マルチメディア情報処理

チェックポイント１

・視神経の桿体と錐体の役割と特徴

桿体　色よりも明暗に対しての感度が強い　暗い環境に適応する

錐体　明るい環境下で作動し、色と明るさを判別　三原色によって色を知覚する

・錐体の感度と視覚の関係

錐体には青に反応するS錐体、緑に反応するM錐体、黄緑に反応するL錐体がある

色覚とはこれら３つによってもたらされる色の感じ方

C型　＝　すべての錐体がそろっている



P型　＝　L錐体がない、またはずれてM錐体に近い



D型　＝　錐体がない、またはずれてL錐体に近い

T型　＝　S錐体がない



A型　＝　すべての錐体がないか、錐体が１つしかない



・画像の量子化のビット数について



輝度やRGB値は８ビット（高精細画像では超える場合あり）



フルカラーはRGBそれぞれについて８ビットより、２４ビット＝３バイト



例：FHDの１０秒間（３００フレーム）のフルカラー映像の情報量



＝1920×1080　×　3　×　300 　　=　 1.87ギガバイト



　　　　（画素数）　（バイト）（フレーム）



ギガバイト＝1024×1024×1024バイト



・各種色空間について

RGB色空間　加法混合の色の表現法（各値最大で白）

YCbCr色空間　輝度と色差で表現



CMY色空間　減法混合の色の表現法（各値最大で黒）

HSV/HSB/HSL色空間　色相、彩度、明度の軸で表現

・YCbCr色空間を利用した情報圧縮の原理と各形式

輝度に対する感度は高く、色に対して低い

色差成分を空間的に間引きして解像度を落とし、画質を維持する

間引きの方法に応じて形式が異なる

非圧縮の4:4:4形式

2/3に圧縮する4:2:2形式

1/2に圧縮する4:2:0形式

・変換符号化の原理

画像データを効率的に量子化できる値に変換し、符号化する



・変換符号化で代表的に利用される変換方法



離散コサイン変換（DCT）



・予測符号化の原理

ある画素の値を、すでに伝送済み（入力済み）の画素値から予測し、その画素の実際の値との差分を符号化する方法。



・DCPMについて

予測に直前の値（隣接する画素の値）を使用する方法

・離散コサイン変換（DCT）の結果（DCT係数）の解釈について

入力画像にどれくらい細やかな成分がどれくらい入っているか

・JPEGの３方式

基本方式（DCTによる変換符号化）

拡張方式

DPCM方式（DPCMによる予測符号化）

・JPEGの基本方式ではどのような方法で圧縮を実現しているか

「DCTによる変換符号化」と、「高周波成分を粗くする量子化」で圧縮する

直流成分の符号化：各ブロックの直流成分はまとめて縮小画像を作成し、DPCM符号化

交流成分の符号化：「ジグザグスキャン」を行い一次元化

・JPEGのDPCM方式ではどのような方式で圧縮を実現しているか

「DPCMによる予測符号化で圧縮」

隣接画素のみを利用した予測式を利用

差分値をエントロピー符号化

チェックポイント２

・JPEGの基本方式とDPCM方式の、画質、圧縮率の比較

基本　　情報に欠損が生じるので画質は劣化するが、圧縮率は高い（不可逆符号化）

DPCM　情報の欠損がなく圧縮できるが、圧縮率は低い（可逆符号化）

・平滑化に利用されるフィルタの種類とその特徴について

平均化フィルタ：付近の画素の平均値に変換する

加重平均フィルタ：重みを２次元ガウス分布に基づいて与えたものをガウシアンフィルタ。付近画素値の「重み付き平均」に変換する

平均化 ガウシアン

1/9 1/9 1/9　　1/16 2/16 1/16

1/9 1/9 1/9 2/16 4/16 2/16

1/9 1/9 1/9 1/16 2/16 1/16

・平滑化の目的

ノイズの除去などに用いられ、画像の濃淡の変化を滑らかにする

・エッジ抽出に利用されるフィルタの種類とその特徴

１次空間微分フィルタ：プレウィットフィルタ、ソーベルフィルタ

２次空間微分フィルタ：ラプラシアンフィルタ

プレウィットフィルタ　　(1次微分値を求めるフィルタ)

水平　　　　　　垂直

-1 0 1 -1 -1 -1

-1 0 1 0 0 0

-1 0 1 1 1 1

ソーベルフィルタ　　　(注目画素に近い画素の重みを大きくしたフィルタ)

水平　　　　　　垂直

-1 0 1 -1 -2 -1

-1 0 2 0 0 0

-1 0 1 1 2 1

ラプラシアンフィルタ(2次微分値を求めるフィルタ → 画像中の輪郭が強調される)

４近傍 ８近傍

0 1 0　　　　 　1 1 1

1 -4 1　　　　 　1 -8 1

0 1 0　　　　　　1 1 1

・ハフ変換の目的

エッジ抽出画像などの入力画像から、直線や円などの形状要素を抽出

・ハフ変換による直線成分抽出の原理

C = -x \* m + y によって（x,y）画素をmc平面で直線の式にする。

直線の最も交わる座標の（m,c）を指揮に代入し、xy平面での直線の式にする

・ハフ変換の利点欠点

利点：画像中の雑音に比較的頑健

　　　隠ぺいなどによって直線が不完全になっていても検出可能

　　　複数の直線要素を同時に検出することも可能

欠点：計算量が多い

　　　パラメータ空間を表現するために比較的大きなメモリ空間が必要

　　　直線検出後に、その直線から必要な部分を抽出する方法が不明確

・ラベリング処理は何を行う手法？

2値化画像について「領域を分割する手法」（点在している領域の個数や、面積や重心などを求めるのに用いる）

・4連結ラベリング、８連結ラベリングのアルゴリズム

４連結は連結の判定が、垂直水平の４つ

８連結は連結の判定が、周囲８つすべて

１：左上から走査し、ラベルを付与。

２：同じ輝度値の画素が連結していれば同じラベルを付与

３：連結が見当たらなくなれば再度走査、ラベルを付与しなおす

・背景差分法の処理と目的

目的：固定されたカメラで撮影された動画から。移動物体を検出するため

処理：事前に登録された背景画像と現時点の入力フレーム画像との差を計算

・オプティカルフローとは

画像中のある点の動きを「速度ベクトル」としてあらわしたもの

・オプティカルフローの抽出の問題点、またその一部を解決する手法

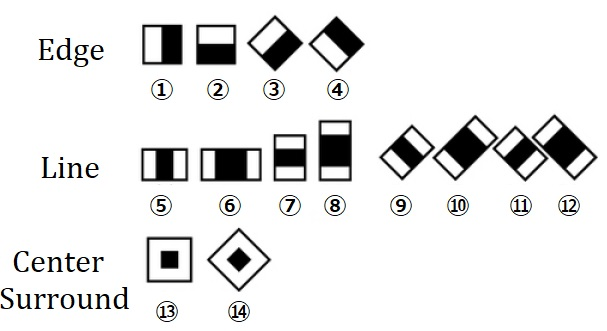
運動の速度が大きな画像の解析には不向き

濃度勾配の小さな領域ではノイズに弱い

照明が時間、空間的に変動する環境では誤検出が起きやすい

解決法：ノイズなどの悪影響などは平滑化によって軽減可能

チェックポイント３

* 特定物体認識と一般物体認識を説明できる？
  + 特定物体認識：特定の既知の物体が画像中に出現しているか否かを判断する技術
  + 一般物体認識：画像中に出現する物体を一般的な名称で認識する技術
* 特定物体認識で用いられる「Haar-like特徴量」って何？
  + いくつかの短径領域のパターンに当てはめ、黒領域と白領域の平均輝度の差を特徴量として計測することで、画素値をそのまま用いる場合よりも照明やノイズの影響を受けづらくした特徴量
  + 顔検出などに用いられる(ex.目玉は暗い、目元は明るい 等の共通情報から顔検出)
* Haar-like特徴量を用いた特定物体認識で「カスケード型の分類器(識別器)」が利用されることがあるけれど、何故？
  + カスケード型の分類器：複数の識別器を連結させた識別器。「第一の識別器で”正解”と判定された場合第二の識別器へ」を繰り返し、途中で”非正解”と判断されれば以降の処理を破棄し識別を終了する。前段階ほど識別の判定を緩く(誤検出率高)し、早い段階で”非正解”と判断された場合以降に用意された高度な識別(処理)を破棄することができるため、高速化、処理の簡略化を図れる。
  + 一般的に人物を撮影した画像には顔以外の領域が多く含まれるため、画像全体を全てhaar-like特徴量を用いた識別機で検出する場合あまりにも時間がかかり非効率的である。そのためカスケード型の分類器を用いることで、早い段階で顔ではないと判断された領域はそれ以降の処理を破棄することができる(全ての段階で顔であると判定された場合のみ顔と判断される)ため、検出の高速化を図れるから。
* 一般物体認識で用いられる「Bag-of-keypoints」って何？
  + 画像中に存在するキーポイントをヒストグラム表現したもの
  + 位置情報を利用しないことが特徴、どのような特徴のキーポイントがどの程度含まれているかをヒストグラムにする。

※キーポイント：輝度変化の特徴的な点

* 「Bag-of-keypointsによる認識」の流れを説明できる？

1. 画像の様々な点(キーポイント)から大量の「局所特徴量｣(ベクトル)を抽出
2. 各特徴量を事前に作成しておいた「コードブック」を用いて記号化(量子化)
3. 各画像をどの記号がどの程度含まれているかを表す「ヒストグラム」に変換

※ヒストグラムが似ているかどうかを「識別器」で判断することで対象とする物体が存在しているかどうかを判断する。

* 物体認識で利用される「局所特徴量」って何？
  + キーポイント上で計算したエッジ勾配のヒストグラムを表した記述子
  + キーポイント上の範囲の輝度値を用いて、もっとも優勢な輝度の勾配方向(オリエンテーション)を計算、決定した後、オリエンテーションを上方向に固定して細かい勾配情報を数値化したもの
* 局所特徴量SIFTは、どんな特徴をもっている？(この特徴量を用いた画像認識手法はどんな利点をもっている？)
  + 対象の大きさや傾きが変化しても対応が取れる
* 「Bag-of-keypointsによる認識」で利用される分類器(識別器)にはどんなものがある？
  + 識別モデル(SVM（Support Vector Machine）)：「ライオン」と「それ以外」の学習画像を用いて、「対象物かそうでないか」識別する分類器を作成。入力画像を様々な種類の分類器にかけて複数の結果を統合し、最終結果とする。
  + 生成モデル：「猫」「ライオン」などのそれぞれの対象物を個別に確率モデル化し、入力画像がどのモデルと一致するか計算し、確率が最大となるものを認識結果とする
* 現在、主流になりつつある画像認識の手法は？
  + 深層学習(統計的機械学習)を用いた一般物体認識
* 「深層学習」って何？
  + ニューラルネットワークを用いた人工知能構築技術
  + ニューラルネットワーク：人間の脳神経系を模した数学モデル
* 「深層学習」を利用すると何で高性能な画像認識を実現できるの？
  + 特徴抽出や分類(識別)などに用いるデータを人間が全て入力するのではなく、大量の画像(=実例)を読み込み、統計的に学んだ人工知能が判断する。そのため、一つの項目に対する情報(判定基準)を人間の手入力よりも早く多く確保でき、より正確な画像認識を実現できる。