CENTRO UNIVERSITÁRIO DA FEI

RODOLFO COURA DE BRITO

IMPLEMENTAÇÃO DE UM SISTEMA DE INTERPRETAÇÃO DE SEQÜÊNCIAS DE IMAGENS BASEADO EM UMA SEMÂNTICA DE CAMINHOS

São Bernardo do Campo

2008

RODOLFO COURA DE BRITO

IMPLEMENTAÇÃO DE UM SISTEMA DE INTERPRETAÇÃO DE SEQÜÊNCIAS DE IMAGENS BASEADO EM UMA SEMÂNTICA DE CAMINHOS

Dissertação de Mestrado apresentada ao Centro Universitário da FEI como parte dos requisitos necessários para a obtenção do título de Mestre em Engenharia Elétrica.
Orientador: Prof. Dr. Paulo Eduardo Santos.

São Bernardo do Campo

2008

Brito, Rodolfo Coura de.
Implementação de um sistema de interpretação de seqüências de imagens baseado em uma semântica de caminhos / Rodolfo Coura de Brito. São Bernardo do Campo, 2008.
63 f. : il.

Dissertação de Mestrado – Centro Universitário da FEI.

Orientador: Prof. Dr. Paulo Eduardo Santos

1. Interpretação de seqüências de imagens. 2. Raciocínio espacial qualitativo. 3. Lógica de Transações. 4. T-Logic. I. Título.

CDU 681.3

Dedico este trabalho ao meu pai Gilberto Coura de Brito, a minha mãe Maria das Graças de Brito e a minha namorada Cristine Rodrigues pelo incentivo e apoio em todo este trabalho e em toda minha vida.

AGRADECIMENTOS

Agradeço primeiramente ao meu pai Gilberto Coura de Brito, a minha mãe Maria das Graças de Brito, pelo apoio, incentivo, atenção, exemplo e amor que me deram em toda minha vida.

Ao meu irmão Gilberto Coura de Brito Júnior, a minha irmã Pérola Maria Coura de Brito e a minha cunhada Ana Paula da Silva Coura pelo apoio, e por tentarem impedir meu sobrinho de me atrapalhar com os estudos; e ao meu sobrinho Neto, de 1 ano e 4 meses, que mesmo com todos brigando, ele insistia em vir onde eu estava com um belo e puro sorriso me dando mais ânimo para tudo.

Um agradecimento especial ao meu orientador Professor Dr. Paulo Eduardo Santos, que foi um grande amigo quando professor e um grande professor quando amigo, pelos ensinamentos, incentivo e dedicação.

Um outro agradecimento especial a minha namorada Cristine Rodrigues, pois, fora o meu orientador Paulo, foi a pessoa que mais me ouviu, incentivou, ajudou e sempre esteve ao meu lado, em todo o período deste curso.

Aos professores Dr. Reinaldo Bianchi e Dr. Carlos Eduardo Thomaz, pelas belas aulas ministradas no curso, com muito empenho e dedicação.

Ao professor Dr. Flávio Tonidandel, que foi meu professor na graduação, e desde então, se mostrou um grande amigo; e no curso de mestrado confirmou isto. Também por ter aceitado ser membro da banca de qualificação e defesa de minha dissertação.

Ao professor Dr. Fleury por ter sido membro de minha banca de qualificação e à professora Renata Wassermann por ter aceitado fazer parte da banca de defesa de minha dissertação e enriquecer este trabalho com suas sugestões, assim como o Flávio.

A tia, professora Dra. Angiliete Moraes, pelo apoio e dedicação em ajudar.

Aos amigos do mestrado: Nelson Aguiar, Julio Sgarbi, Edson Kitani, Sérgio Henry, Luiz Celiberto, Murilo Martins, Marcel Lira e Lendro Demari, pelo companheirismo durante todo o curso. À Adriana e Rejane da secretária do mestrado.

Aos grandes amigos: Roberto, Denis, Carlos, Edson e Uarllei, pelo imenso apoio, lembrando que amigos são irmãos que moram em casas separadas. A todos amigos da FEI, da Pedreira, do Gáia, da Fundação CESP e da Accenture.

A Deus por ter colocado todos aqui citados em minha vida.

A todos meus familiares e amigos, e a todos aqueles que me ajudaram diretamente e indiretamente com este trabalho, um sincero muito obrigado.

“Se queremos progredir, não devemos repetir a história, mas fazer uma história nova”

Mahatma Gandhi

“A verdadeira medida de um homem não é como ele se comporta em momentos de conforto e conveniência, mas como ele se mantém em tempos de controvérsia e desafio“

Martin Luther King Jr

RESUMO

Neste trabalho abordamos o problema de interpretação de seqüências de imagens e sua principal contribuição é apresentar e discutir a implementação de um sistema computacional capaz de interpretar seqüências de imagens baseado em raciocínio espacial qualitativo. Nossa principal motivação para este desenvolvimento é interpretar seqüências de imagens, executando inferências lógicas sobre mudanças ocorridas nas cenas. Este sistema tem a habilidade de extrair informações de objetos de uma seqüência de imagens, obtidas por um sensor, associando relações entre objetos na cena e mudanças ao longo da seqüência, fornecendo assim, uma interpretação de alto nível. O sistema é dividido em dois módulos, o Modulo 1: Extração de informações e o Módulo 2: Interpretação de Alto Nível. No Módulo 1 tem-se a entrada das cenas no sistema. O Módulo 2 é responsável por interpretar os movimentos dos objetos na seqüência de imagens. Para isto, utilizaremos o formalismo chamado *T-logic* (SANTOS, P. e SANTOS, M., 2005), que é uma instância da Lógica de Transações (BONNER e KIFER, 1993). Este formalismo utiliza dois oráculos baseados em raciocínio espacial qualitativo. Os testes efetuados no sistema mostram é possível interpretar seqüências de imagens utilizando raciocínio espacial qualitativo através do formalismo *T-Logic*.

Palavras-chave: Interpretação de seqüências de imagens. Raciocínio espacial qualitativo. Lógica de Transações. *T-Logic*

ABSTRACT

In this work we tackle the problem of image sequence interpretation, where our main contribution is to present and discuss an implementation of a system for interpreting image sequences based on qualitative spatial reasoning. Our motivation for this is to do the interpretation executing logical inference methods on changes that occurred in scenes. The system developed has the ability to extract information of objects picked out in images describing the changes that occurred in the environment, thus providing a high level interpretation about objects` movement. The system has two modules, the first module is Information Extraction and second is High-Level Interpretation. Module 1 has as input snapshots of the world. Module 2 is responsible for the interpretation of object movements in images sequences provided to the system. For this, we use the logical formalism called T-Logic (SANTOS, P. and SANTOS, M., 2005), that is an instance of Transaction Logic (BONNER and KIFER, 1993). This logical formalism uses two oracles based on qualitative spatial reasoning. We present some experiments that show that the system developed is capable to interpret image sequences using qualitative spatial reasoning encoded within T-Logic.

Keywords: Images sequences interpretation. Qualitative spatial reasoning. Transaction Logic. T-Logic.

SUMÁRIO

1 INTRODUÇÃO 14

2 REVISÃO BIBLIOGRÁFICA 19

2.1 SISTEMAS DE INTERPRETAÇÃO DE SEQÜÊNCIAS DE IMAGENS 19

2.2 RACIOCÍNIO ESPACIAL QUALITATIVO 21

2.3 LÓGICA DE TRANSAÇÕES 24

3 EXTRAÇÃO DE INFORMAÇÕES 29

4 SEQÜÊNCIA DE IMAGENS COMO CAMINHOS 34

4.1 ORÁCULO DE ESTADOS 34

4.2 ORÁCULO DE TRANSIÇÕES 35

4.3 PROVANDO FÓRMULAS USANDO O SISTEMA DE INFERÊNCIA 37

4.4 OBTENDO UMA EXPLANAÇÃO ATRAVÉS DAS MUDANÇAS NO CAMINHO 37

5 DEFINIÇÃO E IMPLEMENTAÇÃO DO SISTEMA 40

5.1 NOVAS REGRAS DO ORÁCULO DE ESTADOS 40

5.2 NOVAS REGRAS DO ORÁCULO DE TRANSIÇÕES 41

5.3 IMPLEMENTAÇÃO DO SISTEMA 43

5.3.1 IMPLEMENTAÇÃO DOS ORÁCULOS 43

5.3.2 IMPLEMENTAÇÃO DA REGRA DE INFERÊNCIA 44

5.3.3 O SISTEMA 45

5.4 TRADUÇÃO DE FÓRMULAS DA T-LOGIC EM MÁQUINA DE ESTADOS 47

5.5 FUNCIONAMENTO DO SISTEMA 47

6 ANÁLISE DOS RESULTADOS 52

7 CONCLUSÃO E TRABALHOS FUTUROS 58

REFERÊNCIAS 60

LISTA DE FIGURAS

Figura 1 - Seqüência de imagens. 14
Figura 2 - Esquema do sistema. 15
Figura 3 - Esquema do movimento de rotação dos objetos. 17
Figura 4 - Esquema do movimento de anti-rotação dos objetos. 17
Figura 5 – Diagrama conceitual de transição. 23
Figura 6 – 12 das 14 relações do cálculo lines-of-sight. 24
Figura 7 - Imagem (quadrado no centro da matriz) e seu respectivo mapa de regiões. 30
Figura 8 - Ilustração do mapa de regiões de cada cena da seqüência apresenta da na figura 1. 31
Figura 9 - Informações extraídas da cena 1 da figura 1 32
Figura 10 - Pseudo-código da Etapa 1 do Sistema: algoritmo implementado para a segmentação da imagem por rotulação de cores. 33
Figura 11 - Pseudo-código da Etapa 2 do Sistema: algoritmo implementado para a extração das informação da área das regiões dos objetos 33
Figura 12 - Relações espaciais. 35
Figura 13 - Relações espaciais em ambiente real. 35
Figura 14 - Transição oclusao(i(a),i(b)). 36
Figura 15 – Máquina de estados regras (5) e (6). 45
Figura 16 – Transição b1 atualizou o estado corrente de 1 para 2. 48
Figura 17 – Transição b2 atualizou o estado corrente de 2 para 3. 49
Figura 18 – Transição b3 atualizou o estado corrente de 3 para 6. 50
Figura 19 – Transição b8 atualizou o estado corrente de 6 para 8. 51
Figura 20 - Dois objetos de mesmo tamanho, cores distintas e os objetos distantes entre si na cena 1. 53
Figura 21 - Dois objetos de mesmo tamanho,cores distintas e próximos nas cenas 53
Figura 22 - Dois objetos de tamanhos diferentes, cores distintas e distantes entre si na cena inicial. 54
Figura 23 - Dois objetos de mesmo tamanho, mesma cor e os objetos distantes um do outro na cena inicial. 55
Figura 24 - Objetos de mesmo tamanho, mesma cor e os objetos próximos entre si nas cenas. 56

Figura 25 - Objetos de tamanhos diferentes, mesma cor e os objetos distantes um do outro na cena inicial. 56

LISTA DE TABELAS

Tabela 1 – Síntese dos resultados..... 52

LISTA DE SÍMBOLOS

$\equiv_{def}$	Definição clássica.
$\neg$	Negação lógica.
$\wedge$	Conjunção clássica.
$\vee$	Disjunção clássica.
$\exists$	Existe.
$\otimes$	Conjunção serial.
$\oplus$	Disjunção paralela.
$\leftarrow$	Implicação.
$\models$	Conseqüência serial.
$\models_c$	Conseqüência lógica clássica.
$\circ$	Unificação de caminhos.
$\vdash$	Dedução.
π	Caminho, representam histórias das mudanças de estados dos objetos nas cenas.
D	Estado.
ϕ	Transição de estados.
δ	Valor de distância pré-definido.
O^d	Oráculo de estados.
O^t	Oráculo de transições.
Δ	Fórmula que representa transição verdadeira no caminho (fórmula objetivo).
k	Tamanho do caminho.
P	Par de oráculos.
$\in$	Pertence.
$\geq$	Maior ou igual.

LISTA DE ABREVIATURAS

FEI – Fundação Educacional Inaciana.
CAA – *Causal arrhythmia analysis.*
TR – Lógica de transações (do inglês t*ransaction logic*).
DC – Desconectado (do inglês d*isconnected*).
P – Parte (do inglês *part*).
O – Sober (do inglês *overlap*).
EC – Externamente conectado (do inglês *externally connected).*
CO - Amalgamado (do inglês *colescing*)
PO – Parcialmente sobre (do inglês *proper overlap*).
PP – Parte de (do inglês *proper part*).
TPP – Parte própria tangencial (do inglês *tangencial proper part*).
NTPP – Não tangencial (do inglês *nontangencial proper part*).
C – Desconectado (do inglês c*lear*).
JC – Externamente conectado (do inglês *just clear*).
PH – Parcialmente sobre (do inglês *partially hides*).
PHI – Parcialmente atrás (do inglês *partially hidden*).
JH – Sobre externamente conectado (do inglês *just hides*).
JHI – Atrás externamente conectado (do inglês *just hidden*).
H – Sobre (do inglês *hides*).
HI – Atrás (do inglês *hidden*).
EH – Exatamente igual sobre (do inglês *exactly hides*).
EHI – Exatamente igual atrás (do inglês *exactly hidden*).
F – A frente (do inglês *front*).
FI – Atrás (do inglês *front of it*).
JF – A frente externamente conectado (do inglês *just in front*).
JFI – Atrás (do inglês *just in front of it*).
RGB – Padrão de cores vermelho, verde e azul (do inglês *Red, Green and Blue*).

INTRODUÇÃO

Nós humanos somos capazes de descrever características do que vemos, por exemplo: cor, forma e tamanho. Podemos também, interpretar movimentos e ações. O problema que abordaremos neste trabalho é interpretar seqüências de imagens, onde interpretar é designar um conceito sobre o que é visto, utilizando inferências lógicas para explanar mudanças nas cenas. Mudanças estas que são representações de diferentes regiões nas cenas da seqüência de imagens, sendo que, neste trabalho, seqüência de imagens são fotografias em ordem cronológica de uma mesma cena.

O objetivo principal deste trabalho é apresentar e discutir a implementação de um sistema de computador que tenha a habilidade de extrair informações, baixa abstração, de objetos de uma seqüência de imagens obtidas pelo sensor, associando relações entre objetos na cena e mudanças ao longo da seqüência, fornecendo, assim, uma interpretação de alto nível, alta abstração, sobre o movimento dos objetos.

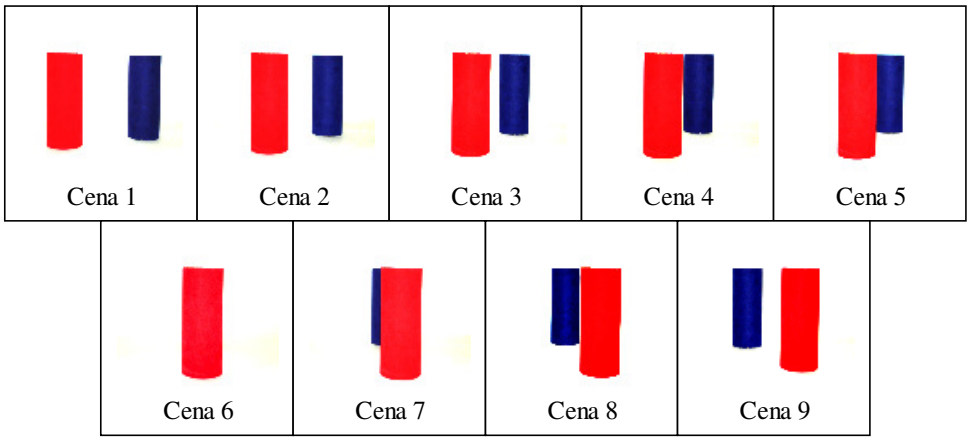


Figura 1 - Seqüência de imagens.

Na figura 1 temos um exemplo de uma seqüência de imagens composta por nove cenas, fotografias, obtidas pelo sensor. A motivação deste trabalho é interpretar seqüências como a figura 1, executando inferências lógicas sobre as mudanças ocorridas nas cenas da seguinte forma: primeiro, atribuímos os símbolos p e q para as regiões cinza claro e cinza escuro encontradas na Cena 1, na transição entre a Cena 1 e a Cena 2 observamos que a região p aumentou e região q diminuiu suas áreas, além de que ambas regiões se aproximaram. Similarmente, o mesmo ocorre nas transições da Cena 2 para Cena 3 e Cena 3 para Cena 4. Na transição da Cena 4 para Cena 5 as regiões p e q se colidem, na Cena 5 para a Cena 6 as suas regiões se fundem em uma e na Cena 6 para a Cena 7 as regiões se afastam.

Interpretando todas estas transições em seqüência, podemos deduzir que os objetos, no decorrer das cenas, estão em movimento de rotação entre si. Obter uma descrição conceitual do movimento dos objetos na seqüência a partir de inferências lógicas motiva o desenvolvimento deste sistema.

O sistema de interpretação de seqüências de imagens desenvolvido neste trabalho é dividido em dois módulos, o Modulo 1: Extração de Informações e o Módulo 2: Interpretação de Alto Nível. A figura 2 mostra os componentes do sistema.

No Módulo 1 temos a entrada das cenas no sistema, estas cenas são as já mencionadas fotografias extraídas pelo sensor. O sistema segmenta cada cena por fusão de suas regiões, a implementação do respectivo algoritmo é feita utilizando rotulação de regiões do tipo rotulação por cores (*Blob Coloring*) (BALLARD e BROWN, 1982). Este algoritmo cria um mapa das regiões encontradas na imagem, contendo as informações de todas as regiões e posições dos objetos de cada cena. Estas informações são os dados de entrada do Módulo 2.

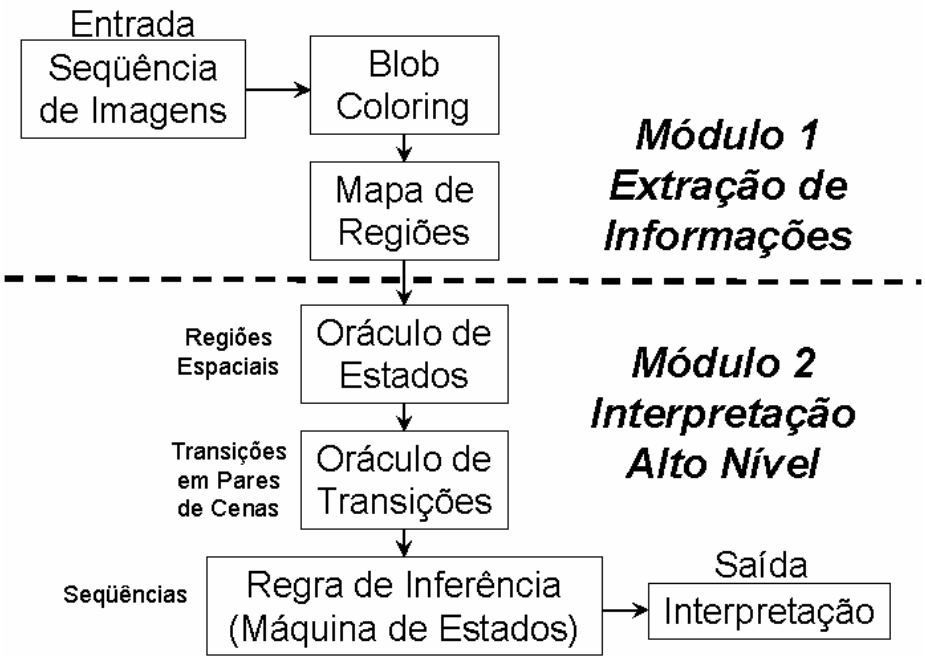


Figura 2 - Esquema do sistema.

A partir das informações extraídas no Módulo 1, o Módulo 2 é responsável por interpretar os movimentos dos objetos na seqüência de imagens apresentada ao sistema. Para isto, utilizaremos um formalismo chamado *T-logic* (SANTOS, P. e SANTOS, M., 2005), que é uma instância da Lógica de Transações (BONNER e KIFER, 1993). Este formalismo de interpretação de seqüências de imagens utiliza dois oráculos baseados em raciocínio espacial qualitativo: o Oráculo de Estados é responsável por relacionar as regiões espaciais em

conceitos de espaços livres e conectividade dos objetos de cada cena; e o Oráculo de Transições fornece informações de transições dos objetos entre cenas da seqüência. Estas informações são dados de entrada para uma Máquina de Estados que implementa as regras de inferências para a interpretação dos movimentos entre objetos ao longo de uma seqüência. A Máquina de Estados fornece a saída do sistema, que é uma descrição conceitual do movimento dos objetos.

Este trabalho é a primeira aproximação para uma interpretação completa da *T-Logic* em um ambiente real. Para abordar o problema, assumimos as seguintes simplificações:

1 - Os objetos são cilíndricos para reduzir a complexidade das cenas e facilitar sua identificação, em (HUANG, 1990) mostra ser possível generalizar corpos de pessoas em formas cilíndricas.

2-Somente dois únicos movimentos foram implementados no sistema para o desenvolvimento deste estudo; para relaxar esta solução para demais movimentos basta modelar o movimento de acordo com o formalismo que será apresentado.

3- O ambiente tem somente dois objetos, a luz é constante e o fundo branco sem texturas, para a captura das imagens, criamos um ambiente bem comportado, isto é, com a iluminação controlada e fundo monotônico. Mais especificamente, cada cena é composta por uma base e um fundo de cor branca, onde situam objetos de formato cilíndrico. É importante salientar que as formas dos objetos não são importantes, o ambiente pode ser composto por outros objetos de formas convexas quaisquer.

4 - O intervalo de tempo entre as cenas subseqüentes de uma seqüência de imagens é muito pequeno. Com isto garantimos a não perda de transições dos objetos.

Neste trabalho iremos discutir a implementação da interpretação de dois movimentos dos objetos que possuem mesmo tamanho, o movimento de rotação e o movimento ambíguo da rotação, que chamaremos de movimento de anti-rotação.

A rotação consiste em dois objetos cilíndricos que se movimentam em torno de um eixo fixo comum a eles, conforme figura 3. A anti-rotação consiste em um movimento similar ao movimento de rotação até o momento da oclusão de um objeto com o outro. A partir deste ponto, a desoclusão do objeto na anti-rotação ocorre do mesmo lado de onde o objeto foi ocluso, enquanto no movimento de rotação a desoclusão do objeto ocorre do lado oposto de onde o objeto foi ocluso, conforme se vê na figura 4. Para demonstrar o comportamento do sistema diante de movimentos ambíguos o movimento de anti-rotação foi inserido junto a eventos de rotação.

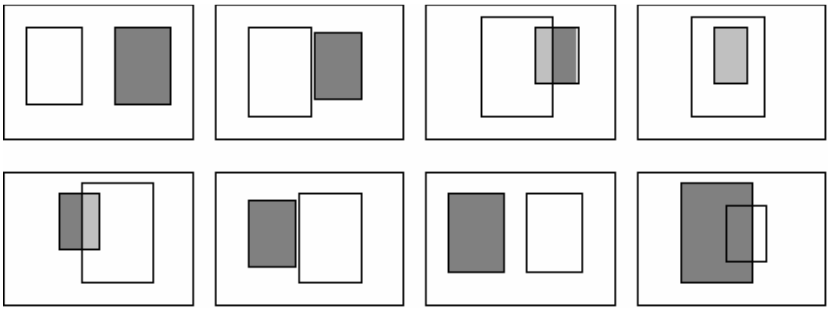


Figura 3 - Esquema do movimento de rotação dos objetos.

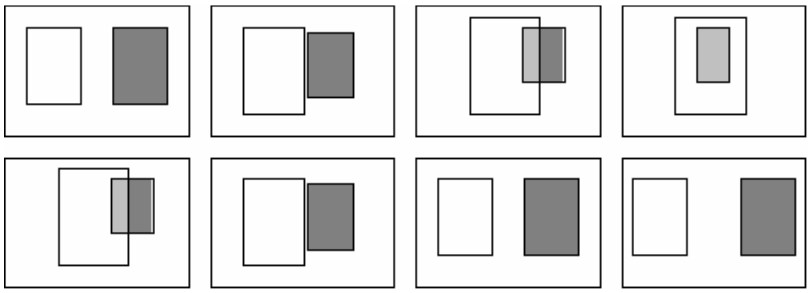


Figura 4 - Esquema do movimento de anti-rotação dos objetos.

As seqüências de imagens a serem interpretadas foram obtidas por uma câmera fotográfica digital Sony, 3 mega-pixels, sendo a câmera estática, portando, o que se move, entre as cenas, são os objetos.

Para o desenvolvimento e validação das informações de interpretação dos movimentos geradas pelo sistema, adotamos como experimentos quatro situações entre objetos de mesmo tamanho:

Situação um - dois objetos de cores distintas, e distantes entre si na cena inicial.

Situação dois - dois objetos de mesma cor, e distantes entre si na cena inicial.

Situação três - dois objetos de cores distintas, e próximos entre si na cena inicial.

Situação quatro - objetos de mesma cor, e próximos entre si na cena inicial.

Adotamos outras duas situações para objetos com alturas diferentes e mesma largura, sendo objetos de mesma forma. As duas seguintes situações são para validar o funcionamento do sistema para objetos que possuem alturas diferentes.

Situação cinco - dois objetos de alturas diferentes, de cores distintas, distantes entrei si na cena inicial.

Situação seis - dois objetos de alturas diferentes, de mesma cor, e distantes entre si na cena inicial.

Para cada movimento e cada situação entre objetos utilizaremos nos experimentos cinco seqüências diferentes de cenas, sendo duas destas utilizadas para o desenvolvimento do sistema e as outras três seqüências para seus testes.

Os testes mostram que a partir das informações extraídas pelo algoritmo de rotulação por cores é possível interpretar seqüências de imagens utilizando raciocínio espacial qualitativo através do formalismo *T-Logic*, deduzindo informações sobre mudanças ocorridas no decorrer das cenas.

A contribuição deste trabalho é apresentar e discutir a implementação de um sistema de interpretação de seqüências de imagens baseado em raciocínio espacial qualitativo utilizando a *T-Logic* como formalismo lógico.

Esta dissertação é organizada da seguinte forma. No capítulo 2 é apresentada uma revisão bibliográfica, onde abordaremos assuntos como: sistemas existentes de interpretação de seqüências de imagens, raciocínio espacial qualitativo e Lógica de Transações. O capítulo 3 discute a implementação do Módulo 1 do sistema, Extração de Informações em Imagens. O capítulo 4 introduz a *T-Logic* como um formalismo de interpretação de seqüências de imagens. No capítulo 5 discutiremos a implementação do Módulo 2 do sistema, Interpretação de Alto Nível. No capítulo 6 os resultados dos testes do sistema são analisados, e o capítulo 7 conclui esta dissertação e propõe trabalhos futuros.

REVISÃO BIBLIOGRÁFICA

Neste capítulo faremos uma revisão bibliográfica onde abordaremos três assuntos: sistemas existentes de interpretação de seqüências de imagens, raciocínio espacial qualitativo e Lógica de Transações.

Sistemas de Interpretação de Seqüências de Imagens

O propósito de um sistema de interpretação de seqüências de imagens é derivar descrições conceituais de uma seqüência de cenas observadas, que estejam de acordo com suposições gerais sobre mudanças causadas por movimento (NAGEL, 1988) (NAGEL, 2000).

O propósito básico de um sistema de interpretação de seqüência de imagens pode ser reduzido a três requerimentos: qualquer sistema deve ter uma finalidade clara definida; deve ser dotado de uma representação interna exaustiva para todas as tarefas e condições ambientais; e, deve mostrar ser capaz de interpretação explícita dos limites de suas capacidades (NAGEL, 1988).

A partir da idéia de criar sistemas capazes de fornecer descrições de mudanças ocorridas no decorrer de uma seqüência de imagens, foram desenvolvidos sistemas como o Mório (NAGEL, 1988), o NAOS (NEUMANN e NOVAK, 1988 apud NAGEL, 1986), o Epex (NAGEL, 1988), o ALVEN (TSOTSOS e SHIBAHARA, 1988), o CAA (TSOTSOS e SHIBAHARA, 1988), o VITRA (HERZOG e WAZINSKI, 1994), o SPROCKET (BRAND, 1997).

O Morio (NAGEL, 1988) é um sistema de interpretação de movimento de objetos rígidos a partir de uma seqüência de imagens monoculares. O sistema processa a seqüência de imagens *off-line*, isto é, as imagens não são processadas no instante em que são obtidas, mas sim após terem sido capturadas as trajetórias completas dos movimentos observados.

Outro sistema que possui o mesmo objetivo do Morio é o sistema NAOS (NAGEL, 1988), que é capaz de gerar descrições das mudanças ocorridas nas cenas observadas. Estas descrições permitem que uma pessoa, ao lê-las, possa imaginar a atividade interpretada pelo sistema. O NAOS também possui uma análise de uma seqüência de imagens baseada em trajetórias completas como o Morio.

O sistema VITRA (HERZOG e WAZINSKI, 1994), ao contrário do Morio e do NAOS, analisa a seqüência de imagens de forma incremental, ou seja, analisa a seqüência de

imagens logo que captura a cena. Portanto as informações sobre cada cena presente são fornecidas imediatamente à sua entrada no sistema.

O VITRA (HERZOG e WAZINSKI, 1994) é um sistema desenvolvido para narrar uma partida de futebol que descreve, em linguagem natural, as mudanças ocorridas dos objetos nas cenas analisadas, seu processo resulta em uma representação geométrica da cena indicando a localização do objeto observado em instantes consecutivos. Este sistema gera uma descrição geométrica dos objetos que compõem a cena, em que o processamento do sistema extrai relações espaciais entre os objetos, conceitos de movimento, hipóteses sobre as intenções e possíveis planos dos agentes. Esta estrutura conceitual é uma forma de organização do conhecimento que relaciona os dados visuais em linguagem natural, utilizando para tal, proposições espaciais, verbos de movimento, advérbios temporais ou cláusulas de causa.

Outro sistema de interpretação de seqüências de imagens que propõe organizar o conhecimento é o ALVEN (TSOTSOS e SHIBAHARA, 1983), o qual descreve detalhadamente o desempenho do ventrículo esquerdo do coração humano analisando geometricamente seqüências de imagens de Raios-X. A característica mais importante deste sistema é sua base de conhecimento organizada em classes, que são responsáveis por prover uma definição generalizada de componentes, atributos e relações de todo seu conhecimento. Sua base de conhecimento é composta por uma série de exemplos de funcionamento de corações; tanto funcionamento de corações com ótimo desempenho a outros exemplos de corações com desempenho irregular.

O sistema CAA (*Causal Arrhythmia Analysis*) (TSOTSOS e SHIBAHARA, 1983) também organiza sua base de conhecimento em forma de classes. O objetivo deste sistema é analisar resultados de eletrocardiogramas de seres humanos e detectar e classificar anomalias de arritmia cardíaca. O CAA compara o eletrocardiograma analisado com sua base de conhecimento, identificando e classificando possíveis anomalias.

Particularmente (BRAND, BIRNBAUM e COOPER, 1989) diz não ser possível extrair a estrutura causal de uma seqüência de imagens sem primeiro recuperar a sua estrutura física. Este método proposto extrai conceitos sobre causalidade entre objetos, raciocinando sobre física qualitativa. Baseado nesse raciocínio surgiu o sistema SPROCKET.

O SPROCKET (BRAND, 1997) é um sistema de visão que executa uma análise causal de mecanismos físicos usando o conhecimento sobre a dinâmica de corpos rígidos. A cena modelo é representada como a conectividade dos elementos da cena e as junções entre eles. Junções representam relações de contato, fricção, penetração etc. O módulo de extração de

características da cena utiliza dois tipos de algoritmos: um baseado em cores e outro em texturas.

Fundamentado no princípio de transições entre seqüências de cenas, foi desenvolvido um modelo de sistema de extração de episódios chamado Epex (NAGEL, 1988). Ele atribui conceitos à atividades que ocorrem na cena, por exemplo: em uma seqüência de imagens onde dois objetos se afastam, podemos atribuir o conceito *afastar* para essa atividade. O Epex é uma tentativa de ordenar os conceitos, atividades nas cenas, atribuindo somente os conceitos mais complexos correspondentes das atividades para interpretar uma simples seqüência.

Um sistema com o objetivo de descrever atividades humanas foi desenvolvido por (KOJIMA, TAMURA e FUKUNAGA, 2001), e é baseado em conceitos hierárquicos de ações. Ele generaliza descrições conceituais a partir da posição do corpo, orientação da cabeça do ser humano e sua postura. Isto porque através da posição da cabeça, o sistema determina para onde a pessoa olha, o que, por sua vez, determinando a direção da ação. Saber a posição das mãos implica em relacionar os gestos e interações com objetos. As regiões da cabeça e postura são extraídas através de rotulação por cores; a orientação da cabeça é encontrada por correlação de regiões com modelos em múltiplos aspectos da base de conhecimento do sistema. Os objetos nas cenas são encontrados por comparação de vértices de histogramas de cores das regiões e modelos.

Raciocínio Espacial Qualitativo

Um outro assunto que abordaremos nesta revisão bibliográfica é raciocínio espacial qualitativo, pois o sistema desenvolvido neste trabalho raciocina sobre os objetos nas imagens de forma qualitativa. Raciocínio espacial qualitativo são relações lógicas de regiões do espaço.

Nosso trabalho é influenciado por um formalismo desenvolvido por (RANDELL, CUI e COHN, 1992), pois utilizaremos intervalos lógicos para relacionar os objetos encontrados nas cenas das seqüências de imagens. Este formalismo descreve as seguintes relações de espaço entre regiões sobre a relação dinâmica *C(x,y)* que se lê "*x* está conectado a *y*". Na figura 5 demonstramos o diagrama conceitual de transição das relações citadas acima.

$DC(x,y) \equiv_{def} \neg C(x,y)$

x está desconectado de *y*

(do inglês *disconnected*)

$P(x,y) \equiv_{def} \forall z[C(z,x) \rightarrow C(z,x)]$

x é parte de *y*

(do inglês *part*)

$O(x,y) \equiv_{def} \exists z[P(z,x) \wedge P(z,x)]$

x está sobre *y*

(do inglês *overlaps*)

$EC(x,y) \equiv_{def} C(x,y) \wedge O(x,y)$

x está externamente conecta a *y*

(do inglês *externally connected*)

$PO(x,y) \equiv_{def} O(x,y) \wedge \neg P(x,y) \wedge \neg P(y,x)$

x está parcialmente sobre *y*

(do inglês *partially overlaps*)

$PP(x,y) \equiv_{def} P(x,y) \wedge \neg P(y,x)$

x é parte de *y*

(do inglês *proper part*)

$TPP(x,y) \equiv_{def} PP(x,y) \wedge \exists z[EC(z,x) \wedge EC(z,y)]$

x é parte propia tangencial a *y*

(do inglês *tangential proper part*)

$NTPP(x,y) \equiv_{def} PP(x,y) \wedge \neg\exists z[EC(z,x) \wedge EC(z,y)]$

x é não tangencial *y*

(do inglês *nontangential proper part*)

$x=y \equiv_{def} P(x,y) \wedge P(y,x)$

x é identico a *y*

$PP^{-1}(x,y) \equiv_{def} PP(y,x)$

$TPP^{-1}(x,y) \equiv_{def} TPP(y,x)$

$NTPP^{-1}(x,y) \equiv_{def} NTPP(y,x)$

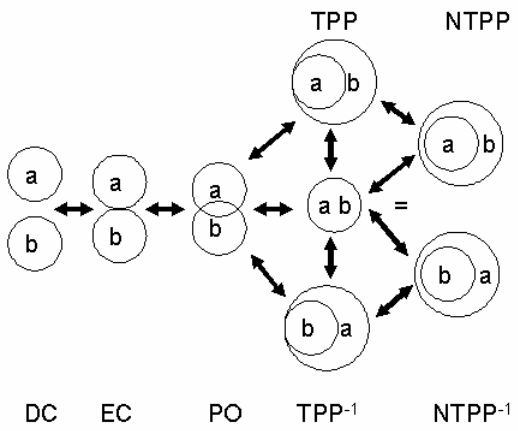


Figura 5 – Diagrama conceitual de transição.

Baseado nesta ontologia, Galton (1994) desenvolveu um cálculo de relações chamado *lines-of-sight* para raciocínio espacial qualitativo de corpos convexos. Este cálculo difere relações qualitativas de objetos de acordo com o campo visual de uma pessoa perante os objetos. Considerando dois objetos *a* e *b* se movendo livremente, o cálculo *line-of-sight* distingue as seguintes 14 relações qualitativas. As definições são análogas as do (RANDELL, CUI e COHN, 1992) e, portanto, foram omitidas.

C	A está desconectado de B	(do inglês *clear)*
JC	A está externamente conectado a B	(do inglês *just clear)*
PH	A está parcialmente sobre B	(do inglês *partially hides)*
PHI	B esta parcialmente sobre A	(do inglês *partially hidden)*
JH	A está sobre B externamente conectado	(do inglês *just hides)*
JHI	B está sobre A externamente conectado	(do inglês *just hidden)*
H	A está sobre B	(do inglês *hides)*
HI	B está sobre A	(do inglês *hidden)*
EH	A é exatamente igual a B	(do inglês *exactly hides)*
EHI	A é exatamente igual a B e esta sobre ele	(do inglês *exactly hidden)*
F	A está à frente de B	(do inglês *front)*
FI	B está à frente de A	(do inglês *front of it)*
JF	A está à frente de B externamente conectado	(do inglês *just in front)*
JFI	B está à frente de A externamente conectado	(do inglês *just in front of it)*

A figura 6 mostra 12 das 14 relações do cálculo *lines-of-sight*. As relações EH e EHI não se encontram na figura, pois os objetos *a* e *b* são exatamente idênticos.

Baseados nos trabalhos de (RANDELL, CUI e COHN, 1992) e (GALTON, 1994) foi desenvolvido um trabalho que descreve um cálculo sobre oclusão de regiões (RANDELL, WITKOWSIK e SHANAHAN, 1997), cujo propósito é modelar oclusão espacial e os efeitos de movimentos paralelos de objetos arbitrários.

Sustentados por estes formalismos utilizaremos raciocínio espacial qualitativo para interpretação de seqüências de imagens. Um estudo mais aprofundado no assunto de raciocínio espacial qualitativo pode ser encontrado em (SANTOS, 2007). A fim de desenvolver um método de raciocínio qualitativo utilizaremos uma lógica chamada Lógica de Transações, pois é uma lógica que raciocina sobre mudanças de estados, e estes, então, serão as cenas da nossa seqüência.

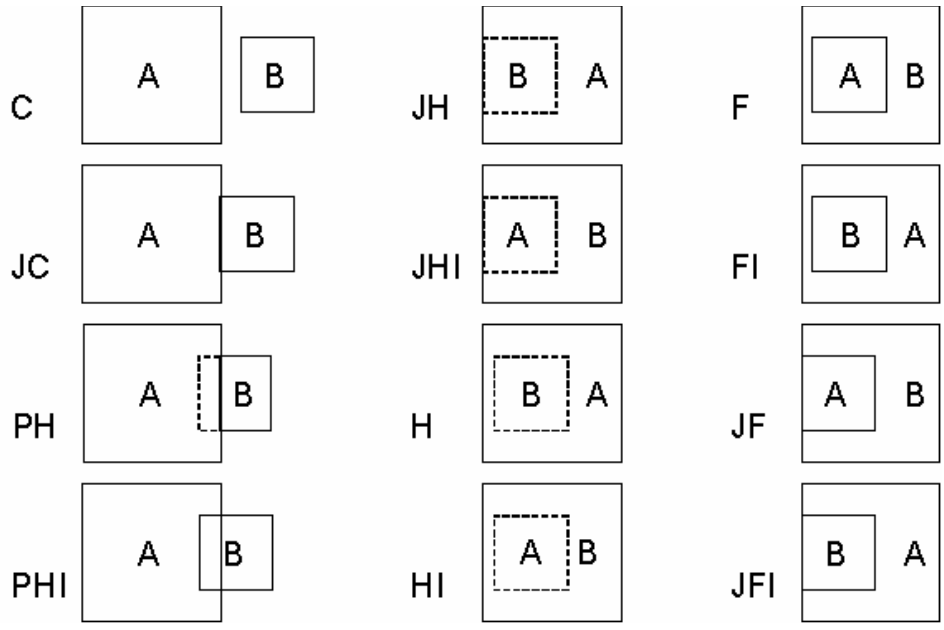


Figura 6 – 12 das 14 relações do cálculo *lines-of-sight.*

Lógica de Transações

A Lógica de Transações, TR do inglês *Transaction Logic,* (BONNER e KIFER, 1993) é uma extensão da lógica clássica de 1ª ordem, fornecendo uma fundamentação lógica para o fenômeno de mudança de estados em programas lógicos e bases de conhecimento. Em linhas gerais, a TR é uma lógica de mudanças de estados incluindo um modelo natural lógico, e ela tem um fragmento Horn, um corpo procedimental e semântica declarativa; seus programas

podem ser executados por um procedimento chamado *SDL-style*. O resultado é uma linguagem baseada em regras com uma semântica e teoria de provas correta e completa.

A TR, analogamente à lógica de primeira ordem, incluí símbolos de funções e símbolos de predicados. Além de conter conectivos da lógica clássica, a *TR* se apropria dos conectivos não clássicos estendendo a lógica clássica com um novo conectivo, $\otimes$, chamado conjunção serial (BONNER e KIFER, 1993). Com a conjunção serial, com a qual podemos formar regras como:

$$p \leftarrow a_1 \otimes a_2 \otimes a_3 \otimes ... \otimes a_n \qquad (1)$$

onde cada a_i é uma fórmula atômica, e $a_1 \otimes ... \otimes a_n$ é chamada conjunção serial. Podemos ler a fórmula (1) da seguinte maneira: para p acontecer é necessário executar primeiro a_1, depois a_2, depois a_3, e assim seqüencialmente até executar a_n. A fórmula (1) é chamada serial-Horn. Isto é, uma regra serial-Horn é uma fórmula de transação da forma $b \leftarrow \phi$, onde b é uma fórmula atômica e ϕ é uma conjunção serial.

Em sua semântica os conectivos clássicos são interpretados da seguinte maneira, $\alpha \wedge \beta$, onde α e β são verdadeiro no caminho. Os conectivos não clássicos são interpretadas por partes do caminho. Assim, $\alpha \otimes \beta$, é verdade se e somente se: α é verdadeiro em um prefixo do caminho e β é verdade no sufixo. A TR interpreta fórmulas como caminhos. A seguir um exemplo de interpretação de fórmulas da TR. (Sendo $\models$ conseqüência serial, $\models_c$ conseqüência lógica clássica e $\circ$ união de caminho). Exemplo:

$D \models \alpha \wedge \beta$, se e somente se existir D $\models_c \alpha$ e D $\models_c \beta$

$D \models \alpha \otimes \beta$, se e somente se existir D=π, tal que $\pi = \pi_1 \circ \pi_2$ e $\pi_1 \models_c \alpha$ e $\pi_2 \models_c \beta$

A TR utiliza oráculos que isolam operações lógicas da base de dados usadas para: combinar, programar e raciocinar.

Como veremos no capítulo 5, para a interpretação de regras é necessário inferir com base nas informações dos oráculos as mudanças ocorridas em todo o caminho dado. Para isto, a TR utiliza seu método de inferência chamado resolução *SDL-Style* (BONNER e KIFER, 1993).

O sistema de inferência da TR manipula expressões chamadas seqüentes, que têm a forma $P, \pi_1 \circ \pi_2 \vdash (\exists)\phi$, onde P consiste no par de oráculos; $\pi_1 = (D_1...D_2)$ e $\pi_2 = (D_3...D_n)$ são itens do caminho π, (representando a seqüência de estados) e ϕ é uma transição de estados, por exemplo: ϕ é verdade na transição do estado D_2 em π_1 a D_3 em π_2.

Para a interpretação destas regras, a resolução *SDL-Style*, a partir da conjunção serial, substitui e unifica os átomos das fórmulas estado a estado, até que o objetivo seja alcançado. Esse método de inferência retorna como correto a seqüência de estados π em que todos os átomos ϕ são satisfeitos de acordo com a fórmula.

A resolução *SDL-Style* possui três regras de inferências. Nas regras 1-3 a seguir, σ é uma substituição, *a* e *b* são fórmulas atômicas, e ϕ e *rest* são conjunções seriais. As regras são as seguintes:

1. *Aplicando as definições*: Supondo que $a \leftarrow \phi$ é uma regra em *P* e *b* e *a* unificam com o *m.g.u.* σ, então:

$$\frac{P, \pi_1 \circ \pi_2 \models (\exists)(\phi \otimes rest)\sigma}{P, \pi_1 \circ \pi_2 \models (\exists)(b\phi \otimes rest)}$$

2. *Questiona a base de dados, Oráculo de Estados*: Se *a* e *rest* σ não possuem variáveis em comum e $O^d(D_i) \models_c (\exists)b\sigma$, então:

$$\frac{P, \pi_1 \circ \pi_2 \models (\exists)\ rest\sigma}{P, \pi_1 \circ \pi_2 \models (\exists)(b \otimes rest)}$$

3. *Executando as transições elementares, Oráculo de Transições*: Se $b\sigma$ e *rest* σ não possuem variáveis em comum e $O^t(D_1,D_2) \models_c (\exists)b\sigma$, então:

$$\frac{P, \pi'_1 \circ \pi'_2 \models (\exists)\ rest\sigma}{P, \pi_1 \circ \pi_2 \models (\exists)(b \otimes rest)},$$

Onde $\pi_1 = (D_1...D_i)$, $\pi_2 = (D_{i+1}, D_{i+2}...D_n)$, $\pi'_1=(D_1...D_i, D_{i+1})$ *e* $\pi'_2 = (D_{i+2}...D_n)$

A regra 1 acima é a regra clássica de inferência, essa regra substitui uma instância da cabeça da regra por uma instância de um corpo de regra em *P*, note que no caminho não houve mudança. Esta regra diz que se *b* é definido em ϕ, e $\phi \otimes$ *rest* segue de *D*, no qual *D* é todo o caminho $\pi_1 \circ \pi_2$, então $b \otimes$ *rest* também segue de *D*.

A regra 2 é a inferência através do oráculo de estados, consulta sobre os estados do mundo. Segundo esta regra uma condição *b*, satisfeita em D_i, pode ser adicionada à fórmula

rest. A regra de inferência 2 trata dos testes que não causam mudança na base de dados. Ou seja, se *b* é verdadeiro em *D* e *rest* sucede de *D*, então $b \otimes rest$ também sucede de *D*. Informalmente, *b* é uma pré-condição que é satisfeita por *D*, portanto ela pode ser posta antes da transação *rest*.

A tarefa de verificação das transições de estados, através do oráculo de transições, é executada pela regra 3. Informalmente, a regra 3 indica que se *b* transforma a base de dados de D_1 em D_2, e se *rest* sucede em D_2, então $b \otimes rest$ pode ser inferido de $\pi_1 \circ \pi_2$.

Cada regra acima consiste em dois seqüentes passíveis da seguinte interpretação: se o seqüente superior, G_{i+1}, pode ser inferido, então a seqüente inferior, G_i, também pode. Assim, a TR a partir dos oráculos e com a inferência da resolução *SDL-Style*, formaliza o processo subjacente à um sistema capaz de executar transições de estados.

O sistema capaz de interpretar seqüências de imagens que desenvolvemos, cuja implementação discutiremos neste trabalho, utiliza o módulo de visão similar ao do SPROCKET, pois utilizaremos um algoritmo baseado em cores para a detecção das regiões que representam os objetos nas cenas; em nosso trabalho utilizaremos o algoritmo *Blob Coloring* (BALLARD e BROWN, 1982), que segmenta a imagem por similaridade, isto é, varre a imagem inteira comparando pixel à pixel selecionando regiões comuns, o que discutiremos a implementação deste algoritmo no capítulo 3 deste trabalho.

Iremos raciocinar sobre os objetos das seqüências de imagens, através raciocínio espacial qualitativo, conforme os formalismos desenvolvidos por (RANDELL, CUI e COHN, 1992), que descreve intervalos lógicos sobre relações de espaços entre regiões, e (GALTON, 1994) chamado *lines-of-sight*, que são cálculos de relações entre espaços de corpos convexos. Estes formalismos foram apresentados neste capítulo, uma referência completa sobre o assunto pode ser encontrada em (SANTOS, 2007).

A análise das seqüências de cenas em nosso trabalho será incremental como encontrado no sistema VITRA (HERZOG e WAZINSKI, 1994), ou seja, conforme há o progresso das cenas das seqüências de imagens, o sistema já analisa o ocorrido e fornece uma interpretação quando possível.

Nosso sistema utilizará a idéia do sistema Epex (NAGEL, 1988), que é atribuir um conceito às atividades que ocorrem na cena, porém neste trabalho utilizaremos o formalismo baseado na Lógica de Transações (BONNER e KIFER, 1993) chamado *T-Logic* (SANTOS, P. e SANTOS, M., 2005) que é uma teoria especializada em interpretação de seqüências de imagens. Esta teoria se baseia em um par de oráculos chamados oráculo de estados e oráculo de transições. O oráculo de estados informa quais relações espaciais são verdadeiras ou falsas

em cada cena. O oráculo de transições informa quais transições ocorrem em estados consecutivos em relação às regiões espaciais. A *T-Logic* é uma teoria para interpretação de seqüências de cenas que utiliza uma formalização de uma porção do conhecimento de senso comum.

A *T-Logic* foi baseada no trabalho (SANTOS e SHANAHAN, 2002), que descreve a construção de um sistema de raciocínio espacial qualitativo para informações extraídas por um robô móvel e será discutida no capítulo 4.

EXTRAÇÃO DE INFORMAÇÕES

Neste capítulo iremos discutir a implementação do Módulo 1 do sistema de interpretação de seqüências de imagens, figura 2, capítulo 1. Este módulo é responsável por extrair informações das cenas por segmentação da imagem, brevemente apresentado no capítulo 1. Estas informações serão utilizadas na interpretação do movimento de toda a seqüência de imagens pelo Módulo 2. A implementação do Módulo 2 será descrita e discutida no capítulo 5.

No Módulo 1 temos a entrada da seqüência de imagens do sistema, onde as imagens serão segmentadas pela aplicação do algoritmo *Blob Coloring* (BALLARD e BROWN, 1982). Este algoritmo segmenta a imagem por similaridade, isto é, ele varre a imagem inteira comparando pixel à pixel selecionando regiões comuns.

A detecção de objetos em imagens poderia ser feita por outros algoritmos, como o Sobel (FORSYTH e PONCE, 2003) - utiliza operadores de gradiente para a detecção de bordas, ou Canny (FORSYTH e PONCE, 2003) - baseado em operadores gaussianos, ambos algoritmos de detecção de bordas.

Devido à complexidade da implementação dos algoritmos de detecção de bordas Sobel e Canny, optamos pelo algoritmo *Blob Coloring*, que se mostrou satisfatório na detecção dos objetos e também nos cálculos de: área, região, perímetro, centróide, entre outros, conforme mostraremos nos resultados dos testes.

Estas informações são importantes para o sistema de interpretação de seqüências de imagens que propomos nesse trabalho. A seguir discutiremos a implementação do Módulo 1 do sistema, extração de informações em seqüências de imagens.

Todo sistema é implementado na linguagem *C++*.

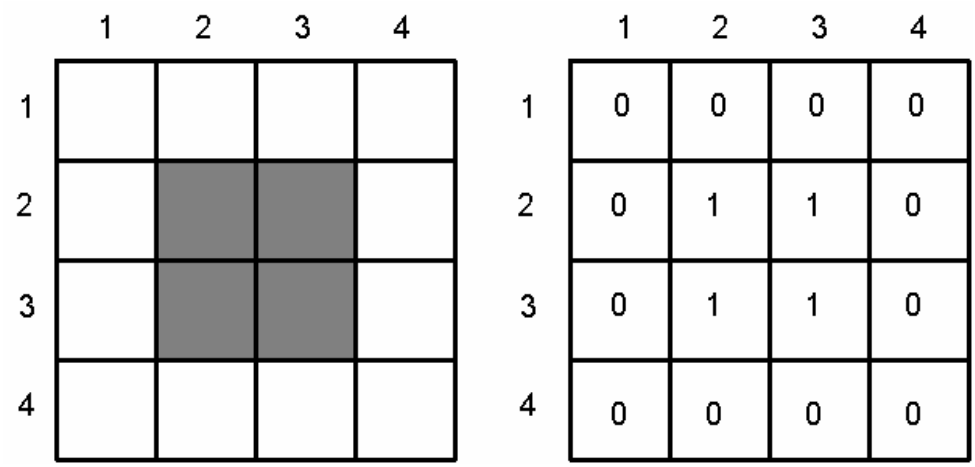
O processo de segmentação de imagens desenvolvido neste trabalho analisa cada cena da seqüência individualmente em duas etapas:

1. Aplicação do algoritmo de rotulação por cores;
2. Extração das informações das regiões encontradas: área, cor e posição.

A primeira etapa do sistema aplica o algoritmo de rotulação por cores, em que a imagem inteira é rastreada pixel à pixel, da esquerda para a direita, e de cima para baixo, comparando as cores RGB, vermelho, verde e azul, do pixel corrente com as cores RGB do pixel anterior, nesta ordem. Se os valores das cores do pixel corrente estiver em um intervalo comum com as cores do anterior, ambos mantém-se na mesma região. Se o pixel não estiver

nessa região, o algoritmo o compara com o pixel acima. Se essa comparação estiver em um intervalo comum, o sistema mantém o pixel corrente na região do pixel acima, caso contrário, cria-se uma nova região. Essa diferença entre os pixels consecutivos da imagem significa que foi detectado um novo objeto, por isso uma nova região é criada. Os mapas de regiões são obtidos com base nesta nova seqüência de imagens.

Mapa de regiões é uma matriz de mesma dimensão da imagem analisada, ou seja, uma matriz com a mesma altura e largura da imagem. Cada célula da matriz representa seu respectivo pixel na imagem. O conteúdo desta matriz é a região a qual o pixel pertence. Na figura 7 ilustramos um exemplo de uma imagem com as dimensões de 4 X 4 pixels e seu respectivo mapa de regiões.



(a) Imagem 4 x 4 pixels (b) Mapa de regiões da imagem

Figura 7 - Imagem (quadrado no centro da matriz) e seu respectivo mapa de regiões.

Ainda nesta etapa do algoritmo são tratadas as regiões de equivalência, que ocorre quando o sistema em um determinado ponto na varredura da imagem encontra uma nova região, sendo que, no decorrer de sua execução, detecta que essa nova região é igual a uma região existente. Nesta interação, o sistema troca o valor da nova região pelo valor da região antiga. Com o fim da varredura de toda a imagem termina a Etapa 1 do sistema.

Por exemplo, tendo como entrada do sistema a imagem (a) da figura 7 o sistema inicia a varredura da imagem da esquerda para direita, de cima para baixo. A célula (1,1), (altura,largura) da matriz do mapa de regiões recebe o valor zero, conforme figura 7 (b), porque a região do mapa de regiões inicia-se com zero. Após o sistema atribuir o valor da região do pixel no mapa de regiões, o sistema atualiza como pixel corrente o próximo pixel (1,2). Neste momento, o sistema compara o valor da cor do pixel corrente com a cor do pixel

à sua esquerda; como se trata da mesma cor, o sistema atribui o mesmo valor da região do pixel anterior para a região do pixel atual no mapa de regiões. Isto ocorre para os pixels: (1,3); (1,4) e (2,1). No caso do pixel (2,1) o sistema compara com o pixel acima dele, pois o pixel (2,1) não possui pixel à sua esquerda. Como anteriormente, após o sistema atribuir o valor da região do pixel no mapa de regiões, o sistema atualiza como pixel corrente o próximo pixel (2,2) e compara com o valor da cor do pixel à sua esquerda; o resultado desta comparação é negativo: a cor do pixel (2,2) é diferente da cor do pixel (2,1). Sendo assim, o sistema compara o valor da cor do pixel (2,2) com a cor do pixel (1,2); o pixel que está acima, que também é diferente, neste caso o sistema incrementa 1 a região do pixel anterior e atribui, neste valor, a região do pixel no mapa de regiões, atualizando para 1, porque foi encontrada uma nova região, conforme figura 7 (b). Assim, sucessivamente, o sistema analisa toda a imagem criando seu respectivo mapa de regiões.

A fim de ilustrar os mapas de regiões criados pela Etapa 1, criamos novas imagens no formato RGB a partir da figura 1 demonstradas pela figura 8. A partir das informações destes mapas de regiões, a Etapa 2 do sistema é responsável por extrair informações, da área, cor e posição dos objetos, que servirão de entrada para o módulo de interpretação de alto nível, conforme figura 1 apresentada no capítulo 1.

O cálculo das áreas dos objetos é feito pela somatória das quantidades de pixels de cada região já segmentada pela etapa anterior. Informações sobre a cor de cada uma destas regiões são obtida correlacionando a região ao respectivo pixel do objeto na imagem original.

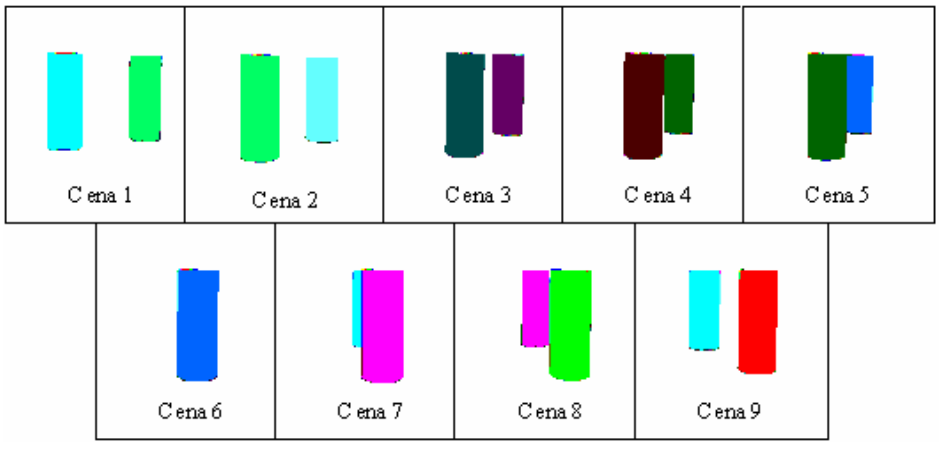


Figura 8 - Ilustração do mapa de regiões de cada cena da seqüência apresenta da na figura 1.

A posição de cada objeto na imagem é fornecida pelos quatro vértices de sua respectiva região no mapa de regiões. Para localizar a posição do vértice superior esquerdo da região

encontramos o pixel de menor linha e coluna desta região; para localizar o vértice superior direito encontramos o pixel de menor linha e maior coluna desta região; para localizar o vértice inferior esquerdo encontramos o pixel de maior linha e menor coluna desta região; e para localizar o vértice inferior direito encontramos o pixel de maior linha e maior coluna desta região; figura 11. Estas informações são gravadas em uma nova matriz, a qual daremos o nome de *MatrizInfo*. No capítulo 5, a partir das informações extraídas das cenas, *MatrizInfo*, iremos discutir a implementação do sistema de interpretação de seqüências de imagens.

A figura 9 mostra um exemplo das informações extraídas pelo sistema de interpretação de seqüências de imagens. Neste exemplo temos as informações extraídas da Cena 1, da seqüência de imagens da figura 1; as informações são as regiões encontradas na cena e seus respectivos valores de área, cor e posição. Podemos notar que foram encontradas onze regiões, sendo as regiões: 1, 2, 3, 4, 7, 8, 10 e 11 ruídos na imagem por possuir um valor pequeno de área. Estas regiões serão desprezadas no Módulo 2 do sistema, Interpretação de Alto Nível, que discutiremos no capítulo 5.

A seguir, nas figuras 10 e 11, apresentamos os pseudo-códigos do algoritmo implementado para a segmentação da imagem por rotulação de cores, Etapa 1; e extração de informação do mapa de regiões, Etapa 2, respectivamente.

REGIAO: 0	**AREA: 7663**	**VERM: 255**	**VERDE: 255**	**AZUL: 255**	**POSL: 0**	**C: 0**
REGIAO: 1	AREA: 1	VERM: 139	VERDE: 106	AZUL: 215	POS L: 76	C: 34
REGIAO: 2	AREA: 3	VERM: 177	VERDE: 172	AZUL: 255	POS L: 76	C: 31
REGIAO: 3	AREA: 9	VERM: 105	VERDE: 97	AZUL: 255	POS L: 76	C: 22
REGIAO: 4	AREA: 1	VERM: 56	VERDE: 60	AZUL: 255	POS L: 76	C: 21
REGIAO: 6	**AREA: 1275**	**VERM: 32**	**VERDE: 23**	**AZUL: 255**	**POS L: 18**	**C: 18**
REGIAO: 7	AREA: 2	VERM: 168	VERDE: 142	AZUL: 106	POS L: 73	C: 86
REGIAO: 8	**AREA: 1018**	**VERM: 82**	**VERDE: 30**	**AZUL: 26**	**POS L: 23**	**C: 71**
REGIAO: 9	AREA: 2	VERM: 198	VERDE: 168	AZUL: 142	POS L: 71	C: 86
REGIAO: 10	AREA: 1	VERM: 255	VERDE: 215	AZUL: 209	POS L: 71	C: 66
REGIAO: 11	AREA: 1	VERM: 232	VERDE: 213	AZUL: 181	POS L: 70	C: 86

Figura 9 - Informações extraídas da cena 1 da figura 1

```
Região = 0
Rastrear a imagem inteira da esquerda para a direita de cima para
baixo
      Se o pixel corrente for diferente ao pixel à esquerda Então
            Se o pixel corrente for diferente ao pixel acima Então
                  Região do pixel corrente = Região + 1;
            Senão
                  Região = Região do pixel abaixo;
      Senão
            Região do pixel corrente = Região do pixel à direita;
```

Figura 10 - Pseudo-código da Etapa 1 do Sistema: algoritmo implementado para a segmentação da imagem por rotulação de cores.

Para todas as imagens que apresentamos ao sistema, o Módulo 1 foi capaz de extrair informações de área, cor e posição das regiões dos respectivos objetos pertencentes as imagens analisadas, sendo importante salientar que a imagens possuem luz constante e fundo branco sem texturas, para a captura das imagens criamos um ambiente bem comportado, isto é com a iluminação controlada e fundo monotônico. Estas informações, extraídas pelo Módulo 1 do sistema, são parâmetros de entrada do Módulo 2.

O Módulo 2, figura 2, utiliza um formalismo chamado *T-Logic* para de interpretação de seqüências de imagens, que será descrito no próximo capítulo.

```
Para Região = 0 até Região = Maior Região Encontrada
      Rastrear o mapa de regiões inteiro da direita para a esquerda
de baixo para cima
      Se a região corrente for igual a Região Então
            Área Região = Área Região + 1;
      Se final do mapa de região da Região Então
            Cor Região = Valor da cor do pixel corrente na imagem
original;
            Posição Região = Valor da posição do pixel corrente;
```

Figura 11 - Pseudo-código da Etapa 2 do Sistema: algoritmo implementado para a extração das informação da área das regiões dos objetos

SEQÜÊNCIA DE IMAGENS COMO CAMINHOS

Neste capítulo apresentamos o formalismo que utilizaremos para interpretar seqüências de imagens, a partir de conectividade e espaços livres entre os elementos das cenas, objetos. Este formalismo se chama *T-Logic* e é uma instância da Lógica de Transações – TR, discutida no capítulo 2.

A *T-Logic* (SANTOS, P. e SANTOS, M., 2005) é uma teoria especializada em interpretação de seqüências de imagens. Esta teoria se baseia em um par de oráculos chamados oráculo de estados e oráculo de transições. O oráculo de estados informa quais relações espaciais são verdadeiras ou falsas em cada cena. O oráculo de transições informa quais transições ocorrem em estados consecutivos em relação às regiões espaciais. A *T-Logic* é uma teoria para interpretação de seqüências de cenas que utiliza uma formalização de uma porção do conhecimento de senso comum.

Por se tratar de uma instância da TR, a *T-Logic* possui mesma sintaxe e semântica. Portanto, para a interpretação de seqüências de imagens assumiremos a interpretação de fórmulas como caminhos da TR. Caminhos, neste caso, representam a história das mudanças de estados dos objetos nas cenas.

Oráculo de Estados

A partir das informações recebidas do sensor, o oráculo de estados, O^d, tem como objetivo informar qual a relação espacial é verdadeira entre objetos em cada cena. Logo, ele relaciona as regiões espaciais a conceitos de espaços livres e conectividade dos objetos.

Em (SANTOS, P. e SANTOS, M., 2005) foram identificadas três relações entre objetos em uma imagem conforme demonstrado na figura 7, onde *dc(x,y)*, significa "*x* está desconectado de *y*", do inglês *disconnected*; *ec(x,y)*, lê-se como: "*x* externamente conectado com *y*", do inglês *connected*, e *co(x,y)* quer dizer: "*x* está se amalgamando com *y*", do inglês *coalescing*.

Dado δ, que é um valor de distância pré-definido e a função *dist(x,y,D)* que retorna o valor da distância entre os objetos *x* e *y* da Cena *D*, o oráculo de estados relaciona as regiões espaciais dos objetos da cena aos seguintes conceitos:

$dc(x,y) \in O^d(D) \leftrightarrow dist(x,y,D) > \delta$

$ec(x,y) \in O^d(D) \leftrightarrow dist(x,y,D) \leq \delta \wedge dist(x,y,D) \neq 0$

$co(x,y) \in O^d(D) \leftrightarrow dist(x,y,D) = 0$

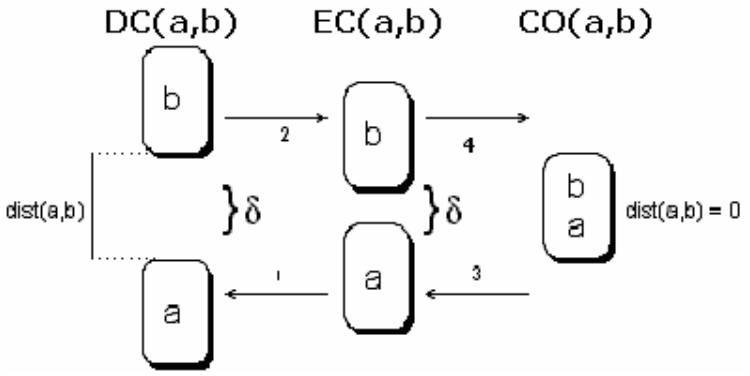
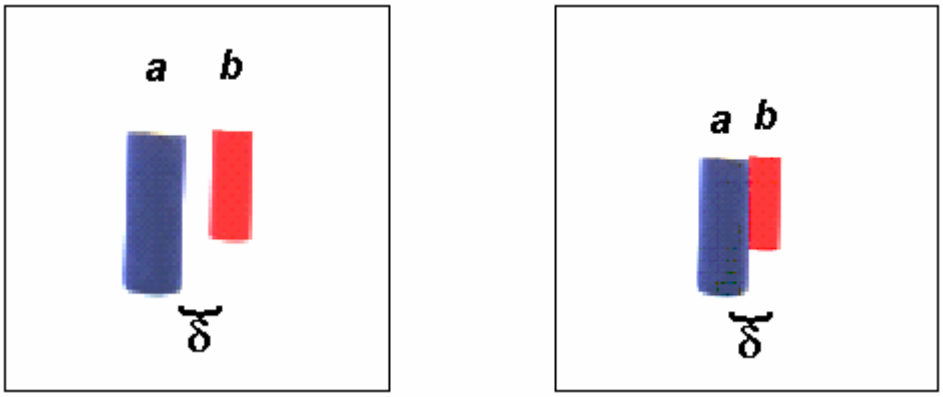


Figura 12 - Relações espaciais.

Logo, o Oráculo de Estados relaciona as regiões espaciais identificadas na cena *D* com as relações dinâmicas apresentadas na figura 13. Abaixo um exemplo em ambiente real.



(a) Relação EC(a,b) (b) Relação CO(a,b)

Figura 13 - Relações espaciais em ambiente real.

Oráculo de Transições

A partir de pares consecutivos de cenas de uma seqüência de imagens, o objetivo do oráculo de transições, O^t, é fornecer quais transições ocorreram nas relações espaciais das cenas em questão. No decorrer deste trabalho utilizaremos os termos cenas e estados sem distinção.

A seguir descrevemos algumas definições de transições entre regiões de uma imagem conforme (SANTOS, P. e SANTOS, M., 2005). Dado os estados D_1 e D_2, regiões da imagem $i(a)$ e $i(b)$ representando, respectivamente, os objetos a e b em D_1 e D_2, temos:

- **naosemove(x)**, que significa "o objeto x não se move":

$naosemove(i(a)) \in O^t(D_1,D_2)$ *se e somente se* $d_1=dist(i(a),i(b)) \in D_1$ $d_2=dist(i(a),i(b)) \in D_2 \wedge d_1= d_2$

- **aproxima(x,y)**, lê-se como "objeto x se aproxima do objeto y":

$aproxima(i(a),i(b)) \in O^t(D_1,D_2)$ *se e somente se* $ec(i(a),i(b)) \in D_1 \wedge \neg co(i(a),i(b)) \in D_2$ $(d_1=dist(i(a),i(b)) \in D_1)$ $(d_2= dist(i(a),i(b)) \in D_2) \wedge d_1> d_2$

- **colide (x,y)**, lê-se como "objeto x se colide com o objeto y":

$colide\ (i(a),i(b)) \in O^t(D_1,D_2)$ *se e somente se* $dc(i(a),i(b)) \in D_1 \wedge ec(i(a),i(b)) \in D_2$ $(d_1=dist(i(a),i(b)) \in D_1)$ $(d_2= dist(i(a),i(b)) \in D_2) \wedge d_1> d_2$

- **oclusao(x,y)**, "objeto x se oclui no objeto y":

$oclusao(i(a),i(b)) \in Ot(D_1,D_2)$ *se e somente se* $ec(i(a),i(b)) \in D_1 \wedge co(i(a),i(b)) \in D_2 \wedge ar(x)=0 \in D_2$

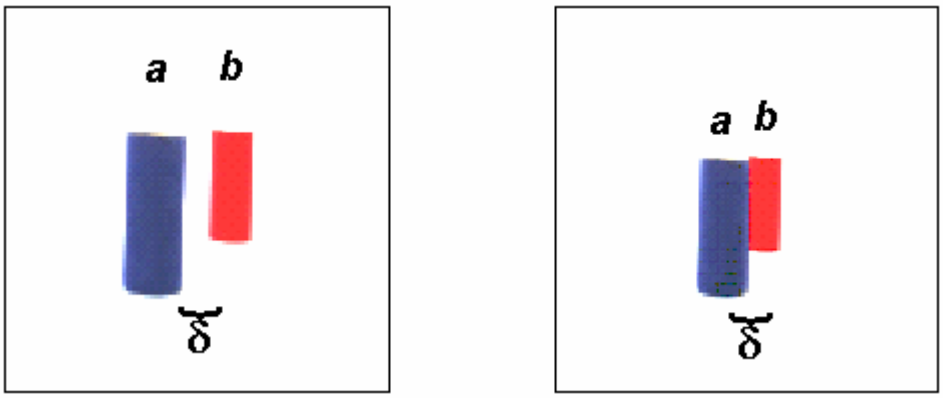
- **desoclusao(x,y)**, "objeto x se desoclui no objeto y":

$desoclusao(i(a),i(b)) \in O^t(D_1,D_2)$ *se e somente se* $co(i(a),i(b)) \in D_1 \wedge (ec(i(a),i(b)) \in D_2) \wedge ar(x)=0 \in D_1$

- **afasta(x,y)**, "objeto x se afasta do objeto y":

$afasta(i(a),i(b)) \in O^t(D_1,D_2)$ *se e somente se* $dc(i(a),i(b)) \in D_1 \wedge dc(i(a),i(b)) \in D_2$ $(d_1=dist(i(a),i(b)) \in D_1)$ $(d_2= dist(i(a),i(b)) \in D_2) \wedge d_1< d_2$

A partir destas definições conseguimos saber quais foram as mudanças das relações espaciais em cenas consecutivas. Conforme exemplo a seguir (figura 14).



(a) ec(i(a),i(b)) ∈ D1 (b) co(i(a),i(b)) ∈ D2

Figura 14 - Transição oclusao(i(a),i(b)).

Para a interpretação de longas seqüências é necessário inferir com base nas mudanças ocorridas em todo o caminho dado, através das informações dos oráculos, para isto, a *T-Logic* utiliza as três regras de inferência do método *SDL-Style* da TR.

Provando fórmulas usando o sistema de inferência

Supomos uma seqüência de imagens formada por três cenas contendo dois objetos *p* e *q*. Sendo assim, o nosso caminho é $\langle D_1,D_2,D_3\rangle$. Assumimos no estado D_1, *p* desconectado de *q*; no estado D_2, *p* e *q* continuam desconectados, mas a distância que os separa é menor; e no estado D_3 *p* e *q* estão externamente conectados, isto é:

$$\begin{aligned} &dc(p,q) \in O^d(D_1) \\ &dc(p,q) \in O^d(D_2) \\ &ec(p,q) \in O^d(D_3) \\ &dist(p,q,D_1) > dist(p,q,D_2) \end{aligned} \qquad (2)$$

A partir destas condições utilizaremos o sistema de inferência *SDL-Style* para provar que $aproxima(p,q) \otimes colide(p,q)$ é verdade no caminho $\langle D_1,D_2,D_3\rangle$, isto é provar que inicialmente os objetos *p* e *q* estão se aproximando e depois se colidem.

$P, \langle D_1\rangle \circ \langle D_2,D_3\rangle \models aproxima(p,q) \otimes colide(p,q)$

Se $P, \langle D_1,D_2\rangle \circ \langle D_3\rangle \models colide(p,q)$ por inferência da Regra 3 e (2)

Se $P, \langle D_1,D_2,D_3\rangle \circ \langle\ \rangle \models (\ )$ por inferência da Regra 3 e (2)

Através do sistema de inferência *SDL-Style*, foi possível deduzir a fórmula.

Obtendo uma explanação através das mudanças no caminho

Na seção 4.3 nós utilizamos o sistema de inferência para provar que a fórmula $aproxima(p,q) \otimes colide(p,q)$ é verdade dado o caminho $\langle D_1,D_2,D_3\rangle$. Agora veremos como conseguir a fórmula através das mudanças dos estados à medida que estas mudanças forem ocorrendo (SANTOS, P. e SANTOS, M., 2005). Isto é, iremos procurar uma fórmula Δ, que seja verdade em todo o caminho, a qual representa alguma transição, possibilitando interpretar, por dedução, qual o movimento dos objetos no caminho.

A partir do caminho dado π, nós obtemos $^{k} \otimes state \equiv \underbrace{\text{state} \otimes ... \otimes \text{state}}_{k}$, que é uma fórmula também verdadeira em π, onde k é o tamanho do caminho π. Nós também obtemos a clausula objetivo $\leftarrow G_0$, isto é uma seqüente do tipo:

$$P,\pi \models {}^{k}\otimes state \qquad (3)$$

onde $\pi = \langle D_1 \rangle \circ \langle D_2 \ldots D_k \rangle$.

Para procurar a fórmula objetivo Δ_n, que é verdade no caminho π, uma refutação da forma $\leftarrow G_0, \Delta_0 \ldots \leftarrow G_n, \Delta_n$, é construída na qual:

- cada G_i é uma seqüente;
- cada Δ_i é uma fórmula objetivo;
- G_n é uma seqüente $P,\pi \models ( )$, exemplo, o axioma do sistema de inferência apresentando no capítulo 2;
- Δ_0 é uma fórmula vazia.

Cada $\ldots \leftarrow G_{i+1}, \Delta_{i+1}$ é obtido de $\leftarrow G_i, \Delta_i$, sendo assim assumimos:

$$G_i = P, \quad \langle D_1 \ldots D_j \rangle \circ \langle D_{j+1} \ldots D_n \rangle \models {}^{m}\otimes state,\ m \geq 1$$

Nós assumimos as seguintes regras para obter Δ_{i+1}:

Regra A: Se existe um predicado d tal que $d \in O^t(D_j, D_{j+1})$, então $\Delta_{i+1} = \Delta_i \otimes d$

Regra B: Se a teoria inclui uma fórmula $d \leftarrow \phi$ e ϕ produz uma mudança no caminho $\langle D_1 \ldots D_j \rangle$ é Δ_i, exemplo P, $\langle D_1 \ldots D_j \rangle \models \Delta_i \wedge \phi$, então $\Delta_{i+1} = d$.

A seguir um exemplo para ilustrar o procedimento deste trabalho. A teoria P inclui o par de oráculos e P a regra abaixo:

$$oscila(O_1,O_2) \leftarrow aproxima(O_1,O_2) \otimes afasta(O_1,O_2) \otimes aproxima(O_1,O_2) \qquad (4)$$

Dado também os objetos distintos a e b, a seqüência pode ser interpretada da seguinte forma:

$$P, \langle D_1 \rangle \circ \langle D_2\ D_3\ D_4 \rangle \models {}^{4}\otimes state, ( )$$

Se $P, \langle D_1\ D_2 \rangle \circ \langle D_3\ D_4 \rangle \models {}^{3}\otimes state$, $aproxima(a,b)$, a partir da **Regra A**, assumimos $O^t(D_1,D_2) = aproxima(a,b)$

Se $P, \langle D_1\ D_2\ D_3 \rangle \circ \langle D_4 \rangle \models {}^{2}\otimes state, aproxima(a,b) \otimes afasta(a,b)$, a partir da **Regra A**, assumimos $O^t(D_2,D_3) = afasta(a,b)$

Se $P, \langle D_1\ D_2\ D_3\ D_4 \rangle \circ \langle\ \rangle \models \otimes state$, $aproxima(a,b) \otimes afasta(a,b) \otimes aproxima(a,b)$, a partir da **Regra A**, assumimos $O^t(D_3,D_4) = aproxima(a,b)$

Se $P, \langle D_1\ D_2\ D_3\ D_4 \rangle \circ \langle\ \rangle \models ( ), oscila(a,b)$, a partir da **Regra B** e da fórmula 4.

Portanto, dado o caminho π, por exemplo uma seqüências de imagens de uma mesma cena, e a teoria *P*, a *T-Logic* fornece um método para explanar as mudanças ocorridas em seqüências de imagens encontrando fórmulas Δ que descrevem movimentos entre objetos em todo o caminho.

Neste capítulo apresentamos a *T-Logic* (SANTOS, P. e SANTOS, M., 2005), que é o formalismo para interpretar seqüências de imagens a partir do raciocínio espacial qualitativo, que foi usado do desenvolvimento deste trabalho.

DEFINIÇÃO E IMPLEMENTAÇÃO DO SISTEMA

Neste capítulo iremos apresentar a definição e implementação do sistema de interpretação de seqüências de imagens, ilustrada com o problema de interpretação do movimento de rotação entre dois objetos em torno de um eixo fixo comum a eles. Este capítulo é organizado da seguinte forma. Na seção 5.1 apresentaremos as novas regras do Oráculo de Estados necessárias para a interpretação do movimento de rotação. Na seção 5.2 apresentaremos as novas regras do Oráculo de Transições. Também nesta seção, definiremos as regras de interpretação dos movimentos de rotação e anti-rotação, conforme discutido no capítulo 1. Na seção 5.3 iremos discutir a implementação do Oráculo de Estados, Oráculo de Transições; e das regras de rotação e anti-rotação, sendo que estas regras foram executadas utilizando máquina de estados. Na seção 5.4 provaremos que as regras responsáveis por interpretar longas seqüências de imagens podem ser traduzidas em máquinas de estados. E por fim, na seção 5.5 discutiremos o funcionamento do sistema capaz de interpretar movimentos de rotação em seqüências de imagens.

Este capítulo apresenta as principais contribuições deste trabalho de mestrado. Validações desta idéias serão apresentadas no capítulo 6.

Novas regras do Oráculo de Estados

Para a interpretação de movimentos de rotação identificamos duas novas relações espaciais: *esquerda(x,y)*, que se lê "o objeto *x* está a esquerda do objeto *y*", e identifica qual dos objetos está a esquerda na cena; e o *ar(x)*, que se lê: "área do objeto *x*". A relação *ar(x)* tem o seu princípio diferente os axiomas criados em (SANTOS, P. e SANTOS, M., 2005) e citados na seção 4.1, pois este axioma não está relacionado a nenhum outro objeto, o *ar(x)* apenas representa o valor da área da região em seu argumento.

Dado δ, que é um valor de distância pré definido entre os objetos *x* e *y* da cena *D; área* é o valor da área da região na cena D_i, e *vert_sup_esq* é a posição do vértice superior esquerdo objeto na imagem D_i. O oráculo de Estados é composto pelas seguintes relações espaciais:

- dc(x,y), significa "*x* está desconectado de *y*", *dc* do inglês *disconnected*.

$$dc(x,y) \in O^d(D) \leftrightarrow dist(x,y,D) > \delta$$

- ec(x,y), lê-se como: "*x* externamente conectado com *y*", *ec* do inglês *externally connected*.

$$ec(x,y) \in O^d(D) \leftrightarrow dist(x,y,D) \leq \delta \wedge dist(x,y,D) \neq 0$$

- co(x,y), quer dizer: "*x* está se amalgamando com *y*" , *co* do *inglês coalescing*

$$co(x,y) \in O^d(D) \leftrightarrow dist(x,y,D) = 0$$

-ar(x), que se lê: "área do objeto *x"*.

$$ar(x) \in O^d(D) \leftrightarrow área(x,D)$$

- esquerda(x,y), significa *"o objeto x está a esquerda do objeto y"*.

$$esquerda(x,y) \in O^d(D) \leftrightarrow vert\_sup\_esq(x,D) < vert\_sup\_esq(y,D)$$

Por tanto *esquerda(x,y)* é transitivo, irreflexivo e anti-simétrico

As relações *dc*, *ec* e *co* foram repetidas neste capítulo por conveniência. As relações *ar* e *esquerda* são introduzidas pela primeira vez.

Novas regras do Oráculo de Transições

O Oráculo de Transições, apresentado na seção 4.2, foi acrescido de dois novos axiomas, para interpretar movimentos de rotação. Estes axiomas utilizam a relação espacial *ar(x).* As relações responsáveis pelas transições de pares de cenas são: *aproxobs(x)*, que significa objeto *x* se aproxima do observador, e o *afastobs(x)*, lê-se como objeto *x* se afasta do observador.

Estes novos axiomas fornecem as transições dos objetos em relação à sua profundidade na cena, isto é, se no decorrer do caminho a área do objeto diminuir, significa que o objeto está mais distante do sensor; inversamente, se a área do objeto aumentar, significa que o objeto se aproxima do sensor. Vale salientar que este trabalho assume os objetos observados são rígidos.

A seguir descrevemos as definições de todas as regras de transições entre regiões de uma imagem que compõe o Oráculo de Transições. Dado os estados D_1 e D_2, regiões *i(a)* e *i(b)* representando, respectivamente, as imagens dos objetos *a* e *b* na cena, temos:

- **aproxobs(x)**, que significa objeto *x* se aproxima do observador:

$$aproxobs(i(a)) \in O^t(D_1,D_2) \text{ se somente se } ar(i(a)) \in D_1 \wedge ar(i(a)) \in D_2$$
$$ar_1 = área(i(a)) \in D_1 \wedge ar_2 = área(i(a)) \in D_2 \wedge ar_1 < ar_2$$

- **afastobs(x)**, lê-se como objeto *x* se afasta do observador:

afastobs(i(a)) ∈ $O^t(D_1,D_2)$ se somente se ar(i(a)) ∈ D_1 ∧ ar(i(a)) ∈ D_2 (ar_1=área(i(a),D_1)) (ar_2=área(i(a),D_2)) ∧ ar_1>ar_2

- **oclusaodir(x,y)**, objeto *x* se oclui pelo lado direito do objeto *y*:

oclusaodir(i(a),i(b)) ∈ Ot(D_1,D_2) se somente se dc(i(a),i(b)) ∈ D_1 ∨

ec(i(a),i(b)) ∈ D_1) ∧ esquerda(i(a), i(b)) ∈ D_1 ∧ co(i(a),i(b)) ∈ D_2 ∧ ar(x)=0 ∈ D_2

- **oclusaoesq(x,y)**, objeto *x* se oclui pelo lado direito do objeto *y*:

oclusaoesq(i(a),i(b)) ∈ Ot(D_1,D_2) se somente se ec(i(a),i(b)) ∈ D_1 ∧

esquerda(i(b), i(a)) ∈ D_1 ∧ co(i(a),i(b)) ∈ D_2 ∧ ar(x)=0 ∈ D_2

- **desoclusaodir(x,y)**, objeto *x* se desoclui do lado direito do objeto *y*:

desoclusaodir(i(a),i(b)) ∈ $O^t(D_1,D_2)$ se somente se co(i(a),i(b)) ∈ D_1 ∧

ec(i(a),i(b)) ∈ D_2 ∧ esquerda(i(b), i(a)) ∈ D_1 ∧ ar(x)=0 ∈ D_1

- **desoclusaoesq(x,y)**, objeto *x* se desoclui do lado direito do objeto *y*:

desoclusaodir(i(a),i(b)) ∈ $O^t(D_1,D_2)$ se somente se co(i(a),i(b)) ∈ D_1 ∧

ec(i(a),i(b)) ∈ D_2 ∧ esquerda(i(a), i(b)) ∈ D_1 ∧ ar(x)=0 ∈ D_1

Estas são as regras que compõem o Oráculo de Transições, que retorna as transições ocorridas em pares de cenas. Também compõem o Oráculo de Transições, as regras de interpretação de movimentos de rotação e anti-rotação. Contudo, sua análise é feita em longos caminhos, não só em pares de cenas. Portando, criamos as regras: *rotacao(x,y)*, que se lê: os objetos *x* e *y* se rotacionam entre si e *antirotacao(x,y)*, que se lê: os objetos *x* e *y* anti-rotacionam entre si. A seguir suas definições:

rotacao(i(a),i(b)) ← *elementar1(i(a),i(b)) ⊗ oclusaodir(i(b),i(a))* (5)

⊗ desoclusaoesq(i(b),i(a)) ⊗ elementar2(i(a),i(b))

rotacao(i(a),i(b)) ← *elementar1(i(a),i(b)) ⊗ oclusaoesq(i(b),i(a))* (5)

⊗ desoclusaodir(i(b),i(a)) ⊗ elementar2(i(a),i(b))

antirotacao(i(a),i(b)) ← *elementar1(i(a),i(b)) ⊗ oclusaodir(i(b),i(a))* (6)

⊗ desoclusaodir(i(b),i(a)) ⊗ elementar2(i(a),i(b))

antirotacao(i(a),i(b)) ← *elementar1(i(a),i(b)) ⊗ oclusaoesq(i(b),i(a))* (6)

⊗ desoclusaoesq(i(b),i(a)) ⊗ elementar2(i(a),i(b))

Sendo:

elementar1(i(a),i(b)) ← *aproxobs(i(a)) ∧ afastobs(i(b))*

elementar2(i(a),i(b)) ← *aproxobs(i(b)) ∧ afastobs(i(a))*

A partir deste formalismo, composto pelas regras do Oráculo de Estados e do Oráculo de Transições, implementamos um sistema capaz de interpretar movimento de rotação de dois objetos em torno de um eixo fixo comum, que passamos a discutir a seguir.

Implementação do Sistema

Nessa seção discutiremos a implementação do Módulo 2 do sistema, figura 2, capítulo 1. Para isso, discutiremos a implementação das regras dos oráculos descritos na seção anterior. Também apresentaremos a inferência do movimento dos objetos pela resolução *SDL-Style* descrito na seção 2.3, utilizando a nova regra de inferência introduzida na seção 5.2.

1.1.1 Implementação dos Oráculos

A implementação dos oráculos no sistema parte das relações espaciais entre os objetos na cena. Identificamos e implementamos no sistema três procedimentos: o primeiro é o *dist*, que retorna o valor da distância entre o vértice inferior direito da região do objeto que esta à esquerda na cena e o vértice inferior direito da região do objeto a esquerda na cena, o segundo é o *área*, que retorna o valor da área da região que representa os objetos e o terceiro é o *vert_sup_esq*, que retorna o valor do vértice esquerdo superior da região do objeto.

Dado δ, que é um valor não nulo de distância pré-definido entre os objetos x e y, conforme apresentado na seção 4.1, se o procedimento *dist* retornar o valor zero, o sistema identifica a cena pelo prefixo *co(x,y)* em que um objeto x está se amalgamando ao objeto y. Para o valor de *dist* maior que δ, o sistema então identifica a cena pelo prefixo *dc(x,y)*, representando que o objeto x está desconectado do objeto y. Se o valor de *dist* for maior que zero mas menor que δ identificamos a cena pelo prefixo *ec(x,y)*, representando que o objeto x está externamente conectado do objeto y.

Para a relação *ar(x)*, utilizamos o procedimento *área* que utiliza as informações extraídas pela segmentação da imagem no Módulo 1 do sistema, figura 2, que retorna o valor da área da região do objeto, conforme discutido no capítulo 3.

A partir do valor que o procedimento *vert_sup_esq* fornece, conseguimos identificar qual objeto está à esquerda na cena, isto é, se o valor de *vert_sup_esq* da região x é menor que o valor de *vert_sup_esq* da região y, significa o que o objeto x está a esquerda na cena, caso contrário, o objeto y está a esquerda do objeto x.

A partir destas funções, implementamos as relações que analisam as transições nas cenas, as relações que compõem o Oráculo de Transições, conforme descrito nas Seções 3.1.3 e 4.2. As funções que criamos para os axiomas que compões o Oráculo de Transições na implementação dos sistemas são:

aproxobs(x): $ar(x) \in D_i > ar(x) \in D_{i-1}$

afastobs(x): $ar(x) \in D_i < ar(x) \in D_{i-1}$

oclusaodir(x,y): $dist(x,y) \in D_i > dist(x,y) \in D_{i-1}$ *e* $ar(x) = 0 \in D_i$ *e* $ar(x) \neq 0 \in D_{i-1}$ *e esquerda(x,y)* $\in D_{i-1}$

oclusaoesq(x,y): $dist(x,y) \in D_i > dist(x,y) \in D_{i-1}$ *e* $ar(x) = 0 \in D_i$ *e* $ar(x) \neq 0 \in D_{i-1}$ *e esquerda(y.x)* $\in D_{i-1}$

desoclusaodir(x,y): $dist(x,y) \in D_i < dist(x,y) \in D_{i-1}$ *e* $ar(x) \neq 0 \in D_i$ *e* $ar(x) = 0 \in D_{i-1}$ *e esquerda(x,y)* $\in D_{i-1}$

desoclusaoesq(x,y): $dist(x,y) \in D_i < dist(x,y) \in D_{i-1}$ *e* $ar(x) \neq 0 \in D_i$ *e* $ar(x) = 0 \in D_{i-1}$ *e esquerda(y,x)* $\in D_{i-1}$

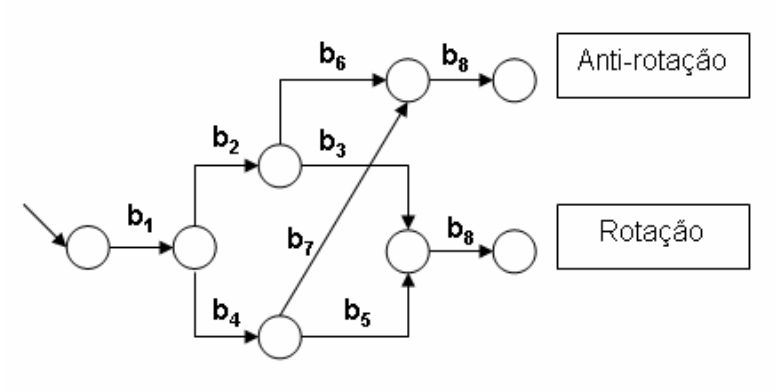
No sistema foram também implementadas. como procedimentos as regras do Oráculo de Estados e as regras do Oráculo de Transições que analisam pares de cenas. As regras do Oráculo de Transições que analisam longas seqüências foram implementadas utilizando uma máquina de estados conforme descrita a seguir.

1.1.2 Implementação da Regra de Inferência

Na Seção 5.2 definimos ao Oráculo de Transições as regras que interpretam movimentos de rotação e anti-rotação, que são implementadas como uma máquina de estados com 8 estados, figura 15, sendo que cada estado representa uma transição necessária para identificar os movimentos. Utilizamos máquina de estados para a implementação de longas seqüências, porque a máquina de estados pode prever qual será a próxima transição entre cenas, possibilitando inferir qual será a relação espacial dos objetos na próxima cena, pois, de acordo com o estado corrente da máquina de estados, podemos saber qual a próxima transição poderá vir a satisfazer a regra de inferência e, conseqüentemente, a próxima relação espacial.

Para cada transição identificada pelo oráculo de transições, o sistema procura na máquina de estados, representada na figura 15, se tal transição faz parte da regra de inferência do movimento. Se a transição satisfizer parte da conjunção serial, o sistema incrementa a máquina de estados mudando-a para um estado subseqüente e aguardando uma nova

transição. Assim que o sistema identificar as 5 transições, ou melhor, assim que o sistema percorrer os 5 estados, identifica que os objetos da seqüência de imagens estão em movimento de rotação ou anti-rotação. A figura 15 mostra a máquina de estados do sistema de interpretação de movimento de rotação, regra (5), e anti-rotação, regra (6), em seqüências de imagens.



Sendo:

- $b_{1:}$ *elementar1(i(a),i(b))*
- $b_{2:}$ *oclusaodir(i(b),i(a))*
- $b_{3:}$ *desoclusaoesq(i(b),i(a))*
- $b_{4:}$ *oclusaoesq(i(b),i(a))*
- $b_{5:}$ *desoclusaodir(i(b),i(a))*
- $b_{6:}$ *desoclusaodir(i(b),i(a))*
- $b_{7:}$ *desoclusaoesq(i(b),i(a))*
- $b_{8:}$ *elementar2(i(a),i(b))*

Figura 15 – Máquina de estados regras (5) e (6).

Os estados do mundo se relacionam com o corpo das regras dos oráculos por mudanças ocorridas no grau de conectividades dos objetos presentes nestes estados.

Mudanças ocorridas nos estados do mundo, relacionam as regras dos oráculos em transições que fazem parte da regra de inferência implementada por meio de uma máquina de estados. Na seqüência discutimos o funcionamento geral do sistema.

1.1.3 O Sistema

No capítulo 3, discutimos a implementação do módulo de extração de informações da seqüência de imagens, as quais foram gravadas na matriz *MatrizInfo* e, nesta encontram-se os dados de: regiões, áreas, cores e posições de todos os objetos das cenas, que são parâmetros de entrada para os oráculos.

Para a interpretação de seqüências de imagens, o sistema analisa, cena a cena, de acordo com os procedimentos descritos a seguir.

O sistema filtra as regiões representadas na *MatrizInfo*, selecionando apenas as regiões cujos os valores de áreas são maiores que 200 pixels e menores que 5000 pixels de cada cena. As regiões com áreas menores que 200 pixels significam ruído da imagem e, por isso, devem ser desprezadas. O sistema também despreza as áreas maiores que 5000 pixels, que são consideradas regiões do fundo das imagens. Desta forma o módulo de interpretação de alto nível considera somente as regiões dos objetos contidos nas cenas.

Após selecionar os objetos da cena, suas informações são parâmetros de entrada para o Oráculo de Estados, que relaciona as regiões espaciais nas imagens aos conceitos sobre espaço livre e conectividade ,seção 5.1.

Uma etapa importante do sistema é identificar consistentemente regiões em cenas subseqüentes de uma seqüência, ou seja, identificar os objetos no decorrer das cenas. Isso é feito comparando as cores dos objetos e também comparando as áreas e as suas posições. Na comparação por cor o sistema respeita um intervalo neste atributo das regiões, pois pode existir uma variância na cor dos objetos no decorrer das cenas devido a influência da luminosidade. O sistema também considera que estes parâmetros de área e posição não podem sofrer grandes variações com relação ao mesmo objeto durante a seqüência. Isto se faz válido, assumindo um conceito de bom senso e continuidade, conforme discutido em (MANN, JEPSON e SISKIND, 1996), em que cada objeto não pode dar grandes saltos, nem mesmo deixar de existir, durante uma mesma seqüência de imagem.

Por exemplo, na figura 1 do capítulo 1, temos dois objetos em movimento de rotação, na transição da Cena 1 para a Cena 2, o Objeto 1, objeto cinza claro, se aproxima do observador, isso significa que existe uma alteração no valor de área; tal mudança no valor da área do Objeto 1 não pode ser muito grande, pois isso significaria que na Cena 2 o objeto estaria bem próximo ao sensor, dando um grande salto na transição de cenas da seqüência e isto não por ocorrer. Sendo assim, limitamos a variância possível de área e posição dos objetos de uma imagem para a outra a um valor pré-determinado.

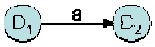
A partir da segunda cena analisada, o Oráculo de Transições com as relações espaciais das regiões dos objetos fornecidas pelo Oráculo de Estados, verifica e armazena quais foram as transições ocorridas em pares de cenas.

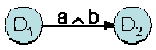
Através de uma máquina de estados, o sistema analisa a possibilidade de as transições ocorridas em pares de cenas fazerem parte do corpo de alguma das regras definidas no Oráculo de Transições.

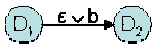
Na seção 5.5 discutiremos o funcionamento deste processo de previsão do sistema. A seguir, conjeturaremos que é possível traduzir axiomas da *T-Logic* em máquina de estados.

Tradução de Fórmulas da T-Logic em Máquina de Estados

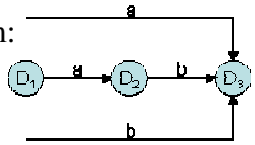
Para interpretar longas seqüências e poder prever quais as próximas relações espaciais das regiões dos objetos no decorrer da seqüência, implementamos as regras dos movimentos através de uma máquina de estados. Isto é possível porque as fórmulas da *T-Logic* podem ser traduzidas em uma máquina de estados, sendo as regras de tradução:

1 – A fórmula $D_1 \circ D_2 \models a$, na qual a transição a ocorre do estado D_1 a D_2, pode ser traduzida em: 

2 – A fórmulas $D_1 \circ D_2 \models a \wedge b$, na qual onde a transição $a \wedge b$ ocorre do estado D_1 a D_2, pode ser traduzida em: 

3 – A fórmulas $D_1 \circ D_2 \models a \vee b$, na qual onde a transição $a \vee b$ ocorre do estado D_1 a D_2, pode ser traduzida em: 

5 – A fórmulas $D_1,D_2 \circ D_3 \models a \otimes b$, na qual onde a transição a ocorre do estado D_1 a D_2 e a transição b ocorre do estado D_2 a D_3 pode ser traduzida em: 

6 – A fórmulas $D_1,D_2 \circ D_3 \models a \oplus b$, na qual onde a transição a pode ocorrer do estado D_1 a D_2 e a transição b pode ocorrer do estado D_2 a D_3 pode ser traduzida em: 

As regras de tradução acima formam o caso base de uma indução no tamanho das fórmulas TR, o que prova que qualquer fórmula TR pode ser traduzida em uma máquina de estados.

Funcionamento do Sistema

Vamos descrever um exemplo do funcionamento do sistema de interpretação de seqüências de imagens para a seqüência de imagens apresentadas na figura 1, capítulo 1.

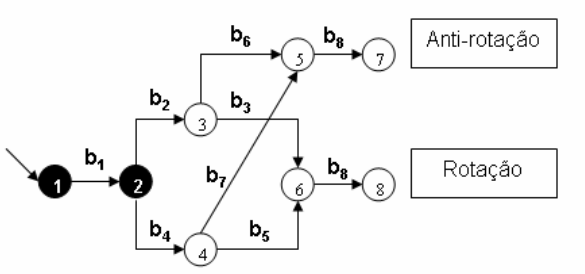
O Módulo 1 do sistema, figura 2, extrai as informações de área, cor e posição dos objetos da imagem, conforme descrito na capítulo 3. Estas informações são gravadas na *MatrizInfo*.

Com estas informações inicia-se a interpretação: o sistema varre a *MatrizInfo* selecionando as regiões com área maior de 200 pixels e menor de que 5000, atendo-se apenas

às regiões correspondentes aos objetos na cena. As informações de área e posição dos objetos são dados de entrada das funções implementadas no oráculo de estados. Este, por sua vez, relaciona as regiões espaciais a conceitos de espaço livre e conectividade dos objetos na cena. Por exemplo: da Cena 1 e Cena 2 conforme figura 1 apresentada na capítulo 1, o oráculo de estados relaciona as regiões espaciais às fórmulas: $dc(a,b,D_1)$.

A partir da segunda cena analisada pelo sistema, esta relação é dado de entrada para as funções que compõem o Oráculo de Transições, que analisa as transições das mudanças ocorridas em pares de cenas. O sistema seleciona as funções das transições encontradas nas cenas consecutivas. Por exemplo: para a Cena 1 e Cena 2 o Oráculo identifica a seguinte transição: o *i(a)*, cinza escuro, se afasta do observador; e, o *i(b)*, objeto cinza claro, se aproxima do observador.

A partir das transições encontradas, o sistema verifica se as transições satisfazem a regra do movimento de rotação ou anti-rotação usando a máquina de estados, na qual, estas regras foram implementadas. As transições encontradas entre a mudança ocorrida da Cena 1 para a Cena 2, figura 1, satisfazem o início na regra, portanto a máquina de estados é atualizada: do Estado 1 para o Estado 2 com a transição *elementar1* que é composta pela transição $aproxobs(i(a)) \wedge afastobs(i(b))$, Este processo pode ser visto na representação da figura 16.



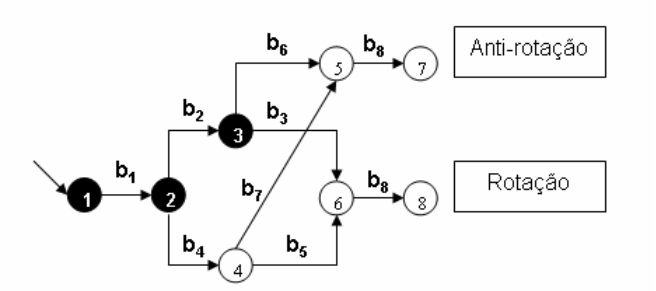
Sendo:

- **b_1: *elementar1(i(a),i(b))***
- b_2: *oclusaodir(i(b),i(a))*
- b_3: *desoclusaoesq(i(b),i(a))*
- b_4: *oclusaoesq(i(b),i(a))*
- b_5: *desoclusaodir(i(b),i(a))*
- b_6: *desoclusaodir(i(b),i(a))*
- b_7: *desoclusaoesq(i(b),i(a))*
- b_8: *elementar2(i(a),i(b))*

Figura 16 – Transição b1 atualizou o estado corrente de 1 para 2.

Também na figura 16, demonstramos que o sistema aguarda entre a transição *oclusaodir(i(b),i(a))* ou *oclusaoesq(i(b),i(a))* para atualizar a máquina de estados na qual foi implementada a regra de inferência de rotação e anti-rotação. Ocorrendo qualquer outra transição de estado no mundo, o sistema continua seu funcionamento normal, contudo sem atualizar os estados da regra, pois a partir deste momento a informação relevante para a inferência do movimento é a transição oclusão do objeto.

De acordo com a nossa seqüência de imagem, figura 1, a transição *oclusaodir(i(b),i(a))* apenas ocorre da Cena 5 para a Cena 6, na qual, também ocorrem as transições: o *i(a)*, objeto cinza escuro, se afasta do observador e se oclui a direita do objeto *i(b)* que por sua vez se aproxima do observador. Com estas transições encontradas, o sistema atualiza a máquina de estados do Estado 2 para o Estado 3, como mostra a figura 17. Neste momento o sistema aguarda a transição *desoclusaoesq(i(a),i(b))* ou *desoclusaodir(i(a),i(b))*.



Sendo:

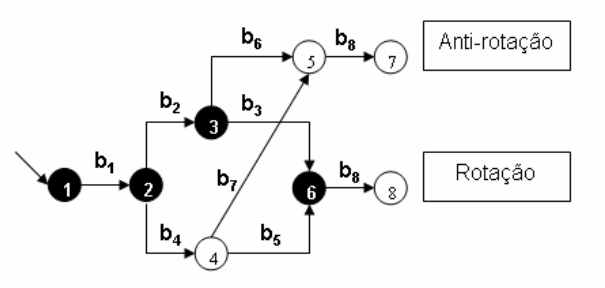
- **b_1: *elementar1(i(a),i(b))***
- **b_2: *oclusaodir(i(b),i(a))***
- b_3: *desoclusaoesq(i(b),i(a))*
- b_4: *oclusaoesq(i(b),i(a))*
- b_5: *desoclusaodir(i(b),i(a))*
- b_6: *desoclusaodir(i(b),i(a))*
- b_7: *desoclusaoesq(i(b),i(a))*
- b_8: *elementar2(i(a),i(b))*

Figura 17 – Transição b2 atualizou o estado corrente de 2 para 3.

No decorrer da análise da seqüência temos as seguintes transições encontradas pelo sistema: da Cena 6 para a Cena 7 o objetos *i(b)* se afasta do observador, o *i(a)*, que por sua vez se desoclui pela esquerda do *i(b)* e se aproxima do observador. A partir destas novas transições encontradas, o sistema atualiza a máquina de estados, do Estado 3 para o 6, com a transição *desoclusãoesq(i(b),i(b)),* conforme figura 18, aguardando neste momento a transição *elementar2(i(a),i(b)),* composta pela transição *aproxobs(i(b)) ∧ afastobs(i(a))*

A transição *elementar2(i(a),i(b))* ocorre da Cena 7 para a Cena 8, atualizando a máquina de estados, do Estado 6 para o 8, figura 19, fornecendo como saída do sistema a

descrição conceitual do movimento: *de rotação entre dois objetos em torno de um eixo fixo comum*. Esta saída é a descrição da interpretação, até este momento, do movimento dos objetos no decorrer da seqüência de imagens. A partir deste momento, o sistema atualiza o estado da máquina es estados para 1, dando início à nova análise para o decorrer da seqüência de imagens.



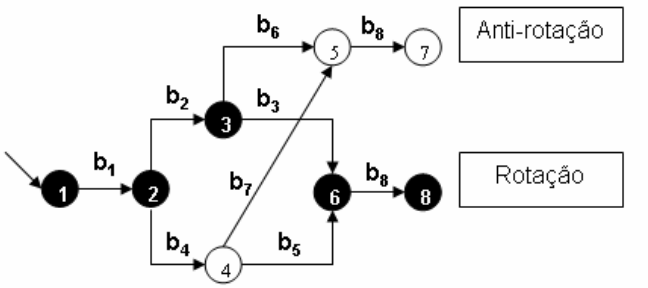
Sendo:

- **b_1: *elementar1(i(a),i(b))***
- **b_2: *oclusaodir(i(b),i(a))***
- **b_3: *desoclusaoesq(i(b),i(a))***
- b_4: *oclusaoesq(i(b),i(a))*
- b_5: *desoclusaodir(i(b),i(a))*
- b_6: *desoclusaodir(i(b),i(a))*
- b_7: *desoclusaoesq(i(b),i(a))*
- b_8: *elementar2(i(a),i(b))*

Figura 18 – Transição b3 atualizou o estado corrente de 3 para 6.

Um detalhe importante do funcionamento do sistema é quando o sistema analisa as cenas, mas não encontra as transições que completam a máquina de estados a partir de estados maiores que 1, o mesmo atualiza o estado atual da máquina para 1 e começa novamente o raciocínio sobre as cenas, vejamos um exemplo. O sistema identifica por exemplo, as transições: *elementar1(i(a),i(b))* entre a Cena 1 e a Cena 2, atualizando o estados atual da máquina de estados para 2. A partir deste momento o sistema aguarda a transição *oclusãodir(i(b),i(a))* ou *oclusãoesq(i(b),i(a))* para continuar a interpretação do movimento, caso o sistema ao analisar mais 5 pares cenas e não encontrar as transições *oclusãodir(i(b),i(a))* ou *oclusãoesq(i(b),i(a))*, o sistema atualiza o estado da máquina de estados para 1, para iniciar a nova análise para o decorrer da seqüência de imagens. Com mais este exemplo, demonstramos o funcionamento do sistema, que é capaz de interpretar movimento de rotação entre dois objetos sobre um eixo fixo, utilizando a *T-Logic* como formalismo de interpretação de seqüências de imagens utilizando raciocínio espacial qualitativo.

As principais contribuições deste sistema são: discutir a implementação de um sistema responsável por interpretar seqüências de imagens utilizando a *T-logic* como formalismo lógico; discutir como devem ser modeladas e construídas as regras para interpretação de longas seqüências - neste caso, exemplificamos com o movimento de rotação entre dois objetos com eixo fixo comum; construir um sistema capaz de raciocinar de forma qualitativa a interpretação de seqüências de imagens; mostrar uma parte da inteligência artificial voltada para problemas de senso comum.



Sendo:

- **b_1:** ***elementar1(i(a),i(b))***
- **b_2:** ***oclusaodir(i(b),i(a))***
- **b_3:** ***desoclusaoesq(i(b),i(a))***
- b_4: *oclusaoesq(i(b),i(a))*
- b_5: *desoclusaodir(i(b),i(a))*
- b_6: *desoclusaodir(i(b),i(a))*
- b_7: *desoclusaoesq(i(b),i(a))*
- **b_8:** ***elementar2(i(a),i(b))***

Figura 19 – Transição b8 atualizou o estado corrente de 6 para 8.

As restrições deste sistema, devido a simples implementação do Módulo 1, capítulo 3, são: o ambiente das cenas deve possuir fundo e iluminação controlada; a utilização de objetos rígidos; e a seqüência de imagens apresentadas ao sistema deve conter todas a transições modelas no movimento, caso contrário, o sistema não interpreta o movimento.

ANÁLISE DOS RESULTADOS

Neste capítulo discutiremos os resultados dos experimentos do sistema descrito no capítulo anterior, com o objetivo de demonstrar o seu funcionamento e limitações

Efetuamos os experimentos a partir de dois movimentos implementados no sistema: o movimento de rotação e o movimento ambíguo da rotação, que chamamos de movimento de anti-rotação. A rotação consiste em dois objetos cilíndricos que se rotacionam em torno de um eixo fixo comum aos objetos. A anti-rotação consiste em um movimento similar ao movimento de rotação até o momento da oclusão de um dos objetos; a partir deste ponto a desoclusão do objeto, na anti-rotação, ocorre do mesmo lado de onde o objeto foi ocluso; enquanto no movimento de rotação a desoclusão do objeto ocorre do lado oposto de onde o objeto foi ocluso. A anti-rotação é utilizada para demonstrar o comportamento do sistema diante de movimentos ambíguos.

A partir dos dois movimentos implementados, adotamos como experimentos quatro situações entre objetos de mesmo tamanho:

1- Dois objetos cores distintas e os objetos distantes um do outro na cena inicial;
2- Dois objetos, cores distintas e próximos entre si;
3- Dois objetos de mesma cor e mesmo tamanho e distantes um do outro na cena inicial;
4- Dois objetos de mesma cor e os objetos próximos um do outro nas cenas.

Adotamos outras duas situações para objetos com alturas diferentes e mesma largura, sendo objetos de mesma forma, apenas para analisar o comportamento do sistema:

1- Alturas diferentes, de cores distintas, distantes entrei si na cena inicial.
2- Alturas diferentes, de mesma cor, e distantes entre si na cena inicial.

Para cada movimento e situação entre objetos, utilizaremos nos experimentos cinco seqüência diferentes de cenas.

A tabela 1 representa os resultados obtidos em que *sim* significa que o sistema interpretou corretamente a seqüência analisada.

Tabela 1 – Síntese dos resultados

		Seqüência 1		Seqüência 2		Seqüência 3		Seqüência 4		Seqüência 5	
		Rotação	Anti Rotação	Rotação	Anti Rotação	Rotação	Anti Rotação	Rotação	Anti Rotação	Rotação	Anti Rotação
Cor Diferente	Objetos Distantes	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
	Objetos Próximos	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
	Objetos Tamanho Diferente	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim	Sim
Mesma Cor	Objetos Distantes	Não	Não	Sim	Sim	Sim	Sim	Não	Não	Sim	Sim
	Objetos Próximos	Não	Não	Não	Não	Não	Não	Não	Não	Não	Não
	Objetos Tamanho Diferente	Sim	Sim	Não	Não	Não	Não	Não	Não	Sim	Sim

Os experimentos efetuados no sistema são discutidos em duas etapas: na primeiras trataremos das cenas com objetos de cores distintas; e, na segunda etapa, das cenas com objetos de mesma cor.

No primeiro experimento da primeira etapa, figura 20, utilizamos cinco seqüências de cenas contendo dois objetos de mesmo tamanho, cores distintas e os objetos distantes entre si na cena inicial. Neste teste o sistema foi capaz de interpretar o movimento de rotação, assim como o seu ambíguo, para todas a cinco seqüências.

Este resultado mostra a eficiência do sistema para interpretação dos movimentos de rotação e anti-rotação para, objetos distantes entre si na cena inicial, sendo os objetos de cores distintas.

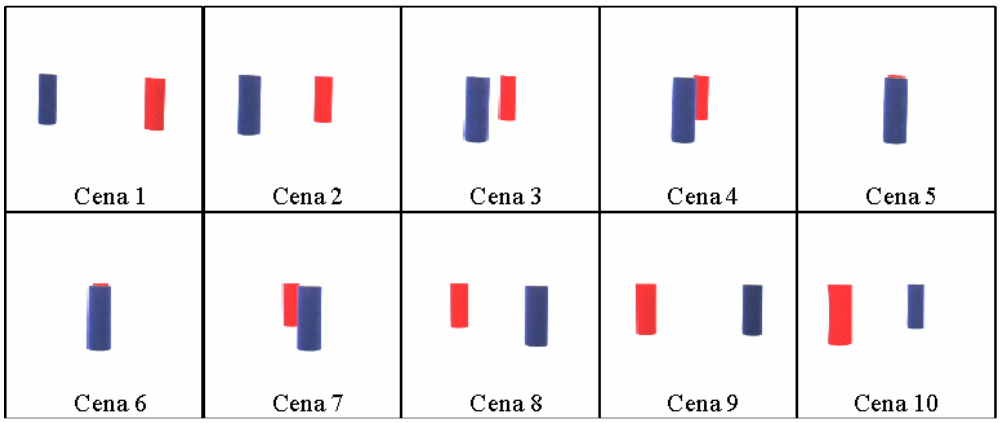


Figura 20 - Dois objetos de mesmo tamanho, cores distintas e os objetos distantes entre si na cena 1.

Efetuamos um novo teste com objetos próximos entre si na cena inicial. Neste segundo experimento, em que as seqüências possuem dois objetos de mesmo tamanho, de cores distintas e próximos entre si, figura 21, também utilizamos cinco seqüências distintas

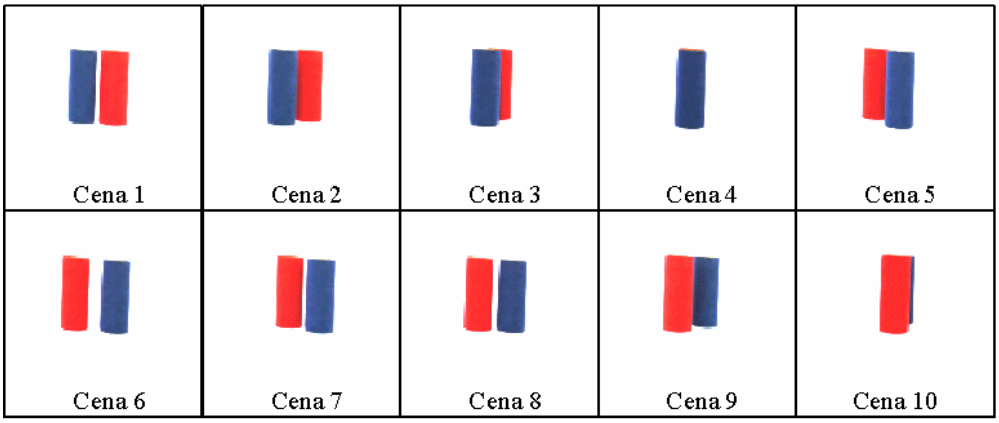


Figura 21 - Dois objetos de mesmo tamanho,cores distintas e próximos nas cenas

como entrada do sistema. Neste caso o sistema também foi capaz de interpretar os movimentos de rotação e anti-rotação.

Os resultados destes dois experimentos mostram que o sistema de interpretação de seqüências de imagens é capaz de interpretar os movimentos das seqüências das imagens independentemente das distâncias entre os objetos nas cenas, tratando-se apenas de objetos de mesmo tamanho e cores distintas.

Efetuamos um novo experimento no sistema, agora visando analisar o comportamento do sistema em relação a altura dos objetos diferentes, figura 22. Neste experimento utilizamos novas cinco seqüências distintas de imagens contendo objetos de alturas diferentes, mesma largura e cores distintas. Nesta experiência, o sistema também se mostrou capaz de interpretar os movimentos de rotação e anti-rotação.

Com o resultado deste terceiro experimento notamos que o sistema tem um funcionamento excelente para interpretação de seqüências de imagens contendo objetos de cores distintas, independente da distância e altura dos objetos nas cenas. Com este terceiro teste, terminamos a primeira etapa dos experimentos. Iniciaremos, então, a segunda etapa dos experimentos em que utilizaremos cenas contendo objetos de mesma cor.

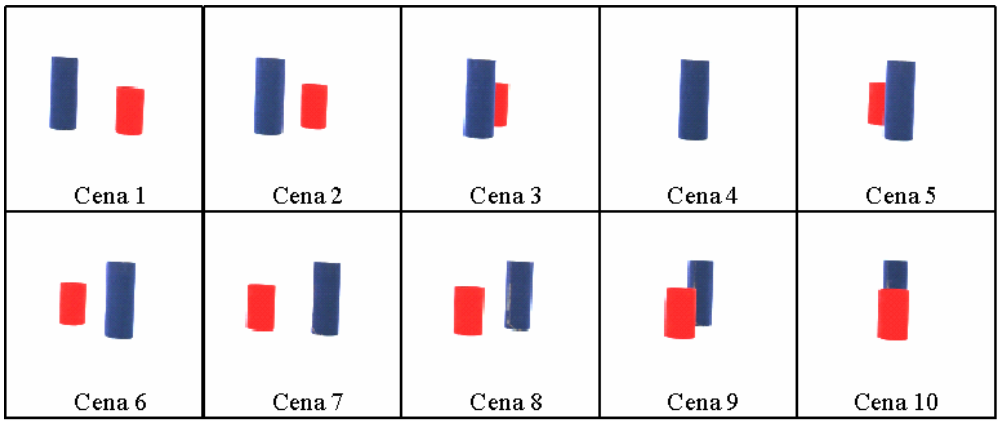


Figura 22 - Dois objetos de tamanhos diferentes, cores distintas e distantes entre si na cena inicial.

No primeiro experimento da segunda etapa utilizamos cinco seqüências de cenas contendo dois objetos de mesmo tamanho, mesma cor e inicialmente distantes entre si, figura 23.

Nesta situação o sistema conseguiu interpretar os movimentos de rotação e anti-rotação de três seqüências de um total de cinco seqüências a ele apresentado. O sistema não foi capaz de interpretar os movimentos de duas seqüências, pois, as cenas não foram extraídas

em intervalos de tempo constante, existindo transições nas imagens em que os objetos dão grandes saltos entre cenas. Isto significa que as áreas dos objetos diminuem ou aumentam consideravelmente, fazendo com que o sistema não identifique os objetos correspondentes no decorrer da seqüência. Este é um bom resultado para objetos de mesma cor, pois mesmo as cenas das seqüências não sendo extraídas em intervalo constante, contendo apenas as principais transições para a identificação dos movimentos, o sistema foi capaz de interpretar corretamente a maioria das seqüências para cenas com objetos de mesma cor, mesmo tamanho e objetos distantes entre si na cena inicial apresentadas.

Para identificarmos se as distâncias entre os objetos nas cenas influenciam a interpretação das seqüências, efetuamos um outro experimento para objetos de mesma cor e mesmo tamanho, diminuindo a distância entre os objetos.

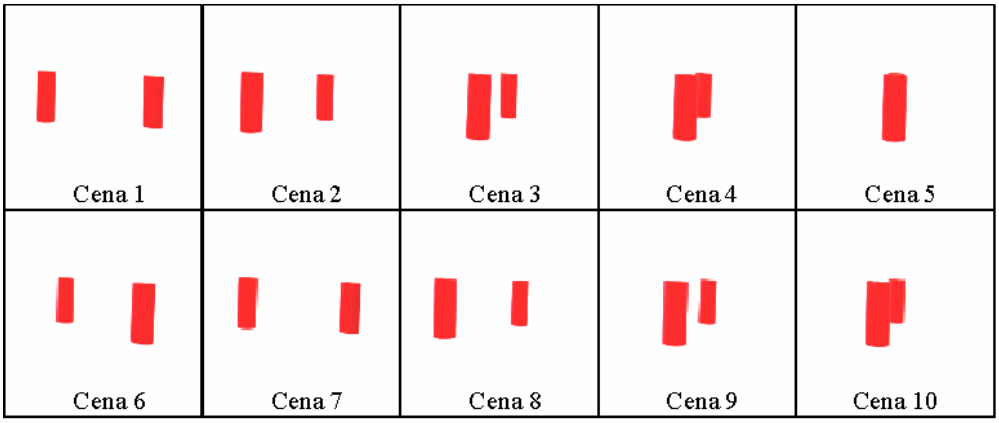


Figura 23 - Dois objetos de mesmo tamanho, mesma cor e os objetos distantes um do outro na cena inicial.

Neste novo teste também utilizamos cinco seqüências de imagens distintas para análise do comportamento do sistema com cenas contendo dois objetos de mesma cor, mesmo tamanho e objetos próximos, figura 24. O comportamento do sistema para este teste foi ruim, pois não interpretou para nenhuma das cinco seqüências e nenhum dos dois movimentos implementados no sistema: rotação e anti-rotação. Neste experimento também existiu uma influência negativa das cenas não serem extraídas em tempos constantes, pois o sistema não identifica os mesmos objetos no decorrer das cenas . Um outro problema encontrado neste experimento é o fato dos objetos estarem inicialmente muito próximos, implicando em uma menor variação de área e posição dos objetos entre as cenas, dificultando, assim, a identificação dos objetos no decorrer nas imagens de toda a seqüência.

Com este resultado temos que, para objetos de mesma cor e mesmo tamanho a distância entre os objetos influencia diretamente na interpretação de seus movimentos.

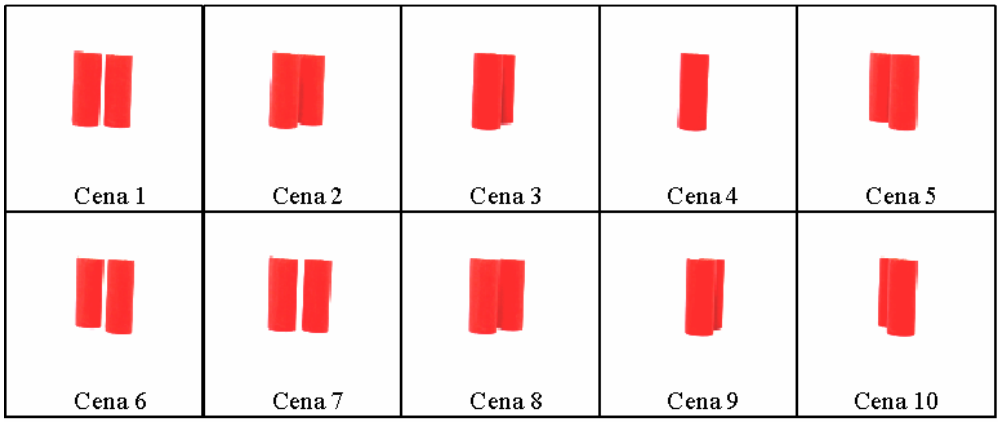


Figura 24 - Objetos de mesmo tamanho, mesma cor e os objetos próximos entre si nas cenas.

Efetuamos um novo teste para investigar se a altura dos objetos de mesma cor influência na interpretação de seus movimentos. Neste experimento utilizamos novas cinco seqüências de imagens contendo objetos de mesma cor e altura diferentes, mesma largura e mesmo formato, e, aqui, o sistema foi capaz de interpretar duas das cinco seqüências.

Este resultado, assim como o resultado do primeiro teste desta segunda etapa, o que influenciou foram as cenas não serem extraídas em tempos constantes, dificultando o sistema identificar os objetos no decorrer das cenas. Este resultado mostra que interpretação de cenas com objetos de mesma cor pouco influencia o tamanho dos objetos, pois o sistema primeiramente identifica os objetos no decorrer das cenas pelas suas cores.

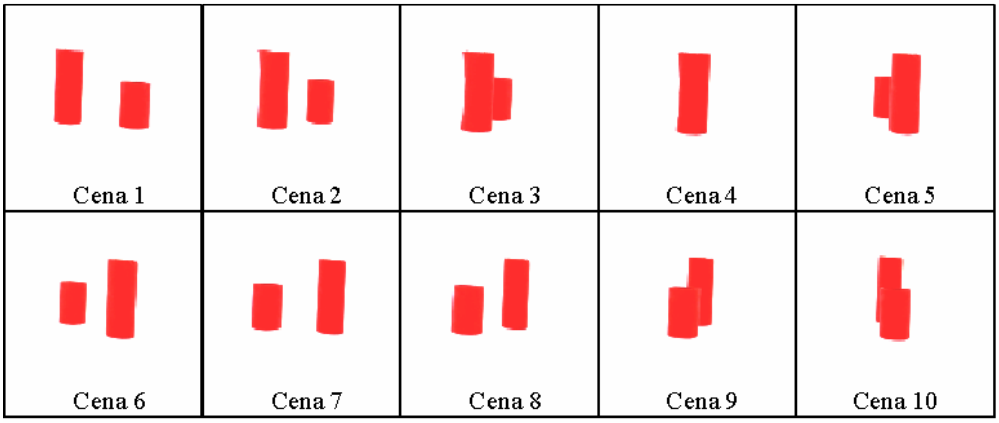


Figura 25 - Objetos de tamanhos diferentes, mesma cor e os objetos distantes um do outro na cena inicial.

Após todos os experimentos da primeira e segunda etapa temos elementos para dizer que para interpretação de seqüências de cenas contendo objetos de cores distintas nada

influencia as cenas serem extraídas em tempo constante, pois o sistema utiliza a cor para a identificação dos objetos no decorrer das cenas de cada seqüência.

Tratando-se de cenas com objetos de mesma cor o que mais influencia para a interpretação dos movimentos é a simetria das transições dos objetos em toda a seqüência, implicando em uma variância não constante de área e posição dos objetos no decorrer das cenas.

Um outro fator que implica na interpretação dos movimentos da seqüência de objetos de mesma cor é a distância entre os objetos, quanto mais próximos os objetos, menor são suas variâncias de áreas e posição, dificultado a identificação dos objetos.

CONCLUSÃO E TRABALHOS FUTUROS

Neste trabalho abordamos o problema de interpretação de seqüências de imagens, sendo que consideramos como sua principal contribuição a apresentação e discussão da implementação de um sistema computacional, baseado em raciocínio espacial qualitativo, capaz de interpretar seqüências de imagens.

Nossa principal motivação para este desenvolvimento do sistema aqui apresentado é interpretar seqüências de imagens, como as demonstradas na figura 1, página 2; executando inferências lógicas sobre mudanças ocorridas nas cenas operando percepções de espaço e como estes espaços podem ser representados com facilidade por inferências espaciais.

O sistema tem a habilidade de extrair informações de objetos de uma seqüência de imagens obtidas por um sensor, baixa abstração, e associar relações entre objetos na cena e mudanças ao longo da seqüência, fornecendo assim, uma interpretação de alto nível, alta abstração, sobre o movimento dos objetos.

Em particular, este trabalho é uma extensão das teorias apresentada nos trabalhos: (SANTOS e SHANAHAN, 2002), o qual descreve a construção de um sistema de raciocínio espacial qualitativo para informações extraídas por um robô móvel; e (SANTOS, P. e SANTOS, M., 2005) que introduz a *T-Logic* que é uma teoria especializada em interpretação de seqüências de imagens. Fundamentado nestas teorias é que desenvolvemos a presente pesquisa e discussão de implementação do sistema capaz de interpretar movimentos de rotação e anti-rotação entre dois objetos em torno de um eixo fixo comum.

O sistema de interpretação de seqüências de imagens desenvolvido neste trabalho é dividido em dois módulos, o Modulo 1: Extração de Informações; e o Módulo 2: Interpretação de Alto Nível.

No Módulo 1, tem-se a entrada das cenas no sistema, estas cenas são fotografias extraídas pelo sensor. O sistema segmenta cada cena por fusão de suas regiões, a implementação do respectivo algoritmo é feita utilizando rotulação de regiões do tipo rotulação por cores (*Blob Coloring*) (BALLARD e BROWN, 1982). Este algoritmo cria um mapa das regiões encontradas na imagem, contendo as informações de todas as regiões e posições dos objetos de cada cena. Estas informações são os dados de entrada do Módulo 2.

A partir das informações extraídas no Módulo 1, o Módulo 2 é responsável por interpretar os movimentos dos objetos na seqüência de imagens apresentada ao sistema. Para isto, utilizamos o formalismo chamado *T-logic* (SANTOS, P. e SANTOS, M., 2005), que é

uma instância da Lógica de Transações (BONNER e KIFER, 1993). Este formalismo de interpretação de seqüências de imagens utiliza dois oráculos baseados em raciocínio espacial qualitativo. O Oráculo de Estados responsável por relacionar as regiões espaciais em conceitos de espaços livres e conectividade dos objetos de cada cena, e Oráculo de Transições que fornece informações de transições dos objetos entre cenas da seqüência. Estas informações são dados de entrada para uma máquina de estados que implementa as regras de inferências para a interpretação dos movimentos entre objetos ao longo de uma seqüência. A máquina de estados fornece a saída do sistema, que é uma descrição conceitual do movimento dos objetos.

O sistema raciocina sobre os objetos das seqüências de imagens através raciocínio espacial qualitativo.

Com base nos testes efetuados neste trabalho, mostramos ser possível a interpretação de seqüências de imagens utilizando raciocínio espacial qualitativo e a *T-Logic* como formalismo lógico, pois de acordo com os resultados apresentados no capítulo 6, chega-se a resultados que apontam que o sistema é capaz de interpretar os movimentos implementados. Tratando-se de cenas com objetos de mesma cor, o que mais influencia para a interpretação dos movimentos é a simetria das transições dos objetos em toda a seqüência, implicando na variância de área e posição dos objetos no decorrer das cenas. Um outro fator que implica na interpretação dos movimentos da seqüência de objetos de mesma cor é a distância entre os objetos: quanto mais próximos, menor são suas variâncias de áreas e posição, dificultando a identificação dos objetos.

Apresentamos, aqui, uma proposta de trabalhos futuros, com a finalidade de modelar e acrescentar a este sistema outros movimentos, criando, assim, uma única máquina de estados capaz de interpretar inúmeros movimentos entre objetos.Propõe-se, também, acrescentar um número maior de objetos interagindo nas cenas. Para isto, faz-se necessário desenvolvimento te uma maneira mais eficiente de referenciar os objetos, já que, atualmente, o sistema referências todos objetos entre si.

Uma proposta adicional, ainda mais interessante para este trabalho: poderia se extrapolar o ambiente dos objetos para o mundo real para e, assim, tornar possível interpretar, por exemplo, movimentos entre pessoas.

Para viabilizar tal proposta, é necessário melhorar o Módulo 1 do sistema, que é responsável por extração de informações em seqüências de imagens. Com isto, podemos utilizar raciocínio espacial qualitativo para interpretar movimentos complexos no mundo real.

REFERÊNCIAS

BALLARD, D. H. e BROWN, C. M. Computer Vision. Englewood Cliffs, Prentice-Hall. ICPR, pages 1149-1158, Barcelona, Spain, 1982

BONNER, A. e KIFER, M. Transaction logic programming, Procedings of the Tenth International Conference on Logic Programming (ICLP), MIT Press, pp. 257-279, 1993

BRAND, M., BIRNBAUM, L. e COOPER, P. Sensible Scenes: Visual Understanding of Complex Strutures through Causal Analysis. National Science Foundation, 1986.

BRAND, M. Physics-base visual understanding. Computer Vision and Image Understanting, 65(2):192-205, 1997.

FORSYTH, D. A. e PONCE, J. Computer Vision: A Modern Approach. New Jersey: Prentice Hall, 2003.

GALTON, A. “Lines of sight”, in Proc. of the Seventh Annual Conference of IA and Cognitive Science, (Dublin, Ireland), pp.103-113, 1994.

HERZOG, G. e WAZINSKI, P. VIsual TRAnslator: Linking perceptions and natural language descriptions. Artificial Intelligence Review, 8(2/3):175-187, 1994

KOJIMA, A., TAMURA, T. e FUKUNAGA, K, Natural Language Description of Human Activites from Video Images Based on Concept Hierarchy of Actions. Kluwer Academic Publishers. Printed in the Netherlands, 2001.

MANN, R., JEPSON, A. e SISKIND, J. M. The Computational Perception of Scene Dynamics, Proceedings of the European Conference on Computer Vision (ECCV), pp. 528-53, 1996.

HUANG, C. L. Contour generation and shape restoration of the straight homogeneous generalized cylinder. IEEE, 1990.

NAGEL, H.-H. From image sequences towards conceptual descriptions. Image and Vision Computing, 6(2):59-74, 1988.

NAGEL, H.-H. Image sequence evaluation: 30 years and still going strong. Em Proc. of ICPR, 2000.

NEUMANN, B e NOVAK, H.-J. NAOS: Ein System zur natürlichen Beschreibung zeitveränderter Szenen. Informatik: Forschung und Entwicklung, 1:83--92, 1988.

RANDELL, D., CUI, Z. e COHN, A. A spatial logic based on regions and connection. 3rd Int.Conf. on Knowledge Representation and Reasoning, Morgan Kaufmann, 1992.

RANDELL, D., WITKOWSIK, M. e SHANAHAN, M. From images to bodies: Modeling and exploiting spatial occlusion and motion parallax. In Proc. of IJCAI, (Seattle, U.S.), pp 57-63, 2001.

SANTOS, E. P. e SANTOS, V. M. A Path Semantics For Image Sequence Interpretation. VII Simpósio Brasileiro de Automação Inteligente, Maranhão, Brazil, 2005.

SANTOS, E. P. e SHANAHAN, M. A logic-based algorithm for image sequence interpretation and anchoring. In Proc. of IJCAI, (Acapulco, México), pp. 14-25, 2002.

SANTOS, P. Reasoning aboult depth and motion from an observer's viewpoint. Spatial Cognition and Computation, vol. 7, no. 2, pp 133-178, 2007.

TSOTSOS, J. K. e SHIBAHARA, T. Knowledge Organization and Its Role in Temporal and Causal Signal Understanding: The ALVEN and CAA Projects. COMPUTER, volume 16, Number 10, October, 1983.