

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

○ 草野 孝幸†, 出口 大輔†, 井手 一郎†, 村瀬 洋†

○ Takayuki KUSANO †, Daisuke DEGUCHI †, Ichiro IDE † and Hiroshi MURASE †

†: 名古屋大学大学院情報科学研究科, tkusano@murase.m.is.nagoya-u.ac.jp

本発表では、デジタルカメラで撮影した画像から猫の顔を高精度に検出する手法を提案する。現在、デジタルカメラを所有している人の多くが猫をはじめとするペットや動物の撮影にカメラを使用している。しかしながら、猫の顔には複雑かつ多様な形状、色、模様があるため、人の顔検出手法をそのまま適用しても高い検出性能を得ることはできない。そこで、本発表では猫の顔を構成する複数の猫パーツを個別に検出し、それらの検出結果を組み合わせることで高精度に猫の顔検出を行う。さまざまな組み合わせ方法を検討した結果、従来の人顔検出手法を適用した場合と比較して、大幅な検出性能の向上を確認した。

<キーワード> 猫検出, パーツ統合, 顔検出, Haar-like 特徴, AdaBoost

1. はじめに

内閣府の調査[1]によると、現在デジタルカメラは全人口の7割程度に普及している。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。これらのうち、子供や動物は撮影者の意図通りに静止した状態で撮影することが難しい。よって、これらの写真をきれいに撮影するためには、機械による補助が必要である。これまでに、人間の顔検出に関する研究は盛んに行われており、現在のデジタルカメラのほとんどにその機能が備わっている。しかしながら、ペットや動物の顔の様子は多様なため、これらの手法を利用したとしても高い検出性能を得ることはできない。

猫は古くから人間の愛玩動物として扱われてきており、研究の対象としても興味を持たれている。たとえば、猫を介して地域のコミュニケーションを活性化させようとする研究[2]がある。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性化させる。また、Cat@Logプロジェクト[3]では猫のライフログを記録し、飼い主とペットの交流を活性化させる研究がなされている。このように、猫は我々の生活を豊かにするペットとして広く興味を持たれてきている。

そこで我々は、猫の顔検出技術に着目し、猫の顔を高精度に検出する手法を提案する。Cat@Logプロジェクトでは、Haar-like特徴を用いたカスケード型の識別器で猫の顔検出を行っており、代表的な人間の顔検出と同じ手法が用いられている。しかし、Fig.1に示すように猫の顔は、顔の形状、色、模様が多様なため、人間の顔検出手法をそのまま用いたとしても、猫の顔を検出することは難しい。また、我々が行った予備的検討においても、Haar-like特徴のみでは、猫の顔の複雑な模様を正確にとらえることが難しいという結果を得ている[4]。

我々が一般的に用いる用語には、「猫目」、「猫耳」といった猫の特徴をあらわすものが存在する。これらは多くの種類の猫において共通の特徴であり、特に形状や模様の変化が少ない。本研究ではこのような猫の顔の特徴づけるものを猫パーツと呼ぶ。提案

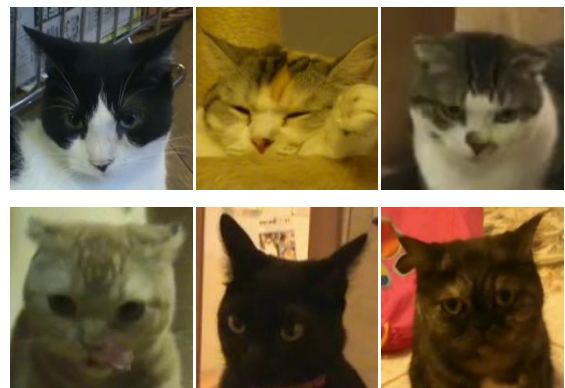


Fig.1 猫の顔の例

手法ではこれらを個別に検出し、組み合わせることで猫の顔を検出する。

2. 猫の顔検出手法

2.1 手法概要

猫の顔は、個体によって形状、色、模様が大きく異なる。そのため、猫の顔全体を正しく学習することは難しい。そこで顔領域に加え、猫の顔を構成する特徴的な猫パーツを検出し、それらを組み合わせることで高精度に猫の顔を検出する。関連する研究として、人物検出をパーツベースに行う研究がある[5]。これは、人物の顔、頭、上半身と包含関係にある部位を高精度に検出し、組み合わせの結果から人物らしい物体を検出する方法である。本研究では、これを猫の顔内部のパーツベースで行う。

提案手法で検出する猫パーツを Fig.2 に示す。具体的には右目 (P^1)、左目 (P^2)、両目 (P^3)、右耳 (P^4)、左耳 (P^5)、両耳 (P^6)、顔 (P^7)、鼻 (P^8) の8箇所を猫パーツとして用いる。これら8個の猫パーツそれぞれを検出する検出器を構築する。以降は猫パーツ k の検出器により得られる領域を P^k と表す。そして、検出結果を猫パーツの位置関係に基づいて組み合わせることで猫の顔全体を検出する。以下で検出器の構成、猫パーツの抽出、猫パーツの組み合わせについて述べる。

2.2 猫パーツ検出器の構成

提案手法では、カスケード型 AdaBoost 識別器を用いた人の顔検出手法[6]を利用し、猫の顔を構成する猫パーツの検出を行う。カスケード型 AdaBoost 識別器では Haar-like 特徴を用いる。Haar-like 特徴は人の顔検出に一般的に用いられている特徴量であり、隣接領域の輝度値の差を表す特徴量である。

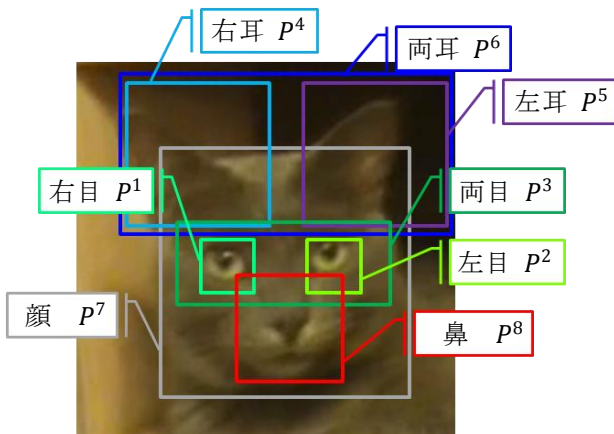


Fig.2 用いる猫パーツ

2.3 猫パーツの抽出

入力された画像に対して、2.2の猫パーツ検出器を用いて猫パーツの抽出を行う。具体的には、入力画像に対して検出窓を走査し、それぞれの猫パーツ検出器により猫パーツを抽出する。最後に検出結果に対して MeanShift クラスタリング[7]を行うことによって統合する。ここで、最終的に抽出された猫パーツのうち、猫パーツ 1 (右目) の検出器で得られた i 番目の領域を p_i^1 、猫パーツ 6 (両耳) 検出器で得られた j 番目の領域ならば p_j^6 のように表す。

2.4 猫パーツの組み合わせによる猫の顔検出

Fig.2 のような猫パーツの位置関係を用いて、抽出された猫パーツの組み合わせを行う。処理は以下の順で行われる。

- (1) 抽出された猫パーツに対応する顔領域の推定
- (2) 猫パーツの組み合わせ
- (3) 猫の顔判定

2.4.1 抽出された猫パーツに対応する顔領域推定

まず、それぞれの猫パーツに対応する顔の領域を求めるため、事前に用意した学習画像から、各猫パーツと顔領域の相対的な X 座標、 Y 座標、拡大率の関係を求める。この関係から検出された猫パーツ k の領域 P を顔領域に変換する関数を $R^k(P)$ とする。例えば、Fig.3 において検出された右目から推定される顔領域は $R^1(P_i^1)$ である。次節では、この顔の推定領域の重なり率を評価することで、パーツを組み合わせるかを判断する。

2.4.2 猫パーツの組み合わせ

ここでは、検出された猫パーツの組み合わせ方法を説明する。ここで、説明を簡単にするため猫パーツ 1 の検出器、猫パーツ 2 の検出器で検出した領域 P_i^1 と P_j^2 について考える。まず、 P_i^1 と P_j^2 のそれぞれに対応した顔領域の重なり率を計算する。重なり率は、それぞれの猫パーツの顔の推定領域 $R^1(P_i^1)$ 、 $R^2(P_j^2)$

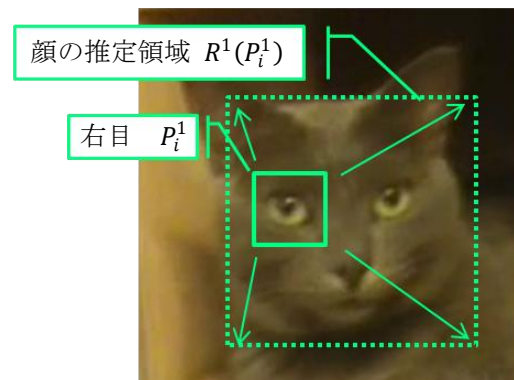


Fig.3 猫パーツから顔領域の推定

Table 1. 猫パーツ別の学習画像の大きさ

猫パーツ	サイズ[pix]
右目	16×16
左目	16×16
両目	32×16
右耳	24×24
左耳	24×24
両耳	32×16
顔	24×24
鼻口	24×24

Table 2. 猫パーツ検出器の性能

猫パーツ	適合率	再現率	F 値
右目	0.248	0.592	0.349
左目	0.291	0.510	0.370
両目	0.581	0.714	0.641
右耳	0.185	0.469	0.266
左耳	0.237	0.531	0.328
両耳	0.481	0.265	0.342
顔	0.829	0.592	0.690
鼻口	0.351	0.694	0.466

を用いて,

$$L(P_k^1, P_l^2) = \frac{R^1(P_k^1) \cap R^2(P_l^2)}{R^1(P_k^1) \cup R^2(P_l^2)} \quad (1)$$

で求める. このとき, 式(1)の値がしきい値以上であれば P_k^1 と P_l^2 の 2 つの猫パーツは同一の猫を表す猫パーツとみなす. これを全ての k と l の組み合わせについて行う.

上記の処理を全ての異なる猫パーツ間で行うことで猫パーツの組み合わせを行う.

2.4.3 猫の顔判定

最終的な猫の顔検出は, 2.4.2 で組み合わせられた猫パーツの総数で判断する. 猫パーツを t 個以上含む猫の顔領域を最終的な猫の顔として検出する. また, 最終的な猫の顔領域は, 組み合わせられた猫パーツの顔領域の平均を出力する.

3. 実験および考察

3.1 実験データ

提案手法の検出性能に関する評価実験を行った. 猫パーツそれぞれの検出器の学習に用いた画像の大きさを Table 1 に示す. 学習画像は, 200 匹の猫から切り出した 5,200 枚 (切り出し位置, 回転を含む) の画像を用いた. また, 検出には学習に使用していない画像 45 枚 (49 匹の猫を含む) を用いた. 猫パ

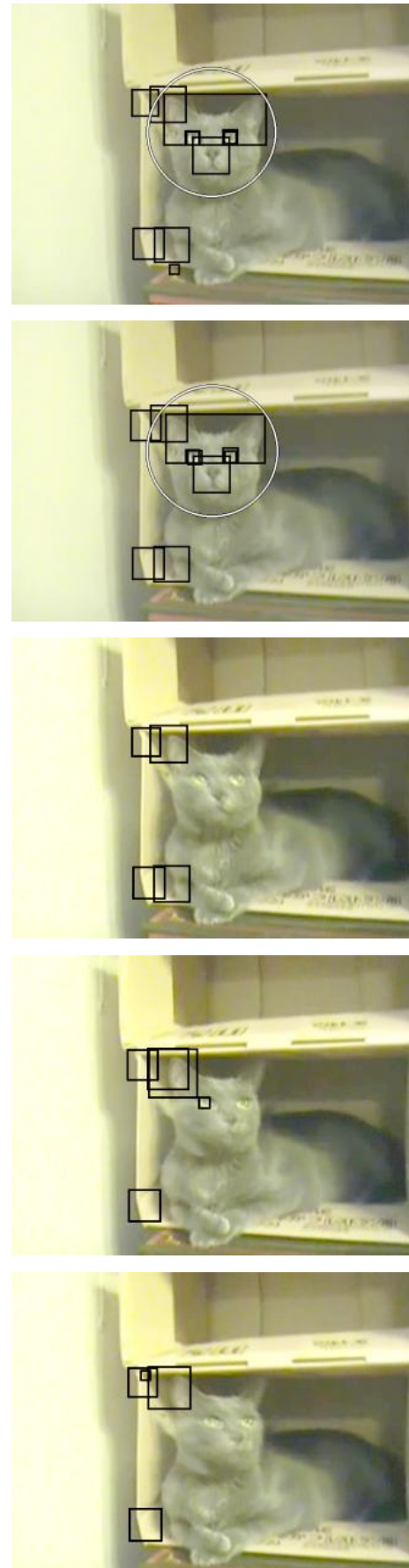
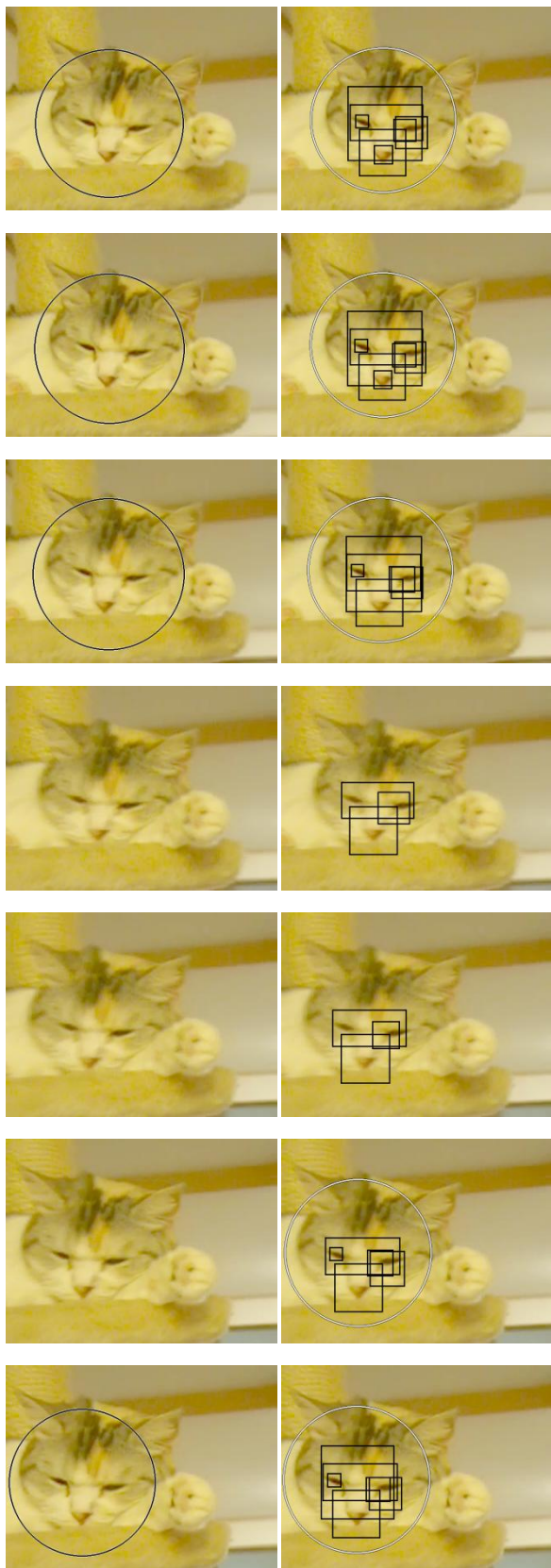
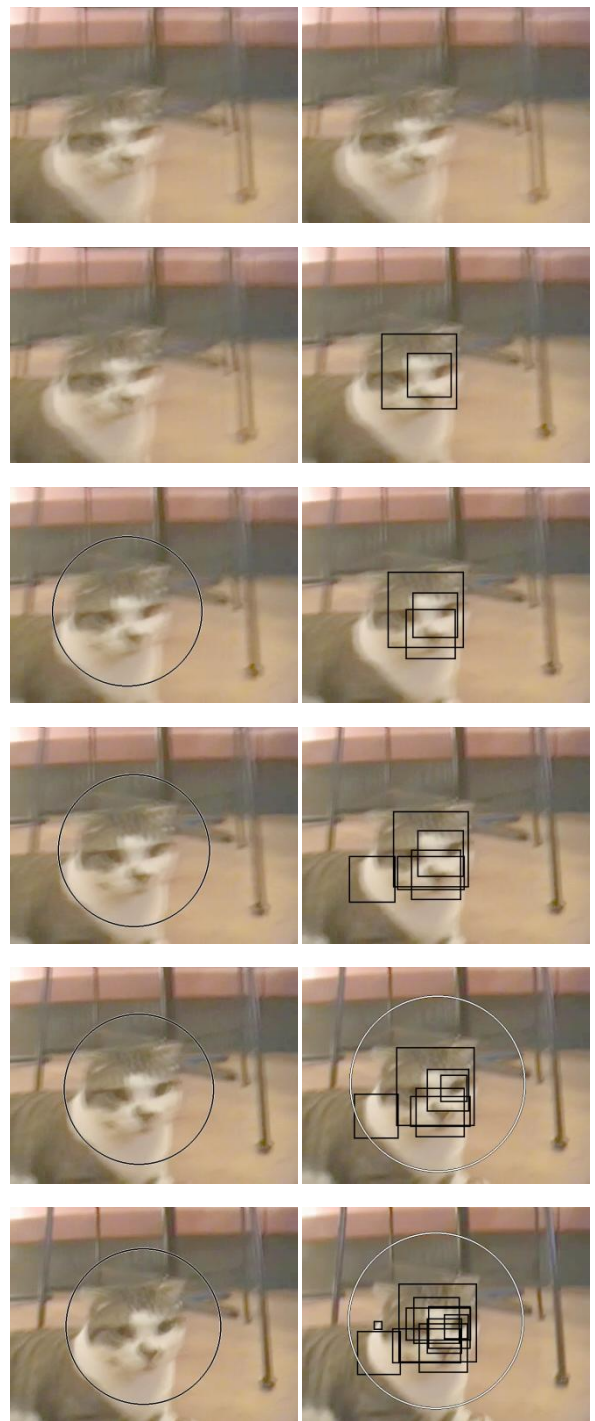


Fig.3 結果例



従来法 提案手法

Fig.4 検出結果例



従来法 提案手法

Fig.5 検出結果例

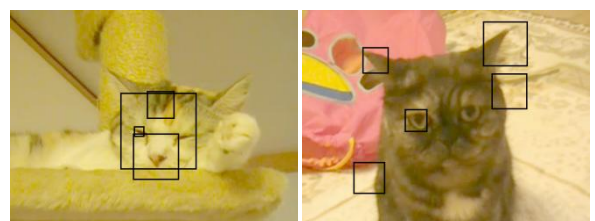


Fig.6 失敗例

Table 3. 組み合わせ手法の性能

	従来手法	提案手法
適合率	0.829	0.907
再現率	0.592	0.776
F 値	0.690	0.836

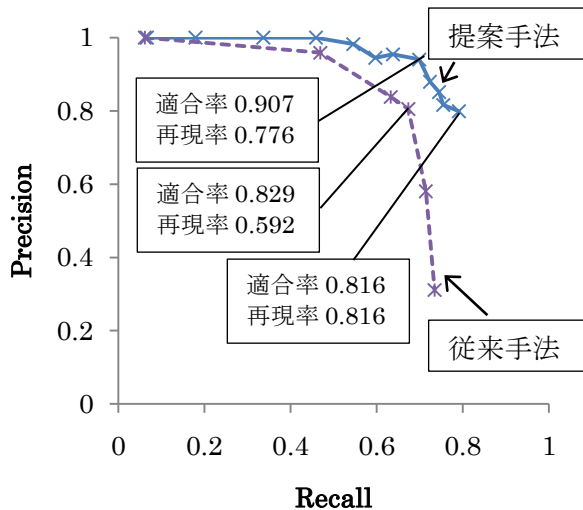


Fig.7 組み合わせの実験結果

一つ検出に用いた検出器の性能を Table 2 に示す.

3.2 組み合わせの評価

従来手法は Haar-like 特徴を用いたカスケード型 AdaBoost 識別器によって猫の顔を学習したものを 用いた. 提案手法による検出結果の一例を Fig.3~6 に示す. Fig.3, Fig.4, Fig.5 は連続した動画像に提案手法を適用した結果であり, Fig.6 は検出の失敗例である. 検出された猫パーツは四角形の枠で囲まれており, 最終的に顔と判断された猫の顔は丸で囲まれている.

提案手法による猫の顔検出と従来法の検出結果の比較を Table 3, Fig.7 に示す. 顔を学習した検出器単体と比べ, 猫パーツを組み合わせることで検出を行う提案手法は, 再現率, 適合率ともに高いことが分かる.

3.3 実験結果の考察

Fig.3 に示す結果と同じデータに対して従来法を適用したところ, 猫の顔を一つも検出することができなかった. 一方, 提案手法を用いることで正面を向いている時に正しく猫の顔を検出することが可能であった. また, Fig.4 も提案手法が多く猫の顔を検出できた例である. これより, 提案手法が猫の顔検出に有効であることがわかる.

Table 3 を見ると, 従来手法の再現率は 60%未満である. これは, 猫の顔は個体によって模様や形状が大きく異なるため, Haar-like 特徴を用いた単独の猫顔検出器では高い再現率が得られないからである. 一方提案手法は, 猫を特徴づける猫パーツを抽出することにより, 78%の再現率を得ることが可能であった. また, Fig.7 を見ると, 従来手法では最大で 80%の再現率しか得られない. しかし, 提案手法を用いることで, 再現率を 80%以上にすることが可能となった.

また, Table 2 の猫パーツの識別器の適合率を見ると, 多くが適合率 50%以下となっている. しかし, これらを組み合わせることにより, 従来手法よりも高い適合率を得ることが可能であった(Table 3). 適合率の低い検出器ではあるが, 複数組み合わせることで性能を改善できた点は, 提案手法の特徴が活かされた結果である.

一方 Fig.5, Fig.6 は検出に失敗した例である. Fig.5 は猫パーツの顔領域は検出できたが, その他のパーツと組み合わせ判定に失敗したため, 検出がうまくいかなかった例である. これは, 組み合わせ時の重なり率のしきい値を低くすることで検出漏れを防ぐことはできるが, 同時に誤検出が増えてしまう. また, Fig.6 左は顔, 鼻, 右目が検出されたが組み合わせに失敗した例である. 鼻は正解よりも大きく, 右目は正解よりも小さい検出枠となっている. これは, 右目・左目は他の猫パーツに比べて大きさが小さいため, 検出枠の大きさの変化に対して, 顔の推定領域が敏感に反応するためである. Fig.6 右の例は左耳と右目が検出されたものの, 組み合わせに失敗した例である. これは, 猫の耳は種類によって大きさや位置の変化が大きいため, あらかじめ用意した猫パーツの位置関係では対応が難しいことを示している. このような検出の失敗を防ぐためには, 単純な重なり率のみではなく, 各猫パーツの位置変動を考慮した組み合わせ手法にする必要がある.

4. むすび

猫パーツの抽出とその組み合わせによる猫の顔検出手法を提案した. 猫の顔を構成する猫パーツを検出し, それらを組み合わせることにより, 適合率 91%, 再現率 78%の高精度かつロバストな猫の顔検出が可能であることを示した.

今後の課題として, Fig.4 のような失敗を防ぐための最適なパーツ組み合わせ方法の検討, 猫パーツ

検出器の性能を改善，猫の模様の変化に頑健な特徴量の検討，などが挙げられる．

謝辞

日頃より熱心にご討論頂く名古屋大学村瀬研究室諸氏に感謝する．本研究の一部は文部科学省研究費補助金による．本研究では，画像処理にMIST ライブラリを使用した．

(<http://mist.s.m.is.nagoya-u.ac.jp/>)

参考文献

- [1] 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/>
- [2] 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,” SICE 北海道支部学術講演会, pp.107–100, 2007
- [3] 米澤 香子, 味八木 崇, 暦本 純一, “Cat@Log: Human Pet Interaction のための猫ウェアラブルセンシング,” 第 17 回インタラクティブシステムとソフトウェアに関するワークショップ (WISS2009), pp.47–52, 2009
- [4] 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”, 2009 電子情報通信学会総合大会講演論文集, vol.2 , p219, 2009
- [5] K. Mikolajczyk, C. Schmid, A. Zisserman, “ Human Detection Based on a Probabilistic Assembly of Robust Part Detectors ” , European Conference on Computer Vision 2004, pp69–82, 2004
- [6] P. Viola, M. J. Jones, “Rapid object detection using a boosted cascade of simple features,” International Journal of Computer Vision, 57(2), 137–154, 2004
- [7] 岡田和典, “ミーンシフトの原理と応用,” 電子情報通信学会技術研究報告, PRMU2007–308, 2008