

ベイズ的動画顔検出における顔候補領域の逐次予測

Sequential Importance Search for Bayesian Face Detection in Video

松井 淳†, ‡
Atsushi Matsuiサイモン クリッピングデル†
Simon Clippingdale松本 隆‡
Takashi Matsumoto

1. まえがき

顔画像検出の問題は、顔とその他の画像を正しく判別する関数(判別器)を事前に「学習」する段階と、入力画像上の様々な位置・大きさの顔領域の候補を逐一検証する「探索」の段階の2つに分けて考えることができる。顔画像検出の処理時間は、探索する候補領域の個数に比例して増大するが、入力画像が動画である場合、その連続性を利用して、探索する顔候補領域を絞り込むことが可能である。しかし、実際の問題において与えられる動画は、放送映像の編集点など、連続性が破綻する箇所が不規則に含まれる場合が多いため、それら不連続点を検出する機構が別途、必要であった。

本稿では、次の時刻の顔候補領域に付随したパラメータの確率分布を逐次モンテカルロ法を用いて予測し、予測確率が低い候補領域を探索空間から削除することによって、顔検出処理を高速化する手法を提案する。さらに、推定した予測確率分布から導かれる周辺尤度を用いて、入力映像の不連続性の判定、および、検出した不連続点における予測確率分布の再初期化を自動的におこなう手法を提案する。

2. 従来手法

Violaら[1]は、多数の矩形差分特徴[2]で構成した判別器を多段接続することで、高速かつ高精度な顔検出器が実現であることを示した。Lienhartら[3]は、用いる特徴量の拡張による検出精度の改善の余地を示した。これらの手法では、顔判別器は T 個の判別機の線形結合で表現される。

$$H(\mathbf{u}, D) = \sum_{j=1}^T \alpha_j h_j(\mathbf{u}, D) \quad (1)$$

$$h_j(\mathbf{u}, D) = \begin{cases} 1 & \text{if } p_j f_j(\mathbf{u}, D) < p_j \theta_j \\ -1 & \text{otherwise} \end{cases} \quad (2)$$

ここで、 $\mathbf{u} = \{x, y, r\}$ は顔領域の中心位置 (x, y) と大きさ r をあらわし、 D は与えられた画像をあらわす。また、各々の判別機は、画像 D の部分領域 \mathbf{u} について計算した特徴量 f_j 、閾値 θ_j 、および、式(2)の不等号の向きを決定するパリティ p_j から構成される。 $H(\mathbf{u}, D)$ の各構成要素は AdaBoost アルゴリズム[4]を用いて学習する。

未知の画像 D が与えられたとき、 $H(\mathbf{u}, D)$ の出力値が正となる領域 \mathbf{u} を探索し顔の候補とする。顔候補の探索は、多くの場合、与えられた各々の画像について一様かつ独立に行われる。 $H(\mathbf{u}, D) > 0$ なる顔候補は、正しい顔領域の近傍に空間的に重複して検出され、かつ、顔が存在しない領域において孤立して検出される傾向がある。そこで、検出した顔候補全体を一度クラスタリングし、各々のグループ G_j

に属する要素の数が閾値 η 以上となった場合、 G_j を顔として受理する処理を施す。最終的な顔領域は、受理されたグループ G_j に属する \mathbf{u} の平均値として与えられる。

3. 提案手法

3.1 逐次モンテカルロ

一連の画像 $D_{1:t} = \{D_1, D_2, \dots, D_t\}$ と、各々の時刻における顔領域パラメータの系列 $\mathbf{u}_{0:t} = \{\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_t\}$ について考える。 $\mathbf{u}_{0:t}$ を確率変数として捉えたとき、動画を処理対象とした顔画像検出の問題は、パラメータ系列 $\mathbf{u}_{0:t}$ の確率分布を与えられた画像列 $D_{1:t}$ から推定する問題として定式化することができる。時刻 t におけるパラメータ \mathbf{u}_t と画像 D_t との関係が確率分布関数のかたちで表現可能であるとき、 $\mathbf{u}_{0:t}$ の事後確率分布は、Bayes 公式より自然なかたちで導かれる。

本稿では、Platt[5]の手法に習い、 \mathbf{u}_t に対する D_t の尤度関係を次式で定義する：

$$P(D_t | \mathbf{u}_t) = \frac{1}{1 + \exp\{-\beta H(\mathbf{u}_t, D_t) + \gamma\}} \quad (3)$$

パラメータ β, γ の設定値は、顔・非顔のサンプルから最尤推定により求まる[5]。

\mathbf{u}_t の確率分布が過去のパラメータ系列と画像列から予想可能な場合、 $\mathbf{u}_{0:t}$ の事後確率分布は、Bayes 公式より、次の逐次更新式で与えられる。

$$P(\mathbf{u}_{0:t} | D_{1:t}) = \frac{P(D_t | \mathbf{u}_t) P(\mathbf{u}_t | \mathbf{u}_{t-1}, D_{1:t-1})}{P(D_t | D_{1:t-1})} P(\mathbf{u}_{0:t-1} | D_{1:t-1}) \quad (4)$$

ただし、 $P(D_1 | D_0) = P(D_1) = 1$ とする。

$\mathbf{u}_{0:t}$ に関するある確率分布 π から N 個の独立なサンプルが抽出可能なとき、顔候補の系列 $\mathbf{u}_{1:t}$ の結合事後確率分布は次式で近似される[7]。

$$\hat{P}(\mathbf{u}_{0:t} | D_{1:t}) = \sum_{i=1}^N \tilde{w}_{0:t}^{(i)} \delta(\|\mathbf{u}_{0:t} - \mathbf{u}_{0:t}^{(i)}\|) \quad (5)$$

$$\tilde{w}_{0:t}^{(i)} = \frac{w_{0:t}^{(i)}}{\sum_{j=1}^N w_{0:t}^{(j)}} \quad (6)$$

$$w_{0:t}^{(i)} = \frac{P(D_t | \mathbf{u}_{0:t}^{(i)}) P(\mathbf{u}_{0:t}^{(i)} | D_{1:t-1})}{\pi(\mathbf{u}_{0:t}^{(i)} | D_{1:t-1})}, \quad i = 1, \dots, N \quad (7)$$

このとき、顔領域の推定値は事後確率分布を用いた期待値で定義される。ただし、尤度が小さいサンプルの寄与はほとんど無視できるとみなし、尤度がある閾値 ϵ_1 以上であるサンプルについてのみ G_j のクラスタリングと式(9)右辺の和を実行し、計算コストの削減を図る。

$$\hat{\mathbf{u}}_t^{[j]} = \frac{1}{P(D_t | D_{1:t-1})_{G_j}} \int_{\mathbf{u}_t \in G_j} \mathbf{u}_t P(\mathbf{u}_t | D_{1:t}) d\mathbf{u}_t$$

$$\cong \frac{\sum_{\mathbf{u}_t^{(i)} \in G_j} \mathbf{u}_t^{(i)} \tilde{w}_t^{(i)}}{\sum_{\mathbf{u}_t^{(i)} \in G_j} \tilde{w}_t^{(i)}} \quad (8)$$

† NHK 放送技術研究所 (人間・情報)

‡ 早稲田大学理工学術院

3.2 変化点検出

式(8)の分母に登場する規格化定数 $P(D_i|D_{1:t-1})_{G_j}$ は、過去の画像列 $D_{1:t-1}$ に対する現画像 D_t の整合性に関する一種の統計的指標（周辺尤度）である。 $D_{1:t}$ の背後に潜むダイナミクスがある時刻において不連続な挙動を示すとき、周辺尤度の値はそれに対応して瞬間的に減少する[8]。本稿では、この逐次周辺尤度を用いて入力映像の不連続点を判定し、検出した不連続点においてパラメータ u_t の確率分布を一様分布で再初期化する。一つの画像に複数の顔が同時に出現しうする場合、それぞれが対応するグループ G_j ごとに不連続性の判定と再初期化を実行する必要があるが、本稿では処理の簡略化のため、全てのグループについての周辺尤度の値が閾値 ε_2 を下回った場合に、パラメータ u_t の確率分布全体を一様分布でおきかえる。

$$\pi(u_t | u_{0:t-1}^{(i)}, D_{1:t-1}) = \begin{cases} N(u_{t-1}^{(i)}; \Sigma) & \text{if } \exists G_j; P(D_t | D_{1:t-1})_{G_j} > \varepsilon_2, \\ 1/|\Omega| & \text{otherwise} \end{cases} \quad (9)$$

4. 実験

提案手法（以下、Proposedと呼ぶ）の有効性を検証するため、OpenCV（以下、Baselineと呼ぶ）との比較実験を行った。評価データには、TRECVID2007 development data [9]から抽出した50ショットを編集し、49カ所の不連続点を含む計500フレームの動画像を用意した。評価データに出現した正面顔の領域を手動で抽出した結果を正解データとし、位置と大きさについて10%の誤差を許容する条件で各々の顔検出結果を評価した。SMCの粒子の総数は500個とした。なお、尤度関数のパラメータの値は、NHKニュース映像から抽出した顔のサンプル画像25,000枚、非顔のサンプル画像50,000枚にもとづく最尤推定値（ $\beta = -3.14$, $\gamma = 3.21$ ）に設定した。

評価データに対して、各々の手法を様々な閾値で実行した結果を図1・図2に示す。また、それぞれの結果のうち、F値が最大となった設定における再現率、適合率、F値、処理時間をまとめ表1に示す。図2・表1より、提案手法は従来手法とほぼ同じ検出精度を約2.5倍の処理速度で実現できていることがわかる。なお、表1に示した提案手法の検出処理の履歴を詳細に調べたところ、49個存在する不連続点（編集点）の全ての箇所ですべて再初期化が行われたことがわかった。

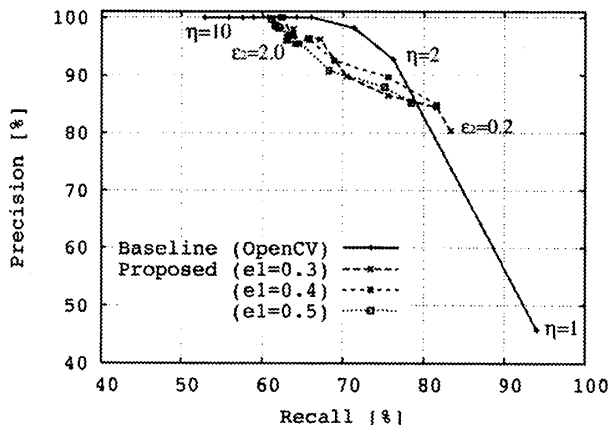


図1：従来手法と提案手法の顔検出結果

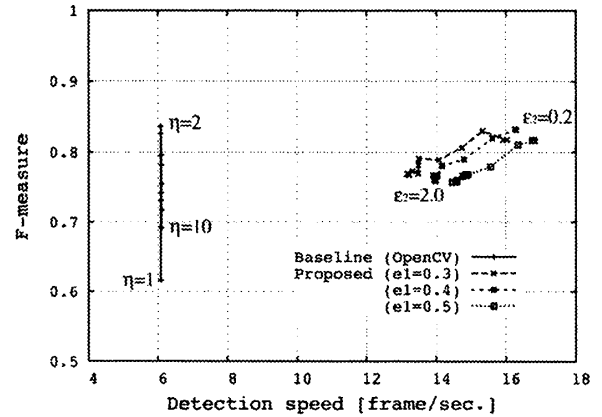


図2：従来手法と提案手法の顔検出結果

表1：顔検出実験の結果（F値の最高値）

| | Baseline ($\eta = 2$) | Proposed ($\varepsilon_1 = \varepsilon_2 = 0.4$) |
|------|----------------------------|---|
| 再現率 | 76.2 % | 81.6 % |
| 適合率 | 92.7 % | 85.0 % |
| F値 | 0.84 | 0.83 |
| 処理速度 | 6.1 fps | 16.3 fps |

5. まとめ

動画像を対象とした顔検出の問題において、観測データの時間的連続性を反映した逐次周辺尤度を導出し、シーン境界の自動検知と確率分布の自動再初期化を実現した。放送映像から抽出した評価データを用いた比較実験の結果、各フレームを独立かつ一様に走査した場合と比べて同等の顔検出精度を保ったまま処理速度を約2.5倍に改善した。

今後は、シーン変化点直後の確率分布の設計、複数の顔への対応、新規に出現した顔の検出方法、および、様々な顔向きへの対応などを検討したい。

文献

- [1] P.Viola and M.Jones, "Robust Object Detection using a Boosted Cascade of Simple Features," Proc. CVPR2001, pp.119-124 (2006).
- [2] C.Papageorgiou, M.Oren, and T.Poggio, "A general framework for object detection," Proc. ICCV'98, pp.555-562 (1998).
- [3] R.Lienhart and J.Maydt, "An extended set of Haar-like features for rapid object detection," Proc. ICIP2002, vol.1, pp.900-903(2002).
- [4] Y.Freund and R.E.Scaphire, "A decisiontheoretic generalization of on-line learning and application to Boosting," Journal of Computer and System Sciences, vol.55(1), pp.119-139 (1997).
- [5] J.Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," In *Advances in Large Margin Classifiers*, pp.61-74, MIT Press (1999).
- [6] Intel Corporation "Open Source Computer Vision Library" <http://www.intel.com/technology/computing/opencv/> (2006).
- [7] A.Doucet, "On sequential simulation-based methods for Bayesian filtering," CUED/F-INFENG/TR-03-06 (1998).
- [8] T.Matsumoto, "Marginal likelihood change detection: particle filter approach," Proc. MaxEnt2005, vol.803, pp.129-136 (2005).
- [9] National Institute of Standards and Technology, The TRECVID 2007 evaluation, <http://www-nlpir.nist.gov/projects/tv2007/tv2007.html> (2007).