

ユーザとロボットとのインタラクションにおける共同注意の形成手法

K-22 Algorithm to Form Shared Attention in Human-Robot Interaction

陳 彬*
Bin Chen

目黒 光彦*
Mitsuhiko Meguro

金子 正秀*
Masahide Kaneko

1. はじめに

「共同注意」とは、複数の個体(人物、ロボット)が同一の目標体に注意することを意味する。ユーザの視線や顔の向きには、他者の注意先を誘導し、自らの注視している物体に注意を向けさせる機能もある。従って、ユーザの視線や顔の向きの情報は、共同注意の形成に密接に係わっている。これまで提案されて来た共同注意の形成メカニズムでは、設定された環境条件に依存する、3次元情報が活用されていないため3次元シーンに適用しにくいなどの点で問題が残されている [1]。

これに対して、本論文では、ユーザを含むシーンに対する2次元画像と3次元の距離情報に基づいて、共同注意の形成における誘目度マップを作成し、このマップを利用してユーザとロボットとの共同注意を形成する方法を提案する。誘目度マップは、ある個体(ここではロボット)が他者(ここではユーザ)の身体情報を読み取り、シーンにおいて注意される部分のある確信値を持って推測した結果の平面分布である。3次元情報を活用して得られた誘目度マップを利用することによって、事前知識や制約を必要とすることなしに、ユーザとロボットの間での共同注意を形成させることが可能となる。

2. モデルベースによる顔姿勢推定

ステレオカメラで取得したテクスチャ画像と距離画像を用いて、頭部の3次元ワイヤフレームモデルを利用して、2次元の動きから3次元の顔動きを最小二乗法に基づいて推定する。次に、隣接するフレームにおいて、モデルの対応する可視頂点間のテクスチャ差分とステレオ画像による頂点の3D座標の実測値と推定値の差分の和を評価関数とし、トラッキングエラーの累積和を最小とするようにSimplex Downhill法によって顔の動きパラメータを調整する。

実験では、初期フレームにおいて、ユーザの顔の初期姿勢がカメラに対して真正面であることを条件とし、シーン中の顔部分を検出する。予め登録してあるユーザ頭部の3次元ワイヤフレームモデルを、検出した顔部分の大きさに合うように調整する。その後、上述の方法によって顔姿勢値を自動的に推定する。このようにして得られた結果の例をFig.1に示す。Fig.1では、入力画像中の人物の顔姿勢の推定結果に基づいて回転、移動させた3次元顔モデルを入力画像にはめ込む形で表示しているが、これらより顔姿勢推定が適切に行われていることが分かる。

3. 視力空間分布の定式化

2. で求めた顔の姿勢値によって、初期フレームの姿勢状態から現フレームの姿勢状態に変換するアフィン行列

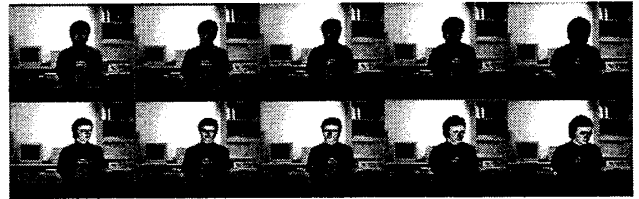


Fig.1: Estimated results of face orientation. Upper: Input images; Lower: Reconstructed face orientation corresponding to its upper image.

が求められる。この行列によって、顔のワイヤフレームモデルを初期の姿勢状態から回転・移動し、ステレオカメラで計測した3次元実空間に融合させる。

次に、視野の範囲を近似する。人間の左右視角を $\theta = \pm 60^\circ$ 、上下視角を $\varphi = \pm 40^\circ$ とし、空間視野の領域を楕円錐体で近似する。楕円錐体の頂点を両目を結ぶ直線の中点に置き、直交する両頂角 (θ, φ) を人間の上下・左右の視角と相等させる。また、楕円錐体のローカル座標系の原点は幾何学的頂点に一致させ、座標軸は頭部のワイヤフレームモデルのローカル座標系の軸と平行にとる。楕円錐体の軸が視線に対応し、軸付近の点に高い感度値を、軸から離れた点に低い感度値を与えることにより、人間の網膜の生理的視力分布の特性を定式化する。実空間の点に対して、ユーザの網膜における視力相対値(感度値を0~1に正規化した値)を計算することができ、画像と同じサイズを持つ「視力分布マップ」と呼ばれる2次元の行列が得られる(Fig.2)。視力分布マップにおいて、明るい領域はユーザの網膜に高い刺激を与えていることを表す。

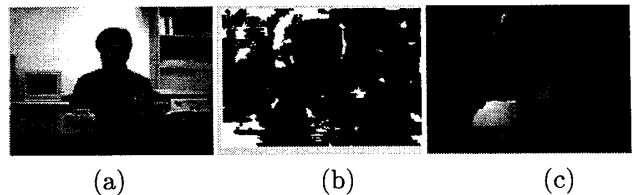


Fig.2: (a) Input image; (b) Disparity map; (c) Visual acuity map.

4. 注視可能な領域の追跡と時空間表現

ユーザの注視方向は必ずしも顔の向きと同じではないため、目を観測できない場合に、単に顔の向きだけで注視物体を正確に抽出することは困難である。このため、時間軸において、顔の向き・環境などの変化を総合的に考慮する必要がある。そこで、視力分布マップ上の値が正となる画像領域を分割し、フレーム内の領域配置関係とフレーム間の領域対応関係を時空間の3次元グラフで表現する。これにより、動物体の画像領域における視力

*電気通信大学 大学院電気通信学研究科電子工学専攻

相対値を、領域の動きに従ってフレーム間で伝搬させる。伝搬された視力分布マップの時間累積和によって、5.1で述べる誘目度マップを算出することができる。

ここで、領域ベース法を利用して動領域のトラッキングを行う。まず、視差画像において、視力相対値が0でない領域について、色相、彩度、動き、距離情報を特徴量として *K-means* 法で領域分割・統合を行い、領域グラフを作成する。次に、頂点が保持している動きベクトルによって、動き特徴の類似する頂点を統合し、モーションベースのグラフを生成する。最後に、各フレームで得られたモーションベースのグラフ同士に対して、頂点が保持している特徴とグラフ構造によりフレーム間でマッチングを行い、時間方向において対応する頂点を辺で連結し、時空間の3次元グラフを構築する。フレーム間の辺は、頂点で示した各領域が時間軸において連続的に追跡されていることを表している。このような時空間グラフを利用することによって、次のフレームで対応する領域へ各領域の視力分布値を時間的に遷移させることが可能になる。

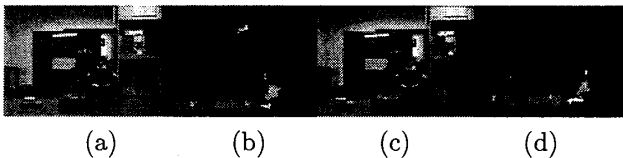


Fig.3: (a)(c) Two consecutive frames; (b)(d) Segmentation and tracking of the possible gaze area, in which the same color is assigned to the corresponding regions between two frames.

Fig.3は、ユーザの目の前にある前景領域について、領域分割・追跡を行った結果の例である。このように領域を分割・追跡することによって、フレーム間の領域対応情報を保有する時空間の3次元グラフを構築することができる。

5. ユーザとロボットとの共同注意の形成

5.1 共同注意形成の誘目度マップの算出

t を現時刻、 F を重視する過去のフレーム数とする。時間区間 $[t-F, t]$ 内の各フレームにおいて、3. で求めた視力分布マップについて時間累積和を計算する。また、視線が空間内を移動する可能性があるため、累積和を計算する時に、異なる物体への視線移動が累積和の計算結果に影響を与えられるように、時間区間 $[t-F, t]$ 内の各フレームにおける視力分布マップに対して、時間的減衰重みを導入する。ただし、3. で追跡した注視可能な動的領域の視力分布値は、領域の動きに従って次のフレームへ遷移させる。この遷移によって、視力相対値が集中した目立つ領域を生成することができ、ロボットに注意を向けさせる対象が明確になる。ここで得られたマップは、ユーザの注意対象を推測したロボットの内部確信値を表しており、ユーザとロボットとの「共同注意形成の誘目度マップ」と呼ぶことにする。

5.2 ユーザが注意しているオブジェクトの抽出

誘目度マップを確率分布マップと見なして、CamShift アルゴリズム [2] により、物体の領域範囲を同定する。

CamShift アルゴリズムに入力する初期枠の中心位置は、最も目立つ画素の画面座標である。CamShift の結果としての四角形のサイズは初期枠の大きさによらず、確率分布マップだけに依存するので、枠サイズの初期値は前フレームの枠をやや拡大して設定する。

5.3 共同注意形成の実験

ユーザが物体を操作するシーンをロボットに見せ、ユーザの注意対象を検出させて、共同注意を形成する実験を行う。Fig.4 の下段は計算した誘目度マップを示す。目立つ領域を CamShift アルゴリズムで抽出した結果を上段に四角形の枠で示す。誘目度マップ上の値は、ユーザがその画面座標の空間対応点に注意を向けていることをロボットが推測した結果としての確信値であるため、ロボットの視覚機構を CamShift で抽出した枠領域へ誘導させることによって、ユーザとロボットとの共同注意を形成できる。Fig.5 は各シーンにおいて、ユーザの注意対象を抽出した結果を枠で示したものである。ユーザの顔姿勢(注意の向き)に対応して、ユーザが注意を向けている物体が的確に抽出されていることが分かる。

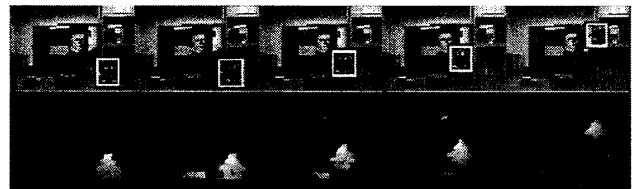


Fig.4: Result of forming human-robot shared attention. Upper: Input image sequence, in which the CG model represents the user's face orientation and a white rectangle indicates an object noticed by him; Lower: The saliency maps.



Fig.5: Result of forming human-robot shared attention.

6. むすび

本論文では、ユーザとロボットとの共同注意の形成メカニズムを提案した。共同注意の形成アルゴリズムを利用することによって、人間・ロボット系では、人間が作業している時の視線停留点の分布と停留時間、その軌跡などから、ユーザが注意を向けている情報を特定して、インタラクションをより柔軟に行うことが可能となる。

参考文献

- [1] R. Stiefelhagen, J. Yang, and A. Waibel: "Estimating focus of attention based on gaze and sound," Proceedings of the Workshop on Perceptive User Interfaces (PUI'01), Nov. 2001.
- [2] G. R. Bradski: "Computer vision face tracking for use in a perceptual user interface," Intel Technology Journal, Q2, 1998.