

隠れに頑健な顔検出法 A Robust Face Detection Method Under Partial Occlusion

堀田 一弘[†]
Kazuhiro Hotta

1. はじめに

画像中から顔を見つけ出す顔検出の研究は世界中で活発に行われ、今日では、制限された環境下においては高い精度で顔を検出できるようになってきている [1, 2]. しかし、実環境下では様々な環境変化があり、コンピュータにとってはまだ難しい。特に隠れがある状況下での顔検出は、隠れる部分が状況によって変わるので、非常に難しい問題である。だが、実環境下ではサングラスやマスク等により顔の一部が隠されることは珍しいことではない。さらに、光源変化により生じる影も隠れの一つと見なすことができるので、隠れに頑健な顔検出が実現できれば、実環境下での顔検出の精度を飛躍的に向上させることができると考えられる。

近年、Support Vector Machine (SVM) の顔検出への有効性が報告されているが、従来手法では画像から抽出した大局的な特徴に対してカーネルを適用していた [1, 2]. 部分的な隠れ等がある場合、大局的な特徴は全体的な影響を受けるので、従来の大局的なカーネルに基づく方法は隠れの影響を受けやすいと考えられる。一方、局所的な特徴に基づく方法では、部分的な隠れが一部の局所特徴にしか影響を与えない。従って、局所特徴をうまく統合することができれば、隠れに頑健になると考えられる。例えば、Martinez は局所的な類似度の重みつき投票をすることで隠れに頑健になることを報告している [3]. これらのことから、SVM に隠れに対する頑健性を付与するためには局所特徴をうまく扱う必要があると考えられる。そこで、本論文では SVM に局所カーネルを導入する。そして、局所カーネルに基づく SVM を用いて隠れに頑健な顔検出を実現する。

提案手法の有効性を示すために、従来の大局カーネルに基づく SVM との比較を行った。その結果、従来の大局カーネルに基づく SVM は隠れがある状況下で精度を劇的に低下させるが、提案手法は高い精度を保つことができることを確認した。また、サングラスやマフラーにより顔の一部が隠されている顔画像からの顔検出が可能であることも確認した。

2. 隠れに頑健な顔検出法

本節では隠れに頑健な顔検出法を説明する。提案手法は局所カーネルに基づいているので、本論文では局所的な見えの情報を抽出できるガボールフィルタを用いる。まず、ガボールフィルタの特性について簡単に説明する。ガボールフィルタは第一次視覚野にある単純型細胞の受容野を良く近似することが知られている。ガボールフィルタは方位選択性を持ち、特定の位置に特定の方向成分がある時だけ大きな出力を出す。得られた特徴はスパースな特徴であり、対象認識に対する有効性が示されてい

る。以下の実験では、4方向のガボールフィルタ(フィルタの大きさ:9×9画素)を用いる。

次に、提案する局所カーネルに基づく SVM を説明する。まず初めに局所カーネルのタイプについて考える。提案手法ではカーネルを認識対象の局所領域に配置するので、各局所カーネルは認識対象の局所領域に特化した細胞と見なすことができる。局所的な特徴を効率的にモデリングし、認識対象の局所領域に特化した細胞を生成するためには Gaussian の反応選択性が有効であると考えられる。そこで、本論文では局所カーネルとして Gaussian カーネルを用いる。局所的な Gaussian カーネルは

$$K_p(\mathbf{x}(p), \mathbf{y}(p)) = \exp\left(-\frac{\|\mathbf{x}(p) - \mathbf{y}(p)\|^2}{\sigma_p^2}\right), \quad (1)$$

となる。ここで、 p は場所のラベルであり、 $\mathbf{x}(p)$ と $\mathbf{y}(p)$ は場所 p を中心とする局所領域の特徴ベクトルを表している。

認識対象の局所領域に配置したカーネルを SVM の枠組で用いるためには、局所カーネルの出力値 $K_p(\mathbf{x}(p), \mathbf{y}(p))$ から1つのカーネル値 $K(\mathbf{x}, \mathbf{y})$ を計算する必要がある。上述のように、局所特徴(カーネル)を如何に統合するかが隠れに対する頑健性を実現するための鍵となる。Mercer の定理を満たす単純な局所カーネルの統合法として和と積が考えられる [4] が、この2つを比べた場合、和の方が隠れに頑健であると考えられる。それは、局所カーネルの積の場合、隠れ等の影響により1つの局所カーネルが小さな値を出すと、全体のカーネル値が小さな値となるからである。つまり、局所カーネルの積は隠れの影響を受けやすいと考えられる。ここで注目すべきは、全ての局所 Gaussian カーネルの分散が一定であり、局所特徴 $\mathbf{x}(p), \mathbf{y}(p)$ が場所 p のスカラー特徴である場合、局所カーネルの積は大局的 Gaussian カーネルに一致するということである。即ち、従来手法で頻繁に利用されている大局的 Gaussian カーネルは局所 Gaussian カーネルの積の特別な場合であり、隠れの影響を受けやすい。一方、局所カーネルの和は、いくつかの局所カーネルが小さな値を出してもその影響を受けにくい。従って、本論文では局所カーネルの和を利用する。局所カーネルの和を用いた SVM は

$$f(\mathbf{x}) = \sum_i^L \alpha_i y_i \frac{1}{N} \sum_p^N K_p(\mathbf{x}_i(p), \mathbf{x}(p)), \quad (2)$$

となる。ここで、 N は局所カーネルの数を表している。式(2)から、局所カーネルの平均値をカーネル値として利用していることが分かる。以下の実験では、提案手法の最も単純な場合として局所特徴 $\mathbf{x}(p), \mathbf{y}(p)$ を場所 p のスカラー特徴とし、分散 σ_p^2 を場所に依らず一定とした。

[†]電気通信大学, The University of Electro-Communications

3. 評価実験

本節では、提案手法の有効性を示すための評価実験を行う。まず、実験で用いるデータベースについて3.1節で説明する。次に、大局カーネルに基づくSVMとの比較結果を3.2節に示す。

3.1 データベース

本論文では、HOIP 顔画像データベース[†]と CMU のテスト用顔画像データベース [6] から切り出した 38×38 画素の顔画像を用いた。また、PIE 顔画像データベース [7] に含まれる光源変化による影の影響の少ない顔画像も利用した。以下の実験では、これらの顔画像を3つに分割し、学習用、パラメタ設定用、評価用として用いた。3つに分割してしまうと各々の画像枚数が少なくなるため、上下左右に1画素ずつ平行移動した顔画像も利用した。また、HOIPの顔画像に関しては鏡面画像も利用した。実験に用いる顔画像の例を図1(a)に示す。様々な環境下で撮影された顔画像が含まれていることが分かる。以下の実験では、これらの顔画像から1画素置きに抽出した $900 (= 15 \times 15 \times 4$ (方向)) 次元のガボール特徴を用いた。一方、顔以外の画像として、PICS データベース [8] や WWW から得た画像からランダム及びブートストラップにより選択した 17,750 枚の画像を用いる。図1(b)に顔以外の画像例を示す。これらの画像を学習用とパラメタ設定用に分割して用いた。

顔検出には、顔以外の対象を顔クラスであると誤識別する False Positive Rate (FPR) と顔領域を正しく顔クラスとして識別する True Positive Rate (TPR) の2つの評価基準がある。本論文では2つの基準を同時に評価するために、Receiver Operating Characteristic (ROC) 曲線 [5] を用いる。以下の実験では、FPRの評価のために pbic データベース [9] から得た 100 枚の顔を含まない画像を用いた。学習した顔検出器を上記 100 枚の1画素置きに適用すると、計 7,670,478 個の顔以外の領域に対して評価することになる。この中から顔クラスとして誤識別される割合を FPR とした。一方、TPRの評価には上述の3番目の顔画像セットを用いた。ただし、ここでは隠れに対する頑健性を評価したいので、評価用の顔画像上に $M \times M$ 画素の白い正方形をランダムに配置し、それらを評価に用いた。以下の実験では、 $M = 0, 5, 10$ の場合について評価を行った。ここで、 $M = 0$ は隠れが全くないことを意味する。 $M = 5$ 及び 10 の白い正方形による隠れを含む顔画像の例を図2に示す。

3.2 大局カーネルに基づくSVMとの比較実験

提案手法の有効性を示すために大局カーネルに基づくSVMとの比較を行う。ここでは、大局カーネルとして Gaussian カーネルと多項式カーネルを用いた。大局的 Gaussian カーネルは局所カーネルの積の特別な場合に一致するので、提案手法との比較は積と和による統合法の比較でもある。また、多項式カーネル $K(x, y) = (1 + \sum_p x(p) \cdot y(p))^d$ は局所特徴の積の和を含んでいるので、大局的 Gaussian カーネルよりも隠れに頑健で

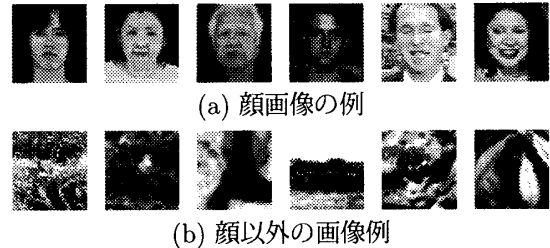


図1: 顔と顔以外の画像例



図2: 隠れを含む顔画像の例

あると期待される。多項式カーネルの d は、パラメタ設定用のデータセットに対して最も高い識別率が得られた2に設定した。比較結果のROC曲線を図3に示す。図の横軸はFPRを表し、縦軸はTPRを表している。TPRが高く、FPRが低いものが良い検出器であるので、図の左上にある曲線が最も精度の高い顔検出器となる。図3(a)は隠れがない状況下でのROC曲線である。FPRが高い場合には大局カーネルに基づくSVMの方が精度が高く、FPRが低い場合には提案手法の方が精度が高いので、この場合には優越の判断がつけにくい。

図3(b),(c)は 5×5 及び 10×10 画素の正方形による隠れがある場合の結果を示している。隠れがある状況では、従来の大局カーネルに基づくSVMは精度を低下させることが分かる。特に、局所カーネルの積の特別な場合に一致する大局的 Gaussian カーネルは隠れの影響を大きく受け、精度を劇的に低下させている。実環境下では、サングラスやマスク等により顔の一部が隠されることは珍しいことではない。従って、従来手法で頻りに利用されている大局的 Gaussian カーネルは実環境下での認識に適さないと言える。一方、提案手法は隠れがある場合でも高い精度を保っている。従って、提案手法は実環境下での顔検出に適した方法であると言える。大局的 Gaussian カーネルとの比較結果から局所カーネルを和で統合することの有効性が実証された。また、大局的多項式カーネルは提案手法には劣るが、大局的 Gaussian カーネルよりも精度が高いことが分かる。これは、上述のように多項式カーネルが局所特徴の積和を含んでいるからであると考えられる。しかし、局所特徴を Gaussian で表現している提案手法に比べ、多項式カーネルでは局所特徴の表現が不十分であるので、提案手法よりも精度が低くなったと考えられる。多項式カーネルとの比較から局所特徴の表現方法も重要であることが分かる。これら結果から、局所的に Gaussian カーネルを配置し、その和で統合する提案手法の有効性が実証された。

最後に、提案手法による顔検出の結果を示す。提案手法の有効性を示すために、AR 顔画像データベース [10] に含まれるサングラスやマフラーにより顔の一部が隠さ

[†]HOIP 顔画像データは財団法人ソフトピアジャパン 研究開発部 地域結集型共同研究推進室から使用許諾を受けたものです。権利者に無断で複写、利用、配布等を行うことは禁じられています。

れている顔画像を用いた。結果を図4の上段に示す。この結果から提案手法が実際の隠れに対しても頑健であることが分かる。また、Yale顔画像データベースB [11]に含まれる影のある顔画像からの検出例も示す。光源変化により生じる影も隠れの一様であるので、隠れに頑健である提案手法により正しく検出されることが分かる。また、隠れを含む複数人が写った画像からも正しく顔を検出できることが分かる。これらの実験により、提案手法が隠れに頑健であることが実証された。

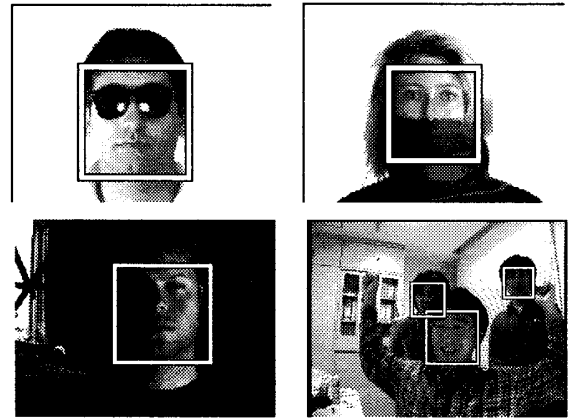
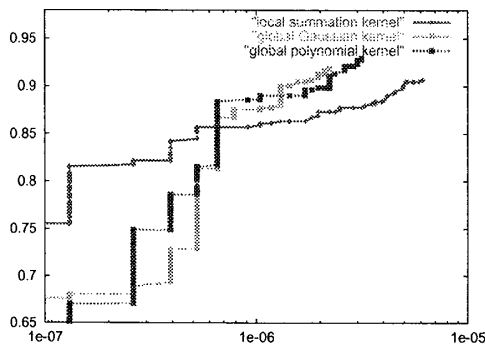
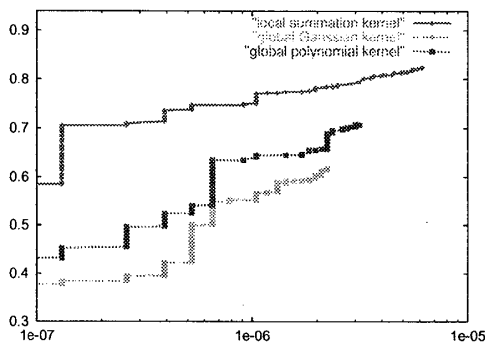


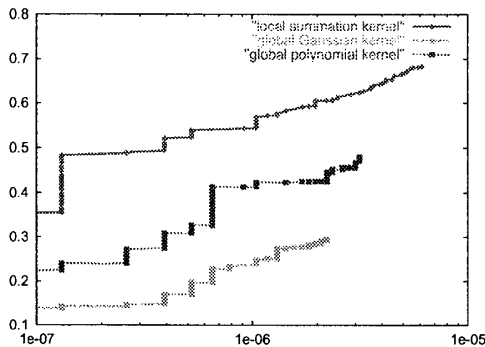
図4: 顔検出結果の例



(a) 隠れがない場合



(b) 5 × 5 画素の正方形による隠れがある場合



(c) 10 × 10 画素の正方形による隠れがある場合

図3: 比較実験の結果

4. おわりに

本論文では、局所カーネルの和に基づくSVMを提案し、それを用いて隠れに頑健な顔検出を実現した。提

案手法の有効性及び隠れに対する頑健性は、大局的なGaussianカーネルや多項式カーネルとの比較により実証された。顔に隠れがある場合、大局カーネルに基づくSVMではその精度を劇的に低下させたが、提案手法では高い精度を保持した。さらに、実環境下で起こりうるサングラスやマフラー等による隠れや光源変化により生じる影に対する頑健性も確認できた。

参考文献

- [1] E.Hjelmas and B.K.Low, "Face detection: A survey," *Computer Vision and Image Understanding*, Vol.83, No.2, pp. 236-274, 2001.
- [2] M.-H.Yang, D.Kriegman, and N.Ahuja, "Detecting faces in images: A survey," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.24, No.1, pp. 34-58, 2002.
- [3] A.M.Martinez, "Recognizing Imprecisely Localized, Partially Occluded, and Expression Variant Faces from a Single Sample per Class," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.24, No.6, pp.748-763, 2002.
- [4] D.Haussler, "Convolution Kernels on Discrete Structures," UCSC-CRL-99-10, 1999.
- [5] B.Heisele and T.Serre and M.Pontil and T.Poggio, "Component-based Face Detection," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 657-662, 2001.
- [6] H.A.Rowley, S.Baluja, and T.Kanade, "Neural network-based face detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.20, No.1, pp. 23-38, 1998.
- [7] T.Sim, S.Baker, and M.Bsat. The cmu pose, illumination, and expression (pie) database. In *Proc. fifth IEEE International Conference on Automatic Face Gesture Recognition*, pages 53-58, 2002.
- [8] *The Psychological Image Collection at Stirling University*. <http://pics.psych.stir.ac.uk/>.
- [9] *Pedestrian and Bicycle Information Center Image Library*. <http://www.pedebikeimages.org/> Dan Burden.
- [10] A.M.Martinez and R.Benavente, "The AR face database," *CVC Tech.Rep.*24, 1998.
- [11] A.S.Georghiadis, P.N.Belhumeur, and D.J.Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.23, No.6, pp. 643-660, 2001.