

# 史跡等屋外展示施設におけるAR案内システムの開発

津波古 正寿<sup>1,a)</sup> 赤嶺 有平<sup>2</sup> 根路銘 もえ子<sup>3</sup>,

**概要:** 本稿では、史跡等に代表される屋外展示施設における利用を想定し、普及型のモバイル端末上で実行可能な、AR（拡張現実）を用いた展示物の自動案内システムについて述べる。ARシステムを実現するためには「重ね合わせ」の問題を解決する必要があるが、モバイル端末の内蔵センサのみでは十分な精度が得られないため画像認識を用いる必要がある。しかしながら、屋外環境においてARを行う際には、光源位置の移動に伴う陰影変化による物体認識の精度低下が問題となる。本研究では、屋外環境下においてモバイル端末上で利用可能な速度で物体検出が可能なアルゴリズムを目指し開発を行った。さらに、文化施設の案内で必要となる説明文等の「長いテキスト」を視界を妨げることなくかつ直感的に対象物との関連を理解できるように考慮したビューマネージメント手法を提案し、評価を行った。

キーワード：AR, 自動案内システム, 屋外環境, ビューマネージメント

TSUHAKE MASATOSHI<sup>1,a)</sup> AKAMINE YUHEI<sup>2</sup> NEROME MOEKO<sup>3</sup>,

## 1. はじめに

観光は、ARの応用が大きく期待される分野の一つである。観光客に対して、モノ、場所等の情報をARを用いて提示することや、3DCG等を用いてその場所に適したデジタルコンテンツを視覚情報に重畳表示することで、ユーザ体験の大きな向上が期待でき、あらゆる観光地の価値向上が実現できる。観光への応用では屋外環境下での安定的な動作が必要不可欠である一方で、後に述べる理由により屋内よりも困難な課題である。

ARを実現する上で最も重要な技術は、「位置合わせ (alignment)」である。ARの最も大きなメリットは、対象物と情報 (アノテーション) との関係を直感的に認識できることにあり、位置合わせでは、リアルタイム性と正確性が特に重視される。提示の遅れはユーザ体験 (UX) を大きく低減し、不正確な位置への提示は、ユーザに大きな混乱をもたらす。

位置合わせを行うには、カメラと対象物の相対的な位置関係を推定する必要がある。位置関係の推定には、二つの代表的なアプローチがある。一つは、センサーを用いて直接測定する手法である。GPS、ジャイロ、加速度センサーを複合的に用いて実装するのが一般的である。これらのセンサーは、スマートホン等ほとんどのモバイルデバイスで利用可能であり、CPUの計算負荷も小さいことから多くのARアプリケーションで利用されている。しかしながら、GPSの測位精度は数m程度であり、cmレベルの物体を識別することはできない。また、地磁気センサは頻繁にキャリブレーションを行う必要があり利便性に課題がある。

他方は、画像処理を用いる手法である。2次元の平面的な情報を重畳表示するアプリケーションにおいては、キャプチャイメージ上の対象物体の2次元位置を検出するだけで実現可能である。しかしながら、屋外のように方向が大きく変化する強光源下では影や陰影のコントラストが非常に高く、「見え」の変化が大きいため、推定精度が大きく低下する。

特定検出の手法として、畳み込みニューラルネットワークを始めとする深層学習 [1] を応用した手法が数多く提案されており、高い性能を持つことが示されている。しかしながら、deep learning は、1) 訓練データを大量に必要とする、2) 学習時間が膨大である、3) 検出時の計算量が大きく、リアルタイムな実行にはGPU等の消費電力の大きいハー

<sup>1</sup> 琉球大学院 理工学研究科  
UNIVERSITY OF THE RYUKYUS, Nishihara, Okinawa  
903-0213, Japan

<sup>2</sup> 琉球大学 工学部  
UNIVERSITY OF THE RYUKYUS, Nishihara, Okinawa  
903-0213, Japan

<sup>3</sup> 沖縄国際大学 経済学部  
OKINAWA INTERNATIONAL UNIVERSITY, GI-  
NOWAN, OKINAWA 901-2701, JAPAN

a) k178581@ie.u-ryukyu.ac.jp

ドウェアを必要とする、等の問題がある。特に 1),2) の学習に関する課題は、利用者の大きな負担となることから、現時点では AR アプリケーションにおける物体検出手法として実用性に問題がある。

本研究の目的は、屋外のような方向が変化する強光源下において、必要最小限の訓練データと学習時間、かつ CPU のみでリアルタイム実行可能な特定物体検出アルゴリズムの開発および、屋外観光施設における案内システムへの応用である。本稿では、計算量の少ない、軽量の画像特徴の空間的な分布を統計的に用いることで見えの変化にロバストな特定物体検出アルゴリズムである、SHISAR[2] について述べ、検出物体への情報提示を行うための「ビューマネージメント」を含めて実装した観光施設ガイダンスシステムについて述べる。SHISAR では、画像特徴は量子化して情報量の削減を行うことで、照明変化に対するロバスト性を確保する。

## 2. 軽量特徴を用いた物体検出アルゴリズム (SHISAR:SHape of Image features as Spatial information for AR)

SHISAR は、画像内における色 [3] や輝度勾配方向 [4] といった特徴から算出した軽量特徴を用いて物体検出を行うアルゴリズムである。学習、検出時の処理を次で述べる。

### 2.1 量子化特徴画像の算出と学習

画像中の任意の特徴（色や輝度勾配方向等）に対して量子化を行ったものを、計算量が小さい軽量特徴とする。ここでは、軽量特徴からなる量子化特徴画像を用いて、対象物体の学習を行う処理について述べる。処理の様子を図 1 と次に示す。

- (1) 対象物体を様々な位置、方向から撮影した画像集合  $I_0$  を用意する
- (2) 画像集合  $I_0$  のそれぞれの画像について
  - (a) 物体の形状が残る程度に特徴の量子化を行うことで、量子化特徴画像  $Q$  を作成する
  - (b) 量子化特徴画像  $Q$  の要素  $Q(x, y)$  を特徴コード  $c$  とした時、要素ごとに各  $c$  の出現回数を表した頻度分布  $D(c)$  を算出する ( $D(c)$  は量子化特徴画像と同じサイズの二次元配列)。
  - (c)  $I_0$  全ての画像に対して同様に処理を行う
  - (d)  $D(c)$  を  $x, y$  軸に対してそれぞれ反転する

### 2.2 スコア分布の算出と対象物体の検出

学習結果を用いて、物体の存在位置を求めるためのスコア分布を算出する。スコア分布の各要素は学習した対象物体である可能性の大きさを表しており、これをスコア値とする。値が大きい位置ほど、対象物体が存在する可能性が

大きい位置であると判断することができる。処理の流れを次に示す。

- (1) 入力画像から、学習処理時と同様に特徴を量子化した量子化特徴画像  $Q'$  を作成する
- (2) 入力画像と同じサイズで各特徴コードに対応するスコア分布  $W(c)$  を用意する ( $W(c)$  は二次元配列であり 0 で初期化)
- (3)  $Q'$  の要素  $Q'(x, y)$  から得られる  $c$  について、図 2 のように、対応する頻度分布  $D(c)$  を  $W(c)$  の原点  $x, y$  から  $D(c)$  と同サイズの部分配列へ加算する
- (4) 上記 (3) を画像内の要素を走査しながら行う
- (5) 入力画像における座標ごとの検出対象物体のスコア値  $l(x, y)$  を式 (1) から算出する。出現数が 0 の特徴コードがある際に結果が 0 にならないように  $\alpha$  を加算する。

$$l(x, y) = \prod_c W(c, x, y) + \alpha \quad (1)$$

- (6)  $l$  の中で最も値の大きい点  $(x, y)$  を検出位置とする  
算出されたスコア分布の例を図 3 に示す。実際の入力画像と重ね合わせたものが図 4 のようになっており、学習した物体と似た特徴を持つ部分により強く反応していることがわかる。

### 2.3 検出対象物体の存在有無判定

スコア分布  $l$  の最大値の位置は、対象物体が存在する可能性の最も高い位置を表しているが、その座標上における物体の存在を保証するわけではない。よって、検出候補位置に対象物体が存在しているかどうかを判断するための処理を行う必要がある。以後この処理を検出対象物体の存在有無判定と呼ぶ。

$l$  のスコア値には、学習に用いるデータセットの量、学習や検出時における各パラメータの値によって最大値にばらつきがある。このため、対象物体の存在有無判定を行う手法として、単純にスコア値に閾値を設ける方法だとまづいかなない。そこで、スコア値の度数分布を利用することで一意の閾値を定め、判定を行う手法を提案した。この手法ではスコア値から取得したヒストグラムの形状から判定を行う。処理の流れは次のようになる。

- (1) スコア分布の要素から、スコア値のヒストグラムを作成する
- (2) ヒストグラムの分散、またはエントロピーを算出する
- (3) 経験則的に定めた閾値により存在有無判定を行う

上記 (2) について、度数分布から分散またはエントロピーを取得する際の算出方法を説明する。分散の場合は、 $n$  をビンの総数、 $i$  をビンの番号、 $x_i$  を各ビンに収まるスコア分布の要素数、 $\bar{x}$  を 1 つのビンあたりに収まるスコア分布の平均の要素数とすると式 (2) のようになる。



図 1 学習処理の様子  
Fig. 1 Learning process.

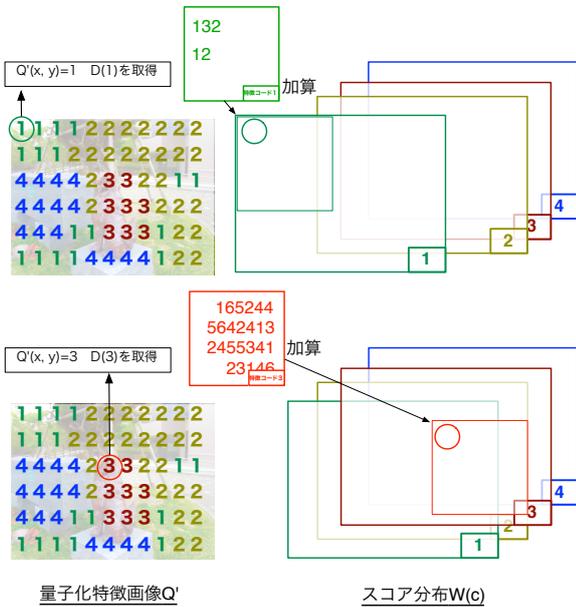


図 2 スコア分布の算出の様子  
Fig. 2 Calculate Score distribution.



図 3 スコア分布例

Fig. 3 Score distribution example.

図 4 入力画像との重ね合わせ

Fig. 4 Overlaying with the input image.



図 5 動画 A1  
Fig. 5 Movie A1.

エントロピー  $H$  が次の式 (3) から求められる。

$$H = - \sum_{i=1}^n p_i \log p_i \quad (3)$$

## 2.4 検出物体の存在有無判定の評価

次の三つのパターンを想定して評価を行った。

- (a) 対象物体が画像内に存在している、かつ検出位置が対象物体上にある
- (b) 対象物体は画像内に存在しているが、検出位置が対象物体上にない
- (c) 対象物体が画像内に存在していない

上記のパターンにおける存在有無の判定結果として、(a) は“存在あり”，(b) では“存在なし”，(c) では“存在なし”とそれぞれ判定できている状態を成功とし、そうでない状態を失敗とする。本節の評価実験ではほこれらの動画を学習し、学習動画とは異なる日時、日照条件でほこれを含むシーンを撮影した動画 (A1, A2, A3) 及びほこれら全く写っていない動画 (B1) を用いた。動画はそれぞれ図 5, 図 6, 図 7, 図 8 に示す。予備実験の結果から、エントロピーは 0.1 を下回った時、分散値は 0.75 を上回った時に“存在あり”として判定結果をつけた。図 9, 図 10 は、エントロピーを用いた検出結果の様子とその時のヒストグラムの形状の例である。また、物体検出には特徴として輝度勾配方向を用いる。

表 1, 表 2 に実験結果の一部を示す。

分母を各パターンに該当するフレーム数、分子をそのパ

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \quad (2)$$

エントロピーを用いる場合は、スコア分布中の特定の位置の要素が  $i$  番目のビンに属する生起確率を  $p_i$  とすると、



図 6 動画 A2  
Fig. 6 Movie A2.



図 7 動画 A3  
Fig. 7 Movie A3.



図 8 動画 B1  
Fig. 8 Movie B1.



図 9 対象物体あり  
Fig. 9 With the target object.



図 10 対象物体なし  
Fig. 10 No the target object.

ターンにおける成功のフレーム数として分数で表している。全体的に正しい判定が行えていることがわかる。分散

表 1 分散を用いた評価

Table 1 Evaluation using dispersion.

動画 No	(a) 学習対象あり	(b) 検出失敗	(c) 学習対象なし	正解率
A1	17/18	0/0	17/19	92 %
A2	10/12	1/2	25/30	82 %
A3	7/7	0/0	18/20	93 %
B1	0/0	0/0	23/30	77 %

表 2 エントロピーを用いた評価

Table 2 Evaluation using dispersion.

動画 No	(a) 学習対象あり	(b) 検出失敗	(c) 学習対象なし	正解率
A1	17/18	0/0	19/19	97%
A2	7/12	2/2	30/30	89%
A3	7/7	0/0	20/20	100%
B1	0/0	0/0	28/30	93%

表 3 処理速度比較

Table 3 Comparison of processing speed

	処理時間
HOG+SVM	5.65
SHISAR	0.153

とエントロピーで多少の差は見られたが、パラメータを調整することで同じような結果が得られると思われる。

## 2.5 既存手法との比較

物体検出の性能を比較するために、HOG 特徴量 [5] と SVM 識別器 [6] による物体検出手法を用いた。評価項目は処理速度とし、結果を表 3 に示す。

この実験では、入力する画像サイズを 960x540、ストライドを 1 としている。ストライドは、検出時においてスコア分布を算出する際に画像を走査する間隔を表す。HOG+SVM による検出の精度に関して、2.4 節で行った実験と同様な状況だとほとんどの場合で 98% 程度の精度が得られたが、要する処理時間が SHISAR と比べて圧倒的に多く必要とされることが表 3 からわかる。これらの点から、モバイルデバイス等のように計算資源に余裕がない状況において、本検出手法は従来の検出手法に対して有利であると考えられる。

## 3. ビューマネジメント

AR コンテンツの付加情報の位置合わせについて、AR システムの用途によって適切な提示方法が異なる。例えば AR によるカーナビゲーションでは、ドライバーの視界を遮らないような配慮が非常に重要である。表示させる情報の種類、数、位置、大きさ、タイミング、透過度、色等、考慮すべき項目は様々である。

本研究のようなガイダンスシステムでは、名称のような短いラベル以外にも、史跡に関する説明文といった、長い文章を表示することも想定する必要がある。説明文は画面上で大きな面積を占めるため、他の注釈や対象物体と被る状況が多くなってしまいう可能性がある。このような視認性

の問題に対して、ビューマネジメントと呼ばれる概念が提唱されている。

### 3.1 関連研究

牧田ら [7],[8] は、注釈を表示させる位置を決定するためにペナルティという概念を取り入れている。注釈同士や対象物体との重なり部分、注釈と対象物体との距離、フレーム間での注釈の移動量をペナルティとして計算しており、ペナルティの最小化を測ることで適切な表示位置を定めている。画面上の位置  $(x, y)$  に注釈を配置する際のペナルティ  $P(x, y)$  は、次の式 (4) から算出される。

$$P(x, y) = w_o P_o(x, y) + w_l P_l(x, y) + w_m P_m(x, y) \quad (4)$$

ここで  $w_o, w_l, w_m$  は各ペナルティにかかる重みの係数を表し、 $P_o, P_l, P_m$  は次のものであるとされる。

$P_o(x, y)$

注釈と線を描画する際に発生する重なりに関するペナルティ項

$P_l(x, y)$

注釈と対象物体を結ぶ線の長さに関するペナルティ項

$P_m(x, y)$

前フレームにおける注釈の位置からの移動量に関するペナルティ項

この手法をもとに、本システムにおいて注釈を表示するためのシステムを実装した。

牧田らの手法では、表示する注釈として物体の名称程度の情報量を想定している。しかし本システムでは、名称以外にも、史跡に関する説明文のようなサイズの大きい注釈を表示することも想定している。このため説明文が他の物体の注釈に被ってしまう恐れがあり、表示位置に関して UI に工夫する必要がある。

そこで対象物体の名称、説明文それぞれを異なる注釈としてわけて扱った。さらにユーザーが得たい情報としてどの物体に関心があるかを推定することで、表示する注釈を絞ることにした。これによって複数の展示物があるような状況においても、ユーザーが表示される余計な情報を排除したり、注視する位置に混乱が生じないようにした。

### 3.2 視認性を考慮した表示方法の提案

対象物体の名称をラベル (小さな注釈)、説明文をテキスト (大きな注釈) と定義し、2種類の注釈を使い分けることで視認性の向上を図る。

始めに、牧田らの手法を参考に画面内で検出した物体のラベルを提示し、ラベルと物体を線で結ぶ (図 11)。次に、関心を持つと判定された物体のテキストを画面の端に提示する。この時、テキストと物体との距離が長くなり相互の結びつきが把握しづらくなってしまいうため、ラベルをテキストに向かってゆっくり移動させる (図 12)。これによ

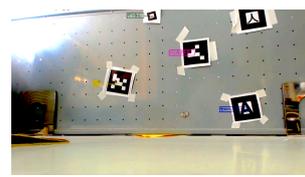


図 11 ラベルの提示例

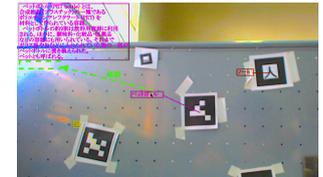


図 12 テキストの提示例

Fig. 11 Label display example Fig. 12 Text display example.

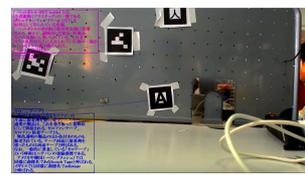


図 13 複数のテキストの提示例

Fig. 13 Multiple texts display example.

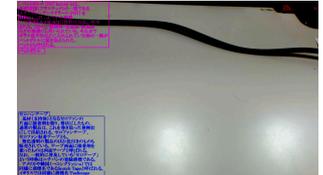


図 14 関心の保持が続いた場合

Fig. 14 The keeping of the target interest.

表 4 実験環境

Table 4 Environment of experiment.

ディスプレイサイズ	13 インチ (1280×800)
フォントサイズ	18pt
テキストの最大提示数	2
関心を持つまでに要する時間	0.8s

てユーザは意識することなくテキストに注目を向けるようになり、直感的に関係性を把握することができる。さらに、画面をテキストで埋め尽くしてユーザの混乱を招かないように、同時に提示するテキストの数を決めておく必要がある (図 13)。また、他の物体に関心が移らない限りは同じ物体に関心が保持されるものとし、物体が画面から外れてもテキストは画面内に提示され続ける (図 14)。

### 3.3 実験

成人男性 5 人を対象として、提案したビューマネジメントの評価実験を行った。壁に貼り付けた AR マーカーを展示物として見立て、牧田らの手法のみを用いた UI、提案手法による UI を比較してもらい、感想を聞いた。この時、オブジェクトの関心判定として、ユーザが携帯端末の画面中央付近で数秒間捉えたオブジェクトに、関心を持つものとして判定する。

#### 牧田らの手法のみを用いた UI 実装

関心判定によるフィルタリングを施さず、図 15 のように、検出した物体のテキストのみを牧田らの手法をもとに提示するバージョン

#### 提案手法による UI 実装

提案手法を実装したバージョン (図 11, 図 12, 図 13, 図 14)

本実験では、web カメラからのライブ映像を MacBook Pro に表示し、表 4 に示すような環境で評価を行った。



図 15 テキストのみを提示したバージョン  
 Fig. 15 Version for displaying only texts.



図 16 城跡を想定したジオラマ  
 Fig. 16 The diorama assuming castle ruins.

### 3.4 アンケート結果

サンプル数が少ないため定量的な評価はできていないが、提案したビューマネジメントに対する意見として、

- 既存手法と比較して、注釈がどの展示物のものなのかがわかりやすかった。
- テキストが提示されている物体が画面外に移動し再び画面内に映された時、結びつきが分からなくなった。
- しばしば意図しないテキストが提示された

等が得られた。物体が再び画面内に表示された時の結びつきの分かりづらさ、関心判定の妥当性に対して解決すべき課題点が挙げられたが、注釈を使い分ける手法に対して好意的な意見もあった。

## 4. ガイダンスシステムの評価実験

物体の注釈に関する視認性を評価する実験においていくつかの改善点は挙げたが、物体検出に SHISAR を用いた場合及び、HOG 特徴量と SVM 検出器の組み合わせによる手法を用いた場合において、携帯端末上でのパフォーマンスを評価するために、実機テストを行った。

### 4.1 実験

今回は、城跡のような凹凸が多く光源変化の影響を受けやすい環境を想定したジオラマ(図 16)を用意し、学習用の動画と検出実験をそれぞれ異なる光源環境においた際の実機テストを iPhone7 を用いて行った。

対象オブジェクトとして、シーサー、井戸、城門等を用意して 1m 程度離れた位置からジオラマ全体を回り込みながら撮影した動画から、それぞれのオブジェクトをトラッキング及び学習し、光源方向を変えた環境で学習動画と同じ方向から撮影した実機での検出テストを行った。

### 4.2 実験結果

表示例を図 17 に示す。SHISAR では、検出対象とするオブジェクトごとに約 500ms の検出時間を要した。HOG 特徴量を SVM で学習した検出器を用いた場合、オブジェクトの数に対する検出時間の増加は比較的少ない。ただ

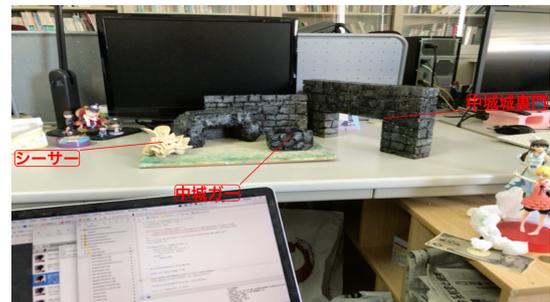


図 17 ジオラマに対する表示例  
 Fig. 17 A result for miniature exhibition.

表 5 実験結果 (総フレーム数:30)

Table 5 Experiment result (total frames:30) .

	SHISAR	HOG+SVM
シーサー	30	30
井戸	30	30
城門	30	29

し、SVM でストライドを 1 とした場合、実用的な速度での検出はできなかったため、SHISAR と同程度の検出速度になるようにストライドを調節し 4 ピクセルとした。従って、本実験の設定においては、SHISAR の位置検出精度は SVM に比べて 4 倍となる。正しく検出できているかを目視で確認したところ、表 5 の通りになった。

学習するパラメータとして複数のスケールを調整することによって、SHISAR を用いた手法と HOG 特徴量を用いた SVM による検出手法、いずれの場合でも、学習した角度とその角度におけるスケールの変化に対してロバストな検出ができた。

### 4.3 屋外における実験

ジオラマでの検出・表示に関する実験では、撮影者と対象物体の距離が近く、撮影者の移動が物体の見えの変化に対して大きな影響を及ぼす。移動による影響を考慮し、より実際の撮影環境に近づけた実験として、屋外にある物体

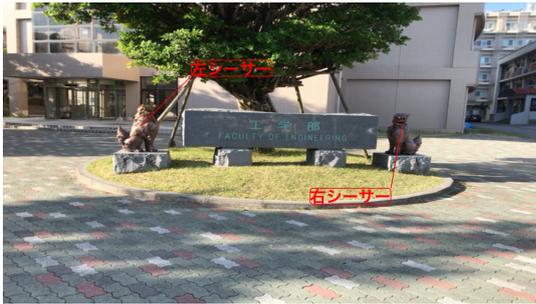


図 18 屋外における検出例

Fig. 18 The detection example at outdoor.

に対して検出を行った。

屋外での検出実験の例(図 18)から、撮影者の移動に対して安定的に検出ができることがわかった。

## 5. まとめ

本稿では、屋外環境での使用を想定した物体検出手法 SHISAR の概要について述べ、実機テストとして SVM による検出手法と共に物体検出に関する実験を行った。さらに、本研究で想定している AR ガイダンスシステムにおける注釈の表示方法について提案、実験を行なった。ビューマネジメントにおける実験では、提案した手法に対する好意的な意見も見られたが、視認性や関心判定手法に対して解決すべき課題が挙げられた。

今後は説明文の表示方法や視認性の評価等、ビューマネジメントの改善案の提案を含め、定量的な評価を受けるために、より多くの被験者を対象とした評価実験を行う必要がある。

## 参考文献

- [1] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi: *You Only Look Once: Unified, Real-Time Object Detection*
- [2] 津波古正寿, 赤嶺有平, 根路銘もえ子: “軽量な特徴量の空間分布を用いた特定物体検出手法”, 第 22 回日本バーチャルリアリティ学会大会論文集 (2017 年 9 月)
- [3] 稲福也美, 赤嶺有平, 根路銘もえ子: “屋外環境における色情報を用いた立体物の検出手法に関する研究”, 第 20 回日本バーチャルリアリティ学会大会論文集, pp.236-237, 2015
- [4] 吉田侃又, 赤嶺有平, 根路銘もえ子: “特徴の空間分布を利用した特定物体検出手法の提案”,
- [5] Dalal, Navneet, and Bill Triggs. *Histograms of oriented gradients for human detection*. Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 1. IEEE, 2005.
- [6] Burges, Christopher JC. *A tutorial on support vector machines for pattern recognition*. Data mining and knowledge discovery 2.2 (1998): 121-167.
- [7] 蔵田武志, 酒田信親, 牧田孝嗣: “展望 3 AR のインタフェース”, 情報処理, Vol.51, No.4, pp. 425-430, Apr.2010
- [8] 牧田 孝嗣, 神原 誠之, 横矢 直和: “ネットワーク型ウェアラブル AR のための動的環境における注釈のビューマネジ