

対象検出問題における注意の伝播に関する研究

坂平 星弘[†] 和田 俊和[†]

[†]和歌山大学 システム工学部

〒640-8510 和歌山県 和歌山市 栄谷 930

E-mail : s_saka,twada@vrl.sys.wakayama-u.ac.jp

あらまし画像からの対象検出技術は、追跡、認識などの前提となる基本的処理である。従来、対象の検出は静止画像内の探索によって行われ、時間軸方向の情報は用いられていなかった。しかし、人間が対象検出を行う場合、一度調べて何も発見できなかった部分は再度調べることは殆どなく、直前に対象が検出された部分や新たに観測された部分については重点的に対象検出を行うというように、対象検出においても過去の情報を用いていると考えられる。本研究では対象検出の新たな手法として、このような注意の伝播メカニズムに着目し、画像内での対象検出を重点的に行う部分とそうでない部分をメンテナンスするアルゴリズムを提案するとともに、実験を通じて顔検出の効率化が行えることを示す。

キーワード AdaBoost, カスケード型識別器, Haar 特徴

Attention Propagation for Object Detection.

Seihiro Sakahira[†], Toshikazu Wada[†]

[†]Graduate School of Systems Engineering, Wakayama University,

930 Sakaedani, Wakayama, 640-8510, Japan

E-mail : [†]{s_saka,twada}@vrl.sys.wakayama-u.ac.jp

Abstract Visual object detection is one of the most basic technology for object tracking and recognition. Existing methods performs detection for independent image frames. In the case of human vision system, however, exploits the previous detection results, i.e., detection will not be performed within the regions where no object has been detected in the previous image frame, and will be performed within other regions. This report proposes a detection framework that exploits previous detection results by propagating the potential values representing how much shall the detection pay attention. The algorithm for maintaining this potential values and for detecting the objects by referring the potential values are proposed and examined through the face detection experiments.

Keywords AdaBoost, Cascade classifiers, Haar features

1 はじめに

画像内から対象を検出する技術は、追跡、認識の前処理として行われる基本的処理である。従来、対象検出は静止画像内で行われる処理であり、時間軸方向の情報は用いられない。しかし、人間が対象を検出しようとする場合、一度見たところや観察したところは再度見ることは無く、まだ見ていないところを無意識に探し、より重要だと思われる場所を重点的に見ようとする。

本研究ではこのような人間の視覚特性を考えた

対象検出法を提案する。人間の目は映像中の全ての領域を探索して、対象を見つけるということはしておらず、どこが重要か、どこを見れば良いのかを無意識に決めている。このような特性は、決して静止画像内の特徴のみで実現されるものではなく、時間軸方向の情報も用いていると考えるのが自然である。従来考えられてきた対象検出は静止画像内だけの問題であったが、本研究では、対象検出を動画処理の問題としてとらえ、あるフレームで探索した場所よりも未探索の場所や過去に対象が検出された場所を重点的に探索するというアルゴ

リズムを提案する。

この実現のためには、どこを見たのか（探索したのか）という情報を時間的に連続するフレーム間に伝播させることが必要になる。本研究ではこの方法として、オプティカルフローを用いた伝播を行う。オプティカルフローは本来物体の移動ベクトルを求める手法であるが、その物体の移動に従って探索した範囲を伝播させるために利用する。

時間軸方向に探索済み領域の情報を伝播させる場合、わずかのオプティカルフローの誤差が徐々に拡大され、大きな誤差となることが考えられる。このため、探索済み領域の情報は時間的に減衰させるように取り扱う。このような注意の払い方は、人間も同様に行っている。これは、人間はかなり過去の情報を信じることはなく、一度注意を払った部分についても一定時間が経過すると再び注意を払い見に行くということをしばしば行うことから也正当化することができる。

以上の議論に従い本研究では、一度見た部分を再び見ないという固い基準ではなく、一度見た部分が対象であったか否かを確率的に表現し、その確率ポテンシャルをオプティカルフローで伝播させていくという方式を採用する。

確率として表現すべき本質的事象は、「対象である」「対象ではない」「未観測」の3つである。しかし、具体的な計算に即した別の見方をすれば、「注意を払うべきか」「注意を払わないべきか」の2事象に分けることができ、この立場から見ると、一つの確率値で表現することが可能になる。

この立場から見ると、ある対象検出機構を用いて、部分画像の検出を行った結果、

1. 対象と判定されなかった場合、その部分に注意を払う確率を下げる。
2. 対象と判定された場合、注意を払う確率を上げる。

という計算を行えば良いことになる。またオプティカルフロー計算を行った結果、

3. 未観測の領域が現れた場合注意を払うべき確率を1とする。

という計算により注意を払うべき確率マップの更新が行える。この確率マップを用い、確率値の高い部分を重点的に調べるだけで、効率的に対象検出が行える。本研究では、さらに「対象である」確率を表すマップも導入し、その確率値に応じて動的に対象検出機構の閾値を調整するという方式を提案する。

以下2章では本研究で取り扱う顔検出の基本手法である Viola らの手法について簡単に説明し、3章で提案手法の詳しい説明を行い、4章の実験を通じて、従来手法に比べて識別回数を削減出来ることを示す。

2 従来研究

本章では、顔検出技術として注目されている Viola らの顔検出手法 [1] について簡単に説明する。彼等が提案した手法が高速かつ識別精度が高いのは、AdaBoost [2] によって学習された識別面をカスケード構造で接続し、識別の早い段階で非顔と判断する Early Reject 機能や、Integral Image による Haar 特徴の計算を高速に行っているからである。ここでは、本研究と関係が深いカスケード構造と実際の検出過程の流れについて説明する。

2.1 カスケード型識別器

カスケード型識別器とは、AdaBoost で学習した強識別器を図1のようにカスケード構造で接続したものである。このような構造をしている識別器に、画像からさまざまな大きさで矩形領域を切り出し、入力として識別を行う。画像中にある明らかに顔ではない入力を与えられた場合、識別器の早い段階で非顔と判定し、識別を終了する。これを Early Reject という。また、顔の部分は複数の識別器を通すことによって高い識別精度を保持しながら高速に検出を行うことが出来る。

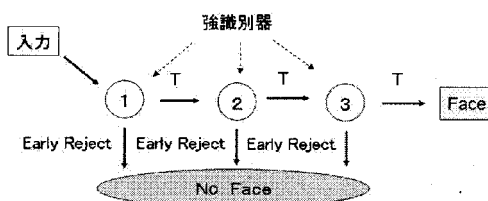


図1: カスケード型識別器

2.2 従来の検出法

実際の顔検出法について述べる。識別器に与える入力サイズは一定でなければならない。様々なスケールに対応させるため、従来の識別手法では、画像が与えられると、最初に画像のスケールを変化させて対応している（図2）。

スケールを変化させた画像全てに対して全走査を行い、矩形を切り出す。この方法で選出された矩形を、識別器へ与える入力とする。各強識別器を構成する弱識別器それぞれが出力する値に重みを乗じた結果が、顔か非顔かの評価基準になり、強識別器それぞれに設けられた閾値よりも低ければ非顔、高ければ顔と判定する。

3 注意の伝播による顔検出

本章では、提案手法である注意の伝播による対象検出について説明する。初めに注意すべき場所



図 2: ピラミッド構造でのスキャン

について論述し、次にその注意すべき場所の伝播を述べ、最後に数フレームを用いることによって得られる利点を述べる。

3.1 注意すべき場所の決定

先行研究として我々は Genetic Algorithm(以下 GA) による対象検出法について提案し、その中で個体生成確率変動法を提案している [3]. これは GA の交叉による個体発生の過程において、過去の世代で発生させた場所には個体を発生させ難くする手法である。

このために画像上の各座標に個体発生確率を与え、これを更新しながら GA の個体発生をコントロールする (図 3). これにより何度も同じ場所の探索を防ぐことが出来る。

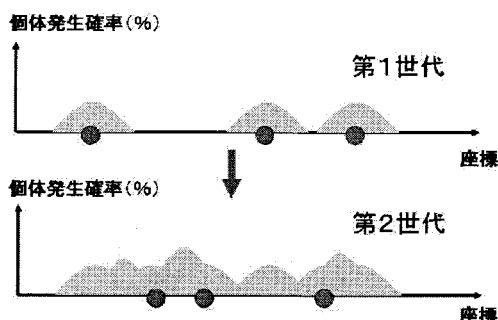


図 3: 個体発生確率の一例

この手法は従来の静止画像内での対象検出法よりも人間が行う対象検出に近いと考えられる。

この手法の基本となる考え方は一度探索したところは次回以降、探索し難くするものである。人間が行う注意すべき点は、まだ見ていないところと、何らかのイベントがある場所であると考えられる。換言すると既に見たところ (注意を向けたところ) は次の瞬間はあまり見ることが無い (注意を払わない) ということである。この考えは個体発生確率変動法と似ていると言える。

今回の報告では GA による対象検出ではなく Viola らが提案した顔検出手法に注意のメカニズムを組み込む方法を提案する。従来の検出手法では、ある画像が与えられた場合、解像度を段階的に落としたピラミッド画像を生成し、この内部から切り出した固定サイズのブロック全てについて顔の識別が行われる (図 2)。これによって、場所と大きさに依存しない顔検出が行える。

この手法を動画内で適用することを考えた場合、毎フレーム全探索を行わなければならない、非常に手間がかかる。実際の人間の視覚による検出を考えた場合、過去に識別して人の顔ではないと判断したところは殆ど見ることがない。そこで識別器にかける場所とそうでないところを分けることを考える。

図 4 は、注意すべき場所を示している (簡単のため注意確率 (後述) は考えず、どこを注視すべきかだけを示す)。過去に顔であると判断した場所や、画像内へと入ってくる物体に対して注意を向けている。このような場所に注意を向け、識別器にかけるなければならない。



図 4: 注意すべき場所例

3.2 注意の伝播

3.2.1 伝播の概要

実際に注意すべき場所を決定した後に (注意しなくても良い場所も含めて) 次フレームへ伝播させなければならないが、単純に伝播させるだけでは上手くいかない。例えば、対象が人の顔であったならば、当然人が動くことが考えられる。これにあわせて、注意すべき場所も動く必要がある。また、背景が変わると、変化した背景にも注意を向ける必要がある。

これらのことを考慮し本研究では注意の伝播方法にオプティカルフロー (図 5) を用いることを提案する。

前フレームでは存在しなかった新しい背景、もしくは物体の出現に対しての検出には画像の外枠からのフローがある場所を見れば良い。つまり、フローが画像外枠から発生している部分に関しては、新しい部分と判断し、探索を行う。これによって、



図 5: オプティカルフロー図

新規に入ってきた物体に対しても注意を向けさせることが出来る。

3.2.2 注意確率の定義

注意する場所、しなくて良い場所を伝播させるが、人間の視覚による検出を考えた場合、たとえ一度見た場所もある程度時間が経てば再度、探索することは十分行われている。このような人間の視覚動作は、一度探索したところは二度と見ない、という制約では実現が難しい。どのタイミングで再探索を行えばよいのか定義し難いからである。そこで、注意を払う確率を考える。

仮に一度探索した場所は注意確率を 0 % と考え、次のフレームでは探索しない。新しく入ってきた部分については注意確率を 100 % と考え、次のフレームでは確実に探索する。注意確率 50 % ならば $1/2$ の確率で探索する。

図 7 は注意する確率を示している。色が濃い程注意確率が高く薄ければ注意確率が低い。

一度探索した場所でも対象が検出された場合は注意確率を 100 % にすることで、次のフレームでも注意を払うことが出来る。具体的なイメージを図 6 と図 7 で示す。前フレームで対象を検出した場所、画像外から物体が入ってきたところ、そしてカメラの移動などによって新たな背景が現れた場所などの注意確率が高くなる。

こうして定義した注意確率を時間と共に伝播させるたびに、確率を増大させ再び見に行くという動作を実現する。

以上の考えをまとめたアルゴリズムを以下に示す。

注意確率伝播アルゴリズム

現時刻 t と $t-1$ で得られた画像 I_t, I_{t-1} を用いて計算されるオプティカルフローで座標 (x_{t-1}, y_{t-1}) が座標 (x_t, y_t) へ動いたとする。以下のアルゴリズム内で、 $I_t.width$ は画像 I_t の width, $I_t.height$ は画像 I_t の height を表し、 $AP_t(s, x_t, y_t)$ ($0 \leq AP_t(s, x_t, y_t) \leq 1$) をスケール s , 時刻 t , 座標 (x_t, y_t) における注意確率とする。

1. if $x_{t-1} \leq 0$ または $I_t.width \leq x_{t-1}$ または $y_{t-1} \leq 0$ または $I_t.height \leq y_{t-1}$ ならば

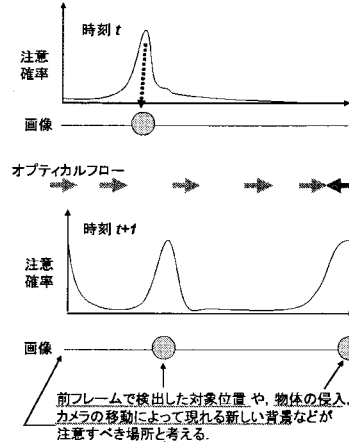


図 6: 注意確率



図 7: 注意確率 (色が濃い程、注意確率高)

- $AP_t(s, x_t, y_t) = 1$

2. else

- $AP_t(s, x_t, y_t) = AP_{t-1}(s, x_{t-1}, y_{t-1})w_e$

w_e は重み (伝播していくたびに、注意確率を増大させるため)。

3.3 強識別器の閾値調整

これまで述べてきたように注意をどこに向けるか、という考えを用いれば画像を全走査、かつ全スケール識別する必要が無く、識別回数を削減出来る。

更に、本研究では過去の識別結果を保持している利点を用い、カスケード識別器の精度を上げられることを示す。

本手法では、過去に顔を検出したところが識別器のミスによって現在フレームで検出しなくなった場合や、過去で顔検出しなかったところが誤検出によって現在フレームで検出してしまった場合等の状況を観測することが出来る。こういった情報を元に識別器の結果に修正を行う。具体的には顔か非顔か判定する閾値を調整することを考える。

3.3.1 対象確率の定義

はじめに対象であるかそうでないかという確率を考える。

これはある部分画像に対して対象と判断した場合には、その部分の対象確率を上げ、逆に非対象と判断した場合は対象確率を下げる。例えばある部分が常に対象であった場合の対象確率は 100 % であり、どれだけ時間が経っても対象が無ければ対象確率は 0 % である。このように定義したものを対象確率 $fp_{(s,x,y)}$ ($0 \leq fp_{(s,x,y)} \leq 1$) と定義する (ここで s は識別枠のスケールサイズ, x, y は画像上での座標)。

この対象確率は注意確率と同じようにオプティカルフローによって伝播させる。対象確率はあくまで次で述べる識別器の、顔か非顔か判定する閾値を修正する時に用いるだけであり、実際顔か非顔かは、対象確率に関係なく、識別器の出力に従う。

3.3.2 閾値更新

カスケード識別器は各強識別器の閾値に従い、入力顔か非顔か判定するが、閾値設定が適切でないとい顔検出が上手くいかない。

ここでは、その閾値を調整することによってより適切な閾値を設定することが出来ることを示す。

閾値更新は対象確率に従ってそれぞれ設定した重みを閾値に足しこんでいくことにより実現する。例えば対象確率が 100 % の位置で非顔と判定された場合には、閾値を大幅に下げて顔と判定し易くし、逆に対象確率が 0 % の位置で顔と判定された場合は閾値を大幅に上げて非顔と判定し易くすることにより検出精度を上げる。

以下に閾値更新のアルゴリズムを示す。

閾値更新アルゴリズム

座標 (x, y) からスケール s で切り出された入力画像 $N_{(s,x,y)}$ を強識別器 $G_t (t = 0, \dots, n)$ で構成されるカスケード識別器で識別する。なお、 u 番目の強識別器の出力結果を $OutTH_u$ と表記し、強識別器の閾値を TH_u とする。

1. スケール s , 座標 (x, y) での対象確率 $fp_{(s,x,y)}$ を計算。
2. その地点での強識別器の結果が non-face or face
 - non-face ならば non-face と判定した強識別器の閾値 TH_t を

$$TH_t = TH_t - fp_{(s,x,y)} \cdot w \quad (1)$$

- face ならば

$$\arg \min_{t=0, \dots, n} |TH_t - OutTH_t| = a \quad (2)$$

$$TH_a = TH_a + (1 - fp_{(s,x,y)}) \cdot w \quad (3)$$

このアルゴリズム内における w は重みであり、手動で設定している。non-face の場合は非顔と判定した強識別器の閾値だけを下げ、face の場合は出力と閾値の絶対値が一番小さい強識別器の閾値だけ上げている。

以上の全体的な流れをフローチャートで図 8 に示す。顔検出フェイズに入る前処理としてオプティカルフローによる移動ベクトル計算とそれに伴って注意確率と対象確率の伝播を行う。顔検出の処理を終えた後に、それらの注意確率と対象確率の更新を行って 1 フレーム内の処理を終える。

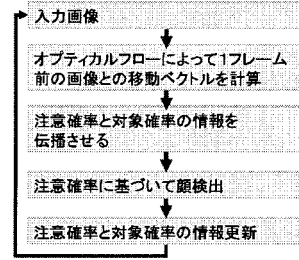


図 8: 処理フローチャート

4 実験結果

これまでの提案手法の有用性を示すためにいくつか実験を行った。

図 9 の左画像は従来手法の顔検出例、右画像が提案手法の顔検出例である。提案手法では過去の結果を用いて閾値の調整を行っている。これにより従来手法で発生していた誤検出を削減することが出来る。

図 10 の左画像も従来手法の顔検出例、右画像が提案手法の顔検出例である。図 9 に比べて検出漏れが生じている。今回の実験ではこのような検出漏れが比較多く見受けられている。これはオプティカルフローによる注意の伝播 (対象確率と注意確率) が正確に行われていないことと、閾値の調整が上手く行っていないことが考えられる。この点については今後の課題としている。

図 11 は 1 フレーム当りの Haar 特徴の顔識別回数をプロットしたものである。従来手法では 1 フレーム当たり平均 11 万回程度の識別回数を、3 万回程度まで削減出来ていることが判る。

図 12 は 1 フレーム当りの処理時間を示している。結果から、提案手法の方が従来手法に比べて明らかに処理時間が多いことが判る。識別回数が減っているにも拘らず処理時間が大幅に増えた理由としてオプティカルフローの計算と、注意の伝播の部分で大きく時間を取られているのではないかと考えられる。この処理時間が掛かってしまう問題についても今後の課題としている。



図 9: 顔検出比較 (成功例)



図 10: 顔検出比較 (失敗例)

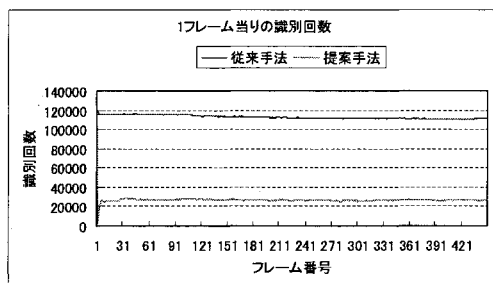


図 11: Haar 特徴による顔識別回数

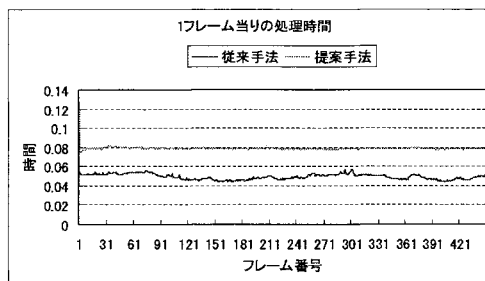


図 12: 1 フレーム当りの処理時間

5 まとめ

本報告では、対象検出の新たな手法として、人間の視覚に応じた挙動を実現し、どの部分を注意して探索すればよいか、ということ論じた。これによって 1 フレーム当りの識別回数を削減することが出来ることを証明した。また、Haar 特徴による顔検出で、過去の情報を用いて強識別器の閾値を修正することによって誤検出を軽減させることが出来たことを示した。

今後の課題としては閾値の調整方法、処理時間の高速化、伝播方法の精密化が挙げられ、今後これらの問題に取り組む予定である。

参考文献

- [1] Viola,P.and Jones,M, "Rapid Object Detection using a Boosted Cascade of Simple Features", Im Proc.IEEE Conf. on Computer Vision and Pattern Recognition,Kauai,USA 2001.
- [2] Y.Freund and R.Schapir, "Decision-theoretic generalization of on-line learning and an application to boosting", Journal of Computer and System Sciences, vol.55, No.1, pp.119-139,1997.
- [3] 坂平星弘, 和田俊和: "Genetic Algorithm を用いた対象検出法", MIRU2007