UNIVERSIDADE DO VALE DO RIO DOS SINOS

CIÊNCIAS EXATAS E TECNOLÓGICAS

PROGRAMA INTERDISCIPLINAR DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA

DETECÇÃO DE MOVIMENTOS SUSPEITOS EM SEQÜÊNCIAS DE VÍDEO

por
John Soldera

Dissertação

São Leopoldo, Fevereiro de 2007

UNIVERSIDADE DO VALE DO RIO DOS SINOS

CIÊNCIAS EXATAS E TECNOLÓGICAS

PROGRAMA INTERDISCIPLINAR DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA

LINHA DE PESQUISA EM
COMPUTAÇÃO GRÁFICA E PROCESSAMENTO DE IMAGENS

DISSERTAÇÃO DE MESTRADO

Detecção de Movimentos Suspeitos em Seqüências de Vídeo

Dissertação submetida à avaliação como requisito parcial para a obtenção do grau de mestre em computação aplicada.

Orientada por Dr. Cláudio R. Jung

RESUMO

É proposto neste trabalho novas técnicas para automatizar processos em vigilância eletrônica. A entrada (*input*) do algoritmo descrito são trajetórias das pessoas capturadas de seqüências de vídeo filmadas, as quais são utilizadas para definição de padrões de comportamentos de pedestres. O modelo proposto se baseia em dois critérios para a definição de comportamentos como usuais ou não-usuais: a ocupação espacial e as relações entre as pessoas. O critério de ocupação espacial inclui um determinado tempo de treinamento onde a cena é avaliada para gerar uma base de dados que contabiliza a ocupação espacial em cada região da cena. Através desta base de dados, novas trajetórias são classificadas como usuais ou não-usuais. As trajetórias usuais são aquelas coerentes com o banco de dados gerado pelo treinamento, correspondendo às áreas mais ocupadas, enquanto que as trajetórias não-usuais são aquelas nas quais andaram em regiões de baixa ocupação espacial. O critério das relações interpessoais utiliza Diagramas de Voronoi para avaliar uma série de parâmetros (como distância entre vizinhos, além de outras características psicossociais, como espaço pessoal). Essas características são utilizadas para detectar comportamentos de aproximação, afastamento e agrupamento, que podem ser detectados através de um autômato finito, cuja regra é especificada pelo usuário.

Palavras-Chave: Visão Computacional, Reconhecimento de Padrões, Vigilância Eletrônica, Ocupação Espacial, Trajetória, SOM.

ABSTRACT

It's proposed in this work new techniques to automate processes in electronic surveillance systems. The input of the algorithm described is the trajectories of people captured from real filmed sequences, which are used to define pedestrian behavior patterns. The proposed model is based in two criteria to define behaviors as usual or unusual: spatial occupancy and relations among people. The spatial occupancy criterion includes a certain training period when the scene is evaluated in order to generate a database which accounts for the spatial occupancy in each scene region. Through this database, new trajectories are classified as usual or unusual. Usual trajectories are those coherent with the training database, corresponding to the most occupied areas; whereas unusual trajectories are those that occur in low spatial occupancy regions. The interpersonal relation analysis criterion employs Voronoi Diagrams in order to evaluate a set of parameters (such as distance between neighbors and other psychosocial characteristics as personal space). These characteristics are used to detect grouping, separating and intercepting behaviors that can be detected by a finite automaton, which rules are defined by user.

Key-Words: Computational Vision, Pattern Recognition, Electronic Surveillance, Spatial Occupancy, Trajectory, SOM.

ÍNDICE DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS

SOM – Mapa de Ocupação Espacial (*Spatial Occupancy Map)*

DV – Diagrama de Voronoi

DVD - Diagramas de Voronoi Dinâmicos

TD – Transformada Distância

EPP - Espaço Pessoal Percebido

AFD – Autômato Finito Determinístico

AFND – Autômato Finito Não-Determinístico

SUMÁRIO

1. Introdução

Na atualidade, cada vez mais é necessário utilizar-se de vigilância eletrônica para monitorar áreas de movimentação de pedestres para evitar atos violentos ou assaltos. Uma forma bastante comum de vigilância eletrônica é utilizar circuitos fechados de TV. Dessa forma, um operador de vídeo observa diversas câmeras que monitoram áreas específicas de segurança, como bancos, caixas de lojas, metrôs ou vias públicas.

O operador de vídeo tem a tarefa de acompanhar o movimento de cada pessoa nas câmeras de vigilância em várias áreas filmadas. Essa tarefa pode se tornar fatigante e ineficiente caso existam muitos monitores e o movimento de pessoas seja intenso. Por causa disso e vários outros fatores, podem passar desapercebidos eventos perigosos, como casos de roubo e vandalismo.

Uma forma de tornar essa tarefa mais eficiente (ou até automatizada) é desenvolver tecnologias de vigilância eletrônica onde o computador observa o movimento das pessoas nas áreas filmadas e determina o comportamento de cada pessoa com base nas características da dinâmica de seu movimento, guiado por um modelo que descreve a cena. Quando ocorrer um evento perigoso em alguma câmera, é desejável que o computador dispare um alarme, que chame a atenção do operador de vídeo para a câmera na qual ocorreu o evento. Nesse caso, o sistema identifica as trajetórias das pessoas envolvidas e mostra o tipo do evento ocorrido.

É fato que o olhar humano tem grande facilidade para interpretar o comportamento das pessoas enquanto o vídeo transcorre. A execução dessa tarefa pela máquina não garante a mesma eficiência devido às características subjetivas do comportamento humano tanto das pessoas filmadas, como pela visão de mundo do observador. Por outro lado, conforme o número de câmeras de vigilância aumenta no ambiente, também são acrescidos monitores a serem observados, o que pode tornar inviável a observação humana. Por isso, a área de segurança e vigilância, usando técnicas automáticas e semi-automáticas, está começando a ganhar destaque nas pesquisas ao redor do mundo, com diversas publicações recentes em conferências e periódicos especializados.

Existem muitas aplicações diferentes em visão computacional e segurança eletrônica que são pesquisadas na atualidade, como detecção de rotas de colisão no trânsito e no espaço aéreo, reconhecimento de faces humanas e de formas, interpretação de gestos humanos, determinação do comportamento humano, interpretação das trajetórias humanas em áreas de segurança, entre outras.

No entanto, o principal desafio da área é a devida compreensão do comportamento dos indivíduos numa área filmada a ser realizada de forma automática por meios eletrônicos. Por exemplo, interpretações de mais alto nível (com uma pessoa levantando sutilmente uma arma) podem ser facilmente realizadas por um observador humano atento, mas difíceis de serem feitas automaticamente por sistemas computacionais.

Além disso, a resolução (em *pixels*) das câmeras utilizadas para gravar as seqüências de vídeo, por mais alta que seja, pode oferecer poucos detalhes das pessoas em movimento, visto que a cena filmada pode ocupar uma grande área. Conforme a distância das pessoas à câmera aumenta, sua projeção em coordenadas de imagem diminui, gerando *pixels* não necessariamente contínuos em tonalidade e forma. Entretanto, o olhar humano tem o poder de observar a cena e interpretar esse mesmo conjunto de *pixels* mais facilmente, e mesmo em condições precárias de visualização, consegue determinar a forma e tipo de movimento de cada pessoa na cena. Isso ocorre, em grande parte, por que abstraímos as limitações do vídeo, processando os eventos em alto nível de acordo com nossa visão de mundo.

No processamento de seqüências de vídeo através de visão computacional, vários processos de mais baixo nível são necessários para a interpretação da cena. Em geral, é necessário extrair informações individuais de cada pessoa (como posição ao longo do tempo, forma, posição de cada parte do corpo, etc.) para então extrair alguma informação semântica mais relevante (como interações entre pessoas, movimentos suspeitos, etc.). Em particular, câmeras em movimento, variação de iluminação, sombras e ambientes densos são condições que atrapalham a correta identificação das pessoas na cena, e são objeto de estudo por vários pesquisadores envolvidos com o acompanhamento de objetos.

Entretanto, o foco deste trabalho está voltado para a detecção dos eventos de mais alto nível, que tomam como base os resultados do acompanhamento de pessoas, em vista, que há uma separação entre o acompanhamento de pessoas (determinação da posição exata de cada pessoa ao longo do tempo) do estudo do comportamento das pessoas, onde eventos suspeitos são detectados.

1.1 O Problema

O termo evento suspeito pode sugerir diversos tipos distintos de iterações entre pessoas, como por exemplo, uma pessoa que está trapaceando num jogo de cartas, uma pessoa que deixa

uma sacola no chão, alguém que está retirando uma faca do bolso, uma pessoa que está subindo numa árvore, ou ainda um indivíduo com comportamento de interceptação a uma outra pessoa. De um modo geral, um evento suspeito é aquele que pode ser interpretado como nocivo. Como a classificação de um evento como suspeito é bastante subjetiva, será utilizado neste trabalho o termo "evento não-usual", que é aquele diferente de um conjunto de eventos encontrado em uma base de dados considerada "normal".

Em particular, uma sub-classe de eventos não-usuais está relacionada ao movimento das pessoas em uma cena. Nesse contexto, um possível fator a analisar é o histórico de movimento das regiões da cena por onde as pessoas passam. Pode ser considerada não-usual uma trajetória que passou por uma região pouco utilizada pelas pessoas, como arredores de uma janela ou uma parede. Normalmente, em calçadas ou corredores, os pedestres andam em áreas bem definidas, que em geral são aquelas que geram uma menor distância percorrida. Portanto, um histórico de movimentação na cena pode ser utilizado para classificar as trajetórias dos pedestres.

Outro fator a ser considerado são as relações interpessoais, onde, através do movimento das pessoas em pares ou grupos, o seu comportamento entre elas é identificado. O comportamento interpessoal pode ser determinado através da avaliação do movimento relativo entre pares de pessoas ou grupos. Por exemplo, determinar se a trajetória de uma pessoa é de interceptação em relação à outra. Um possível caso de risco pode ser considerado quando uma trajetória intercepta outra e depois se mantêm juntas, possivelmente caracterizando um roubo ou seqüestro.

Um terceiro fator a ser considerado é a detecção de movimentos em direção contrária à grande maioria das pessoas. Por exemplo, numa área de saída de um estádio, pode haver uma pessoa caminhando em sentido contrário à multidão, evento digno de atenção. Além disso, pode-se analisar diversos outros fatores, como a postura das pessoas (em pé, deitado, etc.), se estão carregando objetos, entre outras possibilidades. O foco principal deste trabalho será definido a seguir.

1.2 Objetivos

O objetivo principal deste trabalho é desenvolver métodos de detecção de trajetórias não-usuais em uma determinada cena com base na análise do movimento das pessoas filmadas e de um histórico de movimentação. Pretende-se basear a detecção de trajetórias não-usuais em

dois aspectos fundamentais: a análise das trajetórias pelo critério da ocupação espacial da cena e a avaliação das relações interpessoais entre as pessoas filmadas.

Ocupação espacial: o primeiro aspecto a ser analisado para determinar se as trajetórias das pessoas que estão passando na área filmada são usuais ou não tem como base a ocupação espacial esperada do cenário filmado. Tal ocupação espacial deve ser calculada com base em um determinado período de treinamento, que pode ser feito através da avaliação de uma seqüência de vídeo da área filmada em um certo intervalo de tempo. Essa análise gera uma base de dados que mantém um histórico da quantidade de movimento que normalmente há em cada parte da cena filmada. Nesse caso, a trajetória não-usual é aquela que foi desenvolvida numa região, onde, historicamente, houve pouco movimento, como arredores de uma janela ou de uma parede.

No trabalho de Jacques Jr. [JAC 2006b], é proposto o conceito de medição da ocupação espacial que denota o histórico de ocupação do espaço na cena, sendo usado como um critério para verificar se simulações de humanos virtuais condizem com os humanos reais numa mesma cena. Nessa dissertação, as contribuições apresentadas sobre [JAC 2006b], são o uso da ocupação espacial para fins de segurança, além de propor aperfeiçoamentos sobre os conceitos apresentados que serão vistos no modelo proposto.

Relações interpessoais: o segundo aspecto a ser analisado neste trabalho para detecção de movimentos não-usuais se baseia nas relações interpessoais entre as pessoas da cena. Em particular, o conceito de agrupamento será fortemente explorado neste trabalho, dando continuidade ao trabalho de Jacques Jr. [JAC 2006a], onde diagramas de Voronoi [SOI 2002] são empregados para calcular a distância entre cada pessoa e suas vizinhas ao longo da seqüência de vídeo, sendo usado como base para detectar grupos de pessoas na cena. Nessa dissertação, esses fatores comportamentais e outros fatores novos são integrados na classificação de trajetórias suspeitas.

Finalmente, uma outra contribuição apresentada nessa dissertação é o desenvolvido de um autômato finito para detectar ocorrências concomitantes e seqüenciais de diversos eventos (por exemplo, uma pessoa se aproxima da outra por trás e ambas mantém-se em agrupamento). As regras utilizadas no autômato podem ser customizadas para detecção das combinações de eventos desejadas pelo usuário, combinando aspectos da ocupação espacial com relações interpessoais (por exemplo, agrupamento seguido de afastamento entre as pessoas em uma região de ocupação espacial inválida).

O restante deste trabalho está organizado da seguinte maneira. O Capítulo seguinte apresenta alguns trabalhos importantes na detecção de eventos não-usuais, com foco em técnicas que utilizam os conceitos de ocupação espacial e interações entre as pessoas. Também são apresentadas algumas técnicas para o acompanhamento de pessoas em seqüências de vídeo, já que as trajetórias das pessoas são os dados de entrada para a técnicas proposta. O método proposto para detecção de movimentos não-usuais é descrito no Capítulo 3, incluindo a avaliação da ocupação espacial, das interações entre as pessoas, e a combinação dessas duas características através do autômato finito. Alguns resultados experimentais são mostrados e discutidos no Capítulo 4, e as conclusões são apresentadas no Capítulo 5.

2. Revisão Bibliográfica

Na literatura, existem diversos trabalhos recentes acerca da detecção do comportamento de pessoas, automóveis, robôs, aviões e outros em seqüências de vídeo, para fins de segurança. Nesse trabalho, um foco maior será dado à classificação das trajetórias das pessoas, no modo onde a vigilância da cena é feita por apenas uma câmera e a mesma é estática.

Dentre esses trabalhos, tem-se o trabalho de Perera et al. [PER 2006], onde é apresentado um modelo para determinar e manter a identidade de veículos em rodovias mesmo após longos períodos de oclusão ocasionados em geral por pontes ou árvores. Através de um algoritmo de acompanhamento de objetos robusto, a trajetória de cada veículo é acompanhada desde o início da cena até a sua saída da área filmada, mesmo se o veículo sofresse períodos de oclusão no percurso.

Na figura 2.1, têm-se várias trajetórias analisadas numa área de intersecção de rodovias (cada trajetória distinta foi identificada com uma cor diferente). Mesmo após períodos de oclusão causados pela sobreposição das rodovias, cada trajetória manteve a sua cor, denotando a sua identidade.

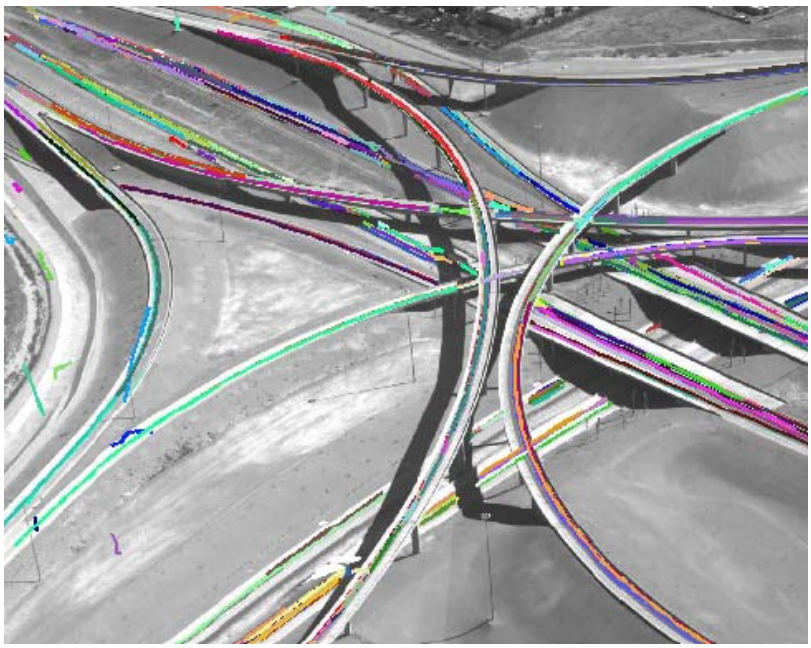

Figura 2.1 – Cada trajetória identificada é colorida com uma cor distinta.

No trabalho de Stauffer e Grimson [STA 2000], a cena é observada durante um certo período de tempo, executando um treinamento que tem a função de preencher uma base de dados hierárquica com trajetórias dos objetos, formas e dados relativos ao seu movimento. Após o acompanhamento de objetos, as trajetórias obtidas com características semelhantes são agrupadas para formar padrões na matriz de co-ocorrências através de sua forma, posição e velocidade.

Como a subtração de *background* é baseada em uma mistura de Gaussianas para cada *pixel* da cena, pequenas oscilações locais (como folhas balançando ao vento) são ignoradas. A cada quadro, as gaussianas de um mesmo *pixel* são atualizadas para determinar qual o *background* mais provável naquele instante. As gaussianas com maior probabilidade indicam o histórico dos *foregrounds*.

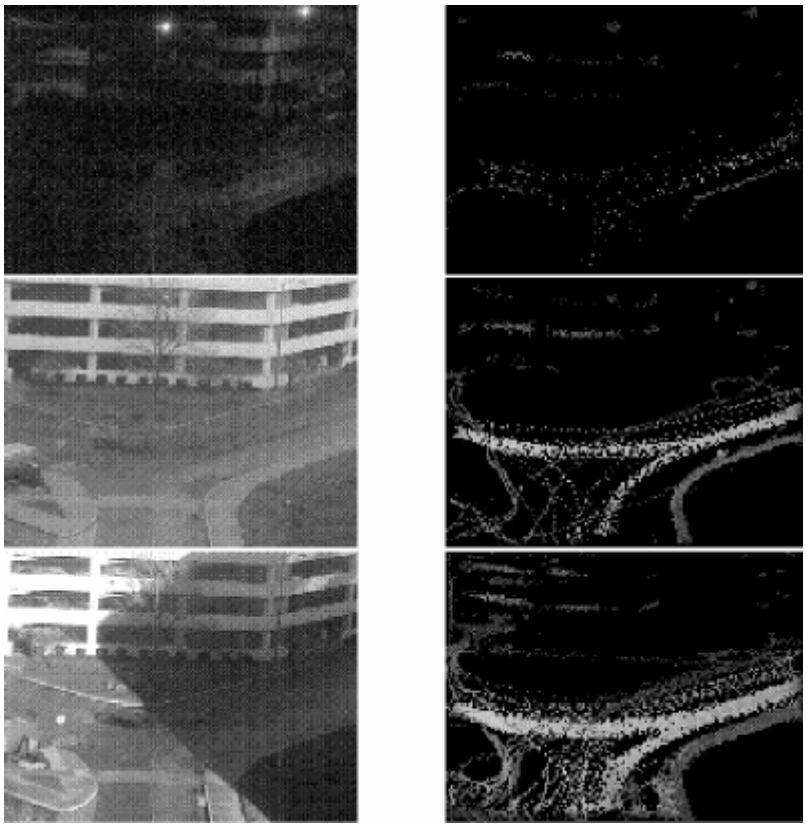

Figura 2.2 – À esquerda tem-se uma imagem da cena em cada período do dia e à direita, tem-se uma representação da base de dados no instante da foto. Á direita, pode-se ver os caminhos mais comumente usados pelos móveis e quanto mais claro, mais usado o caminho.

Novos objetos são classificados de acordo com os outros objetos gravados na base de dados através de sua forma e características do seu movimento. Além de detectar e classificar os objetos envolvidos na cena como carro ou pedestre, o modelo é capaz de detectar trajetórias não-usuais que são as trajetórias que não se enquadram na base de dados, por serem trajetórias cuja forma não foi vista ainda na cena filmada ou foi raramente desenvolvida. E ainda outras características como velocidade incompatível com as outras trajetórias de mesma forma podem determinar trajetórias atípicas.

Nos trabalhos apresentados a seguir, o objetivo principal é procurar por movimentos suspeitos (trajetórias não-usuais), com base em um modelo matemático no qual uma determinada base de dados é preenchida durante um período de treinamento. Além disso, alguns trabalhos necessitam de uma descrição física prévia da cena, ou seja, uma descrição do relevo da cena ou do tipo de movimento feito em cada parte da cena.

Os trabalhos revisados brevemente a seguir, apresentam técnicas distintas que analisam trajetórias de acordo com os padrões das trajetórias anteriores obtidas na cena. Outros trabalhos citados visam determinar o comportamento das pessoas, classificando as relações interpessoais de suas trajetórias. O capítulo encerra com uma breve revisão sobre algoritmos de acompanhamento de objetos, necessários para extração automática de trajetórias.

2.1. Análise do Comportamento de Pessoas em relação às Trajetórias

Muitos trabalhos desenvolvidos nessa área de pesquisa avaliam o comportamento das pessoas a partir de uma base de dados criada durante certo período de treinamento. As estruturas de dados internas de cada um desses modelos são alimentadas durante o treinamento, de acordo com a lógica do próprio modelo, gerando uma base de dados que permite determinar o comportamento das pessoas utilizando os padrões de suas trajetórias.

O método desenvolvido no trabalho de Junejo [JUN 2004] também necessita de um período de treinamento. No treinamento, a cena é filmada para que sejam capturadas diversas trajetórias. As trajetórias similares são agrupadas para formar as arestas de um grafo que indica os caminhos mais comuns usados pelas pessoas. Além disso, cada aresta desse grafo abrange uma área ao seu redor chamada de “envelope de caminhos”, que indica a extensão da área de movimento ao redor daquele caminho, como mostrado na figura 2.3.

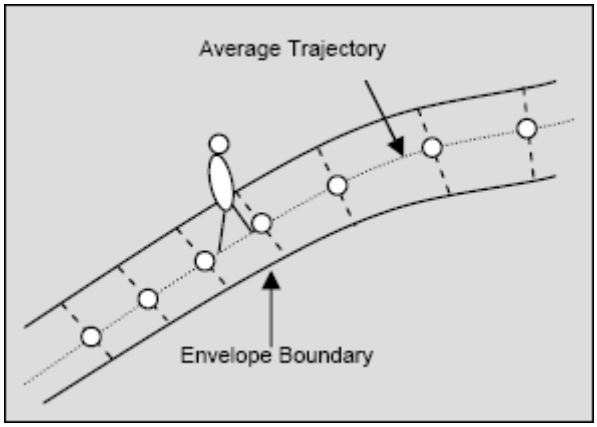

Figura 2.3 – Envelope de Caminhos

Nesse método, a avaliação de novas trajetórias ocorre hierarquicamente segundo três critérios: as características espaciais, velocidade e curvatura. Primeiramente, são avaliadas as características espaciais da trajetória. Nesse critério, as trajetórias inconformes são aquelas com 90% dos pontos fora do envelope de caminhos correspondente, sendo feito um teste com a distância de Hausdorff. Se, com essa avaliação, a trajetória não foi considerada inconforme, é avaliada a velocidade da trajetória usando-se o teste de Mahalanobis. No último teste, é calculada a curvatura da trajetória, que determina se esta é reta ou bastante irregular. Portanto, se a trajetória falhar em algum teste, ela é considerada não-usual.

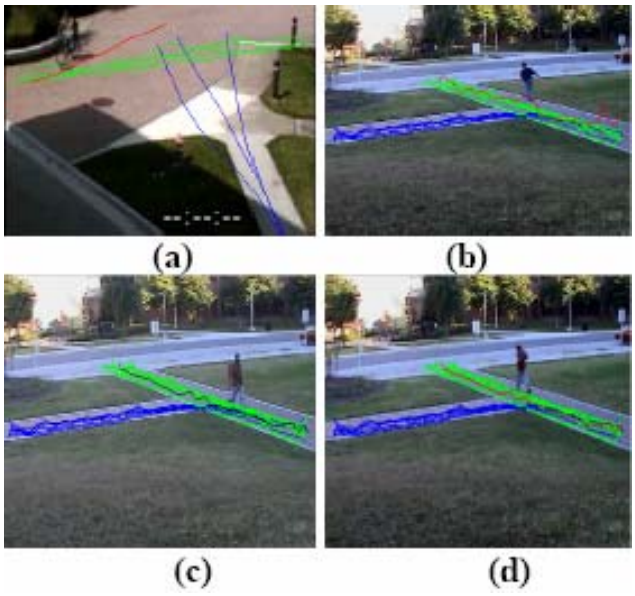

Figura 2.4 – Trajetórias e Envelopes de Caminhos

Na figura 2.4, as trajetórias com velocidade semelhante são desenhadas com a mesma cor, porém a cor vermelha determina as trajetórias não-usuais. Na parte (a) dessa figura, a trajetória em vermelho é a trajetória de um ciclista que está fora do envelope de caminhos criado naquela área pelas trajetórias em verde. Na parte (b), foi detectada uma pessoa bêbada desenvolvendo uma trajetória com curvatura anormal (a sua trajetória aparece em vermelho na borda). Na parte (c), foi detectada uma trajetória normal de uma pessoa caminhando pela calçada. E na parte (d) foi detectada uma pessoa correndo numa área com velocidade incompatível com a velocidade gravada no envelope de caminhos.

Uma abordagem semelhante é apresentada no trabalho de Makris e Ellis [MAK 2005], que utiliza um modelo físico da cena que pode ser informado pelo usuário ou aprendido pelo modelo durante um certo período de treinamento. O modelo descreve semanticamente o movimento em cada parte da cena, classificando-as como entradas, junções, caminhos, rotas, ou pontos de parada, e determinando os "envelopes de caminhos", como na figura 2.5. Essas características ajudam a classificar as trajetórias, pois espera-se que a estrutura da cena afete o comportamento das pessoas em suas trajetórias.

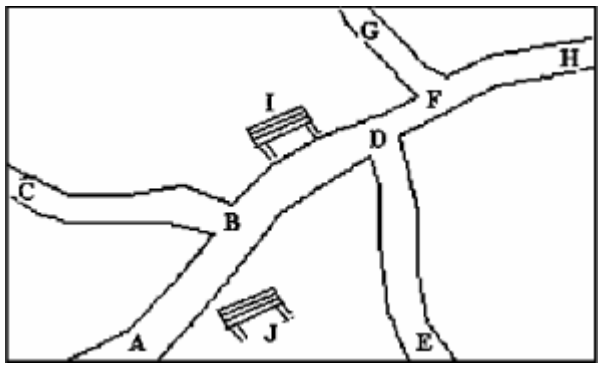

Figura 2.5 – Exemplo de Modelo: Entradas: A, C, E, G, H
Junções: B, D, F
Caminhos: AB, CB, BD, DE, FH
Rotas: ABDFH, EDFG, CBDE
Zonas de Parada: I, J

Ainda sobre o trabalho de Makris et al., pode-se observar que, no tráfego de veículos, os carros andam num envelope de caminhos fixo; entretanto, nas áreas de caminhamento de pedestres, o mesmo padrão é percebido, mesmo não havendo limites precisos para determinar os envelopes de caminhos. Na figura 2.6, tem-se um exemplo de envelope de caminhos que foi reconhecido durante o treinamento.

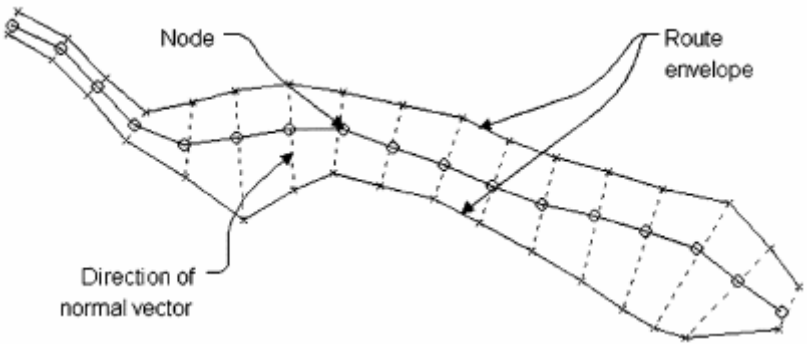

Figura 2.6 – Envelope de Caminhos

No treinamento, são usados dois parâmetros: o fator de amostra, que define a distância entre dois nodos consecutivos num mesmo caminho, e a distância mínima permitida entre duas rotas. O fator de amostra define a resolução do envelope de caminhos enquanto que a distância mínima permitida determina quando dois envelopes de caminhos distintos devem ser combinados para formar apenas um, devido à sua proximidade. Os outros elementos do modelo também são detectados pelo treinamento através da observação da forma de movimento que ocorre onde eles estão localizados.

Depois de feito o treinamento, para a avaliação de novas trajetórias, é utilizada uma cadeia de Markov, que permite avaliar os caminhos pelos quais cada trajetória pode passar. Em outras palavras, ela permite calcular a probabilidade de uma trajetória prosseguir por cada caminho disponível à frente. Através desse modelo, as novas trajetórias são classificadas como típicas ou atípicas. Na figura 2.7, têm-se os envelopes de caminhos determinados pelo treinamento.

Figura 2.7 Envelopes de caminhos obtidos pelo treinamento.

No trabalho de Niu et al. [NIU 2004] é apresentada uma abordagem distinta, no qual a mesma cena é filmada com várias câmeras em ângulos diferentes, sendo a cada instante da trajetória, formado um vetor com a posição da mesma pessoa em cada câmera. Primeiramente, o *tracking* das pessoas é feito através de uma implementação de algoritmos do paradigma de hipótese e verificação (filtros de Monte Carlo, filtro de partículas, algoritmos genéticos, propagação condicional de densidade e condensação baseada na importância). A vantagem desses algoritmos sobre outros, como p.ex. filtros de Kalman, é que não exigem distribuições gaussianas para remover ruído e a propagação do estado não precisa ser unimodal, ou seja, são permitidas múltiplas hipóteses de *background* e de *foreground*, cada uma com uma probabilidade associada de ser a mais condizente com o momento atual. Um "estimador de Bayes" permite calcular essas probabilidades.

Cada quadro pode ser interpretado como um estado (instante) de um *blob* (pessoa) se locomovendo pela área filmada. Para cada estado do *blob*, são avaliadas sua posição, velocidade e aceleração instantâneas. Para cada pessoa em movimento, sua posição é prevista de acordo com a sua velocidade e aceleração anteriores. Através disso, o problema da oclusão é reduzido, pelo fato que pode-se prever as trajetórias oclusas utilizando-se de informações dos estados anteriores e com informações derivadas de outras câmeras quando o objeto está ocluso na câmera, além das informações provenientes do modelo de Markov.

Para reconhecer o comportamento individual ou de grupo, ao invés de ser feito o parsing de cada vetor de estados nas cadeias de Markov, são utilizadas as propriedades estatísticas das cadeias de Markov, que permitem distinguir entre diversos tipos de comportamento: por exemplo, para comportamento entre duas pessoas, tem-se várias situações, como: "Comportamento de Perseguição": posição relativa quase constante e velocidade relativa nula. "Comportamento de Colisão": variação linear da distância entre os dois objetos e com velocidade relativa constante, mas não-nula. E o tipo de comportamento no qual há grande variação da posição e velocidade relativas indica a não-correspondência entre duas trajetórias.

Em conclusão, nesse modelo, o comportamento é determinado pela análise da diferenciação e correlação dos *pixels* entre trajetórias de pessoas distintas que estão no modelo estatístico do programa. Utilizando-se dados de treinamentos, o programa desenvolvido teve um bom desempenho em classificar os tipos de comportamentos. O programa teve 100% de sucesso em detectar "Comportamentos de Perseguição" e 80% para "Comportamentos de Colisão". Na figura 2.8 são exibidas diversas trajetórias, sendo que a primeira figura (parte superior), indica duas trajetórias paralelas. A segunda figura exemplifica o caso de trajetórias

de colisão, onde duas pessoas se cruzam. E na última figura, na parte inferior, são mostradas duas trajetórias que não tem relação entre si.

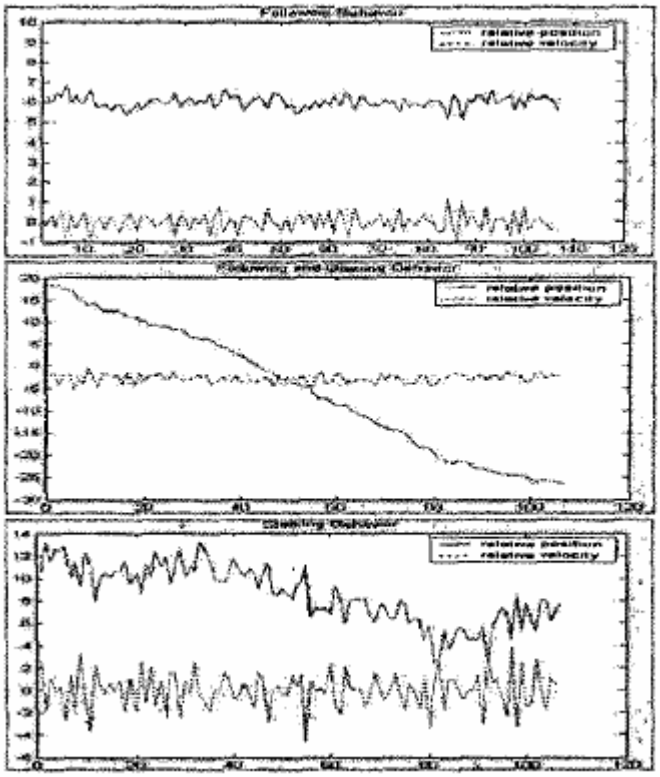

Figura 2.8 – Exemplos de Trajetórias

No trabalho de Cupillard et al. [CUP 2004], várias câmeras são utilizadas sobre a mesma cena para monitorar as atividades ocorridas dentro de estações do metrô. Primeiramente, as áreas de interesse são informadas ao sistema, para a determinação do comportamento e o modelo tridimensional da cena para que o sistema esteja apto a juntar a visão de cada câmera num único modelo tridimensional da cena. Nesse modelo é estimada a forma tridimensional de cada objeto.

Através de comparações com outros objetos gravados na base de dados, o sistema apresentado por Cupillard et al. está apto a classificar novos objetos a partir de sua forma. Os tipos de objetos possíveis são: pessoa, pessoa oclusa, grupo, multidão, metrô, objeto da cena, ruído ou objeto desconhecido. Essa classificação é feita quadro à quadro e permite o sistema manter um grafo dos objetos móveis da cena obtido pela união do grafo de objetos de cada câmera. O sistema usa um algoritmo de acompanhamento de objetos (*Tracker*) para cada tipo diferente de atores, como pessoas sozinhas, grupos ou multidões. O *Tracker* de pessoas sozinhas e o *Tracker* de grupos trabalham em conjunto, monitorando cada pessoa e

contabilizando as pessoas em cada grupo, até que a cena fique superlotada com 2/3 da área visível sobreposta por objetos em movimento. A partir disso, o *Tracker* de multidões é acionado, pois somente ele terá precisão nessa situação, já que é muito difícil separar as pessoas umas das outras devido ao excesso de oclusões ocorrido na cena.

O objetivo do trabalho desenvolvido por Cupillard et al. é determinar certos comportamentos específicos que estejam ocorrendo no metrô. Para tal, foi definido um formalismo que permite escrever e reutilizar todos os métodos necessários para a descrição e o reconhecimento de comportamentos. Esse formalismo é flexível e declarativo, permitindo expressar as conclusões das avaliações sobre as trajetórias, bem como as condições para uma trajetória ser suspeita.

No reconhecimento do comportamento, o modelo confere aos atores da cena várias características, como velocidade, forma e trajetória desenvolvida. Além disso, um ator (pessoa, grupo ou multidão) pode ter um determinado estado, por exemplo, estado de agitado, onde há grande variação na forma. Eventos são situações que ocorrem com os atores, como por exemplo, um grupo entrando numa zona de interesse, como a máquina de tickets. Um cenário é tudo o que está ocorrendo na cena em um determinado momento, contendo estados, eventos e sub-cenários. Comportamentos são seqüências de cenários específicos, ou seja, uma seqüência de eventos especificados através de uma expressão formal. Na aplicação do trem, há os seguintes comportamentos possíveis: trapaça, luta, bloqueio de caminho, vandalismo e ambiente superpopulado. Na figura 2.9, tem-se um exemplo de comportamento onde algumas pessoas posicionam-se na saída da máquina de tickets, bloqueando-a.

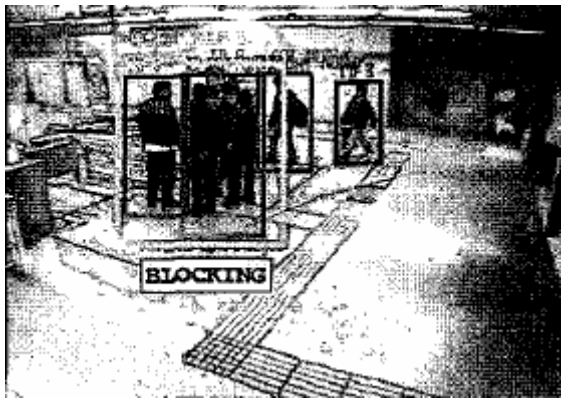

Figura 2.9 – Comportamento de “bloqueio de caminho”.

Durante o processamento do vídeo, para reconhecer os comportamentos procurados, é avaliado se cada cenário procurado está ativo ou não no sistema, e quando um cenário passar a estar ativo, a busca passa para o cenário seguinte definido no formalismo do comportamento, em seqüência, até se encaixar em algum comportamento, ou seja, sendo reconhecendo como sentença do formalismo. O método desenvolvido usa os mesmos formalismos para descrever a cena para o usuário e tem obtido sucesso para reconhecer os comportamentos pré-definidos dentro do metrô, mas ainda perde um pouco para o resultado obtido por um observador humano, segundo as conclusões apresentadas em seu artigo.

Em Fung e Jerrat [FUN 2000], a diferença entre quadros consecutivos é calculada em blocos durante a separação do *background* do *foreground*. Se a média das diferenças entre dois blocos de quadros consecutivos for maior que o limiar, é considerado que há atividade no bloco corrente. Os blocos ativos que são vizinhos entre si são agrupados para formar *blobs*, que correspondem aos objetos em movimento na cena. No processo, geralmente são usados blocos de 5 x 5 *pixels*, mas podem ser usados também blocos do tamanho de um *pixel* para cálculos com maior precisão (com maior tempo de processamento). Os blocos muito oscilatórios, como blocos de árvores ou grama, são reconhecidos e eliminados, gerando um possível problema no qual um intruso poderia não ser detectado se caminhar nessa região. Depois de removido o *background* oscilatório, são utilizadas várias técnicas para detectar o comportamento suspeito. A primeira delas é avaliar se a área do *blob* varia muito, ou seja, se a quantidade de pixels do *blob* tem grande variação no decorrer da trajetória. Cada *pixel* da câmera pode ter um fator de peso que corresponde à distância da câmera à área filmada por esse *pixel*. Esse peso soluciona o problema de trajetórias que se afastam da câmera, causando a diminuição gradual do tamanho do *blob*.

Em outra parte do algoritmo, é verificado se a intensidade de cor dos *pixels* do *blob* varia muito. Se a média das diferenças de cor de cada *pixel* de *blobs* entre quadros consecutivos for maior que um limiar, é considerado movimento suspeito. Em outro teste, é detectado se houve grande variação na forma do objeto através do teste de Freer, que utiliza uma relação entre volume e perímetro. Se o valor obtido no teste for maior que um limiar específico, é considerado suspeito devido à anormalidade das características físicas da pessoa projetada na câmera.

É feito também um teste nas trajetórias detectadas utilizando uma rede neural. As trajetórias são avaliadas quanto à sucessão de zonas pelas quais a pessoa se deslocou. Cada zona tem um grau de suspeitabilidade diferente, que vai afetar o resultado da avaliação. No início do processo, a rede neural necessita um período de treinamento, como nos outros

métodos apresentados. A rede neural é capaz de detectar tipos de comportamentos pré-definidos classificados como suspeitos.

No trabalho de Lou et al. [LOU 2002], durante o treinamento, as trajetórias semelhantes em velocidade, forma e localização física são agrupadas para formar *clusters*. As trajetórias-modelo de cada *cluster* são mantidas numa rede Bayesiana para determinar a quais *clusters* as novas trajetórias filmadas pertencem. Se uma nova trajetória tem probabilidade insuficiente de pertencer a qualquer *cluster*, ela é considerada suspeita, caso contrário ela é aceita por algum *cluster* e o modelo é atualizado. Além disso, são usados formalismos para descrever as trajetórias em linguagem natural, obtendo resultados, como por exemplo: "a trajetória X está entrando numa nova região".

2.2 Análise do Comportamento com Base nas Interações Interpessoais

No trabalho de Hosie et al. [HOS 98], é analisado o comportamento de grupos de pessoas. É determinada a relação do movimento de cada pessoa com cada outra, por exemplo, uma pessoa seguindo outra, ou ambos parados ou pessoas se cruzando. Esses relacionamentos são chamados de Pares e são fortemente determinados pela análise da direção do movimento e pouco pela análise da sua velocidade.

No decorrer do vídeo, o comportamento do grupo pode variar. O sistema suporta consultas com base nesse histórico. Por exemplo, o usuário pode fazer a seguinte consulta: "Procure por um par de pessoas que no inicio uma delas caminhou em direção à outra e depois passaram a andarem juntas". Dessa forma, formalismos podem ajudar a definir com exatidão comportamentos mais complexos em função de comportamentos mais simples, como comportamentos de pares.

Para evitar o alto custo de processamento em cenas muito populadas, a cena é dividida em partes (*clusters*). São calculados os relacionamentos de grupos dentro de cada *cluster* e também com os *clusters* vizinhos. De acordo com as regras estabelecidas para formar grupos, os grupos podem ser classificados como convergentes, divergentes, seguindo, em paralelo ou estacionário. Outros membros com o mesmo comportamento podem se juntar ao grupo depois de um tempo do grupo ter sido detectado. O foco desse trabalho é determinar o comportamento intergrupos e não avaliar trajetórias suspeitas.

No trabalho de Jorge et al. [JOR 2004], as áreas em movimento são detectadas por um algoritmo de acompanhamento de objetos, e após isso, cada *blob* é isolado e rotulado. Os rótulos são determinados por uma Rede Bayesiana, que é usada para representar melhor a iteração entre objetos e para diminuir a oclusão. Portanto, cada pessoa em movimento é classificada por um rótulo, e a rede Bayesiana mantém o mesmo rótulo na pessoa, mesmo em momentos de oclusão.

Quando o *blob* se divide em dois, é determinado que havia um grupo de pessoas desde que esse *blob* entrou na cena e novos rótulos são designados para cada integrante. Outros tipos de formação de grupos podem ser detectados, como pessoas que se encontram e passam a andar juntas e depois se separam novamente, caracterizando eventos de "formação de grupo" e "separação de grupo". Na figura 2.10, têm-se exemplos de trajetórias rotuladas, e na figura 2.11, tem-se um exemplo de formação de grupo por um determinado tempo.

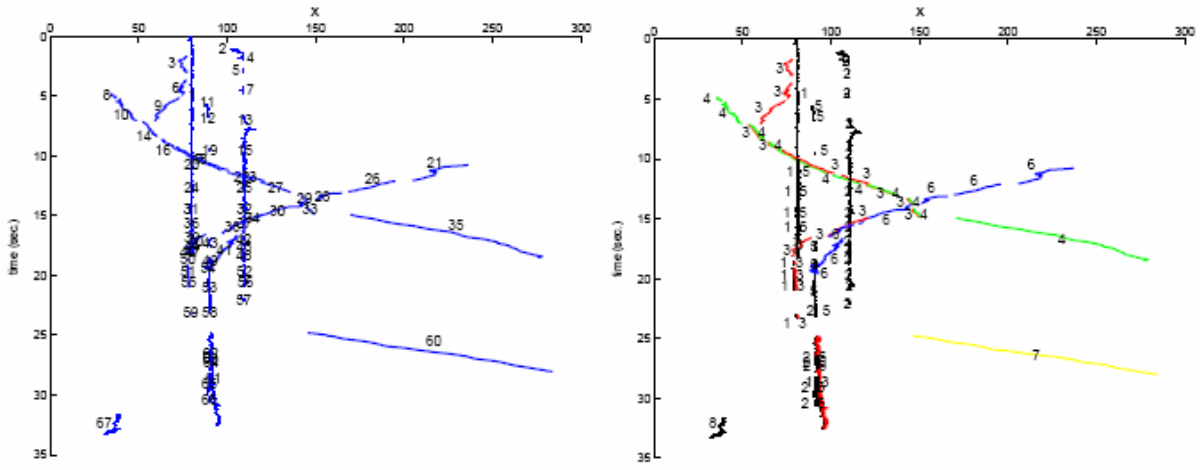

Figura 2.10 – Rotulamento de Trajetórias. Na Figura da esquerda, têm-se trajetórias não rotuladas em azul. Na figura da direita as trajetórias são rotuladas; cada número representa uma trajetória distinta, que é pintada com uma cor diferente.

Nesse modelo, é usado um tempo de treinamento para preencher a rede de Bayes com todas as trajetórias que são lidas. Depois de feito o treinamento, a rede pode crescer muito, causando um elevado tempo de processamento para definir o caminho na rede que mais se encaixa com uma nova trajetória. Portanto, após o treinamento são usadas algumas técnicas para reduzir o tamanho da rede, como remover os nodos de trajetórias muito antigas a fim de manter um total fixo de nodos na rede, evitando a complexidade exponencial do método utilizado.

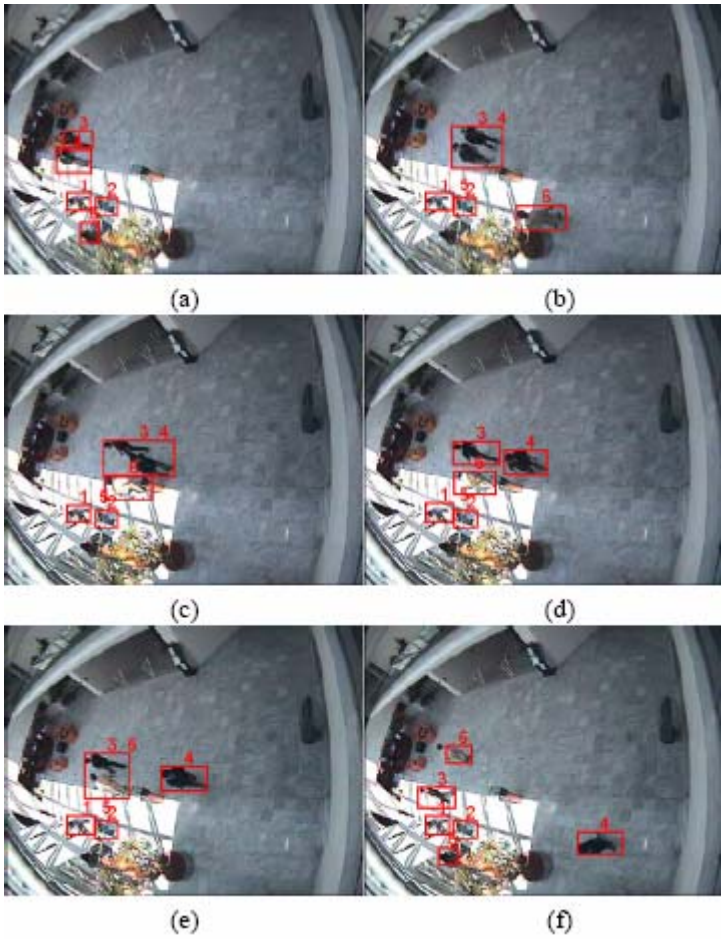

Figura 2.11 – Exemplo de formação de grupo, onde duas pessoas iniciam a seqüência de vídeo separadas, passam a andar juntas e depois se separam.

Através dessas técnicas o modelo está apto a detectar a formação e a separação de grupos, suportando oclusão de membros do grupo. No trabalho de Jorge, o foco não está em detectar se determinadas trajetórias são suspeitas, mas sim determinar quando grupos ocorrem através de uma abordagem probabilística.

No trabalho desenvolvido por Du et al. [DU 2006], é apresentado um modelo para reconhecer iterações entre as pessoas. Para tal, as características de cada pessoa são determinadas em dois conjuntos: características globais e locais. As características globais descrevem as características do próprio movimento da pessoa e as características do movimento entre pessoas, como velocidade, direção, distância e ângulo entre duas pessoas. As características locais descrevem características individuais como contorno, movimento, ângulo do tronco, largura, comprimento do objeto e outros.

Figura 2.12 – Exemplo de aplicação do *tracker* definido no modelo. Cada indivíduo é reconhecido, rotulado e suas características são determinadas.

A modelagem das atividades de iteração é feita através de Redes de Bayes, as quais tem demonstrado alta performance na fusão das informações. O modelo necessita de um tempo de treinamento para alimentar a Rede de Bayes, a qual é criada utilizando-se de parâmetros estimados inicialmente. Novas trajetórias são avaliadas pela rede de bayes utilizando uma função probabilística.

Diversos experimentos foram realizados para detectar um determinado conjunto de comportamentos, os quais são:

- Duas pessoas andam no mesmo caminho com distancia relativa constante.
- Duas pessoas andam no mesmo caminho em direções opostas.
- Duas pessoas correm no mesmo caminho em direções opostas.
- Duas pessoas andam no mesmo caminho, se encontram, e depois seguem em direções opostas.
- Duas pessoas andam no mesmo caminho em direções opostas, uma deixa um objeto no chão e a outra pega o objeto e vai embora.

No modelo de Du et al., os comportamentos procurados são fixos, ou seja, eles são reproduzidos para treinamento da Rede de Bayes e depois os mesmo são avaliados pela rede. Portanto, o usuário não pode especificar um novo comportamento para ser procurado posteriormente.

Oliver et al. [OLI 2000] apresenta um trabalho no qual Cadeias de Markov (*Coupled Hidden Markov Models*) são utilizadas para modelar e reconhecer tarefas humanas. O sistema apresentado por Oliver necessita de um tempo de treinamento supervisionado para reconhecer comportamentos normais entre pares de pessoas, além de comportamentos individuais. Após

feito o treinamento, novos modelos de comportamentos podem ocorrer na cena, e portanto não seriam reconhecidos. Para resolver tal problema, o sistema foi adaptado para reconhecer e modelar novos comportamentos que ocorrem na cena após o treinamento.

Para resolver o problema da falta de dados de treinamento em certas aplicações, foi desenvolvida uma técnica que usa agentes inteligentes que imitam humanos reais para construir e treinar modelos de comportamentos de interesse. Os dados gerados pelos agentes formam uma base de dados sintética que pode ser acoplada às Cadeias de Markov como forma de realizar um treinamento inicial numa nova cena. Com isso, há mais precisão para detectar iterações entre pessoas em cenas em que há poucos dados reais para treinamento.

Durante o treinamento com agentes sintetizados, diversos tipos de comportamentos interpessoais são simulados e contabilizados, dentre eles, alguns são citados a seguir:

- Uma pessoa que segue outra, alcança-a e, após, andam juntos.
- Uma pessoa que caminha em direção à outra pessoa, encontram-se, conversam parados, e finalmente, vão em direções opostas.
- Uma pessoa que caminha em direção à outra pessoa, encontram-se, conversam parados, e finalmente, passam a andarem juntos.
- Uma pessoa que muda a direção de seu movimento para encontrar outra, encontram-se e conversam parados e, após, seguem na mesma direção.

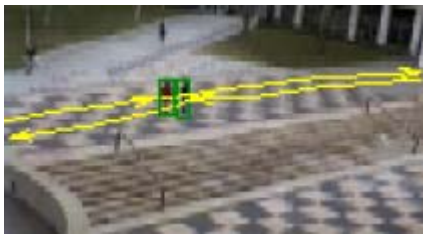

Figura 2.13 – Exemplo de comportamento procurado na cena real.

No trabalho de Oliver et al., não há direcionamento do modelo para detecção de trajetórias suspeitas. Nesse trabalho, as trajetórias que geram novos padrões de comportamentos não são recusadas, gerando novos padrões na base de dados, não havendo uma quantificação de quanto essa trajetória difere das demais, nem outras características das trajetórias são investigadas.

Outro trabalho que utiliza Cadeias de Markov é o trabalho de Liu e Chua [LIU 2006].

Em sua abordagem a rede é treinada para reconhecer diferentes tipos de comportamentos, sendo tanto os comportamentos normais, como os suspeitos. Depois de feito o treinamento, a cena é filmada em busca dos padrões gravados na rede.

O trabalho apresentado por Buxton e Gong [BUX 97] é voltado para fins de segurança e dispara um alarme quando trajetórias não-usuais são detectadas. No primeiro passo do processo, a câmera é calibrada e em seguida um modelo geométrico da cena é informado ao sistema para simplificar a interpretação das trajetórias pelo algoritmo de acompanhamento de pessoas utilizado. Para armazenar as informações obtidas pela supervisão da cena, uma Rede de Bayes é utilizada, guardando informações estatísticas que correspondem a dados espaço-temporais acerca das trajetórias desenvolvidas pelos objetos (carros ou pessoas).

Bases de dados anteriores podem ser usadas como forma de alimentar Redes de Bayes em novas cenas sem dados de treinamento. O comportamento de novas trajetórias é identificado através da rede, de forma a classificar tais trajetórias como: comportamento de agrupamento, um objeto seguindo outro, objetos em fila e comportamento desconhecido, ou seja, nenhum dos outros. Outras características dos objetos em movimento são armazenadas, como velocidade e posição. A disposição espacial entre pares de objetos é contabilizada, podendo um objeto estar na frente, atrás ou ao lado de outro.

Os pares de objetos também são classificados quanto à sua proximidade, como muito distante, distante, próximo, muito próximo e em contato. Todas as características individuais ou de grupos podem ser combinadas utilizando um formalismo criado para determinar quais comportamentos serão procurados durante o desenrolar do vídeo. Dessa forma comportamentos especificados em alto nível podem ser procurados na cena, sendo disparado um alarme quando ocorrerem.

O trabalho apresentado por Fuentes e Velastin [FUE 2004] objetiva procurar por eventos suspeitos em seqüências de vídeo. Seu modelo necessita de uma descrição prévia da cena, onde são informados os pontos que há escadas, entradas, corredores e portas. O modelo ainda associa uma determinada densidade populacional a cada cena filmada, tal densidade populacional calculada em pessoas.

Alguns tipos de comportamentos pré-definidos podem ser procurados na cena, como:

- Pessoa deixando objeto na cena: o *blob* inicial de uma pessoa se divide em dois, e uma parte fica parada sem movimento e a outra continua em movimento.
- Quedas: Quando a largura do *blob* passa a ser maior que a sua altura.
- Pessoas se escondendo da câmera: *blobs* que desaparecem várias vezes durante a filmagem. *Blob* sai da cena em uma região que não é um ponto de saída. Outro caso

é o *blob* ficar muito tempo na cena.

- Vandalismo: Somente um grupo ou pessoa na cena com centróides do *blob* muito irregulares e possíveis mudanças no *background* após evento.
- Lutas: Pessoas em luta se movem juntas e com movimentos rápidos. Centróides de *blobs* se movem correlacionados. Freqüente mistura de *blobs*. Mudanças rápidas nas características dos *blobs*.
- Intrusos: Centróide do *blob* em área proibida. Excesso de tempo permanência em tal área.
- Ataques: *Blobs* novos se aproximam muito dos *blobs* mais antigos na cena. Algum dos *blobs* pode estar estático na cena. Os *blobs* se encontram e iniciam um comportamento de luta.

Outra característica dos *blobs* que é analisada no modelo apresentado por Fuentes et al. é o tipo de distância entre as pessoas. Duas pessoas mantêm uma distância íntima entre si se a sua distância for menor que 1 metro. Para a distância social, têm-se 1,2 metros. Para a distância social têm-se 3 metros. E, por último, uma distância maior que isso é considerada uma distância pública. Tais conceitos estão ilustrados na figura 2.14.

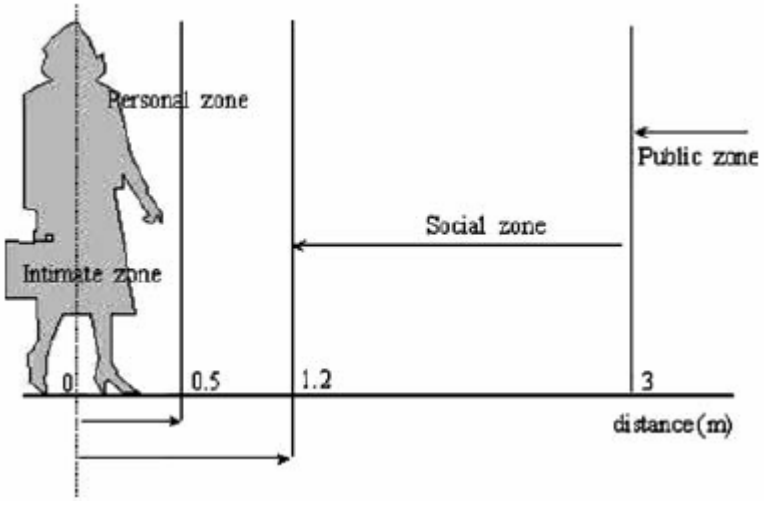

Figura 2.14 – Zonas de distância que as pessoas mantém entre si.

A cada novo quadro, são calculadas todas as características dos *blobs* envolvidos na cena. Em cada quadro, para que um determinado comportamento procurado seja reconhecido, o sistema verifica as características atuais e as anteriores do mesmo *blob* (e os que estão

relacionados com ele). Quando um comportamento for reconhecido, é considerado que houve um evento suspeito, e portanto será disparado um alarme. No artigo de Fuentes et al., não é comentado o método que realiza a consulta dos comportamentos.

2.3 Algoritmos para Acompanhamento de Pessoas

O acompanhamento de objetos não é o foco desta dissertação, mas sim é uma ferramenta para obter a posição de cada pessoa em cada quadro de uma seqüência de vídeo, que posteriormente será utilizada na detecção de movimentos não-usuais. Dessa forma, essa Seção apresentará uma breve revisão sobre os problemas principais encontrados no acompanhamento de objetos (em particular, no acompanhamento de pessoas) utilizando visão computacional, e algumas técnicas amplamentes referenciadas na literatura da área.

Há diversos algoritmos para o acompanhamento de pessoas (ou objetos em geral) descritos na literatura especializada [STA 2000, MCK 2000, CUC 2003, KAE 2003, PAI 2004, ADA 2006, CHE 2006, POR 2006]. A grande maioria dessas técnicas utiliza uma câmera estática para monitorar o ambiente, e então o passo inicial no acompanhamento de objetos é normalmente a remoção (ou subtração) do *background*. A remoção do *background* consiste em obter um modelo matemático do fundo da cena, e cada novo quadro da seqüência de vídeo é comparado com esse modelo. As discrepâncias geralmente são associadas a objetos em movimento, que formam grupos de *pixels* conexos (*blobs*). Entretanto, há diversos fatores que prejudicam a remoção do *background*: por exemplo, sombras tendem a serem detectadas como objetos em movimento, e variações (locais ou globais) na iluminação da cena causam diversos falsos positivos. Vários algoritmos incluem tratamento para sombras e/ou adaptação do modelo de *background* com respeito a variações na iluminação. Uma breve descrição de alguns algoritmos de subtração de *background* é dada a seguir.

No trabalho de McKenna et al. [MCK 2000] é proposto um método de subtração de *background* (método que compara quadros consecutivos para remover as partes estáticas) que combina informação da cor e do gradiente para detectar os objetos em movimento e para remover a sombra dos mesmos. No trabalho apresentado por Haritaoglu et al. [HAR 2000], é proposto um modelo estatístico de modelagem do *background* utilizando imagens em tons de cinza obtidas por uma câmera comum ou de infravermelho, mas não realiza nenhum tratamento para remover sombras. Outra técnica distinta que realiza o acompanhamento de pessoas é

proposta no trabalho de Elgammal et al. [ELG 2002], onde é utilizado um modelo paramétrico de *background* da cena utilizando imagens coloridas ou em tons de cinza. Sua vantagem é detectar corretamente os objetos mesmo sob árvores e mudanças de iluminação.

No trabalho de Cucchiara et al. [CUC 2003], além de utilizar um modelo estatístico da cena, são detectados e removidos os ghosts (conjunto de *pixels* que apresentam movimento médio nulo, como no caso do balançar de árvores). Um filtro temporal da mediana no espaço RGB é utilizado para detectar os objetos em movimento, além de realizar a remoção da sombra.Uma outra abordagem comum para modelar o *background* da cena é utilizar misturas de Gaussianas, provavelmente introduzida nesse contexto por Stauffer e Grimson [STA 2000]. No trabalho de Jacques et al [JAC 2006c], um modelo com base na mediana temporal dos *pixels* é apresentado, com tratamento de sombras e apdatação à variação de iluminação.

Após a extração dos *blobs*, a idéia central da maioria dos algoritmos de acompanhamento de objetos é exatamente analisar a evolução temporal dos *blobs*, que descreve a trajetória do objeto rastreado. Para tal, são utilizadas informações de forma, cor, textura e consistência de movimento (entre outras) para identificar um mesmo *blob* em quadros adjacentes de uma seqüência de vídeo. No sistema W4 [HAR 2000], a esperada projeção vertical dos *blobs* é utilizada para a identificação das pessoas, e informações de forma (curvatura), intensidade luminosa (câmeras monocromáticas são utilizadas neste trabalho) e textura são exploradas para acompanhar cada objeto. Pai et al. [PAI 2004] utiliza uma abordagem de casamento de grafos dinâmico para acompanhar os *blobs* das pessoas, adotando a métrica de Kullback-Leibler para calcular a similaridade de histogramas coloridos. KaewTrakulPong e Bowen [KAE 2003] exploram informações de consistência de movimento, forma e cor. Filtros de Kalman são utilizados na implementação do modelo de movimento, as dimensões do *bounding box* são empregadas como dados de forma, e o casamento histogramas é adotado para explorar a informação de cor.

Cheng e Chen [CHE 2006] realizam o acompanhamento de objetos no domínio das wavelets. Os autores exploram a remoção de ruído inerente à transformada wavelet para descartar movimentos de alta freqüência (como folhas se movendo ao vento), e exploram o desvio padrão em cada canal de cor e dimensões do *bounding box* para realizar o acompanhamento de objetos. No trabalho de Peursum et al. [PEU 2006], o modelo proposto visa determinar a posição da cabeça da pessoa e a sua postura através de Redes Bayesianas. Além disso, o modelo inclui o tratamento de oclusões.

Salienta-se que também há algoritmos focados para a correlação de características em quadros adjacentes que não necessitam da subtração de *background*, como o acompanhamento

baseado na matriz de covariância [ADA 2006] e o casamento dos histogramas integrais [POR 2006], recentemente propostos na literatura. Também podem ser utilizados algoritmos de fluxo óptico, que são empregados para descrever movimentos coerentes de pontos ou características semelhantes entre quadros diferentes do vídeo. Tal abordagem pode ser usada com a câmera em movimento, porém seu custo computacional é alto e é bastante sensível a ruídos, [BAR 94], [WAN 2003].

Neste trabalho, as câmeras de vídeo utilizadas estão instaladas no topo de prédios, fornecendo uma visão aproximadamente perpendicular ao plano do chão. A vantagem de tal tipo de câmera é a minimização de oclusões, e a facilitação no mapeamento de coordenadas de câmera para coordenadas de mundo (que denotam a posição real das pessoas fimadas). A maioria dos algoritmos encontrados na literatura para o acompanhamento de objetos utiliza câmeras com visão lateral/oblíqua da cena, e a perspectiva das pessoas que é explorada nesse tipo de câmera claramente não se aplica para vistas de topo [JAC 2006c]. Assim, decidiu-se neste trabalho utilizar o algoritmo de acompanhamento de pessoas de Jacques Jr. [JAC 2006b], [JAC 2006c].

Resumidamente, o algoritmo escolhido recebe o vídeo em escalas de cinza (a utilização de seqüências de vídeo monocromáticas confere uma maior flexibilidade ao modelo) e remove o *background* da cena, deixando apenas *pixels* do *foreground* (*blobs*). Na visão de topo, a perspectiva das cabeças das pessoas é praticamente invariante, e um algoritmo de correlação de máscaras é aplicado para acompanhar uma pessoa em quadros consecutivos.

Assim, para cada pessoa *i*, a sua trajetória é formada por uma lista de pontos $(x_i(t), y_i(t))$, onde *t* representa o tempo. Tal lista de pontos é utilizada como dados de entrada para o método proposto de detecção de movimentação não-usual, como será descrito no próximo capítulo.

Deve-se salientar que este capítulo apresenta apenas algumas técnicas para detecção de eventos não-usuais e acompanhamento de pessoas. Para obter informações adicionais sobre a detecção de movimentos suspeitos e demais técnicas de visão computacional relacionadas, pode-se ler os *surveys* dos grupos de Hu [HU 2004], Valera e Velastin [VAL 2005] e Moeslund [MOE 2006]. Em tais *surveys*, são comparados diversos métodos para realizar o acompanhamento de objetos na cena, para criar modelos físicos da cena, para analisar o fluxo de pessoas e, finalmente, para determinar o comportamento das pessoas.

2.4 Sistemas Comerciais no Mercado

Existem diversos sistemas comerciais presentes no mercado que realizam processamento do vídeo em várias aplicações na área de visão computacional. A maioria atua para fins de segurança, enquanto outros destinam-se a extrair informações da área filmada. A seguir alguns desses softwares são revisados brevemente.

O software *Movimento* da Empresa *Realviz*[1], oferece um tracker para objetos não-rígidos, como pessoas ou animais. Mais de uma câmera podem ser usadas para acompanhar a cena, e o resultado são as trajetórias em coordenadas tridimensionais. Além disso, as câmeras podem estar em movimento, e podem ser usadas câmeras de diferentes resoluções e modelos.

A empresa *Evitech*[2] desenvolve produtos na área de segurança em visão computacional utilizando sistemas integrados de câmeras. O sistema *Eagle* está apto a detectar trajetórias suspeitas e disparar uma alarme. Há diversos critérios para identificar trajetórias suspeitas, como uma pessoa que anda e pára freqüentemente, movimento na direção errada, velocidade incompatível com o padrão, pessoa que permanece tempo demais numa área, além de condições customizáveis. Além disso, a empresa oferece software para identificar intrusos em áreas de segurança, para acompanhar objetos em movimento, para detectar objetos que são abandonados pelas pessoas, e outras aplicações.

A empresa *L1 Identity Solutions*[3], desenvolve software para controle de acesso, com identificação de pessoas por biometria (leitura de impressão digital) e identificação de documentos fraudulentos por visão computacional. Outro tipo de controle de acesso é oferecido pela empresa *A4Vision*[4], que desenvolve software e hardware integrados para controle de acesso por reconhecimento facial e outros recursos. São empregadas câmeras 3D para reconhecimento facial e cartões podem ser integrados no sistema.

A empresa *Aimetrics*[5] oferece software para reconhecimento de trajetórias suspeitas em áreas de segurança. Várias câmeras podem ser integradas no sistema e quando um evento suspeito é detectado, é disparado um alarme. Quando um evento suspeito ocorrer o sistema grava um trecho de vídeo com *zoom* no infrator para fins posteriores. Eventos especificados pelo usuário podem ser consultados rapidamente no sistema após a gravação do vídeo.[1]

1 – http://realviz.com/
2 – http://evitech.com.br/
3 – http://www.l1id.com/
4 – http://www.a4vision.com/
5 – http://www.aimetis.com/

3. O Modelo Proposto

Como indicado no Capítulo anterior, o algoritmo de acompanhamento de pessoas retorna, para cada pessoa *i*, uma lista de pontos $(x_i(t), y_i(t))$ que denota sua evolução ao longo do tempo (ou seja, a trajetória da pessoa). Tais trajetórias são então avaliadas com relação à ocupação espacial e à interação com outras pessoas, conforme descrito a seguir.

3.1. Movimentos Não-Usuais com Respeito à Ocupação Espacial

Inicialmente, a trajetória de cada pessoa é extraída durante um período de treinamento. Este período é utilizado para gerar um histórico de ocupação na cena (Mapa de Ocupação Espacial), que fornece pontos de alta e baixa ocupação espacial. No período de teste, cada nova trajetória é comparada com esses mapas, e trajetórias não-usuais são detectadas quando a trajetória apresentar baixa ocupação espacial.

3.1.1. Mapas de Ocupação Espacial

O conceito central para a detecção de movimentos suspeitos com base na ocupação espacial é o Mapa de Ocupação Espacial (SOM – *Spatial Occupancy Map*). O conceito de SOM foi introduzido por Jacques Jr. [JAC 2006b], com o objetivo de comparar resultados de simulação de multidões. Neste trabalho, o conceito de SOM foi aprimorado e explorado para a detecção de movimentos não-usuais.

Basicamente, o SOM é uma imagem com o mesmo tamanho da área filmada, e que representa a ocupação espacial em cada *pixel* da área filmada durante um período de tempo (período de treinamento). De forma prática, o SOM é representado por uma matriz, inicializada com zeros. Para preenchê-la, cada ponto de uma trajetória incrementa em uma unidade o *pixel* correspondente à matriz do SOM. Dessa forma, áreas movimentadas são caracterizadas por terem uma alta ocupação espacial, enquanto que as áreas pouco movimentadas têm uma ocupação espacial nula ou quase nula. Tal propriedade será explorada na detecção de

movimentos não-usuais.

Na figura 3.1, tem-se um exemplo de área filmada e na figura 3.2, tem-se um exemplo de SOM (segundo a definição de Jacques Jr.) calculado durante um período de treinamento na mesma área. Além disso, na figura 3.1, pode-se ver as trajetórias capturadas durante o treinamento que gerou o SOM ilustrado na figura 3.2.

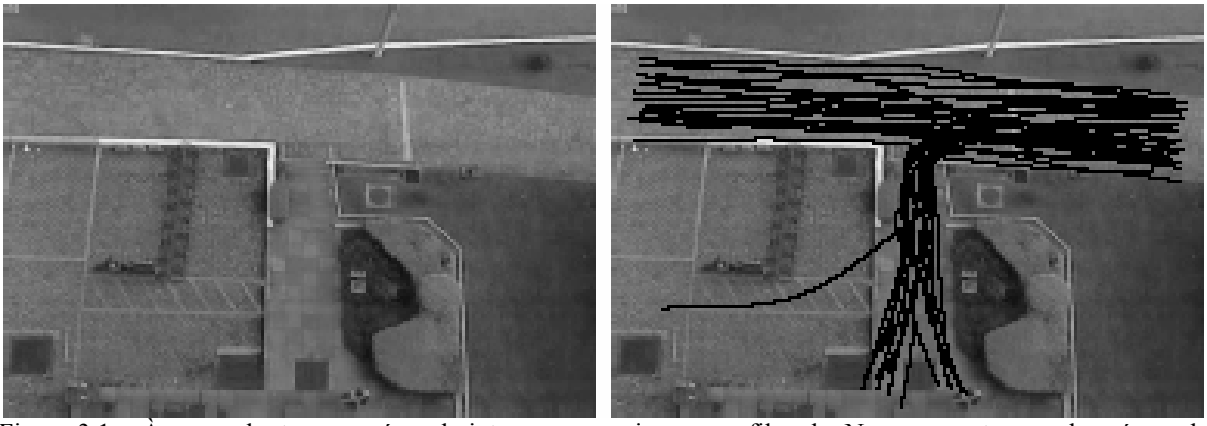

Figura 3.1 – À esquerda, tem-se a área de interesse, ou seja, a cena filmada. Nessa cena, tem-se duas áreas de movimentação, uma calçada (disposta na parte superior da imagem) e uma escada (na parte central da imagem). À direita, têm-se as trajetórias capturadas durante o treinamento.

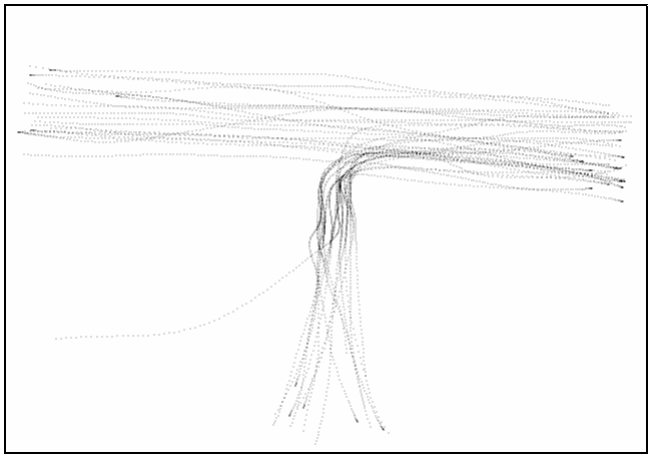

Figura 3.2 – SOM calculado durante um período de treinamento.

Um problema no conceito original do SOM é que pequenas imprecisões no algoritmo de acompanhamento de pessoas (que, por simplicidade, será chamado a partir de agora de *Tracker*) podem ocasionar o incremento do SOM em *pixels* distintos. Além disso, a lista de pontos (trajetória) retornada pelo *Tracker* se refere apenas ao centro da pessoa rastreada. Entretanto, o diâmetro das pessoas é maior do que um *pixel* (e depende do tipo de câmera

utilizada e seu posicionamento). Com isso, duas pessoas podem apresentar o mesmo movimento, mas incrementarem posições diferentes no SOM.

Uma idéia simples e eficiente para resolver esses problemas é "espalhar" a influência de cada ponto de uma trajetória à sua vizinhança, de forma a reproduzir o diâmetro padrão ocupado por uma pessoa e reduzir o efeito imprecisão no acompanhamento das pessoas. Para tal, foi utilizada uma máscara baseada na função gaussiana discreta bidimensional, que é centrada no ponto corrente da trajetória. A máscara é normalizada para ter soma unitária, a fim de causar um incremento total unitário ao SOM. Seu desvio padrão σ é utilizado para determinar a abrangência da máscara, ou seja, é utilizado para causar incremento de forma proporcional ao diâmetro de uma pessoa que é filmada na câmera de topo. Alternativamente, pode-se interpretar o parâmetro σ como a escala de análise do problema: valores pequenos de σ provêm uma análise local, enquanto que valores maiores de σ resultam em uma análise mais global da cena.

Na figura 3.3, tem-se um exemplo de SOM calculado com $\sigma = 5$ (esquerda) e outro com o desvio padrão do tamanho de uma pessoa., nesse arranjo de câmera, com $\sigma = 15$ (direita). Percebe-se que o SOM calculado com $\sigma = 15$ fornece um resultado visual mais coerente com o esperado, com menos lacunas entre as trajetórias individuais acompanhadas. Além disso, percebe-se que uma única pessoa andou no estacionamento durante o período de treinamento. Como essa ocupação espacial não é corroborada por outras pessoas, o SOM nessa região é baixo (a influência dessa pessoa no estacionamento foi diluída com a convolução, especialmente no caso $\sigma = 15$).

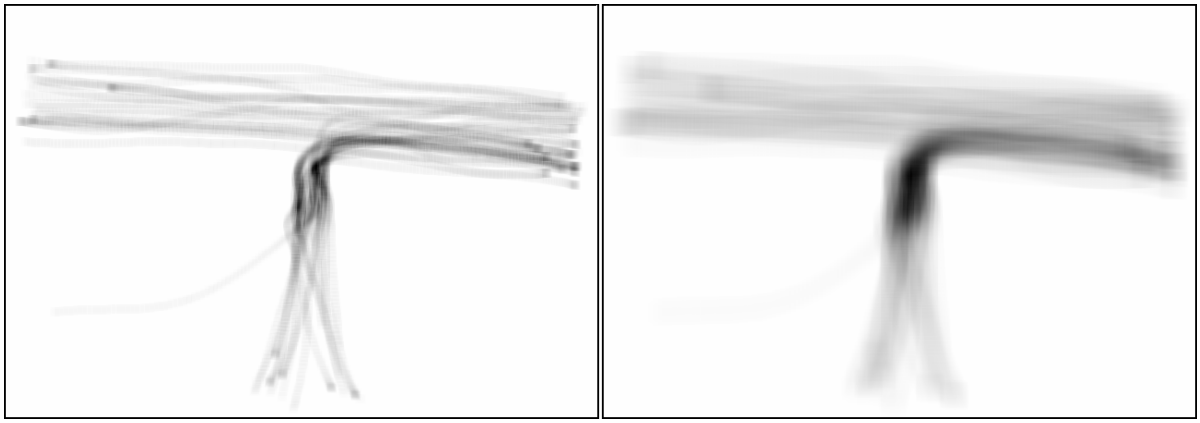

Figura 3.3 –Exemplos de SOMs calculados com diferentes desvios padrões. Na figura da esquerda, o desvio padrão é pequeno, enquanto que a figura da direita foi obtida com um desvio padrão do tamanho de uma pessoa.

Matematicamente, o SOM $S_\sigma(x,y)$ calculado com desvio padrão σ (da gaussiana) é definido na equação (1), que coincide com a convolução do SOM (de incrementos unitários) com a máscara gaussiana. A equação (2) representa a gaussiana bidimensional discreta (truncada), com desvio padrão σ.

$$S_\sigma(x,y) = \sum_{i=1}^{N} \sum_{t=1}^{N_F(i)} g_\sigma(x - x_i(t), y - y_i(t)), \quad \textbf{(1)}$$

$$g_\sigma(x,y) = \begin{cases} \frac{1}{c} e^{\frac{-x^2-y^2}{2\sigma^2}} & \text{se } -2\sigma \le x,y \le 2\sigma \\ 0 & \text{caso contrário} \end{cases}, \quad \textbf{(2)}$$

A constante *c* é obtida através da normalização da máscara da gaussiana, de tal modo que $\sum_{(x,y)} g_\sigma(x,y) = 1$. Na equação (1), *N* é o número de pessoas contabilizadas pelo *Tracker*, $N_F(i)$ é a duração (em número de quadros) da trajetória *i,* e $g_\sigma(x,y)$ é a gaussiana bidimensional discreta truncada.

3.1.2. Avaliação de Trajetórias Através do SOM

O SOM $S_\sigma(x,y)$ é calculado durante o período de treinamento utilizando um determinado desvio padrão σ que corresponde ao diâmetro aproximado de uma pessoa em coordenadas da imagem, gerando o espalhamento da área de influência. Considerando $(x_i(t), y_i(t))$ a trajetória da pessoa *i* a cada instante *t*, as trajetórias usuais são aquelas com $S_\sigma(x,y)$ suficientemente grande ao longo da curva $(x_i(t), y_i(t))$.

A função $S_{\sigma,\mathrm{i}}(t) = S_\sigma(x_i(t), y_i(t))$ determina o valor do SOM no instante *t* para uma pessoa *i*. Através dessa função, a trajetória pode ser separada em partes usuais e não-usuais, segundo o parâmetro T_{SOM}, que é o limiar da avaliação do SOM. As partes da curva $S_{\sigma,\mathrm{i}}(t)$ que apresentarem uma ocupação espacial menor que T_{SOM} são consideradas não-usuais.

Uma escolha bastante simples seria definir $T_{SOM} = 0$, para que cada *pixel* não-nulo de

$S_\sigma(x,y)$ pertença a uma região de ocupação válida. Entretanto, a convolução com a gaussiana usada para calcular $S_\sigma(x,y)$ tende a aumentar a região que foi efetivamente ocupada pelas pessoas contabilizadas no período de treinamento. Este trabalho propõe obter o limiar T_{SOM} automaticamente a partir do SOM, removendo a porção *r* dos menores valores de $S_\sigma(x,y)$ que estão associados com a cauda da gaussiana.

Considere Q um vetor contendo todos os valores do SOM não-nulos em ordem crescente, eliminando os valores repetidos. Se *n* é o tamanho desse vetor, então o limiar T_{SOM} pode ser obtido através da seguinte expressão:

$$T_{SOM} = Q([rn]), \quad \textbf{(3)}$$

Onde [] representa a parte inteira de um número. Em outras palavras, a equação (3) obtém o valor de Q que está no percentil *r* da distribuição de *Q*. Nos experimentos apresentados foi utilizado $r = 0.4$. Na figura 3.4, o gráfico superior apresenta uma trajetória usual em toda sua duração, enquanto que no gráfico inferior é apresentada uma trajetória que começa em uma região usual, mas depois de algum tempo entra em uma região pouco ocupada do SOM (sendo, assim, considerada não-usual). As partes em azul são usuais e as partes não-usuais estão em vermelho.

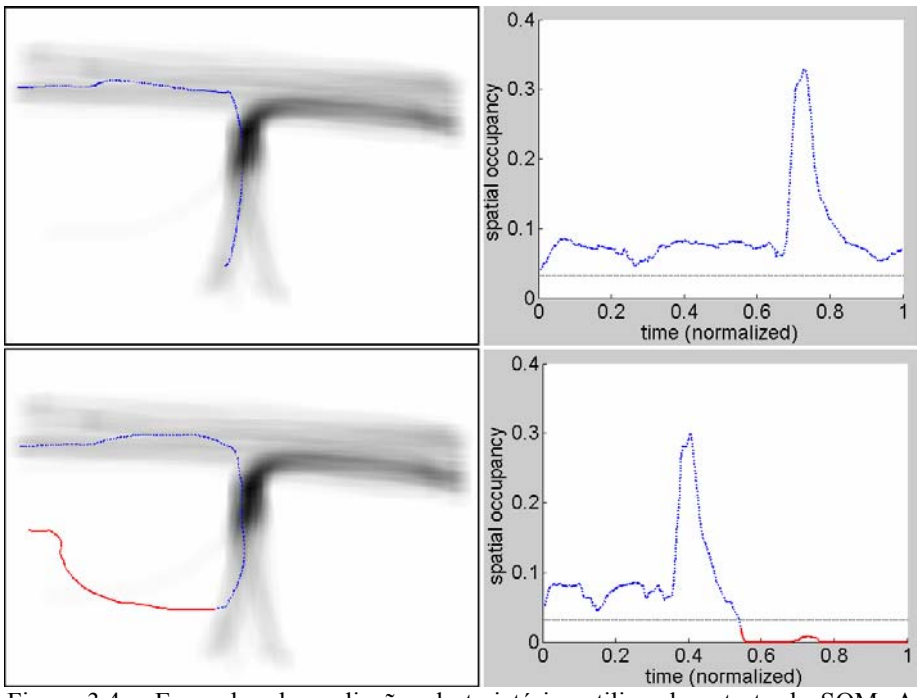

Figura 3.4 – Exemplos de avaliações de trajetórias utilizando o teste do SOM. A figura de cima é a avaliação de uma trajetória normal, com o gráfico do SOM à direita. Na figura de baixo, a trajetória é não-usual, pois, no final do percurso foi passado numa região de baixíssima ocupação espacial (em vermelho).

3.1.3. Transformada Distância

Uma situação comum que pode ocorrer é o caso de uma pessoa caminhar muito próxima da borda (ou um pouco fora) da região ocupada obtida durante o período de treinamento. Nesse caso, a função $S_\sigma(x, y)$ pode retornar valores quase nulos ou nulos em uma grande parte da trajetória, que poderia fazê-la ser classificada como não-usual. Por exemplo, a trajetória mostrada na figura 3.5 apresenta uma ocupação espacial abaixo do limiar T_{SOM} em todos os pontos da trajetória (a figura da esquerda mostra o SOM limiarizado conforme T_{SOM}, e a figura da direita mostra o SOM original). Embora intuitivamente essa trajetória aparente ser usual, o algoritmo proposto até então a classificaria com não-usual em todos seus pontos.

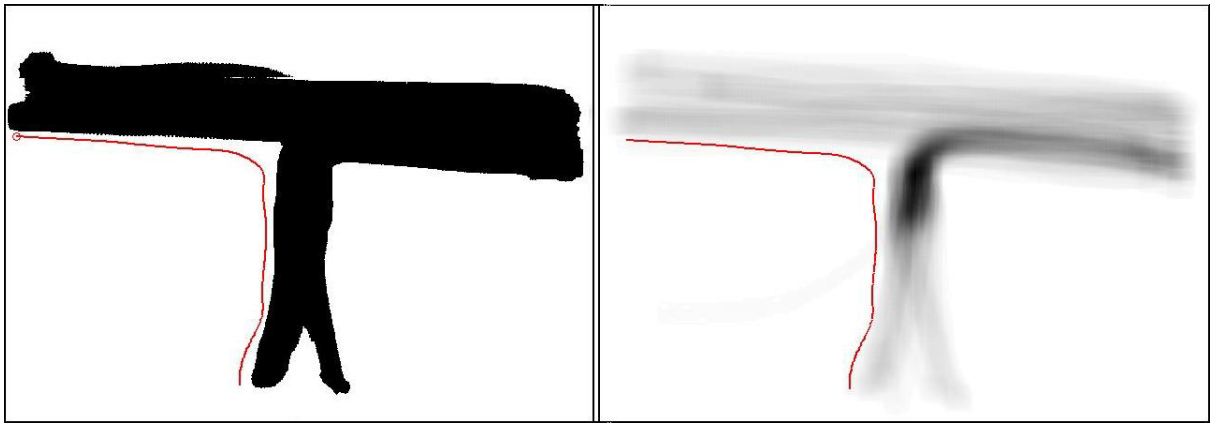

Figura 3.5 – Exemplo de trajetória (em vermelho). Na figura da direita tem-se o SOM e na figura da esquerda o SOM binário correspondente.

Um teste alternativo à comparação direta com o SOM é proposto com base na Transformada Distância (TD), que é uma ferramenta bem conhecida em geometria computacional e processamento de imagens [SOI 2002]. Dada uma imagem binária *I*, a transformada distância é uma imagem *D* com o mesmo tamanho que *I*, tal que o valor $D(x, y)$ em certo ponto (x, y) corresponde à menor distância desse ponto a qualquer outro não-nulo da imagem *I*.

Para utilizar a transformada distância, é necessário criar uma imagem binária a partir do SOM, chamada de SOM Binário. O SOM Binário representa as partes com ocupação relevante do SOM. Para calculá-lo, é utilizado o limiar T_{SOM} definido anteriormente sobre a imagem

SOM $S_\sigma(x,y)$ utilizando o teste a seguir:

$$I(x,y) = \begin{cases} 1 \text{ se } S_\sigma(x,y) \geq T_{SOM} \\ 0 \quad \text{caso contrário} \end{cases} . \quad \textbf{(4)}$$

Seja $D(x,y)$ a transformada distância do SOM binário $I(x,y)$. A evolução da distância mínima entre a trajetória $(x_i(t), y_i(t))$ e a região ocupada (representada na imagem *I*) pode ser calculada por: 4

$$d_i(t) = D(x_i(t), y_i(t)) , \quad \textbf{(5)}$$

Assim, partes não-usuais da trajetória são detectadas quando a trajetória estiver suficientemente longe da região válida (ou seja, $d_i(t) > T_{dist}$, onde T_{dist} é a distância máxima aceitável entre a trajetória e a região ocupada válida). Embora T_{dist} seja dependente do contexto, um valor padrão para T_{dist} pode ser obtido automaticamente utilizando o desvio padrão σ através de $T_{dist} = 2\sigma$, ou seja, a máxima distância aceitável é aproximadamente o dobro do diâmetro esperado das pessoas. Na figura 3.5, tem-se uma trajetória aparentemente usual que é considerada não-usual pelo critério SOM em todos os pontos. De fato, a imagem à esquerda mostra que toda a trajetória está fora do SOM Binário. Na figura 3.6 é mostrado o mesmo SOM binário à esquerda, e os valores da Transformada Distância à direita.

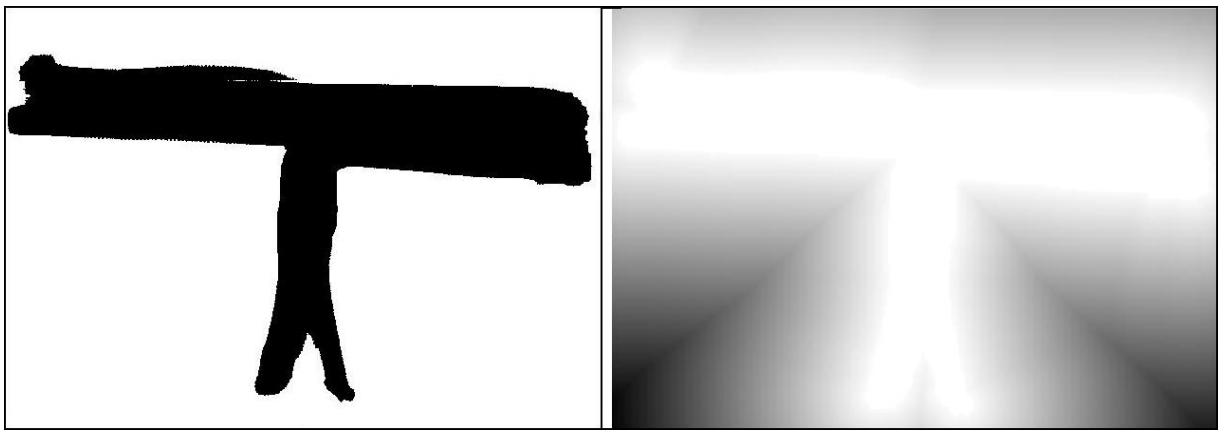

Figura 3.6 – Na figura da esquerda, tem-se o SOM Binário e na figura da direita tem-se a transformada distância calculada a partir do Som Binário. Na transformada distância, quanto mais escuro o *pixel*, mais distante da área movimentada relevante.

Um exemplo do uso da transformada distância está nas figura 3.7 e 3.8. Na figura 3.7, tem-se o SOM calculado no treinamento à esquerda, e na direita tem-se uma trajetória que é rejeitada pelo teste do SOM (como pode ser visto na figura 3.8 à esquerda, a ocupação espacial da trajetória está abaixo do limiar T_{SOM}). À direita, na figura 3.8, tem-se o teste equivalente com a transformada distância. Pode-se ver que os valores da transformada na trajetória ficaram abaixo do limiar T_{dist}, portanto a trajetória foi considerada usual em todo seu percurso (de fato, ela apresenta uma pequena distância com relação à área efetivamente ocupada no período de treinamento).

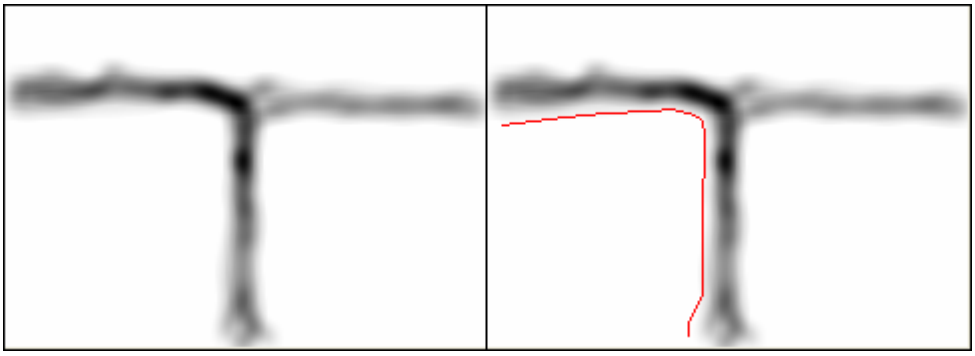

Figura 3.7 – Outro exemplo de trajetória. Um observador humano pode considerar a trajetória em vermelho como usual, porém o teste do SOM a rejeita.

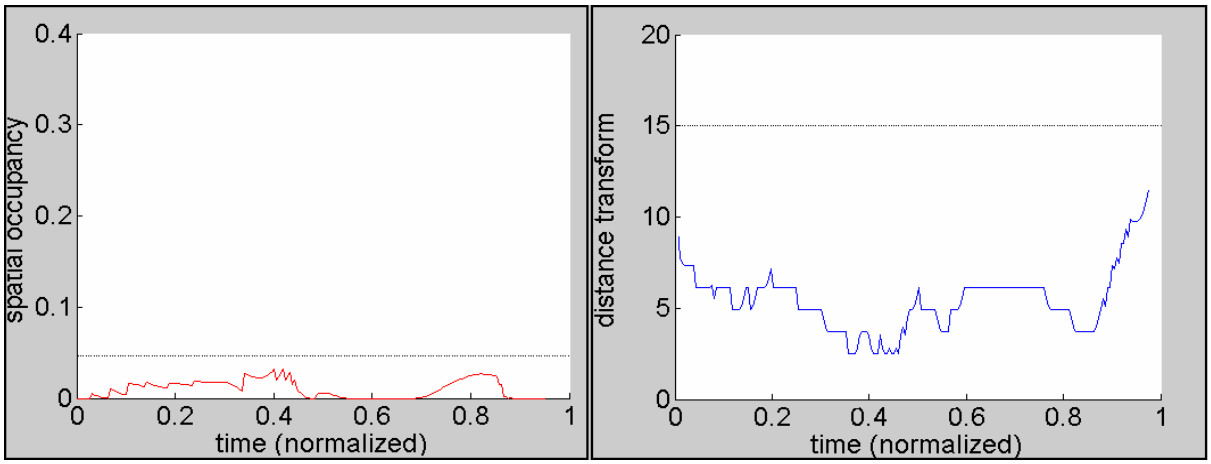

Figura 3.8 – Avaliação obtida com o SOM (esquerda) e transformada distância (direita) da trajetória mostrada na Figura 3.7.

Até agora foram apresentadas duas formas de se classificar como usuais ou não-usuais partes das trajetórias: uma através do teste do SOM e outra através do teste da TD. Além disso, um outro teste pode ser feito, onde a trajetória inteira pode ser classificada, comparando a

duração de partes não-usuais com o total da duração da trajetória. Se $D^i_{unusual}$ representa a duração das partes não-usuais da trajetória *i*, a trajetória inteira é considerada não-usual se:

$$D^i_{unusual} < T_{unusual} N_f(i), \quad \textbf{(6)}$$

Onde $0 \le T_{unusual} \le 1$ é um limiar e $N_F(i)$ representa a duração da trajetória (em quadros). Embora $T_{unusual}$ seja dependente de contexto e aplicação, foi usado um limiar $T_{unusual} = 0.3$ em todos os exemplos neste trabalho. Isso significa que, se em pelo menos 30% do tempo que uma trajetória foi classificada como não-usual, então a trajetória inteira é dita não-usual.

É interessante salientar que os testes do SOM e da TD são complementares. O teste do SOM fornece valores nulos fora da região válida, e a ocupação espacial no interior da região válida. Por outro lado, o teste da TD fornece as distâncias entre a trajetória e a região válida no exterior da região válida, e valores nulos no interior da região válida. Os resultados obtidos com os testes do SOM e da TD foram publicados em [JAC 2006a].

3.2. Movimentos Não-Usuais com Respeito às Relações Interpessoais

Além de avaliar comportamentos suspeitos com base apenas na trajetória (ocupação espacial), neste trabalho também é avaliada a relação entre as pessoas que transitam concomitantemente na cena. De fato, a avaliação das relações entre as pessoas fornece um tipo de informação complementar à simples ocupação espacial do ambiente.

Para determinar as relações interpessoais entre as pessoas na cena, o conceito de *proxemics*, presente na sociologia e psicologia, é empregado nesse trabalho. O termo *proxemics* foi proposto primeiramente por Edward Hall [HAL 73] para descrever o uso social do espaço, ou seja, para se referir ao espaço pessoal de cada indivíduo. Tal espaço corresponde a uma área invisível ao redor do mesmo, que em condições normais não é invadida por outras pessoas.

Hall propõe quatro faixas de distâncias interpessoais mais comumente usadas pelas pessoas, sendo cada faixa caracterizada por um determinado tipo de interação entre elas. O tamanho de cada faixa e o comportamento esperado em cada faixa pode variar de acordo com a

cultura das pessoas envolvidas, e a tabela 1 abaixo mostra as classificações padrão propostas por Hall e suas características.

Classificação	**Faixa de Distância**	**Tipo de Iteração**
Distância Íntima	Até 0,5 metros	Flerte, ameaças
Distância Pessoal	0,5 metros até 1,25 metro	Conversa entre amigos
Distância Social	1,25 metros até 3,5 metros	Negócios, Atendimento
Distância Pública	Acima de 3,5 metros	Caminhando numa multidão

Tabela 1: Tipos de distância pessoal

Jacques Jr. [JAC 2006a] propôs uma técnica para extrair e quantificar características psicológicas e sociológicas de cada pessoa utilizando Diagramas de Voronoi (DV), e assim determinar qual o tipo de interação entre cada par de pessoas na cena de acordo com a classificação das faixas de distâncias propostas por Hall. A evolução de tais interações ao longo do tempo foi então utilizada para detectar e classificar a formação de grupos. Neste trabalho, a dinâmica da formação de grupos é explorada na detecção de movimentos não-usuais, que são detectados através de um autômato finito não-determinístico. Por exemplo, um comportamento de aproximação, seguido de agrupamento e posteriormente de afastamento pode caracterizar um assalto. Os parágrafos seguintes descrevem brevemente o algoritmo de detecção de grupos de Jacques Jr, e na seqüência a técnica proposta para detecção de movimentos não usuais é explicada.

A idéia central do método de Jacques Jr. é utilizar, a cada quadro da seqüência de vídeo, a posição de cada pessoa como um *site* para geração do DV. Uma característica do DV é que qualquer ponto interno a um certo polígono está mais próximo do *site* gerador do que qualquer outro *site*. Dessa forma, os polígonos podem ser considerados aproximações do espaço pessoal de cada pessoa. Além disso, o DV permite facilmente determinar os vizinhos de cada pessoa, além das respectivas distâncias.

A cada instante temos um novo quadro do vídeo, portanto, nesse trabalho, é utilizado o conceito de Diagramas de Voronoi Dinâmicos (DVDs), onde a cada quadro do vídeo, é gerado um novo DV, permitindo identificar a evolução em função do tempo das características de agrupamento das pessoas. Em resumo, duas pessoas são agrupadas se a distância entre elas for menor do que um certo limiar (obtido pelas distâncias interpessoais de Hall) durante um certo período de tempo (chamado de tempo de agrupamento, ou T_g). Na prática, um grupo é

detectado se a distância entre as pessoas em uma fração *p* do tempo de agrupamento é menor do que o limiar, pois o algoritmo de acompanhamento de pessoas apresenta pequenas imprecisões, e até mesmo grupos fortes (como um casal) podem se afastar momentaneamente (por exemplo, para desviar de obstáculos).

No trabalho de Jacques Jr., também foi desenvolvido o conceito de espaço pessoal percebido (EPP), que corresponde à região do polígono de Voronoi que está no campo de visão da pessoa (tal campo de visão é aproximado por um setor circular). O EPP é então utilizado para classificar um grupo como voluntário ou involuntário: em um grupo voluntário, assume-se que as pessoas estão juntas por vontade própria, ou seja, ficam próximas mesmo tendo espaço livre à frente (seus EPPs são altos); por outro lado, grupos involuntários podem ocorrer devido a restrições do espaço, ou seja, as pessoas ficam próximas por falta de espaço (seus EPPs são baixos).

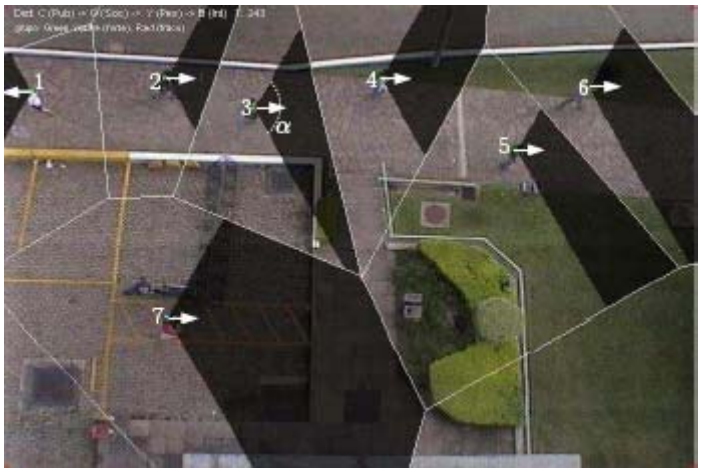

Figura 3.9 – No trabalho de Jacques Jr, as áreas escuras correspondem ao EPP de cada pessoa na cena. Ela é centrada no vetor direção, desenhado em branco.

Neste trabalho, diversas outras métricas são extraídas e armazenadas a partir dos DVDs, como a evolução do movimento relativo entre duas pessoas (afastamento ou aproximação, estas podendo ocorrer pela frente ou por trás). Em particular, a abordagem para afastamento ou aproximação (pela frente ou por trás) é avaliada comparando o vetor velocidade das duas pessoas envolvidas. Se o produto escalar dos vetores for maior que zero, então é considerada aproximação por trás (duas pessoas apresentam aproximadamente o mesmo sentido de movimento), caso contrário, é considerado movimento de aproximação pela frente. Também é

possível quantificar o sentido de afastamento/aproximação em mais direções (frente, trás e laterais), avaliando o valor numérico do produto escalar (que fornece o ângulo entre os vetores envolvidos).

Na vida real, cada tipo de comportamento interpessoal está relacionado com as características do movimento relativo entre as pessoas. Por exemplo, seqüestros podem ser determinados por comportamento de interceptação (aproximação) e depois de agrupamento de pessoas. Em outro exemplo, um comportamento de furto (batedor de carteira) pode ser detectado por uma aproximação por trás, distância relativa pequena (distância íntima) sem agrupamento, e afastamento. Além disso, pode-se combinar as informações de relações interpessoais com informações de ocupação espacial (SOM). Por exemplo, um agrupamento em uma região com SOM alto pode ser considerado normal, enquanto que um agrupamento similar em uma região pouco ocupada do ambiente pode ser considerada suspeita. Também deve-se salientar que há diversos comportamentos "normais" com características similares aos não-usuais. Por exemplo, ao invés de um seqüestro, a aproximação seguida de agrupamento pode representar simplesmente um encontro de amigos que seguem na mesma direção.

3.2.1. Autômato Finito Não-Determinístico

Durante o desenrolar do vídeo, é gerada uma grande quantidade de informações relativas às características do comportamento das pessoas, que ficam armazenadas na forma de uma base de dados que pode ser consultada mais tarde em procura de comportamentos específicos pré-estabelecidos. Tais consultas são feitas na base de dados através de um autômato finito não-determinístico (AFND). É necessário um autômato distinto para cada comportamento a ser procurado, sendo que o mesmo pode ser decomposto em vários sub-comportamentos.

Antes de apresentar-se a definição de AFND, é necessário apresentar-se as definições de gramática e alfabeto. Um alfabeto é um conjunto de símbolos (por exemplo $\Sigma = \{a, b, c,..., x, z\}$, que compreende as letras da língua portuguesa). Sentenças (ou palavras) são seqüências de símbolos de um alfabeto (por exemplo $S = \{cidade\}$, onde os símbolos são agrupados numa seqüência ordenada formando a palavra *cidade*).

Uma linguagem é um conjunto de palavras provenientes de um alfabeto. Por exemplo, pode-se ter uma linguagem *L* cujas sentenças que a compõem são todas as palavras que iniciam

por *ci*. Dessa forma, *cidade* e *cimento* são elementos da linguagem *L*. Finalmente, uma gramática é uma regra formal para descrever como são obtidas todas as sentenças de uma determinada linguagem.

De forma genérica, uma gramática é uma *4-tupla* ordenada *G = (V, T, P, S)*, onde *V* é o conjunto de símbolos não-terminais. *T* é o conjunto de símbolos terminais, sendo que *V* e *T* não podem ter nenhum elemento em comum. *P* é o conjunto das relações de cada não terminal pertencente a *V* com uma seqüência ordenada de símbolos $(V \cup T)^*$, onde * representa a operação de *fecho de conjunto*. Por último, *S* é o não-terminal inicial da gramática.

Um exemplo de gramática pode o seguinte: *G = (V, T, P, S)*, onde:

- $\boldsymbol{V} = \{A, B, C\}$
- $\boldsymbol{T} = \{0,1, 2, 3, 4, 5, 6, 7, 8, 9\}$
- $\boldsymbol{P} = \{ C \rightarrow A\ B\ ,\ A \rightarrow 01\ ,\ B \rightarrow 0|1|2|3|4|5|6|7|8|9\}$
- $\boldsymbol{S} = \{C\}$

Uma derivação de uma gramática é a formação de uma sentença através da substituição consecutiva dos símbolos não-terminais gerados através do não-terminal inicial $S = (V \cup T)^*$ pelas suas próprias relações até restarem apenas elementos terminais na sentença. No exemplo de gramática visto anteriormente, uma derivação possível que gera a palavra *018* é $S = C \rightarrow AB \rightarrow 01B \rightarrow 018$. Como visto no exemplo, a gramática *G* é a gramática que descreve a linguagem cujas sentenças contém apenas três números e começam por *01* e terminam por um outro número do alfabeto.

Em geral, um autômato é uma máquina de estados que tem a função de executar um programa. Um autômato é basicamente composto de 3 partes:

- Fita: É a entrada que contém a informação a ser processada. A fita é finita de ambos os lados e é dividida em células, onde cada célula guarda um símbolo do alfabeto.
- Unidade de Controle: Guarda o estado corrente do autômato. É uma unidade de leitura (cabeça da fita) que acessa cada célula da fita, lê o seu valor, troca de estado e se movimenta para a direita. No início, a unidade de controle está posicionada na célula mais à esquerda da fita.
- Programa ou Função Transição: Função que determina as mudanças de estado do autômato de acordo com os valores lidos em cada célula da fita.

Um autômato finito determinístico pode ser definido por uma *5-tupla* ordenada $M = (\Sigma, Q, \delta, q_0, F)$. Onde Σ é o alfabeto dos símbolos de entrada. Q é o conjunto finito de estados. δ é a função programa tal que: $\delta: Q \times \Sigma \rightarrow Q$, sendo esta parcial. q_0 é o estado inicial do autômato, sendo elemento de Q. E por último, F é o conjunto de estados finais, também pertencentes a Q.

Os autômatos podem ser utilizados para reconhecer sentenças de uma gramática. Nesse caso a fita é a sentença, e o autômato avança na fita trocando de estado de acordo com a função transição. Quando o autômato chegar ao final da fita, se o seu estado atual for um estado final, a sentença é dita como pertencente à referida gramática. Para cada gramática, é necessário construir um autômato específico.

Em outras palavras, o autômato inicia o processamento sobre a célula mais à esquerda da fita, tendo como estado atual o estado inicial q_0. A função transição recebe o estado corrente do autômato e o elemento corrente da fita e retorna o novo estado do autômato, em seguida o autômato avança para a posição seguinte à direita. O processo é repetido até chegar ao último elemento da fita. Se, após o processamento do elemento final, o estado atual do autômato pertencer ao conjunto de estados finais, a sentença na fita será dada como reconhecida, ou seja, ela pertence a gramática. A seguir, têm-se um exemplo de gramática e autômato:

Tem-se a gramática $G = (V, T, P, S)$, onde:

- $\boldsymbol{V} = \{A, B, C\}$
- $\boldsymbol{T} = \{0,1, 2, 3, 4, 5, 6, 7, 8, 9\}$
- $\boldsymbol{P} = \{ C \rightarrow A\ B\ ,\ A \rightarrow 01\ ,\ B \rightarrow 0|1|2|3|4|5|6|7|8|9\}$
- $\boldsymbol{S} = \{C\}$

E o autômato $M = (\Sigma, Q, \delta, q_0, F)$, onde:

- $\boldsymbol{\Sigma} = \{0,1, 2, 3, 4, 5, 6, 7, 8, 9\}$
- $\boldsymbol{Q} = \{q_0, q_1, q_2, q_3\}$
- $\boldsymbol{\delta} = \{(0, q_0) \rightarrow q_1, (1, q_1) \rightarrow q_2, (x, q_2) \rightarrow q_3\}$, *onde* $x \in \Sigma$.
- $\boldsymbol{q_0} = q_0$
- $\boldsymbol{F} = \{q_3\}$

Para reconhecer a sentença *012*, o autômato finito determinístico seguirá os seguintes passos:

1. Símbolo corrente: *0*, estado atual: q_0, transição $(0, q_0) \rightarrow q_1$
2. Símbolo corrente: *1*, estado atual: q_1, transição $(1, q_1) \rightarrow q_2$
3. Símbolo corrente: *2*, estado atual: q_2, transição $(2, q_2) \rightarrow q_3$
4. Estado atual q_3, como a sentença acabou e $q_3 \in F$, então a sentença é aceita.

O autômato utilizado nesse trabalho é o autômato finito não-determinístico (AFND), que difere do autômato finito determinístico (AFD) por aceitar mais de um estado como estado corrente do autômato. Assim como no AFD, o AFND também é definido por uma *5-tupla* ordenada $M = (\Sigma, Q, \delta, q_0, F)$, com a diferença que a função programa é definida da seguinte forma: $\delta(q, a) = \{q_1, q_2, \ldots, q_n\}$, onde $q \in Q$, $a \in \Sigma$ e $q_1, q_2, \ldots, q_n \in Q$. Em um AFND, uma sentença é reconhecida quando a fita terminar e pelo menos um dos estados atuais do autômato pertencer ao conjunto de estados finais.

Nesse trabalho, as sentenças reconhecidas pelo autômato correspondem a seqüências de comportamentos (símbolos) obtidos por uma mesma pessoa na cena. A cada quadro de uma pessoa é criado um novo símbolo e ele é processado pelo autômato da mesma. Quando o autômato chegar a um estado final (mesmo sem ter terminado a seqüência), é considerado reconhecido o comportamento procurado e é disparado um alarme, pois não é necessário processar até o fim da trajetória para reconhecer um evento suspeito. A função transição do autômato é responsável por fazer as trocas de estados, sendo sempre troca de um determinado estado por outro quando lê determinado símbolo (comportamento).

O autômato pode ser executado enquanto o vídeo transcorre, detectando eventos suspeitos em tempo real, ou pode ser acionado mais tarde, no intuito de realizar uma consulta ao histórico de comportamentos que houveram na cena (a base de dados). Dessa forma, se o operador de vídeo esteve ausente, ele pode procurar por comportamentos suspeitos que ocorreram no momento anterior.

Por exemplo, para o comportamento de seqüestro, que pode ser descrito como a seqüência dos seguintes comportamentos: movimento de interceptação (aproximação) e depois de agrupamento de ambas as pessoas, pode-se criar um arquivo de texto que contém todas as definições do autômato que reconhece essa seqüência de comportamentos e informado ao programa para procurar na base de dados. Os comportamentos possíveis numa seqüência são combinações das letras exibidas na tabela 2.

Aproximação = *P*	Afastamento = *F*
Frente = *E*	Atrás = *A*
SOM Válido = *V*	Som Inválido = *I*
TD Válida = *T*	TD Inválida = *D*
Agrupamento Íntimo = *N*	Agrupamento Pessoal = *S*
Agrupamento Social = *C*	Agrupamento Público = *L*

Tabela 2: Alfabeto da gramática

Por exemplo, para o comportamento de seqüestro visto no exemplo anterior, forma-se a seqüência de comportamentos: $P\ N$, ou seja, aproximação seguido de agrupamento. Uma trajetória é uma fita (seqüência de comportamentos) obtida da avaliação de cada Diagrama de Voronoi de um DVD enquanto o vídeo transcorre. Para cada pessoa será mantida uma instância distinta do seguinte autômato que tem a função de reconhecer a seqüência $P\ N$ em uma fita. O autômato a seguir reconhece o comportamento $P\ N$.

Autômato $M = (\Sigma, Q, \delta, q_0, F)$, onde:

- $\Sigma = \{P, N, X\}$, *onde X são os demais símbolos possíveis.*
- $Q = \{q_0, q_1, q_2\}$
- $\delta = \{(P, q_0) \rightarrow q_1, (N, q_1) \rightarrow q_2\}$
- $q_0 = q_0$
- $F = \{q_2\}$

A tabela 2 mostra o alfabeto que é utilizado pelos autômatos. Para procurar por comportamentos simultâneos no DVD, como aproximação pela frente, deve-se concatenar todos os símbolos simultâneos envolvidos. Eles serão tratados pelo sistema como um único símbolo durante a consulta ao histórico, pois o histórico guarda a cada quadro todas as informações que podem ser extraídas. Esse tipo de comportamento pode ser escrito como $PE\ NI$, utilizando-se dos símbolos. Os comportamentos concomitantes são concatenados e os consecutivos são separados por espaço. Por exemplo, o seguinte autômato é possível para reconhecer tal comportamento:

Tem-se o autômato $M = (\Sigma, Q, \delta, q_0, F)$, onde:

- $\Sigma = \{PE, NI, X\}$, *onde X são os demais símbolos (ou combinações) possíveis.*
- $Q = \{q_0, q_1, q_2\}$
- $\delta = \{(PE, q_0) \rightarrow q_1, (NI, q_1) \rightarrow q_2\}$
- $q_0 = q_0$
- $F = \{q_2\}$

Uma característica útil do autômato é que ele suporta buscas mais complexas, como por exemplo, "*(PA | N) F*": aproximação por trás ou agrupamento, seguido de afastamento. Esse tipo de comportamento pode gerar autômatos bastante complexos. Porém é desejável obter-se uma forma mais simples de se formalizar as consultas ao histórico. Para resolver tal problema, ao invés de se informar um autômato específico para reconhecer um comportamento, pode-se informar diretamente a seqüência de comportamento como forma de realizar a mesma consulta, por exemplo, informa-se a seqüência: *(PA | N) F*.

O programa faz a tradução automática da sentença informada para um AFND equivalente, simplificando a forma de interação do operador de vídeo com o programa. Através do AFND, cada vez que o comportamento procurado for encontrado, será disparado um alarme. Ao disparar o alarme as trajetórias envolvidas são detectadas e exibidas.

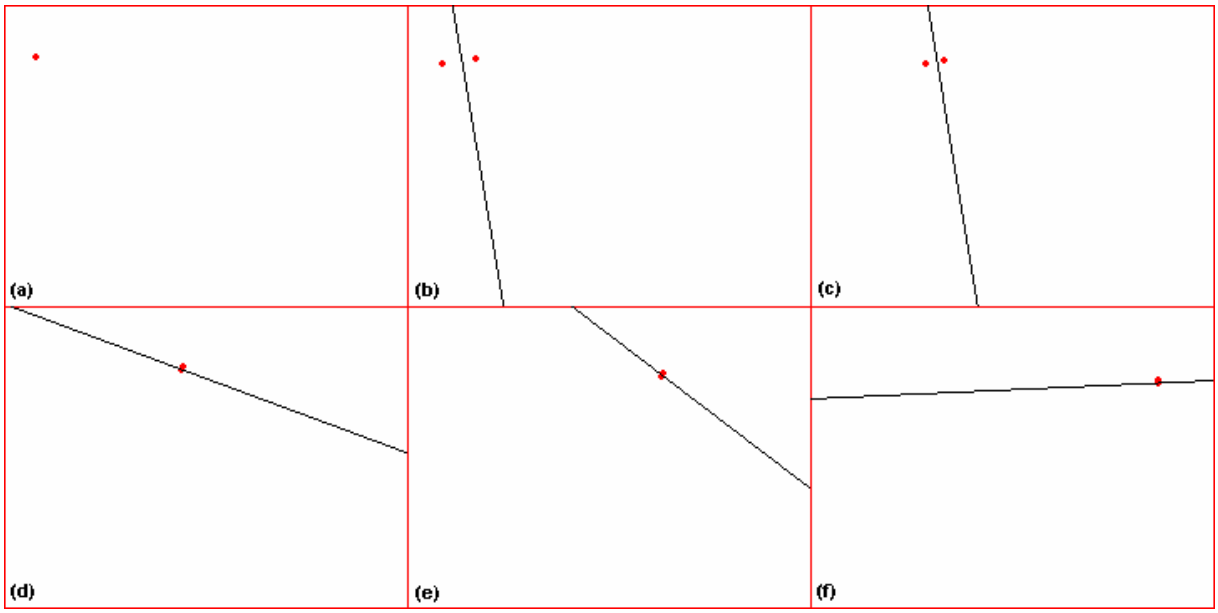

Figura 3.10 – Exemplo de resultado de uma consulta "*PA N*", ou seja, aproximação por trás seguida de agrupamento íntimo. São exibidos vários quadros do DVD.

Na figura 3.10, é exibido um exemplo de uso do autômato, onde é procurado por um comportamento que pode representar um seqüestro. Durante o desenrolar do vídeo, foram procurados por comportamentos especificados pela consulta: *PA N*, ou seja, aproximação por trás seguida de agrupamento íntimo. A Figura 3.10 exibe alguns quadros da seqüência onde a consulta foi verdadeira, e as pessoas envolvidas (*sites* do DV, marcados como círculos) foram pintadas de vermelho. A linha mostrada entre os nodos representa uma aresta do polígono de Voronoi.

3.2.2. Autômato de Conversão

Ao invés de informar-se um autômato para procurar por um comportamento, uma descrição do comportamento composta por símbolos pode ser informada ao programa para realizar a consulta. O programa faz a tradução automática dessa sentença para um AFND equivalente. Além da sentença ser composta pelos símbolos da tabela 2, ela pode conter expressões entre parênteses e expressões condicionais exclusivas através do operador | (ou).

Foi criada uma gramática para descrever essas seqüências de comportamentos que podem ser consultadas pelo programa. Portanto, através dessa gramática, o programa recebe a seqüência de comportamentos do usuário e a converte para o autômato que realizará a consulta ao histórico. Para tal conversão ser possível, foi criado um autômato especial que tem a função de realizar tal conversão.

Uma sentença da gramática corresponde á um determinado comportamento a ser procurado. Entretanto, esse comportamento é composto de uma seqüência de sub-comportamentos em ordem, sendo cada sub-comportamento um conjunto de símbolos concomitantes. Cada conjunto de símbolos (sub-comportamento) deve especificar um tipo de movimento válido, ou seja, ele deve conter ou um símbolo de movimento (*P* ou *F*) ou um símbolo de agrupamento (*N, S, C* ou *L*), mas nunca mais de um desses símbolos.

Em um sub-comportamento, quando o movimento for de aproximação (*P*) ou afastamento (*F*), pode-se concatenar a esse sub-comportamento o tipo de aproximação/afastamento: (*E*) para aproximação/afastamento pela frente ou (*A*) para aproximação/afastamento por trás, mas nunca ambos. Se nenhum desses símbolos for informado, a avaliação do autômato resultante será indiferente a direção da aproximação/afastamento.

Outros símbolos que podem ser adicionados a um sub-comportamento são os símbolos do SOM, portanto, (*V*) ou (*I*) podem ser adicionados a um sub-comportamento para determinar o tipo de sensibilidade ao SOM que esse sub-comportamento requisitará para ser reconhecido. Se nenhum desses símbolos for informado, o autômato resultante avaliará o sub-comportamento indiferentemente ao SOM. Em outras palavras, os símbolos do SOM permitem a integração da análise da ocupação espacial com o método apresentado que estuda as relações interpessoais das pessoas.

Por exemplo, um comportamento suspeito procurado pode ser a seqüência de dois sub-comportamentos: o movimento de encontro entre duas pessoas numa área com SOM válido *(PV)* e depois ambas passam a andar em agrupamento íntimo numa região de ocupação espacial baixa *(NI)*, nesse caso, formando uma sentença final *(PV NI)*.

Além dos símbolos do SOM, têm-se os símbolos para teste da transformada distância. Da mesma forma que o teste do SOM, pode-se restringir a busca por sub-comportamentos de acordo com a aceitação do teste da transformada distância, utilizando a letra *(T)* para TD válida e *(D)* para TD inválida. Em um sub-comportamento, se esse teste for omitido, a avaliação do autômato será indiferente ao teste da transformada distância.

Para que a consulta dirigida por uma sentença seja executada na base de dados, o sistema realiza a conversão da sentença da gramática para um AFND equivalente. Essa conversão exige a aplicação de um analisador (tradutor) que recebe a sentença e converte-a em autômato. Para ser possível tal tradução, foi necessário desenvolver a seguinte linguagem formal para descrever as sentenças:

```
G → E G
  | E
  | O G
  | O

E → T
  | "(" G ")"

O → E "|" E

T → x ; x ∈ R
```

A gramática utilizada no autômato de conversão é a seguinte: *G = (V, Z, P, S)*, onde:

- ***Não-terminais: V** = {G, E, O, T}*
- ***Terminais: Z** = { | , (,) } ∪ R, onde R representa todas as combinações válidas de símbolos da tabela 2}*
- ***Derivações: P** = {todas as produções apresentadas na linguagem acima}*
- ***Terminal Inicial: S** = {G}*

Na gramática apresentada, os não-terminais *G* e *E* permitem derivar múltiplas produções, porém só uma derivação pode ser escolhida como derivação. Na gramática, os símbolos entre aspas são elementos terminais da gramática, ou seja, pertencem ao conjunto de terminais *Z*. O não-terminal *T* representa todas as combinações válidas de símbolos da tabela, ou seja, um elemento de *Z*.

O processamento do analisador (tradutor) é feito em três etapas. Na primeira etapa, a partir da sentença informada pelo usuário, o analisador léxico gera a lista de *tokens* correspondente. Cada *token* representa um elemento terminal da gramática pertencente a *Z*. Por exemplo, para a sentença "*(PA | FA) NI"*, são gerados os seguintes tokens em ordem "*{ (, PA , | , FA ,) , NI}"*.

No próximo passo, o analisador sintático recebe a lista de *tokens* e gera a árvore de derivação, que é a entrada para o próximo passo. Para criar a árvore de derivação, o não-terminal inicial foi derivado sucessivamente até obter a sentença analisada. Nesse processo, todos os não-terminais que foram derivados foram inseridos de forma estruturada na árvore de derivação e os elementos terminais são inseridos como nodos-folha da árvore.

Na análise semântica, a árvore de derivação é percorrida, e enquanto que ela é percorrida, o autômato vai sendo gerado através de um conjunto de ações semânticas, que estão atreladas aos elementos não-terminais. As ações semânticas são pequenos comandos de um programa que é executado para gerar o autômato correspondente à sentença.

Na tabela 3, é mostrada a gramática com suas ações semânticas. Cada linha da tabela contém um não-terminal e sua derivação. Os comandos (ações semânticas) da coluna da esquerda são executados antes da avaliação (execução) dos não-terminais que compõem essa produção (gerando atributos propagados), e as ações semânticas da coluna da direita são executadas após a avaliação de todos os não-terminais que compõem essa produção (gerando atributos sintetizados).

G :- E G_1
 	
E.psef=falso
G1.psef= G.psef
	G.tr=E.tr
	G.tf=G_1.tf
	
	Para cada E.tf
	  Para cada G_1.tr
	  Cria Transição : T(G_1.tr.co, E.tf.eo)= G_1.tr.ed

G :- E
 	
E.psef= G.psef
	G.tr=E.tr
	G.tf=E.tf
	
	se G.psef
	  Para cada G.tf.eo
	  Define G.tf.eo como estado final do autômato

G :- O G_1
 	
O.psef=falso
G1.psef= G.psef
	G.tr=O.tr
	G.tf=G_1.tf
	
	Para cada O.tf
	  Para cada G_1.tr
	  Cria transição : T(G_1.tr.co, O.tf.eo)= G_1.tr.ed

G :- O
 	
O.psef= G.psef
	G.tr=O.tr
	G.tf=O.tf
	
	se G.psef
	  Para cada G.tf.eo
	  Define G.tf.eo como estado final do autômato

E :- T	Cria Estado E1 Cria Estado E2 Cria Comportamento C a partir do valor de T Cria Transição1 : T(C,E1)=E2 Cria Transição2 : T(C,E2)=E2 E.tr=Transição1 E.tf=Transição2
E :- "(" G ")" G.psef=E.psef	E.tr=G.tr E.tf=G.tf

O :- E_1 "|" E_2
 	
E1.psef=O.psef
E2.psef=O.psef
	Para cada E_1.tf
	  Para cada E_2.tr
	  Cria transição : T(E_2.tr.co, E_1.tf.eo)= E_2.tr.ed
	
	Para cada E_2.tf
	  Para cada E_1.tr
	  Cria transição : T(E_1.tr.co, E_2.tf.eo)= E_1.tr.ed
	
	O.tr=E_1.tr+E_2.tr
	O.tf=E_1.tf+E_2.tf

Tabela 3: Ações semânticas

A execução das ações semânticas é a execução de um programa que gera o autômato. A cada não-terminal que ele percorre foram vinculados atributos, que são as variáveis desse programa. Essas variáveis permitem tanto propagar atributos para dentro de outros não-terminais, como receber valores sintetizados pelos mesmos. Os atributos vistos na gramática acima estão listados abaixo:

- O atributo *"psef"* pode ter valor *verdadeiro* ou *falso* e é associado a qualquer não-terminal *X* da gramática. Ele indica se *X* e seus filhos na árvore de derivação podem ser representados como estados finais do autômato. Esse atributo é falso quando *X* é interno á outros nodos na árvore, assim como seus filhos. No início da análise, esse atributo é definido como *verdadeiro* no *G* raiz da árvore.
- *"tr"*, que significa *"transições raiz"*, é um vetor de transições que representa todas as transições que levam a essa própria seção do autômato criada pelo não-terminal *X*.
- *"tf"*, que significa *"transições folha"*, é um vetor de transições que representa todas as "últimas transições" que ocorrem nessa própria seção ao autômato criada pelo não-terminal *X*.

Os atributos *"transições raiz"* e *"transições folha"* armazenam vetores (ou listas) de transições. Cada transição tem 3 atributos, pois uma transição é uma função que recebe um comportamento e um estado origem e retorna o estado destino do autômato. A seguir são descritos os atributos das transições:

- *"co"* representa um comportamento lido numa transição.
- *"eo"*, representa o estado origem de uma transição.
- *"ed"*, representa o estado destino de uma transição.

Dessa forma, ao término do processamento do autômato de conversão, tem-se o autômato correspondente a sentença informada. O autômato gerado analisará o histórico ou os dados que estão sendo obtidos em tempo real pelos DVDs. A cada novo quadro é gerado um DV e através dele o comportamento interpessoal (e individual) das pessoas é identificado. O autômato gerado utilizará essa base de dados como fita, sendo um autômato por pessoa. Quando o autômato chegar a um estado final, ele disparará um alarme pois um evento suspeito foi identificado.

Nesse trabalho, o uso dos mapas de ocupação espacial permite quantificar o quanto são incomuns as trajetórias que foram desenvolvidas em áreas de caminhamento não-usuais. Nos

trabalhos de Junejo [JUN 2004] e Makris e Ellis [MAK 2005], as áreas usuais correspondem ás áreas internas aos envelopes de caminhamento detectados pelo modelo, pois em tais técnicas, os envelopes de caminhamento são formados pelo agrupamento das trajetórias semelhantes em forma e localização. Em ambas abordagens, não é quantificado o grau de *usualidade* de cada trajetória que é comparada com os envelopes de caminhamento. Entretanto, nessa dissertação, foram apresentadas duas técnicas distintas e complementares para avaliar as trajetórias individualmente: a análise do SOM (que concentra informação na região usual da cena) e a análise da TD (que concentra informação na região não-usual).

No trabalho apresentado por Stauffer e Grimson [STA 2000], as trajetórias com forma, posição e velocidade semelhantes são agrupadas em uma matriz de co-ocorrências, como forma de sintetizar os caminhos (padrões) mais utilizadas pelos objetos em movimento, e o tipo de móvel é identificado através de uma base de imagens hierárquica. Através dessas técnicas, algumas aplicações iniciais em detecção de movimentos suspeitos foram exploradas. O sistema está apto a detectar trajetórias não-usuais de acordo com a análise do padrão de forma, posição e velocidade de tais trajetórias através da matriz. Analisando por esse modelo, uma trajetória que apresente variações em algum desses parâmetros pode ser considerada não-usual, como uma trajetória de uma pessoa que andou normalmente, deu meia volta e retornou pelo mesmo caminho, pois esse tipo de situação tem uma natureza aleatória que altera a forma da trajetória, fazendo com que seja recusada uma trajetória que pode ser considerada usual. O teste do SOM é indiferente a essas variações de comportamento.

No trabalho de Cupillard et al. [CUP 2004], uma abordagem distinta é apresentada, onde, através de um modelo prévio da cena que descreve zonas, entradas, saídas e demais elementos, são procurados por comportamentos suspeitos no metrô (e nas catracas), através da análise de informações obtidas das pessoas na cena, como postura das pessoas, zona de localização, se a pessoa está parada ou em movimento, o grau de variação da forma do *blob* e outras informações. Porém o critério de ocupação espacial não é explorado, provavelmente por causa do tipo de cenário filmado e das aplicações desejadas. Em comparação, o SOM não considera a postura das pessoas, porém informa as regiões que não são utilizadas pelas pessoas para se locomoverem, que é uma informação relevante.

Nessa dissertação, além da análise individual das trajetórias, o estudo das relações interpessoais das pessoas permite determinar o comportamento de cada pessoa em relação a sua vizinha, adicionando mais um fator para classificar trajetórias além da ocupação espacial. Em outros trabalhos citados, como em Hosie et al. [HOS 98], são analisados fatores sociológicos, porém não são estudadas as características individuais das pessoas. O mesmo ocorre em

Fuentes e Velastin [FUE 2004], onde o conceito do proxemics é explorado para criar zonas de distância. Em contrapartida, nessa dissertação, a análise das características interpessoais das pessoas pode ser realizada em conjunto com a análise das características individuais das trajetórias obtidas pelo SOM através de um autômato.

Em alguns trabalhos apresentados, como o de Cupillard et al. [CUP 2004] e Lou et al. [LOU 2002], são empregados formalismos para definir linguagens de consulta sobre os dados obtidos na cena através da aplicação de seus modelos a fim de tornar possível consultar-se por certos comportamentos ocorridos. Nessa dissertação, um modelo abrangente é apresentado, onde a análise das características individuais do movimento através da análise da ocupação espacial e as características interpessoais do comportamento das pessoas são integradas através de um autômato, possibilitando explorar o conceito de *proxemics*, agrupamento e tipos de aproximação integrados com informação do SOM ou da TD.

O autômato realiza esssa integração e cria uma ótima flexibilidade para que sejam criadas novas consultas, pois ele permite que sejam especificados novos tipos de comportamentos a serem procurados na cena através da utilização de cada informação obtida através do modelo apresentado, ainda possibilitando que essas consultas possam ser alteradas mais tarde pelo operador de vídeo, sem que haja necessidade de realizar-se um novo treinamento para essa situação específica nova, nem havendo necessidade de atualizar-se a base de dados, pois o SOM é invariável a essa situação.

4. Resultados Experimentais

Nesse capítulo serão demonstrados resultados obtidos através de aplicações do modelo proposto, que foi implementado na linguagem de programação *Java*. Para melhor efeito visual, alguns gráficos foram desenhados com o software *Matlab*. Para facilitar alguns testes, foram geradas artificialmente algumas trajetórias.

Considere a cena ilustrada na Figura 3.1, que será utilizada nos próximos exemplos, nessa cena há várias áreas "caminháveis" (corredor e escada) e outras "não caminháveis" (estacionamento, jardim e gramado), conforme analisado no capítulo anterior.

Na figura 4.1, são apresentadas duas trajetórias: uma que pode ser considerada usual (a) (relacionada ao indivíduo A), e outra que pode ser interpretada visualmente como não-usual (b) (relacionada ao indivíduo B). A trajetória usual foi desenvolvida sobre o corredor e a escada, enquanto que a não-usual corresponde a uma pessoa que veio pelo corredor, subiu a escada, andou paralelamente ao jardim, e logo após, entrou no gramado. Embora esse comportamento também possa ser considerado usual, ele não está condizente com o período de treinamento. As linhas em vermelho representam as trajetórias desenvolvidas.

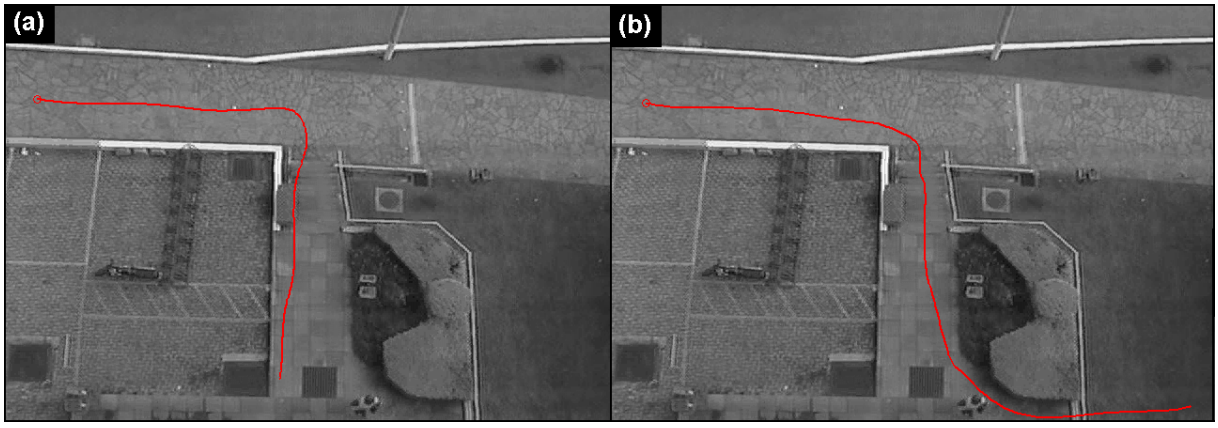

Figura 4.1 – Trajetória usual (a) e trajetória não-usual (b).

Na figura 4.2, ambas trajetórias são ilustradas sobrepostas ao SOM (a) e (b) e sobrepostas ao mapa da Transformada Distância (c) e (d). Para a geração do SOM, foi utilizado um desvio padrão $\sigma = 15$, como forma de representar o diâmetro médio de uma pessoa durante o treinamento. Para calcular o SOM apresentado a seguir, foram avaliadas 19 trajetórias no

período de treinamento sobre a cena. Porções das trajetórias consideradas usuais são mostradas em azul, enquanto que porções não usuais são marcadas em vermelho. Em particular, percebe-se nas Figuras 4.2 (b) e 4.2 (d) que o movimento do indivíduo B foi considerado não-usual quando ele entrou na área do gramado, tanto no teste do SOM quanto no teste da TD.

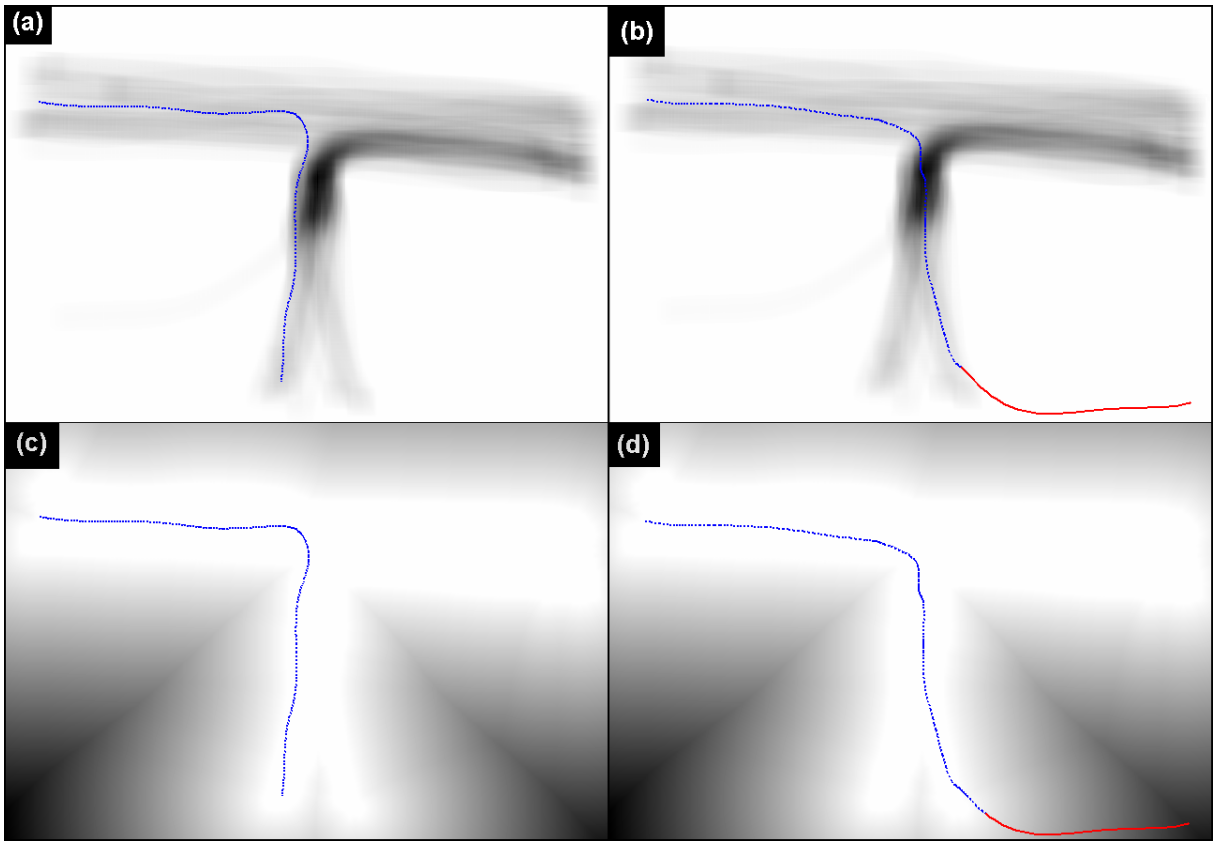

Figura 4.2 –Avaliação com SOM e com TD.

Na figura 4.3, os mesmos resultados são apresentados de forma diferente. Nas Figuras 4.3 (a) e 4.3 (b) são ilustradas, respectivamente, a evolução do SOM ao longo das trajetórias dos indivíduos A e B. As Figuras 4.3 (c) e 4.3 (d) mostram resultados análogos, mas para a evolução da Transformada Distância ao longo das trajetórias dos dois indivíduos. Novamente, parcelas em azul representam partes usuais das trajetórias, e parcelas vermelhas porções não-usuais. Em particular, percebe-se na Figura 4.3 (d) que o indivíduo B caminha boa parte de seu percurso em uma região válida do ambiente, mas no final sai dessa região e se afasta dela (pois o gráfico da TD aumenta). Outra observação é que na figura 4.3 (c) ilustra que a transformada distância é nula, pois o movimento ocorreu sobre a área de movimento válido. Como mencionado no

capítulo anterior, as informações do SOM e da TD são complementares: no interior do SOM válido, o SOM retorna a respectiva ocupação espacial (assim, pode-se detectar se a pessoa transitou por pontos de alta ocupação espacial), enquanto que q TD retorna zero (independente da ocupação espacial); por outro lado, no exterior do SOM válido, o SOM retorna zero (ou valores abaixo do limiar de aceitação $T_{SOM)}$, enquanto que a TD retorna exatamente a menor distância entre cada ponto da trajetória e o SOM válido.

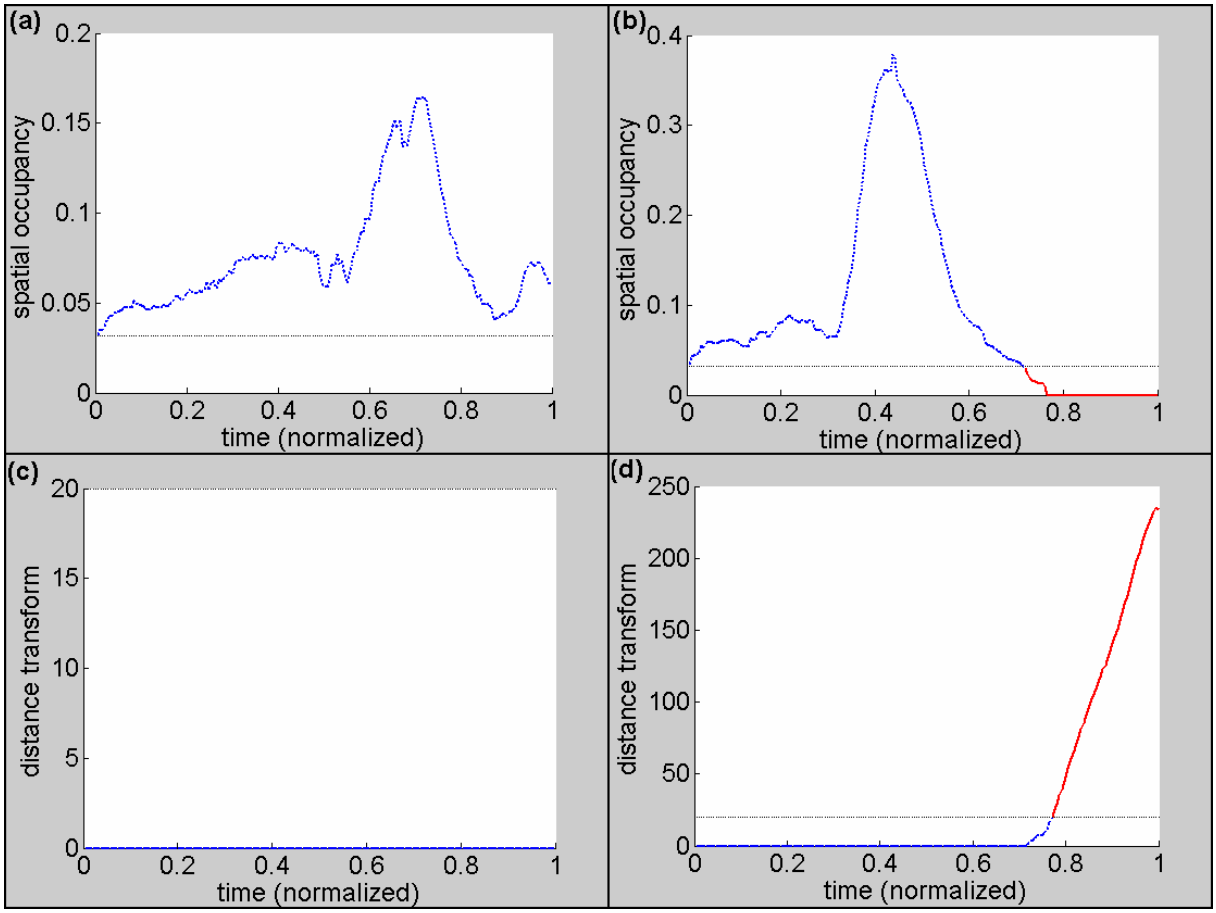

Figura 4.3 – Variação do SOM e da TD nas trajetória ilustradas na Figura 4.1.

A seguir são apresentados os resultados da implementação do modelo proposto sobre as consultas ao autômato finito, que podem conter informações sobre as relações interpessoais e/ou ocupação espacial. Deve-se salientar que esta etapa do método proposto retorna os resultados da consulta referentes a um padrão informado pelo usuário (como aproximação seguida de agrupamento, por exemplo). Assim, é o usuário quem deve determinar quais são os padrões que podem representar movimentos suspeitos/não-usuais, e inseri-los na busca.

Na figura 4.4, é apresentado o resultado de uma consulta que procura por um comportamento que tenha o padrão de aproximação por trás seguida de afastamento (que pode representar, por exemplo, um roubo). A sentença *"PA F"* representa o comportamento procurado, e é informada ao sistema para realizar a consulta, identificando em verde as ocorrências encontradas.

Nesse caso, nem sempre que for detectado na cena um comportamento de *"PA F"* pode ser considerado de fato um caso de roubo, pois o comportamento humano é subjetivo, e o alarme dado pelo sistema, poderia ser referente a um caso de dois amigos que se encontraram rapidamente e que se saudaram.

Por outro lado, ao invés de um encontro rápido, pode-se ter um encontro mais demorado, caracterizando um agrupamento, nesse caso, a expressão *"PA F"* poderia ser alterada para reconhecer esse caso, que é o caso de *"PA N F"*. E ainda, para manter as duas sentenças na consulta pode-se ter: *"(PA F) | (FA N FE)"*. Um dos usos do operador *ou*, representado por "|", é permitir expressar múltiplas sentenças. A sentença é convertida para um AFND, que lida corretamente com essas expressões, por ser não-determinístico.

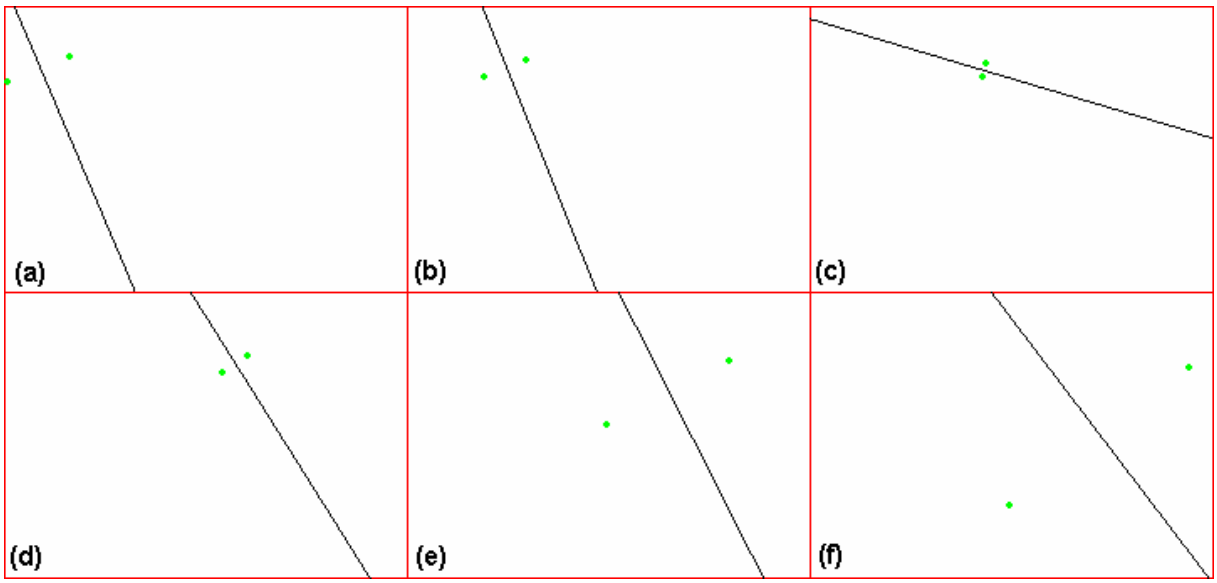

Figura 4.4 – Seqüência de quadros do comportamento *"PA F"*, que pode estar associado a um roubo.

Embora o modelo proposto neste trabalho seja focado para a detecção de comportamentos não-usuais, qualquer tipo de comportamento que possa ser descrito pela gramática proposta pode ser procurado. Por exemplo, procurar por encontros de amigos, ou

seja, o comportamento no qual ocorre a seguinte seqüência de sub-comportamentos: aproximação pela frente seguida de agrupamento íntimo seguido de afastamento pela frente. Para determinar quais pares de pessoas se enquadram nesse padrão, é feita uma consulta utilizando-se a seguinte sentença: *"PE N FE"*.

A figura 4.5 mostra uma seqüência de quadros onde a consulta foi satisfeita, sendo que as pessoas envolvidas são marcadas em verde. O movimento de aproximação entre os dois nodos em verde é mostrada nos quadros (a), (b) e (c), pois sua distância diminui com o tempo. Em seguida, o agrupamento pode ser percebido nos quadros (d) e (e), onde ambos nodos verdes mantém uma distância íntima entre si num tempo suficiente de acordo com o modelo. E por último, no quadro (f) percebe-se que houve afastamento entre os nodos.

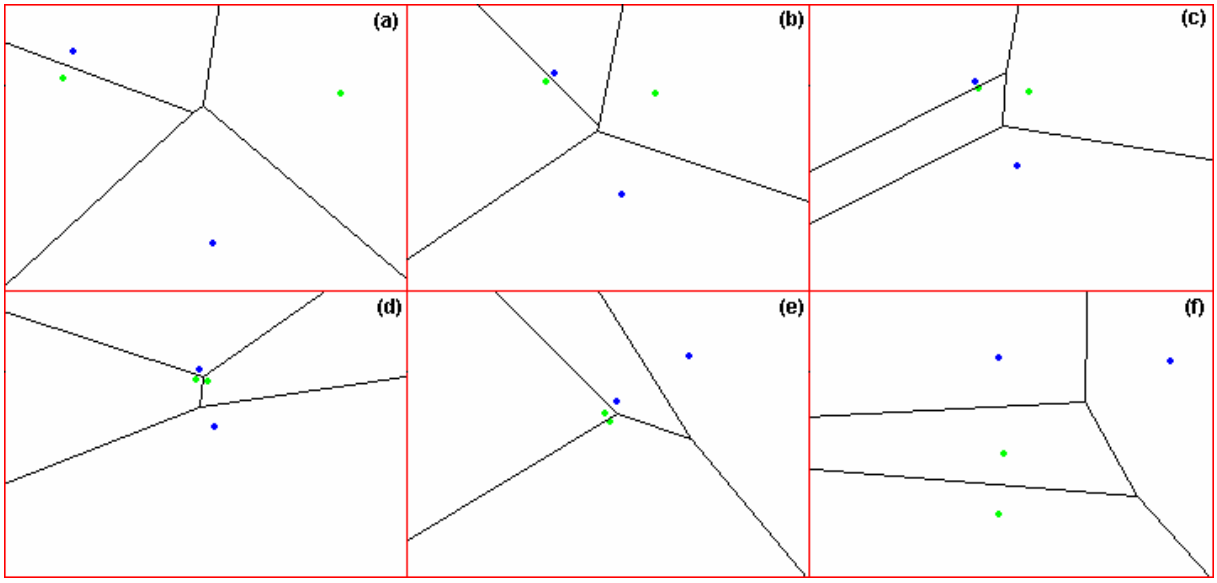

Figura 4.5 – Seqüência de quadros do comportamento: "encontro de amigos".

Outros elementos do modelo proposto podem ser agregados na consulta de maneira concomitante ou seqüencial, como os testes do SOM ou da TD. Assim, pode-se avaliar se os itens de consulta das relações interpessoais (aproximação, afastamento, agrupamento, etc.) estão ocorrendo em regiões válidas de acordo com o teste do SOM ou da TD. Por exemplo, a sentença: *"PA N (FI | FD)"* representa aproximação por trás, seguida de agrupamento íntimo, e, por fim, afastamento numa região de SOM não-usual ou afastamento numa região de TD não-usual. Tal sentença pode ser interpretada com uma possível situação de roubo.

A Figura 4.6 ilustra um resultado de busca com a sentença *"PA N (FI | FD)"*. Assim

como nos exemplos anteriores, as pessoas envolvidas na consulta são marcadas em verde, e as linhas representam arestas dos polígonos de Voronoi. Além disso, os resultados foram sobrepostos ao SOM, para uma melhor visualização dos resultados. Na figura 4.6 (a) e (b) percebe-se o movimento de aproximação entre as duas pessoas. Na figura 4.6 (b), (c) e (d) percebe-se o agrupamento formado pelas duas pessoas. Quando as duas pessoas se afastam, como visto na figura 4.6 (e) e (f), é disparado o alarme, pois ocorre o movimento de afastamento e uma delas está numa região de baixa ocupação espacial.

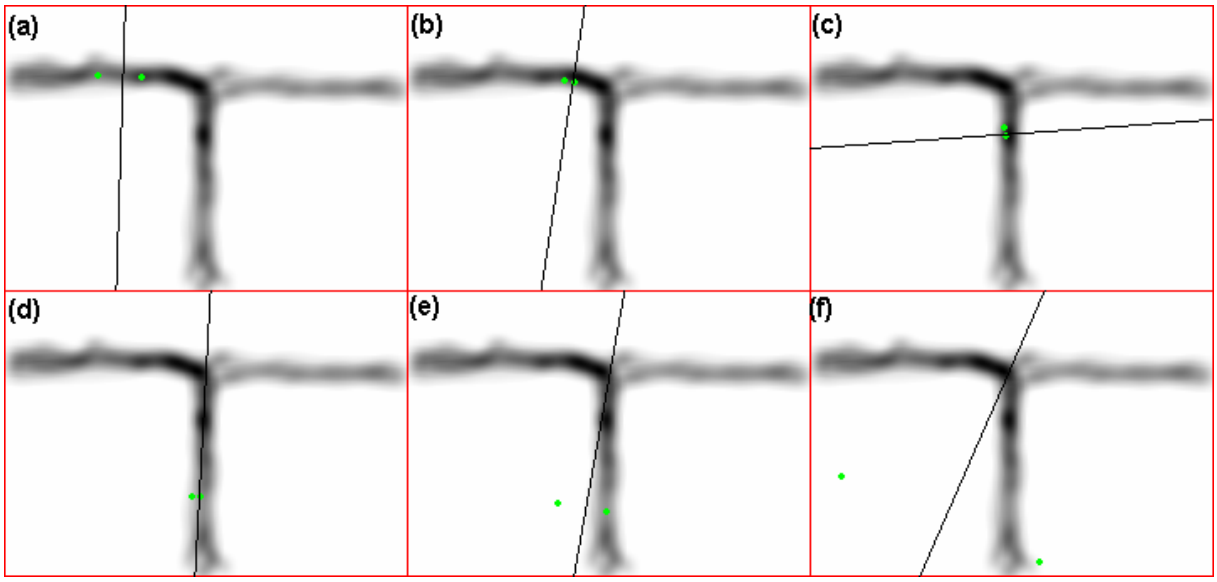

Figura 4.6 – Seqüência de quadros do comportamento "roubo" sobre o SOM.

A figura 4.7, mostra o mesmo exemplo ilustrado na figura 4.6, com a diferença que os DVDs são desenhados sobre a transformada distância. Na figura 4.6 (f) percebe-se com clareza que uma das pessoas está numa área escura da TD, representando uma distância grande à região ocupada (e, conseqüentemente, região não-usual de acordo com o teste da TD).

Os experimentos ilustrados nesse capítulo mostram uma total integração das técnicas apresentadas no modelo. Novas consultas podem ser realizadas apenas formando-se novas sentenças a partir da gramática apresentada no modelo. As buscas podem incluir comportamentos concomitantes (por exemplo, *"PV"* denota aproximação em uma região válida do SOM), seqüenciais (por exemplo, *"P N"* denota aproximação seguida por agrupamento íntimo), ou mesmo alternativas (por exemplo, *"P|N"* denota comportamento de aproximação ou de agrupamento íntimo). Portanto, buscas mais complexas podem ser feitas

pela concatenação desses comportamentos mais simples. Na figura 4.7, tem-se os mesmos dados da figura 4.6, com a diferença que estão sobrepostos sobre a transformada distância.

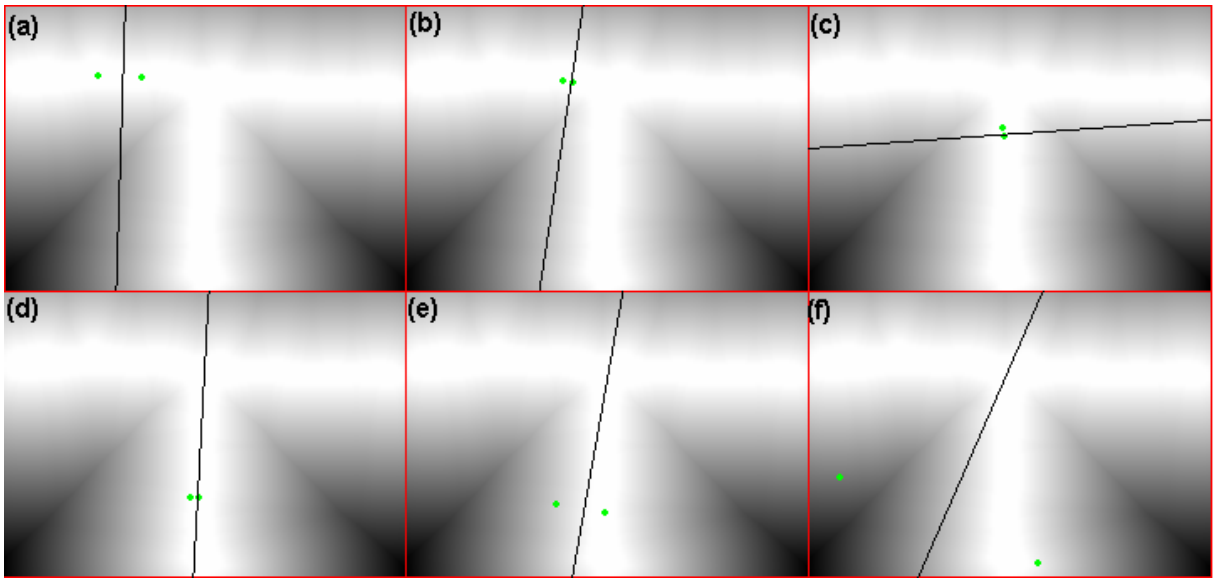

Figura 4.7 – Seqüência de quadros do comportamento “roubo” sobre a TD.

5. Conclusões e Trabalhos Futuros

Este trabalho apresentou um método para a detecção de movimentos não-usuais em seqüências de vídeo, baseado em dois critérios: a ocupação espacial e a relação entre as pessoas filmadas.

Para o critério de ocupação espacial, o método apresentado necessita de um período de treinamento para gerar uma base de dados que representa padrões de normalidade da cena no período filmado. No período de teste, cada nova trajetória é comparada com o banco de dados do treinamento, e porções não-usuais da trajetória são detectadas se a pessoa filmada trafegar em uma região pouco ocupada do espaço (teste do SOM), ou se a pessoa se afastar significativamente da região válida computada no período de treinamento (teste da TD).

Deve-se salientar que o período de teste é completamente dependente do período de treinamento, ou seja, o algoritmo deve ser aplicado em condições similares àquelas encontradas no período de treinamento. Dessa forma, o algoritmo irá retornar uma grande quantidade de alarmes em ambientes cujo "padrão de normalidade" varie ao longo do dia (ou em diferentes dias da semana). Por exemplo, a região em torno do restaurante em um campus universitário deve ter uma ocupação alta perto do meio-dia, mas baixa no período da tarde. Se o treinamento for realizado à tarde, o sistema retornaria diversos alarmos perto do meio-dia.

Para o critério das relações interpessoais, a posição de cada pessoa a cada quadro da seqüência de vídeo foi utilizada como um *site* na geração do Diagrama de Voronoi. E evolução dos Diagramas de Voronoi ao longo do tempo é explorada para calcular a variação temporal diversos parâmetros, como as distâncias entre vizinhos, o espaço pessoal, e o espaço pessoal percebido. Esses parâmetros, por sua vez, foram utilizados na detecção e classificação de grupos, além de fornecerem informações como afastamento ou aproximação entre pessoas, entre outros. A concatenação entre esses eventos de baixo nível pôde ser utilizada para detecção de eventos de mais alto nível (possivelmente suspeitos). Por exemplo, uma possível situação de roubo pode ser aproximação, seguida de agrupamento e após afastamento.

A busca pela concatenação dos eventos de baixo nível foi implementada através de um autômato finito, cuja sentença de busca é informada via uma gramática. No processo de busca, os eventos de ocupação espacial (testes do SOM e da TD) podem ser concatenados com os eventos de relações interpessoais (afastamento, aproximação, agrupamento, etc.), possibilitando buscas a sentenças complexas.

Os resultados experimentais obtidos foram considerados satisfatórios, embora não houvesse uma grande quantidade de seqüências de vídeo disponíveis. Pôde-se detectar com sucesso diversos tipos de busca, concatenando as informações de ocupação espacial e relações interpessoais. Entretanto, é importante salientar que a definição de comportamento não-usual é subjetiva e dependente de contexto, e todos os resultados apresentados neste trabalho se baseiam na noção de "normalidade" apresentada ao longo do texto.

Como trabalhos futuros, o método aqui apresentado pode sofrer diversas modificações e melhorias no intuito de detectar automaticamente comportamentos não-usuais. Por exemplo, pode-se desenvolver um modelo que suporta múltiplas hipóteses de padrões de normalidade no critério de ocupação espacial, de forma que o modelo apresentado se adapte com o respectivo período do dia no qual a câmera está gravando. Dessa forma, a base de dados seria direcionada de acordo com o período do dia, e o treinamento abrangeria todos os padrões diferentes de movimentação na cena. Ainda no contexto da ocupação espacial, poderia se incluir informações sobre a velocidade das pessoas na cena, de modo que movimentos contrários à grande maioria das pessoas poderiam ser detectados.

Também caberia realizar uma consulta a especialistas na área de segurança (como policiais), para determinar quais seqüências de movimentos podem geral um alarme de comportamento não-usual. Além disso, outro elementos poderiam ser incorporados à gramática atual, para fornecer uma gama maior de eventos.

Finalmente, deve-se enfatizar que o sistema proposto recebe como entrada as trajetórias computadas por um sistema de visão computacional, e a partir disso processa os dados de entrada. O seu processamento é em tempo real, ou seja, na medida que vai lendo cada quadro, aplica todos os passos de processamento descritos no modelo. Porém, a integração com o *Tracker* não está consolidada em tempo real. Outra possibilidade de trabalho futuro é realizar uma integração entre o *Tracker* e o sistema desenvolvido, de forma que o sistema completo (acompanhamento de pessoas juntamente com detecção de eventos) seja executado em tempo real.

Convém destacar que podem ser utilizados outros *Trackers* além do *Tracker* de câmera de topo, pois o modelo apresentado precisa apenas utilizar os resultados obtidos pelo *Tracker*, ou seja, as trajetórias. Por exemplo, um *Tracker* para câmera oblíqua pode ser utilizado desde que ele retorne corretamente a posição de cada pessoa em função do tempo indiferentemente à sombra ou variações de iluminação.

REFERÊNCIAS

[ADA 2006] ADAM, A.; RIVLIN, E.; SHIMSHONI, I.. *Robust fragments-based tracking using the integral histogram*. In CVPR '06: proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Volume 1, p. 98–805, Washington, DC, USA, Junho, 2006.

[BAR 94] BARRON, J. L.; FLEET, D. J.; BEAUNCHEMIN, S. S.. *Performance of optical flow techniques*. International Journal of Computer Vision (IJCV1994), Volume 12, p. 43-77, Fevereiro, 1994.

[BUX 97] BUXTON, H.; GONG, S.; *Advanced Visual Surveillance Using Bayesian Networks*. IEEE International Conference on Computer Vision, Cambridge, Massachusetts, p. 9/1-9/5, Junho, 1995.

[CHE 2006] CHENG F.; CHEN, Y.. *Real time multiple objects tracking and identification based on discrete wavelet transform*. Pattern Recognition, Volume 39, p. 1126–1139, Junho, 2006.

[CUC 2003] CUCCHIARA, R.; GRANA, C.; PICCARDI, M.; PRATI, A.. *Detecting moving objects, ghosts, and shadows in video streams*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 25, p. 1337–1342, Outubro, 2003.

[CUP 2004] CUPILLARD, F.; AVANZI, A.; BREMOND, F.; THONNAT, M.. *Video Understanding For Metro Surveillance*. The IEEE ICNSC 2004 in the special session on Intelligent Transportation Systems, Volume 1, p. 186-191, Taiwan, Março, 2004.

[DU 2006] DU, Y.; CHEN, F.; XU, W.; Li, Y.. *Recognizing Interaction Activities using Dynamic Bayesian Network*. 18th International Conference on Pattern Recognition, Volume 1, p. 618-621, Agosto, 2006.

[ELG 2002] ELGAMMAL, A. M.; DURAISWAMI, R.; HARWOOD, D.; DAVIS, L. S.. *Background and foreground modeling using nonparametric kernel density estimation for visual surveillance.* Proceeedings of the IEEE, Volume 90, p. 1151–1163, Julho, 2002.

[FUE 2004] FUENTES, L. M.; VELASTIN, A.. *Tracking-based event detection for CCTV systems.* Pattern Analysis and Applications, Volume 7, p. 356–364, Dezembro, 2004.

[FUN 2000] FUNG C.C.; JERRAT, N. *A Neural Network based Intelligent Intruders Detection and Tracking System using CCTV Images*. Proceedings of the IEEE Region 10 Conference on Intelligent Systems and Techologies for the Next Millennium (Tencon' 2000), Kaula Lumpur, Malaysia, Volume 2, p. 409-414, Setembro, 2000.

[HAL 73] HALL, E. T. *La Dimension Oculta - Tradução de Joaquin Hernandez Orozco do original: The Hidden Dimension*. Madrid: Instituto de estudios de administracion local, 1973.

[HAR 2000] HARITAOGLU, I.; HARWOOD, D.; DAVIS, L. S.. *W4: Realtime surveillance of people and their activities*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 22, p. 809–830, Agosto, 2000.

[HOS 98] HOSIE, R.; VENKATESH, S.; WEST, G.. *Classifying and Detecting Group Behaviour from Visual Surveillance Data*. IEEE Computer Society - ICPR '98: Proceedings of the 14th International Conference on Pattern Recognition, Washington, DC, USA, Volume 1, p. 602-604, Agosto, 1998.

[HU 2004] HU, W.; TAN, T. F.; WANG, L.; MAYBANK, S.. *A Survey on Visual Surveillance of Object Motion and Behaviors*, IEEE Transactions on Systems, Man, and cybernetics — Part C: Applications and Reviews, Volume 34, p. 334-352, Agosto, 2004.

[JAC 2006a] JACQUES, C. S. J. J.; JUNG, C. R.; SOLDERA, J.; MUSSE, S. R.. *Detection of Unusual Motion Using Computer Vision.* 19th Brazilian Symposium on Computer Graphics and Image Processing, 2006 - SIBGRAPI '06, p. 349-356, Outubro, 2006.

[JAC 2006b] JACQUES, C. S. J. J.. *Utilizando Visão Computacional para Simular e Validar Comportamentos de Multidões de Humanos Virtuais*. Dissertação (Mestrado) — Universidade do Vale dos Sinos, 2006.

[JAC 2006c] JACQUES, C. S. J. J.; JUNG, C. R.; MUSSE, S. R.. *A background subtraction model adapted to illumination changes.* In IEEE International Conference on Image Processing, p. 1817–1820. Atlanta, USA, Outubro, 2006.

[JOR 2004] JORGE, P. M.; MARQUES, J. S.; ABRANTES, A. J.. *On-line Tracking Groups of Pedestrians with Bayesian Networks*. 6th International Workshop on Performance Evaluation for tracking and Surveillance (PETS, ECCV), p. 65-72, Prague, Maio, 2004.

[JUN 2004] JUNEJO, I. N.; JAVED, O.; SHAH M.. *Multi Feature Path Modeling for Video Surveillance*. icpr, 17th International Conference on Pattern Recognition (ICPR'04), Volume 2, p. 716-719, Agosto, 2004.

[KAE 2003] KAEWTRAKULPONG, P.; BOWDEN, R.. *A real time adaptive visual surveillance system for tracking low-resolution colour targets in dynamically changing scenes*. Image and Vision Computing, Volume 21, p. 913-929, Setembro, 2003.

[LIU 2006] LIU, X.; CHUA, X. S.. *Multi-agent activity recognition using observation decomposed hidden markov models.* Image and Vision Computing, Volume 24, p. 166–175, Fevereiro, 2006.

[LOU 2002] LOU, J.; LIU, Q.; TAN, T.; HU, W.. *Semantic interpretation of object activities in a surveillance system*. Pattern Recognition, Volume 3, p. 777-780, Agosto, 2002.

[MAK 2005] MAKRIS, D.; ELLIS, T.. *Learning semantic scene models from observing activity in visual surveillance.* IEEE Transactions on Systems, Man, and Cybernetics, Part B 35, Volume 3, p. 397-408, Junho, 2005.

[MCK 2000] MCKENNA, S.; JABRI, S.; DURIC, Z.; ROSENFELD, A.; WECHSLER, H.. *Tracking groups of people.* Computer Vision and Image Understanding, Volume 80, p. 42-56, Outubro, 2000.

[MOE 2006] MOESLUND, T. B.; HILTON A.; KRUGER, V.. *A survey of advances in vision-based human motion capture and analysis*. Computer Vision and Image Understanding, Volume 103, p. 90-126, Novembro, 2006.

[NIU 2004] NIU, W.; LONG, J.; HAN, D.; WANG, Y.. *Human activity detection and recognition for video surveillance*. In Proceedings of the IEEE International Conference on Multimedia and Expo, ICME'04, Volume 1, p. 719-722, Taipei, Taiwan, Junho, 2004.

[OLI 2000] OLIVER, N.; ROSARIO, B.; PENTLAND, A.. *A Bayesian Computer Vision System for Modeling Human Interactions*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 22, p. 831-843, Agosto, 2000.

[PAI 2004] PAI, C. J.; TYAN H. R.; LIANG, Y. M.; LIAO, H. Y.; CHEN, S. W.. *Pedestrian detection and tracking at crossroads*. Pattern Recognition, Volume 37, p. 1025–1034, Setembro, 2004.

[PER 2006] PERERA, A. G. A.; SRINIVAS, C.; HOOGS, A.; BROOBSKY, G.; HU, W.. *Multi-Object Tracking Through Simultaneous Long Occlusions and Split-Merge Conditions*. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Volume 1, p. 666-673, Junho, 2006.

[PEU 2006] PEURSUM, P.; VENKATESH, S.; WEST, G.. *Observation-Switching Linear Dynamic Systems for Tracking Humans Through Unexpected Partial Occlusions by Scene Objects*. 18th International Conference on Pattern Recognition, 2006, ICPR 2006, Volume 4, p. 929-934, Agosto, 2006.

[POR 2006] PORIKLI, F.; TUZEL, O.; MEER, P.. *Covariance tracking using model update based on lie algebra*. In IEEE Computer Vision and Pattern Recognition, Volume 1, p. 728–735, Junho, 2006.

[SOI 2002] SOILLE, P. *Morphological Image Analysis*, 2002.

[STA 2000] STAUFFER, C.; GRIMSON, W. E.. *Learning Patterns of Activity using Real-Time Tracking.* IEEE Transactions on Pattern Analysis and Machine Intelligence, Artificial Intelligence Lab., MIT, Cambridge, MA, Volume 22, p. 747-757, Agosto, 2000.

[VAL 2005] VALERA, M.; VELASTIN, S. A.. *Intelligent distributed surveillance systems: a review*. IEEE Proc. Vis. Image Signal Process, Volume 152, p. 192-204, Abril, 2005.

[WAN 2003] WANG, J. J.; SINGH, S.. *Video analysis of human dynamics: a survey*. Real-time imaging, Volume 9, p. 321–346, Outubro, 2003.