

平成 26 年度  
学士学位論文

# ORB 特徴を用いた画像の一部の類似性に 基づく画像検索

Region-Based Image Retrieval using Oriented FAST  
and Rotated BRIEF

1150321 竹中 浩貴

指導教員 吉田 真一

2015 年 2 月 27 日

高知工科大学 情報学群

# 要 旨

## ORB 特徴を用いた画像の一部の類似性に基づく画像検索

竹中 浩貴

従来の画像の一部の類似性に基づく画像検索 (部分類似画像検索) では, 大域特徴である色特徴や形状特徴などを用いたものが提案されている. 近年では, 画像間の対応付けは局所特徴を用いる方法が有効であると考えられており, 部分類似画像検索に SURF 特徴を用いたものが提案されている. しかし, SIFT 特徴や SURF 特徴などの特徴量を高次元ベクトルで表現する局所特徴は認識率の向上に有効である反面, メモリ使用量が多く, また類似度計算が遅いといった問題がある. そこで, 本研究では特徴量をバイナリコードで表現する ORB 特徴を用いた部分類似画像検索を提案する. 評価実験は, 被験者 10 名による画像 200 枚での検索実験を行う. 試行は ORB 特徴を用いた場合と SURF を用いた場合の 2 種類行い, クエリとして提示されている画像 20 枚から最低 1 枚以上を選択して検索を行う. 検索結果から求められる適合率・再現率・F 値と特徴抽出にかかる計算時間で評価を行う. ORB 特徴を用いた手法では, 再現率 10%, 適合率 41%, F 値 0.13 であったのに対し, SURF 特徴を用いた手法では, 再現率 12%, 適合率 46%, F 値 0.14 を示しており, 精度に減少傾向が見られるが, 特徴抽出にかかる計算時間では, SURF 特徴が 12 秒だったのに対し, ORB 特徴は 2.94 秒と 4.08 倍の高速化を確認している.

キーワード 画像の一部の類似性に基づく画像検索, ORB 特徴

# Abstract

## Region-Based Image Retrieval using Oriented FAST and Rotated BRIEF

Hiroki TAKENAKA

Region-based image retrieval(RBIR) has been studied to retrieve partly similar images. Conventional RBIR has been proposed and they use color features and shape features, global features. Recently, local features have been studied for Content-based image retrieval(CBIR) and image matching. Kamimura has proposed an application of the Speeded-Up Robust Features(SURF) algorithm to RBIR in 2013. However, some of the features are represented by high-dimensional vector such as SIFT and SURF. These features require a lot of time for feature description and the similarity calculation of image and increase memory usage. In this study, a region-based image retrieval using Oriented FAST and Rotated BRIEF(ORB) is proposed. Proposed RBIR system is evaluated using precision, recall, and F-measure obtained by subjective experiment using 200 images by 10 subjects. Moreover, we calculate the time required for image feature extraction. The precision, recall, and F-measure of proposed RBIR system using ORB are 10%, 41%, and 0.13, respectively, whereas conventional RBIR system using SURF are 12%, 46%, and 0.14, respectively. The processing time for feature extraction of the ORB is 2.94 seconds in average, whereas SURF is 12 seconds in average. In comparison, accuracy of the RBIR system using SURF is better than the RBIR system using ORB, however, ORB could be 4.08 times faster processing time for feature extraction.

**key words**      region-based image retrieval, Oriented FAST and Rotated BRIEF(ORB)

# 目次

|       |  |    |
|-------|--|----|
| 第 1 章 | 序論   | 1  |
| 第 2 章 | 関連研究   | 4  |
| 2.1   | 検索手法 . . . . .   | 4  |
| 2.1.1 | Text-Based Image Retrieval . . . . .                           | 4  |
| 2.1.2 | Content-Based Image Retrieval . . . . .                        | 5  |
| 2.1.3 | Region-Based Image Retrieval . . . . .                         | 5  |
| 2.2   | 画像認識に用いる特徴量 . . . . .  | 5  |
| 2.2.1 | Features from Accelerated Segment Test(FAST) . . . . .         | 6  |
| 2.2.2 | Binary Robust Independent Elementary Features(BRIEF) . . . . . | 7  |
| 2.2.3 | Oriented FAST and Rotated BRIEF(ORB) . . . . .                 | 8  |
|       | 特徴点検出 . . . . .  | 8  |
|       | 特徴量記述 . . . . .  | 9  |
| 2.3   | クラスタリング手法 . . . . .  | 9  |
| 2.3.1 | K 平均法 (K-means clustering) . . . . .                           | 9  |
| 2.3.2 | Ward 法 . . . . .   | 10 |
| 2.4   | Bag-of-Visual-Word . . . . .                                   | 10 |
| 第 3 章 | <b>ORB</b> 特徴を用いた部分類似画像検索                                      | 11 |
| 3.1   | Key-Image の作成 . . . . .  | 12 |
| 3.1.1 | 画像の分割 . . . . .  | 12 |
| 3.1.2 | Visual Word の選定 . . . . .                                      | 13 |
| 3.1.3 | VW ヒストグラムの作成 . . . . .   | 13 |
| 3.1.4 | Key-Image の選定 . . . . .  | 14 |

## 目次

|       |                                 |    |
|-------|---------------------------------|----|
| 3.1.5 | 検索結果の表示 . . . . .               | 15 |
| 第 4 章 | 構築システムの評価                       | 16 |
| 4.1   | 実験条件 . . . . .                  | 16 |
| 4.2   | 評価項目 . . . . .                  | 17 |
| 4.3   | 実験結果 . . . . .                  | 18 |
| 4.4   | 考察 . . . . .                    | 19 |
| 第 5 章 | まとめ                             | 20 |
|       | 謝辞                              | 21 |
|       | 参考文献                            | 23 |
| 付録 A  | 各画像の平均適合率, 平均再現率, 平均 F 値 (ORB)  | 25 |
| 付録 B  | 各画像の平均適合率, 平均再現率, 平均 F 値 (SURF) | 27 |

# 目次

|     |                                   |    |
|-----|-----------------------------------|----|
| 2.1 | 注目画素 $p$ のコーナー判定 . . . . .        | 6  |
| 2.2 | 画像ピラミッド . . . . .                 | 8  |
| 3.1 | 提案システムの全体図 . . . . .              | 11 |
| 3.2 | 画像分割までの流れ . . . . .               | 12 |
| 3.3 | VW 選定までの流れ . . . . .              | 13 |
| 3.4 | VW ヒストグラム作成までの流れ . . . . .        | 13 |
| 3.5 | Key-Image の選定の流れ . . . . .        | 14 |
| 3.6 | 索引情報データベースのテーブル . . . . .         | 14 |
| 3.7 | Key-Image 選択画面 . . . . .          | 15 |
| 3.8 | 結果表示画面 . . . . .                  | 15 |
| 4.1 | 検索対象画像 . . . . .                  | 17 |
| 4.2 | 適合率, 再現率, F 値の比較 (全体平均) . . . . . | 18 |
| 4.3 | 特徴抽出の計算時間 . . . . .               | 19 |
| A.1 | ORB の平均適合率 . . . . .              | 25 |
| A.2 | ORB の平均再現率 . . . . .              | 26 |
| A.3 | ORB の平均 F 値 . . . . .             | 26 |
| B.1 | ORB の平均適合率 . . . . .              | 27 |
| B.2 | SURF の平均再現率 . . . . .             | 28 |
| B.3 | SURF の平均 F 値 . . . . .            | 28 |

# 表目次

|                    |    |
|--------------------|----|
| 4.1 実験条件 . . . . . | 16 |
|--------------------|----|

# 第 1 章

## 序論

スマートフォンやタブレット PC のような多機能情報端末の普及，それに伴ったソーシャル・ネットワーキング・サービスの利用者の増加などにより，膨大な数の画像データが我々の周囲にあふれており，多くの画像の中から自分の望む必要な画像を迅速に探す手段として画像検索の研究が行われている．一般的に画像検索は Text-Based Image Retrieval(TBIR) と Content-Based Image Retrieval(CBIR) に大きく分けられる．

大量の画像の中から目的の画像を探す手段としては，画像に予めタグ付けされたキーワードに基いて画像検索を行う TBIR を用いることが一般的である．TBIR を実現するためには，画像に対して予めキーワードを付与しておく必要がある．計算機の高速化・大容量化により一般物体認識の研究が大きく発展し，一般物体認識を用いることで事前に人手によるキーワード付けを行う必要がなくなり，画像や映像に自動的にキーワードを付与することが可能となっている [1]．しかし，索引付けの段階で必ずしも正しいキーワードが画像に付与されていると限らないという問題があり，検索精度に課題を残していると言える [2]．

TBIR とは異なるアプローチとして，検索したい画像を入力することで，画像から抽出される色や形状などの画像特徴量に基いて画像の類似度を定義し，類似している画像を検索するもの CBIR である．CBIR では，画像を入力として扱うため，入力画像が持っている情報と類似した情報を持っている画像を検索する精度が重要となる．しかし，CBIR のような単純な画像特徴量に基づく画像検索では，特徴量が画像の意味内容に直接結びつかないため，見た目の類似性が意味の類似性と一致しないという問題 (セマンティックギャップ) がある．CBIR の研究例は数多く報告されているものの，セマンティックギャップの問題から実用的に使われている例はあまり多くない．



そこで、セマンティックギャップを解決するアプローチとして、画像の一部分の内容に着目する Region-Based Image Retrieval(RBIR) が研究されている。RBIR は、CBIR 手法の一種である。画像を決められた領域ごとに分割し、分割されたそれぞれの領域から特徴を抽出する。抽出した特徴をもとに各領域にインデックスを付与する。これにより 1 枚の画像に対して複数のインデックスが付与でき、セマンティックギャップを減少させつつ検索することが期待されている [3]。

RBIR の精度を向上させるものとして、特徴抽出にアフィン変化や照明変化に頑健である Speeded-Up Robust Features(SURF) 特徴を用いた部分類似画像検索が提案されている [4]。しかし、Scale-Invariant Feature Transform(SIFT) 特徴や SURF 特徴などに代表される局所特徴量は、特徴量を高次元のベクトルで表現しているため、高い識別能力を持っている反面、計算時間やメモリ使用量が多く、類似度計算の速度が遅いと指摘されている [5]。

そこで本研究では、スケール不変と回転不変を持ちながら、SURF 特徴の 10 倍以上、SIFT 特徴の 100 倍以上の速度での計算が可能と言われているバイナリ特徴量 ORB(Oriented FAST and Rotated BRIEF) 特徴 [6] を用いた部分類似画像検索を提案する。また本研究では、類似画像を検索する手法として、RBIR 手法の一種であり、分割画像をビジュアルキーとして用いて検索を行う Visual-Key Image Retrieval(VKIR)[7] を取り扱う。

提案システムの評価は、被験者 10 名に検索目的画像を 20 枚提示し、ORB 特徴を用いた場合と SURF 特徴を用いた場合の 2 種類の画像検索を行い、適合率、再現率、F 値、95%信頼区間を求め比較する。被験者実験の結果、ORB 特徴の適合率の平均は 0.10、再現率の平均は 0.41、F 値は 0.13 となり、SURF 特徴の適合率の平均は 0.12、再現率の平均は 0.46、F 値の平均は 0.14 となっており、SURF 特徴と比べて精度は減少したが、特徴量の計算時間は、ORB 特徴が 2.94 秒、SURF 特徴が 12.00 秒と 76%削減し、高速化できることを確認している。

本論文は以下の構成をとる。第 2 章では基礎知識と関連研究について説明し、第 3 章では ORB 特徴を用いた部分類似画像検索の提案手法について述べる。第 4 章では被験者実験に

よる評価を行い，第 5 章で本研究をまとめ，今後の展望を述べる．

## 第 2 章

# 関連研究

本章では，画像検索に用いる手法，局所特徴量，クラスタリング手法についての説明を行う．

### 2.1 検索手法

画像検索に用いられる手法は，画像に対して予めタグ付けされたキーワードに基いて画像検索を行う手法 (Text-Based Image Retrieval:TBIR) と入力された画像から抽出された特徴量に基いて検索を行う手法 (Content-Based Image Retrieval:CBIR) に大きく分けられる．また CBIR の一種として，画像の一部分から抽出された複数の特徴量に基いて検索を行う手法 (Region-Based Image Retrieval:RBIR) がある．

#### 2.1.1 Text-Based Image Retrieval

Text-Based Image Retrieval(TBIR) は，画像に対して予め対応付けされたテキスト (キーワード) に基いて，入力されたキーワードから検索を行う手法である．TBIR の検索精度は，画像に正しいテキストによるタグ付けがされているかに影響される．

予め画像にテキストを付与する方法は人手によって行うことが主流であったが，近年の機械学習手法の進歩や Web の普及によるデータセット構築の容易化，計算機の高高速化や大容量化などによって一般物体認識の研究が発展し，一般物体認識を用いることで画像に自動的にテキストを付与することが可能となっている．しかし，索引付けの段階で必ずしも正しいキーワードが付与されているとは言えず，検索精度に課題を残している．

## 2.2 画像認識に用いる特徴量

### 2.1.2 Content-Based Image Retrieval

Content-Based Image Retrieval(CBIR) は、クエリとして入力された画像から色や形状などの特徴を抽出し、特徴同士の類似度に基づいて、類似する画像を検索する手法である。CBIR は、画像データに含まれる特徴量を索引とするため、TBIR のように事前に画像にテキストを付与する必要がない。特徴量は画像の色合いや形状情報などをもとに多次元ベクトルで表現され、特徴を数値化することで画像間の類似度を測ることができる。しかしながら、CBIR のような単純な画像特徴量に基づく画像検索では、特徴量が画像の意味内容に直せず結びつかないため、見た目の類似性と意味の類似性が一致しないという問題 (セマンティックギャップ) がある。現在では、一般物体認識技術の発展によって、画像間のセマンティックギャップは減少しつつあると言えるが、実用的に使われている例はあまり多くはない。

### 2.1.3 Region-Based Image Retrieval

Region-Based Image Retrieval(RBIR) は、CBIR の手法の一つである。画像を決められた領域ごとに分割し、分割されたそれぞれの領域から特徴を抽出され、抽出した特徴をもとに各領域にインデックスが付与される。RBIR は、1 枚の画像から複数のインデックスを付与することができるため、CBIR と比べて、セマンティックギャップを減少させ、適用範囲を大きく広げることができると考えられている。

## 2.2 画像認識に用いる特徴量

一般に画像認識に用いられる特徴量は、大域特徴と局所特徴の 2 つに分けられる。大域特徴量とは、画像の全体から抽出される特徴量であり、代表的な例として CBIR の色ヒストグラムがあげられる。一方、局所特徴とは、画像の一部分から抽出される特徴量であり、代表的な例として、SIFT 特徴や SURF 特徴などがあげられる。

大域特徴は、抽出される特徴量が画像 1 枚に対して 1 つであるため、画像の索引付けに用いる特徴量としてデータ量を大幅に少なくすることができる。一方で、画像の一部に隠れが

## 2.2 画像認識に用いる特徴量

おきると特徴抽出の際に同じ値を得ることが不可能となる問題もある [8]. 局所特徴は、抽出される特徴量が画像に対して複数であるため、隠れていない部分から同じ特徴量を得ることができ、変化に強いという利点がある.

### 2.2.1 Features from Accelerated Segment Test(FAST)

Features from Accelerated Segment Test(FAST) は、コーナーのみを特徴点の対象とし、決定木により高速かつ効率的に検出する特徴点検出手法である [11]. FAST によるコーナー検出は、図 2.1 のように注目画素  $p$  の周囲 16 箇所の画素を観測する. 16 個の画素のうち連続して  $n$  個以上の画素値が  $p$  と比較して、予め定めた閾値  $t$  以上明るい、もしくは暗い時に  $p$  をコーナーとして判定する.

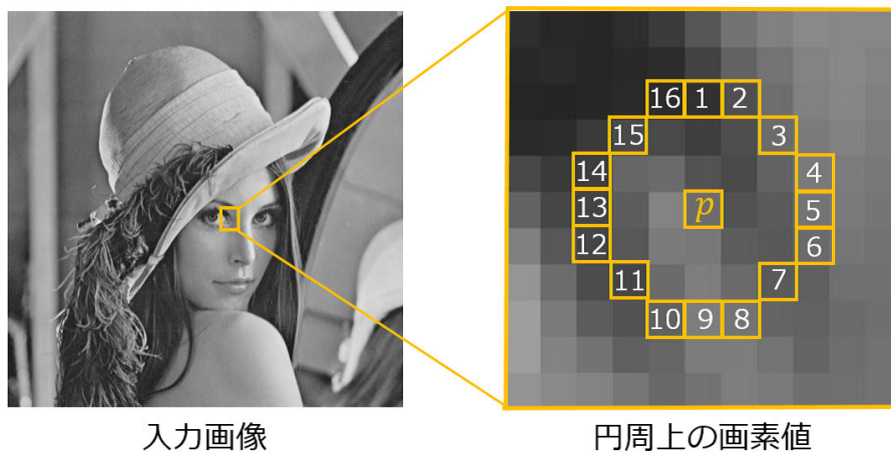


図 2.1 注目画素  $p$  のコーナー判定

FAST は、再現性の高い特徴点を高速かつ効率的に検出するために決定木を用いてコーナー検出を行う. まず、 $p$  の周囲 16 箇所の画素を明るい (brighter), 類似 (similar), 暗い (darker) の 3 値に分類する. 基準となる式を以下に示す.

$$S_{p \rightarrow x} = \begin{cases} \text{brighter} & I_p + t \leq I_{p \rightarrow x} \\ \text{similar} & I_p - t < I_{p \rightarrow x} < I_p + t \\ \text{darker} & I_{p \rightarrow x} \leq I_p - t \end{cases} \quad (2.1)$$

## 2.2 画像認識に用いる特徴量

$I_p$  は注目画素の輝度値,  $x$  は円上の画素の位置,  $I_{p \rightarrow x}$  は円周上の画素の輝度値,  $t$  は閾値を表している. 3 値に分類した 16 個の画素を特徴ベクトルとし, 画素値が連続して  $n$  個以上が brighter, もしくは darker の時にコーナー  $c$  とし, 条件を満たさない場合非コーナーとなる. 決定木は ID3 アルゴリズムに従う決定木を構築する. コーナーと非コーナーを最も分類できる円上の画素を分岐ノードとして選択する処理を再帰的に行うことで全体の木構造を得る.

多くの特徴点検出手法では, 特徴点らしさを表すレスポンス値の局所最大値を抽出することで, 隣接して検出される特徴点の中で有効な画素のみを検出する. しかし, FAST による決定木を用いたコーナー検出アルゴリズムはレスポンス値を出力しないため, 以下の式よりレスポンス値  $V$  を算出する. 隣接して検出されたコーナー点の中で, 最もレスポンス値  $V$  が高い画素をコーナー点として得ることで, 互いに離れたコーナー点を取得できる.

$$V = \max\left(\sum_{x: S_{p \rightarrow x} = \text{brighter}} |I_{p \rightarrow x} - I_p| - t, \sum_{x: S_{p \rightarrow x} = \text{darker}} |I_p - I_{p \rightarrow x}| - t\right) \quad (2.2)$$

### 2.2.2 Binary Robust Independent Elementary Features(BRIEF)

Binary Robust Independent Elementary Features(BRIEF) は, 特徴点周辺領域 (パッチ) 内においてランダムに選択された 2 点の画素値の比較によってバイナリコードを生成する特徴記述手法である. ノイズに対する耐性を高めるために, あらかじめパッチをガウシアンフィルタによって平滑化する. 以下にバイナリ列を生成するための式を示す. パッチ  $p$  上の  $n_d$  個のバイナリテスト  $\tau$  は, 点  $x$  と点  $y$  の画素地の比較によって  $n_d$  個のバイナリを生成する.

$$\tau(p; x, y) = \begin{cases} 1 & \text{if } p(x) < p(y) \\ 2 & \text{otherwise} \end{cases} \quad (2.3)$$

## 2.2 画像認識に用いる特徴量

BRIEF では，生成された  $n_d$  個のバイナリをシフトさせることで， $n_d$  次元のバイナリ列を生成する．

$$f_{n_d}(p) := \sum_{1 \leq i \leq n_d} 2^{i-1} \tau(p; x_i, y_i) \quad (2.4)$$

### 2.2.3 Oriented FAST and Rotated BRIEF(ORB)

Oriented FAST and Rotated BRIEF(ORB) は，2 点の画素値の比較によってバイナリコードを生成するという BRIEF の考え方にスケール不変性と回転不変性を導入したものである．

#### 特徴点検出

ORB の特徴点検出には FAST を用いる．また，スケール不変性を導入するために，図 2.2 のような入力画像を多段階に縮小した画像ピラミッドを作成する．それぞれのレイヤに対して FAST を適用して全ての画素に関して特徴点らしさのレスポンス値を算出する．レスポンス値が閾値以上であり，かつ  $x$ ,  $y$  方向およびスケール方向の 3 次元空間において局所最大値となる点を特徴点として抽出する．

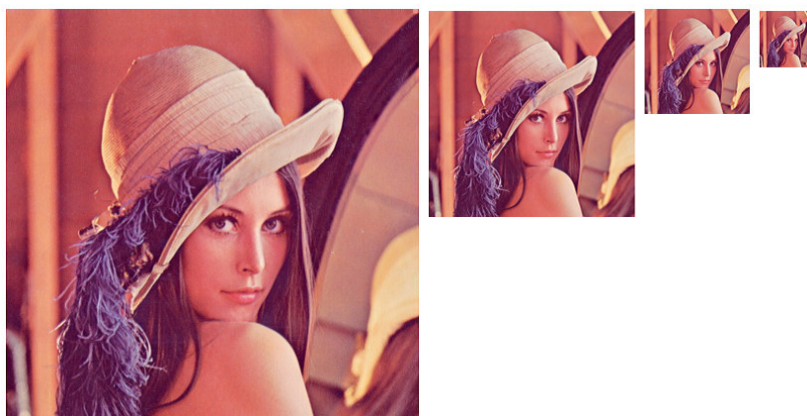


図 2.2 画像ピラミッド

## 2.3 クラスタリング手法

### 特徴量記述

ORB の特徴量記述は、BRIEF と同様に 2 点の画素値の比較によってバイナリコードを生成する。BRIEF では画素値の比較を行う 2 点のペアをランダムに選択しているが、ORB では学習によって選択する。良いペアの選択基準として、ビット分散が大きくなるペアは特徴量の記述能力が高く、選択された  $N$  組のペア同士の相関が低いほど無駄のないコードが生成できることが定義されている。そこで、ORB はパッチ内であり得る全てのペア候補を列挙し、ビット分散が大きく、かつペア同士の相関が低くなるようなペアを貪欲法で選択して絞り込むことで記述能力の高いバイナリコードを生成する。

ORB では、オリエンテーション推定に勾配を用いず、パッチの画素値に関する 0, 1 次モーメントから重心を算出し、特徴点中心と重心を結ぶ直線の傾きをオリエンテーション  $\alpha$  として用いる。バイナリコードを生成する際に、画素値を比較する 2 点のペアの座標位置を  $\alpha$  だけ回転することにより回転不変性を得ている。

## 2.3 クラスタリング手法

### 2.3.1 K 平均法 (K-means clustering)

K-means 法は非階層的クラスタリングの一種であり、適切なクラスタ中心を求めて分類するという簡単なアルゴリズムであるため、他のクラスタリングアルゴリズムと比べ計算コストが小さいことからよく用いられる手法である [12][13]。K-means 法は、中心となるベクトル (セントロイド) をクラスタの代表点とし、各データと各クラスタ中心との距離を求め、データ点を最も近い中心のクラスタに割り当て直し、全てのデータのクラスタ割り当てに変化がなくなるまで処理を繰り返すことで最適解の探索を行う。以下のアルゴリズムを示す。

1. 任意の  $k$  個のクラスタ中心  $c_i$  をランダムに選択する。
2. 全てのデータを各データ点  $x_j (j = 1, \dots, n)$  から最も近いクラスタ  $i$  に割り当てる。



## 2.4 Bag-of-Visual-Word

3. 割り振ったデータをもとに各クラスタの中心を以下の式にしたがって求める． $C_i$  は各クラスタ  $i$  に含まれるデータの集合であり， $|C_i|$  に含まれるデータ数である．

$$c_i = \frac{1}{|C_i|} \sum_{x_j \in C_i} x_j \quad (2.5)$$

4. クラスタに変化がなくなるまで 2, 3 の処理を繰り返す．

### 2.3.2 Ward 法

Ward 法は階層的クラスタリングの一種であり，クラスターとしてサンプルをまとめる際に生じる各サンプルの情報の損失量の増加分をクラスター距離とする手法である．全てのクラスター内の平方和の増加量を最小になるよう組み合わせていくため，クラスタリング感度が良く，階層的クラスタリングの中でよく用いられる手法である．しかし，データ数  $N$  個に対する計算量は  $O(N^2)$  であるため，大量のデータ解析には適していない．

## 2.4 Bag-of-Visual-Word

Bag-of-Visual-Word(BoVW) とは，文書分類手法である Bag-of-Words の考え方を画像に適用した手法である [14]．BoVW では局所特徴の特徴量を量子化し，その出現頻度で画像表現を行う．そのため，BoVW 表現を用いることで画像を 1 つの特徴ベクトルとして表現することができる．処理の流れを以下に示す．

1. 対象画像から局所特徴を抽出する．
2. 抽出した局所特徴を  $K$  個のクラスタにクラスタリングする．
3.  $K$  個のクラスタの各セントロイドを Visual Word(VW) として選定する．
4. 画像中の各局所特徴量について一番近い VW を検索し，その VW に投票することで全画像を VW を次元とするヒストグラムに変換する．

## 第 3 章

# ORB 特徴を用いた部分類似画像検索

本章では、提案する ORB 特徴を用いた部分類似画像検索システムについて説明する。システムの全体図を図 3.1 に示す。この検索システムは、仮想化ソフトウェア VirtualBox 上で動作する CentOS 上に構築し、データベース SQLite に画像の索引情報を格納し、Apache HTTP サービスから呼び出される検索を行う。ユーザは Web ブラウザから Web サーバ上の本システムにアクセスし、提示された KI を選択して画像検索を行う。検索システムは索引情報データベースを参照して KI を含む画像を結果として出力する。

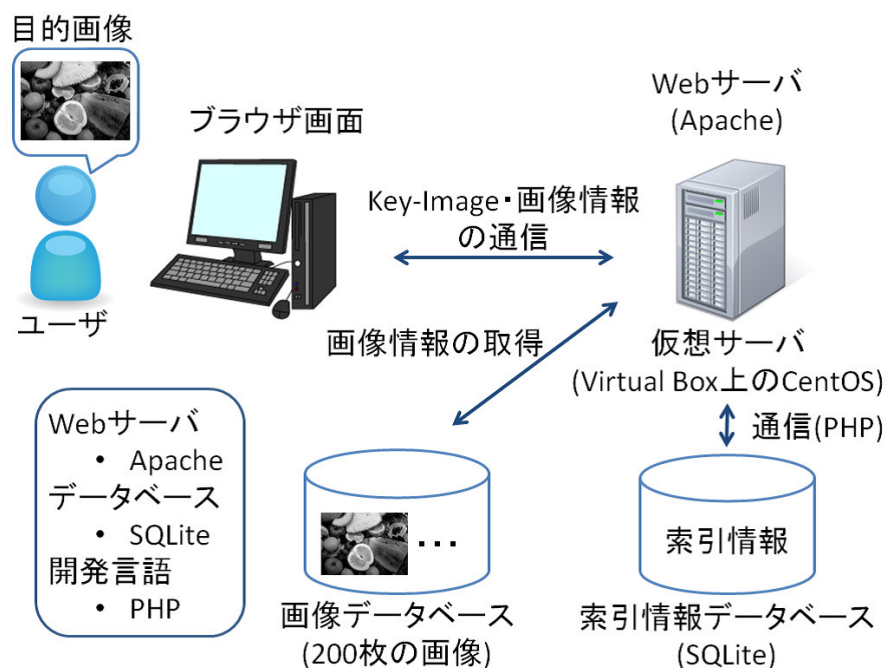


図 3.1 提案システムの全体図

### 3.1 Key-Image の作成

## 3.1 Key-Image の作成

本節では、ORB 特徴を用いた部分類似画像検索のデータベース構築について説明する。画像は、ArtExplision の 10 カテゴリから 20 枚ずつ抽出した 200 枚の画像を用いる。ただし、ORB 特徴は色特徴を使用せず特徴量を抽出するため、ユーザが KI 選択の際に色によって判断し、選択を行わないように全画像をグレースケールに変換する。下記にデータベース構築の流れを示す。

1.  $3 \times 3$  に分割された全画像から ORB 特徴を抽出する。
2. 抽出された ORB 特徴を K-means 法を用いてクラスタリングする。
3. 各クラスターの最も平均値に近い画像を VW として選定する。
4. 選定された VW をもとに VW ヒストグラムを作成する。
5. VW ヒストグラムのクラスタリングを行い KI を選定する。
6. データベースを構築する。

### 3.1.1 画像の分割

画像分割の流れを図 3.2 に示す。検索対象となる全ての画像はそれぞれ高さや幅が異なるため、全て同じ大きさに揃える必要がある。そのため、全ての画像を  $300 \times 300$  ピクセルのスケールに変更した後、 $3 \times 3$  に画像を分割する。

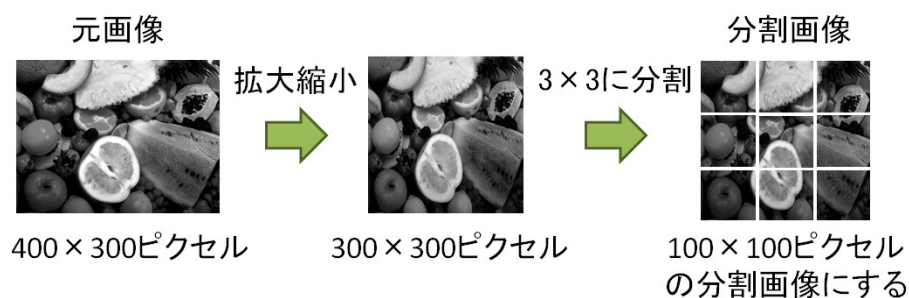


図 3.2 画像分割までの流れ

### 3.1 Key-Image の作成

#### 3.1.2 Visual Word の選定

VW の選定までの流れを図 3.3 に示す．まず，全分割画像から抽出された ORB 特徴量を K-means 法でクラスタリングして 10000 個のクラスタを生成する．この時，特徴量を持たない画像は意味を持たない画像として取り除く．次に，分類された各クラスタの平均値を算出し，各クラスタの各重心に最も近い値を持つ画像を VW として選定する．

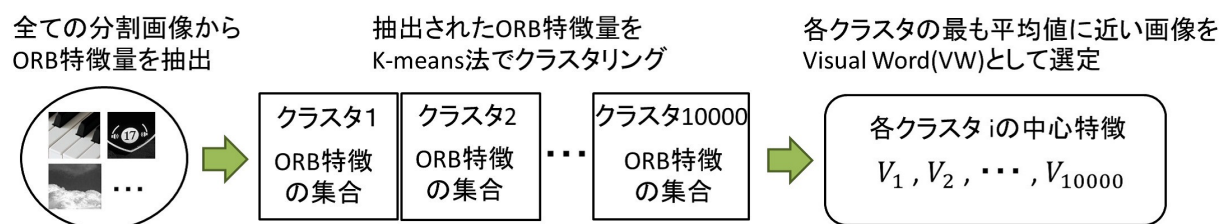


図 3.3 VW 選定までの流れ

#### 3.1.3 VW ヒストグラムの作成

VW ヒストグラムの作成の流れを図 3.4 に示す．部分画像中に含まれる各特徴量について各 VW とのユークリッド距離を算出することで一番近い VW を検索し，その VW に投票していくことでヒストグラムとして表現する．ヒストグラムは，VW を次元とするため，各特徴量を 10000 次元のベクトルとして表現する．

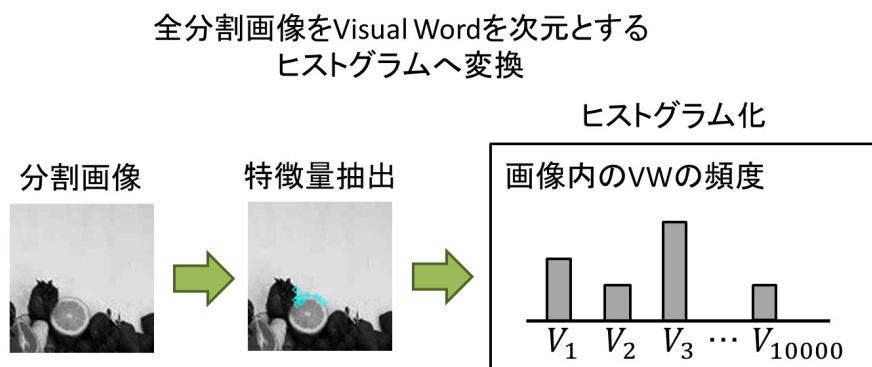


図 3.4 VW ヒストグラム作成までの流れ

## 3.1 Key-Image の作成

### 3.1.4 Key-Image の選定

Key-Image の選定について図 3.5 に示す．作成した全ての VW ヒストグラムを特徴ベクトルとして，Ward 法でクラスタリングし，デンドログラムを作成する．作成したデンドログラムを 20 個に分割し，分割された各クラスターの各重心に最も近い値を持つ画像を KI として選定する．選定した KI をもとにデータベースを構築する．図 3.6 に索引データベースの一部を示す．索引情報データベースには，ID，元画像，部分画像，部分画像が属するクラス番号を格納する．

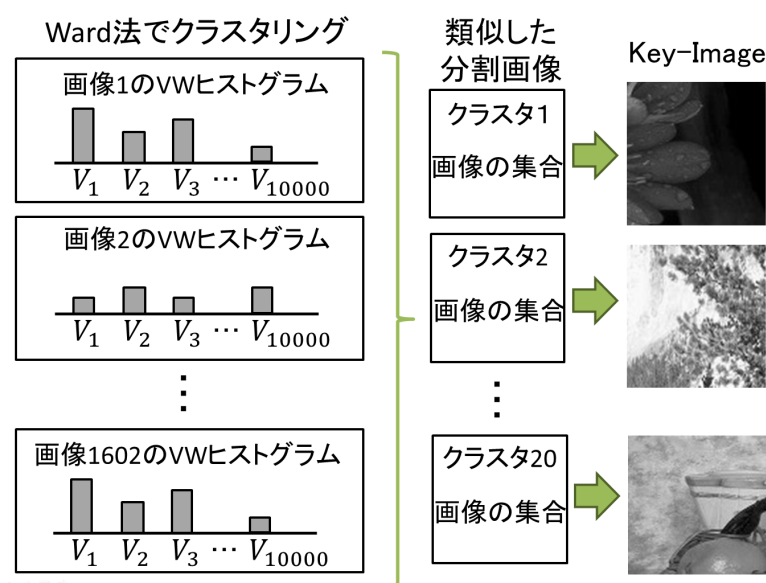


図 3.5 Key-Image の選定の流れ

|      | id   | img           | block     | class |
|------|------|---------------|-----------|-------|
| 1    | 1    | image_140.jpg | 14005.jpg | 1     |
| 2    | 2    | image_102.jpg | 10205.jpg | 1     |
| 3    | 3    | image_058.jpg | 05804.jpg | 2     |
| ...  |      |               |           |       |
| 1599 | 1599 | image_110.jpg | 11001.jpg | 1     |
| 1600 | 1600 | image_071.jpg | 07107.jpg | 4     |
| 1601 | 1601 | image_053.jpg | 05307.jpg | 2     |
| 1602 | 1602 | image_096.jpg | 09609.jpg | 1     |

図 3.6 索引情報データベースのテーブル

## 3.1 Key-Image の作成

### 3.1.5 検索結果の表示

ユーザは図 3.7 の画面に提示されている 20 枚の Key-Image から目的画像のイメージに類似するものを選択することで検索を行う。クエリとなる Key-Image を選択した後、「Retrieve」ボタンを押すことで結果を図 3.8 の画面に表示する。

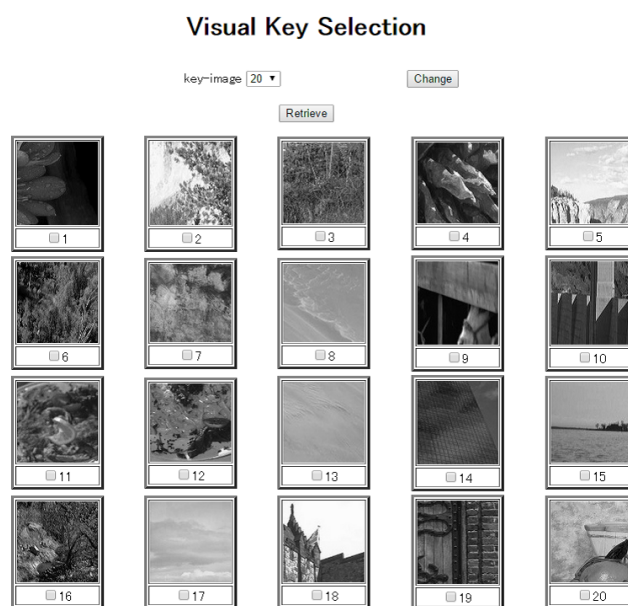


図 3.7 Key-Image 選択画面

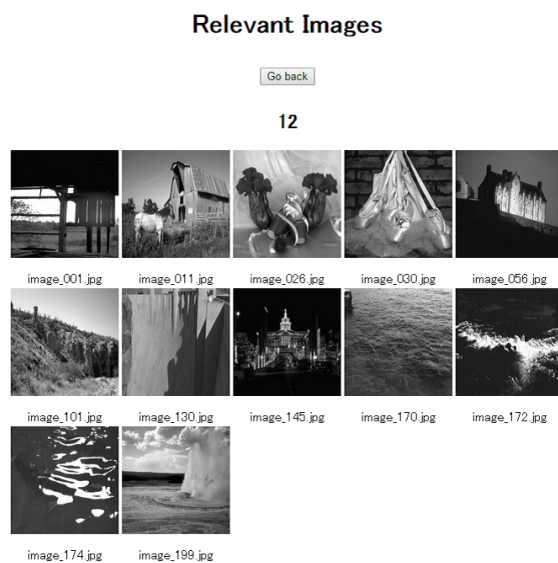


図 3.8 結果表示画面

## 第 4 章

# 構築システムの評価

### 4.1 実験条件

実験条件を表 4.1, 検索目的画像を図 4.1 に示す. 構築したシステムの評価は被験者実験にて行う. 評価は, 被験者 10 名に検索目的画像を 20 枚提示し, 20 枚の KI を用いて画像の検索を行う. 選択する KI の数は最低 1 個, 最大数は限定せず, ORB 特徴を用いた場合と SURF 特徴の特徴を用いた場合の 2 種類の実験を行う.

画像データベースには, ArtExption の 10 カテゴリから 20 枚ずつ選定した 200 枚の画像を用いる. 検索対象画像は 10 カテゴリから 2 枚ずつ選定した 20 枚の画像を用いる. また, 実験に用いている ORB 特徴と SURF 特徴は, 色情報ではなく輝度情報を用いて特徴量を抽出しているため, ユーザが実験の際に色情報を考慮することのないように Key-Image と目的画像をグレースケールで提示する.

表 4.1 実験条件

|             | ORB 特徴                                    | SURF 特徴 |
|-------------|---|---------|
| 被験者         | 10 名                                      |         |
| 画像データ       | ArtExplosion の 10 カテゴリから 20 枚ずつ選定した 200 枚 |         |
| 目的画像        | 20 枚 (グレースケール)                            |         |
| Key-Image 数 | 20 枚 (グレースケール)                            |         |
| 適合画像        | 被験者が類似すると判断した画像                           |         |

## 4.2 評価項目

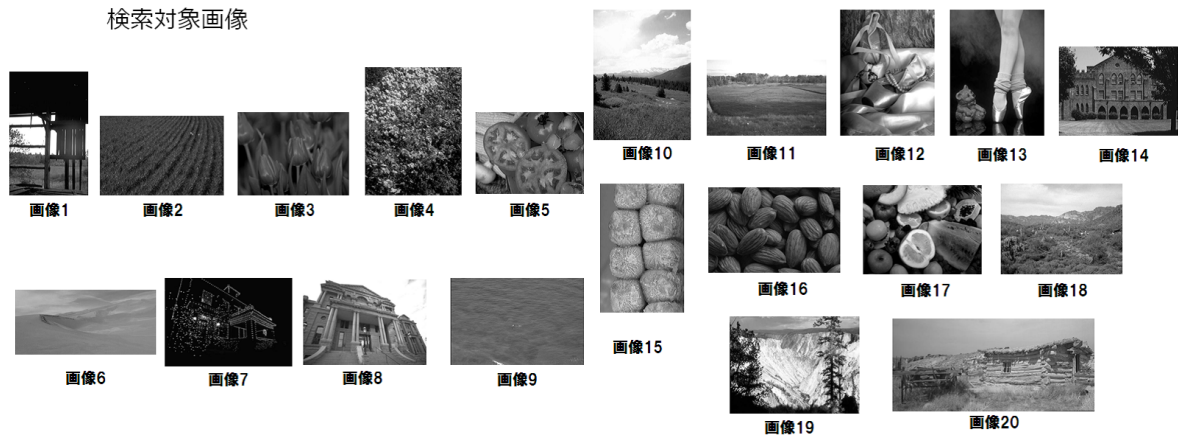


図 4.1 検索対象画像

## 4.2 評価項目

構築したシステムの評価として、被験者ごとの各条件での実験結果から適合率，再現率，F 値を算出する．また，特徴の評価として，特徴抽出にかかる計算時間を算出する．適合率，再現率，F 値は以下の式を用いて算出する．再現率の全画像中の適合画像枚数は，各試行ごとにユーザーが全 200 枚の画像から適切だと判断した枚数を用いる．適合率は正確性に関する指標，再現率は網羅性に関する指標であり，F 値は正確性と網羅性の総合的な評価に用いられる尺度である．動作環境は，メモリ 2GB，CPU に Intel core i3 530(2.93GHz)，OS は Ubuntu14.10，ライブラリとして OpenCV2.4.9 を使用する．

$$\text{適合率} = \frac{\text{検索結果に含まれる適合画像の枚数}}{\text{検索結果の画像枚数}} \quad (4.1)$$

$$\text{再現率} = \frac{\text{検索結果に含まれる適合画像の枚数}}{\text{全画像中の適合画像枚数}} \quad (4.2)$$

$$F \text{ 値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}} \quad (4.3)$$



## 4.3 実験結果

図 4.2 に ORB 特徴, SURF 特徴を用いた実験結果の適合率, 再現率, F 値の平均および 95%信頼区間の幅を示す. また, 図 4.3 に特徴抽出の計算時間の平均を示す. 特徴抽出時間はインプットとアウトプット時間を含まず, 特徴点検出と特徴量記述時間を合わせた時間を測定した. ORB は適合率 10%, 再現率 41%, F 値 0.13, 特徴抽出時間は 2.94 秒となった. SURF は適合率 12%, 再現率 46%, F 値 0.14, 特徴抽出時間は 12 秒となった. ORB 特徴を用いた結果は, SURF 特徴を用いた結果と比べて, 適合率, 再現率, F 値の項目について減少傾向が見られるが, 特徴抽出計算の時間に関しては 4.08 倍の高速化を確認した.

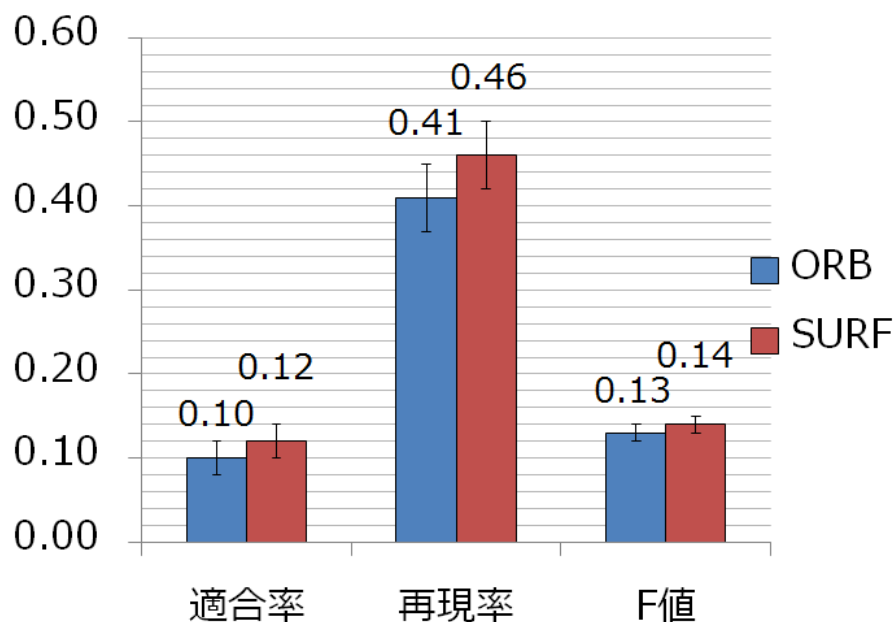


図 4.2 適合率, 再現率, F 値の比較 (全体平均)

## 4.4 考察

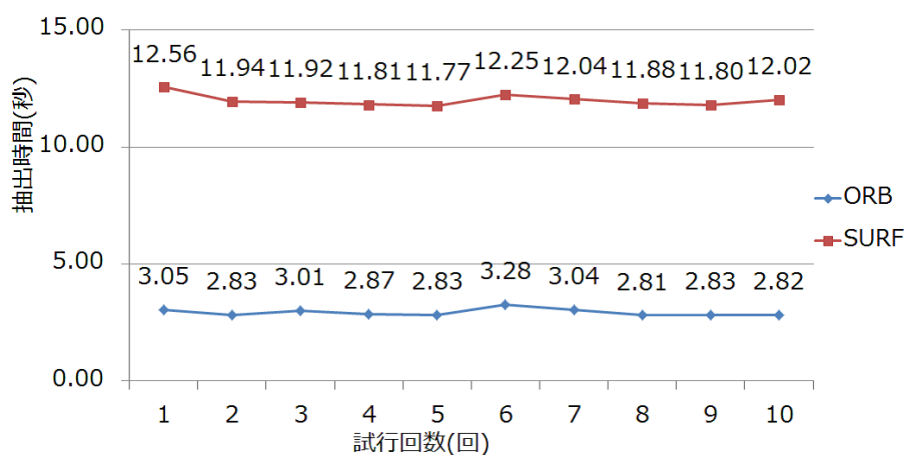


図 4.3 特徴抽出の計算時間

## 4.4 考察

ORB 特徴と SURF 特徴の特徴抽出にかかる計算時間は約 4 倍高速であることを示している。これは、解像度  $640 \times 480$  でフレームレート 30fps の 10 分の動画から特徴を抽出する場合、SURF 特徴で約 60 分かかる処理時間を ORB 特徴ならば約 14 分で処理することができる。また、画像枚数 140 万枚の大規模データセット「ImageNet」の全画像を  $300 \times 300$  ピクセルにスケール変化させたものから特徴を抽出する場合、SURF 特徴では約 23 時間かかる処理時間を約 5.6 時間で処理することができる。今回の結果から、ORB 特徴はフリーの特徴でありながら、リアルタイム性に富んだ高速な処理を可能とする特徴であり、大規模なデータベースに対適用することの有効性を示せていると考える。近年、フル HD や 4K 解像度など高解像度な規格が登場しており、今後このような画像や動画を対象に特徴抽出を行う場合、高速な特徴抽出を行うことができる ORB 特徴は有効であると考ええる。

## 第 5 章

### まとめ

本研究では，局所特徴の一つであり，バイナリベースの特徴量である ORB 特徴を用いた部分類似画像検索を提案し，先行研究である SURF 特徴を用いた部分類似画像検索との比較を行った．ユーザにクエリとして提示する Key-Image の生成するために，2 回のクラスタリング (Visual Word の選定と Key-Image の選定) を行った．

構築したシステムの評価は被験者実験で行い，被験者 10 名に ORB 特徴を用いた提案手法と SURF 特徴を用いた従来手法の 2 つの場合において，固定された 20 枚の KI を用いて目的画像 20 枚を検索させ，各条件での結果から適合率，再現率，F 値を求めた．また，特徴量の評価として特徴抽出にかかった時間を算出した．実験結果は，SURF 特徴を用いた従来手法と比較して，適合率が-2 ポイント，再現率-5 ポイント，F 値が-1 ポイントと減少傾向が見られたが，特徴抽出にかかった計算時間は，SURF 特徴に比べて 4.08 倍高速であることを確認した．今回の実験結果から ORB 特徴はリアルタイム性に優れた高速な特徴量であることを示し，近年登場している高解像度の画像や動画，大規模な画像データセットに対して ORB 特徴を用いることの有効性を示すことができたと考える．

今後は，Key-Image を可変にした場合との比較やシステム評価について有意差を出すために画像のサンプル数と被験者数を増やして実験を行うことなどが考えられる．

# 謝辞

本研究を行うにあたり、丁寧なご指導をして頂いた高知工科大学情報学群 吉田真一准教授に心から感謝いたします。吉田准教授には、本論文を作成するにあたり丁寧かつ熱心なご指導を賜りました。また研究内容にとどまらず、発表やプレゼンテーションの効果的な作り方など多くのことを学ばせて頂きました。また、本研究の副査を引き受けて頂いた高知工科大学情報学群 福本昌弘教授と高知工科大学情報学群 栗原徹准教授には研究内容に対して的確なご指南を頂き、より研究に対する理解を深めることができました。深く感謝いたします。高知工科大学情報学群 横山和俊教授と横山研究室のメンバーには研究内容以外のことで刺激と癒やしを頂き、精神的にも支えられました。ありがとうございます。

また、本研究の趣旨を理解し、貴重な時間を割いて快く協力して頂いた同研究室の皆様、および被験者を引き受けて下さった皆様に深く感謝いたします。本当にありがとうございます。

同研究室の皆様には感謝の念が絶えません。今日まで研究活動が続けてこられたのも、ひとえに皆様の支えがあってこそだと感じております。奥山氏とは、研究内容が近いことから研究を進めるにあたり、局所特徴やその詳細についての議論を通して、多くの知識や示唆を頂きました。あなたの A 棟の 1 階から 5 階まで響き渡るような活気溢れる声と自爆を恐れず何事にも積極的に取り組む姿勢には本当に元気付けられました。奥村氏とは、日常会話を通してささやかな癒やしを提供して頂きました。また、あなたの丁寧かつ迅速な仕事の進め方や普段の姿からは考えられない腹黒さを見せる意外性など見習いたいと思う点が多くありました。なお、今でも奥山氏と奥村氏の名前を頻繁に間違えてしまうことをここに深くお詫び申し上げます。中路氏とは、研究内容以外の点で様々な討論を交わし、よく笑いを提供して頂きました。研究のことで行き詰り心身とも参っていた時、あなたのさっぱりとした考え方や前向きな姿勢には何度も助けられました。山本氏とは、よくわからないテンションで一緒に盛り上がっていたことを覚えています。あなたが持つ独特なペースに多々困らせられる

## 謝辞

こともありましたが、何よりもあなたの物事に対する直向きさや鋭い着眼点にはいつも感服させられていました。高橋氏とは、机が隣なことから研究内容や発表資料などについて積極的に意見を交わし、完成度を高め合いました。お互い笑いのツボが浅いことから些細な会話でよく盛り上がったと思います。江木氏とは、よく深夜遅くまで研究内容について議論を交わしました。あなたの常に限界寸前を歩く姿勢は、スケジュール管理や睡眠の重要性を心底感じさせてくれました。

修士課程のお三方には、吉田准教授と同様に本論文を作成するにあたり厳しくも温かいご指導を賜りました。また、同研究室の3年生の皆様には研究室活動やイベントなどでお世話になりました。

最後に、大学4年間で様々な形で支えてくださった両親に心から感謝します。

## 参考文献

- [1] 柳井啓示, “[招待講演] 画像・映像の認識と意味的検索,” 電子情報通信学会技術研究報告:信学技法, pp.25–30, 2010.
- [2] 竹内勤治, 黄瀬浩一, “Earth Mover’s Distance に基づく Text-Based Image Retrieval,” 情報処理学会研究報告, NL-177-5, pp.33–38, 2007.
- [3] Feng Jing, Mingjing Li, Hong-Jiang Zhang, Bo Zhang, “An Efficient and Effective Region-Based Image Retrieval Framework,” IEEE TRANSACTIONS ON IMAGE PROCESSING, Vol.13, pp.699–709, 2004.
- [4] 上村香樹, “SURF 特徴を用いた部分画像に基づく類似画像検索,” 平成 25 年度高知工科大学学士学位論文, 2014
- [5] 安倍満, 吉田悠一, “高速かつメモリ消費量の少ない局所特徴量 CARD:Compact And Real-time Descriptors,” デンソーテクニカルレビュー, Vol.18, pp.66-76, 2013.
- [6] 望月宏史, 柳井啓司, “物体認識技術を用いたモバイル物品管理システム,” DEIM Forum, 2012.
- [7] 岡本一志, “ビジュアルキー型画像検索に適した DCT 領域における特徴抽出,” 平成 19 年度高知工科大学修士学位論文, 2008.
- [8] 黄瀬浩一, 岩村雅一, 中居友弘, 野口和人, “局所特徴量のハッシングによる大規模画像検索,” DBSJ Journal, Vol.8, 2009.
- [9] 藤吉弘宣, 安倍満, “局所勾配特徴抽出技術 -SIFT 以降のアプローチ-,” 精密工学会誌, Vol.77, No.12, pp.1109–1116, 2011.
- [10] 馬場口登, “マルチメディア検索の技術動向,” 映像メディア学会誌, Vol.64, No.1, pp.58–63, 2010.
- [11] Edward Rosten, Tom Drummond, “Machine learning for high-speed corner detection,” European Conference on Computer Vision, pp.430–443, 2006.

## 参考文献

- [12] 小野田崇, 坂井美帆, 山田誠二, “初期値設定法の違いによる k-means 法の性能比較,” 第 27 回ファジィシステムシンポジウム, pp.231–236, 2011.
- [13] 倉橋和子, 森井藤樹, “分割・併合機能を有する K-Means アルゴリズムによるクラスタリング,” 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解, Vol.106, No.470, pp.67–71, 2007.
- [14] 内田祐介, 酒澤茂之, “大規模特定物体認識の最新動向,” 電子情報通信学会誌, Vol.96, No.3, pp.207–213, 2013.

## 付録 A

# 各画像の平均適合率， 平均再現率， 平均 F 値 (ORB)

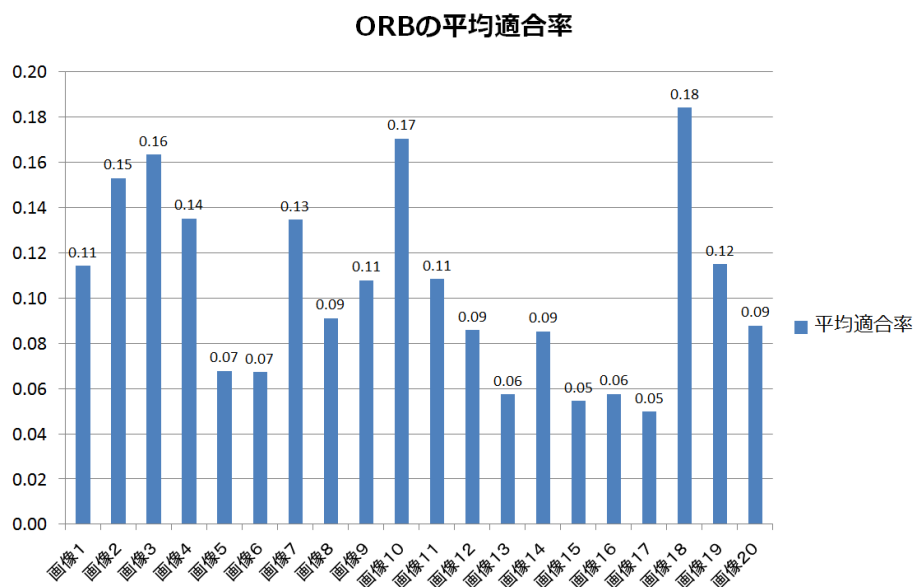


図 A.1 ORB の平均適合率



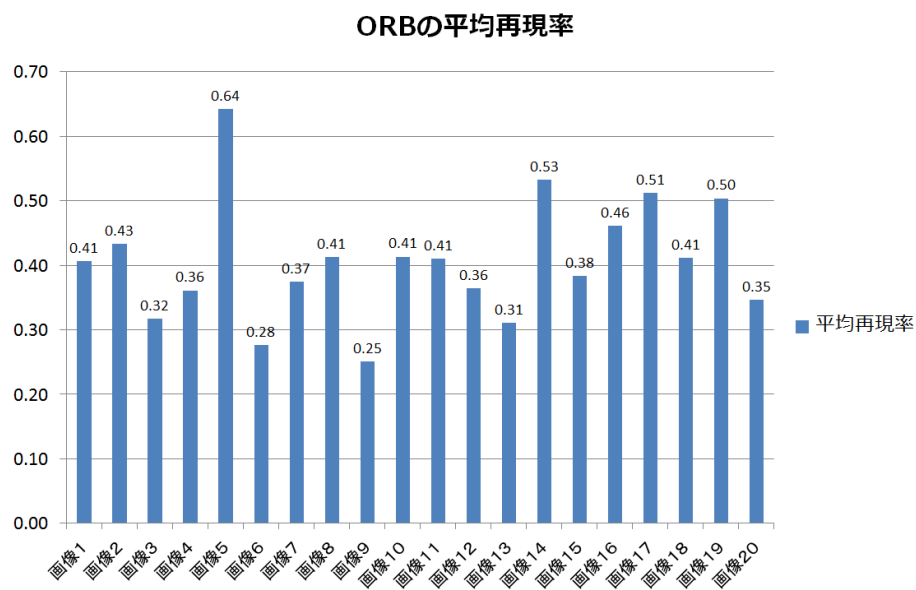


図 A.2 ORB の平均再現率

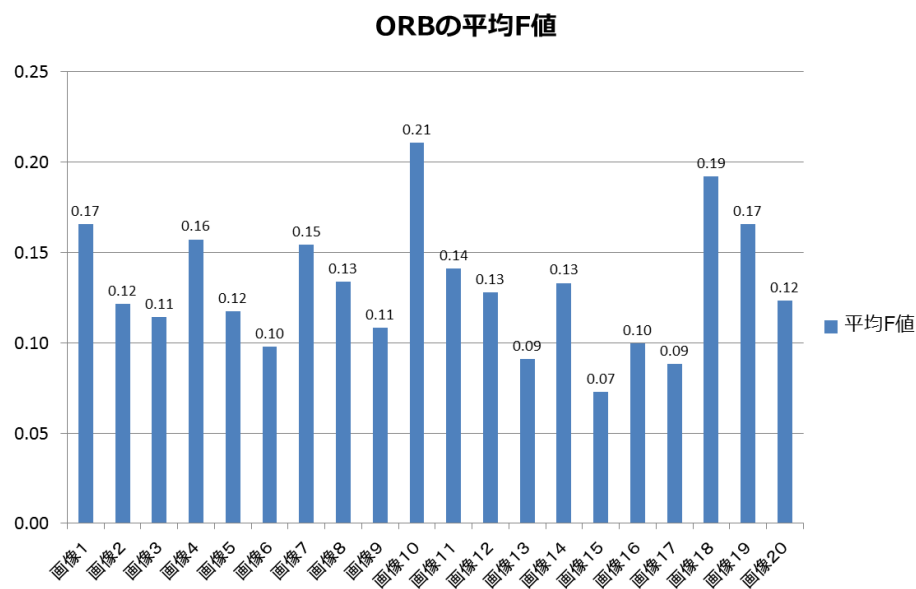


図 A.3 ORB の平均 F 値

## 付録 B

### 各画像の平均適合率， 平均再現率， 平均 F 値 (SURF)

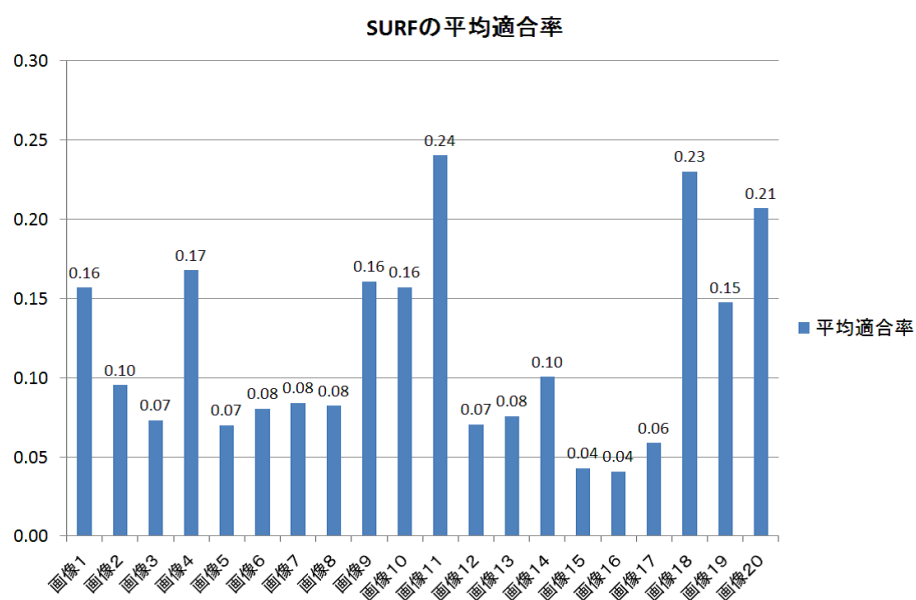


図 B.1 ORB の平均適合率

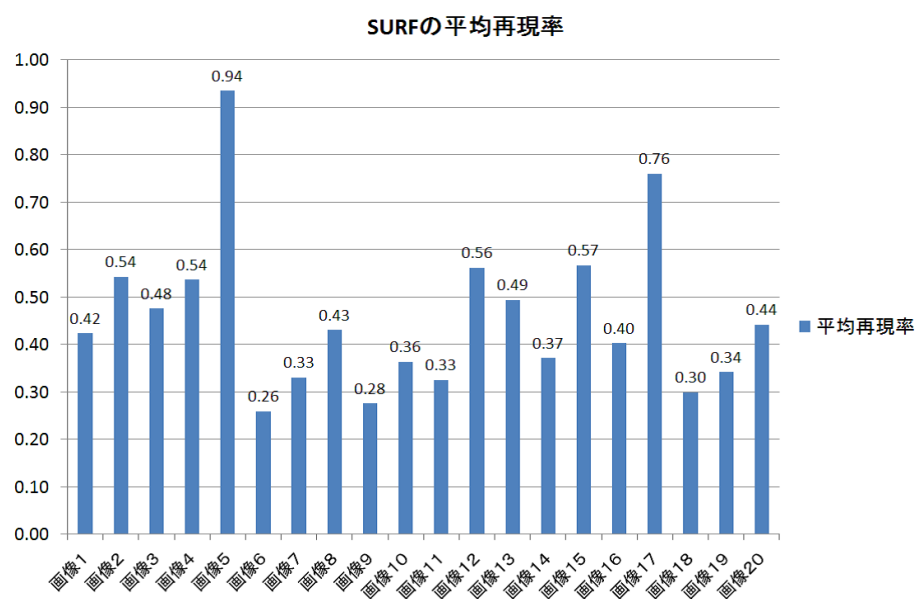


図 B.2 SURF の平均再現率

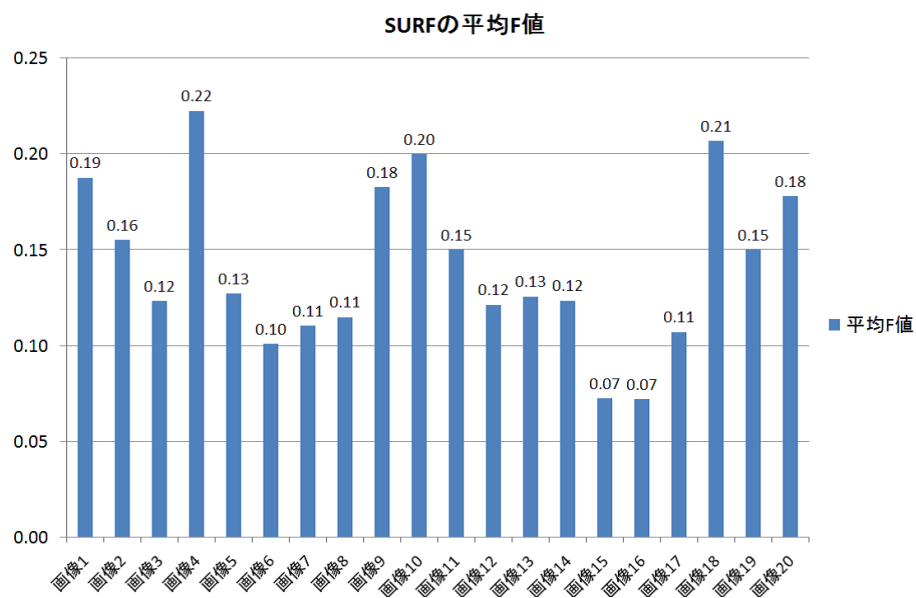


図 B.3 SURF の平均 F 値