

服飾画像マイニングのための衣類領域からの色情報抽出

相田 優[†] 柳井 啓司[†] 柴原 一友^{††} 藤本 浩司^{††}

[†] 電気通信大学大学院 情報理工学研究科 総合情報学専攻

〒 182-8585 東京都調布市調布ヶ丘 1-5-1

^{††} テンソル・コンサルティング株式会社

〒 102-0083 東京都千代田区麹町 2-10-1 クレアーレ麹町 2F

E-mail: [†]aita-m@mm.inf.uec.ac.jp, ^{††}yanai@cs.uec.ac.jp, ^{†††}{kazutomo.shibahara,koji.fujimoto}@tensor.co.jp

あらまし 本論文では、大量の画像データから衣類の視覚情報を抽出するために、画像中の衣類領域からの色情報抽出手法を提案する。提案手法では衣類領域の指定を行う必要がなく、顔検出と領域分割によって衣類領域の推定をし、色分類が行われる。その後、割り当てられた色名が同じとなる領域を統合し、最終的な色名の判定は面積ベースによる決定法で行われる。実験では、118人、277つの衣類に対する色分類の精度を求め、評価を行った。提案手法は、衣類領域の指定が行われた場合と比較しても精度の低下が少ないことが確認され、衣類を含む画像からの色情報抽出に関して有効であることが示された。

キーワード 色分類 服飾画像 衣類画像 画像マイニング

Color Information Extraction for Clothing Image Mining

Masaru AITA[†], Keiji YANAI[†], Kazutomo SHIBAHARA^{††}, and Koji FUJIMOTO^{††}

[†] The University of Electro-Communications

1-5-1 Chofugaoka, Chofu, Tokyo 182-8585 Japan

^{††} Tensor Consulting Co.Ltd.

2-10-1 Kouji, Chiyoda, Tokyo 102-0083 Japan

E-mail: [†]aita-m@mm.inf.uec.ac.jp, ^{††}yanai@cs.uec.ac.jp, ^{†††}{kazutomo.shibahara,koji.fujimoto}@tensor.co.jp

Key words Color classification, Clothing image, Image mining

1. はじめに

1.1 背景

近年、Flickr^(注1)、Panoramio^(注2)といった写真投稿サイトの登場によって Web 上の写真はますます増大しており、画像データを大量に入手することも容易となっている。それらの画像の中には人物が写っているものも多く、衣類の情報を含むデータを大量に集める事が可能である。また、大量のデータを解析することで、データの中に隠れている知識やパターンを取り出すといった、データマイニング手法がマーケティングの分野で重要視されている。画像の視覚情報を用いたデータマイニングに関する研究も盛んに行われており、そうした技術の有用性は高まっている。衣類の視覚情報を抽出することが可能となれば、

データマイニング手法により有用な情報が得られる可能性が非常に高い。仮に、画像データから衣類の色や模様といった視覚情報が分かれば、ファッションの流行や地域性等に関する分析を行うことが可能である。そのため、画像中に含まれる衣類の視覚情報の抽出は重要な事項となっており、大量のデータを処理するためにも、画像中の衣類領域からの視覚情報抽出手法の確立が期待されている。

1.2 目的

本研究の目的は、画像中に含まれる衣類領域から、既存のデータマイニング手法で有効となるような視覚情報を抽出することである。視覚情報の中でも特に色に着目し、画像中の衣類の色を 12 色に分類する手法について研究を行った。提案手法では、入力として画像を与えると、その画像に含まれる衣類の色名を出力する。通常、色の分類を行う際はマスクの形で対象領域を指定するのが一般的だが、本研究では大量データを処理するために、それらの処理を自動化し、色名の出力を行う。画

(注1): <http://www.flickr.com/>

(注2): <http://www.panoramio.com/>

像中の衣類の領域を推定し、色分類を行うことで衣類の色名がラベルとして出力されるが、最終的なラベルの決定は衣類領域の面積ベースによる決定法となる。背景と人物領域の分離、人物領域の細分化による細かな衣類領域への分割、の2種類の領域分割を行うことで、衣類領域に対する色分類を実現した。

2. 関連研究

画像中に含まれる衣類の情報を扱った既存の研究として、個人認識・同一人物の特定に関する研究が挙げられる。Gallagherらの研究では、人物の写真から、衣類の情報を利用して個人の認識を行う[1]。Gallagherらは、2つの写真の人物が同じ服を着ている場合、同一人物である可能性が高い、という仮定のもと、個人の認識を行う。複数枚の画像データを利用し、衣類の領域分割を行ったあと、個人認識のための特徴抽出を行っている。Farenzenaらの研究では、衣類の単位で画像を分解し、同一人物の特定を行う[2]。Farenzenaらは人物の画像を、頭・胴・足の3つのパーツに分解し、特徴抽出を行い、同一人物の特定を行う。パーツの分解では、シャツ・ズボン・顔では画素値が大きく異なる、という仮定を立て画像の分解を行っており、各衣類領域からの特徴抽出を行うことで同一人物の特定の精度を向上させている。これらの研究の目的は、多くの画像から特定の人物を検索する、という点にある。しかし、既存の研究では検索する際に衣類の視覚情報を用いるものの、衣類自体が何色であるか、という事に関しては研究の対象外となっている。こうした個人認識の研究とは異なり、本研究においては、衣類の色を分類する、ということが主な仕事となる。

動画中の人物が着ているシャツの色分類を行ったものとしてWuiらの研究やVaqueroらの研究が挙げられる。Wuiらは動画中の人物が着ているシャツの色分類を行った[3]。色空間上での11色の分布を学習し、ピクセルレベル、フレームレベル、シーケンスレベルでの色分類を行っている。物体追跡手法により人物の位置を特定し、顔の下の領域を矩形領域として割り出すことによって、シャツの領域の推定を行っている。Vaqueroらは属性ベースでの人物検索を行う際に、衣類の色分類を行った[4]。Vaqueroらの研究においても、色分類の対象は動画中のシャツとなる。また、シャツの色の情報は、属性ベースでの人物検索に利用されることを目的としているため、精度を計る際に使用されたデータセットの色分布に大きな偏りが生じている。これらの研究の目的は、防犯・監視システムでの応用や、個人検索を行うという点にあり、そのためにシャツの色分類を行っている。本研究で用いるデータは動画ではなく静止画となる。また、シャツだけでなく、ズボン、ブーツ、バッグといった、人物が着用している物体全てが色分類の対象となる。

色分類を行った研究としてWeijerらの研究が挙げられる[5]。カラーチップに対する命名と実世界での色の命名には違いがあり、代表的な色名であっても、その色名の色空間上での範囲は一意には定まらない。Weijerらは、代表的な色名11個について、画素値の分布から色名を学習させることで色分類を実現した。eBayデータセットを使用し、PLSAやSVMといった手法に関して色分類の精度比較を行っており、いずれも高い分類

精度を示した。Weijerらの実験では色分類を行う際、物体に対するマスクが与えられるが、本研究ではマスクは与えられず、衣類領域を推定することで色分類を行う。また、eBayデータセットでは平坦なテクスチャとなる物体が多く見られるが、本研究で用いるデータセットは複雑なテクスチャを含むものも多く存在することとなる。加えて、衣類の色は中間色となるようなものも多く存在し、それらの要因は色分類問題をより難しいものとしている。また、Weijerらはピクセル単位での色分類精度も示しており、その結果から周囲画素の情報を利用した方が精度がよいことが分かっている。本手法においてもこの性質を利用し、色分類を行う際は小領域単位での情報を用いる。

3. 提案手法

3.1 概要

本研究で取り扱う問題は、人物画像を入力すると、その画像中の人物が着ている衣類の色名を返す、というものになる。提案手法では図1の手順に従い、画像中に含まれる衣類の色分類を行う。このとき、入力する画像は、リサイジングやクリッピングの処理が必要であれば前処理として行っておく。また、顔検出～領域細分化の処理の概要を図2に示す。



図1 処理手順の概要図

3.1.1 顔検出

本研究で扱う衣類の画像は着用されているものとし、衣類領域の推定は顔検出から始まる。ここでは、Haar-like特徴を用いたブースティングによる分類器を利用し、顔検出を行っている。また、後述する領域分割手法では、物体と背景との大まかな境界を与える必要がある。そのため、顔検出から得られる顔の領域範囲をもとにし、衣類領域を大まかに決定する。

3.1.2 人物・背景の分離

ここでは領域分割手法としてGrabCutを用いる[6]。Grab-



図 2 顔検出～領域細分化の処理の流れ

Cut は 2 つのエネルギー関数を最小化することで、物体と背景に画像を分割する。エネルギー関数は、隣接画素の差分値と、画素が物体・背景に属する確率モデルを使って定義される。背景に属する画素集合を指定し、それらの値をガウス混合分布によって表現することで確率モデルが決定される。前述した顔検出ステップで大まかな領域を決定しなければならないのは、このためである。

3.1.3 領域細分化

衣類単位での処理を行うため、領域を細分化する。細分化された画像領域はスーパーピクセルと呼ばれる。スーパーピクセルの作成には Felzenszwalb らが提案したグラフベースの領域分割手法を用いる [7]。画素をノードとみなした最小全域木を構成していく過程で、適時、木の分断を行うことで領域分割を達成する手法となる。領域分割数の指定を行う必要がなく、適切な分割数が一意に求まらない問題と相性が良いため、本研究でのスーパーピクセルの作成に使用した。

3.1.4 小領域単位でのラベル付け

領域細分化を行った後は、各スーパーピクセルから画像特徴を抽出し、ラベル付けを行う。ここで、ラベルは代表的な 12 色の色名を表すものとする。すなわち、この手順では各スーパーピクセルに対し色分類を行うことになる。これによって、各スーパーピクセルが 12 色の内のいずれかの色名を持つことになる。

3.1.5 小領域の統合

ここでは同じラベルを持つスーパーピクセル同士を結合していく。この処理により、複数個あったスーパーピクセルが、最大でも 12 個までにまとめられる。12 色の領域はそれぞれ、12 色の内のいずれかの色が占める領域を表すことになる。領域統合の結果、図 3 のようなマスクが得られる。例として使用した画像には blue と white の領域が多く確認できるが、他のマスクを見た場合、ノイズのような割り当てが起きているのが確認できる。最後の手順で、これらのノイズを削除する処理を行う。

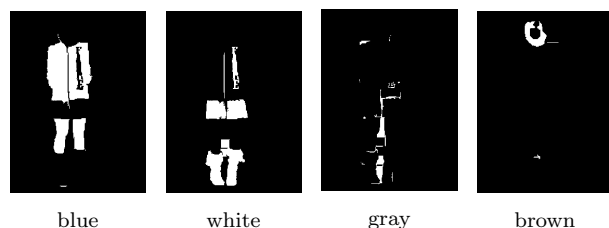


図 3 統合後のスーパーピクセルの例

3.1.6 ラベルの出力

最後の処理として、面積ベースの決定法によってノイズとなるラベルの除去を行い、衣類のラベルを求める。衣類のテクスチャの複雑さが増すほど、ノイズのようなマスク画像の発生が起きやすい。大きな領域として検出されているマスクほど、その画像に含まれる衣類の色情報として優位な色であると判断することができる。そこで、12 個のマスクに含まれる白い領域の総和を定数倍 (0:1) し、それを閾値として、マスクの面積をベースとしたラベルの採択を行う。これにより、図 3 の例からは、gray、brown のようなラベルが破棄され、blue、white のラベルのみが選ばれることが期待できる。

3.2 色分類

本研究では、各スーパーピクセルから画像特徴を抽出し、機械学習法によってそれらを表 1 で示される 12 色のいずれかに分類する。

表 1 分類する 12 色一覧

yellow	orange	green	gray	purple	pink
brown	black	blue	beige	white	red

画像特徴として、各色空間におけるカラーヒストグラムと Bag-of-Color(BoC) ヒストグラムを抽出し、精度比較を行った。カラーヒストグラムは画像中の各画素の出現頻度をカウントしたものである。色空間を等間隔に区切り、そこに属する画素の出現頻度をカウントすることによりヒストグラムが得られる。通常のカラーヒストグラムでは固定した bin で量子化するのに対して、Bag-of-Color(BoC) ではベクトル量子化を用いて bin をデータ分布に基づいた設定により量子化する [8]。データ分布中の代表点の集合はコードブックと呼ばれ、コードブックを基にしてヒストグラムが作成される。カラー画像の場合、各画素の値は 3 次元のベクトルとみなすことができ、コードブックは実際には 3 次元ベクトルの集合となる。通常のカラーヒストグラムでは色空間が等間隔に区切られ投票が行われることになるが、BoC では、色の分布を表現するための重要度に応じて色空間が分割されることになる。

スーパーピクセルに対する色分類では、前述の画像特徴を用い、機械学習法によってラベル付けを行う。分類器として、K-Nearest-Neighbor(KNN) と Support Vector Machine(SVM) を用い、精度比較を行った。KNN はシンプルなアルゴリズムであるが、データ構造を工夫することで高速な検索が可能であり、大規模データでの利用に期待が集まっている。SVM は色分類に関する既存研究でも用いられた手法であり、それらで高精度の分類結果が示されているように、優れた分類器の一つである。

なお、これらの分類手法はスーパーピクセルに対する色分類での使用に止まる。画像中に含まれる衣類に対するラベルは、提案手法で示した通り、その後の処理が必要である。衣類のラベルは、各スーパーピクセルを統合し、面積ベースによるラベルの採択を行うことで、最終的な色名が割り出される。

4. 実験

4.1 データセット

本研究では、ファッションサイト^(注3)から収集した 9592 枚の人物画像を対象とし、実験を行った。収集した画像に対し、提案手法の手順で人物・背景の分離までの処理を行った。今回は、初期処理の人物領域の切り出しが現状の技術で十分できる画像のみを選んで、色分類の精度の実験を行う。Web 上には大量の衣類画像が存在するため、再現率は重視せず、適合率を重視することにし、認識が出来る画像を精度よく色分類できることを目指した。具体的には、以降に示す実験では、こうして得られた画像の内、GrabCut が大方成功しているものから 118 人の画像を選出し、色分類の精度比較を行った。画像の選出に関して、118 人の画像は人物・背景の分離が成功しているかどうかを基準となる。本研究は、実在する大量データに対して処理を行うことを視野に入れており、それらのデータに含まれる衣類は中間色や複雑なテクスチャとなる物が多くを占めている。選出した 118 枚の画像にも、そういった複雑性を持つ画像が多く含まれているため、既存の研究に比べ色分類が難しいデータセットとなっている。選出した画像の衣類領域を自動抽出した例を図 4 に示す。

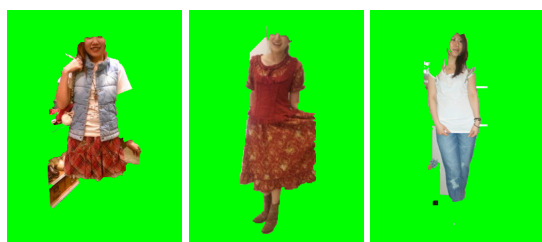


図 4 選出した画像の衣類領域を自動抽出した例

118 人の画像には、手動により衣類や目立つオブジェクト (例えばブーツやバッグ) のマスクを与え、計 277 件のマスクに対し色名によるラベルを付けた。各マスクは、人間の視覚的な感覚をもとに与えられる。つまり、衣類に細かい模様や影が含まれている場合でも、それらは無視できるものとして同じ衣類領域としてマスクをとった。マスクの例を図 5 に示す。また、作成したデータセットの、各ラベルの枚数を表 2 に示す。

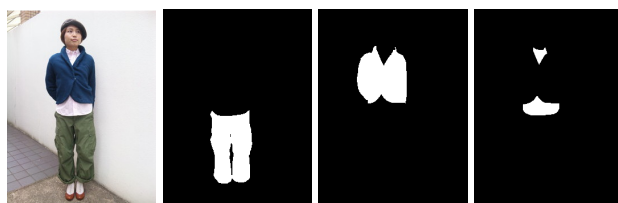


図 5 左から順に、元画像、green、blue、white のマスクとなる

以上のデータセットを学習データとして用いることで機械学習を行い、5-fold のクロスバリデーションにより色分類精度を求めた。

表 2 各色にラベル付けされた件数一覧

yellow: 8	orange: 8	green: 11	gray: 33
purple: 13	pink: 17	brown: 34	black: 50
blue: 23	beige: 29	white: 33	red: 18

4.2 マスクが与えられたときの色分類精度

作成したデータセットを用いて、マスクが与えられたときの色分類精度を求めた。マスクは、衣類や目立つオブジェクト毎のマスクである。マスクで指定された領域から画像特徴を抽出し、12 色の学習・分類を行った。画像特徴、学習手法を変えて実験を行い、5-fold のクロスバリデーションにより 12 色の分類率を求めた。分類率は次式のようにして求める。なお、今回の実験では、分類対象物の総数は常に 277 となる。

$$\text{分類率} = \frac{\text{正しく分類されたものの総数}}{\text{分類対象物の総数}} \quad (1)$$

本研究が提案する手法では衣類領域等の指定は不要であるが、この節で示す実験内容は、衣類領域が確実に分かっているという条件において、どの程度の色分類が可能であるかを確かめるための実験となる。

4.2.1 12 色の分類率

12 色の分類率の結果を図 6 に示す。このときの画像特徴はそれぞれ 1000 次元のヒストグラムとなる。

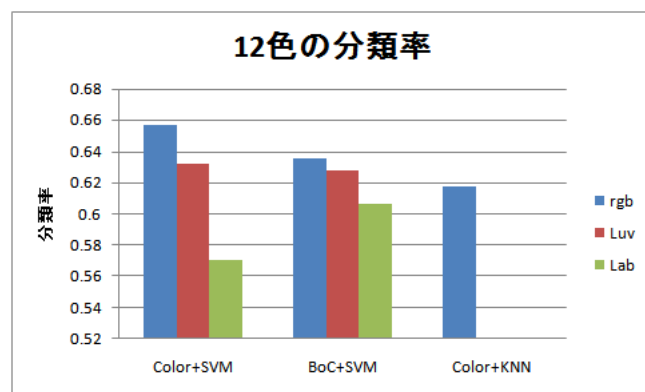


図 6 マスクが与えられたとき

カラーヒストグラムと SVM において、RGB 色空間での結果が最もよく、そのときの分類率は 65.7% となった。BoC と SVM による分類においても RGB 色空間が最もよく、そのときの分類率は 63.5% となった。KNN を用いた分類では、K の値を変化させ実験を行い、最もよい値を使用した。そのときの K の値は K=9 となり、分類率は 61.7% となった。Weijer らによる eBay データセットを用いた実験では、Lab カラーヒストグラムと SVM による精度を出しているが、そのときの精度は 96% となっている。本研究においても同様の手法を試しているが、その精度は 57.0% である。この原因は、使用したデータセットの複雑性に関する影響が大きい。eBay データセットでの衣類は、平坦なテクスチャとなっているものが多く、そのため各衣類領域のヒストグラムは単純な色分布となっていることと考えられる。一方、本研究のデータセットでは、細かい模様や影になっている領域もマスクの領域として含んでいるため、抽

(注3): <http://eruca.jp/>

出されたヒストグラムは複雑な色分布を示すこととなる。一般的に、機械学習法によってラベルが付けによる分類が行われる場合、使用する各ヒストグラムの複雑性が増すほど、分類が難しい問題となる。また、[5] では 11 色での色分類を行っているが、それらのデータセットには中間色が少ない。本研究のデータセットは、ファッションサイトから収集した画像を用いているため、12 色の中での中間色となるような衣類も多く存在する。人間の視覚においても、中間色の判断は難しく、人によって違った判断とされるものも多い。これらの要因が、本研究で作成したデータセットの色分類をより難しい問題としている。

4.3 提案手法の色分類精度

前の節で示した実験は衣類領域がマスクとして与えられるため、色分類手法の性能を確かめるという目的が主であった。本研究の提案手法は、マスクを与える必要のないアルゴリズムとなっている。この節では、提案手法の色分類精度を求め、マスクの指定の有無でどの程度性能が変化するかを確認する。

本研究は、画像中の衣類領域に対し色名によるラベル付けを行うことが目的であるが、大量のデータを処理するためにもそれらの工程は自動化されている必要がある。つまり、本手法では前節で与えたような衣類領域のマスク等は与えられない、という条件になる。本手法では提案手法の章で示したとおり、スーパーピクセル毎に色名を求め、それらの領域を統合し、面積ベースによって色名の決定を行う。

スーパーピクセル毎での色分類では、学習データとして前節で使用した分類器をそのまま用いた。学習データは衣類領域単位での画像特徴となり、テストデータはスーパーピクセル単位での画像特徴となる。各画像特徴は異なる単位で抽出されているため、同じデータになることはほぼないが、公正を期するため、5-fold のクロスバリデーションによる色分類を行った。つまり、同じ人物から得た特徴は、学習・テストの両方で同時に用いられることはない。これによってスーパーピクセル毎に色名によるラベル付を行い、領域の統合を行った。

提案手法で述べたように、領域統合によって得られる 12 個のスーパーピクセル (それぞれ各色名を表わすことになる) は、最終的に面積ベースで選択される。選択されたスーパーピクセルのラベルが、そのまま、画像中に含まれる衣類の色名となる。面積ベースの選択では、12 個の領域の総和に定数倍 (0:1) をかけたものを閾値として用い、それ以下のものを破棄することでラベルの採択が行われる。そのときの定数倍の値を *rate* とし、*rate* の値を変え、再現率 (*recall*) と適合率 (*precision*) を求めた。*recall* と *precision* はそれぞれ次のようにして求める。

$$recall = \frac{\text{正しく分類されたものの総数}}{\text{検出されるべきラベルの総数}} \quad (2)$$

$$precision = \frac{\text{正しく分類されたものの総数}}{\text{検出されたラベルの総数}} \quad (3)$$

また、このときの *F* 値は次式によって求めた。

$$F = \frac{2recall \cdot precision}{recall + precision} \quad (4)$$

今回の実験では、人物画像に対し複数個のラベルが付いた

データセットとなる。ラベルの数は画像によって異なり、例えばシャツ・上着・ズボンがそれぞれ異なる色である場合、ラベルの数は 3 つとなる。すなわち、検出されるべきラベルの総数とは、分類対象物の総数であり、常に 277 である。また、本手法は画像を入力として与えるとその画像中の衣類のラベルが複数個出力されるが、パラメータ *rate* の値によりそのラベルの検出数は変化する。例えば、red・blue・black のラベルが付いている画像に対し、検出されたラベルが blue・yellow であれば、*recall*=1/3、*precision*=1/2 となる。正解・不正解の判定は、検出されたラベルが正解ラベルに含まれているかで判断する。正解ラベルは衣類領域ごとに付けられているが、正解・不正解の判断基準に、検出された衣類領域の位置・面積といった情報は含まれない。ラベル名でのみの比較となるが、これは、領域分割としての精度ではなく、色名検出としての精度を求めためである。

4.3.1 カラーヒストグラムと SVM

カラーヒストグラムと SVM による色分類において、パラメータ *rate* の値を変化させたときの再現率のグラフを図 7、適合率のグラフを図 8 に示す。また、そのときの *F* 値を求めたものを図 9 に示す。カラーヒストグラムは 1000 次元とし、色空間を RGB、Luv、Lab としたときでそれぞれの値を求めた。

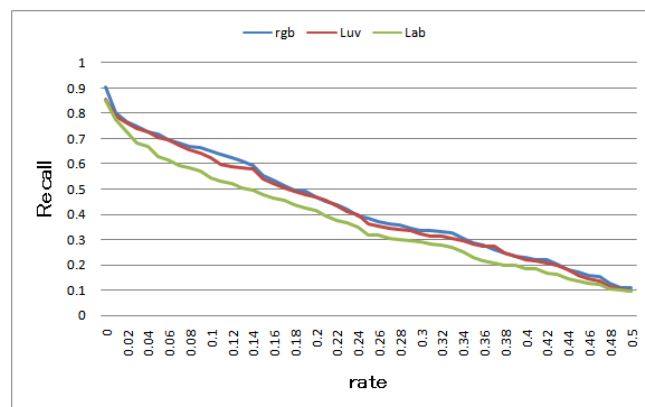


図 7 rate の値を変化させたときの再現率

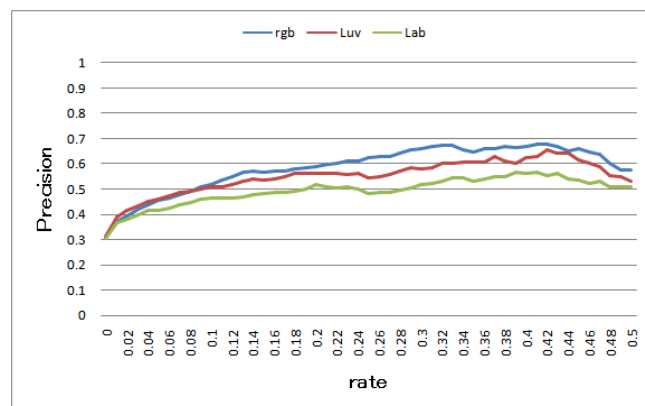


図 8 rate の値を変化させたときの適合率

再現率に関して、*rate* の値を上げるにつれて徐々に減少していくのが確認できる。*rate* は最終的なラベルの選択に係るパ

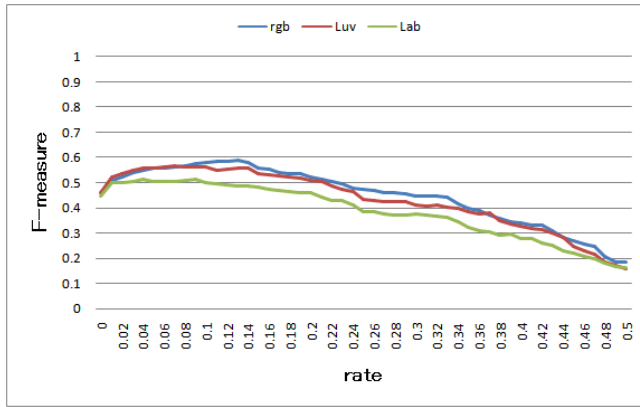


図9 rateの値を変化させたときのF値

ラーメータであり、0に近づくにつれてラベルの検出数が増えることになる。rateが0のときは検出した色すべてを採用したときと同等であり、rateが1のときは、統合されたスーパーピクセルが唯一の色に絞られるときとなる。

適合率に関して、rateの値を上げるにつれて上昇していく傾向にあるが、ある程度の値を境にその性能は頭打ちとなり、その後は不安定ながらも減少傾向にある。rate=0.33のときのRGBでの適合率は67.4%となるが、このとき検出されたラベルの数は135件となり、そのうち正解となったのは91件となる。ここで、rate=0.33というのは、検出された12色の総面積を3で割った値を閾値としたときとほぼ同じであり、そのときの再現率は32.9%となる。正解ラベルの総数は277件であるが、この検出数からも、総面積の1/3を占める衣類領域を閾値とするのは、非常に厳しい設定であるのが分かる。

F値が最高となったのはRGBにおいてrate=0.13としたときであり、そのときのF値は0.5878となった。そのときの再現率は61.0%であり、適合率は56.7%であった。F値はこの後減少していくことになるが、rate=0.13~0.17の間に関しては、適合率にその主な原因があることが分かる。適合率はrate=0.13~0.17の間、ほぼ値の変動がなく、精度の伸びが止まっている状態にある。その間の最大・最小の精度の差は0.7%であり、ここでラベルの検出に関して一時的なピークを迎えていることが確認できる。また、再現率に関してはrateが0.14~0.15と変化する間に約4%の精度の低下がみられる。その結果、0.14~0.15と変化する間のF値においても大きく値が低下する原因となっている。

マスクを与えたときの実験での分類率の最高値は65.7%であった。それに対し、マスクが与えられない条件での色分類となる本手法では、F値が最高のとき、再現率=61.0%、適合率=56.7%となる。これらの精度をみると、マスクなしという厳しい条件においても、色名によるラベル付けの性能低下が少ないことが分かる。マスクを与えたときの分類率の最高値が65.7%である以上、本手法における色分類精度の限界はその付近となってしまうことが予想される。しかし、その性能は、色分類の性能にほぼ依存することとなるため、マスクを与えたときの分類率を高めることができれば、本手法におけるラベル付けの性能も高まるが大いに期待できる。

5. おわりに

本研究では、画像中の衣類領域からの色情報抽出手法の提案を行った。提案した手法では衣類領域の指定を行う必要がなく、顔検出と領域分割によって衣類領域の推定が行われる。画像中に存在する複数の衣類・オブジェクトに対応することができ、それらの色の最終的な判定は面積ベースの決定法で行われる。人間の視覚においても、着用している衣類の色を判断する際、それらの全体に占める割合での大きさは重要な判断材料となっている。そのため、最終的なラベルの決定を行う際に面積ベースで選択を行うのは、人間の視覚に近い判定法だと言える。

今回の実験では全体で9592件ある内の118人、277個の物体に対する色分類実験となったが、Web上に大量の画像データが存在することを考えると、再現率より適合率を重視する方がよい。衣類の色を割り出せた後のステップとして、データマイニング手法によるデータ解析が挙げられるが、このときに色の分類精度が非常に重要となる。顔検出が未検出等の理由で色の判定ができずに画像が破棄されていっても、データが大量にある場合はさほど問題にならないことが予想され、マイニングの性質上、正確さの方が問題となることが多いためである。

本手法では画像からスーパーピクセルを得て、それらに対しラベルを付けるが、領域細分化を行った段階で、小さすぎる領域の削除を行うといった改良案がある。影の領域や、衣類と衣類の境目は非常に小さなスーパーピクセルとなる傾向がある。そうしたスーパーピクセルは、衣類の色の判定においてノイズとなる可能性が高く、これらを削除することで精度向上が見込める。また、色分類は機械学習法による分類となるが、そこで学習として使うデータの質や量を増やすような改良を行うことでも精度向上が見込める。

文献

- [1] A.C. Gallagher and T. Chen, "Clothing cosegmentation for recognizing people," Proc. of IEEE Computer Vision and Pattern Recognition, pp.1-8, 2008.
- [2] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," Proc. of IEEE Computer Vision and Pattern Recognition, pp.2360-2367, 2010.
- [3] G. Wu, A. Rahimi, K. Goh, T. Tsai, A. Jain, Y. Wu, E.Y. Chang, and Y.F. Wang, "Identifying color in motion in video sensors," Proc. of IEEE Computer Vision and Pattern Recognition, pp.561-569, 2006.
- [4] D.A. Vaquero, R.S. Feris, D. Tran, L. Brown, A. Hampapur, and M. Turk, "Attribute-based people search in surveillance environments," Proc. of Workshop on Applications of Computer Vision, pp.1-8, 2009.
- [5] J. Van De Weijer, C. Schmid, and J. Verbeek, "Learning color names from real-world images," Proc. of IEEE Computer Vision and Pattern Recognition, pp.1-8, 2007.
- [6] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," Proc. of ACM SIGGRAPH, pp.309-314, 2004.
- [7] P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient graph-based image segmentation," International Journal of Computer Vision, vol.59, no.2, pp.167-181, 2004.
- [8] C. Wengert, M. Douze, H. Jégou, et al., "Bag-of-colors for improved image search," ACM Multimedia, pp.1437-1440, 2011.