

10 画像認識 1 ～文字の認識～

10.1 画像認識

画像の処理や認識は、製品検査、リモートセンシング、個人認証など、さまざまな場面で利用されています。例えば、地図や図面の読み取りは、かつて、盛んに研究されていた課題の1つです。製品検査も、画像認識の応用例の1つとして実用化されています。例えば、プリント基板の検査や、製品の外観検査、農産物の選別などが例として挙げられます。リモートセンシングも画像認識の応用例の1つです。リモートセンシングは、航空機や人工衛星などから地表の様子を撮影し、植生や土地利用、気象などの情報を測定するものです。測定には、可視光や赤外線、マイクロ波などが利用されます。医用画像の認識も画像認識の応用例として研究が進められています。例えば、結核の場合、胸部X線画像に丸い影が生じる場合があります。このため、X線画像から丸い領域を検出することで、結核の疑いがある場所を抽出する手法が提案されています。また、マンモグラフィー画像から病変の疑いがある場所を抽出する手法についても実用化が進められています。画像認識は、個人認証にも利用されます。例えば、指紋は人により異なることから従来から個人の識別に利用されています。PCのログインなどでも利用されるようになっていきます。顔画像の認識により個人認証を行うシステムも研究が進められており、実用化により、いわゆる顔パスを実現することができるようになります。

10.2 文字認識

ここで、画像認識の例として、文字認識について詳しく見てみましょう。画像認識を行うには、まず、画像をコンピュータに取り込む必要があります。例えば、紙に書かれた文字を読み込む場合、スキャナを利用することができます。スキャナでは、例えば、細長い光センサを用意し、それをスライドさせることで、2次元の画像を取り込むことができます。画像入力には、デジタルカメラなどを使うこともできます。読み込んだ画像の文字の部分を拡大すると、明るさが異なる点の集まりであることが分かります。このようなデータから、どのように文字を読み取るか考えていきます。

もし、文字が活字で印刷されており、大きさも分かっているような場合は、物体検出などで広く利用されるテンプレートマッチングを文字認識に利用することもできます。具体的には、あらかじめ認識しようとする文字の画像を用意しておきます。そして、その画像を、取り込んだ画像上で移動させながら最も一致する場所を探します。

より一般的に文字認識を行う場合、1文字1文字を切り抜いたり、文字の大きさをそろえるといった処理が必要になります。このような処理は、前処理と呼ばれます。例えば、スキャナで、図10-1のような「7,4,1,9」と書いてある画像を読み込んだとします。まず、行ごとに切り出すことを考えてみましょう。この場合、上から順に横方向の黒画素の数を数えてやります。文字が存在する部分では、黒画素の数は多くなっています。行と行の間の部分には、文字が存在しないため、黒画素の数は少なくなっています。一定間隔で文字が印刷されていれば、文字が存在する黒画素が多い領域の縦幅や、行と行の間の黒画素が少ない領域の縦幅は、ほぼ一定になるはずです。これにより、黒画素が多く、縦幅もほぼ一定の領域を抽出することで、1行分の文字を抽出することができます。

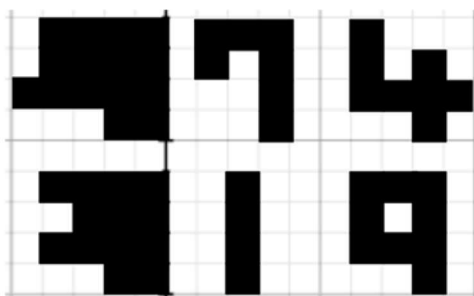


図10-1 文字の切り出し

1文字ごとの切り抜きも同様に行うことができます。先ほどの画像から1行分を抽出し、今度は、縦方向の黒画素の数を数えてやります。すると、文字が存在する部分では、縦方向の黒画素の数は多く、文字と文字の間の部分には、縦方向の黒画素の数は少なくなります。一定間隔で文字が印刷されているとすれば、文字の幅はほぼ一定で、文字と文字の間の幅もほぼ一定になるはずです。これにより、1文字ごとの切り抜きを行うことができます。

一方、郵便番号の場合は、通常、赤い枠の中に郵便番号を書き込みます。このため、カメラで撮影した画像から、赤い四角形の領域を抽出することで、文字の切り出しを行うことができます。ただし、1文字1文字を抽出できたとしても、人によって、文字を上側に書いたり、大きく書いたり、薄く書いたりといったこともあり得ます。このため、これらを一定にそろえる正規化と呼ばれる処理が行われます。

次に特徴量の抽出が行われます。特徴量とは、それぞれの画像を特徴づける数値データのことです。特徴量としては、さまざまなものが提案されており、どういった画像を対象とするか、どのような目的に利用するか等によって使い分ける必要があります。ここでは、1文字ごとに切り出された2値画像として、いくつかの特徴量について見てみます。

図10-2は、ゾンデ法という手法を説明するものです。ゾンデ法は、主に、英数字の認識に用いられます。まず、基準となる点を決めます。この図では、2つの青い点が基準点になります。そして、基準となる点から線を伸ばしていき、文字領域と何回交差するかを数えることで、特徴量の抽出を行います。例えば、まず、上側の点を起点として、上方向に直線を伸ばしていきます。この直線は文字と1度交わるので交差した回数は、1になります。右側の線は1度交わるので1になります。左側の線は1度交わるので1になります。下側の線は1度交わるので1となります。同様に、下側の点を起点として、右側の線は1度も交わらないので0、左側の線は1度交わるので1、下側の線は1度も交わらないので0となります。これらの数値を並べることで、(1,1,1,1,0,1,0)のような特徴量を抽出することができます。文字の画像から7次元の数値データを抽出できたわけです。

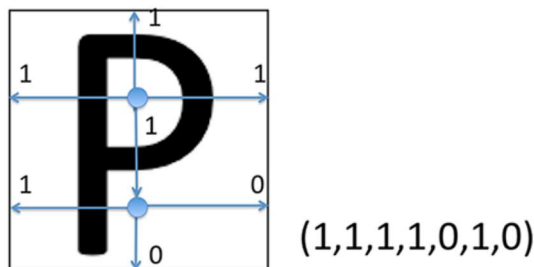


図10-2 ゾンデ法

射影特徴量も、文字の特徴量として利用することができます。文字の切り出しの際、縦方向、横方向の黒画素の数を数えましたが、この方向ごとの黒画素数を文字の特徴量として利用します。また、斜め45度方向の黒画素数を利用する場合もあります。

別の方法として細線化による方法があります。文字は、スキャナ等で読み込んだ場合、黒画素部分が幅を持っています。細線化は、幅1画素の線へと変換する処理です。細線化を行うには、収縮処理のように黒画素部分をどんどん削っていきます。しかし、幅が1になった場合は、それ以上、黒画素部分を削らないようにします。細線化した画像からは、端点、分岐点といった特徴点を抽出することができます。これらの数は、文字を識別するための特徴量として利用できます。

ここまでは、文字画像から、さまざまな特徴量の抽出を行ってきました。では、これらの特徴量を使って、どのように文字の認識を行うのでしょうか。ここからは少し、認識について抽象化して考えてみましょう。もし、抽出された特徴量がN個あったとすれば、それらの数値を並べたN次元のベクトルによって、それぞれの画像は表現することができます。画像認識の最終段階では、このN次元ベクトルが、どの文字に対応するか判断することになります。このためには、この文字は、このように表現できるといった認識モデルを用意し、入力された文字がどのグループに属するかの分類が行われます。

例えば、入力された文字が1であるか2であるか分類するという簡単な例を考えてみます。まず、たくさんの1と2の文字を集めます。そして、それらから、特徴量の抽出を行います。100個の特徴量を抽出すれば100次元のベクトル、1000個の特徴量を抽出すれば1000次元のベクトルになります。図10-3では、これらのベクトルを2次元で表現しています。1と書いているのは、たくさんの1から特徴量の抽出を行い、その位置を示したものです。同じ1という文字であったとしても、抽出される特徴量は必ずしも一致するとは限らないため、1に対応する点は複数存在します。2と書いているのも同様に、たくさんの2から特徴量の抽出を行い、その位置を示したものです。2に対応する点も複数存在します。Xと書いているのは、入力された文字です。入力された文字からも特徴量の抽出を行い、その位置にXと書いています。これからやろうとしていることは、このXが1なのか2なのか判断することです。

このようなデータを分類するための手法としては、さまざまな手法があります。最も簡単な方法の1つがk近傍法です。k近傍法は、多数決により、どのグループに属するかの判断が行われます。k近傍法では、まず、入力された点から近い順にk個のデータを選びます。そして、その多数決により、どのグループに属するかを決めます。最も簡単な場合としてk=1の場合を考えてみましょう。これは、最も近い点が属するグループに、入力されたデータが属するとします。そこで、この図で、Xに最も近い点を探してみます。すると、それは、1になります。よって、Xは、1という文字だと判断されます。ただし、k近傍法の場合、近い順にいくつ選ぶかで、結果が変わってくる場合があります。例えば、k=3の場合を考えてみましょう。Xから、近い順に3つのデータを選びます。そして、多数決によりどのグループに属するかを決めます。この図の例では、一番近いのは1ですが、2番目に近いのは2、3番目に近いのは2になります。このため、多数決をとると、Xは、2という文字だと判断されます。このように、kの値を変えることで、分類結果が変わる場合がありますが、k近傍法は、データの分類を行うための最も簡単な方法の1つとして利用することができます。

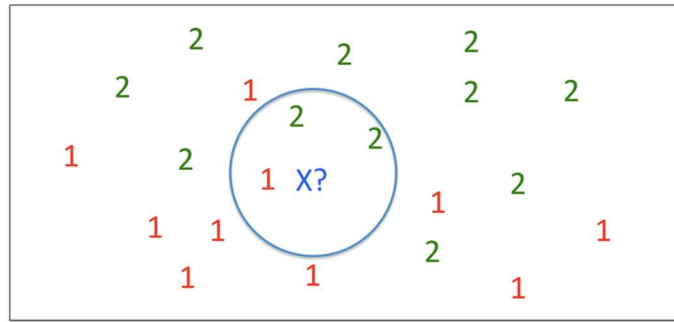


図10-3 k近傍法

また、データの分類には、線形判別分析といった方法も利用できます。図10-4は、線形判別分析の概要を示しています。線形判別分析では、データを直線（一般には平面）で分割し、入力されたデータが、そのどちら側に属するかで、データの分類を行います。この図では、直線が1本引かれていますが、これが最も1と2を分離できる直線にあたります。そして、入力された画像Xが、直線の上にあるか下にあるかで、データの分類を行います。この場合は、直線の下側にXがあります。これは、1が属する側ですから、入力された文字Xは、1であると判断されます。

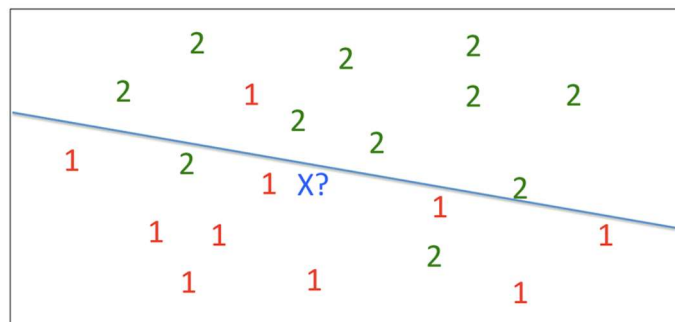


図10-4 線形判別分析

もう少し複雑な場合についても考えてみましょう。1のデータが2のデータに囲まれている場合を考えます。この場合、直線や平面でデータを分割することができません。このような場合に利用されるのがサポートベクトルマシンです。サポートベクトルマシンでは、カーネル関数と呼ばれる関数により、データをより高次元のデータへと変換してやりま。その後、データをうまく分けることができる平面を求め、データの分類を行います。例えば、真ん中を盛り上げるような変換をしてみましょう。これにより、もともと2次元だったデータが、1次元増えて3次元のデータになります。そこで、盛り上がった部分をうまく切るような平面を求めてやれば、1と2をうまく分離できるようになるわけです。

以上述べた手法以外にも、ベイズ推定を利用する方法やニューラルネットワークなど、データの分類には、さまざまな手法を利用することができます。