

平成 25 年度
学士学位論文

**SURF 特徴を用いた部分画像に基づく
類似画像検索**

**Region-Based Image Retrieval using Speeded Up
Robust Features as a Shape Feature**

1140319 上村 香樹

指導教員 吉田 真一

2014 年 2 月 28 日

高知工科大学 情報学群

要 旨

SURF 特徴を用いた部分画像に基づく類似画像検索

上村 香樹

近年，局所特徴を用いた研究が物体認識の分野において主流となっている．また，類似画像検索の分野においても Bag of Visual Word (BoVW) など局所特徴が使われている．従来の画像の一部分の類似性に着目した部分類似画像検索 (Visual Key 型画像検索など) においては，これまで大域特徴である色特徴を中心に用いてきた．また，それと同時に，エッジを用いた形状特徴やテクスチャ特徴を用いるものもあるが，精度の向上にはあまり貢献せず，精度の低下を招く場合もあった．そこで，本研究では，局所特徴の一つであり，アフィン変化や照明変化に頑健である SURF 特徴を用いた部分類似画像検索を提案する．また，岡本や橋詰らにより，ユーザに提示されるクエリとして用いられるキー画像の数は，ユーザや対象画像，あるいは画像データベースなどによって，その適切な数は変化するため，クエリに用いるキー画像数を可変にするシステムの提案，および HTML5 を用いてキー画像数を動的に変化させることができ，かつユーザビリティの高いインターフェースの実装を行う．評価実験では，10 名の被験者による 200 枚の画像での検索実験を行う．検索結果から，適合率・再現率・F 値を求め，従来の Visual Key 型画像検索とキー画像数を 20 枚に固定した提案法，さらにキー画像数をユーザの自由に变化させる場合の 3 つについて比較・評価する．評価実験の結果から，従来手法の平均適合率・平均再現率・平均 F 値はそれぞれ，11%，30%，0.16，であったのに対し，キー画像 20 枚の提案手法はそれぞれ 15%，33%，0.20，キー画像数可変の方法はそれぞれ，21%，21%，0.18 であり，検索精度の向上を確認している．

キーワード 部分画像に基づく画像検索，SURF 特徴，Bag of Visual Word，HTML5

Abstract

Region-Based Image Retrieval using Speeded Up Robust Features as a Shape Feature

Kazuki KAMIMURA

Recently, local image features are widely used in the research area of image recognition. Local image features are also used in the area of content-based image retrieval, for example, a technique of bag of visual word (BoVW). Region-based image retrieval (RBIR) has been studied to retrieve partly similar images. In RBIR, image features are calculated for each sub-region of image. In conventional RBIR, global features, for example color features, has been mainly used. Shape features and texture features are also used. Conventional RBIR uses Sobel edge feature as a shape feature. The accuracy of RBIR, however, is difficult to be improved in conventional research. Conventional study has pointed out that Sobel feature is are of the reasons of low accuracy. In this study, we propose a region-based image retrieval with scale-invariant local features. Speeded-Up Robust Features (SURF) are used for image features for sub-images. Furthermore Okamoto and Hashizume have suggested the appropriate number of key images in RBIR depends on users, target images, and image database. Therefore we also implement the user interface of RBIR that users can change the number of key images directly and easily in real time. In order to obtain high usability, we use HTML5 for the development of the system. Proposed system is evaluated using precision, recall, and F-measure obtained by the experiment using 200 images by 10 subjects. The precision, recall, and F-measure of proposed system are 15%, 33%, and 0.20 for fixed number of

key images, 21%, 21%, 0.18 for flexible number of key images, whereas 11%, 30%, and 0.16 for conventional system respectively.

key words Region-Based Image Retrieval, Speeded Up Robust Features, Bag of Visual Word, HTML5

目次

第 1 章	序論	1
第 2 章	関連研究	3
2.1	大域特徴量	3
2.1.1	Serata らの手法	3
	色モーメント	4
	Texture neighborhood	5
	8 方向 Sobel フィルタ	5
2.1.2	DCT 領域における特徴抽出を用いた研究	6
2.1.3	色ヒストグラムを用いた研究	6
2.2	Visual Key 枚数に関する研究	7
2.3	局所特徴量	7
2.3.1	SURF	7
	特徴点の検出	7
	特徴量の記述	9
2.4	Bag of Visual Word	9
2.5	Visual Key 型画像検索	10
2.6	クラスタリング手法	10
2.6.1	階層的クラスタリング	11
	Ward 法	11
2.6.2	非階層的クラスタリング	11
	k-means 法	12
第 3 章	SURF 特徴を用いた類似画像検索システム	13

目次

3.1	SURF 特徴を用いたデータベースの構築	14
3.1.1	画像の分割	14
3.1.2	Visual Word の選定	15
3.1.3	VW ヒストグラムの作成	15
3.1.4	VW ヒストグラムのクラスタリングと KI の選定	16
3.1.5	データベースの構築	17
3.2	HTML5 によるユーザインターフェースの構築	18
3.2.1	Key-Image の表示	18
3.2.2	Key-Image 数の調節	18
3.2.3	クエリの作成	19
3.2.4	検索結果画像の表示	20
第 4 章	構築システムの被験者実験と評価	22
4.1	実験環境	22
4.2	実験内容	23
4.3	実験結果	24
4.4	考察	25
第 5 章	結論	26
	謝辞	27
	参考文献	29
付録 A	被験者ごとの適合率の比較	31
付録 B	被験者ごとの再現率の比較	37

目次

2.1	色特徴のイメージ	4
2.2	テクスチャ特徴のイメージ	4
2.3	特徴点検出の流れ	8
2.4	Box フィルタ	8
2.5	BoVW による物体認識の流れ	10
3.1	システムの全体図	13
3.2	画像分割の流れ	14
3.3	VW 作成までの流れ	15
3.4	VW ヒストグラム作成までの流れ	16
3.5	VW ヒストグラムのクラスタリングと KI の選定	16
3.6	画像データベース	17
3.7	索引情報データベース	17
3.8	Web ブラウザ画面	18
3.9	Key-Image の表示	19
3.10	Key-Image の拡大	19
3.11	スライダーによる KI の増減	20
3.12	検索結果	21
4.1	目的画像 1	22
4.2	目的画像 2	22
4.3	全体平均適合率, 全体平均再現率, 全体平均 F 値の比較	24
A.1	被験者 1 の適合率の比較	31
A.2	被験者 2 の適合率の比較	32

目次

A.3 被験者 3 の適合率の比較	32
A.4 被験者 4 の適合率の比較	33
A.5 被験者 5 の適合率の比較	33
A.6 被験者 6 の適合率の比較	34
A.7 被験者 7 の適合率の比較	34
A.8 被験者 8 の適合率の比較	35
A.9 被験者 9 の適合率の比較	35
A.10 被験者 10 の適合率の比較	36
B.1 被験者 1 の再現率の比較	37
B.2 被験者 2 の再現率の比較	38
B.3 被験者 3 の再現率の比較	38
B.4 被験者 4 の再現率の比較	39
B.5 被験者 5 の再現率の比較	39
B.6 被験者 6 の再現率の比較	40
B.7 被験者 7 の再現率の比較	40
B.8 被験者 8 の再現率の比較	41
B.9 被験者 9 の再現率の比較	41
B.10 被験者 10 の再現率の比較	42

表目次

4.1 実験条件	23
--------------------	----

第 1 章

序論

近年，Web 上に画像をアップデートする機会が増えており，Web 上の画像データ蓄積量が増加している．それに伴い，画像検索に関する研究も進展しており，画像全体の内容に基づいた Content-Based Image Retrieval (CBIR) や，画像の一部分の類似性のみに着目した Region-Based Image Retrieval (RBIR) [1] といった手法が提案されている．RBIR では，画像全体が類似していなくても，一部分が類似していれば画像を検索することができる．また，クエリに用いる画像はシステム側が提示するため，ユーザが類似する画像を用意する必要がなく、検索したい画像の明確なイメージが無くても提示された画像を組み合わせることでクエリを作成できるため，検索が容易にできる．

RBIR の中でも，部分画像をそのままクエリとして用いる Visual Key 型画像検索が提案されている．部分画像ごとに特徴を抽出し，クラスタリングを行うことで類似する部分画像を分類している．従来の Visual-Key 型画像検索では，大域特徴である色・形状・テクスチャの 3 つの特徴を組み合わせた手法が提案されていた．しかし，近年では色特徴を中心とした研究が多く行われている [2]．これは，形状・テクスチャを特徴として用いることは精度の低下を招いてしまうという問題点が確認されているためである．特に Sobel エッジ抽出フィルターの係数値を用いた形状特徴では単純な特徴しか抽出できないため，検索の適合率を低下させてしまうという問題点が指摘されている [3][4][5]．

そこで，本研究では近年，物体認識の分野で主流となっている局所特徴を用いた RBIR を提案する．ここで，局所特徴は，SIFT 法 (Scale-Invariant Feature Transform) ，SURF 法 (Speeded-Up Robust Feature) ，ORB (Oriented Fast and Rotated Brief) などを指すが，これらは輝度勾配を基にした，局所的な形状の記述に用いることができる．局所特徴

の中でも，アフィン変化や照明変化に頑健である SURF 特徴 [6] を用いて形状特徴による RBIR の精度向上を目指す．SURF 特徴は輝度勾配情報のみを用いた特徴であり，色情報を考慮せず物体の形状に関する特徴を抽出する．

また，ユーザに提示するクエリ画像数は柔軟に考える必要があるという報告が岡本らの研究からされており [7][8]，ユーザに提示するユーザインターフェースを考慮することで検索精度が向上するという報告が橋詰の研究 [9] からされている．そこで，本研究では，クエリに用いる部分画像である Key-Image (KI) 数を 10 枚から 60 枚まで動的に変化させて検索が行えるシステムの提案とシステムに適したユーザインターフェースの実装も行う．ユーザインターフェースには HTML5 を用いて実装を行う．橋詰らの Adobe Flash 技術を用いる方法に比較し，標準化が進められている HTML5 を用いることでより多くのプラットフォームで利用が可能になる．

性能評価は，被験者 10 名による実験を行い，従来手法と KI 数を 20 枚に固定した提案手法と KI 数を可変にした提案手法の 3 つの手法において適合率，再現率，F 値の比較による評価を行っている．実験条件として，従来手法では色特徴をもとにしているため，目的画像と提示画像をカラーで提示する．これは，これまでの研究による報告を本研究のものと同じの被験者，同一の条件で比較するために行う．提案手法では局所特徴をもとにしているため，グレースケール化した画像を提示する．得られた検索結果から被験者全体の平均を求め比較を行っている．被験者実験より，従来手法の適合率，再現率，F 値がそれぞれ 11%，30%，0.16 であるのに対し，KI20 枚の提案手法は 15%，33%，0.20，KI 可変の提案手法は 21%，21%，0.18 となっており，検索精度の向上を確認している．この結果から，SURF による局所特徴を用いる提案手法は，色特徴のみを用いる RBIR と同程度かそれより高い適合率，再現率を達成しており，局所特徴は人間の感覚に近い特徴であると考えられる．また，これまでの色特徴を用いる RBIR に SURF を組み合わせることで，さらに精度向上ができると思われる．

第 2 章

関連研究

本章では，従来の部分画像に基づく類似画像検索手法に用いられている大域特徴，および提案システムで用いる局所特徴量，さらにクラスタリング手法についての説明を行う．

2.1 大域特徴量

大域特徴量とは，画像の全体像から得られる特徴量であり，画像全体から得られた特徴量を特徴ベクトルとし類似度を算出する．大域特徴は一つの画像から一つの特徴量を抽出することができる．

大域特徴の例として，色情報や形状情報，テクスチャ情報や，画像全体の色ヒストグラムなどがある．既存の研究では，大域特徴を用いた研究が行われており，色・形状・テクスチャの組み合わせを用いた手法や色ヒストグラムを用いた手法がある．

2.1.1 Serata らの手法

Serata らが用いた手法 [1] では，画像のピクセル領域から色（9 次元）・形状（8 次元）・テクスチャ（8 次元）の特徴を抽出し，3 つの特徴を組み合わせた 25 次元の特徴を用いて画像の分類を行っている．色特徴には， $L^*a^*b^*$ 色空間における色モーメントを，テクスチャ特徴には，Texture Neighborhood を，形状特徴には Sobel Edge Detector を用いて特徴抽出している．以下に各特徴抽出法について説明する．

2.1 大域特徴量

色モーメント

色モーメント法では、 $L^*a^*b^*$ 色空間上の輝度 L^* 、2つの色情報 a^* 、 b^* の計3つのチャンネルごとに、部分画像 P_c の座標、 (x, y) の画素値 $P_c(x, y)$ の平均 μ 、標準偏差 σ 、歪度 γ を算出する。それぞれの値は以下の式で求める。平均 μ は、次式の加算のみのため1次モーメント、標準偏差は2次式を含む2次モーメント、歪度は3次式の3乗根であり3次モーメントと呼ばれる。

$$\mu_{P_c} = \frac{1}{WH} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} P_c(x, y) \quad (2.1)$$

$$\sigma_{P_c} = \sqrt{\frac{1}{WH} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} (P_c(x, y) - \mu_c)^2} \quad (2.2)$$

$$\gamma_{P_c} = \sqrt[3]{\frac{1}{WH} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} (P_c(x, y) - \mu - c)^3} \quad (2.3)$$

ここでの W は部分画像の横方向のピクセル数、 H は縦方向のピクセル数である（図2.1）。 C はチャンネルを表し、 L^* 、 a^* 、 b^* の3つのどれかである。3つのチャンネルそれぞれ、 μ 、 σ 、 γ の3つの統計量を使うため、計9つの量が特徴量であり、色特徴として9次元のベクトルとしている。

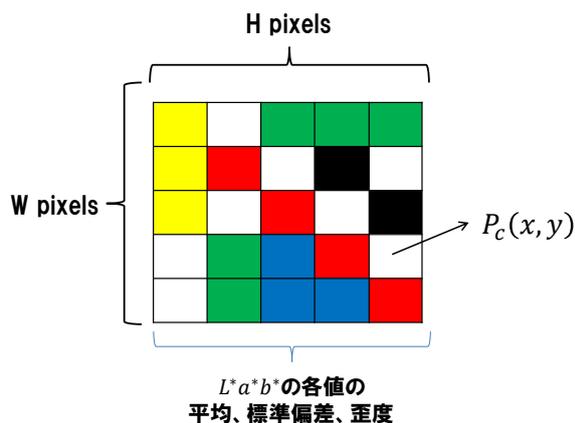


図 2.1 色特徴のイメージ

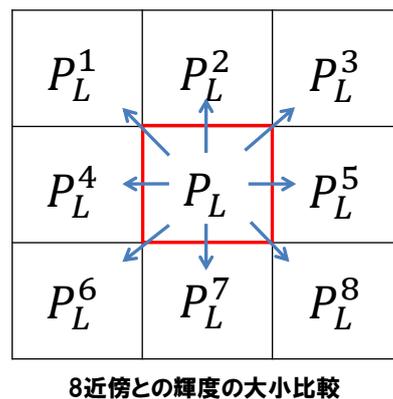


図 2.2 テクスチャ特徴のイメージ

2.1 大域特徴量

Texture neighborhood

Texture neighborhood は, L^* チャンネルにおいて, ある画素 P_L とその 8 近傍画素 $P_L^d (d = 1, \dots, 8)$ を比較し, P_L に対してそれぞれの近傍の画像値 $P_L^d (d = 1, \dots, 8)$ が P_L より大きいかなを求め (図 2.2). 画像中の全ての画素についてこの操作を行い, P_L が $P_L^d (d = 1, \dots, 8)$ より一定以上 (閾値 δ) 大きくなる割合を求める. この割合が特徴量となる. 特徴量 t_d を算出する式を以下に示す. 文献 [1] では $\delta = 0.0015$ として t_d を算出している.

$$t_d = \frac{1}{WH} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} G(P_L(x, y) + \delta, P_L^d(x, y)) \quad (2.4)$$

$$G(a, b) = \begin{cases} 1(a > b) \\ 0(\text{otherwise}) \end{cases} \quad (2.5)$$

8 方向 Sobel フィルタ

8 方向 Sobel フィルタは, HSI 色空間上の彩度 (S) チャンネルと輝度 (I) チャンネルに対して, 8 つの方向のフィルタ演算を行う. S チャンネルのフィルタリング画像 F_{Sd} と I チャンネルのフィルタリング画像 F_{Id} を得る. F_{Sd} と F_{Id} は以下の式で得られる.

$$F_{Sd} = f_{Sd}(x, y)|_{0 \leq x \leq W-1, 0 \leq y \leq H-1} \quad (2.6)$$

$$F_{Id} = f_{Id}(x, y)|_{0 \leq x \leq W-1, 0 \leq y \leq H-1} \quad (2.7)$$

各フィルタリング方向 $d (d \in 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi, \frac{5\pi}{4}, \frac{3\pi}{2}, \frac{7\pi}{4})$ の形状特徴 e_d は,

$$e_d = \frac{1}{WH} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} G(f_{Sd}(x, y), \tau_S \max_{d,x,y} f_{Sd}(x, y)) | G(f_{Id}(x, y), \tau_I \max_{d,x,y} f_{Id}(x, y)) \quad (2.8)$$

の式で得られる, ここで, $|$ は論理和演算を意味しており, $G(f_{Sd}(x, y), \tau_S \max_{d,x,y} f_{Sd}(x, y))$ か $G(f_{Id}(x, y), \tau_I \max_{d,x,y} f_{Id}(x, y))$ の値が 1 の時に式 2.8 は 1 となる. すなわち, 彩度 S について, 最も強く出ているエッジの 0.15 倍以上のものをエッジとしてカウントし, 輝

2.1 大域特徴量

度 I については, 0.35 倍以上のものをエッジとしてカウントする. 8 方向それぞれについてエッジとなるピクセルの割合を e_1 から e_8 に代入する.

また, τ_S と τ_I は閾値であり文献 [1] ではそれぞれ, 0.15 と 0.35 としている. これらを用いて 8 次元の形状ベクトルを得る.

2.1.2 DCT 領域における特徴抽出を用いた研究

岡本らが提案した, DCT 領域における特徴抽出法 [11][12][13] では, 圧縮画像から特徴抽出を行う際に逆離散コサイン変換 (IDCT:Inverse Discrete Cosine Transform) を行わずに圧縮処理に用いられている離散コサイン変換 (DCT:Discrete Cosine Transform) 領域から直接特徴を抽出することで計算量を削減している. 任意の部分領域から, 9 次元の色特徴ベクトルと 5 次元の形状特徴ベクトルを組み合わせた 14 次元の特徴ベクトルを抽出し, 画像の分類を行っている. 特徴ベクトルは全て DCT 係数を用いて算出されるが, IDCT を省略しているため, $L^*a^*b^*$ 色空間および HSI 色空間への変換ができない. そのため, YCrCb 色特徴から特徴を抽出している. 色特徴は色モーメントを形状特徴には H.S.Chang らが提案したラベリング法を用いて各特徴を抽出している [10].

2.1.3 色ヒストグラムを用いた研究

画像を分類する為の特徴に 3 つの特徴の組み合わせを用いるのではなく, 色ヒストグラムを用いた手法を西岡 [14] や篠原 [4] が提案している. 色ヒストグラムとは, 画像内の赤成分 (R), 緑成分 (G), 青成分 (B) の各チャンネルごとの輝度及び色合い分布を示したものである. 画像内に各チャンネルのピクセルがいくつ出現しているのかを表している. 西岡や篠原の研究では, 各色 256 色とする場合, $256^3 = 16777216$ 次元の特徴となるため, 各画像を 64 色に減色して利用している.

2.2 Visual Key 枚数に関する研究

Serata らの提案した手法ではユーザに提示する Visual Key 数は 20 枚とされてきた。そこで、岡本らが Visual Key の枚数を変化させる研究を行い、提示数が 60 枚程度の時が検索結果が最も良いと報告している [7][8]。このことから、Visual Key はユーザや対象画像、画像データベースなどによって適切な数が増えるため、柔軟に考える必要がある。また、この報告から橋詰の研究では、Visual Key 数を変更可能にするシステムを提案している [9]。

2.3 局所特徴量

局所特徴量とは、画像全体からではなく画像の一部分から得られる特徴量である。一つの画像から数百から数千の特徴量を抽出することができる。代表的な局所特徴量に SIFT (Scale-Invariant Feature Transform)、SURF (Speeded-Up Robust Features) などがある。

2.3.1 SURF

SURF (Speeded-Up Robust Features) とは、SIFT のアルゴリズムを改良し、特徴量抽出を高速化したアルゴリズムである。特徴点の検出と特徴量の記述の 2 段階からなる。

特徴点の検出

SURF は、局所特徴の一つであり、ある画素周辺の輝度の状態を記述する。全ての点についての状態を計算することは、コストが大きく、またメリットもないため、注目すべき点を選出して特徴の計算を行う。この選ばれた特徴点をキーポイント、あるいは”interest point”と呼ぶ。キーポイントの選出は基本的には、いくつかの段階で平滑化した画像の差分をとり、差分のあるところは輝度変化が大きいと判断される (DoG: Difference of Gaussian)。

SURF では、この部分を Box フィルタとヘッセ行列を用いて近似計算を行う。

2.3 局所特徴量

Step1. Box フィルタとヘッセ行列による DoG の近似

Step2. スケールスペースの構築

Step3. 極値探索

入力画像から特徴点検出までの流れを図 2.3 に示す。

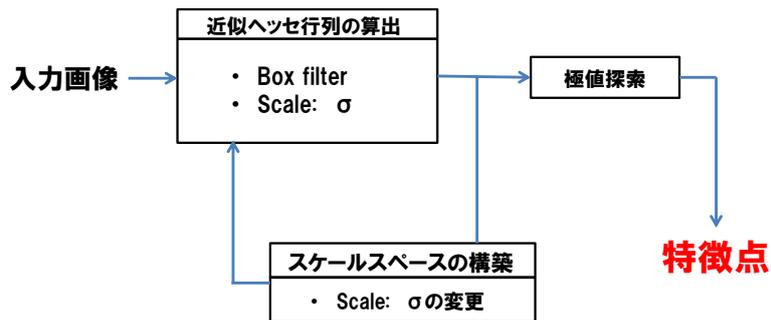


図 2.3 特徴点検出の流れ

SIFT では、ヘッセ行列の行列式が極値であるものを選ぶ。しかし、画像からヘッセ行列を算出する際、ヘッセ行列にはガウシアン²の 2 次微分を用いるため、計算コストが高く、行列値の計算に時間が掛かる。そこで、SURF では、ヘッセ行列の算出に Box フィルタを用いて近似を行う。Box フィルタを用いて算出した近似ヘッセ行列から極値を探索し、特徴点とする。図 2.4 に Box フィルタを示す。

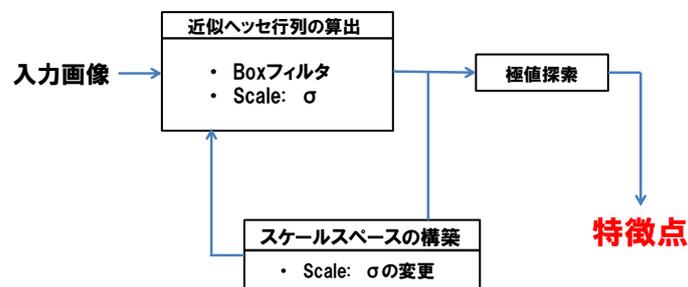


図 2.4 Box フィルタ

2.4 Bag of Visual Word

特徴量の記述

検出した特徴点から特徴量を記述する．特徴点付近の領域から Haar-Wavelet を用いて勾配方向と強度を算出し，勾配強度が最も大きい角度を全体的なキーポイントの方向と定める．これをオリエンテーションと呼ぶ．次にオリエンテーションから各領域の方向を正規化する．その後，各領域を $4 \times 4 (= 16)$ のブロックに分割し，分割した領域から Haar-Wavelet を用いて，輝度勾配ベクトルを算出する．各ブロックで，勾配ベクトルの方向を示す 2 つの成分，および x 方向， y 方向それぞれの勾配の大きさの 4 つの特徴を記述するため，キーポイントでの特徴は， 4×16 ブロックで 64 次元となる．さらに，拡張特徴では，場合訳を行い，倍の 128 次元が特徴ベクトルの次元となる．本研究では，この拡張された SURF 特徴を用いる．

2.4 Bag of Visual Word

Bag of Visual Word (BoVW) とは，局所特徴を用いた特定物体認識のために提案された手法である [15][16]．BoVW では，1 枚の画像から抽出した局所特徴の集合を単一のヒストグラムで表現し，ヒストグラムの比較を行うことで，画像間の類似度を算出している．すなわち，ある性質を持つ局所特徴をどれくらい持っているかで，その画像そのものの特徴とする考えである．BoVW は，一般物体認識の研究において有効性が示されている．BoVW による物体認識の流れを図 2.5 に示す．

青い矢印でデータベース作成の処理，赤い矢印で画像認識の処理を示す．データベース作成には，まず，全データベース画像から抽出した局所特徴ベクトルから，代表的な特徴ベクトルである Visual Word (VW) を選定する．VW の選定では，まず特徴ベクトルの集合全体に対しクラスタリングを行い，類似した特徴ベクトルのクラスタ（部分集合）を作る．同一クラスタ内の特徴ベクトルは全て同じものと見なす．各クラスタの中から 1 つずつ，中心付近の特徴ベクトルを選定し VW とする．次に全データベース画像を VW のヒストグラムで表現し，画像ごとに索引付けを行いデータベースに格納する．画像認識の処理では，デー

2.5 Visual Key 型画像検索

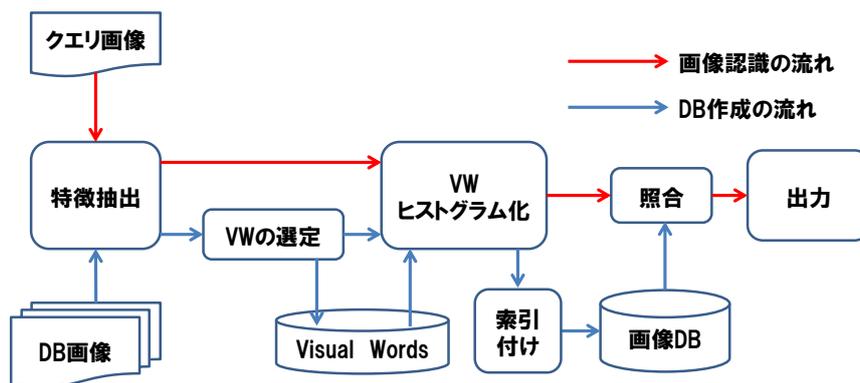


図 2.5 BoVW による物体認識の流れ

データベース作成時と同じく、クエリ画像から局所特徴を抽出し、画像を VW ヒストグラムで表現する。クエリ画像の VW ヒストグラムとデータベースに格納されている VW ヒストグラムを照合し、類似度の高い画像を認識結果として出力する。局所パターンは物体のカテゴリと関係が深く、そのヒストグラムの類似度を測ることでカテゴリの分類が可能となっている。

2.5 Visual Key 型画像検索

Visual Key 型画像検索 (VKIR) では、画像を細かく分割し、分割した部分画像をキーとして索引付けを行う。部分画像の類似度は VW を用いて行う。VW を用いて表現した部分画像に対し、類似したもの同士を集めるためにクラスタリングを行い、各クラスタの中心付近の部分画像を Visual Key (VK) として選定する。ここで VK は VW (Visual Word) とは異なるものである。ユーザはこの VK でクエリを行う。本研究では混同を避けるため、Visual Key を Key Image (KI) と呼ぶ。

2.6 クラスタリング手法

クラスタリングは教師ありクラスタリングと教師なしクラスタリングに分けられる。教師ありクラスタリングは、クラス分けと呼ばれる手法で与えられた情報をもとにクラスタを分

2.6 クラスタリング手法

割する．教師なしクラスタリングは，一般的にクラスタリングと呼ばれており，クラスタリング範囲の要素の分布により，クラスタを分割する．

教師なしクラスタリングは階層的クラスタリングと非階層的クラスタリングに分けられる．階層的クラスタリング，非階層的クラスタリングについて説明する．

2.6.1 階層的クラスタリング

階層的クラスタリングは，クラスタリング範囲の全ての要素を各クラスタとして扱い，クラスタ間の距離が近いものから順番に結合していく．クラスタリング結果はデンドログラム(図??)で表示することができ，グラフの高さによってクラスタ数を自由に变化させることができる．階層的クラスタリングでは，最短距離法，最長距離法，Ward 法などの手法が提案されている手法によって類似度の算出法が異なるが，本研究では Ward 法を用いる．

Ward 法

Ward 法とは，2つのクラスタを結合する際に，新たにできるクラスタ内の平方和の増加量が最小になるクラスタを結合していく手法である．クラスタ間の距離は2つのクラスタを結合した時の群内平方和の増加量となっており，増加量が小さいものが類似度が高いと判断する．Ward 法は群内平方和の増加量を利用していることから，一般的にクラスタリング感度が良いとされており、階層的クラスタリングの代表的な手法とされている．しかし，計算量は $O(n^2)$ であり，大量データには適さない．本研究では，Key Image の選定にこれを用いる．

2.6.2 非階層的クラスタリング

非階層的クラスタリングは，階層的クラスタリング以外の教師なしクラスタリングのことを指す．クラスタリング結果としてデンドログラムが生成されず，単にグループが生成される．非階層型クラスタリングの代表的な手法として k-means 法や k-means 法の派生である

2.6 クラスタリング手法

k-means++法や x-means 法があり，本研究では k-means 法を使用する．k-means 法の計算量は $O(n)$ であり，大量データに適している．本研究では，k-means 法を SURF 特徴のヒストグラムを作るためのクラスタリングに用いる．以下に，本研究で用いた k-means 法について説明する．

k-means 法

k-means 法は，あらかじめ指定した k 個のクラスタに要素をクラスタリングを行う手法である．クラスタの中心と各個体との間のユークリッド距離を類似度とし，距離が近いほど類似度が高いと判断する．以下にアルゴリズムの説明を行う．

Step1. クラスタ数 K を決定する．これが，本研究におけるヒストグラムの Visual Word の数になる．

Step2. 各要素を k 個ランダムで選択し，選択された要素を初期クラスタのシードとする．

Step3. 各要素と各クラスタのシードの距離を計算する．

Step4. 各シードと最も距離の近い要素を対応するクラスタに分類させる．

Step5. 各クラスタに含まれる全要素から重心を計算し，重心を新たなシードとする．

Step6. Step3. , Step4. の処理を行う．

Step7. 終了条件の判定を行う．終了条件は以下の 2 つで，終了条件を満たしていない場合 Step3. から再び処理を繰り返す．

- Step6. の結果クラスタ内の要素に変化が無い場合．
- 指定した繰り返し回数と Step7. の実行回数が同じ場合．

k-means 法では，初期のクラスタがランダムに付加されるため，初期値によってはクラスタリング結果が大きく変化するという問題もある．

第 3 章

SURF 特徴を用いた類似画像検索システム

本章では、提案する SURF 特徴を用いた類似画像検索システムについて説明する。システムは、Web サーバに Apache、データベースに SQLite、開発言語に HTML5、Javascript、サーバサイドに PHP を用いる。システムは全て仮想サーバ Virtual Box で動作する CentOS 上に構築する。システムの全体図を図 3.1 に示す。ユーザは Web ブラウザから Web サーバ上に構築した本システムにアクセスし画像検索を行う。

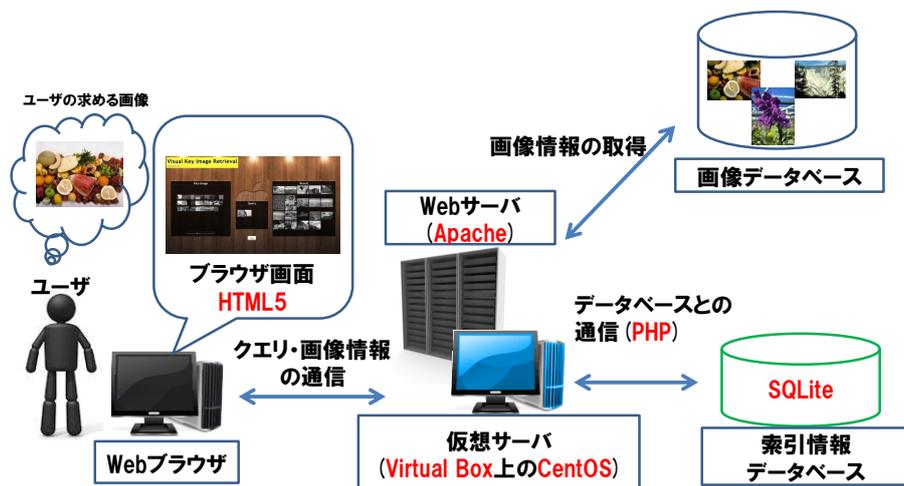


図 3.1 システムの全体図

3.1 SURF 特徴を用いたデータベースの構築

3.1 SURF 特徴を用いたデータベースの構築

本節では，SURF 特徴を用いたデータベースの構築方法について説明する，従来の手法では，大域特徴量を用いてデータベースを構築していた．提案する手法では局所特徴を用いてデータベースを構築する．データベースには，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．データベース構築は，下記のように行う．

1. 検索対象画像の分割
2. Visual Word (VW) の選定
3. VW ヒストグラムの作成
4. VW ヒストグラムのクラスタリングと KI の選定
5. データベース構築

3.1.1 画像の分割

画像の部分領域はデータベースに格納する検索対象画像を 3×3 に等分割することで生成する．検索対象画像は画像の高さや幅が異なるため，縦向きや横向きの画像が存在する．そのため，全て同じ大きさに揃える必要がある．提案するシステムでは，全ての画像を 300×300 ピクセルに拡大縮小を行い， 3×3 に画像を分割する．画像分割の流れを図 3.2 に示す．

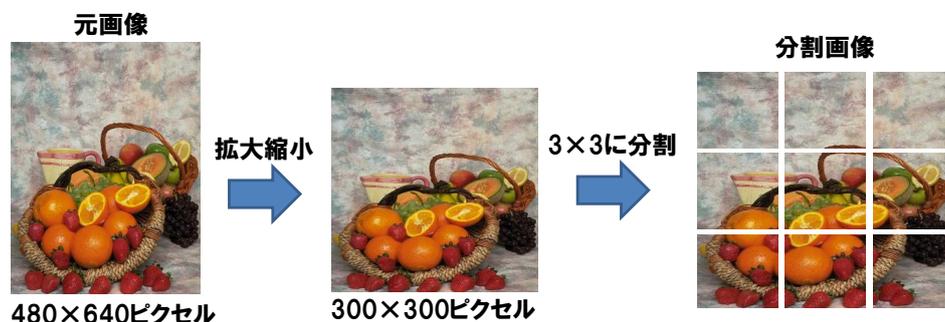


図 3.2 画像分割の流れ

3.1 SURF 特徴を用いたデータベースの構築

3.1.2 Visual Word の選定

Visual Word は全画像から抽出した SURF 特徴から選定する．まず， 300×300 ピクセルに拡大縮小した全検索対象画像から SURF 特徴を抽出する．次に，抽出した SURF 特徴に対しクラスタリングを行い，全特徴点から代表的な特徴点である，Visual-Word を選定する．本研究では，k-means 法を用いてクラスタリングを行い，10000 個のクラスタを生成する．各クラスタの中心を VW とし 10000 個の VW を生成する．VW は SURF 特徴と同じ 128 次元ベクトルとなる．VW 作成までの流れを図 3.3 に示す．

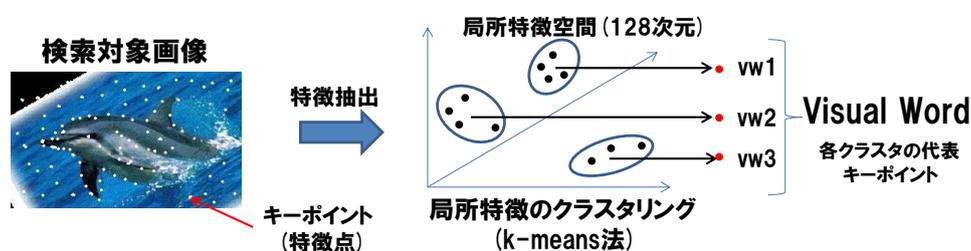


図 3.3 VW 作成までの流れ

3.1.3 VW ヒストグラムの作成

作成した VW をもとに各部分画像を VW ヒストグラムで表現する．部分画像中の各特徴量と各 VW とのユークリッド距離を算出し，最も近い VW に投票していくことでヒストグラムを作成する．最終的に作成したヒストグラムが各画像の特徴となる．本研究では，10000 個の VW を用いるため，各画像を 10000 次元のベクトルで表すことができる．VW ヒストグラム作成の流れを図 3.4 に示す．

3.1 SURF 特徴を用いたデータベースの構築

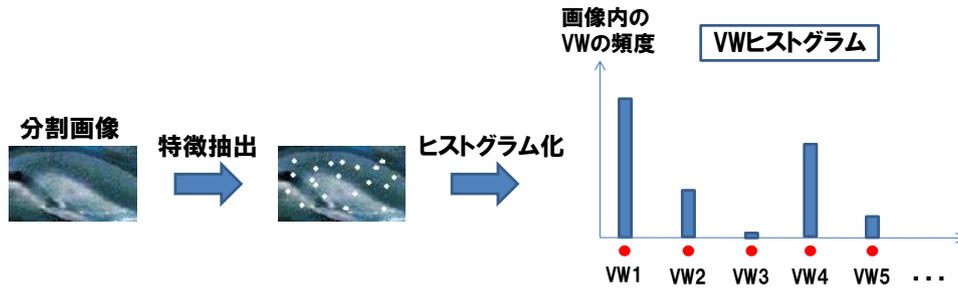


図 3.4 VW ヒストグラム作成までの流れ

3.1.4 VW ヒストグラムのクラスタリングと KI の選定

作成した VW ヒストグラムを各部分画像の特徴とし、VW ヒストグラムをクラスタリングすることで類似する部分画像を分類していく。Ward 法にてクラスタリングを行い、デンドログラムを作成する。作成したデンドログラムを、10 から 60 個のクラスタ数に分割し、各クラスタの中心ベクトルに最も近い画像を KI を選定する。KI 選定までの流れを図 3.5 に示す。

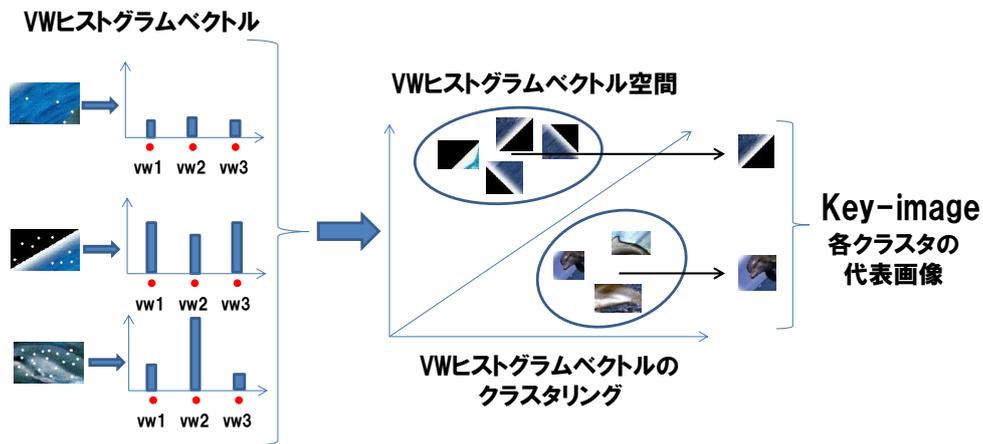


図 3.5 VW ヒストグラムのクラスタリングと KI の選定

3.1 SURF 特徴を用いたデータベースの構築

3.1.5 データベースの構築

選定した KI をもとにデータベースを構築する．画像データベースには，各クラスタ数ごとのディレクトリを用意し，各クラスタに属する KI を格納する（図 3.6）．索引情報データベースには，クラスタ数ごとに ID，元画像，部分画像，部分画像が属するクラスタ番号を格納する．図 3.7 はクラスタ数が 10 の場合の索引情報データベースの一部である．

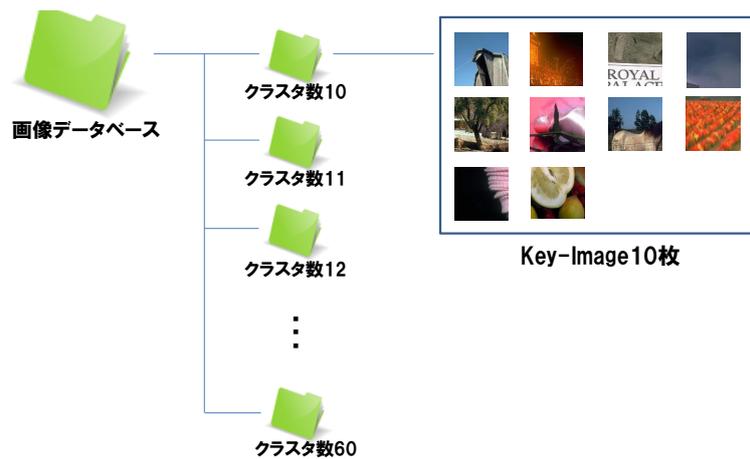


図 3.6 画像データベース

id	img	block	class
1	1 image_001.jpg	00104.jpg	1
2	2 image_001.jpg	00105.jpg	2
3	3 image_001.jpg	00106.jpg	3
4	4 image_001.jpg	00107.jpg	4
5	5 image_001.jpg	00108.jpg	5
6	6 image_001.jpg	00109.jpg	1
7	7 image_002.jpg	00202.jpg	2
8	8 image_002.jpg	00204.jpg	2
9	9 image_002.jpg	00205.jpg	4
10	10 image_002.jpg	00206.jpg	6
11	11 image_002.jpg	00207.jpg	7
12	12 image_002.jpg	00208.jpg	7
13	13 image_002.jpg	00209.jpg	6
14	14 image_003.jpg	00304.jpg	8
15	15 image_003.jpg	00305.jpg	8
16	16 image_003.jpg	00306.jpg	8
17	17 image_003.jpg	00307.jpg	3

図 3.7 索引情報データベース

3.2 HTML5 によるユーザインターフェースの構築

本節では、HTML5 により構築したユーザインターフェースについて説明する。提案するシステムでは、ユーザがクエリに使用する KI 数を動的に変化させることが可能なシステムとなっており、ユーザの判断で自由に KI 数を変化させることができる。構築した Web ブラウザ画面を図 3.8 に示す。

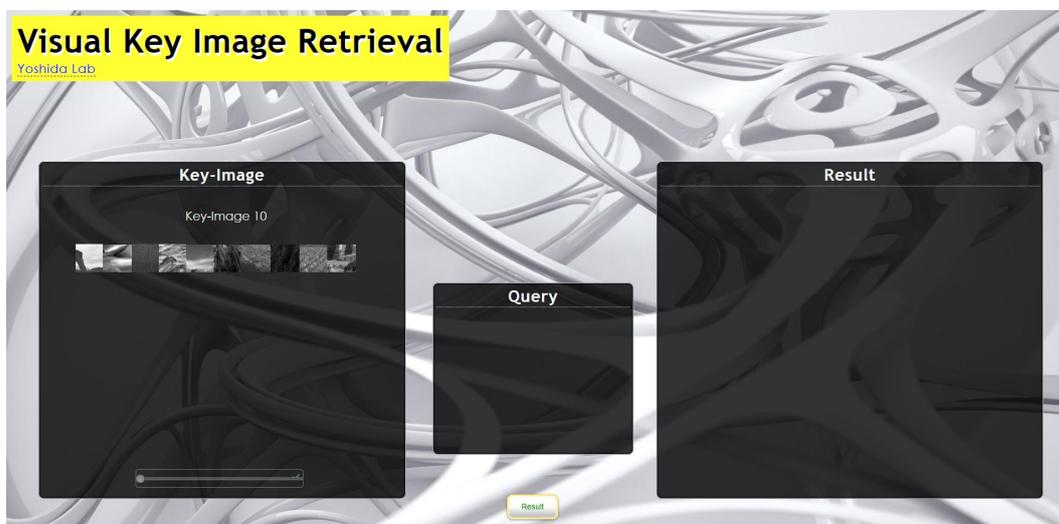


図 3.8 Web ブラウザ画面

3.2.1 Key-Image の表示

構築するシステムでは、図 3.9 のように KI を提示する。最小提示枚数は 10 枚、最大提示枚数は 60 枚となっている。提示する画像は 50 ピクセルで表示されており、図 3.10 のようにマウスカーソルが画像に近寄ると 100 ピクセルに拡大する。

3.2.2 Key-Image 数の調節

Key-Image 数を動的に変更する機能としてスライダー型インターフェースを実装する。図 3.11 のようにスライダーを左に動かすと KI 数を減らし、右に動かすと KI 数を増やすこ

3.2 HTML5 によるユーザインターフェースの構築

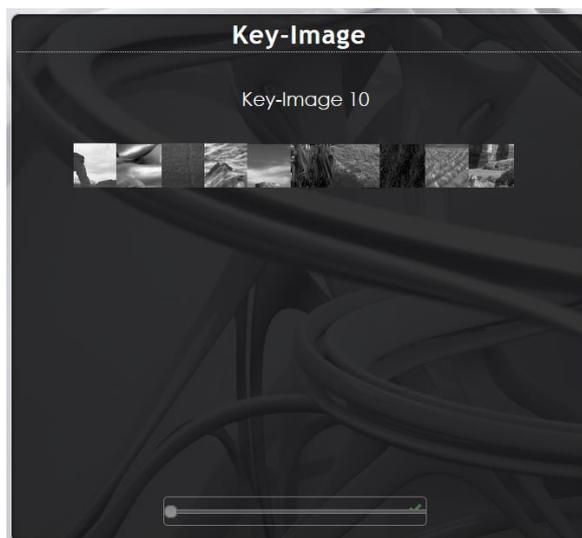


図 3.9 Key-Image の表示



図 3.10 Key-Image の拡大

とができる．初期表示数は 10 枚となっている．現在表示されている KI 数は Key-Image 画面の上部に表示されており，スライダーの位置にあわせて数値が変化する．

3.2.3 クエリの作成

ユーザは提示される Key-Image から検索画像のイメージに類似する Key-Image を選択することでクエリを作成する．ユーザが求める Key-Image が提示されていない場合，スライダーを動かして KI 数を増減させ目的の Key-Image を提示する．Key-Image をクリックすると Query 画面に画像が追加されクエリが作成される．Query 画面上の画像をクリックすると Query 画面から選択された画像は削除される．

3.2 HTML5 によるユーザインターフェースの構築

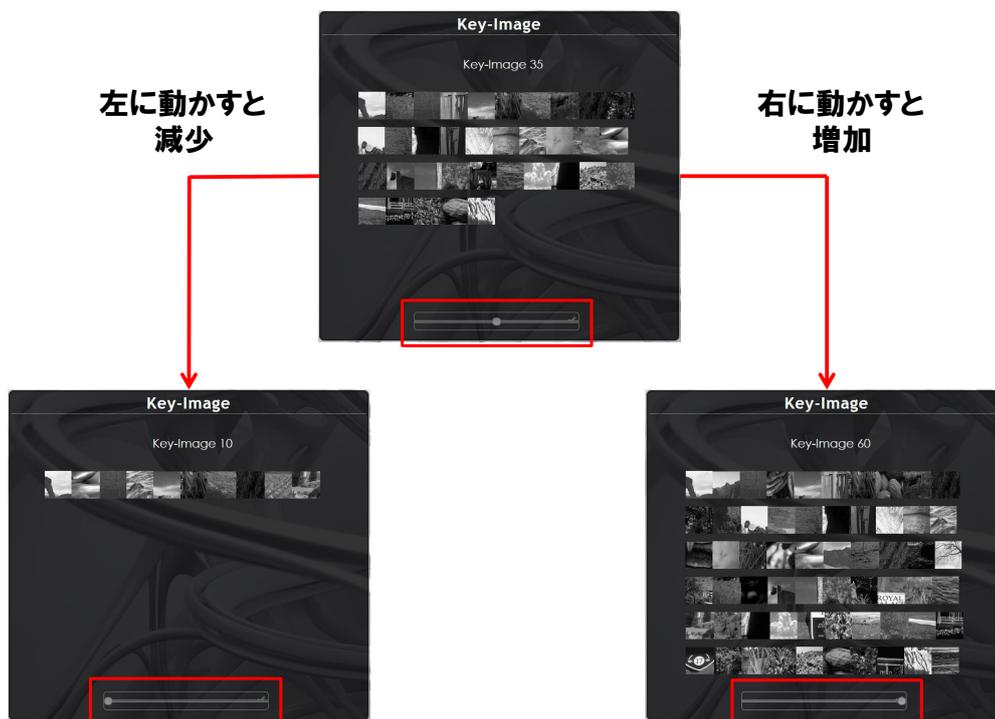


図 3.11 スライダーによる KI の増減

3.2.4 検索結果画像の表示

Query 選択後，Query 画面の下にある Result ボタンをクリックすることで，クエリを Web サーバ上の PHP に送信する．Web サーバから返信された検索結果をもとに画像データベースから検索結果画像を取得し，Result 画面に検索結果が表示される．図 3.12 は，Key-Image を 2 枚選択した時の検索結果である．Result 画面の上部に検索結果枚数が表示され，結果画像は一行に 5 枚ずつ表示される．

3.2 HTML5 によるユーザインターフェースの構築

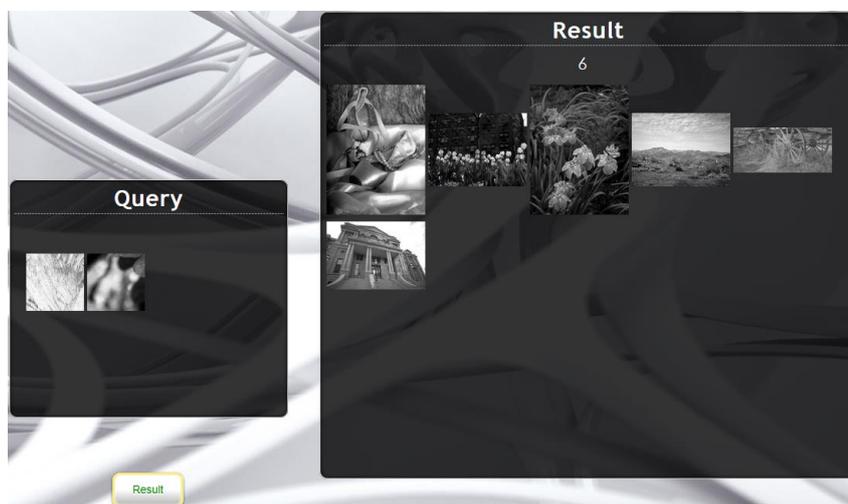


図 3.12 検索結果

第 4 章

構築システムの被験者実験と評価

本章では，被験者実験による従来手法と提案手法の比較を行う．提案手法と従来手法との被験者ごとの適合率，再現率，F 値の比較を行う．

4.1 実験環境

従来手法として，Serata らが提案した色特徴・形状特徴・テクスチャ特徴を組み合わせた手法を用いる．提案手法，従来手法ともにデータベースには，ArtExplosion の 10 カテゴリからそれぞれ 20 枚ずつ抽出した 200 枚の画像を用いる．被験者に検索してもらう目的画像は ArtExplosion の 10 カテゴリから 2 枚ずつ抽出した 20 枚の画像とする（図 4.1，図 4.2）．



図 4.1 目的画像 1

図 4.2 目的画像 2

従来手法の Pixel は部分画像の色特徴をもとに分類しているため，ユーザに提示する Key-Image と目的画像はカラー画像で提示する．提案手法で用いる SURF 特徴は，色情

4.2 実験内容

報ではなく形状の情報を分類しているため，ユーザに提示する Key-Image と目的画像はグレースケールで提示する．表 4.1 に実験条件を示す．

表 4.1 実験条件

	KI20 枚	KI 可変	従来手法
被験者	10 名		
画像データベース	ArtExplosion		
データセット	10 カテゴリから 20 枚ずつ抽出した 200 枚の画像		
目的画像	20 枚 (グレースケール)		20 枚 (カラー)
Key-Image 数	20 枚	10 ~ 60 枚まで可変	20 枚
適合画像	検索結果のうち被験者が類似すると判断した画像		

4.2 実験内容

被験者 10 名は従来手法 (KI20 枚固定)，提案手法 (KI20 枚固定)，提案手法 (KI 可変) の 3 つの手法でそれぞれ 20 枚の目的画像を検索する．従来手法での検索は目的画像に色が類似する Key-Image をクエリとして検索する．提案手法での検索は目的画像と形状が類似する Key-Image をクエリとして検索する．

クエリに使用する KI 数は最低 1 枚とし，上限は指定せずに行う．実験結果から被験者ごとの各条件での適合率，再現率，F 値を算出する．それぞれの算出方法は，下記の式で算出される．

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

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

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

4.3 実験結果

適合率は検索システムの正確性に関する指標であり，再現率は網羅性に関する指標である．これらを総合的に評価する指標に F 値があり，適合率と再現率の調和平均を取った値となる．

4.3 実験結果

図 4.3 に適合率，再現率，F 値の平均および 95%信頼区間の値を示す．結果は，従来手法は全体平均適合率，全体平均再現率，全体平均 F 値はそれぞれ，10%，31%，0.15 となった．それに対して提案手法は，KI20 枚がそれぞれ，16%，35%，0.21，KI 可変が 22%，22%，19%となった．適合率，再現率，F 値すべてにおいて提案手法が従来手法と同等または，それより大きくなるという結果になった．

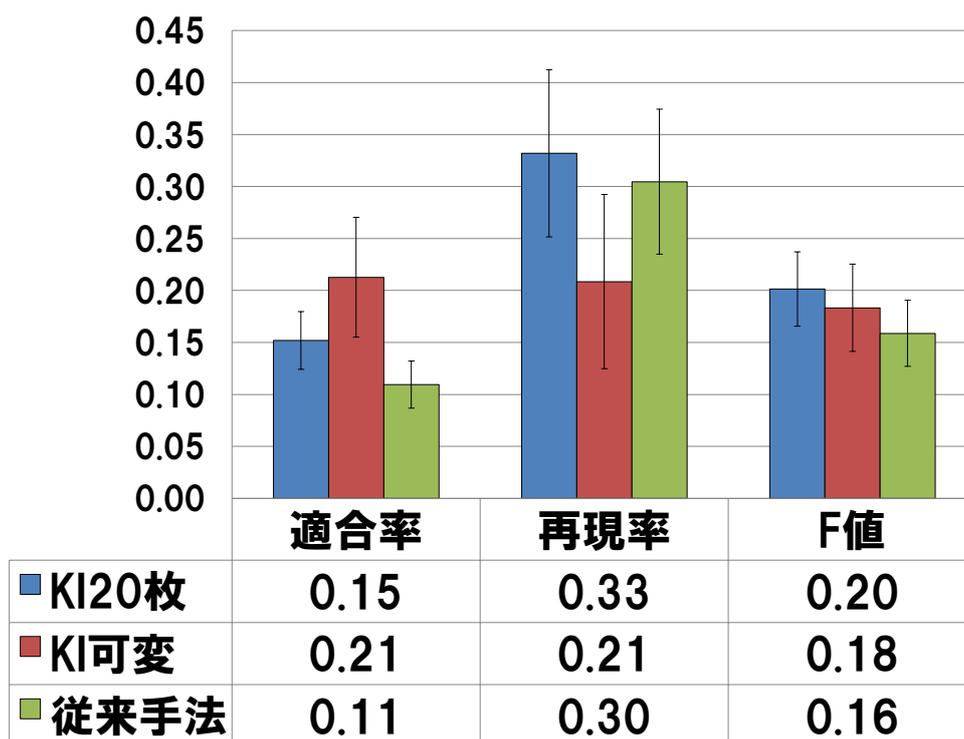


図 4.3 全体平均適合率，全体平均再現率，全体平均 F 値の比較

4.4 考察

従来手法と提案手法 (KI20 枚固定) を比較すると, 同じ KI 枚数でも全ての指標が同等かそれより高い数値を示している. このことから, 画像の分類に用いる特徴を Serata らの手法よりも提案手法の方が大域特徴から局所特徴に変更することで精度が向上しているといえる. 提案手法 (KI 可変) は KI20 枚に比べて 6 ポイントの精度向上となった. これは, 提示する KI 数が増加することで部分画像を細かく分類し, 結果として提示する画像の枚数が減少したためだと考える. しかし, 再現率, F 値については, KI 数を可変にすると低下する. これは, 提示する KI 数を変更することはできるがクエリとして選択する KI 数の変化がなかったため, 絞り込む条件が厳しくなり, 検索結果が 0 枚になるケースが多くなったためだと考える. これらの結果から, SURF 特徴を RBIR に用いることの有効性を示せていると考える.

第 5 章

結論

本研究では、局所特徴の一つである SURF を用いた RBIR の提案とユーザに提示する Key-Image (KI) を動的に変更できるシステムの実装、HTML5 によるユーザインターフェースの構築を行った。SURF 特徴を用いた RBIR では、SURF 特徴から k-means 法によるクラスタリングを行うことで Visual Word (VW) を選定し、ヒストグラムによる特徴ベクトルを作成した。RBIR 検索システムは、200 枚の画像から局所特徴を用いて、Ward 法でクラスタリングを行い 1800 枚の部分画像を 10 から 60 個までのクラスタに分類し、Key-Image とした。システムでは、HTML5 と JavaScript, JQuery を組み合わせて、KI 数を動的に変更できるユーザインターフェースを実装・構築した。

被験者実験では、被験者 10 名に従来手法と KI20 枚に固定した提案手法、KI を可変にした提案手法の 3 つの場合において、20 枚の目的画像を検索させ、各場合の適合率、再現率、F 値の平均を求めた。実験結果は、従来手法と KI20 枚の提案手法を比較すると適合率、再現率、F 値がそれぞれ+4 ポイント、+3 ポイント、+4 ポイントの精度向上となった。また、従来手法と KI 可変の提案手法を比較すると、それぞれ+10 ポイント、-9 ポイント、+2 ポイントと適合率、F 値は精度向上が確認できた。再現率については KI 数は可変になるものの、クエリに選択する画像については制限をかけていなかったため、検索結果画像数が 0 枚になることが多くなり、低下したと考える。適合率、再現率、F 値全ての指標において提案手法が大きくなることから、SURF 特徴を RBIR に用いることの有効性を示せたと考えている。

今後は、色ヒストグラムと SURF 特徴を併用することや、ユーザインターフェースの改良を行うことで更なる精度向上ができると思う。

謝辞

本研究を進めるにあたり、ご指導していただきました高知工科大学 情報学群 吉田真一准教授には心から感謝します。吉田先生には、2年間熱心にご指導していただきました。こうして本研究が形になり、論文執筆までできたのも吉田先生のおかげだと感謝しております。同時に、吉田先生には多大なご迷惑をお掛けしました。特に論文に関しては、日付が変わるまでご指導していただきました。そのおかげでなんとか締め切りまでに論文が完成しました。また、全く英語のできない私に一から英語を教えていただいたおかげで、なんとか英語を書ききることができました。最後の最後までご指導していただき本当にありがとうございました。心より感謝します。

そして、本研究の副査を引き受けていただいた高知工科大学 情報学群 福本昌弘教授と高知工科大学 情報学群 坂本明雄教授に深く感謝致します。福本先生には、本研究だけでなく、研究室活動などでもさまざまな助言を頂きました。私が気がついていないことに助言を頂いたおかげで、大きな問題もなく終わることができました。本当にありがとうございました。坂本先生には、喫煙所で会うたびに相談に乗っていただきました。くだらない相談にも乗っていただき本当にありがとうございました。お二人には本当にお世話になりました。深く感謝します。

同研究室の皆様にもお世話になりました。研究室活動を楽しんでくれたのも皆様のおかげです。ともに卒業する4年生の皆様とは、2年ほどの付き合いでしたが、いつも明るく接していただいたおかげで、毎日楽しく生活できました。松尾達郎氏とは、研究室だけでなく、学校生活全てにおいて、ともに学び遊んだ仲でした。いつもくだらない話に付き合っていたり、深夜に研究室で大声で歌ったりしたおかげでモチベーションを維持でき、研究が完成したと思います。修士課程でもがんばってください。中村大樹氏とは、いつも馬鹿な討論をしていたことを覚えています。あなたに「お前が言うな」という台詞を何回言ったか覚えていません。小林礼佳氏を、私はいつも小馬鹿にしていますが、半分ほどが本心でし

謝辞

た。研究室のムードメーカーとして中村氏，小林氏を利用させて頂いたことを深く感謝します。岩切雄希氏には，毎週月曜日に週刊誌を読ませていただいたおかげで，一週間の始まりを楽しくスタートできました。

修士のお二人にもお世話になりました。3年次からいろいろなことを教わりました。藤森夏輝氏には，反面教師としていろいろなことを教わりました。絶対に遅刻や寝坊をしないようにします。また，よく食事や遊びに誘っていただいたのも藤森氏だったと覚えています。小池規伎氏には，研究の相談だけでなく，誰も知らないマニアックな知識を教えてくださいました。同じマニアックな知識を知っていたのは小池氏だけよく二人で盛り上がったことを覚えています。

3年生の皆様には，研究室活動やイベントでの幹事等，お世話になりました。3年生がいるだけで研究室が明るくなりました。また，実験では，被験者をしていただき，感謝しています。2時間ほどかかる実験に付き合ってくれたおかげで良い結果を得られました。これからも，皆様の個性を生かして研究室をより良いものにしてください。

吉田先生ならび，研究室の皆様。皆様のおかげで，本当に楽しい研究室生活を過ごすことができました。2年間だけでしたが，大事なものを学びました。心から感謝します。

最後に，今までずっと生活面や学費で支えてくれ続けた家族に心から感謝します。

参考文献

- [1] M.serata, Y.Hatakeyama, and K.Hirota, "Designing Image Retrieval System with the Concept of Visual Keys," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 10(2), pp. 136-144, 2006.
- [2] 河野 拓司, "部分領域画像検索における分割粒度の検索精度に与える影響," 平成 24 年度学士学位論文, 高知工科大学.
- [3] 立花 啓海, "ビジュアルキー型画像検索のための低次元色特徴の提案とキー提示法の改良," 平成 20 年度学士学位論文, 高知工科大学.
- [4] 篠原 悠貴, "色ヒストグラム特徴を用いた部分領域画像検索," 平成 24 年度学士学位論文, 高知工科大学.
- [5] K. Okamoto, Fangyan Dong, S. Yoshida, and K. Hirota, "Content-Based Image Retrieval via Combination of Similarity Measures," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 15(6) pp. 687-697, 2011.
- [6] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded Up Robust Features," InECCV, pp. 404-417, 2006.
- [7] K.Okamoto, K.Kawamoto, Fangyan Dong, S.Yoshida, and K.Hirota, "Design Criteria of Querying Interface for Visual Key Image Retrieval on Mobile Devices," Proc. 2nd IWACII, SS6-3, 2011.
- [8] K.Okamoto, K.Kawamoto, Fangyan Dong, S.Yoshida, and K.Hirota, "An Evaluation Strategy for Visual Key Image Retrieval on Mobile Devices," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 16(6), pp. 713-722, 2012. Proc. 2nd IWACII, SS6-3, 2011.
- [9] 橋詰 翔健, "類似画像検索におけるクエリ提示方法とユーザインタフェースの研究," 修士学位論文, 高知工科大学.

参考文献

- [10] H.S.Chang and K.Kang, "A Compressed Domain Scheme for Classifying Block Edge Patterns," *IEEE Transactions on Image Processing*, 14(2), pp.145-151, 2005.
- [11] 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 年度修士学位論文, 高知工科大学.
- [12] 岡本 一志, 董 芳艶, 吉田 真一, 廣田 薫, "離散コサイン変換領域の特徴に基づいた画像索引作成法," 知能と情報(日本知能情報ファジィ学会誌), 21(6), pp. 1092-1102, 2009.
- [13] 岡本 一志, 吉田 真一, "周波数領域における特徴量を用いたビジュアルキー型画像検索," 電子情報通信学会技術研究報告 WBS Wideband System 107(183), pp.37-42, 2007.
- [14] 西岡 孝晃, "ファジィ演算による色ヒストグラム比較を用いた画像領域抽出," 平成 23 年度学位学士論文, 高知工科大学.
- [15] 黄瀬 浩一, "Bag-of-Features に基づく物体認識(1) —特定物体認識—," コンピュータビジョン最先端ガイド 3, アドコム・メディア株式会社, pp. 63-84, (2010-12).
- [16] 柳井 啓司, "Bag-of-Features に基づく物体認識(2) —一般物体認識—," コンピュータビジョン最先端ガイド 3, アドコム・メディア株式会社, pp. 85-117, (2010-12).

付録 A

被験者ごとの適合率の比較

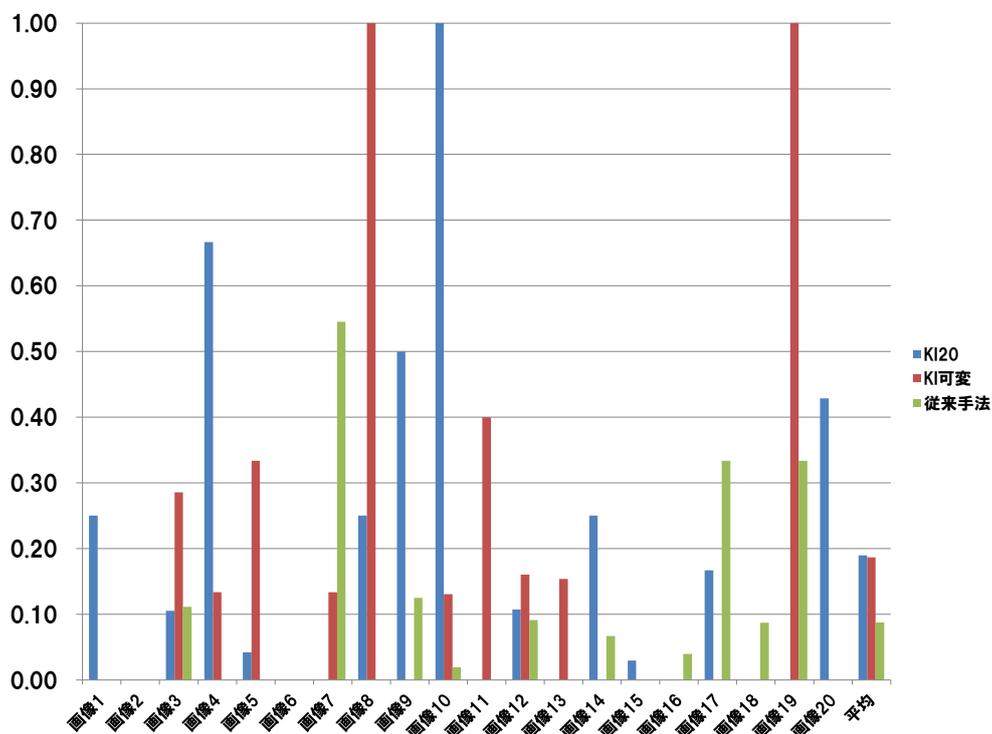


図 A.1 被験者 1 の適合率の比較

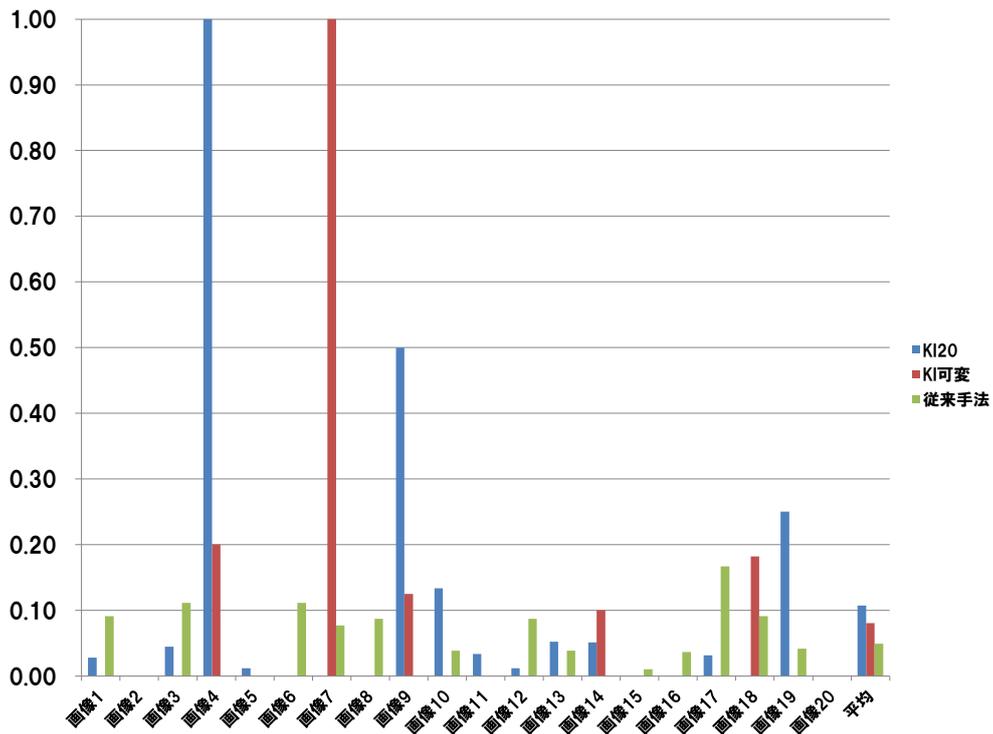


図 A.2 被験者 2 の適合率の比較

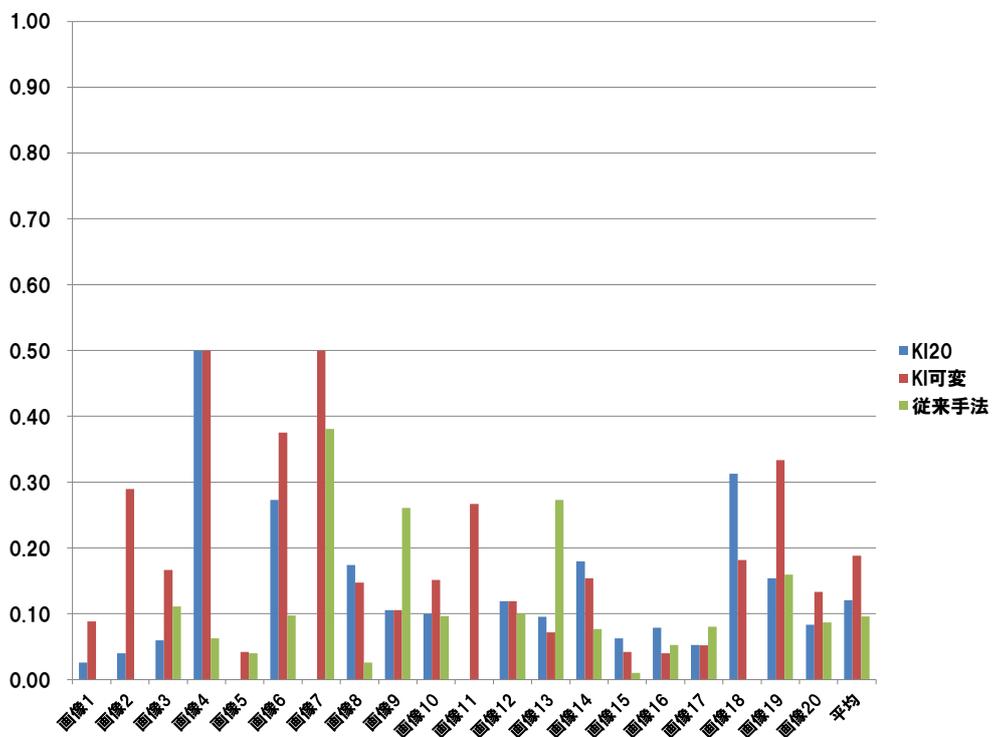


図 A.3 被験者 3 の適合率の比較

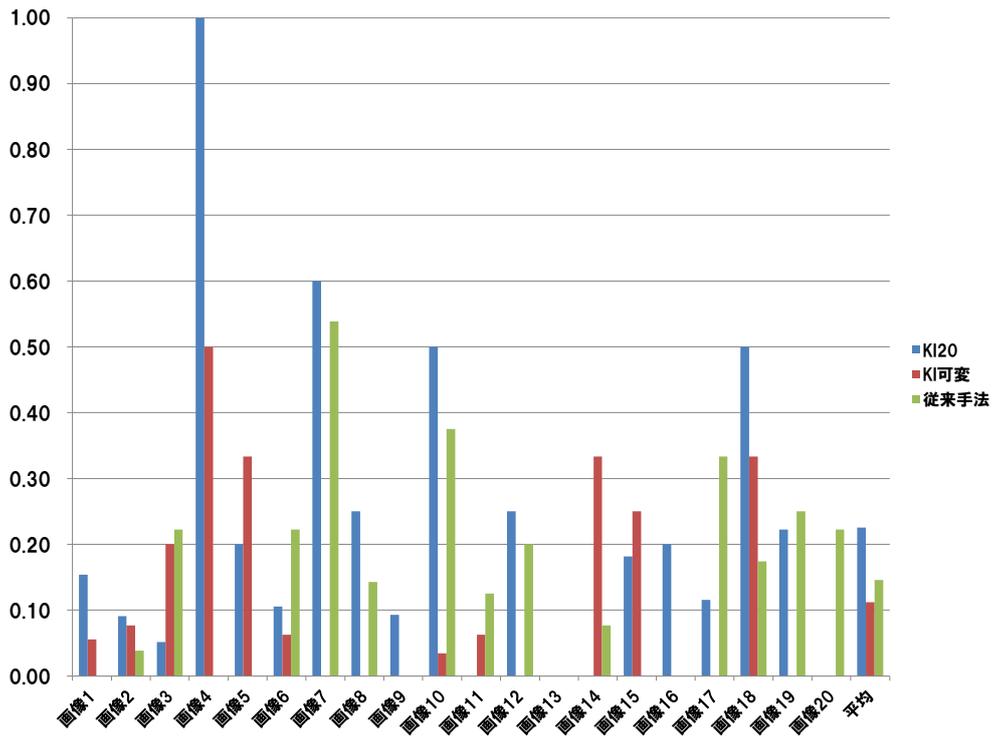


図 A.4 被験者 4 の適合率の比較

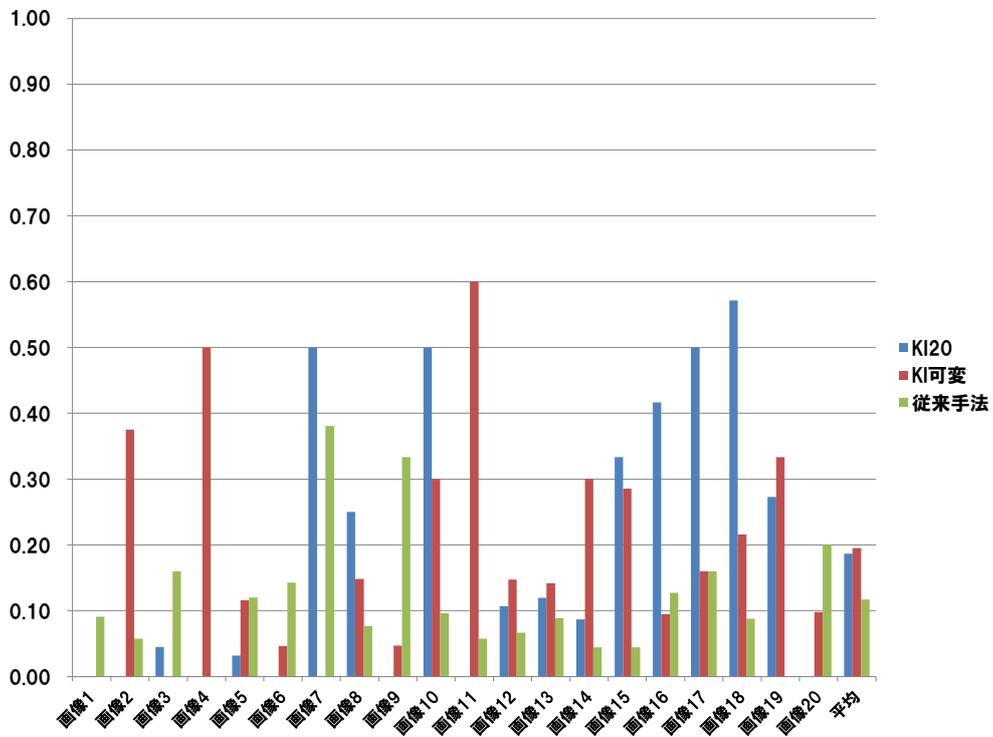


図 A.5 被験者 5 の適合率の比較

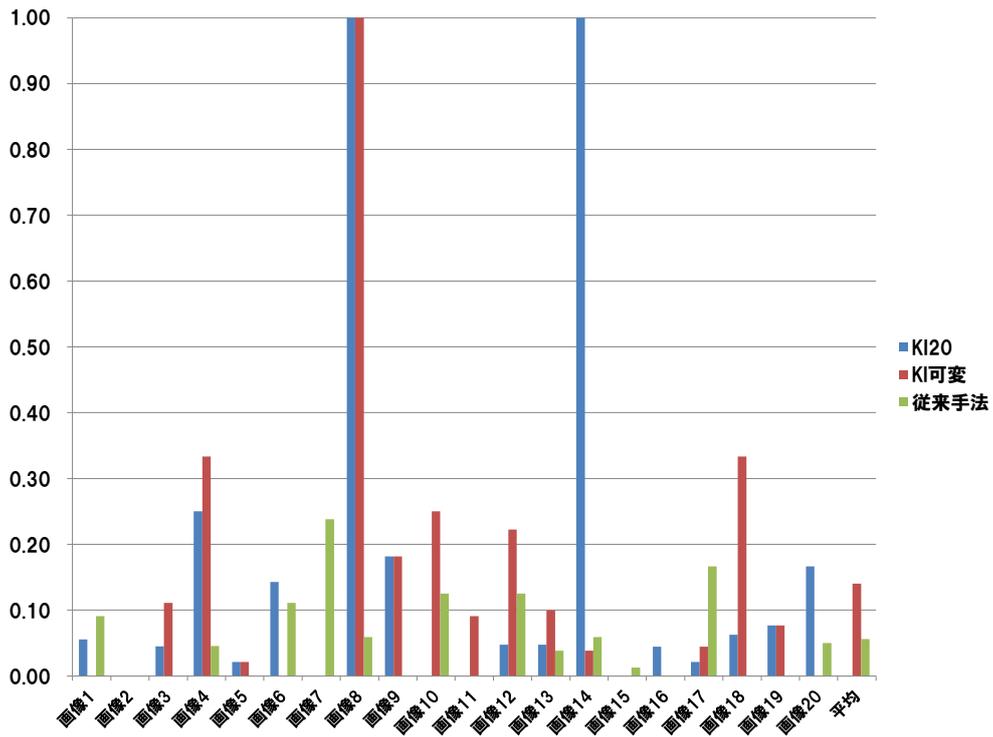


図 A.6 被験者 6 の適合率の比較

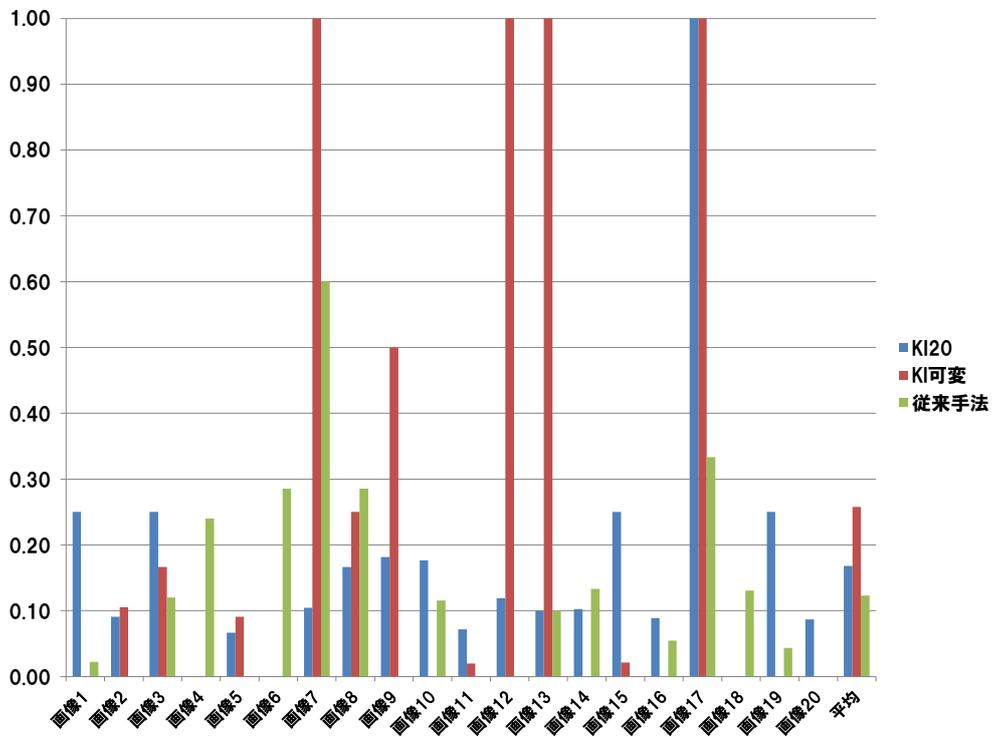


図 A.7 被験者 7 の適合率の比較

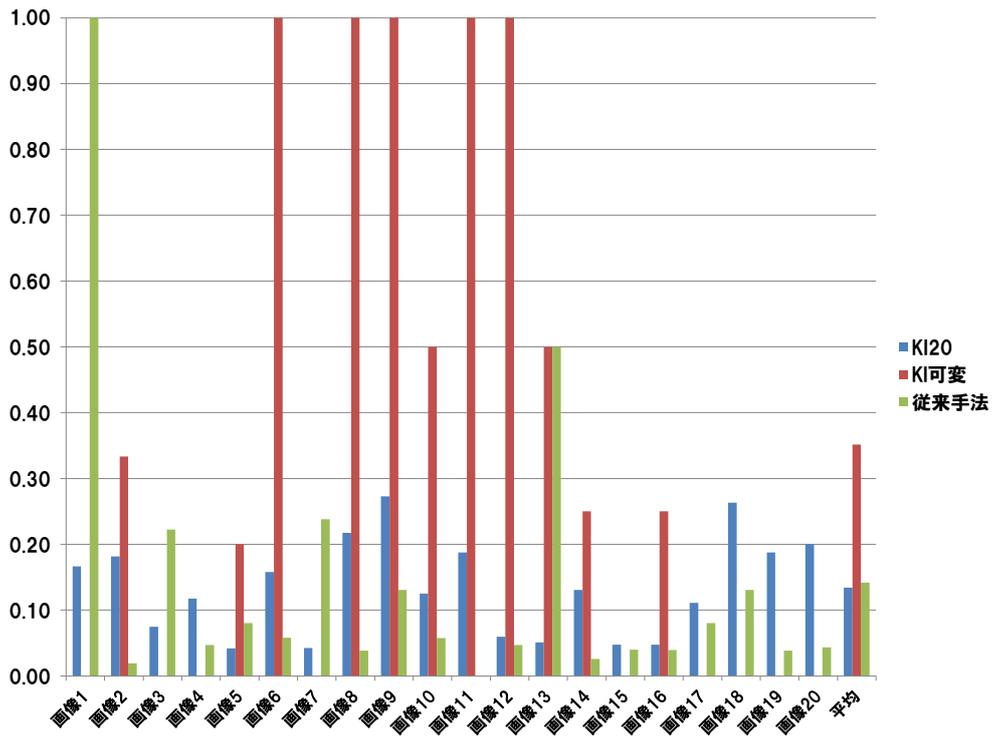


図 A.8 被験者 8 の適合率の比較

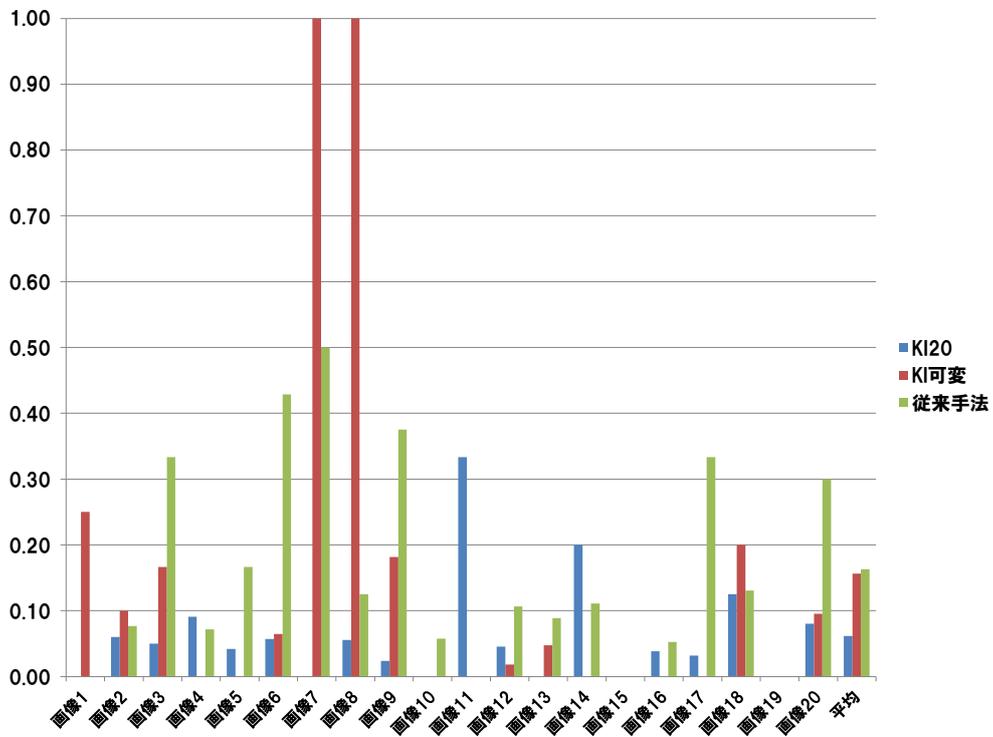


図 A.9 被験者 9 の適合率の比較

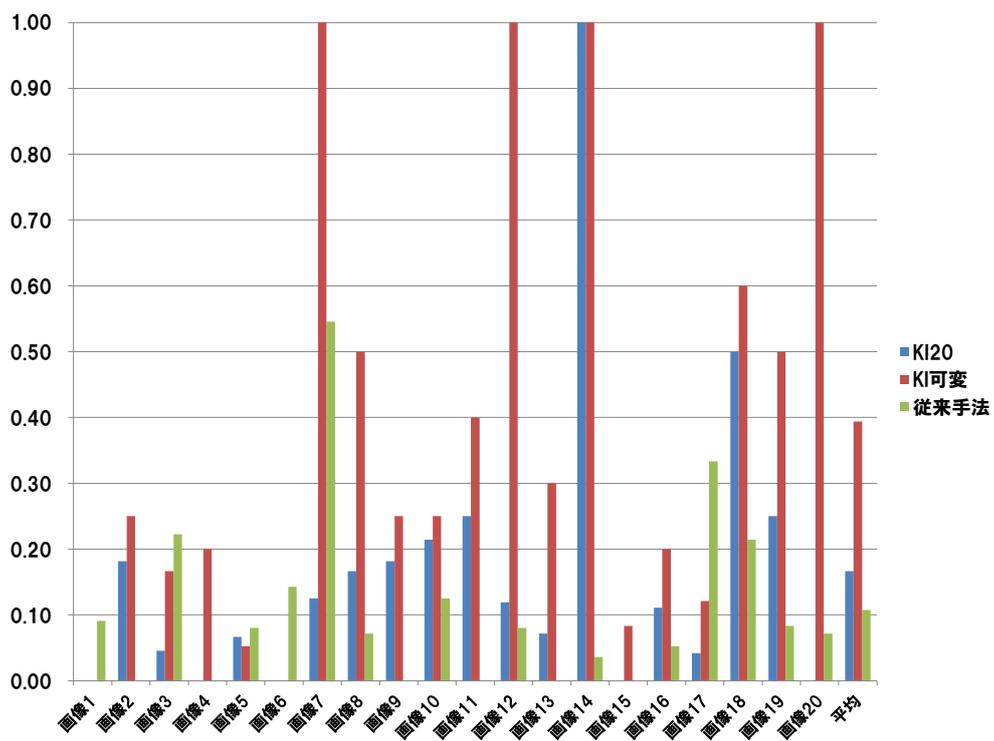


図 A.10 被験者 10 の適合率の比較

付録 B

被験者ごとの再現率の比較

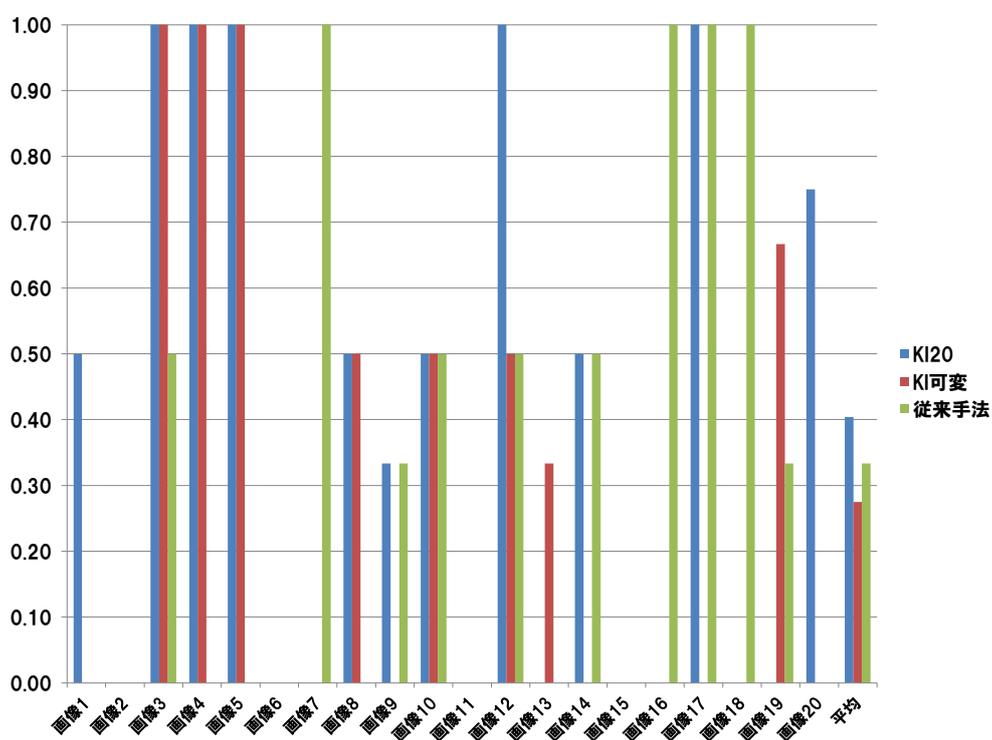


図 B.1 被験者 1 の再現率の比較

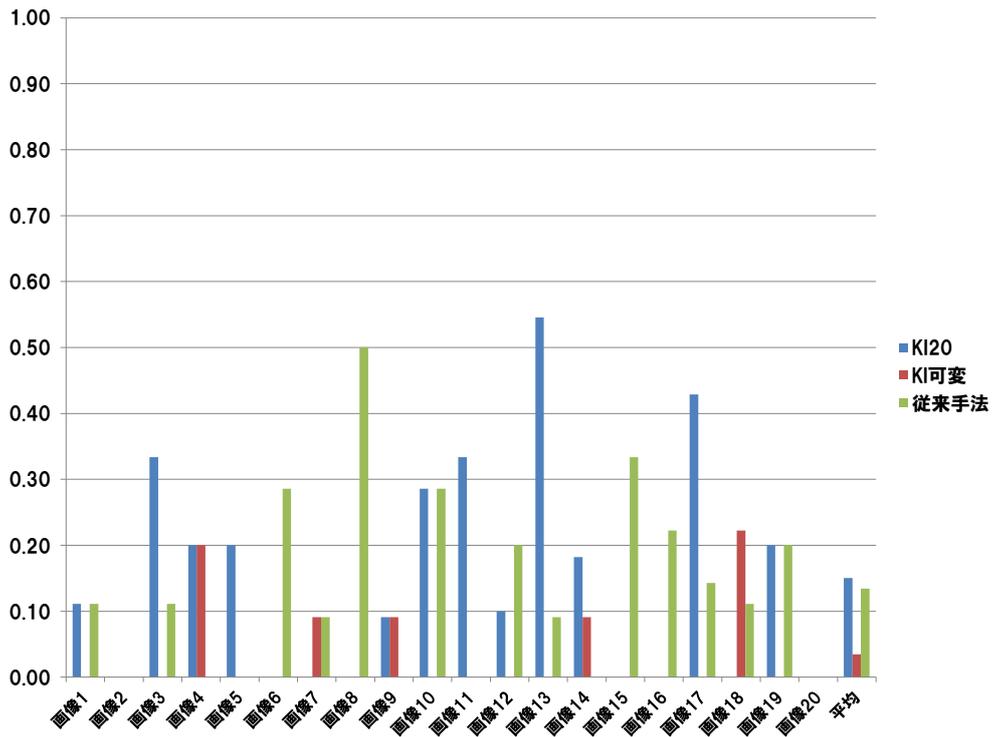


図 B.2 被験者 2 の再現率の比較

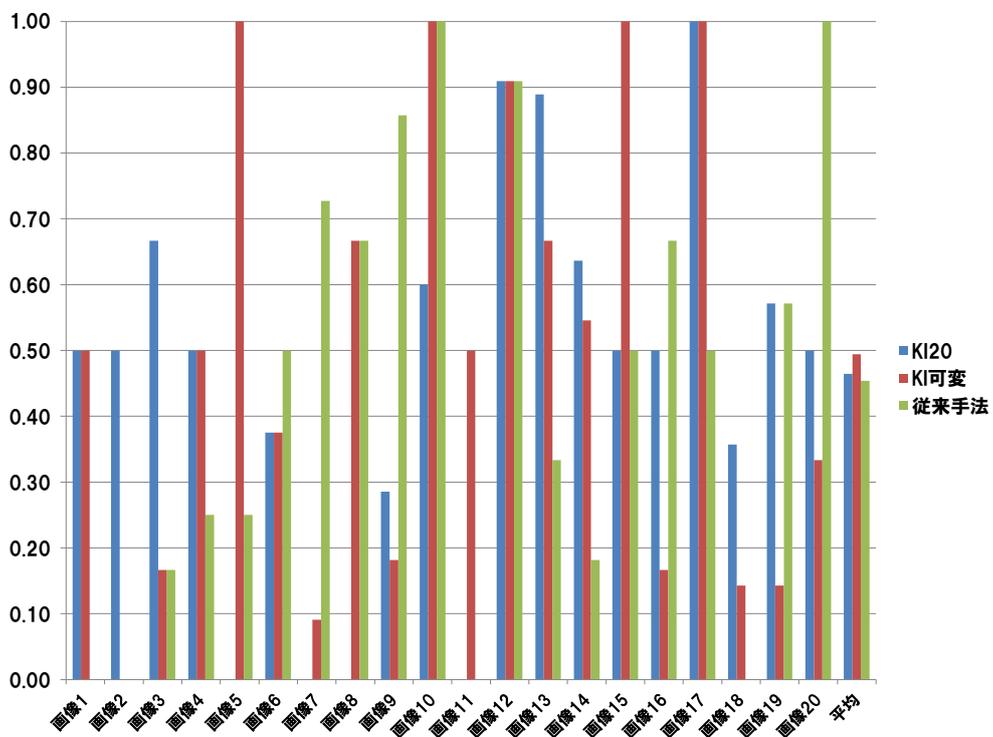


図 B.3 被験者 3 の再現率の比較

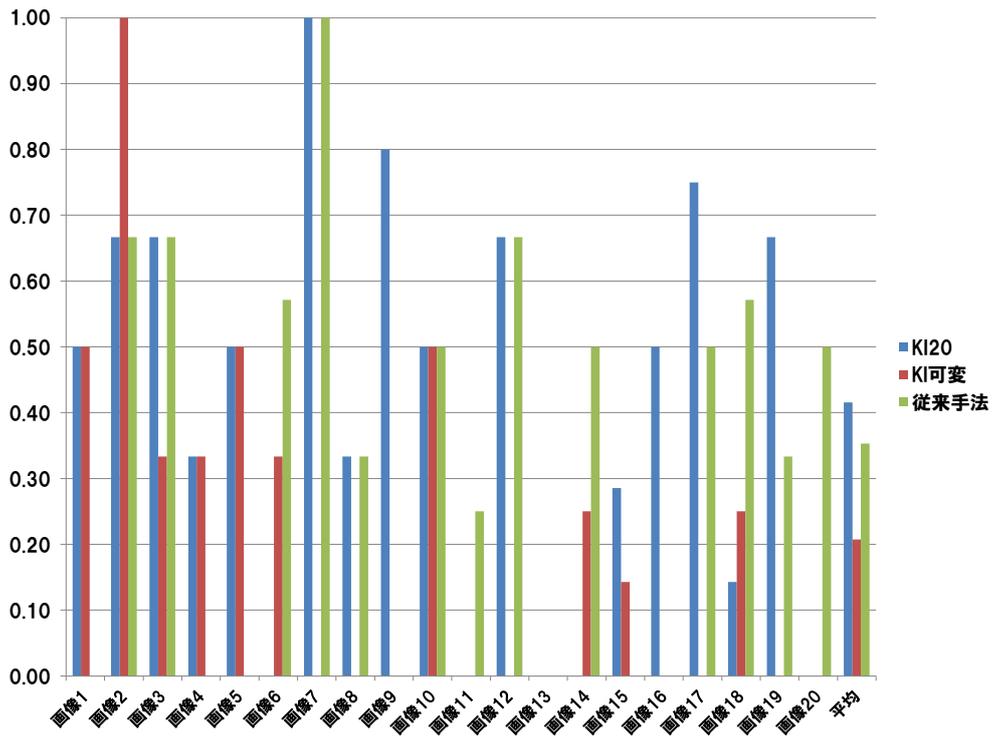


図 B.4 被験者 4 の再現率の比較

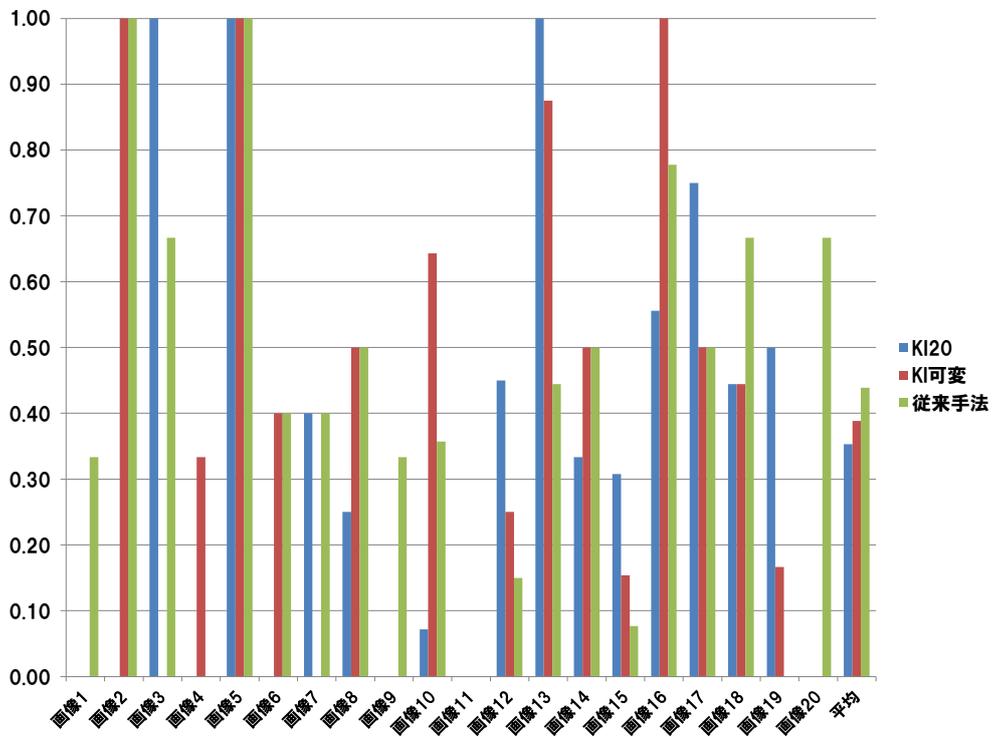


図 B.5 被験者 5 の再現率の比較

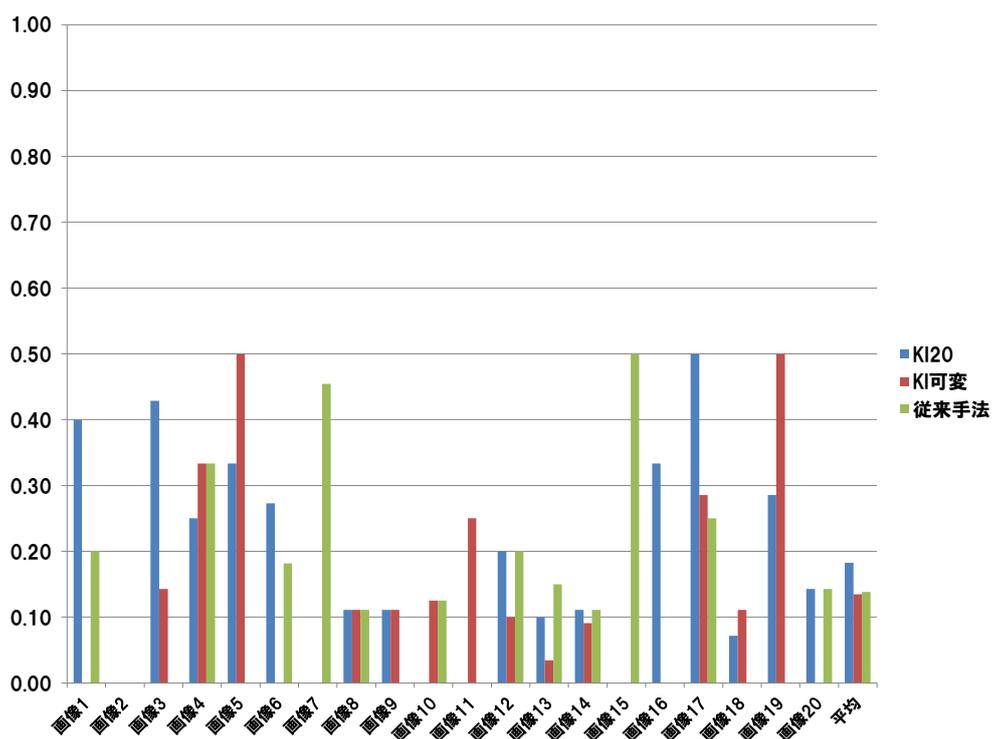


図 B.6 被験者 6 の再現率の比較

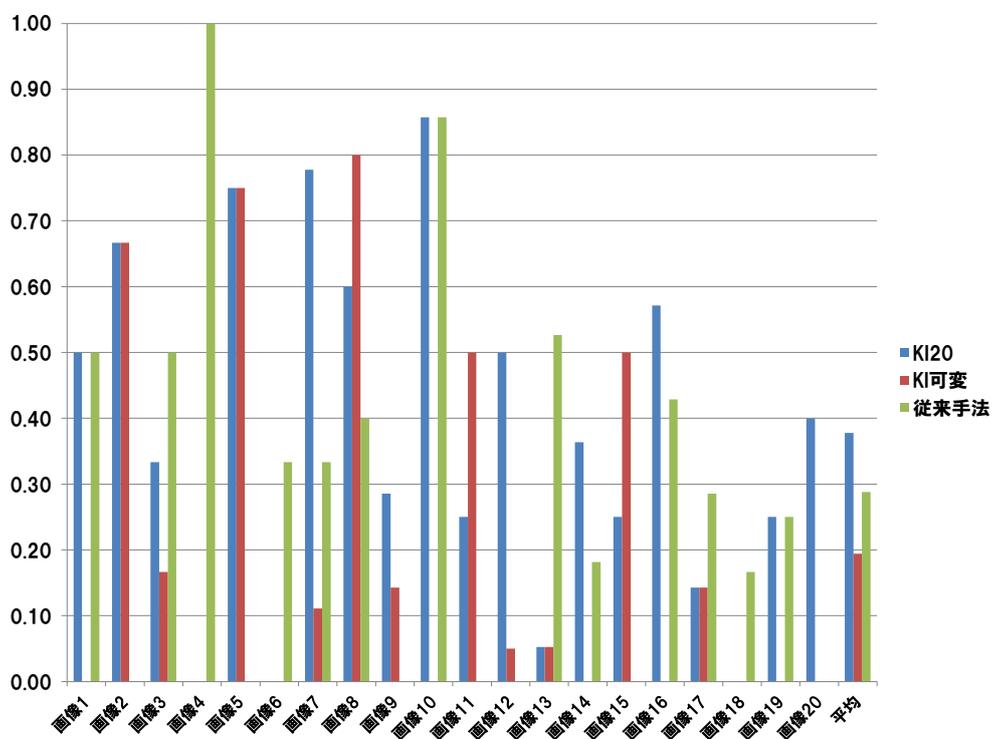


図 B.7 被験者 7 の再現率の比較

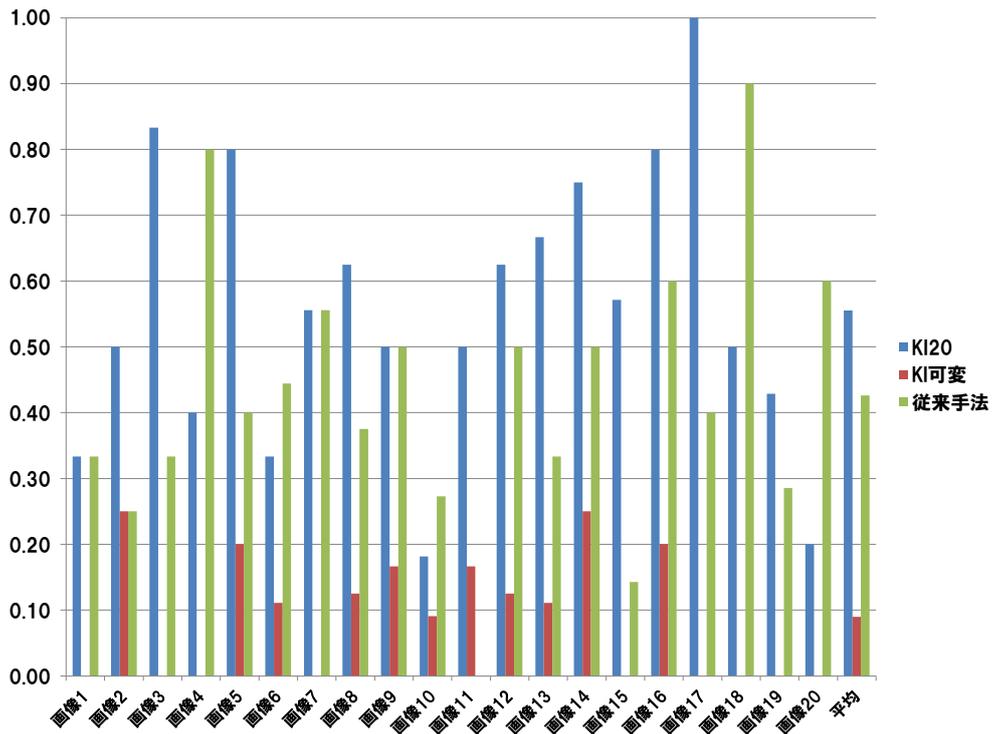


図 B.8 被験者 8 の再現率の比較

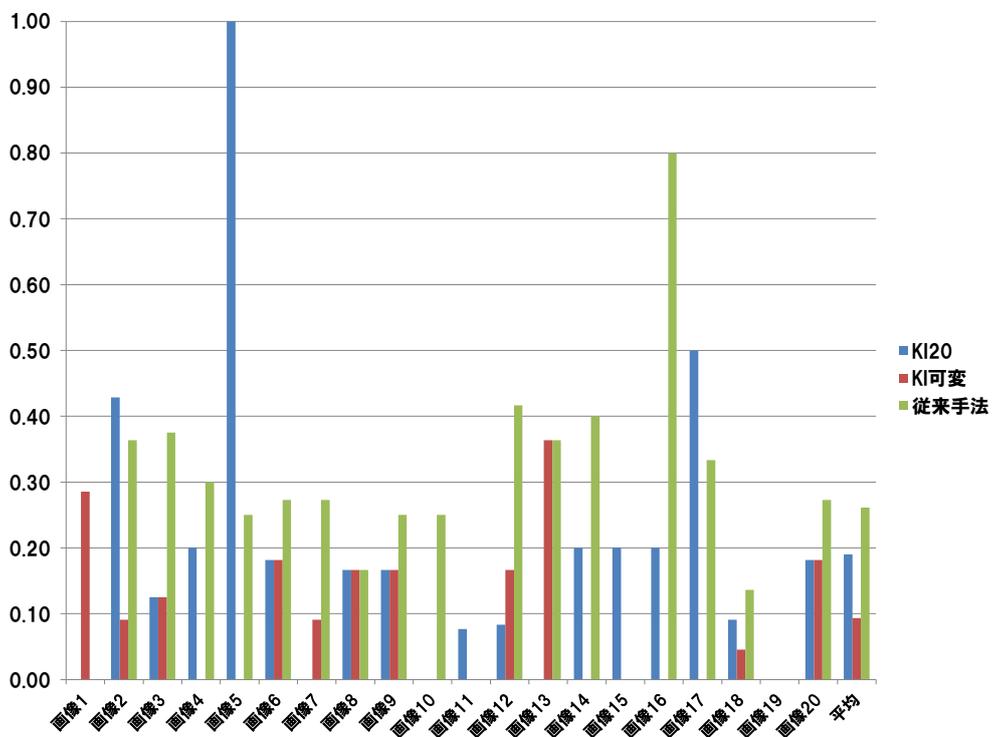


図 B.9 被験者 9 の再現率の比較

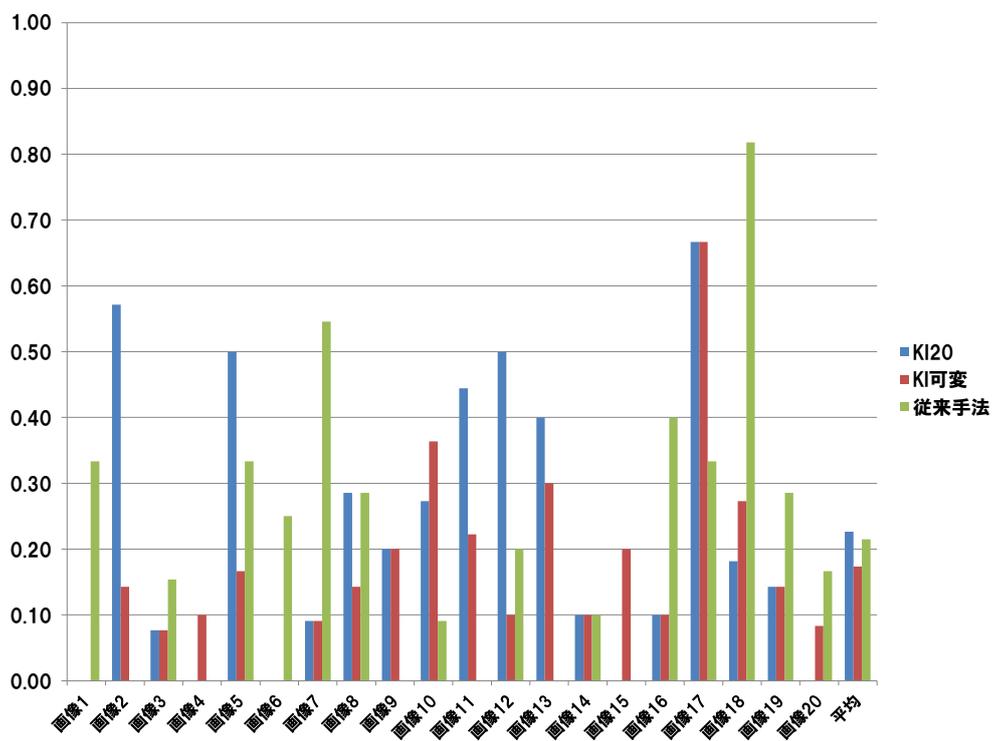


図 B.10 被験者 10 の再現率の比較