

ローカルなアピアランス特徴によるクラスモデル獲得法の検討

5K-1

— 映像へのインデックス生成を目指して — *

藤田 悦郎 安部 伸治 杉村 利明

NTT ヒューマンインタフェース研究所

1 はじめに

我々は先に、テロップ認識技術により抽出したテロップ等の言語情報に基づいて、映像データを言語処理と背景知識により自動的に構造化する映像文脈処理技術について報告した[1]。

この映像文脈処理技術の適用可能な対象は、現状では、テロップが付与された映像データに限られる。故、対象の拡大のため、テロップの全く付与されていない映像データからもテロップに代るインデックス情報を画像認識等により生成する技術が必要である。

現在我々は、映像中の様々な対象を自動的に同定・分類するための認識モデルをできるだけ容易にかつ効率的に獲得する枠組について検討を進めている。本稿では、そのモデル獲得法について、概略と現時点までの結果を報告する。

2 モデル獲得

モデル獲得に関する研究は、コンピュータビジョンの分野においてこれまでに数多くの手法が提案されている。中でも Turk や村瀬らは、対象が写された事例画像より2次元的なアピアランスモデルを作成する手法によって、モデル作成の効率化と高い認識精度を実現している[2, 3]。しかし彼らの手法では、一定の条件（例えば背景の有無等に関する）下で撮影された画像が処理の対象となるため、撮影環境の大きく変動する一般映像に対しても有効な手法とは必ずしも言えない。このため本稿では、複雑なシーンからでも抽出が安定な色・エッジ等の、よりローカルな画像特徴に着目し、これらによって対象クラスのモデルを作成する枠組を試みる。ローカルな画像特徴の利用により、撮影条件の変動への耐性が可能になるとともに、処理の単純化による高速化も実現可能になると考えている。

2.1 モデル記述のための画像特徴

対象クラスのモデルの記述には、事例画像の色特徴とエッジ特徴の組み合わせを利用する。すなわち、画像から抽出した色特徴ベクトル v_{color} とエッジ特徴ベクトル v_{edge} の両者をベクトルとして連結させた特徴ベクトル

$$v = (v_{color}, v_{edge}) \quad (1)$$

を用いて対象を記述する。画像検索等[4, 5]ではこれまでに色特徴やエッジ特徴の単独利用で比較的良好な結果が得られているが、単独での記述ではその表現能力にやはり限界がある。そこでここでは、これらの特徴を結合して相補的に利用することで、より強力な特徴記述を実現する。以下では、ここで用いる色・エッジ双方の特徴について概説する。

2.1.1 色特徴

色特徴としては、画像の色ヒストグラムを用いる。ここでは、人間の視覚に適合しかつ照明変化にロバストな HSI 表色系の色相の16分割による色ヒストグラムを採用する。実際の処

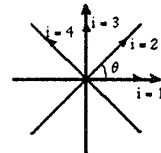


図 1: エッジ方向

理においては、正規化のため、ヒストグラムの各成分を画像の全画素数で割って正規化をする。

2.1.2 エッジ特徴

エッジ特徴としては、エッジの方向ヒストグラム[5]を用いる。これは対象の輪郭特徴等を表現するのに有効である。ここでは、次の手順で求められる、より単純化されたエッジ方向ヒストグラムを用いる。

手順1 画像 f の各画素 (x, y) に対し次のガウス微分の絶対値を求める。

$$I_i(x, y) = |D_i G \otimes f| = \left| \iint D_i G(u-x, v-y) f(u, v) du dv \right| \quad i=1, 2, 3, 4. \quad (2)$$

ここで、 $D_i G$ はそれぞれガウス関数 G の $\theta = 0, \pi/4, \pi/2, 4\pi/3$ に関する方向微分を表す（図1参照）。 $I_i(x, y)$ は、その値が大きいほど、画素 (x, y) が、 i 方向にエッジ方向を持つエッジ点らしいことを示す尺度になっている。

手順2 ある閾値 $th (th > 0)$ に対し、 $I_i(x, y)$ が、

$$\max\{I_i(x, y) : i=1, 2, 3, 4\} > th \quad (3)$$

を満たすならば、その画素 (x, y) はエッジ点であると判定する。さらに、式(3)を最大にする i を求め、エッジ点 (x, y) のエッジ方向は i 方向と定める。

手順3 $i=1, 2, 3, 4$ それぞれの方向について、対応するエッジ点の個数をカウントしヒストグラムを作成する。

色の場合と同様、実際の処理においては、正規化のためヒストグラムの4成分を画像の全画素数で割り正規化をする。

なお、特徴ベクトル v の次元数は、 v_{color} の次元数が16、 v_{edge} の次元数が4となるため20次元である。

2.2 特徴空間におけるクラスモデルの形成

画像の特徴ベクトルにより画像に含まれる対象を同定・分類するためには、特徴空間において各対象クラスのモデルをあらかじめ作成しておく必要がある。これは識別境界の問題に相当し、パターン認識の分野において現在までに数多くの手法が提案されている。その中で Lovell らは、学習データに基づいてマルチスケールで境界を設定していくことで、特徴空間を階層的に分類する手法を提案している[6]。本稿では、彼らの手法を改良したものにより特徴空間を分類する手法を提案する。ただし

*A Model Generation Method using Local-appearance Features

Etsuro FUJITA, Shinji ABE and Toshiaki SUGIMURA
NTT Human Interface Laboratories

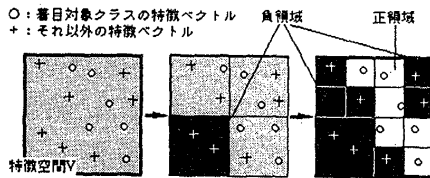


図 2: 特徴空間の分類イメージ

説明の簡単のため、ここでは、ある対象クラスとそれ以外の 2 クラスに分類する場合に話を限定する。提案手法は、 N 個の対象クラスの場合へ容易に拡張されるが、紙面の都合上ここでは説明を省略する。

S_{pos} 、 S_{neg} をそれぞれ、着目する対象クラスの事例画像 (複数枚) から抽出した特徴ベクトルの集合、着目する対象以外のクラスの事例画像 (複数枚) から抽出した特徴ベクトルの集合とする。ただし、 S_{pos} の画像は、画像中の着目する対象の外接矩形領域として与える。以下では、特徴空間の対象クラスの領域を正領域、それ以外の領域を負領域と呼ぶこととする。分類のアルゴリズムは次の通りである (図 2 参照)。

手順 1 特徴空間 V の各軸を 2 等分し、 V を $2^{\dim F} (= 2^{20})$ 個の正方領域に分割する。

手順 2 (a) 分割によりできるある正方領域が S_{pos} の要素のみを含むならば、その領域は正領域と定める。(b) 他方、正方領域が S_{neg} の要素のみを含むならば、その領域は負領域と定める。 S_{pos} 、 S_{neg} いずれの要素も含まない場合も負領域と定める。

手順 3 V の正方領域がすべて正負いずれかに分類されているならば処理を終了する。正負に分類されない正方領域 W (S_{pos} と S_{neg} の要素を同時に含む) がある場合は、 W に対し手順 1 ~ 手順 3 の処理を行い、正負いずれかに分類されるまで W を細分化していく。

3 モデル照合

ここでは、入力された未知画像内に着目する対象が含まれるか否かを判定する照合処理について概説する。

画像中に着目する対象が含まれる場合、その位置やスケールは画像毎に異なる。このため照合過程では、様々なスケールの探索ウィンドウを未知画像全体にスキャンさせながら照合を行う。照合で、探索ウィンドウ内に着目対象が含まれるか否かは、その領域から抽出した特徴ベクトルが、特徴空間 V の正領域 (着目対象の) に含まれるかどうかにより判定する。

V の正負への分類が一種の木構造で表現されることにより、モデル照合は階層的に行なわれる。これにより照合の効率化と高速化が可能であると考えている。

本手法では、照合の際、距離尺度の概念を用いる必要がない。色とエッジのように異なる画像特徴を組み合わせて利用する場合、例えば NN のように何らかの距離尺度により照合する枠組では、「異なる特徴量間の重み付けをいかに最適化するか」という問題が生じる。これに対し、結合特徴空間での一種の分布推定により入力パターンを分類する本枠組では、このような重み付けの問題は考慮する必要がない。この点は本手法の利点であると考えている。なお、分布推定に関しては、例えばガウス混合分布の適用 [7] 等も考えられるが、一般に対象クラスの分布が単純でないことから、混合数の選択等事前の取り扱いが難しい。事前にこのような問題を考慮せずにモデルが作成できる点も本手法の利点と考えている。

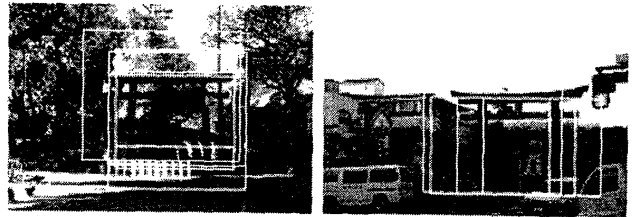


図 3: 照合結果

4 簡単な実験例

ここでは、「鳥居」を対象とした簡単な実験例を示す。実験では、「鳥居」の事例画像 50 枚と、それ以外の対象が写された事例画像 2550 枚を用いてモデルを獲得させた。

学習データ以外の「鳥居」の画像を未知画像として入力し、モデル照合を行った結果を図 3 に示す。ただし実験では、一段階につき 1.3 ~ 1.5 倍程度スケール拡大した 3 段階の探索ウィンドウを約 25 パーセントスキップさせながら画像全体をスキャンし照合を行った。

5 おわりに

本稿では、2 種のローカルなアピアランス特徴、すなわち色ヒストグラムとエッジ方向ヒストグラムに基づいて対象クラスのモデルを獲得する枠組について述べた。その中で、これら 2 つ特徴量の結合化により、対象クラスの特徴を色情報と輪郭情報の両面から相補的に記述するより強力な特徴記述について提案し、さらに、その記述法によるモデルを正負事例の学習により容易にかつ効率的に作成する手法について提案した。また、簡単な実験例によってその可能性を示した。

今後の課題としては、本手法をより客観的に評価することが挙げられる。特に、他の手法との精度や処理速度の面での比較が必要である。この他、本手法の適用対象の明確化等も検討していきたいと考えている。

謝辞

本研究を進めるにあたり、ご指導を賜りました中野博隆メディア応用システム研究部長、並びに、ご討論頂きました研究グループの皆様へ深く感謝します。

参考文献

- [1] 安部 伸治, 松村 季樹, 藤田 悦郎, 小倉 健司: 文書情報を介在させた映像情報文脈編集技術, 人工知能学会合同研究会 AI シンポジウム'97, 1997.
- [2] Turk, M. and Pentland, A.: Eigenfaces for Recognition, Journal of Cognitive Neuroscience, Vol.3, No.1, 1991.
- [3] Murase, H. and Nayar, S.K.: Visual Learning and Recognition of 3-D Objects from Appearance, IJCV, Vol.14, No.1, 1995.
- [4] Swain, M.J. and Ballard, D.H.: Color Indexing, IJCV, Vol.7, No.1, 1991.
- [5] 水野 陽一, 小林 亜樹, 吉田 俊之, 酒井 善則: エッジ方向特徴量による画像検索, 信学技報, IE98-62, 1998.
- [6] Lovell, B.C. and Bradley, A.P.: The Multiscale Classifier, IEEE Trans on PAMI, Vol.18, No.2, 1996.
- [7] Sung, K.K. and Poggio, T.: Example-Based Learning for View-Based Human Face Detection, IEEE Trans on PAMI, Vol.20, No.1, 1998.