

# ペンギンの腹部模様描画による個体検索アルゴリズムの改善と実験室実験および実地実験での検証

高野 閑<sup>1</sup> 中川 由貴<sup>1</sup> 中村 聡史<sup>1</sup>

**概要：**動物園や水族館における生き物を漠然と観察してしまう問題を解決するため、我々は腹部の模様が特徴的なペンギンに着目し、描画をしながら観察および検索を行うシステムを実現してきた。しかし、実地実験などではペンギンの見え方の違いで、描かれる点がグリッド境界線上に配置された結果、ランキング上位に対象とするペンギンが登場しないなど、手法の精度に関する問題があった。そこで本研究では、境界線上の点もカバーするように、グリッドの分割方法を変え、それぞれのスコアを合算して最終スコアとする手法を提案した。また、過去の研究で構築した実験室での描画データセットおよびサンシャイン水族館で実施された実地実験データに対して検証を行った結果、推定したペンギンランキングの上位1位以内正解率が実験室データでは従来手法に比べ約19%、実地実験データでは約10%の精度向上が確認された。

**キーワード：**ペンギン、アルゴリズム、観察、検索、グリッド

## 1. はじめに

水族館や動物園は従来、娯楽施設としての側面が強調されてきたが、生物多様性の危機や社会的要請に応じて教育・研究・保全といった新たな機能を担う社会教育施設へと進化している [1, 2]。実際、多くの水族館・動物園が自然保護への意識啓発や来館者の学習促進を使命に掲げている。一方、来館者は必ずしも展示を学習の機会として活用しておらず、解説パネルなどに目を通す来館者は27%に留まる [3] など、動物を漠然と眺める傾向が指摘されている [4, 5]。

来館者が動物園・水族館でより深い学びを得るには、展示手法の工夫だけでなく、来館者自身が能動的に動物に関わる体験を促すことが重要である。特に、動物を種全体としてではなく「個体」として認識することが、観察への関与を深め、記憶や学習に良い影響を与えることが指摘されている [3]。

ここで、水族館や動物園での人気が高いペンギンは、外見上の個体差が大きいことから、来館者が個体を識別することは容易ではない [6]。ペンギン個体識別の伝統的手法としてはフリッパーバンドが広く用いられてきたが、長期研究により生存率や繁殖成功への負の影響が報告されており [7]、動物福祉の観点からも代替手法の検討が求められている。

我々は、こうした問題を踏まえ、観察者がペンギンの腹部模様を描画することで個体検索を行う手法を提案し、実験室実験、および実地実験で有用性を明らかにしてきた [6]。しかし、ペンギンの腹部領域を3×3グリッドに分割する従来の手法では、特に実地実験においては精度が低いという問題があった [8]。具体的には、描画点がグリッド境界付近に集中した場合に精度が悪化する傾向があった。

そこで本研究では、従来の3×3に加えて、4×4や5×5など複数グリッドを使ってスコアリングを行い、結果を統合する手法を提案する。複数の分割数を組み合わせることで、粗いグリッドでの全体的な斑点分布と、細かいグリッドでの局所的な特徴を同時に扱い、領域境界付近の揺らぎに頑健な類似度評価を実現する。

## 2. 関連研究

### 2.1 ペンギンの腹部模様を用いた個体識別研究

Burghardt ら [9] は、成鳥のケープペンギン（アフリカペンギン）の斑点模様が一羽一羽異なり、季節によって変化しないことに着目し、その胸の斑点模様を生体識別子として利用する手法を提案した。実際に、胸部斑点パターンの抽出によって各ペンギンごとに一意の生体識別子を生成し、用いることで映像や写真から得たペンギンを個体識別することができることを示した。

また、Baciadonna ら [10] は、アフリカペンギンが斑点模様を視覚の手がかりに仲間を識別していることを、行動

<sup>1</sup> 明治大学  
Meiji University

観察実験により明らかにした。従来はペンギンのコミュニケーションは主に聴覚に依存していると考えられていたが、この結論により視覚の手がかりがペンギンにおいても重要な役割を果たしていることが判明した。

これらの先行研究は、ペンギンの腹部・胸部に現れる斑点模様が、生体学的に個体差の大きい特徴であるとともに、個体識別の手がかりとして有効であることを示している。

## 2.2 他の生き物の体表模様を用いた個体識別

動物の体表模様を用いた個体識別の研究は、ペンギン以外を対象としたものも多く存在する。アシカ類では、上顎のひげが生える位置に現れる黒点（ひげ斑点）の配置が個体固有であることを利用し、ひげ斑点パターンから個体を同定する手法が提案されている [11]。またシマウマでは、縞模様を特徴量として扱い、写真から個体を照合する StripeSpotter などの個体識別システムが実現されている [12]。StripeSpotter は、縞模様を StripeCode と呼ばれる識別子に変換し、データベースとの照合によって個体を同定している。さらに魚類においても、Meguro ら [13] は、ゼブラフィッシュの尾びれ・臀びれに現れる縞模様の違いを用いて個体を識別する手法を提案している。このように小型の水生生物に対しても体表パターンが安定した識別手がかりとなることが示されている。

これらの研究は、自然に存在する体表模様は個体差が大きく、写真撮影などにより取得できるため、個体識別に有効な特徴量として利用できるという共通点をもつ。本研究が扱うベクトルも、こうした模様ベースの個体識別研究の流れに位置づけられるが、カメラベースの個体識別はよりよい観察にはつながらないと考え、斑点を描く能動的な方法をとっている。

## 2.3 グリッド分割および多スケール特徴量に基づく類似度計算

画像認識分野では、局所特徴量の出現頻度をヒストグラム化する Bag-of-Features 表現が広く用いられてきたが、画像全体を1つのヒストグラムで表現するだけでは空間配置情報が失われるという課題が指摘されている。この課題に対し、Spatial Pyramid Matching (SPM) [14] は、画像を複数解像度のグリッドに分割し、各領域の特徴量を階層的に統合することで空間的な情報を保持する手法として知られている。また、植物の葉画像認識において、葉画像を複数のグリッドに分割し、各グリッドから局所特徴量を抽出して統合する Multiple-grid ベースのローカル記述子が提案されており [15]、多スケールなグリッド分割とヒストグラム特徴の統合が識別性能向上に有効であることが報告されている。さらに、3D 点群解析においても、多スケールなグリッドを用いて空間を分割し、各セルに意味ラベルや

特徴量を付与することで、物体検出やセグメンテーションの性能を向上させる枠組みが提案されている [16]。

本研究のアルゴリズムは、ペンギンの腹部を複数のグリッド分割で解析し、それぞれのグリッドで得られる特徴量を統合するものであり、SPM や Multiple-grid ベースの手法と同様に位置に依存した頻度情報をベクトルとして表現し、多スケールに統合するという点で概念的な類似性をもつものである。一方で、本研究は画像特徴量ではなく、ひとの描画行為により得られる離散点を扱う点で SPM と異なる。

## 3. 正解データセット

精度を検証するためには、描画された腹部模様がどのペンギン個体に対応するかを示すための正解ラベル付きデータが必要になる。本章では、実験室実験データと実地実験データにおける正解ラベルの付与方法と、それぞれの信頼性について述べる。

### 3.1 実験室実験データにおける正解

実験室実験 [8] では、実験協力者はスクリーンに提示されたペンギン個体の腹部画像を見ながら模写する形式で斑点を描画した。提示される画像にはペンギンの個体名が明示されており、実験協力者は描画後にその個体名を入力するため、正解ラベルは実験手続き上明確であり、誤りの可能性はない。ただし、入力された個体名には表記揺れが数件存在したため、本研究ではこれらを手作業で統一し、表記揺れを取り除いたうえで正解データセットとした。

### 3.2 実地実験データにおける正解

我々が過去に行っていた実地実験 [8] では、描画後に提示されるペンギン候補のランキング結果から実験協力者自身が正解個体を選択するようになっていた。図1にはシステムのログイン画面から検索結果ランキング画面を示す。実験協力者自身はランキングから描いているペンギンを探し、最後に「この子だった!!!」というボタンをクリックしたものを正解としていた。

そのため、実地実験データにおける正解ラベルは、実験協力者の選択に依存しているという問題があった。実際、バックヤードにいるはずのペンギンが正解となっているデータや、斑点が少ないペンギンが正解とされているデータ、点が多数描画されていたデータが存在した。そのため、正解とされたペンギンが観察した個体と一致しているとは限らない。また明らかに不真面目な描画において正解データが定まっているものも存在した。そこで以下の基準でそうしたデータを除外した。

- 実地実験に協力していた研究室の学生が描いている
- バックヤードのペンギンが正解とされている



図 1 上段からログイン画面、おえかき画面、下段から検索結果ランキング画面、個体詳細。「この子だった!!!」のボタンを押すと、選択したペンギンが正解データとなる

- 斑点ではない文字や模様が描画されている
- 正解のペンギンの斑点に対して、斑点数が明らかに多い/少ない
- 正解のペンギンには点が存在しない特徴的な位置に点が描かれている

除外処理の結果、最終的に正解データとして 266 件のデータのうち、127 件を除外し、139 件のデータを分析対象とした。

## 4. 提案手法

### 4.1 従来手法

これまでの研究 [6] では、腹部領域全体を  $3 \times 3$  に分割し、各領域に含まれる斑点数から 9 要素の腹部模様ベクトルを構築し (図 2)、コサイン類似度およびユークリッド距離を用いて描画の類似度を算出していた。また、どのペンギンかについては、実験室実験で得られた複数人の描画データから、個体ごとの  $3 \times 3$  について平均的な斑点分布

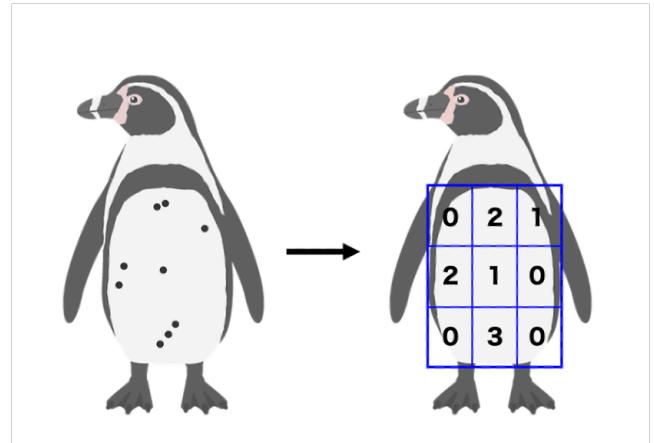


図 2 領域ごとの点数カウントの方法

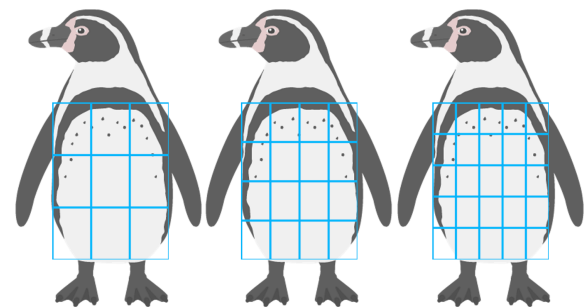


図 3 左から  $3 \times 3$ ,  $4 \times 4$ ,  $5 \times 5$  の分割例

を表すベクトルを作成し、利用していた。

そのため、実験室実験において描画点がグリッド境界付近にある場合に精度が低下するなどの問題があった。

### 4.2 複数グリッド分割の導入

4.1 節で述べた問題に対して、本研究では従来の  $3 \times 3$  に加えて複数の分割数を用いた特徴量を導入する。分割数を変え、 $2 \times 2$ ,  $3 \times 3$  のような粗いグリッドでは大まかな斑点分布を捉え、 $4 \times 4$ ,  $5 \times 5$  のような細かいグリッドでは局所的な密度の違いを反映可能とする。この複数のグリッド分割のスコアを統合することで、境界付近の揺らぎに頑健な表現が可能となる。

図 3 に  $3 \times 3$ ,  $4 \times 4$ ,  $5 \times 5$  の分割例を示す。

### 4.3 スコア算出とランキング付け

本研究では、まず識別対象のペンギン集合を  $P = \{p_1, p_2, \dots, p_M\}$  として、各ペンギン  $p_i$  について、実験室実験で得られた複数人の描画データより、分割数  $n \times n$  のグリッドから得られるペンギンの個体ベクトル  $\mathbf{p}_{i,n}$  が求められる。一方、ユーザが描いた腹部模様  $u$  を  $n \times n$  のグリッドに分割し、得られる特徴ベクトルを  $\mathbf{u}_n$  とする。

スコア算出には、従来手法と同様にコサイン類似度とユークリッド距離を組み合わせた指標 *similarity* を用いる。ここで cosine はコサイン類似度、dist はユークリッド

距離である。類似度スコアは以下で定義する。

$$similarity(\mathbf{u}_n, \mathbf{p}_{i,n}) = \frac{\cosine(\mathbf{u}_n, \mathbf{p}_{i,n})}{\text{dist}(\mathbf{u}_n, \mathbf{p}_{i,n}) + 1}$$

本研究では、複数のグリッド分割で得られた類似度を統合する点が、ここで用いるグリッド集合を  $\mathcal{G}$  とすると、ペンギン  $p_i$  の最終スコア  $Score$  は次式で与えられる：

$$Score(u, p_i) = \sum_{n \in \mathcal{G}} similarity(\mathbf{u}_n, \mathbf{p}_{i,n})$$

腹部模様描画後の検索では、結合スコア  $Score$  の降順に基づいて各ペンギン個体をランキング形式で検索候補として提示する。

## 5. 評価と結果

グリッド分割によるそれぞれのスコアを合計した結合スコアの検索性能を実験室実験データおよび実地実験データを対象として評価する。

### 5.1 評価方法

評価指標には、推定されたランキングの上位  $k$  位以内に正解個体が含まれる割合である Top- $k$  精度を用いる。本研究ではまず、比較的粗い  $2 \times 2$  から細かい  $5 \times 5$  グリッド分割でのスコアを組み合わせて検索精度がどの程度向上するのかを、実験室実験データおよび実地実験データを用いて検証する。

### 5.2 実験室実験データ

実験室実験は、教室内のスクリーンに提示された 44 個体のペンギン腹部画像を参加者が模写する形で描画した斑点データで構成されている。統一された観察環境で収集されているため、描画のばらつきが比較的少ない。

検索性能を検証するため、本研究では 5 分割交差検証を採用した。実験室データ（全 1,687 件）を 5 つに分割し、各分割において 4 分割分をペンギンの個体ベクトル ( $\mathbf{p}_{i,n}$ ) の作成に使用し、残り 1 分割分を評価用データとした。

結果は、表 1 の通りである。まず、従来手法に相当する  $\mathcal{G} = \{3\}$  のみの場合、Top-1 は 0.699 であった。それに対して、 $3 \times 3 + 4 \times 4 + 5 \times 5$  を統合した  $\mathcal{G} = \{3, 4, 5\}$  では、Top-1 精度は 0.889 であり、従来手法よりも約 19% 精度が改善した。この  $\mathcal{G} = \{3, 4, 5\}$  は、本章で扱う全ての組み合わせの中で最も高い精度を示し、Top-2～Top-5 のいずれにおいても、0.948, 0.965, 0.973, 0.978 と高く、いずれの指標においても最良の結果となった。また、分割数を統合することで上位候補に正解個体が含まれる割合が増加した。

一方、全ての組み合わせが精度向上につながるわけではなく、 $\mathcal{G} = \{2\}$  や  $\mathcal{G} = \{2, 4\}$  など、従来手法よりも Top-1 精度が低くなるケースも確認された。特に  $\mathcal{G} = \{2\}$  を含む

表 1 実験室データにおける分割パターン別精度 (Top-1 降順)

$\mathcal{G}$	Top-1	Top-2	Top-3	Top-4	Top-5
$\{3, 4, 5\}$	0.889	0.948	0.965	0.973	0.978
$\{4, 5\}$	0.879	0.936	0.962	0.970	0.975
$\{2, 3, 4, 5\}$	0.853	0.923	0.947	0.962	0.971
$\{3, 5\}$	0.842	0.919	0.946	0.958	0.968
$\{3, 4\}$	0.838	0.912	0.944	0.957	0.966
$\{2, 4, 5\}$	0.815	0.907	0.937	0.957	0.964
$\{2, 3, 5\}$	0.813	0.902	0.934	0.947	0.957
$\{5\}$	0.796	0.878	0.920	0.936	0.945
$\{2, 3, 4\}$	0.790	0.888	0.916	0.941	0.960
$\{4\}$	0.763	0.865	0.904	0.931	0.947
$\{2, 5\}$	0.736	0.870	0.912	0.931	0.947
$\{2, 3\}$	0.722	0.833	0.887	0.917	0.934
$\{3\}$	0.699	0.821	0.876	0.907	0.922
$\{2, 4\}$	0.693	0.834	0.880	0.910	0.932
$\{2\}$	0.534	0.704	0.777	0.816	0.848

表 2 実地実験データにおける分割パターン別精度 (Top-1 降順)

$\mathcal{G}$	Top-1	Top-2	Top-3	Top-4	Top-5
$\{2, 3, 5\}$	0.400	0.550	0.617	0.650	0.700
$\{2, 3, 4, 5\}$	0.400	0.542	0.592	0.675	0.708
$\{3, 4, 5\}$	0.392	0.575	0.617	0.683	0.742
$\{3, 5\}$	0.392	0.475	0.558	0.608	0.658
$\{2, 3\}$	0.375	0.500	0.583	0.633	0.683
$\{3, 4\}$	0.375	0.517	0.592	0.658	0.708
$\{2, 4, 5\}$	0.342	0.500	0.583	0.650	0.700
$\{2, 4\}$	0.333	0.483	0.583	0.625	0.650
$\{2, 5\}$	0.325	0.467	0.542	0.600	0.650
$\{4\}$	0.317	0.433	0.500	0.567	0.633
$\{3\}$	0.308	0.433	0.508	0.550	0.600
$\{4, 5\}$	0.267	0.458	0.550	0.625	0.667
$\{2\}$	0.258	0.408	0.492	0.567	0.608
$\{5\}$	0.217	0.333	0.442	0.517	0.600

パターンでは、分割数が粗すぎることにより特徴量が十分に得られず、精度が低下する傾向が見られた。

### 5.3 実地実験データ

サンシャイン水族館で実施した実地実験データについて、3.2 節で構築した正解データセットをもとに評価した。なお、実験室実験ではデータが取得されなかった 3 個体に関しても、実地実験データには一定数の描画（計 19 件）が存在するが、これらの個体はペンギンの個体ベクトルを構築できないため、検索性能評価からは除外し、120 件を対象とした。検索精度の検証では、全ての個体ベクトルを実験室実験データから構築した。

結果は、表 2 の通りとなった。従来手法に相当する  $\mathcal{G} = \{3\}$  のみの場合、Top-1 は 0.308 であった。一方、 $\mathcal{G} = \{2, 3, 5\}$  の場合には Top-1 精度が 0.400 となり、従来手法より約 10% の改善が見られた。Top-2～Top-5 でも同様に、グリッド分割の追加に伴う改善傾向が確認された。

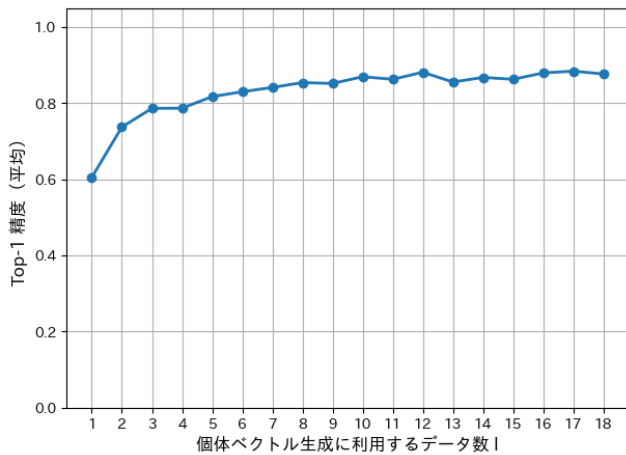


図 4 個体ベクトル生成に利用するデータ数  $l$  ごとの Top-1 精度

また、Top-5 に着目すると、 $G = \{3, 4, 5\}$  が 0.742 と最も高く、実験室実験と同様に全体的に最も高い精度となった。

ただし、全ての組み合わせが一貫して性能を改善するわけではなく、 $G = \{5\}$  や  $G = \{4, 5\}$  では Top-1 精度が  $G = \{3\}$  を下回る結果となった。

## 6. 考察

### 6.1 分割数がランキング精度に与える影響

本研究では、腹部領域を複数のグリッドで分割し、それらのスコアを統合する手法の有効性を検証した。まず  $2 \times 2 \sim 5 \times 5$  の範囲で比較した結果、 $G = \{2\}$  では、分割数が粗すぎたため実験室実験および実地実験のいずれにおいても従来の  $G = \{3\}$  よりも低い精度となった。

一方、 $G = \{3, 4, 5\}$  は最も高い精度を示し、特に実験室実験データでは Top-1 精度は 0.889 と最も高い値となった。また、実地実験データでも他の統合に比べて高い精度を示した。この結果は、異なる大きさのグリッドを組み合わせることで、同一ペンギンに対する描画ごとの斑点配置の揺らぎを捉えることができるためと考えられる。

### 6.2 個体ベクトル作成に必要なデータ数

本研究では、腹部模様を描画することでペンギン個体を識別する枠組みを構築したが、その際にペンギンの個体ベクトルとして何件の描画を用いるべきかが明らかになっておらず、これまでは 20~30 人分のデータを収集していた。個体ベクトルに用いる描画数が少なすぎると特徴が安定せず、逆に多すぎる場合には収集の負担が増大する。今後、様々な動物園・水族館への導入を考えた場合に必要なデータ数がわかっていることが望ましい。そのため、1 個体につきどの程度の描画を集めれば十分な識別精度が得られるのかを明らかにする。

ここでは、実験室実験データを用い、個体ベクトル生成において使用する描画数  $l$  と識別精度の関係を分析した。具体的には、各ペンギン個体の描画データから  $l$  件をラン

ダムに抽出して個体ベクトルを構築し、さらに別の 10 件をテスト用データとして用いた。

試行に用いる描画については、全個体の見本数に対してスコアを算出し、Top-1 精度を求めた。この操作を  $l = 1 \sim 18$  まで変化させながら繰り返し、各  $l$  における識別精度を全ペンギン個体で平均した。

結果を図 4 に示す。個体ベクトル生成に利用するデータ数  $l$  を 1 から 5 まで増加させると、平均 Top-1 精度は約 0.60 から 0.82 まで急激に向上した。一方、 $l = 6$  以降は増加量が小さくなり、 $l = 10 \sim 12$  の範囲では平均 Top-1 精度はおよそ 0.87~0.88 で推移し、飽和傾向が見られた。 $l = 17$  において平均 Top-1 精度は最大値 0.88 程度に達したが、 $l = 12$  と比較した改善量は 0.01 未満であり、実用上は 10~12 件程度の描画があれば十分な識別精度が得られるといえる。

また、ペンギン個体ごとに Top-1 精度が最大となる必要なデータ数を求めたところ、最適なデータ数の平均は  $l = 11.54$  であり、多くの個体が  $l = 8 \sim 13$  の範囲に集中していた。

以上の結果から、個体ベクトル構築に必要な描画数は平均的には 10~12 件程度であり、これより多くの描画を追加しても精度向上は限定的であることが示唆された。

### 6.3 分割数増加に伴う精度の飽和

5 章で述べたように、グリッド分割数を増加させることで特徴量が増え、検索精度が向上することが期待される一方で、過度に細分割を行うと描画点数に対してグリッドが過剰となり、スコア統合の効果が限定的になる可能性がある。そこで本節では、グリッドの分割数を段階的に増加させた場合の検索精度の変化を分析する。 $G = \{3\}$  から  $G = \{3, 4, \dots, 12\}$  まで Top-k 精度を指標として評価した。これにより、分割数増加が検索性能に寄与する範囲、および精度向上が頭打ちとなる飽和点を明らかにする。

まず、実験室実験データでの精度を図 5 に示す。縦軸は Top-k の精度、横軸はグリッド分割の累積を表す。Top-1 精度に着目すると、 $3 \times 3$  に対して、 $4 \times 4$  を追加した時点で最も大きな精度の改善が見られたが、 $7 \times 7$  の累積以上は改善幅が小さくなり、 $10 \times 10$  前後でほぼ頭打ちとなった。

次に実地実験データでの精度を図 6 に表す。実験室実験データと同様に、縦軸は Top-k の精度、横軸はグリッド分割の累積を表す。Top-1 精度に着目すると、実験室実験データと同様に  $3 \times 3$  の Top-1 精度 0.308 に対して、 $4 \times 4$  を追加した時点で 0.375 という最も大きな精度の改善が見られた。その後も  $6 \times 6$  まで累積すると、Top-1 精度は 0.417 に達し、ここまでは一貫した向上が見られた。しかし、 $7 \times 7$  以降では精度が低下し、 $6 \times 6$  を上回る精度が見られなかった。特に、 $12 \times 12$  では 0.383 まで低下してお



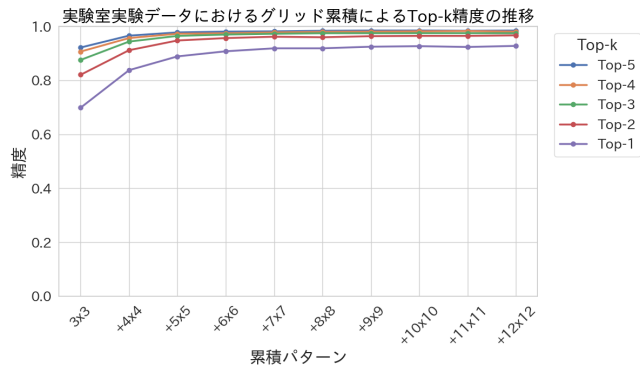


図 5 実験室実験データにおけるグリッド分割数増加に伴うランキング精度の変化 (Top-k)

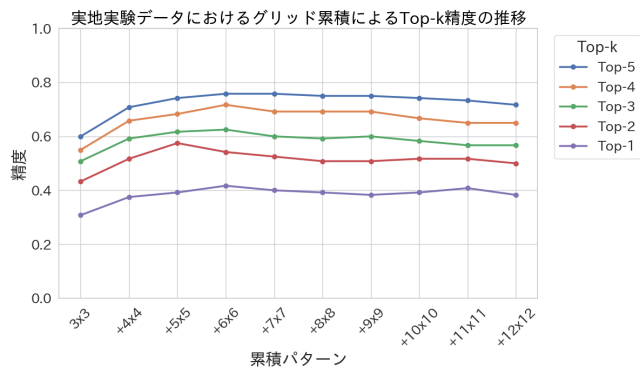


図 6 実地実験データにおけるグリッド分割数増加に伴うランキング精度の変化 (Top-k)

り、実験室実験データとは異なり、Top-1 精度が悪化する傾向が確認された。また、Top-2 に着目すると、5×5 までのものが最も精度が高いことがわかる。

まず、正解データセットの信頼性が異なる。実験室実験では、スクリーンに提示されたペンギン名と腹部画像が対応した状態で描画を行うため信頼性が高い。一方、実地実験では、描画後に実験協力者自身がシステムから提示されるランキング一覧の中から正解だと思う個体を選択する方式を採用している。この方式では、描画が不十分で実験協力者自身が個体を正しく認識できない場合や、ランキング上位が画面上部に固定表示される UI の影響で、スクロールせずに上位候補をそのまま選択してしまう場合がある。さらに、システムの検索精度が十分ではない場合には、本来の正解個体が上位に提示されず、実験協力者が誤った候補を選ばざるを得ない状況も生じる。これら複数の要因が重なることで、実験協力者が最終的に選択した正解ラベル自体に誤りが含まれる可能性があり、実地実験に特有のノイズとして結果に影響を与えと考えられる。正解データに対し、除外を行ったものの、本来の実験協力者の意図を汲み取れなかった可能性がある。この差が実験室と実地の精度差に影響していると考えられる。

次に、観察環境の違いによる描画のばらつきがある。実験室実験では、同一の腹部画像を全員が同条件で模写する

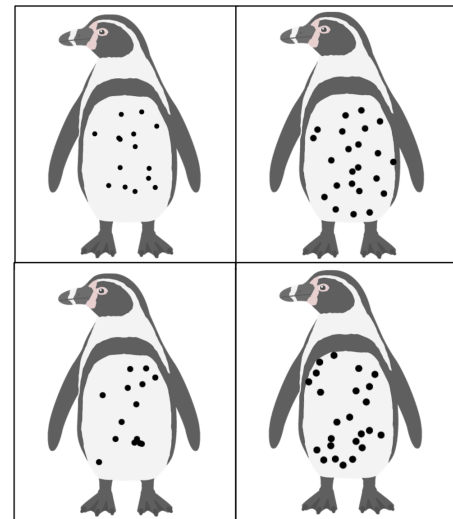


図 7 実地実験での「オレライム」の描画 4 件

ために斑点数のばらつきが小さい。しかし、実地実験では以下のような環境要因が描画の揺らぎを起こしていた。

- ペンギンが動くため腹部が見える時間が限られる
- 観察距離・角度が実験協力者ごとに異なる
- 光の反射や視力などで点が見えにくい場合がある
- 実験協力者ごとに腹部の中心と認識される位置が異なる
- 年齢層が広く、描画の熟練度に差がある

これらの要因により、描画点の位置ずれや点数不足が生じ、提案手法が十分に機能しないケースが見られる。実際に、実地実験での描画に大きな差が出るペンギンがいた。その例として、「オレライム」を挙げる。図 7 に「オレライム」の描画 4 件を示す。左側の 2 件では主に腹部の中央の斑点が観察されており、数が少なめである。対して右側の 2 件では、腹部全体の斑点が観察され 20 点以上描画されていた。このような実験協力者ごとに観察される斑点が大きく異なることが描画の揺らぎとなっている。

以上の結果を踏まえると、実験室実験と実地実験の両方において、3×3 から 5×5 または 6×6 までの累積が安定して高い精度を示しており、両データセットに適していることが明らかとなった。

一方、7×7 以上の分割は、実験室実験では改善幅が小さく、実地実験ではむしろ精度が低下する傾向を示した。これは、動的な観察環境での描画のばらつきが細かいグリッドと相性が悪いと考えられる。そのため、実験室実験および実地実験双方では 6×6 までの累積で留めることが合理的であり、それ以上の細分割は必ずしも有効ではないことがわかる。

#### 6.4 実験室実験と実地実験の違い

また、それぞれのデータセットでのペンギンごとの平均の描画点数を比較したものを表 3 に示す。実験室実験デー

表 3 実験室実験と実地実験における腹部斑点数比較

ペンギン名	実験室	実地	差
青ライム	9.03	11.00	1.97
ピンク	2.02	2.40	0.38
金	5.93	6.00	0.07
白緑	8.10	7.80	-0.30
緑オレンジ	2.03	1.50	-0.53
オレンジ	7.91	7.25	-0.66
黒ピンク	7.80	7.00	-0.80
オレライム	20.12	19.29	-0.84
白ライム	7.93	7.00	-0.93
しろ	7.93	7.00	-0.93
みい	4.07	3.00	-1.07
テント	12.07	11.00	-1.07
ピンクライム	6.58	5.00	-1.58
ピンク黄	5.38	3.75	-1.62
緑ライム	7.87	6.20	-1.67
青オレンジ	6.67	5.00	-1.67
青ピンク	5.53	3.80	-1.73
くるみ	9.11	7.33	-1.77
おいも	4.78	3.00	-1.78
ダイスケ	10.82	9.00	-1.82
あやさ	4.57	2.00	-2.57
赤ピンク	5.93	3.33	-2.60
武蔵	11.60	8.84	-2.76
のんちゃん	17.63	14.69	-2.94
マリオ	6.39	3.00	-3.39
渚	6.50	3.00	-3.50
金緑	9.40	5.00	-4.40
青紫	10.43	6.00	-4.43
オレンジ銀	22.51	18.00	-4.51
赤白	14.37	9.00	-5.37
ペン太	15.11	7.00	-8.11
青	18.07	8.00	-10.07

タよりも実地実験データの方が平均の斑点の数が少なく、実験室に比べて描画が簡略化される傾向が見られた。これは実地環境ではペンギンの腹部を観察しづらい場面が多く、十分に書き込めない状況が発生することに起因すると考えられる。図 8 に「武蔵」の画像と実験室実験および実地実験での描画例を示す。実験室実験では、左の「武蔵」の画像が実際にスクリーンに映し出され、実験協力者の見本となっていた。実験室実験と実地実験の描画例を比較すると、実験室実験は 18 点で、実地実験は首元を含めずに数えると 9 点であった。このように、位置的に一致する点はあるが、描画が簡略化されていることがわかる。

ペンギンの展示構造そのものが描画対象の偏りを生み、実地実験のデータ品質に大きく影響している可能性もある。サンシャイン水族館のペンギンにおける飼育エリアは「草原エリア」「天空エリア」「バックヤード」の 3 つに分類され、実地実験の際に飼育されている個体数は草原 16 匹、天空 24 匹、バックヤード 7 匹であった。表 4 に実地実験におけるペンギンおよび展示エリアごとの描画回数を示す。

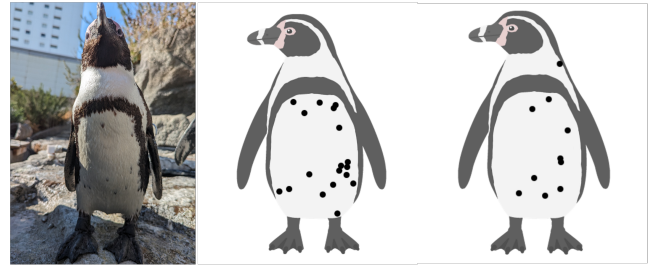


図 8 左：実験室実験で表示された「武蔵」の写真。中央：実験室実験での描画例。右：実地実験での描画例

草原エリアでは 79 件の描画が得られたのに対し、天空エリアでは 41 件に留まった。これは草原エリアでは地上で静止している個体も多く観察しやすい一方、天空エリアでは頭上を高速に泳ぐ個体が多く、腹部の模様を観察しながら描画することが困難であるためである。実際に天空エリアでシステムを使用した参加者の中には、2 人組で 1 人がスマートフォンで撮影し、もう 1 人がその画像を参照しながら描画する方法をとっているケースもあった。このように 1 人での観察では描画が難しいことから、天空エリアでは描画データが十分に収集されなかったと考えられる。

こうした描画環境の差は検索精度にも表れている。表 5 に展示エリア別の Top-k 精度の比較を示す。 $G = \{3, 4, 5, 6\}$  の Top-k 精度の比較では、草原エリアの Top-1 精度が 0.443 であったのに対し、天空エリアは 0.366 であった。Top-2, Top-3 においても草原エリアの精度が天空エリアを上回っており、観察環境による描画品質の低下が個体識別精度に影響していることが示唆された。

以上のように、展示エリアごとの観察条件の違いが描画点数・描画対象の偏りを生み、実験室実験と実地実験の精度差を引き起こす主要因となっている。特に、天空エリアにおける観察困難性が実地実験のデータ品質を不均一にし、検索精度を不安定にする方向に作用していると考えられる。

## 7. おわりに

本研究では、ペンギンの腹部模様の描画に基づく個体検索手法において、従来の  $3 \times 3$  グリッドのみを用いるアルゴリズムの課題に対し、複数のグリッド分割で得られたスコアを統合する手法を提案した。実験室実験データおよびサンシャイン水族館で取得した実地実験データに対して評価を行った結果、実験室データでは  $G = \{3, 4, 5\}$  を用いた場合に Top-1 が 0.889 と従来手法から大きく改善した。実地実験データにおいても  $G = \{2, 3, 5\}$  または  $G = \{3, 4, 5\}$  で従来手法より高い精度が得られることが明らかとなった。

また、グリッド分割数を段階的に増加させた分割数増加に伴う精度の分析から、実験室環境では  $10 \times 10$  付近まで精度向上が続く一方、実地環境では  $6 \times 6$  程度で早期に飽和し、それ以上の細分割では精度がむしろ低下することが示された。この結果は、静止画像を十分に観察できる実験

表 4 実地実験におけるペンギンごとの描画回数（回数降順）

ペンギン名（エリア）	回数	ペンギン名（エリア）	回数	ペンギン名（エリア）	回数
武蔵（草原）	19	オレンジ（草原）	4	白桃（天空）	1
ぼてと（草原）	18	おいも（天空）	4	あやさ（草原）	1
オレライム（草原）	14	くるみ（天空）	3	ピンクライム（天空）	1
のんちゃん（草原）	13	オレンジ銀（草原）	3	金緑（天空）	1
赤ピンク（草原）	9	しろ（天空）	2	ペン太（天空）	1
ピンク（天空）	5	渚（天空）	2	青ライム（天空）	1
金（天空）	5	みい（天空）	2	マリオ（天空）	1
白緑（草原）	5	白ライム（草原）	2	青紫（草原）	1
緑ライム（草原）	5	黒ピンク（天空）	2	ダイスケ（草原）	1
青ピンク（天空）	5	テント（草原）	2		
ピンク黄（天空）	4	緑オレンジ（天空）	2		

表 5 展示エリア別の Top-k 精度比較（3×3～6×6 累積スコア）

エリア	件数	Top-1	Top-2	Top-3	Top-4	Top-5
草原	79	0.443	0.570	0.646	0.722	0.772
天空	41	0.366	0.488	0.585	0.707	0.732

室と、動き回るペンギンを限られた視認条件で観察する実地環境とでは、適切なグリッド分割設計が異なることを示しており、描画型検索システムのアルゴリズム設計に対して1つの指針を与えるものである。

今後は、実地実験データにおける正解ラベルの不確実性を低減するための再ラベリングや描画順序を組み込んだ設計、および観察者の描画行動に基づくアルゴリズムの適用が挙げられる。これらを発展させることで、水族館における来館者の能動的な観察体験をさらに支援しつつ、より精度の高い個体検索システムの実現を予定している。

謝辞

実験にあたり、実地環境やペンギンの個体情報の提供、ならびにご協力をしてくださったサンシャイン水族館の皆様へ感謝致します。

参考文献

[1] World Association of Zoos and Aquariums (WAZA): Social Change For Conservation – The World Zoo and Aquarium Conservation Education Strategy, <https://www.waza.org/priorities/community-conservation/the-ize-waza-education-strategy/> (2020). Accessed: 2025-11-13.

[2] Spooner, S. L., Walker, S. L., Dowell, S. and Moss, A.: The Value of Zoos for Species and Society: The Need for a New Model, *Journal of Zoological and Botanical Gardens*, Vol. 4, No. 1, pp. 100–110 (2023).

[3] Clayton, S., Fraser, J. and Saunders, C. D.: Zoo Experiences: Conversations, Connections, and Concern for Animals, *Zoo Biology*, Vol. 28, No. 5, pp. 377–397 (2009).

[4] Godinez, A. M. and Fernandez, E. J.: What Is the Zoo Experience? How Zoos Impact a Visitor's Perception, *Frontiers in Psychology*, Vol. 10, p. 1746 (2019).

[5] Davey, G.: Visitor behavior in Zoos: A review, *Anthrozoös*, Vol. 19, No. 2, pp. 143–157 (2006).

[6] 中川由貴, 中村聡史: ペンギンの腹部模様注目した斑点描画型検索・観察手法とその検証, 情報処理学会論文誌, Vol. 65, No. 12, pp. 1842–1853 (2024).

[7] Saraux, C., Le Bohec, C., Durant, J. M., Viblanc, V. A., Gauthier-Clerc, M., Beaune, D., Park, Y.-H., Yoccoz, N. G., Stenseth, N. C. and Le Maho, Y.: Reliability of Flipper-banded Penguins as Indicators of Climate Change, *Nature*, Vol. 469, No. 7329, pp. 203–206 (2011).

[8] 中川由貴, 中村聡史, 芦刈治将, 板東恵理子, 與倉陵太, 渡邊果南, 岩永七海: ペンさく: 描いて探すことで深まるペンギン観察手法と水族館来訪者を対象とした実証実験, 第33回インタラクティブシステムとソフトウェアに関するワークショップ (WISS 2025) (2025).

[9] Burghardt, T., Thomas, B., Barham, P. and Calic, J.: Automated Visual Recognition of Individual African Penguins, *Proceedings of the 5th International Penguin Conference* (2004).

[10] Baciadonna, L., Solvi, C., Terranova, F., Godi, C., Pilenga, C. and Favaro, L.: African Penguins Utilize their Ventral Dot Patterns for Individual Recognition, *Animal Behaviour*, Vol. 210, pp. 75–87 (2024).

[11] Osterrieder, S. K., Salgado Kent, C., Anderson, C. J. R., Parnum, I. M. and Robinson, R. W.: Whisker Spot Patterns: A Noninvasive Method of Individual Identification of Australian Sea Lions (*Neophoca cinerea*), *Journal of Mammalogy*, Vol. 96, No. 5, pp. 988–997 (2015).

[12] Miller, B. K.: New Technology Identifies Zebras by Their Stripes, <https://www.audubon.org/news/new-technology-identifies-zebras-their-stripes> (2012). Accessed: 2025-12-04.

[13] Meguro, S., Miyazaki, T., Takeuchi, H., Naruse, K. and Shiraki, T.: Stripe Pattern Differences Can Be Used to Distinguish Individual Adult Zebrafish, *PLOS ONE*, Vol. 19, No. 10, p. e0311372 (2024).

[14] Lazebnik, S., Schmid, C. and Ponce, J.: Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 2, pp. 2169–2178 (2006).

[15] Chompookham, T., Gonwirat, S., Lata, S., Phiphiphat-phaisit, S. and Surinta, O.: Plant Leaf Image Recognition Using Multiple-grid Based Local Descriptor, *International Journal of Multimedia Data Engineering and Management* (2020).

[16] Chen, X.-F., Lee, C.-C., Lo, J.-H., Chuang, C.-H. and Fan, K.-C.: Multi-Scale Grid-Based Semantic Surface Point Generation for 3D Object Detection, *Electronics*, Vol. 14, No. 2, pp. 123–140 (2025).