

シーン認識の周波数依存に関する実験的研究

坂本真人¹⁾・細川弘江²⁾・友添直子²⁾Experimental Study for Frequency-Dependence
of Scene RecognitionMakoto SAKAMOTO¹⁾, Hiroe HOSOKAWA²⁾, and Naoko TOMOZOE²⁾

Abstract

In general, we can recognize various complex scenes as fast as isolated objects. How can we achieve this efficiency, in spite of the large number of objects and interfering factors in scene? In 1994, P.G.Schyns and A.Oliva published their paper against this open problem, and showed whether the respective roles of coarse blobs and fine boundary edges in scene recognition at a glance were used in a particular sequence over the course of fast and leisurely scene processing [5]. In this paper, we demonstrate their statement, and show that in very fast recognition tasks, scenes are identified by high-frequency and low-frequency components.

Key Word : FFT, frequency component, picture processing, psychology, scene recognition

1 はじめに

ディスプレイに表示される非常に認識タスクの多いシーンを、単体の物体を認識するのと同じくらいの速さで私達は識別する。そのシーンの中には影を閉じたり、閉塞したりといった構成要素を干渉する要因がかなり多く存在するのにも関わらず、どのようにしてこの効率的な動きを達成するのか。シーンカテゴリは、それらの主要構成要素の空間組織を異なるものにした、典型的なものにする。もし、人間の知覚的な構造が処理の段階で、この情報を速く引用し、調整するとしたら、*CtF*((Coarse-to-Fine)処理:低周波成分のぼんやりとしたパターンから入り、次第に細かな高周波成分を認識していく処理)がシーン認識の結果を生じる原因を説明することができる[7,8]。*P.G.Schyns*と*A.Oliva*は粗密な情報がシーン認識の処理にどのような役割を担うかを研究し、1994年にその結果を発表した。つまり、例えば低周波であるぼんやりした線をまず入力し、視覚システムは記憶にあるシーンスキーマを促進させながら判断することで高周波のはっきりとした線を習得していく。画像の密な情報を処理する段階であいまいな情報を改良して認識していくのである。本稿では、自然なシーンを素早く同一化する事において、周波数領域における低周波と高周波の情報それぞれの役割を対比する実験的研究を行い、彼らの研究成果を実証した[5,6]。実験で使用するプログラムはC言語で作成し、PC上で動作させた[1-3]。

2 周波数領域における画像処理の原理

画像を与えた時、それが単位長さあたりの正弦波状の濃淡変化の繰り返し回数、つまり空間周波数がどのような成分から構成されているかを求めると、原画像をながめただけではわからない情報を得ることができ。ここでは、本研究を進めるにあたりプログラムの主要な部分になる画像のフーリエ変換についてまとめる[1,2,4]。

2.1 高速フーリエ変換

画像を周波数で処理する場合は、まず画像を周波数領域に変換する必要がある。この変換を担うのが、フーリエ変換である。周波数領域への変換とは「任意の波形は、単純な正弦波の和で表現できる」ということが基礎となっており、異なる周波数を持った正弦波に分解できる。フーリエ変換とは直交変換の仲間で、振幅 A と位相 ϕ を一度に表すために複素数表示を使う。フーリエ変換を複素関数 $X(f)$ を使って表すと、

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt$$

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft} dt$$

である。

コンピュータを使って画像処理する為に行うフーリエ変換では、扱う信号が標準化されたデジタル信号である事と、数学上では無限大という考えが通用するが、コンピュータでは有限回の計算で終わらなければならないという制約を与えたものを「離散的フーリエ変換(DFT: Discrete Fourier Transform)」と呼ぶ。

任意の関数 $f(x)$ は、互いに直交する正規直交関数系の1次結合により表現することができる。 x が0から

1) 情報システム工学科助教授
2) 情報システム工学科学部生

$N-1$ の離散整数値をとる関数 $f(x)$ の DFT および離散的フーリエ逆変換 (*Inverse DFT : IDFT*) はそれぞれ式 (2.1), (2.2) のように表される。ただし j は虚数単位であり、 $j^2 = -1$ である。

$$F(k) = \sum_{x=0}^{N-1} f(x) e^{-j \frac{2\pi}{N} kx} \\ = \sum_{x=0}^{N-1} f(x) \left(\cos \frac{2\pi}{N} kx - j \sin \frac{2\pi}{N} kx \right) \quad (2.1)$$

$$f(x) = \frac{1}{N} \sum_{k=0}^{N-1} F(k) e^{j \frac{2\pi}{N} kx} \\ = \frac{1}{N} \sum_{k=0}^{N-1} F(k) \left(\cos \frac{2\pi}{N} kx + j \sin \frac{2\pi}{N} kx \right) \quad (2.2)$$

これらの式からわかるように、任意の関数 $f(x)$ は、基本周波数成分とその高調波の和によって表現することができる。式 (2.1), (2.2) で $k=0$ を直流成分、 $k=1$ を基本周波数成分、 $k>1$ を高調波成分という。元の関数 $f(x)$ に含まれるいくつかの周期の波の強さに応じて、直流成分、基本周波数成分、高調波成分の大きさが決定される。

元の関数が画像のような2次元関数 $f(x, y)$ に対する DFT 、逆 DFT は、それぞれ式 (2.3), (2.4) のように表される。ただし、 x, y の取り得る範囲をそれぞれ $[0, N_x-1]$ 、 $[0, N_y-1]$ とする。

$$F(u, v) = \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} f(x, y) e^{-j \frac{2\pi}{N} (ux+vy)} \\ = \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} f(x, y) \left(\cos \frac{2\pi}{N} (ux+vy) - j \sin \frac{2\pi}{N} (ux+vy) \right) \quad (2.3)$$

$$f(x, y) = \frac{1}{N_x} \frac{1}{N_y} \sum_{u=0}^{N_x-1} \sum_{v=0}^{N_y-1} F(u, v) e^{j \frac{2\pi}{N} (ux+vy)} \\ = \frac{1}{N_x} \frac{1}{N_y} \sum_{u=0}^{N_x-1} \sum_{v=0}^{N_y-1} F(u, v) \left(\cos \frac{2\pi}{N} (ux+vy) + j \sin \frac{2\pi}{N} (ux+vy) \right) \quad (2.4)$$

式 (2.1) に従い1次元 DFT を計算する場合、多数回の計算、特に掛け算が必要である。これに対し、 DFT の計算を高速に行う高速フーリエ変換 (*Fast Fourier Transform : FFT*) 及び高速フーリエ逆変換 (*Inverse Fast Fourier Transform : IFFT*) が提案されている。 FFT には、時間間引きの方法と、周波数間引きの方法

があるが、ここでは時間間引きによる FFT について述べる。式 (2.1) は、 $W = e^{-j \frac{2\pi}{N}}$ と表記すると次式のように書くことができる。

$$F(k) = \sum_{x=0}^{N-1} f(x) e^{-j \frac{2\pi}{N} kx} = \sum_{x=0}^{N-1} f(x) W^{kx} \quad (2.5)$$

時間間引きによる FFT では、元の関数 $f()$ の要素を並べ替える必要がある。この並べ替えでは、データ列の奇数個目 (1番目, 3番目, ...) を前半、偶数個目 (2番目, 4番目, ...) を後半に並べ替える処理を順次行う。

1次元 $IDFT$ も、 FFT と全く同様な考え方で高速化することができ、次の手順に従い FFT を実行することで、 $F(k)$ から $f(x)$ を得ることができる。

1次元 $IDFT$ の計算手順：

1. $F(k)$ から $\frac{1}{N} F^*(k)$ を求める。
(ただし $F^*(k)$ は $F(k)$ の複素共役)
2. $\frac{1}{N} F^*(k)$ に対して前述の FFT と同一の方法を適用する。
3. 得られた $f^*(x)$ の複素共役をとって $f(x)$ を求める。

一方、式 (2.3) の2次元 DFT は、 x, y 方向のデータの定義域を共に $[0, N-1]$ (ただし N は2のべき乗) とすると、1次元 FFT の場合と全く同様に、 $W = e^{-j \frac{2\pi}{N}}$ を用いて次式のように書くことができる。

$$F(u, v) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) e^{-j \frac{2\pi}{N} (ux+vy)} \\ = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) W^{(ux+vy)} \quad (2.6)$$

式 (2.6) を y について展開すると次式のようにになる。

$$F(u, v) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) W^{(ux+vy)} \\ = W^0 \sum_{x=0}^{N-1} f(x, 0) W^{ux} + W^v \sum_{x=0}^{N-1} f(x, 1) W^{ux} \\ + \dots + W^{(N-1)v} \sum_{x=0}^{N-1} f(x, N-1) W^{ux} \\ = W^0 (DFT_0) + W^1 (DFT_1) + W^2 (DFT_2) \\ + \dots + W^{(N-1)v} (DFT_{N-1}) \quad (2.7)$$

ただし、 (DFT_i) は y の第 i 番目のラインにおける1次元 DFT である。

このため、関数 $f(x, y)$ に対する2次元 FFT は y が一定の各ラインにおける1次元 FFT の結果を用いて計

算することができる。 $N \times N$ の2次元データに対して2次元 DFT を用いると計算量は $O(N^4)$ 必要であるが、2次元 FFT を用いることにより計算量は $O(2N^2 \log_2 N)$ になり大幅に削減される [2,4]。

2.2 フィルタ処理

画像を周波数領域に変換させたところでフィルタ処理を行う。フィルタとは、あるものは通過させ、あるものは阻止する役目を持つものをいい、周波数領域におけるフィルタは、「ある周波数は通過させ、ある周波数は阻止する」という一元的なものとなる。

N 個の1次元データ列 (実数部: $data[0] \sim [N-1]$, 虚数部: $jdata[0] \sim [N-1]$) に対する FFT の結果として得られたフーリエ係数 (実数部: $data[0] \sim [N-1]$, 虚数部: $jdata[0] \sim [N-1]$) は、元のデータの周波数成分を表しており、添え字 = 0: 直流成分、添え字 = 1 と $N-1$: 基本周波数成分、添え字 = 2 と $N-2$: 基本周波数の2倍の高調波成分... 添え字 = $(N/2)$: 基本周波数の $(N/2)$ 倍の高調波成分という意味を持つ。

このため、これらの係数の値を変更する処理、例えば、高周波成分に相当する係数を0にする処理 (低域通過フィルタ) などを施した後、逆 FFT を実行することによって周波数領域におけるフィルタリングを行うことができる。周波数領域におけるフィルタリングでは、特定の周波数成分の強調や除去など、データ空間領域のフィルタリングでは実行できない処理を行うことができる。

2次元 FFT の結果得られたフーリエ係数 $F(u, v)$ は、 $F(0,0)$ が直流成分であり、図1のように分割・再結合することにより、中心が直流成分で中心から離れるに従い高周波成分になる分布が得られる。このため、同図のように分布の中心付近だけを通すフィルタは低域通過フィルタになる [2]。

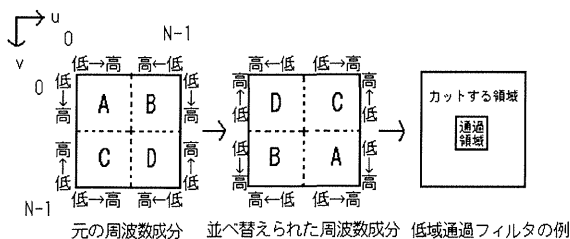


図 1: 2次元周波数成分と低域通過フィルタの例。

3 実験

ここでは、*P.G.Schyns* らが行った研究を実証するために行った2つの実験について述べる。1つ目は、 CtF 処理、つまり粗野なパターンを認識してから次第に細かい部分に気づいていくようにしてシーン認識を行って

いるのかを確かめる実験である。もう1つは、動画のように微小時間でシーンを切り替えても同様の結果を得るか否かの実験である。

3.1 環境

- 使用機器 : *PC(DELL computer)*
- 使用 OS : *Windows 2000*
- プログラミング言語 : *C*
- 被験者 : 20人 (成人)
- 画像 : 4シーンの各低周波画像, 高周波画像, 合成画 (ある画像の低周波と違う画像の高周波を組み合わせたもの)

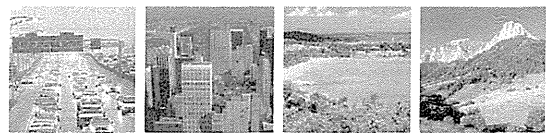


図 2: 道路, ビル街, 海, 山のノーマルシーン [6].

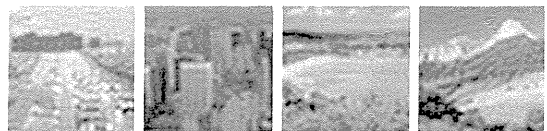


図 3: 道路, ビル街, 海, 山の LF シーン.

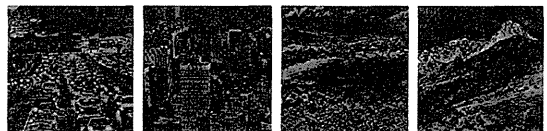


図 4: 道路, ビル街, 海, 山の HF シーン.

3.2 実験 1

画像を2つの時間で表示することにより、 CtF が異なる段階においてそれぞれの情報の特有な使用法を示すことで、解釈の違いが現れる。

3.2.1 予想

まず頭の中でおおざっぱな形を描いて、段々細かいパーツを描いていくように、低周波で形を創造し、細かい高周波の情報を認識していくと考える。

3.2.2 方法

サンプルをコンピュータ画面に 30 ms と 150 ms の速さで表示する。被験者はサンプルが目的のシーンと合うかどうか、知らせる。画像のシーンは道路、山、海、ビル街の 4 つで 256 グレーレベルの絵である。この 4 つのシーンは、全般的なコントラストが類似しているという制約により選んだ。サンプルはノーマルシーン、低周波シーン、高周波シーン、合成シーンのどれかである。合成シーンは目的シーンの低周波か高周波のどちらかを組み合わせる。目的シーンは常にノーマルシーンである。ランダムに一人 5 回の計 200 回試行し、サンプルと目的シーンが一致したものと、そうでないものに統計で分ける。

3.2.3 結果

統計		Sample				N
		LF-hybrid	HF-hybrid	LF	HF	
0.03秒	正解率	66.70%	90.00%	85.00%	80.00%	100%
	誤解答率	33.30%	10.00%	15.00%	20.00%	0%
0.15秒	正解率	100%	100%	90.00%	40.00%	100%
	誤解答率	0%	0%	10.00%	60.00%	0%

表 1: 実験 1 の結果.

情報が *CtF* を行うことを確立するために、*LF* (Low Frequency: 低周波数) と *HF* (High Frequency: 高周波数) の対照実験で短時間見せた時と長時間見せた時において正しく理解できるか、合成物が粗密な情報を利用できるのにも関わらず、別々に解釈できるかどうかを示す必要がある。表 1 は、*LF* 混合物、*HF* 混合物、*LF*、*HF* と *N* サンプルについてのマッチング試行の結果をまとめたものである。実験を開始する前に表示される画像のノーマルシーンを見せておいた場合と見せてない場合では、0.03 秒の画像に対する評価は変わる。見せない場合は全く分からなかったという意見が多く、ノーマルシーンを見せておくとその情報を元に判別をする事ができた。図 2 から分かるように *CtF* 処理は低周波において情報がシーンカテゴリを開始するのに十分であることを示すが、0.03 秒では、やはり個人の動体視力にも左右される。

3.3 実験 2

CtF 処理により素早いシーンカテゴリを説明できるかどうか

3.3.1 予想

実験 1 の結果から、先に低周波で表示された画像を認識すると考える。

3.3.2 方法

2 つの合成画 (道路の低周波と街の高周波の合成画と道路の高周波と街の低周波の合成画) を連続して表示する。その結果、認識は粗密な空間スケールで生じる事を示唆する。初めに、低周波であるぼんやりとした線を結果として生じる事により、視覚のシステムは記憶にあるシーンスキーマを促進しながら情報を入力し、判断することで、高周波のはっきりした線を習得していく。つまり画像の密な情報を処理する段階で、曖昧であった情報を改良し、認識する。

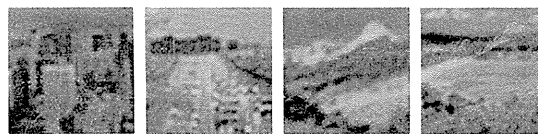


図 5: 道路 *HF* とビル街 *LF* の合成画, ビル街 *HF* と道路 *LF* の合成画, 海 *HF* と山 *LF* の合成画, 山 *HF* と海 *LF* の合成画.

3.3.3 結果

<i>CtF</i>	<i>FtC</i>
63.3 %	36.7 %

表 2: 実験 2 の結果.

この実験により *CtF* と *FtC* ((Fine-to-Coarse) 処理: *CtF* と逆に高周波から低周波へと処理していく。) は同等の条件で利用できるが、結果により *FtC* よりも *CtF* による解釈が行われている事が分かる。これらの結果、カテゴリ処理の分野で実験 1 の結果、*CtF* で使用する情報は速いシーン処理の証明となる。実験 1 で、境界エッジは高周波を見つけることにより表され、それゆえに処理の速い段階から認識の為に使用できるが、この実験において、境界エッジは、時々速いカテゴリの基礎であるという説明は立証できるが、シーンが知られていない時は、低周波より信用されない。今回は実験 1 と 2 を時間を空けず続けて試行したが、実験 1 の情報が被験者に印象を与えていたとしたら、実験 2 に影響が出ている事も考えられる。また実験をするにあたり、低周波と高周波の合成画において、同等な濃さや見え方になるように試みたが、個人の見え方の違いでも差が出ている事も考えられる。

4 考察

実験 1 の結果は *CtF* 処理の方法を示し、2 つ目の実験は曖昧な刺激の連続の自発的な分類により、*CtF* 解

積が FtC 解釈より組織的に好まれることを示す。相互的にこれらの結果は、初めの段階でシーンの明確な情報を信頼し、あとの段階において物体に基づいている時空間規模に依存するシーン認識処理であることを示している。シーンが記憶でシーンスキーマを活動的にする原因は低周波であり、物体に基づいた認識は高周波を利用する事がわかった。視覚に十分な時間を与えることで、シーンの物体の輪郭から細かい情報を認識する。高周波における空間スケールで、エッジは広範囲にわたる処理によりのみで取り除かれる、かなりのノイズレベルと構成要素を干渉する要因の認識に役に立つ。もし、知らないシーンで、とても速くシーンを分類しないとイケないならば、不確かであるが、高周波がシーンの初めに認識した低周波を能率的にするに違いない。

5 おわりに

本稿では、*P.G.Schyns* らの研究に興味を持ち、シーン認識する上において一般にあいまいな低周波成分から入り、次第に細かな高周波成分を認識していくという彼らが示した研究成果を実証した。この研究を通して、人間の認識の奥深さをあらためて感じ取ることができた。実験では 0.03 秒というとても短い表示時間で行ったが、その一瞬で画像を理解する事はとても難しいが、人間の脳にたくさんの情報が送られ被験者に大きな印象を与えていた。本稿では人間の周波数における認識を研究したが、この問題はまだまだ不明な点が多く存在する。人間だけではなく、イルカやコウモリなど他生物のシーン認識に対する周波数特性も大変興味深く思う。今後は、さらなる周波数の解析を通し、人間がどのようにシーンを認識していくか、また様々な風景や色彩などを比べ、その認識の違いや感じ方を細かく分析したい。

最後に、本研究にあたりご助言を頂いた伊達章助教授、ならびに実験にご協力を頂いた関係各位に深く感謝の意を表す。

参考文献

- [1] 安居院 猛, 中嶋 正之, 木見尻 秀子, C 言語による画像処理, 昭晃堂, 1990.
- [2] 安居院 猛, 長尾 智晴, C 言語による画像処理入門, 昭晃堂, 2000.
- [3] 井上 誠喜, 八木 伸行, 林 正樹, 中須 英輔, 三谷 公二, 奥井 誠人, C 言語で学ぶ実践画像処理, オーム, 1999.
- [4] A. Rosenfeld and A.C. Kak, Digital Picture Processing, Academic Press, Inc., 1976.
- [5] P.G. Schyns and A. Oliva, FROM BLOBS TO BOUNDARY EDGES : Evidence for Time-and Spatial-Scale-Dependent Scene Recognition, Psychological Science, Vol.5, No.4, 1994.
- [6] COMPUTATIONAL VISUAL COGNITION LAB AUDE OLIVA, (<http://cvcl.mit.edu/new.htm>).
- [7] 人間の初期視覚系を参考にした特徴, (<http://www.neurosci.aist.go.jp/kurita/lecture/st-atface/node3.html>).
- [8] 視覚による認識, (<http://home4.highway.ne.jp/nicholas/x73020/b/3-.html>).