

拡張類似 Haar フィルタを用いた物体認識

東島 由佳[†], 高野 茂^{††}, 新島 耕一^{††}

[†]九州大学大学院システム情報科学府情報理学専攻

^{††}九州大学大学院システム情報科学部門

1 はじめに

デジタル機器の普及とストレージの大容量化により、ユーザーは膨大なデジタル画像を手軽に取得することができるようになった。それに伴い、それらの画像の中から、高速かつ正確に物体を認識する技術が極めて重要になってきている。これまで多くの画像認識技術が提案されているが、依然として頑健性や認識速度等に問題点が残っている。本研究では、拡張類似 Haar フィルタの高速な学習法と特徴ベクトルの作成法を提案し、顔画像認識へ応用する。拡張類似 Haar フィルタとは、Haar フィルタ $(1, -1)$ を多次元ベクトルに拡張したものであり、要素は 1 または -1 からなる。提案手法では、複数の訓練画像を用意し、各訓練画像を横ベクトルの集合と考え、同じ位置にある横ベクトルと拡張類似 Haar フィルタの内積の分散が最大になるように拡張類似 Haar フィルタを学習する。そして、学習した拡張類似 Haar フィルタを用いて、質問画像と訓練画像の特徴ベクトルを計算し、質問画像に最も類似している訓練画像を選ぶことによって物体認識を行う。

2 拡張類似 Haar フィルタ

Haar フィルタは $(1, -1)^T$ の 2 次元ベクトルの形をしている。提案する拡張類似 Haar フィルタは、Haar フィルタを M 次元ベクトル

$$\mathbf{g} = (g_1, g_2, \dots, g_M)^T$$

に拡張したものである。ただし、 g_j は次の条件を満たす。

$$g_j = 1 \quad \text{or} \quad g_j = -1, \quad (1)$$

$$\sum_{j=1}^M g_j \approx 0. \quad (2)$$

3 拡張類似 Haar フィルタの学習

訓練画像データベースから、拡張類似 Haar フィルタを学習する手法について説明する。まず、正方形に正規化された N 個の訓練画像 $C^\nu, \nu = 1, \dots, N$

のそれぞれを $M \times M$ の正方形ブロックに分割する。次に、 C^ν の各ブロックの画素値の総和を計算し、 $C_{i,j}^\nu$ で表す。そして、 i 行めの $C_{i,j}^\nu$ の並びをベクトル

$$\mathbf{C}_i^\nu = (C_{i,1}^\nu, C_{i,2}^\nu, \dots, C_{i,M}^\nu)^T$$

で表す。訓練画像 C^ν は M 個のベクトルの集合 $\{\mathbf{C}_i^\nu | i = 1, 2, \dots, M\}$ とみなすことができる。

一方、 \mathbf{C}_i^ν に適用する拡張類似 Haar フィルタを

$$\mathbf{g}_i = (g_{i,1}, g_{i,2}, \dots, g_{i,M})^T$$

で表す。提案する学習法は、フィルタ \mathbf{g}_i と訓練画像 \mathbf{C}_i^ν の内積

$$\mathbf{g}_i^T \mathbf{C}_i^\nu = \sum_{j=1}^M g_{i,j} C_{i,j}^\nu$$

の分散

$$\sigma_i^2 = \mathbf{g}_i^T \mathbf{V}_i \mathbf{g}_i \quad (3)$$

が最大となるようにフィルタ \mathbf{g}_i を決めることである。ただし、 \mathbf{V}_i は分散共分散行列を表し、次式で与えられる。

$$\mathbf{V}_i = \frac{1}{N} \sum_{\nu=1}^N (\mathbf{C}_i^\nu - \bar{\mathbf{C}}_i)(\mathbf{C}_i^\nu - \bar{\mathbf{C}}_i)^T. \quad (4)$$

この問題は、 2^M 通りの \mathbf{g}_i の中から (3) を最大にする \mathbf{g}_i を選ぶ組み合わせ最適化問題であり、解くのに膨大な時間がかかる。

本研究では、この問題を実ベクトル空間における最小化問題ととらえ、次の最小化問題を解く。

$$J_i = -\mathbf{g}_i^T \mathbf{V}_i \mathbf{g}_i + \frac{K_1}{2} \left(\sum_{j=1}^M g_{i,j} \right)^2 + \frac{K_2}{4} \sum_{j=1}^M (g_{i,j}^2 - 1)^2 \rightarrow \min. \quad (5)$$

右辺の第 2 項は条件 (2) を考慮したものであり、第 3 項は条件 (1) を考慮したものである。 K_1, K_2 は処罰係数である。この最小化問題は多くの極小点をもちうる。これらの極小点 \mathbf{g}_i を高速に求めるために、 J_i の停留点が満たす方程式

$$\frac{\partial J_i}{\partial g_{i,j}} = 0, \quad j = 1, \dots, M \quad (6)$$

を Newton 法を用いて解く。こうして得られた拡張類似 Haar フィルタを M 個並べ、フィルタパターンを作成する。

4 物体認識

物体認識は次の手順にしたがって行う。

1. N 枚の訓練画像を準備し、前節で述べたアルゴリズムを用いてフィルタ g_i を学習する。そして、訓練画像から作成されたベクトル $C_i^\nu = (C_{i,1}^\nu, C_{i,2}^\nu, \dots, C_{i,M}^\nu)^T$ とフィルタ g_i との内積 $p_i^\nu = g_i^T C_i^\nu$ を計算し、画像の特徴ベクトル $p^\nu = (p_1^\nu, p_2^\nu, \dots, p_M^\nu)^T$ を作成する。
2. 質問画像 C から、ベクトル $C_i = (C_{i,1}, C_{i,2}, \dots, C_{i,M})^T$ を作成し、フィルタ g_i との内積 $q_i = g_i^T C_i$ を計算して、質問画像の特徴ベクトル $q = (q_1, q_2, \dots, q_M)^T$ を作成する。
3. q と p^ν から $\cos \theta^\nu = q^T p^\nu / |q| |p^\nu|$ を計算し、最大となる p^ν を見つけ、 ν 番めの画像を q に最も類似した画像と認識する。

5 実験結果

実験では、4 節のアルゴリズムを用いて顔画像認識を行った。まず、人物が写った質問画像と訓練画像の集合から、スキンカラーセグメンテーション法 [1] を用いて正方形の顔画像を抽出し、 128×128 のサイズに正規化する。次に、各訓練画像を 8×8 サイズの正方形ブロックに分割し、3 節で述べたアルゴリズムを用いて拡張類似 Haar フィルタを学習する。Newton 法の初期値と、処罰係数を変えながら、複数の拡張類似 Haar フィルタを構成した。それらを極小値の小さい順に並べ、5 個のフィルタパターンを作成した。

認識実験では、これらのフィルタパターンを用いて、4 節で述べたアルゴリズムにより、10 枚の訓練画像と質問画像を比較した。その結果を以下に示す。



図 1: 訓練画像





















フィルタパターン	類似度		
	第 1 位	第 2 位	第 3 位
FP1	 0.296697	 0.608558	 0.850232
FP2	 0.223537	 0.491467	 0.957461
FP3	 0.246460	 0.386540	 0.923641
FP4	 0.247386	 0.441208	 0.827966
FP5	 0.313475	 0.434741	 0.828817
総和	 1.327675	 2.362504	 4.806272

図 2: 5 枚のフィルタパターンによる認識結果。白色のセルは 1 を表し、黒色のセルは -1 を表し、それぞれの数字は類似度 $1 - \cos \theta^\nu$ を表す。

6 結論

本研究では、分散最大化法を用いて拡張類似 Haar フィルタを決定し、顔画像認識を高速に行う方法を提案した。しかしながら、Newton 法のアルゴリズムが、初期値や処罰係数の選び方によっては発散する状況が起こる。今後の研究課題は、このような不安定性を取り除くことと、拡大、縮小、回転に頑健な拡張類似 Haar フィルタを構成することである。

参考文献

[1] D.Chai and K.N.Ngan, Locating facial region of a head-and-shoulders color image, Proceedings of the Third International Conference on Automatic Face and Gesture Recognition, pp.124-129,1998.