

INTEGRANDO VISÃO E COMPORTAMENTO: UMA APLICAÇÃO DE RECONSTRUÇÃO PROPOSITIVA.

Anna H. R. C. Rillo, Reinaldo A. C. Bianchi, Briand Moreira Jr e Fábio Ferraz.

*Divisão de Automação e Inteligência Artificial - Laboratório de Sistemas Integráveis
Universidade de São Paulo (DAIA / LSI - EPUSP)
Av. Prof. Luciano Gualberto, travessa 3, 158.
05508-900 São Paulo, SP, Brasil.
E-mail: arillo-, rbianchi-, briand- e fferraz@lsi.usp.br*

Resumo: Este artigo discute a dicotomia existente entre dois paradigmas dominantes hoje na área de visão computacional, o reconstrutivista e o propositivo. Defendemos que eles são complementares e que devem ser combinados para apresentarem bons resultados, quando aplicados em agentes operando em ambientes complexos. A abordagem combinada recebe o nome de Reconstrução Propositiva e foi testada no contexto de um sistema de visão para um manipulador, realizando tarefas simples, onde as interrelações entre ações, tarefas, comportamentos e intenções foram exploradas, visando otimizar a utilização de uma forte integração entre ação e percepção.

Palavras chave: Visão Computacional, Inteligência Artificial.

Abstract: In this paper we address the reconstructive versus purposive dichotomy in computer vision paradigms. We argue that they are complementary and must be combined to provide a framework for operating in an agent embedded in a complex environment. Such a combined approach is called Purposive Reconstruction. We have tested the approach in the context of vision for a manipulator performing simple tasks in a simple domain, exploring the roles of actions, behaviors and intentions and taking advantage of a tight integration between perception and action.

Key words: Computer Vision, Artificial Intelligence.

1. INTRODUÇÃO.

Existe uma discussão bastante atual na área de visão computacional questionando dois paradigmas dominantes: reconstrutivista e propositivo [2].

O paradigma reconstrutivista [5], [6] foi o primeiro a surgir e dominou por décadas as pesquisas em visão computacional. Segundo este paradigma, o objetivo da visão é “construir uma descrição simbólica de uma cena do mundo real em um modelo, que possa ser usado como interface para outros processos cognitivos (raciocínio, planejamento inteligente de atividades), fazendo isto através de descrições funcionais do mundo visível” [11]. Assim, sua finalidade principal está voltada para a reconstrução da representação completa do mundo. Esta abordagem levou a muitas contribuições (principalmente teóricas), resultando em algumas técnicas matemáticas, por exemplo, relacionadas a descontinuidades e regularização.

O paradigma reconstrutivista justifica a necessidade de representações intermediárias como modo de reduzir o custo computacional da visão, afirmando que comportamentos complexos em ambientes desconhecidos só se tornam possíveis através do uso de representações poderosas. Reconstrutivismo normalmente objetiva visão genérica (isto é, possibilidade de realizar comportamentos complexos) por uma abordagem *bottom-up, data-driven*.

Já o paradigma propositivo [1], [2], [5], [9], [10] acredita que a visão deve ser considerada dentro de um contexto de tarefas que um agente deve realizar. Deste modo, ao tentar solucionar o problema da visão, deve-se questionar o que se deseja reconhecer, levando a uma questão diretamente ligada às tarefas visuais, isto é, ao propósito. Assim, o pensamento propositivo coloca questões cujas respostas levam a soluções de tarefas específicas e não de uso geral, tornando-se um paralelo à teoria computacional proposta pelos reconstrutivistas.

Propositivismo objetiva visão genérica por uma abordagem *top-down*, *goal-oriented*, *task-driven*. Com isso, visão genérica emergirá de uma organização de diversas soluções dedicadas a diferentes tarefas visuais. Assim, o problema se torna como organizá-las e quais são estas tarefas primitivas, voltando o enfoque para arquiteturas de integração dos sistemas visuais.

A principal crítica da comunidade reconstrutivista em relação à propositiva tem sido a de que soluções propositivas não possibilitam fácil expansão para tratamento de problemas mais sofisticados, isto é, questionando a possibilidade de adquirir comportamentos mais complexos com a simples montagem de comportamentos mais simples. Outra questão que surge é quanto à provável duplicação de esforços tentando construir módulos distintos para cada comportamento diverso. Por outro lado, os propositivos argumentam que, com o paradigma reconstrutivista, surge o problema de, por não conhecer o objetivo final da tarefa visual, assumir restrições incoerentes com a tarefa, prejudicando o resultado.

Esta discussão entre reconstrutivistas e propositivos é paralela a outra semelhante que ocorre em outros domínios da Inteligência Artificial. Por exemplo, Brooks [3] tem sido um entusiasta proponente de uma abordagem baseada em comportamentos em AI (*behavior-based view*), a qual é consistente com o paradigma propositivo. A popularidade destas abordagens propositivas tem crescido bastante, uma vez que elas alcançam melhores resultados práticos na interação de robôs inteligentes com ambientes complexos.

Atualmente, a comunidade de visão computacional está entendendo que nem um paradigma nem outro deve ser a resposta para o problema de visão [5], mas sim que eles devem ser integrados e combinados, numa nova abordagem denominada reconstrução propositada (*purposive reconstruction*) [4].

O presente trabalho propõe um compromisso entre ambos paradigmas, através da definição de tarefas primitivas e dos processos que as realizam, juntamente com uma arquitetura que possibilite o controle tanto por parte das informações fornecidas pela tarefa quanto pelos dados, e aplicar esta arquitetura em um agente dotado de um manipulador e percepção visual.

2. UMA APLICAÇÃO DE RECONSTRUÇÃO PROPOSITIVA.

A construção de um sistema inteligente que apresente integração de percepção visual, ação e aprendizado necessita, inicialmente, do estudo, definição e modelamento do mundo e de tarefas primitivas que serão por ele executadas. Permitir a representação das diversas informações sobre o mundo, relevantes para as tarefas a serem realizadas pelo sistema de visão computacional, exige o estudo de modelamento e estratégias de controle da percepção, envolvendo modelamento dos sensores, dos objetos, do domínio e da interação entre esses componentes, para um dado propósito.

Para um primeiro estudo dentro do paradigma de reconstrução propositiva foi definido um agente, possuindo capacidades de ação e percepção visual, com um propósito bastante simples: jogar o Jogo da Velha de modo totalmente autônomo, contra um oponente, no caso, humano. O agente foi implementado em um sistema dotado de um manipulador e visão computacional. O propósito usado foi bastante simples, para permitir que as rotinas visuais e de controle do manipulador fossem fáceis, possibilitando que maiores esforços fossem dispendidos no objetivo principal do trabalho, que é o mapeamento de intenções em comportamentos e decomposição e gerenciamento destes comportamentos. Além disso, vale ressaltar que as ações definidas neste sistema são básicas e necessárias em tarefas de, por exemplo, montagem, manipulação e mobilização.

A reconstrução propositiva é alcançada na medida em que não se busca a reconstrução visual total do ambiente, mas sim representações simbólicas intermediárias mínimas (por exemplo, o estado atual do tabuleiro, uma peça disponível para a jogada, etc), determinadas em função da intenção.

Foram definidos quatro níveis de abstração baseados no aspecto comportamental do sistema. O primeiro nível é o da intenção, definindo os propósitos do sistema de maneira global; o segundo é o dos comportamentos existentes em cada intenção; o terceiro é o das tarefas nas quais se dividem os comportamentos; e o quarto, em um nível de abstração mais baixo, é o das ações que compõem cada tarefa. A definição destes níveis comportamentais para o propósito escolhido é dada na tabela 1.

Nível	Descrição	Aplicação
<i>Intenções</i>	O propósito do sistema.	Jogar “Velha” contra um oponente.
<i>Comportamentos</i>	Comportamentos necessários para realizar o objetivo.	Efetuar uma jogada. Esperar jogada do oponente.
<i>Tarefas</i>	Tarefas em que se dividem os comportamentos.	Reconhecer as peças disponíveis. Reconhecer o tabuleiro. Determinar o estado atual do tabuleiro. Verificar se estado atual do tabuleiro é válido. Decidir jogada. Mover a peça. Perceber término da jogada do oponente.
<i>Ações</i>	Módulos funcionais básicos.	Segmentar imagem. Determinar tamanho de objetos. Determinar forma de objetos. Localizar objetos. Mover manipulador. Fechar/abrir garra do manipulador. Decidir nova jogada.

Tabela 1 - Níveis de abstração comportamental do sistema.

3. IMPLEMENTAÇÃO.

O sistema possui um controle geral e planos pré-definidos para cada comportamento. A figura 1 esquematiza o controle geral do sistema, onde a “jogada do robô” é o resultado do comportamento “efetuar uma jogada” e a “jogada do oponente”, ao comportamento “esperar jogada do oponente”.

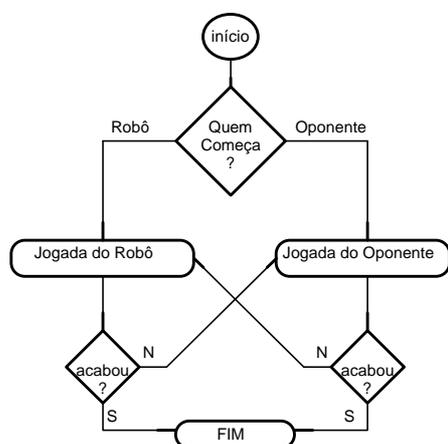


Figura 1 - Diagrama de blocos do controle geral do sistema.

O comportamento “efetuar uma jogada” é conseguido a partir da execução do plano de atividades mostrado na figura 2, enquanto o comportamento “esperar jogada do oponente” é obtido a partir da análise de seqüência de imagens, determinando quando a jogada foi terminada.

As tarefas definidas na Tabela 1 e na figura 2 são decompostas em ações, que constituem as primitivas comportamentais do sistema. Assim, o reconhecimento do tabuleiro envolve segmentação da imagem, determinação de forma (quadrada) e tamanho do objeto (tabuleiro), além de sua localização no campo visual. Com os limites estabelecidos pelo reconhecimento do tabuleiro, pode-se determinar as peças sobre ele e pelo menos uma peça fora dele, para realizar a jogada. As peças são localizadas e identificadas, envolvendo também segmentação de imagem, determinação de forma (redonda) e tamanho e localização.

Uma vez definida a posição de uma peça fora do tabuleiro, pode-se comandar o manipulador para mover e agarrar a peça que será utilizada na próxima jogada.

Paralelamente, com a informação da localização das peças sobre o tabuleiro, pode-se determinar seu estado atual e conferir a validade com base no estado anterior do mesmo.

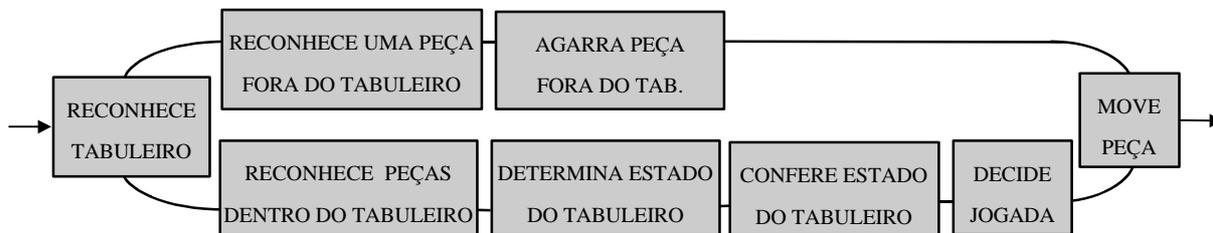
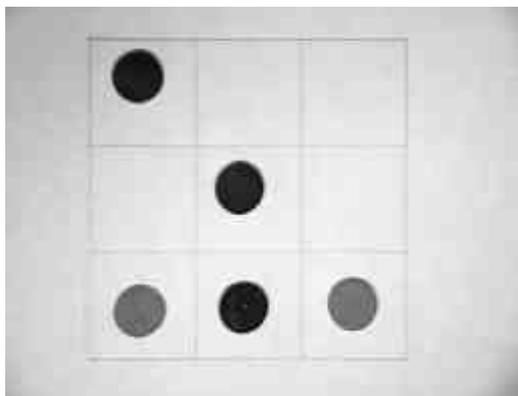


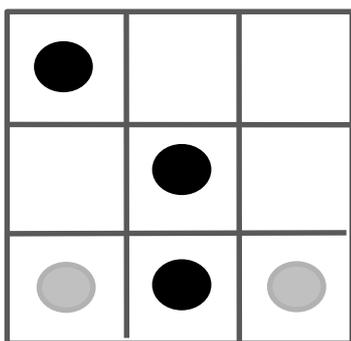
Figura 2 - Tarefas para comportamento “efetuar uma jogada”.

De posse do estado atual do tabuleiro (validado), aplica-se um procedimento decisório, determinando a próxima jogada. Decidido o movimento a ser realizado, e de posse de uma peça disponível, o manipulador realiza a jogada.

A segmentação da imagem foi feita por processos de limiarização de acordo com os níveis de cinza da mesma. A figura 3 mostra uma imagem adquirida pela câmera e o resultado da limiarização a ela aplicada.



(a) Imagem em níveis de cinza.



(b) Imagem segmentada

Figura 3 - Imagens do tabuleiro, antes e depois da segmentação.

Na imagem segmentada, para cada patamar de limiarização, identifica-se as peças por suas formas circulares e área conhecida e, pelo centro de área de cada círculo, determina-se a posição de cada peça. Do mesmo modo o tabuleiro é identificado por sua composição de segmentos de reta e localizado (posição e orientação) [8].

Usando este mesmo procedimento, considerando que o oponente é um ser humano e que, ao retirar sua mão do campo visual da câmera terminou sua jogada, o sistema identifica o término da jogada do oponente pelo reconhecimento de uma forma semelhante a uma mão e seu respectivo desaparecimento.

Para a realização do procedimento decisório foi utilizado o algoritmo Minimax de busca em uma árvore de estados possíveis [7]. O sistema permite três diferentes níveis de expansão da árvore de busca, refletindo a profundidade das previsões das próximas jogadas: jogo fácil, um nível de profundidade; jogo médio, dois níveis de profundidade; e jogo difícil, com três níveis de profundidade.

A heurística usada para determinação da função de avaliação estática (fae) analisa o tabuleiro pelas possibilidades de vitória nas linhas, colunas e diagonais, tanto para o sistema quanto para o oponente. A figura 4 mostra dois exemplos de cálculo da fae para um nível de expansão da árvore de busca.

A movimentação do manipulador é realizada enviando instruções pelo microcomputador através da sua porta serial para sua unidade de controle. A figura 5 mostra o sistema em operação, com o manipulador realizando uma jogada.

O manipulador usado é articulado verticalmente possuindo 5 graus de liberdade, com um motor de passo em cada articulação, e uma garra motorizada. A imagem é adquirida através de uma câmera e de uma placa de digitalização de imagens, que capta as imagens de 640 x 480 pixels em 256 níveis de cinza.

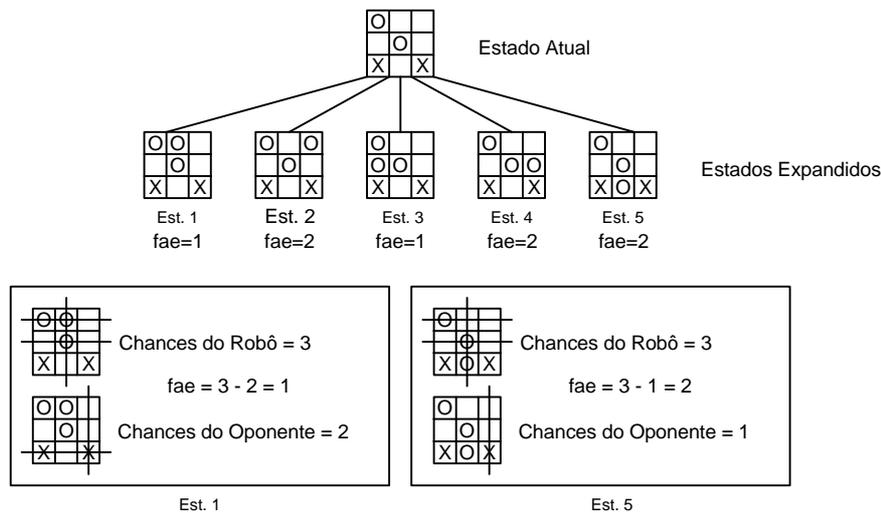


Figura 4 - Heurística utilizada para o cálculo da função de avaliação estática.

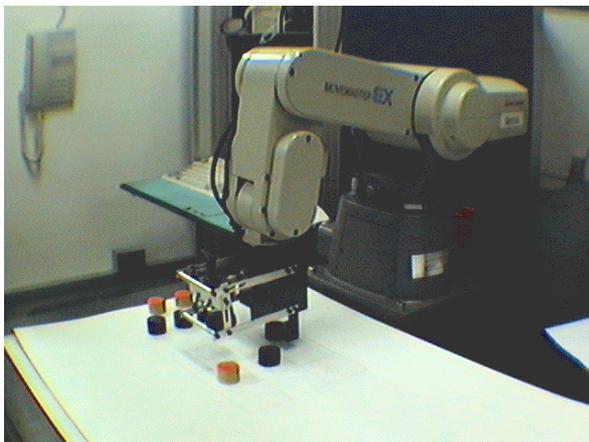


Figura 5 - Manipulador realizando uma jogada.

4. CONCLUSÃO E PROPOSTAS FUTURAS.

O paradigma de reconstrução propositiva proposto demonstrou poder ser utilizado em aplicações imersas em domínios complexos, que exijam cooperação entre percepção e ação e coordenação entre comportamentos distintos, como, por exemplo, manipulação e montagem em uma célula flexível e tarefas típicas de um robô móvel autônomo.

Um primeiro passo na realização deste objetivo foi dado neste trabalho, através da construção de um sistema que joga, de modo totalmente autônomo, o Jogo da Velha. O sistema, dotado de percepção visual e manipulador, apresentou resultados bastante promissores, motivando um grande interesse em

direcionar esforços na implementação de outros sistemas, dentro do paradigma proposto, em tarefas e ambientes mais complexos.

A aplicação usada como exemplo foi simples, permitindo que as diversas ações fossem facilmente implementadas, possibilitando melhor estudo da decomposição e gerenciamento de comportamentos, objetivo primeiro neste teste de abordagem de reconstrução propositiva. Deve-se ressaltar, ainda, que as ações utilizadas nesta aplicação são ações básicas e típicas de uma montagem.

Enfim, o trabalho realizado veio corroborar com a idéia de que visão não é um problema contido em si próprio, mas que ele está interligado com planejamento, raciocínio e aprendizado. O aprendizado é necessário para compensar a imprevisibilidade do mundo. Este aprendizado deve se dar de modo autônomo, no sentido de que o sistema de visão governa a si próprio e organiza sua própria estrutura para ter comportamento apropriado para realizar seus propósitos. O aprendizado autônomo não é supervisionado e é baseado na realimentação (normalmente, com atraso) da sua atuação e desempenho [13].

Assim, para realizar tarefas complexas, o sistema deve permitir a introdução de modificações em suas representações internas, através do aprendizado, visando maior precisão em seus modelos. Além disso, é interessante utilizar um planejador inteligente de atividades para criar, de modo autônomo, os planos das tarefas a serem realizadas.

Deve-se salientar que, em ambientes e intenções complexas, uma dificuldade a ser

considerada é a definição das ações primitivas e a composição destas em tarefas e comportamentos. Torna-se também de grande interesse e importância a busca por métodos e linguagens mais apropriados para o mapeamento de intenções em comportamentos.

5. REFERÊNCIAS BIBLIOGRÁFICAS.

- [1] ALOIMONOS, Y. What I have learned. **CVGIP: Image Understanding**, v.60, n.1, p.74-85, July 1994.
- [2] BLACK, M. J.; ALOIMONOS, J.; HORSWILL, I.; SANDINI, G.; BROWN, C.M.; MALIK, J.; TARR, M.J. Action, representation, and purpose: re-evaluating the foundations of computational vision. In: IJCAI, p. 1661-1666, 1993.
- [3] BROOKS, R. A. Intelligence without representation. **Artificial Intelligence**, v.47, p. 139-59, 1991.
- [4] CHRISTENSEN, H. I.; MADSEN, C. B. Purposive Reconstruction: A Reply to "A Computational and Evolutionary Perspective on the Role of Representation in Vision" by M. J. Tarr and M. J. Black. **CVGIP: Image Understanding**, v.60, n.1, p.103-8, July 1994.
- [5] JOLION, J.M. Computer vision methodologies. **CVGIP: Image Understanding**, v.59, n.1, p.53-71, Jan. 1994.
- [6] MARR, D. **Vision**. New York, Freeman, 1982.
- [7] RICH, E.; KNIGHT, K **Inteligência Artificial**. 2. ed, São Paulo, Makron Books, 1994.
- [8] RILLO, A.H.R.C. Um Sistema de Visão Binária para Reconhecimento de Peças Isoladas e Parcialmente Oclusas. In: Simpósio Brasileiro de Computação Gráfica e Processamento de Images (III SIBGRAPI), Gramado, RS, 1990, pp. 236-245.
- [9] RIVLIN, E.; ALOIMONOS, Y.; ROSENFELD, A. **Purposive Recognition: a Framework**. Technical Report CAR-TR-597, CS-TR-2811, DACA76-89-C-0019, IRI-9057934, University of Maryland, Dec. 1991.
- [10] SWAIN, M.J.; STRICKER, M. **Promising Directions in Active Vision**. Technical Report CS 91-27, University of Chicago, Nov. 1991.
- [11] TARR, M. J.; BLACK, M. J. A computational and evolutionary perspective on the role of representation in vision. **CVGIP: Image Understanding**, v.60, n.1, p.65-73, July 1994.

- [12] TARR, M. J.; BLACK, M. J. Reconstruction and purpose. **CVGIP: Image Understanding**, v.60, n.1, p.113-8, July 1994b.
- [13] VAN DE VELDE, W. Toward learning robots. **Robotics and Autonomous Systems**, v.8, n.1, pp 1-6, Nov. 1991.

6. AGRADECIMENTOS

Os autores Fábio Ferraz e Briand Moreira Jr contam com o apoio da CAPES, e o autor Reinaldo A. C. Bianchi com o apoio do CNPq.