

定性的な位置関係による実画像の把握

Real-image Recognition Using Qualitative Space Representation

菅野 裕 (名大院) 正 渡辺 崇 (名大)

Yutaka SUGANO, Nagoya University, Furo-cho, Chikusa-ku, Nagoya

Takashi WATANABE, Nagoya University

Keywords: Image Recognition, Qualitative Space Representation, Common Understanding, Active Camera

1 はじめに

自然言語表現を受けて注目点を切り替えるカメラワークが求められるスタジオ撮影や遠隔講義、自律型ロボットといったシーンでは、ユーザとの定性的なコミュニケーションの結果として映像を抽出する機能が求められる。

これまでも、アクティブカメラの制御を目的に、自然言語を用いたカメラワークの理解⁽¹⁾や、与えられた撮影対象にふさわしいズーム率を自動的に決定する方法⁽²⁾などが報告されている。しかし、これらでは、自然言語の持つ曖昧性と空間キャプチャのための定量的変換の方法が十分に検討されていなかったり、取り込まれる空間情報が理解されてカメラの行動に反映されているわけではなかった。

映像の内容の理解としては、宮崎ら⁽³⁾が、教室での講義を対象に、5つの状況を捉えて、それぞれの状況において各ユーザがみたい映像を調停の形で決定する方法を提案しているが、カメラの動きは予め規定されたユーザの映像化ルールに基づいている。

本研究では、人間と計算機との空間情報の共有と円滑なコミュニケーションを目指して、パンチルト機能とズーム機能を備えたカメラを用いて、視野内あるいは視野外に存在する複数の物体を認識し、それらの間の位置関係を理解することについて検討する。

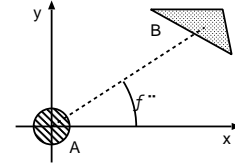


Fig. 1 orientation of object

2 定性的表現の定量化

本節では、人間の定性的な表現による入力を定量的に評価する手法について考える。コミュニケーションに用いる表現として、物体間の位置関係と、色、形状、大きさ、物体間の距離で表す物体の特徴量について考える。

2.1 物体間の位置関係

ユーザとカメラは空間を共有しており、同一の方向を向いている状況を想定する。

物体の相対的な位置を表す「～の右」「～の前」といった定性的な表現は離散的に表せない曖昧さを持つ。Fig.1では、BはAに対して右にあるとも言えるし、上にあるとも言える。「とても右らしい」、「あまり右らしくない」という量を定量的に表したものを右指向性と呼び、BのAに対する右指向性 $D_{B \rightarrow A}^{right}$ を次式で定義する。

$$D_{B \rightarrow A}^{right} = \begin{cases} 1 - \frac{1}{(1 + e^{a(-|\theta| + \pi/4)})} & |\theta| \leq \pi/2 \text{ のとき} \\ 0 & |\theta| > \pi/2 \text{ のとき} \end{cases} \quad (1)$$

θ は B の A に対する角度 (Fig.1), a は定数である。 $0 \leq D^{right} < 1$ の値をとる。

Fig.2は (x, y) 平面原点に対して D^{right} を図式化したものである。同様に左、上、下など、総称して向き指向性 ($D^{orientation}$) と呼び、(1)式と同様に求められる。

2.2 物体の特徴量

ある物体が、ユーザの期待する物体とどれだけ特徴が類似しているかを表す物体の特徴量 A を以下の要素を用いて表す。

- 色指向性 a_1
色についての定量的評価を与える量。ユーザの望む色と

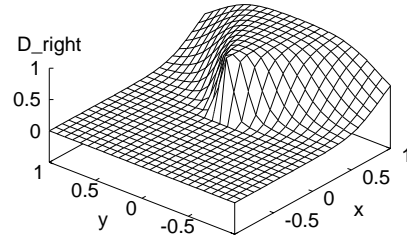


Fig. 2 D_{right} ("right" directivity)

どれだけ近いかを表す。物体の RGB 値を事前に用意した標本の RGB 値との距離を出して算出する。

- 形状整合度 a_2
形状についての定量的評価を与える量。ユーザの望む形状とどれだけ近いかを表す。多角形についてはハフ変換による線分抽出により辺の数を評価し、円については複雑度を用いる。
- 大きさ a_3
他物体との相対的な大きさ。面積を用いる。
- 物体間の距離の近さ a_4
対象物体が視野内にあるかの判定と、視野内での物体間の距離の定量的評価を与える量。物体の重心間の距離を用いる。 (x_1, y_1) にある物体に対して、 (x_2, y_2) にある物体の a_4 を以下で定義する。

$$a_4 = \begin{cases} a \left(1 - \frac{1}{(1 + e^{b(1 - (x_2 - x_1)^2 + (y_2 - y_1)^2)})} \right) + (1 - a) & (x_2, y_2) \text{ が視野の中にあるとき} \\ 0 & (x_2, y_2) \text{ が視野の外にあるとき} \end{cases} \quad (2)$$

a, b は定数である。ただし、 $0 < a \leq 1$ 。

Fig.3は $a = 0.7$, 40×40 の大きさの視野の中心 $(0, 0)$ に対する a_4 を図式化したものである。

これら a_n の要素を全て $0 \sim 1$ で規格化し、物体 i の特徴量 A_i を次式で定義する。

$$A_i = \sum_{k=1}^4 w_k a_k \quad (3)$$

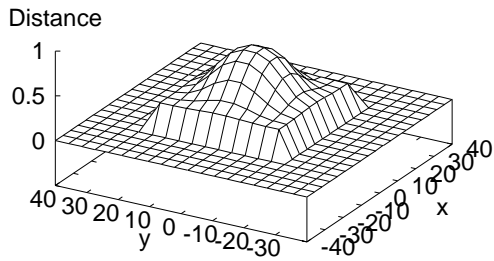


Fig. 3 distance from (0,0)

w_k は重みで、条件に応じて適当な値をあてる。

3 候補決定の手法

本節では、ユーザの要望に対して最適な物体を選出する手法について述べる。

ユーザはカメラから得られる映像をもとに、見たい物体を要求する。ユーザの要望は次のようなものである。

- 例 1: 「右にある三角形を見よ」
- 例 2: 「左の緑の円の上にある赤いものを見よ」

カメラを動かす過程で発見した物体は逐次的にデータベースに追加される。データベース内のある物体 i の j (但し, $i, j \in N$, N はデータベース内の物体の集合) に対する向き指向性 $D_{i \rightarrow j}^{orientation}$ と特徴量 A_i との積 $M_{i \rightarrow j}^{orientation}$ をユーザの要望に対する適合度とする。

$$M_{i \rightarrow j}^{orientation} = D_{i \rightarrow j}^{orientation} A_i \quad (4)$$

$M_{i \rightarrow j}^{orientation}$ を最大にする i が最適な物体として選出される。

例 1 の場合、「右」とはカメラ視野中心に対する向きなので、

$$M_{i \rightarrow center}^{right}$$

が適合度であり、これを最大にする i を選出する。

例 2 のような 2 物体間の位置関係を含む場合は、カメラ視野中心に対する「左の緑の円」の適合度 $M_{i \rightarrow center}^{left}$ とその i に対して「上にある赤いもの」の適合度 $M_{j \rightarrow i}^{above}$ の積

$$M_{i \rightarrow center}^{left} M_{j \rightarrow i}^{above}$$

を最大にする j を選出する。

適合度が極端に小さいときは、データベース内には合致する物体がないものとし、指示された向きにある未知の領域からより大きな適合度の物体を探す。

4 検証

SONY のパンチルトカメラ EVI-D30 を用いて実空間に対して実験、検証した。

ユーザはカメラから得られる映像をもとに、見たい物体を要求する。ユーザの要求は、物体の特徴や他物体に対する位置関係に定性的な表現を用いる。

初期状態 → 「右を見よ」 → 「三角の下の円を注目せよ」という手順で命令した。

Fig.4 は、左がカメラからの入力動画、右がそれに対して個々の物体の色、形状、位置を認知したデータベース空間を示している。5 つの物体が検出され、0 から 4 の番号でラベリングされている。

Fig.5 は「右を見よ」という命令を受けた後の状態である。右のデータベース空間に新たに発見された物体が 5, 6 でラベリングされて加わっている。

Fig.6 は「三角の下の円を注目せよ」という命令を受けた後の状態である。右のデータベース空間では、話題となった物体として最適と認識された物体の番号がチェックされている。複数ある円の中から、3 の三角の下にある 4 の円を的確に注目している。

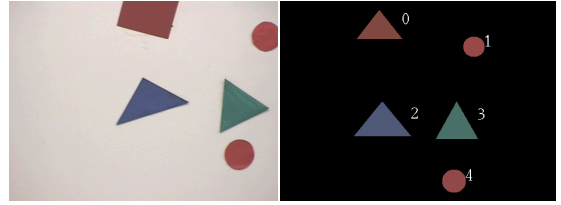


Fig. 4 Real-image(left) and cognitive space(right)

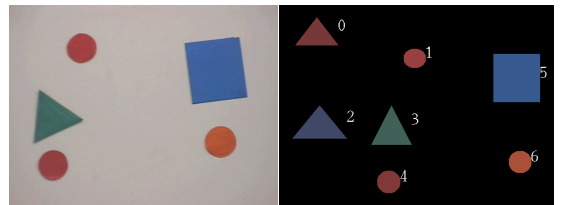


Fig. 5 Detection new objects

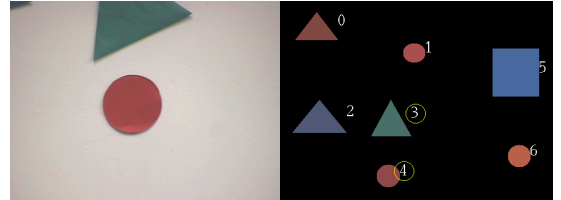


Fig. 6 Focusing the object

5 おわりに

本研究は、人間と計算機との空間情報の共有と円滑なコミュニケーションを目指し、定性的な表現による位置関係の理解と、その位置関係を用い、空間情報からユーザの望む物体を選出する手法について検討した。また、これを実空間に適用することでの有用性を検証した。

ユーザの要求する物体への指示が多いほど、特に位置関係を用いた表現を使うほど物体の候補選出の正確さが増した。

今後、物体の形状の認識にパターンマッチなどを取り入れることで、物体の特徴をより正確に捉えることが可能になり、適用範囲をより一般的にすることが期待できる。

参考文献

- (1) 新山祐介, 秋山英久, 鈴木泰山, 徳永健信, 田中穂積, 1999 年度人工知能学会全国大会論文集, 217-220
- (2) 戸田真志, 川嶋稔夫, 青木由直, 信学論, Vol.J81-D-II(1998),84-92
- (3) 宮崎英明, 亀田能成, 美濃導彦, 信学論, Vol.J82-D-II(1999),1598-1605