

One-String : CNNを用いたあやとりの工程認識手法と競技性を付加したあやとりゲームの実現

大橋 錬¹ 塚本 啓太¹ 杉本 隼斗² 濱川 礼¹

概要: 本論文では, CNN を用いてあやとりの技の工程認識をおこない, あやとりに競技性を付加したゲームを提案する. 伝承遊びは近年再注目されている分野であり, あやとりもその一種である. しかし, その爽快感と競技性から近年人気を集めているけん玉と比較して, あやとりは注目度が低い. そこで各工程の写真をもとに CNN 学習モデルを作成し, エフェクト表示, 対戦ゲーム化することであやとりに爽快感を付与し, 広い世代で継続して遊びたいと思われるゲーム「One-String」を開発した.

One-String : Realization of a string figure game that adds competitiveness and a process recognition method for string figures using CNN

1. はじめに

あやとりは, 日本において古くから遊ばれている伝承遊びの一種である. あやとりの情景を描いた最古の文献 [1] では, 二人の女の子があやとりをしている姿が描かれており, 17 世紀には存在していたと考えられ, その歴史は古い.

古くから遊ばれているあやとりだが, 大人になってからも遊び続けている人は少ない. 評価実験の際に大学生 10 人にアンケートを行ったところ, あやとりを経験したことのある人は 8 人だった. その内最後に遊んだ経験が保育園・幼稚園と答えた人が 5 人, 小学校と答えた人が 2 人という結果となり, 8 人中 7 人は中学校以降は遊んでいなかった. このアンケートの結果から, 中学校以降もあやとりを遊び続ける人は少ないと考えられる.

伝承遊びの中でも, けん玉は 1975 年に日本けん玉協会が設立され, 国内を対象としたものから海外選手も含めた世界規模のものまで多くの大会が毎年行われ, 昔ながらの遊びという枠から離れて発展を遂げている. 一方あやとりも同様に歴史は古く, 1978 年には協会が設立されているが, 競技性の高い催しは行われていない. 我々は, ストリートスポーツとしても親しまれているけん玉とそうでないあやとりの差は, 競技性にあると考えた.

他者と対戦しているという思い込みと熱中の度合いの関係性を調べた研究 [2] では, 思い込みが熱中の度合いを高めるという結果が得られている. 本研究では, CNN を用いて技の工程認識を行い, あやとりに競技性を付加したゲームを開発することで, 従来のあやとりよりも熱中し, 広い世代で継続して遊びたいと思われるゲームの創出を目指した.

2. 関連研究

2.1 伝承遊びに関する研究

近年, 伝承遊びと呼ばれる古くから遊ばれている遊びが再注目されている.

2.1.1 影絵を用いた半自動アニマルインスタンスの生成

影絵に新たな遊び方を創出するため, 坂本らの研究では, 作った影絵を認識し, プロジェクター上で動かす研究を行っている [3].

デブスカメラで生成された疑似的な影の画像を入力とし, パターン認識を用いて作成した学習モデルで影絵の部分のラベリング, 位置, 向きや大きさを出力する.

例えば犬の影絵を作った場合, プロジェクター上で影絵が犬全体の姿になり勝手に動き出す.

影絵のような道具を一切使わない伝承遊びにおいても, パターン認識を用いて新たな遊び方を創出する研究が行われている. あやとりも影絵と同様に具体的な形を作って楽しむ伝承遊びであるため, 技を認識して絵や写真の表示を

¹ 中京大学 工学部 情報工学科

² 中京大学 工学研究科 情報工学専攻

行うことは One-String にも適している機能だと考える。坂本らの研究ではプロジェクターを用いるが、手軽に遊べるように One-String ではアプリ化を目指した。

2.1.2 けん玉の IoT 製品化『電玉』

「KDDI ∞ Labo」第 9 期のハードウェアプログラムに参加した平野ら 5 人のチームは、けん玉をインターネットと融合した IoT 製品を開発した [4]。

玉の部分にはアルミが使われ、大皿、中皿、小皿、剣先に内蔵されたセンサーと LED が感知する。さらに、剣には姿勢検知センサーを入れることで地面に対して剣先がどちらを向いているかがわかる。本体にはモーターが内蔵されているため、対戦状況においてパイプする機能もある。

これらを利用し、電玉では技を成功させる毎に陣地を獲得し、相手の陣地を減らすことも出来る陣取り合戦形式で対戦して遊べる。

けん玉は、あやとりと同様に道具を使って技を作る伝承遊びのため、陣取り合戦形式の対戦は One-String にも適した対戦方式だと考えた。だが、One-String では、気軽に遊べることを重視し、紐自体に手を加えることはせず、市販の紐でも遊べることを目指した。

2.1.3 ケンケンパの場所の制約を無くした遊びの提案

小澤らは、センサーを取り付けた枠を作成し、枠内に物体が入ったら音や光が出るようなデバイスを開発した [5]。

デバイスは自由に配置することを目指し、体験者の体力や周囲の環境によって自由に配置する。それによって、自発的な運動による発達を促し、暗い場所での使用やエンターテインメントとしての楽しみ方等様々な遊ぶ場所や遊び方の提案が可能となる。

ケンケンパのような体を動かさず遊びでも、デバイスを用いることで新しい遊び方の創出を目指した研究が行われている。

2.2 あやとり完成図からの手順推定法

大澤らは、あやとり完成図を用いて作成手順を文章で出力する研究を行っている [6]。

二次元画像で表現されたあやとり完成図のデータ構造を入力とし、入力したデータ構造の紐をほどいていき、その過程を記録する。この記録を逆にたどることによってあやとり完成図までの手順を出力する。

この研究を参考に、カメラであやとりしている状態を撮影し、紐の状態を写真から二次元画像に変換して認識する手法を検討した。だが、画像内の服の柄や背景によるノイズの影響により、あやとりを遊ぶ画像から正確に紐のみを認識することは困難であると考えた。そのため、背景等の部分をトリミングして取り除き、ノイズを減らした画像で学習を行った。

2.3 ディープラーニングを用いた人物の表情推定

Yu らは、CNN によるディープラーニングを用いて画像ベースで顔検出及び 7 表情の識別を行っている [7]。表情毎の精度のばらつきはあるが、最高約 8 割の精度で表情識別を可能としている。

この研究から、画像の識別において CNN は有効だと考えられる。

3. 提案手法

「One-String」では、ユーザーがあやとりを遊んでいる映像を入力とし、CNN を用いて各工程の判別を行う。判別結果に応じてゲージで進行度を可視化し、全行程が終了した時点のタイムを表示する。データセットの作成には、あやとりを遊んでいる映像を図 1 のように正方形にトリミングした画像を使用する。このデータセットを使用して CNN の構築を行い、あやとり判別モデルを作成する。ゲームにおいてはデバイスのカメラからの映像をトリミング・リサイズした画像を入力として作成したモデルを用いて推論を行う。システムの流れを図 2 に示す。

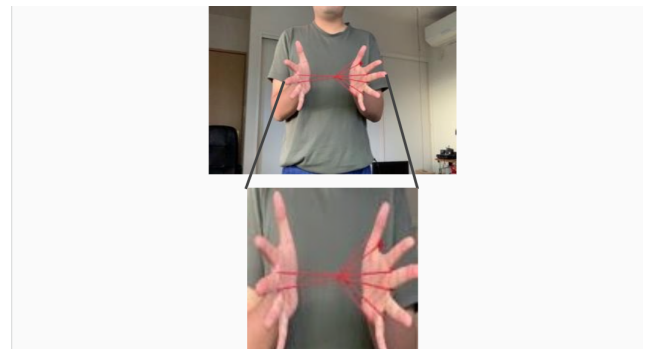


図 1 トリミング例

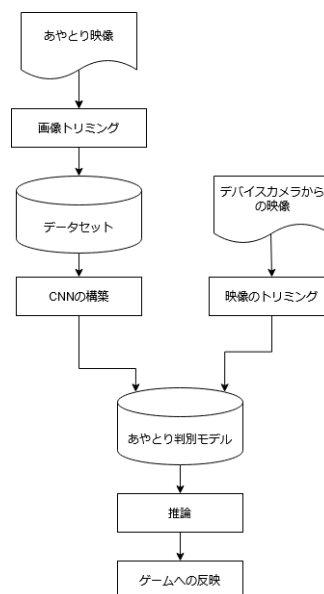


図 2 システムの流れ

あやとりを遊ぶ際、各工程にはたるんでいる紐を張る作業が存在する。その作業が終わったのを認識した時点を経程の完了とする。今回はあやとりの中でも基本的な技である「ほうき」を取り扱う。ほうきは冊子 [8] を参考にし、図 3 に示す 4 工程とした。

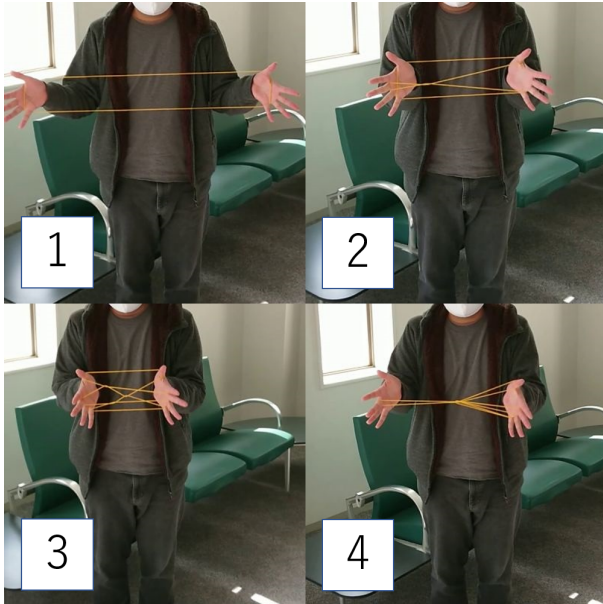


図 3 「ほうき」工程

3.1 画像トリミング

3.1.1 動画からの画像取得

CNN による工程認識モデルのデータセットとして、あやとりをしている様子を動画で撮影し、トリミングをおこなった画像を作成した。トリミングの際は、あやとりの部分のみを図 1 のように両手と紐が収まるように 512×512 の正方形になるようにした。動画は、筆者を含めた 25 人がほうきを作る様子を撮影し、4 工程それぞれで、水平状態、傾けた状態を 5 秒ずつ撮影し、1 工程につき 15 秒ずつ撮影した。学習には多くの画像が必要になるため、OpenCV を用いて 0.1 秒に 1 枚を画像として取得した。

3.1.2 カスケード分類器によるトリミングの自動化

取得した約 15000 枚の画像をトリミングするために OpenCV のカスケード型分類器 [9] を利用して自動化を行った。カスケードファイルの作成には、正解画像を 300 枚、不正解画像を 931 枚使用した。正解画像の作成には、1 人 1 工程につき水平状態、傾けた状態と逆方向に傾けた状態から 1 枚ずつ選択して手動でトリミングをし、服や背景の色が精度に影響を与えないようにグレースケール変換したものを用意した。不正解画像には、あやとりをしている様子が写っていないものを用意した。服装や背景によって輝度は安定しないが、手を使ってあやとりをすることには変わらないため、手を前に出すことによって生じる影と周囲の明

暗差は服装等に関わらず生じると考え、あやとり部分の検出には Haar-Like 特徴量による検出を行っている。

カスケードファイルを使用した各工程のトリミング精度の平均を表 1 に示す。それぞれの工程の精度の平均は、第 1 工程が 0.54、第 2 工程が 0.67、第 3 工程が 0.73、第 4 工程が 0.65 だった。服の柄やアクセサリ、背景のノイズが影響したと考えられ、図 4 の左のように正しくトリミングされた画像もあれば、右のように服の水玉模様に対応してしまった画像もあった。また、特に第 1 工程の精度が低いことの原因は、紐が四角形を成し、紐の交点や斜めになる紐が無いため、工程中で最も特徴の少ない形だったことも影響したと考えられる。



図 4 トリミング結果例

第 1 工程の形はあやとりの基本形であり、ほとんどの技で最初に作る形であるため、第 1 工程の認識はゲームを作るうえで必要性が高いかどうか、必要であればノイズの影響を最小限に抑えて認識する手法を考える必要がある。

表 1 平均トリミング精度

1 工程	2 工程	3 工程	4 工程
0.54	0.67	0.73	0.65

上記のカスケードファイルを使用してトリミングした画像を正解画像として 4 工程各 3000 枚、不正解画像を 3000 枚の計 15000 枚用意した。不正解画像にはトリミング失敗画像のほかに、「ほうき」の各工程に至るまでの紐がたるんでいる状態を用意した。

3.2 CNN 構築

前項までのあやとり画像を利用し CNN によるあやとり工程認識モデルの構築を行う。開発環境として Keras を利用した。モデルのベースとして MobileNetV2 を利用して転移学習を行った。転移学習とは、学習済みネットワークを利用して学習を行うことで、学習コストを抑えられるという特徴がある。ベースのモデルとして MobileNetV2 を採用した理由は、スマートフォン等のモバイル端末で実行することを考慮し、軽量さと精度を重視したからである。使用画像は Data Augmentation によるデータ拡張を行った。表 2 に ImageDataGenerator のパラメーター、図 5,6 に学習を行っ

た正解率と損失関数を示す。

表 2 ImageDataGenerator におけるパラメーター

項目	数値
batch_size	32
rotation_range	30
width_shift_range	0.3
height_shift_range	0.3
shear_range	0.6
zoom_range	0.2
channel_shift_range	100

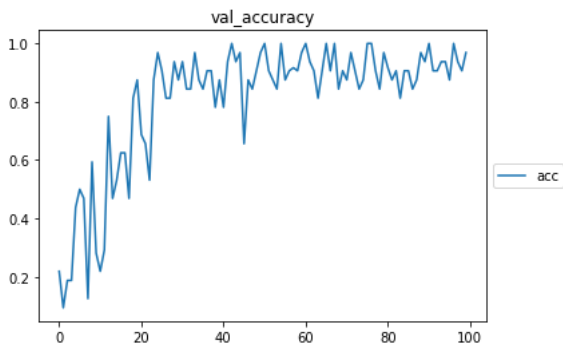


図 5 検証用画像での精度

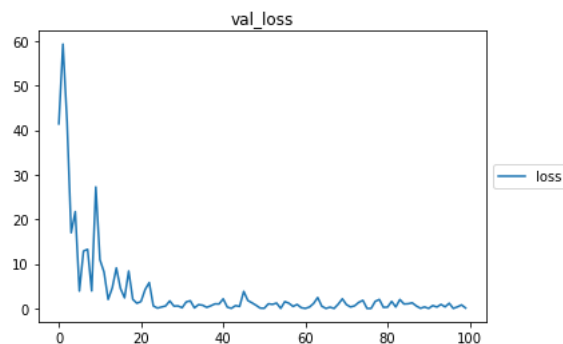


図 6 検証用画像での損失関数

3.3 システムへの実装

3.3.1 推論環境

本研究では、手軽に様々なプラットフォームで体験してもらうことを考え Unity 向けの TensorFlowLite プラグインを用いて実装を行った。対応したモデル形式である .tflite への変換を行うために Keras から h5 形式で出力を行い TensorFlow の pb 形式に変換した後に tflite 形式に変換を行っている。変換には TensorFlow2.0 の API である tf.lite.TFLiteConverter.from_saved_model を用いた。

3.3.2 カメラ映像のトリミング

Unity 上では端末のカメラで取得した映像を入力映像の縦幅の長さに合わせて中心をトリミングしたのち 512×512 にリサイズしている。作成したモデルの入力に合わせるためにこの処理をおこなう。

3.3.3 推論と結果の反映

リサイズ後の画像を入力とし 0.5 秒間隔で推論を実行する。直近 3 回の出力結果をもとに現在の工程と合っているのが確認出来たらゲージの表記を進める。最後の工程が認識されたら、経過タイムを表示する。図 7 が実際のアプリ画面である。

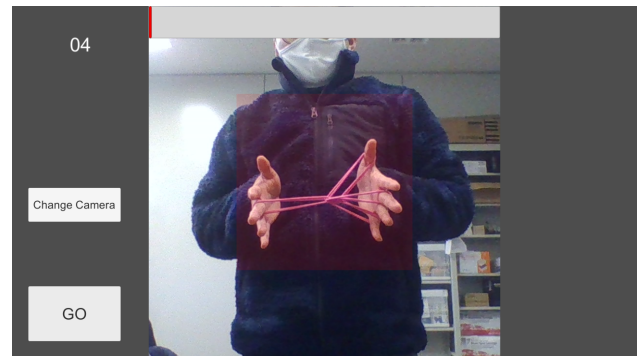


図 7 ゲーム画面

実際に Unity で実装を行い Android アプリとしてビルドをして検証した結果、推論速度は実用に耐えうると判断した。検証をした端末である Xperia XZ1 における 1 回の推論速度の平均値は 0.38 秒であった。

4. 評価実験

4.1 評価方法

評価実験は、被験者に実際にアプリを体験してもらうことで行った。大学生 10 人に Android 端末でアプリを体験してもらった後、アンケートを取った。

4.2 結果と考察

認識精度についての質問では、5 段階評価で 5 が 20%、4 が 40% であり、被験者は概ね満足のいく認識ができていたことが分かった。課題点としては、「地味な紐の色では反応しにくい」といった回答があり環境によっては認識精度が落ちてしまうという点が挙げられる。システムの楽しさについての評価では 5 段階評価で 3 以下が 6 人であり、判定が出来る状態のみでは満足度が低いといえる。また、画面の小さいスマートフォンにおいて、現 UI では画面を確認しながらあやとりを遊ぶのが難しいという回答が得られたので、視認性の改善が必要だということが分かった。

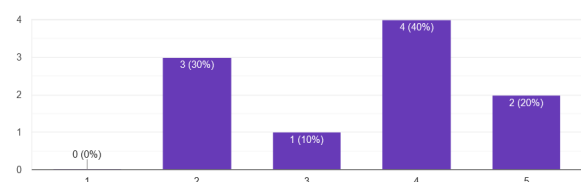


図 8 認識精度についての評価

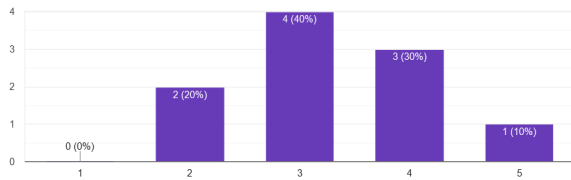


図 9 システムの楽しさについての評価

5. 今後の展望

今後は、トリミング精度、認識精度の向上、「ほうき」以外の技の実装のための汎用的なモデルの開発に取り組んでいきたいと考えている。

あやとり画像の自動トリミングではさらなる精度向上のために、紐自体の検出を行ってトリミングをする手法に取り組みたい。

ゲームの体験面では、エフェクトやBGMによるエンタメ性の向上、自分と対戦相手の進行度合いを同時表示して対戦ゲーム化させることによる競技性の付加に取り組みたい。

評価実験後のアンケート結果には「あやとりの仕方を提示しつつ、行程の確認が出来たらいい」というコメントもあったため、初心者への練習支援にも取り組みたいと考えている。

参考文献

- [1] 井原西鶴：諸艶大鑑 (1684 年)。
- [2] 川合 伸幸, 中田 龍三郎, 久保 賢太：誰かと対戦しているとの思い込みがビデオゲームを楽しくさせる:事象関連電位 (P300) による検討, 『The 33rd Annual Conference of the Japanese Society for Artificial Intelligence』 (2019)。
- [3] 坂本 賢哉, 高山 裕介, 目黒 僚, 日栄 航：影絵を用いた半自動アニマルインスタンスの生成, 『エンタテインメントコンピューティングシンポジウム (EC2019)』 (2019)。
- [4] 【検証】「けん玉」が IoT で『電玉』に進化！アメリカ人トップ・プレイヤーを召還して開発チームがガチバトル, <https://time-space.kddi.com/kddi-now/tsushin-chikara/20160222/> (2016)。
- [5] 小澤 靖子, 金丸 紫乃, ジョ ビュンヒュン, 串山 久美子：こどもの運動能力向上のためのケンケンパ遊びの提案, 『情報処理学会 インタラクション』 (2018)。
- [6] 大澤 優人, 西田 誠幸：あやとり完成図からの手順推定法, 『情報処理学会研究会』 (2011)。
- [7] Zhiding Yu, Cha Zhang : image based static facial expression recognition with multiple deep network learning, Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (2005)。
- [8] あやとり探検隊：あやとりしようよ！ ((実用 BEST BOOKS)), 日本文芸社 (2005), pp.22
- [9] OpenCV 2.2 (r4295) C リファレンス:カスケード型分類器, http://opencv.jp/opencv-2.2/c/objectdetect_cascade_classification.html。