

# 多人数が映る動画からの表情解析: 子どもの興味の対象を見つけるシステムの作成に向けて Analysis of Facial Expressions from Video Image of Many People: Toward a System That Find Things of Interest to a Child

吉崎 美紗<sup>\*1</sup>  
Misa YOSHIZAKI

早川 博章<sup>\*1</sup>  
Hirofumi HAYAKAWA

岡 夏樹<sup>\*1</sup>  
Natsuki OKA

大森 隆司<sup>\*2</sup>  
Takashi OMORI

長井 隆行<sup>\*3</sup>  
Takayuki NAGAI

<sup>\*1</sup>京都工芸繊維大学  
Kyoto Institute of Technology

<sup>\*2</sup>玉川大学  
Tamagawa University

<sup>\*3</sup>電気通信大学  
The University of Electro-Communications

In childcare services, it is said that teaching fit for each child is important. Childcare workers hence need to understand children's personality from usual nursery life. However, we think it is difficult for fresh leaders to understand individual character of children. We may be able to reduce their work by building a system to find things children are interested in. We set the final goal of creating a system that discovers things in which children have interest. As the first step, we analyzed facial expressions from video image of many people in this paper. We used histograms of oriented gradients (HOG) and SVM with RBF kernel for the analysis. As a result, f-measure of face detector was 0.813 and facial expression analyzer was 0.861. However, we couldn't build an analyzer that suits each child. We thus have a plan to create an analyzer that fits each child by personal identification.

## 1. はじめに

近年の女性の社会進出の増加や少子化対策にとって、待機児童の存在は大きな障害となっている [内閣府 09]。そのため社会における保育の需要は増々高まってきており、指導者の数を増加させるという試みも実施されている [厚生労働省 15]。また保育の現場では指導者には子ども一人一人に合わせて指導することが要求されており [文部科学省 08][厚生労働省 08]、普段から子どもたちを観察し、それぞれの子どもの傾向や性格を掴まなければいけない。しかし、指導者の仕事は観察することだけに留まらないため、経験がある指導者であっても常時大人数の子どもを観察し続けることは難しい上、子ども達の性格というものは指導者の感覚に頼る部分が大きい情報であるため指導者の経験によって掴める情報に差があること [吉田 15] から、指導者間の情報共有も以前ほど円滑には行えない可能性がある。よって、今後も以前のように子ども一人一人に対応した指導を継続することは難しくなると考えられる。

このような状況の中で、指導者の代わりに子ども達を常に観察しその情報を指導者に提供するシステムがあれば、指導者全員が経験の有無に関わらず客観的な情報を得ることが出来るため、指導者の負担が軽減しより効果的に保育を行うことができるのではないかと考えた。必要な情報の一つとして今回は子ども達の好きな遊びという情報に着目した。子ども達の一日の多くは遊びの時間となっていることや、大人が遊びに介入することによって遊びをよりよく出来る [飯島 08] といわれているためである。そこで遊んでいる子ども達の表情をビデオカメラで撮影し映像を解析することによって、表情から興味の対象を発見しようと考えた。しかし、ビデオカメラと対面しながら遊ぶという状況は普段の遊びからは考えられず、本研究で重要と考えられる自然な表情が現れない可能性が高い。そこでカメラと対面せずに遊んでいる子供たちの表情解析を行わなければならないが、従来の表情解析では子どもとカメラが対面していることを前提としてなされているものが多く、今回のようなカメラを気にせずに遊んでいる子ども達を対象としている表情解析

には適していない。そのためこのような環境に合わせた表情解析システムを独自に作成する必要がある。そこで本研究では、子どもの興味の対象を見つけるシステムを作るための足掛かりとして、多人数がカメラと対面せずに映っている動画からの表情解析を行う。

## 2. 提案システム

本研究で提案する表情解析システムを図解したものが図 1 となる。

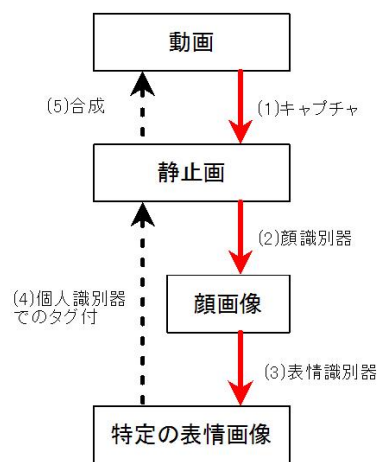


図 1: 提案システムの図解

- (1) 一秒に三枚程度のペースで動画を静止画にする。この時、何秒の時点での静止画なのか分かるよう、秒数を記載した状態で静止画にする。
- (2) 静止画全てを顔識別器にかけ、画像の中から人の顔を検出する。顔識別器作成過程の詳細は五章で述べる。
- (3) 切り取った顔画像を先ほどと同じように表情識別器にかけ、その時に検出したい表情の顔画像とそうでない顔画像に分類する。表情識別器作成過程の詳細は六章で述べる。

連絡先: 〒 606-8585 京都市左京区松ヶ崎橋上町 1  
 京都工芸繊維大学大学院工芸科学研究科 情報工学専攻  
 インタラクティブ知能研究室, misa@ii.is.kit.ac.jp

- (4) 検出したい表情を個人識別器にかけ、元の静止画に映っている誰の顔なのか検出し、座標を記録する。
- (5) 記録した座標と静止画に記載された秒数から、動画に直接タグ付けを行う。

以上が提案システムの概要である。

### 3. 動画の取得

本研究では、リトミックをしている子ども達を遊んでいる子ども達として撮影させてもらい、その動画を用いて解析を行った。リトミックを採用した理由としては、子ども達の個性に合わせた様々なプログラムが用意されているため、より子ども達の好きなものの傾向が現れやすいのではないかと考えたからである。リトミックとは、リズムによってもたらされる全体的調和をもとにして、子どもの音楽的素質の向上をはかった教育法である[村山 66]。具体的には全身を使いリズムを感じることで、音楽の基礎となるリズムを身につけさせようというものである。

#### 3.1 リトミック時の状況

リトミックを行った環境の見取り図は図 2 の様になっている。図 2 の○と●はカメラの位置に対応しており、○はビデオカメラ（計 3 台）、●は kinect v2<sup>\*1</sup>（計 1 台）である。加えてリトミックを主導する先生には Tobii Pro Glasses 2<sup>\*2</sup> を装着してもらい、全部で 5 台のカメラで実験を撮影した。また、参加者全員に加速度測定器を内蔵したハーネスを装着してもらい、個人ごとの動きも測定した。今回の実験には二歳児クラスと三歳児クラスの子ども達に参加してもらった。以降各クラスのリトミックの詳細を述べる。尚、両クラスのリトミックを主導する先生は同一人物である。

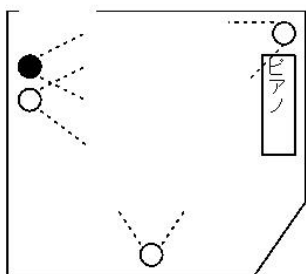


図 2: リトミックを行った環境

##### (1) 二歳児クラス

二歳児クラスでは、子ども達は 13 人（内、男 5 人、女 8 人）、先生はリトミックを主導する先生含め 3 人参加した。子ども達の年齢は二歳～三歳で、リトミックを行った時間は 30 分。その時間中すべてのカメラでリトミックを撮影していた。

30 分の間に行ったプログラムは

- (1) 太鼓のリズムに合わせて歩いたり走ったり。音が止まったら止まる。
- (2) 絵本の登場人物になりきってみんなで動く。

<sup>\*1</sup> ジェスチャーや音声で入力が出る周辺機器。今回は動画と人のモーションを検出するために使用

<sup>\*2</sup> メガネ型のカメラになっており、装着者の視野範囲を撮影することが出来る。 <http://www.tobiipro.com/>

- (3) ピアノの音に合わせてボールを上投げる。
- (4) ピアノの音に合わせてボールを転がす。
- (5) マラカスをじょうろに見立てて持って歩く。時々花に見立てた布に水をあげる。
- (6) みんなで円になって座り、動きを交えて歌う。

となっていた。

##### (2) 三歳児クラス

三歳児クラスでは、子ども達は 10 人（内、男 4 人、女 6 人）、先生はリトミックを主導する先生含め 3 人参加した。子ども達の年齢は三歳～四歳で、リトミックを行った時間は 30 分。その時間中すべてのカメラでリトミックを撮影していた。

30 分の間に行ったプログラムは

- (1) 一枚の布をみんなで持って、布を波に見立てて曲に合わせて遊ぶ。
- (2) 太鼓のリズムに合わせて歩いたり走ったり。音が止まったら指示した体勢になる。
- (3) 上記の太鼓の音がピアノに変わる。
- (4) 紙芝居の登場人物になりきってみんなで動く。
- (5) 小さな布を紙芝居の中に出てきた蛇に見立てて動かす。
- (6) ピアノの音に合わせてボールを上投げる。
- (7) みんなで円になって座り、動きを交えて歌う

となっていた。

#### 3.2 解析手法

顔識別器、表情識別器を作成する際に用いた解析手法について説明する。

#### 3.3 Histograms of Oriented Gradients

Histograms of Oriented Gradients(HOG) とは、物体検出のための代表的な局所特徴量で、局所領域での輝度の勾配方向をヒストグラム化した特徴量である [Dalal 05]。照明の変化に強いといわれており、画像の大まかな概形を掴むのに適しているため、人認識や物体認識に用いられることが多い。算出の手順は (1) 輝度勾配の算出 (2) 局所ヒストグラム化 (3) 局所ブロックによる正規化の 3 つに分かれる。

今回は HOG 特徴量をソフトマージン SVM にかけることで学習を行った。尚、SVM は RBF カーネルを用い、コストパラメータを調節し、最も高い精度となるパラメータを求めた。

#### 3.4 Dlib について

Dlib は主に C++ の機械学習アルゴリズムとツールが入っているオープンソースのツールキットである [Davis 09]。アルゴリズムが使いやすいうようにサンプルコードが多く提供されており、様々な機械学習が簡単に利用できる。また、C++だけでなく Python にも対応している。SVM は前述したすべての手法が対応しており、解析したいデータの特徴に合わせて、適切な手法を選択することが出来る。

本研究では HOG 特徴量の算出から SVM での学習器の作成までこのキットで行った。SVM ではソフトマージンと RBF カーネルを用いた。

## 4. 顔識別器の作成

### 4.1 手続き

顔識別器作成の手順を説明する。

- (1) 実験動画を一秒に三枚のペースで静止画にし、6000×8枚の静止画にする。
- (2) 顔がはっきりと写っている画像を170枚厳選する。
- (3) 厳選した画像のどこに顔があるのかを記録するため、図3のように顔の左上の座標、その座標からの幅、高さをすべての画像に対してタグ付けをする。記録する顔は両目、鼻、口が全て見えているものとし、どれか一つでも欠けているものは記録しない。
- (4) タグ付けを終えた画像170枚のタグ付けを行った範囲のHOG特徴量を算出する。
- (5) 特徴量データを用いて10-foldクロスバリデーションを行う。
- (6) コストパラメータを変更しクロスバリデーションを繰り返す。
- (7) 最も精度の良かったパラメータで識別器を作成する。

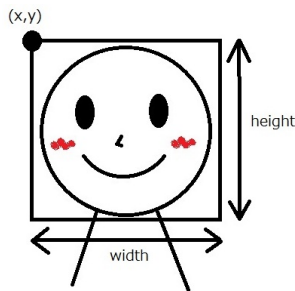


図 3: タグ付けに必要な特徴

### 4.2 結果

#### (1) コストパラメータを変化させた際の推移

コストパラメータを変化させた際の precision, recall, F 値<sup>\*3</sup>の推移を図4に示す。

#### (2) 最終精度

最も F 値が高かったのは  $C = 108$  と設定した時であった。その時の precision が 0.832, recall は 0.794. よって F 値は 0.813 となった。

## 5. 表情識別器の作成

### 5.1 手続き

表情識別器作成の手順を説明する。

- (1) 顔識別器で切り抜いた画像から、顔の方向を揃えた画像を150枚厳選する。
- (2) 厳選した画像のどれが笑顔なのかを記録するため、笑顔の部分に顔識別器と同様に図5の特徴をタグ付けする。

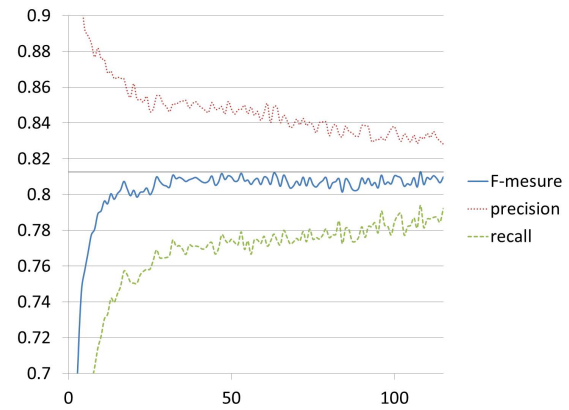


図 4: コストパラメータを変化させた際の推移

- (3) タグ付けを終えた画像150枚のタグ付けを行った範囲のHOG特徴量を算出する。
- (4) 特徴量データを用いて10-foldクロスバリデーションを行う。
- (5) コストパラメータを変更しクロスバリデーションを繰り返す。
- (6) 最も精度の良かったパラメータで識別器を作成する。

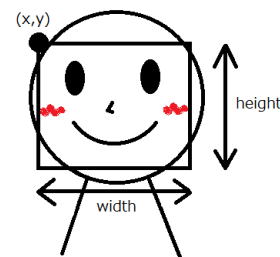


図 5: タグ付けに必要な特徴

### 5.2 結果

#### (1) コストパラメータを変化させたときの推移

コストパラメータを変化させた際の precision, recall, F 値の推移を図6に示す。

#### (2) 最終精度

最も F 値が高かったのは  $C = 60$  と設定した時であった。その時の precision が 0.882, recall は 0.84. よって F 値は 0.861 となった。

## 6. 考察

今回、パラメータを調節した結果、顔識別器の精度 (F 値) は 0.813, 表情識別器の精度 (F 値) は 0.861 にすることが出来た。しかし、識別器を作成している中でシステムとして成立させるための課題が見つかった。図7はそれぞれの子どもの表情識別器で笑顔と判定された枚数と笑顔であるのに笑顔と判

\*3  $\frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$  で算出できる

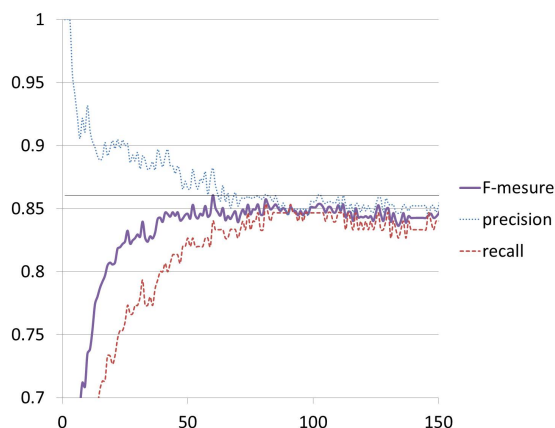


図 6: コストパラメータを変化させた際の推移

定されなかった枚数である。この図からも分かるように、子どもによって笑顔と判定される数、されなかった数が大きく偏っている。偏った識別器ではシステムの目標でもある各子ども達の特徴を正しく掴むことは不可能であると考えられる。何故かと考えるため、表情識別器を作成する際に用いた学習データについて検討する。図 8 は学習データに用いたそれぞれの子ども達の画像枚数である。学習データも結果と同様に偏っていることが分かる。これが判定された枚数が偏った原因であると考えられる。つまり偏らない識別器を作成するためには学習データを均等にすることが必要である。よって今後は学習データを均等にすることで、検出された顔がどの子どものかを自動分類できるようにしていきたいと考えている。また、このシステムは子ども達の興味を見つけることを最終目標としている。興味というのは人間の基本表情の中には無く、細かい特徴から検出していかねばいけなことが予想される。その際、すべての子ども達を同様の検出器にかけることが適切ではない可能性が高い。その二つを解決するためにも今後は個人識別器を導入し、子どもごとの表情検出器の作成を行いたいと考えている。最終的にはすべての作業を自動で行えるようにし、システムとして完成させていきたい。

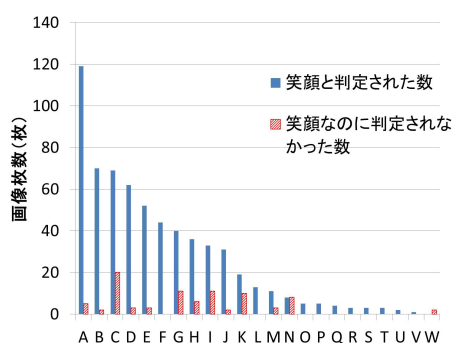


図 7: 表情識別器での判定結果

## 7. まとめ

本研究は、子どもの興味の対象を見つけるシステムを作ることを見込みとしている。作成するシステム全体の流れとし

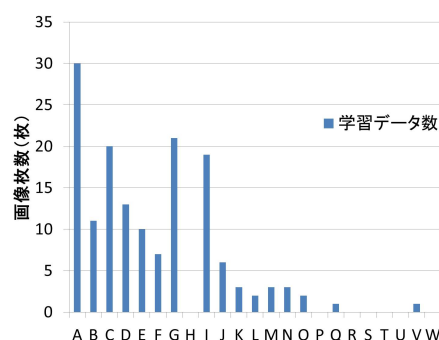


図 8: 表情識別器に用いた学習データの枚数

ては、まず動画を静止画にしたものから顔の部分だけを切り取り、表情に着目して解析を行う。その結果を動画に反映させるというものである。本論文ではそのための足掛かりとして、システムの前半部分である多人数がカメラと対面せずに映っている動画からの表情解析を行った。解析を行うための特徴量として輝度勾配 (HOG) 特徴量を算出し、その特徴量を SVM にかけることで目標とする表情の識別を行った。SVM は RBF カーネル、ソフトマージン法を用いてコストパラメータを調節し、高い精度となるパラメータ値を求めた。結果顔識別の F 値は 0.813、表情認識の F 値は 0.861 となった。結果をみると検出されにくい子どもがいることが分かった。これはシステムの本来の目的である子ども達一人一人に合わせた表情解析に適していない。よって今後は個人識別を行い、個人ごとに表情解析を行っていききたい。

## 参考文献

- [内閣府 09] 内閣府: 少子化社会白書, ぎょうせい (2009).
- [厚生労働省 15] 厚生労働省: 厚生労働省 保育士確保プラン : 2015(平成 27) 年 1 月 14 日, 保育情報, No. 459, pp. 35-50 (2015).
- [文部科学省 08] 文部科学省, 厚生労働省: 幼稚園教育要領: 平成 20 年告示: 原本, チャイルド本社 (2008).
- [厚生労働省 08] 厚生労働省: 保育所保育指針: 平成 20 年告示, フレーベル館 (2008).
- [吉田 15] 吉田 満穂, 片山 美香, 高橋 敏之, 西山 修: 保育経験からみた気づき体験の特徴, 岡山大学教師教育開発センター紀要, Vol. 5, pp. 9-18 (2015).
- [飯島 08] 飯島 典子: 「気になる」 子どもの遊びの成立を促す保育者の働きかけ, 東北大学大学院教育学研究科研究年報, Vol. 57, No. 1, pp. 328-338 (2008).
- [Dalal 05] N Dalal, B Triggs: Histograms of oriented gradients for human detection, Computer Vision and Pattern Recognition, Vol. 1, pp. 886-893 (2005).
- [Davis 09] Davis E. King: Dlib-ml: A Machine Learning Toolkit, Journal of Machine Learning Research, Vol. 10, pp. 1755-1758 (2009).
- [村山 66] 村山 和: 幼児の音楽教育、特にリトミックについて, 札幌大谷短期大学紀要, Vol. 3, pp. 95-111 (1966).