

情報工学序説 コンピュータによる自動認識 (3) : 画像と情報

1. 画像と音声

音は1次元の信号, 画像は2次元の広がりをもつ信号.
違いはあるが, 信号処理の立場から見ると, 多くの性質を共有している.

2. 画像のデジタル表現 . アナログ信号 デジタル信号 (AD変換)

- CCD カメラ (Charge Coupled Device, 光に反応する微少な素子)
- RGB カラー画像 :
赤 : レッド (R), 緑 : グリーン (G), 青 : ブルー (B) の三色の光の点 .
すべての色は RGB の明るさを調整することによって作ることができる .
- ピクセル (画素) : 画像の最小単位 . RGB の三つの点をまとめたも .
- グレースケール画像 : 色に関する情報を含まない画像 .
以下では説明を簡単にするため, グレースケール画像を考える .

3. 画像の数学的表現 : 【線形代数】【応用数学2】【マルチメディア情報処理】

- 標本化 (サンプリング) と量子化

- 画像 $f(x, y)$, $x = 0, 1, 2, \dots, 7$, $y = 0, 1, 2, \dots, 7$ を 64 個の基底ベクトル $e_k(x, y)$ の重ね合わせで表現 (7 や 64 という数字は一例) :

$$f(x, y) = \sum_{k=1}^{64} a_k e_k(x, y) = \sum_{k=1}^{64} b_k e'_k(x, y)$$
$$x = 0, 1, 2, \dots, 7, \quad y = 0, 1, 2, \dots, 7$$

画像 $f(x, y)$ は 64 次元ベクトル $\mathbf{a} = (a_1, a_2, \dots, a_{64})$ で指定できる .
 $f(x, y)$ の表現の仕方は, 基底ベクトル $e_k(x, y)$ の選び方に依存して無数にある .

- (例) 各ピクセルごとに表現
- (例) ウォルシュ基底

- (例) 周波数領域での表現 (周波数スペクトル): 正弦波の重ね合わせによる表現
フーリエ級数, フーリエ変換
空間周波数: 空間的に変化する信号の周波数

4. 動画: 連続した静止画像を高速に映し出している. 映画 24 枚/1 秒. テレビ 30 枚/1 秒.

5. 情報圧縮と誤り訂正 【確率論と情報理論】【マルチメディア情報処理】

情報源圧縮: データの量をいかに減らすか.

誤り訂正方式: 通信の際に混入する雑音をどう減らすか.

例: 画素数: 720×480 pixel. 階調: 24 bit (RGB 各 8 bit). フレーム数: 30/秒.

1 秒あたりのデータ量: $720 \times 480 \times 24 \times 30 = 248,832,000$ bit = 31,104,000 byte.

DVD-ROM: 一枚で約 4.7 Gb (giga byte) のデータを保存できる.

圧縮しないで動画画像だけを保存した場合, $4700 \text{ Mb} / 31.104 \text{ Mb/sec} = \text{約 } 151 \text{ 秒}$.

MPEG: 動画画像の符合化 (圧縮) アルゴリズムの規格

6. 物体認識, コンピュータビジョン 【パターン認識】

例: 画像 $a = (a_1, a_2, \dots, a_n)$ を観測し, そこにある人の顔, 人物, 感情, など y を求める. 人間には比較的簡単な問題ではあるが...

7. Compositionality: 絵画, CG 【コンピュータグラフィックス】...