3O1-OS-3a-5

連想に基づいた心象映像表出によるエージェントの内部状態表示

Displaying robot's internal state in the form of a mental image

小川　貴弘*1　藤原　菜々美*1　尾関　基行*1　岡　夏樹*1
Takahiro Ogawa　Nanami Fujiwara　Motoyuki Ozeki　Natsuki Oka

*1京都工芸繊維大学大学院　工芸科学研究科
Graduate School of Science and Technology, Kyoto Institute of Technology

We have been developing a system based on the concept of "See-through working memory" which displays an internal state of the robot in the form of a mental image. Our current system is composed of simple modules —working memory, visual attention, long term memory, and mental image display— for the robot in an idling state. In this paper, we focus on the association method to recall visual information from the long term memory to the working memory. Although the method is the simplest one, the recalled information seem to be useful for a person in front of the robot. We show some examples suggesting the possibility of our system in the experiment.

1. はじめに

展覧会や企業広告のための高機能なロボットが世間を騒がせるようになって久しいが，一般家庭や身近な場所でロボットが活躍する姿はほとんど見かけない．費用対効果や認識技術・人工知能技術の問題も大きいが，ここではロボットの情報伝達能力の問題に着目する．ロボットを普及させるにはコストダウンを図る必要があるが，人同士のコミュニケーションで重要な役割を果たすノンバーバルな情報（表情や韻律など）を表出するための機能や構造がまず省略される．このことは人との対話を通して各種のサービスを提供するタイプのロボット（以降，"エージェント" と呼ぶ）の場合は特に問題であり，エージェントの言動の思惑（"内部状態"）が相手に伝わりにくく，対話が円滑に進まなかったり，時には危険を招くことも考えられる．よって，できるだけ廉価なデバイスによって効果的にエージェントの内部状態の伝達する手段を検討する必要がある．

スピーカや LED パネルは最も廉価なタイプのエージェントにも装備可能なので，エージェントの内部状態を音声やテキストメッセージなどのバーバルな情報で伝えることも考えられる．しかし，内部状態は必ずしも相手に伝わる必要はなく，相手が必要とするときにだけ伝わればよい．言い換えれば，相手が無視したいときには無視できることが重要である．バーバルな伝達手段は相手に情報を明確に伝えたいときに使われるものであり，そこに内部状態の情報を織り交ぜると無視することができず情報過多となる．ノンバーバルで廉価な伝達手段として，LED やビープ音を使った研究がある [小松 10]．しかし，これらは人同士で普段やり取りしているノンバーバル情報とかけ離れているため，エージェントの擬人化の効果を損なう（道具であることが強調される）恐れがある．人からエージェントが擬人化されることは，人同士のコミュニケーション技術を転用するための重要な前提条件である．

この問題に対して，尾関らは，エージェントに搭載した小型ディスプレイに内部状態を映像化して表出する "シースルーワーキングメモリ" というコンセプトを提案している [尾関 10]．映像を伝達手段に使うため擬人化の効果を損ないかねないが，普段から我々が経験している「心象」を思わせるような映像表現を用いることでこの問題を軽減する．シースルーワーキングメモリは，エージェントを動かしているプロセスの内部状態（現在処理している内容）を画像の形で受け取って映像化するという情報の流れを規定するものである（図 1）．内部状態を画像として受け取ると同時にその画像の時制と直前の内部状態との関係性を受けとり，それらに対応した映像効果を適用して心象映像を生成する．

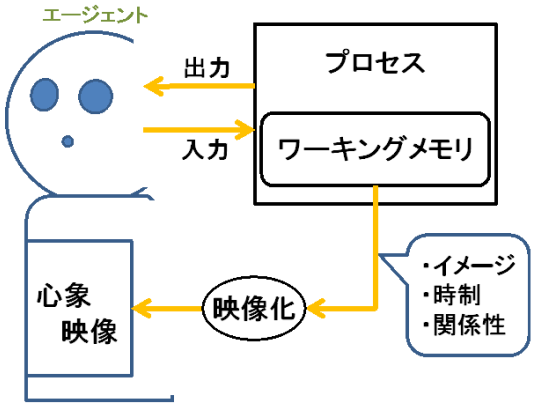

図 1: シースルーワーキングメモリのコンセプト

このシースルーワーキングメモリのコンセプトに則り，藤原らは，エージェントが注視した物体の視覚情報をワーキングメモリに保持し，そのイメージ（背景から切り出した物体の画像）を単純に表出するシステムを構築している [藤原 11]．前述のシースルーワーキングメモリの枠組みでは，イメージの時制として「現在」，直前のイメージとの関係性として「共起」のみを扱ったものとなる．このエージェントは USB カメラとディスプレイのみで構成され，シースルーワーキングメモリの一実装という位置づけで作られた．そのため，目的を伴った行動というものはなく，目立つ場所に注意を向け，飽きたら他の場所に注意を移動する．

本研究では，「過去」の時制を持ったイメージも扱えるように藤原らのシースルーワーキングメモリの実装を拡張する．藤原らのエージェントに長期記憶（視覚に関する宣言的記憶）を加え，長期記憶からワーキングメモリに視覚情報を呼び出す仕組みを実装する（図 2）．本手法では，ワーキングメモリ内の

連絡先: 京都工芸繊維大学情報工学部門
〒 606-8585 京都市左京区松ヶ崎橋上町
E-mail: ogawa@ii.is.kit.ac.jp, ozeki@kit.ac.jp

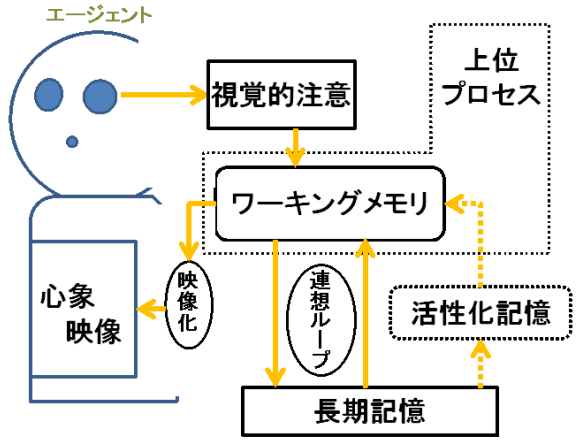

図 2: 構築したシースルーワーキングメモリの実装

視覚情報をクエリとして長期記憶にある視覚情報を次々とワーキングメモリに呼び出す（連想する）ことで，芋づる式に過去の記憶を表出する．本研究も藤原らの実装と同じくシースルーワーキングメモリの一実装という位置づけであり，エージェントは具体的な目的を伴った行動や連想は行わない．連想手法としても最も単純な実装方法である．しかし，「同時期に注視したものを関連付けて記憶し，芋づる式に連想してくる」だけでも，目的を持ったエージェントに役立ちそうな情報が得られそうなことがわかってきた．本実験ではこの点について実例を示しながら議論する．

本稿では，シースルーワーキングメモリの機能のうち，長期記憶からの連想に焦点を絞って議論する．2. では，シースルーワーキングメモリについて述べた後，その一実装として現在我々が構築しているシステムについて説明する．3. では長期記憶からの連想手法を概説し，4. では実例を通して本手法の可能性を議論する．

2. シースルーワーキングメモリ

シースルーワーキングメモリとは，エージェントの内部状態を心象を模した映像（心象映像）として相手に表出するという考え方である．ここでいう内部状態とは，エージェント内のプロセスが処理中のデータや現在の状態，次の行動などを指し，画像で表現できるものを対象とする．何らかの画像処理を行うエージェントであればその処理過程がそのまま利用できる．画像処理を行わないエージェントでも，状態や行動が有限である場合は，それらに対応した画像を用意しておけばシースルーワーキングメモリの枠組みに乗せることができる．

エージェントの内部状態を心象映像としてうまく表現するには，内部状態を示す画像だけでなく，その画像を特徴づける付加情報が必要である．例えば，現在の処理しているのが「記憶の検索結果」であれば，その画像は過去の記憶に見えるよう加工したほうがよい．同様に，直前の内部状態の結果として現在の内部状態が決まったのだとすると，「結果」を思わせる画像切り替え表現にしたい．幅広いエージェントを対象とするには，画像を特徴付ける付加情報をできるだけ汎用的なものにしておく必要がある．現時点では，付加情報として “時制” と “関係性” というカテゴリを用意し，それぞれフィルタ効果（ガウシアンフィルタなど，画像単体に対する処理）とトランジション効果（クロスフェードやワイプなど，画像切替に対する処理）に対応させている．

時制： 現在…注視している外界の物体や現在の状態など，過去…記憶の検索結果や現在の行動の原因となった事象など，未来…次の行動や現在のタスクのゴールなど．

関係性： 共起…時間的もしくは空間的に同時に存在する，結果…直前の内部状態の結果として現在の内部状態がある，理由…直前の内部状態の理由として現在の内部状態がある，その他，類似など．

以上のようなシースルーワーキングメモリのコンセプトに則って我々が構築しているシステムを次に紹介する．システムの概要は図 2 に既に示した．本システムは，視覚的注意・長期記憶・ワーキングメモリ・心象映像生成部で構成されており，現時点ではエージェントに具体的なタスクを行わせるための上位プロセスは組み込んでいない．つまり，エージェントが特に目的を持たない時のアイドリング状態であり，目立つものを飽きるまで注視し，過去に同時期に注視したものを単純に連想している[*1]．

以下，各モジュール・機能について説明する．

視覚的注意： エージェントが注視している領域（物体）を視覚的注意モデルによって選び出し，背景から切り出す．切り出した部分を “イメージ” と呼ぶ．視覚的注意モデルには Itti らの Saliency Map と粒子フィルタを組み合わせた尾関らの手法 [尾関 11] を，背景からの切り出しには OpenCV の Grabcut[Rother 04] に基づく手法を用いている．上位プロセスからの指令がない限り，目立つところを飽きるまで注視し，飽きたら付近の目立つ場所に注意を移動させるという処理を繰り返す．

ワーキングメモリ： 注視された物体，若しくは連想によって呼び出された物体の視覚情報が格納され，一定時間保持される．視覚情報は最大で四つまで保持することができ，四つ以上になると一番古い視覚情報が削除される．注視／連想された物体が既にワーキングメモリにある場合は，新しく入ってきた物体として順位が更新される．現時点では視覚情報として，心象映像の素材となるイメージと，物体同定に用いられる画像特徴量を扱っている．

長期記憶： 注視によって初めてワーキングメモリに入ってきた物体の視覚情報が記録される[*2]．その際，同時にワーキングメモリに含まれている物体との関連性も長期記憶に保持される．詳しくは 3. で述べる．

心象映像生成部（映像化・心象映像）： ワーキングメモリに保持されている視覚情報からイメージを順に受け取り，フィルタ効果とトランジション効果を適用して心象映像を生成する．シースルーワーキングメモリの核となる機能であるが，本稿では長期記憶からの連想に焦点を絞る．

以下のモジュールは現時点では実装していないが，目的を持って行動するエージェントを構築するためには必要となる機能である．図 2 では破線で示されている．

活性化記憶： 人工知能では必要な情報を必要なときにだけデータベース（長期記憶）から検索することが普通だが，本研究では，長期記憶に対する連想がバックグラウンドで

*1 上位プロセスからの指令を受けるためのインタフェースは各モジュールで備えている．
*2 本来は注視されるたびに画像特徴量が追加・更新されるべきだが，現在の実装はそこまで至っていない．

働いており，そうして収集（活性化）された記憶に対して人工知能が検索をかけるという構造を想定する．容量はワーキングメモリよりも大きく，人とのコミュニケーションに差支えのない程度（数秒以内）に検索できる程度の大きさとする．現在は活性化記憶を実装していないため，連想された視覚情報は直接ワーキングメモリに格納される．

上位プロセス： エージェントに目的を持った行動をさせるためのプロセスであり，シースルーワーキングメモリの上位に位置する．ワーキングメモリは，本来，この上位プロセスに含まれる．図 1 の定義では，視覚的注意と長期記憶，活性化記憶もシースルーワーキングメモリの上位プロセスに含まれるが，本研究ではこれらを意図的に区別している*3．エージェントの目的や状況に応じて，この上位プロセスが各モジュールをトップダウンに制御する．例えば，アイドリング状態では「同時期に注視したもの」を連想しているが，上位プロセスの指示によって「形が似ているもの」「同じ用途に使えるもの」などが連想されて活性化記憶に格納されていく．

3. 長期記憶からの連想手法

長期記憶には，各視覚情報の ID・イメージ・特徴量がセットで記憶されている．特徴量は，注視された物体と記憶されている物体を同定するためのもので，イメージから求められた局所特徴量（SURF[Herbert 08]）を用いている．

視覚情報に加えて，長期記憶には，各視覚情報の間の関連度を数値で記録したリスト（関連度リスト）が複数保持されている．連想のトリガがかかると，関連度リストの一つが選ばれ，ワーキングメモリの先頭に保持されている視覚情報との関連度が最も大きい視覚情報がワーキングメモリの先頭に呼び出される．関連度として共起の度合いや各種の類似度など様々なものが考えられるが，現時点では，共起の度合い（ワーキングメモリに同時に含まれていた時間の長さ）のみが実装されている．

関連度の更新は次のように行われる．まず，注視された物体の視覚情報が長期記憶に存在しない場合，その視覚情報に関するエントリが関連度リストに追加される．ワーキングメモリに保持されていた視覚情報が古い情報として消える際，それまで共にワーキングメモリに含まれていた視覚情報との間の関連度を更新する．関連度の大きさは同時にワーキングメモリに存在していた時間に比例した値とする．ただし，ここで関連度が更新されるのは，注視によってワーキングメモリに格納された視覚情報と，上位プロセスによって用いられた視覚情報のみとする．つまり，連想によって呼び出された視覚情報同士の関連度は更新しない．

連想のトリガは一定時間毎にかかり，注視処理とは独立してワーキングメモリの先頭に視覚情報を呼び出す．一方，同じ物体が注視されている間は，その視覚情報がワーキングメモリに送られ続け，ワーキングメモリの先頭を独占しようとする．ただし，同じ物体を注視し続けている時間に比例して視覚情報をワーキングメモリに送る間隔を長くするため，次第に連想によって呼び出される視覚情報がワーキングメモリを占めるようになる．別の物体に注意が移ると，またしばらくの間はその物体の視覚情報がワーキングメモリの先頭を独占する．

*3 つまり，視覚的注意及び長期記憶と連想の機能は，エージェントの目的に依存した上位プロセスとは独立して，共通して利用できるモジュールであると我々は考えている．

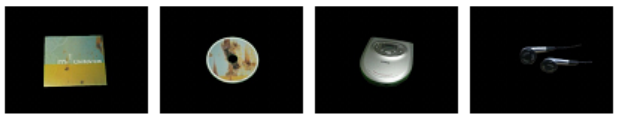

図 3: プレーヤーで音楽を聴くタスクの画像と注視時間

4. 実験

4.1 目的と方法

「ワーキングメモリに同時期に保持されていた」という関連度のみでどのような連想が行われるかについて調べる．本実験では連想手法のみを検証するため，エージェントへの入力に視覚的注意モジュールを用いず，手作業で背景から切り抜いたイメージを用意し，注視したものとして直接ワーキングメモリに与えた．例えば，コーヒーを用意する様子をエージェントが見ていたと想定する場合，コーヒー豆→コーヒーサーバー→やかん→コーヒーカップ→…という順番で，用意しておいたイメージを適当な時間ずつ注視したとしてワーキングメモリに入力していく．よって，視覚的注意モジュールを用いた場合に起こる諸問題（見るべき対象を注視しない，物体がきれいに切り出せない，物体の見えが変わって同定に失敗するなど）も今回の実験では無視できる．

実験手順は次のとおりである．まず，事前に一定の入力をエージェントに与えて長期記憶を構築する．入力は，前述のコーヒーの例のように，エージェントがある作業を見たり実施したことを想定して作られた一連の画像群（以下，記憶タスクと呼ぶ）を複数用意した．長期記憶を構築した後，特定のタスク（実験タスクと呼ぶ）を入力して，その連想結果を観察した．

事前に記憶させておいたタスクと実験タスクの関係は次の 3 通り用意した．

設定 1： 記憶タスク＝実験タスク

設定 2： 記憶タスク＝実験タスクを含む 2 個のタスク

設定 3： 記憶タスク＝実験タスクを含む 300 個のタスク

これらのタスクは，筆者の一人の 1 週間の行動を洗い出し，日常的に使われている物体を約 190 個用いて作成した．一つのタスクで注視する物体数が 4 個± 3 個になるようにタスクを区切り，注視する時間は実際の作業を想定して与えた．タスクとしては，例えば，アイロンをかける，歯磨きをする，りんごを切る，ホットケーキを食べる，パフェを食べるなどである．以降では，プレーヤーで音楽を聴くタスクとコーヒーを用意するタスクを実例として取り上げて議論する．

4.2 記憶タスク＝実験タスク

プレーヤーで音楽を聴くタスクのみを事前にエージェントに与え，その後，同じタスクをエージェントに入力したときの連想結果を観察した．このタスクで入力した画像と注視時間を図 3 に，連想結果を表 1 に示す．注視の合間を縫って連想された視覚情報がワーキングメモリに呼び出されている様子がわかる．同じ物体を注視している時間が長くなるほど，連想によって呼び出された視覚情報がワーキングメモリを占めるようになる．

同様に，コーヒーを用意するタスクを記憶させてコーヒーを用意するタスクを入力として与えると，「コーヒー豆→コーヒー

	ワーキングメモリ			
イベント	WM1	WM2	WM3	WM4
ケースを注視	ケース			
CD を連想	CD	ケース		
ケースを注視	ケース	CD		
CD を連想	CD	ケース		
プレーヤーを連想	プレーヤー	CD	ケース	
ケースを注視	ケース	プレーヤー	CD	
CD を連想	CD	プレーヤー	ケース	
CD を注視	CD	プレーヤー	ケース	
プレーヤーを連想	プレーヤー	CD	ケース	
CD を注視	CD	プレーヤー	ケース	
プレーヤーを連想	プレーヤー	CD	ケース	
イヤホンを連想	イヤホン	プレーヤー	CD	ケース
CD を注視	CD	イヤホン	プレーヤー	ケース
プレーヤーを連想	プレーヤー	CD	イヤホン	ケース
プレーヤーを注視	プレーヤー	CD	イヤホン	ケース
イヤホンを連想	イヤホン	プレーヤー	CD	ケース
プレーヤーを注視	プレーヤー	イヤホン	CD	ケース
イヤホンを連想	イヤホン	プレーヤー	CD	ケース
プレーヤーを注視	プレーヤー	イヤホン	CD	ケース
イヤホンを連想	イヤホン	プレーヤー	CD	ケース
イヤホンを注視	イヤホン	プレーヤー	CD	ケース

表 1: プレーヤーで音楽を聴くタスクの連想結果

サーバー→やかん→コーヒーカップ→砂糖→スプーン→牛乳→…」というように連想が進んでいく．単純な連想であるが，例えば飲み物をサーブしてくれるロボットがこのような映像を表出していれば，相手は（ゆっくりと動く）ロボットがどこまで用意しようとしてくれているのか推測することができる*4．

4.3 記憶タスク＝実験タスクを含む 2 個のタスク

コーヒーを用意するタスクとパフェを食べるタスクを記憶させた後，コーヒーを用意するタスクを入力した．前述の実験と同様，「コーヒー豆→コーヒーサーバ→…」と連想が進んでいくが，「砂糖→スプーン→牛乳」ではなく，「砂糖→スプーン→フォーク」という連想になった．これは，パフェをスプーンとフォークで食べたときの関連度のほうが大きかった（注視時間が長かった）からである．これは連想に失敗したとも言えるが，コーヒーをかき混ぜるための代替品を想起したという意味では興味深い結果である．

ただし，より大きな記憶を持たせたときにもこのように何かしら意味のある結果が連想されるかどうかはわからない．実際，2 個のタスクしか記憶させなかった場合でもフォーク以外に多くの連想候補があり，とても混ぜるのに使えないものまで連想される可能性がある．現時点の実装では連想したものをそのままワーキングメモリに呼び出すが，やはり間に活性化記憶を置き（図 2），そこからワーキングメモリに呼び出す視覚情報は上位プロセスの高度な判断に任せなくてはならない．

4.4 記憶タスク＝実験タスクを含む 300 個のタスク

最後に，300 個のタスクを記憶させておいた場合の連想結果について述べる．先ほどと同じコーヒーを用意するタスクを入力すると，「コーヒー豆→コーヒーサーバー→やかん」と連想が進むが，「やかん→お茶の葉→コップ→」というようにお茶の葉に関する連想に飛んでしまう．連想が飛んでしまっても，目の前で起こっているタスク内の物体を注視することによって再び現実に引き戻されるので問題はないが，「やかん」から連想されるのが必ず「お茶の葉」であることは問題である（この問題は前述のスプーン→フォークでも同様）．これはワーキングメモリの先頭の視覚情報だけをクエリとしていることに原因がある．ワーキングメモリには複数の視覚情報が保持されているので，それらの情報を併用した連想を行うことで，より眼の前の状況に即した連想が行われると考えられる．

*4 ただし，表出されている内容がエージェントの次の行動であるとは限らない．表出されている心象映像が単なる記憶探索の過程なのか，次に取ろうとしている行動なのか，別途与えられた時制と関係性に基づいてこれらを区別できる映像表現が必要とされる．

5. まとめ

本研究では，シースルーワーキングメモリのコンセプトの実装として，視覚的注意と長期記憶を持ったエージェントを構築している．本稿では，長期記憶からワーキングメモリに視覚情報を呼び出すための連想部分について説明し，手動で作成したタスクを用いた連想結果をいくつか示した．我々の手法は「ワーキングメモリに同時期に保持された」という関連度による単純な連想であるが，エージェントの前にいる人にもエージェント自身にも役立ちそうな情報が取り出せる可能性を見出すことができた．今後はまず，ワーキングメモリ内の複数の視覚情報から Association rule や N-gram などによって連想する手法に取り組む．更に，活性化記憶モジュールを実装し，具体的な目的を持ったエージェントを構築することでシースルーワーキングメモリの有効性を検証していきたい．

参考文献

[藤原 11] 藤原菜々美, 尾関基行, 岡夏樹: 心象映像によるエージェントの内部状態表現，人工知能全国大会論文集, 2011.

[Herbert 08] Herbert Bay,Andreas Ess,Tinne Tuytelaars,Luc Van Gool: Speeded-Up Robust Features（SURF）, Computer Vision and Image Understanding, vol.110, pp.346-359, 2008.

[尾関 11] 尾関基行, 柏木康寛, 井上茉莉子, 岡夏樹: トップダウン注意制御による人とエージェントの共同注意，HAI シンポジウム, 2011

[小松 10] 小松孝徳, 山田誠二, 小林一樹, 船越孝太郎, 中野幹生: Artificial Subtle Expressions:エージェントの内部状態を直観的に伝達する手段の提案， 人工知能学会論文誌, vol.25, no.6, pp.733-741, 2010

[尾関 10] 尾関基行, 藤原菜々美, 岡夏樹: シースルーワーキングメモリ～エージェントの心象表示による新しいコミュニケーションに向けて～，HAI シンポジウム, 2010

[Rother 04] C.Rother, V.Kolmogorov, A.Blake: Grabcut: interactive foreground extraction using iterated graph cuts, ACM Transaction on Graphics, vol.23, pp.309-314, Aug.2004