

# コード化点字ブロックの仕組みを活用した ピクトグラムの音声化・多言語化の研究

松井 くにお<sup>\*</sup>, 坂本 真仁<sup>†</sup>, 堀田 耕一郎<sup>‡</sup>

2024年3月18日

## 概要

ピクトグラム（案内用図記号）は、不特定多数の人々が利用する公共交通機関や公共施設、観光施設等において、文字・言語によらず対象物、概念または状態に関する情報を提供する図記号である。識字が困難な方や外国人観光客等にも理解が容易な情報提供手法として、日本を含め世界中の公共交通機関、観光施設等で広く掲示されている。その一方で、ピクトグラムは爆発的に種類が増え、その結果ピクトグラムだけでは正確に伝わらない場合がある。また、現状のピクトグラムから得られる情報だけでは不十分な場合もある。たとえばピクトグラムを確認しただけでは、対象物までの方向や距離などの情報は得られない。そこでコード化点字ブロックの音声提供の仕組みを利用して、ピクトグラムの音声化・多言語化を行ない、ピクトグラムから更に付加的な情報をユーザに与えるシステムを提案する。提案システムの実現に向けて、ピクトグラム判別システムの実装に取り組んだ。本稿では実装したピクトグラム判別システムのプロトタイプ二種類について解説する。評価の結果シンプルなプロトタイプはピクトグラム判別に約 625 ms 要し、判別精度向上の工夫を取り入れたプロトタイプは判別精度は向上したが、ピクトグラム判別に約 1578 ms 要することが明らかになった。

キーワード：ピクトグラム判別、特徴点マッチング、ホモグラフィ行列、オイラー角

## 1 はじめに

ピクトグラム（案内用図記号）は、言語に依存せず対象物、概念または状態に関する理解が容易な情報提供手段として用いられている [1]。

たとえば図 1 に示すピクトグラムは、公共施設や公共交通機関などで実際に使用されており、とくに図 1(a) に示す非常口のピクトグラムはよく知られた、わかり易いピクトグラムの一例である。

一方で図 1(b) や図 1(c) のように、何を表しているか明瞭でないピクトグラムや、図 1(d) のように、家の中で靴を脱ぐ文化的な知識が必要なものなど、立場によっては理解が容易でないピクトグラムもある。また、ピクトグラムだけでは対象物までの距離

や方向といった情報が伝わらない。

上記の問題に対して、しゃべるピクトグラムを提案する。これはピクトグラムから更に付加的な情報をユーザに対して与えるシステムである。これは文化的知識が無い人や目が見えづらい人も、ピクトグラムを利用可能にするユニバーサルかつインクルーシブなシステムである。

本稿では JIS 規格のピクトグラムに絞ってピクトグラム判別システムを実装および評価する。これによりピクトグラムの音声化・多言語化に向けたシステム基盤の構築を目指す。

## 2 関連研究

### 2.1 コード化点字ブロック

よく知られた案内用設備に視覚障がい者誘導ブロック（以下、点字ブロック）がある。点字ブロッ

<sup>\*</sup> Kunio Matsui, Kanazawa Institute of Technology

<sup>†</sup> Shinji Sakamoto, Kanazawa Institute of Technology

<sup>‡</sup> Kohichiro Hotta, Professional University of Information and Management for Innovation



図 1: ピクトグラムの例

クは、線状の突起が進行方向に沿って並んで配列されている誘導ブロックと、点状の突起が配列されている警告ブロックがある。しかし、警告ブロックだけでは何が警告されているかわからず、従来の点字ブロックはピクトグラムと同様に十分な情報が提供できているとは言えない。

上記の問題に対して、コード化点字ブロック [2-4] が提案されている。コード化点字ブロックとは、向きを明らかにするための印と、突起それぞれに対して印アリと印ナシの二者択一を割り当てることで、コードを識別できる点字ブロックおよびその技術を指す<sup>1)</sup>。また、コード化点字ブロックを専用アプリで読み取ることにより、設置箇所に応じた案内文を読み上げる仕組みも広義のコード化点字ブロックである。

コード化点字ブロックは、識別番号のためのコードが付与された点字ブロックであり、アプリを利用してコードを読み取らせることでその場に適した情報を提供できる。本研究ではコード化点字ブロックの音声化・多言語化の仕組みを活用して、ピクトグラムの利用価値向上につなげる。

1) JIS 規格の警告ブロックは  $5 \times 5$  の突起と定められているため、JIS 規格の警告ブロックをコード化する場合、前後左右の 4 パターンと  $2^{25}$  の組み合わせとなり、1 億 3000 万通りを超えるコードが発行できる。

## 2.2 物体認識技術

画像から物体を検出・認識する技術は、その実用性から広く研究されている。たとえば伊藤らは、画像をフーリエ変換して得られる位相情報を用いて画像をマッチングする手法を提案している [5]。提案された手法は従来手法と比較して高精度な画像マッチング手法であるが、特徴点マッチングによるシンプルな画像マッチングと比べて低速である。

ピクトグラムを認識するシステムに関しても関連する研究がある。久野らは CNN (畳み込みニューラルネットワーク) によるピクトグラム認識手法を提案している [6]。CNN では判別すべきピクトグラムが増えるたびに全データを使った学習のやり直しが必要である。

## 3 実装システム

実装する二種類のプロトタイプについて述べる。

### プロトタイプ A

カメラ画像と正解画像の特徴点マッチングによるシンプルなピクトグラム判別システム [7]

### プロトタイプ B

特徴点のパターンからホモグラフィ行列を推定し、推定した行列とカメラ行列からオイラー角を求めることによってピクトグラムを判別するシステム [8]

どちらのプロトタイプも共通して特徴点検出と特徴量記述を行う。特徴点検出と特徴量記述には、商用/非商用を問わず利用できる A-KAZE (Accelerated KAZE) を用いる [9]。

A-KAZE は画像から頂点や線の交わりといった特徴的な座標を検出し、検出した座標の固有性を特徴量として算出する手法である。特徴点検出と特徴量記述に係る従来手法である SIFT (Scale-Invariant Feature Transform) や SURF (Speeded-Up Robust Features) の欠点を改善した KAZE [10] がさらに改良された手法が A-KAZE であり、KAZE と同様に拡大や縮小、回転に強い特徴を持ったまま KAZE に比べて計算時間を短縮している。

### 3.1 プロトタイプ A

特徴点マッチングによってピクトグラムを判別するシステムがプロトタイプ A である。プロトタイプ A はよく知られた物体認識技術を用いて実装され、ある意味で教科書通りの正攻法といえるシンプルなピクトグラム判別システムである。

正解画像の特徴点とカメラ画像の特徴点を比較し、類似度の高い特徴点同士を対応付けることで、どの正解画像のピクトグラムがカメラ画像に写っているかどうか総当たりで処理する。

類似度の計算にはハミング距離を用いており、類似度が高いかどうかの判断にはレシオテスト<sup>2)</sup>を行う。

### 3.2 プロトタイプ B

プロトタイプ A では、カメラ画像内に正解画像以外の被写体が存在する場合、正解画像とは関係のない特徴点同士が対応付けられてしまうことがあり、その結果判別誤りが発生する問題点がある。

そこで、特徴点のパターンからホモグラフィ行列を推定し、推定した行列とカメラ行列からオイラー角を求めることによってピクトグラムを判別するシステムを提案する。

ホモグラフィ行列とは、二次元平面上の点の座標を異なる二次元平面上の点の座標に射影変換（ホモグラフィ変換）するのに用いられる行列であり、二次元座標点の平行移動、回転、拡大縮小、せん断等の変換ができる。ただしホモグラフィ行列には三次元の回転情報が含まれていないため、ホモグラフィ行列からはオイラー角を算出できない。そこで、ホモグラフィ推定によって得られたホモグラフィ行列とカメラ行列から三次元回転行列を算出する。

カメラ行列とは、カメラの内部パラメータを表す行列であり、焦点距離や画像の中心座標、レンズの歪み等の情報を持つ。カメラ行列は通常、カメラのキャリブレーションによって算出されるが、プロトタイプ B においては、カメラ行列を既知であるもの

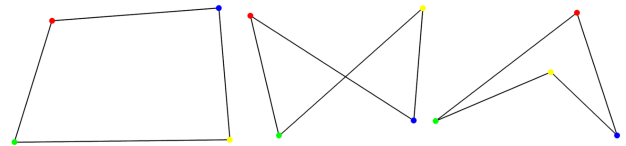


図 2: 対応する四隅の全パターン

として扱う。

オイラー角は、三次元空間内の各軸に対する回転を表す角度であり、三次元回転行列から変換が可能である。このオイラー角の各軸に対する角度の絶対値の合計が小さいほど、カメラに対してピクトグラムが正面を向いていることを表す。反対に、オイラー角の各軸に対する角度の絶対値の合計が大きいほど、カメラに対してピクトグラムが斜めに傾いていることを表す。プロトタイプ B では、得られたオイラー角の各軸に対する角度の絶対値の合計に基づいて判別を行う。

オイラー角を算出する過程で、現実にはありえない変形が行われるようなホモグラフィ行列や角度が算出される場合がある。その場合には不適切であると判断し、それ以降の処理を行わないことで、誤った判別結果の出力を防ぐとともに、パフォーマンスの向上を図るべく次の工夫を施す。

カメラ画像に含まれるピクトグラムを、正解画像のように正面を向けるための変換において、対応する四隅に注目すると、本質的には図 2 に示すような、三パターンに分類できる。このうち、すべての角が 180 度未満の四角形を成していない対応パターンは現実にはありえない。すなわち誤った判別である。実際の対応パターンの例を図 3 に示す。図 3(a) は適切であり問題ないが、図 3(b) や図 3(c) は、現実にはありえない変形であるため、その後の処理を行わず判別結果を無効とする。図 3(a) に示すような正解画像候補の中から最も回転角の絶対値の和が小さな正解画像を判別結果とする。

## 4 比較評価

一枚のカメラ画像からピクトグラムを判別するための平均処理時間とその精度によって実装したプロトタイプを比較評価する。

2) レシオテストとは、最も類似度が高い特徴点と、2 番目に類似度が高い特徴点との特徴量差を求め、しきい値未満の場合は対応付けないようにする手法である。

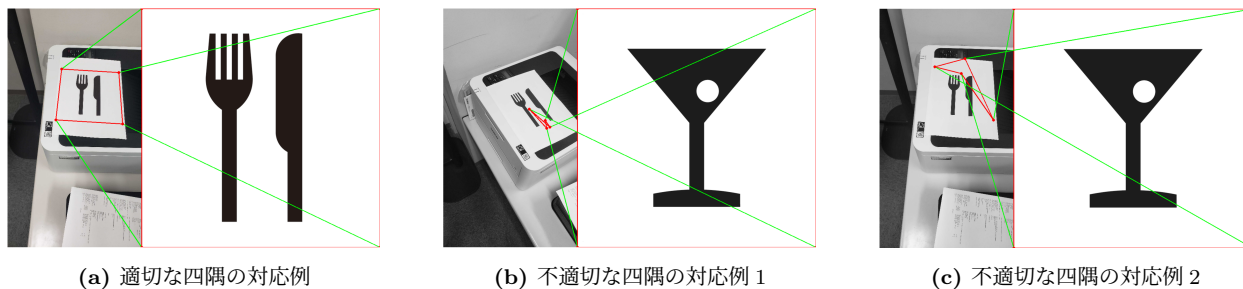


図 3: 対応する四隅の全パターンの実例

## 4.1 判別対象とするピクトグラム

JIS Z 8210 [1] を含む標準案内用図記号 [11] に定められた 156 種類を判別可能なピクトグラムの対象とする。評価のために商業施設として登録されている 11 種に対して 156 種類のうちのピクトグラムであるか判別する。判別処理時間および判別精度の評価には、正解画像を加工して生成したデータセットを用いる。

## 4.2 判別精度評価方法

判別精度の比較評価のために式 (1) に示す評価指標を用いる。

$$S = \frac{\sum_{\theta} \sum_{\varphi} \frac{\cos \theta \cos \varphi}{\text{rank}(\theta, \varphi)}}{\sum_{\theta} \sum_{\varphi} \cos \theta \cos \varphi} \quad (1)$$

ここで  $S$  は角度を考慮した判別精度スコアであり、0 から 1 の実数値をとる。 $\theta$  と  $\varphi$  はカメラ画像内のピクトグラムの横方向と縦方向の角度をそれぞれ表し、 $-90$  度から  $+90$  度までの値となる。 $\text{rank}(\theta, \varphi)$  は角度  $\theta, \varphi$  における正解画像の候補番号である。たとえば、角度  $\theta, \varphi$  において正解画像が第一候補の場合は  $\text{rank}(\theta, \varphi) = 1$  であり、第二候補の場合は  $\text{rank}(\theta, \varphi) = 2$  である。本評価シナリオにおいては 156 種類のピクトグラムから判別を行うため、 $\text{rank}(\theta, \varphi)$  は 1 から 156 までの整数値をとる。

式 (1) では、ピクトグラムを判別する際の角度を重みとして考慮する。すなわち正面 ( $\theta = \varphi = 0$ ) から撮影されたピクトグラムはスコア  $S$  への影響が大きく、角度が加わると影響は小さくなる。

表 1: プロトタイプ A における判別処理時間

項目	時間 [ms]	割合 [%]
特徴点検出と特徴量記述	102.0	16.3
マッチングおよびその他	522.9	83.7
全体	624.9	100.0

表 2: プロトタイプ B における判別処理時間

項目	時間 [ms]	割合 [%]
特徴点検出と特徴量記述	219.2	13.9
マッチングおよびその他	1359.0	86.1
全体	1578.2	100.0

## 4.3 評価結果

判別処理時間を表 1 と表 2 にそれぞれ示す。プロトタイプ A では、一枚のカメラ画像を判別するための平均処理時間は約 625 ms である。一方プロトタイプ B では、平均処理時間は約 1578 ms であり、判別に要する処理時間はプロトタイプ A の方が短い。

プロトタイプ A の判別精度および抽出された特徴点数を図 4 に示す。同様にプロトタイプ B の判別精度を図 5 に示す。プロトタイプ A の全体平均判別精度は 0.411 であり、プロトタイプ B は 0.469 である。プロトタイプ B の方がプロトタイプ A と比較して平均的に良い結果が得られた。

## 5 考察

ピクトグラム判別に要する時間は短ければ短いほど好ましい。ユーザ中心のパフォーマンスモデルである RAIL モデル [12] によると、ユーザが不快に感じない応答時間は一秒以下であるとされている。そのため、処理時間を一秒以下にすることが望まし

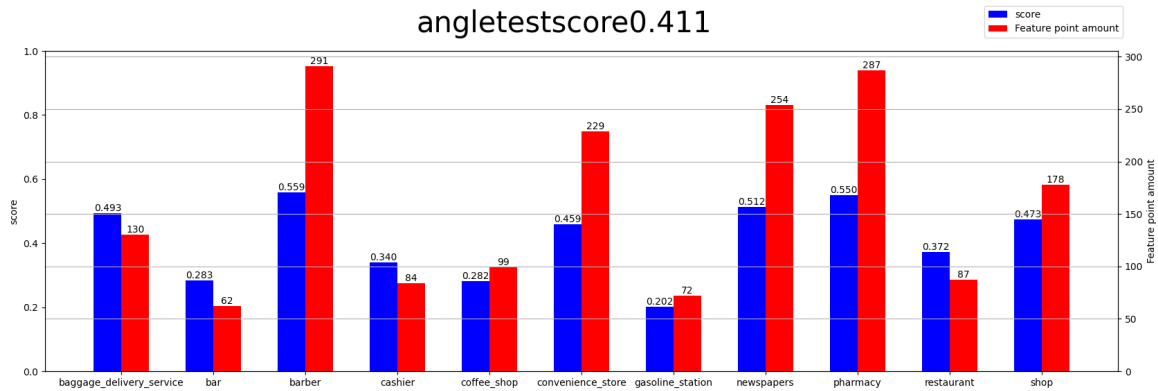


図 4: プロトタイプ A の判別精度および抽出された特徴点数

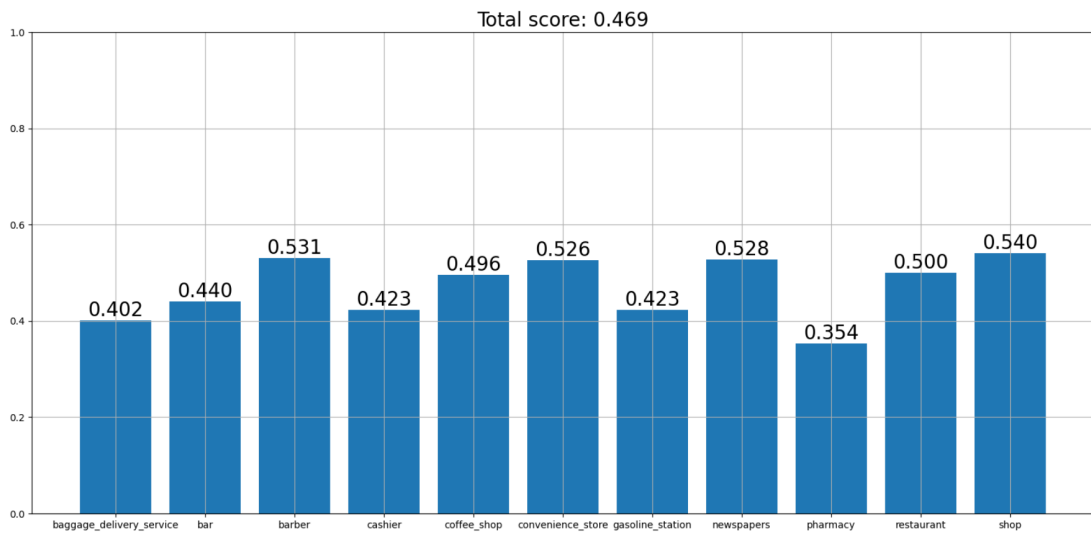


図 5: プロトタイプ B の判別精度



(a) 配達サービス



(b) 手荷物受取所

図 6: あるピクトグラムが別のピクトグラムを包含する例

い。ピクトグラム判別に要する処理時間の観点では、プロトタイプ B は判別所要時間の観点から改善が必要であり、現状ではプロトタイプ A の方が実用に適している。

判別精度に関して、プロトタイプ B の方がプロトタイプ A と比較して平均的に良い結果が得られたが、プロトタイプ A においても抽出された特徴点数が多ければプロトタイプ B より良い場合もある。そ

のため、特徴点を多く検出できるような工夫を施すことでプロトタイプ A の判別精度向上が図れるかもしれない。また、図 6 のようにあるピクトグラムが別のピクトグラムを包含するような場合に精度が低下することがわかった。この問題解決のために、特徴点マッチングの際にこれまで使わなかったマッチしていない特徴点を利用するアイデアがある。

## 6 まとめ

ピクトグラムの画像を判別してコード化し、コード化点字ブロックの音声提供の仕組みを利用して、ピクトグラムの音声化・多言語化を行なうため、ピクトグラム判別システムの実装に取り組んだ。

本稿では実装したピクトグラム判別システムのプロトタイプ二種類について解説した。実装したプロ

トタイプを評価した結果、シンプルなプロトタイプ A はピクトグラム判別に約 625 ms 要する。判別精度向上の工夫を取り入れたプロトタイプ B はプロトタイプ A と比較して判別精度は向上したが、一方でピクトグラム判別に約 1578 ms 要し、RAIL モデルによるユーザが快適に感じるとされる処理時間を超える結果となった。

今後は撮影したピクトグラムに応じて、多言語に対応した音声案内を行うシステムを実装したい。そのために判別したピクトグラムとその位置情報をサーバで集約し、その場に適した音声案内を提供するシステムを構築する。また、あるピクトグラムが別のピクトグラムを包含するような場合に精度が低下する問題の解決のために、マッチしていない特徴点を利用するアイデアを実装して評価したい。

## 謝辞

本研究は北陸地域づくり協会による第 28 回「北陸地域の活性化」助成番号 2023 地域 20 の助成により実現した。ここに謝意を示す。また、報告者らと共に本研究に取り組んだ金沢工業大学の知野凌大氏、藤田和希氏、中村涼氏、および情報経営イノベーション専門職大学の長井悠氏、熊田有里氏に感謝する。

## 参考文献

- [1] バリアフリー：案内用図記号 (JIS Z8210) (令和元年 7 月 20 日) - 国土交通省. [https://www.mlit.go.jp/sogoseisaku/barrierfree/sosei\\_barrierfree\\_tk\\_000145.html](https://www.mlit.go.jp/sogoseisaku/barrierfree/sosei_barrierfree_tk_000145.html). (最終アクセス日時 2024 年 3 月 17 日).
- [2] 飯田優也, 青木貴也, 早川智之, 澤野弘明, 鈴木裕利, 千葉和也. コード化点字ブロックを利用した弱視者向け周辺情報提示手法の検討. 電子情報通信学会技術研究報告; 信学技報, Vol. 119, No. 86, pp. 13–18, 2019.
- [3] 埴田翔. 工学系私大で産学連携が活性化 金沢工業大学における産官学連携による社会実装型研究への取り組み. 産学官連携ジャーナル, Vol. 17, No. 3, pp. 10–13, 2021.
- [4] 松井くにお. コード化点字ブロックによる AI 音声情報案内サービス. 研究結果報告書集: 交通安全等・高齢者福祉, Vol. 27, pp. 31–35, 2021.
- [5] 伊藤康一, 高橋徹, 青木孝文. 高精度な画像マッチング手法の検討. 第 25 回信号処理シンポジウム, No. C5-1, pp. 547–552, 2010.
- [6] 久野真輝, 熊谷瞭, 竹本修, 野崎佑典, 吉川雅弥. CNN によるピクトグラムの認識率の定量的評価とその改善手法. 電気学会研究会資料. IS/情報システム研究会 [編], Vol. 2023, 19-24・26-35, pp. 11–16, 2023.
- [7] 藤田和希, 知野凌大, 坂本真仁, 堀田耕一郎, 松井くにお. 特徴点マッチングによる JIS 規格ピクトグラム判別システム. 情報処理学会第 86 回全国大会講演論文集, pp. 5ZH-04, 2024.
- [8] 知野凌大, 藤田和希, 坂本真仁, 堀田耕一郎, 松井くにお. ホモグラフィ推定による JIS 規格のピクトグラム判別システム. 情報処理学会第 86 回全国大会講演論文集, pp. 5ZH-05, 2024.
- [9] Pablo F Alcantarilla and T Solutions. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans. Patt. Anal. Mach. Intell.*, Vol. 34, No. 7, pp. 1281–1298, 2011.
- [10] Pablo Fernández Alcantarilla, Adrien Bartoli, and Andrew J Davison. KAZE features. In *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7–13, 2012, Proceedings, Part VI 12*, pp. 214–227. Springer, 2012.
- [11] 公益財団法人交通エコロジー・モビリティ財団 | バリアフリー推進事業 > 標準案内用図記号 > 標準案内用図記号使用上の注意. [https://www.ecomo.or.jp/barrierfree/pictogram/picto\\_attention%202021.html](https://www.ecomo.or.jp/barrierfree/pictogram/picto_attention%202021.html). (最終アクセス日時 2024 年 3 月 17 日).
- [12] Google Web Fundamentals, “Measure performance with the RAIL model”. <https://web.dev/articles/rail>. (最終アクセス日時 2024 年 3 月 17 日).