TUMSAT-OACIS Repository - Tokyo University of Marine Science and Technology (東京海洋大学)

小型ROVに搭載するウニ認識システムの開発

メタデータ	言語: jpn 出版者: 公開日: 2020-06-12 キーワード (Ja): キーワード (En): 作成者: 齋藤, 幹大 メールアドレス: 所属:
URL	https://oacis.repo.nii.ac.jp/records/1933

修士学位論文

小型 ROV に搭載する
ウニ認識システムの開発

2019 年度

(2020 年 3 月)

東京海洋大学大学院

海洋科学技術研究科

海洋システム工学専攻

齋藤　幹大

まえがき

東日本大震災以降，宮城県志津川湾の一部地域において磯やけが発生しており，漁業者に被害をもたらしている．磯やけ発生の一因に，大量発生したウニによる藻類の食害が考えられており，それにより藻場が喪失したと考えられている．対策として小型 ROV を用いてウニを回収・駆除を行うシステムの開発が進んでいる．本システムでは，船上から ROV を投入して海底 10m から 20m に生息する殻径 φ20 から 60mm のウニを搭載されたカメラを介してオペレータが確認する．その後 ROV に搭載されたホースによりウニを吸い上げ，カゴに回収する．ウニは，画像認識手法を用いたシステムにより発見し，ROV オペレータの補助を行う．本システムでは ROV のカメラから取得した映像を船上にてリアルタイムで画像認識させる．さらにカメラ映像を他の PC 等で情報共有することによりウニの発見を補助する．

画像認識には学習器と呼ばれる画像識別器が必要だが，現地での仕様変更を考え，1 日で作成することを目標とする．画像認識にはパターンマッチングとディープラーニングを用いて行う 2 つの手法が主流だが，ディープラーニングと比べてパターンマッチングは短時間で学習器を作成でき，ワークステーションのような高性能 PC も不要なため，本システムにはパターンマッチングを採用した．パターンマッチングにおける画像認識は，正解画像(ポジティブサンプル)と不正解画像(ネガティブサンプル)の 2 種類からなる教師データを用いて作成された学習器により，画像から対象のモノの特徴量を捉えて発見するものである．

学習器作成の際に用いた教師データは，ポジティブサンプルは発見対象であるウニ，ネガティブサンプルはヒトデや貝，岩などのウニではないものとした．本研究では，まずポジティブサンプルとネガティブサンプルの最適な比率を調査した．次にサンプルに用いる画像の種類の選定をした．現地画像，ウェブサイトから取得した画像から適切な種類を評価した．また，教師データに使用する画像枚数を 10,000 枚集め，現地での仕様変更を考慮した 1 日程度で作成可能な成果の高い学習器を検討した．その結果，本システムに用いる学習器の教師データは，ポジティブサンプル 3 に対してネガティブサンプル 7 の比率で，かつ現地で取得した画像を 2,000 枚用いることが適していることがわかった．

本論文では，ウニ認識システムに搭載する，1 日程度で作成可能な学習器の作成を目標として行ったそれぞれの調査結果及び考察について述べる．

第 1 章では磯焼け対策として進めている ROV システムの概要と本研究のつながりについて，第 2 章では ROV に搭載する画像認識の基礎について述べる．第 3 章と第 4 章では画像認識に必要な教師データに用いる画像群について調査した結果を述べる．第 5 章では教師データの画像数について調査した結果について述べる．第 6 章では第 5 章までのまとめと今後の展開について述べる．

目次

1. はじめに 1
1.1. 磯やけ 1
1.2. ウニ駆除システム(ROV) 1
1.3. ウニ自動認識システム(画像認識) 1
1.4. ウニのための画像認識 1
1.5. 識別器作成時間目標 2
2. 画像認識アルゴリズム 6
2.1. 画像認識手法の選定 6
2.2. Viola-Jones 法を用いた画像認識手法 6
2.3. 学習器作成方法 7
2.4. サンプル設計 9
3. サンプル収集 20
3.1. a のサンプル収集 20
3.2. b のサンプル収集 20
3.3. c のサンプル収集 21
3.4. d のサンプル収集 21
4. 画像群による画像認識結果の違い 23
4.1. 画像認識評価 23
4.2. 認識結果 23
4.3. 認識結果まとめ 23
5. 学習器に用いるサンプル数の調査 27
5.1. 10,000 枚のサンプル収集 27
5.2. 学習器作成 27
5.3. 画像認識結果 27
5.4. 2,000 枚の学習器 28
5.5. 作成時間 28
5.6. 評価値 CV を用いた学習器評価 28
5.7. サンプル群再調査 29
6. まとめ 35
6.1. 学習器について 35
6.2. 画像認識システムについて 35
参考文献 36
謝辞 37

1. はじめに

1.1. 磯やけ

東日本大震災以降，宮城県志津川湾の一部地域において磯やけが発生している[1](図 1-1)．震災以降大量発生したウニによる藻類の食害が原因の一つだと考えられている．藻類の減少により藻類の水揚げだけで無く，藻類を用いて行うアワビの生育が阻害されアワビの水揚げ量も減少しており漁業に被害を及ぼしている[2]．また，ウニは藻類の無い環境にあるため身入りの少ないウニ(図 1-2)となっており，早急に駆除することが求められている．現在，駆除にはダイバーを海域に投入して行われている．ただしダイビングでは，長時間の潜水や 10～20m の水深での作業は非常に身体的に負担がかかり現実的ではない．これを水中ロボット(ROV)で解決する．

1.2. ウニ駆除システム(ROV)

現在, 1.1.の問題点から，人を使った回収・駆除ではなく，小型 ROV を用いたウニの回収・駆除システムの開発を進めている[3](図 1-3)．小型 ROV の諸元表は表 1-1 に示す．本システムでは船上から ROV を投入しオペレータが操作，ウニをポンプにより吸引し，カゴに回収する．回収するウニは，殻径 20-60mm 程度である[4]．本システムではウニを効率的に回収するためにオペレータのウニ回収作業を補助する画像認識を用いたウニ発見システムを導入する．本システムによりダイバーが直接潜ってウニを回収・駆除せずにすみ効率的かつ安全に身入りの少ないウニを回収することを目指す.

1.3. ウニ自動認識システム(画像認識)

本システムでは ROV に搭載したカメラから NTSC 形式で画像を船上の PC に送り画像認識する．また，カメラ映像をその他の PC に Gstreamer と呼ばれる映像伝送のオープンフレームワーク用いて映像を分配する．これによりオペレータだけでなく船上の作業員がウニの回収状況を逐一共有できる(図 1-4).

本システムに搭載している Gstreamer とは映像伝送に用いられる一般的なオープンフレームワークで言語は C であり，映像の伝送はパイプラインを作成して行う．パイプラインは送信側と受信側で立てる．伝送形式は TTP 通信や RTP 通信があり，1 対 1 のピアツーピア通信だけでなく 1 対多のキャスト通信も可能である．現在までに両方の通信を確認した．このとき，VLC を用いて行う場合 SDP ファイルと呼ばれるファイルを使用する．これらの通信概念図を図 1-5 に示す.
現在は Gstreamer による拡張は行っておらず, ROV の水中カメラ映像をテザーにより船上の PC に送信し, PC 上で画像認識をしているのみである.

1.4. ウニのための画像認識

本システムに搭載する画像認識システムは，多少の誤認識があってもウニを多く発見することが重要である．そのため，群集から特定の人の顔を認識するような顔認証のようなものではなく，ビル街のような町並みの風景から人を発見してその数を知るような認識方法となる．ただし，誤認識は

できるだけ避ける必要がある．目標は 90 パーセントほどを目指す．

1.5.　識別器作成時間目標

画像認識に必要な画像識別器(2.で示す)の作成時間は 1 日程度を目指している．ウニは現在までにデータが揃い，ウニの個体数調査や回収には準備ができているが，現地調査の使用用途に合わせてナマコ，ヒトデ，タコなどを対象にした学習器を急に作成する必要がある．その場合に対応できるように 1 日程度で作成することを目標にする．

図 1-1　磯やけ

図 1-2　身入りの少ないウニ

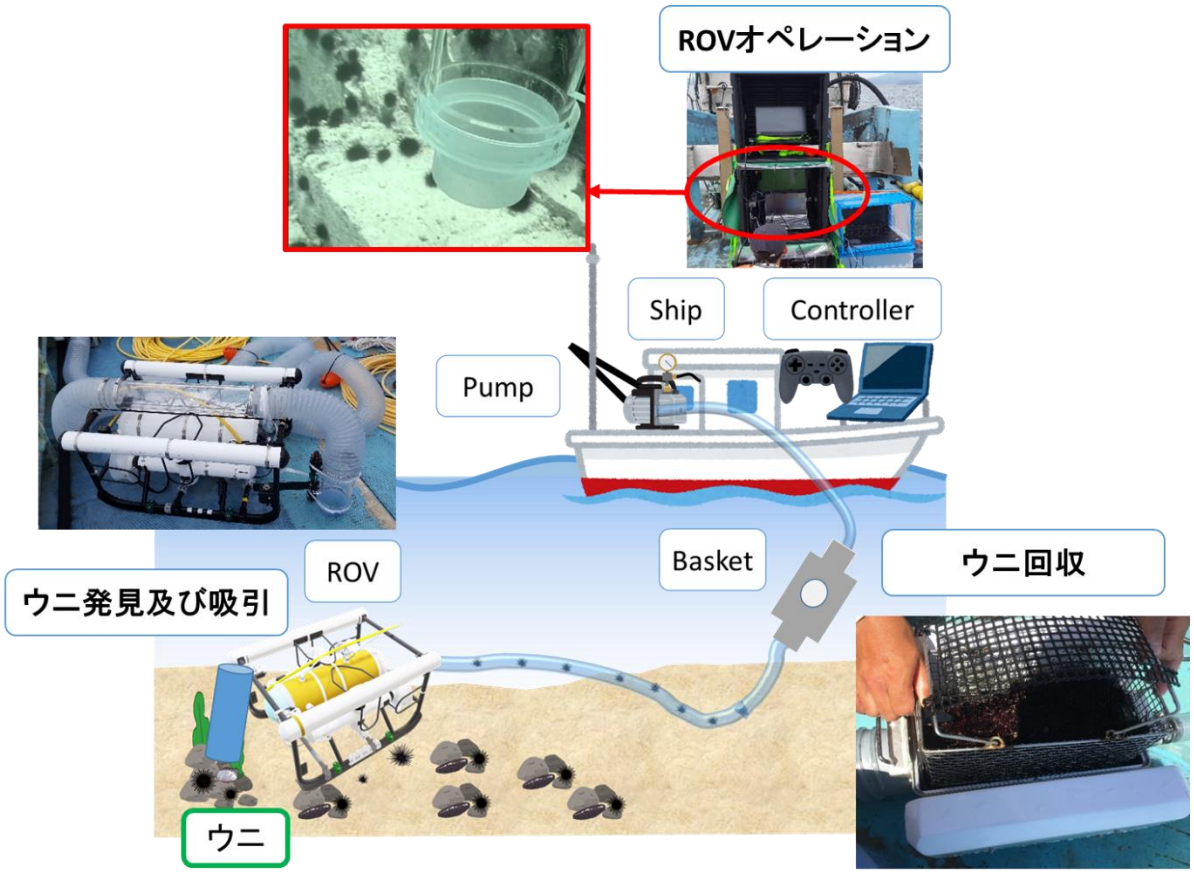

図 1-3　ウニ回収システム

表 1-1　小型 ROV 諸元表

項目	内容
外形寸法	W635×L840×H490mm
重量	45.4kg
カメラ	NTSC 形式
メインスラスタ	水平 2 基、垂直 2 基(200W)
サブスラスタ	垂直 2 基(200W)
ロボットハンド	1 基
センサ	深度計、方位計、高度計

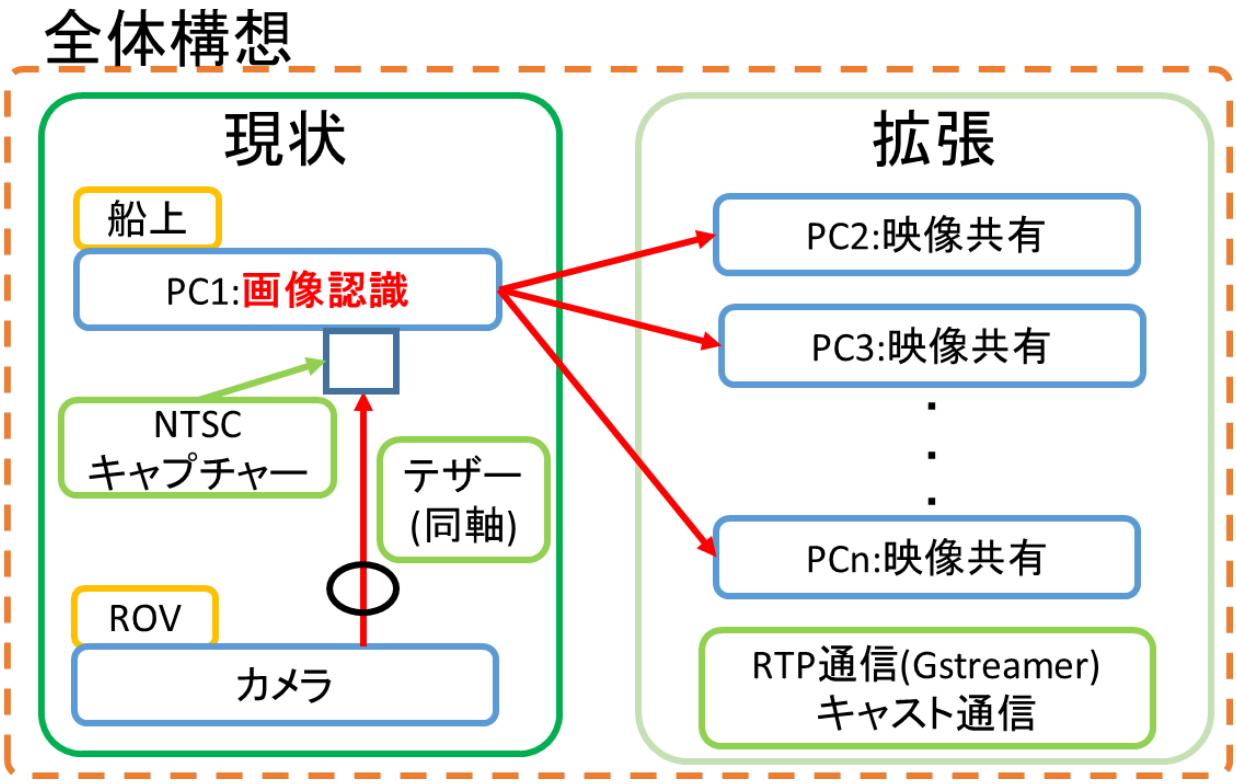

図 1-4　画像認識システム

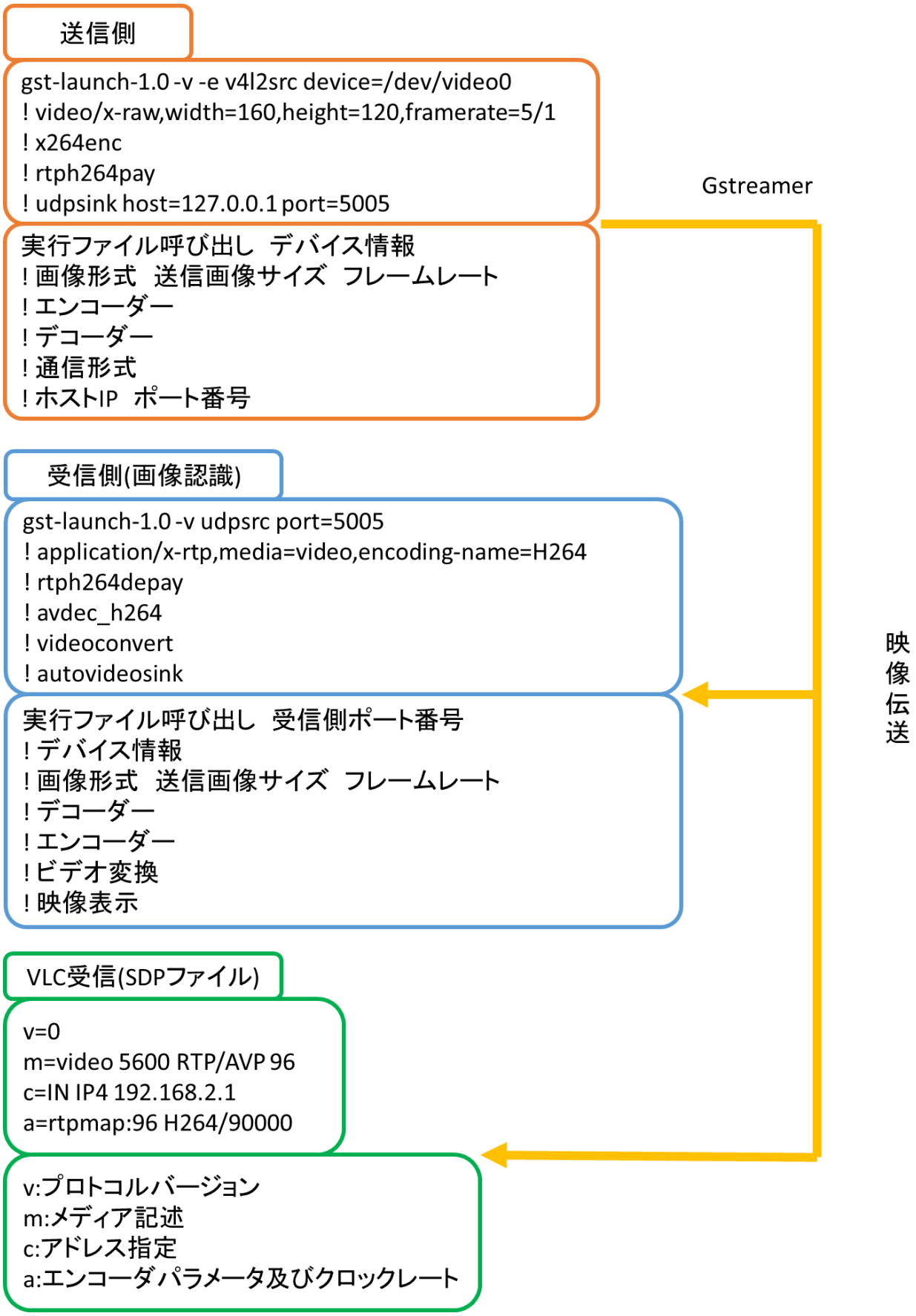

図 1-5　Gstreamer 映像伝送ネットワーク図

2. 画像認識アルゴリズム

2.1. 画像認識手法の選定

本画像認識システムでは風景の中にある特定のものを発見し数をカウントするような認識方法のため，認識率を大きく重視はしない．認識率よりも多少のエラーがあってもすぐにウニを発見できることと，ウニ以外の海洋生物への対応も想定し，短い作成時間で識別器を作成することを重視する．具体的には 1 日程度を想定している．

識別器は画像認識手法における学習器であり，機械学習を行い作成する．現在，機械学習手法としてディープラーニングとパターンマッチング法を用いた手法がある．ここではそれぞれの学習手法を用いた画像認識について述べたうえで本研究に適した手法の選定を行った．

2.1.1. ディープラーニング

機械学習手法のうち現在よく使用されているのはディープラーニング手法である．本手法は非常に高い認識手法を誇り，よく顔認証に使用されることが多い．ディープラーニングでの学習器は，人間の脳内ネットワーク構造に似たニューラルネットワークになっており，画像中にあるモノの輪郭から抽出した特徴量は自動で設計される．そのため複雑なものを認識できる利点や認識対象物のみを学習させればいいため枚数が少なくても十分に認識できる利点がある．

ディープラーニングを用いた画像認識では，フリーのソフトウェアである DarknetYOLO を用いて行うことが多い．しかし，本ソフトウェアのデータセットには海洋生物はほとんどいないためウニの認識は不可能である．そのため，本研究で扱う場合には学習器を一から作る必要がある．しかし，学習器の作成時間は膨大な時間がかかり，高性能 CPU と大容量 GPU を搭載したワークステーション PC が必要である．そのため本研究には採用しない

2.1.2. パターンマッチング法を用いた機械学習手法(Viola-jones 法)

ディープラーニングがよく使われるようになる前に主流だった機械学習手法のひとつであり，認識精度はあまりよくない．本手法ではニューラルネットワークのようなものは無く，分岐ネットワークであり，学習器は画像中にあるモノの輪郭から特徴量を自分で設計する必要がある．そのため一般的に認識対象物と非認識対象物を合わせて 10,000 枚など大量の画像が必要となるが，学習器作成時間は表 2-1 のゲーミング PC 程度で，数日で終わる程度である．

本手法は比較的容易に作成できる学習器で, OpenCV を用いることが一般的である. OpenCV とはインテルが開発・公開したオープンソースのコンピュータビジョン向けライブラリで C や Java, Python が言語として用いられる．また，OpenCV は行列演算だけでなく，機械学習や画像認識にも使用される．

多少の誤認識を容認し，識別器の作成時間の短い Viola-Jones 法[5]を本研究では用いることとした．また，使用 PC は表 2-1 に示すノート PC である．

2.2. Viola-Jones 法を用いた画像認識手法

Viola-Jones 法とは, Viola,Jones によって開発された手法で，現在は, Lienhart, Maydt らによって

改良されている[6]．本手法では，判別の粗い強識別器から判別の細かい強識別器を多数連ねた学習器によって入力された画像を画像認識させる．強識別器には認識対象物と非認識対象物の特徴量を持っており，判別時にはこの特徴量で入力された画像が認識対象物か非認識対象物かを判別する．すべての強識別器で画像を判別するわけではなく，判別の粗い強識別器から判別の細かい強識別器に順に判別を行い，早い段階で非認識対象物をはじくことができる(図 2-1).

次項で強識別器，学習器について触れる．

2.3. 学習器作成方法

2.3.1. サンプル作成方法

学習器の作成に必要なものは，認識したいものであるポジティブサンプルと認識させたくないものであるネガティブサンプルからなる教師データ，画像の特徴量の抽出方法や学習回数等を決定するパラメータである．本論文では，ポジティブサンプルはウニ(図 2-2)，ネガティブサンプルはウニ以外のヒトデ，貝，岩場等の画像とした(図 2-3)．なおポジティブサンプルは必要数の 1.1～1.2 倍程度で画像を収集する必要がある[7]．これは，後述する強識別器を作成する際にポジティブサンプルの不足を補うためである．具体的には複数の強識別器を作る過程で，ポジティブサンプルに不良画像があった場合にはそのサンプルははじかれるが，一度はじかれたサンプルは強識別器に用いられることはないためである．ただし，はじかれるサンプル数はランダムである．また，教師データのサンプルを使いきってしまうと学習が止まってしまうため注意が必要である．現状 1.1 倍～1.2 倍程度集めているが，本研究において止まることはなかった．

また，ネガティブサンプルは画像名のリストを作成する必要があるが，ポジティブサンプルはリスト化したうえでベクトルファイルにする必要がある．ベクトルファイルは，画像を曲線や直線で近似したベクトルデータであり，配布されている OpenCV の一般的なベクトルファイル作成実行ソフトである，opencv-createsamples.exe を用いて自動で生成される．

2.3.2. 特徴量抽出

教師データに用いた画像から特徴量を抽出するには Haar-Like 特徴量を用いて行う．

Haar-Like 特徴は，画像上の明暗差から特徴量を抽出できるものである．具体的には，画像上に探索窓を自動で生成し，図 2-4 で示す 14 種類の矩形パターンが教師データの画像上を自動で走査し特徴量抽出を行う(図 2-5)．矩形パターンは画像の明暗差から輪郭を認識し，画像の特徴量を抽出する．14 種類のうち最も画像の特徴量として適したものを弱識別器と呼ぶ．

OpenCV は他にも LBP(Local Binary Pattern)特徴量や HOG(Histograms of Oriented Gradients)特徴量をサポートしているが，本論文では，Viola-Jones 法を採用するため用いない．

2.3.3. 強識別器

強識別器は弱識別器が多数集まったものである．弱識別器は Adaboost[8]を用いて多数連結され強識別器が生成される(図 2-6)．Adaboost は一つ一つの識別精度が低いが多数連結することによって一つの強識別器を作成する学習アルゴリズムである．

また，弱識別器はサンプルの特徴量を持っており，その特徴量で入力された画像の一部を判別す

る．そして強識別器は，複数の弱識別器の判別結果をもとに入力画像の一部が正解かどうかを判断する．なお弱識別器には重みが付けられる．弱識別器は，総サンプル数だけ作成されるわけではなく，ランダムで作成される数は決定する．それはすべての強識別器作成の際にも同様である．強識別器は判別の粗いものから判別の細かいものまで作成され，その数は，学習回数と同数かそれ未満である．学習回数よりも強識別器の数が小さくなる場合は，過学習を防ぐために学習が途中で終了した場合である．

学習回数は設計者により決めることができ，デフォルトでは 20 回学習が行われるようにパラメータが指定されている．本論文でも 20 回を学習回数とした．

2.3.4. カスケードファイル

学習器作成の際には配布されている OpenCV の一般的なカスケードファイル作成実行ソフトである，opencv-traincascade.exe を用いて行う．実行時には教師データ，使用サンプル数，および学習回数や特徴量指定等のパラメーターを用いる(図 2-7)．設定できるパラメータ一覧を表 2-2 に示す．本論文ではサンプル数のみを変更した．

なお，学習回数一回につき一つ，強識別器の情報が記載されたファイル(xml 形式)が生成される．具体的には，図 2-8 に示すように学習回数を 20，ポジティブサンプルを 100，ネガティブサンプルを 200 と指定した場合を例にとると以下になる．なお，ポジティブサンプルについては 1.1 から 1.2 倍の枚数が必要となるため，120 枚の画像を教師データとした．

・学習 1 回目(Step1)：教師データのうちポジティブサンプル 120 枚のうち，番号順に 100 枚，ネガティブサンプルは 200 枚選別され，各種パラメータを用いて学習が開始される．学習完了後に強識別器が 1 つ作成される．

・学習 2 回目(Step2)：まず，教師データのうちポジティブサンプルは番号順に 100 枚，ネガティブサンプルは 200 枚が選別される．ポジティブサンプルについては，選別段階で不良と判断されるもサンプルがある．判断されればはじかれるので，新たにポジティブサンプル群からサンプルを引き出す．例えば図 2-8 にあるように，3 枚のサンプルがはじかれた後に新たに 3 枚を補填している。学習に必要なポジティブサンプルの数が 100 になったら各種パラメータを用いて学習が開始される．学習完了後 1 つ強識別器が作成される．

・学習 3 回目(Step3)：3 回目では教師データのうち，まずポジティブサンプルは 2 回目の学習時に使用された 100 枚，ネガティブサンプルは 200 枚選択される．ポジティブサンプルについては，2 回目に用いたサンプルから不良画像がはじかれ，新たにポジティブサンプル群からサンプルが引き出される．例えば， 図 2-8 のように 2 枚のサンプルがはじかれたあとに新たに 2 枚を補填している．学習に必要なポジティブサンプル数が 100 になったら学習が開始される．学習完了後に 1 つ強識別器が作成される．

基本的には指定した学習回数分この処理が行われ，学習が行われるが，学習回数が強識別器過学習を防ぐために指定した学習回数よりも少ない回数で終了することがある．

すべての学習完了後，すべての強識別器の情報とパラメータファイルを一つにしたカスケードフ

ファイルが作成される．このファイルはxml形式で，ネットワーク構造が記載されている．このカスケードファイルによって画像認識を行う．

2.4. サンプル設計

2.4.1. サンプル比率

教師データに使われるポジティブサンプルとネガティブサンプルの比率は，7:3[9]，3:7[10]と論文によっても様々である．本論文で採用する比率を調査するため，総サンプル数を300とし，3:7, 5:5, 7:3で調査した．教師データに用いた画像は2018年5月の現地調査時に取得した計1分程度の動画(HD形式:解像度1920×1080pixel)から0.5秒刻みで画像を切り出し，その画像から取得したウニ(図2-9)やヒトデ，貝など(図 2-10)の画像である．認識に使用した画像は 2019 年 12 月の現地調査時に取得した計3時間程度の動画から計30秒程度の動画(NTSC形式:解像度720×480pixel)を切り出し，0.5秒刻みで切り出した計60枚の画像(図2-11)である．

このときの枚数は表 2-3 に示す．認識率は式(1)で評価した．認識率の概念図を図 2-12 に示す．ここで，図に示す四角い枠は機械が認識したウニ箇所である．四角い枠の数，機械が認識したウニの数をここではmとする．また，三角形で示す箇所は機械が認識したウニのうち実際にウニだった，ウニ認識成功数をkとする．

$$R = \frac{\mathrm{k}}{\mathrm{m}} \quad \cdot\cdot\cdot(1)$$

図 2-12 を例にとると四角い枠で示された機械が認識したウニの数は 7．また三角形で示されたウニ認識成功数kは5であった．それぞれの数値を式(1)に代入するとRは以下のようになる．

・機械が認識したウニの数 m：7
・ウニ認識成功数 k：5
・認識率 R：0.6

7:3, 5:5, 3:7 で調査した結果，3:7 の比率であると認識率が 100 パーセント近くで非常に認識率が高い(表 2-4)．また，学習器のポジティブサンプル比率による認識率の変化を図 2-13 に示す．

また，画像認識結果を図 2-14 に示す．

図 2-13 より，2:8 や 1:9 と比率を変更しても認識率上昇への効果は少ないため，今回は 3:7 で学習器を設計することとする．

2.4.2. サンプルの種類

学習器に用いるポジティブサンプルはウニ，ネガティブサンプルにはヒトデや貝，岩場等の志津川の海域にいるウニ以外のものとした．サンプル画像は2018年の画像から取得した計1時間30分程度の動画(NTSC 形式:解像度 720×480pixel)から 0.5 秒ごとに画像を 30,000 枚程度，ペイントや自作アノテーションソフトウェア(図 2-15)で切り出した．なおポジティブサンプルは28,000枚程度，

ネガティブサンプル 8,000 枚程度(図 2-2, 図 2-3)となっている.

このアノテーションソフトウェアは python と processing(java)で作成したもので GUI となっており, Linux 上で動作する. ここで processing とは Java 言語を単純化し, グラフィックのみに特化させたプログラミング言語であり, GUI の作成によく使われるものである.

このアノテーションソフトを用いることで, 1 時間当たり 512 枚程度で切り抜くことができる. ポジティブサンプルは 28,000 枚で 55 時間程度, 8,000 枚のネガティブサンプルの切り出しには 15 時間程度かかり, 合計で 70 時間程度かかった.

なお, ポジティブサンプルはウニのみを抜き出すので簡単なため, 多く抜き出せたが, ネガティブサンプルはウニ以外のものを集める必要があり難しく少なかった. そのためデータセットを作成する人によりネガティブサンプルの中身は変わることがある.

表 2-1　用いた PC スペック

項目	スペック
CPU	AMD　Ryzen7　2700X
コア数	8
クロックレート	3.7GHz
CPU メモリ	16GB
GPU	NVIDIA1070
OS	Ubunutu16.04

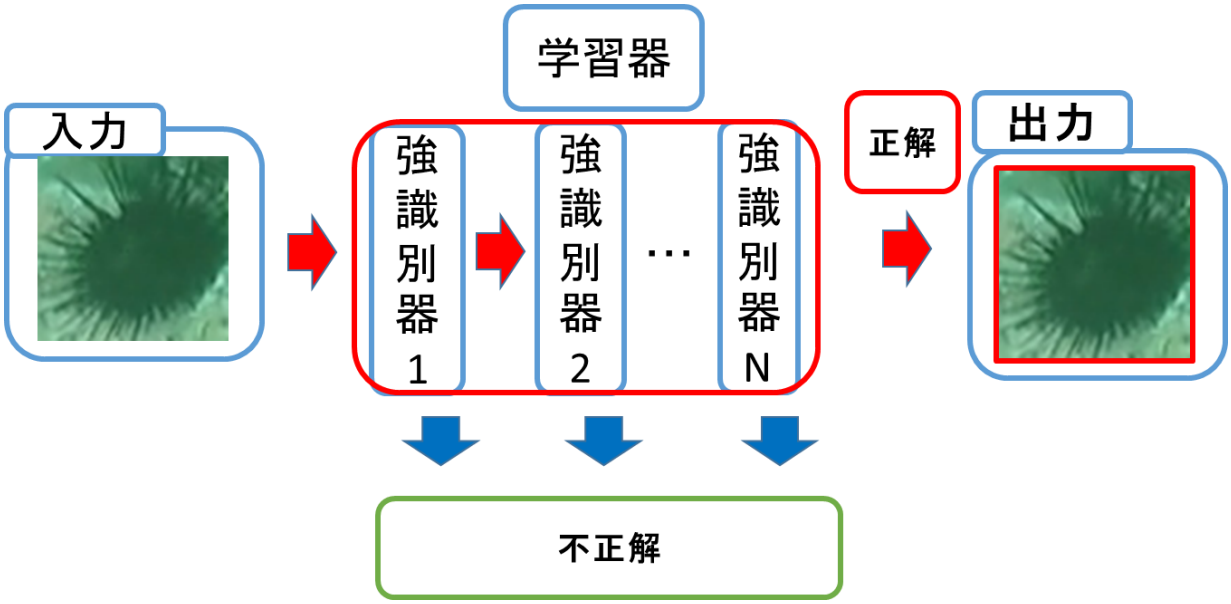

図 2-1 画像認識フロー

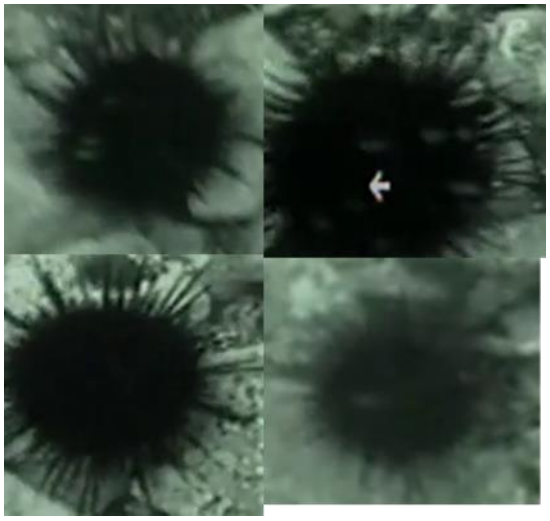

図 2-2　ポジティブサンプル(ウニ)

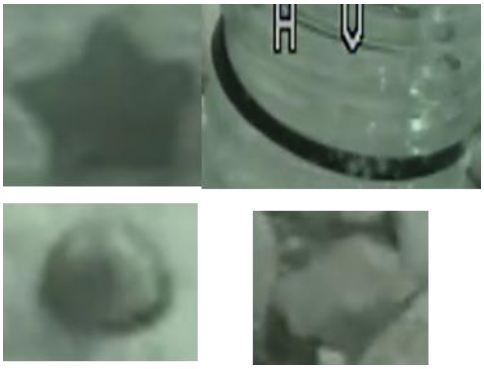

図 2-3　ネガティブサンプル(ヒトデ，貝，岩場)

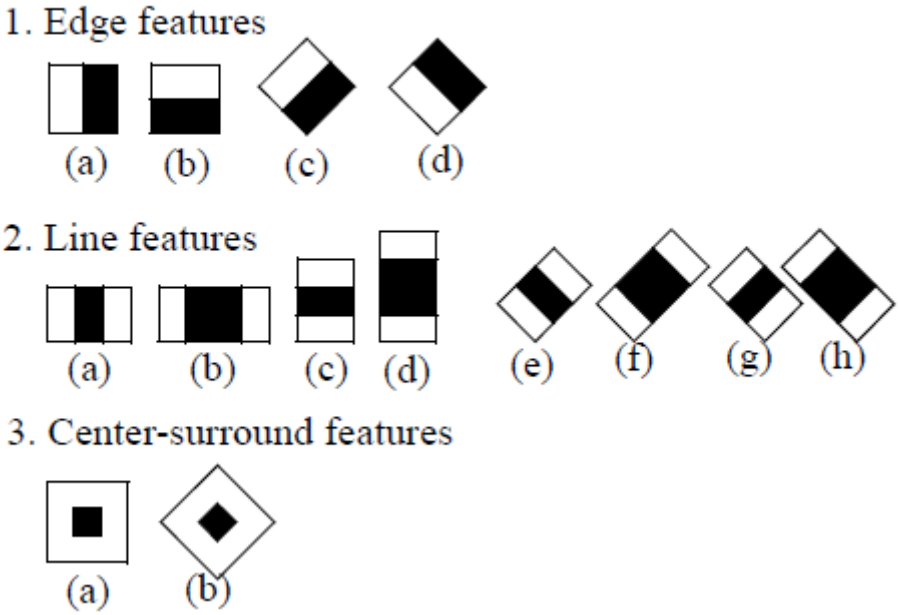

図 2-4　矩形パターン(出展:[2])

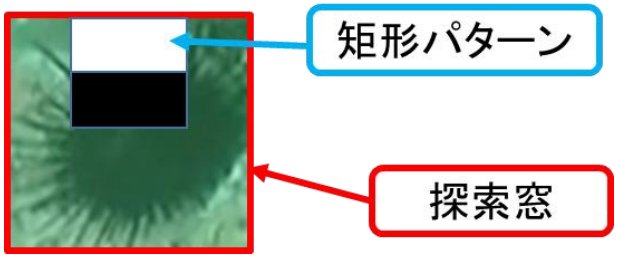

図 2-5 特徴量の図

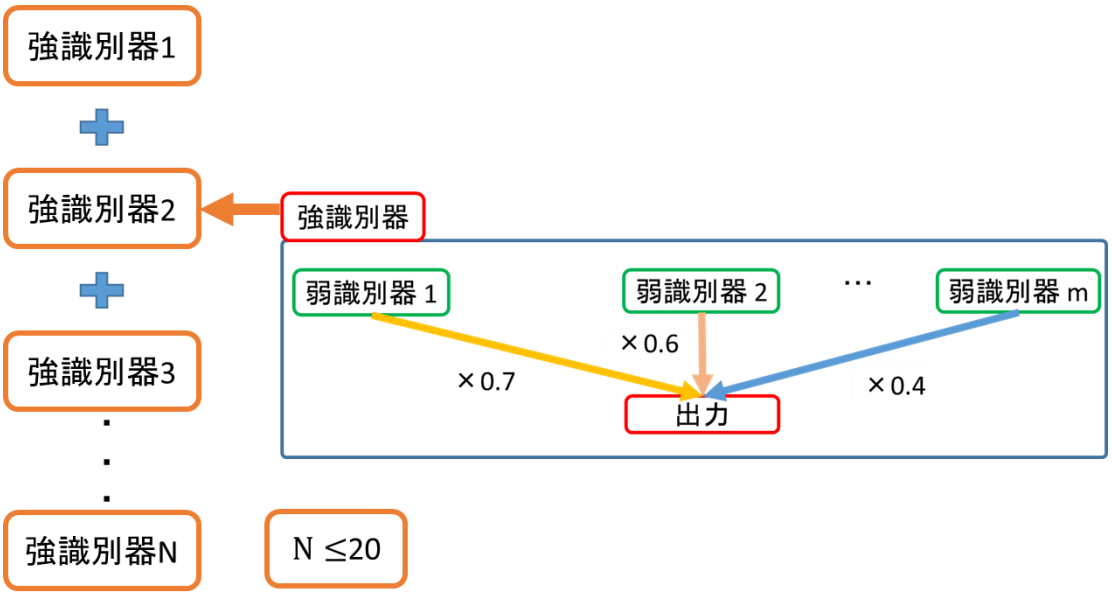

図 2-6 弱識別器と強識別器

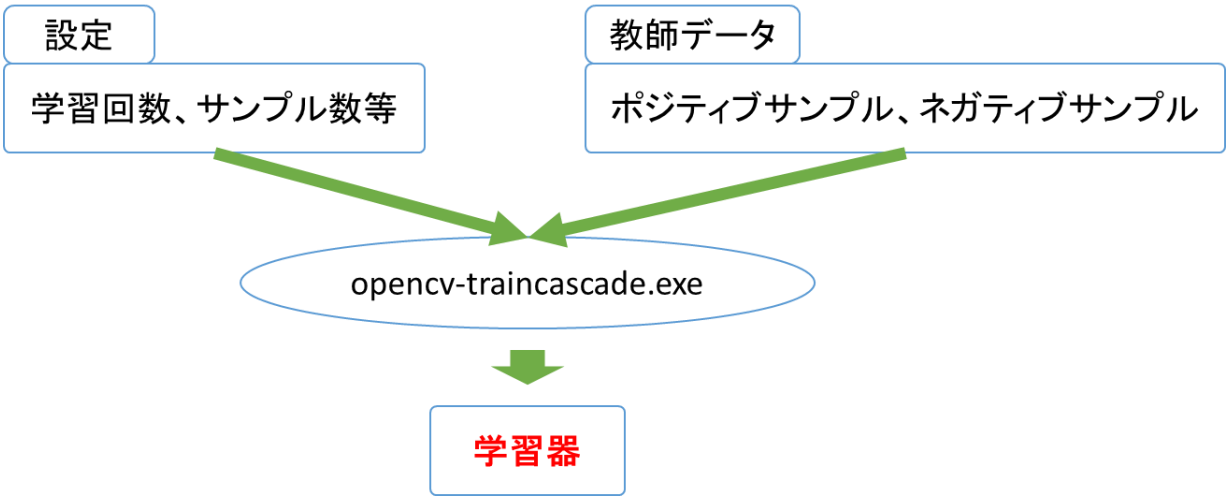

図 2-7 学習器作成概要

表 2-2 カスケードパラメータ(一部)

引数	デフォルト	説明
data	NULL	カスケード分類器ファイルの作成先フォルダの指定
vec	NULL	ポジティブサンプルファイルの指定
bg	NULL	ネガティブサンプルファイルの指定
numPos	1000	使用するポジティブサンプルの数
numNeg	2000	使用するネガティブサンプルの数
feature Type	HAAR	サンプル画像から特徴量を探す方法の指定
w	24	サンプルと同じ高さを指定
h	24	サンプルと同じ幅を指定

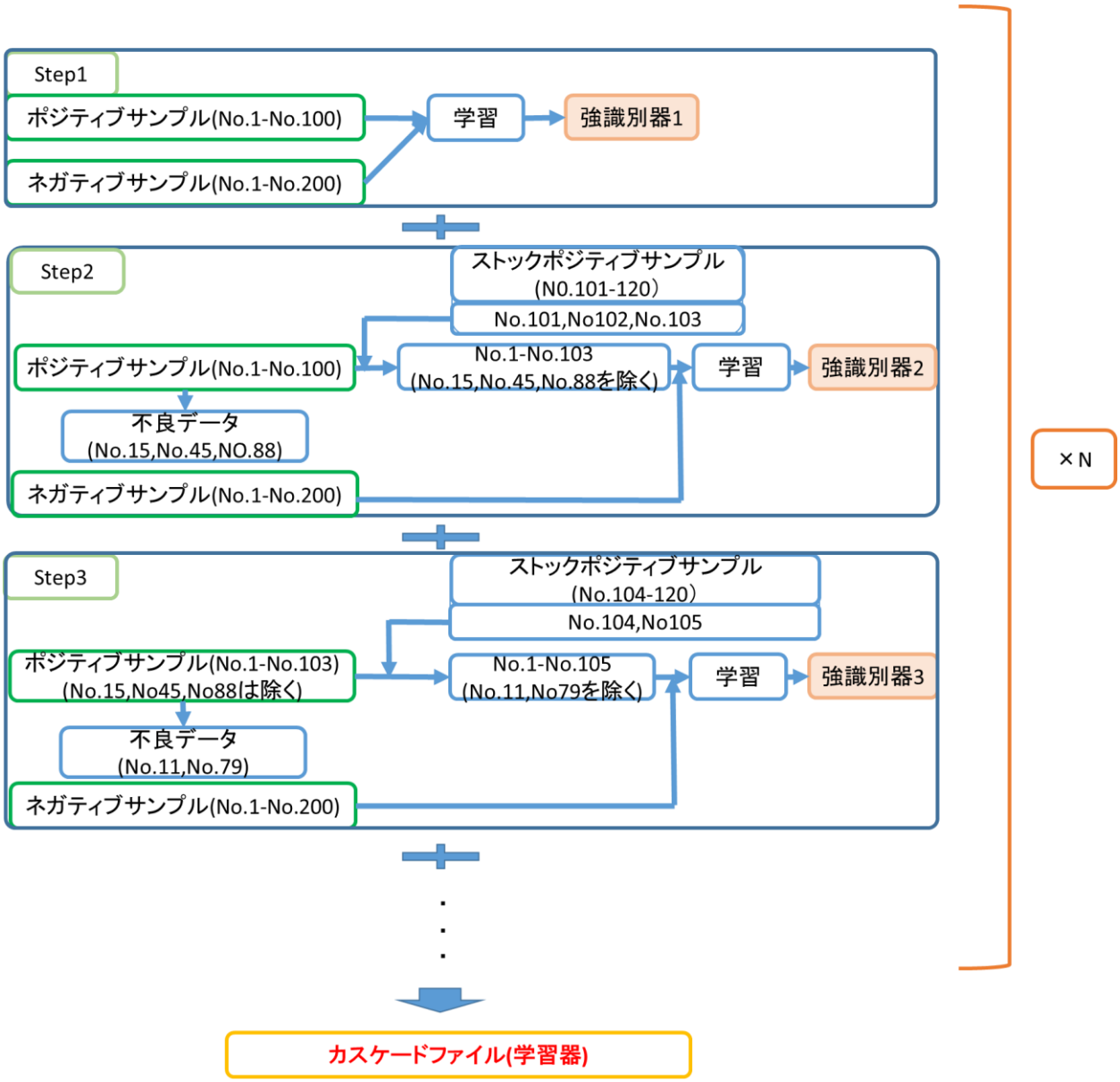

図 2-8　カスケードファイル作成手順

図 2-9 ウニ(2018 年 5 月取得画像)

図 2-10 ヒトデ，貝，岩肌(2018 年 5 月取得画像)

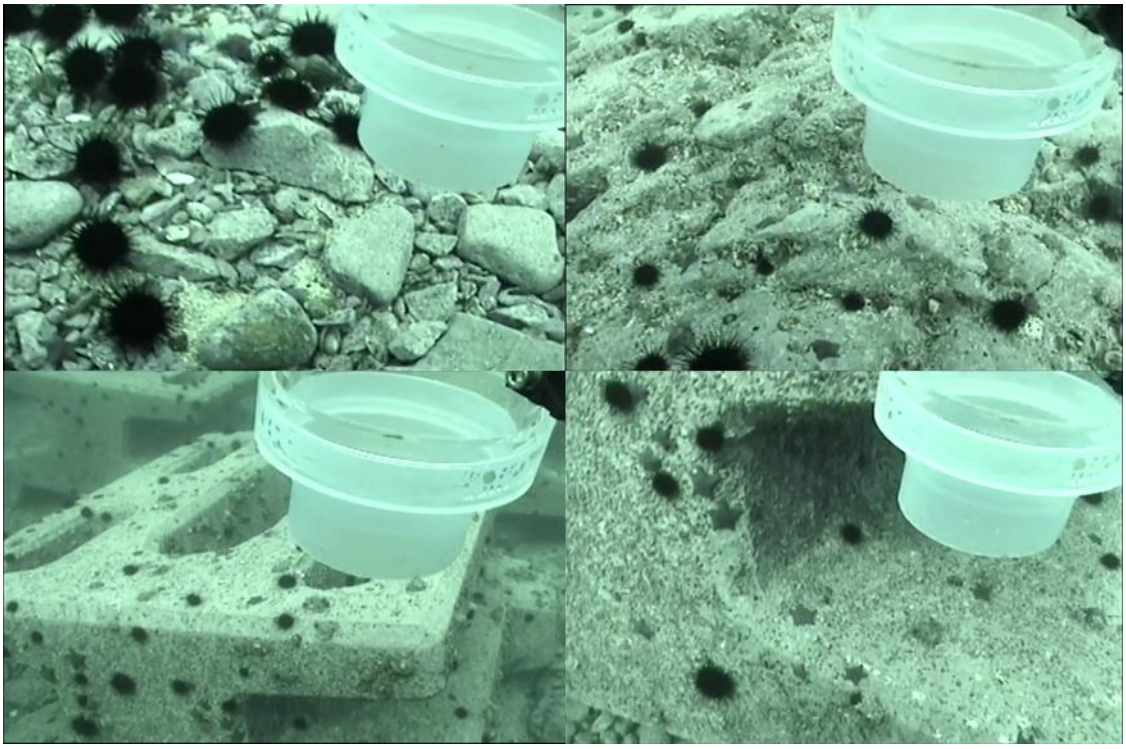

図 2-11　2019 年 12 月の画像

表 2-3 サンプル数

学習器	サンプル比	ポジティブサンプル(枚)	ネガティブサンプル（枚）
1	ポジティブ7:ネガティブ 3	210	90
2	ポジティブ 5:ネガティブ 5	150	150
3	ポジティブ 3:ネガティブ 7	90	210

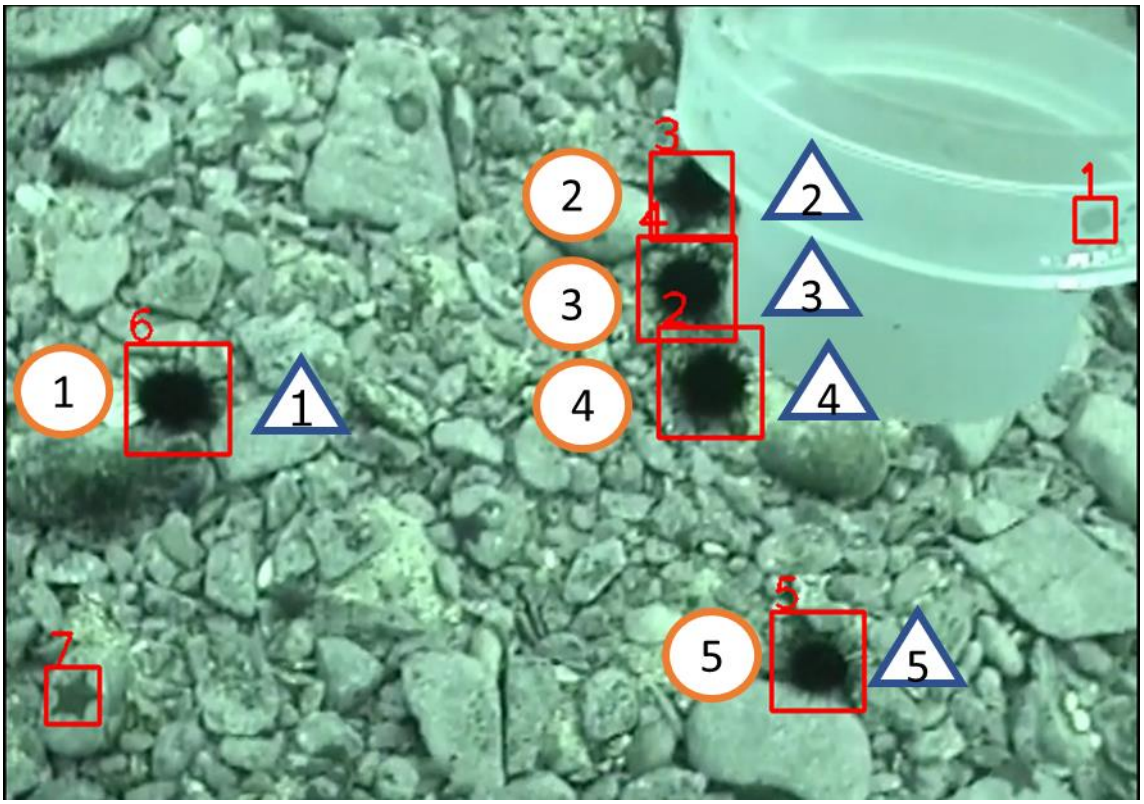

図 2-12 画像認識概念図

表 2-4　画像認識結果

枚数比率	動画	認識率
ポジティブ 7:ネガティブ 3	A	0.710
	B	0.667
	C	0.521
	avr.	0.633
ポジティブ 5:ネガティブ 5	A	0.858
	B	0.828
	C	0.906
	avr.	0.864
ポジティブ 3:ネガティブ 7	A	0.992
	B	0.965
	C	0.967
	avr.	0.975

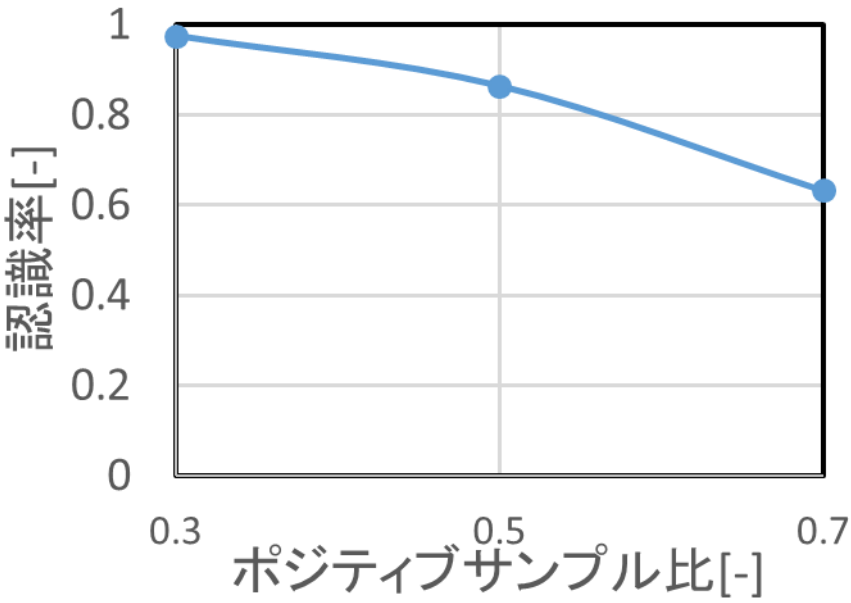

図 2-13 ポジティブサンプル比による認識率の変化

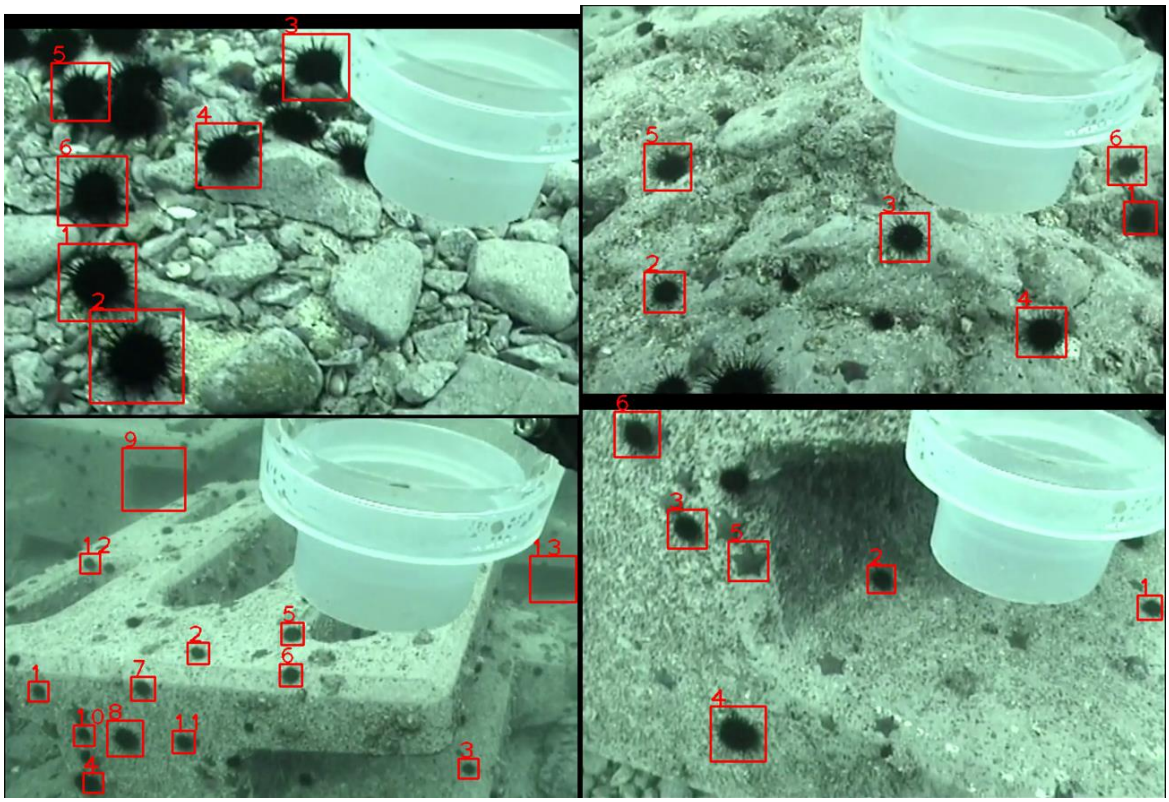

図 2-14　画像認識結果(2019 年 12 月の画像)

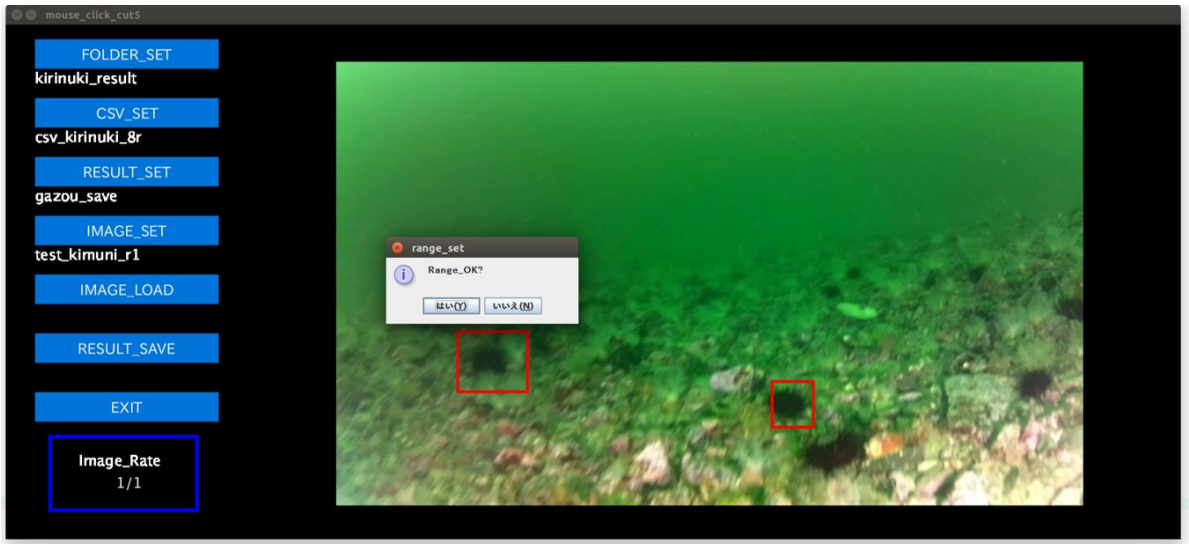

図 2-15　アノテーションソフトウェア

3.　サンプル収集

認識率は 90 パーセント程度で，発見率が高く作成時間の短い学習器を目標とする．学習時間についてはウニ以外を調査対象にした場合を想定し，1 日を限度とする．

これらの目標を達成する学習器作成のためにまず用いるサンプル群を a～d の 4 種類集めた．

a. 現地画像(実験海域)のみ
b. ウェブサイト取得画像のみ
c. ポジティブサンプルには現地画像，ネガティブサンプルにはウェブサイト取得画像
d. ポジティブサンプルにはウェブサイト取得画像，ネガティブサンプルには現地画像

3.1.　a のサンプル収集

宮城県志津川湾の動画データから計 400 枚程度集めた．データの切り出し方は以下の手順で行った．

・Step 1　動画データの取得
宮城県志津川湾にて 2018 年 5 月の現地調査より取得した動画データ(HD 形式:解像度 1920×1080 pixel)から計 1 分程度の動画に切り出し．

・Step2　画像データの取得
切り出した計 1 分程度の動画から 0.5 秒刻みで画像を切り出し．

・Step3　サンプルの取得
切り出した画像を Windows のペイントソフトウェアにて切り出し．切り出したポジティブサンプルは 120 枚，ネガティブサンプルは 255 枚の計 375 枚であった．
　切り出し枚数は 1 時間当たり 30 枚程度であったため，サンプル数の総切り出し時間は約 13 時間であった．

切り出したポジティブサンプル群の一例を図 2-9 に，ネガティブサンプル群の一例を図 2-10 に示す．

3.2.　b のサンプル収集

Google の画像検索により画像を 400 枚程度取得した．取得手順を以下に示す．

・Step 1　画像取得
Google により検索したキーワードにより画像を検索しダウンロードした．

・Step2　サンプル取得
ポジティブサンプルの場合にはウニを見つけて切り出しを行い，サンプルを取得したが，ネガティブサンプルの場合には切り出しは行わなかった．切り出したポジティブサンプルは 113 枚，ネガティブサンプルは 255 枚の計 368 枚であった．

切り出し枚数は 1 時間あたり 20 枚程度であったため，サンプル数の総切り出し時間は約 18 時間であった．ポジティブサンプルは画像の切り出し，ネガティブサンプルについては検索するところに時間が多くかかった．

切り出したポジティブサンプル群の一部を図 3-1 に，ネガティブサンプル群の一部を図 3-2 に示す．

3.3.　c のサンプル収集

ポジティブサンプルは 3.1.で示した画像取得方法で取得した現地画像群 120 枚，ネガティブサンプルは 3.2.で示した画像取得方法で取得したウェブサイト取得画像 255 枚とした．

取得したサンプル群の一部をポジティブサンプル群は図 3-1，ネガティブサンプル群は図 2-9 に示した．

3.4.　d のサンプル収集

ポジティブサンプルは 3.2.で示した画像取得方法で取得したウェブサイト取得画像群 255 枚，ネガティブサンプルは 3.1.示した画像取得方法で取得した現地取得画像 112 枚とした．

取得したサンプル群の一部をポジティブサンプル群は図 2-10，ネガティブサンプル群は図 3-2 に示す．

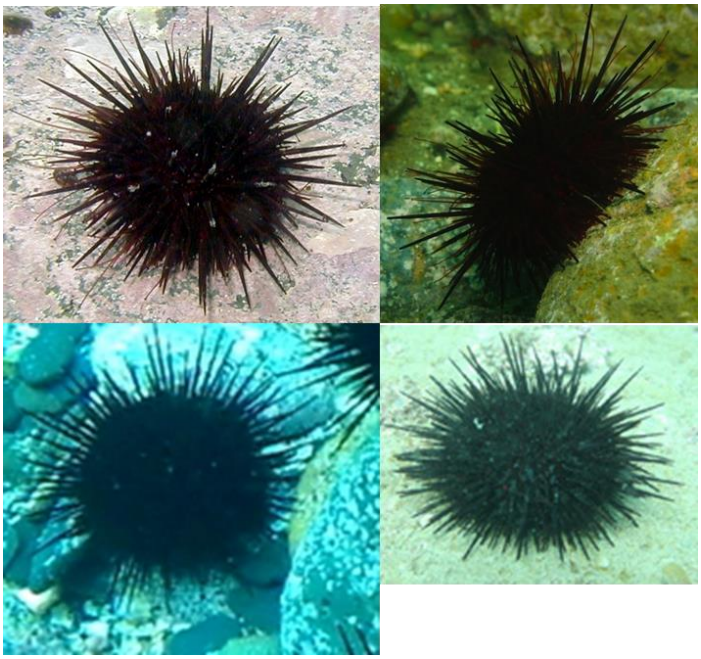

図 3-1　ウェブサイト取得ネガティブサンプル群

図 3-2　ウェブサイト取得ネガティブサンプル群

4.　画像群による画像認識結果の違い

4.1.　画像認識評価

3 章における，切り出した画像群で作成された a～d の学習器の評価方法として，認識率 R の式(式(1))に加えて発見率 D の式(式(2))を導入した.ここで，概念図は図 2-12 に示す．丸形で示した箇所は認識結果画像に映ったすべてのウニの数で n とする．また図 2-12 で四角い枠で表示された箇所，機械が認識したウニの数 m を式(2)に用いている.

$$D = 1 - \frac{|n-m|}{n} \quad \cdot\cdot\cdot (2)$$

図 2-12 を例にとって数値を代入して計算すると D は以下になる.

・画像中に映るウニの数 n：5
・機械が認識したウニの数 m：7
・発見率 D：0.4

ここでは，認識率 R 及び発見率 D を総合的に評価する評価値 V(式(3))を導入することとした．認識率 R が高い値を示しても発見率 D は必ずしも高い値をとらない可能性がある．そのため認識率 R と発見率 D，両方の異なる性質のものを総合的に評価できるように評価値 V を導入する.

$$V = R \times D \quad \cdot\cdot\cdot (3)$$

4.2.　認識結果

3 章で示した a から d のそれぞれ 400 枚程度のサンプル群で学習器を作成し画像認識を行った[11]. 画像認識には 2018 年 3,7,9 月の計 3 時間程度の現地動画データ(HD 形式:解像度 1020×760pixel, NTSC 形式:解像度 720×480pixel)から取得した 15 枚の画像を用いた(図 4-1).

画像認識結果を a は図 4-2, b は図 4-3, c は図 4-4, d は図 4-5 に示す.

4.3.　認識結果まとめ

a から d の学習器の画像認識結果を表 4-1 に示す．表 4-1 より，現地画像のみで作成した a の学習器の評価が一番高かった．そのため，本研究では現地画像のみで学習器を作成することとする.

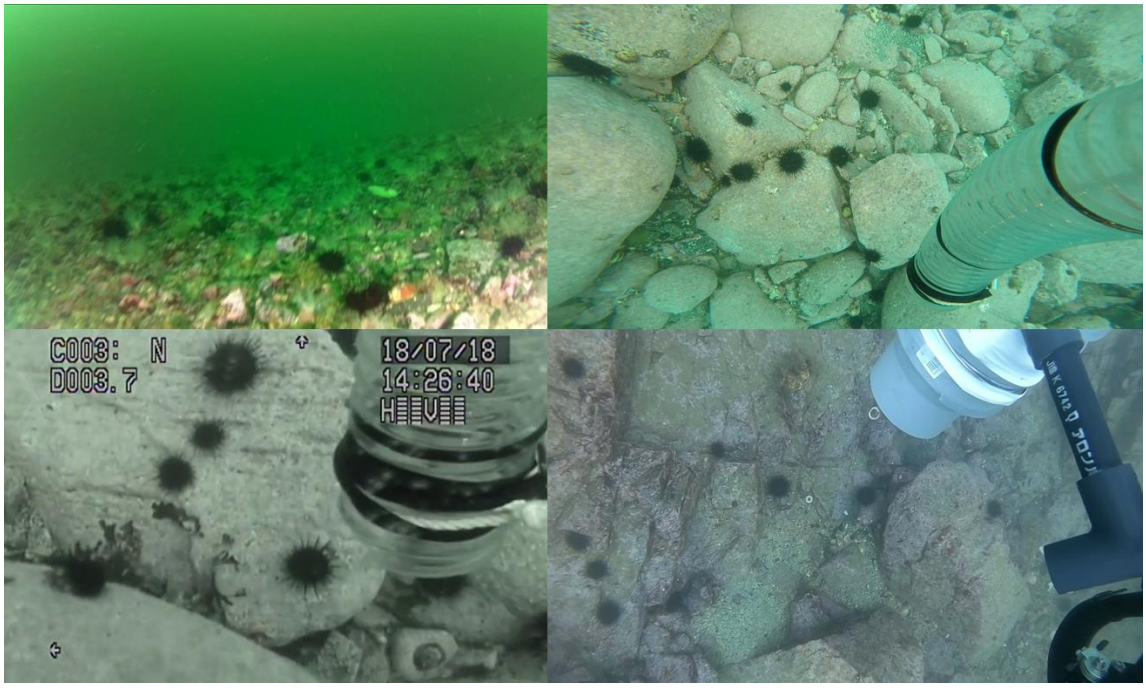

図 4-1　2018 年 3,7,9 月の現地取得画像

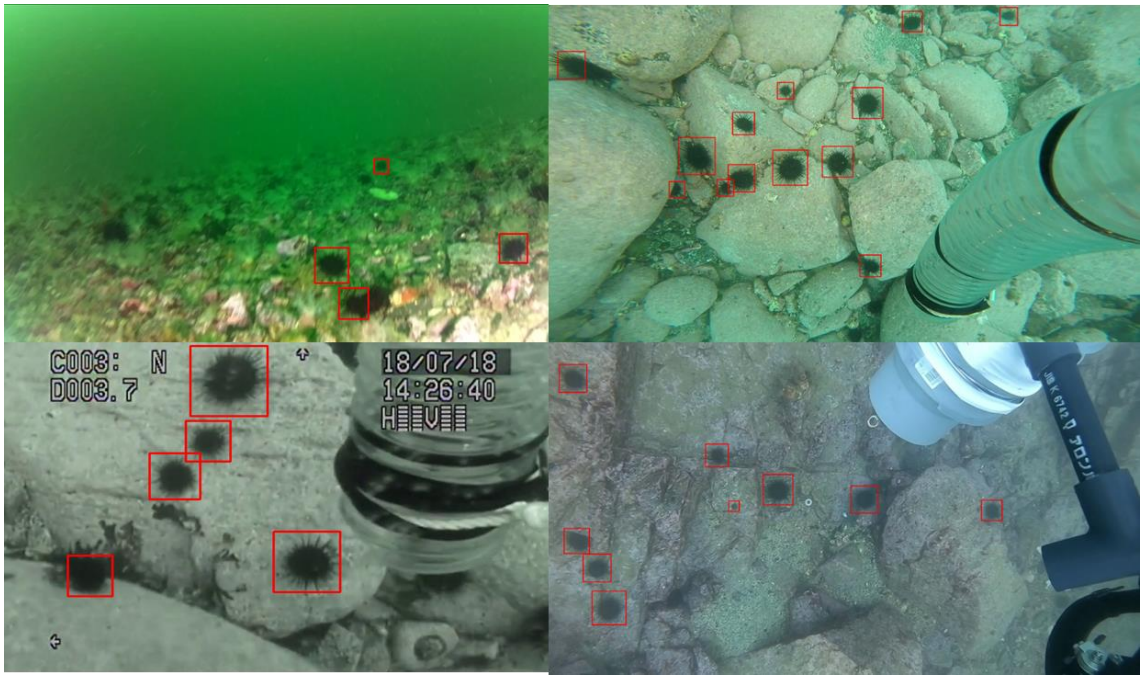

図 4-2　学習器 a を用いて画像認識した結果

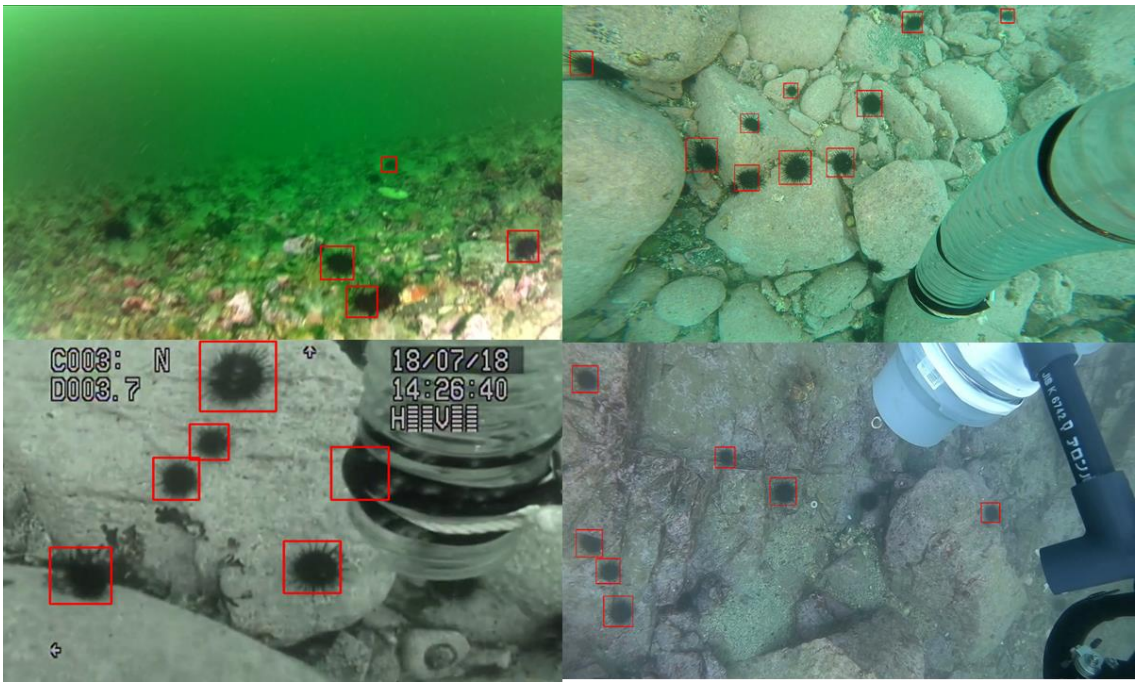

図 4-3　学習器 b を用いて画像認識した結果

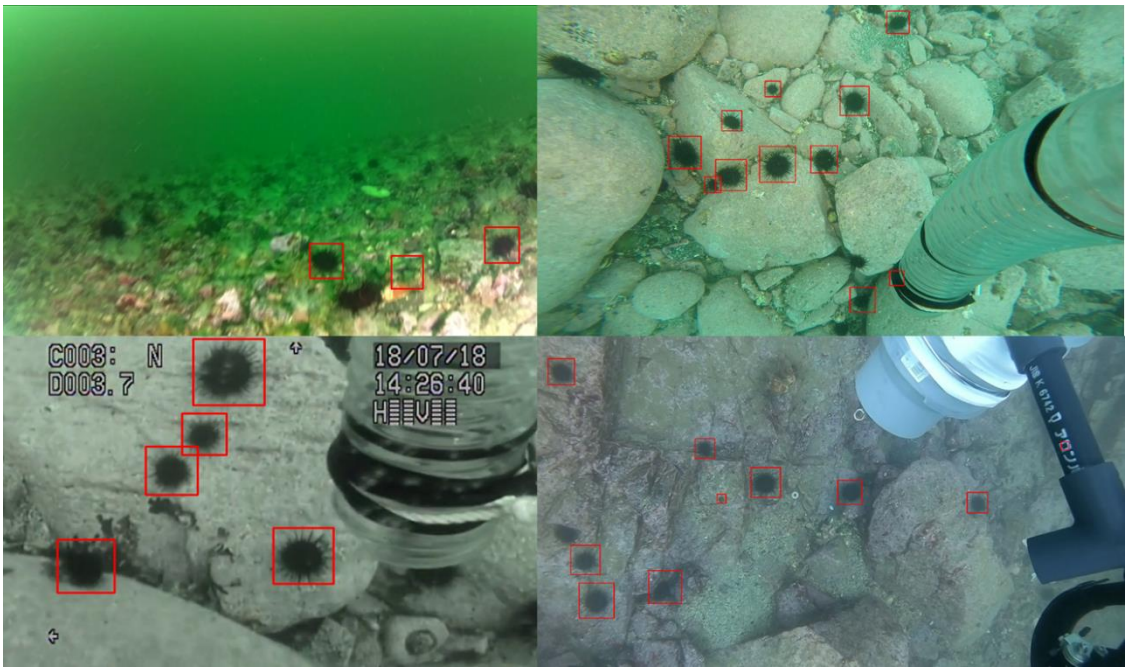

図 4-4　学習器 c を用いて画像認識した結果

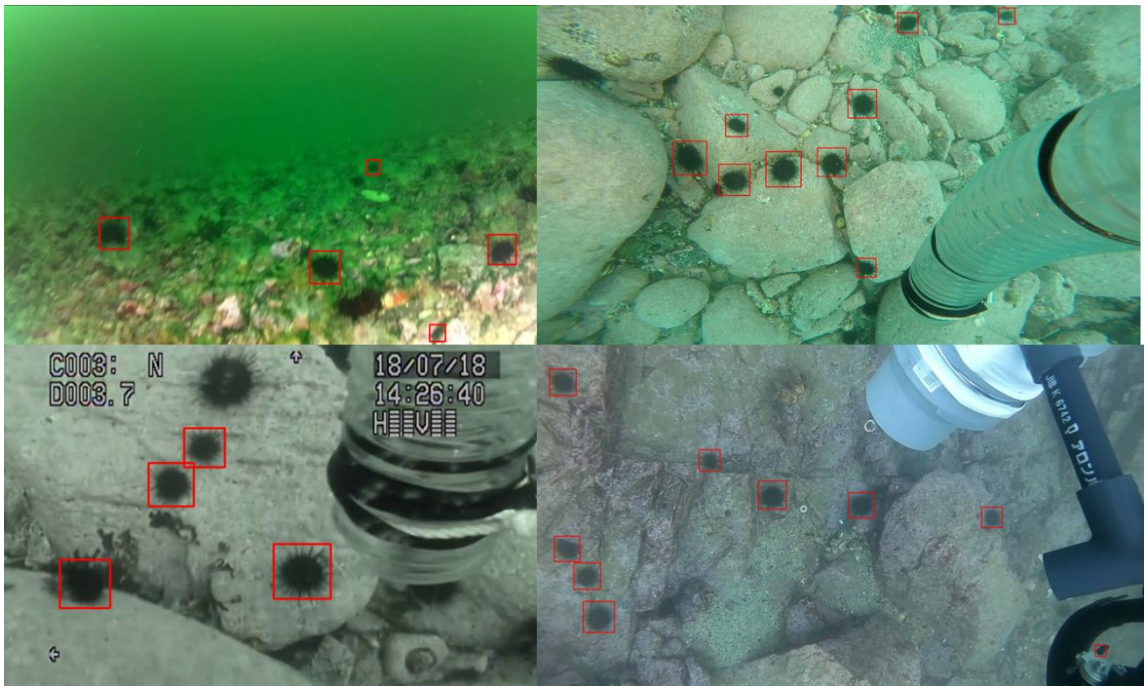

図 4-5　学習器 d を用いて画像認識した結果

表 4-1 各学習器の画像認識結果

学習器	ポジティブ	ネガティブ	D	R	V
a	現地アノテーション画像	現地アノテーション画像	0.829	0.928	0.770
b	ウェブサイト取得画像	ウェブサイト取得画像	0.585	0.968	0.566
c	現地アノテーション画像	ウェブサイト取得画像	0.577	0.986	0.569
d	ウェブサイト取得画像	現地アノテーション画像	0.820	0.766	0.629

5.　学習器に用いるサンプル数の調査

学習器に用いる画像群は現地画像が適していることがわかったため，本章ではその画像群の適切な数について調査した.

5.1.　10,000 枚のサンプル収集

最小サンプル数 400 での学習器作成において現地画像で教師データの作成が適していることがわかったため，次に 10,000 枚のサンプル数を集めた.

400 枚程度のサンプルを収集した際, windows のペイントソフトウェアを用いたが，非常に時間がかかるため，自作のアノテーションソフトを導入した.

10,000 枚の画像群の切り出しはアノテーションソフト以下の手順で行った.

・Step 1　動画データの取得

宮城県志津川湾にて 2018 年 7 月の現地調査より取得した動画データ(NTSC 形式：解像度 720×480pixel)から計 1 時間 30 分程度の動画に切り出し.

・Step2　画像データの取得

切り出した計 1 時間 30 分程度の動画から 0.5 秒刻みで画像を切り出し

・Step3　サンプルの取得

3.1 で 1 時間当たり 30 枚程度しか切り取れなかったため, processing にて作成したアノテーションソフトウェア(図 2-15)により切り出しを行った．ソフトの使用方法は付録に記載する．切り出したポジティブサンプルは 3,600 枚，ネガティブサンプルは 7,000 枚であった.

切り出し枚数は 1 時間当たり 512 枚であったため，サンプル数 10,000 枚の総切り出し時間は 19 時間ほどであった.

切り出したポジティブサンプル群の一部を図 2.2 に，ネガティブサンプル群の一部を図 2.3 に示した.

5.2.　学習器作成

5.1.で取得したサンプルを用いて，ポジティブサンプル 3：ネガティブサンプル 7 の比を固定とし，1,000 枚から 1,000 枚ずつ 7,000 枚までと 10,000 枚の 8 つの学習器を作成した．なお，10,000 枚以外はランダムで画像を選択して作成した.

5.3.　画像認識結果

認識率及び発見率，作成時間について調査した．調査に用いた画像は, 2019 年 7,8 月に現地で取得した計 3 時間程度の動画(NTSC 形式:解像度 720×480pixel, HD 形式:解像度 1020×760pixel)から切り出した 9 枚の画像(図 5-1)である.

画像認識した結果，学習器の枚数が変化しても認識率，発見率の変化はほぼ横ばいであった(図 5-2)．そのため，枚数を増やしても両値は変わらないと考えられる．しかし, 5,000 枚以下だと上下に両

値が振れていることがわかった．特に発見率については 2,000 枚で大きく飛び出していることがわかった．これはランダムに画像を選んで学習器を作成しているため，画像群の質が大きく影響したのではないかと考えられる．画像の質とは，ポジティブサンプル及びネガティブサンプルのどちらにもいえることだが，画像の特徴量によるものが大きい．質の良い画像とは，画像の輪郭がよくわかるようなもので，具体的には，ウニやヒトデと画像を見てはっきりわかるものを指す．反対に質の悪い画像とは，画像の輪郭を捉えにくい，岩場や海中画像等を指す．教師データに用いる質の良い画像と悪い画像の比率により上下にそれぞれの数値が振れたのではないかと考えられる．また，認識率と発見率は一方が上がれば一方が下がるというような逆の関係性がみられた．これは式でも表すことができ，まず式(1)を'm='の形にすると式(4)になる．次に式(4)を式(2)に代入すると式(5)となる．

$$\mathrm{m} = \frac{k}{R} \quad \cdot\cdot\cdot\cdot\cdot (4)$$

$$\mathrm{D} = 1 - \frac{n - \frac{k}{R}}{n}$$

$$\mathrm{D} = \frac{k}{nR} \cdot\cdot\cdot\cdot\cdot (5)$$

5.4.　2,000 枚の学習器

図 5-2 に示すとおり，2,000 枚の学習器の発見率および認識率が大きく外れていた．5.3.に示すように，画像の質が影響していると考え，他に 2 つの 2,000 枚の学習器をランダムで作成し，3 つの 2,000 枚の学習器の発見率を比較した．5.2.で作成した 2,000 枚の学習器を A とし，他の 2 つの学習器を B, C とした(表 5-1).

3 つの 2,000 枚の学習器の発見率の平均をとったところ，他の枚数で作成された学習器とほぼ同じような値となった(図 5-3)．つまり学習器に用いる画像の質は重要で，画像の質により学習器に影響を及ぼすことがわかった．

5.5.　作成時間

枚数変化による学習器作成時間，アノテーション時間を図 5-4 に示す．図 5-4 より学習器作成に用いる教師データの枚数が増えると，指数的に学習器作成時間が上昇しているが，7,000 枚のところで大きく作成時間が飛び出ている．これについても画像の質が大きく影響していると考えられ，2,000 枚学習器の発見率調査時と同様，複数の 7,000 枚の学習器を作成して平均化すれば他の学習器の作成時間の間を補完するような値となると考えられる．しかし本論文では実施をしなかった．

5.6　評価値 CV を用いた学習器評価

学習器に用いるサンプル数を決定するため，発見率および認識率，作成時間を総合的に評価するために評価値 CV を作成した(式(6))．式(6)において，T は学習器作成時間および画像アノテーション画像を含めた総合作成時間を式(7)により無次元化したものである．なお，アノテーション時間は 1 時間あたり 512 枚で計算を行う．

各重みは，本研究では発見率と時間を重視したいため発見率重みは 200，認識率重みは 100，無次

元化時間重みは 200 とした．また，式(5)の分母の 0.01 は T が 0 を取りえないため導入した．

$$\mathrm{CV} = \sqrt{W_D D^2 + W_R R^2 + W_T T^2} \quad \cdot\cdot\cdot \quad (6)$$

$$\mathrm{T} = 1 - \left(\frac{t}{tmax+0.01}\right) \quad \cdot\cdot\cdot \quad (7)$$

t ：総合作成時間
D：発見率
R：認識率
T：無次元化時間
W_D：発見率重み(200)
W_R：認識率重み(100)
W_T：無次元化時間重み(200)

調査した結果，2,000 枚で学習器を作成したほうが，評価値が一番高かった(表 5-2)．これは学習器のサンプル数を増やしても発見率及び認識率の値の変化が小さかったため，学習器の評価が高い傾向にあったためだと考えられる．

また，1000 枚，2000 枚，10000 枚で作成した学習器の画像認識結果を図 5-5 に示す．図 5-5 より，1000 枚と比べて 2000 枚の学習器では認識したウニの数が増えていることがわかる．10000 枚の学習器ではさらにウニの認識した数が増えていることがわかるが，10000 枚だと 1 日で学習器の作成が終了しない．そのため評価値 CV(表 5-2)のとおり 1 日で学習器の作成が終了する 2000 枚の学習器が適当だと考えられる．

5.7. サンプル群再調査

学生により集められた画像群を用いた学習器では認識率が低かった．認識率は画像の質に大きく影響すると考えられるため，学習器に用いたサンプルを確認した．ポジティブサンプルについてはウニであったが，ネガティブサンプルにはヒトデや貝が少なく岩場や ROV に搭載されているホースの画像が多かった．比率にすると 1:9 程度であり，非常に少ないことがわかった．そのためネガティブサンプルの質が重要であると考え，ネガティブサンプルの質を向上させ，評価を行った．

5.7.1. 良質な画像 1,000 枚による学習器再構成

画像の質を向上させるためにヒトデや貝の割合を増やして 1,000 枚の学習器を作成した．画像の選別には 700 枚行い，2 時間ほどを要した．

2019 年の 7，8 月に取得した計 3 時間程度の動画(NTSC 形式：解像度 720×480pixel，HD 形式：解像度 1020×760pixel)から 9 枚の画像(図 5-1)を切り出し画像認識させた結果，ネガティブサンプル

の質を高めると表 5.3 に示したとおり, 10 パーセントほど認識率が向上した(図 5-6)

5.7.2. 良質な画像 2,000 枚による学習器再構成

ヒトデや貝の枚数が少なくネガティブサンプルの必要枚数に満たないため ImageNet[12]と呼ばれる画像収集 Web サイトから画像を集めて 2,000 枚の学習器を作成した. 現地画像の選別に 2 時間, ImgaeNet による画像収集はネガティブサンプルを 700 枚集めるのに 3 時間の計 5 時間を要した. 2019 年の 7, 8 月に取得した計 3 時間程度の動画(NTSC 形式：解像度 720×480pixel, HD 形式：解像度 1020×760pixel)から 9 枚の画像(図 5-1)を切り出し画像認識させた結果, ネガティブサンプルの質を高めると表 5.4 に示したとおり, 12 パーセントほど認識率が向上した(図 5-7).

5.7.3. 1,000 枚および 2,000 枚の学習器変化

1,000 枚及び 2,000 枚の画像の質による学習器の変化を図 5-8 に示す. 図 5-8 より 2,000 枚の学習器のほうが認識率の上昇率が 1,000 枚の学習器よりも 2 パーセントほど高いことがわかった.

5.7.4. 画像の質

ネガティブサンプルの質を向上させたことにより認識率が向上した. そのため認識率を重視したい場合にサンプル群の調整を行うことは効果が高いと考えられる.

しかし, 時間が多くかかることと, 切り出す人により画像の質が左右されてしまうことが問題である. そのため GA 等で, 自動で質の良い画像を選別するようなシステムが必要であると考えられる.

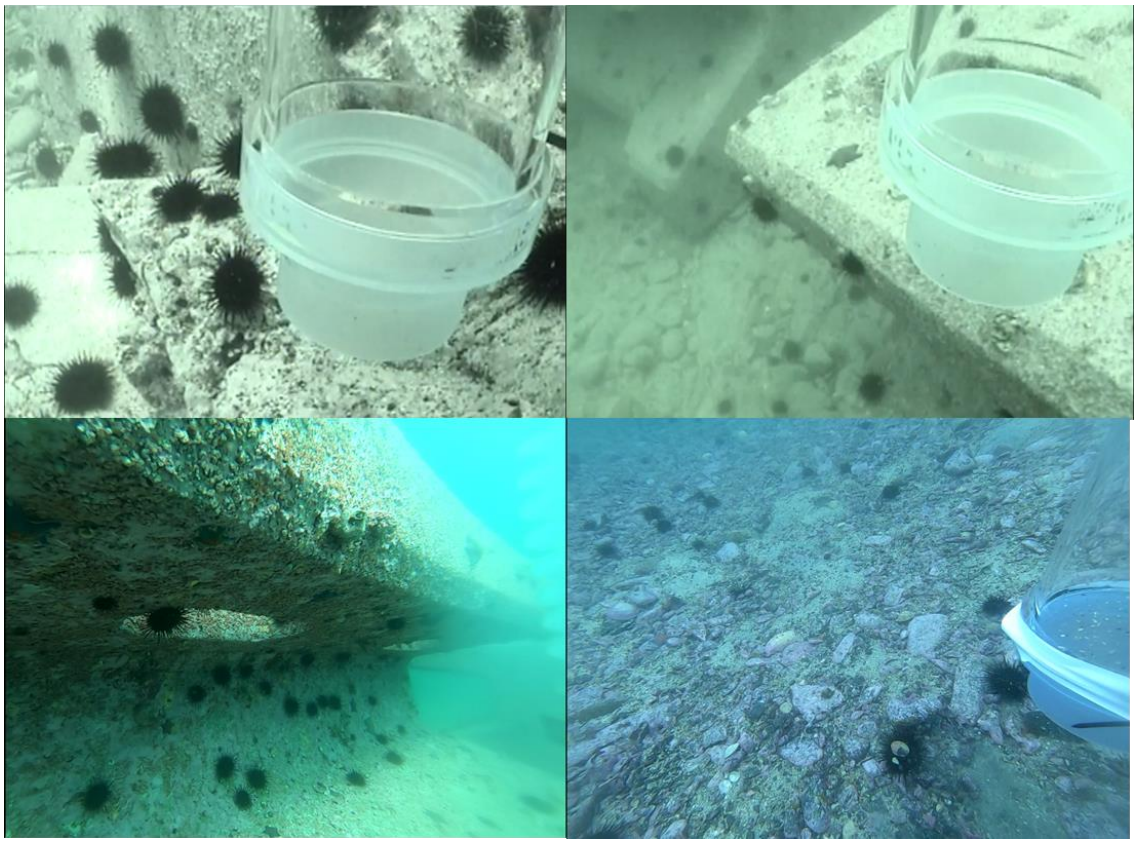

図 5-1　2019 年 7,8 月の現地取得画像

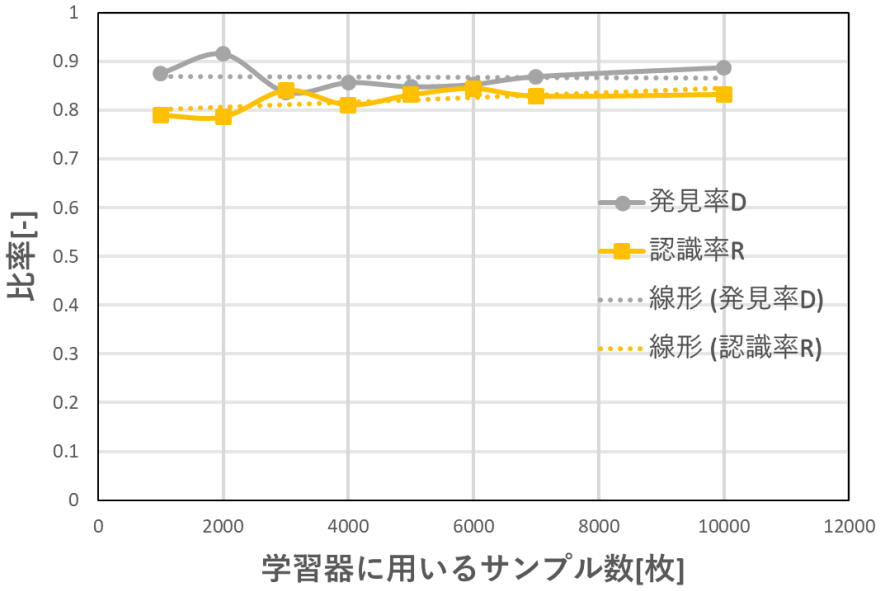

図 5-2　発見率及び認識率

表 5-1 発見率

学習器	発見率 D
2000 枚 A	0.916
2000 枚 B	0.808
2000 枚 C	0.924
avr.	0.862

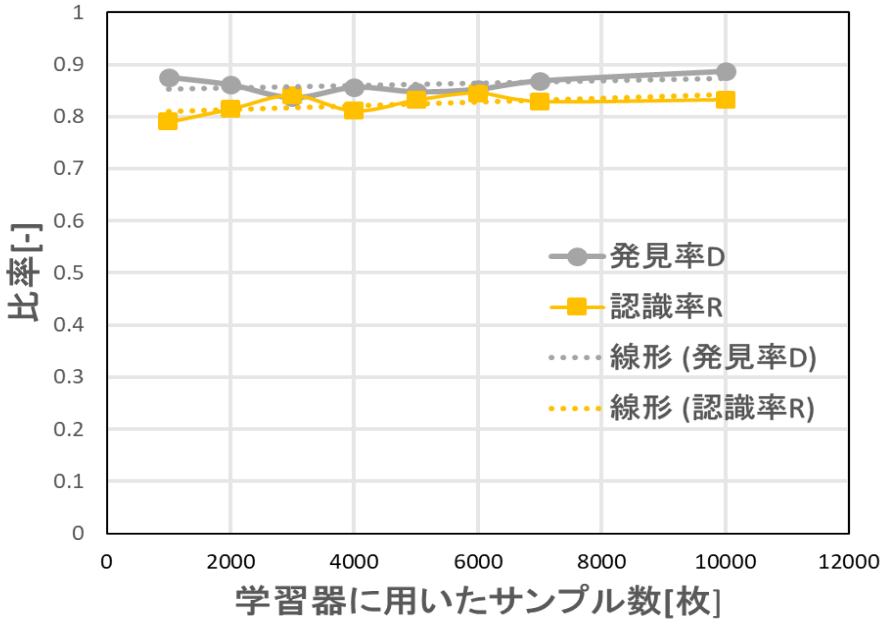

図 5-3　3 つの 2000 枚の学習器の平均値を用いた場合

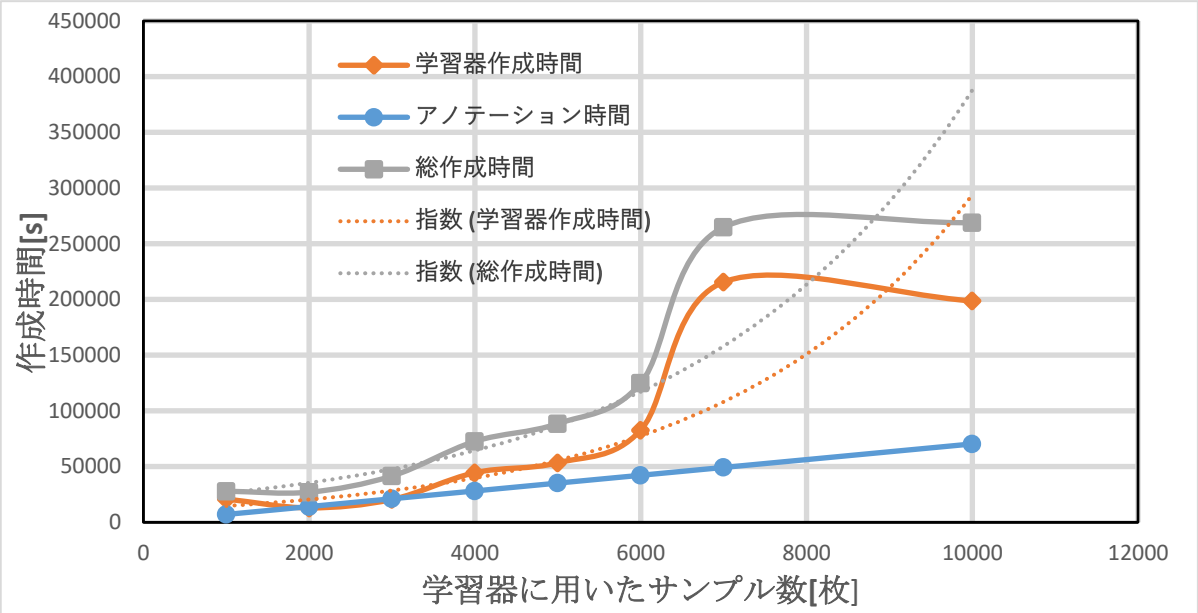

図 5-4　学習器に用いるサンプル数による作成時間の変化

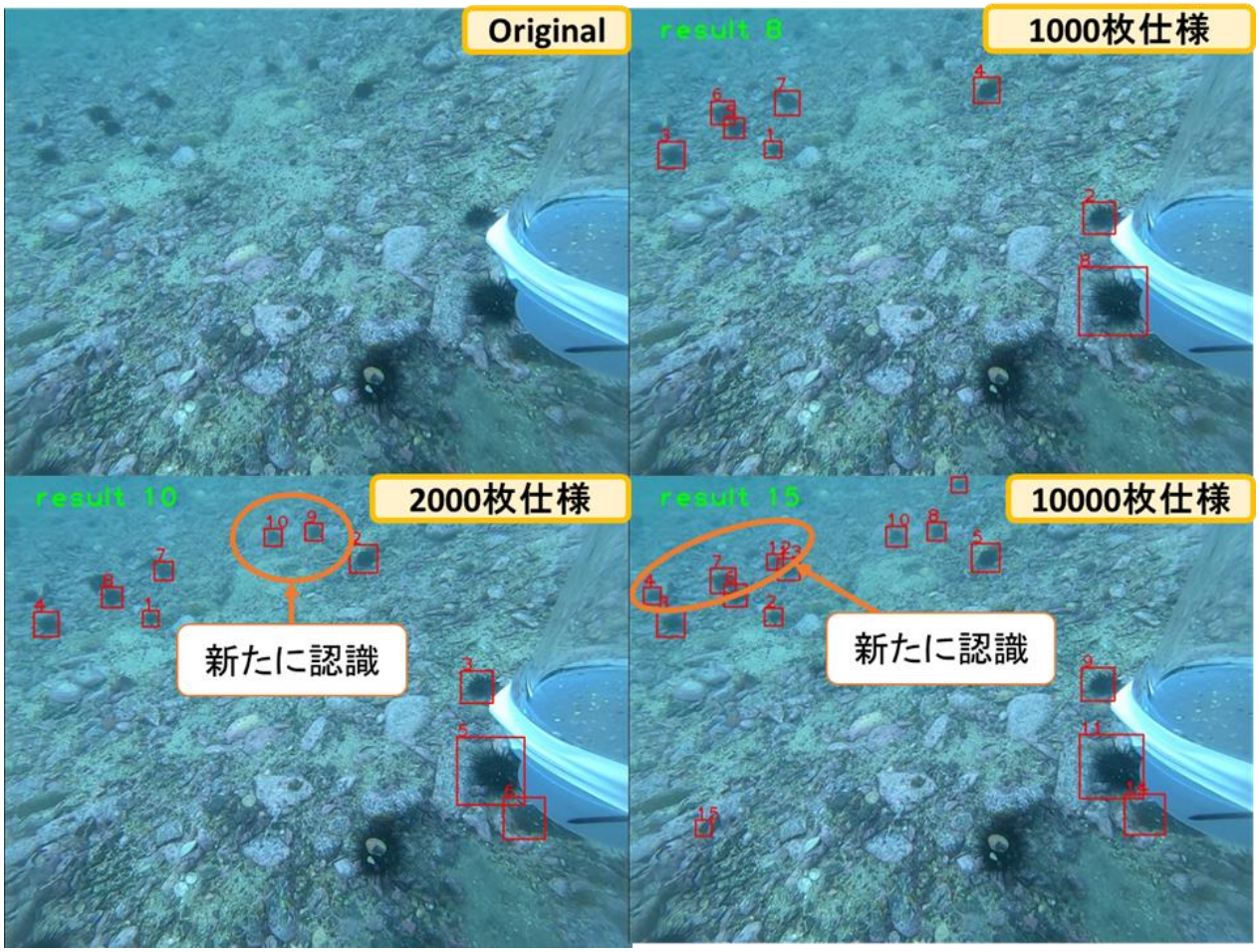

図 5-5　学習器による認識結果の違い

表 5-2　評価値

枚数	発見率	認識率	学習器作成時間[s]	アノテーション時間[s]	総合作成時間[s]	総合作成時間無次元化	総合評価
1000	0.876	0.790	20866	7031	27897	0.896	19.404
2000	0.916	0.787	12924	14063	26987	0.900	19.784
3000	0.837	0.840	20467	21094	41561	0.845	18.805
4000	0.857	0.811	44534	28125	72659	0.730	17.862
5000	0.848	0.831	53376	35156	88532	0.671	17.406
6000	0.853	0.845	82712	42188	124900	0.536	16.560
7000	0.869	0.829	215583	49219	264802	0.015	14.825
10000	0.888	0.832	198617	70313	268930	3.72E−08	15.061

表 5-3　学習器に用いるサンプル数による認識率(1000 枚)

学習器	認識率_R
不良 1000 枚	0.790
良 1000 枚	0.887

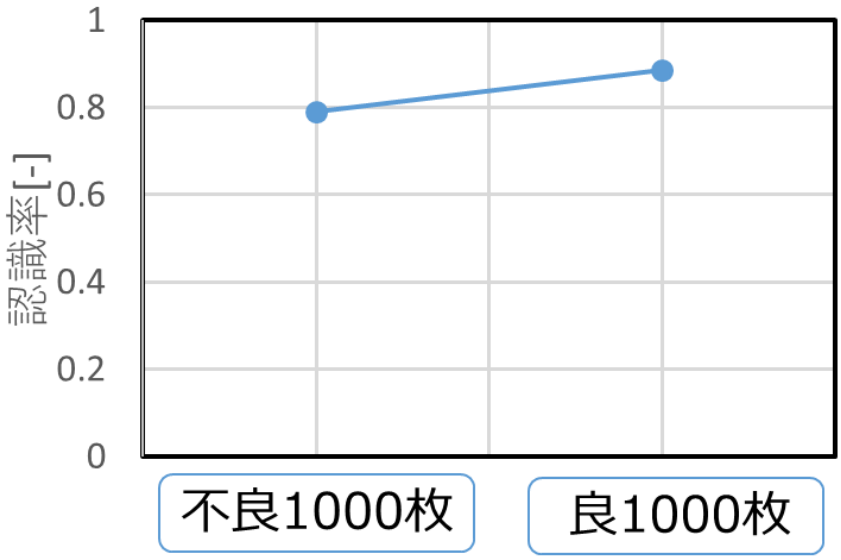

図 5-6　1000 枚の学習器の比較

表 5-4　学習器に用いるサンプル数による認識率(2000 枚)

学習器	認識率_R
不良 2000 枚	0.787
良 2000 枚	0.920

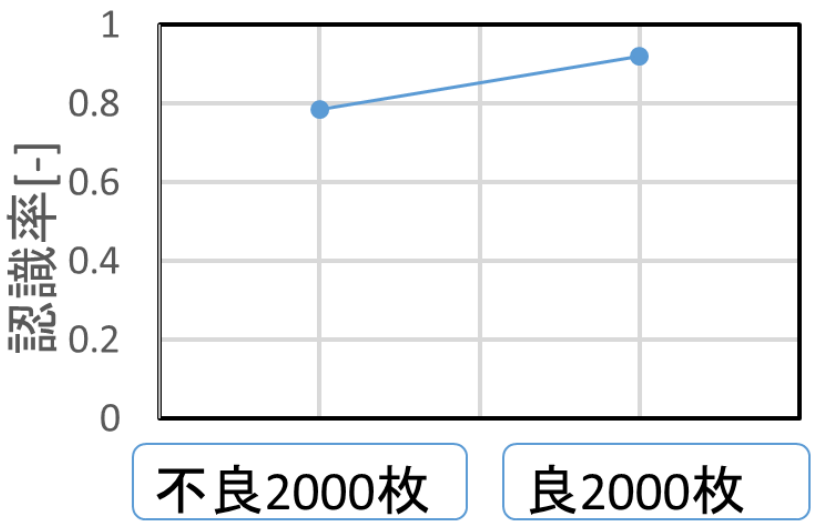

図 5-7　2000 枚の学習器の比較

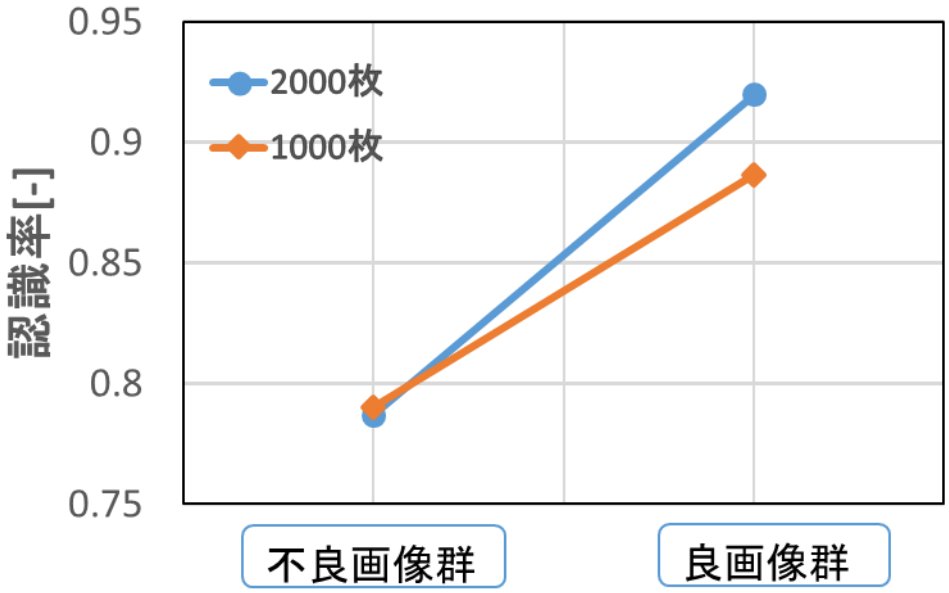

図 5-8　1000 枚, 2000 枚の学習器比較

6. まとめ

6.1.　学習器について

1 日で出来る学習器を作成するには，2,000 枚が適当であることがわかった．なおサンプル比率は 3:7，画像の種類はできるだけ現地画像で集めたほうがいいという結果となった．本研究で集められた画像はネガティブサンプルについて質の悪い画像であったと思われ，今後質のいい画像を現地画像から多く切り出して学習器を作成する必要があると考えられる．ただし，現状では質のいい画像を集めるのは困難であるため，今後は GA 等，新規アルゴリズムにより現地で取得した画像から質のいい画像を収集する必要があると考えられる．

質のいい画像を収集するには，質のいい画像を収集する人を雇うことや画像収集を専門にする業者に依頼することが考えられる．また，遺伝的アルゴリズム(GA)を用いた画像収集も考えられる.この手法では，まず座標から自動で画像を切り出すソフトウェアを作成する．さらに認識ソフトを用いてウニを発見し，ウニ部のみを切り出す．こうすることでウニを切り出さないようにする．次に特徴量のはっきりした画像をいくつか用意し，特徴量を数値化．最後にヒトデや貝を発見する学習器，自動で画像を切り出すソフトと GA を用いて，数値化した値に近いもの(評価の高いもの)を切り出すように進化をさせる．このように行うことで多くの質のいい画像が入手できると考えられる．

6.2.　画像認識システムについて

現在のシステムでは ROV のカメラ映像を，テザーを介して船上 PC で受信し画像認識を行っているが，今後はその結果を複数の PC で共有することを目指す．

また，追加機能としてレーザースリットとステレオカメラによりウニのサイズ測定を目指す．ステレオカメラによるステレオ視とレーザースリットによる格子表示をすることでサイズを知ることができる(図 6-1)．回収するウニは殻径がφ20mm から 60mm となっており，本機能を用いればターゲットのサイズのみを回収できるようになるため有用だと考えている．

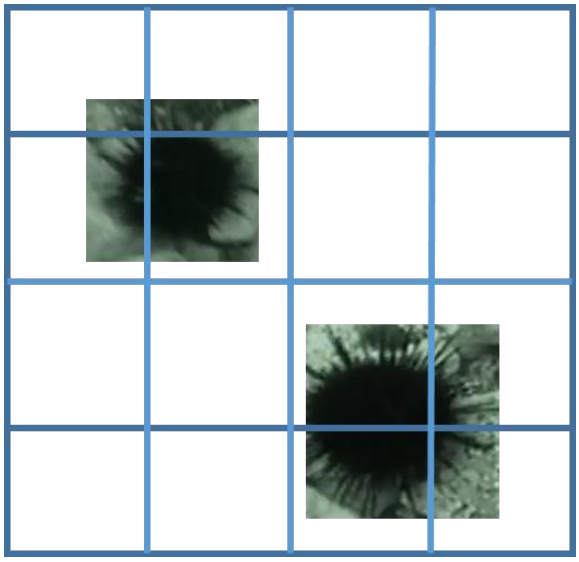

図 6-1　レーザースリットによるウニ測定

参考文献

[1] 川村大和，田原淳一郎，和泉充，井田徹哉，阿部拓三, “小型 ROV を用いた生物調査システム”，海洋理工学会平成 30 年度春季大会, pp39-40, 5.2018

[2] 加藤哲，川村大和，田原淳一郎，和泉充，井田徹哉，阿部拓三, “小型 ROV システムの動作特性”，海洋理工学会平成 30 年度春季大会, pp65-68, 5.2018

[3] 川村大和，SonMunseong，齋藤幹大，伊藤魁，加藤哲，田原淳一郎，和泉充, “ROV を用いたウニ駆除システム”，海洋理工学会 2019 年秋季大会, 11.2019

[4] 伊藤魁，後藤慎平，和泉充，井田徹哉，田原淳一郎, “ウニ回収 ROV の回収システムの概要と実地試験結果”，海洋調査技術学会第 31 回研究成果発表会, pp25-26, 11.2019

[5] P. Viola , M. Jones, “Rapid object detection using a boosted cascade of simple features”, “Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001”., 8-14 Dec. 2001

[6] R. Lienhart , Maydt , “An extended set of Haar-like features for rapid object detection” , “ Proceedings. International Conference on Image Processing” , 22-25 Sept. 2002

[7] MathWoks ドキュメンテーション,オブジェクトのカスケード検出器の学習
https://jp.mathworks.com/help/vision/ug/train-a-cascade-object-detector.html

[8] Yoav Freund , Robert E. Schapire , “Experiments with a new boosting algorithm” , “Proceedings of the Thirteenth International Conference on International Conference on Machine Learning” , July 1996

[9] 森のどか，會澤要，鈴木拓央，小林邦和, “カスケード型分類器を用いた照明環境変化にロバストなサッカーボール認識” , 人工知能学会 AI チャレンジ研究会 , 2018

[10] Feras Dayoub , Matthew Dunbabin , Peter Corke , “Robotic detection and tracking of Crown-of-Thorns starfish” , 28 Sept.-2 Oct. 2015

[11] 齋藤幹大，田原淳一郎，加藤哲，川村大和, “機械学習手法を用いた画像認識システムの開発”，ロボティクス・メカトロニクス学会, p29, 6.2019

[12] ImgageNET, http://www.image-net.org/

謝辞

本研究の成果は農林水産省食料生産地域再生のための先端技術展開事業「異常発生したウニの効率的駆除及び有効利用に関する実証研究」で助成を受け，「ウニと藻場の豊かな漁場再生コンソーシアム」が行ったものです．ここに感謝の意を表します．

また，本研究において多くの助言をくださった指導教員である田原淳一郎准教授には感謝いたします．さらに，研究を進めていくなかで相談にのっていただいた川村大和，加藤哲，孫ムン成修士学生，小野聡太郎，森戸誠両学部生にも感謝の意を表します．

付録

以下に本研究で使用したプログラムリストを示す.

List1　画像認識する際に１枚ずつ画像認識するプログラム群

１枚ずつ画像認識行う場合の画像認識プログラムは以下の No.1 から No.2 で構成される.

No	ファイル名	内容
1	uni_ninsiki_dn2.py	ウニを認識するプログラム
2	gazou_ninsiki_x_y_2.py	No1 における関数プログラム
場所	https://www.dropbox.com/sh/z3hr0mxuxs8x31h/AADF0frUW9I53alxQVX9CpfMa?dl=0	

List2　動画を画像認識するプログラム群

動画を認識したい場合に用いる画像認識プログラムは以下の No.1 から No.3 で構成される.

No	ファイル名	内容
1	python_test_button_text_r6.pde	GUI による動画の認識ソフトウェア
2	python_comand_pros_5.py	ウニを動画認識するプログラム
3	gazouninsiki_x_y_5,py	No2 における関数ファイル
4	動画解析 GUI マニュアル.docx	GUI の使い方マニュアル
場所	https://www.dropbox.com/sh/m21xpcmr17qimr4/AAC2LyIkuiQirxdpEfaAvBRza?dl=0	

List3　リアルタイムでカメラ映像を認識するプログラム群

カメラ映像を認識した場合に用いる画像認識プログラムは以下の No.1 から No.5 で構成される.

No	ファイル名	内容
1	device_list_hyouji4_.sh	PC に接続された USB 情報の取得
2	count_pros_f_r4.pde	GUI によるリアルタイム画像認識ソフトウェア
3	camera_config.py	デバイス情報の読み込みを行うプログラム
4	python_comand_pros_4.py	ウニをリアルタイムで認識するプログラム
5	gazou_ninsiki_xy4.py	No4 における関数プログラム
6	リアルタイム認識 GUI マニュアル	GUI の使い方マニュアル
場所	https://www.dropbox.com/sh/l9365pxjex9uug4/AADNei42FEpXFM99d6DXjO9ta?dl=0	

List4　教師データ作成プログラム群

画像認識に必要な学習器(カスケードファイル)作成の際に用いるプログラムは以下の No.1 から No.6 で構成される．ただし，No.1 から No.3 はポジティブサンプル，No.4 から No.6 はネガティブサンプルの作成プログラム群である．

No.	ファイル名	内容
1	uni_image_size_text_before.py	ポジティブサンプルの画像のナンバリングするプログラム
2	uni_image_size_text_int2.py	ポジティブサンプルのリスト化するプログラム
3	cascade_pos_test_r3.pde	GUI によるポジティブサンプル作成ソフトウェア
4	ポジティブサンプル作成 GUI マニュアル	ポジティブサンプル作成 GUI マニュアル
5	vec_make.sh	ポジティブサンプルをベクトルファイル化するプログラム
6	uni_neg_text_int.py	python，ネガティブサンプルをリスト化
7	neg_make.pde	GUI によるネガティブサンプル作成ソフトウェア
8	ネガティブサンプル作成 GUI マニュアル	ポジティブサンプル作成 GUI マニュアル
場所	https://www.dropbox.com/sh/f91gu2imblaq17u/AADEMtAnOtsFpItfDC9zfvLka?dl=0	

List5　画像切り出し(アノテーション)ソフトウェアの使い方

画像の切り出しに用いるアノテーションソフトウェアのプログラム及びマニュアル．

No	ファイル名	内容
1	mouse_click_cut6.pde	GUI による画像の切り出しソフトウェア
2	anotemanual.pptx	アノテーションソフトウェアの使い方マニュアル
場所	https://www.dropbox.com/sh/i3u4fttbn4zyzou/AADmPxHg_upr8AArJJL8NAw5a?dl=0	