

画像と地図情報を利用した建物の認識

3K-10

劉佩林[†]池内克史[†]坂内正夫[†][†]東京大学生産技術研究所

1 始めに

近年、通信基盤や情報処理機器の発達に伴い、ナビゲーションシステムに代表されるようなリアルタイムな実世界情報を利用する技術への要求が高まっている。実世界情報を利用するシステムを構築するには、市街地シーンの認識が必要である。本文では、市街地シーンにおける建物を認識することによって、市街地シーンの理解について、検討を行う。

市街地シーンにおいては、広い範囲にわたって建物が映っている遠景のシーンでは、オブジェクトのほとんどは建物であるため、シーンを理解するためには建物を認識すればよい。一方、地図データベースには、実世界における建物の2次元的位置情報を示す空間位置情報とポリゴンや階数などのような幾何的な情報が存在する。そこで、画像から得られる建物の情報と、地図データベースに存在する情報のマッチングをとることで、市街地シーンの認識を行う。

カメラの位置と向きのパラメータがわかる場合、投影変換を用いれば、地図データベースにおける建物のモデルを画像に投影し、建物を検出することが可能となるが、カメラの投影変換が完全かつ精密にわかるのであれば、精密にオブジェクトとモデルのマッチングがとれるはずである。しかし、現実には、投影変換のパラメータの推定で誤差が生じるため、投影されたオブジェクトと実際に画像に映っているオブジェクトが精密に合えない可能性が高い。そこで、オブジェクトをロバストに検出するため、2DTM(2D Template Matching)を利用して、建物を検出する。2DTMとは、3次元オブジェクトのモデルから投影変換で生成されたテンプレートと探索する画像領域のオブジェクトの部分との類似度を評価することにより、画像中のオブジェクトと一致する箇所を求

め、オブジェクトの姿勢を検出する手法である [1] [2]。画像から2DTMを用いて建物を検出する場合、まず、画像からエッジを抽出する。次に、建物のモデルを生成する。そして、カメラの投影変換を用いて建物のモデルを画像に投影する。投影されたモデルをテンプレートとして、エッジ画像におけるテンプレートの周辺のエッジとマッチングを行う。

2 地図情報から建物の3Dモデルの生成

2DTMにおけるオブジェクトのモデルは、オブジェクトの3Dエッジで記述される。各エッジはそのエッジ上の点の位置、そのエッジの方向、そのエッジが構成する面の垂直方向で表現する。これらのエッジを用いてオブジェクトの見え方のテンプレートを計算する。

地図データベースには建物のポリゴンと階数の情報が存在するため、2DTMに対する建物の3次元のモデルが生成可能である。具体的には、まず、建物の高さの情報を階数を用いて推定する。次に、建物のポリゴン情報に基づき、立方体を生成する。最後に各立方体の辺で建物のモデルを記述する。

3 テンプレート・マッチングによる建物の検出

求められたカメラの投影変換を用いて、生成された建物の立方体のモデルを画像中に投影し、投影された部分(テンプレート)をエッジ画像領域における近くのエッジと比較し類似度を調べることにより、画像上の建物を検出する。

画像とテンプレートとの類似度を測定する目標関数はオブジェクトの姿勢のパラメータで定義する。オブジェクト姿勢は三つの回転パラメー

タと三つの移動パラメータによって決定される。目標関数の最小値を与えるパラメータがオブジェクトの最適な姿勢としてとられる。目標関数が式(1)によって与えられる。zはモデルの点と対応する画像の点との誤差を表し、zの独立な分布 $P(z) \propto e^{-\rho(z)}$ と仮定する。qはオブジェクトの姿勢を表すとする。

$$E(q) = \sum_i \rho(z_i(q)) \quad (1)$$

ここで、 z_i はi番目のモデルの点と近くの画像部分の点との誤差を表し、式2で表すとしている。

$$z_i(q) = \min_{\vec{a} \in D} \|\vec{x}_i - \vec{a}\| \quad (2)$$

ここで、Dは画像における2D点の集合であり、 \vec{x}_i はqを用いて変換されたi番目モデルの点ワールド座標である。モデルの点と近くの画像部分の点である \vec{a} はK-dimensional nearest-neighbor searchを用いて計算される。目標関数の最小値を求めることにより、qの最大事後確率を推定できる。例えば、誤差の分布がGaussian関数と仮定できる場合、パラメータの最大事後確率の推定は最小2乗誤差 ($\rho_z = z^2$)となる。最小の誤差を求めるため、勾配 gradient-descent update rule を利用すると考えられる。

$$\Delta q \propto -\frac{\partial E}{\partial q} = \sum_{i=1}^n \Psi(z_i) \frac{\partial z_i}{\partial q} \quad (3)$$

ここで、 $\Psi(z_i) = \frac{d\rho(z)}{dz}$

目標関数の最小値を与えるqが建物の最適な姿勢として得られる。この最適な姿勢に基づき、画像中の精密な建物の位置を求めることができる。

4 実験

この画像に対して、テンプレート・マッチングを利用して画像内部におけるホンダビルを検出する実験を行った。図1は地図情報に基づいて構築したモデルを示す。図2は投影した結果であり、図2はマッチングの結果である。

参考文献

[1] Mark D. Wheeler and Katsushi Ikeuchi " Sensor Modeling, Probabilistic Hypothesis

Generation, and Robust Localization for Object Recognition" IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.17, No.3,pp252-265,1995

[2] Mark Damon Wheeler, " Automatic Modeling and Localization for Object Recognition", Ph.D dissertation ,CMU., 1996

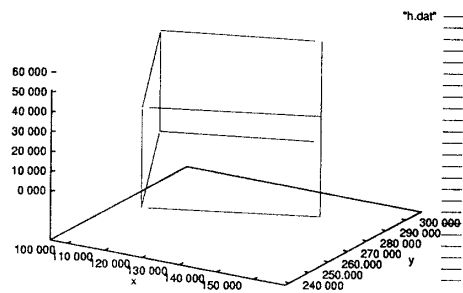


図1: ホンダビルの立方体のモデル

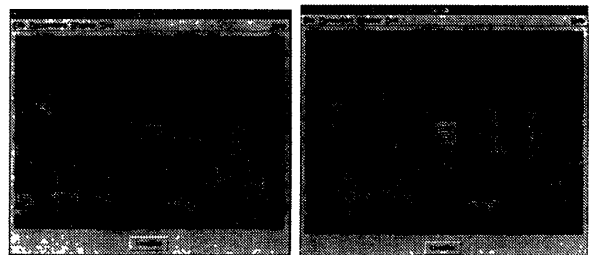


図2: 2DTMによる建物の認識