

位置ずれを許容するエッジ特徴による物体識別

大平 英貴¹ 関根 真弘¹ 西山 正志¹

概要：近年、身の回りの「もの」とデジタルデータを繋げるサービスに期待が集まっている。我々は、「もの」に携帯端末をかざすことで、対応する関連データを提示する新しいサービス形態を実現する技術として、かざす UI を提案している。この UI を実現するためには物体を素早く認識する機能が必須である。しかし、従来の技術では物体の特徴を抽出する処理時間が長いという問題があった。そこで、短い処理時間で物体の特徴を抽出できる領域ベースエッジ特徴を開発し、高い認識率、かつ、短い処理時間の認識機能を実現した。これにより、物体を認識率 99% で見分ける機能と、ARM Cortex-A9 1.3GHz の携帯端末上で応答速度 0.5 秒で動作する UI を実現した。

1. はじめに

近年、拡張現実感の技術を用いて「もの」とそれに対応する関連データを繋げるサービスに期待が集まっている。例えば、目の前の雑誌や新聞記事に携帯端末をかざすと、それらに関連するインターネット動画を再生したり、ウェブサイトを表示したりすることで、ユーザにこれまでにない驚き、面白さ、手軽さを与えることができる。

我々は、拡張現実感を用いた新しいサービスを実現する技術として、かざす UI [1] を提案している。かざす UI では、ユーザが身の回りの「もの」に携帯端末をかざすと、事前に登録した関連データを提示することを想定している (図 1 参照)。

このかざす UI を実現するには、目の前の物体をカメラで素早く認識する技術が必要となる。カメラを使った物体認識は、画像記述子 [2-9] が性能を左右する。画像記述子は局所特徴 (local feature) と大域特徴 (global feature) に分類できる。局所特徴とは画像の一部から抽出した特徴量を使って物体を認識する手法であり、SIFT [2] や ORB [3] が一般的に用いられる (図 2(a) 参照)。局所特徴では、特徴量を抽出する位置を決定する処理と、その位置から特徴量を抽出する処理と、特徴量同士をマッチングする処理が必要になる。局所特徴は物体の見え方の変化に対して頑健である一方、多くの物体を認識しようとする特徴量同士をマッチングするための処理時間が長くなり、限られた計算性能の携帯端末上で快適に動作しなくなるという課題がある。大域特徴とは画像全体から抽出した特徴量を使って物体を認識する手法であり、HOG [4] が一般的に用いられる



図 1 かざす UI の例

(図 2(b) 参照)。大域特徴では、特徴量を抽出するという処理と、特徴量同士をマッチングする処理が必要になる。大域特徴は、局所特徴と比較して特徴量同士をマッチングするための処理時間は短い、物体の見え方が少しでも変化すると認識できなくなるという課題がある。かざす UI では、携帯端末上で多くの物体を素早く認識するため、大域特徴を用いる。

そこで、従来の大域特徴の課題に対し、高い認識率、かつ、短い処理時間の領域ベースエッジ特徴を開発した。これにより、20 種類の雑誌を認識率 99% で見分ける認識機能と ARM Cortex-A9 1.3GHz の携帯端末上で応答速度 0.5 秒で動作する UI を実現した。

2. 位置ずれの概要

2.1 位置ずれによる問題

かざす UI では様々な見え方で撮影された物体を認識す

¹ 株式会社東芝 研究開発センター
Research and Development Center, TOSHIBA Corporation



(a) 局所特徴 (b) 大域特徴

図 2 画像記述子の例．図中の四角は特徴量を抽出する範囲（パッチ），矢印はパッチの方向を意味する．



図 3 物体の見え方が変化する例

る必要がある（図 3 参照）．見え方の変化には水平方向の変化，角度の変化，距離の変化がある．

従来のエッジ算出手法として，Sobel がある．Sobel はサンプリング点の隣接画素から算出したエッジ強度を使ってマッチングする．ここで隣接画素とは，サンプリング点を中心とした周囲の 8 つの画素のことである．エッジ強度は，これらの画素に対し図 4 に示すような係数をそれぞれ乗算して求めた x 方向の輝度差と y 方向の輝度差から算出する．この手法は，各サンプリング点が狭い範囲の形状しか抽出できないため，物体の見え方が変化すると，認識時と登録時でエッジ強度の位置がずれて認識できなくなるという問題がある（図 5 参照）．また，このエッジ算出方法は，x 方向の輝度差と y 方向の輝度差のみでエッジを算出するため，エッジを頑健に算出できないという問題もある．

HOG はサンプリング点の隣接画素から算出したエッジの方向をヒストグラム化したものを使ってマッチングする．この手法はヒストグラム化によってエッジの位置ずれを許容しているため，Sobel よりも高い認識率を実現できる一方，物体の角度の変化すると，認識時と登録時でエ

-1	1	-1	-2	-1
-2	2	2	2	2
-1	1	1	2	1

x 方向のエッジ y 方向のエッジ

図 4 Sobel のエッジ算出方法

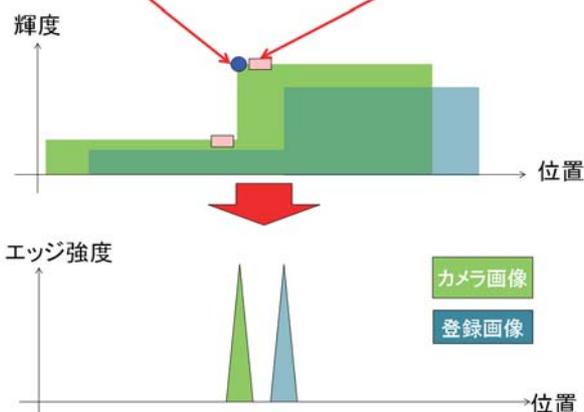
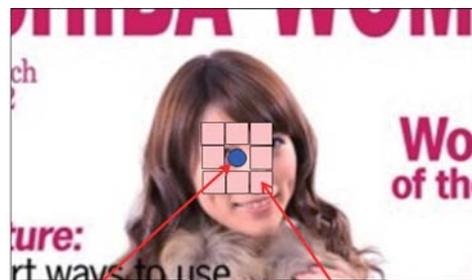


図 5 隣接画素からエッジを算出する方法の概念図

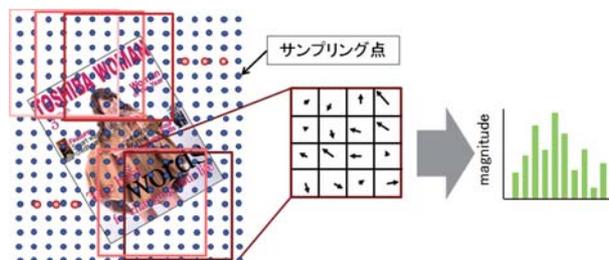


図 6 多数のサンプリング点からのヒストグラム化の概念図

ジの方向がずれて正しくヒストグラム化できなくなるという問題がある．また，この手法は各サンプリング点が狭い範囲の形状しか抽出できないため，物体全体の形状を抽出するには多くのサンプリング点が必要となり，処理時間が長くなるという問題もある（図 6 参照）．

2.2 位置ずれへの対応

これらの問題に対し，広範囲の形状抽出と少数のサンプリング点によるヒストグラム化が可能な領域ベースエッジ特徴を提案する．領域ベースエッジ特徴では，

- サンプリング点の周辺領域を使った広範囲の形状抽出による水平方向の変化への対応（図 7 参照），
- 少数のサンプリング点による短い処理時間でのヒスト

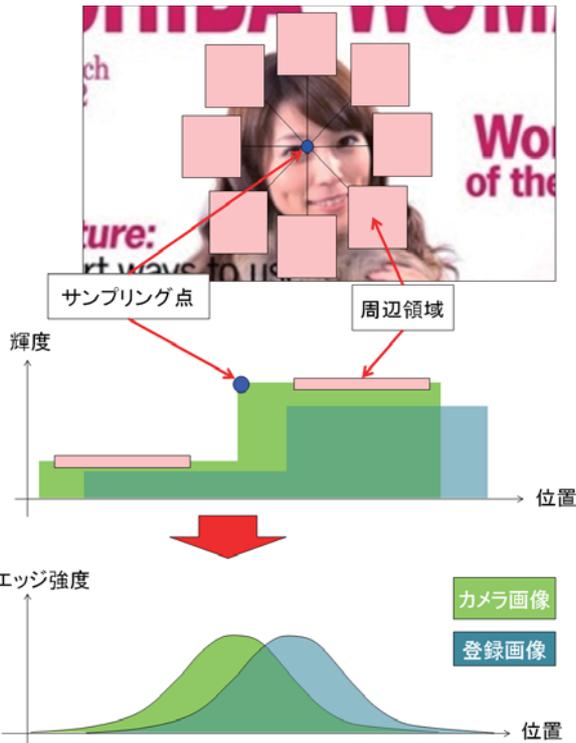


図 7 周辺領域からエッジを算出する方法の概念図

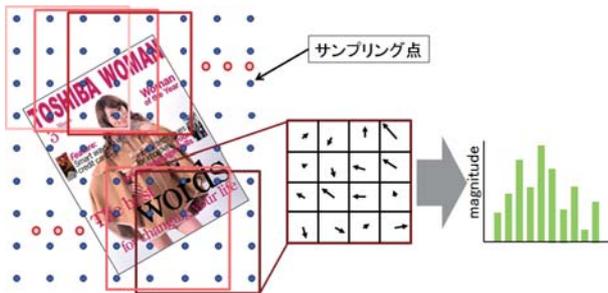


図 8 少数のサンプリング点からのヒストグラム化の概念図

グラム化 (図 8 参照) ,

- 物体の面内回転の変化推定による角度の変化への対応 (図 9 参照) ,

- マルチスケール探索による距離の変化への対応

を行う。これにより、エッジの位置や角度のずれの許容と短い処理時間での特徴量抽出が可能となる。

3. 領域ベースエッジ特徴

3.1 位置ずれに頑健なエッジ算出方法

領域ベースエッジ特徴は、図 10 に示すようなサンプリング点の周辺に円形に配置した領域からエッジを算出する。周辺領域から取得した画素情報を使ってエッジを算出することで広範囲の形状を抽出可能になり、エッジの位置ずれに対して頑健になる (図 7 参照)。また、周辺領域を円形に配置し、 x 方向の輝度差と y 方向の輝度差に加えて $x-\pi/4$ 方向の輝度差と $y-\pi/4$ 方向の輝度差も算出することで、エッジを頑健に算出可能にする。これにより、エッジ



図 9 面内回転推定の概念図

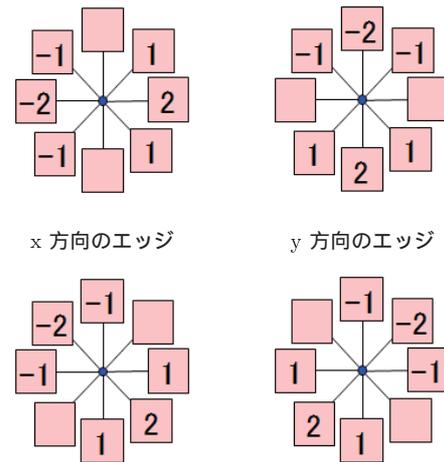
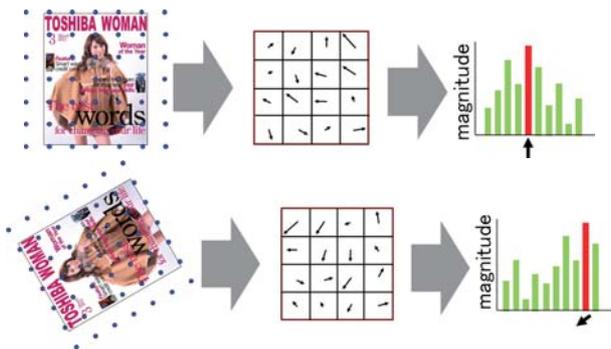


図 10 領域ベースエッジ特徴のエッジ算出方法

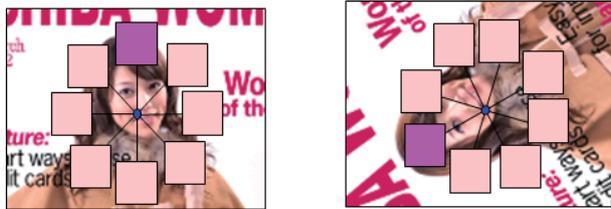
強度によるマッチングが位置ずれに対して頑健になる。高い認識率を実現できるヒストグラムによるマッチングでは、領域ベースエッジ特徴によって各サンプリング点が広範囲の形状を抽出できるため、少数のサンプリング点で物体全体の形状を抽出できるようになり、ヒストグラム化の処理時間を短くすることができる。また、周辺領域の形状を矩形にすることで積分画像 (Integral Image) を使ったエッジ算出が可能となり処理時間を短くすることができる。

3.2 面内回転の変化推定

物体の角度の変化に対応するために、物体の面内回転方向を推定し、パッチを回転させる。面内回転方向は、エッジ方向のヒストグラムの中で最も頻度の高い方向 (Dominant Orientation) を基準に決定する (図 11(a) 参照)。そして、その方向にサンプリング点の位置と周辺領域を回転させる。サンプリング点はパッチの中心位置を基準に回転させる (図 11(b) 参照)。領域ベースエッジ特徴はサンプリング点が少数であるため、高速に回転することができる。周辺領域は、その領域に対応するサンプリング点の位置を基準に回転させる (図 11(c) 参照)。撮影する物体の角度が変化した場合でも同じようにエッジ方向を算出できる。



(a) Dominant Orientation



(c) 周辺領域の回転制御



(b) サンプリング点の回転制御

図 11 パッチの回転制御

3.3 マルチスケール探索

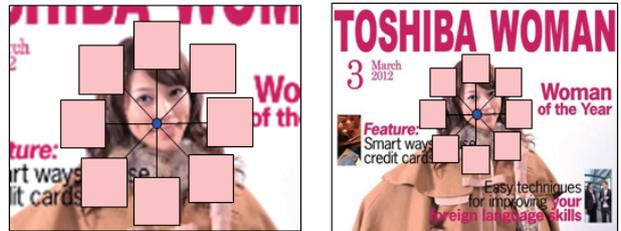
カメラと物体との距離の変化に対応するために、パッチのスケールを変えて物体を探索する。特徴量を抽出する範囲のスケールは、サンプリング点の間隔、サンプリング点と周辺領域との距離、周辺領域の大きさによって変化させることができる。例えば、カメラと物体との距離が遠いために物体のスケールが小さく撮影されている場合は、サンプリング点の間隔を狭く、サンプリング点と周辺領域との距離を短く、周辺領域の大きさを小さくすることで、物体のスケールにパッチを合わせることができる(図 12 参照)。領域ベースエッジ特徴はサンプリング点が少数であることと、周辺領域の大きさを領域の 4 隅の位置で設定できることから、少ないメモリアクセス回数によるスケール変化が可能となる。

4. 実験による性能評価

領域ベースエッジ特徴の有効性を確認するために、面内回転の変化に対する頑健性の評価実験、仰俯角の変化に対する頑健性の評価実験、距離の変化に対する頑健性の評価実験、特徴量抽出に要する処理時間の評価実験、登録物体の認識に要する応答速度の評価実験を行った。



(a) サンプリング点のスケール制御



(b) 周辺領域のスケール制御

図 12 パッチのスケール制御

面内回転の変化に対する頑健性の評価実験、仰俯角の変化に対する頑健性の評価実験、距離の変化に対する頑健性の評価実験では、領域ベースエッジ特徴、HOG、Sobel の 3 つの手法で物体の見え方の変化に対する認識率を測定した。ここで認識率とは、同物体の受入率と他物体の排除率が等しくなるように設定した場合の値である。特に、領域ベースエッジ特徴は次の 3 つの手法を使って評価した。

- 提案手法 ①
ヒストグラムを使った領域ベースエッジ特徴
- 提案手法 ②
面内回転の変化推定とマルチスケール探索が無いヒストグラムを使った領域ベースエッジ特徴
- 提案手法 ③
面内回転の変化推定とマルチスケール探索が無いエッジ強度を使った領域ベースエッジ特徴

認識率は回転角 0° 、仰俯角 0° 、距離 40cm の雑誌を 20 種類登録し、それぞれの雑誌に対し回転角と仰俯角を 10° ごと、距離を 5cm ごとに变化させたデータベースを用いて評価した(図 13 参照)。ここで仰俯角とは水平を基準とした上下方向の角度のことをいう。実験では各雑誌に付き 1 枚の画像を登録した。特徴量抽出に要する処理時間の評価実験では、特徴量抽出時間を従来手法と比較した。

登録物体の認識に要する応答速度の評価実験では、提案手法 ① と ORB で登録物体数の違いに対する応答速度の変化を測定した。

処理時間と応答速度は Intel Core i5 CPU 2.67GHz の PC を使って測定した。

4.1 面内回転の変化に対する頑健性の評価実験

面内回転の変化に対する頑健性を検証するために、雑誌

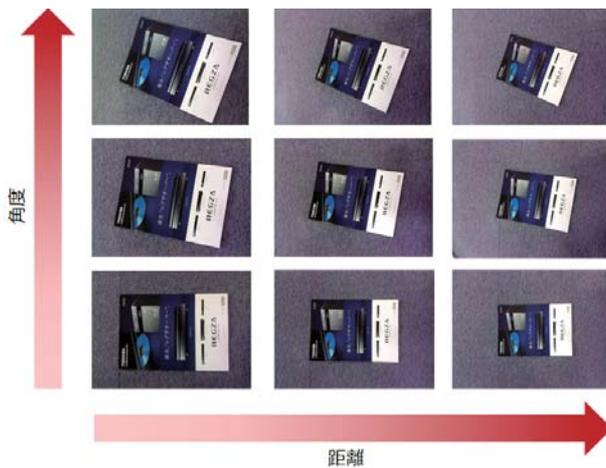


図 13 評価用データの例

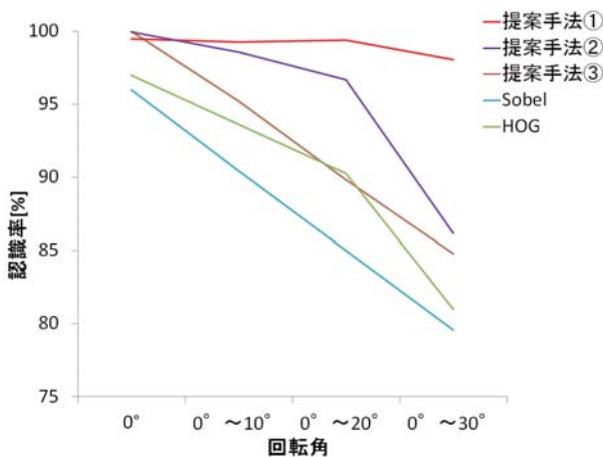


図 14 回転角を変化させた時の認識率

の回転角を変化させた時の各手法の認識率を測定した。実験では、回転角が 0° の画像 100 枚、 0° から 10° の画像 220 枚、 0° から 20° の画像 340 枚、 0° から 30° の画像 460 枚の 4 つの条件で認識率を測定した。いずれの条件でも、仰俯角は 0° 、距離は 40cm である。実験結果を図 14 に示す。回転角が大きい場合でも提案手法①の認識率が高いことから、面内回転の変化推定によって回転角の変化に頑健になることが確認できた。また、提案手法②が HOG よりも高い認識率であることと、提案手法③が Sobel よりも高い認識率であることから、周辺領域からエッジを算出することによって回転角の変化になることが確認できた。特に提案手法②と HOG の認識率が回転角の変化に伴い著しく低下した理由として、雑誌の回転によってエッジの方向が変わり、エッジのヒストグラムが認識時と登録時とで異なる特徴になったためと考えられる。

4.2 仰俯角の変化に対する頑健性の評価実験

仰俯角の変化に対する頑健性を検証するために、雑誌の仰俯角を変化させた時の各手法の認識率を測定した。実験では、仰俯角が 0° の画像 100 枚、 0° から 10° の画像 220

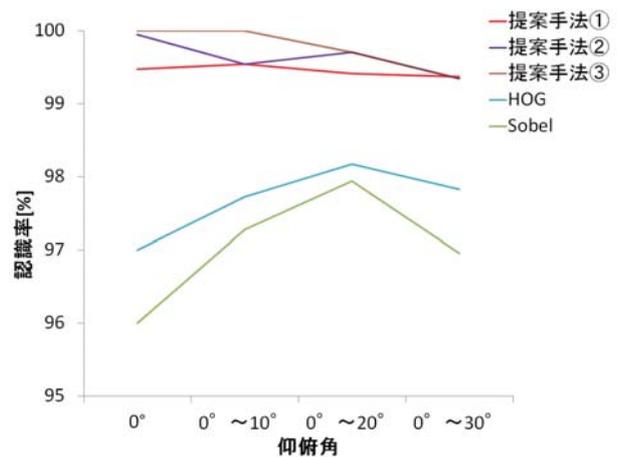


図 15 仰俯角を変化させた時の認識率

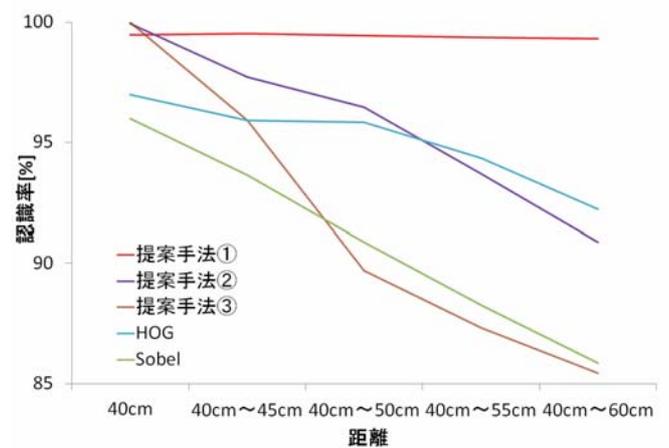


図 16 距離を変化させた時の認識率

枚、 0° から 20° の画像 340 枚、 0° から 30° の画像 460 枚の 4 つの条件で認識率を測定した。いずれの条件でも、回転角は 0° 、距離は 40cm である。実験結果を図 15 に示す。提案手法①、提案手法②、提案手法③の全てにおいて識別率は 99% 以上であり、仰俯角の変化に対して頑健であることが確認できた。

4.3 距離の変化に対する特徴量の性能評価

距離の変化に対する頑健性を検証するために、雑誌の距離を変化させた時の各手法の認識率を測定した。実験では、距離が 40cm の画像 100 枚、40cm から 45cm の画像 220 枚、40cm から 50cm の画像 340 枚、50cm から 55cm の画像 460 枚、55cm から 60cm の画像 580 枚の 5 つの条件で認識率を測定した。いずれの条件でも、回転角と仰俯角は 0° である。距離が大きく変化した場合でも提案手法①の認識率が高いことから、マルチスケール探索によってエッジの角度のずれに頑健になることが確認できた。

4.4 特徴量抽出に要する処理時間の評価実験

特徴量抽出に要する処理時間を評価するために、各手法

表 1 各手法の特徴量抽出の処理時間

認識手法	提案 手法 ①	提案 手法 ②	提案 手法 ③	HOG	Sobel
処理時間 [ms]	2.2	1.82	1.76	13.82	1.58

の特徴量抽出の処理時間を測定した。各手法の処理時間を表 1 に示す。提案手法 ① や提案手法 ② が HOG に比べて高速である理由として、ヒストグラム化の際にメモリを参照する回数は HOG が 289444 回であることに對し、領域ベースエッジ特徴が 1000 回であることから、少ないメモリ参照回数によって短い処理時間でヒストグラム化できることが確認できた。また、提案手法 ① と提案手法 ② の処理時間に大きな差が無いことから、短い処理時間で面内回転の変化推定とマルチスケール探索が可能であることが確認できた。

4.5 登録物体の認識に要する応答速度の評価実験

登録物体の認識に要する応答速度を評価するために、提案手法 ① と ORB を使って登録物体数の違いに対する応答速度の変化を測定した。ORB は下記の式を使ってマッチングした、

$$\frac{\text{登録画像の特徴点と撮影画像の特徴点の対応数}}{\text{登録画像の特徴点の数}} \quad (1)$$

特徴点同士の対応付けは全探索によって行う。

実験結果を図 17 に示す。ORB は登録数の増加に伴いマッチングの処理時間が大幅に長くなるため応答速度が低下する一方、提案手法 ① は登録物体数が増加した場合でも速い応答速度で認識できることが確認できた。

また、回転角を 0° から 20°、仰俯角を 0° から 20°、距離を 40cm から 60cm まで変化させた画像 5380 枚を使って認識率を評価した所、提案手法 ① は 99.3[%]、ORB は 98.4[%] であり、提案手法 ① は局所特徴と比較しても高い認識精度であることが確認できた。

ARM Cortex-A9 1.3GHz の CPU を搭載した携帯端末上でかざす UI を動作させた時の応答速度は 20 物体を登録した場合、340.5ms であり、携帯端末上で高速に動作することを確認した。

5. おわりに

「もの」とその関連データを繋げるかざす UI を実現するために、高い認識率、かつ、短い処理時間で物体を認識する領域ベースエッジ特徴を提案した。実験により、領域ベースエッジ特徴は物体の見え方が変化した場合でも頑健に認識できることを確認した。今後は、さらに高い認識率を実現するとともに、マッチングに要する処理時間をさらに短くする。また、かざす UI の利便性を定量評価によって確認することで、かざす文化の創造を目指していく。

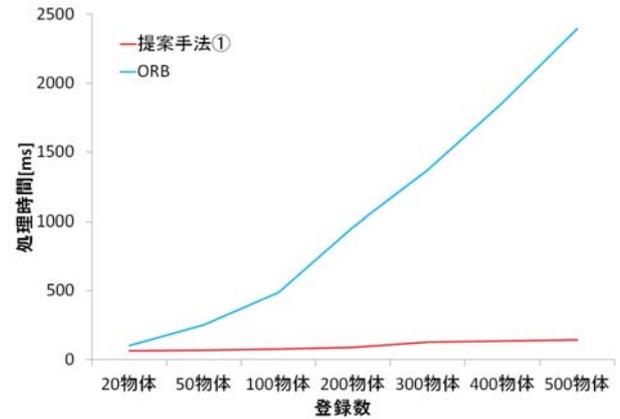


図 17 登録数と応答速度との関係

参考文献

- [1] 大平 英貴, 広畑 誠, 関根 真弘, 西山 正志, 山内 康晋, : かざす UI ~ 携帯端末で高速な物体認識 ~ , 第 15 回 画像の認識・理解シンポジウム論文集 (2012).
- [2] D. G. Lowe.: *Distinctive image features from scale-invariant keypoints*. International Journal of Computer Vision, 60:91-110 (2004).
- [3] E.Rublee, V.Rabaud, K.Konolige, and G.Bradschi.: *Orb: an efficient alternative to SIFT or SURF*, IEEE International Conference on Computer Vision, pp.2564-2571 (2011).
- [4] N. Dalal and B. Triggs.: *Histograms of oriented gradients for human detection*, Int. Conf. on Computer Vision and Pattern Recognition, Vol.2, pp.886-893 (2005).
- [5] H. Bay, T. Tuytelaars, and L. Van Gool.: *Surf: Speeded up robust features*. In European Conference on Computer Vision, pp.404-417 (2006).
- [6] M. Calonder, V. Lepetit, C. Strecha and P. Fua.: *Brief: Binary robust independent elementary features*, In European Conference on Computer Vision, pp.778-792 (2010).
- [7] S. Leutenegger, M. Chli, and R. Siegwart.: *Brisk: Binary robust invariant scalable keypoints*. IEEE International Conference on Computer Vision, pp.2548-2555 (2011).
- [8] M. Ambai and Y. Yoshida.: *Card: Compact and real-time descriptors*. IEEE International Conference on Computer Vision, pp.97-104 (2011).
- [9] A.Alahi, R.Ortiz, and P.Vanderghenst.: *Freak: Fast retina keypoint*, Int. Conf. on Computer Vision and Pattern Recognition, pp.510-517 (2012).