社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

多様な属性に柔軟に対応できる 人物属性認識の準教師付き学習フレームワーク

井尻　善久[†,††]　勞　　世竑[†]　　村瀬　　洋[††]

†† 名古屋大学大学院情報科学研究科　〒 464-8603　愛知県名古屋市千種区不老町
† オムロン株式会社技術本部コアテクノロジーセンター　〒 619-0283 京都府木津川市木津川台 9 丁目 1 番
E-mail: †{joyport,lao}@ari.ncl.omron.co.jp, ††murase@is.nagoya-u.ac.jp

あらまし　人物属性 (メガネ等の装着物，髪型等) 認識手法の新たなフレームワークを提案する．属性認識においては，属性が非常に多様であるため，使用時に必要とされる属性を追加変更しなければならないという問題がある．従って少数の (ラベル有) 学習データにより短時間で新たな属性に対応する必要がある．提案手法では，事前にラベル無データを用いて実際生じ得る属性の局所的パターンを集約した Spatial Codebook を抽出しておき，新たな属性に対応する際には，少数の (ラベル有) データと Codebook との類似度を特徴量として各属性の識別器を学習する．この過程は Codebook との簡単な類似度計算とその結果得られた低次元の特徴量を学習するだけであるので，データ量の多いピクセルレベルから学習を行う従来法に比べ高速に新たな属性への対応ができる．また低次元の特徴量を用いるので，少数のラベル有データによる学習でも高い性能を得ることができる．提案手法の有効性は実験により実証する．
キーワード　人，顔，属性，認識，準教師付き学習

A Semi-supervised Framework for Human Attributes Recognition to Deal with A Large Number of Diversity in Attributes

Yoshihisa IJIRI[†,††], Shihong LAO[†], and Hiroshi MURASE[††]

†† Graduate School of Information Science, Nagoya Univ.　Furo-cho Chikusa-ku Nagoya, 464-8603 Japan
† Core Technology Center, Corporate R&D, OMRON Corp.　Kizugawadai9-1 Kizugawa-city, 619-0283 Japan
E-mail: †{joyport,lao}@ari.ncl.omron.co.jp, ††murase@is.nagoya-u.ac.jp

Abstract　A novel framework for recognizing facial attributes such as glasses, hairstyles, etc, is presented. Difficulty is that facial attributes are so diverse and thus attributes of interest can be often changed. In the proposed framework, "Spatial Codebook", which consists of small number of representative local patterns, is learned with unlabeled data in offline process. When learning new attributes, similarities between the codebook and small number of labeled training data are used as features. This process is computationally more efficient than the baseline method which uses high dimensional pixel level features, since it is based on simple computation of the similarities and training with low dimensional features of the resulting similarities. Moreover, the low dimensional representation enables high accuracy even with small number of labeled training data. The effectiveness of the proposed framework is shown by experiments.
Key words　human, face, attribute, recognition, semi-supervised

1. はじめに

公共の場での人々の安心や安全のために，多くの監視カメラが公共の施設に設置されるようになってきた．これに伴い，それら全てを人間が目視確認するのが困難となってきている．このために自動監視技術もしくは監視支援技術の重要性が増大している．

一方，捜査支援や複数カメラによる人物追跡等において，監視カメラ映像から特定の人を見つけることは重要な要素である．ここで問題となるのは，少ない監視カメラ数で広い領域を監視するために，人物の顔が鮮明に見えない状態で運用されている監視カメラが多いことである．こうした条件下で人物特定を行

う際には，サングラス・マスク等の装着物や髪型・髪色等の属性の方が有効な手掛かりである場合が多い．本論文で論じるのはこうした監視カメラ条件で人物特定するのに必要とされる人物属性認識のフレームワークである．

人物属性認識の設計においては，二つの考え方がある．(1) 必要な属性全ての識別器を予め学習しておくアプローチ，もう一方は，(2) 必要な属性の識別器は必要に応じその場で追加変更するアプローチである．前者は注目する属性が少数で固定的である場合 (性別・年齢等) に有効であるが，予め用意された属性しか認識できない不自由さがあり，新たな属性を用いる必要性が生じるような場合には適さない．また属性は非常に多様でありそれら全てを設計時に考慮しておくのは困難であるという問題もある．従って，属性の多様性を考慮し新たな属性に柔軟に対応できる (2) のアプローチが適していると思われ，本論文で扱うのも (2) のアプローチである．一方このアプローチを成功させるためには，新たな属性の追加や変更を簡単に行う必要がある．従って学習に基づく手法を用いる場合「学習が短時間で済むこと」および「学習に必要なデータ数を簡単に用意できる」ことが重要となる．

本論文では上記のような柔軟なシステムを実現するため，図.1 に示したような新たなフレームワークを提案する．短時間で新たな属性に対応するためには，できる限りの計算をオフラインで行っておくことが有効である．従って提案手法では学習プロセスを，オフラインで (大量の) ラベル無データを用いて行っておく学習と，属性の追加変更時に比較的少数のラベル有データを用いて行う学習の，二つに分けている．オフライン学習では，大量のラベル無データを用い実際に生じ得る属性を構成する基本的パターンを抽出しておく．この際に各属性を構成する変動は局所的な領域に表れることが多いので局所的なパターン抽出を行う (2.1.1 章に詳述)．このようにして抽出された各局所的基本パターンのことを Spatial Codewords と呼び，Spatial Codewords の集合を Spatial Codebook と呼ぶ．これらは実際に生じ得る (ラベル無学習データに含まれる) 人物属性の基本的な変動パターンをある程度集約したものとなっている．さらにオフライン学習では，抽出された基本的変動パターンを正確に検出できるようにするために，Codewords パターンの検出器 (Codewords 検出器) を学習しておく (2.1.2 章に詳述)．属性の追加変更時には，対応したい属性の少数のラベル有学習データに Codewords 検出器を適用し，その出力を特徴量として属性識別器の学習を行う (2.2 章に詳述)．ここでは，オフラインで Spatial Codebook により属性を構成する基本パターン集約が行われているので，各属性変動がどの Spatial Codewords と特に関係があるのかを学習するだけで良い．実験結果により示すが，この学習は比較的少ないラベル有データで行うことができる．属性識別時には，識別用すべき入力画像に対し，属性追加変更時と同じ Codewords 検出器を適用し特徴量を得る．この特徴量に属性追加変更時に学習しておいた各属性識別器を適用し識別を行う．

顔属性認識に関しては多くの手法が既に提案されている．Moghaddam [1]，Hosoi [2]，Hayashi [3]，Zhuang ら [4] は性別，

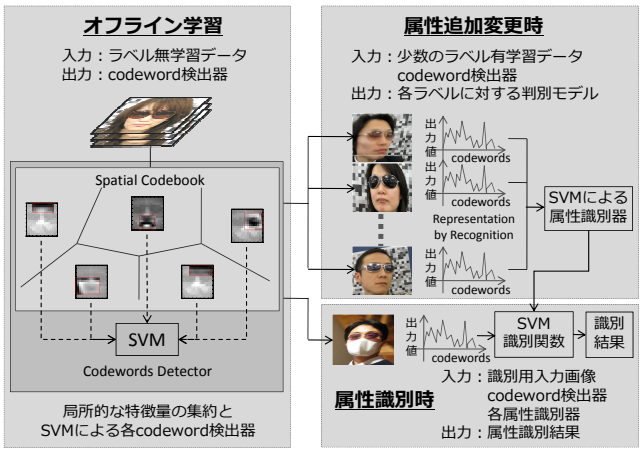

図 1　提案手法の流れ

年齢，人種等を個別に推定している．これに対し Lyon ら [5] は性別，人種，表情の複数属性に同時対応し，Wilhelm ら [6] は性別，年齢，表情の推定に加え顔認識を同時に実現している．これら初期の手法は，特定の属性に特化した特徴量を用いており，また大量のラベル有データを用いて学習を行うので，ユーザレベルで新たな属性を追加変更することはできない．比較的最近では多様な人物属性に対応したフレームワークとして Kumar ら [7] は FaceTracer を提案している．これは，予め決めた 10 個の局所領域で合計 450 種類の特徴量を抽出し，それぞれの特徴量に対する 450 個の SVM を訓練し，さらに Adaboost により精度の高い SVM のみを選択し，選択された SVM の出力をさらに上位の SVM で統合する手法である．この手法では特定の属性に特化した特徴量を用いる代わりに 450 種類の冗長な特徴を用いることにより様々な属性に対応している．一方で新たな属性の追加変更時には，450 種類の特徴抽出と識別器学習，識別器選択と統合を，一から行うので，計算コストが非常に高い上に，ラベル有学習データが大量に必要となる．従ってこの手法も，ユーザレベルで属性を追加変更するには適した方法とは言えない．

提案手法は多様な属性に，少数のラベル有データのみで対応するために Spatial Codebook を用いてパターンの集約を行っており，これは一般物体認識のフレームワークに類似である．一般物体認識の代表的フレームワークとして，画像内の特徴部位検出により得られた局所的特徴を，位置情報を使わず生起頻度のみで特徴量表現する Bag of Keypoints [8], [9](BoK) が提案されている．位置情報を使わないのは，隠れが生じたり見え方が異なったりする対象に対して，いつも同じ特徴部位検出が行われる保証がないために必ずしも同じ特徴領域間の比較が行えないからだと思われる．物体の位置決めが行われており局所領域同士の比較をある程度正確にできるときには，位置情報は有効であると考えられる．また BoK における特徴量の生起頻度計算にはベクトル量子化 [10]，すなわち得られた特徴量から最も近い Codewords を一つ選びそのクラスに割り付けることが行われる．しかし，複数の Codewords に近い特徴量の場合，一つを選ぶことにより複数の Codewords に類似しているという情報は失われ符号化誤差が生じてしまう [11].

提案手法は，認識対象を人物に絞っており人体検出や顔検出器により対象の位置決めをある程度行うことができるので，位置情報を含む Spatial Codewords を用いている．また得られた特徴量を最近傍 Codewords 一つに対応させるのではなく，各局所的な Codewords の検出器を学習しておき，その検出器の出力を特徴量表現として用いることで，符号化誤差が生じないように (Representation by Recognition) する．以上をまとめると，提案手法の貢献は次のとおりである．(1) 位置情報を用いた Spatial Codewords を提案する，(2)Spatial Codewords 検出器の出力による特徴量表現 (RbR) を提案する，(3) 上記二つの応用により，ユーザが少数のラベル有学習データだけで高速に属性を追加変更できる顔属性認識のフレームワークを提案する．

2. 提 案 手 法

提案手法の流れは第 1 章で簡単に述べた (図.1)．オフライン学習において，大量のラベル無データ $X = \{x_n; n = 1, \ldots, N\}$ を用い，顔内部の局所的基本パターンを表現する Spatial Codewords $\{(s_t); t = 1, \ldots, T\}$ からなる Spatial Codebook$S$ を構成する．さらにこれら局所的な基本的パターンを検出するための Codewords 検出器 $\{g_t; t = 1, \ldots, T\}$ を学習しておく．ここでは教師無データをクラスタリングした際のクラスタメンバを用いて学習するのでラベル有データは必要ない．次いで，属性追加変更時には，これら基本的パターンの検出器 g_t をが教師有学習データ $Y = \{y_m; m = 1, \ldots, M\}$ に適用しその出力 $\{f_t; t = 1, \ldots, T\}$ を特徴量として SVM で学習することにより各属性 ω_i の識別器 $H = \{h_i; i = 1, \ldots, I\}$ を構成している．属性識別時には，属性追加変更時と同じ方法で特徴量抽出した後に，属性追加変更時に学習しておいた識別器 H を適用し，識別結果 $\hat{\omega}$ を出力する．以下の各章で詳述する．

2.1 オフライン学習

2.1.1 Spatial Codebook の構成方法

以降の記述では，人物領域が顔検出もしくは人物検出等によりある程度正確に抽出されていると仮定する．多様な人物属性を表現するために，それらの属性を構成する局所的基本パターンを大量のラベル無データから抽出する．局所的基本パターンは，人物領域を小さな領域に分割し (領域間の重なりがあってもよい)，その各微小領域に対して K-means クラスタリングを適用することにより抽出する [9], [10]．ここで局所領域の数を L，各局所領域におけるクラスタ数を K とすると，全体で KL 個のクラスタ $\{C_{lk}\}(k = 1, \ldots, K; l = 1, \ldots, L)$ を得る．得られた全てのクラスタを用いてクラスタリングを行うと，各局所領域において同じようなクラスタリング結果しか得られない場合がある．例えば，左目の領域でサングラスのクラスタを抽出しておき，右目の領域で同じくサングラスのクラスタを抽出したとすれば，これらのクラスタのメンバとなるデータは計算誤差を除けば同じになる可能性が高い．このような場合，どちらかを除去しても各属性が出現したことを表現できる．従って，KL 個全てのクラスタ全て用いるのは冗長であることが分かる．このように各クラスタができる限り別々の属性と対応するように，クラスタ集合を求めることが重要となる．そのような観点で，KL 個のクラスタからそれを用いてクラスタリングされるメンバができるだけ直交となるようにクラスタを選択することを考える．また一方で，ほとんどのデータをメンバとして含むようなクラスタは，属性の識別問題を考えたときに望ましくない．一方で，メンバ数が非常に少ないクラスタは一つの属性と対応していない場合が多い．このように，全体のデータ分布を考えたときに中程度の出現確率を与えるクラスタが何らかの属性と対応している可能性が高い [11].

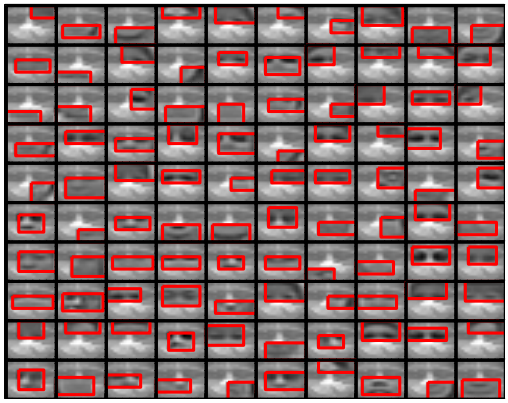

図 2　Spatial Codebook の一例: 各ブロックが顔領域を表わす

これら望ましい性質を実現するために各クラスタ C_{lk} に対して，そのクラスタへのメンバシップを表わす $\phi_{lk}(x_n)$

$$\phi_{lk}(x_n) = \begin{cases} 1 & x_n \in C_{lk} \text{ のとき} \\ 0 & x_n \notin C_{lk} \text{ のとき} \end{cases} \tag{1}$$

を定義する．ここで，$l = 1, \ldots, L$ は局所的な位置，$k = 1, \ldots, K$ は各局所領域における各クラスタ，$n = 1, \ldots, N$ でありラベル無学習データ x のインデックスである．これを用い，何らかの属性と対応したクラスタを選択するために ϕ_{lk} のエントロピー $-\sum P(\phi_{lk}) \log(P(\phi_{lk}))$ が最大 (maximum entropy) であるものを選択する．また，できる限り個別の属性を抽出するために，すでに選択された $\phi_{\hat{l}\hat{k}}$ に直交する ϕ_{lk} を持つ C_{lk} を順番に選択していく．実際には選択が進むにつれ，完全に直交する ϕ_{lk} を求めることは困難になるので，内積が最小である C_{lk} を選択することにより最も直交性が高いものを選択する (maximum orthogonality)．この基準を MEMO 基準 (Maximum Entropy Maximum Orthogonality) と呼ぶことにする．MEMO 基準を使うことにより，独立な属性変動を表わすクラスタを抽出することができる．Algorithm.2. 1. 1 に構成法をまとめる．またこうして選択された Spatial Codebook の一例を図.2 に示している．各画像内太枠が入力領域全体に対する Spatial Codeword の位置を表している．これを見ると，顔内部の比較的小さな領域のクラスタが Codewords になっており，それらがある程度，別々の装着物等と対応している様子がわかる．

2.1.2 Representation by Recognition: Spatial Codewords 検出器による特徴量表現

本章においては，前章で得られた Spatial Codebook の利用法を説明する．従来の物体認識等においては，keypoint 検出器

Algorithm 1 MEMO 基準による Spatial Codebook

Require: ラベル無学習用データ X; Spatial Codewords の数 T;
Ensure: MEMO 基準を満たす Spatial Codebook S

1: 初期化: $S = \{\emptyset\}$
2: 学習データセットから L 個の局所領域を抽出
3: **for** 局所領域 $l = 1, \ldots, L$ **do**
4: 各局所領域 l において K-means を実行し K 個のクラスタ中心 $C_{lk}(k = 1, \ldots, K)$ を求める.
5: **end for**
6: ラベル無学習データ X に対して，メンバシップ $\phi_{lk}(X)$ を計算
7: 得られた LK 個のクラスタ中心 C_{lk} から $\phi_{lk}(X)$ のエントロピが最大になるような $C_{\hat{l}_1\hat{k}_1}$ を選択し S に追加．すなわち
$(\hat{l}_1, \hat{k}_1) = \arg\max_{(l,k)} \left\{ -\sum \mathrm{P}(\phi_{lk}) \log(\mathrm{P}(\phi_{lk})) \right\}$,
$s_1 = C_{\hat{l}_1\hat{k}_1}$, $S = S \cup s_1$
8: **for** $t = 2, \ldots, T$ **do**
9: すでに選択された S と直交性が最大でありエントロピができる限り大きくなるような新しい C_{lk} を選択し S に追加．すなわち
$\{\hat{l}_t, \hat{k}_t\} = \arg\max_{(l,k)\setminus(\hat{l},\hat{k})} \left\{ \frac{1}{||C_{lk}^T S||} + \alpha \left(-\sum \mathrm{P}(\phi_{lk}) \log(\mathrm{P}(\phi_{lk})) \right) \right\}$,
$s_t = C_{\hat{l}_t\hat{k}_t}$, $S = S \cup s_t$.
ただし $(\hat{l}, \hat{k}) = ((\hat{l}_1, \hat{k}_1), (\hat{l}_2, \hat{k}_2), \ldots, (\hat{l}_{t-1}, \hat{k}_{t-1}))$, α は適当なバランスパラメータ.
10: **end for**

等により得られた局所領域を，ユークリッド距離により最も近い Codewords を一つ選択しそのクラスインデックスを特徴量とするのが一般的である．結果的に，例えば二つの Codewords のちょうど中間に位置するような局所領域パターンに関しては，どちらかに強制的に割りつけられることになり，両方に類似しているという情報を表現できない問題があった．こうした符号化誤差を解決するのには，前記中間に位置する局所領域パターンを含むような比較的多くの Codewords を用意しておくしかなく，冗長な Codebook が必要となっていた．

提案手法においては，各 Spatial Codewords と入力画像との類似度により特徴抽出を行う．この類似度を多少の位置ずれや，輝度変化等にロバストに出力するためには単純な距離計算より高度な検出手段が必要となる．このために，各 Spatial Codewords s_t をとらえるための検出器 g_t を学習しておく．この際に，用いる正サンプルは，与えられたラベル無学習データ x から抽出した領域 $x(\hat{l}_t)$ のうち各 Spatial Codewords s_t によりクラスタリングされたクラスメンバのデータ $\{x(\hat{l}_t) \in s_t\}$ である．負サンプルはそれ以外のデータ $\{x(\hat{l}_t) \notin s_t\}$ からランダムに選択する．このようにここでの学習はラベル無データをクラスタリングした結果のクラスタメンバを利用するので教師有学習データは必要ない．

後に詳述する属性追加変更時および属性認識時の入力画像 ξ に対する特徴量 f は，$f_t = g_t(\xi(\hat{l}_t))$ により得られる．ただし，$t = 1, \ldots, T$, $f \in \mathcal{R}^T$ である．後で詳述するが実験においては $T = 50 - 100$ 程度を利用した．このように，ピクセルレベルの特徴量からそれらを集約した中間的な表現である Spatial Codewords を構成し，その Codewords 検出器出力をさらに上位の識別のための特徴量表現とする特徴量抽出方法を，本論文においては Representation by Recognition(RbR) と呼ぶ．図.3 は，図内左側の各入力画像 ξ に対してどのように特徴量表現が行われるかを示している．横軸は特徴量のインデックス t を表わしており，縦軸は各 Spatial Codewords 検出器 $g_t(\xi)$ の出力を表している．各横軸の下に示しているのは，検出器出力が高かった Spatial Codeword である．

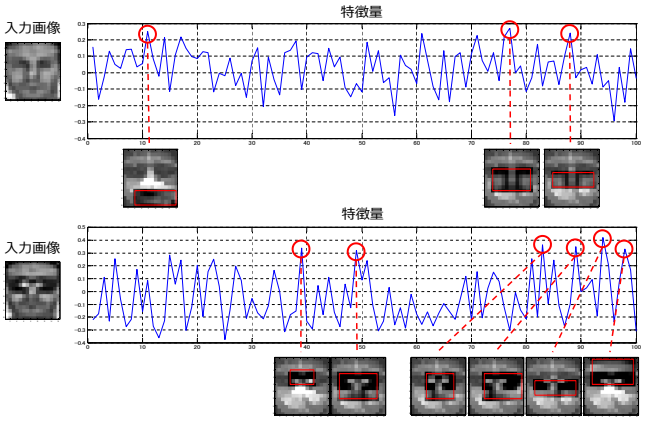

図 3　特徴量表現の一例

2.2　属性の追加変更

属性の追加変更時には，少数のラベル有学習データ y を利用する．前章で示した RbR をラベル有学習データに適用して (すなわち g(y)) 特徴量を抽出し，追加変更する各属性 ω_i を識別するための識別器 H_i を構成する．このために様々な識別器を利用することができるが，本論文では SVM を利用し対応したい属性とそれ以外の属性を識別する「1 vs. All」型の識別器 H_i を訓練した．

Face Tracer のように，ピクセルレベルもしくはそれと同程度の数百から数千次元の特徴量を用いた場合には，過適合を防ぐために，特徴量次元数に見合った数の学習サンプルを用意しなければならない．しかし提案手法においては，属性を表現するのに必要な基本パターンが数十程度に集約しているので RbR による特徴量も 50-100 次元程度となった．従ってラベル有学習データも次元数と同程度の枚数 (100 枚程度) で安定した性能を得ることができた．これにより，比較的少数の事例を集め数秒間の学習をするだけで新しい属性を追加変更ができる (学習時間は実験で詳述).

2.3　人物属性認識

属性の認識時には，識別用の入力画像 z に対して RbR を適用することにより特徴量を抽出し，前章において属性の追加変更時に学習しておいた各属性の識別器 H_i を適用し $\hat{\omega} = \arg\max_{\omega_i} H_i(g(z); \omega_i)$ により属性クラス ω を推定する．

3.　実　　験

本手法の有効性を検証するために，二つの異なる DB(OMRON DB，AR DB) に対して実験を行った．

OMRON DB は，装着物なしの人物，サングラスやマスクを装着した多様な人物が含まれるデータベースである．また多様な照明条件・表情・顔向きの条件が含まれている．含まれる画

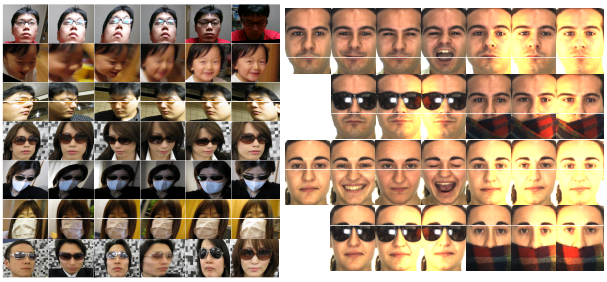

図 4　各 DB 含まれる画像の例: (左)OMRON DB，(右)AR DB

像の例を図.4 に示した．この DB には顔位置等の人物位置を特定するための Ground-Truth が存在しないので OKAO Vision ライブラリ [12] の顔検出 [13](サングラスやマスクを装着した顔でも顔検出可能) および顔器官検出 [14] を用いて自動的に人物位置を特定した．大きな顔向きのデータも顔検出できるがこれらに関しては，アフィン変換により顔と口の位置が一定になるように変換を適用した．この際に器官検出が失敗した場合も同じ正規化手法により正規化したため，歪んだ画像も含まれているが，学習および実験にはそのまま含めた．図 4 において顔向きが大きい時に歪んで見えるのはこのためである．テストでは (1) 装着物なし (2) サングラス (3) マスク (4) およびサングラスとマスクの両方を装着した人物の 4 クラス識別を行った．

一方，AR DB [15], [16] は Martinez らにより撮影された顔認識用 DB である．この DB には，通常の顔のほかサングラスやスカーフを装着した顔，また極端な表情として「叫び (Scream)」等が含まれているので，属性認識等の評価にも用いることができる．顔向きは正面のみであり，安定した照明条件下で撮影されている．一例を図.4 に示した．テストでは (1) 装着物なし (2) サングラス (3) スカーフを装着した顔 (4)「叫び」の，4 クラス識別を行った．なおこの DB では Ground-Truth を用いて正規化をした．

なお，両方の DB において，画像の正規化サイズは，64×75[pix]（両目間のサイズは 28[pix] 程度）とした．

次に提案手法の実装について説明する．Spatial Codebook の学習 (2. 1. 1 章記載) には，大量に用意できるラベル無データを想定し，OMRON DB の場合 2032 枚，AR DB の場合 2600 枚を用いた．提案手法の Codewords 検出器の学習 (2. 1. 2 章記載) において用いる各クラスタのメンバ数は，正サンプル，負サンプルそれぞれに対して 25 枚程度であり，それぞれ Codewords のクラスタメンバおよびそれ以外のデータからランダムに選択して用いた．また両方の DB での実験において，属性の追加変更 (2. 2 章記載) には，各属性 100 枚ずつ合計 400 枚のラベル有データを利用した．

比較手法には Face Tracer を用いた．入力ピクセル表現として，提案されているグレースケール，カラー (RGB, HSV)，エッジ方向・強度の 5 種類のうち，提案手法とできる限り同条件で比較するためグレースケールのみを用いた(注1)．Face Tracer の訓練には，ラベル無データは必要なく，ラベル有のデータのみが必要となる．

(注1)：本来 FaceTracer は 450 種類の特徴量を用いるがここではグレースケールのみのため 90 種類となった．

表 1　属性認識精度

手法	OMRON DB	AR DB
Baseline: Face Tracer	86.04	95.54
PCA + SVM	68.00	93.12
Sparse Coding + SVM	83.02	93.65
NMF + SVM	85.07	95.58
NNSC + SVM	81.52	96.53
Spatial Codebook + SVM	85.61	96.38

提案手法は，教師無学習により得られた局所的基本パターン (Spatial Codebook) を用いて特徴抽出する．これと対立する概念として入力領域全体を用いた教師無学習による基本パターン抽出があるが，局所的基本パターンの効果を確認するため，そうした手法の代表的手法である PCA [17] との比較も行った．さらに顔認識や一般物体認識等で，しばしば Part-based な認識を目的として Non-negative Matrix Factorization(NMF) [18], [19] や，Sparse Coding(SC) [20]〜[22]，さらにそれらを拡張した Non-negative Sparse Coding(NNSC) [23], [24] が提案されているのでそれらとも比較を行った．

比較実験の結果は，表.1 に示した．提案手法は表内に示した”Spatial Codebook + SVM”であるが，FaceTracer と比べ同等以上精度を実現していることが分かる．また顔全体に PCA を適用した結果に比べ提案手法や局所領域に注目する NMF や Sparse coding，NNSC 等の手法の精度が高いことから，局所的基本パターン抽出がこの種の問題に適していることが分かる．一方で局所領域に注目する NMF, SC, NNSC は局所特徴に注目する意味で提案手法と類似であるが，特徴量の計算時に対象とする全領域 (ここでは顔全体) を用いるのに対し提案手法は局所領域のみを明示的に切り出して用いるのでメモリ使用量や処理速度の面で優れている．

以上の結果により提案手法の精度が Face Tracer と同等であることが分かった．しかしながら提案手法の利点は，精度そのものというより非常に多様な属性を簡単に追加変更できるところにある．そこでこの面で有効性を示すために学習時間と学習に必要なデータ数を AR DB において 100 枚のラベル有データを用い調べた．

まず学習時間を計測するに当たり，提案手法では属性追加変更の際に必要となるラベル有データに対する RbR と SVM 訓練に要する処理時間を計測した．Face Tracer は，前記の評価と同様，本来 450 種類の特徴量を抽出するがこのうちカラーとエッジに基づくものを省きグレースケールに基づくもののみを用い，90 種類の特徴量を抽出する時間および各特徴量に対する SVM の学習，adaboost による識別器選択さらに選択された識別器のさらに後段の SVM 学習に必要な処理時間を計測した(注2)(注3)．この条件での学習時間は，表.2 に示した通りである．Face Tracer は 2 時間近く要しているのに対し，提案手法は 10 秒程度で学習が終了しユーザサイドで属性を追加変更す

(注2)：90 個の SVM 全てについてパラメータ最適化するのは困難であるので，予め求めておいた経験的に良いパラメータを用いた．

(注3)：その他の実験条件；CPU：Core 2 Duo 3.00 GHz，メモリ：2GByte，実装言語：MATLAB，SVM の実装には SVM^{light} [25] を使用

表 2　学習に要する時間

手法	学習時間 [sec]
Baseline: Face Tracer	6060 程度
提案手法	11.5

るに現実的な速度になっていることが分かる．FaceTracer の学習時間が長いのは，Adaboost の特徴選択により最終的に選択されない特徴量も含めて 90 種類もの冗長な特徴量抽出を行い，そのすべてに対して SVM の学習を行っていることが，主な原因である．一方ユーザサイドで簡単に属性を追加変更するためには，必要な (ラベル有) データを簡単に用意できることも重要である．これに関して学習に必要なデータ数を調べた結果は表.3 のとおりである．結果を見ると提案手法はラベル有学習

表 3　ラベル有学習データ数による精度変化

学習サンプル数	10	25	50	75	100	200
FaceTracer	85.04	89.35	94.38	95.42	95.54	95.85
提案手法	88.88	92.81	95.54	95.92	96.38	97.23

データが少ない時にも FaceTracer に比べ精度低下が少ないことが分かる．従って，提案手法は属性の追加変更においてあまり多くのデータが用意できない時でもよりよい精度を与えることが実証されたと言える．

4. 結　　論

装着物等による属性による人物特定は，監視カメラにおいて人物の監視を行う際に重要な技術の一つである．特に人物のサイズが小さく顔を鮮明に撮影できないことが多い監視カメラ条件においてその重要性は高くなる．一方，人物属性は非常に多様である故，それら全てに設計時に対応することはほぼ不可能であり，属性の追加変更が簡単に行えることが重要であることから，属性追加変更時に少数のデータを集めるだけで短時間に属性を学習できるフレームワークを提案した．また，このために様々な属性を構成する局所的な基本パターンからなる Spatial Codebook を提案した．一方で，Spatial Codebook を用いた新たな特徴量表現として各 Spatial Codewords 検出器を用いて特徴量表現を行う Representation by Recognition を提案した．これにより従来手法と比べ，飛躍的に短時間に，また少ないラベル有データで新たな属性に対応できるようになった．また精度面でも従来手法と同等以上であることを示した．

謝辞

本研究の一部を支援して頂いた独立行政法人 NEDO 次世代ロボット知能化技術開発プロジェクトに感謝します．

文　　献

[1] B. Moghaddam and M.-H. Yang: “Learning gender with support faces”, IEEE Trans. Pattern Anal. Mach. Intell., **24**, 5, pp. 707–711 (2002).
[2] S. Hosoi, E. Takikawa and M. Kawade: “Ethnicity estimation with facial images”, Proc. of FGR, p. 195 (2004).
[3] J. ichiro Hayashi, H. Koshimizu and S. Hata: “Age and gender estimation based on facial image analysis”, Knowledge-Based Intelligent Information and Engineering Systems, pp. 863–869 (2003).
[4] X. Zhuang, X. Zhou, M. Hasegawa-Johnson and T. Huang: “Face age estimation using patch-based hidden markov model supervectors”, Proc. of ICPR, pp. 1–4 (2008).
[5] M. J. Lyons, J. Budynek, A. Plantey and S. Akamatsu: “Classifying facial attributes using a 2-d gabor wavelet and discriminant analysis”, Proc. of FGR, p. 202 (2000).
[6] T. Wilhelm, H.-J. Bohme, and H.-M. Gross: “Classification of face images for gender, age, facial expression, and identity”, pp. 569–574 (2005).
[7] N. Kumar, P. Belhumeur and S. Nayar: “FaceTracer: A Search Engine for Large Collections of Images with Faces”, Proc. of Europican Conference on Computer Vision (2008).
[8] G. Csurka, C. R. Dance, L. Fan, J. Willamowski and C. Bray: “Visual categorization with bags of keypoints”, ECCV Workshop on Statistical Learning in Computer Vision (2004).
[9] F. Li and P. Perona: “A bayesian hierarchical model for learning natural scene categories”, Proc. of CVPR, pp. 524–531 (2005).
[10] T. Leung and J. Malik: “Representing and recognizing the visual appearance of materials using three-dimensional textons”, IJCV, **43**, pp. 29–44 (2001).
[11] A. Agarwal and B. Triggs: “Hyperfeatures Multilevel Local Coding for Visual Recognition”, INRIA Tech. Report (2007).
[12] OMRON Corporation: “OKAO Vision”, http://www.omron.com/r_d/coretech/vision/okao.html.
[13] C. Huang, H. Ai, Y. Li and S. Lao: “High-performance rotation invariant multiview face detection”, IEEE trans. on Pat. Anal. and Mach. Intel., **29**, 4, pp. 671–686 (2007).
[14] 木下, 小西, 勞, 川出：“3D モデル高速フィッティングによる顔特徴点検出・頭部姿勢推定”, 画像の認識・理解シンポジウム (2008).
[15] A. M. Martinez and R. Benavente: “The AR Face Database”, CVC Tech. Report, p. 202 (1998).
[16] “The AR Face Database”, http://www.ece.osu.edu/ aleix/ARdatabase.html.
[17] M. Turk and A. Pentland: “Eigenfaces for Recognition”, Journal of Cognitive Neuroscience, **3**, 1, pp. 71–96 (1991).
[18] D. Guillamet and J. Vitria: “Classifying Faces with Non-Negative Matrix Factorization”, Proc. of the Catalan Conference for Artificial Intelligence, pp. 24–31 (2002).
[19] D. Lee and H. Seung: “Learning the parts of objects by non-negative matrix factorization”, Nature, **401**, pp. 788–791 (1999).
[20] B. A. Olshausen and D. J. Field: “Natural image statistics and efficient coding”, Workshop on Information Theory and the Brain, pp. 524–531 (1995).
[21] H. Lee, A. Battle, R. Raina and A. Y. Ng: “Efficient sparse coding algorithms”, Proc. of NIPS (2007).
[22] R. Raina, A. Battle, H. Lee, B. Packer and A. Y. Ng: “Self-taught learning: Transfer learning from unlabeled data”, Proc. of ICML (2007).
[23] P. O. Hoyer: “Natural image statistics and efficient coding”, IEEE Workshop on Neural Networks for Signal Processing, pp. 557–565 (2002).
[24] B. J. Shastri and M. D. Levines: “Face recognition using localized features based on non-negative sparse coding”, Machine Vision and Applications, **18**, pp. 107–122 (2007).
[25] T. Joachims: “Learning to Classify Text using Support Vector Machines”, Kluwer (2002).