情報処理学会論文誌　デジタルコンテンツ　Vol.5　No.2　11–19 (Aug. 2017)

推薦研究論文

お絵かきサウンドシステム「らくがっきー」におけるオブジェクト解析手法の改良

加藤 里美[1,a)]　水野 慎士[1,b)]

受付日 2017年1月15日, 採録日 2017年5月26日

概要：「らくがっきー」は絵を描きながらインタラクティブにサウンドを生成できるメディアシステムである．サウンドの生成は描かれた絵に含まれるオブジェクトを検出することで実現するが，従来システムは描かれたオブジェクトの検出に少数のサンプルに基づく単純な形状特徴量を用いていたため，バラエティのあるオブジェクトの検出は困難であった．そこで，本論文ではオブジェクトの検出のため，大量のサンプルから共通する特徴を抽出して各オブジェクトの検出に用いる機械学習の手法を取り入れる．改良したシステムは従来システムに比べてバラエティに富む手描きオブジェクトを精度良く安定的に検出して，絵に適したサウンドを生成することが可能となった．「らくがっきー」を一般の人に使ってもらった実験では，多くの人からお絵描きが楽しくなったという評価を得た．

キーワード：お絵かき，サウンド，インタラクション，Haar-Like 特徴量，機会学習

Improvement of the Analyzing Method for Drawing and Sound Generation System "RAKUGACKY"

Satomi Kato[1,a)]　Shinji Mizuno[1,b)]

Received: January 15, 2017, Accepted: May 26, 2017

Abstract: "RAKUGACKY" is an interactive media system that could generate sounds from a hand-drawn sketch. "RAKUGACKY" is a media system that could generate sounds through sketching. Sounds are generated based on objects detected from a hand-drawn sketch. The former system used simple features of shapes selected from a small number of samples, and it was difficult to recognize objects that have many variations. In this paper, we apply a machine learning method to recognize hand-drawn objects, which uses common features detected from a large number of samples. The improved system could recognize hand-drawn objects more accurately and stably than the former system. In our experiment, many users felt sketching with "RAKUGACKY" more pleasant than usual sketching.

Keywords: sketch, sound, interaction, Haar-Like feature, machine learning

1. はじめに

コンピュータ技術の発展にともない，デジタル技術を用いて映像やサウンドを提示したり，人の操作などに対して対話的に反応したりするインタラクティブメディアアートが数多く開発されている．デジタルコンテンツでは画像とサウンドを扱うものが多い．画像と音声は密接に関連しており，音により強調されたり，画像の印象を変えたりすることがある．そのため画像から音楽を生成するような手法がいくつも研究されている．たとえば，ユーザが実際に描いた絵をもとにユーザが実際に音を吹き込んで楽しむデジタルコンテンツ [1]，ユーザの動きにより音を生成するコンテンツ [2]，ユーザが描いた曲線をもとに作曲の支援をするアプリケーション [3]，ユーザの描いた線やディスプレイをタッチするなどの動作をもとに音楽を奏でるコンテン

1 愛知工業大学大学院経営情報科学研究科
Graduate School of Business Administration and Computer Science, Aichi Institute of Technology, Toyota, Aichi 470–0392, Japan
a) b15709bb@gmail.com
b) s_mizuno@aitech.ac.jp

ツ [4]，など人の操作に対して対話的に反応し音を生成するコンテンツが開発されている．これらは，デジタルサイネージやエンタテイメント，教育などの分野で活用されており，今後の進展も大きく期待されている．

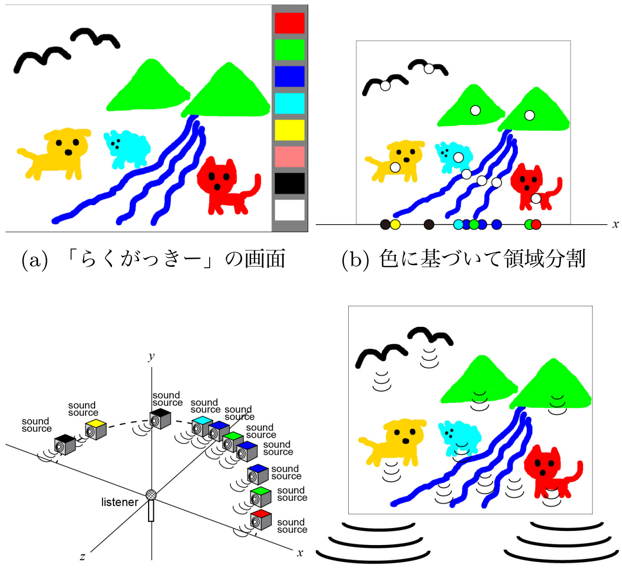
このような背景の中，著者らは絵と音を融合したインタラクティブデジタルコンテンツである「らくがっきー/RAKUGACKY」を提案して開発してきた [5]．このシステムでは，ユーザはスクリーン内にお絵描きをすることで，システムがそのお絵描きに対応するサウンドを自動的に生成する．システムはユーザが絵を描いている間に絵の解析を行い，その解析結果に基づいて対話的なサウンドの生成や変更を行う．ユーザはサウンドをともなったお絵描きを対話的に楽しむことができる．しかし，現状のシステムでは少数のサンプルに基づいて開発者が選択した特徴量に基づいて絵の解析を行っていたため，描かれた絵に対するサウンド生成の際に想定外の絵が描かれた場合，正しいサウンドが生成されない場合がある．

そこで「らくがっきー」の描かれた絵に対する新たな検出手法を提案する．本論文では特にデフォルメされたイラストに含まれる特徴的な形状の描画に着目して，機械学習に基づく手法で大量の動物や自動車などのイラスト画像から抽出した共通的な特徴量を選択して，オブジェクトの検出に用いる手法を提案する．提案手法を適用して改良した「らくがっきー」は，従来システムよりも手描きの絵から適切なサウンドが生成されることが期待できる．

2. 「らくがっきー」の概要

「らくがっきー」とは絵と音を融合したインタラクティブデジタルコンテンツである．図 **1** は「らくがっきー」のプロセスを示している．ユーザは PC 画面内に表示されたキャンバスにペンタブレットやマウスを用いてパレットで色を選択しながら，一般的なペイントツールと同様に自由にオブジェクトを描いていく．このとき，描かれたオブジェクトに応じて様々なサウンドを対話的に生成する．これによりお絵描きをしながらサウンドの生成を楽しむことができる．

生成されるサウンドを描いた絵によって変化させるため，システムは絵の中に描かれたオブジェクトの検出を行う．検出対象のオブジェクトとしては，不思議なスケッチブック [6] を用いた子供向けワークショップなどでの経験に基づいて（図 **2**），子供たちが描きそうな猫，ひよこ，山，池，川など 10 種類とした．まず，システムはキャンバス上に描かれた絵を色に基づいて領域に分割する．そして，各領域の形状を面積，長さ，円形度，傾斜，湾曲など複数の形状特徴量を計算して，オブジェクトの色と形状特徴量に基づいて各領域を識別する．たとえば，青色領域は水に関連するオブジェクトとして，描く領域の面積，長さ，円形度，傾きを用いて雨粒，池，川，海の 4 種類のオブジェクトの検出を試みる．



(a) 「らくがっきー」の画面　(b) 色に基づいて領域分割

(c) 領域ごとに音源を空間に配置　(d) 各位置で音を発生

図 1　「らくがっきー」の概要

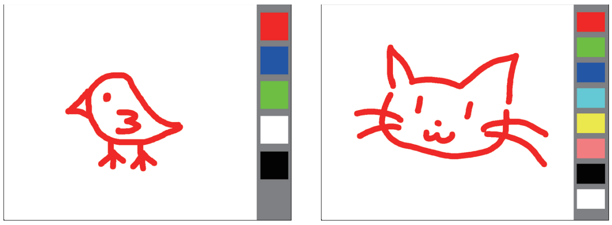
Fig. 1　Overview of “RAKUGACKY”.

図 2　「不思議なスケッチブック」で子供が描いた絵

Fig. 2　Pictures drawn by children in “Amazing Sketchbook”.

対象の絵からオブジェクトを検出した後に，各オブジェクトの位置に基づいて音源を仮想空間内に配置する．各オブジェクトに対応する音源はあらかじめ WAV ファイルとして用意されており，オブジェクト領域の形状特徴量に基づいて音源のピッチは変更される．システムは，仮想空間中に配置された各音源を合成することでステレオサウンドを生成する．

オブジェクトの検出とサウンド生成は繰り返し行われるため，ユーザが描画中に音は発生し続けて，オブジェクトを追加描画するたびに音源も加えられてサウンドも変化する．これにより，ユーザは対話的に絵を描くことを楽しむことができる．図 1 の例では，猫（赤），犬（黄），羊（シアン），鳥（黒），山（緑），川（青）の 6 種類のオブジェクトが描かれており，それぞれのオブジェクトに対して猫，犬，羊，鳥，鈴虫の鳴き声，川のせせらぎ音の音源が配置されて，すべての音源を合わせたサウンドが生成されている．



(a) 猫の鳴き声を生成するひよこの絵　(b) 輪郭欠損で鳴き声が生成されない猫の絵

図 3　従来システムで検出に失敗するオブジェクト例

Fig. 3　Examples of objects that fail to be detected in the former system.

3.　従来の「らくがっきー」の問題点

前述のとおり，「らくがっきー」では描いた絵を解析してオブジェクトを検出することで配置する音源の種類を選択する．オブジェクトの検出はまず色に基づく領域分割を行い，次に各領域の形状解析を行う．このとき，従来システムにおける形状解析は，少数のサンプル画像に基づいて選択・決定したオブジェクトの面積，周囲長，円形度など単純な形状特徴量や，左右端点と領域当分点との位置関係に基づく発見的特徴量に基づいていた．そのため，オブジェクトの検出精度は必ずしも十分でなく，描いたオブジェクトに適切なサウンドが生成されなかったり，わずかな変化でオブジェクトが検出されなかったりすることが多かった．

たとえば，赤色で猫の顔を描くと猫の鳴き声が生成されるが，検出条件はオブジェクトの面積と円形度だけを用いている．そのため，**図 3** (a) に示したようなひよこの絵を描いた場合でも猫の検出条件を満たしてしまい，猫の鳴き声のサウンドが生成されてしまう場合がある．また，猫の顔の輪郭がわずかに切れていたり（図 3 (b)），顔に胴体を描き加えたりすると，猫として識別されなかった．

4.　「らくがっきー」の改良

4.1　概要

従来の「らくがっきー」の問題点を解決するには，単純な形状特徴量や発見的特徴量に加えて，手描きオブジェクトをより詳細に分析したり，見本となるオブジェクトとの比較を行ったりしながら，オブジェクトを検出する必要がある．

オブジェクトの検出や識別のためにしばしば用いられる手法としてはテンプレートマッチングがあげられる．テンプレートマッチングでは検索対象画像からテンプレート画像と類似する領域を検出する手法である．類似度の計算は輝度値の差や相関係数などが用いられるほか，近年ではSIFT や SURF といった局所特徴量を用いた手法も提案されており，文字や記号など形状のバラエティが少ないオブジェクトや特定の画像を検出するには非常に有効な手法である．しかし，手描きオブジェクトの場合，猫を描いた場合でも様々な描き方が考えられるため，テンプレートマッチングを「らくがっきー」に適用することは困難である．

種別が同じでもバラエティに富んだオブジェクトを認識するには，多数のサンプルから共通する特徴を見つけ出して用いることが必要になり，いくつかの手法が提案されている．たとえば，X 線画像からの腫瘍検出に関する研究では特徴量ベクトルとして注目領域のすべての画素値を用い，主成分分析で次元を下げることで多数のサンプルに共通する画素値の特徴を抽出している [7]．

そして，最近では画像認識に深層学習を適用する研究が数多く報告されているが，手描きスケッチへの適用例は発展途上であるといえる．たとえば，写真と手描きスケッチの認識に応用した研究例が報告されているが [8]，インタラクティブコンテンツへの応用は行われていない．また，手描きの絵の内容をリアルタイムで判定する「Quick Draw」というコンテンツが開発されているが [9]，現状では単独のオブジェクトにのみ対応している．

バラエティに富んだ代表的なオブジェクトとして顔がある．顔認識では特徴量として画像の局所的な明暗差の情報である Haar-like 特徴を用い，多数の弱検出器を生成しながら多数のサンプルに合わせて各検出器の重みを調整して全体的な検出器を構成していく Adaboost が有効である [10], [11]．これらの手法では画像中に含まれる複数の顔をリアルタイムで正確に検出することができる．そして，Haar-like 特徴を用いて動物の顔の検出を行った研究 [12] や手描き顔画像の認識を行った研究 [13] も報告されている．

そこで，本論文では「らくがっきー」で描かれた絵からのオブジェクト検出に Haar-like 特徴を用いることを提案する．そして adaboost によって検出したいオブジェクトの特徴量を学習させた検出器を生成して，検出器を「らくがっきー」で描かれた絵に適用して，絵に含まれる複数オブジェクトを検出することを試みる．

4.2　検出器の生成

本論文では，従来システムで子供の絵によく見られるとして検出対象としていたオブジェクトのうち，猫，羊，ひよこを Haar-like 特徴に基づく検出器の対象とする，また新たにライオン，自動車，木，イルカを検出器の対象オブジェクトとする．生成されるサウンドは，自動車はクラクション音，木は風音で，それ以外の動物は鳴き声である．実装は文献 [11] の手法に基づく．

提案手法では初めに学習によって検出器を生成する必要がある．そこで各検出対象オブジェクトの学習用画像をそれぞれ 1,000 枚以上用意する．これらの画像は手描きイラストやイラスト風の画像で，グレースケール化を行っている．**図 4** に学習用画像の例を示す．また，検出対象でないオブジェクトとして，人の顔や動物のイラスト，ボールな

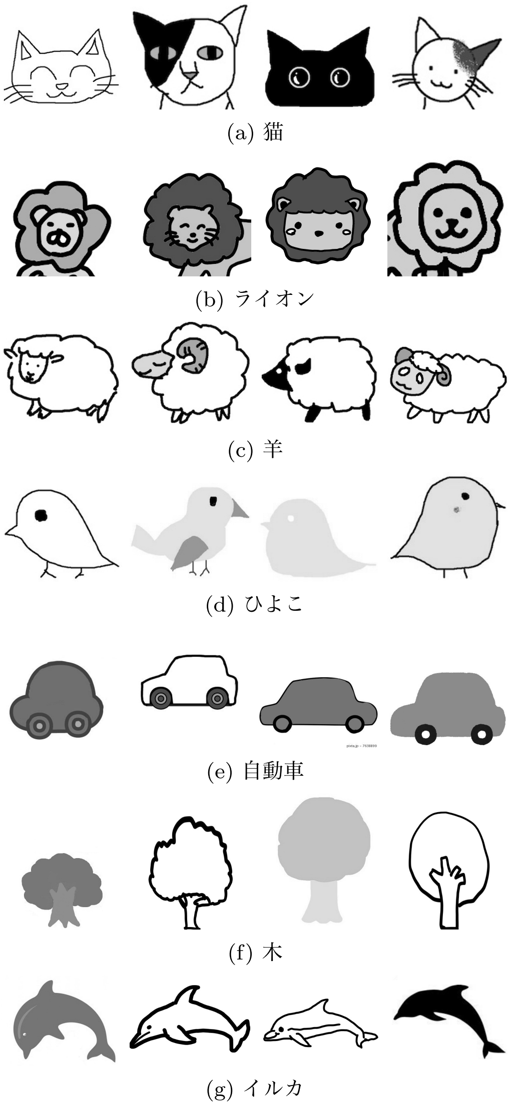


図 **4**　学習用画像一例
Fig. 4　Examples of images for learning.

どの物体のグレースケール画像を 2,000 枚以上用意する．

検出器は各対象オブジェクトに対して個別に生成する．1 つの対象オブジェクトの検出器を生成するため，その検出対象オブジェクトの学習用画像をポジティブ画像として使用して，それ以外の対象オブジェクトと非検出対象オブジェクトの画像をネガティブ画像として使用する．そして，Haar-like 特徴を用いて Adaboost で検出器を学習させる．他の対象オブジェクトに対しても同様の処理を行って，8 つの対象オブジェクトの検出器がそれぞれ生成される．

4.3　検出器を用いた手描きオブジェクトの検出

「らくがっきー」ではユーザがキャンバスに様々なオブジェクトを含む絵を描画しながら，絵に応じたサウンドを逐次対話的に生成する．本論文で提案する検出手法を用いた場合でも，従来システムと同様にユーザが描画操作を止めた直後に絵に対してオブジェクト検出処理を行って，その結果に応じてサウンドを生成する．

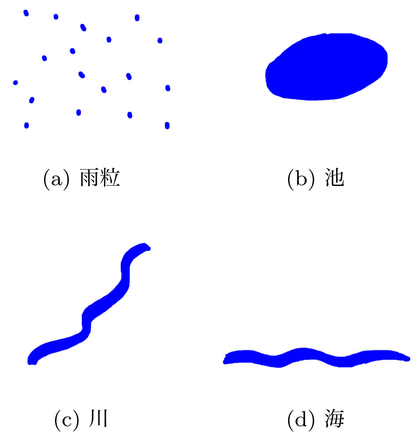


図 **5**　青色領域における形状特徴に基づいた分類
Fig. 5　Classification based on shape features of blue regions.

オブジェクトの検出のため，まずキャンバスの絵をペンの色ごとに分離して，ペン色と同数のグレースケール画像を生成して二値化する．そのため，各オブジェクトは 1 つの色のペンで描かれることを前提にしている．そして各二値化画像に対して，前節で生成した Haar-like 特徴に基づく検出器を順次適用していく．検出器は画像スケールを変えながら何度もオブジェクト検出処理を行い，検出場所を矩形で提示する．このとき対象オブジェクトが存在する場合には同じ場所で重複して検出する傾向がある．そこで，しきい値回数以上の検出があった矩形領域を最終的なオブジェクト検出領域と判定する．これらのオブジェクト検出処理は，検出器を切り替えながら順次行うことで，描かれた絵からすべての対象オブジェクトの検出を試みる．

各検出器での検出結果は他の検出器の結果に影響させていない．そのため，同じ領域で複数のオブジェクトが検出される結果になる場合もある．たとえば，ライオンを描いたときに，ライオンに加えて猫として検出される場合があり，このときにはライオンの吠える音と猫の鳴き声が合わせて生成される．

本論文で述べた検出器での処理を行ったあと，従来システムで行っていたオブジェクト検出手法を用いて，雨粒，川，池，海，山，鳥の検出を試みる．たとえば青色の領域は各領域の形状特徴である s：面積，l：長さ，c：円形度，i：傾きを使用して雨粒，池，川，海の 4 種類のオブジェクトの検出を試みる（図 **5**）．

- 雨粒：$s < s_0$ (s_0: a threshold)
- 池：$s > s_1$ and $c > c_0$ (s_1, c_0: thresholds)
- 川：$l > l_0$, $|i| > i_0$ and $c < c_1$ (l_0, i_0, c_1: thresholds)
- 海：$l > l_1$, $|i| < i_1$ and $c < c_1$ (l_1, i_1, c_1: thresholds)

本論文で新たに導入した検出器による処理，および従来システムの検出器による処理は独立して行われており，お互いに干渉しない，そのため，たとえば，猫，自動車，川，山が同時に検出されることもあり，その場合には川のせせ

らぎ音や虫の音が聞こえる中で猫の鳴き声や自動車のクラクション音が聞こえる．

本論文の手法によって改良した「らくがっきー」は，従来システムのサウンド生成機能を踏襲しつつ，オブジェクト検出精度と種類を大きく向上させることで，より多彩なサウンドを生成しやすくなっている．それにより，単なるお絵描きシステムではないことをより気付きやすくさせて，ユーザの興味を引くことを目指している．

5. 実験

5.1 検出器によるオブジェクト検出実験

提案した認識方法を検証するため予備実験を行った．実際に使用した PC は Mac OS X 10.10.5，2.6 GHz Intel Core i7，16 GB 1600 MHz DDR3 である．実装には C++ を用いており，画像処理や検出器構築のために Open CV，描画のために Open GL を使用した．

検出器による検出は 8 つのオブジェクト（猫，ライオン，羊（左右），ひよこ（左右），木，自動車，イルカ（左向き））を対象とする．そのため，Haar-like 特徴に基づく 8 つの検出器を生成する．各オブジェクトの検出器作成のためのサンプル画像の枚数は以下のとおりである．

- 猫，ライオン，羊（左右），ひよこ（左右）：ポジティブサンプル 1,100 枚，ネガティブサンプル 2,500 枚
- 自動車：ポジティブサンプル 1,100 枚，ネガティブサンプル 200 枚
- 木：ポジティブサンプル 1,100 枚，ネガティブサンプル 500 枚
- イルカ：ポジティブサンプル 1,100 枚，ネガティブサンプル 1,000 枚

また，画像サイズや検出器作成条件は以下のとおりである．

- サンプル画像サイズ：24 *24
- 各ステージにおける満たすべき最小認識率：0.995
- 各ステージにおける満たすべき最大誤認識率：0.5
- 学習ステージ数：20

初めに，検出器の検出能力を検証するため，検出対象の手描きオブジェクト各 20 枚を含む，320 枚の手描きイラストに対して 9 つの検出器を適用する実験を行った．**図 6** に検出実験結果の例を示す．色付きの矩形（赤：猫，黄緑：羊，黄色：ひよこ，オレンジ：ライオン，黒：自動車，緑：木，イルカ：水色）により各対象オブジェクトが検出されたことを示す．また，**表 1** に実験の正解数と誤検出数を示す．検出判定の閾値は検出器ごとに実験的に決定した．なお，誤検出数は検出対象でない 300 枚のオブジェクトから誤って検出した数である．

全体としては，50%以上の検出対象物体が正しく検出された．しかし，対象オブジェクトによって精度のばらつきが大きく，その中でもライオンと自動車の検出率が低かった．ライオンのサンプルイラストでは，タテガミの描き方に非常に多くの種類があったため，学習時に共通の特徴を検出することが困難であった可能性がある．また，自動車の場合にはサンプルの描き方が単調なため，学習に必要な

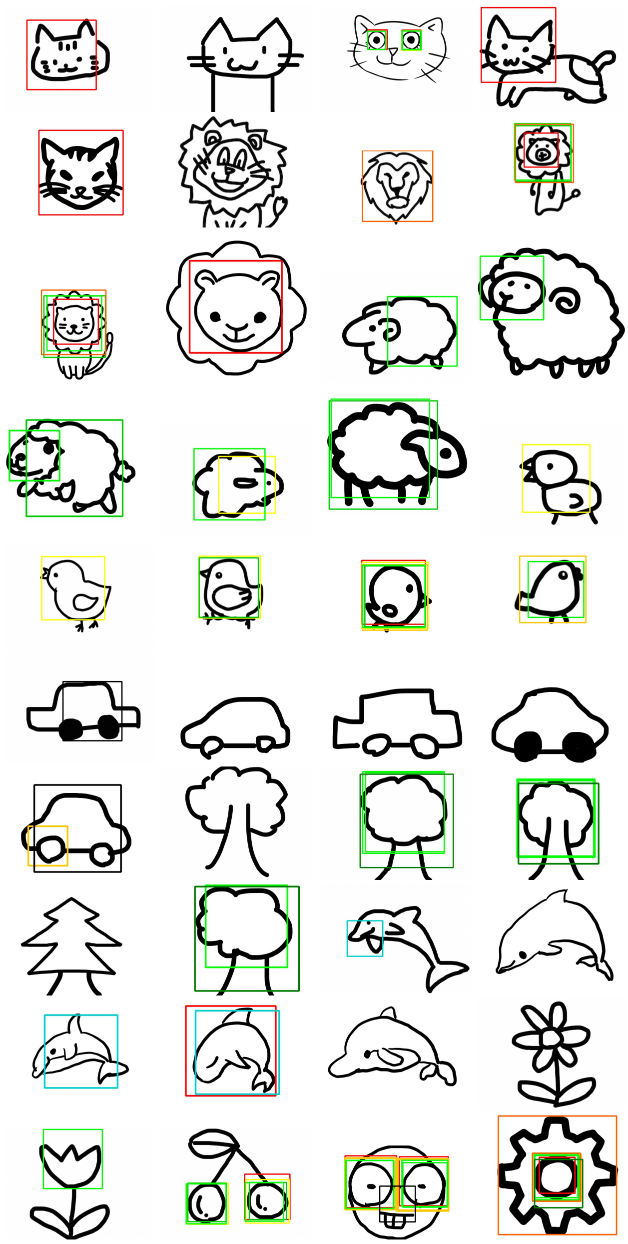


図 6　対象物体のイラスト検出例

Fig. 6　Examples of detecting target objects from illustrations.

表 1　検出実験結果

Table 1　Results of a detection experiment.

イラスト	正解数（正解率）	誤検出数
猫（20）	15（75%）	70
ライオン（20）	3（15%）	9
羊（20）	16（80%）	112
ひよこ（20）	16（80%）	90
自動車（20）	4（20%）	12
木（20）	8（40%）	20
イルカ（20）	12（60%）	8

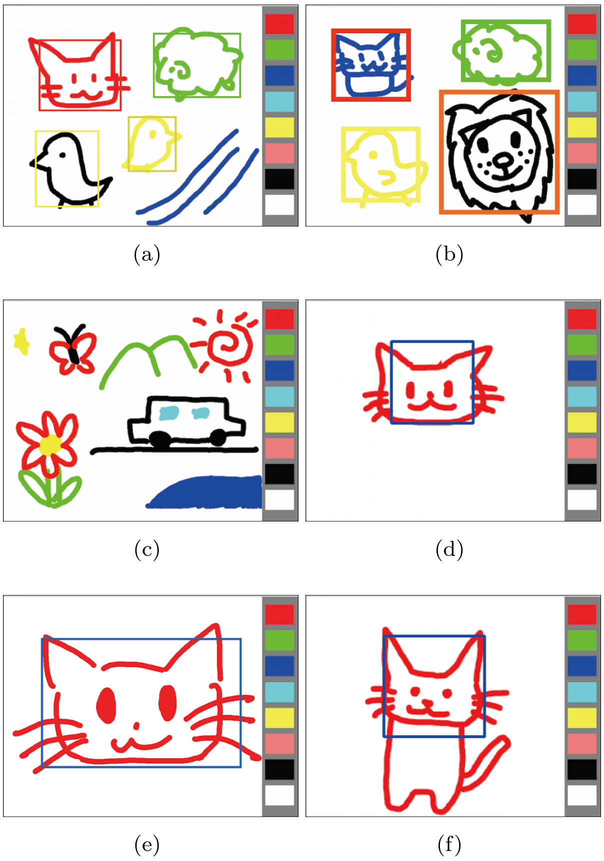


図 7　「らくがっきー」実装実験結果
Fig. 7　Experimental results of implementing "RAKUGACKY".

図 8　検出精度比較実験結果
Fig. 8　Results of comparing detection accuracy.

共通的な特徴が不十分だったと考えられる．羊は誤検出数が多かったが，これは胴体の形状を羊に共通する特徴として学習しており，木の葉が生い茂る部分やひよこ全体の形状を誤って羊として検出している例が多く見られた．

8 つの検出器の適用による対象オブジェクト検出の処理時間は 1 つのイラストに対して約 0.2〜0.8 秒であった．これは従来の「らくがっきー」とほぼ同等の処理速度であり，お絵描きによるサウンド生成には問題ないと考える．

次に，生成した検出器を「らくがっきー」のオブジェクト検出処理に組み込んだ．そして，従来システムと同様に，ユーザの絵を描く操作が止まるたびに各検出器を順次適用してオブジェクトの検出を行う．**図 7** に実装実験の結果を示す．システムが対象物体を検出するとリアルタイムで音源が生成され，本論文で提案した手法を組み込んだ「らくがっきー」でも従来システムと同様にお絵描きとサウンド生成を対話的に行うことができることを確認した．

そして，提案手法を組み込んだ「らくがっきー」の検出精度は従来システムより向上していることを確認した．たとえば図 7 (e)，(f) のイラストのように輪郭線が途切れた猫や胴体のついた猫は従来システムでは検出できなかったが，改良システムでは正しく検出することができた．オブジェクトの検出精度は対話的に絵を描いて音を発生させることを楽しむには十分な検出精度であると考える．

表 2　検出実験結果（手描き猫画像 30 枚）
Table 2　Results of a detection experiment (with 30 hand-drawn sketches of cat).

検出方法	正解数	正解率
発見的形状特徴量	3	10%
Haar-like 特徴量を用いた検出器	21	70%

5.2　従来システムとの検出精度比較実験

本論文で提案した Haar-like 特徴を用いた手法と従来システムの発見的形状特徴量による検出精度の比較実験を行った．実験では猫の絵を対象にした．20 歳から 25 歳までの 30 人の被験者に対して，見本などは見せずに「猫を正面から見た絵を描く」という指示を出した．そして得られた手描きの猫の絵 30 枚に対して検出実験を行った．

図 8 に検出実験での検出例を示す．本論文の検出器による検出結果は，赤色の矩形で示している．また，従来手法による検出結果は，青色の輪郭で示している．

表 2 に実験の正解数と正解率を示す．この実験では，本論文の手法を取り入れたシステムは従来システムより検出精度が大きく向上していることが確認できた．従来システ

図 9　体験時にユーザに示した見本 (1)
Fig. 9　Samples shown to users in the experience (1).

図 10 「らくがっきー」を体験するユーザ (1)
Fig. 10　User experiencing “RAKUGACKY” (1).

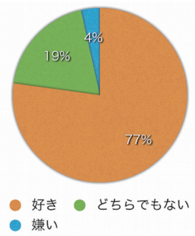
ムでは輪郭が切れたり追加されたりした場合や想定したサイズで描かれていない場合に検出がほとんど失敗する．それに対して，本論文の手法ではそれらに影響される場合は少ないため，従来システムより高い検出精度となった．

5.3　一般の人による「らくがっきー」体験実験 (1)

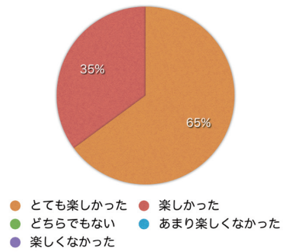
本論文で提案した手法の有効性を検証するため，一般の人による「らくがっきー」体験実験を行った．愛知工業大学オープンキャンパスにおいて 70 人の方に「らくがっきー」を体験してもらい，その体験の様子を観察しながらアンケートを実施した．実験実施の際に，システムを体験する前に見本（**図 9**）を見てもらい，色と描くものをあらかじめ説明してからシステムの体験を行った．**図 10** に体験者たちによる体験の様子を示す．体験者は「らくがっきー」による絵から音が生成される体験を楽しんでいる様子であった．**図 11** には体験者による作品例を示す．

体験後に実施したアンケートに対して 10 歳～34 歳の 57 人から回答があった．回答結果を**図 12** に示す．お絵描きが好きな体験者は 57 人中 44 人であったが，「らくがっきー」については 57 人全員が楽しかったと答えた．また，98%の体験者が普通のお絵描きよりも「らくがっきー」の方が楽しいと回答した．また，85%の体験者が描いた絵に対して適切な音が出たと回答しており，対話的に絵を描きながらその絵に関連する音を発生させる「らくがっきー」の特徴を十分に楽しんでいると考えられる．しかし，12%の体験者は思いどおりのサウンドが生成されなかったと回答した．これは，誤認識によるものと検出対象外のオブジェクトを描画したものとに原因が分かれるが，検出器の生成手法を改良してより多くのオブジェクトを高精度に検出することが必要であると思われる．

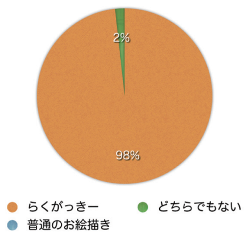
図 11　体験者が描いた作品 (1)
Fig. 11　Sketches drawn by experiences (1).



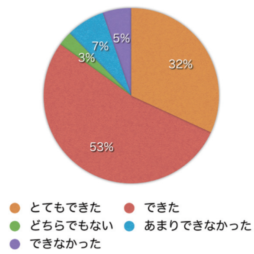
(a) お絵描きは好きですか？



(b) 「らくがっきー」は楽しかったですか？



(c) 「らくがっきー」と普通のお絵描きどっちが楽しいですか？



(d) 描いた絵から適切な音が生成されましたか？

図 12　アンケート結果
Fig. 12　Questionnaire results.

図 13 体験時にユーザに示した見本 (2)
Fig. 13 Samples shown to users in the experience (2).

図 14 「らくがっきー」を体験するユーザ (2)
Fig. 14 User experiencing “RAKUGACKY” (2).

5.4 一般の人による「らくがっきー」体験実験 (2)

次に，愛知工業大学本山キャンパスで実施された少年少女おもしろ体験教室において約 20 人の小学生に「らくがっきー」のデモ体験を実施した．実験実施の際に，システムを体験する前に見本（**図 13**）を見てもらい，色と描くものをあらかじめ説明してからシステムの体験を行った．**図 14** に体験者たちによる体験の様子を示す．体験者は「らくがっきー」による絵から音が生成される体験を楽しんでおり，様々な絵を描いてどんな音が生成されるかを何度も試していた．**図 15** には体験者による作品例を示す．音が生成されなかった場合は，絵を書き足したり，削ったりして音が生成されるように絵を書き換えながら絵から音を生成することを楽しんでいた．どのように描けば音が生成されるかを検証するユーザもいた．

5.5 一般の人による体験実験に関する考察

一般の人に対する実験では，ほとんどの体験者が「らくがっきー」を大いに楽しんでいる様子が確認できた．これは，絵を描くことでサウンドが生成されること，そして描く内容が変わることで生成されるサウンドの種類が変化することが最も大きな要因であると考えられる．「らくがっきー」は非常にシンプルなお絵描きアプリのような操作で

図 15 体験者が描いた作品 (2)
Fig. 15 Sketches drawn by experiences (2).

ありながら，サウンドを生成するという大きな特徴を持つ点が，多くの人の興味と支持が得られた．そして，表 1 で示したように現状の検出精度は必ずしも高いものではないが，描いた絵に関連するサウンドが生成されることで知的好奇心も刺激したのではないかと考えられる．

現状のシステムでは，生成されるサウンドは動物などの鳴き声か，自然など環境音の 2 つに大別され，いずれもある意味一般的なサウンドといえる．そのため，生成されるサウンドに対する反応についてオブジェクトごとに大きな違いは見られなかった．なお，子供に人気の某有名キャラクタの鳴き声を試用した際には，他のサウンドに比べて驚きと喜びの反応が大きい傾向が見られた．そのことから，個性的，独創的なオブジェクトとそのサウンド生成を取り扱う場合には，高い精度のオブジェクト検出性能が要求されると思われる．

また，今回の実験では 1 人あたりの体験時間は 5～15 分間程度であり，比較的短いものであった．「らくがっきー」をより長い時間で継続的に使用してもらうためには，検出精度の向上や検出できるオブジェクトの種類の拡張は重要であると考える．

6. まとめ

本研究では，インタラクティブメディアシステム「らくがっきー」の手描き絵の検出精度の向上によるシステムの改善を行った．提案手法により，猫，ライオン，羊（左右），ひよこ（左右），自動車，木，イルカの 9 種類のオブジェクト検出を目的として，Haar-like 特徴を用いた検出器を「らくがっきー」へ適用した．実験では猫，ライオン，羊，ひよ

こ，自動車，木，イルカの検出器を生成して，各検出器の検出率は 15〜80%，全体では 50%以上の検出率となった．一般の人に対する実験では大部分の人が「らくがっきー」を楽しんでおり，本論文で提案した手法はシステム使用に関しては必要なオブジェクト検出精度を持つことが確認できた．また，体験者は，自分が描いた絵から音が生成されることを楽しんでいた．描かれた絵に対して適切な音が生成されること以外にも描いている途中の部分的な絵からも音が生成されると驚き，さらに描き進める様子が見られた．ただし，本システムをより多くの人に継続的に使用してもらうためには，より精度の高い検出が必要であると考える．

今後の課題としては，検出対象のオブジェクトの種類が十分ではないと考えられるため，現在は検出対象となっていないオブジェクトを検出対象として取り入れる拡張を行う必要がある．また，ライオンや自動車を対象とした検出器などは精度が不十分なため，新たな特徴量の使用や手法の検討が必要である．たとえば文献 [8] では Deep Learning を適用して手描きスケッチの認識を行って 70%程度の検出精度であったと報告しており，「らくがっきー」に対しても処理時間なども考慮しながら Deep Learning などの適用を検討している．

謝辞 本研究の一部は科研費基盤研究（C）（26330420）による．

参考文献

[1] Raffle, H., Vaucelle, C., Wang, R. and Ishii, H.: Jabberstamp: Embedding sound and voice in traditional drawings, *Proc. IDC 2007*, pp.137–144 (2007).

[2] Levin, G. and Lieberman, Z.: Sounds from Shapes: Audiovisual Performance with Hand Silhouette Contours in The Manual Input Sessions, *Proc. NIME 2005* (2005).

[3] Ichino, J., Pon, A., Sharlin, E., Eagle, D. and Vuzik, S.C.: Creative Music Expression for Children thorough Whole Body Interaction, *J. Information Precessing Society of Japan*, Vol.53, No.12, pp.2773–2786 (2012).

[4] teamLab：スケッチピストン，入手先 ⟨http://www.team-lab.com/sketchplayingmusic.⟩

[5] Goto, S., Kondo, N. and Mizuno, S.: RAKUGACKY: Making sounds with drawing, *Proc. ACM SIGGRAPH 2013 Posters* (2013).

[6] 近藤菜々子，水野慎士：スケッチブックでのお絵描きを三次元 CG で拡張する映像ツールの提案とその実現方法，情報処理学会論文誌・デジタルコンテンツ，Vol.1, No,1, pp.1–9 (2013).

[7] 深野元太郎，中村嘉彦，滝沢穂高，山本眞司，松本 徹，舘野之男，飯沼 武：“Eigen Nodule”：部分空間法を用いた胸部 X 線 CT 画像からの肺結節認識，電子情報通信学会技術研究報告，MI，医用画像，Vol.103, No.319, pp.15–20 (2003).

[8] 山川まどか，関口香菜，佐々木一磨，尾形哲也：Convolutional Neural Network による写真と手描きスケッチの認識，第 30 回人工知能学会全国大会論文集，4L1-4 (2016).

[9] Google: Quick, Draw!, available from ⟨https://quickdraw.withgoogle.com/⟩

[10] Viola, P. and Jones, M.J.: Rapid Object Detection using a Boosted Cascade of Simple Features, *Proc. IEEE CVPR 2001*, Vol.1, pp.511–518 (2001).

[11] Lienhart, R. and Maydt, J.: An Extended Set of Haar-like Features for Rapid Object Detection, *Proc. IEEE ICIP 2002*, Vol.1, pp.900–903 (2002).

[12] 草野孝幸，出口大輔，井出一郎，村瀬 洋：猫パーツの抽出とその組み合わせによる猫の顔検出の高精度化，動的画像処理実利用化ワークショップ（DIA2014）公演論文集，pp.137–142 (2014).

[13] 島田真衣，馬場哲晃，串山久美子：検出器を用いた手描き顔検出システムの提案，情報処理学会インタラクション 2013 論文集，pp.768–769 (2013).

加藤 里美 （正会員）

2015 年愛知工業大学情報科学部卒業．2017 年愛知工業大学大学院経営情報科学研究科博士前期課程修了．在学中はコンピュータグラフィックス，画像処理，機械学習に関する研究に従事．

水野 慎士 （正会員）

1998 年名古屋大学大学院工学研究科博士後期課程修了．博士（工学）．1999 年豊橋技術科学大学情報処理センター助手，2009 年愛知工業大学情報科学部講師，2010 年同准教授を経て，2014 年同教授，現在に至る．コンピュータグラフィックス，画像処理，マルチメディア等に関する技術の開発やそれらを応用した教育用やエンタテイメント用のインタラクティブデジタルコンテンツに関する研究に従事．2017 年山下記念研究賞受賞．画像電子学会，芸術科学会，日本バーチャルリアリティ学会各会員．