Vol.2016-CG-165 No.22
Vol.2016-DCC-14 No.22
Vol.2016-CVIM-204 No.22
2016/11/10

お絵かきサウンドシステム「らくがっきー」の改良

加藤 里美[1,a)]　水野 慎士[1]

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，本稿ではオブジェクトの検出に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識別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よ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．

Improvement of an Interactive Media System “RAKUGACKY”

Satomi Kato[1,a)]　Shinji Mizuno[1]

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.

1. はじめに

コンピュータ技術の発展に伴い，デジタル技術を用いて映像やサウンドを提示，人の操作などに対して対話的に反応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幾つも研究されている．ユーザが実際に描いた絵を元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元に作曲の支援をするアプリケーション [3] 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．これらは様々な場所で活用され，デジタルサイネージやエンターテイメント，教育などに役立ち今後の進展も大きく期待されるものである．

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 [4]．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．システムはユーザの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伴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

[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
[a)] b15709bb@aitech.ac.jp

Vol.2016-CG-165 No.22
Vol.2016-DCC-14 No.22
Vol.2016-CVIM-204 No.22
2016/11/10

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．

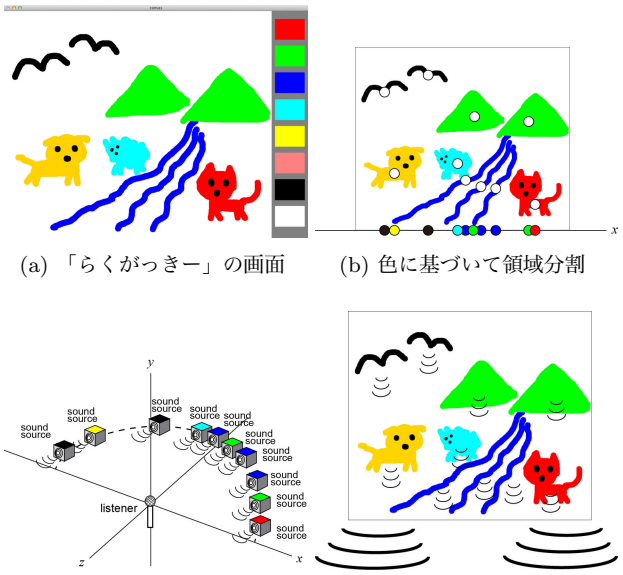
そこで「らくがっきー」の描かれた絵に対する新たな検出手法を提案する．本論文では特にデフォルメされた動物イラストに含まれる顔や体の特徴的な部位などの描画に着目して，機械学習に基づく手法で大量の動物のイラスト画像から抽出した共通的な特徴量を選択して，オブジェクトの識別に用いる手法を提案する．提案手法を適用して改良した「らくがっきー」は，従来システムよりも手描きの絵から適切なサウンドが生成されることが期待できる．

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．この時，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を生成を楽しむことができる．

生成されるサウンドは描いたオブジェクトによって変化させる必要がある．キャンバス上で絵を描くと描画時の色に基づいて幾つかの領域に分割される．描く領域の形状は面積，長さ，円形度，傾斜，湾曲など複数の形状特徴量を計算することによって分析される．分析対象の絵として子供たちが描きそうな猫，ひよこ，山，池，川など 10 種類の絵を対象とした．各対象の音源はモノラルの wav ファイルをあらかじめシステムに用意してある．各動物の鳴き声，虫のさえずり，川のせせらぎ音などがある．システムはその色と形状特徴量に基づいて各領域を分類する．例えば，水に関連する青色領域は，描く領域の面積，長さ，円形度，傾きを用いて雨粒，池，川，海など 4 種類に分類される．赤色の領域で一定の面積と円形度を持っている絵は猫に分類される．黄色の領域で一定の面積と円形度を持つと犬やひよこに分類される．

対象の絵の領域を分類した後に，各領域の位置に基づいて音源を空間内に配置する．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，それぞれの位置で音を生成し，合成することでステレオサウンドを生成する．描画中も音は発生し続け，ユーザが追加オブジェクトを描画するたびに音源が加えられて変更される．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 2 に「らくがっきー」での描画の様子を示す．この絵には猫（赤），山（緑），川（青）の 3 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の鳴き声，鈴虫の鳴き声，川のせせらぎ音の音源が配置されて全ての音源を合わせたサウンドが生成されている．



(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

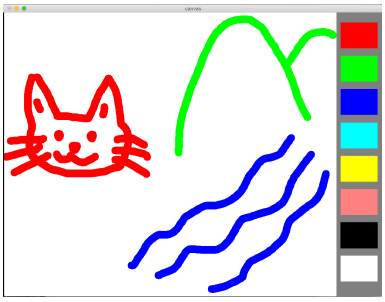


図 2: 「らくがっきー」

3. 「らくがっきー」の改良

3.1 従来手法の問題点

前述の通り「らくがっきー」では描いたオブジェクトを解析することで配置する音源の種類を選択する．オブジェクトの解析はまず色の分類を行い，次に形状解析を行う．このとき，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る．そのため，必ずしも描いたオブジェクトに適切なサウンドが生成されるとは限らない．例えば赤色で猫の顔を描くと猫の鳴き声が生成されるが，識別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 に示したようなひよこの絵を描いた場合でも猫の識別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．

3.2 絵の認識手法の改良

前節で述べた「らくがっきー」の問題点を解決するためには，単純な形状特徴量や発見的特徴量に加えて，手描き

Vol.2016-CG-165 No.22
Vol.2016-DCC-14 No.22
Vol.2016-CVIM-204 No.22
2016/11/10

図 3: 猫の鳴き声が生成されるひよこ

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いながらオブジェクトを識別する必要がある．

オブジェクトの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挙げられる．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他，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．例えば，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全ての画素値を用い，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 [5]．また，顔認識など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識別機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 [6][7]．そして，Haar-like 特徴を用いて動物の顔の検出を行った研究 [8] や手描き顔画像の認識を行った研究 [9] も報告されている．

そこで，本研究では「らくがっきー」で描かれたオブジェクトの識別に Haar-like 特徴を用いることとする．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ることで，「らくがっきー」で描かれたオブジェクトを識別することを試みる．

3.3 検出器の生成

本研究では，子供が描いた絵によく見られるオブジェクトのうち，猫，ライオン，羊，ひよこを提案手法で検出する対象オブジェクトとする．

まず，各対象オブジェクトの学習用画像をそれぞれ 1000 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，非対象オブジェクトである人の顔や動物のイラスト，ボールなどの物体のグレースケール画像を 2000 枚以上用意する．

検出器は各対象オブジェクトに個別に生成する．一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，6 つの対象オブジェクトの検出器がそれぞれ生成される．

3.4 「らくがっきー」での手描きオブジェクトの識別

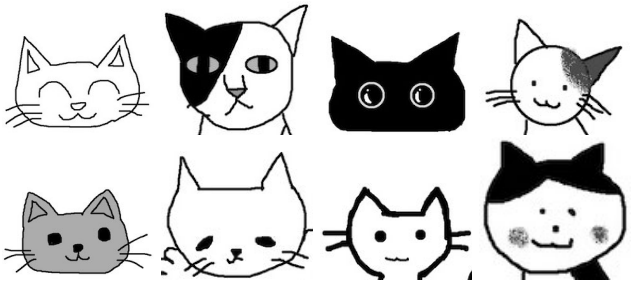
「らくがっきー」ではユーザがキャンバスに様々なオブジェクトを含む絵を描画しながら，絵に応じたサウンドを逐次対話的に生成する．本研究で提案する検出手法を用いた場合でも，従来システムと同様にユーザが描画操作を止めた直後に絵に対してオブジェクト検出処理を行って，その結果に応じてサウンドを生成する．

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，オブジェクト付近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を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．

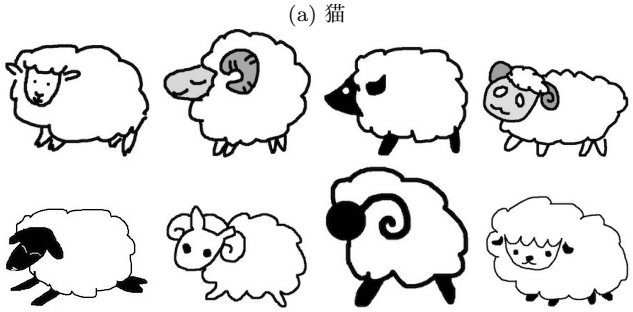
なお，本論文で述べた検出器での処理を行ったあと，従来システムで行っていたオブジェクト検出手法を用いて，雨粒，川，池，海，山の検出を試みる．例えば青色の領域は各領域の形状特徴である s：面積，l：長さ，c：円形度，i：傾きを使用して雨粒，池，川，海の 4 種類のオブジェクトに分類する (図 5)．

- rain drop: $s < s_0$ (s_0: a threshold)
- pond: $s > s_1$ and $c > c_0$ (s_1, c_0: thresholds)
- river: $l > l_0$, $|i| > i_0$ and $c < c_1$ (l_0, i_0, c_1: thresholds)
- sea: $l > l_1$, $|i| < i_1$ and $c < c_1$ (l_1, i_1, c_1: thresholds)

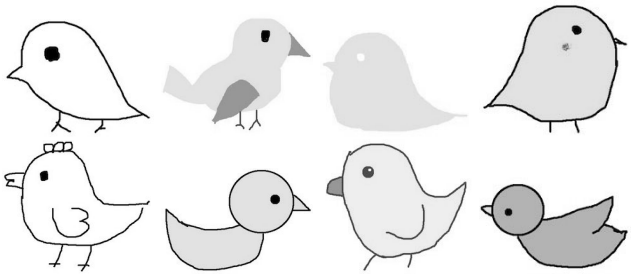
Vol.2016-CG-165 No.22
Vol.2016-DCC-14 No.22
Vol.2016-CVIM-204 No.22
2016/11/10



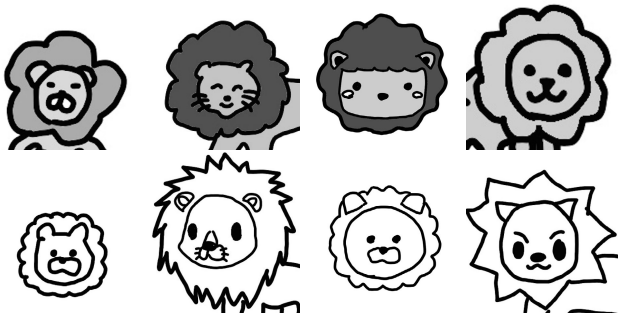
(a) 猫



(b) 羊

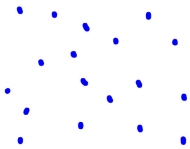


(c) ひよこ



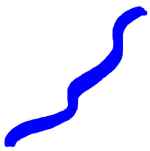
(d) ライオン

図 4: 学習用画像一例



(a) 雨粒

(b) 池



(c) 川

(d) 海

図 5: 青色領域における形状特徴に基づいた分類

4. 実験

4.1 検出器による識別実験

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 10.10.5，2.6GHz Intel Core i7，16GB 1600MHz 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．

検出は 6 つのオブジェクト（猫，ライオン，羊 (左右)，ひよこ (左右)）を対象とする．そのため，Haar-like 特徴に基づく 6 つの検出器を生成する．各オブジェクトの検出器を生成するため，ポジティブサンプルとして対象物体のサンプルイラストを 1100 枚，ネガティブサンプルとして非対象物体のサンプルイラストを 2500 枚を各検出器に定義した．羊とひよこは向きに対応した検出器の作成のため左向き用と右向き用を用意した．そして，生成した検出器を「らくがっきー」に組み込む．「らくがっきー」ではユーザが絵を描いている間にペン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．このとき　検出判定の閾値は検出器毎に実験的に決定した．

初めに，検出器の識別能力を検証するため，検出対象物体の手描きイラストを 20 枚を含む，260 枚の手描きイラスト対し検出器による検出実験を実施した．図 6 に検出実験結果の例を示す．色付きの矩形 (赤：猫，緑：羊，黄色：ひよこ，オレンジ：ライオン) により各対象物体の位置を検出し示す．対象物体検出の処理時間は約 0.2 秒であった．表 1 に実験結果を示す．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の検出率は低かった．図 7(b) に示すように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．

次に「らくがっきー」に検出器を実装し，実装実験を行った．図 7 に結果を示す．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」は従来システムと同様に対話的に描画を楽しむことができる上，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．例えば図 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ている．

Vol.2016-CG-165 No.22
Vol.2016-DCC-14 No.22
Vol.2016-CVIM-204 No.22
2016/11/10

4.2 一般の人による「らくがっきー」体験実験

提案した手法による「らくがっきー」を愛知工業大学オープンキャンパスにて 70 名の方に体験してもらい，その体験の様子を観察しながらアンケートを実施した．図 8 に体験者たちによる体験の様子を示す．多くの人はお絵描きが好きであり「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だ．図 9 には体験者による作品例を示す．

また，体験後に実施したアンケートの回答結果を図 10 に示す．10 歳〜34 歳の方 57 名にアンケートに回答していただいた．アンケートの結果を見ると普通のお絵描きよりも「らくがっきー」の方が楽しいという回答結果となった．また，描いた絵に対して適切な音が出たという方が 85 % となっており対話的に絵を描いて音を発生させて楽しめていると考えられる．しかし，15 % の人からは思い通りのサウンドが生成されなかったという回答があった．これについては，検出器の生成手法を改良しながら，より多くのオブジェクトを識別対象にする必要があると思われる．

5. 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) の 6 種類の動物の検出を目的として，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，ひよこの検出器を生成し，各検出について有効な結果が得られた．

今後の課題としては，検出対象のオブジェクトが十分ではないと考えられるため，その他の子供が絵描きそうなオブジェクト対象とした追加作成を行う必要がある．また，特徴量として色を用いることで検出精度を向上することを考えている．さらに deeplearning などの新たな識別手法の適応も検討している．

なお，本研究の一部は科研費基盤研究 (C)(26330420) による．

表 1: 検出実験結果

Illustration	True positive	False positive
cat (20)	15 (75%)	54
sheep (20)	14 (70%)	105
chick (20)	10 (50%)	32
lion (20)	3 (15%)	7

参考文献

[1] H. Raffle, C. Vaucelle, R. Wang, H. Ishii, Jabberstamp: Embedding sound and voice in traditional drawings, Proc. of IDC 2007, pp.137–144, 2007.
[2] G. Levin, Z. Lieberman, Sounds from Shapes: Audiovisual Performance with Hand Silhouette Contours in The Manual Input Sessions Proc. of NIME 2005, 2005.
[3] J. Ichino, A. Pon, E. Sharlin, D. Eagle, S. Carpendale Vuzik: Creative Music Expression for Children thorough Whole Body Interaction, J. of Information Precessing Society of Japan, Vol. 53, No. 12, pp. 2773–2786, 2012.
[4] S. Goto, N. Kondo, S. Mizuno, RAKUGACKY: making sounds with drawing, Proc. of ACM SIGGRAPH 2013 Posters, 2013.
[5] 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, 103(319), pp. 15-20, 2003.
[6] P. Viola, M. J. Jones, Rapid Object Detection using a Boosted Cascade of Simple Features, Proc. of IEEE CVPR 2001, Vol. 1, pp. 511–518, 2001.
[7] R. Lienhart, J. Maydt, An Extended Set of Haar-like Features for Rapid Object Detection, Proc. of IEEE ICIP 2002, Vol. 1, pp. 900–903, 2002.
[8] 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.
[9] 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.

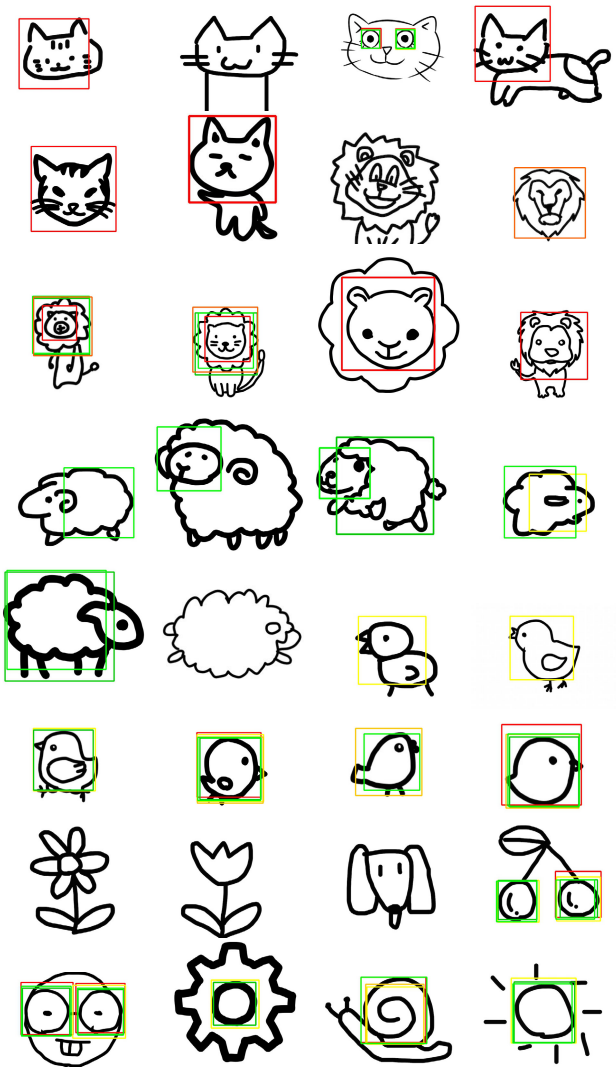
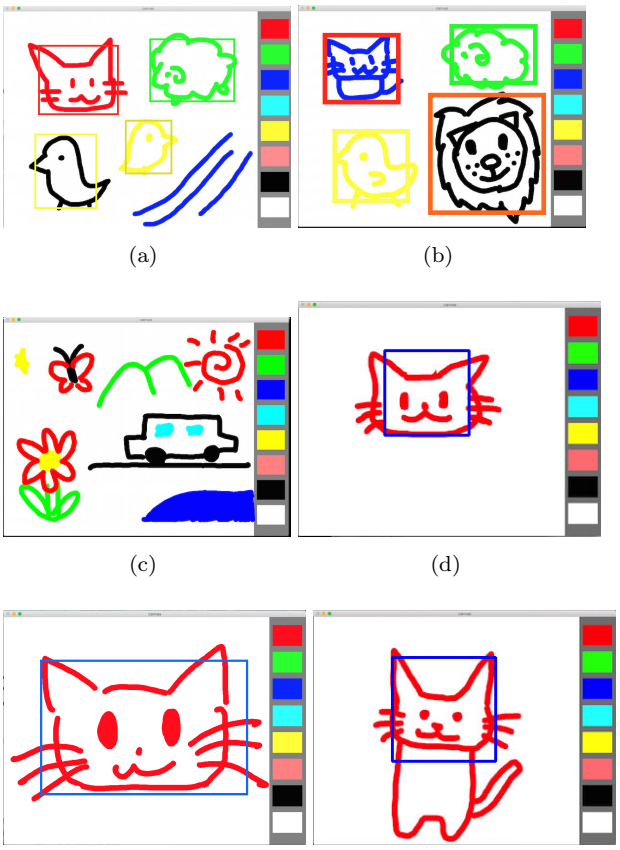


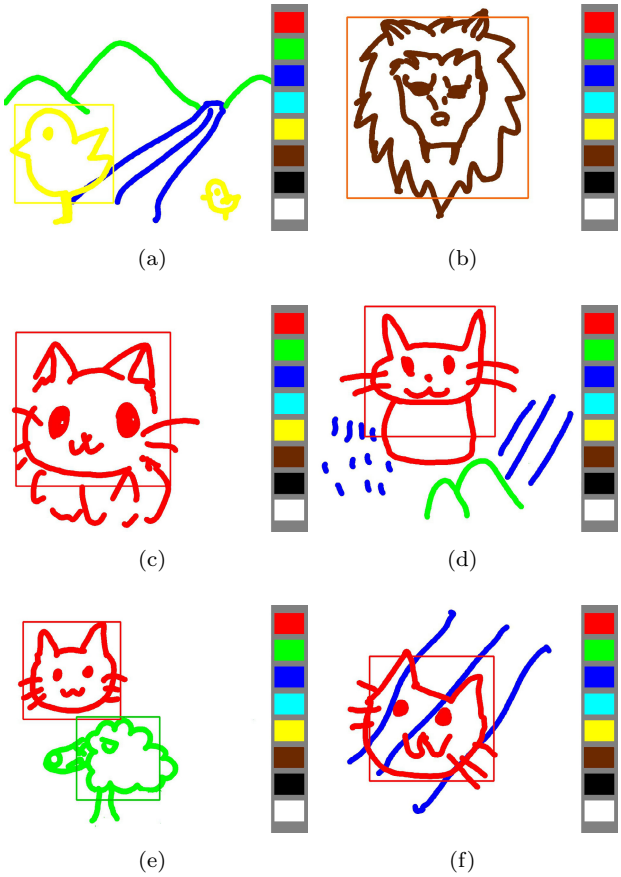
図 6: 対象物体のイラスト検出例

Vol.2016-CG-165 No.22
Vol.2016-DCC-14 No.22
Vol.2016-CVIM-204 No.22
2016/11/10



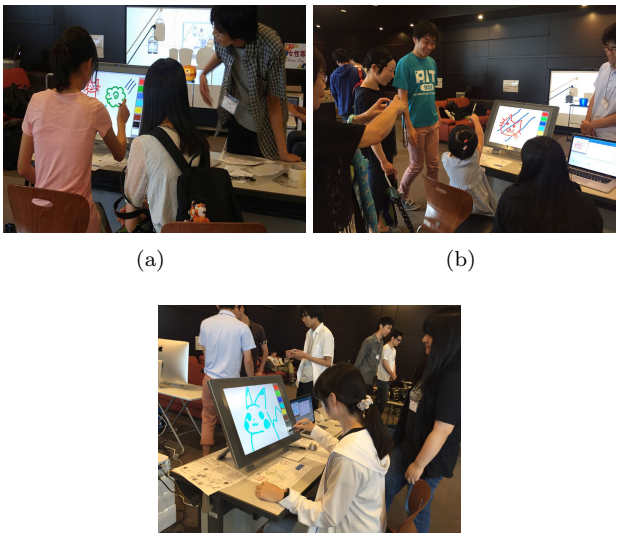
(a) (b) (c) (d) (e) (f)

図 7: 「らくがっきー」実装実験結果



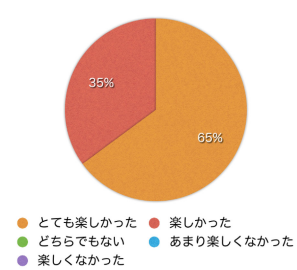
(a) (b) (c) (d) (e) (f)

図 9: 体験者に描いてもらった作品

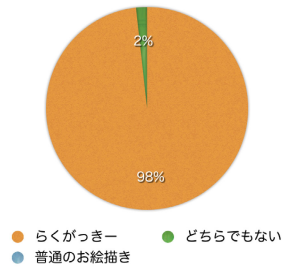


(a) (b) (c)

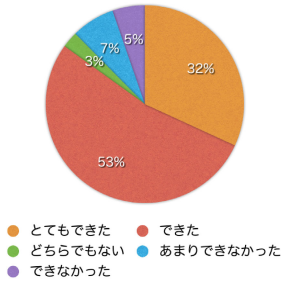
図 8: 「らくがっきー」体験時の様子



(a) 「らくがっきー」は楽しかったですか？



(b) 「らくがっきー」と普通のお絵描きどっちが楽しいですか？



(c) 描いた絵から適切な音が生成されましたか？

図 10: アンケート結果