

平成 24 年度
学士学位論文

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

Effectiveness of Sub-Image Scale to the Precision of
Region-Based Image Retrieval

1130336 河野拓司

指導教員 吉田 真一

2013 年 3 月 1 日

高知工科大学 情報学群

要 旨

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

河野拓司

類似画像検索の中で画像部分領域の類似性に基づくビジュアルキー型画像検索が提案されている。しかし、従来の検索方法ではユーザは、要求する画像に提示されているビジュアルキーが使われているか否かの選択しかできず、類似するビジュアルキーが要求する画像のどの部分にあるかという位置関係は考慮されていなかった。そこで位置情報を含むビジュアルキーによりクエリを構成するシステムが提案されたが、 3×3 に分割数を固定しているため、分割数を変更することによる影響は明らかになっていない。そのため、被験者と画像における視覚的ギャップがビジュアルキー選択の誤りの原因になり、検索精度に影響を及ぼしている可能性があると考えられる。本研究では、画像の分割数を変更した場合の検索精度の変化を検証し、ビジュアルキーの生成方法についての基礎的な知見を得ることを目指す。本研究の実験内容は、画像を様々な分割数で分割してビジュアルキーを生成し、分割数の違いによる検索精度の影響を測定する。性能評価では、被験者 10 名による実験を行い、従来手法のビジュアルキー型画像検索と提案手法による検索の場合との適合率、再現率、F 値を比較する。実験の結果、 5×5 の分割数の検索精度が高いことを示す。従来手法のビジュアルキー型画像検索の場合、適合率 11%、再現率 24%、F 値 0.15 であるのに対し、 5×5 の分割数では適合率 36%、再現率 33%、F 値 0.35 となり、検索精度が向上していることを確認している。

キーワード ビジュアルキー，類似画像検索

Abstract

Effectiveness of Sub-Image Scale to the Precision of Region-Based Image Retrieval

KONO , Takushi

Visual-key image retrieval based on the similarity of partial area of image , called sub-image , has been studied as a region-based image retrieval . Conventional visual-key image retrieval measures a similarity between sub-image and query , which is input by a user regardless the location of the sub-image in images . Location based visual-key image retrieval has been proposed , however the size of scale of sub-images are fixed to 3×3 division . The size of sub-images or the number of division is an important issue to the accuracy of image retrieval . The size of sub-image affects the semantic gap between user's impression and therefore this affect the precision of retrieval . In this thesis , all images in the image database are divided to 2×2 , 3×3 , 4×4 , 5×5 and sub-images are constructed . These sub-images are clustered by Ward clustering algorithm . Twenty visual-keys are produced . Precision , recall , and F-values are derived by image search assessment . Image search assessment are conducted with ten subjects . Precision , recall , and F-values are compared with those of conventional method , 3×3 division . The result shows that the accuracy is the most highest in the case of division of 5×5 . The precision is 36% , the recall is 33% , and F-value is 0.35 for 5×5 division , while the precision is 11% , the recall is 24% , and the F-value is 0.15 for conventional 3×3 division .

key words visual-key , content-based image retrieval

目次

第 1 章	序論	1
第 2 章	関連研究	3
2.1	画像検索について	3
2.1.1	Text-Based Image Retrieval	3
2.1.2	Content-Based Image Retrieval	4
2.1.3	Region-Based Image Retrieval	4
第 3 章	ビジュアルキー型画像検索における画像の部分領域の処理や検索の流れ	5
3.1	ビジュアルキー型画像検索	5
3.1.1	ビジュアルキーの選定方法	7
3.1.2	特徴抽出	8
3.1.3	色モーメント法	8
3.1.4	Texture neighborhood	8
3.1.5	8 方向 Sobel フィルタ	9
3.1.6	特徴ベクトルの削減	9
3.1.7	クラスタリング	10
3.1.8	Ward 法	10
3.1.9	画像検索の流れ	11
第 4 章	位置情報を考慮した部分画像検索	13
4.1	位置情報を含むクエリを構成する検索システム	13
4.1.1	ビジュアルキーの生成	18
4.1.2	分割の流れ	18
4.1.3	ビジュアルキーの選定	19

目次

4.1.4	位置情報の付加処理	19
4.1.5	画像検索の流れ	21
第 5 章	分割粒度の検索精度に与える影響	23
5.1	実験環境	23
5.1.1	検索精度	26
5.1.2	実験結果	26
5.1.3	考察	29
第 6 章	結論	31
	謝辞	33
	参考文献	35
付録 A	被験者ごとの平均適合率	36
付録 B	被験者 1 人 1 人に対する画像ごとの適合率	37

目次

3.1	ビジュアルキー型画像検索のシステムの仕組み	6
3.2	ビジュアルキー選定	7
3.3	検索画面	12
3.4	検索結果	12
4.1	ビジュアルキー型画像検索の仕組み	15
4.2	位置情報を考慮した部分画像検索の仕組み	16
4.3	位置情報を考慮したビジュアルキー型画像検索	16
4.4	システム概念図	17
4.5	画像を分割するまでの流れ	18
4.6	インデックス情報	20
4.7	クラスと位置情報	20
4.8	位置情報を考慮したシステムの仕組み	22
5.1	画像データベース	24
5.2	検索目的画像	25
5.3	実験の条件	25
5.4	従来手法との比較	27
5.5	ビジュアルキー 1 枚の時の実験結果	27
5.6	ビジュアルキー 2 枚の時の実験結果	28
5.7	各画像ごとの平均適合率	28
5.8	奇数と偶数の分割数の違い	30
A.1	被験者ごとの平均適合率	36

目次

B.1 被験者 A	37
B.2 被験者 B	38
B.3 被験者 C	38
B.4 被験者 D	39
B.5 被験者 E	39
B.6 被験者 F	40
B.7 被験者 G	40
B.8 被験者 H	41
B.9 被験者 I	41
B.10 被験者 J	42

第 1 章

序論

近年，デジタルカメラや携帯電話の高性能化や急激な増加と記憶装置の大容量化に伴ってホームページや SNS などを通じて誰でも気軽にインターネット上に画像をアップロードできるようになり，インターネット上にある画像データの蓄積量が増加している．その中から自分の望む情報をいち早く探し出す手段として画像検索の研究が行われている．そこで，画像に対してあらかじめ対応付けされたテキストに基づき，キーワードをクエリとして検索を行う Text-Based Image Retrieval(TBIR) や画像内容に基づき画像の特徴などで検索を行う Content-Based Image Retrieval(CBIR) がある [2]．しかし，これらは画像を探す適切なキーワードが必要であったり，類似した画像でないと目的画像が見つからないという問題がある．そこで，画像の部分領域で特徴を抽出し，クエリを生成する Region-Based Image Retrieval(RBIR)[2] がある．RBIR の手法の一つとして，分割された画像をクエリに用いて画像検索を行うビジュアルキー型画像検索が提案されている [1]．ビジュアルキー型画像検索では，データベース中の画像を複数の部分領域に分割する．分割された部分領域の特徴を抽出 [4] し，クラスタリング [3][7] を適用する．クラスタリングによりそれぞれ類似する部分領域をまとめる．これら部分領域による類似画像の集合から最も平均的な特徴を持つ画像を代表画像としてビジュアルキーに選定する．この選定された部分領域をビジュアルキーと呼び，組み合わせることでクエリを生成する．生成されたクエリの各ビジュアルキーに索引付けられている画像をデータベースより参照し，参照された中で共通する画像を出力する．しかし，従来の検索方法では，ユーザは，要求する画像に提示されているビジュアルキーが使われているか否かの選択しかできず，類似するビジュアルキーが要求する画像のどの部分にあるかという位置関係を考慮していなかった．そのため，検索目的の画像と類似した画像

がデータベース上に複数存在する場合，目的画像をそこから絞り込んで見つけ出すのは困難であり，時間を要してしまう．そこで，位置情報を含むビジュアルキーによりクエリを構成するシステムが提案されている [9]．しかし， 3×3 に分割数を固定しているため，分割数を変更することによる影響は明らかになっていない．そのため，被験者と画像における視覚的ギャップがビジュアルキー選択の誤りの原因になり，検索精度に影響を及ぼしている可能性があると考えられる．本研究では，画像の分割数を変更した場合の検索精度の評価を検証し，ビジュアルキーの生成方法を見直す．性能評価は，被験者 10 名による実験を行い，従来手法のビジュアルキー型画像検索と提案手法の各分割数の適合率，再現率，F 値 [5][6] の比較により行っている．実験の条件として，ビジュアルキーの配置数は最大 2 枚までとし，配置する位置はブロック内において自由であり，配置位置が異なっていれば 2 枚とも同じビジュアルキーを選択してよいものとする．被験者は提案手法の 2×2 ， 3×3 ， 4×4 ， 5×5 の分割数で検索を行う．得られた検索結果から被験者全体での平均を求め，比較する．実験の結果， 5×5 の分割数の検索精度が高いことが分かった．従来手法のビジュアルキー型画像検索の適合率 11%，再現率 24%，F 値 0.15 に対し， 5×5 の分割数は適合率 36%，再現率 33%，F 値 0.35 であり，検索精度の向上を確認している．本研究により，膨大な画像データが存在する Web 上での画像検索において，よりユーザの目的画像と類似した画像を検索結果に表示できるようになると考えられる．また，絵も描けない [8]，キーワードも出てこない，手元に類似する画像もないといった時，自分のイメージする画像に対してビジュアルキーをジグソーパズルのように貼り付けていくことで目的の画像を安易に探し出せると考えられる．

第 2 章

関連研究

本章では、現在の画像検索技術について述べる。

2.1 画像検索について

本節では、目的画像を見いだすために研究されている画像検索の手法である TBIR, CBIR, RBIR について述べる。

2.1.1 Text-Based Image Retrieval

本節では、一般的な画像検索手法である Text-Based Image Retrieval(TBIR) について説明する。TBIR はキーワードにより検索を行う日常的に最も使用する Web 上の画像検索である。TBIR は検索に用いるメタデータとなるキーワードを登録時、画像に付加する必要がある。しかし、画像が急増するにつれ、画像の登録時にキーワードを人手により付加する方法では労力がかかる。また、1 枚の画像に対して、付加されるキーワードには個人的な主観に影響されるため、客観的な索引付けが求められるが、人手による方法ではどうしても人間の主観に依存してしまう問題がある。そのため、画像への自動索引付けを行う研究も行われているが、TBIR では検索したい画像のイメージはあるが検索キーワードが思いつかない時などには画像を取得することができない。

2.1 画像検索について

2.1.2 Content-Based Image Retrieval

本節では、Content-Based Image Retrieval (CBIR) について説明する。CBIR はテキストから探したい画像を検索するのではなく、検索したい画像を入力する事でそれに類似する画像を検索する手法である。CBIR では、画像にテキストデータ等のタグ付けからインデックスを作成する手法とは違い、画像から色、テクスチャ、エッジ等の特徴を数値として多次元のベクトルを抽出する。画像の特徴ベクトルから画像同士の類似度を計測し、類似している画像を検索する事ができる。しかし、ユーザがイメージする画像に類似する画像を所持していなければ、検索ができない問題がある。また、Web 上においてユーザが絵を描き、それを検索要求画像としてシステムに問い合わせを行って画像を検索する手法などが提案されているが、検索精度がユーザの描画能力に左右されるといった問題がある。また、ユーザが画像から感じる印象や類似度とシステムが処理する画像の類似度に差が存在する。

2.1.3 Region-Based Image Retrieval

本節では、Region-Based Image Retrieval(RBIR) について説明する。RBIR は、CBIR の手法の1つで分割した画像領域に基づいた画像検索システムである。RBIR では、画像の一部を類似する領域や特徴となる領域ごとに分け、それぞれの領域から特徴を求める。部分画像には、求めた特徴ベクトルからクラスタリングを行い形成されるクラスタに沿ってインデックスが付与される。RBIR は画像の全体象を入力する必要がなく、画像に含まれると思われる領域を入力する事が可能となる。よってシステムがデータベース中に含まれる代表的な領域を提示することにより、検索が行える。そのため、ユーザはシステムから提示される部分画像を見て、自分のイメージに近い画像を選択するだけで画像を検索できる。

第 3 章

ビジュアルキー型画像検索における 画像の部分領域の処理や検索の流れ

本章では、ビジュアルキー型画像検索の検索方法や検索で使われる特徴抽出、クラスタリングといった基本概念について説明する。

3.1 ビジュアルキー型画像検索

ビジュアルキー型画像検索は、部分画像をクエリとして用いる RBIR の 1 つである。ビジュアルキー型画像検索は、画像データベース中の画像全てを部分領域に分割し、そこで得られた部分領域の画像の特徴を抽出し、クラスタリングにより複数の集合に分類する。各集合において特徴が平均値に最も近い画像を代表とし、これをビジュアルキーとする。ここで得られたビジュアルキーは、テキスト検索におけるキーワードの役割を果たす。データベース中の全画像はビジュアルキーにより索引付けされており、ユーザがビジュアルキーをクエリとして生成した時に、データベース側では、ビジュアルキーによって索引付けられている画像を検索結果としてユーザに返す。図 3.1 にその流れを示す。図 3.1 において、ユーザはビジュアルキー群から 2 枚のビジュアルキーを選定し、クエリを構成することで検索を行っている。システム側では、クエリが含んでいる各ビジュアルキーと索引付けされている画像を画像データベースより候補としてあげる。図中では、クエリで構成した青のビジュアルキーと緑のビジュアルキーに索引付けされている候補画像のうち、赤色の枠で囲まれている画像が共通する画像であるため、共通した画像 3 枚が検索結果としてユーザに表示される。

3.1 ビジュアルキー型画像検索

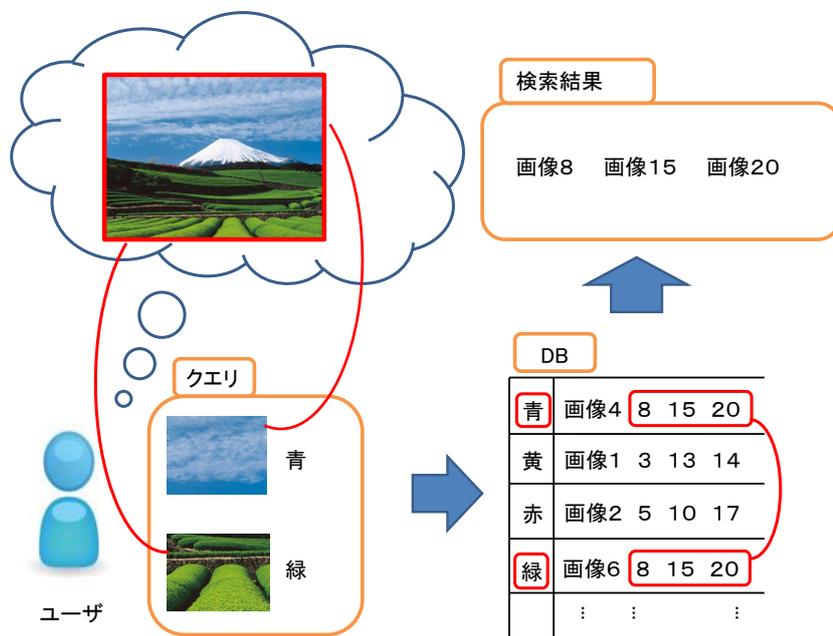


図 3.1 ビジュアルキー型画像検索のシステムの仕組み

3.1 ビジュアルキー型画像検索

3.1.1 ビジュアルキーの選定方法

本節では、ビジュアルキーに選定するまでの処理を示す。この処理を図 3.2 に示す。

1. 画像データベース上にある画像全てを部分領域に分割する。
2. 部分領域の画像の特徴を抽出する。
3. 抽出した特徴に対し、クラスタリングを行う。
4. クラスタリングの結果から得られた各クラスタの代表画像の集合を算出し、これをビジュアルキーとして選定する

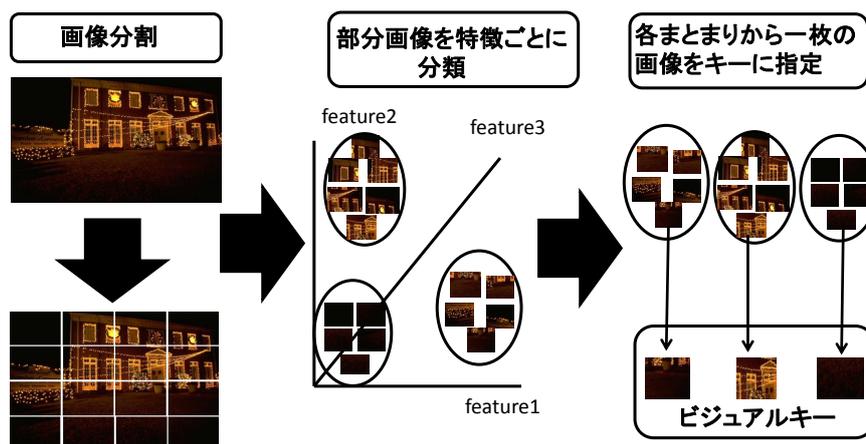


図 3.2 ビジュアルキー選定

3.1 ビジュアルキー型画像検索

3.1.2 特徴抽出

本節では，ビジュアルキー型画像検索に用いられる特徴抽出について述べる．画像から抽出した特徴は，大きく分けて，色特徴である色モーメント法，テクスチャ特徴である Texture neighborhood，形状特徴である 8 方向 Sobel フィルタの 3 種類がある．本手法では，色モーメント法の $L^*a^*b^*$ 色空間の平均，分散，歪度の 9 次元ベクトルを用いる．以下で，色モーメント法，Texture neighborhood，8 方向 Sobel フィルタについて述べる．

3.1.3 色モーメント法

色モーメント法では，表色系は $L^*a^*b^*$ を用いており，色空間上の各チャンネルごとに，画素値の平均 μ ，標準偏差 σ ，歪度 γ を算出しており，以下の式で求めることができる．

$$\mu_{P_c} = \frac{1}{WH} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} P_c(x, y) \quad (3.1)$$

$$\sigma_{P_c} = \sqrt{\frac{1}{WH} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} (P_c(x, y) - \mu_c)^2} \quad (3.2)$$

$$\gamma_{P_c} = \sqrt[3]{\frac{1}{WH} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} (P_c(x, y) - \mu_c)^3} \quad (3.3)$$

W は画像の横， H は画像の縦， P_c は対象の部分画像である．これらを用いて，色特徴は 9 次元のベクトルとしている．

3.1.4 Texture neighborhood

Texture neighborhood は L^* (luminance) チャンネルにおいて，全画素 P_L と 8 近傍画素 P_L^d を比較し，ある画素に対しそれぞれの近傍の画素値がその画素値の値より大きくなる確率により次のように算出している． δ は閾値とし， $\delta = 0.015$ である．

3.1 ビジュアルキー型画像検索

$$t_d = \frac{1}{WH} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} G(P_L, P_L^d + \delta) \quad (3.4)$$

$$G(a, b) = \begin{cases} 1 & \text{if } a > b \\ 0 & \text{otherwise} \end{cases} \quad (3.5)$$

これらを用いて 8 次元のテクスチャ特徴ベクトルとしている

3.1.5 8 方向 Sobel フィルタ

8 方向 Sobel フィルタでは、画像を HSV 色空間に変換し、色相 (H) チャンネルを除いた、彩度 (S)、明度 (V) に対して 8 つの Sobel フィルタ演算を行う。S チャンネル、V チャンネルに対してフィルタリングをかけた $f_{Sd}(x, y)$ と $f_{Vd}(x, y)$ を用いてフィルタリング方向 $d(d \in 0, \pi/4, \pi/2, 3\pi/4, \pi, 5\pi/4, 3\pi/2, 7\pi/4)$ の形状特徴 e_d は次のように求める。

$$e_d = \frac{1}{WH} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} G(f_{Sd}(x, y), \gamma_S \max_{d,x,y} f_{Sd}(x, y)) | G(f_{Vd}(x, y), \gamma_I \max_{d,x,y} f_{Vd}(x, y)) \quad (3.6)$$

ここで、 $|$ は論理和演算を意味しており、例えば $G(f_{Sd}(x, y), \gamma_S \max_{d,x,y} f_{Sd}(x, y))$ が $G(f_{Vd}(x, y), \gamma_I \max_{d,x,y} f_{Vd}(x, y))$ の値が 1 の時に式 3.6 の値は 1 となる。また、閾値 γ_I と γ_S 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としている。

3.1.6 特徴ベクトルの削減

従来のビジュアルキー型画像検索は色、テクスチャ、形状の 25 次元の全ての特徴を用いてクラスタリングを行っている。クラスタリングにより得られるクラスは、ユークリッド距離を用いて全てのベクトルの値を統合し、重心に近いビジュアルキーをクラスタに入れているため、クラスタリングによりできるクラスは、色特徴の一部のベクトルとテクスチャ特徴の一部のベクトルが類似し、同じクラスタに所属させることが起こり得る。また、提示されるビジュアルキーは、クラスタリングされた代表のビジュアルキーのみであり、どの特徴により算出されたかはユーザは分からない。そのため、色特徴として選んだビジュアル

3.1 ビジュアルキー型画像検索

キーが形状特徴により算出されていたということが起こる．よって本研究では，特徴ベクトルを削減し，人間の目で見えて似ているか，似ていないかを判断しやすい色特徴のみを用いてクラスタリングを行っている．

3.1.7 クラスタリング

本節では，ビジュアルキー型画像検索にも用いられているクラスタリングについて述べる．クラスタリングとは，対象間の距離情報を基にし，距離の近さによって対象の分類を行う統計手法の1つである．クラスタリングはデータ分類手法であり，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，事前に情報を与え，それによりデータを分類する．教師なしクラスタリングは，事前に教師となるデータがなく，データの要素の範囲により分類を行う．ビジュアルキー型画像検索では，分割した画像の数に加え，各画像が色や形状などの多くの次元を持つため，クラスタが満たすべき条件を事前に与えることは困難であるため，教師なしクラスタリングを用いている．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．以下に階層的クラスタリングである Ward 法について説明する．

3.1.8 Ward 法

本研究では，類似した部分画像の分類法として一般的に分類感度が高いとされている階層的クラスタリングの Ward 法を用いる．階層的クラスタリングの処理の初期状態では，クラスタリング範囲である全ての要素を各クラスタとする．各クラスタ間の距離を求め，最も近いものを結合し新しいクラスタとする．この処理を全てのクラスタが1つのクラスタに結合されるまで繰り返す．以下に Ward 法のアルゴリズムを示す．

1. 各要素をクラスタとする．
2. 各クラスタの重心を計算する．
3. 各クラスタごとに要素の重心との距離の差を求める．

3.1 ビジュアルキー型画像検索

4. 全てのクラスタの結合した場合の要素の重心との距離の差を求める。
5. 最も重心の変化がないクラスタ同士を結合する。
6. クラスタ数が1になると終了。そうでなければ2の処理に戻る。

Ward 法では、各要素をクラスタとし、クラスタ G における重心と各要素との距離の差を求める。群内平方和 $E(G)$ を

$$E(G) = \sum_{x \in G} \|x_i - M(G)\|^2 \quad (3.7)$$

で算出する。

この群内平方和を用いたクラスタ G_i と G_j の群内平方和の増加量 $\Delta E(G_i, G_j)$ が最も少ないクラスタ同士を結合していく。群内平方和の増加量 $\Delta E(G_i, G_j)$ は以下の式で定義されている。

$$\Delta E(G_i, G_j) = E(G_i \cup G_j) - E(G_i) - E(G_j) \quad (3.8)$$

これをクラスタ数が1になるまで繰り返す。

3.1.9 画像検索の流れ

本節では、ビジュアルキー型画像検索による検索の流れを以下に示す。

1. ユーザは、自分のイメージする画像に対して提示されているビジュアルキーを選択し、クエリを構成する。
2. 各ビジュアルキーに索引付けされている画像全てをデータベースより候補としてあげる。
3. データベースより候補としてあげられた画像の中で各ビジュアルキーの共通する画像を検索結果としてユーザに提示する。

図 3.3 は、ビジュアルキー型画像検索の検索画面であり、図 3.4 は、検索結果の実例をあげている。

3.1 ビジュアルキー型画像検索



図 3.3 検索画面



図 3.4 検索結果

第 4 章

位置情報を考慮した部分画像検索

本章では、位置情報を考慮した部分画像検索の説明とシステムの構成を述べる。

4.1 位置情報を含むクエリを構成する検索システム

従来手法のビジュアルキー型画像検索と位置情報を含むビジュアルキー型画像検索の仕組みの違いについて図 4.1 と図 4.2 を用いて説明する。図 4.1 は従来手法のビジュアルキー型画像検索の概念を示したものである。図の左側は、ユーザがビジュアルキー 1 をクエリとして生成したものであり、これを検索にかけると、図の右側の検索結果にあるようなビジュアルキー 1 を右上に含む画像 1 や中中付近に含む画像 2 といったビジュアルキー 1 を含む全ての画像が検索結果で表示される。図 4.2 は、位置情報を含むビジュアルキー型画像検索の概念を示したものである。図の左側は、ユーザが 25 個のブロックの中から右上にビジュアルキー 1 をクエリとして生成したものであり、これを検索にかけると、図の右側の検索結果にあるようなビジュアルキー 1 を右上の位置に含む画像 1 は適合画像とし、ユーザに出力結果として提示されるが、画像 2 は同じ要素を含んでいるものの同じ位置に索引付けされていないため、不適合画像となり、検索結果として出力されない。図 4.3 に位置情報を含むビジュアルキー型画像検索のシステムの画面を示す。画面の左の画像は実験に用に被験者に提示する目的画像である。ユーザはシステムの右側に表示されている 20 個のビジュアルキーをシステムの下部にある No. 1 から No. 25 の画像を 5 × 5 に分割した 25 個のブロック体の位置に指定する。ビジュアルキーは各ブロックに指定可能なため、ブロックの数だけ指定可能である。システムは、世界中で最も普及しているオープンソフトである Apache を

4.1 位置情報を含むクエリを構成する検索システム

Web サーバとして用い、ビジュアルキーと画像データベースの索引付けは SQLite により構築し、これらを実行させるための Web サーバとデータベースとの通信には PHP を用いた (図 4.4) .

4.1 位置情報を含むクエリを構成する検索システム

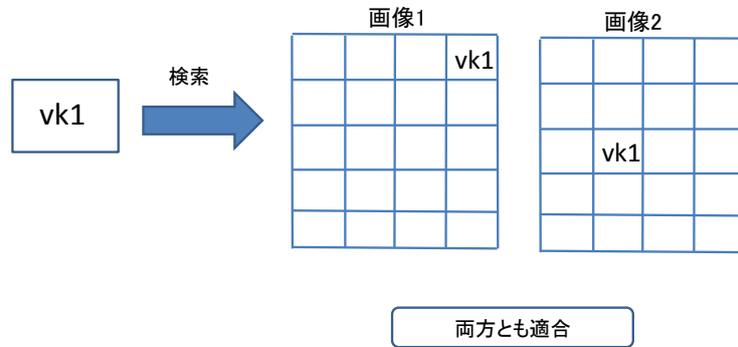


図 4.1 ビジュアルキー型画像検索の仕組み

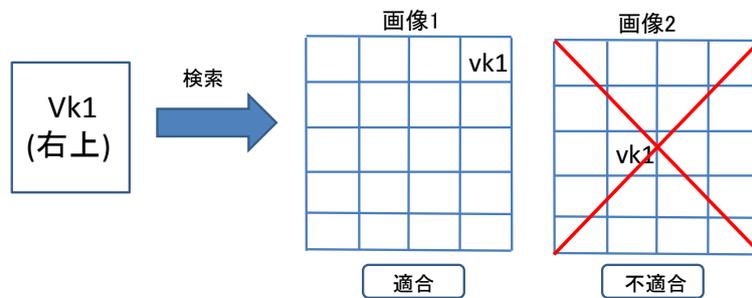


図 4.2 位置情報を考慮した部分画像検索の仕組み

4.1 位置情報を含むクエリを構成する検索システム

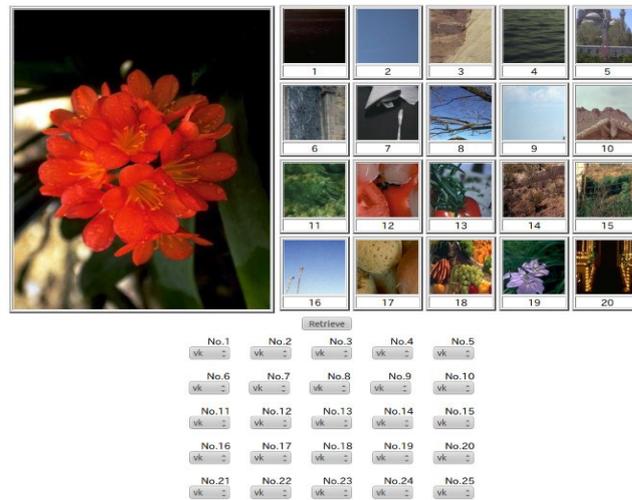


図 4.3 位置情報を考慮したビジュアルキー型画像検索

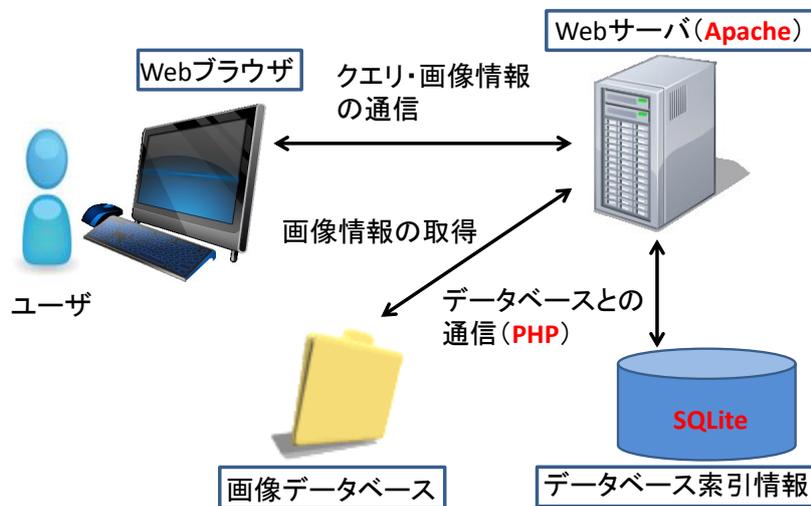


図 4.4 システムの概念図

4.1 位置情報を含むクエリを構成する検索システム

4.1.1 ビジュアルキーの生成

本節では、ビジュアルキーの生成方法について述べる。ユーザに提示するビジュアルキーはデータベース内の画像により生成する。ユーザに提示するまでの生成の過程は、画像の分割、分割された各部分領域の特徴抽出、クラスタリング、ビジュアルキーの選定により行う。以下でその流れを示す。

4.1.2 分割の流れ

本節では、画像分割の処理について述べる。ビジュアルキー型画像検索の画像の部分領域は、データベース内の画像をそれぞれ指定の分割数で分割し、生成する。本研究では、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を調べるため、画像をそれぞれ 2×2 から 5×5 までの分割を行った。また、画像の分割を行う際、データベース内の各画像は画像の大きさが違うため、データベースの全ての画像を 480×480 ピクセルの同じ大きさに揃えた。画像を分割するまでの処理を図 4.5 に示す。

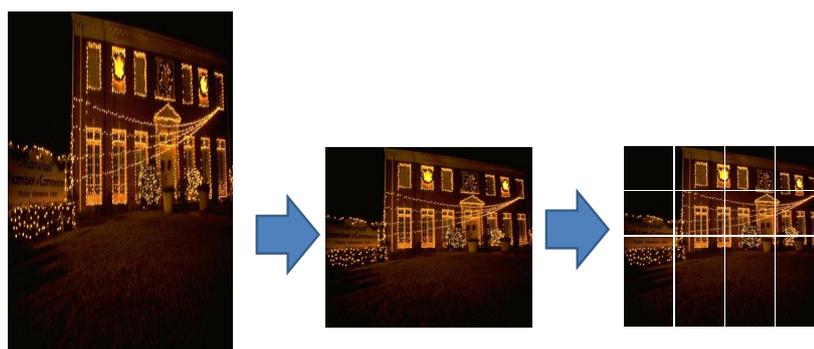


図 4.5 画像を分割するまでの流れ

4.1 位置情報を含むクエリを構成する検索システム

4.1.3 ビジュアルキーの選定

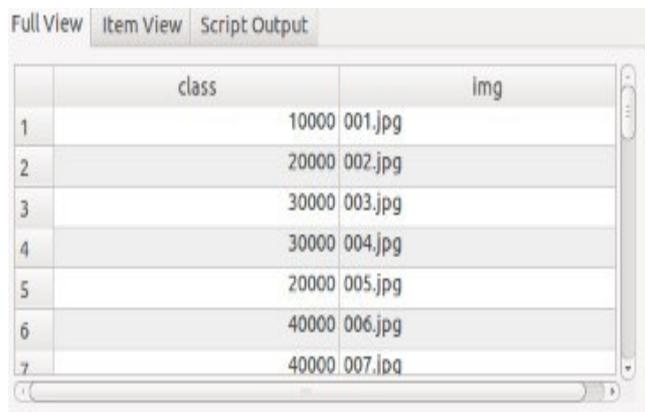
本節では，等分割された画像からビジュアルキーに選定するまでの処理を示す．

1. 画像を各 2×2 から 5×5 まで等分割する．
2. 各部分領域の色特徴である $L^*a^*b^*$ 色空間の平均，分散，歪度により特徴を抽出する．
3. 抽出した特徴に対し，階層的クラスタリングの Ward 法によりクラスタリングを行う．
4. 各クラスタの最も平均値に近い部分領域をビジュアルキーとして選定する．

4.1.4 位置情報の付加処理

本節では，データベースにおける画像と位置情報のインデクシングについて述べる．SQLite により構築されたデータベースでは，クラスタリングにより分類された部分領域のクラス，分割された各ブロックにおける位置情報，画像データベースにある画像の情報を索引付けする．位置情報の索引付けを図 4.6 と図 4.7 に示す．図 4.6 はデータベースのインデックス情報である．データベースの class の 4 桁目までの数値には，ブロック番号と照らし合わせた位置情報を書き込み，5 桁目以降をクラスタリングにより分類された部分領域のクラスを書き込む．また，データベースの img には画像データベースにある原画像の情報を書き込み，これらを索引付けしている．図 4.7 に画像を 4×4 に分割した場合のブロック番号と位置情報を示す．

4.1 位置情報を含むクエリを構成する検索システム



The screenshot shows a software window with three tabs: 'Full View', 'Item View', and 'Script Output'. The 'Item View' tab is active, displaying a table with two columns: 'class' and 'img'. The table contains seven rows of data, with the first row highlighted in grey.

	class	img
1	10000	001.jpg
2	20000	002.jpg
3	30000	003.jpg
4	30000	004.jpg
5	20000	005.jpg
6	40000	006.jpg
7	40000	007.jpg

図 4.6 インデックス情報

① 0000	② 0001	③ 0002	④ 0003
⑤ 0100	⑥ 0101	⑦ 0102	⑧ 0103
⑨ 0200	⑩ 0201	⑪ 0202	⑫ 0203
⑬ 0300	⑭ 0301	⑮ 0302	⑯ 0303

図 4.7 クラスと位置情報

4.1 位置情報を含むクエリを構成する検索システム

4.1.5 画像検索の流れ

本節では、画像検索の流れを以下に示す。

1. クラスタの数 ($k=20$) を選定したビジュアルキーをユーザに提示する。
2. ユーザはイメージしている画像に対し、ビジュアルキーをブロック番号に指定し、クエリを生成する。
3. クエリを構成するビジュアルキーに索引付けされている画像全てをデータベースより候補としてあげる。
4. データベースより候補としてあげられた画像の中で各ビジュアルキーの共通する画像を検索結果としてユーザに提示する。

図 4.8 で位置情報を考慮したシステムの仕組みの図を示す。風景の画像を目的画像とした時、ユーザは 3×3 のブロックの右上の領域に空をイメージした青色のビジュアルキーを指定し、中下の領域に緑色のビジュアルキーを指定し、クエリを構成する。システム側では、クエリが含んでいる各ビジュアルキーと索引付けされている画像を画像データベースより候補としてあげる。図中では、クエリで構成した右上の領域にある青のビジュアルキーと中下の領域にある緑のビジュアルキーに索引付けされている候補画像のうち、赤色の枠で囲まれている画像が共通する画像であるため、共通した画像 2 枚が検索結果としてユーザに表示される。

4.1 位置情報を含むクエリを構成する検索システム

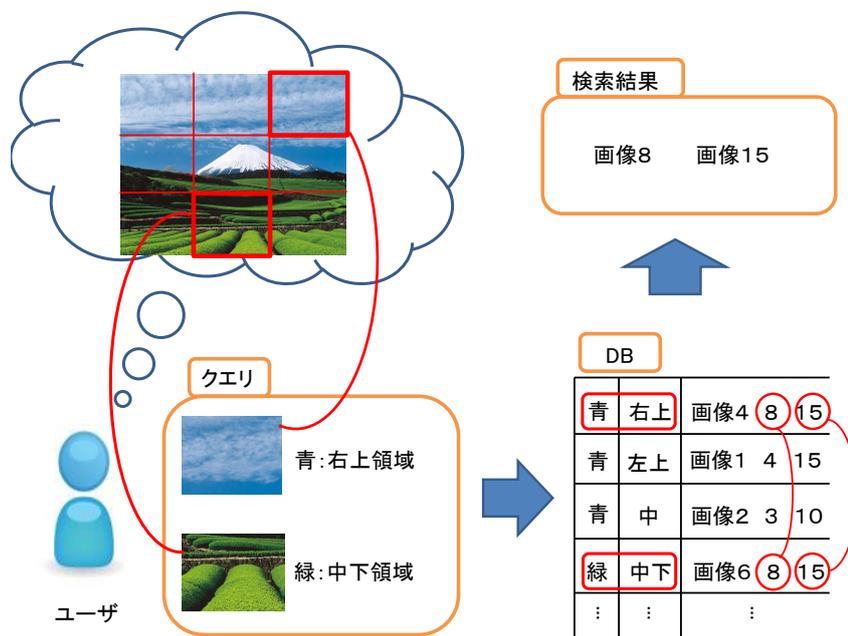


図 4.8 位置情報を考慮したシステムの仕組み

第 5 章

分割粒度の検索精度に与える影響

豊田らは、 3×3 の位置情報を含むビジュアルキーによりクエリを構成するシステムを提案している。しかし、 3×3 に限ってのみ検索精度を明らかにしていることから、分割数による影響は明らかにされていない。本研究では、画像の分割数による影響から最適な分割数を見だし、精度向上を目指す。本章では、従来のビジュアルキー型画像検索と、提案手法の性能評価を行う。検索精度については、被験者実験により得られた適合率、再現率、F 値により評価する。以下に実験環境と実験結果、考察を示す。

5.1 実験環境

比較の対象は従来のビジュアルキー型画像検索と位置情報を考慮した画像検索の 2×2 , 3×3 , 4×4 , 5×5 のシステムである。なお、従来のビジュアルキー型画像検索の適合率、再現率、F 値はすでに検証されている適合率 11%、再現率 24%、F 値 0.15 を用いる。データベースには ArtExplision 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 (図 5.1)。被験者に検索してもらう目的画像には ArtExplision の 10 カテゴリから各 1 枚を抽出した画像 10 枚を用いる (図 5.2)。被験者に提示するビジュアルキーは人間の目で見えて似ているか似ていないかを判断しやすい色特徴にクラスタリングを行い選定している。ビジュアルキーの枚数は 20 枚であり、ユーザはブロック番号にビジュアルキーを配置して検索を行う。ただし、ブロック全てに画像データベースと索引付けられたビジュアルキーを完全に一致させて配置するのは困難なため、今回の実験ではビジュアルキーの配置数を 1 枚のビジュアルキーを選択した場合と 2 枚のビジュアルキーを選択した場合で検証す

5.1 実験環境

る．ビジュアルキーの配置する位置に関しては，ブロック内において自由であり，2枚とも同一のビジュアルキーでも配置位置が異なっていれば選択してもよいものとする．検索精度については，適合率，再現率，F 値の被験者全体での平均を求め，比較した．実験の条件を図 5.3 で示す．

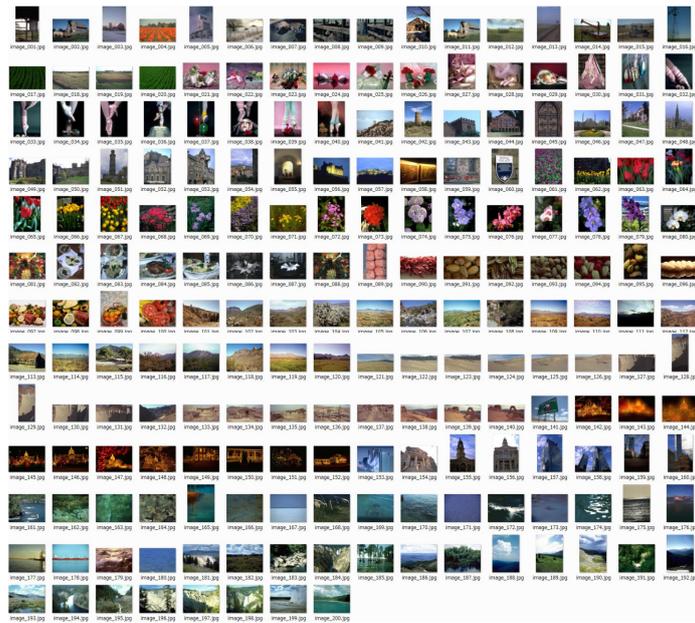


図 5.1 画像データベース

5.1 実験環境



図 5.2 検索目的画像

被験者	10名
検索目的画像	ArtExplosion10カテゴリから各1枚抽出した画像10枚(画像データベースには10カテゴリから各20枚抽出した画像200枚)
ビジュアルキー配置数	1枚のビジュアルキーを選択した場合と2枚のビジュアルキーを選択した場合を検証
ビジュアルキーの配置	ブロック内で自由に配置(2枚とも同一のビジュアルキーでも配置が異なっていれば選択可能)
比較対象	従来手法によるビジュアルキー型画像検索の平均適合率(約11%)と本手法である $2 \times 2, 3 \times 3, 4 \times 4, 5 \times 5$ を比較
適合画像	検索結果のうち被験者が目的画像と類似すると判断した画像
適合率	検索結果の適合画像数を検索結果に表示された画像数で割る
再現率	検索結果の適合画像数を全画像中(データベース200枚)の適合画像数で割る

図 5.3 実験の条件

5.1 実験環境

5.1.1 検索精度

検索精度は被験者実験で得られた適合率，再現率，F 値により評価している．適合率とは，検索結果の画像中の適合画像の数の割合である．検索結果の適合画像数を R ，検索された画像数を N とした場合，適合率 (Precision) は以下の式で算出される．

$$\text{Precision} = \frac{R}{N} \quad (5.1)$$

再現率とは，検索範囲の中にある全適合画像の数と検索結果として得られた適合画像の数の割合である．検索範囲全ての適合画像数を C とした場合，再現率 (Recall) は以下の式で算出される．

$$\text{Recall} = \frac{R}{C} \quad (5.2)$$

F 値とは，適合率と再現率の調和平均によって導かれた情報検索の精度を測る指標である．F 値 (F-measure) は以下の式で算出される．

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5.3)$$

5.1.2 実験結果

被験者 10 人に目的画像 10 枚を提示し，それぞれの画像に対して検索を行った．結果はビジュアルキーの配置数が 1 枚の場合 3×3 の分割数の検索精度が高かった．算出した平均の値は適合率 32%，再現率 39%，F 値 0.35 となった．また，ビジュアルキー 2 枚を選択した場合には 5×5 の分割数の検索精度が高かった．算出した平均の値は適合率 36%，再現率 33%，F 値 0.35 となった．従来手法 11% に比べ 25 ポイントの精度向上となった (図 5.4)．ビジュアルキー 1 枚の場合の適合率，再現率，F 値の平均と標準誤差を算出したものを図 5.5 に示し，ビジュアルキー 2 枚の場合を図 5.6 に示す．また，各画像ごとの平均適合率を図??に示す．

5.1 実験環境

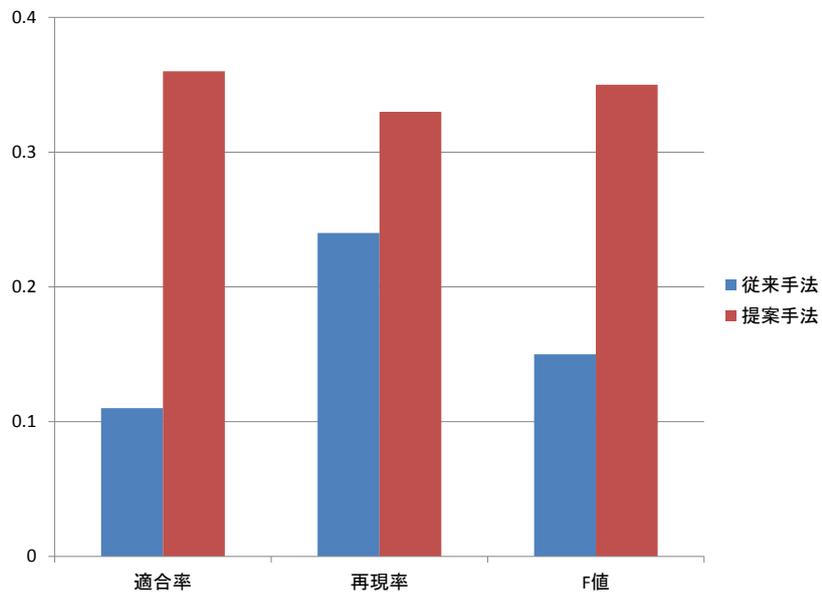


図 5.4 従来手法との比較

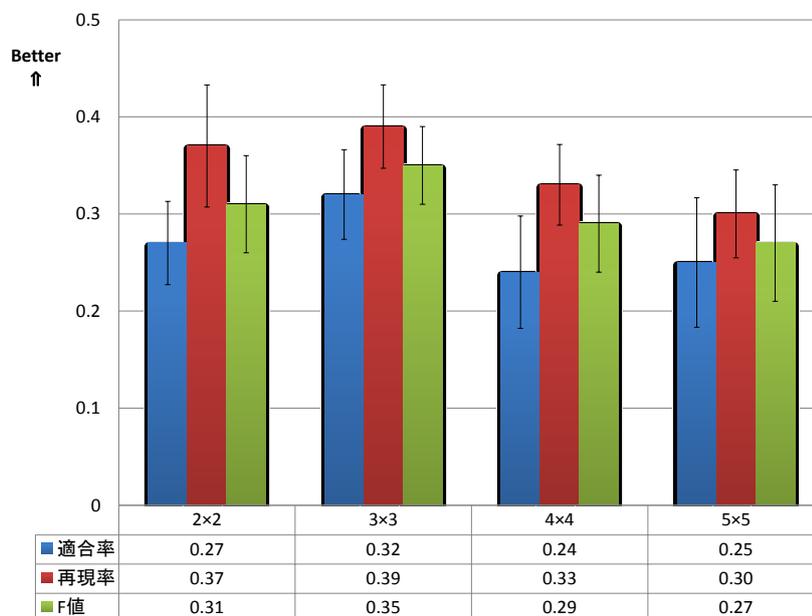


図 5.5 ビジュアルキー 1 枚の時の実験結果

5.1 実験環境

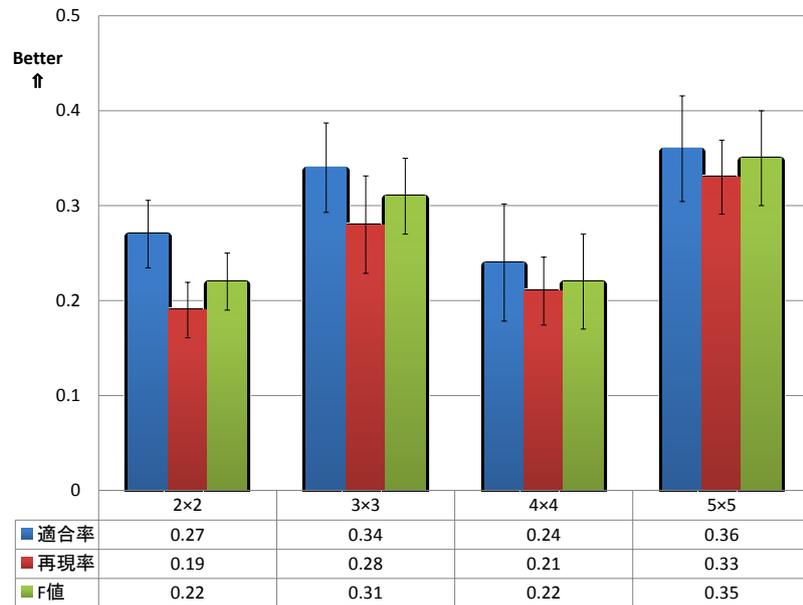


図 5.6 ビジュアルキー 2 枚の時の実験結果

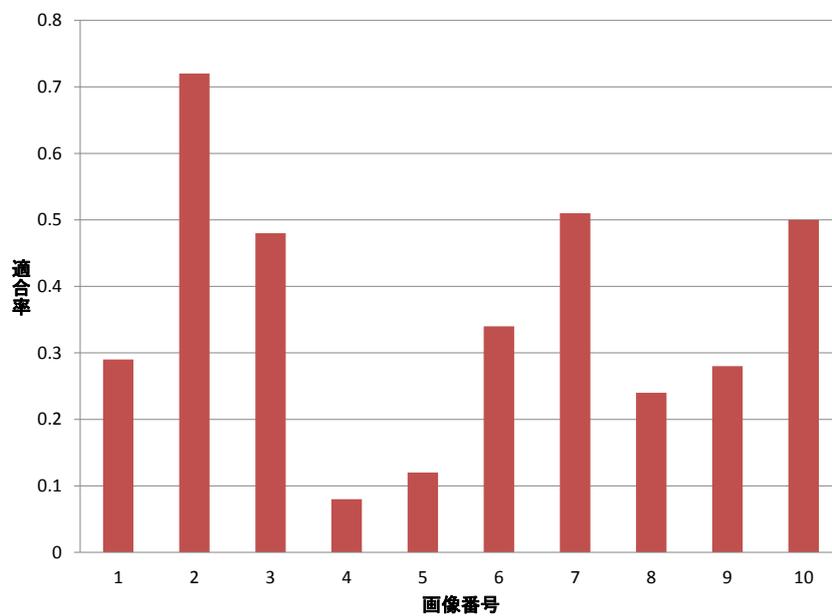


図 5.7 各画像ごとの平均適合率

5.1 実験環境

5.1.3 考察

平均適合率は 5×5 の分割数の 36% が最大となり、従来手法に比べて 25 ポイントの優位性がある結果となった。この適合率の向上に繋がった理由として、適合率を求める計算は、検索結果のうち被験者が目的画像に類似していると判断した画像を検索結果に表示された画像数で割ることにより算出できるため、本システムにより検索結果の出力画像枚数を絞り込みつつ、適合画像を算出できたからだと考えられる。しかし、図 5.7 のように画像によって適合率の差が激しいという問題も見えた。特に適合率の低かった画像は図 5.2 の 4 番と 5 番である。この 2 枚の目的画像は被験者のクエリの生成時間も他の画像に比べ、長い時間を要していた。原因として考えられるのは 2 通りある。1 つ目はビジュアルキーに黄色単体のものがなかったため、被験者は間違ったビジュアルキーを選択していたことが原因と考えられる。2 つ目はビジュアルキーの配置位置の誤りだと考えられる。例えば画像 4 番を検索する際、被験者は紫の花の色を配置したつもりが、実際はその位置には葉の緑のビジュアルキーが属するといったものである。この被験者と画像における視覚的ギャップがビジュアルキー選択の誤りの原因となり、画像ごとの適合率の差や検索結果に画像が出力されないといった結果に繋がったと考えられる。このような傾向は、分割数が少ない場合に生じていた。分割数が多い程、被験者は的確な場所に配置できていた。このことから分割数をこまめにするこゝとで被験者とシステム間のビジュアルキーの捉え方のギャップを少なくでき、正確なビジュアルキーの配置ができると考える。 4×4 や 5×5 といった分割数になってくると分割されただけ繊細にビジュアルキーを当てはめていかなければならないので検索精度は悪くなると筆者は考えたが、実験を行った被験者から分割をこまめにした方が的確なビジュアルキーを配置しやすく、検索結果の画像枚数が多すぎないため目的画像も見つけやすいという意見があった。一方で 2×2 などの単純な分割数の場合は、分割領域が大きいので的確なビジュアルキーを指定しにくい。また、検索結果の出力枚数が多いため、目的画像を探し出すのが困難であり、適合率の結果を落としたといえる。そのほか、実験結果としてビジュアルキーが 1 枚の場合は 3×3 の精度が高く、ビジュアルキーが 2 枚の時は 5×5 の精度が高かった。

5.1 実験環境

このようにビジュアルキーの配置数が実験結果に大きく影響を及ぼしていると考えられる。しかし、ビジュアルキー 1 枚の時は絞り込みがほぼされず、検索結果に多くの画像が出力されるため本システムを扱う場合、最低 2 枚のビジュアルキーを配置する必要があると考えられる。また、実験結果から 3×3 や 5×5 などの奇数の分割数の結果が良かった理由として、画像の中心にキーポイントができるので人の目で見て画像の位置が分かりやすかったからだと考えられる。目的画像には分割された線が表示されていないため、分割数が偶数の 2×2 や 4×4 などは、位置のずれによりビジュアルキーの配置を誤っている可能性が高いと考えられる (図 5.8)

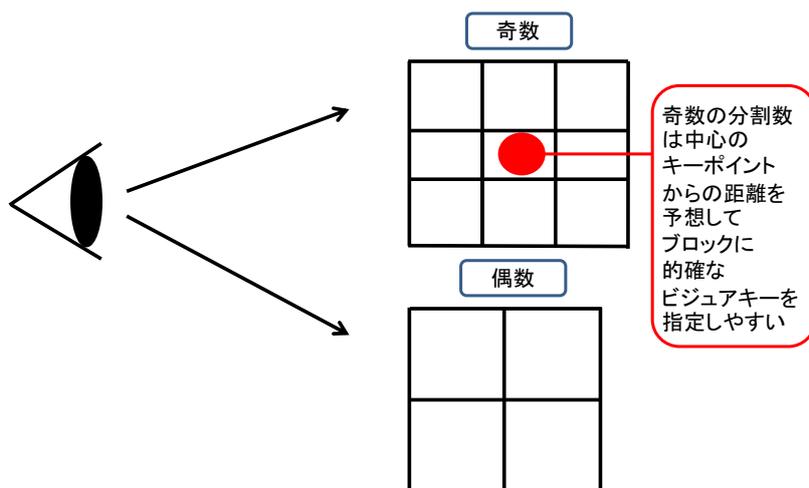


図 5.8 奇数と偶数の分割数の違い

第 6 章

結論

位置情報を含むビジュアルキー型画像検索の分割数を変更した場合の検索精度の変化を検証することで、ビジュアルキーの生成方法についての基礎的な知見を得ることができた。また、等分割の場合の最適分割数を見いだせたことで、従来のシステムと比較して検索精度が向上していることも確認している。これらはクラスタリングにより分類されたクラスとビジュアルキーを配置するための 2×2 , 3×3 , 4×4 , 5×5 の位置に合わせた情報と画像データベースにある原画像情報との索引付けしたデータベースを構築することで実現している。実験結果から位置情報を含むビジュアルキーによりクエリを構成することで、検索精度の向上が実現しているのが分かる。理由としては 2 つ考えられる。1 つ目は、検索結果に似たような画像が多く出力される場合でも、ビジュアルキーを配置する位置と数を考慮することで、検索結果の絞り込みがなされたからである。2 つ目は、分割により絞り込むことで 1 つ 1 つの情報は単純になり、画像全体で検索を行うよりはビジュアルキーを指定しやすくなったからだと考える。検索精度を検証するため、適合率、再現率、F 値の被験者全体の実験結果の平均を求め、比較した。提案手法は従来手法のビジュアルキー型画像検索の適合率 11%、再現率 24%、F 値 0.15 という結果をどれも上回っていた。また最適な分割数としては実験結果から 5×5 の分割数だということが分かった。 5×5 の分割数の検索精度が高かった理由としては、細かい分割によりユーザとシステム間のビジュアルキーの捉え方のギャップを少なくでき、正確なビジュアルキーの配置を行えたことで、システムに与える情報を絞り込めたからだと考える。一方で 2×2 などの単純な分割数の場合は、分割領域が大きいので的確なビジュアルキーを指定しにくい。また、検索結果の出力枚数が多いため、目的画像を探し出すのが困難であり、適合率の結果を落としていた。 5×5 の分割数の実験

結果の平均は適合率 36% , 再現率 33% , F 値 0.35 である . これにより , 従来のビジュアルキー型画像検索と比べ , 25 ポイント適合率が向上する結果となった . また , 6 × 6 以上の分割数では , ビジュアルキーの組み合わせによっては検索結果に画像が 1 枚も出力されない場合が増えるのでこれは今後の課題といえる . 課題解決のためには , 検索結果に柔軟性を持たせる必要がある . 本研究で用いた階層的クラスタリングの Ward 法は , ある画像の特徴が唯一のクラスタに完全に属しているため検索結果に柔軟性はない . よって , 複数のクラスタにある程度ずつ属るといったようなファジィクラスタリングを用いてクラスタへの所属の度合いをあいまいに表すことで , ビジュアルキーとインデクシングに幅を持たせ , 検索結果に柔軟性を持たせることができると考える .

謝辞

本研究を進めるにあたり、ご指導をして頂きました高知工科大学 情報学群 吉田 真一講師には大変お世話になりました。吉田先生には、週 2 回の研究の進捗の場を設けていただき、これが大変効率の良い方法で非常に研究が捗りました。うまく結果がでない時が多々ありましたが、その度に御知恵をお貸しいただき、熱心なご指導をしていただきました。振り返れば吉田先生には終始頼りっきりだったかのように思います。今回の研究がまとまり、論文を書き終えたのも吉田先生の御蔭であると感じており、御指導に深く感謝致します。また、吉田先生には研究活動以外の面でも大変お世話になりました。学校生活や就職活動、課外活動や飲み会の場などで様々な助言や礼儀なども教えていただきました。吉田先生の今後の IT 社会のユニークな話や今までやってきたこと全てに意味があると色々なことに関連付けして話して頂いたこと、そして就職活動の際に心が折れた時には親身になって話を聞いていただいた御蔭で自信を持って行動でき、無事に就職先も決めることができたと思っております。研究室配属から約 2 年間丁寧かつ熱心なご指導をしていただき、多くのことを学びました。深く感謝致します。本研究の副査を担当していただきました、高知工科大学 情報学群 福本昌弘教授ならびに高知工科大学 情報学群 植田 和憲講師に深く感謝致します。急なお願いにも関わらず、快く副査を引き受けていただきました。お忙しい中、植田先生には論文の御指導、助言をしていただきました。御蔭様でより良い卒業論文を作成することができました。福本先生には卒業研究の中間発表の合宿や研究室合同の芋煮会でも大変お世話になりました。深く感謝致します。同研究室の同期である岡尾 浩気氏、小野 陽平氏、小池 規伎氏、篠原 悠貴氏、藤森 夏輝氏、山口 祐輝氏には、研究を共に行い、苦楽を共にしてきました。ここまで進んでこられたのもお互いが切磋琢磨しながら研究に励み、困った時には必ず手助けをするといった非常にチームワークのいいメンバーだったからだと思います。また、学校以外でも同期とは交流が深く、皆様と過ごした学校生活は楽しいものでした。同じ研究室で過ごせたことを大変嬉しく思います。これから皆様は社会人や大学院に進学とそれぞれの道を

謝辞

歩んでいくことになりましたが、今以上に頑張ってください。同研究室の3年生の方々には、持ち前の明るさで大変元気づけられました。また、飲み会の幹事などをやっていただき大変助かりました。礼儀の振る舞いや、まめに気が付いたり、安心して任せられました。これからも個性を大切に頑張っていってください。また、被験者実験に協力していただいた方々に深く感謝致します。最後に、生活や学費の面で支えてくれた家族に心から感謝致します。

参考文献

- [1] M.Serata et al., "Designing Image Retrieval System with the Concept of Visual Keys", JACIII, 10(2).
- [2] 多田 昌裕, 加藤 俊一, "階層的分類を用いた視覚感性のモデル化と類似画像検索への応用", 情報処理学会論文誌, 2003.
- [3] 宮本 定明, "クラスター分析入門 ファジィクラスタリングの理論と応用", 森下出版株式会社, 1999.
- [4] 橋詰 翔健, "遺伝的プログラミングを用いた集約演算子の最適化と画像検索への応用", 平成 21 年度学士学位論文.
- [5] 藤田 雄大, 野口 達也, 石田 崇, 平澤 茂一, "色情報に対する人間の感性を考慮した類似画像検索", 情報科学技術フォーラム一般講演論文集, 2007.
- [6] 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.
- [7] 前田 豊文, "ビジュアルキー型画像検索のためのクラスタリング法の提案", 平成 21 年度学士学位論文.
- [8] E.Di Sciascio,G.Mingolla,M.Mongiello, "Content-based Image Retrieval over the Web using Query by sketch and Relevance Feedback"
- [9] 豊田 佑介, "画像の構成を考慮した部分領域に基づく画像検索", 平成 23 年度学士学位論文.

付録 A

被験者ごとの平均適合率

図 A に被験者ごとの平均適合率を示す。人によって画像の見方は変わってくるため、被験者によって適合率のばらつきはあるが、全体を平均すると 5 × 5 の精度が高いことが分かる。

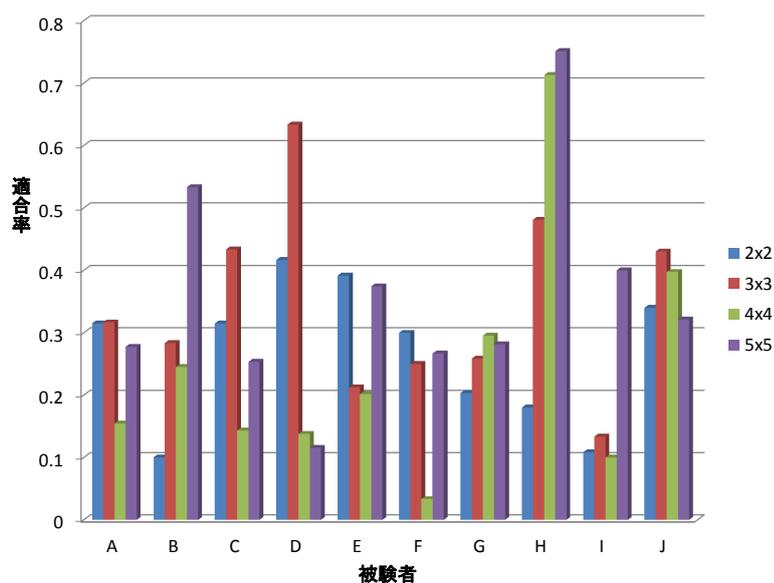


図 A.1 被験者ごとの平均適合率

付録 B

被験者 1 人 1 人に対する画像ごとの適合率

図 B に被験者 1 人 1 人に対する画像ごと画像ごとの適合率を示す。10 人の被験者の結果を見ると、画像番号 4 と 5 の適合率が全体的に低いことが分かる。これは、黄色単体のビジュアルキーがなかったため、ユーザによるビジュアルキーの選択ミスが原因と考えられる。

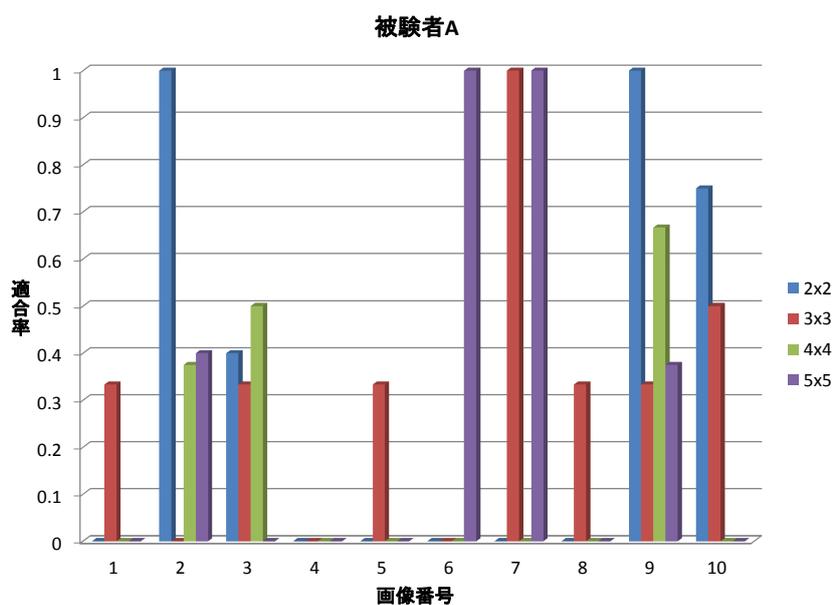


図 B.1 被験者 A

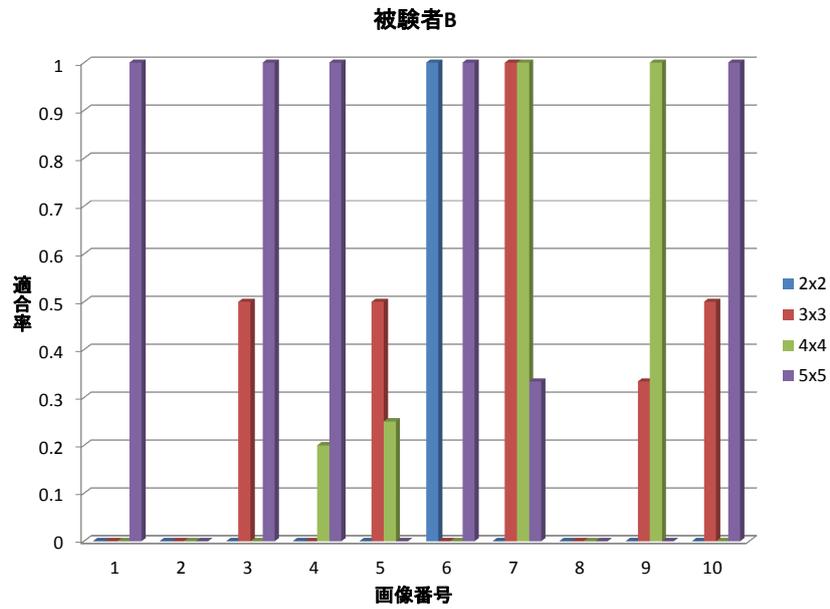


図 B.2 被験者 B

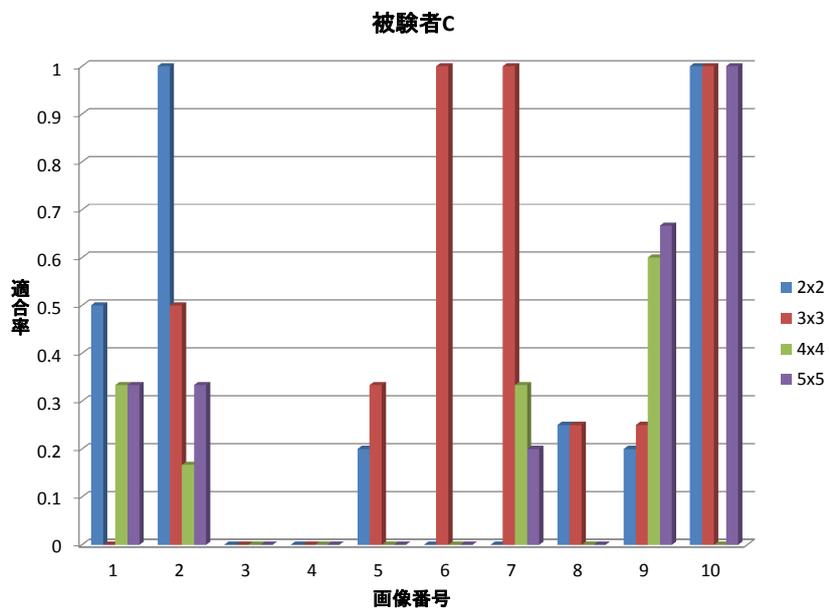


図 B.3 被験者 C

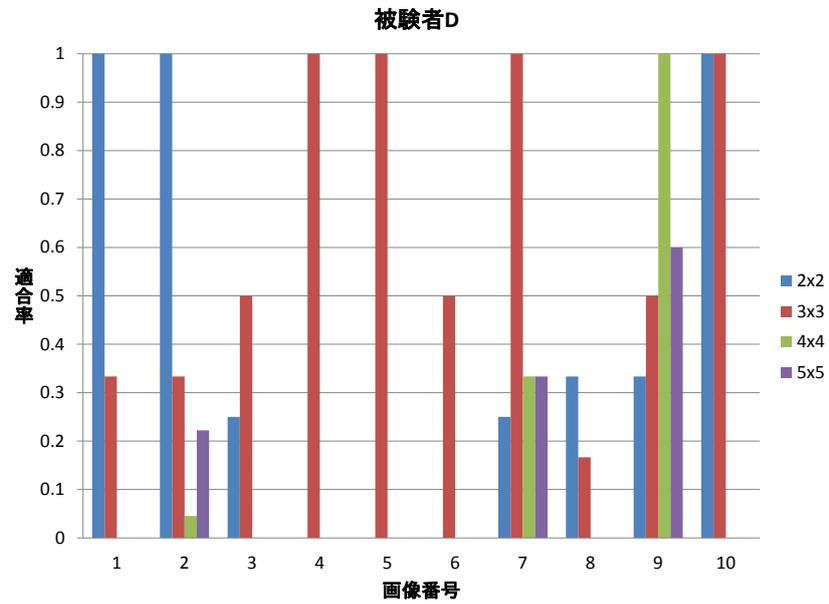


図 B.4 被験者 D

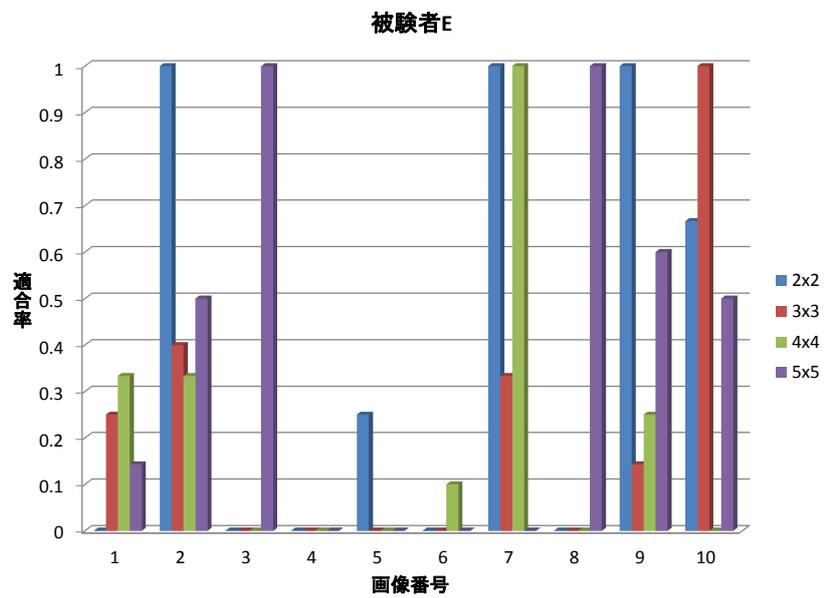


図 B.5 被験者 E

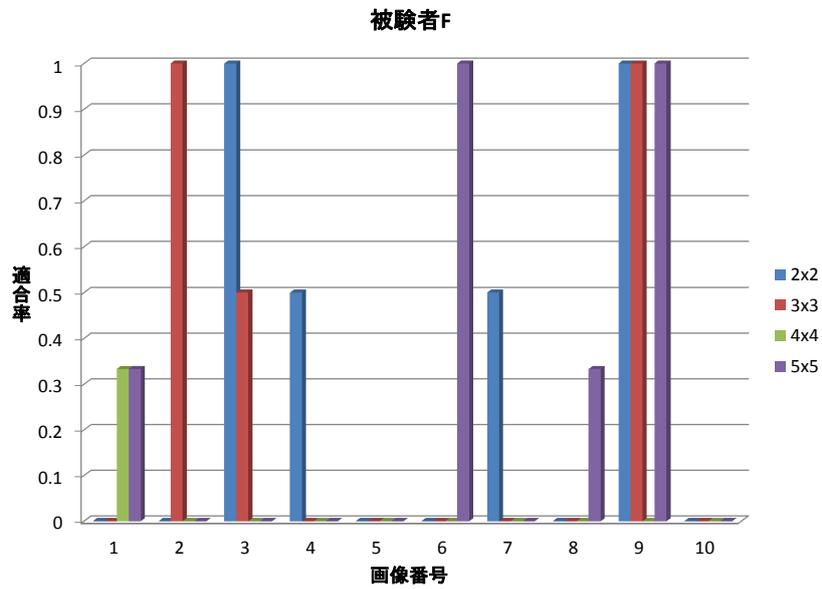


図 B.6 被験者 F

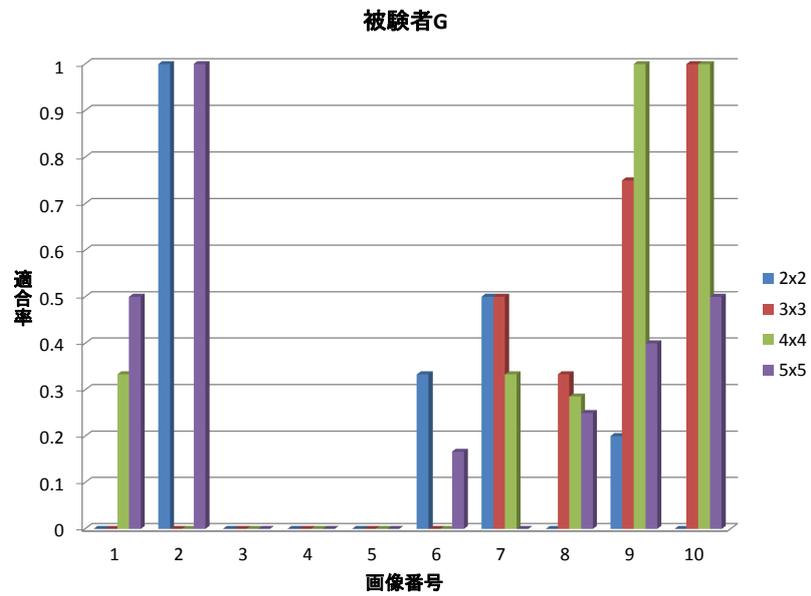


図 B.7 被験者 G

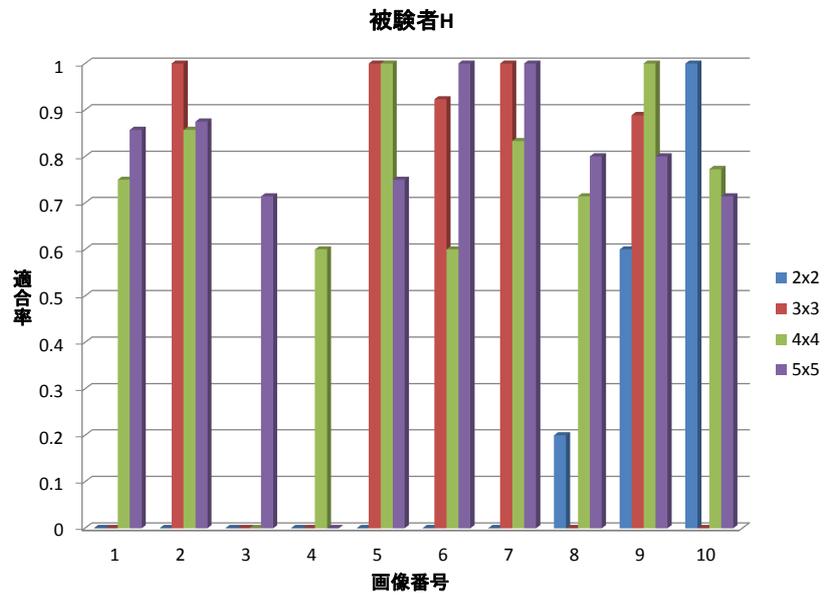


図 B.8 被験者 H

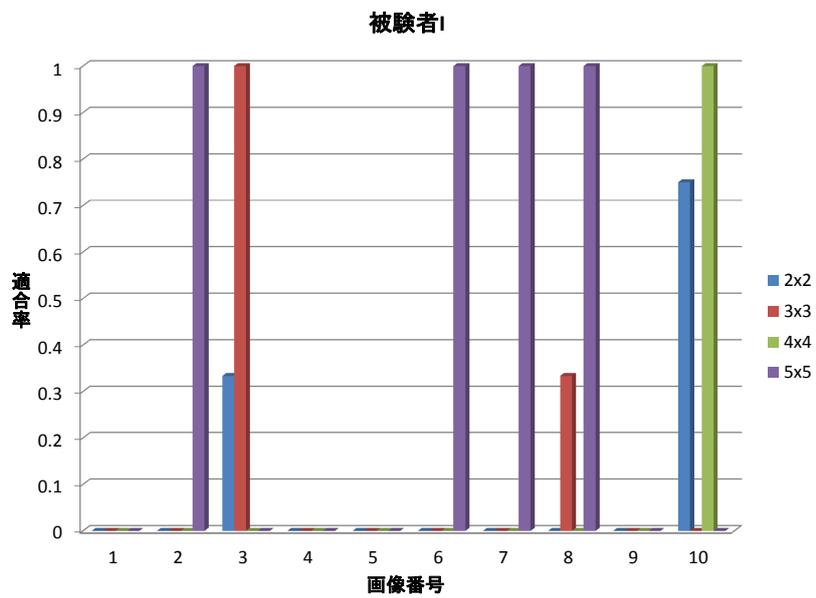


図 B.9 被験者 I

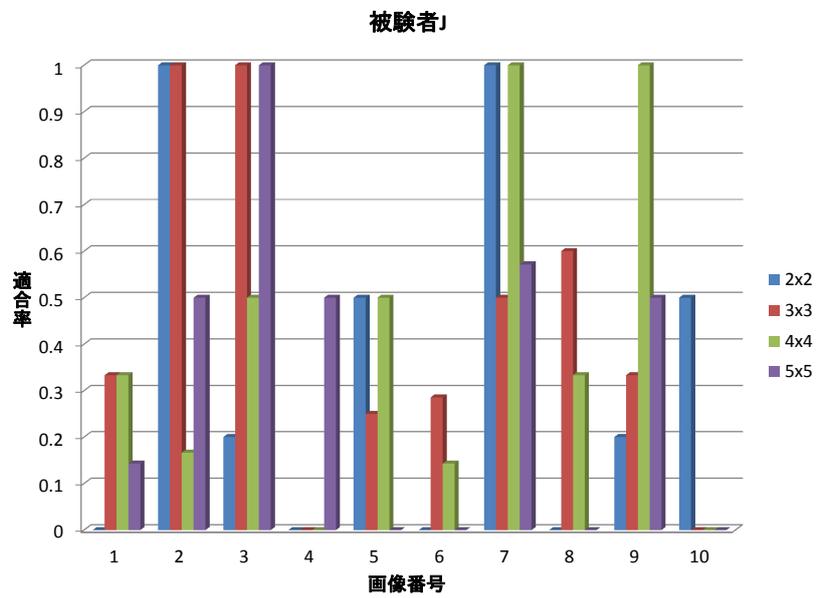


図 B.10 被験者 J