

H-049 文脈情報を用いた対象検出のための事前確率の推定 Prior Probability for Object Detection Based on Context Information

鈴山 有紀[†]
Yuki Suzuyama

堀田 一弘[†]
Kazuhiro Hotta

高橋 治久[†]
Haruhisa Takahasi

1. はじめに

静止画像や動画から特定の対象を検出する対象検出は、コンピュータによる対象の自動認識の第一段階としてその重要性が増してきている [1]。静止画像中には人物、動物、車、花などの対象物、そして背景として道路、建物、植物、空などが存在する。このような画像からある1つの対象を検出したいとする。このとき、従来は検出したい対象の学習サンプルから前もって対象に特有の特徴を抽出し、それらを基に学習した検出器を用いて対象を検出するのが一般的である。このような手法で自動的に対象を検出する場合、入力画像中で利用される情報は、局所的な領域のものだけである。しかし、人間が対象を認識し、検出する際には画像全体の情報も利用している。例えば、その画像が街頭を写した物ならば車が存在しているかもしれないし、室内ならば存在しないだろう。また、画像全体が劣化し、ぼやけている場合などには、その一部分である対象そのものだけを取り出して見せられてもそれが何であるか認識し難い。だが、例えばやけたものであったとしても画像全体を与えられれば、それがどのようなシーンであるかという情報が得られるため、その物体が何であるかを認識することができる。

このように、周囲の景色とそこに存在する物体との間には強い相関関係があり、人間はその関係も用いて対象を認識している。本研究では、対象も含めた画像全体の情報、つまり文脈情報を Gabor フィルタ [2] を利用して抽出し、得られた特徴を基に画像中に対象が存在するかどうかの事前確率、及び存在する位置の推定を行う。文脈情報が混合ガウス分布に従うと仮定し、EM アルゴリズムを用いて推定する方法 [3] が提案された。本研究ではそれに対し、文脈情報を基に非線形な学習データや、未知のデータに対して高い汎化性能を持つ学習機械であるサポートベクトルマシン [4] を用いて対象が存在する事前確率を推定する方法を提案する。提案手法では、存在確率の推定に対してサポートベクトルマシン (Support Vector Machine, SVM) を、存在位置の推定に対してサポートベクトル回帰 (Support Vector Regression, SVR) を用いる。

実験では、対象を車とし、存在確率の推定及び存在位置の推定を行った。車が存在する画像、存在しない画像共に Label Me [5] に含まれる 256×256 画素の画像を $1/2$ サイズの 128×128 画素にしたものである。文脈情報が混合ガウス分布に従うと仮定し、EM アルゴリズムを用



図 1: 元の画像の例

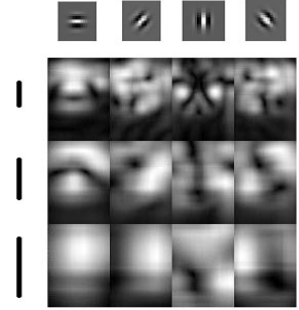


図 2: Gabor フィルタをかけた後の画像の例

いて推定する方法とサポートベクトルマシンを用いて直接推定する方法の2種類の方法で実験した結果、存在確率の推定、位置の推定共に提案手法であるサポートベクトルマシンを用いて直接推定する提案手法の方が良い推定結果が得られた。

2. 特徴の抽出

本研究では、対象も含めた画像全体の情報、つまり文脈情報を利用して対象が存在する事前確率の推定を行う。まず、画像全体の特徴を抽出するために原画像に Gabor フィルタを適用した。Gabor フィルタとは人間の脳内における初期視覚情報処理のモデルであり、人の感覚に近い情報を抽出することができる。以下に Gabor フィルタの式を示す。

$$Gab(x, y) = g_0 \exp\left(-\frac{x^2 + y^2}{\sigma^2}\right) \exp(2\pi j \langle f, (x, y)^T \rangle) \quad (1)$$

周波数 f を変化させると、抽出する Gabor 特徴の大きさや向きを変えることができる。また、分散 σ は Gabor フィルタの大きさを決める変数である。図 1 に示す画像から様々な周波数の Gabor 特徴を抽出した例を示す。図 2 は 12 種類 (大きさ: 3 段階, 向き: 4 方向) の Gabor 特徴を並べて表示したものである。各行は 3 段階の大きさに対応し、各列は 4 方向の向きに対応している。

このようにして得た多様な周波数の Gabor 特徴は次元数が膨大である。そこで、得られた Gabor 特徴を主成分分析し、次元を圧縮する。そして、求めた固有空間に各画像の Gabor 特徴を射影することにより得られた特徴量を文脈情報 v_c とし、以降の処理で利用する。

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

3. 対象の存在確率の推定

前節で得られた文脈情報 v_c を用いて各画像における対象の存在確率を推定する方法について説明する。

3.1 ガウス分布のパラメタの推定

まず、文脈情報 v_c が与えられたときに対象 o の存在する確率 $P(o|v_c)$ は、対象の存在確率を $P(o) = 0.5$ と仮定すると以下のように表すことができる。

$$P(o|v_c) = \frac{P(v_c|o)P(o)}{P(v_c)} \quad (2)$$

$$= \frac{P(v_c|o)P(o)}{P(v_c|o)P(o) + P(v_c|\bar{o})P(\bar{o})} \quad (3)$$

$$\approx \frac{P(v_c|o)}{P(v_c|o) + P(v_c|\bar{o})} \quad (4)$$

よって、尤度 $P(v_c|o)$, $P(v_c|\bar{o})$ がわかれば $P(o|v_c)$ が求められることになる。そこで、対象が存在する文脈情報の分布である尤度 $P(v_c|o)$ を $P(v_c|o) = \sum_{i=1}^M b_i G(v_c; \mu_i, \Sigma_i)$ のように混合ガウス分布で近似できると仮定し、パラメタである重み b_i , 平均 μ_i , 共分散 Σ_i を次に示す EM アルゴリズムを用いて推定することによって求める。

E-step: クラスタ $h_i^k(t) = \frac{b_i^k G(v_t; \mu_i^k, \Sigma_i^k)}{\sum_{i=1}^M b_i^k G(v_t; \mu_i^k, \Sigma_i^k)}$ を推定。

M-step: E ステップで求めた $h_i^k(t)$ に基づいて混合ガウス分布のパラメタである重み b_i , 平均 μ_i , 共分散 Σ_i を更新。

これにより、尤度 $P(v_c|o) = \sum_{i=1}^M b_i G(v_c; \mu_i, \Sigma_i)$ が求められる。また、同様にして尤度 $P(v_c|\bar{o})$ も求めることができる。

3.2 対象の存在の識別

前節において尤度 $P(v_c|o)$, $P(v_c|\bar{o})$ が求められたので、式 (4) を用いれば対象の存在確率 $P(o|v_c)$ を得ることができる。上記の方法では対象の事前確率を $P(o) = 0.5$ と仮定して存在確率を求めていたが、一般には事前確率はわからない。そこで、識別率の観点から事前確率を推定する方法について説明する。

文脈情報が与えられたときの対象の存在する確率 $P(o|v_c)$ は式 (2), (3) のように表された。このとき、 $P(o|v_c) > \frac{1}{2}$ となる場合に対象が存在すると考えることができる。そこで、式 (3) を変形していくと以下のようになる。

$$P(o|v_c) = \frac{P(v_c|o)P(o)}{P(v_c|o)P(o) + P(v_c|\bar{o})P(\bar{o})} \quad (5)$$

$$= \frac{P(v_c|o)}{P(v_c|o) + P(v_c|\bar{o}) \frac{P(\bar{o})}{P(o)}} > \frac{1}{2} \quad (6)$$

$$\frac{P(v_c|o)}{P(v_c|\bar{o})} > \frac{P(\bar{o})}{P(o)} \quad (7)$$

すると、尤度比について式 (7) が得られた。学習データから識別率が最も高くなるように $\frac{P(\bar{o})}{P(o)}$ を決定すれば、事前確率の比を閾値として利用することができる。テストデータに対しては、求めた尤度比が閾値より高ければ車が存在する、低ければ車は存在しない、と判断し、その正誤から識別率を求める。

4. 対象の存在位置の推定

次に、文脈情報を基に各画像において検出対象が存在する確率の高い場所を推定する。

4.1 各画素における対象の存在確率の推定

学習データの文脈情報 v_c と位置情報 x を与え、それを元に各画素における対象の存在確率を求める。その際、対象の位置の分布を $G(x; x_i, X_i)$, 文脈情報の分布を $G(v_c; v_i, V_i)$ というように、それぞれがガウス分布に従うと仮定する。このとき、検出したい対象 o と文脈情報 v_c が与えられたときの、ある画素 x における対象の存在する確率 $P(x|o, v_c)$ は、混合ガウス分布を用いて

$$P(x|o, v_c) = \frac{\sum_{i=1}^M b_i G(x; x_i, X_i) G(v_c; v_i, V_i)}{\sum_{i=1}^M b_i G(v_c; v_i, V_i)} \quad (8)$$

と表すことができる。ここで、対象の位置の平均 x_i は、各クラスタの文脈特徴に対して一次従属な性質を持つため、 $x_i = a_i + A_i^T (v_c - v_i)$ と表すことができる。

以上より、混合ガウス分布のクラスタの重み b_i , 位置の平均 x_i を求めるために必要な A_i と a_i , 位置の共分散 X_i , 文脈情報の平均 v_i と共分散 V_i を EM アルゴリズムを用いて推定する。

E-step: クラスタ $h_i^k(t) = \frac{b_i^k G(x_t; x_i^k, X_i^k) G(v_t; v_i^k, V_i^k)}{\sum_{i=1}^M b_i^k G(x_t; x_i^k, X_i^k) G(v_t; v_i^k, V_i^k)}$ を推定。

M-step: E ステップで求めた $h_i^k(t)$ に基づいて、混合ガウス分布のパラメタである重み b_i , 位置の平均 x_i を求めるために必要なパラメタ A_i と a_i , 位置の共分散 X_i , 文脈情報の平均 v_i と共分散 V_i を更新。

このようにして混合ガウス分布のパラメタが推定できると、式 (8) に代入することによって各画素における対象の存在確率 $P(x|o, v_c)$ が求められる。

4.2 対象の存在確率が高い領域の中心点の推定

一方、EM アルゴリズムによって求めた混合ガウス分布のパラメタを用いて、対象が存在する確率の高い領域の中心の一点を推定することもできる。中心点 (\bar{x}, \bar{y}) を定めると、前もって指定した点との平均二乗誤差を求めることができるので、精度の評価に用いることができる。

5. サポートベクトルマシンによる方法

この節では、サポートベクトルマシン (SVM) を用いて対象の存在確率の推定を行う方法、サポートベクトル

回帰 (SVR) を用いて対象の存在位置の推定を行う方法について説明する。SVM を用いるとより直接的な推定を行うことができる。

5.1 SVM を用いた対象の存在確率の推定

SVM は識別面と学習データとの距離の最小値 (マージン) が最大になるような超平面を構築する 2 クラス識別器である。学習は凸二次計画問題に帰着されるので、局所最適解の問題は発生しない。また、線形分離が不可能な場合でも、カーネル関数を利用することによって写像自体を直接計算することなく高次元の特徴空間における最適な分離超平面を求めることができる。

SVM では 2 クラスの識別問題を解く際に、各テストデータの識別超平面からの距離に対してシグモイド関数 $f(x) = \frac{1}{1 + \exp(-x)}$ を利用することで、どちらのクラスに近いかを確率として表すことができる。これを利用することにより、対象が存在するか否かの 2 値の結果だけでなく、各テスト画像に対する対象の存在確率を求めることができる。

5.2 SVR を用いた対象の存在位置の推定

SVM は最大マージンアルゴリズムの主要なすべての特徴を保ちながら回帰にも適用できる。SVR では真の値から一定距離内の誤差を無視する損失関数を定義する。この損失関数を用いることで、凸関数の最小値として大域的最適解が存在することと信頼できる汎化限界の最小化が保証される。

ここでは、SVR を用いて対象の存在位置の中心点を求める。学習画像に対してあらかじめ見積もった位置との平均二乗誤差を求め、結果の評価に利用する。また、同様に存在位置の分散も求める。位置と分散を合わせることで結果を図示することができる。

6. 実験および考察

実験では対象を車とし、 128×128 画素の車が存在する画像 990 枚、存在しない画像 990 枚を用いて学習を行い、学習に用いていない各々 330 枚の画像に対して評価を行った。

まず、原画像から特徴量を抽出する際、Gabor フィルタの周波数の向きは 4 方向、フィルタの大きさは 9×9 , 13×13 , 17×17 画素の 3 サイズとした。そして、得られた多様な周波数の Gabor 特徴の集合を局所的な 4×4 画素の領域で平滑化して主成分分析し、上位 20 次元を文脈情報として利用することにした。

6.1 実験 1. 車の存在確率の推定

各テスト画像に対して車が存在する確率を求め、実際に車が存在する画像、存在しない画像それぞれに対する識別率を求めた。その際、以下に示す 2 つの方法、8 種類のパラメタで実験を行った。

方法 1: 尤度 $P(v_c|o)$ が混合ガウス分布に従うと仮定し、

EM アルゴリズムを用いて推定する方法。混合ガウス分布の個数を $M = 2 \sim 6$ と変化させた。

方法 2: 文脈情報 v_c を SVM に入れて存在確率を直接推定する方法。SVM は多項式カーネル $K(x, z) = (\langle x \cdot z \rangle + 1)^d$ を使い、次元数を $d = 2, 3$ と変化させた。

それぞれの方法によって求めた識別率を表 1, 2 に示す。この結果を見ると、手法 1 では平均すると 80% に届かないのに対して提案手法である方法 2 は 80% を超えている。方法 1 よりも識別率が高くなっていることがわかる。方法 1 では文脈情報の分布がガウス分布に従うものとする仮定の上で識別率を求めているが、方法 2 においてはそのような仮定を用いることなく、直接的に識別を行っているためにこのような結果が得られたと考えられる。また、方法を問わず全体的に車の方が識別率が高くなっているが、これは車特有の文脈情報が得られたためだと考えられる。学習用、テスト用の画像共に、車が存在する画像は全体的に道路や建物が写っている画像が多く特徴が統一されていたのに対して、車が存在しない画像は人、山、海など写っている物体や景色があまり統一されていなかったためである。

6.2 実験 2. 車の存在位置の推定

文脈情報と学習画像中の車の位置情報を利用してテスト画像における車の存在位置を推定した。また、平均二乗誤差を求めるための準備としてテスト画像に対しても対象が存在する領域の中心点を事前に見積もっておく。

実験 1 と同じように、以下に示す 2 通りの方法で位置の推定を行った。

方法 1: 対象の位置と文脈情報が混合ガウス分布に従うと仮定し、各画素における対象の存在確率を EM アルゴリズムを用いて推定する方法。混合ガウス分布の個数は $M = 4$ とした。

表 1: 方法 1 によるテスト画像の識別率

	M=2	M=3	M=4	M=5	M=6
車あり	82.73	69.70	79.39	84.24	78.18
車なし	73.03	81.82	79.09	73.64	76.06
平均	77.88	75.76	79.24	78.94	77.11

表 2: 方法 2 によるテスト画像の識別率

	d=2	d=3	d=4
車あり	82.12	82.42	76.06
車なし	80.91	82.73	82.42
平均	81.52	82.58	79.24

表 3: 平均二乗誤差

	縦方向 [pixel]	横方向 [pixel]
方法 1	11.04	26.61
方法 2	8.68	22.23



図 3: 実験に使用した画像の例

方法 2: 文脈情報 v_c を SVR に入れて車の存在位置の中心の一点を直接推定する方法。また、位置の図示に用いるために同様に存在位置の分散も推定する。SVR は実験 1 と同様に多項式カーネルを用い、次元数は $d = 2$ とした。

テスト画像である図 3 に対して方法 1 によって推定された確率をかけて表示したのが図 4 である。明るく表示されている部分が車の存在する確率が高い領域である。また、方法 2 によって推定された中心の一点と分散を用いて存在確率の高い領域を楕円で表示したのが図 5 である。これらを見ると、両方の方法で概ね良好な推定結果が得られていることがわかるが、左側の画像のように元々の存在範囲が狭い画像では SVR を用いた方法 2 の方が良い結果が得られている。



図 4: 方法 1 で求めた推定位置の画像例

2 つ目の結果として、それぞれの方法によって推定された中心の一点の座標とテスト画像に対してあらかじめ手動で求めておいた座標との平均二乗誤差の値を表 3 に示す。表 3 から、SVR を用いた方法 2 の方が誤差が少ないことがわかる。これは、実験 1 の結果と同様の理由によるものと考えられる。また、縦方向の誤差よりも横方向の誤差の方が圧倒的に大きいことも見て取れる。図 3 を見ればわかるが、実世界において車は横に連なっていることが多い(左側の画像のように正面を向いていても、右側の画像のように横を向いていても)ため、横方向は分散が大きくなってしまいうも仕方がないといえる。なお、この結果は図 4,5 の結果とも一致する。

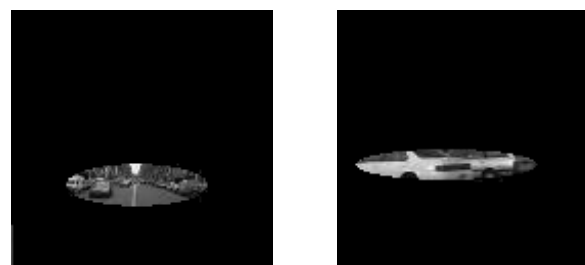


図 5: 方法 2 で求めた推定位置の画像例

7. おわりに

本論文では、画像全体から得た特徴量である文脈情報を元に対象が存在する事前確率を求め、さらに画像中の位置の推定も行った。その際、従来手法では文脈情報の分布がガウス分布に従うものとする仮定を用いて推定を行うのに対し、提案手法ではサポートベクトルマシンを用いて文脈情報から直接的に推定を行った。その結果、対象の存在確率の推定では SVM によって適切な識別超平面が構築されたことにより従来法よりも高い 80% を超える精度が得られた。また、存在位置の中心点の推定でも平均二乗誤差を抑え、より適切な存在領域を推定することができた。

参考文献

- [1] M-H. Yang, D. Kriegman, and N. Ahuja, "Detecting Faces in Images: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24, No.1, pp.34-58, 2002.
- [2] J. Daugman, "Entropy reduction and decorrelation in visual coding by oriented neural receptive fields," Trans. on Biomedical Engineering, Vol.36, No.1, pp.107-114, 1989.
- [3] A. Torralba, "Contextual Priming for Object Detection," International Journal of Computer Vision, Vol.53, No.2, pp.169-191, 2003.
- [4] V. N. Vapnik, "Statistical Learning Theory", John Wiley & Sons, Inc., 1998.
- [5] Label Me, "http://labelme.csail.mit.edu/".