

DPM及びHOG特徴量による犬猫顔検出手法の性能評価

Performance evaluation of dog/cat face detection using DPM and HOG features

早稲田大学大学院 国際情報通信研究科

4213A334

張 海蒙

Graduate School of Global Information and Telecommunication Studies

Waseda University

4213A334

Heimeng Zhang

2015 年 7 月

July 2015

目次

第1章 序論

1. 1	研究の背景	3
1. 2	ペット顔認識の現状	3
1. 3	研究の目的	5

第2章 画像特徴量

2. 1	HOG特徴量	6
2. 2	LBP特徴量	7
2. 3	Deformable Part Model	8

第3章 犬猫顔検出実験

3. 1	評価実験	10
3. 2	検出結果	11
3. 3	考察	14

第4章 結論

参考文献

付録 A DPM による犬顔画像検出結果

付録 B HOG による犬顔画像検出結果

付録 C DPM による猫顔画像検出結果

付録 D HOG による猫顔画像検出結果

第1章 序論

1. 1 研究の背景

近年の犯罪者やテロリストの増加に伴い、バイオメトリクス個人認証技術に関する必要性は年々高まっている。顔認証技術は、被撮影者が認証動作を必要としない点や、履歴として残された画像から本人特定が容易にできる点など、他のバイオメトリクス個人認証技術にはない優位性がある。例えば、港などで群集の中から犯罪者を見分けることができる。このような利点は、指紋、虹彩、音声などによる認証では得られていない。

顔認証システムは、1960年代に主に軍事・セキュリティ目的で研究が始まった[1]。日本では、1973年当時京都大学に在籍した金出教授の研究から始まる。1990年代には、顔認証システムは入退出ゲート・セキュリティカメラなどで実用化された。コンシューマ向けでは、2005年2月にNikonがIdentics社の顔検出技術をデジタルカメラに初採用している。そして、翌年キヤノン、富士フィルム、ペンタックスが顔検出機能を搭載したデジタルカメラを発売している。以後、各社デジタルカメラ・携帯電話に採用されて身近なものとなっている。

現在では、デジタルカメラの普及率は全人口の8割を超えており、デジタルカメラ所有者の約2割がペットを被写体として撮影している[2]。デジタルカメラの使用目的を調べると、景色、子供、ペット、動物などが撮影対象となっている。さらに、ペット監視カメラも登場し、我々の身の周りではペットをデジタルカメラで撮影する機会が増加している。このことから、動物を認識するための映像処理技術に対する期待が大きくなっている。

1. 2 ペット顔認識の現状

猫や犬は古くから人間の愛玩動物として扱われてきて、研究の対象としても興味を持たれている。人以外の動物では、近年愛犬、愛猫の顔を認識するためのデジタルカメラが開発されている。例えば、猫を介して地域のコミュニケーションを活発化させようとする研究がある[3]。猫には外猫という猫特有の飼育形態があり、猫のテリトリー内の複数の家で飼育されることがある。この研究ではこの飼育形態を利用し、猫に記録メディアやQRコードを付与することで、地域のコミュニケーションを活発化させている。

一方、2010 年 2 月に富士フィルムが犬や猫を認識して、正面を向いた時に自動で撮影する機能を備えたデジタルカメラを発売した。また、ペンタックスが登録した犬や猫の顔を認識する製品を発売した[4]。さらに具体的な事例としては、「Bistro」という世界初、猫顔認識機能付きの自動餌やり器がある[5]。

Bistro は複数の猫の顔を認識し、それぞれに合った食事を提供、さらに体重測定まで行う機能を持ち、さまざまなメディアに取り上げられている。各パーツの詳細を図 1 に示す。縦に伸びているパーツの上部(food hopper)がキャットフードを入れておくスペースで、「猫がどれだけエサを食べているのか」を検出し自動でエサを追加する。その下にはエサを置いておくトレイ(feeding tray with digital scale)がある。これは重量の計測が可能で、猫がどれくらいの量を食べたかを飼主に知らせることもできる。エサ置きトレイの横には飲み水タンク(water tank with digital scale)がある。これにも重量計が付いておりどれくらいの水を飲んでいるか計測可能である。またフィルターがついているのでいつでもきれいな水が飲めるようになっている。



図 1： 顔認識機能付きの自動餌やり器(文献[5]より引用)

これは猫顔画像認識機能とカメラが搭載されているからで、装置がエサを食べに来た猫の顔を把握して顔の特徴から（例を図 2 に示す），“このネコは「オレオ」”といった具合にどの猫が現在エサを食べているか、体重がどれくらいかなどを正確に計測できるようになっている。



図 2：猫認識（文献[5]より引用）

ところで犬の顔画像認識については、犬種により目鼻口などの顔パーツの形状がかなり違うことから、Bistro では犬顔画像認識機能は採用されなかった。現状では、犬顔認識技術は発展途上にあると考えられる。

1. 3 研究の目的

従来、画像から人物を検出や顔認識するなど技術については、活発的に研究が行われているが、犬や猫などのような特定の動物を検出する技術に関する研究はあまり行われていない。人間と比較して犬、猫は個体によって顔、耳の形状や模様が様々であり、向きの変化に伴って、様々な要因が同時発生して見えの変動が大きいので、検出が困難である。

猫顔検出に関する研究として、名古屋大学情報科学研究科の草野らは局所特徴を用いた猫の顔を検出する手法を提案している[6]。草野らは、Haar-like 特徴量と HOG 特徴量を用いた猫の顔検出比較実験を行っている。実験結果によると、猫顔の検出成功率はあまり高くなく、Haar-like 特徴量を用いた場合は適合率と再現率それぞれは 0.17 と 0.42 である。Haar-like 特徴量よりも HOG 特徴量の方が比較的認識結果は良く、適合率と再現率はそれぞれ 0.35 と 0.75 となっている。これにより、猫に対しては HOG 特徴が有効であることがわかるが、HOG 特徴量を用いて捉えるエッジ情報の性質から、顔の向きに対する安定性が弱いという問題点がある。

一方、犬の顔検出に関する研究については、ほとんど有効な手法が提案されていない。McGill 大学コンピュータサイエンス学科の博士課程学生 Chu Wang が Deformable Part Model を用いて、犬か猫かの識別実験結果を報告しているのみである[7]。

本研究では、より高い検出成功率を得るために、HOG 特徴量による手法に加えて、HOG 特徴量と DPM を組み合わせた手法を用いる。これにより、猫及び犬顔画像のそれぞれの特性に合った手法となることが期待できる。実験により検出精度について検討する。

第2章 画像特徴量

2. 1 HOG特徴量

Histograms of Oriented Gradients 特徴量は、画像の局所領域から輝度勾配を求め、さらに区分して勾配方向のヒストグラムを取り出す特徴量である。輝度差が捉えられなくても顔の形状を解析の手掛かりとして利用できると考える。また、HOG 特徴量は幾何学的変換に強く対応できる。

輝度の勾配方向は式(1)で算される。

$$\theta(x, y) = \tan^{-1} \frac{fy(x, y)}{fx(x, y)} \quad (1)$$

$$\begin{cases} fx(x, y) = L(x + 1, y) - L(x - 1, y) \\ fy(x, y) = L(x, y + 1) - L(x, y - 1) \end{cases} \quad (2)$$

勾配強度は式(2)で定義される。

$$m(x, y) = \sqrt{fx(x, y)^2 + fy(x, y)^2} \quad (3)$$

ヒストグラムは図3に示すように作成できる。

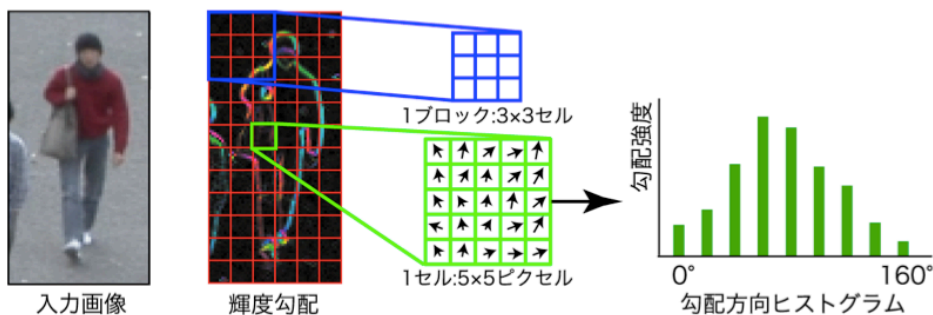


図3：HOG 特徴のヒストグラムの例(文献[9]より引用)

入力画像をグレースケール画像に変換し、色によるバリエーションを除去する。画像を5ピクセルの正方形領域に分割する。一つのセル領域には25ピクセル分のエッジベクトルが含まれている。各セル領域に含まれるピクセルのエッジ方向、エッジ強度から、ヒストグラムを作成する。0° から 180° まで、20° 毎の9次元ベクトルとして表される。

3セル×3セルの領域をブロック領域と呼ぶ。このブロック領域には9個のヒストグラムが存在し、この9個のヒストグラムを用いて各セルのヒストグラムを正規化する。正規化を行ったブロック領域を1セルずつスライドさせ、各セルは複数回正規化が行われる。

最終的なHOGの次元数 V の算出は、以下の式(4)で表される。

$$V = (w \div c - b + 1) \times (h \div c - b + 1) \times b^2 \times v \quad (4)$$

ここで、 v はヒストグラムの次元数、 w 、 h は画像の横幅と縦幅、 c はセルサイズ、 b はブロックサイズである。

例：250×250の入力画像の場合、1セル：5×5ピクセル、1ブロック：3×3セル、 $V=(4 \times 10) \times (3 \times 3) \times 9=3240$ となる。

2. 2 LBP特徴量

Local Binary Pattern 特徴量は、T・Ojalaによって提案された画像の認識や分類に用いられる特徴量の一つである。通常3×3の画素領域で計算され、画像の局所的な特徴を抽出できる。特に、画像の照明変化の影響を受けにくい。そして、数値の大小比較と2進数から10進数への変換が主な計算で、高速に特徴計算ができる。

LBPの計算には3×3画素領域内の画素値が使われ、入力画像の階調数に関わらずLBP画像は256階調(0-255の数値)の画像になる。まずは、256階調グレースケール画像へ正規化される。次に、画像を複数のブロックに分割し、ブロック内の画素に対して計算する。LBP特徴量は論理的に注目画素の周囲の半径と近傍領域の画素数に制限はないが、特徴量の次元数が爆発するため、比較実験で半径は1、画素数は8に設定する。

例えば、注目点の画素値 F_r とその8近傍にある画素値 F_n を比較してLBP特徴量を計算する。

$$f(x) = \begin{cases} 0, & Fn < Fr \\ 1, & Fn \geq Fr \end{cases}$$

注目画素と近傍領域

72	64	115
88	84	130
80	80	89



大小比較

0	0	1
1		1
0	0	1

LBP パターン : 00111001 (二進法)

LBP バリュー : 140 (十進法)

2. 3 Deformable Part Model

Deformable Part Model は、一つの物体を各パーツから構成されるツリー構造として捉え、その各パーツの位置が隠れ変数となっているモデルである[8]。従来の手法では、検出対象のパーツの位置が固定されており、人や動物といった物体を検出する際に姿勢が大きく変化した場合に対応できないという問題があった。DPMは、画像によって姿勢が大きく変化する対象物に対しても高いパフォーマンスを示すことが可能である。この方式は2006年のPASCAL Person Detection Challengeにおいて、従来手法の2倍の平均精度を成し遂げ、20のカテゴリ中10の部門で最高の結果を出している[9]。

検出モデルは、検出対象全体を捉えるルートフィルタと、検出対象が持ついくつかのパートフィルタから構成される。図4はDeformable Part Modelを用いた人検出の例である。

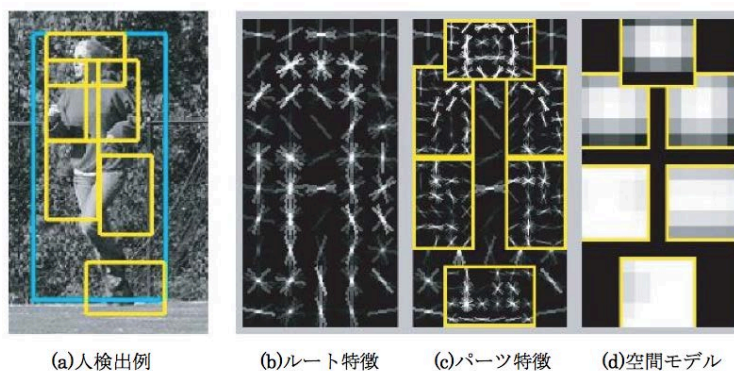


図 4 : Deformable Part Model の人検出モデル(文献[10]より引用)

Deformable Part Model のスコア算出は、まず、入力画像から HOG 特徴量マップを求め、ルートフィルタとの内積を求める。次に、入力画像から 2 倍の解像度での HOG 特徴量マップを求めて、各パートフィルタについて内積の合計が最も大きくなるようにマッチングを行なう。更に、各パートフィルタについてルートフィルタに対する位置のずれを計算し、スコアから減算する。最後にルートフィルタと各パートフィルタの合計スコアが閾値以上になった箇所を対象物体として検出する。図 5 に Deformable Part Model により物体検出を行うための処理全体を示す。

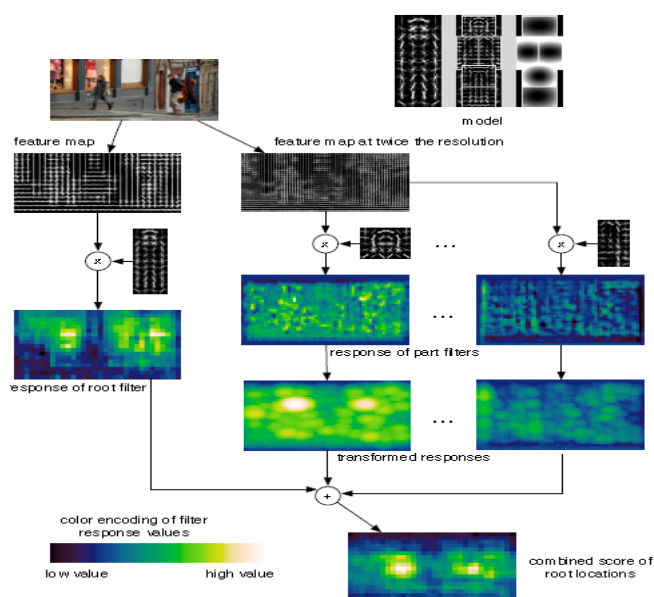


図 5 : スコア算出(文献[11]より引用)

第3章 犬猫顔検出

3.1 評価実験

実験では、犬及び猫の顔検出に対して、HOG 特徴量及び Deformable Part Model の2種類の識別器を用いて検出精度を比較した。データセットには The Oxford-IIIT-Pet dataset を使用する。学習画像は猫画像の枚数と犬画像の枚数を同じにして、ポジティブ学習画像 1188 枚、ネガティブ学習画像 3578 枚として学習を行う。

Deformable Part Model によって生成された顔領域の検出モデルを図 6 に示す。また、既知画像及び未知画像について検出を行った結果の例を図 7 に示す。

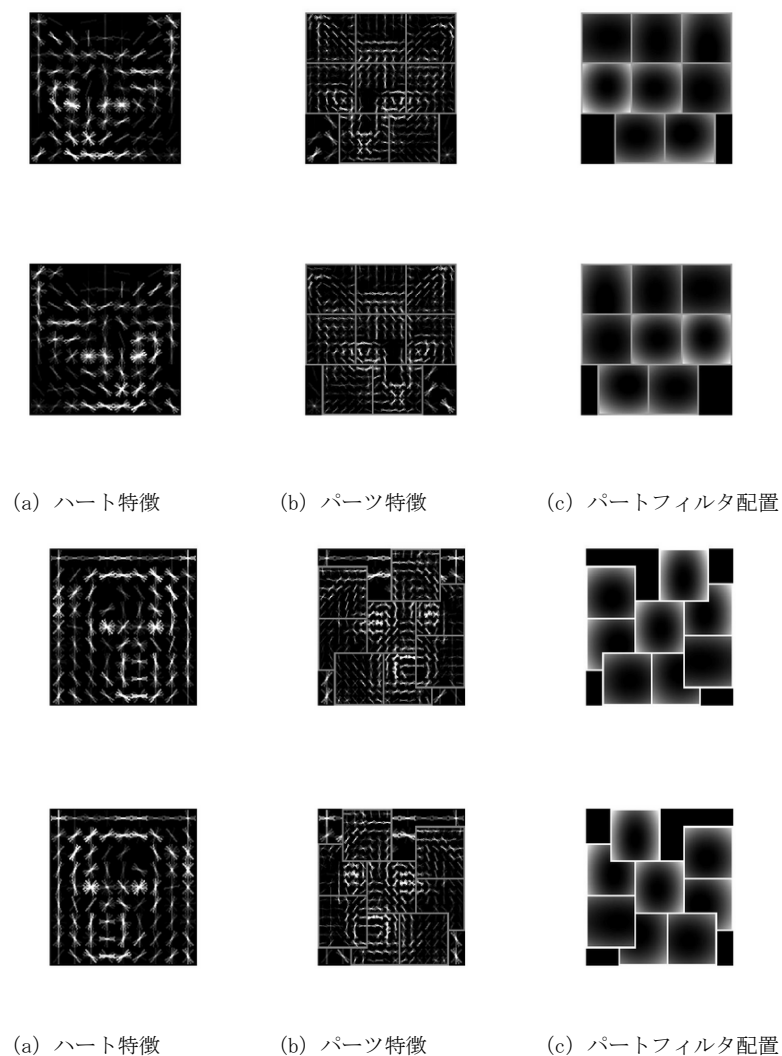


図 6 : 顔検出モデル (上側は猫顔モデル、下側は犬顔モデル)



図 7：検出結果の例（赤枠が顔部分として検出された箇所）

3. 2 検出結果

まず Deformable Part Model を用いて、既知画像の犬と猫のデータの各 1188 個に対して検出実験を行った。その結果、犬顔画像の平均適合率（AP）は 0.715、猫顔画像の平均適合率（AP）は 0.750 であった。犬と猫の PR 曲線を図 8 に示す。

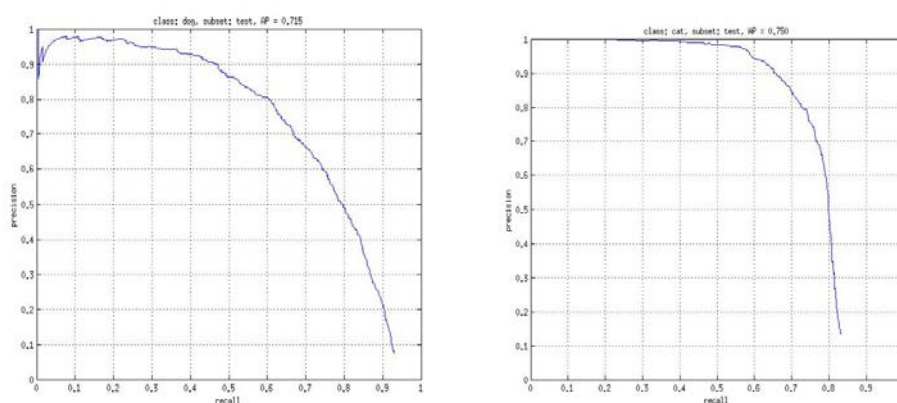


図 8：Precision-Recall 曲線（左側犬、右側猫）

適合率は検索結果のデータに検索された正解データの割合、再現率とは、全データ中の正解データに検索された正解データの割合である。計算式を以下の図 9 に示す。

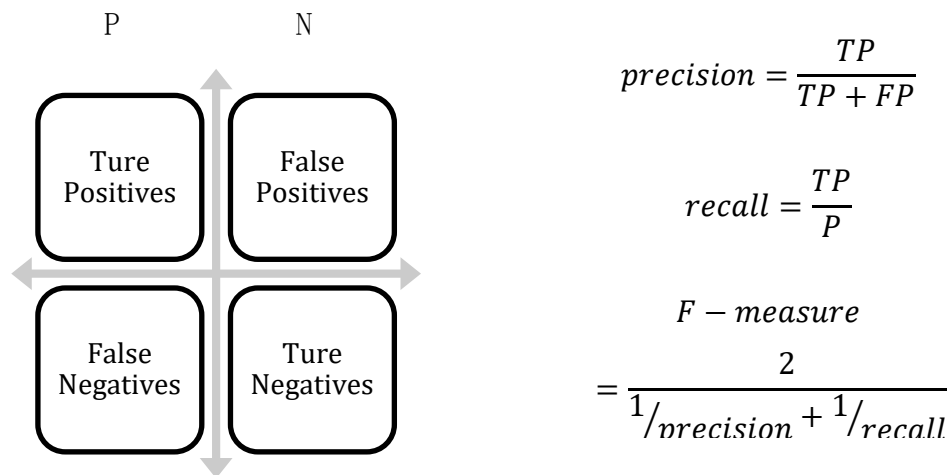


図 9 : 適合率と再現率の計算

P と R はトレードオフの関係である。検索結果として全ての画像を返せば、 $R \rightarrow 1$ となるが $P \rightarrow 0$ となる。検索結果を画像枚数によって、再現率と適合率が変わる。再現率と適合率を総合的な観点から評価するため、平均適合率 (Average Precision, AP) を評価尺度とする。これは検索結果における各再現率レベルでの適合率の平均である。

未知画像に対しては、犬と猫画像データは 30 枚ずつ入力した。結果を以下の表 1 と表 2 に示す。対象となる画像には複数の犬猫が含まれる。そのため、犬画像 61 個、猫画像 59 個であり、その内検出されたものは、犬画像 56 個、猫画像 44 個であった、この結果、犬顔画像の検出率で約 91.8%、猫顔画像の検出率で約 74.5% と高い値が得られた。

顔領域として検出された領域の数は、犬画像 80 個、猫画像 44 個であった、このうち顔領域を全て含んでいるものを全検出、顔領域を一部分含んでいるものを一部検出とする。顔領域を全て含んでいないものが誤検出となる。表 2 の結果より、検出された領域の正解率は、犬画像で 70%、猫画像で 100% となった。

表 1 : Deformable Part Model で検出された顔領域の個数

	対象となる顔画像	検出された顔画像
未知犬画像	61	56
未知猫画像	59	44

表 2 : Deformable Part Model で正しく検出された領域の個数

	検出した領域	完全検出	一部検出	誤検出
未知犬画像	80	56	14	10
未知猫画像	44	44	0	0

次に HOG 特徴量を用いて同様の実験を行った。Deformable Part Model と同じ学習画像を入力し、Adaboost 学習アルゴリズムで犬顔と猫顔を別々に学習させ、1000 回繰り返し学習させた画像より各特徴量を算出し、データベースを構築した。次に未知画像を入力し、検出実験を行った。

今回対象となる未知画像は犬画像 61 個、猫画像 59 個であり、その内検出されたものは、犬画像 43 個、猫画像 29 個であった、この結果より、犬顔画像の検出率で約 70.4%、猫顔画像の検出率で約 49.1%となった。顔領域として検出された領域の数は、犬画像 85 個、猫画像 46 個であった、このうち完全検出した個数、一部検出個数、誤検出個数が表 4 を示す。この結果より、検出された領域の正解率は、犬画像で 48.2%、猫画像で 63.0%となった。

表 3:HOG 特徴量で検出された顔領域の個数

	対象となる顔画像	検出された顔画像
未知犬画像	61	43
未知猫画像	59	29

表 4: HOG 特徴量で正しく検出された領域の個数

	検出した領域	完全検出	一部検出	誤検出
未知犬画像	85	41	32	12
未知猫画像	46	29	0	17

Deformable Part Model と HOG 特徴量による検出性能の比較を表 5 に示す。

表 5 : Deformable Part Model と HOG 特徴量による検出性能の比較

	猫顔検出率	犬顔検出率	猫顔正解率	犬顔正解率
DPM	74.5%	91.8%	100%	70%
HOG 特徴量	49.1%	70.4%	63.0%	48.2%

3. 3 考察

検出結果により、Deformable Part Model が HOG 特徴量より未知画像に対する検出結果が良いことが分かった。その上で、Deformable Part Model と HOG 特徴量のいずれも、犬顔検出率が猫顔検出より検出率約 20%を超えていた。逆に、猫顔正解率は犬顔正解率より約 20%を超えることが分かった。犬顔検出率が猫顔検出より良くなる結果の原因としては、犬顔検出率を計算する時、完全検出された画像と一部分検出された画像二つのタイプを含めて、計算された。その上、HOG 特徴量の場合は一部分検出と完全検出が大体同じ枚数である。

HOG 特徴量を用いた手法では顔や瞳といった対象となる領域以外に、背景といった対象とは異なる領域について誤検出が存在したことに対して、Deformable Part Model では、猫画像に対する検出された領域は全て顔だけの部分である。誤検出率は 0 になった。犬の場合は検索された領域は顔領域の近傍に存在しているから、顔領域とそれ以外の領域の判別について、より正確な判別を行っている。

Deformable Part Model を使って、犬顔検出率と正解率 21.4%、21.8%を改善された、猫顔検出率と正解率の場合は 25.4%、37%を改善された。

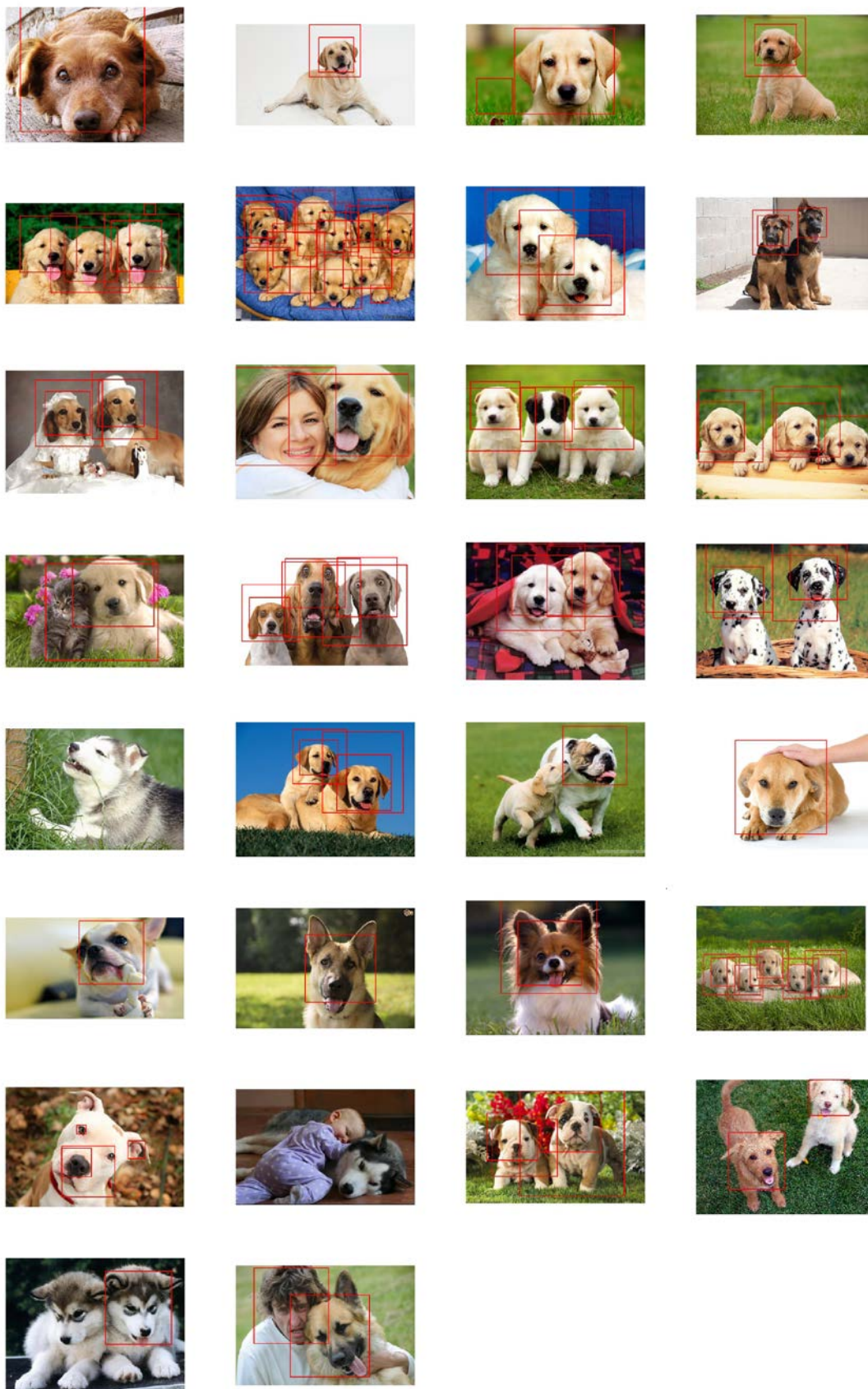
第4章 結論

Deformable Part Model は HOG 特徴量に比べ、犬猫顔画検出において未知画像に対する検出率及び正解率が約 20%改善された。特に犬の顔画像検出率は 91.8%となり、想定したとおりパーツモデルが犬の顔パーツの変動を吸収していると考えられる。しかし、顔パーツモデルの分散が小さいはずの猫顔画像に対する検出率は 74.5%であり、犬顔画像検出率より低い値となった。この理由については明確ではないが、Deformable Part Model のスコア判定時のしきい値設定により大きく検出性能が変化することが影響していると考えられる。

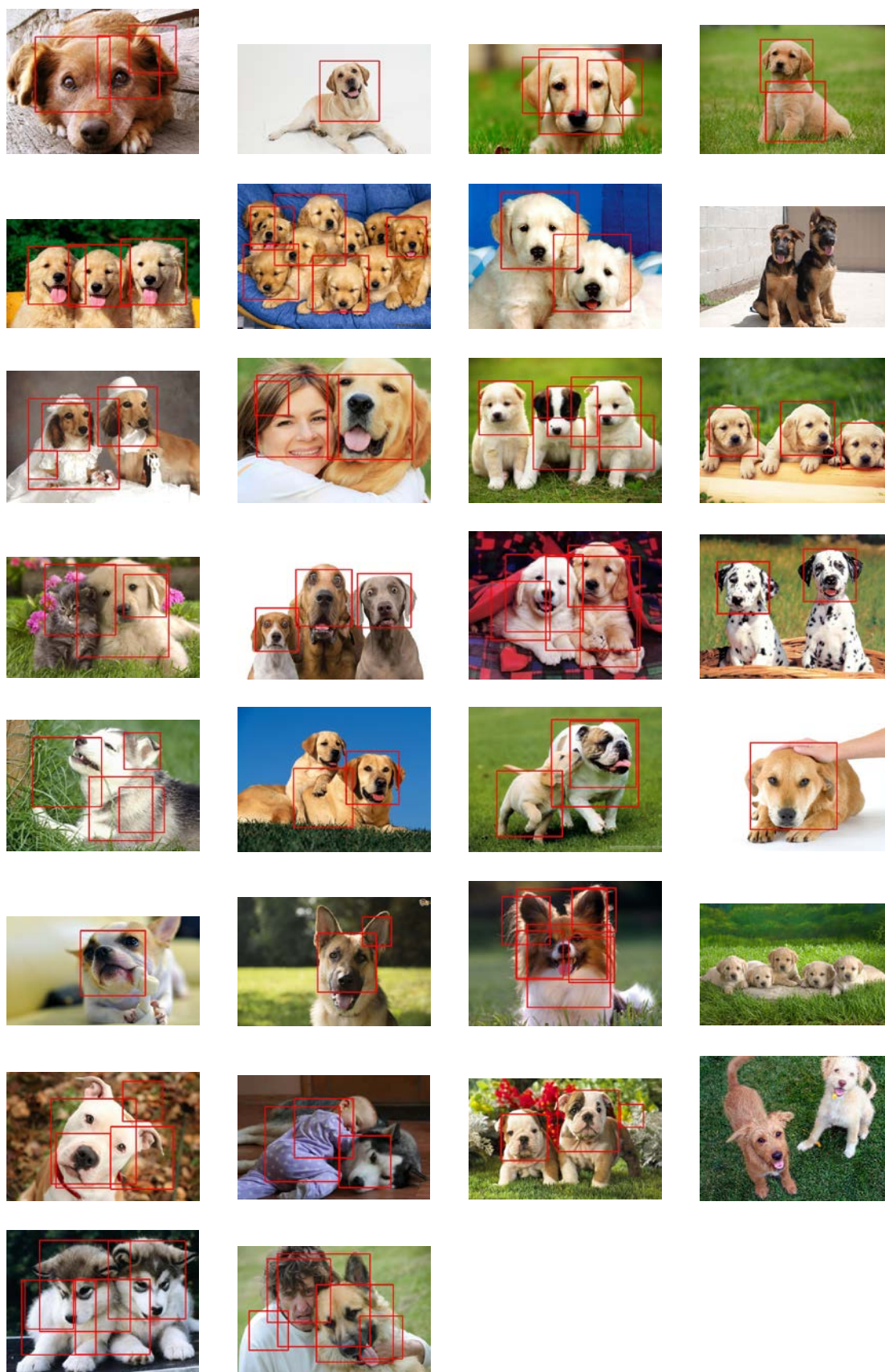
参考文献

- [1] Wikipedia “顔認識システム” <https://ja.wikipedia.org/wiki/顔認識システム>
- [2] 総務省: “消費動向調査平成 21 年 3 月実施調査結果” <http://www.esri.cao.go.jp/>
- [3] 小田 智大, 須藤 秀紹, 小北 麻記子: “猫メディア・エンジンのための基礎研究”, 第 39 回計測自動制御学会北海道支部学術講演会, pp.107-100, 2007
- [4] Oricon Life: “2010 年高まる愛犬撮影のニーズ、「ペット “顔” 認識機能」付きデジカメ発売相次ぐ”, <http://life.oricon.co.jp/73055/>
- [5] Bistro: A Smart Feeder Recognizes Your Cat’s Face (顔認識機能付き自動餌やり器) <https://www.indiegogo.com/projects/bistro-a-smart-feeder-recognizes-your-cat-s-face#/story>
- [6] 草野 孝幸, 出口 大輔, 井手 一郎, 村瀬 洋: “局所特徴を用いた猫の顔検出に関する予備的検討”, 電子情報通信学会総合大会, D-12-110, 2009.
- [7] C. Wang, L. Jiang: “Recognizing Cats and Dogs with Shape and Appearance based Models,” Research Report, School of Computer Science, McGill University 2015. http://www.cim.mcgill.ca/~chuwang/files/cat_dog/cat_dog_pdf.pdf
- [8] P. E. Felzenswalb, et al.: “Object Detection with Discriminatively Trained Part Based Models,” IEEE Trans. PAMI, vol.32, No.9, pp.1627-1645, 2009.
- [9] 藤吉研究室(中部大学工学部情報工学科): “HOG 特徴量と Boosting を用いた人検出” http://www.vision.cs.chubu.ac.jp/joint_hog/pdf/HOG+Boosting_LN.pdf
- [10] 三井 相和: “CV 輪講 A Discriminatively Trained, Multiscale, Deformable Part Model”, FLAB Work Document Jul 15, 2008, pp. 1-10, 2008. <http://www.vision.cs.chubu.ac.jp/CV-R/jpdf/PedroCVPR2008.pdf>
- [11] P. E. Felzenswalb, et al.: “A Discriminatively Trained, Multiscale, Deformable Part Model,” Proceedings of the IEEE CVPR, pp.1-8, 2008.

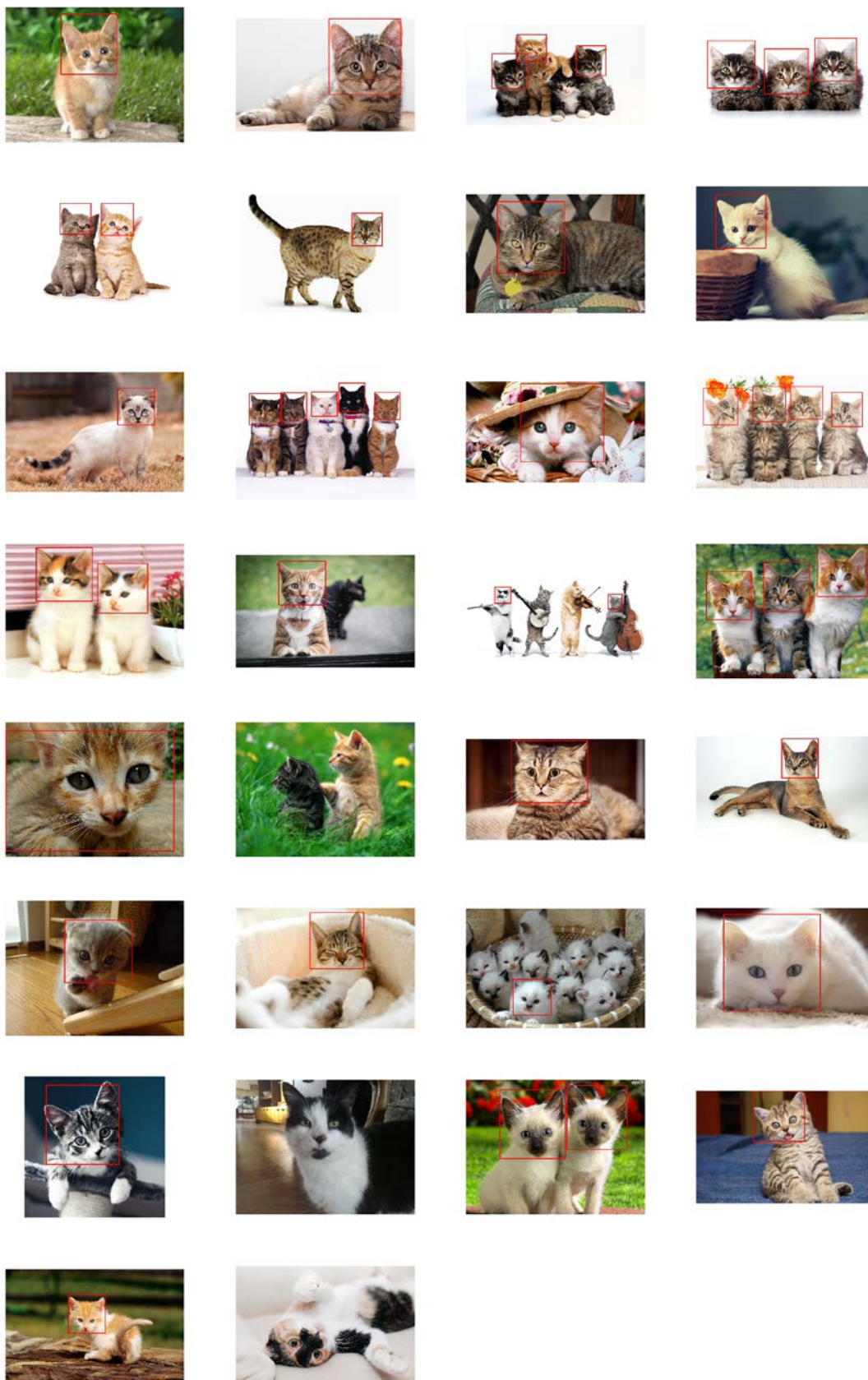
付録 A DPM による犬顔画像検出結果



付録 B HOG による犬顔画像検出結果



付録 C DPM による猫顔画像検出結果



付録 D HOG による猫顔画像検出結果

