文献紹介

Lampert, C. H., Blaschko, M. B. and Hofmann, T.: Beyond sliding windows: Object localization by efficient subwindow search（写真の中から猫をすばやく見つける方法）, *Proc. IEEE Computer Vision and Pattern Recognition（CVPR）2008*, pp. 1-8（2008）

1. 画面のどこに猫がいるかを探す

カメラで撮影した画像から人の顔や，泥棒，自動車，猫などを探すシステムがある．このような応用では「部分画像を画面内で探す」（部分画像探索）というタスクが基本的かつ重要である．問題を式で簡単に表すと以下となる．

$$R_{\mathrm{obj}} = \arg\max_{R \subseteq I} f(R)$$

R_{obj}が目的の部分画像もしくは領域，Rは領域，Iは画像である．fを品質関数（quality function）と呼び，fの出力が大きいほど，見つけたい対象と似ていることを表す．さまざまな品質関数が提案されており，目的に応じて，また見つけたい対象に応じて適した関数は異なる．

さて，この探索の最も素朴な解決策は，画面の左上から，少しずつ探索窓をずらして（sliding window，以下，滑走窓方式と呼ぶ），一つ一つfの出力を調べるという方法である．もちろん拡大縮小を考慮すれば，窓の大きさも変えながら調べることになる．一般には$n \times n$画素の画像で$O(n^4)$の計算量が必要である．

計算機が速くなったとはいえ，これでは時間がかかりすぎる．そのため動画像を扱う場合には，前後のコマとの差や，背景との差などを用いて，画面内で変化があった領域を限定し，その周辺のみで探索を行うのが一般的である．しかし静止画像の場合にはこの手段は使えず，動画像であっても静止している対象を見つけ出すことはできない．

なお探索対象を限定すれば，高速化の手段は多数提案されている．見つけたい対象特有の情報を用いてフィルタリングし，探索範囲を限定すればよい．例えば，特徴的な色や模様があれば，それらがない場所を探す必要はない．

ただ探索対象や品質関数を限定しない一般的な手法としては，滑走窓方式を超える提案はなかった．この論文では滑走窓方式ではない，高速な部分画像探索手法を提案している．その計算量は，$n \times n$画素の画像で$O(n^2)$である．

2. 分枝限定法の利用

基本的な考えは古典的な分枝限定法（Branch-and-Bound Search）である．分枝限定法は，問題を複数の部分問題に分け，それぞれの部分問題での目的関数の上限（最小化の場合は下限）を計算して，枝刈りをする方法である．

この論文では解もしくは解候補を，長方形領域（top, bottom, left, right）$=(t, b, l, r)$とする．ここで，例えば上端の要素tの最小値と最大値の組を$T=[t_{\mathrm{low}}, t_{\mathrm{high}}]$と表現すると，部分画像探索は$[T, B, L, R]$という探索範囲（図1）に対する，$f$を目的関数とした最適化問題となる．なお，長方形領域に制約があれば部分問題分割を効果的に行える．

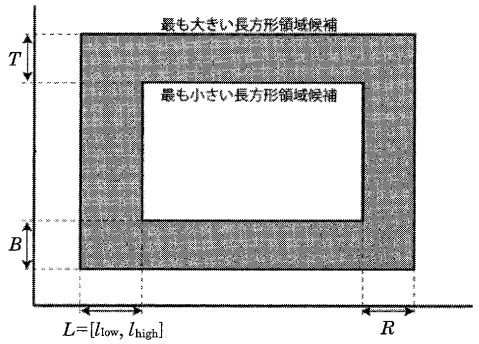

図1 探索領域

ここで問題となるのは上限の計算である．fの上限を与える関数を$\hat{f}$とすると，$\hat{f}$は次の条件を満たす必要がある．

i）$\hat{f}$は，その探索範囲内の任意領域のf以上，

ii）探索範囲内に領域が一つの場合はfに等しい

例えば，$\hat{f}$を極端に大きな定数とすれば，当然ながら条件iを満たす．また全件探索した結果から$\hat{f}$を計算すれば，最も正確な上限を返すことができる．しかし，いずれも無意味である．適切な$\hat{f}$は，計算量が少なく，それでいて，なるべく正確な上限を返す関数である．この論文では三つの適用例に関して具体的に$\hat{f}$を提案し，実験した結果を示している．

3. 適　用　例

(1) 犬猫探索

最初の適用例は，画像中から犬もしくは猫を探す問題である．なお，紙面の都合からこの適用例に関してのみ詳細に述べ，残り二つに関しては概要に留める．また，この犬猫探索で用いた方法が，彼らの手法の最もシンプ

ルな適用例であり，残り二つは，その拡張という位置づけにある．

実験に用いた画像はPASCAL VOC[*1]というベンチマークであり，屋内，屋外の多数の写真に，写っているオブジェクトの領域とカテゴリーが人手でタグ付けされている．PASCAL VOC 2006（5 304枚）と2007（9 963枚）をそれぞれ用いている．

特徴量には一般物体認識で有効性が示されている，bag of visual words（bovw）を用いている．これは画面内から注目に値する特徴的なポイントを多数抽出し，それら注目点を表現する模様などの画像特徴量を計算，さらにベクトル量子化を行って，コードブックを作成しヒストグラム化するというものである．具体的にはSURF[Bay 06]が用いられている．SURFの説明は省略する．

品質関数には線形SVMの出力を用いる．SVMの識別関数を$f(I)=\beta+\sum_i\alpha_i\langle h, h^i\rangle$とする．$h$はbovwで計算されたヒストグラム，$\langle,\rangle$は内積である．ここで$W_j=\sum_i\alpha_i h^i_j$とすると

$$f(I)=\beta+\sum_{j=1}^{n}w_{c_j}$$

となり，それぞれのwはコードブック中の各インデックスの，品質関数に対する寄与とみなすことができる．

ここでbovwヒストグラムが，注目点数に重みを掛けて加算したものになっていることに着目すると，上限を以下のように計算できる．

$$\hat{f}(I):=f^+(R_{\max})+f^-(R_{\min})$$

ここで，f^+は正のwだけの合計，f^-は負のwだけの合計である．探索の最大領域は最多の注目点を含んでいる．最小領域は，最も少ない注目点を含んでいる．よって，最大領域の正のwの合計から，最小領域の負のwの合計を引けば，それを超える値が得られることはない．図2に例を示す．図2の外側の最大領域での，重みが正の注目点は5個，内側の最小領域内にある，重みが負の注目点は3個である．重みの絶対値を仮にすべて1とすれば，$\hat{f}=5-3=2$となる．グレーの範囲にどう枠線を引こうが，これを超える値を返すことはない．

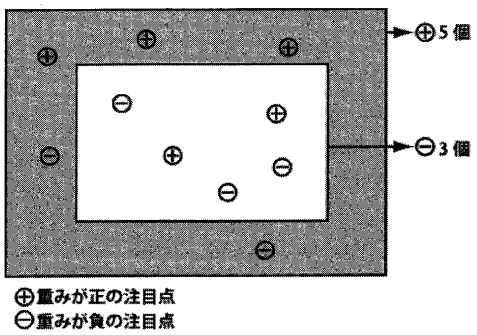

図2　上限の求め方

実験結果では，猫探索の精度22.3％，犬14.8％で，従来手法のVOC 2006への適用結果を上回る精度を達成している．なおこの精度は，全画像中からの猫・犬画像の探索も含めた精度であり，猫・犬画像のみを扱った場合は40％以上の領域特定精度である．VOC 2007への適用ではVOC 2006の場合よりも，精度が向上した結果が得られている．なお，精度が従来法を上回ったのは，従来法では計算時間の関係上，探索空間全体を探索できていないのに対し，提案手法が滑走窓方式に匹敵する網羅性をもつためと思われる．

もともとの目的であった速度に関してはどうかというと，2.4 GHzのPCで，一画像（500×333画素～640×480画素）当たり40 msと十分高速である．これは滑走窓方式で探索窓を20×20画素に固定した場合に匹敵する．

(2) 自動車探索

自由に撮影された画像から多種の犬や猫を見つけるのは，そもそも難しい設定の問題であるため，従来よりも高いとはいえ，2割程度と低い精度に留まっていた．そこで次に，もっと限定的な適用例として，側面から見た自動車を探索する問題を扱っている．

用いた画像群はUIUC自動車データセット[*2]である．数百枚の白黒の自動車画像であり，自動車は横から撮影されたものに限定されている．訓練用の画像と，テスト用の画像に分かれている．

特徴量としてはbovwをベースとした上で，ピラミッド構造の階層型ヒストグラム[Lazebnik 06]を用いている．まず，画像を縦横に均等に，例えば2×2や4×4に区分し，それぞれの領域でbovwを計算する．次にすべての階層の個々のヒストグラムでの品質関数の合計を，品質関数とする．この階層型ヒストグラムは，形状変形の少ない対象に対して効果的な照合を可能とする．上限関数は，多少複雑にはなるが，犬猫検索と全く同じ考え方で構成できる．

実験結果では，1×1～10×10分割までの階層型ヒストグラムを用いて，適合率と再現率が一致する結果で，エラー率1.4％と従来法を大幅に上回る結果を得ている．

(3) ロゴマーク探索

ビデオ映像のキーフレーム画像中から，あるロゴマークを探索する問題に適用している．質問画像としてロゴマークが与えられ，それが出てくるシーンを映像中から探索するという応用である．特徴量はbovwヒストグラムであり，類似度にはχ^2検定を用いている．

実験では類似度が高いほうからN個の候補を見つけ

*1 Pattern Analysis, Statistical Modeling and Computational Learning Visual Objects Classes Challenge, http://pascallin.ecs.soton.ac.uk/challenges/VOC/

*2 UIUC Car dataset, http://l2r.cs.uiuc.edu/~cogcomp/Data/Car/

るタスクを扱っている．画像一枚一枚を個別に探索するのではなく，候補画像を全部一度に分枝限定法の対象とすることで，40倍～70倍の，大幅な高速化を実現している．1万枚のキーフレーム画像に対して，候補を一つ見つけるのに要する時間は2秒未満であった．

4. 論文の位置づけと意義

部分画像探索で，上限を使ってスピードアップするというと，日本の画像研究者なら，1998年のアクティブ探索法 [村瀬 98] を思い出す人は多いだろう．アクティブ探索法は，ヒストグラムインタセクション（ヒストグラムの重なりの多少で類似度を判定）を前提とした場合に，上限値を計算して探索を効率化する手法である．

また分枝限定法自体は古典的な方法で，広く知られている．つまり画像探索で上限を求めて探索効率を上げるという考え方自体は，古くから多くの人が知っていたことになる．また，この紹介論文で使われている技術，例えばbovwやSVM，階層型ヒストグラムは，どれも有名な手法であり，すでにパーツは出そろっていた．著者らの貢献は，これらのパーツを緻密に組み上げて，多くの人が納得する素直な枠組みとした点にある．たぶん類似の発想でトライしていた研究者は世界中に山ほどいるだろう．しかし彼らはほんの少し長じていた．

筆者個人として，最も興味を覚えたのは，分割可能で積算型の特徴量の有用性である．これまでに画像分野では，膨大な数の特徴量が提案されている．その中にはヒストグラムのような積算型の特徴量と，全体的な形状特徴のような分割不可能な特徴量がある．この論文ではbovwという特徴量が活用されているが，上限を容易に求められるのは，この特徴量の性質に依存するところが大きい．

実は，筆者は人間とのインタラクションを考えた場合にも，このタイプの特徴量が重要な位置を占めているのではないかと考えている．情報検索や，画像認識システムのラベリングなど，人間が介在することで目的を達するしくみは多い．その場合，人間はシステムの振舞いを何らかの形で推定して，先読みをしながら操作方法を変える．その際に，このタイプの特徴量がベースになっていると，理解をステップアップしやすいのではないか．それは一種，上限や下限を容易に見積もれることにも関係しているのかもしれない．思いつきに過ぎないが．

この性質の良い論文を読むと，それまでの混沌としていた思考にまとまりを得て，階段を一歩上ったようなクリアな視野を得ることができる．それがCVPRでベストペーパとなった理由の一つではないかと思う．

◇ 参 考 文 献 ◇

[Bay 06] Bay, H., Tuytelaars, T. and Van Gool, L. J.: SURF: Speeded up robust features, *ECCV*, pp. 404-417（2006）

[Lazebnik 06] Lazebnik, S., Schmid, C. and Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, *CVPR*, pp. 2169-2178（2006）

[村瀬 98] 村瀬 洋，Vinod, V. V.：局所色情報を用いた高速物体探索—アクティブ探索法—，信学論（D-II），Vol. J81-DII, No. 9, pp. 2035-2042（1998）

〔堤　富士雄（電力中央研究所）〕