

画像特徴を用いたオノマトペ表現による紙の材質感分析

上村 純^{1,a)} 田川 聖^{1,b)} 松下 康之^{1,c)} 八木 康史^{1,d)}

概要：我々人間は視覚を通して、対象物が金属、プラスチックなどのどのような素材で構成されているかを簡単に識別できる。また、その物体の手触りや柔らかさ、温度といった状態についても瞬時に判断ができる。このように質感の僅かな違いに対しても人間は敏感であるため、材料作成の現場において顧客の求める質感や使用感の実現や製品の品質管理などが大きな課題となっている。そのため、質感を定量評価するシステムを構築し、生産者が利用可能な質感に関するデータベースを作成することが必要となる。そこで、まず人の質感評価の基準を明らかにすることを目的とし、本研究では、紙を質感分析における対象物とし撮影実験を行い、得られた画像群に対して主成分分析と離散コサイン変換を用いて画像特徴量を抽出し、それらに基づく識別器に主観的質感評価を学習させ各識別器の性能評価を行う。

1. はじめに

我々の身の回りには様々な質感を持つものが存在している。人間は質感を知覚することによって、物体の素材やその状態などといった生きていくために必要不可欠な情報を得ている。質感は、視覚、触覚、聴覚といった複数の感覚を通して感じるが、視覚、触覚、聴覚などといった単一の感覚から感じることができる質感も存在する。例えば、我々は視覚だけを通して物体を認識するとき、プラスチックや金属などの物体を構成する素材だけではなく、物体の手触りや柔らかさなどといった物体の表面の状態まで判断することができる。

また、そのような質感認知は対象物に対する情動や嗜好に深く関わる。つまり、質感をどのように感じるかで人がその対象物に対して与える価値判断や意思決定に変化を及ぼす。例えば、同じ素材であっても光沢感がある方に対しては高級感を感じたり、本物近似の質感を再現された素材に対しては異なる素材のものであっても本物感を感じることがある。加えて、人間は質感の違いについて非常に敏感であるため、材質の加工の些細な違いが質感において大きな差を生み出す。このことが材料作成の現場において顧客が求める質感を実現することを難しくする要因となっている。そのため、質感を定量評価するための手法を確立することが課題となっており、そのために生産者が利用可能な

質感に関するデータベースを作成することが重要である。

視覚に関する質感の研究として、大槻ら [1] はターンテーブルを用いて様々な角度からの画像を撮影するような複雑な装置を用いて物体形状や反射特性を計測している。一方で、下田ら [2] は質感を表現する言葉に対応した画像を Improved Fisher Vector (IFV) [3] や Deep Convolutional Neural Network (DCNN) [4] といった画像特徴を用いることで学習させ、質感の認識可能性の分析を行っている。下田らの方法では、真値となる画像のデータセットを Web 画像のマイニングによって構築しており、人手を一切介さずに行っている。しかし、そのような方法ではデータセット内にノイズ画像が含まれるという可能性がある。そのため、本研究の方針としては、クラウドソーシングといった人手を介する方法で精度の高いデータセットを構築することが挙げられる。それらを踏まえ、本研究の特色として「人間から得られる主観的なデータ」と「コンピュータが画像から得る特徴」を関連付けるという点にある。人間は視覚から得られる情報のみで質感を認知できることから、本研究では、コンピュータビジョンの技術を用いて、複雑な計測はせず、人間の視覚から得る情報と同等以上の解像度の画像から得られる情報を元に人が質感を評価する際の基準を明らかにする。そのため、様々な手法で抽出した画像特徴量と各画像に対する主観データを機械学習させた識別器の性能を評価することで、人間が画像から得られるどのような特徴を質感評価の基準としているかを考察する。

¹ 大阪大学
Osaka University

a) j-kamimura@am.sanken.osaka-u.ac.jp

b) tagawa@am.sanken.osaka-u.ac.jp

c) yasumat@ist.osaka-u.ac.jp

d) yagi@am.sanken.osaka-u.ac.jp

2. 画像からの材質感分析手法

前章で述べたように、人は物体から得られる視覚情報のみで、質感に関する様々な情報を得ることができる。同様に画像に含まれる情報だけでも機械が人間と同じように質感を識別できると考えられる。そこで、画像から得られる人間の質感に関する主観的評価を機械学習させ、人間の質感評価の基準を明らかにする。

2.1 材質感とは

人は素材の材質感を形容詞やオノマトペを用いて表現している。言い換えれば、人間は各素材に対して様々なオノマトペが素材の質感の表現に適切であるかないかの2値で表現している。すなわち、図1のように素材の材質感は様々なオノマトペを要素とした2値ベクトルとして表現できる。本研究の目的は、画像が入力して与えられたときに材質感を表した2値ベクトルが出力として得られる関数を求めることである。

	(ざらざら, つるつる, さらさら, ..., かさかさ)
素材1 (1, 0, 0, ..., 1)
素材2 (0, 0, 1, ..., 0)
.	.
.	.
.	.

図1 材質感の表現方法

2.2 画像特徴の抽出のための手法

機械学習に取り組むにあたり、質感を識別するために適切な画像特徴の抽出を行う必要がある。人間が視覚情報だけで質感を判断するとき、画像の画素1つ1つの値を見るような微視的な視点ではなく、素材表面全体からの証明場からの光の反射による明暗のパターンが深く関わる。本研究では、主成分分析 (Principal Component Analysis: PCA) から得られる全素材のテクスチャのパターンを用いて抽出した特徴量と離散コサイン変換 (Discrete Cosine Transform: DCT) で使われる周波数パターンを用いて抽出した特徴量から、それぞれに基づいた識別器の評価を行うことで質感を感じさせるような特徴パターンを明らかにする。

主成分分析

主成分分析とは、多次元データのもつ情報をできるだけ損なわずに低次元空間に情報を縮約する方法である。主成分分析の目的として、もとのデータの情報の損失ができるだけ小さくなるような基底を次元空間内に探すことが挙げられる。射影したデータのばらつきが大きいほど、もと

のデータの情報を多く含んでいると考えられるため、射影したデータの分散が最大となるような基底を探す。このようにして得られた基底は主成分と呼ばれ、求められた順にデータに共通する特徴が現れる。このことから、材料ごとのテクスチャのパターンの類似や差異に注目し、主成分分析を用いた特徴抽出を行う。

離散コサイン変換

離散コサイン変換とは、信号を周波数成分に直交変換する手法の一つである。離散コサイン変換は、画像の圧縮形式である JPEG や MPEG などの国際標準に採用されている。 N 次元の画像信号を離散コサイン変換することで、画像を N 個の周波数成分の足し合わせで表現できる。このように画像の周期パターンが質感の表現に与える影響に注目し、離散コサイン変換を用いた特徴抽出を行う。

2.3 主観的評価の学習手法

パターン認識において、抽出した特徴が識別器でクラシフィケーションするのに本質的であるかを評価することが必要である。すなわち、各特徴の数値を表すための座標軸で構成される特徴空間において、特徴量を要素とする各画像の特徴ベクトルがそれぞれ適切なクラスに分類できているかを判定することが必要である。本研究では、パターン識別器を作成するにあたってサポートベクターマシン (Support Vector Machine: SVM) を用いてパターン識別器に学習させる。各画像から得られる質感に関する主観的評価、ここでは質感を表すようなオノマトペに対して当てはまるか当てはまらないかというデータを学習させる。サポートベクターマシンでは、学習データが d 次元の特徴を持っていた時 $d-1$ 次元の超平面で2クラスに分離する。この時、ポジティブとネガティブという2クラスの学習サンプルの距離が最大となる超平面を設定することで識別器の汎化能力が高くなる。サポートベクターマシンは様々なカーネル関数と組み合わせることで非線形識別器に拡張することが可能であるが、今回は事例となるデータ数に対して特徴量の次元数が大きいため、高次元に写像しても精度向上が見込まれないと考え、線形なものを利用する。

3. 材質感データの取得と解析

本研究では、「紙」を対象とした撮影実験を行い、それらを解析し、紙の質感に関する識別器の作成をする。デジタル時代において「紙」は、情報伝達を行うための媒体としての役割だけではなく、感性を伝えるメディアとして、より広い可能性を秘めた素材である。それによって、様々な質感をもったものが存在するため、学習や評価を行うためのデータセットを揃えるのに適していると考えられる。

3.1 撮影実験

本実験では、紙の持つ繊維の構造や表面の加工から質感

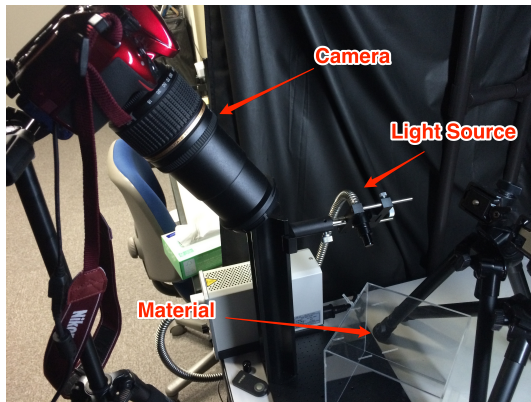


図 2 質感解析のため画像撮影装置

カメラと光源を 1 台ずつ固定した状態で、アクリル台の上に紙をセットし撮影する。

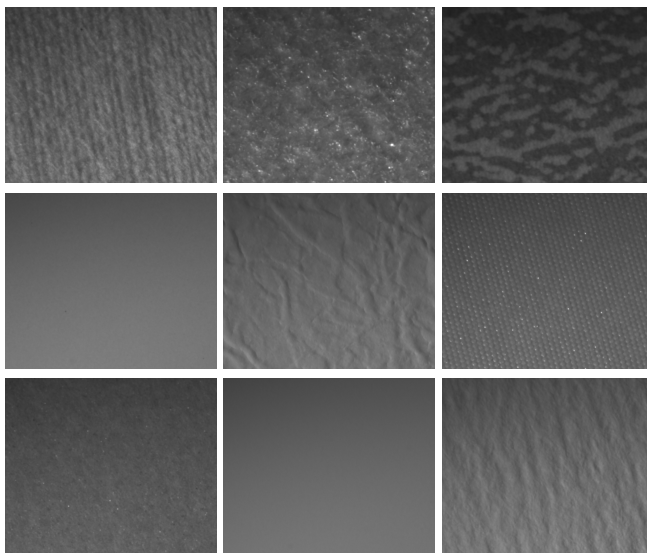


図 3 様々な質感を示す紙の撮影画像

ブレンな表面な紙や粗い質感でラフな手触りの紙、エンボス加工を施した紙、毛羽立った表面を持つ紙など様々な質感を反映する。

の違いが現れると考え、白色を基調とした 94 枚の紙を収集し実験を行った。

撮影装置

紙の撮影方法として、図 2 のような撮影装置を用いた。撮影対象に対して、カメラ (Nikon D5200) と光源 (シグマ光機 SLA-100) を 1 台ずつ固定し、撮影するシステムを用いる。カメラの設定は、ISO 感度は 100、F 値は 25、シャッターは 1 秒と設定する。

次にデータから特徴を抽出しやすくするために、撮影した画像データに処理を加える。紙の質感は、繊維質の状態や表面の加工の具合などにより画像上に現れる陰影のパターンから判断されると考えられる。そのため、12mm 角程度の小領域を 960×800 の解像度で撮影した画像をグレースケールに変換する。処理後の画像の一部を図 3 に示す。

3.2 各画像の特徴抽出

パターン認識を行うためには、得られたデータセットから対象を区別できるような特徴を抽出しなければならない。オリジナルの画像が持つ情報が冗長である場合、すなわち特徴の次元が大きい時、識別器の性能の向上が見込めない。そのため、オリジナルの画像から本質的な性質を損なうことなく特徴の次元を小さくすることが必要である。前章で述べたように、主成分分析と離散コサイン変換という 2 つの別々な手法を用いて各画像から特徴量の抽出を行う。

3.2.1 主成分分析を用いた特徴抽出

基底関数の計算

主成分分析を行うことで、もとのデータの情報の損失を抑えつつ、本質的な情報を取り出すことができる。これによって多次元のデータを低次元空間に縮約できる。まず、94 枚の画像群から空間上のデータの分散が大きくなるように画像群から主成分の基底を求める。各主成分 $Z_i (i = 1, \dots, 94)$ は、式 (1) のように各変数 $X_j (j = 1, \dots, 768000)$ の多項式で表すことができる。ここでは、変数 X_j は各画像における j 番目のピクセルの輝度値を表す。

$$Z_i = \sum_{j=1}^{768000} w_{i,j} X_j \quad (1)$$

基底画像の作成

求めた各主成分から基底となる画像を作成する。基底画像は、式 (1) の各ピクセルに対する重み w_i を $0 \sim 255$ の値に正規化し、 800×960 の形のマトリックスに変換することで作成する。作成した基底画像は合計で 94 枚である。第 6 主成分までの基底画像を図 4 に示す。

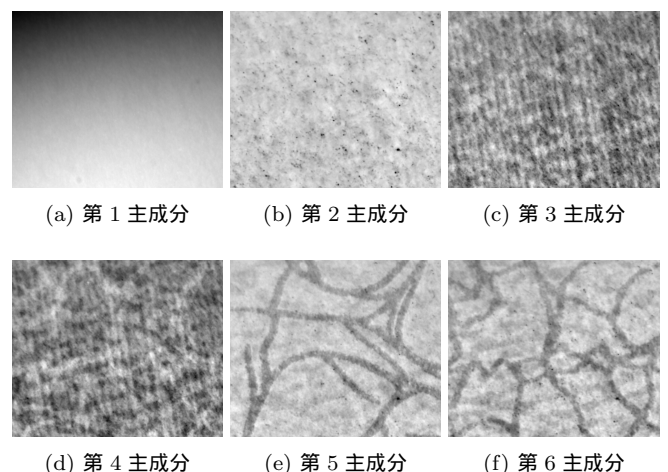


図 4 PCA により得られた主成分パターンを表す基底画像
各主成分の基底画像は元の画像群に対し固有のものであり、番号が小さい主成分の基底ほどデータの分散が大きくなる方向に軸をとる。例えば、第 1 主成分の基底画像は元の画像群に共通するような陰影のパターンを表現している。

基底画像によるオリジナル画像の再構成

$$\mathbf{I}_m = \sum_{n=1}^{94} a_{m,n} \mathbf{i}_n \quad (2)$$

$$[a_{m,1}, a_{m,2}, \dots, a_{m,94}] = \mathbf{I}_m \mathbf{A}^+ \quad (3)$$

作成した基底画像を用いてオリジナルの画像の特徴抽出を行う。式(2)で示すように、オリジナル画像 \mathbf{I}_m を第 n 主成分の基底画像 \mathbf{i}_n の線形結合で表現し、各係数 $a_{m,n}$ を要素とする 94 次元のベクトルを各オリジナル画像の特徴量として利用する。特徴量は式(3)のようにオリジナル画像 \mathbf{I}_n に擬似逆行列 $\mathbf{A}^+ (= [\mathbf{i}_1, \mathbf{i}_2, \dots, \mathbf{i}_{94}]^+)$ を掛け合わせることで求めることができる。

3.2.2 2次元離散コサイン変換を用いた特徴抽出

N 点における 1次元離散コサイン変換、逆変換はそれぞれ式(4)、(5)で表される。

$$F(u) = \sqrt{\frac{2}{N}} \sum_{x=0}^{N-1} f(x) \cos \frac{(2x+1)\pi}{2N} u \quad (4)$$

$(u = 0, 1, \dots, N-1)$

$$f(x) = \sqrt{\frac{2}{N}} \sum_{u=0}^{N-1} F(u) \cos \frac{(2x+1)\pi}{2N} u \quad (5)$$

$(x = 0, 1, \dots, N-1)$

2次元離散コサイン変換は、1次元離散コサイン変換の拡張であり、水平方向、垂直方向について1次元離散コサイン変換を行うことで実現できる。 $N \times N$ のブロックに対する2次元離散コサイン変換、逆変換はそれぞれ式(6),(7)で表せる。

$$F(u, v) = \frac{2C_u C_v}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \cos \left(\frac{(2x+1)\pi}{2N} u \right) \cos \left(\frac{(2y+1)\pi}{2N} v \right) \quad (6)$$

$(u, v = 0, 1, \dots, N-1)$

$$f(x, y) = \frac{2}{N} \sum_{u=0}^{N-1} \sum_{v=2}^{N-1} C_u C_v F(u, v) \cos \left(\frac{(2x+1)\pi}{2N} u \right) \cos \left(\frac{(2y+1)\pi}{2N} v \right) \quad (7)$$

$(x, y = 0, 1, \dots, N-1)$

但し、

$$C_u, C_v = \begin{cases} \frac{1}{\sqrt{2}} & (u = 0, v = 0) \\ 1 & (u \neq 0, v \neq 0) \end{cases}$$

画像信号に対し、2次元離散コサイン変換を行うことで画

像の信号波形をいくつかの成分に分解し、それぞれの大きさを情報として符号化することで、データの次元を減らすことが可能である。離散コサイン変換の処理を図5に示す。今回、画像を小領域に分割して離散コサイン変換を行わず、画像全体に対して離散コサイン変換を行った。オリジナル画像 $f(x, y)$ に離散コサイン変換することで、周波数画像 $F(u, v)$ を生成することができる。周波数画像は、各画素の値が各周波数成分の係数を表しており、左上ほど低周波成分、右下ほど高周波成分を表している。

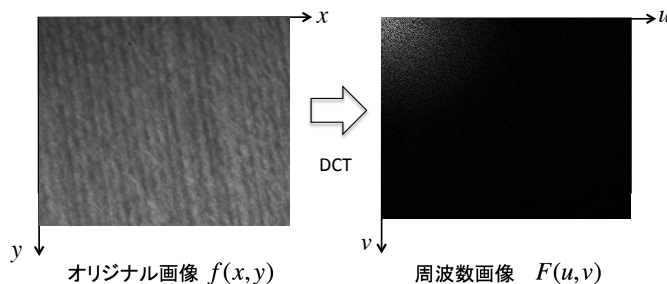


図5 離散コサイン変換の処理

今回、画像サイズが 960×800 であるため、周波数画像のすべての情報を利用すると特徴量が 768000 次元という膨大な情報量となる。本研究では、撮影画像の高周波成分がほとんど見られなかったことから、オリジナル画像にローパスフィルタをかけた後の特徴量を用いる。すなわち、周波数画像の低周波成分だけを用いる。そこで問題となるのが、ローパスフィルタをどのくらいかけるのが適切であるかということであるが、図6のようにローパスフィルタをかけた周波数画像を逆変換して得られた復元画像を元に、特徴量の次元の目安を決める。今回 100×100 サイズの低周波成分だけを用いた場合、オリジナル画像と同等近くのものが見られたので、10000 次元を上限の目安として考える。

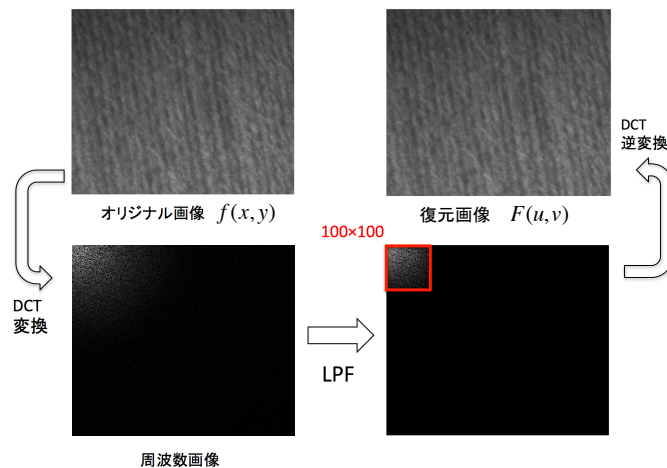


図6 高周波成分除去の一例

また、基底関数の波長パターンが半波長だけ位相がずれている場合、つまり関数の正負が逆になっている場合に対しても、人間が認識するパターンとしては同じものであると考えることができるため、DCT 係数の絶対値を取ったものを特徴量の成分として用いる。

3.3 質感に関する主観的評価データ

サポートベクターマシンは 2 クラスのパターン認識のための手法であるため、予め正解のデータとしてデータセットの各画像を質感を表すオノマトペに対し 2 クラスに分類する必要がある。ここでの 2 クラスとは、紙の質感を表すオノマトペに対して各画像が適切であるか不適切であるかを意味する。例えば、「ざらざら」というオノマトペに対して、画像から質感が「ざらざらである」と判断できればポジティブなクラス、「ざらざらでない」と判断できればネガティブなクラスに分類される。質感に関する主観的データを作成するにあたって、11 人の被験者に協力のもと、各人が各画像を 10 個の質感を表すオノマトペに対して 2 つのクラスに分類する。質感を表すオノマトペとしては、「ざらざら」、「さらさら」、「しわしわ」、「つるつる」、「ごわごわ」、「かさかさ」、「ぱりぱり」、「へろへろ」、「きらきら」、「へろへろ」という 10 個のキーワードを用いた。次に、得られた全員分の結果から主観的評価に関するコンセンサスを取る。それぞれのオノマトペに対し、ポジティブだと判断した人が 4 人以上いる画像をポジティブなクラスに分類し、それ以外のもをネガティブなクラスに分類する。ここから、一定数のポジティブ数を得られた「ごわごわ」、「かさかさ」、「さらさら」、「しわしわ」、「つるつる」、「ざらざら」のデータを識別器の性能評価に用いる。

4. 識別器の性能評価と考察

本章では、サポートベクターマシンを用いて識別器に画像特徴と主観的評価データを学習させ、その識別器の性能評価を行う。

4.1 性能評価

PCA 特徴を用いた識別器と DCT 特徴を用いた識別器の性能評価を行うためにそれぞれの識別器に対し Leave-one-out 交差検証を行う。Leave-one-out 交差検証を用いた理由は、主観的評価データのコンセンサスを取ったところ、全画像数に対してポジティブなクラスに分類された画像数の割合が低かったためである。Leave-one-out 交差検証は以下の手順で行われる。

- (1) 学習データから 1 つの事例だけを抜き出す
- (2) 取り除いた事例を除く学習データをサポートベクターマシンに学習させる
- (3) 取り除いた事例を評価データとして入力する
- (4) 1~3 の手順を全事例について行う

PCA 特徴, DCT 特徴 (100 次元), DCT 特徴 (10000 次元) を用いた時の Leave-one-out 交差検証おける以下の項目の結果を表 1, 表 2, 表 3 に示す。また、識別器の Accuracy, Precision, Recall を式 (8), (9), (10) のように定義する。Accuracy は正解データと比較した時の正確にポジティブなクラスとネガティブのクラスに分類できた割合を表している。Precision は正しいと識別した中に、どれだけ正答が含まれているかを表しており、Recall は全正例のうち正しく正例と識別できた割合を表している。また、システムの信頼性を表す指標として F 値というものがあり、式 (11) のように Precision と Recall の調和平均で定義できる。Precision と Recall の両方のバランスがよく高いほど高信頼性のシステムであると言える。

	実際が正例	実際が負例
予測が正例	True Positive(TP)	False Positive(FP)
予測が負例	False Negative(FN)	True Negative(TN)

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (8)$$

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

$$F = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (11)$$

4.2 考察

まず、全体の結果から、「かさかさ」、「しわしわ」を正しくポジティブと判断した事例がないことから、再現率が 0 % であり、これらの質感を表す特徴として今回用いた PCA 特徴と DCT 特徴は不適切であったと考えられる。一方で、「さらさら」をポジティブと判断した割合が高く、識別器の信頼性を表す F 値の値も大きいことから、これらの特徴は「さらさら」を表す特徴として適切であると考えられる。また、「つるつる」、「ごわごわ」を表す特徴として PCA 特徴よりも DCT 特徴の方が適切であると言える。特に、「ごわごわ」に関する識別では、PCA 特徴の場合が全く再現できていないのに対して、DCT 特徴の方がやや信頼性があると言える。「つるつる」に関する識別では、100 次元の DCT 特徴を用いた時が最も精度がよく、人間が「つるつる」と感じる基準が低周波の周期パターンに現れていると考えられる。

オノマトベ	ポジティブ数	Accuracy(%)	Precision	Recall	F 値	TP	FP	TN	FN
ごわごわ	10/94	89.3	-	0	-	0	0	84	10
かさかさ	8/94	91.4	-	0	-	0	0	86	8
さらさら	58/94	74.4	0.76	0.86	0.81	50	16	20	8
しわしわ	8/94	91.4	-	0	-	0	0	86	8
つるつる	25/94	75.5	0.57	0.32	0.41	8	6	63	17
ざらざら	31/94	68.0	0.53	0.29	0.38	9	8	55	22

表 1 PCA 特徴

オノマトベ	ポジティブ数	Accuracy(%)	Precision	Recall	F 値	TP	FP	TN	FN
ごわごわ	10/94	90.4	0.67	0.2	0.31	2	1	83	8
かさかさ	8/94	91.4	-	0	-	0	0	86	8
さらさら	58/94	82.9	0.8	0.97	0.88	56	14	22	2
しわしわ	8/94	91.4	-	0	-	0	0	86	8
つるつる	25/94	89.3	0.89	0.68	0.77	17	2	67	8
ざらざら	31/94	71.2	0.64	0.29	0.4	9	5	58	22

表 2 DCT 特徴 (100 次元)

オノマトベ	ポジティブ数	Accuracy(%)	Precision	Recall	F 値	TP	FP	TN	FN
ごわごわ	10/94	86.1	0.33	0.3	0.32	3	6	78	7
かさかさ	8/94	86.1	1.0	0.13	0.22	1	0	86	7
さらさら	58/94	76.5	0.8	0.83	0.81	48	12	24	10
しわしわ	8/94	87.2	0	0	-	0	4	82	8
つるつる	25/94	77.6	0.6	0.48	0.53	12	8	61	13
ざらざら	31/94	69.1	0.53	0.52	0.52	16	14	49	15

表 3 DCT 特徴 (10000 次元)

5. おわりに

本論文では、デジタル画像から得られる情報から人間が質感を評価するときの基準を明らかにするために、画像特徴に基づく識別器に質感に関する主観的評価のデータを学習させ、性能評価を行った。主観的評価の学習手法としては、質感を表すオノマトベが当てはまるか当てはまらないかの2値問題に帰着させサポートベクターマシンを用いた。次に、識別器に学習させるための材質感を表すデータとして画像特徴と人間の主観的評価のデータを作成した。画像特徴では、材料ごとのパターンの類似や差異に注目した主成分分析を用いた手法と画像から周期パターンに注目した離散コサイン変換を用いた手法の2つの手法を用いた。主観的評価は、11人の評価者に全画像について10種類の質感に関する各オノマトベが当てはまるか当てはまらないかアンケートを取り、その結果のコンセンサスを取ることで回答者ごとの基準によらない平均となる主観的評価のデータを作成した。

これらのデータを用い識別器の学習と性能評価を行い、その結果から、「さらさら」などのオノマトベに対しては、今回用いた特徴量による識別機の学習に適している一方で、「かさかさ」や「しわしわ」などといったオノマトベに関しては主成分分析や離散コサイン変換で得られる特徴量では識別機を十分に学習させられなかった。

今後の課題として、今回うまくいかなかった質感を表

す特徴を検討することが挙げられる。また、Accuracyの値が高い割に True Positive の割合が低い。これは、特にデータに偏りがある場合に線形サポートベクターマシンが Accuracy 上げる方向に学習しているためである。今後は True Positive の割合が向上できるような学習手法を検討したい。それと並行して、全標本に対するポジティブクラスの要素数とネガティブクラスの要素数の偏りが Accuracy に影響を与えているため、多数の人間の主観的評価に基づいた偏りの少ない材料のデータセットを作成する必要がある。

参考文献

- [1] 大槻 正樹, 三浦 淳, 佐藤 幸男. 多方向測定による物体形状と表面反射特性の計測. 電子情報情報通信学会論文誌. Vol.J76-D2. No.8. pp.1536-1543
- [2] 下田 和, 柳井 啓司. DCNN 特徴を用いた Web からの質感画像の収集と分析. 信学技報, vol. 114, no. 409, PRMU2014-89, pp. 67-72, 2015 年 1 月.
- [3] Perronnin, Florent, Jorge Sanchez, and Thomas Mensink. Improving the fisher kernel for large-scale image classification. Computer Vision ECCV 2010. Springer Berlin Heidelberg, 2010. 143-156.
- [4] Donahue, Jeff, et al. A deep convolutional activation feature for generic visual recognition. arXiv preprint arXiv:1310.1531 (2013).
- [5] 津田 宏治. サポートベクターマシンとは何か. 電子情報通信学会誌. 2000.
- [6] 後藤 富郎. 直交変換によるデジタル画像の高圧縮・高精細化に関する研究. 2007.