

## 自己相関特徴変化による学習法の提案とサンプリングの違いによる学習効果の比較

村山 誠      加藤 邦人      山本 和彦

岐阜大学工学部応用情報学科  
〒501-11 岐阜市柳戸1-1

近年、動画像から人物を同定する研究が各方面でおこなわれている。これらは通常パターンマッチングにより行なわれており、辞書の作成方法が認識率に影響を与える。ある一定時間、人物を撮影した動画像から特定フレームを学習データとして選択する場合、等間隔にサンプリングする方法や画像の変化によりサンプリングする方法が考えられるが、これらは入力画像によっては偏りを持った学習データが作成されてしまう。そこで自己相関特徴量の変化に合わせて辞書を作成する方法を提案する。さらに、入力画像によってこれら2つの手法による認識率にどのような違いが現れるかを実験・比較した。

### **Sampling method to make dictionary and compare effect of learning**

**Makoto Murayama and Kunihiro Kato  
and Kazuhiko Yamamoto**

Department of Information Science,  
Gifu university, 1-1 Yanagido, Gifu, 501-11 Japan

Recently, many organizations were doing research about corresponding image. Usually, this type of research uses pattern matching. The produce of recognition rate of pattern matching is influenced by the use of dictionary of pattern. There are "fix period" and "variable period" sampling methods of making pattern matching dictionary. The aspect of the database influences in sampling methods. This paper suggests a new method of sampling a feature of the autocorrelation and compares the recognition rate of it with the mentioneds methods.

# 1 まえがき

近年、パターンマッチングを用い人物を認識する研究が各方面で行われている。[1][2][3][4] そこで認識辞書をつくる際、データによっては、人物の偏った表情や似通った画像ばかりを学習してしまう事が起こり得る。そのような辞書で、認識を行うと、認識率は低下すると考えられる。そこで、そのような現象を避けるために、学習データを作成する際に、似通った画像をできるだけ避け、多種多様なデータを集めた辞書を作成する手法が必要である。一般的にはデータから特徴量の変化を算出し、変化の大きいものをサンプリングし学習データを構成する方法が考えられるが、本稿では、データから特徴量の変化を算出し、変化の大きいものをサンプリングし学習データを構成する方法を提案する。さらに、これら2つの手法について比較を行いそれぞれの方法の性質について述べる。

## 2 認識システムの概要

### 2.1 高次局所自己相関関数特徴による人物判別

自己相関関数は平行移動に対して不変であることが知られている。その高次への拡張が高次自己相関関数である。画面内の対象画像を  $f(r)$  とすると、 $N$  次自己相関関数は、変位方向  $(a_1, a_2, \dots, a_N)$  に対して

$$x^N(a_1, a_2, \dots, a_N) = \int f(r)f(r+a_1)\dots f(r+a_N)dr$$
 で定義される。従って、高次自己相関関数は、次数や変位方向  $(a_1, \dots, a_N)$  の取り方により無数に考えられる。しかし、画面に関する加法性の条件を満たすためには、変位方向は参照点  $r$  の局所領域に限定されなければならない。ここでは、簡単のため高次自己関数の次数  $N$  を2までとする。

また変位方向を参照点  $r$  のまわりの  $3 \times 3$  画素の領域に限定する。平行移動により等価な特徴を除くと、2値画像に対して、特徴の数は全部で25個になる。[5][7] 各特徴の計算は、局所パターンに対応する画素の値の積を全画像に対して足し合わせればよい。

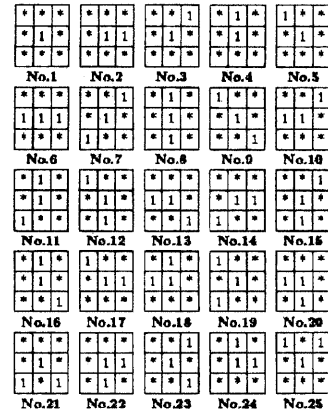


図 1: マスクパターン

濃淡画像 (320x240 pixels) に Prewitt オペレータによる微分処理と2値化を施し、エッジを抽出した画像を図1のマスクパターンでサーチしていき、得られた画像面上の平行移動に対して不変な25種の自己相関特徴を得る。ここで得られた25種の特徴の分布パターンは、個々を反映した固有の物となる。個人識別は、この分布パターンの違いを利用して行っている。[1]

### 2.2 多重解像度

高次局所自己相関関数に基づく特徴は、近傍の画素の濃淡値の積を画像全体に足し合わせて得られる。これは、画像の局所的な特徴である。従って、高解像度の画像から抽出された画像は認識のためには、細かすぎるかもしれない。そこで、今回は3種類の解像度の画像をつくり、特徴を抽出している。図2に画素の対応関係を表す。簡単化のため低解像度の画像は、高解像

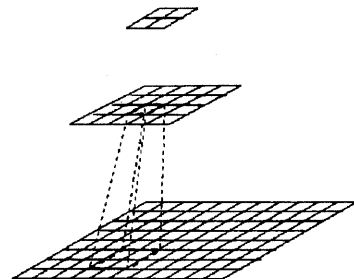


図 2: 解像度の対応関係

度の4画素の平均値としている。今回は、解像度1/4、1/16の画像も使い、計75種の特徴量を算出している。[6]

### 2.3 線形判別分析

判別には線形判別分析を用いた。線形判別分析は、より低次元で、与えられたクラスを良好に判別する線形写像、を構成する多変量解析の手法である。

$$y = A^T x \quad (1)$$

$x$ : 初期特徴ベクトル、 $y$ : 新特徴、 $A$ : 係数行列

本研究では、最適な係数行列  $A$  を学習用画像データから算出している。[8][7]

なお、学習過程で用いない画像の識別では、式(1)から  $y$  を計算し、最も近い平均ベクトルを持つクラスに属するとした。

今回の実験では、このシステムを用い実験・比較を行った。

## 3 データ

今回、実験データとして2種類のデータセットを用意した。

データセット1はデジタルビデオカメラにより6人の顔を各人につき1分間撮影した物を使用した。背景は白色の単純背景である。

一分間のうち、始めの45秒を正面を向いたまま静止し、残り15秒は、自由に動いている。

1秒間に30フレーム計10800フレームのうち、各人の前後50フレームを削除した合計10200のフレームから各サンプリング方法を適用した。図3にこのデータのフレーム間でフレーム間差分を取っていき、差分領域により動きの状態を表したグラフを示す。

このグラフは  $x$  軸に10200フレームまでの時間軸、 $y$  軸に各フレーム間の差分領域を取ったものである。

データセット2は、同様にデジタルビデオカメラにより6人をそれぞれ一人1分30秒撮影したデータである。

このデータにおいても、背景は単純背景で白色であり、動きに関してはデータ1のような差をつけた動きではなく1分30秒間動きつづけて

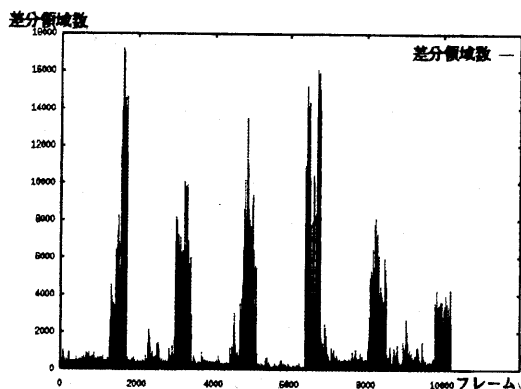


図3: 使用データの差分量

いるデータである。

## 4 サンプリング方法の提案

### 4.1 提案方法

本論文では用意したデータに対して次の3種類のサンプリング手法を適用した。

#### (1) 等間隔にサンプリング (等間隔法)

使用するデータから、ある決まった間隔でデータをサンプリングし学習データを作成する方法。間隔は学習に採用する枚数でデータ時間を割った値である。

#### (2) 自己相関特徴変化時サンプリング (要素変化量法)

全てのフレームに対し特徴を抽出し、75種の要素  $x_i$  の前にサンプリングされたデータ  $X_i$  と最大の変化をしたものを変化量  $\epsilon$  とし、この  $\epsilon$  がある閾値以上の変化があった時に学習データとして加えていく。  $\epsilon$  の計算を式(2)に示す。この操作を繰り返し、これを学習データとする。ここで、閾値は等間隔でサンプリングされたデータと比較するために学習に採用する枚数をデータ内でとりきれるように変化させ決定した。

$$\epsilon = \max |X_i - x_i| \quad (0 < i < 75) \quad (2)$$

#### (3) 自己相関特徴変化時にサンプリング (全体変化量法)

全てのフレームに対し特徴を抽出し、75種の要素の  $x_i$  が前にサンプリングされたデータ  $X_i$  との距離を算出し、それを変化量  $\epsilon$  とする。これ

がある閾値以上の変化があった時に学習データとして加えていく操作を繰り返し、学習データを作成する。ここで閾値は等間隔でサンプリングされたデータと比較するために学習に採用する枚数をデータ内でとりきれないように変化させ決定した。この変化量 $\epsilon$ は、次の式(3)で決定される。

$$\epsilon = \sum_{i=0}^{75} |X_i - x_i| \quad (0 < i < 75) \quad (3)$$

図4、図5に一人分のデータに上のそれぞれの学習データ選出の方法を適用しサンプリング点を得た図を示す。プロットされているサンプリング点数は、それぞれ100点、300点である。

差分領域数

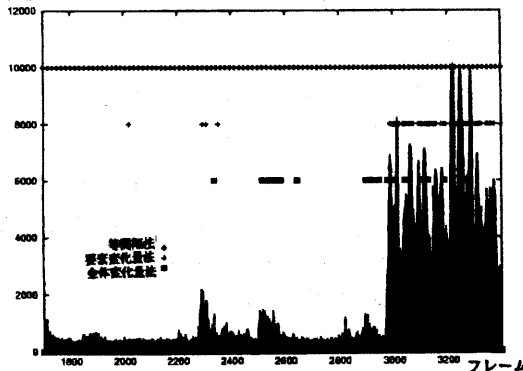


図4: データの差分量とサンプリング点(100点)

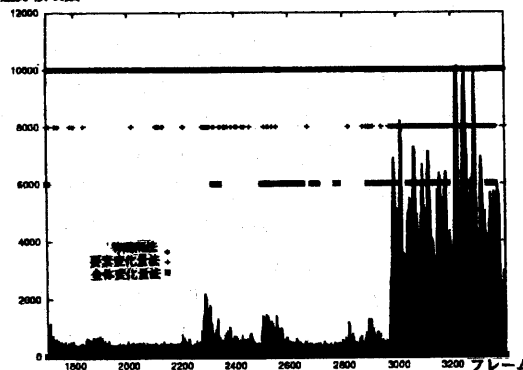


図5: データの差分量とサンプリング点(300点)

## 4.2 提案方法の意味

### (1) 等間隔法

等間隔にサンプリングする方法では、データセット1のように正面顔ばかりがデータの75

%を占めたような偏ったデータから学習すると、必然的に学習データの75%は、正面の顔ばかりとなってしまっている。

### (2) 要素変化量法

特徴量変化時サンプリングの方法では、正面ばかりを撮影したデータの75%からは少量しか学習データに採用せず動きをつけた残り15%の中から多くの学習データを採用した事がわかる。多種多様なデータが学習データのために選出されている。

### (3) 全体変化量法

特徴量変化時サンプリング1と違うのは、特徴量変化時サンプリング1が75種の各要素に大きな変化が現れた時にサンプリングするのに対し、この方法では、75種の合計、つまり個々の要素の変化を考えずにサンプリングする。

## 5 実験

データセット1、2を共に奇数フレーム、偶数フレームの2つに分け片方を学習データ用、もう片方を未知データとして学習データの量を変化させて認識率を計測した。ここでは、各クラスのデータに対してそれぞれ適切な変化値を搜索し、それぞれのクラスが、ほぼ同等の学習データを持つように設定した。各々の手法による認識率をグラフに示す。

データセット1に対する実験結果を図6、データセット2に対する実験結果を図7に示す。

認識率 (%)

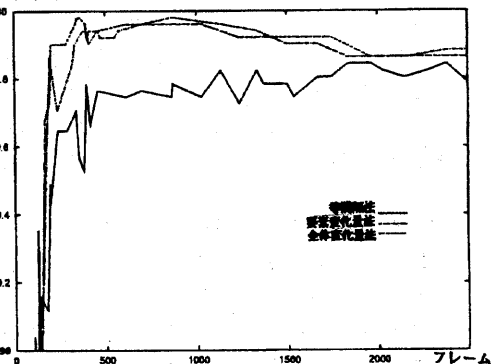


図6: データセット1における認識率の推移

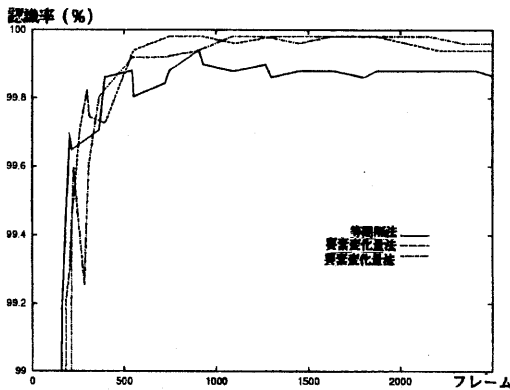


図 7: データセット 2 における認識率の推移

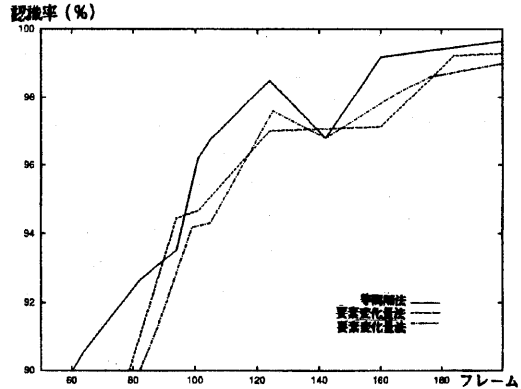


図 9: 学習が少ない時の認識率 (データセット 2)

## 6 考察

### 6.1 認識率についての考察

データセット 1、2 に対しての今回の実験では学習データ 50 枚以上の時、変化量を用いた方法の方が、等間隔法よりも、良い認識率を得る事ができた。しかし、学習データ個数が、少ない時には、変化量法の認識率が等間隔法の認識率より劣る現象が確認された。その時点のデータセット 1、2 の結果を図 8、図 9 に示す。

これは、等間隔法が少ない枚数の時でもデータ全体から均一にデータを収集するのに対し、変化量を用いた方法ではある点に偏ってサンプリングされてしまうためだと考えられる。

データセット 1 の場合では、等間隔法では動きの少ないデータの 75 % 分の部分、動きのあった残り 25 % 分の部分で、共にサンプリングされるのに対し、変化量法では、動きのあった変化の激しい場所数点でしかサンプリングされなかったために、動いてない変化の少ない場所

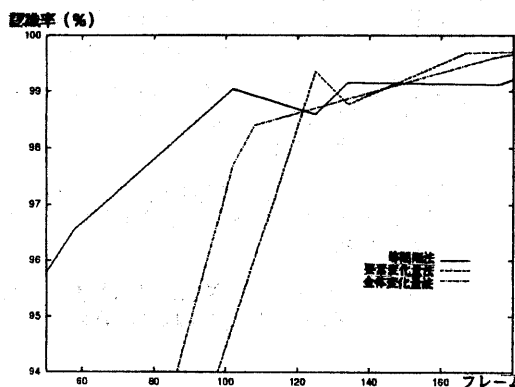


図 8: 学習が少ない時の認識率 (データセット 1)

の認識力が甘く、また変化の大きい地点も少ない枚数では、全ての認識はできなかったものと思われる。

また学習データを増やした時に、認識率は一定に収束すると考えらるが変化量を用いた方法は等間隔法よりも早く収束していくという事が結果より考えられる。これは、データセット 1 に対して言えば等間隔法では新たに学習データに加えられる時、そのデータの 75 % が以前学習したデータと同様な似通った画像、25 % が新たに違うバリエーションの画像であると考えられるためである。変化量を用いた方法では、新たに加わるデータの大半が今までに学習データに加えたデータとは違う画像であるためだと考えられる。

### 6.2 サンプリング法についての考察

完全に学習枚数を決定して実験を行った結果、どちらのデータセットに対しても、変化量を用いた方法よりも、等間隔法が優れた結果となった。これには、変化量の閾値を徐々に下げてサンプル数を決定する際に、指定した枚数で丁度終了することができず、数フレーム余分に抽出されるデータが増えてしまい、そこでの処理が問題となってしまったためである。そこで、違う変化量法についての実験も試みた。これは、ある規定値を設定しておき、前にサンプリングされたデータとの差が規定値以上であれば、サンプリングするという方法で、1 回の走査で設定枚数サンプリングできなかった場合は、規定値を下げて同様の操作を行いサンプリングする数を

増やして行く方法である。これを繰り返すことにより、学習データを作成する方法である。この方法で、実験したところデータセット1では、等間隔法よりも、良い認識率を得る事ができた。しかし、データセット2に対して実験を行ったところ、等間隔法、変化量を用いた方法、双方よりも、劣った認識率であった。これには、序々に変化の閾値を下げて行くために、ある点がサンプリングされ、次に2回目の走査で少し閾値を下げた点をサンプリングするという事は、前回サンプリングされた点周辺をサンプリングする事になってしまい、結果的に一点に集中して学習データをサンプリングするという現象が起きてしまうために、データセット1のような、後半に動きが集中した様なデータ等、特別なデータでは、良い結果を得られたが、データセット2のような通常の動作をするようなデータに使用するのには適切ではないと考えられる。

### 6.3 変化量を用いた方法の問題点の考察

変化量を用いた方法では、データセット1のような特別な動きがあったデータに対しては、動いた点を集中してサンプリングするために、あまり変化がなかった点、もしくは変化が少ない点では枚数に制限があった時などでは、学習データとして加えられないことも起こる。そのために動きのあった点に対しては、等間隔でサンプリングした学習データよりもより良い辞書を構成する事ができるが、全体の認識率を考えた時には、認識データの動いた点の全体における割合にも、影響してくるために、評価が難しくなる。しかし、多種多様なバリエーションを学習し汎用性を持たせられる学習方法だと思われる。全体特徴量法では、75特徴の全てを積み上げるために、要素特徴量法よりも、変化が激しくなる。そういった理由でサンプリング量が少ない時には、要素特徴量法よりも、より一層集中した箇所をサンプリングしてしまうために、認識率が悪くなったと考えられる。また変化量を用いた法とも、変化した時に学習サンプリングに加えるという点から、データセットにミスデータ、もしくは学習データとして加えたくないデータ等が入っている場合にはそのデータを拾ってしまう確率が高いという問題が残っ

ている。

## 7 むすび

本論文では学習データ作成の際、効率良く学習データを作成する方法を提案した。さらに、多種多様な学習データを集中的にサンプリングする提案方法の有効性が実験により示された。また提案方法では、不良データを含んだデータベースからのサンプリングには、適してない事が確認された。

今後の課題としては、様々なデータセットについて実験を行い、定量化する事である。

謝辞 本実験を行うにあたり御指導頂いた電子技術総合研究所の栗田多喜夫博士に感謝致します。

## 参考文献

- [1] 村山誠, 山本和彦, 長谷川修: “手の表面情報認識による人物同定”, 96年信学総大. (1996)
- [2] 赤松茂: “コンピュータによる顔の認識の研究動向”, 電子情報通信学会誌, vol.80.No.3 pp.257-266, March (1997)
- [3] M.Turk, and, A.Pentland: “Eigenfaces for recognition,” J.Cognitive.Neuroscience.vol.3.No.1, pp.71-86, March. (1991)
- [4] A.Pentland, B.Moghaddam, Thad Starner: “View-Based and Modular Eigenspaces for Face Recognition”, IEEE Conference on Computer Vision and Pattern Recognition. (1994)
- [5] 栗田多喜夫, 大津展之: “高次局所自己相関特徴に基づく適応的画像計測”, 第4回画像センシング技術シンポジウム講演論文集, (1988)
- [6] 栗田多喜夫, 大津展之, 佐藤辰雄: “高次局所自己相関特徴を用いた顔画像の認識実験について”, 電総研研究速報, (1992)
- [7] 栗田多喜夫: “柔らかな情報処理のための統計的手法の応用に関する研究”, 電総研報告, 第957号. (1993)
- [8] 大津展之: “パターン認識における特徴抽出に関する数理的研究”, 電総研報告, 第811号, (1981)