表したものである.

適合率は,

$$\frac{\text{正解数}}{\text{正解数} + \text{非正解数}}$$

再現率は,

$$\frac{\text{正解数}}{\text{画像総数}(226)}$$

により求める.

高い適合率を要求すると正解数の割合は大きくなるが、そもそもの検出数が減る。一方、高い再現率を要求すると検出数は増えるが、非正解数の割合も増える。

表 2 確率と正解数, 非正解数

確率(以上)	正解数	非正解数
0.9	37	3
0.8	60	7
0.7	75	11
0.6	92	11
0.5	97	12
0.4	115	14
0.3	127	20
0.2	137	27
0.1	150	40
0	164	62

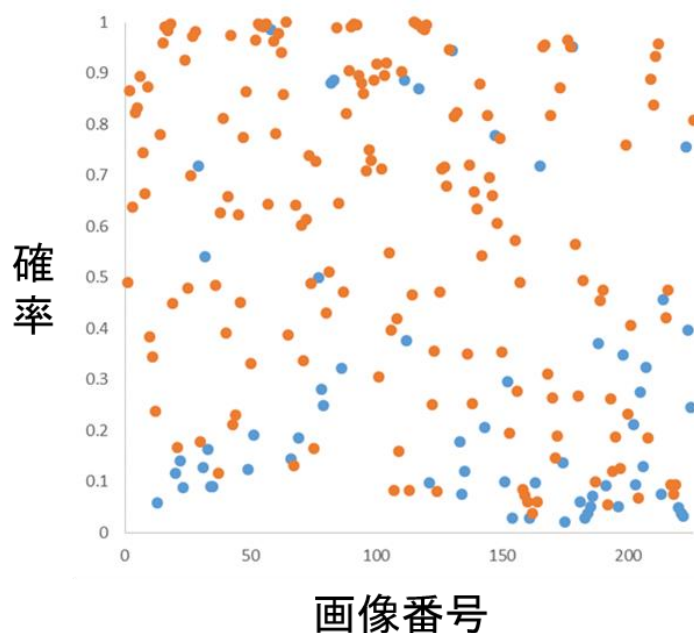


図 11 最も確からしい画像として選択された画像の番号と SVM による確率値

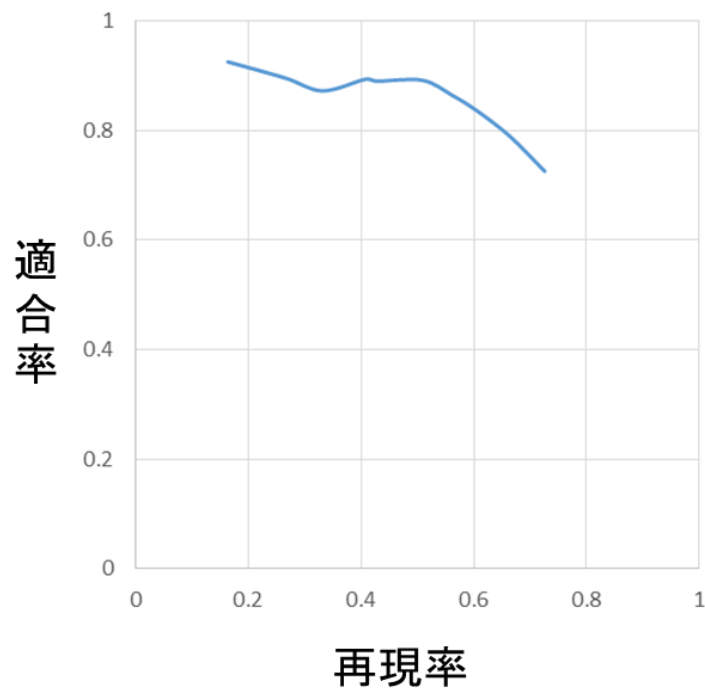


図 12 適合率と再現率

4.4 一般背景での比較実験

本研究では、特に複雑背景を想定して手法を考案した。検出性能の向上が複雑背景特有であるか評価するために、一般背景に対する比較実験を行った。

4.4.1 比較実験の手順

セミの評価用原画像のセミ以外の部分を一般背景に変え、同じ画像変換、同じ検出器・識別器を適用する。一般背景として、セミと比較的区別が付きやすい 12 種類の草原画像を用意した。この画像に対して、4.2 節のセミの正事例 100 領域を複雑背景の時と同じ位置、大きさに貼りつけた画像を作成し、一般背景の画像とした。作成した画像の例を図 13, 図 14 に示す。

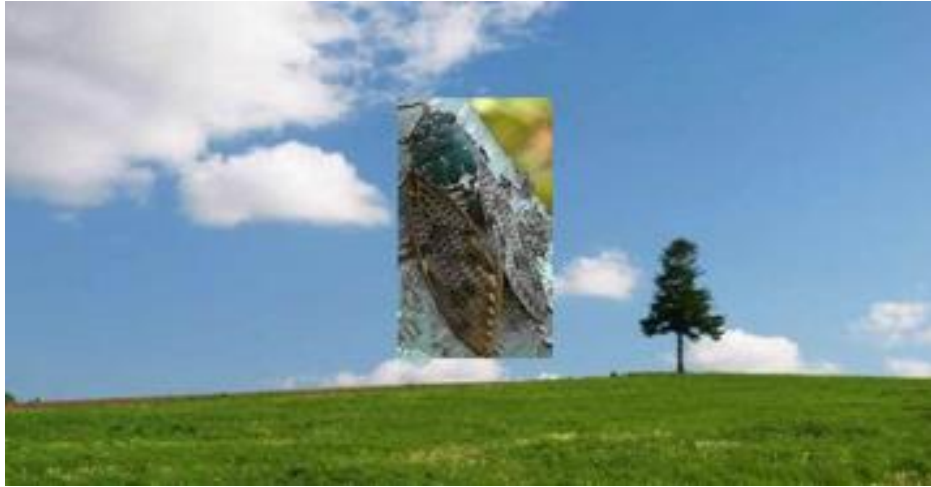


図 13 作成した一般背景画像の例(1)



図 14 作成した一般背景画像の例(2)

4.4.2 実験結果

一般背景における候補領域抽出の結果を，表 3 の上段に示す．複雑背景と同じく検出数は約 2 倍になった．次に，候補領域を検証して得られた物体検出結果を表 3 の下段に示す．検出率は 30.0 ポイント向上し，複雑背景よりも一般背景の方が若干検出率はよくなった．

表 3 一般背景での実験結果

候補抽出	原画像のみ	画像変換後
候補領域数	2694	65663
正解検出数 (総数226枚)	123	224
検証		
最終検出数 (総数226枚)	112	180
最終検出率	49.6%	79.6%

4.4.3 考察

複雑背景に対する物体検出の実験結果から、画像変換を行うことで検出率が増加し、複雑背景に対して効果があることが分かった。しかし、比較実験として一般背景に対して同じ処理を行った結果、複雑背景の場合と同じように検出率は増加した。つまり、本手法は、複雑背景だけでなく、一般背景に対しても検出率が上がることが分かった。

表 4 画像変換により検出できるようになった画像数

変換種類	1	2	3	4	5	6
画像数	22	16	29	18	34	35
7	8	9	10	11	12	13
15	10	26	29	29	22	32
14	15	16	17	18	19	
22	12	1	2	44	24	

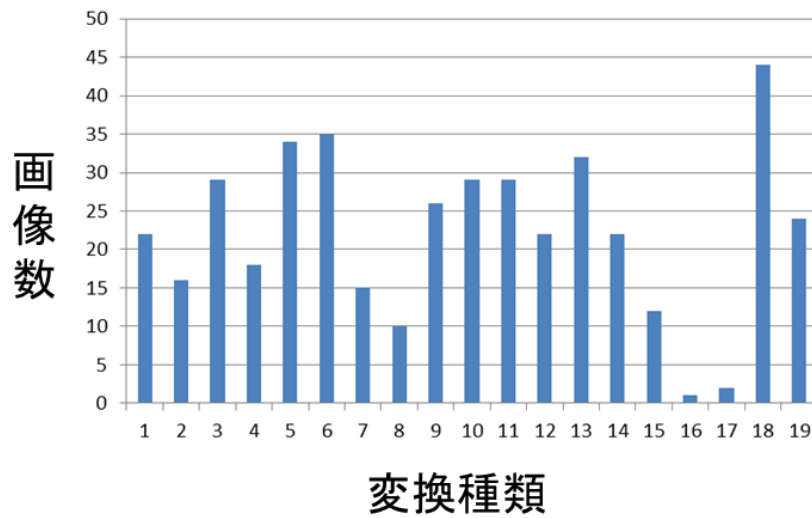


図 15 画像変換により検出できるようになった画像数

表 5 画像変換種類

1	モノクロ	11	階調変換(64階調)
2	補色変換	12	階調変換(128階調)
3	コントラスト強調	13	median(3x3)
4	エンボス処理	14	median(5x5)
5	均等化	15	median(7x7)
6	スムージング(3x3)	16	回転(右回転90度)
7	スムージング(5x5)	17	回転(左回転90度)
8	スムージング(7x7)	18	拡大(200%)
9	階調変換(16階調)	19	縮小(50%)
10	階調変換(32階調)		

4.5 各画像変換の有効性評価

画像変換により検出率が上がることが分かった。そこでどの画像変換が有効であるのかを評価するために、候補抽出の段階でそれぞれの画像変換ごとに原画像のみと比べて新たに検出できるようになった数を集計した。結果を表 4、図 15 に示す。表中の変換種類の番号に対応する処理を表 5 に示す。表 4、図 15 より、検出数が一番増えているのは拡大で、次がスムージング(3x3)や均等

化，メディアン処理(3x3)の場合である．変換種類と総検出数を表 6 と図 16 に示す．図 16 中の 0 は原画像である．表 4，図 15 と見比べると，最も検出できるようになった画像数が多いのが拡大であるが，総検出数も増えている．拡大では，画像を拡大することにより 3.3.2 節における領域抽出の際の「一定の割合」が疑似的に細くなり，候補抽出数が増えたことが要因だと考えられる．スムージングや均等化，メディアン処理については，背景とセミの微妙な差異が処理を施すことで強調されるようになり，検出できるようになるのではないかと考えられる．この 3 つの画像変換画像と原画像の LBP 特徴量の差分を表した画像を図 17~19 に示す．図 17~19 において赤線でくくっているところがセミのいる場所である．少しではあるが，セミの部分が他の部分と異なっている．つまり，画像変換によってセミが強調され，検出器によるセミの抽出向上につながったといえる．また，均等化は，検出数が増えたのに比べて総検出数はあまり変わらず，スムージング(3x3)，メディアン(3x3)は検出数が増えたのに比べて総検出数は減っており，検出数を考えた場合，これら 3 つは画像変換として効果があるといえる．モノクロ，エンボス処理は総検出数が 2 倍以上になったのに対して，検出できるようになった画像数は，最も検出数の多い拡大に比べてそれぞれ 0.5 倍，0.41 倍となっており，検出数を考えた場合，あまり効果的ではないと言える．

表 6 変換種類と総検出数

変換種類	原画像	1	2	3	4	5
総検出数	3564	8744	3727	3549	7664	3465
6	7	8	9	10	11	12
2335	1286	906	4007	3678	3618	3585
13	14	15	16	17	18	19
2442	1399	900	2683	2692	7957	710

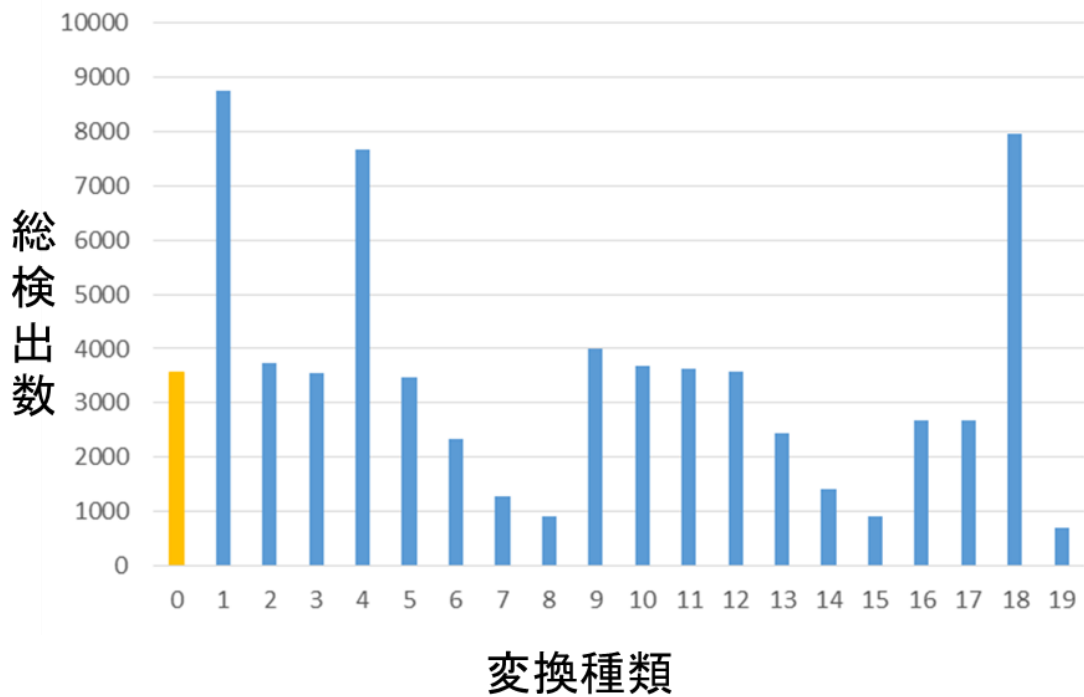


図 16 変換種類と総検出数

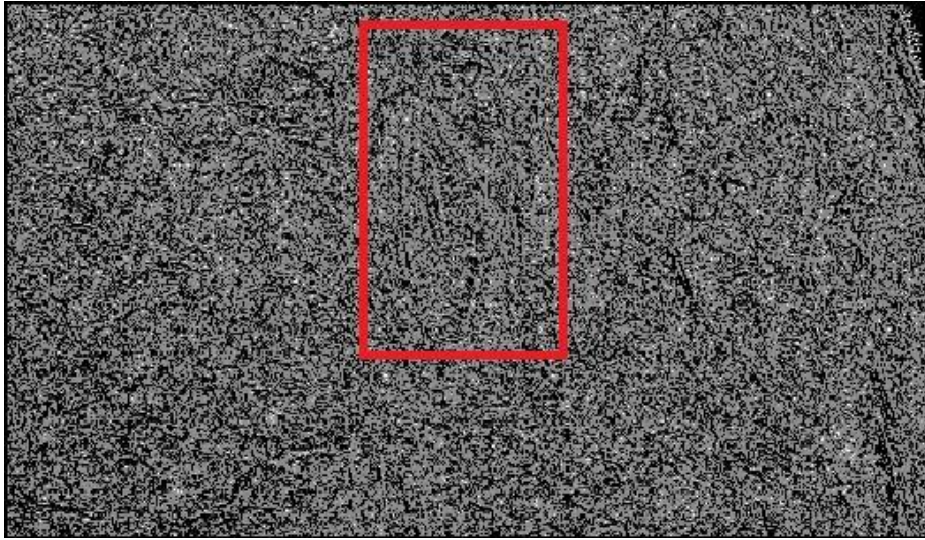


図 17 LBP 特徴量差分(原画像とスムージング)

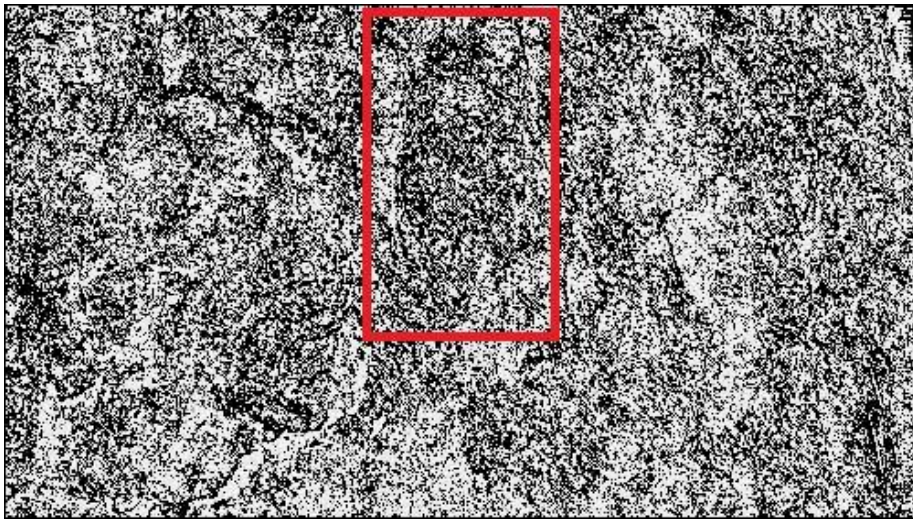


図 18 LBP 特徴量差分(原画像と均等化)

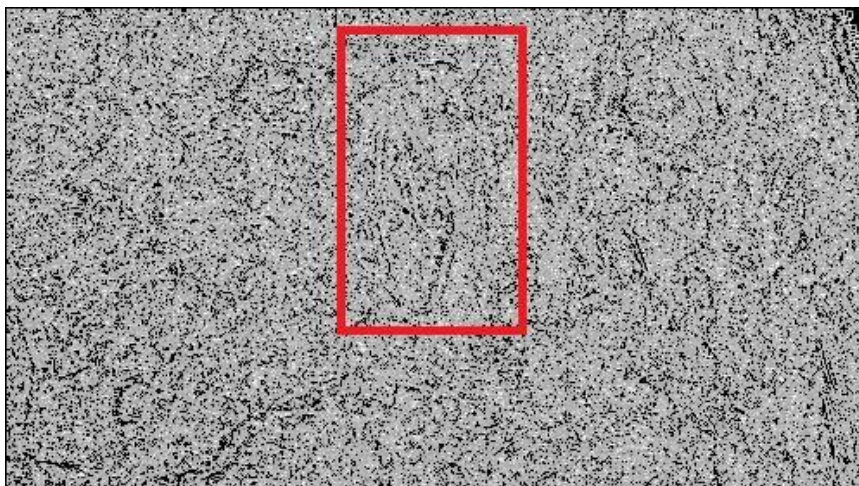


図 19 LBP 特徴量差分(原画像とメジアン処理)

5. おわりに

本研究では、画像変換を用いて多数候補抽出を行い、得られた候補に対して検証を行うことで誤検出を減らす手法を提案した。候補抽出において、検出漏れを少なくすることができ、最終的な物体検出率もよくなった。しかし、複雑背景だけでなく一般背景に対しても同じような結果となった。今回は対象物体が必ず1つ、画像内に存在するという前提をおいたが、実利用の場面では、存在しない場合や複数存在する場合もある。そのような場合にも対処できるようにすることが今後の課題である。

謝辞

最後に、本研究を行うにあたりご指導いただいた古谷博史教授， 椋木雅之教授
そのほか古谷研究室をはじめとする多くの方々にお礼申し上げます。

参考文献

- [1]岩井儀雄, 勞世竝, 山口修平, 山高嗣 “画像処理による顔検出と顔認識”, 情処研報 CVIM-149, 2005.
- [2]三谷重知, 吳海元, 塩山忠義 “ガボール変換 による自動車検出”, 信学論 J83D-II (12), pp.2641-2651, 2000.
- [3] 佐藤雄隆,金子俊一,丹羽義典,山本和彦 “Radial Reach Filter (RRF)によるロバストな物体検出”, 信学論 J88D-II (5), pp .616-624, 2003.
- [4] 傍士竜百, 山田悠太, 清水明宏 “認識対象の回転・大きさの変化に対応した機械学習に関する研究”, 信学技報 ISEC2011-47, 2011.
- [5] Y. Jia, et.al. “Caffe : Convolutional Architecture for Fast Feature Embedding”, ACM Conference on Multimedia, pp.675-678, 2014.
- [6]Jo Chang-yeon “Face Detection using LBP features”, CS 229 Final roject Report, 2008.
- [7] “サポートベクターマシーン(SVM)”
<http://www.sist.ac.jp/~kanakubo/research/neuro/supportvectormachine.html>
- [8] 倉爪亮 他 “コンピュータビジョン最先端ガイド 1”, アドコム・メディア株式会社, 2008.