C2-08

SIFT と GraphCuts を用いた物体認識及びセグメンテーション

Object Recognition and Segmentation Using SIFT and GraphCuts

須賀 晃† Akira Suga　福田 恵太† Keita Fukuda　滝口 哲也‡ Tetsuya Takiguchi　有木 康雄‡ Yasuo Ariki

1． はじめに

本稿では，SIFT[1]と GraphCuts[2]を用いた物体の認識・セセグメンテーション法を提案する．従来の SIFT を用いた物体認識手法では，詳細な物体領域の切り出しまでは行うことができなかった．一方，物体領域のセグメンテーションとしては GraphCuts による領域抽出法が提案されているが，手動でシードを与えなければならないという問題があった．本手法では，認識に用いた SIFT 特徴をそのまま GraphCuts のシードとして用いることで，認識とセグメンテーションの両方を自動で行う．実験の結果，オクルージョンを含む画像においても対象物体を認識し，その領域を切りだすことができ，本手法の有効性が確認できた．

2． 認識

2.1 SIFT 特徴点抽出処理

特徴点は，DoG 画像 $D(x,y,\sigma)=L(x,y,k\sigma)-L(x,y,\sigma)$ から求める． $L(x,y,\sigma)$ はスケールの異なるガウシアンフィルタと入力画像 $I(x,y)$ を畳み込みにより求められる平滑化画像である．DoG 画像 3 枚一組とし，注目画素の DoG 値を，隣接する上下スケールを含めた 26 近傍の画素と比較し，極値であれば特徴点として検出する．検出された平滑化画像の各画素において，勾配強度 $m(x,y)$ と勾配方向 $\theta(x,y)$ を以下の式により求める．

$$m(x,y)=\sqrt{f_x(x,y)^2+f_y(x,y)^2} \tag{1}$$

$$\theta(x,y)=\tan^{-1}\frac{f_y(x,y)}{f_x(x,y)} \tag{2}$$

$$\begin{cases} f_x(x,y)=L(x+1,y)-L(x-1,y) \\ f_y(x,y)=L(x,y+1)-L(x,y-1) \end{cases} \tag{3}$$

以上により得られた $m(x,y)$ と $\theta(x,y)$ を用いて，36 方向の勾配方向ヒストグラムを作成する．このヒストグラムの最大値から 80%以上となるピークを，その特徴点の代表オリエンテーションとして割り当てる．そして，特徴点周辺領域を代表オリエンテーションを基準とした軸に回転させて 4x4 ブロックに分割し，各ブロック毎に 8 方向の輝度勾配ヒストグラムを作成する(4x4x8=128 次元).

2.2 マッチング処理

あるモデル画像には M 個の特徴点，入力画像からは N 個の特徴点が抽出されたとすると，それぞれ次のように表わせる．

$$S_{model}=\{s^m\} \qquad (m=1,\cdots,M) \tag{4}$$

$$W_{input}=\{w^n\} \qquad (n=1,\cdots,N) \tag{5}$$

†神戸大学大学院工学研究科

‡ 神戸大学自然科学系先端融合研究環

$$s^m=(s_1^m,\cdots s_{128}^m)^T \qquad w^n=(w_1^n,\cdots w_{128}^n)^T \tag{6}$$

モデル画像中の m 番目の特徴点と，入力画像中の n 番目の特徴点のユークリッド距離が最小となる特徴点 n' を以下の式から求めることで，マッチングを行う．

$$n'=\arg\min_{n\in N}\sqrt{\sum_i^{128}(s_i^m-w_i^n)^2} \tag{7}$$

2.3 投票処理

予め，モデル画像中での物体の中心点(基準点)を定め，各特徴点 (x_{model},y_{model}) に対して基準点からの位置ベクトル $(\Delta x,\Delta y)$ を計算しておく(図 1)．次に，その特徴点とマッチングのとれた入力画像中の特徴点に対する中心候補点 (X,Y) を，位置ベクトル，スケール $\sigma_{\{input,mode\}}$，オリエンテーション $\theta_{\{input,model\}}$ を用いて次の式から求める．

$$\begin{cases} X=x_{input}+\dfrac{\sigma_{input}}{\sigma_{model}}\times\sqrt{\Delta x^2+\Delta y^2}\times\cos(\theta+\theta_{model}-\theta_{input}) \\ Y=y_{input}+\dfrac{\sigma_{input}}{\sigma_{model}}\times\sqrt{\Delta x^2+\Delta y^2}\times\sin(\theta+\theta_{model}-\theta_{input}) \end{cases} \tag{8}$$

ここで， $\theta=\arctan(\Delta y/\Delta x)$ であり，マッチングした全特徴点についてこの中心候補点に投票を行う．この処理により，入力画像中に対象物体が存在すると，求められる中心候補点は同じ位置に集まるため，その位置に多くの投票が集まることになる．そこでこれらの候補点をクラスタリングし，各クラスタにおいて閾値以上の投票が得られれば，対象物体が存在すると判定する．[3]

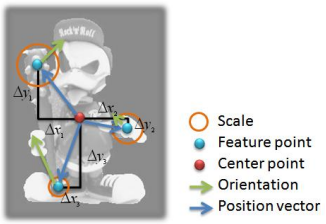


図 1　位置ベクトルの算出

3． セグメンテーション

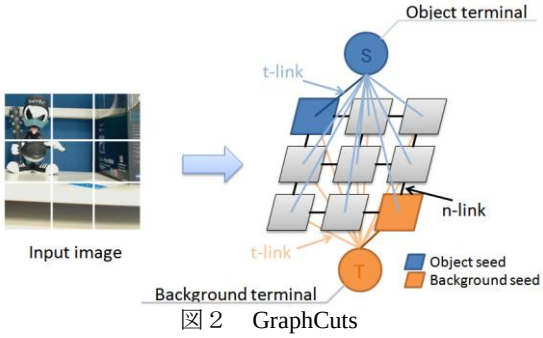


図 2　GraphCuts

GraphCuts では，画像 P の各ピクセル $p\in P$ に対応するノードと，sink(T)と source(S)というターミナルからなる

グラフを作成する(図２)．ノード間を結ぶエッジは n-link，ノードとターミナルを結ぶエッジは t-link と呼ばれ，それぞれ表１及び式(9)-(11)に示すコストが与えられる．

表１　エッジコスト

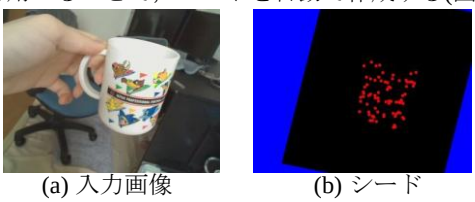
edge		cost	For
n-link	{p,q}	$B_{\{p,q\}}$	{p,q}∈N
t-link	{p,S}	λ・R_p(“bkg”)	p∈P, p∉O∪B
		K	p∈O
		0	p∈B
	{p,T}	λ・R_p(“obj”)	p∈P, p∉O∪B
		0	p∈O
		K	p∈B

$$\begin{cases} R_p(\text{"}obj\text{"}) = -\ln \Pr(C_p \mid O) \\ R_p(\text{"}bkg\text{"}) = -\ln \Pr(C_p \mid B) \end{cases} \quad (9)$$

$$B_{\{p,q\}} \propto \exp\left(-\frac{(I_p - I_q)^2}{2\sigma^2}\right) \cdot \frac{1}{dist(p,q)} \quad (10)$$

$$K = 1 + \max_{p \in P} \sum_{q:\{p,q\} \in N} B_{\{p,q\}} \quad (11)$$

ここで，C_p，I_p はピクセル p における RGB 値と輝度値である．$\Pr(C_p \mid O)$，$\Pr(C_p \mid B)$ はシード以外の物体と背景の尤度であり，色情報を GMM に適用することで推定する．p∈O，p∈B はそれぞれ物体シードと背景シードのことであり，本研究では投票処理において認識されたクラスタに投票した特徴点を物体シードとし，それらを用いてモデル画像をアフィン変換した外側領域を背景シードとして用いることで，シードを自動で作成する(図３)．



(a) 入力画像　(b) シード

図３　シードの自動作成

以上のように作成されたグラフに対し，Min cut/Max flow algorithm を適用し，コスト総和が最小になるようにグラフをカットすることで，セグメンテーションを行う．

４．実験

20 個のモデル物体に対して，100 枚のテスト画像を用意し認識とセグメンテーションの実験を行った．各モデル物体はそれぞれ 45° ずつの角度から撮影されたものである．認識精度は再現率と適合率で評価を行った．結果を表２に示す．セグメンテーション精度は，正解マスクデータに対しどれだけ誤検出ピクセルがあったかを示すエラー率により評価した．結果を表３に示す．

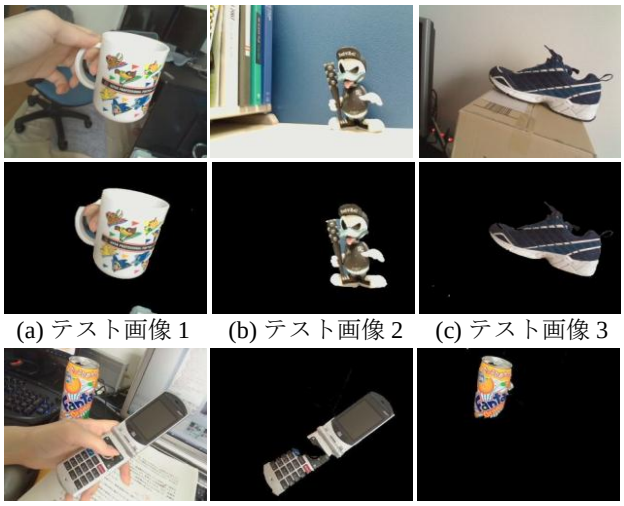
表２　認識精度

再現率	適合率
82.1%	100%

表３　セグメンテーション精度

物体領域エラー	背景領域エラー	全体エラー
3.73%	6.21%	9.94%

セグメンテーション結果の例を図４に示す．



(a) テスト画像 1　(b) テスト画像 2　(c) テスト画像 3

(d) テスト画像 4

図４　セグメンテーション結果

５．考察とまとめ

本論文では，SIFT と GraphCuts を組み合わせることで，認識とセグメンテーションを自動で行う手法を提案した．認識精度については，投票処理による認識のため，オクルージョンを含む場合でも見えている部分からの投票により認識でき，且つ誤った対応点は除外されるため誤認識が見られず，有効性が確認できた．セグメンテーション精度については全体的には良い結果が得られたが，シードに偏りがあるものについては，エラー率が高かった．これについては SIFT 以外の特徴量を用いるなどしてなるべく偏りのないシードを生成する方法を検討している．

本手法では特徴点の全探索をするため，マッチング処理に時間がかかるという問題がある．これに対して，特徴ベクトルをビット化して特徴量空間を分割することで，探索の際に大幅に検索領域を絞り込んでマッチングを行う手法が提案されている[4]．この手法を適用することで，実験で約 30 秒かかっていたマッチング処理を約 0.1 秒で行うことができた．近似によりマッチング精度は下がるが認識は投票で行うため影響が少ないというメリットもある．しかしシードの精度が下がるとセグメンテーションの精度に影響が出てしまうため，モデル画像で予め物体シードを学習しておくなどして，認識とセグメンテーションの両方を高速且つ精度よく行える手法を現在考察中である．

参考文献

[1] D.G.Lowe, “Distinctive image features from scale-invariant keypoints,” Journal of Computer Vision, 60, 2, pp.91-110, 2004.

[2] Y.Boykov, M.P.Jolly, “Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images,” ICCV, vol. 2, pp.731-738, 2004.

[3] 高木雅成,藤吉弘亘,”SIFT 特徴量を用いた交通道路標識認識,”画像センシングシンポジウム, LD-206, 2007.

[4] 野口和人,黄瀬浩一,岩村雅一, ” 近似最近傍探索の多段階化による物体の高速認識.”, MIRU, pp.111-118, 2007.