

組合せ最適化アルゴリズムに基づくオブジェクトとラベルのレイアウト認識 Layout Recognition of Objects and Labels Based on Combination Optimization Algorithm

大浦 淳貴
Junki Oura

辻 亮弥
Ryoya Tsuji

黒木 修隆
Nobutaka Kuroki

廣瀬 哲也
Tetsuya Hirose

沼 昌宏
Masahiro Numa

1. はじめに

近年、自立行動能力をもつロボットの実現が求められている。しかし実環境において人間のために設計された機器を理解・操作することは極めて困難である。例えば図 1 のリモコンによってテレビの電源を入れる操作を考える。これを計算機の視覚情報、すなわち画像を用いて理解・実行するためには、次の 3 つのプロセスが必要である。

- (1) 文字 (ラベル) を読み取る。
- (2) その文字が指すオブジェクトを検出する。
- (3) そのオブジェクトの動作 (押す/回す/スライドする) を理解する。

このうち、(1) は文字認識の技術が利用可能である。(3) については各種ボタンの形状とその動作を "知識" として事前に持っておく必要がある。その上でパターンマッチング等の画像処理によって実現できると考えられる。しかし、(2) について、人間の常識的な判断と同等の判断を計算機に行わせることは大きな課題である。

本稿では (2) に注目し、様々な画像中からオブジェクトとラベルを正しく関連づけることを目的とする。

2. オブジェクトとラベルのレイアウト認識

図 1 からわかるように、オブジェクトとラベルを正しく関連付けるためには、局所的な距離にとらわれず、全体のバランスを考慮する必要がある。そこで、本稿では、組み合わせ最適化アルゴリズムを用いる。



図 1 テレビのリモコン

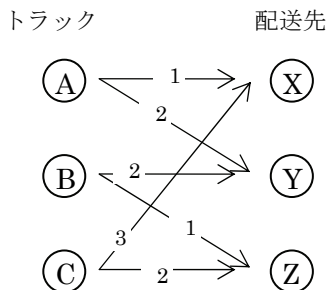


図 2 組合せ最適化の例 (物資の配送問題)

2.1 問題設定

オブジェクトとラベルは画像処理によって検出できると仮定する。オブジェクトとラベルの数は必ずしも一致しないが、一つのラベルに対応するオブジェクトは必ず画像中に存在するものとする。

2.2 組合せ最適化アルゴリズム

図 2 に組合せ最適化の例を示す。複数考えられる組合せの中で、輸送コストの合計が最も小さくなるような組合せを探し出すアルゴリズムを組合せ最適化アルゴリズムという。

2.3 手法の流れ

オブジェクト m ($1 \leq m \leq M$) とラベル n ($1 \leq n \leq N$) 間のコストとして、それぞれの距離 $L(m, n)$ を用いる。 M 個のオブジェクトと N 個のラベル ($M \geq N$) が存在する場合、 $M \cdot N$ 個の組合せが存在する。その組合せ全てに対して、コストの合計値

$$C = \sum_{n=1}^N L(m, n) \quad (1)$$

を求め、その合計値 C がもっとも小さくなる場合の組合せを求める。

3. 実験

3.1 実験条件

リモコン、路線図、テレビ番組のテロップを例題として実験する。各例題におけるオブジェクトとラベルを表 1 に示す。

表 1 各例題におけるオブジェクトとラベル

画像名	オブジェクト	ラベル
リモコン	ボタン	ボタンの動作
路線図	駅	駅名
テレビ番組テロップ	人物の顔	人物名

距離 $L(m, n)$ の定義については、図 3 に示すように、重心間距離を用いる方法と、外接矩形同士の最短距離を用いる方法の 2 通りで実験する。

3.2 実験結果

実験結果を図 4~図 8 に示す。

リモコンでは、路線図など他の画像に比べて全て規則正しく配置されており、ラベルやオブジェクトの大きさも同じものが多く、重心間距離と最短距離のどちらを用いる場合でも、図 4 のように正しい結果が得られた

路線図では、ラベルがオブジェクトの右側にある場合と、左側にある場合があり、駅名の長さによってラベルの大きさも様々である。そのため、重心位置が駅名の長さによって変わり、重心間距離をコストとして用いた場合は図 5 に示すように、実際とは異なるオブジェクトに結びつけられる結果となった。最短距離をコストとして用いた場合はラベルの大きさに依存しないため、図 6 に示すように全て正しい結果となった。

テレビ番組においては、図 7 では、重心間距離と最短距離のどちらをコストとしても、正しい結果が得られた。それに対し図 8 のように複数の人物が映っている場合には、重心間距離と最短距離のどちらをコストとして用いても、誤った結果が出力された。この画像ではラベルが 3 次元的に一番手前の人物を表していると考えるのが自然である。このような判断を行うためにはオブジェクトの前後関係を理解する必要がある。

4. まとめ

本稿では画像中のラベルとオブジェクトの最適な組合せを求める方法について検討を行った。重心間距離をコストとするより、最短距離をコストとする方がラベルの大きさに依存せず、高い認識率を得ることができた。今後はより複雑なレイアウトに対して、3 次元的な位置関係や全体のバランスを考慮することで、より人間の判断に近い対応付け方法を検討していく必要がある。

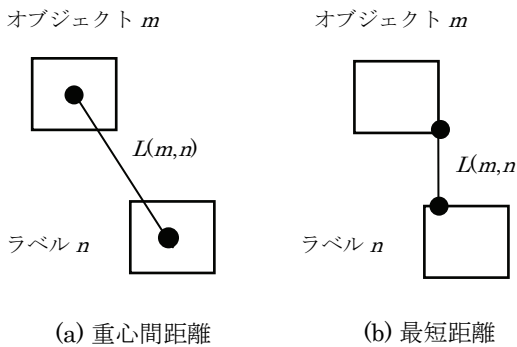


図 3 オブジェクトとラベル間の距離の定義

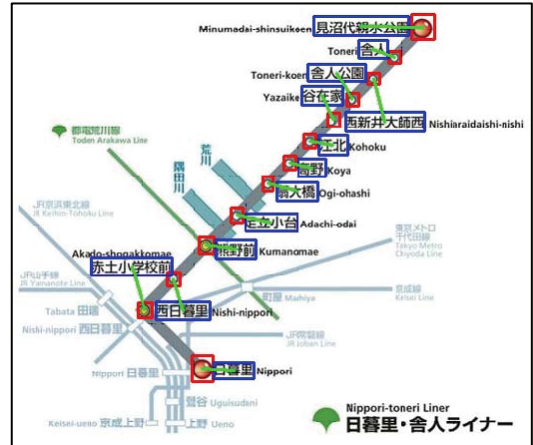


図 5 重心間距離をコストとした場合の実験結果 (61.5%)

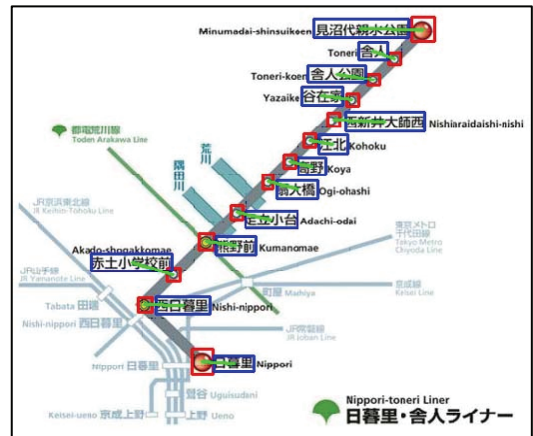


図 6 最短距離をコストとした場合の実験結果 (100%)



図 7 テレビ番組テロップにおける対応付けの結果



図 4 リモコンにおける対応付けの結果



図 8 テレビ番組テロップにおける対応付けの失敗例