連続講座

断層映像法の基礎　第20回
画像の相互情報量

篠原　広行[1)]、伊藤　猛[1)]、今井　貴祐[1)]、伊藤　賢司[1)]、橋本　雄幸[2)]

1) 首都大学東京大学院人間健康科学研究科放射線科学系
2) 横浜創英短期大学情報学科

はじめに

画像の位置合わせにおいて画像の変形を伴わない剛体モデルを仮定するとMRIとPETなどイメージング手段の異なる画像間の位置合わせについては、相互情報量（mutual information）による画像位置合わせが臨床で用いられている。相互情報量は通信分野のエントロピーの考えを利用したものである。第20回では画像における相互情報量について解説する。

まず、相互情報量を求めるための1次元ヒストグラムと2次元ヒストグラムの考え方を解説する。その後、2次元ヒストグラムから相互情報量を算出する方法を解説し、異なる画像の場合と同じ画像が平行移動や回転移動した場合に、2次元ヒストグラムとそこから算出される相互情報量の値がどのように変化するかを示す。

1. 1次元ヒストグラムと2次元ヒストグラム
2. 相互情報量
3. 相互情報量の画像への応用
4. 画像の平行移動と回転移動による相互情報量の変化

1. 1次元ヒストグラムと2次元ヒストグラム

1.1　1次元ヒストグラム

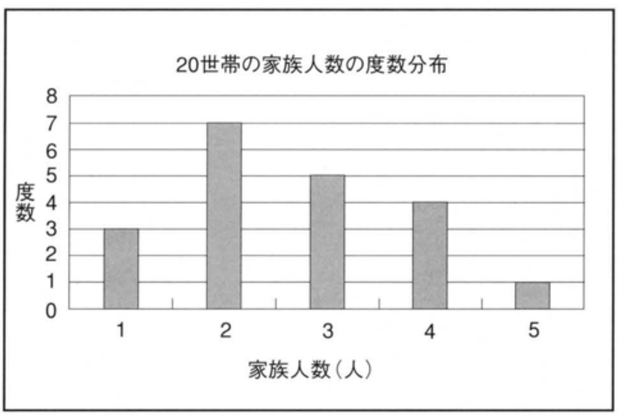
ヒストグラムは画像の濃度情報を表すもので、同じ濃度値の画素数を数えて度数を求め、濃度値とそれに

表1.
20世帯の家族人数のデータ

20世帯の家族人数	
世帯番号	家族人数
1	3人
2	1人
3	4人
4	2人
5	4人
6	3人
7	3人
8	2人
9	3人
10	2人
11	1人
12	2人
13	4人
14	2人
15	1人
16	5人
17	2人
18	3人
19	2人
20	4人

表2. 表1の度数分布表

家族人数	度数
1	3
2	7
3	5
4	4
5	1



別刷請求先：〒116-8551 東京都荒川区東尾久7-2-10　首都大学東京人間健康科学研究科放射線科学系　篠原広行
TEL：03-3819-1211　FAX：03-3819-1406

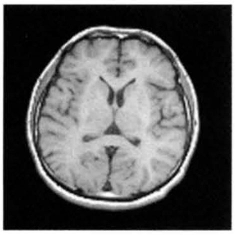


図2.
MRIの画像(1)(128×128画素)

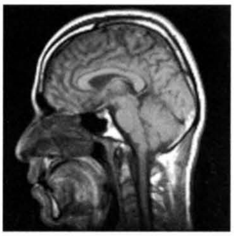


図4.
MRIの画像(2)(128×128画素)

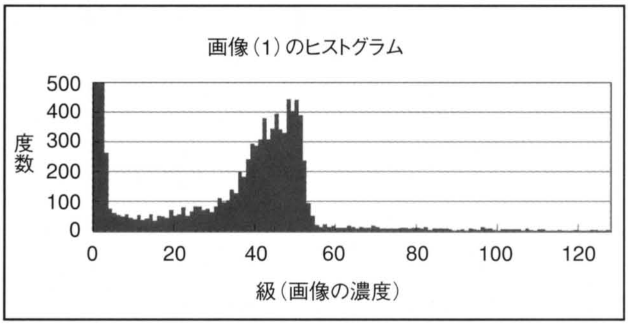


図3. MRIの画像(1)から作成したヒストグラム

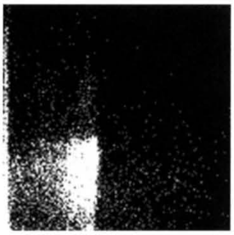


図5.
MRIの画像(1)と画像(2)から作成した2次元ヒストグラム
128×128画素ヒストグラムは画像全体に分布している。

図6.
MRIの画像(1)を2つの画像として作成した2次元ヒストグラム(128×128画素)。ヒストグラムは45度の直線上に分布する。

対応する度数の関係をグラフで表す。一般的に、あるデータを値ごとに度数を数え、値と度数の関係を求めたものを度数分布と呼び、棒グラフで表す。たとえば、**表1**に示すような20世帯の家族人数のデータがあるとき、これを度数分布にすると**表2**のようになり、度数分布の棒グラフは**図1**のようになる。医用画像のように濃度値が多い場合は、濃度値を等間隔の範囲に分けて適当な級を作り、その級の範囲内で画素数を数えて度数を求める。濃度値を等間隔に分けるのは、画像をある階調に分ける計算と等しくなる。n階調に分ける場合、画像の画素値の最大値をf_{max}、最小値をf_{min}、画素値をf(i)とすると、そのときの級C_a(i)は

$$C_a(i)=\mathrm{int}\left(n\times\frac{f(i)-f_{min}}{f_{max}-f_{min}}\right) \tag{1}$$

で求まる。ここで、簡単のために画素の番号は1次元でi番目という形にする。nの値は、ヒストグラムを作成するときの級の総数となる。この級の総数はbin数とも呼ばれる。この計算をすべての画素値で行い、C_a(i)の値ごとに度数を求めると、級aごとの度数分布h(a)は

$$h(a)=\sum_{i=1}^{pix}\delta(C_a(i)-a) \tag{2}$$

と表せる。ここで、pixは画像の総画素数で、δ(x)はクロネッカのデルタ関数を意味し

$$\delta(x)=\begin{cases}1 & x=0\\ 0 & \text{otherwise}\end{cases} \tag{3}$$

とする。級の値aに対する度数分布h(a)のグラフを作成すれば、1次元のヒストグラムとなる。
図2で示す画像を128個のbin数でヒストグラムにしたものを**図3**に示す。**図3**のグラフは、横軸が級で縦軸が度数になっている。

1.2 2次元ヒストグラム

2次元ヒストグラムは2つの画像からヒストグラムを作成する。1つ目の画像f(i)のi番目の画素をn階調に変換したものをC_a(i)とする。また、2つ目の画像g(i)のi番目の画素をn階調に変換したものをC_b(i)とする。C_a(i)とC_b(i)は、それぞれ1からnまでの値をとる。2次元の度数分布h(a,b)は

$$h(a,b)=\sum_{i=1}^{pix}\delta(C_a(i)-a)\cdot\delta(C_b(i)-b) \tag{4}$$

となる。2次元の度数分布は、i番目の画素において2次元位置である(C_a(i), C_b(i))ごとに度数を数えていくことになる。この2次元度数分布は横軸をaの級、縦軸をbの級とした2次元画像のヒストグラムとして表示することができる。2次元ヒストグラム画像の画素値

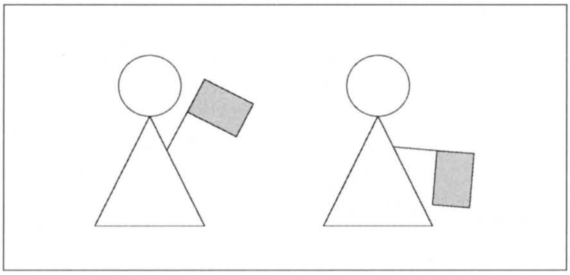


図7. 2つの事象を手旗信号で伝える様子
旗は1つあれば情報を伝えることができる。

が度数に相当する。**図2**と**図4**に示す画像から作成した2次元ヒストグラムの画像を**図5**に示す。値が2次元ヒストグラムのさまざまな場所に分布する。2つの画像が同じ場合は、$C_a(i) = C_b(i)$となるので2次元ヒストグラムが45度の直線に分布する。**図2**の画像を2つの画像として2次元ヒストグラムを作成すると**図6**のようになり、45度の直線に分布しているのが分かる。

2. 相互情報量

画像の位置合わせで使われる相互情報量は、2次元ヒストグラムから計算される。もともと、情報量や相互情報量という概念は、情報通信の分野で用いられているものである。ここでは、情報量や相互情報量がどのようなものであるかを解説する。

2.1 情報量

情報という形のないものの量をどのように数値で表すかを考える。ある事柄を離れた相手に伝えるにはどのような手段があるかを考えると、たとえばのろしを上げたり、光を照らして知らせたりということがある。ある事象が起こったか起こっていないかは、2つの事象が同じ確率で起こると考えると、その確率は1/2となる。それを知らせるには、たとえば手旗信号で伝えるとすると、**図7**のように旗を上げるか下げるかで伝えられる。このとき旗は1つあれば伝えることができる。このときの情報量がもっとも小さい場合で「1」と考える。

では、確率が1/4の事象を伝えようとする場合は、どのようにすればよいだろうか。確率が1/4ということは4つのうちの1つが起こったと考えればよい。それを伝える場合4つの事象を伝えられればよいことになる。手旗信号で考えると、**図8**のように旗を2つ持って、それぞれの旗を上げるか下げるかで4つの事象を伝えることができる。旗は2つ必要となるので、この場合の情報量は「2」と考える。

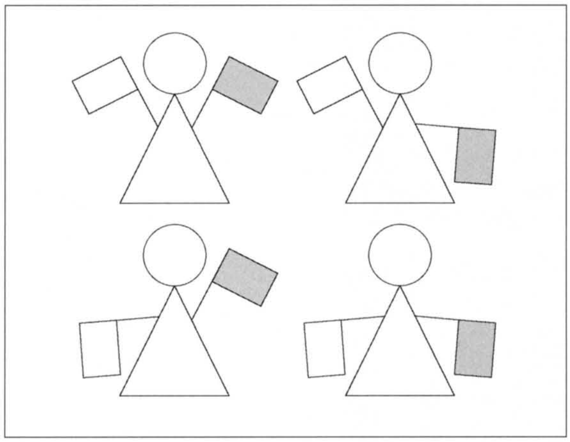


図8. 4つの事象を手旗信号で伝える様子
旗が2つあれば情報を伝えることができる。

同様に考えていくと、確率がその半分の1/8のときは、8つの事象が考えられるので、旗は3つ必要となり、情報量は「3」となる。確率が1/16の場合は、旗は4つ必要となり、情報量は「4」となる。これを、確率をpとして一般式に直すと、情報量Hは

$$H=\log_2 \frac{1}{p}=-\log_2 p \tag{5}$$

と表すことができる。この式が情報量Hの定義となり、その単位は「bit」と呼ばれる。この式は、次の条件を満たす。

1) 情報量Hは発生確率pに対して単調減少する関数である。
2) 新しい情報の追加によって情報量が加算的に増加していく。

$$f(p_1 \times p_2)=f(p_1)+f(p_2) \tag{6}$$

3) 確率が半分(p= 0.5)のときに、1[bit]と定義できる。

2.2 平均情報量(エントロピー)

事象が次々と発生する場合、その事象が発生する確率によって情報量は変わってくる。その系列全体の情報量を考える場合は、情報量の平均値を求める。この値を平均情報量、あるいは、エントロピー(entropy)と呼ぶ。テレビのニュースや新聞を考えた場合、発生する確率の低いものを数多く伝えている。発生する確率が低いということは、情報量が大きくなり、情報量の大きいものほど数多く伝えているので、伝えている

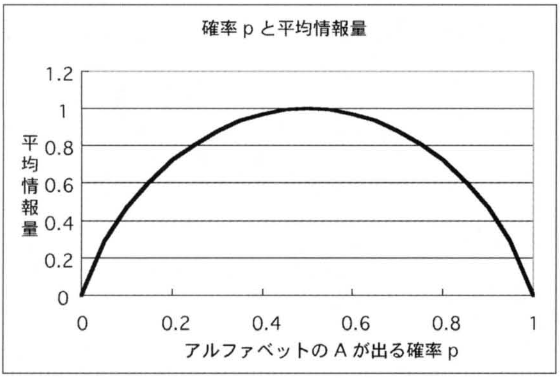


図9.
2つのアルファベットが出る確率と平均情報量との関係

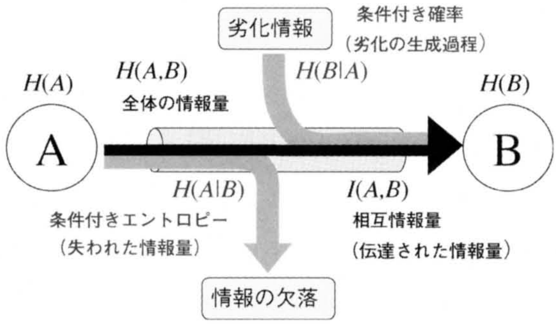


図10. 情報通信論的な解釈で相互情報量の関係

平均情報量も大きいと想像できる。

発生する事象Aをa_i (i=1〜m) とし、個々の事象の確率を$p(a_i)$とする。それぞれの情報量を$H(a_i)$とすると、平均情報量H(A) は重み付け平均（期待値）で表される。式で表すと

$$H(A)=\sum_{i=1}^{m}p(a_i)H(a_i)=-\sum_{i=1}^{m}p(a_i)\log_2 p(a_i) \tag{7}$$

となる。この平均情報量は、その情報源がどれだけ情報を出しているかを測る尺度となる。

たとえば、2つのアルファベットA、Bがランダムに出力されるとする。それぞれの確率をp(A)、p(B)とすると、Aが出力されたことを知った場合、$-\log_2 p(A)$ビットの情報を得る。また、Bが出力されたことを知った場合、$-\log_2 p(B)$ビットの情報を得る。平均すると1アルファベットあたり

$$H=-p(A)\log_2 p(A)-p(B)\log_2 p(B) \tag{8}$$

の情報量を得ることになる。アルファベットのAが出る確率をpとするとBが出る確率は1－pとなるので平均情報量H(p) は、

$$H(p)=-p\log_2 p-(1-p)\log_2(1-p) \tag{9}$$

となり、pの範囲を0から1でグラフにすると**図9**のようになる。確率pが0.5の場合が平均情報量はもっとも大きくなる。

2.3 条件付き平均情報量

全く別系列で、なおかつ互いに関連している2つの事象（結合事象）A、Bを考える。事象Aがa_i (i=1〜m)、事象Bがb_j (j=1〜n) としたとき、それぞれの確率を$p(a_i)$と$p(b_j)$とし、それらが同時に起こる確率である同時確率を$p(a_i \cap b_j)$で表すと、同時確率の平均情報量H(A,B) は、

$$H(A,B)=-\sum_{i=1}^{m}\sum_{j=1}^{n}p(a_i \cap b_j)\log_2 p(a_i \cap b_j) \tag{10}$$

と表される。これを結合エントロピーと呼び、2つの系列全体の平均情報量となる。

また、事象Bが起こった条件下で事象Aが起こる確率を考える。これを条件付き確率という。事象Bのb_jが起こった条件下で事象Aのa_iが起こる確率を$p(a_i \mid b_j)$と書くと、a_iとb_jが同時に起こる確率である同時確率$p(a_i \cap b_j)$は、事象Bが起こる確率に、事象Bが起こってから事象Aが起こる確率を掛け合わせて

$$p(a_i \cap b_j)=p(b_j)\cdot p(a_i|b_j) \tag{11}$$

と表すことができる。

よって、条件付き確率$p(a_i \mid b_j)$は

$$p(a_i|b_j)=\frac{p(a_i \cap b_j)}{p(b_j)} \tag{12}$$

となる。これをもとに、条件付き平均情報量（条件付きエントロピー）を考えると、次のように導出される。まず事象Bがb_jであるとしたときの事象Aの平均情報量は、条件付き確率をもとに平均情報量を求めるので

$$H(A|b_j)=-\sum_{i=1}^{m}p(a_i|b_j)\log_2 p(a_i|b_j) \tag{13}$$

となる。今度は、事象B全体で考えると、その条件付き平均情報量は、この事象Bについての期待値を計算すればよいので

$$H(A|B)=\sum_{j=1}^{n}p(b_j)H(A|b_j) \tag{14}$$

となる。この (14) 式に (13) 式を代入すると

$$H(A|B)=-\sum_{j=1}^{n}p(b_j)\sum_{i=1}^{m}p(a_i|b_j)\log_2 p(a_i|b_j) \\ =-\sum_{i=1}^{m}\sum_{j=1}^{n}p(a_i\cap b_j)\log_2 p(a_i|b_j) \tag{15}$$

と表すことができる。この条件付き平均情報量を

$$H(A|B)=-\sum_{i=1}^{m}\sum_{j=1}^{n}p(a_i\cap b_j)\log_2\frac{p(a_i\cap b_j)}{p(b_j)} \tag{16}$$

$$=-\sum_{i=1}^{m}\sum_{j=1}^{n}p(a_i\cap b_j)\log_2 p(a_i\cap b_j)+\sum_{i=1}^{m}\sum_{j=1}^{n}p(a_i\cap b_j)\log_2 p(b_j)$$

$$=-\sum_{i=1}^{m}\sum_{j=1}^{n}p(a_i\cap b_j)\log_2 p(a_i\cap b_j)+\sum_{j=1}^{n}\left\{\sum_{i=1}^{m}p(a_i\cap b_j)\right\}\log_2 p(b_j)$$

$$=-\sum_{i=1}^{m}\sum_{j=1}^{n}p(a_i\cap b_j)\log_2 p(a_i\cap b_j)+\sum_{j=1}^{n}p(b_j)\log_2 p(b_j)$$

$$=H(A,B)-H(B)$$

と変形して考えると、事象AとB全体の平均情報量H(A, B)から事象Bの平均情報量H(B)を引いているので、事象Aのみが知りえる情報の平均情報量を表している((16) 式の3行目の{　}内は後述の(22)式で表される周辺確率 $p(b_j)$ になる。周辺確率は同時確率 $p(a_i\cap b_j)$ を一方の確率変数について総和したものである。)。これは情報がAからBに伝達する時にその通信路において失われる平均情報量を表す。この関係について、情報通信論的な解釈で図示したものを**図12**に示す。

2.4　相互情報量(mutual information)

相互情報量は、情報量の平均値である平均情報量から計算されるので、平均相互情報量とも呼ばれる。2つの事象A, Bが共通に持っている平均情報量である平均相互情報量I(A,B)を考える。2つの事象A, Bが独立でない場合、それぞれの持っている平均情報量H(A)とH(B)の和に対して、事象A, B全体の平均情報量H(A, B)は異なり、その差が共通に持つ平均情報量となる。よって、平均相互情報量I(A,B)は、それぞれの平均情報量の和から全体の平均情報量を引いて

$$I(A,B)=H(A)+H(B)-H(A,B) \tag{17}$$

と表される。この(17)式と条件付き平均情報量の(16)式から、平均相互情報量は

$$I(A,B)=H(A)-H(A|B) \tag{18}$$

と表すこともできる。**図10**にこれらの関係も示されている。すなわち、送信側Aが元々持っていた平均情報量H(A)から、条件付き平均情報量H(A|B)分だけ失われてI(A,B)のみがBに到達するのである。それに、途中で情報の劣化による条件付き平均情報量H(B|A)が付加され、結局、H(B)がBに受信される。

求めたい平均相互情報量は

$$I(A,B)=H(A)-H(A|B) \tag{19}$$

$$=-\sum_{i=1}^{m}p(a_i)\log_2 p(a_i)+\sum_{i=1}^{m}\sum_{j=1}^{n}p(a_i\cap b_j)\log_2 p(a_i|b_j)$$

$$=-\sum_{i=1}^{m}\left\{\sum_{j=1}^{n}p(a_i\cap b_j)\right\}\sum_{j=1}^{n}\log_2 p(a_i)+\sum_{i=1}^{m}\sum_{j=1}^{n}p(a_i\cap b_j)\log_2 p(a_i|b_j)$$

$$=-\sum_{i=1}^{m}\sum_{j=1}^{n}p(a_i\cap b_j)\log_2 p(a_i)+\sum_{i=1}^{m}\sum_{j=1}^{n}p(a_i\cap b_j)\log_2 p(a_i|b_j)$$

$$=\sum_{i=1}^{m}\sum_{j=1}^{n}p(a_i\cap b_j)\log_2\frac{p(a_i|b_j)}{p(a_i)}$$

$$=\sum_{i=1}^{m}\sum_{j=1}^{n}p(a_i\cap b_j)\log_2\frac{p(a_i\cap b_j)}{p(a_i)p(b_j)}$$

と導かれる((19) 式の3行目の{　}内は後述の(21)式で表される周辺確率 $p(a_i)$ になる)。ここで事象A, Bのそれぞれで起こる確率 $p(a_i)$ と $p(b_j)$ は周辺確率(marginal probability)と呼ばれる。

3. 相互情報量の画像への応用

相互情報量の画像への応用を考える。画像のある値が出現する確率は、ヒストグラムから容易に計算できる。ヒストグラムは、画像の値の度数分布に相当するので、それを総度数で割れば、画像のある値が出現する確率分布となる。

次に2つの画像の結合事象を考える。2つの画像から2次元ヒストグラムが作成でき、2次元ヒストグラムを総度数で割ると、2つの画像のそれぞれの値が出現する同時確率となる。また、2次元ヒストグラムから作成した同時確率の分布をそれぞれ、縦方向、横方向に積分(加算)したものが、個々の画像の確率分布になる。これは、周辺確率分布に相当する。同時確率分布と周辺確率分布が分かれば、前述の式から相互情報量を求めることができる。

2次元ヒストグラムを $h(a_i,b_j)$ として相互情報量を計算する手順を以下に示す。ヒストグラムは、ある程度の値を束ねて級の度数を算出するので、そのときのヒストグラムの級の総数をbinとすると、同時確率 $p(a_i\cap b_j)$ は

$$p(a_i\cap b_j)=\frac{h(a_i,b_j)}{\sum_{i=1}^{bin}\sum_{j=1}^{bin}h(a_i,b_j)} \tag{20}$$

となる。周辺確率 $p(a_i)$ と $p(b_j)$ は、それぞれ

$$p(a_i)=\sum_{j=1}^{bin}p(a_i\cap b_j) \tag{21}$$

$$p(b_j)=\sum_{i=1}^{bin}p(a_i\cap b_j) \tag{22}$$

となる。これらの式と(19)式から、相互情報量I(A,B)は、

$$I(A,B)=\sum_{i=1}^{bin}\sum_{j=1}^{bin}p(a_i\cap b_j)\log_2\frac{p(a_i\cap b_j)}{p(a_i)p(b_j)} \qquad (23)$$

と表される。

2つの画像が同じ場合、相互情報量が最も大きくなる。2つの画像が全く無関係(独立)な場合は、$p(a_i\cap b_j)=p(a_i)p(b_j)$ となり、相互情報量は0となる。情報論的な解釈では、2つの画像が同じということは、双方で失われた情報はないということで、すべて情報が伝達され相互情報量は最大となる。また、2つの画像が無関係なときは、伝達の過程ですべての情報が失われている状態であり、相互情報量はなくなる。**図5**に示した異なる画像間の2次元ヒストグラムから求めた相互情報量の値は、0.743188となる。また、**図8**の同じ画像同士の2次元ヒストグラムから求めた相互情報量の値は4.589811となる。同じ画像同士の2次元ヒストグラムから求めた後者の値のほうが相互情報量は大きくなる。

4. 画像の平行移動と回転移動による相互情報量の変化

画像が平行移動した場合と、回転移動した場合の2次元ヒストグラムの様子と相互情報量の変化を**図11**

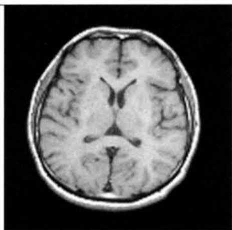
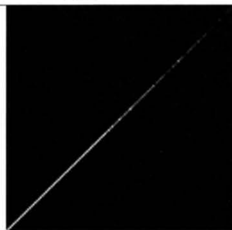
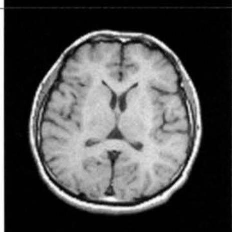
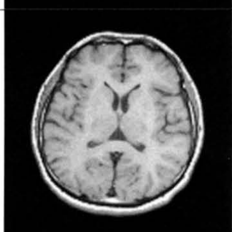
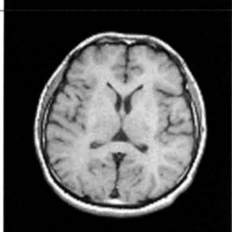
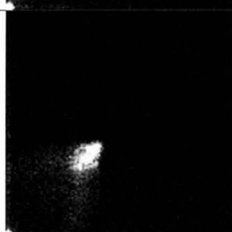
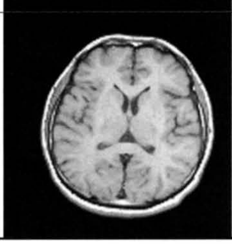
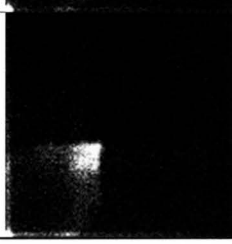
x方向の移動値	画像	2次元ヒストグラム	相互情報量
なし			4.589811
1画素			1.697539
2画素			1.427240
4画素			1.169425
8画素			0.928529

図11.
画像を平行移動したときの2次元ヒストグラムと相互情報量の関係

と**図12**に示す。**図11**には、**図2**の画像をx方向に図に示した画素だけ平行移動し、もとの画像との2次元ヒストグラムの様子と、そのときの相互情報量の値を示す。平行移動すると2次元ヒストグラムは45度の直線から徐々に幅を持つように値が分布し、それに伴って相互情報量も小さくなる。**図12**では、**図2**の画像を図に示す角度だけ回転移動し、もとの画像との2次元ヒストグラムの様子と、そのときの相互情報量の値を示す。回転移動した場合も、2次元ヒストグラムは45度の直線から徐々に幅を持つようになり、相互情報量も小さくなる。相互情報量がもっとも大きくなるときは、2つの画像の位置が一致したときであり、その結果を利用すれば、相互情報量を画像の位置合わせの指標とすることができる。

謝辞：本稿で使用したプログラムの開発は，首都大学東京共同研究（第一ラジオアイソトープ研究所）平成17年度～19年度「体幹部を対象としたマルチモダリティ位置合わせ法（アルゴリズム）の考案とソフトウエアの開発」によるものである．

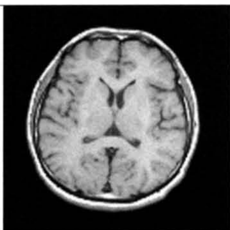
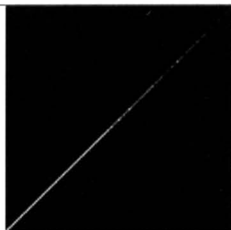
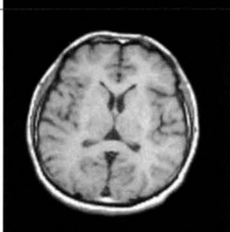
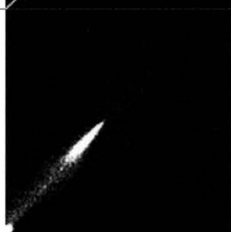
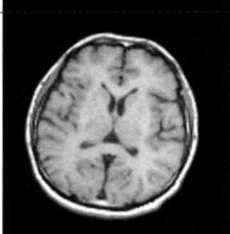
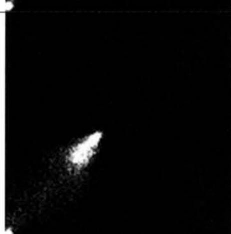
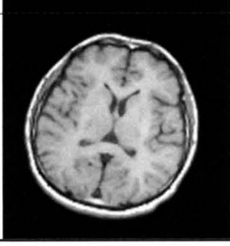
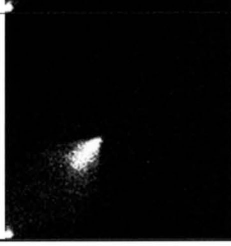
回転角度	画像	2次元ヒストグラム	相互情報量
0度			4.589811
1度			2.181722
2度			1.882536
4度			1.641860
8度			1.458493

図12.
画像を回転移動したときの2次元ヒストグラムと相互情報量の関係