

平成23年度 修士論文

論文題目

イラストの著作権保護のためのHOG 特徴量を用いた複製検出

神奈川大学 大学院 工学研究科 電気電子情報工学専攻

学籍番号 201070080

工藤 護

指導担当者 木下宏揚 教授

目次

第1章	序論	5
第2章	基礎知識	7
2.1	電子透かし	7
2.2	代表的画像検索技術	8
2.2.1	TBIR	8
2.2.2	CBIR	10
2.3	画像処理技術	12
2.3.1	2値化	12
2.3.2	グレースケール	13
2.3.3	ガンマ補正	16
2.4	画像検索の流れ	17
2.5	エッジ検出	18
2.5.1	グラディエント(1次微分)	18
2.5.2	ラプラシアン(2次微分)	18
2.6	対象とする画像の定義	20
2.7	特徴量	22
2.8	局所特徴量	24
2.9	使用する局所特徴量	25
2.9.1	SURF	25
2.9.2	HOG (Histograms of Oriented Gradients)	28
第3章	提案	30
3.1	提案手法	30

3.1.1	処理の流れ	30
3.1.2	線画の抽出	32
3.1.3	実験の条件	32
3.2	実験結果	32
3.3	考察	38
第4章	結論	40
	謝辞	41
	参考文献	42
	発表文献	45

目 次

2.1	画像検索エンジン GazoPa	11
2.2	画像の2値化	12
2.3	中間値法スペクトル	13
2.4	単純平均法スペクトル	14
2.5	加重平均法スペクトル	15
2.6	画像検索の流れ	17
2.7	グラディエントに用いるオペレータ	19
2.8	ラプラシアンに用いるオペレータ	19
2.9	検出対象の画像	20
2.10	輪郭線のトレース画像	21
2.11	Integral imageによる積分	26
2.12	勾配方向の算出	26
2.13	ガウス関数による重み付け	27
2.14	輝度勾配と強度の算出	29
3.1	複製検出の流れ	31
3.2	オリジナルの直接利用の検出	35
3.3	サイズを変えた複製の検出	35
3.4	回転を加えた複製の検出	36
3.5	一部抜き出した複製の検出	36
3.6	トレース画像の検出	37
3.7	各特徴量の複製への適性	38

第1章

序論

近年,インターネット上には音声や画像,動画などのデジタルデータが大量に流通しており,個人レベルでもそれらのデジタルコンテンツを自由に扱えるようになってきている.

しかし,それらの発達に伴い不正なコピーによる著作権の侵害などの問題が発生している. 著作権を保護すべきコンテンツとしては様々なものが考えられるが,その中でも画像コンテンツ,特にイラストコンテンツが,漫画などの電子書籍化, pixiv 等の大型のイラスト投稿サイトの影響で急速に増加している.

イラストの不正利用者はオリジナルの画像をそのまま使うだけでなく,様々な処理によって改変を加えたり,線画部分をトレースして使用したり,線画を元に全く別の着色をすることで自分の作品として用いることが多い.

このような問題に対処するためにコンテンツの著作権の保護の手法として,電子透かしや画像検索に基づく複製検出などの技術が提案されている.

電子透かしは,対象のコンテンツに著作権情報を埋め込む技術で,カラー画像のような冗長性の高いコンテンツについては,人間に知覚できないような形で著作権情報を埋め込むことが可能である. しかし,線画のような冗長性の少ないコンテンツの場合は,人間に知覚できない形で埋め込むことが困難となっている.

一方,画像検索に基づく複製検出では,著作権保護の対象となるコンテンツをデータベースに保持し,複製の疑いがある画像を検索質問として検索をし,

複製の検出を行う。このとき、データベースに保持する情報として、各種変換や部分的複製に対して安定である局所特徴量と呼ばれるものがよく用いられる。

局所特徴量を用いたデータベースで複製を検出する研究として、孫氏らによって提案された線画の著作権保護のための部分的複製検出法がある[1]。この方法では、局所特徴領域の検出器としてMSER (Maximally Stable Extremal Regions) [2]を、局所特徴領域の記述子としてHOG (Histograms of Oriented Gradients) を用いるもので、更に、データベースの容量を減少するために、PCA (Principal Components Analysis) を用いて特徴ベクトルの次元数を削減すると共に、近似最近傍探索により検出時間を短縮したもので、手書きによる線画の複製に対する有効性を示したものである。しかし、対象とする画像が線画像から線画像を複製したものであるため、トレース画像のようにカラーイラストからその輪郭線部分を複製したものや構図を複製したものについてはその有効性が検証されていない。

そこで本論文では、オリジナルのカラーイラストから線画部分を抜き出し、そこから複製されたトレース画像に対する手法を提案する。

提案手法では、計算時間の早いSURFおよび、線画の検出に有効とされるHOG特徴量を用いてデータベースを作成する。

HOG特徴量は、画像の局所領域から輝度勾配・輝度強度を取り出す特徴量で、物体のおおまかな形状を表現することが可能となる。そのため、歩行者や人工物などの物体を検出するのによく用いられる。

また、SURF (Speeded Up Robust Features) は局所特徴量として最も一般的な、SIFT (Scale-Invariant Feature Transform) [3]に比べ、速度が早く、SIFTより異なる画像での安定性が高いとされている、

これらの特徴量で同様の処理をし、実験を行うことでそれぞれの局所特徴量のトレース画像に対する有効性の検証を行う。

第2章

基礎知識

2.1 電子透かし

電子透かしは、画像や動画、音声などのマルチメディアデータに対し、画質や音質にほとんど影響を与えずに特定の情報を埋め込む技術であり、著作権情報を埋め込むために用いられることが多い。

電子透かしには、知覚可能型と知覚困難型の二種類があり、一般的に電子透かしという用語は後者の知覚困難型電子透かしを指す。

透かし情報を埋めこまれた画像などのデータは一見すると元のデータと変わらないように見えるが、専用の電子透かし検出ソフトに読み込ませると、作者名やコピー回数などの埋めこまれた情報が表示される。これにより、不正コピーやデータの改ざんなどを防ぐことが可能となる。更に、データが改ざんされたという事実だけでなく、改ざんされた箇所も具体的に特定できるようにすることも可能である。

画像電子透かしの処理は、元画像の情報を利用して透かし情報を生成し、画像合成を行う。そのため、透かし情報は元画像によって異なり、処理後の画像から透かし情報を除去し、復元することが極めて困難となる。

2.2 代表的画像検索技術

大量の画像データに対応するため、蓄積された多数の画像データに対して、ユーザの所望するデータをより早く、効率的に取り出せる画像検索技術の実用化が必要となる。現在の検索方式には代表的なものとして、テキストをメタデータとした検索方法である TBIR、画像の内容をメタデータとした検索方法である CBIR がある。

2.2.1 TBIR

TBIR[4]はデータベース内の画像データに、画像のタイトルや製作者、内容を形容するキーワードをメタデータとして画像に付与し、それを基にしてユーザの求めるキーワードとのマッチングにより検索を行う方法である。

検索範囲がタイトルや製作者といった一様に統一された分類内容の検索に適しており、静止画像検索に限らず、同画像検索においても主流である。

例として、電子図書館における蔵書検索があげられる。蔵書検索を行うにあたって、所望する蔵書の名称や特徴を定義したものを検索情報として扱う。これらは多様な情報資源を表す共通な要素として基本15エレメント [5]から構成される。

1. タイトル：情報資源に与えられた名前
2. 作成者：情報資源の内容の作成に主たる責任を持つ実体
3. キーワード：情報資源の内容のトピック
4. 内容記述：情報資源の内容の記述
5. 公開者：情報資源を利用可能にするのに対して責任を持つ実体
6. 寄与者：情報資源の内容への寄与に対して責任を持つ実体
7. 日付：情報資源のライフサイクルにおける何らかの事象に対して関連付けられた日付

8. 資源タイプ：情報資源の内容の性質もしくはジャンル
9. 記録形式：物理的表現形式ないしデジタル形式での表現形式
10. 資源識別子：与えられた環境において一意に定まる情報資源に対する参照
11. 出所：現在の情報資源が作り出される源になった情報資源への参照
12. 言語：当該情報資源の内容の言語
13. 関係：関連情報資源への参照
14. 時空間範囲：情報資源の内容が表す範囲あるいは領域
15. 権利管理：情報資源に含まれる、ないしは関わる権利に関する情報

この基本15エレメントはデータベースの内容に準じて項目に加減が施されて、神奈川大学図書館の蔵書検索 [6] にも、これらの項目が見受けられる。

通常、Web上の画像データはHTML文書からリンクがなされており、文書には画像の内容を説明するキーワードが含まれている場合が多いため、画像データをリンクしているHTML文書を解析することでTBIRによるWeb上の画像検索を行うことが可能である。

しかし、TBIRには次のような問題点がある。

画像内容考慮 付加されるキーワードは画像の内容に依らず、HTML文書に含まれるキーワード検索しか行わないため、所望する画像データの他にそのキーワードの含まれる画像データも検索されてしまう。

手作業による索引付け 各画像データに索引を手作業で付加することになるため、膨大な画像データを取り扱う場合においてデータベース所有者への負担が増えてしまう。

追加画像の索引付加 未完成のデータベースに画像データが追加される場合、追加された画像データにおいても索引付けを行わなければならない、更なる負担が伴う。

索引の主観性 データベース所有者の主観によってキーワードの付加が行われるため、ユーザーがデータベース所有者の主観と一致しない場合、効果的な検索を行うことができない。

コスト 大規模なデータベースを管理する場合、上記負担を軽減するための管理費、人件費を考えることでコストがかさんでしまう。

既に膨大な画像データが存在し、増え続けている今日の情報化社会において、これらの問題は致命的であり、所望する画像データを効率的に入手するための画像検索技術としては満足のできるものではない。

2.2.2 CBIR

CBIR[7][8][9] は色や形、構図など、画像の特徴量を利用した検索方式である。前述した TBIR では手作業によってあらかじめ全ての画像データにキーワードを付加する必要があったが、CBIR では画像データに既に存在する画像特徴を計算機によって自動判別して画像間の類似度を判定し、ユーザーが指定した画像に類似した画像を検索することが可能である。また、画像データを視覚的に捉えることができ、よりユーザーの感覚に近い画像の検索が可能である。

用例として、日本発画像検索エンジン「GazoPa」[10]を挙げる。図2.1はGazoPaの画面となる。GazoPaの検索システムはユーザーから提供される画像のアップロードやネットサーフィン中に見かけた画像を検索に利用できる。所望する色や形に近い画像を見つけ、似たようなデザインをもっと探したい時に、その見つけた画像をそのまま検索画像として利用できるのである。更にFlashを利用したdrawerを使用することで、ユーザーが作成した画像を検索画像に転用できるため、ユーザーの想像を直に検索システムへ伝えることができる。



図 2.1 画像検索エンジン GazoPa

しかし、例にも適用されるが CBIR には次のような問題点が挙げられる。

画像間のギャップ ユーザーの所望する想像上の画像と実際の画像の特徴の間にギャップが存在し、その二つが完全に一致することは非常に困難である。

検索速度の遅延 Web を用いるシステムの場合、不必要な画像のダウンロードを行って検索速度が著しく遅くなる場合がある。

2.3 画像処理技術

2.3.1 2値化

白と黒だけによる画像の表現処理[11]である。各画素の明るさを一定の基準値により、白色と黒色の2つの値に変換する処理を行う。図2.2は2値化処理を施した例である。この一定の基準値を閾値という。閾値より小さい値を持つ画素を白、閾値より大きい値の画素を黒として表現する。通常、画像の各画素は0~255のRGB値を持っていて、RGB値の平均値が各画素における明るさとなる。2値化処理は、画素(x,y)の濃淡画像を $f(x,y)$ 、閾値を T とすると、一般的には次項に従って処理される。

$$\begin{cases} \text{白色} & f(x,y) \leq T \\ \text{黒色} & f(x,y) > T \end{cases}$$

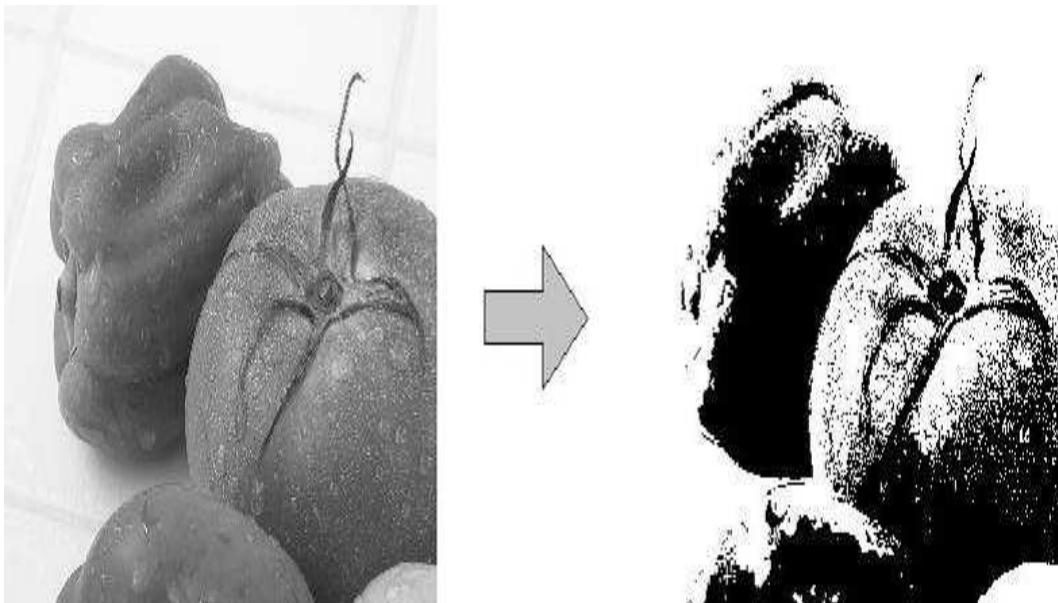


図 2.2 画像の2値化

基準となる閾値の値を変更することで2値化処理後の画像が異なり、2値化処理を行うことで画像からの検索情報の抽出が容易になり、また判定処理なども高速に実行できる。

2.3.2 グレースケール

2値化では白と黒の2色のみで表現したが，グレースケール[12]は画像を白から黒までの明暗だけで色の情報は含まずに表現する．灰色を何階調で表現するかをビット数によって表す．1ビットの場合は白と黒のみで中間色がない状態，つまり前述した2値化と同義となる．8ビットなら(白と黒を含めて)256階調，16ビットなら65536階調の灰色で表現する．

グレースケール処理はその計算方式によりいくつかの計算方法が存在する．

- 中間値法:出力画素=(最大成分+最小成分)/2

対象画素における最大最小成分の平均を画素値とする方法である．カラーパターンが灰色一色に変換され，それぞれの色の違いが全て潰れてしまっている．非常に直感的で簡単ではあるが，意図して使う以外は適切ではない．図 2.3 に中間値法のスペクトルを示す．

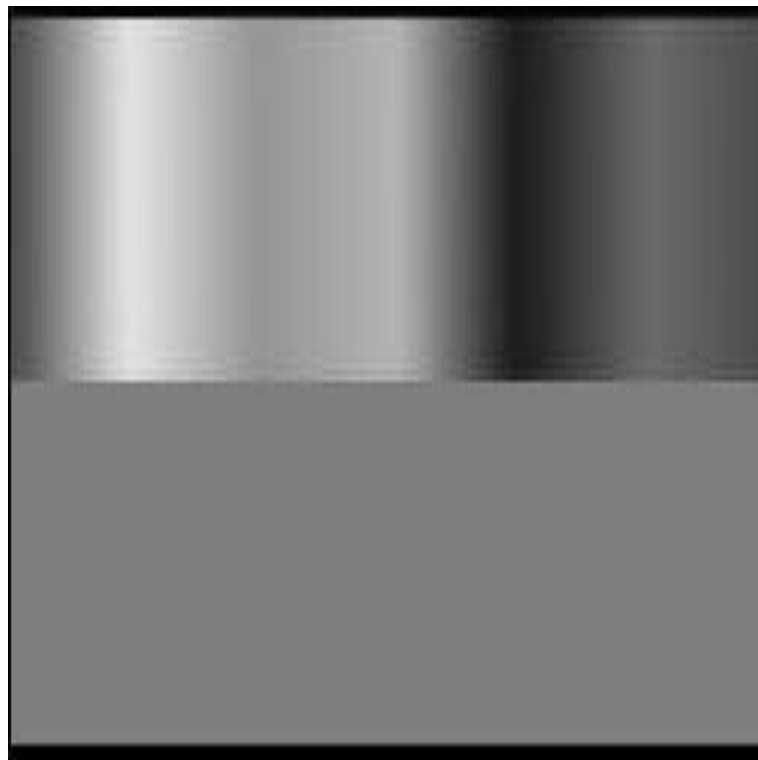


図 2.3 中間値法スペクトル

- 単純平均法:出力画素=(R成分+G成分+B成分)/3

対象画素の RGB 各成分の平均値を画素値とする方法である。RGB 成分の値によってグレーに濃淡ができていますが、異色である青と緑が処理後では同色であったりと、人間が色の違いによって感じる明るさの違いを無視してしまっている。図 2.4 に単純平均法のスペクトルを示す。

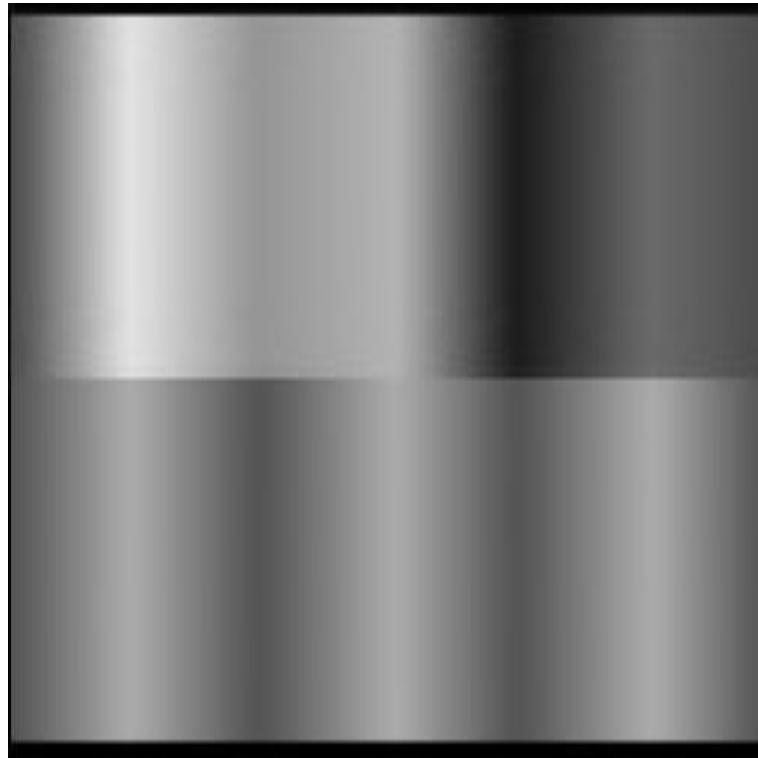


図 2.4 単純平均法スペクトル

- 加重平均法:出力画素= $(0.299 * R \text{ 成分} + 0.587 * G \text{ 成分} + 0.114 * B \text{ 成分})$

RGB 各成分に重み付けを加え, 視覚感度を考慮した計算方法. 単純平均法で問題のあった青と緑にも処理後の差異が生じ, 処理後の多色化が成される. 図 2.5 に加重平均法のスペクトルを示す.

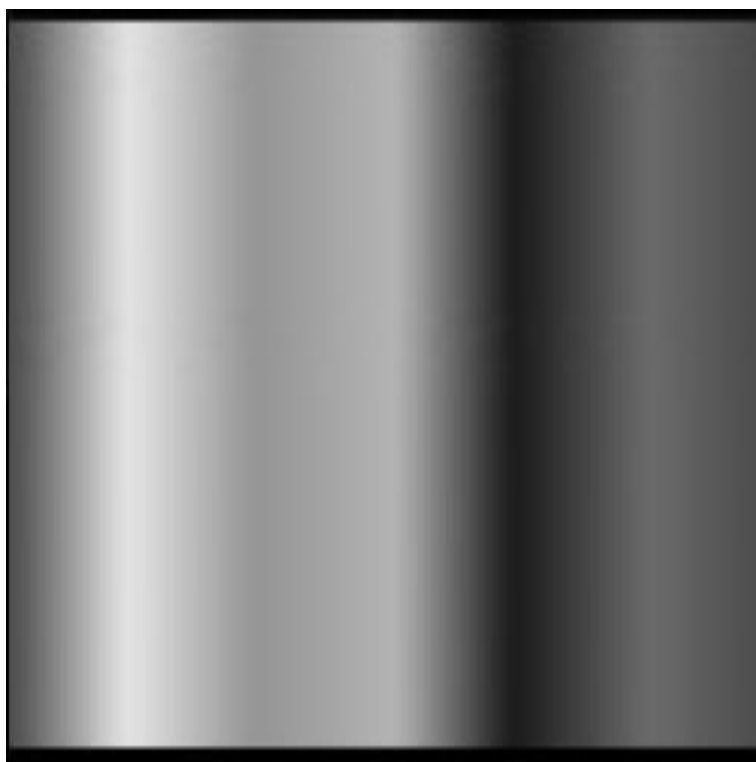


図 2.5 加重平均法スペクトル

2.3.3 ガンマ補正

γ (ガンマ)とは画像の明るさの変化に対する電圧換算値の変化比である。画像の色データと画像が実際に出力される際の信号の相対関係を調節して、より自然に近い表示を取得し、元画像データに忠実な画像の表示を再現するための補正をガンマ補正 [11] という。 γ 補正の数式を (2.1) に示す。

$$\gamma = 255 * (\gamma / 255)^{1/\gamma} \quad (2.1)$$

画像データと出力データとの値が正比例している時、 γ 値は1になるが、実際には正比例しない。スキャナなどの入力装置やプリンタなどの出力装置はそれぞれ特有の γ 値を持っているため、スキャナで入力した画像をそのまま出力してしまうと γ 値が1に近づかないため色合いが違ってしまう。広義に解釈すると、現在のほぼ全ての画像データは特定の出力環境に合わせて作られていると言ってよい。出力環境が変わることでその時出力した画像データは完全な状態で表示することができないということである。そのため、出力機器の違いによる差異を緩衝し、より正確な画像データ取得においてユーザーはガンマ補正を行う必要がある。

2.4 画像検索の流れ

画像検索の流れは、まずキーとなる画像データを用意する。キーとなる画像データとは、ユーザが所望する画像を検索するときの元となる画像データである。キーとなる画像データに画像処理を施し、検索の元となるメタデータを抽出する。また、検索される画像データベースの画像にも処理を施し、メタデータを抽出する。キーとなる画像のメタデータと、画像データベースのメタデータとの類似度を比較し、類似度の高い順に順位付けしたものを所望画像の類似画像群とする。画像検索システムによる画像の検索はここまでで完了である。最終的にはユーザの目による画像の確認によって、所望の画像を入手することができる。図2.6は画像を検索する際の流れを示したものである。

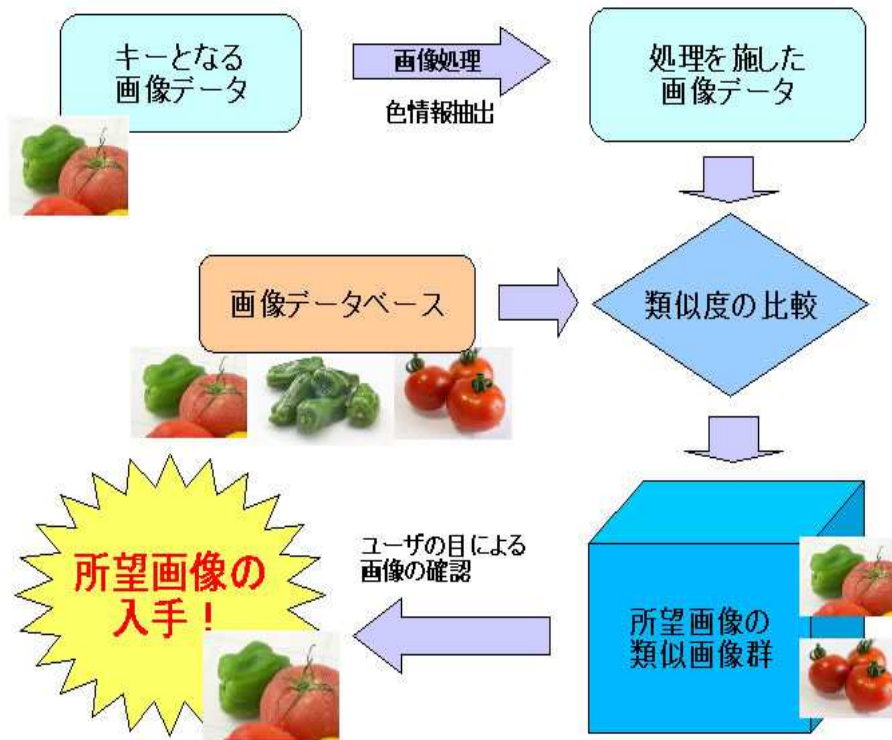


図 2.6 画像検索の流れ

2.5 エッジ検出

画像内のすべての画素は、色情報を持っている。その画像内の明るさが急激に変化する場所がオブジェクトの領域の境界、つまり輪郭となる。エッジ検出とは、画像の明るさの変化によりオブジェクトの輪郭を算出することを指す。明るさの変化値は、微分演算を利用することで算出ができる。

しかし、デジタル画像は連続ではないため、厳密には微分演算は行えない。そこで、隣接する画素の微分値の近似値を差分により算出する必要がある。隣接する画素の差分は、微分パラメータを用いて画素に重み付けすることにより算出ができる。この差分値が隣接する画素の微分値の近似値となる。

2.5.1 グラディエント (1次微分)

画面上の座標 (x,y) における明るさの勾配はグラディエントにより算出ができる。この微分値は、ベクトル $(\Delta x, \Delta y)$ として表現ができる。微分と強さを算出する式を以下に示す。

【微分】

$$\Delta x = f(x + 1, y) - f(x, y) \quad (2.2)$$

$$\Delta y = f(x, y + 1) - f(x, y) \quad (2.3)$$

【強さ】

$$\sqrt{\Delta x^2 + \Delta y^2} \quad (2.4)$$

グラディエントに用いるオペレータを図 2.7 に示す。

2.5.2 ラプラシアン (2次微分)

ラプラシアンはグラディエントをさらに微分したもので、単純に画像上の画素 (x,y) における明るさの差分 $\Delta(x, y)$ のみを表す。以下はラプラシアンの式で

ある.

$$\Delta(x, y) = 4 \cdot f(x, y) - \{f(x, y - 1) + f(x, y + 1) + f(x - 1, y) + f(x + 1, y)\} \quad (2.5)$$

ラプラシアンに用いるオペレータを図 2.8 に示す.

	Gradient	Prewittフィルタ	Sobelフィルタ
Δx算出 オペレータ	0 0 0	1 0 -1	-1 0 1
	0 1 -1	1 0 -1	-2 0 2
	0 0 0	1 0 -1	-1 0 1
Δy算出 オペレータ	0 0 0	1 1 1	-1 -2 1
	0 0 1	0 0 0	0 0 0
	0 -1 0	-1 -1 -1	1 2 1

図 2.7 グラディエントに用いるオペレータ

	ラプラシアン1	ラプラシアン2	ラプラシアン3
Δ(x,y)算出 オペレータ	0 -1 0	-1 -1 -1	1 -2 1
	-1 4 -1	-1 8 -1	-2 4 -2
	0 -1 0	-1 -1 -1	1 -2 1

図 2.8 ラプラシアンに用いるオペレータ

2.6 対象とする画像の定義

今回の実験で、検出を目的とする複製は図 2.9 のようなものとなる。WEB 上などに存在するデータそのものや印刷物などの紙媒体の画像をオリジナルとし、改変などを加えず、オリジナル全体を使用するものを直接利用、オリジナル画像に回転やサイズの変化を加える、部分的な抜き出しを行うといったものを改変を加えた複製、そして手書きによりオリジナルの線画部分（輪郭線や構図など）を複製したもの、それらにオリジナルと全く別の着色を施したもの等をトレース画像と定義する。図 2.10 は、オリジナル画像の輪郭線をトレースした場合の例である。

複製検出の難易度は順に高くなっていき、特にトレース画像はオリジナル画像と比べると、パーツの省略、手書きによる線のぶれ、サイズや角度の変化など様々な違いがあげられる。

本研究では以上の 3 つの複製画像、特にトレース画像の検出を可能とすることを目的とする。

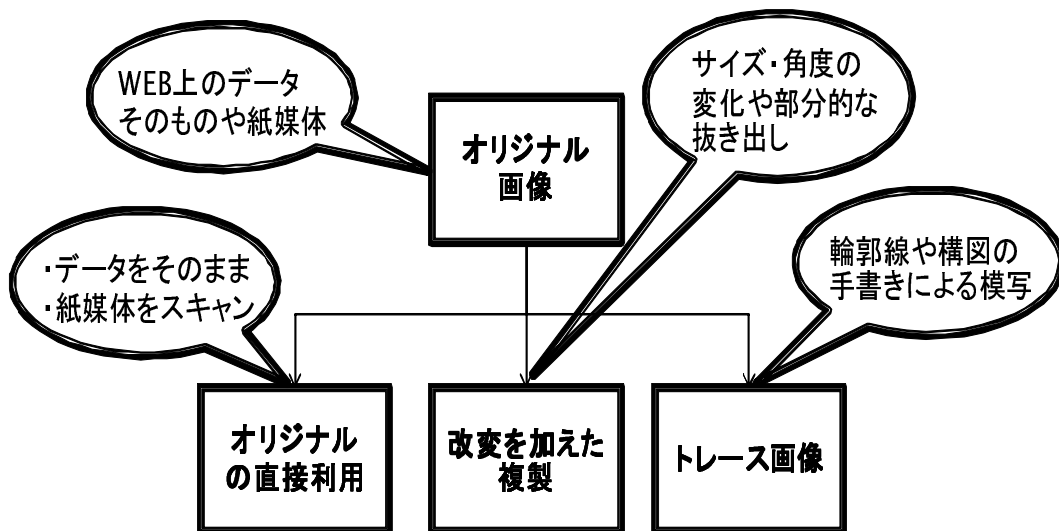


図 2.9 検出対象の画像



オリジナル画像

輪郭線のトレース画像

図 2.10 輪郭線のトレース画像

2.7 特徴量

特徴量とはその画像が持つ特徴を数式や数値によって表したものである。画像の特徴としては以下のようなものがある。

- ・ **面積** 2値画像中の図形 S の面積は S 中の画素の数として定義可能。
- ・ **周囲長** 物体の輪郭線の画素数。斜め方向はデジタル図形特有の誤差を生じるので $\sqrt{2}$ 倍して数える。
- ・ **円形度** 面積, 周囲長を基に, 形状の複雑さを測る特徴量。円形になるほど値が 1 に近づく。
- ・ **伸長度** ある領域が凸であるというのは, その領域の任意の 2 点を結ぶ線分が必ずその領域内を通る。また, 任意の領域に対し, その領域を含む最小の凸領域を凸閉包と呼ぶ。任意の領域に対し, その凸閉包から元の領域を差し引くとその領域の凹部が検出されることになる。この凹部がその領域の重要な形状測度となる。
- ・ **モーメント** 座標 (i, j) における画像 $f(i, j)$ の $(p+q)$ 次のモーメントは

$$m_{pq} = \sum_i \sum_j i^p j^q f(i, j) \quad (2.6)$$

で定義される。

2値画像の場合, 0 次のモーメント m_{00} は面積を表すことになる。また, m_{10}/m_{00} および m_{01}/m_{00} は重心座標 $G(i_G, j_G)$ を表す。重心の周りのモーメント M_{pq} は, 特に重心モーメントと呼ばれ,

$$M_{pq} = \sum_i \sum_j (i - i_G)^p (j - j_G)^q f(i, j) \quad (2.7)$$

で与えられる。

また, 2次モーメントは慣性モーメントとも呼ばれ, 画像 $f(i, j)$ の 2次重心モーメントは式 (2.7) で $p = 2, q = 0$ ならびに $p = 0, q = 2$ で与えられ,

$$M_f = \sum_i \sum_j \{(i - i_G)^2 + (j - j_G)^2\} f(i, j) = M_{02} + M_{20} \quad (2.8)$$

となる.

式(2,8)で重心を原点とした時のx軸,y軸についての2次モーメントをそれぞれ μ_{02}, μ_{20} とすると, 原点の周りの2次モーメントは

$$\mu_2 = \mu_{02} + \mu_{20} = \sum_i \sum_j j^2 f(i, j) + \sum_i \sum_j j i^2 f(i, j) \quad (2.9)$$

さらに, 原点を通る傾き θ の直線

$$j = i \tan \theta \quad (2.10)$$

の周りの2次モーメントを求め, これが最小となる角が慣性主軸 θ で, 次式で表される.

$$\theta = \frac{1}{2} \tan^{-1} \left\{ \frac{2\mu_{11}}{\mu_{20}\mu_{02}} \right\} \quad (2.11)$$

この θ は図形が伸びている方向を表す特徴量となる.

- ・ **フーリエ記述子** ある領域の境界線を追跡していくと一つの閉曲線が得られる. この閉曲線上の任意の点における接線の偏角を, ある基準点(出発点)からの距離 d の関数として $\alpha(d)$ と表すことにする. α_0 を基準点での偏角とすれば, この $\alpha(d)$ は $\alpha_0 + 2\pi$ まで変化する d の関数となる. すなわち, 閉曲線の一周の長さを D とすれば, $\alpha(D) = \alpha_0 + 2\pi$ となる. ここで, 閉曲線は周期性を持つことから

$$\alpha_N(d) = \alpha(d) - 2\pi \frac{d}{D} \quad (2.12)$$

とすると, $\alpha_N(d)$ は周期 D の周期関数となり, フーリエ級数に展開できることになる. $\alpha_N(d)$ のフーリエ級数展開を

$$\alpha_N(d) = \sum_{k=1}^{\infty} \left\{ A_k \cos\left(2\pi k \frac{d}{D}\right) + B_k \sin\left(2\pi k \frac{d}{D}\right) \right\} \quad (2.13)$$

とあらわすと, 係数 A_k, B_k は閉曲線の形状を表す特徴量となる. この係数をフーリエ記述子と呼ぶ.

しかし, 適用できるのは閉曲線だけであり, 開曲線には適応できないという欠点を持つ.

2.8 局所特徴量

画像の照合に用いられる特徴量は大きく大域特徴量(global features)と局所特徴量(local features)に分類できる。大域特徴量は画像全体から抽出される特徴量で、本研究で用いる局所特徴量は、画像の局所領域から抽出される特徴量である。最も著名な局所特徴量としてSIFTがあげられる。

局所特徴量を抽出するには、どの領域から特徴量を取り出すかを決定し、その領域から特徴量を抽出するという2段階の処理が必要となる。前者を行うプログラムをdetector、後者を行うプログラムをdescriptorと呼ぶ。

大域、局所のどちらも特徴量はベクトル(特徴ベクトル)として表現され、これを照合することで画像の検索をすることが可能となる。

局所特徴量のdetectorに対応する処理が大域特徴量の抽出では不要となるので、大域特徴量の抽出は計算的に有利であり、特徴量が画像1枚に対し1つとなるため、画像の索引付けに用いるデータ量としても大幅に少なくて済むという利点もある。

しかし、画像の一部が隠れるなどした場合、大域特徴量では同じ値を得ることが不可能になってしまう。局所特徴量を用いると、隠れたことによって得られない特徴量があっても、隠れていない部分から同じ特徴量を得ることができるという利点がある。この特性が、トレース画像のように画像の輪郭線だけを用いるなど部分的な複製を検出する際に必要であると考えられる。

2.9 使用する局所特徴量

2.9.1 SURF

SURF[13][14] は Herbert 氏らによって提案された局所特徴量で、物体認識に用いる特徴点を求める画像処理のアルゴリズムである。抽出されるデータとして、特徴点の位置、特徴点の輝度勾配の方向、スケール、64次元の特徴を表すベクトルが抽出される。その流れは、まず図 2.11 のように Integral image による積分を適用し、計算量を減らしておく。これを特徴点検出フィルタでは2次元に拡張する。次に、Hessian 応答 $\det(H)$ を近似した次式の Fast-Hessian Detector を様々なスケールでフィルタの大きさを変え適用し、

$$\det(H_{approx}) = D_{xx}D_{yy} - (0.9)^2 \quad (2.14)$$

適当な閾値で捨て、極大を求めることで特徴点を検出する。 D_{xx}, D_{yy}, D_{xy} は画像の 1px あたりの輝度値をそれぞれの方向ごとに2回偏微分をすることで求める。その後、図 2.12 のように Haar wavelet フィルタを適用し、特徴点の傾き d_x, d_y を求める。

$$d_x = d_y = 2(E - F) - C + D - A + B \quad (2.15)$$

そして、図 2.13 のように特徴点から一定の範囲にある特徴点の傾き d_x, d_y をガウス関数で重み付けすることでプロットし、着色部分から見えるベクトルを加算していき、最も長くなった時のベクトルを平均の向きとする。そして、求めた向きを基準に 4×4 のブロックに分割し、各ブロックで総和を取り $\Sigma d_x, \Sigma d_y, \Sigma |d_x|, \Sigma |d_y|$ の値を求める。このため、64次元の値で特徴量が求まる。

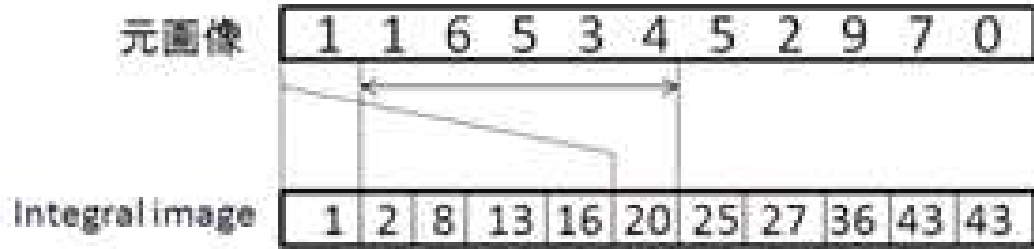


図 2.11 Integral image による積分

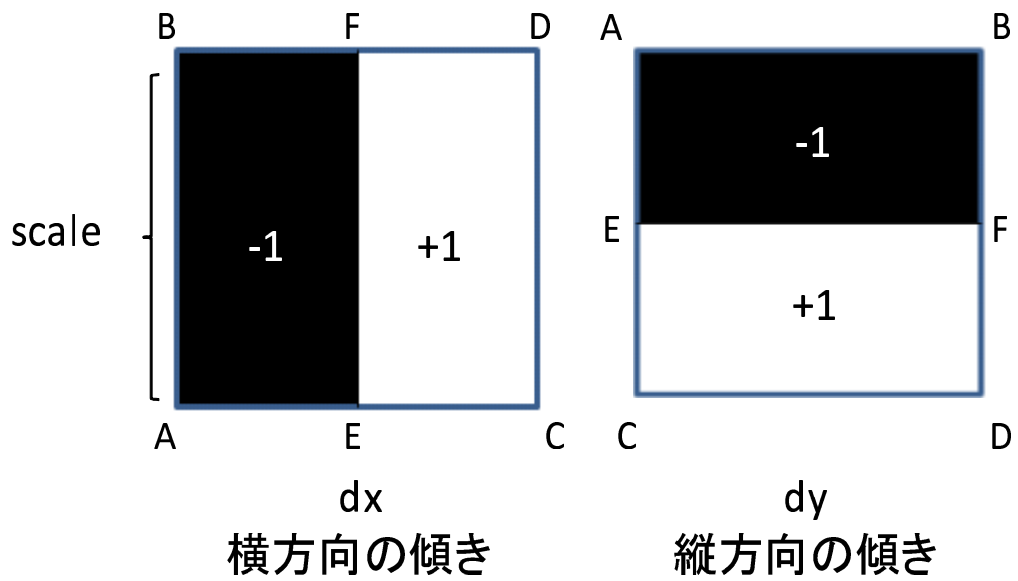


図 2.12 勾配方向の算出

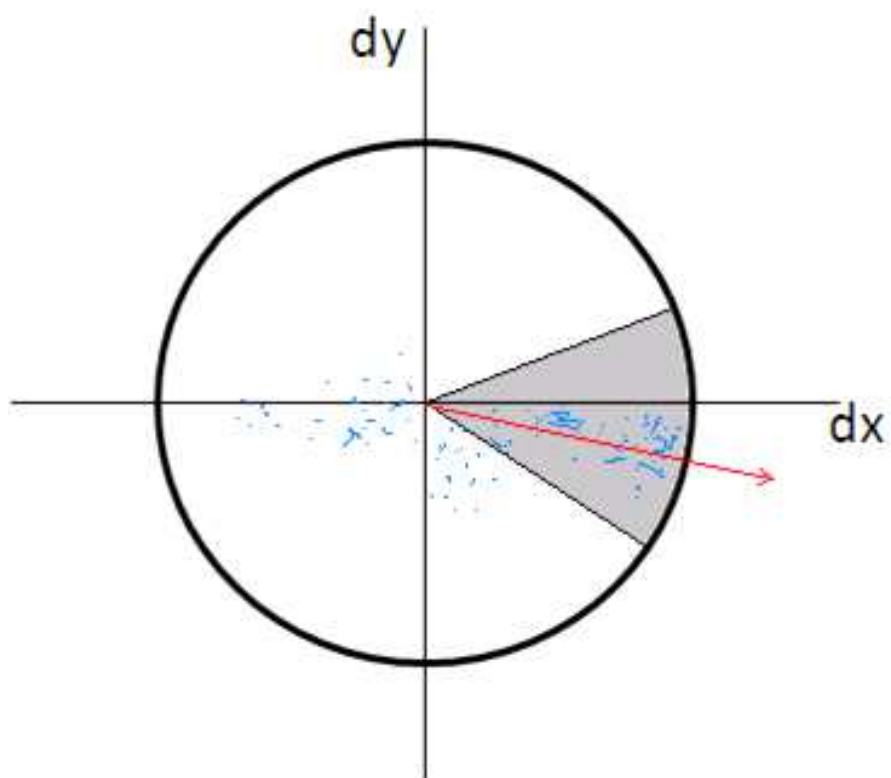


図 2.13 ガウス関数による重み付け

2.9.2 HOG (Histograms of Oriented Gradients)

HOG[15]はDalas氏らによって提案された、画像の局所領域から輝度勾配・輝度強度を取り出す特徴量で、物体のおおまかな形状を表現することが可能である。そのため、一般に歩行者の検出や人工物などの物体の検出に用いられる。

まず、局所領域を3*3のブロックに分割しそのブロックを5*5のセルに分割する。分割したセルごとに次式を用いて輝度Lから勾配強度mと勾配方向 θ を図2.14のようにセルの座標(x,y)から計算する。

$$m(x, y) = \sqrt{f_x(x, y)^2 + f_y(x, y)^2} \quad (2.16)$$

$$\theta(x, y) = \tan^{-1} \frac{f_y(x, y)}{f_x(x, y)} \quad (2.17)$$

$$\begin{cases} f_x(x, y) = L(x+1, y) - L(x-1, y) \\ f_y(x, y) = L(x, y+1) - L(x, y-1) \end{cases} \quad (2.18)$$

そして、0度から180度までを20度ずつ9方向に分割し輝度勾配のヒストグラムを作り、ブロックごとに正規化を行う。最終的にすべてのヒストグラムを統合して特徴量とする。

こうして得られた特徴量は、全ブロックのベクトルから構成される多次元ベクトルで、互いに重なりあったブロックから抽出されているので位置ずれに対する安定性が高いと考えられる。この特性は、トレース画像の手書きによる線のぶれに対応し複製の検出を行う際に重要な性質であると考えられる。また、HOG特徴量は、手書き文字認識で用いられている局所方向ヒストグラム特徴とほぼ同じものであるため、トレース画像の検出にも有効性が期待できる。

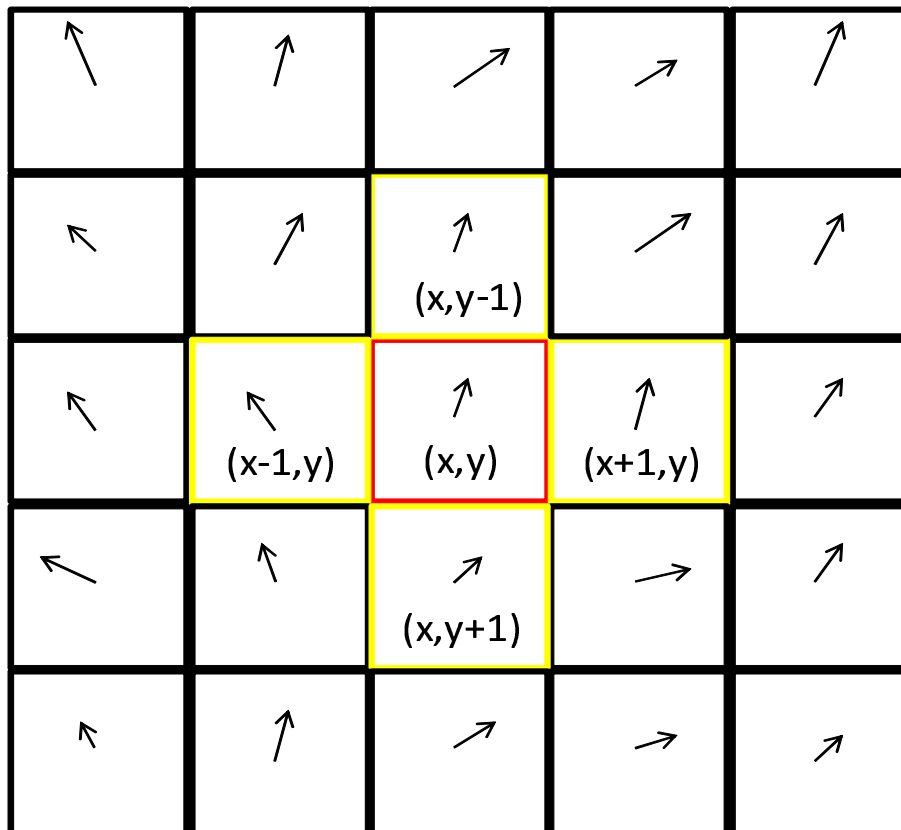


図 2.14 輝度勾配と強度の算出

第3章

提案

3.1 提案手法

3.1.1 処理の流れ

提案手法では図3.1のように、2つの処理からなる流れで複製の検出を行う。

まず、データベース側の処理では著作権保護の対象となるオリジナル画像から線画部分を抜き出し、線画像から局所特徴量を抽出し、データベースに登録する。局所特徴量の抽出アルゴリズムとして、SURFとHOGを用いる。また、クエリ側の処理では著作権を侵害する可能性がある画像をクエリとしてデータベース側の処理と同様に局所特徴量を抽出して、クエリの特徴量とデータベース側の特徴量を照合し、著作権保護対象の画像に一致する特徴量の個数分投票を行う。

クエリとデータベースに登録された画像からは、局所特徴量の種類による数百から数万の特徴量が抽出され、その数だけ投票の処理が行われる。

投票処理の終了後、得票数が多い画像から一定数までを著作権侵害の可能性がある画像として報告する。

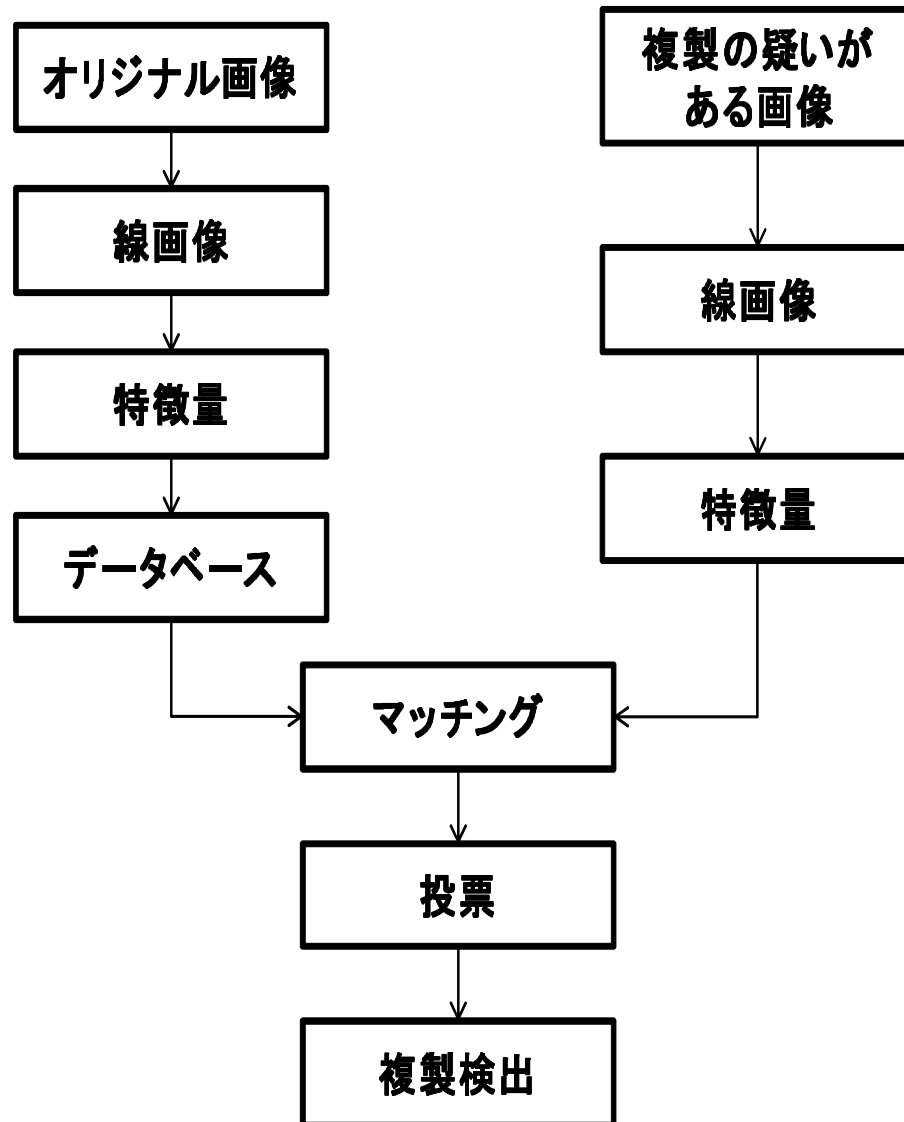


図 3.1 複製検出の流れ

3.1.2 線画の抽出

トレース画像に対応するために、オリジナル画像から線画を抽出する。

オリジナル画像に閾値を設定し、コントラストの変化がある部分を輪郭線として、線画の抽出を行う。この際、閾値を一定にするとオリジナル画像の彩色方法次第で輪郭部分がうまく抽出できない場合があるので、ユーザーが実際に閾値を適宜変化させ、最も輪郭が正確に抽出される画像を使用する。

本実験ではソーベルフィルタを用いて輪郭線抽出を行う。

3.1.3 実験の条件

本手法の有効性の検証のために、pixivに投稿されている画像から、使用の許可が降りたもの100枚をデータベースに保存する著作権保護の対象の画像とする。クエリとして用いるのは、データベースに登録されている画像をそのままスキャンやコピーをしたオリジナルの直接利用、オリジナル画像のスケールを変化させたものや回転を加えたもの、部分的に画像を切り出すなどした改変が加えられた画像、手書きによるオリジナル画像のトレース画像の3種類を用意した。

3.2 実験結果

・オリジナルの直接利用の検出

まず、改変の加えられていないオリジナルの直接利用の検出に対する有効性の検出を行う。

SURF と HOG それぞれの特徴量で実験を行った結果を図 3.2 に示す。

どちらの特徴量も共に特徴量の一致率は90%を超え、今回の実験では、得票数が1位の画像とクエリが複製の対象とした疑いがあるオリジナル画像は全て一致しており、2位以下の一致率は1%以下、10%程度とSURF, HOGのどちらの特徴量も直接利用の検出に関しては十分な有効性があることがわかった。

特徴量の一致率が若干下がっているのは,LSH (Local Sensitive Hashing) と呼ばれる最近傍探索を用いていることで, 本当の最近傍が得られていないため完全に一致しなかったためだと考えられる.

しかし, 得票数が多い順に5位まで表示するようにしているため, 少数の画像の検査ならばユーザーに負担がかからないと考えられるので, 誤対応のリスクより検出時間の高速化のメリットのほうが大きいと判断し, この手法を用いた.

- 改変を加えた複製の検出

次にオリジナル画像に改変を加えた場合の検出についての有効性の検出を行う.

実験に用いるクエリは, オリジナル画像を元にサイズを1/2倍,2倍にしたもの, 画像を30度,60度,90度の回転を加えたもの, 画像の一部を抜き出したものを用いる.

まず, オリジナル画像に対しサイズの変化を加えた場合の特徴量の一致率を図3.3に示す.SURFの場合は平均して95%程度, 対してHOGの場合では平均して90%程度の一致率となった. 2位以下の一致率についてはこちらも同様に1%以下,10%程度となった.

次に, オリジナル画像に回転を加えた場合の特徴量の一致率を図3.4に示す. こちらも安定してどちらの特徴量も90%程度の一致率が得られることがわかった. そして, 最後にオリジナル画像の一部分を抜き出した場合の特徴量の一致率を図3.5に示す. SURFの場合が平均92%,HOGの場合が平均88%, 2位以下についてはやはり1%以下,10%程度と1位の一致率は若干下がるが, 最終的にユーザーが検査を行うことを考えると十分な精度が得られていると考えられる.

以上の実験から,SURF,HOGともにサイズ変化や回転, 画像の一部分の抜き出しと言った改変にはある程度対応が可能であることがわかる.

- トレース画像の検出

最後にトレース画像に対する有効性の検証を行う.

まず,SURFについて特徴量の一致率の検証を行ったところ,最も多く一致した画像でも30%程度,平均すると20%程度となり,2位以下は1%未満であり,さらに特徴量が一致した画像も複製元のオリジナル画像とは異なっているものが多く,トレース画像の検出には全く有効ではないことがわかった.

このことからSURFはトレース画像のように複雑な変動に対しては安定性が低く,トレース画像を検出するという今回の目的には向かないと言える.

次に,HOGについて特徴量の一致率の検証を行った.こちらは,図2.10の例で示した画像では一致率が76%となり,平均して70%程度の一致率となった,2位以下の画像については構図が似ている画像の場合30%程度に増えたが,平均して10%程度であり,目的の画像が5位以内に入っていたため,トレース画像の検出に対してある程度の有用性はあると考えられる.図3.6は平均の一致率を示したものである.

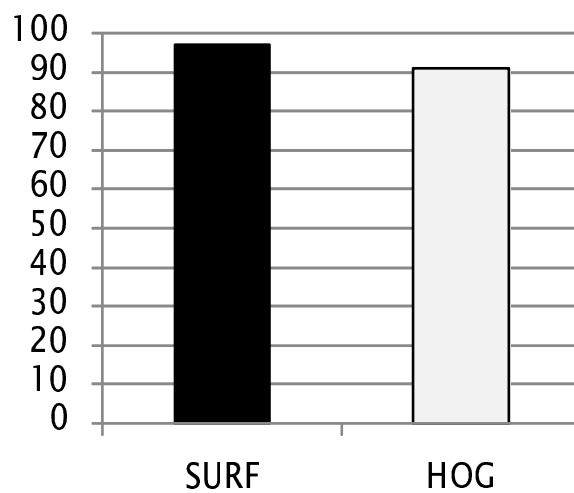


図 3.2 オリジナルの直接利用の検出

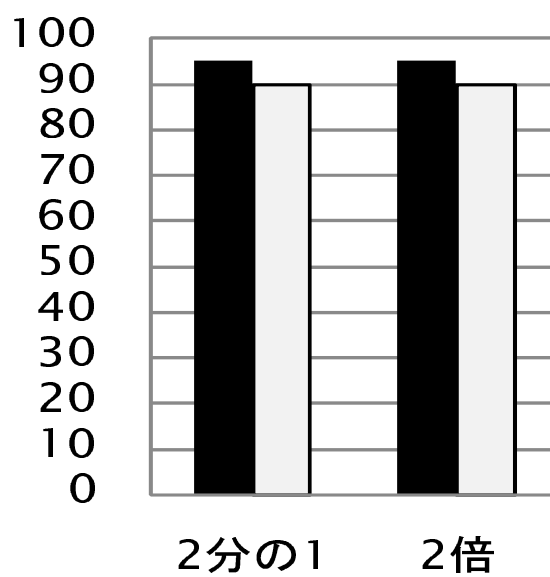


図 3.3 サイズを変えた複製の検出

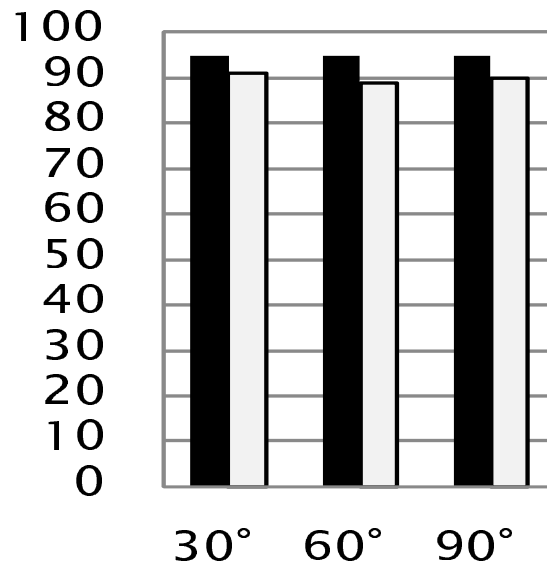


図 3.4 回転を加えた複製の検出

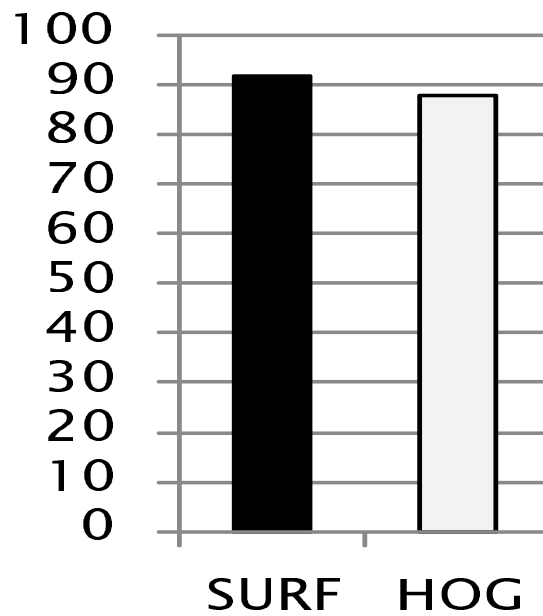


図 3.5 一部抜き出した複製の検出

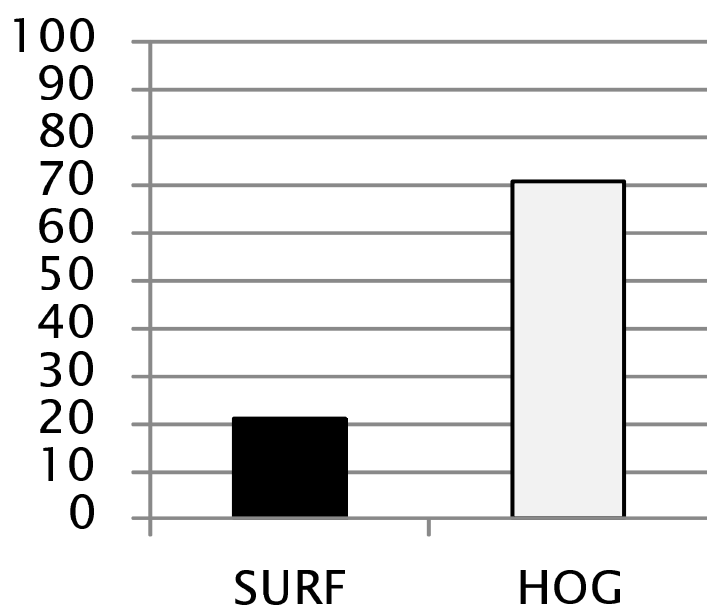


図 3.6 トレース画像の検出

3.3 考察

今回の実験から、以下のことが分かった。SURF、HOG どちらの特徴量も共に、オリジナル画像のコピーや印刷による直接利用や、サイズの変化や回転、部分的な抜き出しといった改変を加えた画像に対しては十分有効であると考えられる。しかし、トレース画像に対してはどちらも決して十分なものとは言えない。

SURF の場合は、オリジナルの直接利用や回転やサイズの変化など局所の変動が小さい画像に対しては有効であるが、トレース画像のように手書き独特の線のぶれに対する不安定さと、ベクトルの算出方法が先の位置ずれに対応できていないためと考えられる。

対してHOGを用いた場合は、それなりの精度ではあるが、複製の検出が可能である事がわかる。今回の実験で精度が下がった原因としては、1枚の画像をブロックとセルに分割した際、得られる局所領域に含まれる線分が十分でなかったためと考えられる。このため、より精度の高い検出を行うには、イラスト全体から十分に線分が含まれる局所領域を抽出し、抽出した領域に対し局所特徴量の算出を行う必要があると考えられる。図3.7は今回の実験でのそれぞれの特徴量の適性を示したものである。

	SURF	HOG
直接利用	○	○
改変を加えた複製	○	○
トレース画像	×	○

図 3.7 各特徴量の複製への適性

また、今回の実験では抽出した特徴量をそのまま使用しているため、データベースのメモリ量や計算時間が大きくなっている。このことからより多くの画像を保護対象とするには特徴量の次元数を効率よく削減し、メモリ量を減らす必要がある。こうすることで、特徴量の照合の際に比べる回数が減らせるので検索にかかる時間も減らせると考えられる。

その方法として、”大規模特定物体認識における認識率、処理時間、メモリ量のバランスに関する実験的検討”[16]や黄瀬氏らによる”局所特徴量のハッシングに基づく大規模画像検索”[17]などでも処理時間やメモリ量を削減する手法が提案されている。このような手法を取り入れることで、データベースのメモリ量や検索時間については改善できると考えられる。

第4章

結論

本論文では,局所特徴量を用いたデータベースによるイラストの複製検出の手法を提案した.今回用いた特徴量はSURFとHOGだが,データのコピーや印刷物のスキャンによる複製,回転やサイズの変化など比較的变化の少ない複製に対してはどちらも有効であるが,検出時間を考えるとSURFが適していると言える.だが,トレース画像を含めどちらにも対応するにはHOG特徴量を用いるのが有効であることが実験によりわかった.

しかし,トレース画像に対する精度は十分なものではないため,線分を十分に含む,効果的な局所領域の抽出法を検討する必要がある.

今後の課題としては,効果的な局所領域の抽出法の検討,メモリ量や検出時間を削減するための手法の検討があげられる.

謝辞

本研究を行うにあたり，終始熱心に御指導していただいた木下宏揚教授と鈴木一弘助手に心から感謝致します。また，公私にわたり良き研究生生活を送らせていただいた木下研究室の方々に感謝致します。

2012年2月

工藤 護

参考文献

- [1] Weihan Sun, and Koichi Kise, “線画の著作権保護のための部分的複製検出法,” 電子情報通信学会論文誌, D Vol. J93-D No.6 pp.909-919, 2010.
- [2] J. Matas, O. Chum, M. Urban, and T. Pajdla, “Robust wide baseline stereo from maximally stable extremal regions,” BMVC, pp.384-393, 2002.
- [3] D. G. Lowe, “Distinctive image features from scale-invariant key-points,” Int. J. Comput. Vis., vol.60, no.2, pp.91-110, 2004
- [4] 小野淳史, 天野督士, 斗谷充宏, 佐藤隆, 坂内正夫:”状態遷移モデルとシーン記述言語による自動キーワード付与機能をもつ画像データベースとその評価,” 信学論 (D-II), vol.J79-D-II, no.4, pp.476-483, 1996年.
- [5] ”Dublin Core , ” < <http://www.kanzaki.com/docs/sw/dublin-core.html> >
- [6] ”神奈川大学図書館蔵書検索,” < <http://opaclib.kanagawa-u.ac.jp/> >
- [7] 斎藤剛史, 金子豊久:”花と葉による野草の自動認識,” 電子情報通信学会論文誌, Vol.J84-D-II No.7, pp.1419-1429, 2001年7月.

- [8] 松崎公紀, 檜村雅章, 小沢慎治:” 概略図を入力とした特徴量グラフに基づく絵画画像検索,” 電子情報通信学会論文誌, Vol.J87-D-II No.2, pp.521-533, 2004年2月.
- [9] 南部藤太郎, 八村広三郎:” 類似判定基準の自動設定を行う画像検索システム,” 「人文科学とコンピュータシンポジウム」, 平成14年9月.
- [10] ” 画像検索エンジン GazoPa, ” < <http://www.gazopa.com/> >
- [11] 田村秀行:” コンピュータ画像処理,” オーム社, 2002年.
- [12] ” osakana.factory グレースケールのひみつ,”
< <http://of.jp/osakana/cgtips/grayscale.phtml> >
- [13] Herbert Bay, ” SURF : Speeded Up Robust Features,” Computer Vision and Image Understanding (CVIU), vol. 110, No.3, pp. 346-359, 2008.
- [14] SURF :” Speeded-Up Robust Features の紹介,”
<http://storage.osdev.info/pub/idmjt/diaryimage/0901/neta09013011.pdf>
- [15] Navneet Dalal and Bill Triggs, “ Histograms of Oriented Gradients for Human Detection,” Proc.of IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp.886-893, 2005
- [16] 野口和人, 黄瀬浩一, 岩村雅一, ” 大規模特定物体認識における認識率, 処理時間, メモリ量のバランスに関する実験的検討,” 電子情報通信学会論

文誌 D Vol. J92-D No.8 pp. 1135-1143, 2009

- [17] 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基づく大規模画像検索,” 日本データベース学会論文誌, vol.8, No.1, pp119-124, 2009

発表文献

1. 工藤護, 工藤敬文, 木下宏揚, 森住哲也, ”イラストの著作権保護のための局所特徴量を用いた複製検出”, SCIS2012,4F1-4,(2012)
2. MORIZUMI Tetsuya, KUDO Mamoru, SUZUKI Kazuhiro, KINOSHITA Hirotsugu : “ An electronic money system as substitute for banknotes ” , International Symposium on Applications and the Internet (IEEE/IPSJ) ,ITeS 2010,pp.316-319(2010-7)