Proceedings of the 29th Annual Conference of the Robotics Society of Japan
1A3-4, September 2011

注視領域における時間的顕著性を用いた目標・経路指向動作の識別

○福嶋雄基 (阪大院)　長井志江 (阪大院)　浅田稔 (大阪大学，JST ERATO)

1.　はじめに

ロボットが模倣能力を獲得することは，専門家でない人でも，ロボットに行わせたいタスクを，実際に呈示することによって，教えることが可能になるという点で重要であると考えられる．動作には，物体の把持や設置のように，手先などの経路ではなく，動作によって環境に生じる変化が重要な動作（目標指向動作）や，バイバイのように運動経路が重要な動作（経路指向動作）があり，動作によって重要な要素が異なっている．従って，目標指向動作と経路指向動作を区別して認識することは，ロボットがその動作の意味を理解するために重要であると考えられる．そのため，ロボットは呈示者の動作だけでなく，その動作によって周囲に生じる環境の変化も観察する必要がある．

模倣の従来研究における動作認識の手法を見てみると，呈示される動作と物体に生じる変化を全て想定し，手・物体の位置情報や，物体のエッジ情報に基づいて動作を定義しておくことで，呈示された動作を認識する方法 [1] がある．しかしながら，この手法では全ての動作をあらかじめ定義しておく必要があるため，かなり限定された場面でしか機能しない．また，操作物体の状態変化に基づいて動作は定義されているため，経路指向動作については認識できない．他の動作認識の手法としては，手・物体の色情報や, あらかじめマーカをつけることで物体の位置や運動の軌跡を追跡し，動作を認識する方法 [2, 3, 4] がある．しかし，この方法では注視対象を物体や呈示者の手の位置情報などに固定している為，動作によって生じる環境の変化を観察することができない．従って，目標指向動作は認識できない．

これらの研究から，ロボットは注視対象を物体の色やマーカなどのタスクに依存する情報を用いて選択するのではなく，自身の知覚情報に基づいて自律的に選択することで，動作だけでなく環境の変化にも対応できる動作認識を行うこと必要があると考えられる．Demiris et al.[5] は視覚的顕著性モデル [6] を用いた階層型学習モデルを提案した．視覚的顕著性モデルとは人間の低次の注意システムをモデル化したもので，入力画像の色相，エッジ，フローなどの低次特徴量について，周囲刺激との比較を行うことで空間的に顕著な領域を検出し，注視対象とする．このモデルでは顕著性を用いて注視点を決定し，その点における情報を元に動作を学習・認識を行っている．しかしながら，このモデルで選択される注視点は画像の微小領域にすぎず，観察可能な範囲がかなり限定されてしまうため，動作によって生じる環境の変化を観察するのは困難である．従って，注視対象を動作の周囲の情報も観察できるような領域として捉えることができるメカニズムが必要であると考えられる．

Nagai[7] は，顕著性に基づいて注視領域を選択することによって，物体操作タスクにおいて動作と手・物体の対応関係の学習メカニズムを提案した．注視領域は物体や呈示者の手の色を指定して選択していないため，動作や周囲の環境に応じて注視対象を動的に変化させることが出来る．例えば，物体を把持する動作が呈示された場合，把持する前は手のみを注視し，物体を把持した後は物体と手の両方を注視することができる．この注視領域内において，動きや物体について時間的連続性を評価している．しかし，この連続性評価は時間的に連続するフレーム間でしか行われておらず，一瞬の環境の変化しか検出することが出来ない．

そこで本研究では，Nagai[7] のモデルで，空間的顕著性により選択された注視領域において，様々な時間スケールで各低次特徴量の比較を行うことで，時間的顕著性の計算を行う．これにより，注視領域内の一瞬の変化から数フレームにわたる変化まで様々に対応することができる．時間的顕著性が大きくなる時刻において動作を分節化し，その時刻において顕著に変化している低次特徴量を検出する．これにより，フローのみが顕著に変化する動作を経路指向動作，色が顕著に変化する動作を目標指向動作と定義することで，これらの動作を識別できる手法を提案する．また，この手法を簡単な物体操作タスクの動画に対して適用することにより有効性を検証する．

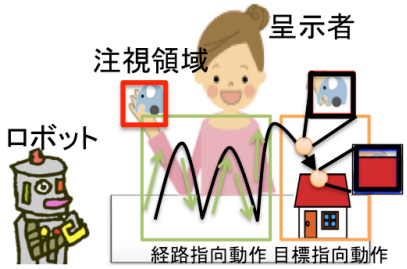


図 1 呈示動作からの目標・経路指向動作の識別

2.　前提条件と問題設定

本研究で提案する手法は以下のような前提条件のもとで稼働する．

- ロボットは，画像全体ではなく，注視領域内のみで動作を観察する．
- 呈示動作は複数の指向性を持つものとする．
- 経路指向動作は，注視領域内の動作のみが顕著に変化する動作とする．
- 目標指向動作は，注視領域の色分布の変化を特徴とする動作を扱う．

以上の前提条件のもと，呈示動作において目標・経路指向動作を識別できるシステムを実現する．図 1 は物体操作タスクを呈示した場合の例である．呈示者は手に持っている物体を，黒矢印のようにホップさせて動かし，家に入れる．このタスクにおいて，黄緑色の枠の動作のように，ホップさせる動作はその経路が重要であると考えられるため，経路指向動作である．また，オレンジ色の枠のように，運動経路よりも家に入れるという環境の変化が重要であるため，目標指向動作である．

注視領域内において観察できるこれらの動作の特徴として，図 1 の黄緑色の矢印のように，経路指向動作では動きが大きく変化しており，他の色や形の大きな変化は見られない．これに対し，目標指向動作では，図 1 の黒枠のように，手に持っていた人形が突然見えなくなることによって，注視領域の色分布が変化する．このように，動作によって，その視覚的特徴が異なることをを利用することで，目標・経路指向動作の識別を行う．

3. 目標・経路指向動作の識別手法

図 2 に目標・経路指向動作の識別を行う手法の概要を示す．空間的顕著性によって，注視領域を選択し，その領域内において低次特徴量の時間変化を計算することで時間的顕著性を計算する．これより，動作や動作の結果生じる環境の変化を検出でき，顕著な変化の生じた時間で動作を分節化する．その時間におけるフロー・色相の時間的顕著性を観察することで目標・経路指向動作を識別する．以下に詳細な計算手順を示す．

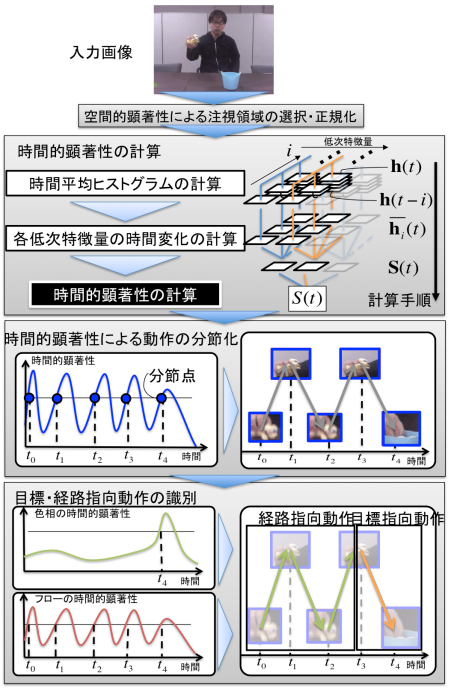


図 2 メカニズム概要

3·1 空間的顕著性による注視領域の選択・正規化

まず，入力画像から空間的顕著性 [6] を計算する．空間的顕著性とは，入力された視覚画像の明度，色相，エッジ，フリッカー（明度の時間差分），フローの 5 つの低次特徴量について，周囲刺激との比較を行うことによって, 画像上において顕著な部分を表すマップである．このマップから最も顕著性の高い部分を注視点として選択し，この注視点の周囲で顕著性の類似性を評価することで注視領域を決定する．この注視領域の選択方法は Nagai のモデル [7] を参考にしている．

時刻 t における注視領域内の $w_0 \times h_0$ の低次特徴量ベクトルを $\boldsymbol{R}(t)$ とする．使用する低次特徴量について，色相は赤・緑, 青・黄の差異の 2 つの要素で表され，明度, フリッカーは 1 つの要素で表される．また，エッジは $\{0°, 45°, 90°, 135°\}$，フローは，$\{0°, 90°, 180°, 270°\}$ のように 4 方向の要素で表される．したがって，$\boldsymbol{R}(t)$ は $w_0 \times h_0$ の画像行列を要素とする 12 次元のベクトルで表される．

ここで，$\boldsymbol{R}(t)$ に対して，横幅 w, 高さ h になるよう正規化を行う．これにより，各低次特徴量をヒストグラム化をする場合のビン値の総和を揃えることが出来る．ここでは $w = 100, h = 100$ とした．正規化した $\boldsymbol{R}(t)$ をヒストグラム化したものを $\boldsymbol{h}(t)$ とする．

3·2 時間的顕著性の計算

次に，図 2 の 3 段目のように，$\boldsymbol{h}(t)$ について，過去 i フレーム間の時間平均ヒストグラム $\overline{\boldsymbol{h}}_i(t)$ を以下の式で計算する．

$$\overline{\boldsymbol{h}}_i(t) = \frac{1}{i} \sum_{k=t-1}^{t-i} \boldsymbol{h}(k) \tag{1}$$

ただし，ここでは $i = 1, 2, 5, 10$ とした．次に注視領域内の各低次特徴量における時間変化を $\boldsymbol{h}_{i_1,i_2}$ として，以下の式で計算する．

$$\boldsymbol{h}_{i_1,i_2} = |\overline{\boldsymbol{h}}_{i_1}(t) - \overline{\boldsymbol{h}}_{i_2}(t)| \tag{2}$$

ただし，$\overline{\boldsymbol{h}}_0(t) = \boldsymbol{h}(t), i_1, i_2 \in \{0, 1, 2, 5, 10\}, i_1 > i_2$ とする．従って，(i_1, i_2) の組み合わせは $(1, 0), (2, 0), (2, 1), (5, 0), (5, 1), (5, 2), (10, 0), (10, 1), (10, 2), (10, 5)$ の 1 0 通りである．

最後に $\boldsymbol{h}_{i_1,i_2}$ のビン値の総和を h_{i_1,i_2} とすると，各低次特徴量の時間的顕著性ベクトル $\boldsymbol{S}(t)$ は，以下のように計算される．

$$\boldsymbol{S}(t) = \sum_{i_1} \sum_{i_2} h_{i_1,i_2} \tag{3}$$

また，$\boldsymbol{S}(t)$ における低次特徴量の時間的顕著性の総和を $S(t)$ とする．今後，単に時間的顕著性を言う場合，$S(t)$ のことを表すこととする．

3·3 時間的顕著性による動作の分節化

次に，時間的顕著性 $S(t)$ の値に対して閾値を引くことで，注視領域内において，視覚的に顕著な変化があったか否かを判別する．顕著な変化が生じていた場合，動作を分節化する．ただし，分節化を行うのは，図 2 の 4 段目左図のように，注視領域内に変化の生じていない状態から最初に顕著な変化の生じた時間である．なお，図中の 4 段目右図のグラフは例として，前章でも示した物体操作タスクを実際に行った場合，どのように動作の分節化が起こると想定されるかを示している．

3·4　目標・経路指向動作の識別

分節化された時間において，顕著に変化している低次特徴量を調べる．顕著な変化は前節と同様にして，各低次特徴量の時間的顕著性に対して閾値を引くことによって判別する．分節化はその直前の動作によって生じているため，この時刻の特徴量の変化は直前の動作の特徴を表している．ここで，フローのみが顕著に変化している動作を経路指向動作と定義する．また，少なくとも色相が顕著に変化している動作を目標指向動作と定義する．これらの定義に基づき，目標・経路指向動作の識別を行う．

例えば，手振りを考えると，注視領域には，手のみが入っており，動きは顕著に変化するが，領域内の色分布は大きな変化しない．従って，フローの変化が顕著になるため，手を振る動作は経路指向動作であると識別される．一方，物体を把持する動作を考えると，物体を把持する前は手のみが注視領域に入っているが，物体を把持した瞬間に注視領域に物体が入ることによって，領域内の色分布が顕著に変化する．従って，色相の変化が顕著になり，物体の把持は目標指向動作と識別できる．(図 2 5 段目参照)

また，一連の呈示動作に対して, 目標・経路指向動作の識別を行うのではなく，分節化したそれぞれの動作に対して識別を行うこととする．これは，別のタスクが呈示されたときでも，部分的に共通すると考えられる動作が含まれている場合には，今回分節化した動作の情報を用いることが出来ると想定される為である．

4.　物体操作タスクを用いた実験

実際にタスクを呈示している 4 つの動画に提案手法を適用することによって，本システムの有効性を検証する．図 3 のように，呈示者は人形を持ち，机の上を (a) ホップさせて移動（ホップ条件），(b) 机に平行に移動（スライド条件），の 2 通りの方法で動作を呈示する．また，実験環境も，図 3 のように，(1) 水色の容器がある状態（バケツあり条件）と，(2) 人形以外の物体は移っていない状態（バケツなし条件）の 2 通り条件を考える．バケツあり条件の場合，呈示者は上で説明したような 2 通りの移動方法のあと，容器へその人形を入れる．また，バケツなし条件では，バケツあり条件で容器が設置されている位置まで移動して動きを止める．以上より，ホップ・バケツあり条件，スライド・バケツあり条件，ホップ・バケツなし条件，スライド・バケツなし条件の 4 つの条件で実験を行う．これらの条件において，ホップ・バケツあり条件とスライド・バケツあり条件では，ホップ・スライド動作が経路指向動作に対応し，バケツに入れる動作は目標指向動作に対応している．ホップ・バケツなし条件，スライド・バケツなし条件では，ホップ・スライド動作から停止まで，全体の動作が経路指向動作に対応している．

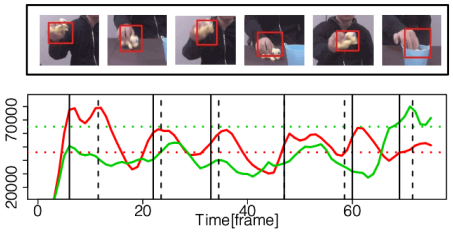
5.　時間的顕著性の計算結果

注視領域における時間的顕著性の計算結果を図 4, 5, 6, 7 に示す．各図中の一番上の枠は，動作の呈示中に実際に選択された注視領域が赤い枠で示されている．また，２つの折れ線グラフはそれぞれ (a) 赤色がフロー，

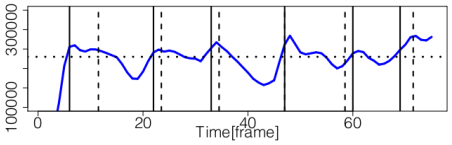
(a) バケツあり・ホップ条件 (b) バケツあり･スライド条件

図 3 実験に用いた呈示タスク

緑色が色相の時間的顕著性，(b) 青色が時間的顕著性 ($S(t)$) を表している．横軸が時間であり，縦軸が変化量である．また，横軸に平行な鎖線は閾値であり，(a) においては色の示す特徴量と閾値が対応している．この閾値の値は実験条件によらず，同じ値を用いている．縦軸に平行な破線は，実際の動画において動きが顕著に変化した時間であり，実験者が決めている．実線はシステムによって，分節化された時間を示している．なお，ノイズを低減する為，移動平均をとっている．

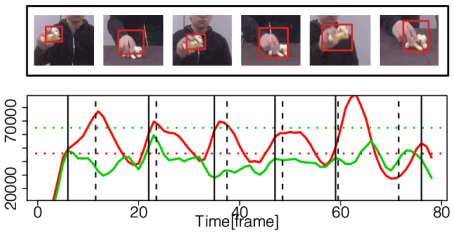


(a) 色相・フローの時間的顕著性

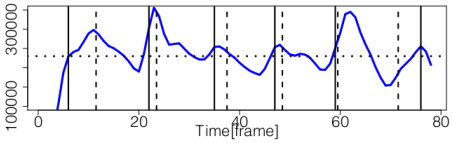


(b) 時間的顕著性

図 4 ホップ・バケツあり条件における実験結果



(a) 色・フローの時間的顕著性



(b) 時間的顕著性

図 5 ホップ・バケツなし条件における実験結果

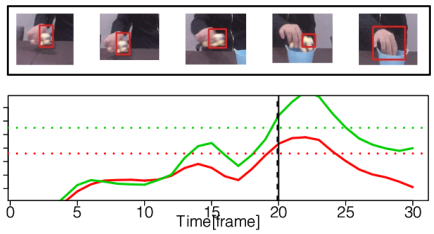
6. 考察

図 4 をみると，ホップ動作により動きが変化するときと，最後にバケツに人形を入れるときに，時間的顕著性が大きくなっており，それぞれで動作が分節化されている．また，ホップ動作では，いずれの分節点においても，フローのみが顕著に変化している特徴量となっているのに対し，バケツに人形を入れる部分では，フロー・色相の両方が顕著に変化した特徴量となっている．従って，ホップ・バケツあり条件では，ホップ動作において，分節化されたそれぞれの動作が経路指向動作と識別され，最後にバケツに入れる動作のみが目標指向動作であると識別される．

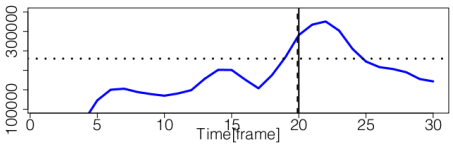
図 5 では，ホップ・家あり条件と同様にホップ動作は分節化され，最後に停止するときにおいても分節化が生じる．このとき，色相・フローの時間的顕著性はいずれの分節点においてもフローのみが顕著に変化した特徴量と判別されている．従って，ホップ・バケツなし条件では，分節化された全ての呈示動作が経路指向動作であると識別される．

一方，図 6 をみると，スライド動作では分節化が生じず，最後にバケツに人形を入れる時にのみ分節化が起こっている．また，この時間において色彩・フローの両方が顕著に変化したと判別されているため，スライド・バケツあり条件では，スライド動作からバケツに入れるまでの動作が一つの目標指向動作と識別される．最後に，図 7 をみると，スライド・家あり条件と同様に，スライド動作では分節化が起こらず，動作の最後に動きを停止したときにのみ，動作が分節化される．この時間における顕著に変化した特徴量はフローのみである為，スライド・家なし条件では，呈示動作全体が 1 つの経路指向動作と識別される．

以上のように，分節化された動作毎に識別を行うことで，例えば，入れる動作が含まれる別のタスクが呈示された場合，今回のタスクの動作の情報を用いるができれば，動作の認識率が向上する可能性がある．



(a) 色相・フローの時間的顕著性

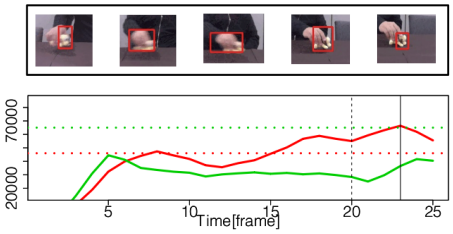


(b) 時間的顕著性

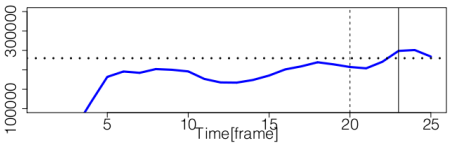
図 6 スライド・バケツあり条件における実験結果

7. まとめ

目標・経路指向動作を識別する能力は動作学習を行う際に重要であると考えられる．本研究では，空間的な顕著性に基づき注視領域を選択し，更にその領域内において，時間的顕著性を計算することによって，簡単な物体操作タスクにおいて，目標・経路指向動作の識別を実現した．今回は目標指向動作として，色相の時間変化のみを用いたが，今後はエッジなどの低次特徴量も用いることで，対応できる動作を拡張する．



(a) 色相・フローの時間的顕著性



(b) 時間的顕著性

図 7 スライド・バケツなし条件における実験結果

参考文献

[1] Y. Kuniyoshi, M. Inaba, and H. Inoue. Learning by watching: extracting reusable task knowledge from visual observation of human performance. *Robotics and Automation, IEEE Transactions on*, Vol. 10, No. 6, pp. 799–822, December 1994.

[2] A. Billard, S. Calinon, and F. Guenter. Discriminative and adaptive imitation in uni-manual and bi-manual tasks. *Robotics and Autonomous Systems*, Vol. 54, No. 5, pp. 370–384, 2006.

[3] S. Calinon, F. Guenter, and A. Billard. Goal-directed imitation in a humanoid robot. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 299–304, 2005.

[4] R. Yokoya, T. Ogata, J. Tani, K. Komatani, and H.G. Okuno. Discovery of other individuals by projecting a self-model through imitation. In *Intelligent Robots and Systems, 2007. IEEE/RSJ International Conference on*, pp. 1009–1014, 2007.

[5] Y. Demiris and B. Khadhouri. Hierarchical attentive multiple models for execution and recognition of actions. *Robotics and Autonomous Systems*, Vol. 54, No. 5, pp. 361–369, 2006.

[6] L. Itti, N. Dhavale, and F. Pighin. Realistic avatar eye and head animation using a neurobiological model of visual attention. In *Proc. SPIE 48th Annual International Symposium on Optical Science and Technology*, pp. 64–78, 2003.

[7] Y. Nagai. From bottom-up visual attention to robot action learning. In *Proceedings of the 8th IEEE International Conference on Development and Learning*, pp. 1–6, 2009.