

人間によるエッジ描写結果を用いたエッジ検出手法の評価

藤本 伸明[†], 大崎 美穂^{††},
杉山 岳弘^{††} 阿部 圭一^{††}

基本的な画像処理であるエッジ検出には、多くの手法が提案されている。しかし、これらに対する評価手法は確立しておらず、エッジ検出手法を使う際は、ユーザが、どの手法をどのようなパラメータ設定で用いるかを個人的な判断で決めるのが現状である。そこで、本研究では、従来の評価手法の問題点を踏まえ、評価の妥当性と実用性を確保できるような評価手法を提案し、有効性を検討することを目的とする。我々は評価の手続きを、(1) 複数の人間が実画像に対して正解画像を作成し、(2) 人間の視知覚を模擬した誤り評価関数を用い、エッジ検出手法の出力画像と正解画像との誤差によってエッジ検出手法のパラメータを設定し、(3) 複数の人間が心理測定法によって出力画像を比較評価するように設計した。そして、本評価手法を用い、自然物が人工物か、テクスチャがあるかないか、の4つ組合せの画像カテゴリに対し、Canny, Deriche, Iverson, Rothwell, Sugiyama-Abeの5手法を評価した。この結果から得られた知見、および従来の評価手法との比較検討より、本手法の有効性を示すとともに、どの画像にどのエッジ検出手法が適するかというエッジ検出手法の特徴を知ることができた。

Evaluation of Edge Detection Algorithms Based on Edges Drawn by Humans

NOBUAKI FUJIMOTO,[†] MIHO OHSAKI,^{††} TAKAHIRO SUGIYAMA^{††}
and KEIICHI ABE^{††}

Edge detection is one of the most fundamental image processing operations, and there have been many approaches to this operation to date. However, methodology for evaluating their performance has yet to be established. Under the current practical state of the art, edge detector users only select an edge detector and adjust parameters based on their empirical judgment. This research proposes an evaluation method based on human edge detection perception and discusses its validity through a case study in which several popular edge detectors were evaluated with our proposed methodology. We designed this framework with the best interests of propriety and practicability in mind as follows: (1) ground truth generation by humans; (2) edge detector parameter optimization by error function simulating human visual perception; and (3) edge detector comparison by humans. We then conducted an experiment to evaluate five different edge detectors — Canny's, Deriche's, Iverson's, Rothwell's, and Sugiyama-Abe's — in four image categories, namely: "natural and textured"; "natural and non-textured"; "artificial and textured"; and "artificial and non-textured". The results showed which edge detector was suitable to which image category, and we confirmed that our proposed method was valid compared with conventional methods for evaluating edge detectors.

1. はじめに

エッジ検出は画像処理に欠かせない重要な処理であり、さまざまなエッジ検出手法が提案されている。しかし、提案した手法を客観的に評価した研究は少なく、その手法が本当に有効かを判断できない、どのような応用問題にどの手法を使えばよいか分からない、といった問題がある¹⁾。そこで、この問題を解決するため、エッジ検出手法の評価に関するさまざまな研究が行われてきた^{2)~16)}。

[†] 静岡大学大学院情報学研究科

Graduate School of Information, Shizuoka University

^{††} 静岡大学情報学部

Faculty of Information, Shizuoka University

現在、アルパイン株式会社

Presently with Alpine Electronics Inc.

現在、同志社大学工学部

Presently with Faculty of Engineering, Doshisha University

しかし、これらの研究にもかかわらず、現状では、評価手法はあまり使われていない。特に、現実的な問題にエッジ検出手法を応用するさいは、画像処理の専門知識を持つユーザがエッジ検出画像を目で見て評価して検出手法を決定することが一般的である。また、どのエッジ検出手法にも調整すべきパラメータがいくつかあるが、これらの値も経験的に決定されている。

評価手法があまり使われない原因の1つとして、従来の評価手法の多くが、エッジ検出手法の用途や入力画像の種類の影響を考慮していない点があげられる。経験的には、エッジ検出手法の良し悪しはこれらに依存すると知られている¹⁷⁾。これらを実験手法に反映するには、(1) エッジ検出手法の用途、(2) エッジ検出に特化した観点での画像分類、(3) 用途や画像分類に応じた評価基準を調査し体系化する必要がある。

評価手法が使用されない最大の原因は、評価結果の妥当性と評価にかかる手間や時間とのトレードオフと考えられる。評価手法には、人間を介入させず、誤り評価関数を評価基準として自動的に評価を行う手法^{2)-10),12),15),16)}、複数の人間を介入させ、心理測定法を用いて評価を行う手法^{11),13),14)}がある。

現在、前者が主流であるが、この手法では評価コストを低減できる反面、評価基準が画像処理専門家の評価基準とはかなり異なるため、結果の妥当性に疑問が残る。一方、後者の場合、逆の問題が考えられる。このように、妥当性の確保とコストの低減にはトレードオフがあり、両者を同時に満たすことが困難なため、実用的に使われる評価手法の確立には至っていない。

そこで、本研究では妥当性と実用性をできる限り確保しうる評価手法を提案し、その有効性を検証することを目的とする。ただし、エッジ検出手法の用途や入力画像の種類に特化した評価には立ち入らず、人間の視覚特性と画像処理の観点から、エッジ検出に対して共通的に要求されるであろう最低限の基準により評価を行う。問題解決のアプローチとしては、人間の視覚による評価をベースとし、次の3点を考慮する。

1. 妥当性の確保のため、画像処理の専門知識を持たない人でも、妥当な評価ができるように教示を作成する。
2. 評価コストの低減のため、一部の作業を計算機に代行させ、評価者の作業や人数を制限する。
3. エッジ検出手法の用途や画像の種類を評価結果に反映させるように、評価手続きを設計する。

本論文の2章では関連研究を解説し、その利点と問題点を議論する。3章では、2章の議論を踏まえ、評価のフレームワークを設計し詳細な条件を決める。4章では、提案する評価手法を用いたエッジ検出手法の

評価実験を行う。最後に、5章で本論文をまとめ、今後の展望について述べる。

2. 関連研究

エッジ検出手法を評価するには、入力画像を選定し、評価基準を明確に定義して、評価手続き(エッジ検出手法のパラメータ設定、エッジ検出手法の出力画像の評価)を設計する必要がある。

過去に提案された評価手法を入力画像で分類すると、人工画像²⁾⁻⁹⁾、実画像¹⁰⁾⁻¹⁶⁾に大別できる。人工画像を使えば、計算機上で検出性能を評価しやすいが、人工画像のモデルに最もマッチするエッジ検出手法の評価が高くなるため、一般的な有効性を示せない。また、人工画像と実画像は大きく異なり、実用的な有効性も明確にならない。一方、実画像を用いる場合、選定された実画像が適切であれば、この問題を解決できると考えられる。

評価手法を評価基準で分類すると、誤り評価関数等の人工的な基準^{2)-10),12),15),16)}、人間の視覚を通じた心理的な基準^{11),13),14)}に大別できる。エッジ検出手法の応用現場では、人工的な基準を用いず、専門家が目で見て性能評価することが多い。これより、従来の人工的な基準は性能評価に十分適していないと推測される。一方、人間による評価では、評価コストは高いが、適切な人数を確保して専門家の評価を反映した教示を与えれば、妥当な結果が得られると考えられる。また、人間の視覚による評価を調べれば、その特性を反映した人工的な基準の考案にもつながる。

以上より、我々は、評価における入力画像には実画像を、評価基準には人間の視覚を用いることにした。ここで、特に関連ある研究を取り上げ、これらの利点と問題点を議論する(表1参照)。

Heathら¹¹⁾は、心理実験を通して、エッジ検出手法のパラメータ設定と性能評価を行う評価手法を提案した。これらの作業には正解画像を使わなかったため、正解画像作成の手間はかからなかった。しかし、被験者が入力画像と正解画像を見ないで、エッジ検出手法の出力エッジ画像を相対的に点数付けし、評価を行ったため、評価手続きが体系化されておらず、評価の結果と信頼性が低いと考えられる。また、画像中央の対象物が認識しやすいほど検出性能が高いと教示したため、被験者が不要なエッジまで検出する手法を高く評価する傾向がみられた。

杉山ら^{13),14)}は、Heathらと同様に人間の視覚をベースとした評価手法を提案した。Heathらとの主な違いは、正解画像を作成し、エッジ検出手法のパラ

表 1 実画像を用いたエッジ検出手法の評価の研究（発表された年代順に記載）

Table 1 Summary of some related work on method for edge detector evaluation.

文献（年代順）	入力画像	評価基準	評価フレームワーク	利点	問題点
Heath ¹¹⁾	実画像 28 枚	人間	正解画像の作成：なし パラメータの設定：手動 性能の評価：心理実験	正解画像が不要	不要なエッジが多くても評価が高い
杉山 ^{13),14)}	実画像 60 枚	人間	正解画像の作成：手動（エッジ選択） パラメータの設定：自動 性能の評価：心理実験	正解作成において位置ずれ調整が不要	エッジ検出手法ごとに正解画像が必要
Bowyer ¹⁵⁾	実画像 60 枚	ROC 曲線	正解画像の作成：手動（エッジ描画） パラメータの設定：なし 性能の評価：自動	評価コストが低い	パラメータを最適に設定していない
Shin ¹⁶⁾	実画像 110 枚	ROC 曲線	正解画像の作成：自動 パラメータの設定：自動 性能の評価：自動	評価コストが低い	パラメータを最適に設定していない

メータ設定に利用したことである。正解画像は、複数の被験者がエッジ点候補から正しいエッジを選択し、これらを平均して得られた。この方法には、選択されたエッジを平均する際、位置ずれの調整が不要という利点があった。その反面、作業が難しく、エッジ検出手法ごとに作業を必要とした。性能評価では、Heath¹¹⁾の問題を解決するため、入力画像と出力画像を見比べながら容易に作業できる、正解画像の例を見せ、誤った評価を防ぐような教示を実施する等の工夫がなされた。

Bowyer¹⁵⁾, Shin¹⁶⁾は、手動、もしくは自動で正解画像を作成し、ROC 曲線（Receiver Operating Characteristic curve）を利用した誤り評価関数でエッジ検出手法の性能を評価した。通常、ROC 曲線は正検出率と検出洩れ率から作成されるが、BowyerらはROC 曲線を誤検出率と検出洩れ率から作成した。この場合ROC 曲線は上下が反転する。評価基準である誤り評価関数には、ROC 曲線と各軸で囲まれた領域の面積、つまり、誤検出率と検出洩れ率をさまざまなパラメータ設定で加算した結果を用いた（詳細は3.3.2項で解説する）。この評価手法には自動評価が可能という利点があるが、評価基準がパラメータ設定の影響を無視しているため、公平な結果が得られるとはいえない。

3. 評価手法の提案

本章では、関連研究の利点と問題点を踏まえ、エッジ検出手法の評価手法を提案する。まず、エッジ検出手法のどのような処理や性能を評価対象とするか、どのパラメータを設定対象とするかを議論し、明確に規定する。次に、入力画像の選定、評価基準の定義を行う。最後に、評価手法のフレームワークと、フレームワーク内の各ステージを詳細に設計する。

3.1 評価のために規定すべき項目

3.1.1 評価対象と設定対象

まず、評価対象とする性能を議論する。エッジ検出手法の性能は、エッジ点の検出性能と位置決め性能に大別される。前者は、検出すべきエッジ点を検出し、検出すべきでないエッジ点を検出しない性能であり、主にエッジ点の検出処理に依存する。一方、後者は、正しい位置でエッジ点を検出する性能であり、主にスケールパラメータの値に依存する。エッジ点が正しければ、位置ずれの修正は比較的容易なため、我々はエッジ点の検出性能をエッジ検出手法の本質的な性能と見なすことにした。

次に、評価対象とする処理を議論する。エッジ検出手法には、雑音処理フィルタ等の前処理やエッジ接続等の後処理といった、他の画像処理手法に共通する処理を含むものがある。したがって、エッジ検出手法をそのまま用いると、公平に評価できない。そこで、エッジ検出手法の本質的な性能（エッジ点の検出性能）に寄与するエッジ点の検出処理のみを、各エッジ検出手法から切り出して評価することにした。

従来の研究では、エッジ検出手法のパラメータを最適に設定するという過程を含まないで評価を行っているものも多い^{15),16)}。しかし、実際にエッジ検出手法を使用するユーザにとっては、エッジ検出手法の最適なパラメータ設定、および、最適な状態でどの手法が良いかという情報が必要である。そこで、我々は、エッジ検出処理のパラメータであるエッジ特徴量のしきい値とスケールパラメータを対象とし、これらの値を最適に設定することにした。

3.1.2 入力画像

エッジ検出手法を公平に評価するには、偏りのない入力画像を複数用いる必要がある。しかし、エッジ検出手法には、特定の画像に特化して設計されたもの、

あるいは、処理の特性から、ある種の画像には適するが他の種の画像には適さないものがある。したがって、多様な画像に対する平均的な性能よりも、画像の種類ごとに性能を調べ、どのような画像にどのエッジ検出手法が適するかを知るほうが有用である。また、評価手法の簡便さの観点からも、一般性を損なわない範囲内で画像数を制限する方がよい。

ただし、現時点では、エッジ検出手法の評価用入力画像という観点で、実画像の種類を調べ分類した研究はみられない。将来的には、このような研究を行う必要があると考えているが、今回は、文献 11) で用いられた認識対象に基づく分類と、各分類に属する実画像を用いた。分類は以下の 4 つ、各分類の画像は 5 枚、全分類で計 20 枚である。

分類 1: 人工物 & テクスチャなし

分類 2: 人工物 & テクスチャあり

分類 3: 自然物 & テクスチャなし

分類 4: 自然物 & テクスチャあり

3.1.3 評価基準

評価手法に関する従来の研究の多くは、誤り評価関数を用いて、それを基準として評価を行っており、結果が画像処理の専門家の評価と同じとは限らない問題を残しつつも、評価コストの低減を優先していた。本研究は逆の立場をとり、評価コストは高くなるが基本的に人間が作業を行い、評価結果が専門家の評価と同様になることを試みる。

エッジ検出手法の良し悪しはその用途に依存するため、今回は画像認識に適用する場合を想定する。また、実用性の観点から、画像処理の専門知識を持たない人であっても、専門家に近い評価ができるように評価基準を定義する。我々は、画像処理研究者である本論文第 3 著者の意見と Heath らの研究¹¹⁾を参照し、また、予備実験を行うことで¹⁸⁾、専門知識がない人でも理解でき、画像認識に適するような評価基準を次のように定義した。

検出すべきエッジ: 対象物の境界線, 対象物を構成する部品の境界線, 面と面の境界, 影を示す線。テクスチャは含まない。

良いエッジ画像: 検出すべきエッジを検出し, 検出すべきでないエッジを検出していない画像。画像中のすべての物体を正しく認識するのに必要な線分で構成されており, 対象物の質感はできる限り含まない。

3.2 評価フレームワークの設計

我々は、図 1 に示す評価のフレームワークを提案する。これは、(1) 正解画像の作成、(2) エッジ検出手法のパラメータ設定、(3) 出力画像の比較評価の 3

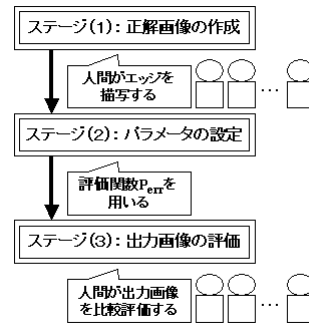


図 1 提案する評価フレームワーク
Fig. 1 Outline of our proposed method.

ステージからなる。ただし、(1)、(3) は人間が行い、(2) は 3.3.2 項で述べる誤り評価関数と最適化アルゴリズム^{13),14)}を用いる。これは、一部の作業の自動化で人間の負担を減らし、評価コストの低減を図るためである。ただし、誤り評価関数中の重みは、事前に人間を使った実験によって決定する。

Heath ら¹¹⁾は、正解画像を用いず、(2) のパラメータ設定を心理実験を通して行ったため、被験者の負担が大きかった。そこで、本評価手法では (1) で正解画像の作成を行い、作成した正解画像を (2) のパラメータ設定に用いた。

3.3 各ステージの設計

3.3.1 ステージ (1): 正解画像の作成

ステージ (1) では、複数の人間が入力画像を手動でなぞり、エッジ画像を描写する。そして、各点におけるエッジを描写した人数を濃度と見なすことで、1 枚の正解画像を得る。杉山ら^{13),14)}は、各エッジ検出手法により得られたエッジ点候補画像からエッジ点を選択することにより正解画像を作成したため、エッジ検出手法ごとに正解画像の作成作業が必要であった。また、Bowyer ら¹⁵⁾においては、人間が正解画像を作成したが、被験者が 1 名のみで客観性に疑問が残った。そこで、我々は、複数の人間による描写を採用し、1 画像あたりの被験者を 5 名とした。

今回は、ペンを使って、紙に印刷した入力画像上にエッジを描写してもらい、描写部分をスキャナで取り込みデジタル化した。なお、この作業はペンタブレットを使い計算機上で行っても問題ない。描写する画像の順序はランダムとし、描写の慣れの影響の偏りを防いだ。描写線には 8 画素程度の太さを持たせ、被験者間のエッジ位置のずれを描写線の太さで吸収し、位置合せを省略可能にした。

3.3.2 ステージ (2) : パラメータの設定

ステージ (2) では、杉山ら^{13),14)} が提案したパラメータ最適化手法を採用した。この手法では、探索領域を分割し、徐々に狭めるパラメータ探索を行い、誤り評価関数を最小化する点を見つける。領域分割回数はパラメータ空間の広さに依存するが、エッジ検出手法のパラメータ空間では 2 回の分割で探索が可能である、と文献 13), 14) では確認されている。また、我々は今回用いるエッジ検出手法のパラメータ空間が単峰性であり、この探索が局所解に陥らないことを確認した。

パラメータ最適化の誤り評価関数として、我々は式 (1) に示す P_{err} を提案する。

$$P_{err} = \alpha P_{FN} + (1 - \alpha) P_{FP} \quad (1)$$

この式は、あるパラメータ設定における検出洩れ率 P_{FN} と誤検出率 P_{FP} の重み付け加算を意味する。今回は、正解画像について、過半数の人が描写したエッジを検出すべきエッジ、一人も描写しなかったエッジを検出すべきでないエッジ、それ以外は検出してもしなくてもよいエッジとして、 P_{FN} と P_{FP} を計算する。

α の値は、検出洩れと誤検出の重みを表しており、この値は求める出力結果、つまり評価基準となるエッジの定義によって異なると考えられる。そこで、今回提案した評価基準 (3.1.3 項参照) に対して α の値を心理実験により決定する。この手続きより、人間にとっての検出洩れと誤検出のバランスを式 (1) に反映し、おおまかではあるが人間の評価の模擬を試みる。

ここで、我々が提案する誤り評価関数 P_{err} と、文献 15) で用いられた ROC 曲線^{15),16)} による評価との関係を述べておく (図 2 参照)。ROC 曲線は、横軸に検出洩れ率 P_{FN} 、縦軸に誤検出率 P_{FP} をとり、エッジ検出手法が持つ各検出パラメータを調整し、そのパラメータにおける P_{FN} と P_{FP} を計算することによって作成することができる。本評価手法のパラメータ設定は、次の 2 段階の手続きを踏んでいる。まず、人間が最適と感じる P_{FN} と P_{FP} のバランス α を求める (α の値により式 (1) の傾きが変化する。図 2 の左側参照)。次に、 P_{err} を最小とするパラメータの最適な設定、つまり、式 (1) と ROC 曲線の接点を求める (図 2 の右側参照)。文献 15), 16) では、ROC 曲線を作成し、その曲線下の面積で各エッジ検出手法を評価している。しかし、我々は係数 α の値を心理実験により求め、それを用いることで各手法ごとに最適なエッジ検出結果を決定する。そして、最適なエッジ検出結果を用いて、各エッジ検出手法の比較を行う。

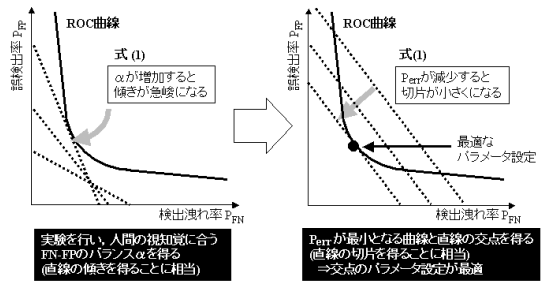


図 2 P_{err} と ROC 曲線の関係

Fig. 2 Relation between P_{err} and ROC curves.

3.3.3 ステージ (3) : 出力画像の評価

ステージ (3) では、2 つのエッジ検出手法の出力画像を対にし、入力画像とともに人間に提示する。ただし、対はエッジ検出手法の全組合せで作成する。そして、出力画像対から良いほうを選択してもらう (同じ場合は無理に選択しない) ようにした^{13),14)}。Heath ら¹¹⁾ は出力画像のみを被験者に提示し、7 段階評価を行った。しかし、比較対象および入力画像なしにエッジ画像を評価し、かつ、7 段階という細かい点数を付けることは、かなり難しいタスクと考えられる。そこで、我々は入力画像、比較対象ありの選択作業という容易なタスクを採用した。また、入力画像に対する理想のエッジ画像を頭に思い浮かべやすくするために、被験者に事前に正解画像の例を見せる等の工夫を施した。

出力画像対を各被験者に同じ順序で提示すると、評価結果に順序効果が現れてしまう。これを防ぐため提示はランダムに行う。また、出力画像対の評価プロセスを、エッジ検出手法ごと、もしくは画像ごとに分けると、対の数が少なすぎるため、評価を公平に行えなくなる恐れがある。そこで、エッジ検出手法と入力画像の全組合せ条件をまとめて、1 つの評価プロセスとした。被験者は 10 名、1 対あたりの評価回数は 20 回 (1 被験者が 2 回ずつ評価) である。

評価作業後、全被験者の選択回数を合計する。そして、このデータに符号検定を施し、出力画像間の選択回数、つまり、エッジ検出手法間の性能に有意差があるかを調べる。なお、符号検定とは、対応のある 2 標本群を用い、これらの母集団平均値に大小の差があるかを調べる検定手法である¹⁹⁾。これより、対にしたエッジ検出手法のうち、どちらの性能が高いかが分かるため、画像分類ごとに全対の検定結果をまとめれば、各画像分類におけるエッジ検出手法の性能順位が得られる。

ただし、この順位付けは、心理測定法の観点から見ても厳密には正しくない¹⁹⁾。なぜなら、2 つの評価対象

表 2 提案する評価手法の概要

Table 2 Outline of our proposed method.

入力画像	評価基準	評価フレームワーク
人工物/自然物	人間にとって	(1) 人間による
テクスチャなし/あり	良いエッジ	正解画像の描画作成
の組み合わせ 4 分類	画像	(2) P_{err} による 最適パラメータ設定
1 分類あたり 画像数 5 枚		(3) 人間による 出力画像の比較評価

を比べた相対的な評価が、全評価対象を順位付けして並べた絶対的な評価と一致するとは限らないからである。しかし、エッジ検出手法間の性能差が十分大きければ、評価手続きの厳密さによって性能順位が大きく異なるとは考えにくい。また、我々は、厳密さを若干犠牲にしても評価結果の分かりやすさを優先すべきと考え、順位付けを試みた。

性能順位は、3.1.2 項で定義した画像の 4 つの分類ごとに得られる。さらに、これらに対して実験データを統合すれば、人工物、自然物、テクスチャなし、テクスチャあり、という新しい 4 つの分類についても性能順位を求められる。たとえば、分類 1: 人工物 & テクスチャなし、分類 2: 人工物 & テクスチャあり、の実験データを統合して、人工物に対する性能順位が得られる。

我々は、評価結果が理解しやすいように、これら合計 8 つの分類ごとのエッジ検出手法の性能順位を可視化することにした。具体的には、8 つの分類を表現する対極軸を使い、各軸上に性能順位をプロットして、レーダチャート形式の図を作成する（実際の図は、4 章の実験結果を参照）。

最後に、本節のまとめとして、3.3.1~3.3.3 項で説明した本評価手法の概要を表 2 に示す。

4. エッジ検出手法の評価実験

我々は、提案する評価手法の 3 つのステージ (1), (2), (3) に対応して、実験 I: 正解画像の作成、実験 II: パラメータの設定、実験 III: 出力画像の評価、を行った。本節では、全実験に共通の条件、実験 I, II, III の条件と結果、考察について述べる。

4.1 共通の実験条件

入力画像は、3.1.2 項で述べた 4 つの画像分類、1 分類あたり 5 枚の計 20 枚とし、この条件に合致する Heath ら¹¹⁾ の実験画像を用いた。図 3 にその一部を示す。評価するエッジ検出手法には、比較的よく用いられている、Canny²⁰⁾、Deriche²¹⁾、Iverson²²⁾、Rothwell²³⁾、Sugiyama-Abe²⁴⁾ の 5 手法を選んだ。各手法のパラメータを表 3 に示す。 T 、 β 、 h_{min} は

表 3 各手法の検出パラメータ

Table 3 Parameters for the detectors analyzed.

手法	パラメータ
Canny	T_1, T_2, w
Deriche	T_1, T_2, w
Iverson	T
Rothwell	T, β, σ
Sugiyama-Abe	$T_1, T_2, h_{min}, \sigma$

設定対象である検出しきい値、 w 、 σ はスケールパラメータである。被験者には、画像処理の専門知識を特に持たず、絵の描写技術も一般的な人を選んだ。

4.1.1 実験 I: 正解画像の作成

実験 I では、3.3.1 項で述べたステージ (1) の手続きに沿って、エッジ画像を描写し正解画像を作成した。ただし、実験時間を短縮するため、本実験では被験者を 5 名ではなく 10 名とし、1 画像を描写する人が 5 名になるように画像をランダムに割り振った。なお、1 被験者あたりの描写枚数は 10 枚、描写時間は計約 40 分であった。図 3 に入力画像と正解画像の一例を示す。正解画像は、描写した人数に比例した濃度で示してある。

4.1.2 実験 II: パラメータの設定

実験 II では、3.3.2 項で述べたステージ (2) の手続きに沿って、各エッジ検出手法の最適な検出結果を作成した。最初に適切な α の値を求める予備実験を行った。その手続きは次のとおりである。 α の値を 0.1, 0.3, 0.5, 0.7, 0.9 の 5 種類に設定し、各 α の値について、誤り評価関数 P_{err} を用いて最適なエッジ検出結果を得る。このようにして、1 つのエッジ検出手法について 5 種類の最適な検出結果が得られる。そして、各エッジ検出手法について、異なる α の値による出力画像を人間に対提示し、良い方を強制選択してもらう。ただし、 α の対数は 10 対、1 つの α の値に対して入力画像は 5 枚、被験者は 12 名、1 画像あたり 24 回の評価（1 被験者が 2 回評価）とした。そして選択結果より、最も高い選択回数を得た α の値を選択する¹⁸⁾。実験の結果、今回我々が定義した評価基準では、 α の値は 0.3, 0.5, 0.7 の 3 種類が良かった。1 被験者あたりの選択時間は約 40 分であった。

そして、これら 3 種類の α の値による出力画像から最も良いものを被験者が選択した。なお、スケールパラメータを持つエッジ検出手法については、3 種類の各 α の値について、4 種類のスケールパラメータによる出力画像を作成し、計 12 枚の画像の中から選択した。それにより、適切な α とスケールパラメータの値を求めた。1 被験者あたりの選択回数は 100 回、



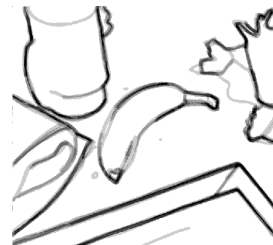
入力画像 a(512×512)



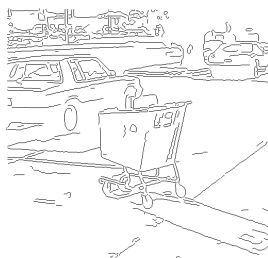
正解画像



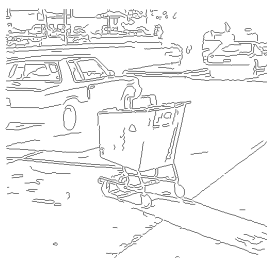
入力画像 b(512×468)



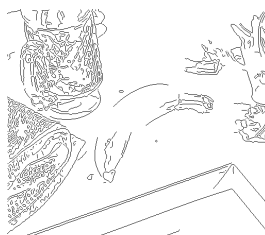
正解画像



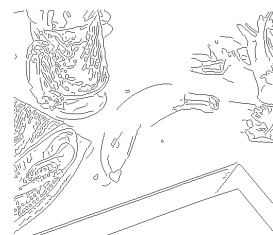
Canny の手法の出力



Deriche の手法の出力



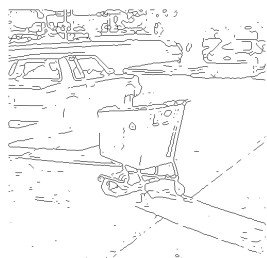
Canny の手法の出力



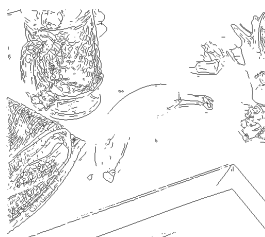
Deriche の手法の出力



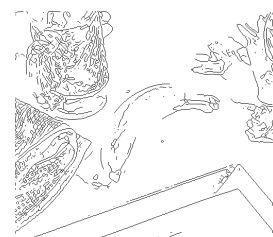
Iverson の手法の出力



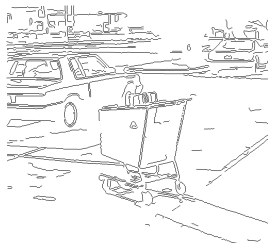
Rothwell の手法の出力



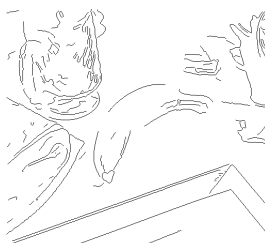
Iverson の手法の出力



Rothwell の手法の出力



Sugiyama-Abe の手法の出力



Sugiyama-Abe の手法の出力

図 3 エッジ検出結果の一例

Fig. 3 The example of input images, ground truths, output images from optimized edge detectors.

選択時間は約 20 分であった。

次に、これらの値と実験 I で作成した正解画像を用い、全エッジ検出手法のパラメータを最適に設定し出力画像を得た。なお、出力画像を求める各エッジ検出手法の処理時間は、Linux の OS、2GHz の CPU の計算機上で 1~5 秒程度であり、大きな差はなかった。図 3 に各エッジ検出手法の出力画像の一例を示す。

4.1.3 実験 III：出力画像の評価

実験 III では、3.3.3 項で述べたステージ (3) の手続きに沿って、出力画像の比較評価を行い、エッジ検出手法の性能順位を得た。なお、1 被験者あたりの選択回数は 400 回、選択時間は約 40 分であった。図 4 に、得られたエッジ検出手法の性能順位を、画像分類ごとにプロットしたレーダチャートを示す。

図 4 より、Canny の手法は画像分類に関係なく性

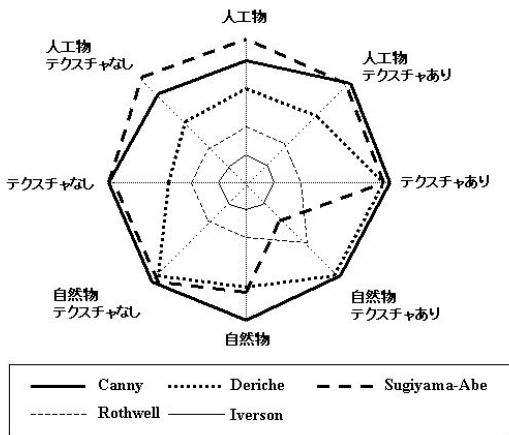


図4 画像分類ごとのエッジ検出手法の性能順位
Fig.4 The effectiveness rank for each category.

能が高かった。Dericheの手法は人工物よりも自然物で性能が高かった。Iversonの手法とRothwellの手法は、どの分類でも性能が低かった。Sugiyama-Abeの手法は自然物 & テクスチャありで性能が低いが、他の分類では高かった。性能の全体傾向は、Canny, Sugiyama-Abe, Deriche, Rothwell, Iversonの順であった。

4.2 評価手法の有効性に対する考察

4.2.1 正解画像の妥当性

正解画像を作成したほとんどの被験者が、実験後の内観報告において、入力画像をなぞるエッジの描写タスクは容易であったと述べた。よって、エッジ描写に特別な技術は不要であるといえる。

著者3名が目視で確認したところ、被験者の描写画像間にはかなりのばらつきがあった。しかし、ばらつきがみられたのは、服のしわや薄い影の輪郭等であり、認識において重要な物体の輪郭や明確な濃淡を持つ箇所では、ばらつきが小さかった。よって、画像処理の専門知識を持たない人でも良いエッジ画像の定義を理解し、それを反映した描写が可能と考えられる。

我々は、描写線の太さを約8画素にすることで、各被験者が描写したエッジ画像のずれを吸収し、位置合せの省略を試みた。実際に、位置合せなく平均して得られた正解画像は、図3に示すように、重要なエッジを取りこぼしなく含み、太さも描写時に近かった。これより、この太さで描写する方法の利便性が示された。Roldanら²⁵⁾は、画素間の距離がどの程度小さいと、人間はそれらを弁別できなくなるか(同一のエッジとして見なすか)を調べ、その値が1画素であることを示した。これは太さ1画素のエッジに対する結果であり、我々の太さを持った正解画像では、2, 3画素の

ずれがあっても同一の連続したエッジと見なすことができると考えられる。よって、今回の描写の太さは、人間の弁別能の観点でも妥当である。しかし、この点については今後検討していく必要がある。

以上より、被験者数が十分でない問題はあるが、本評価手法の正解画像はおおむね妥当であるといえる。

4.2.2 評価基準の妥当性

本評価手法では主に人間が評価を行い、パラメータ設定のみに誤り評価関数 P_{err} を用いた。まず、人間の評価基準の妥当性を検討する。評価実験では、被験者ごとに教示した評価基準の理解度を質問し、評価基準の理解は困難でないことを確認した。実際、実験結果において評価に三すくみ ($a < b, b < c, c < a$ のような状態)が生じることはなく、被験者間でエッジ検出手法の性能順位に大きな差もなかった。さらに、画像処理の専門家である第3著者が同様の簡易な実験を行い、本実験と同様の結果を示した。よって、良いエッジ画像の定義は専門知識がない人でも理解可能であり、これによって専門家に近い安定した評価を行えるといえる。

次に、誤り評価関数 P_{err} の妥当性を検討する。我々は、心理実験の結果から、 P_{err} における検出洩れと誤検出のバランス α を決定した。よって、少なくともこのバランスを考慮しないROC曲線^{15),16)}より、 P_{err} の方が人間の視覚に近いといえる。また、ROC曲線ではパラメータ設定を考慮しないため(2章参照)、パラメータを最適化し、最良の状態のエッジ検出手法を評価できる P_{err} の方が公平な基準といえる。ただし、人間の評価要因は検出洩れと誤検出のみとは限らないため、 P_{err} の式は再検討すべきであろう。

以上より、 P_{err} は再考の余地があるが、ROC曲線よりも公平で人間の視覚を反映する点、被験者が良いエッジ画像の定義を理解しこれに従える点で、本評価手法の評価基準は妥当といえる。

なお、本評価実験の結果は、同じエッジ検出手法を評価したHeathら¹¹⁾の結果と大きく異なる。これは、我々の評価基準(3.1.3項参照)と、Heathらの評価基準(2章参照)の違いによる。Heathらの評価基準は、アート製作でエッジ検出手法を特殊効果に用いる用途には適するかもしれない¹⁷⁾。一方、我々の評価基準は画像認識の用途に適すると考えられる。

4.2.3 評価手法の実用性

評価手法の実用性は、評価にかかる人数と時間で決まる。今回は、5つのエッジ検出手法の評価に、10名、1名あたり2時間が必要であった。Heathら¹¹⁾の評価手法と比較すると、被験者数はほぼ等しく、実験時

間は約 1/2 であった。また、Heath らの手続き（原画像との比較なし、7 段階相対評価）に比べ、我々の手続き（原画像との比較あり、良い方の選択）は容易であり、信頼性は高いであろう。よって、少なくとも Heath らの評価手法より、本評価手法のほうが実用面で望ましい。

しかし、他の評価手法より簡便とはいえ、実用化するには評価者の人数と時間がかかり過ぎる。これを解決するには、一連の評価作業を支援するシステムを開発する、より人間の視知覚に近い誤り評価関数を考案し作業の代行を増やす等が考えられる。また、さまざまな入力画像やエッジ検出手法に対して、本評価手法を適用するケーススタディを積み重ね、その結果をデータベース化して再利用可能にする方法もあげられる。

4.2.1~4.2.3 項の議論より、正解画像と人間の評価基準が妥当であることが示され、本評価手法の有効性が確認された。一方、誤り評価関数 P_{err} については改善の余地がある、本評価手法の実用化には多くの課題が残されている、ということが分かった。

5. ま と め

本研究では、人間の視知覚に基づくエッジ検出手法の評価手法を提案し、その有効性を検証する実験を行った。評価の手続きは、(1) 人間による正解画像の作成、(2) 人間の視知覚を模擬した誤り評価関数によるパラメータの設定、(3) 人間による出力画像の評価、と設計した。実験の結果、画像分類ごとにエッジ検出手法の性能順位が得られ、本評価手法が従来の評価手法よりも妥当であることが示された。

今後は、実用性の向上のため、評価作業支援システムの開発、誤り評価関数の改善、さまざまな入力画像やエッジ検出手法に対するケーススタディと結果のデータベース化、を行う予定である。現時点では多くの問題が残されているが、最終的には、いくつかの画像に対して、検出してほしいエッジを描写するだけで、最適なエッジ検出手法を選定しパラメータ設定をする自動システムを目指したい。

謝辞 懇切なご助言をいただいた査読者の方々に感謝する。本研究の一部は、科学研究費補助金（若手研究（B）、No.14780273）の補助によって行われた。

参 考 文 献

- 1) 阿部：画像処理アルゴリズムの性能評価に関する一考察，信学技報，PRU94-15，pp.71-77 (1994)。
- 2) Deutsch, E.S., et al.: A quantitative study of the orientation bias of some edge detector

- schemes, *IEEE Trans. Comput.*, Vol.27, No.3, pp.205-213 (1978).
- 3) Abdou, I.E., et al.: Quantitative design and evaluation of enhancement/thresholding edge detectors, *Proc. IEEE*, Vol.67, No.5, pp.753-763 (1979).
- 4) Eichel, P.W., et al.: Quantitative analysis of a moment-based edge operator, *IEEE Trans. Systems Man Cybernet.*, Vol.20, No.1, pp.59-66 (1990).
- 5) Venkatesh, S., et al.: Edge evaluation using necessary components, *CVGIP: Graphic. Models Image Process.*, Vol.54, No.1, pp.23-30 (1992).
- 6) Strickland, R.N., et al.: Adaptable edge quality metric, *Opt. Eng.*, Vol.32, No.5, pp.944-951 (1993).
- 7) Kanungo, T., et al.: A methodology for quantitative performance evaluation of detection algorithms, *IEEE Trans. Image Process*, Vol.4, pp.1667-1674 (1995).
- 8) Palmer, P.L., et al.: A performance measure for boundary detection algorithms, *Comput. Vision and Image Understand.*, Vol.63, No.3, pp.476-494 (1996).
- 9) Zhu, Q.: Efficient evaluations of edge connectivity and width uniformity, *Image Vision Comput.*, Vol.14, pp.21-34 (1996).
- 10) Cho, K., et al.: Performance assessment through bootstrap, *IEEE Trans. Pat. Anal. Machine Intell.*, Vol.19, No.11, pp.1185-1198 (1997).
- 11) Heath, M.D., et al.: A robust visual method for assessing the relative performance of edge-detection algorithms, *IEEE Trans. Pat. Anal. Machine Intell.*, Vol.19, No.12, pp.1338-1359 (1997).
- 12) Shin, M., et al.: An objective comparison methodology of edge detection algorithms using a structure from motion task, *Comput. Vision and Pat. Recognit.*, pp.190-195 (1998).
- 13) 杉山, 大崎, 大野, 阿部: 人間による領域境界の検出結果を用いたエッジ検出手法の比較評価, 信学技報, PRMU 2000-39, pp.33-40 (2000).
- 14) Ohsaki, M., et al.: Evaluation of edge detection methods through psychological tests — Is the detected edge really desirable for humans?, *IEEE Proc. Int. Conf. on System, Man, Cybernet.*, pp.671-677 (2000).
- 15) Bowyer, K., et al.: Edge detector evaluation using empirical ROC curves, *Comput. Vision and Image Understand.*, Vol.84, pp.77-103 (2001).
- 16) Shin, M., et al.: Comparison and edge detector

performance through use in an object recognition task, *Comput. Vision and Image Understanding*, Vol.84, pp.160–178 (2001).

- 17) Ohsaki, M., et al.: A research on edge detector applications and definition of edge quality, *Int'l Symposium on Measurement, Analysis and Modeling of Human Functions (ISHF'01)*, pp.322–327 (2001).
- 18) 藤本, 大崎, 杉山, 阿部: 人間による境界線の検出結果を用いたエッジ検出手法の比較評価, 画像の認識・理解シンポジウム (MIRU) 論文集, Vol.II, pp.453–460 (2002).
- 19) 北野 (編): 遺伝的アルゴリズム 4, 産業図書出版, 第 13 章 (2000).
- 20) Canny, J.: A computational approach to edge detection, *IEEE Trans. Pat. Anal. Machine Intell.*, PAMI-8, No.6, pp.679–698 (1986).
- 21) Deriche, R.: Using Canny's criteria to derive an optimal edge detector recursively implemented, *International J. of Computer Vision*, Vol.2, pp.167–187 (1987).
- 22) Iverson, L.A. and Zucker, S.W.: Logical/linear operators for image curves, *IEEE Trans. Pat. Anal. and Machine Intel.*, Vol.17, No.10, pp.682–696 (1993).
- 23) Rothwell, C.A., et al.: Driving vision by topology, *Int'l Symp. Computer Vision*, pp.395–400 (1995).
- 24) 杉山, 阿部: エッジベクトル抽出器と複数エッジの解析, 電子情報通信学会論文誌 D-II, Vol.J80-D-II, No.6, pp.1379–1389 (1997).
- 25) Roldan, R.R., et al.: A measure of quality for evaluating methods of segmentation and edge detection, *Pat. Recognit.*, Vol.34, pp.969–980 (2001).

(平成 15 年 8 月 6 日受付)

(平成 16 年 6 月 8 日採録)



藤本 伸明 (正会員)

平成 14 年静岡大学情報学部情報科学科卒業, 平成 16 年同大学院情報学研究科修士課程修了。同年アルパイン (株) 入社。在学中は画像処理の研究に従事。



大崎 美穂 (正会員)

平成 6 年九州芸術工科大学 (現九州大学) 芸術工学部音響設計学科卒業。平成 11 年同大学院芸術工学研究科博士後期課程修了。平成 11 年静岡大学情報学部助手。平成 16 年から現在, 同志社大学工学部専任講師。博士 (工学)。現在, 医療データマイニング, 聴覚障害補償, 画像処理手法評価の研究に従事。IEEE Computer Society, 人工知能学会, 知能情報ファジィ学会, 音響学会会員。平成 10 年 International Conference on Soft Computing and Information/Intelligent Systems (IIZUKA'98) Best Paper Award 受賞。平成 15 年情報処理学会山下記念研究賞受賞。



杉山 岳弘 (正会員)

平成 4 年静岡大学工学部情報知識工学科卒業。平成 9 年同大学院電子科学研究科博士後期課程修了。現在, 静岡大学情報学部助教授。博士 (工学)。画像処理手法の比較評価, 映像メディア処理, 知識コンテンツに関する研究に従事。電子情報通信学会会員。



阿部 圭一 (正会員)

昭和 38 年名古屋大学工学部電気学科卒業。昭和 43 年名古屋大学大学院工学研究科博士課程単位取得退学。昭和 46 年工学博士。昭和 43 年名古屋大学工学部助手。昭和 46 年静岡大学工学部電気工学科講師。助教授, 教授を経て, 現在静岡大学情報学部情報科学科教授。平成 7 年 10 月から平成 11 年 9 月まで情報学部長。構造的パターン認識, 画像処理アルゴリズム等の研究に従事。情報教育にも深い関心を持つ。著書: 『ソフトウェア入門 第 2 版』(共立出版) ほか。電子情報通信学会, 人工知能学会, 社会情報学会, IEEE 等の会員。IAPR, 電子情報通信学会フェロー。