Universidade Federal do Rio Grande do Sul
Instituto de Informática
Programa de Pós-Graduação em Computação

Detecção e Classificação de Mudança de Comportamento em Multidões Humanas

Igor Rodrigues de Almeida

Porto Alegre, 10 de Novembro de 2014

Universidade Federal do Rio Grande do Sul
Instituto de Informática
Programa de Pós-Graduação em Computação

Detecção e Classificação de Mudança de Comportamento em Multidões Humanas

Igor Rodrigues de Almeida

Trabalho de Conclusão do Mestrado em Computação submetido à avaliação, como requisito parcial à obtenção do título de Mestre de Computação.

Orientador(a): Prof. Dr. Claudio Rosito Jung

Porto Alegre, 10 de Novembro de 2014

Este trabalho foi analisado e julgado adequado para a obtenção do título de Mestre de Computação e aprovado em sua forma final pelo orientador.

_____

Prof. Dr. Claudio Rosito Jung

Banca Examinadora:

Prof. Dr. Claudio Rosito Jung
Inf – UFRGS (Orientador)
Prof. Dr. Marcelo Walter
Inf – UFRGS
Prof. Dr. Paulo Martins Engel
Inf – UFRGS
Prof. Dr. Soraia Raupp Musse
PUCRS

"Nas grandes batalhas da vida, o primeiro passo para a vitória é o desejo de vencer." Mahatma Gandhi

Agradecimentos

Meus sinceros agradecimentos ao meu orientador Prof. Dr. Claudio Jung pela orientação, auxílio e dedicação direcionados à realização deste trabalho.

Aos membros da banca pelo tempo dedicado ao trabalho.

Aos colegas de laboratório pela convivência quase diária.

Aos companheiros de Rotaract Club pela compreensão e apoio durante a realização deste projeto.

Aos amigos que auxiliaram durante o período da construção do trabalho.

Aos meu pai Ricardo, e a minha mãe Rozana, pela educação e princípios passados a mim, e por estarem sempre ao meu lado nesta longa caminhada.

Aos meus irmãos, e grandes amigos, Lucas e Thales, pelo apoio, companheirismo e por me aturarem nos momentos mais críticos no período de realização deste trabalho e sempre.

A todos aqueles que de alguma forma contribuíram para a realização deste trabalho.

Resumo

Este trabalho apresenta um método para detectar mudança de comportamento em multidões humanas baseado em histogramas de velocidade e orientação em coordenadas de mundo. Uma combinação de remoção de fundo e fluxo óptico é usada para extrair o movimento global a cada quadro do vídeo, descartando pequenos vetores de movimento devido artefatos como ruído, pixels de fundo não estacionários e problemas de compressão. Usando uma câmera calibrada, o movimento global pode ser estimado, e é usado para construir um histograma 2D contendo informações de velocidade e direção para todos os quadros. Cada quadro é comparado com um conjunto de quadros anteriores usando uma métrica de comparação de histogramas, resultando em um vetor de similaridade. Este vetor é então utilizado para determinar mudanças no comportamento da multidão, permitindo também uma classificação baseada na natureza da mudança no tempo: mudanças de curto ou longo prazo. Uma extensão do método apresentado é proposta utilizando técnicas de agrupamento para identificar diferentes grupos da cena, em seguida, aplicar o método de detecção em cada grupo. Isso proporciona não apenas detectar, mas também localizar a mudança de comportamento. O método foi testado em conjuntos de dados públicos disponíveis que envolvem cenários lotados.

Abstract

Detection and Classification of Changes in Behavior of Human Crowds

This work presents a method to detect change behavior in human crowds based on histograms of velocities in world coordinates. A combination of background removal and optical flow is used to extract the global motion at each image frame, discarding small motion vectors due artifacts such as noise, non-stationary background pixels and compression issues. Using a calibrated camera, the global motion can be estimated, and it is used to build a 2D histogram containing information of speed and direction for all frames. Each frame is compared with a set of previous frames by using a histogram comparison metric, resulting in a similarity vector. This vector is then used to determine changes in the crowd behavior, also allowing a classification based on the nature of the change in time: short or long-term changes. An extension of the presented method is proposed using clustering techniques to identify different groups in the scene, and then apply the detection method in each group. This provides not just detect but also localize the change behavior. The method was tested on publicly available datasets involving crowded scenarios.

Lista de Figuras

Conteúdo

Capítulo 1

Introdução

Desde a criação dos primeiros centros urbanos, e posteriormente o crescimento destes centros até os dias de hoje, observamos o crescimento de situações em que há multidões. Atualmente temos exemplos de eventos esportivos, shows, shopping centers, ruas com grande quantidade de comércios em horário comercial, boates, entre outras situações em que há alta densidade populacional. Uma multidão pode ser descrita como o comportamento coletivo de um grande número de agentes que interagem com um objetivo de grupo comum (OLFATI-SABER; MURRAY, 2004).

Porém, a multidão nem sempre flui tranquilamente: quando ocorre uma situação de emergência, a multidão pode desenvolver comportamentos anormais e criar uma situação de pânico. Com a grande ocorrência de situaçõesque envolvem multidões humanas, torna-se necessária a vigilância dos mesmos, para assim ser possível evitar que os problemas decorrentes de multidões, como esmagamento devido à alta densidade de pessoas e situações de pânico, possam causar grandes estragos. Neste trabalho apresentamos um método para detecção na alteração do comportamento da multidão, o que pode ser utilizado por especialistas humanos como indícios de situações anormais.

1.1 Motivação

A vigilância de ambientes não é algo simples, pois mesmo quando dotados de câmeras de vigilância, uma pessoa tem dificuldade de manter-se todo o tempo observando os vídeos gerados de fato, como demonstrado no estudo realizado pelo Instituto Nacional de

Justiça dos EUA (GREEN; LABS. ALBUQUERQUE, 1999), após somente 20 minutos, a maioria dos indivíduos testados ficam com a atenção abaixo do aceitável. Neste cenário, a vigilância automatizada pode ser empregada de forma a facilitar e prevenir situações de problemas sinalizando a um observador humano a ocorrência de algo anormal.

A vigilância automática de ambientes tem sido bastante pesquisada na área de visão computacional, pois bons detectores podem identificar rapidamente um problema ocorrido. Em uma subárea dos problemas de detecção de comportamentos está um dos maiores desafios atuais da visão computacional: a detecção de comportamentos em ambientes que envolvem multidões. Eventos esportivos, boates, centros comerciais e shows são exemplos de situações em que multidões aparecem com frequência, ou no interior, como nas boates, ou mesmo aos arredores, como o caso de grandes eventos esportivos. Em ambientes como estes, quando ocorrem situações de pânico na multidão, se não tratados rapidamente, podem causar grandes danos, enquanto que se detectados rapidamente, os danos podem ser prevenidos ou reduzidos.

A área de pesquisa de detecção de comportamentos em multidões ainda esta em aberto, com vários métodos sendo sugeridos recentemente, e com espaço para melhorias nos resultados atuais. Isto torna a detecção de comportamentos um problema interessante, pois além de real utilidade, não há métodos com resultados definitivos.

1.2 Objetivos

O objetivo principal do trabalho é detectar alterações de comportamento em multidões, como por exemplo, em situações decorrentes de pânico. Ainda objetiva-se localizar em qual grupo de pessoas, se houver mais de um, ocorre a alteração de comportamento. Neste trabalho assumimos que o ambiente é monitorado por uma câmera estática e calibrada.

Em ambientes com densidade alta de pessoas, é muito difícil identificar e rastrear cada pedestre de forma individual, sendo comum o uso do fluxo ótico para extrair informação de movimento. Além disso, quando câmeras estáticas são usadas, algoritmos de remoção de fundo são bastante úteis para identificar os objetos do primeiro plano. Salienta-se que ambos problemas (fluxo ótico e remoção de fundo) são problemas ativos nas comunidades de processamento de imagens e visão computacional, e este trabalho não visa contribuir

nessas duas áreas. Neste trabalho, assume-se que o fluxo ótico e a remoção do fundo são conhecidos.

Neste trabalho observamos o comportamento de multidões, que definimos como sendo formado por um ou mais grupos de pessoas. E estes grupos, por sua vez, determinamos como sendo um conjunto de pessoas que possuem características de deslocamento próximos, com velocidade e orientação de deslocamento parecidos entre os membros do grupo.

1.3 Organização do Trabalho

No Capítulo 2 são apresentados trabalhos que buscam a detecção de comportamentos não usuais em multidões, e que servirão como comparativo com o método apresentado nesta dissertação.

No Capítulo 3 é exposta a técnica proposta, que tem como objetivo determinar em qual quadro da sequência de vídeo ocorreu a mudança no comportamento da multidão, utilizando informações de movimento da mesma. Além de apresentar uma extensão do trabalho, que busca separar a multidão da cena em grupos e identificar alterações nestes grupos.

No Capítulo 4 são expostos os resultados obtidos com a técnica apresentada, e compara estes resultados com algumas técnicas apresentadas no Capítulo 2. Ainda são apresentados os resultados obtidos com a extensão apresentada no capítulo anterior.

Por fim, no Capítulo 5 são apresentadas considerações finais sobre o método proposto, e discorre-se sobre possíveis trabalhos futuros tendo em vista os resultados obtidos neste trabalho.

Capítulo 2

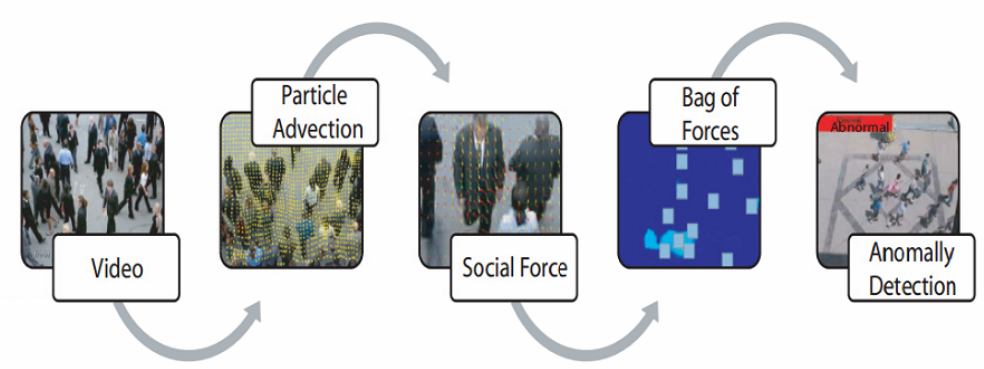
Revisão Bibliográfica

Há várias abordagens para analisar comportamentos de multidão utilizando visão computacional, que podem ser divididos em três classes principais: microscópicas, macroscópicas, e uma combinação dos dois (MEHRAN; OYAMA; SHAH, 2009). Na abordagem microscópica pessoas são analisados como indivíduos distintos, e esta informação é utilizada para inferir o comportamento da multidão. Na abordagem macroscópica a multidão, em vez disso, é analisada como uma única unidade, sem a detecção/rastreamento individual dos pedestres, que é uma maneira de evitar os problemas com a oclusão. Uma combinação de abordagens micro- e macroscópicas podem ser feitas, mantendo a multidão como uma massa homogênea, mas, ao mesmo tempo, considerando uma força interna. Outra maneira é manter as características das pessoas, mantendo uma visão geral de toda a multidão. A seguir, revisaremos alguns métodos de análise de multidão utilizando visão computacional.

Mehran, Oyama e Shah (2009) utilizam o modelo de forças sociais (*social force model*) para detectar e localizar comportamento incomum em cenas de multidões. Em sua abordagem, ilustrada na Fig. 2.1, as interações de partículas orientadas por uma média espaço-temporal do fluxo ótico são estimadas usando o modelo de forças sociais, e um conjunto de abordagens são adotadas para detecção de eventos incomuns (sequências de vídeo selecionados selecionados aleatoriamente são usados para modelar o comportamento normal).

Solmaz, Moore e Shah (2012) também utilizam fluxo ótico e conceitos físicos para avaliar multidões, mas explorando conceitos relacionados com a estabilidade de sistemas

Figura 2.1: O resumo da abordagem proposta para detecção de comportamento anormal nos vídeos multidão proposto por Mehran et al. Mehran, Oyama e Shah (2009)



Fonte: Mehran, Oyama e Shah (2009).

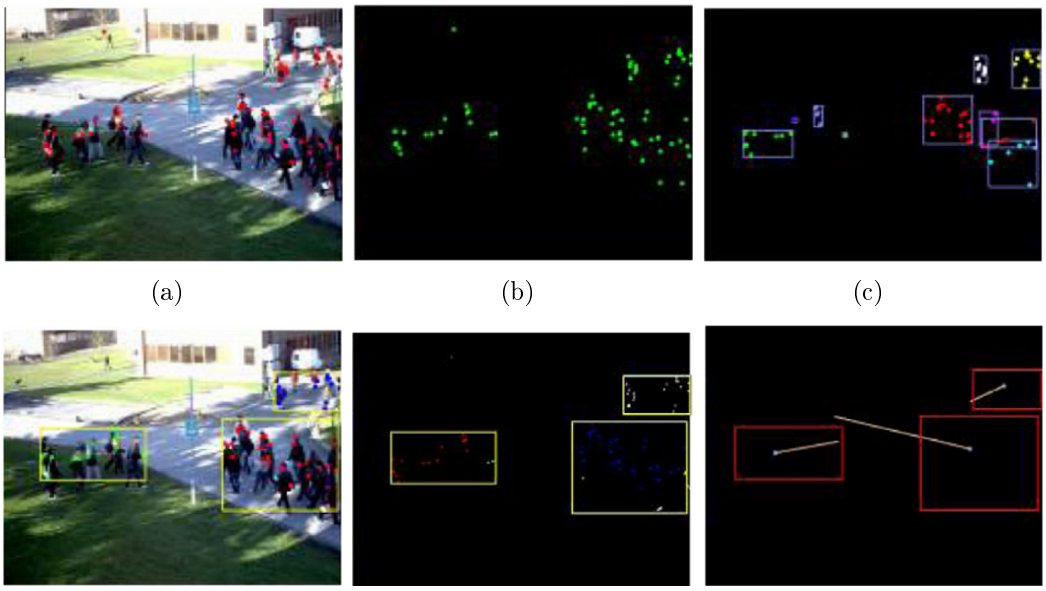
dinâmicos. Em particular, a análise dos autovalores da matriz Jacobiana, computada com o fluxo local em cada ponto, é usada para detectar eventos conhecidos em multidões reais.

Dee e Caplier (2010) apresentou um protótipo de sistema para a análise automatizada de cenas de multidões baseado em histogramas locais de vetores de movimento. Depois de detectar pedestres e faces para estimar a escala local, eles usam o rastreador KLT (SHI; TOMASI, 1994) para obter pedaços de trajetórias (*tracklets*), que são utilizados para estimar o movimento local. Os histogramas de movimento local são computados regionalmente pela divisão de cada quadro em um conjunto de regiões quadradas, e então os histogramas de cada quadro são comparados com os histogramas médios de um conjunto de treinamento.

Em (BROSTOW; CIPOLLA, 2006), Brostow e Cipolla usam um conjunto de dados sem supervisão impulsionado por algoritmo Bayesiano, que tem a detecção de entidades individuais como seu objetivo principal, para observar a informação de movimento dos indivíduos. Eles rastreiam características da imagem e as agrupam usando uma abordagem probabilística, onde cada grupo representa o movimento de uma entidade.

Chen e Huang (2011) usaram o fluxo ótico para agrupar multidões humanas em grupos de uma maneira não supervisionada usando uma abordagem chamada de *adjacency-matrix based clustering* (AMC), ilustrada na Fig. 2.2. Nesta abordagem, cada *cluster* é caracterizado com base no modelo de forças sociais, e eventos incomuns na multidão são detectados quando a orientação de uma multidão é abruptamente alterada ou quando a

Figura 2.2: (a) os pontos característicos detectadas; (b) apresenta pontos mostrados sem fundo; (c) Os aglomerados obtido após a aplicação da primeira fase de aglomeração e (d) agrupamentos obtidos por meio do algoritmo de AMC (e) fluxo ótico de cada *cluster* e (f) as forças dominantes destes *clusters*. (CHEN; HUANG, 2011)



(a) (b) (c)

(d) (e) (f)

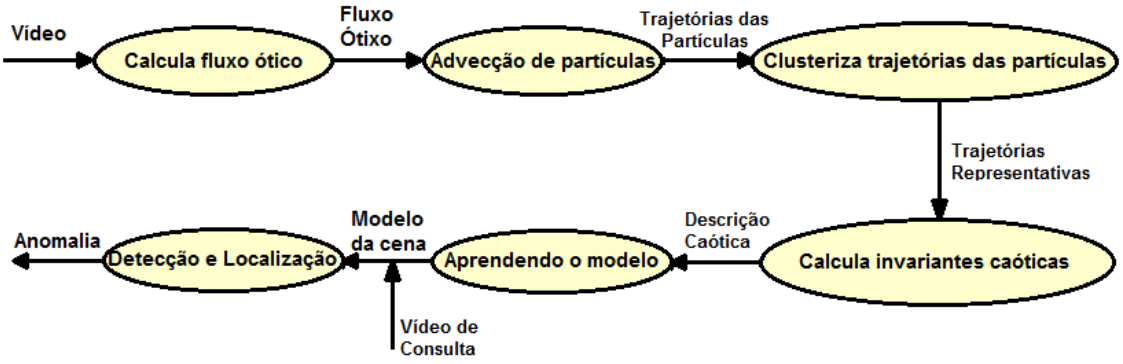
Fonte: Chen e Huang (2011).

interação entre os membros da multidão não é similar ao valor previsto.

Outro método não supervisionado e sem necessidade de treinamento é proposto por Hu (HU; ZHANG; DAVIS, 2013) para detecção de atividade anormal em multidões. Nesta abordagem, é feita a varredura de um vídeo com janelas de forma e tamanho variável. A anormalidade de cada janela é medido por um teste estatístico de razão de verossimilhança, que decide sobre entre duas hipóteses: as observações dentro e fora da janela são semelhantes ou não.

Briassouli e Kompatsiaris (2011) propuseram uma abordagem baseada no domínio de frequência para detecção de novos eventos em multidões, que não exige o cálculo do fluxo ótico explicitamente. Neste método, a distribuição do movimento da multidão é caracterizada no domínio de Fourier, e variações temporais são detectadas usando métodos estatísticos, como a Somas Cumulativa (*Cumulative Sum*, ou CUMSUM). Para incluir informação espacial na detecção, a abordagem é aplicada em diferentes blocos a cada quadro, que são avaliados de forma independente.

Figura 2.3: Estrutura para detecção e localização de anomalias proposta por Wu, Moore e Shah (2010).



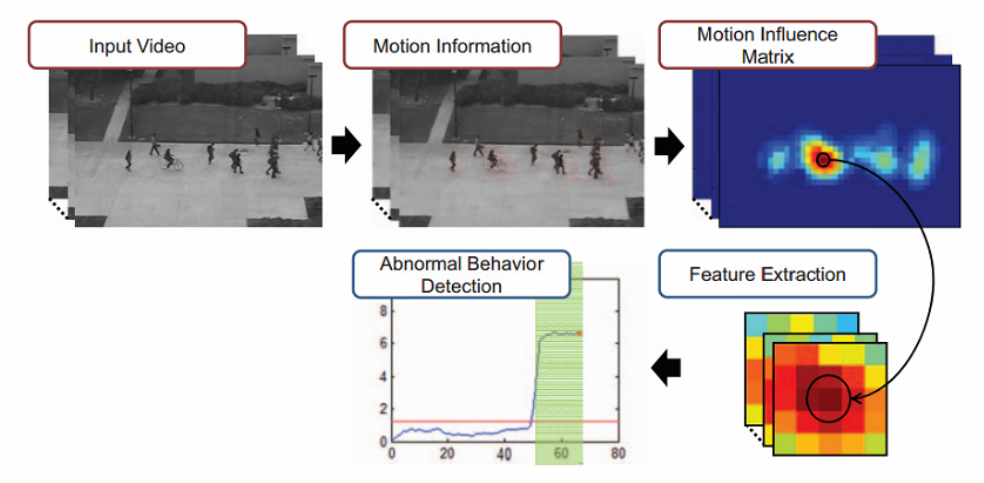
Fonte: Wu, Moore e Shah (2010).

Haque e Murshed (2010) também apresentaram uma abordagem que não se baseia nem em indicações do movimento, nem em trajetórias. Em vez disso, atributos temporais de componentes conexos do primeiro plano são extraídos, e a sua variação temporal, avaliada ao longo de uma janela temporal deslizante, é usada para classificar eventos pré-definidos usando Máquinas de Vetores de Suporte (*Support Vector Machines*, ou SVMs).

O método apresentado por Wu, Moore e Shah (2010) visa detectar e localizar anomalias em multidões complexas, utilizando uma abordagem dinâmica de partículas de Lagrange, em conjunto com modelagem caótica. Nesta abordagem, o fluxo ótico calculado em blocos espaço-temporais, e agrupados usando o algoritmo de aglomeramento K-Means. As trajetórias representativas são então usadas para estimar os expoentes de Lyapunov (usados para caracterizar sistemas caóticos), empregados como atributos para modelar probabilisticamente vídeos "normais" com base em um conjunto de treinamento. Finalmente, o teste de máxima verossimilhança é usado para detectar eventos não-usuais na fase de teste. Na Fig. 2.3 é ilustrada a abordagem utilizada por Wu.

Andersson et al. (2013) apresentam uma abordagem para detecção de padrões de movimentos anormais em multidões. Os autores usam K-Means para a identificação de grupos e modelos ocultos de Markov (HMMs - *Hidden Markov Models*) para modelar o padrão de movimento esperado de grupos densos e calmos. Embora os resultados experimentais apresentados sejam bons, os cenários usadas nos experimentos não são muito densos.

Figura 2.4: Uma visão geral do método proposto por Lee para a detecção de comportamento não-usual em multidões (LEE; SUK; LEE, 2013).



Fonte: Lee, Suk e Lee (2013).

Andrade, Blunsden e Fisher (2006) utilizam em seu trabalho método estatístico para reconhecimento de eventos, mais precisamente através de cadeias de Markov escondidas (HMM) (XIANG; GONG, 2005). A exemplo de Andrade, em (PATHAN; AL-HAMADI; MICHAELIS, 2010), Pathan et al. também utilizam método estatístico na modelagem de eventos, porém alternativamente as cadeias de markov escondidas, é utilizado CRF (*Conditional Random Field*) (LAFFERTY; MCCALLUM; PEREIRA, 2001).

Ullah, Ullah e Conci (2014) propõem uma abordagem para detectar anomalias em multidões baseada na observação das características de cantos. Para cada canto observado é adquirido características do movimento do mesmo. Estas características são utilizadas para treinar uma rede neural MLP na fase de treinamento, e o comportamento da multidão é inferido sobre as amostras de teste. Apesar de realizar a detecção em tempo real, a base de dados dos experimentos continha cenários não muito densos.

Lee, Suk e Lee (2013) utilizam em seu trabalho uma matriz de influência de movimento para representar comportamentos de multidão. A matriz de influência de movimento é utilizada para detectar comportamentos anormais na cena. Dois tipos diferentes de comportamentos anormais são abordados neste trabalho: comportamento anormal global e comportamento anormal local.

Cheng, Chen e Fang (2013) em seu trabalho apresenta um método para detecção e

localização de eventos anormais em multidões. É apresentado um detector de anomalias que estende o classificador de Bayes, a partir de classificação multi-classe para uma classe, para caracterizar eventos normais. Cheng propõe também um sistema de localização para localizar anomalias como um problema de subsequência máxima em uma sequência de vídeo. Greenewald e Hero (2014) propõe uma abordagem de aprender a regra de distribuição normal dos pixels multiquadros e detectar desvios a partir dela através de uma abordagem baseada em probabilidade. É utilizada uma abordagem de média e covariância e consideram métodos de aprendizagem da covariância espaço-temporal no regime de baixa amostra. A covariância é estimada utilizando a redução de parâmetros e modelos esparsos.

Recentemente, Li, Mahadevan e Vasconcelos (2014) atacaram o problema de detecção de localização de comportamentos anômalos em multidões, propondo um detector espaço-temporal. Tal detector é baseado em uma representação do vídeo que considera tanto aparência quanto a dinâmica de movimento, usando um conjunto de modelos de texturas dinâmico.

Como pode ser observado, a maioria das abordagens existentes exploram pistas de movimento da multidão. No entanto, quando os efeitos de perspectiva da câmera são significativos (câmera longe da configuração vista superior), os mesmos vetores de movimento em coordenadas de mundo podem mapear para diferentes vetores de movimento em coordenadas da imagem. Em (DEE; CAPLIER, 2010), um estimador de escala aproximada é realizado por detecção de pedestres, porém que também podem falhar quando efeitos de perspectiva são fortes. Além disso, a utilização de conjuntos de treinamento ou o conjunto pré-definido de eventos limita a aplicação prática destes métodos.

Capítulo 3

Técnica Proposta

Neste capítulo apresentamos um método para a detecção de comportamentos anormais em multidões em nível global, admitindo como sendo apresentada uma única multidão em cena, e em nível de grupos, buscando localizar espacialmente a ocorrência da anomalia usando informação de agrupamento.

3.1 Detecção de Eventos em Nível Global

Este método tem como objetivo determinar em qual quadro da sequência de vídeo ocorre uma mudança no comportamento da multidão. Para isto assumimos que todas as pessoas observadas na cena fazem parte de uma única entidade (um grupo) e analisamos a movimentação desse grupo. Como a detecção e rastreamento individual das pessoas em uma multidão é difícil devido à grande quantidade de oclusões, adotamos uma abordagem macroscópica da cena, organizando as características de movimento da multidão em histogramas, e classificando o comportamento comparando estes histogramas ao longo de uma janela temporal.

3.1.1 Codificando Informações de Movimento

Vamos considerar uma câmera de vigilância estática calibrada, e assumir que a região filmada é aproximadamente planar (o plano de chão é dado por $z = 0$). A abordagem proposta começa a extrair *blobs* do primeiro plano usando um algoritmo de remoção de fundo. Neste trabalho, optou-se por usar a técnica apresentada em (JUNG, 2009), por ser

facilmente codificada, apresentar bons resultados e remover sombras na detecção. Resumidamente, essa técnica cria um modelo de plano de fundo e estimativas locais do ruído usando as estatísticas robustas, de modo que ele pode ser aplicado mesmo quando um movimento forte está presente no conjunto de quadros utilizados para aprender o modelo de plano de fundo. A seguir, cada novo quadro da sequência é comparado com o modelo aprendido, e pixels são marcados como pertencentes ao primeiro plano se forem suficientemente diferentes do modelo de fundo. A seguir, técnicas de morfologia matemática são aplicadas para remover rúidos isolados, e uma abordagem estatística detecta e remove regiões com sombra.

Também estimamos o movimento da multidão na cena usando um algoritmo de fluxo ótico robusto (BROX; MALIK, 2011). Este método emprega uma abordagem variacional que pode lidar com grandes vetores de deslocamento, comuns em câmeras de vigilância de baixa taxa de quadros, gerando um campo vetorial $\boldsymbol{v}(\boldsymbol{x})$. Nesta técnica é estabelecida uma região hierarquica em duas imagens subsequentes. Correspondência de descritores nessas regiões fornecem um conjunto de hipóteses de possíveis correspondências. Estas são integradas em uma abordagem variacional que orienta a otimização local para grandes deslocamentos de pequenas estruturas móveis de forma independente, enquanto as outras restrições no modelo variacional fornecem a precisão dos métodos variacionais. Uma vez que o fluxo de vídeo apresenta artefatos (como o ruído, movimento de sombras, balançar de árvores, etc), estes geram uma série de vetores de movimento espúrios, e então restringimos a saída do fluxo ótico $\boldsymbol{v}(\boldsymbol{x})$ somente para pixels do primeiro plano.

Neste estágio, temos uma máscara binária $F(\boldsymbol{x})$ com pixels que pertencem ao primeiro plano (supostamente membros da multidão), juntamente com os correspondentes vetores de movimento $\boldsymbol{v}(\boldsymbol{x})$ (em pixels), obtidos com o fluxo ótico. Utilizamos estas informações combinadas, ignorando os vetores de movimento dos pixels onde a máscara binária determina que o pixel pertence ao palno de fundo. No entanto, duas pessoas que se deslocam com a mesma velocidade exata em coordenadas de mundo podem ter diferentes vetores de movimento estimados com fluxo ótico, devido à perspectiva da câmera. Para estimar os vetores de velocidade, em coordenadas globais, é necessário conhecer a matriz da câmera e a altura (coordenada z) de cada pixel, e, em seguida, calcular o mapeamento inverso da perspectiva para obter o vetor de movimento (x, y) no plano de chão (em metros), com

base em coordenadas de pixel (u, v). Uma vez que é difícil fornecer uma estimativa da altura do pixel (mas é plausível supor que se encontra na faixa de $[0, h_{max}]$, onde h_{max} é a altura máxima permitida para uma pessoa) , calculamos o mapeamento inverso utilizando a homografia do plano de chão (ou seja, assumimos que $z = 0$), em vista que a matriz de projeção é definida por:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \boldsymbol{K} \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \boldsymbol{P} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}_1 & \boldsymbol{p}_2 & \boldsymbol{p}_3 & \boldsymbol{p}_4 \end{bmatrix} \begin{bmatrix} x \\ y \\ 0 \\ 1 \end{bmatrix}. \tag{3.1}$$

Como utilizamos $z = 0$ então podemos escrever a equação 3.1 como:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}_1 & \boldsymbol{p}_2 & \boldsymbol{p}_4 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{3.2}$$

sendo:

$$\begin{bmatrix} \boldsymbol{p}_1 & \boldsymbol{p}_2 & \boldsymbol{p}_4 \end{bmatrix} = \boldsymbol{K} \begin{bmatrix} \boldsymbol{r}_1 & \boldsymbol{r}_2 & \boldsymbol{t} \end{bmatrix} \Rightarrow \boldsymbol{H} = \boldsymbol{K} \begin{bmatrix} \boldsymbol{r}_1 & \boldsymbol{r}_2 & \boldsymbol{t} \end{bmatrix}, \tag{3.3}$$

onde $\begin{bmatrix} \boldsymbol{r}_1 & \boldsymbol{r}_2 & \boldsymbol{t} \end{bmatrix}$ definem os parâmetros extrínsecos da câmera, $\boldsymbol{r}_1$ e $\boldsymbol{r}_2$ são as duas primeiras colunas da matriz de rotação $\boldsymbol{R}$ e $\boldsymbol{t}$ o vetor de translação, enquanto a matriz $\boldsymbol{K}$ determina os parâmetros intrínsecos da câmera. A partir da homografia $\boldsymbol{H}$, podemos definir a equação de projeção:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \boldsymbol{H} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \tag{3.4}$$

e encontrar a posição das coordenas de pixel (u, v) utilizando a projeção inversa:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \boldsymbol{H}^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}. \tag{3.5}$$

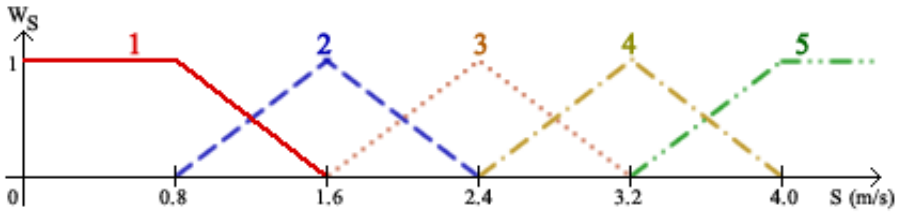
Embora o cálculo do mapeamento inverso usando $z = h_{max}/2$ pudesse potencialmente reduzir o erro de aproximação, adotamos $z = 0$, uma vez que a homografia do plano do chão pode ser calculada facilmente selecionando pontos do plano de chão com coordenadas no mundo conhecidas, ou com base no movimento de pedestres (BOSE; GRIMSON, 2003).

Tendo em conta o campo de vetores movimento $\boldsymbol{v}_w(\boldsymbol{x})$ em coordenadas do mundo em um determinado instante t, o movimento global da multidão é codificada por um histograma 2D, desassociando velocidade e orientação. Para cada pixel $\boldsymbol{x}$ relacionado a uma pessoa no meio da multidão, nós quantificamos a velocidade $s(\boldsymbol{x})$ (em m/s, estimado com base na taxa de quadros da sequência de vídeo) e a orientação $\theta(\boldsymbol{x})$ (em graus) em N_s e N_θ 'bins', respectivamente.

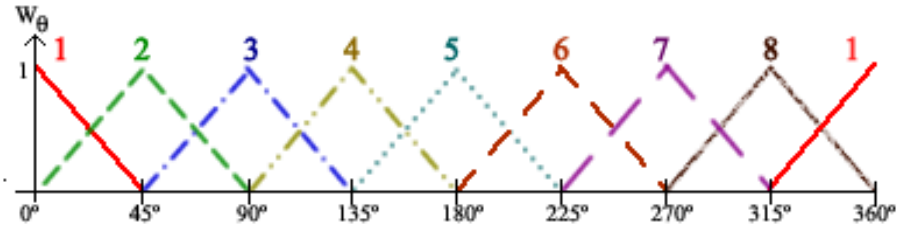
Optamos $N_s = 5$ para quantificar a velocidade em cinco classes: muito lenta, caminhando, caminhando rápido, correndo e correndo rápido, usando um tamanho de bin de $\Delta s = 1,6$ m/s, de modo que a transição entre " caminhando rápido" e " correndo" ocorre a uma velocidade de cerca de $2,4$m/s, em acordo com (ROTSTEIN et al., 2005). Deve-se salientar que se são usadas apenas as coordenadas da imagem, a definição dos tamanhos de 'bins' relacionados com a velocidade torna-se um problema complexo, especialmente quando os efeitos de perspectiva são mais visíveis. Quanto à quantização das orientações, definimos experimentalmente $N_\theta = 8$, com base em direções cardeais e ordinais, levando a um tamanho de bin orientação $\Delta\theta = 45°$.

Para reduzir a influência do ruído e dos problemas de quantização, primeiro estimamos a função de distribuição de probabilidade subjacente (*probability distribution function* - PDF), utilizando estimativa de densidade kernel (*Kernel Density Estimation* - KDE) (HWANG; LAY; LIPPMAN, 1994) em vez de calcular o histograma diretamente de $\boldsymbol{v}_w(\boldsymbol{x})$. No KDE, um núcleo centrado em cada observação é usado para obter uma PDF contínua dos dados, espalhando a sua influência ao longo de mais do que um bin do histograma. Utilizamos kernels triangulares em ambas as dimensões do histograma: velocidade e orientação. Os suportes das duas janelas triangulares foram definidos como o tamanho dos bins de velocidade e orientação (Δs e $\Delta\theta$, respectivamente), de modo que cada vetor movimento $\boldsymbol{v}_w$ com velocidade s e orientação θ contribui para o histograma H de acordo com

Figura 3.1: velocidade (a) orientação e (b) as funções de ponderação usar para obter os histogramas de movimento 2D.



(a)



(b)

Fonte: O próprio autor.

$$H(i,j) = \max\left\{0, 1 - \frac{|s - s_i|}{\Delta s}\right\} \cdot \max\left\{0, 1 - \frac{|\theta - \theta_j|}{\Delta \theta}\right\}, \tag{3.6}$$

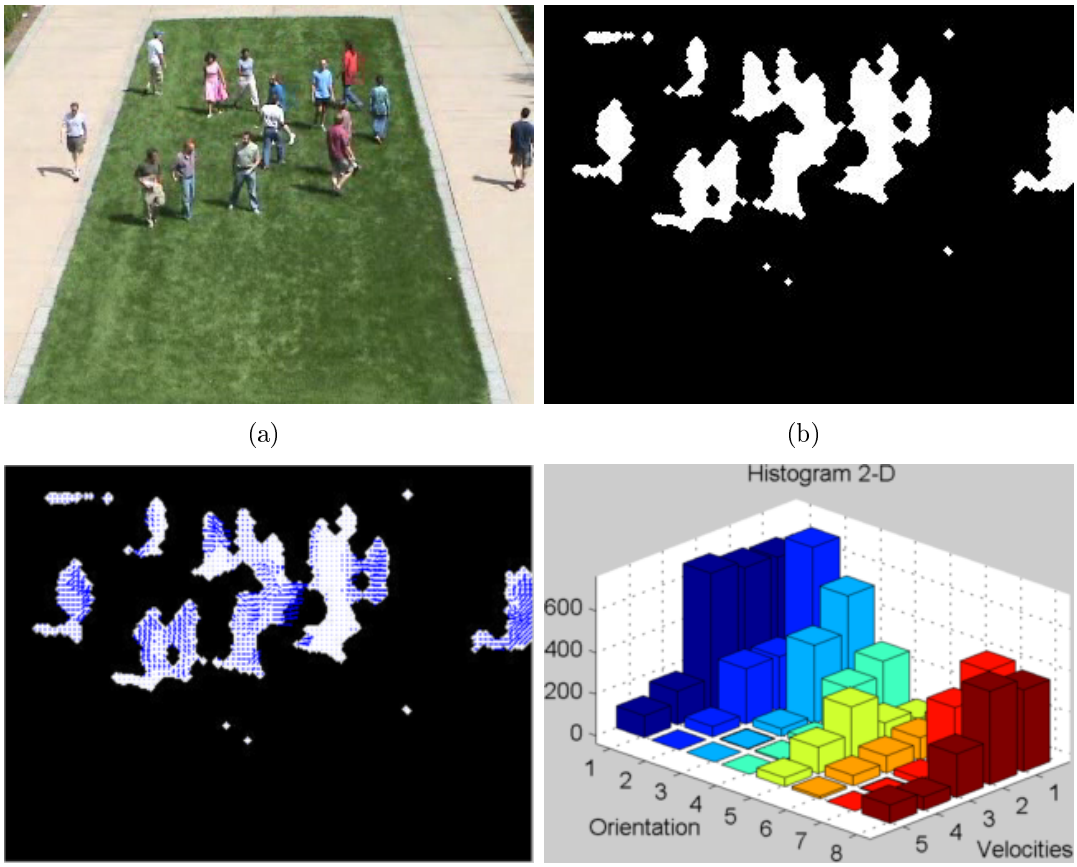
onde s_i w θ_j são os centros dos bins de velocidade e orientação, respectivamente, e i, j os índices dos bins no histograma. Pode ser observado que apenas as quatro bins mais próximos de (s, θ) apresentam pesos diferente de zero. Para fins de ilustração, as funções de ponderação triangulares em s e θ são mostrados na Fig. 3.1.

O pipeline para estimar os histogramas KDE-ponderados é ilustrado na Fig. 3.2. Um quadro típico é mostrado na Fig. 3.2(a), e o resultado da abordagem de remoção de fundo é ilustrado na Fig. 3.2(b). Os vetores do fluxo ótico computados para pixels do primeiro plano válidos são mostrados na Fig. 3.2(c), e o histograma 2D é ilustrado como uma superfície em Fig. 3.2(d).

3.1.2 Análise do Comportamento da Multidão

Para cada quadro t, calculamos o histograma orientação-velocidade de acordo com o procedimento descrito até agora (na verdade, para incluir algumas suavidade temporais nos histogramas, utilizamos amostras de ambos os quadros t e $t-1$ para construir o

Figura 3.2: (a) Quadro selecionado. (b) Pixels do primeiro plano. (c) Vetores do fluxo ótico do primeiro plano. (d) Superfície que ilustra o histograma 2D $H(i, j)$.

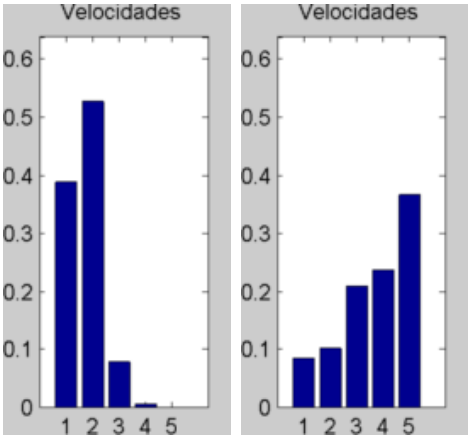


(a) (b)

(c) (d)

Fonte: O próprio autor.

Figura 3.3: histograma de velocidade da cena quando as pessoas estão (a) caminhando e (b) correndo.



(a) (b)

Fonte: O próprio autor.

histograma do quadro t). Em seguida, normalizamos os histogramas (de modo que possam ser tratadas como PDF discretos), obtendo-se os histogramas normalizados $H_t(i, j)$. Se o padrão de movimento da multidão permanece semelhante dentro de um período de tempo, os histogramas correspondentes deverão ser similares. Por outro lado, as alterações no comportamento da multidão espera-se que gerem discrepâncias quando se compara os histogramas, pois a distribuição nos histogramas tende a sofrer alterações, como por exemplo, na alteração do pico do histograma, como pode ser observado na Fig. 3.3 em que o quadro a esquerda mostra o histograma de velocidade da cena, quando as pessoas estão caminhando, enquanto que no quadro da direita é apresentado o histograma de velocidade quando na mesma cena as pessoas estão correndo.

Na abordagem proposta, em vez de gerar um conjunto de treinamento para aprender o "movimento usual", comparamos o movimento em cada quadro com os padrões de movimento de um conjunto de quadros anteriores. Mais precisamente, nós geramos um vetor de similaridade $\boldsymbol{S}_t$ dado por

$$\boldsymbol{S}_t = \Big(C(H_t, H_{t-\Delta t_1}),\ C(H_t, H_{t-\Delta t_2}),\ ..., C(H_t, H_{t-\Delta t_n}) \Big), \tag{3.7}$$

onde n é o número de frames anteriores utilizados na comparação, Δt_i é o intervalo entre os quadros, e C é a métrica de similaridade entre histogramas Embora existam

muitas possibilidades para C, utilizamos a correlação entre histogramas, pois este retorna valores dentro de um intervalo pré-determinado $[0, 1]$ o que facilita na manipulação destes resultados. Formalmente, a correlação entre histogramas H_1 e H_2 é dado por

$$C(H_1, H_2) = \frac{\sum_{i,j}(H_1(i,j) - \overline{H_1}) \sum_{i,j}(H_2(i,j) - \overline{H_2})}{\sqrt{\sum_{i,j}(H_1(i,j) - \overline{H_1})^2} \sqrt{\sum_{i,j}(H_2(i,j) - \overline{H_2})^2}} \tag{3.8}$$

onde $\overline{H}$ é a média dos valores de H.

O vetor de similaridade $\boldsymbol{S}_t$ é então analisado para detectar mudanças no comportamento da multidão, bem como classificá-lo em termos de quão rápido ele ocorreu. A estabilidade temporal, σ_t do comportamento da multidão no quadro t é definida como a média ponderada de $\boldsymbol{S}_t$:

$$\sigma_t = \boldsymbol{w}^T \boldsymbol{S}_t, \tag{3.9}$$

onde $\boldsymbol{w}$ é o vetor de pesos, que apresenta valores mais altos para quadros mais recentes. Nós utilizamos decaimento exponencial dos pesos

$$\boldsymbol{w} = \frac{1}{\sum_{i=1}^{n} e^{-\lambda \Delta t_i}} \left(e^{-\lambda \Delta t_1},\ e^{-\lambda \Delta t_2}, \cdots, e^{-\lambda \Delta t_n} \right), \tag{3.10}$$

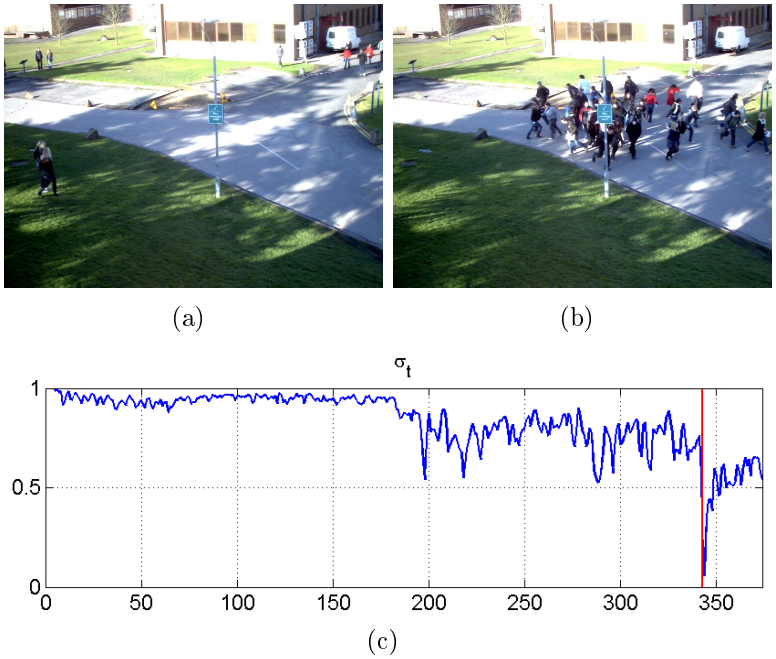
onde λ é a constante de decaimento. Em todos os experimentos, nós utilizamos $n = 14$ quadros para construir $\boldsymbol{S}_t$, e utilizamos $\Delta t_i = i\Delta t$, com Δt constante e definido em $0,57$s, logo avaliamos um período de tempo de 8s. O valor de Δt corresponde a 4 quadros quando a sequência é adquira a uma taxa de 7 quadros por segundo, que é o comum em câmeras de vigilância. A constante de decaimento λ foi definida experimentalmente em $\lambda = 0,52$.

Uma mudança no comportamento da multidão é detectada quando a estabilidade temporal σ_t é baixa, significando que a similaridade entre o frame atual e os anteriores é pequena. Mais precisamente, definimos um limiar adaptativo β_t baseado na historia de σ_t:

$$\beta_t = \frac{1}{2n} \sum_{i=1}^{n} \sigma_{t-\Delta t_i}, \tag{3.11}$$

ou seja, o limiar é metade da média dos valores de estabilidade temporal σ_t numa janela

Figura 3.4: Na imagem à esquerda é o primeiro quadro do vídeo e na imagem da direita o quadro onde a mudança de comportamento das multidões ocorreu. Abaixo mostramos os valores de σ_t em função do tempo, com uma linha vertical vermelha indicando a mudança detectada.



Fonte: O próprio autor.

temporal.

Fig. 3.4 mostra um exemplo de mudança de comportamento da multidão do conjunto de dados PETS2009 S3.Event Recognition[1], sequência 4. Nesta sequência, as pessoas entram em cena a partir de várias direções, e se reúnem no centro. Depois de algum tempo, todos eles começam a correr quase instantaneamente afastando-se do centro, cada um numa direção aleatória. Na parte inferior da Fig. 3.4 mostramos os valores de σ_t ao longo da sequência de vídeo. Podemos ver que σ_t oscila mesmo quando o comportamento das multidões não muda. Quando a mudança no comportamento das multidões ocorre, σ_t cai drasticamente, e a mudança de comportamento detectada é indicado pela linha vertical vermelha.

[1] <http://www.cvg.rdg.ac.uk/PETS2009>

3.1.3 Classificação da mudança em curto ou longo prazo

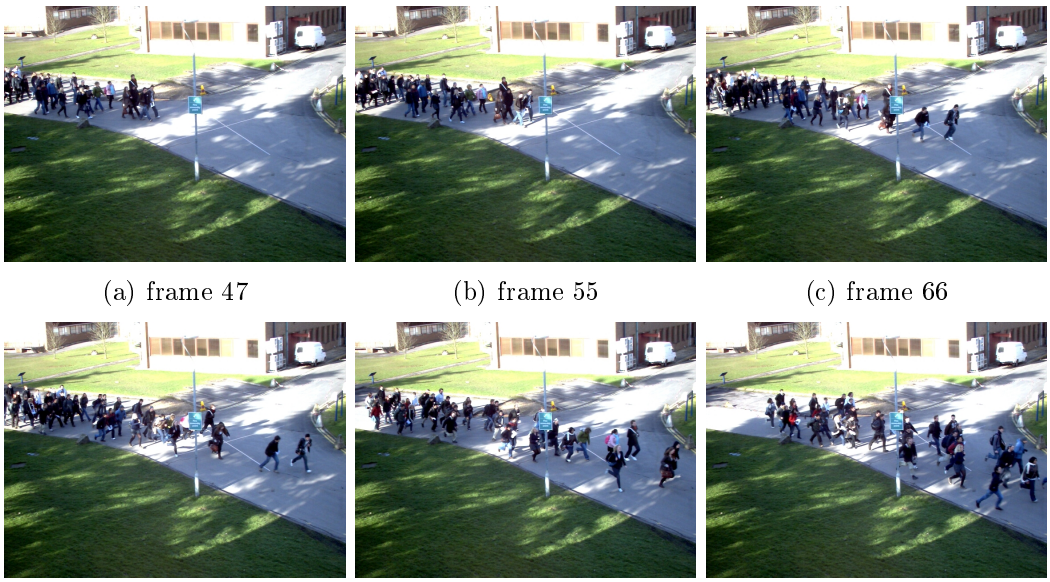
Quando uma mudança de comportamento é detectada, é possível classificá-la ainda com base em quão rápido ela aconteceu. Neste trabalho, definimos dois tipos de mudanças: as mudanças de curto ou de longo prazo. Mudanças de longo prazo ocorrem de forma gradual, o que significa que a semelhança do comportamento entre os quadros temporalmente próximos pode ser grande, mas diminui quando os quadros mais distantes são avaliados. Por exemplo, se um grupo está caminhando para a mesma direção e os membros na frente da multidão começam a correr (e, em seguida, os membros de trás, progressivamente), a mudança de comportamento será gradual (de longo prazo). Por outro lado, as mudanças de curto prazo ocorrem mais abruptamente, como em uma situação de pânico, em que todos os membros da multidão começam a correr repentinamente para diferentes direções. Fig. 3.5 mostra alguns quadros representativos de conjunto de dados PETS2009 S3.Event Recognition (sequência 1), em que as pessoas começam a correr para a mesma direção de forma progressiva (de longo prazo). Fig. 3.6 mostra alguns quadros do conjunto de dados PETS2009 S3.Event Recognition (sequência 4), relacionado a uma mudança de curto prazo (as pessoas de repente começam a correr em diferentes direções).

Em mudanças de longo prazo, vetor $\boldsymbol{S}_t$ tende a apresentar valores que diminuem suavemente (como quadros mais antigos são utilizados na comparação). No segundo caso, a maioria dos valores de $\boldsymbol{S}_t$ tendem a ser menores, uma vez que mudanças repentinas também afetarão o valor de similaridade de quadros recentes. Este comportamento é capturado pelo cálculo das diferenças de primeira ordem de $\boldsymbol{S}_t$ (em valor absoluto), armazenadas em outro vetor $\boldsymbol{y}_t$, dado por

$$\boldsymbol{y}_t = \Big(|C(H_t, H_{t-\Delta t_1})|,\ |C(H_t, H_{t-\Delta t_2})|,\ ..., |C(H_t, H_{t-\Delta t_n})| \Big). \tag{3.12}$$

Quando uma mudança de longo prazo ocorre, os valores de $\boldsymbol{y}_t$ tendem a ser aproximadamente semelhantes, de modo que o valor máximo deve ser próximo da média. Por outro lado, durante as mudanças de curto prazo, a queda acentuada no $\boldsymbol{S}_t$ deverá conduzir a um valor relativamente elevado de $\boldsymbol{y}_t$, mas a média de $\boldsymbol{y}_t$ deve ser baixa. Assim, uma

Figura 3.5: Quadros que ilustram uma mudança a longo prazo: as pessoas estão se movendo para a direita, e começam progressivamente a correr (primeiro as pessoas na frente, e, em seguida, os outros).

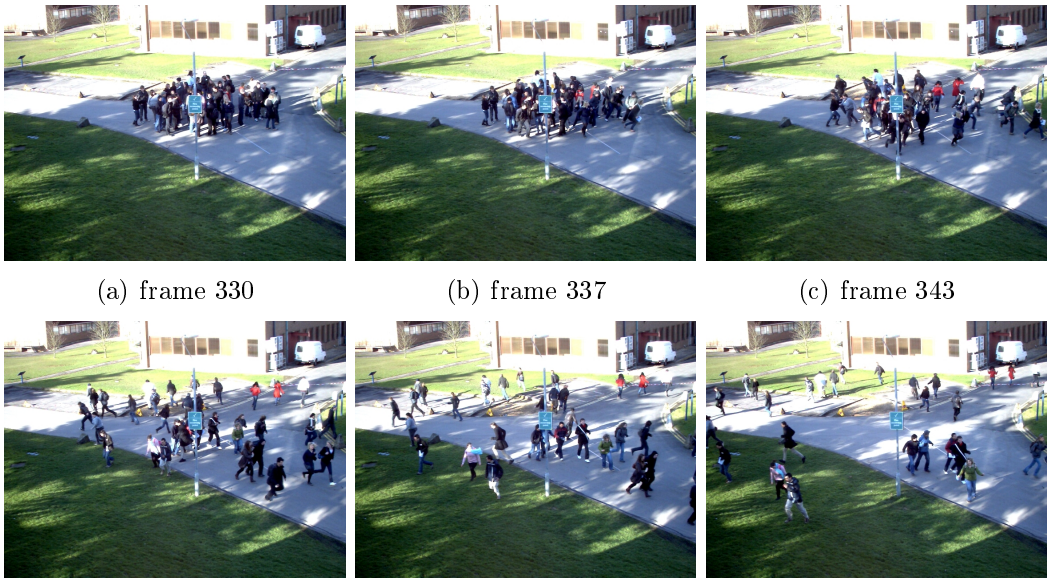


(a) frame 47 (b) frame 55 (c) frame 66

(d) frame 71 (e) frame 78 (f) frame 90

Fonte: Ferryman e Ellis (2010).

Figura 3.6: Quadros que ilustram uma mudança a curto prazo: as pessoas estão em pé no meio da cena e de repente começam a correr.



(a) frame 330 (b) frame 337 (c) frame 343

(d) frame 347 (e) frame 352 (f) frame 359

Fonte: Ferryman e Ellis (2010).

mudança é classificada como de curto prazo se

$$c_t = \frac{\max\{\boldsymbol{y}_t\}}{\overline{\boldsymbol{y}_t}} > \alpha, \tag{3.13}$$

onde α é um limiar definido experimentalmente em $3,7$.

3.2 Detecção de Eventos em Nível Local

A técnica apresentada na Seção 3.1 utiliza a informação de toda a cena, assumindo que todas as pessoas do vídeo fazem parte da mesma multidão. Porém, em algumas cenas, é possível que as pessoas capturadas pela câmera estejam em mais de um grupo, e a alteração ocorra em apenas um dos grupos. Determinamos os grupos como sendo formados por pessoas que possuem movimentação semelhante e que estão próximas espacialmente. A abordagem apresentada até então pode, dependendo da diferença do tamanho dos grupos, não detectar a mudança. E, mesmo que detecte, não consegue localizá-la espacialmente na cena.

Para que seja possível detectar mudanças em diferentes grupos que estão na cena, passamos a separar a multidão observada em grupos e a realizar a análise em cada grupo, estendendo o método até então exposto. É importante observar que a extensão não tem um decréscimo na precisão de detecção da alteração de comportamento da multidão, pois a análise de toda a multidão da cena é realizada em paralelo.

3.2.1 Determinação dos Grupos

Neste trabalho, assume-se que um grupo de pessoas possui vetores de deslocamento coerentes entre si, e estão localizadas espacialmente próximas. Assim, a divisão da multidão em grupos é realizada utilizando algoritmos de clusterização. Embora existam vários algoritmos de clusterização reportados na literatura ((XU; WUNSCH D., 2005; BERKHIN, 2006)), optou-se por uma adaptação do tradicional algoritmo K-means. Como o número de classes é determinado previamente no algoritmo K-means tradicional, foi utilizado um método que avalia as probabilidades a posteriori dos modelos M_K, onde diferentes modelos correspondem a soluções com diferentes quantidades de classes K, para a seleção do

melhor modelo, semelhante ao utilizado no X-means (PELLEG; MOORE, 2000). Para realizar a clusterização da cena, foram utilizados a posição(x, y) de cada pixel da imagem (exceto dos pixels pertencentes ao fundo), o fluxo ótico referente a cada um destes pixels, denotando a velocidade e orientação de deslocamento de cada pixel, de modo que a entrada para o K-means é um vetor 4D.

Na estratégia utilizada, é determinado que K está no intervalo de valores $[n, m]$, e o algoritmo K-Means é executado para cada valor inteiro nesse intervalo. Para cada K é determinado um índice BIC (*Bayesian Information Criterion*), como apresentado por Pelleg (PELLEG; MOORE, 2000), e é utilizado o valor de K com melhor índice. Assumindo um conjunto de dados D referente a cena observada, para cada modelo M_K é calculado o índice BIC utilizando a equação:

$$BIC(M_K) = \hat{l}_K(D) - \frac{p_K}{2} \cdot \log R, \tag{3.14}$$

onde $\hat{l}_K(D)$ é a log-verossimilhança dos dados referente ao modelo e é tomada no ponto de máxima verossimilhança, p_K é o número de parâmetros no modelo M_K, que é dado pelo somatório de $K-1$ classes de probabilidades, $M \times K$ coordenadas de centróides (M sendo a dimensão da base de dados) e uma estimativa de variância, e R é o número total de pontos que pertencem aos centroides sob consideração.

Como espera-se que o número de classes não deve variar consideravelmente em quadros consecutivos, utiliza-se o número de *clusters* K_t no quadro $t-1$ para limitar a faixa de grupos no quadro t. Mais precisamente, seleciona-se $n = K_{t-1} - p$ e $m = K_{t-1} + p$, onde p é um valor pré-determinado, que controla a variação máxima do número de grupos em quadros adjacentes (experimentalmente definido como $p = 1$).

Esta abordagem de determinar um índice BIC para estimar a quantidade de *clusters* tende a priorizar a criação de um número maior de grupos, conforme mencionado em (PELLEG; MOORE, 2000). Como consequência, a aplicação deste critério a cada quadro tende a aumentar o número de *clusters* à medida que o tempo passa, gerando uma clusterização demasiada da cena. Com objetivo de evitar a criação de *clusters* que não melhoram a representação da cena (no contexto desse trabalho, divide um grupo real em dois *clusters*), realizamos um pós processamento nos grupos gerados, verificando a distância entre estes grupos através da distância euclidiana ponderada. Para tal, cada grupo

é representado usando as informações dos seus centroides em coordenadas de mundo, ou seja, a sua posição no mundo, sua velocidade e orientação de movimento.

Desta forma, se a distância entre dois grupos for menor que um certo limiar L, identificamos que o algoritmo dividiu um grupo real em dois *clusters*. O valor de L foi definido empiricamente, e para isto definimos um valor base L_n e um valor teto L_m. Para definir L_n foram utilizadas cenas onde há apenas um grupo durante o vídeo, então utilizamos o algoritmo K-Means com $K = 2$, formando dois *clusters* i e j. Após calculamos a distância l_{ijt} entre os *clusters* i e j para cada quadro t ao longo do vídeo, e então definimos L_n como sendo:

$$L_n = \max l_{ijt}. \tag{3.15}$$

O cálculo de L_m foi feito de forma semelhante ao cálculo de L_n, porém foram utilizadas cenas onde há mais de um grupo na cena, no caso analisado 3, e estes permanecem até o final da análise da cena, de forma que não há alteração no número de grupos. Utilizamos o algoritmo K-Means com $K = 3$, e encontramos as distâncias entre os três *clusters* formados. A difereça entre a definição do L_n para o L_m é que definimos o L_m como sendo a menor distância encontrada ao longo do vídeo. Com o valor base e o valor teto definidos como $L_n = 93.4$ e $L_m = 109.8$, admitimos que o limiar L é um valor dentro deste intervalo $[L_n, L_m]$, então escolhemos nos nossos experimentos $L = 100$ pixels. Embora nos testes realizados neste trabalho foi utilizado um limiar L fixo, este é dependente da resolução do vídeo analisado, pois utilizamos informações extraídas unicamente do plano da imagem para a realização da análise dos grupos.

Depois de realizada a clusterização e o pós processamento da cena, é importante relacionar os *clusters* detectados ao longo do tempo, para que a variação do fluxo de movimento dentro de cada grupo seja possível. Para isto, os centroides dos grupos formados no quadro $t-1$ são utilizados na inicialização do K-Means no quadro t, ao invés da tradicional inicialização aleatória. Além disso, determinamos a distância euclidiana entre os centróides e o tamanho dos *clusters* formados no quadro t e os *clusters* formados no quadro $t-1$, e utilizamos o algoritmo hungáro (KUHN, 1955) para determinar quais os *clusters* em t correspondentes em $t-1$, conforme explicado a seguir.

O centroide $\boldsymbol{W}_{i,t}$ de cada *cluster* i contém a informação de posição (na imagem)

e movimento (vetor deslocamento) do grupo no quadro t, informações estas utilizadas também como dados observados pelo algoritmo K-Means quando efetuada a clusterização. Um vetor de similaridade $\boldsymbol{U}_{i,t}$ é criado para cada *cluster*:

$$\boldsymbol{U}_{i,t} = (\boldsymbol{W}_{i,t}, N_{i,t}), \tag{3.16}$$

onde N_i é o tamanho do *cluster* em pixels. Assim, é calculada a similaridade entre os vetores do quadro t e do quadro $t-1$ através da distância euclidiana entre os vetores $\boldsymbol{L}$ correspondentes aos *clusters* analisados. Os valores de similaridade (quanto mais próximo de zero, mais similares são os *clusters*) são utilizados como dados de entrada para o algoritmo hungáro. O método húngaro é um algoritmo de otimização combinatória de minimização, retornando o menor custo de correspondência entre os *clusters*. Assim, é possível determinar uma coerência temporal dos grupos, possibilitando que seja realizada uma análise de cada grupo.

Se um grupo se mantém ao longo do tempo, a análise de similaridade temporal dos histogramas individuais de cada grupo é usada para identificar variações de comportamento intra-grupo.

3.2.2 Análise do Comportamento dos Grupos

A análise dos grupos é realizada analisando as alterações no histograma gerado para o grupo, com a mesma técnica apresentada na Seção 3.1.2. A diferença é a aplicação da técnica de comparação para cada grupo, e não mais num histograma único representando a multidão. Todos os critérios anteriormente apresentados, são novamente aplicados nesta abordagem em grupos. Se o valor de K inferido for igual a 1, a análise é realizada da mesma forma de quando abordamos a multidão como sendo única em cena. Também é importante salientar que a analise global descrita na seção 3.1.1 pode ser realizada após a detecção de grupos, mesmo que o algoritmo de *clustering* apresente erros. A razão é que o histograma global (usado em 3.1.1) pode ser escrito como a média dos histogramas de cada grupo (ponderada pelo número de elementos do grupo), independentemente da qualidade do *clustering*.

Além da alteração do padrão de movimento dentro de cada grupo, a detecção de agru-

pamentos pode ser explorada também para identificar outros tipos de comportamentos na cena, como a união e separação de grupos. O surgimento de um novo grupo pode se dar por dois motivos: ou o grupo está entrando na cena ou é uma ramificação de um grupo já existente. Para determinar qual das duas causas é a que determinou o surgimento do grupo, é realizado uma análise da similaridade deste grupo com grupos do quadro anterior.

Para isto, novamente é utilizado o centroide $\boldsymbol{W}_{i,t}$ do grupo i, pertencente ao quadro t, e é determinada a similaridade $R_{i,j}$ calculando a distância euclidiana ponderada entre o centroide do grupo analisado e os centroides $\boldsymbol{W}_{j,t-1}$, onde $j = 1, ..., K_{t-1}$ e K_{t-1} é o número de grupos encontrados em $t-1$. Assim como no cálculo da distância entre grupos de um mesmo quadro, são utilizadas as informações em coordenadas de mundo dos centroides.

Após determinar a similaridade $R_{i,j}$, é escolhido o j em que $R_{i,j}$ for menor, ou seja, o grupo em $t-1$ que possui a menor distância do grupo i. Para determinar se i é uma ramificação de j, e é determinado que i é uma ramificação de j se:

$$R_{i,j} < \gamma, \tag{3.17}$$

caso contrário, é determinado que o grupo i está entrando na cena. Nos experimentos utilizamos o valor do limiar γ com o mesmo valor do limiar L. Determinando assim um único limiar que determina a a similaridade entre grupos.

Quando ocorre o desaparecimento de um grupo, novamente há duas explicações, ou a união de dois grupos ou a saída de cena do grupo. Para verificar a causa, é realizado uma abordagem semelhante a apresentada no surgimento de grupos, porém ao invés de analisar a distância de um grupo pertencente ao quadro t, é analisado um grupo j pertencente ao quadro $t-1$ e comparado com os grupos $i = 1, ...m$, onde m é o número de grupos no quadro t.

Capítulo 4

Resultados

Os resultados deste trabalho serão divididos em detecção de eventos globais e na detecção de eventos entre grupos. Na detecção de eventos globais, serão abordadas situações onde há alteração repentina na velocidade da multidão e/ou alteração na orientação da multidão, na tentativa desordenada de evasão, impulsionada pelo desejo único de se afastar do foco do problema. Já as interações entre grupos serão utilizadas para determinar a divisão de um grupo, ou a união de dois ou mais grupos em um grupo.

Nos experimentos realizados foram utilizados os valores para os parâmetros apresentados na Tab. 4.1. Os resultados destes experimentos são apresentados no decorrer deste capítulo.

Tabela 4.1: Valores dos parâmetros utilizados nos experimentos.

Parâmetro	Definição	Valor
n	Quantidade de quadros observados	14 quadros
Δ t	Intervalo utilizado a cada observação	$0,57$ s
λ	Taxa de decaimento dos pesos das observações	$0,52$
α	Limiar que define o tipo da mudança	$3,7$
L	Limiar de similaridade entre grupos no mesmo quadro	100
γ	Limiar de similaridade entre grupos de quadros adjacentes	100

4.1 Detecção de Eventos Globais

Para avaliar o método de detecção global de mudança de comportamento apresentado neste trabalho, foram utilizados nos experimentos dois conjuntos de dados públicos. O

conjunto de dados PETS2009 S3 para análise de multidão e o conjunto de dados público da Univesity of Minnesota (DATASET,) para fuga em cenários de pânico foram utilizados para validar a abordagem proposta, e o objetivo é detectar alterações assim que elas acontecem. O conjunto de dados PETS2009 fornece os parâmetros da câmera, e para o conjunto de dados de Minnesota estimamos a homografia do plano de chão usando estruturas planar geométricas presentes na sequência de vídeo.

Embora o conjunto de dados de Minnesota apresente os valores do *ground truth* para a detecção de eventos, eles parecem ser marcados alguns quadros após o evento ocorrer, como observado em (CHEN; HUANG, 2011). Por isso, usamos os valores de 'ground truth' atualizados propostos no (CHEN; HUANG, 2011) para este banco de dados, e usamos a mesma técnica para encontrar o *ground truth* real nas outras sequências, como pode ser visto na Fig. 4.1.

Em Fig. 4.2, pode ser visto que o nosso método apresenta resultados melhores do que o *social force model* (SFM) (MEHRAN; OYAMA; SHAH, 2009) e o *adjacency-matrix based clustering* (AMC) (CHEN; HUANG, 2011), o que significa que o evento (alteração no comportamento) foi detectado mais rápido.

Fig. 4.3 ilustra nossos resultados para o conjunto de dados PETS2009 S3.Event Recognition – sequência 1. O *ground truth* foi observado por nós, usando a mesma estratégia adotada em (CHEN; HUANG, 2011), ou seja, rotulamos manualmente o quadro no qual a mudança de comportamento acontece (neste caso, quando as primeiras pessoas começam a correr).

Nossa abordagem detectou a alteração 10 quadros depois do valor no *ground truth*, correspondendo a aproximadamente $1,5$ segundos. Isto acontece porque a mudança ocorre de uma forma em longo prazo: as pessoas na frente da multidão começam a correr em primeiro lugar, e em seguida os outros, sucessivamente. De fato, o nosso método detectou este evento como uma mudança em longo prazo utilizando a Equação (3.13), e o valor correspondente de c_t foi de $3,52$. Para efeito de comparação, as cenas mostradas na Fig. 4.4 e Fig. 4.2 são mudanças de curto prazo, e apresentam como resultado $c_t = 3,91$ e $c_t = 4,35$ respectivamente. Na Fig. 4.4 há uma alteração de comportamento da multidão, em que as pessoas diminuem a velocidade até parar no centro da imagem. Esta alteração não é detectada por ser uma mudança que ocorre muito lentamente.

Figura 4.1: Quadros ilustrando a detecção de um evento de curto prazo (detectado no quadro 38) Um evento incomum começando no quadro 38 (c) foi detectado, em que o público realmente começa a correr, portanto utilizamos o quadro 38 como o *ground truth* para essa sequência. No quadro 48 (g) é onde o nosso método detecta a alteração no comportamento da multidão.

(a) frame 36 (b) frame 38 (c) frame 40

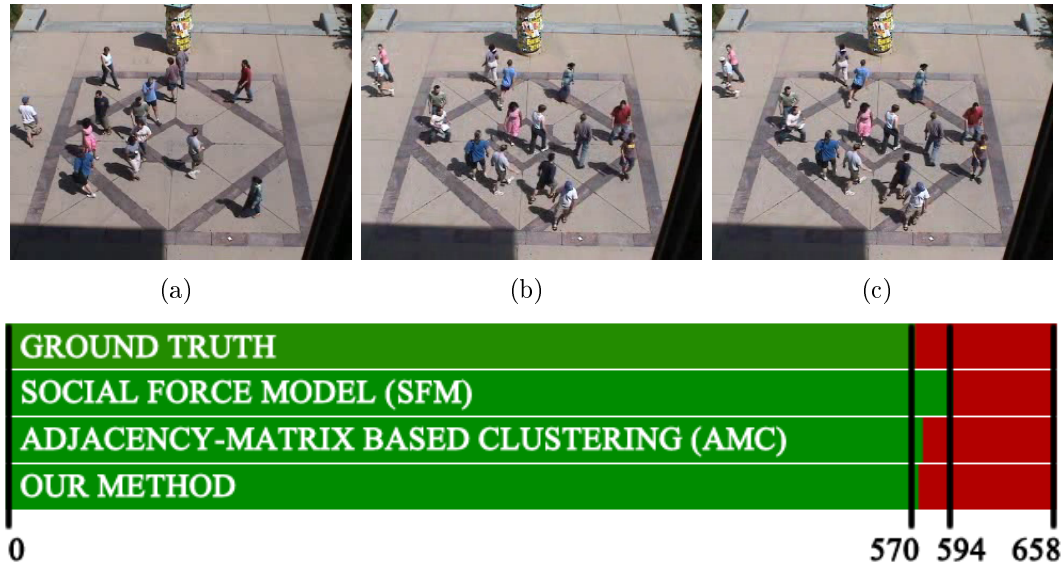
(d) frame 42 (e) frame 44 (f) frame 46

(g) frame 48 (h) frame 50 (i) frame 52

(j) frame 54

Fonte: Ferryman e Ellis (2010).

Figura 4.2: (a) A primeira imagem da sequência. Evento começa no frame 570, indicado (b). Nosso método detecta a alteração no comportamento de multidão no frame 572, mostrado em (c), superando tanto o *social force model* (MEHRAN; OYAMA; SHAH, 2009) quanto o *método adjacency-matrix based clustering* (CHEN; HUANG, 2011), que detectam os eventos quadros 594 e 575, respectivamente.

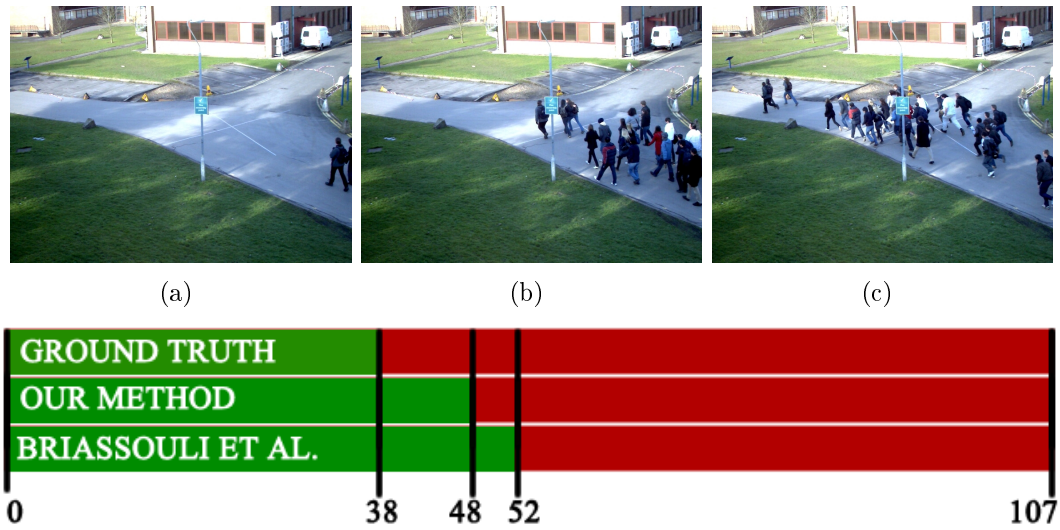


(a) (b) (c)

(d)

Fonte: O próprio autor.

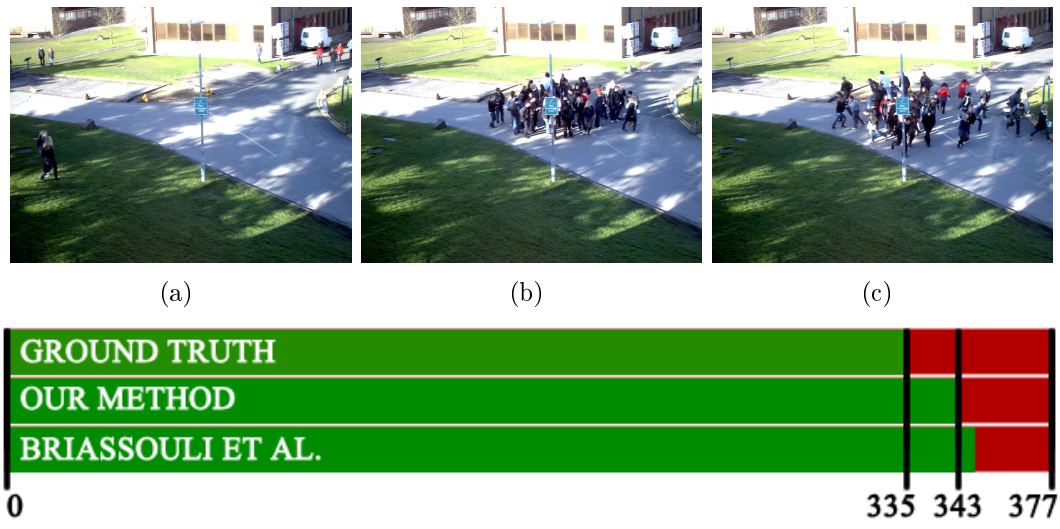
Figura 4.3: (a) A primeira imagem da sequência. (b) O quadro em que o 'ground truth' indica uma mudança (quadro 38). (d) Quadro onde nosso método detecta a mudança no comportamento da multidão (quadro 48). (d) Ilustração esquemática da linha do tempo da sequência de vídeo com o *ground truth* e quadros onde ocorrem a detecção em cada método.



(a) (b) (c) (d)

Fonte: O próprio autor.

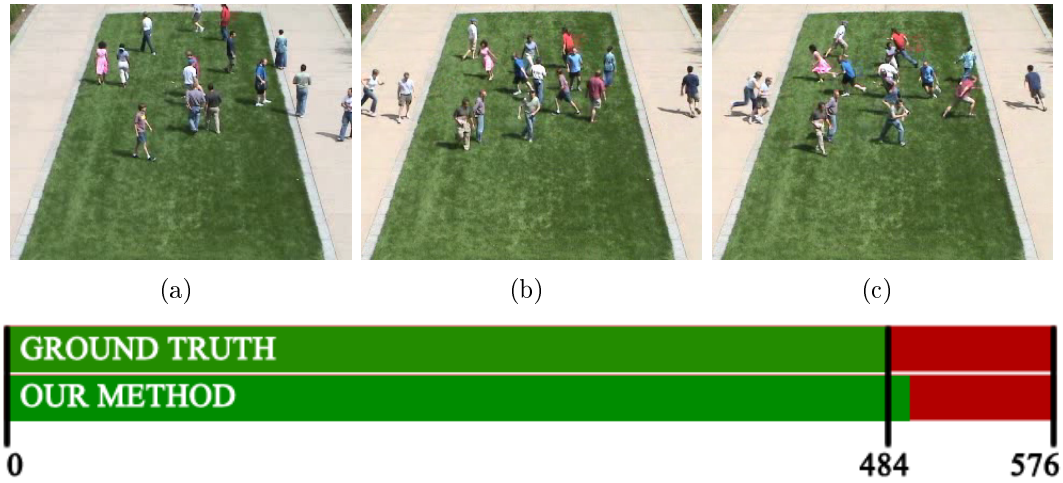
Figura 4.4: (a) A primeira imagem da sequência analisada. (b) Quadro 335, onde o evento começa. (c) Quadro 343, onde nosso método detecta o evento, 1,14 segundo após o início.



(a) (b) (c) (d)

Fonte: O próprio autor.

Figura 4.5: (a) A primeira imagem da sequência. (b) O quadro onde começa evento (quadro 484), (b) Quadro em que o nosso método detectou a mudança de comportamento (quadro 496).



(a) (b) (c)

(d)

Fonte: O próprio autor.

Experimentos também foram realizados em outras sequências de ambos os conjuntos de dados, como exemplificado em Fig. 4.5 e Fig. 4.6 mas sem comparação com outros métodos[1]. Em Fig. 4.5, uma mudança de curto prazo foi detectada pelo nosso classificador 12 quadros após a anotação verdade terrestre.

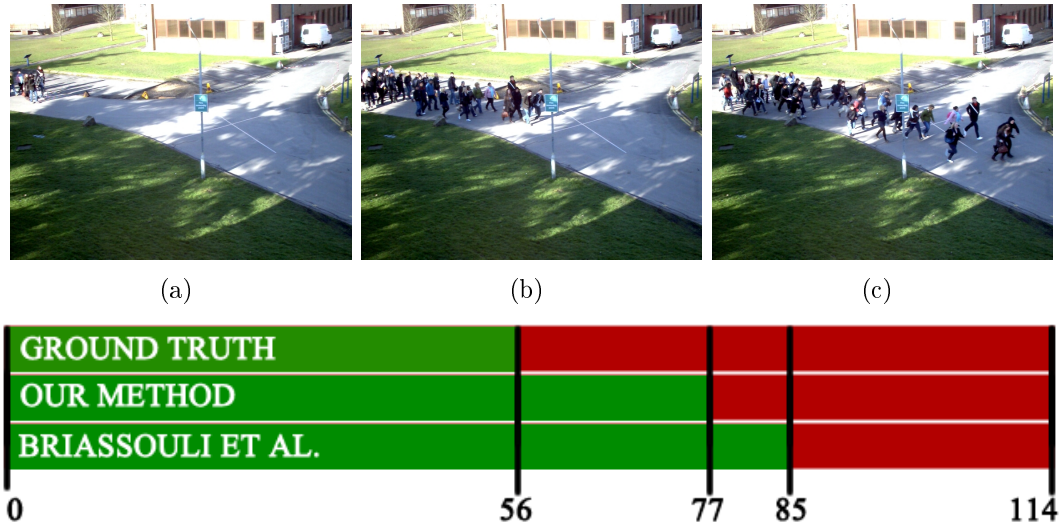
Na Fig. 4.6, a alteração é classificada como uma mudança de longo prazo, e a detecção ocorreu cerca de 3 segundos após a anotação manual. Como já foi explicado anteriormente, mudanças de longo prazo apresentam um atraso de detecção maior, devido à alteração suave do σ_t.

4.2 Detecção de Eventos em Nível Locais

Não foram encontradas sequências de vídeo publicamente disponíveis contendo eventos em grupos específicos. Em vista disso, foram utilizados dados sintéticos, que simulam dois grupos na cena e ocorre uma mudança de comportamento em apenas um deles. Foram utilizados três cenários diferentes para os testes: i) dois grupos se movendo em sentidos opostos, e um deles começa a correr, como ilustrado na Fig. 4.7; ii) outro cenário onde

[1]Eles não mostram resultados relacionados para essas sequências em seus artigos.

Figura 4.6: (a) A primeira imagem da sequência. (b) O quadro em que o *ground truth* indica uma mudança (quadro 56). (c) O quadro em que nosso método detectou o evento (quadro 77).



(a) (b) (c)

(d)

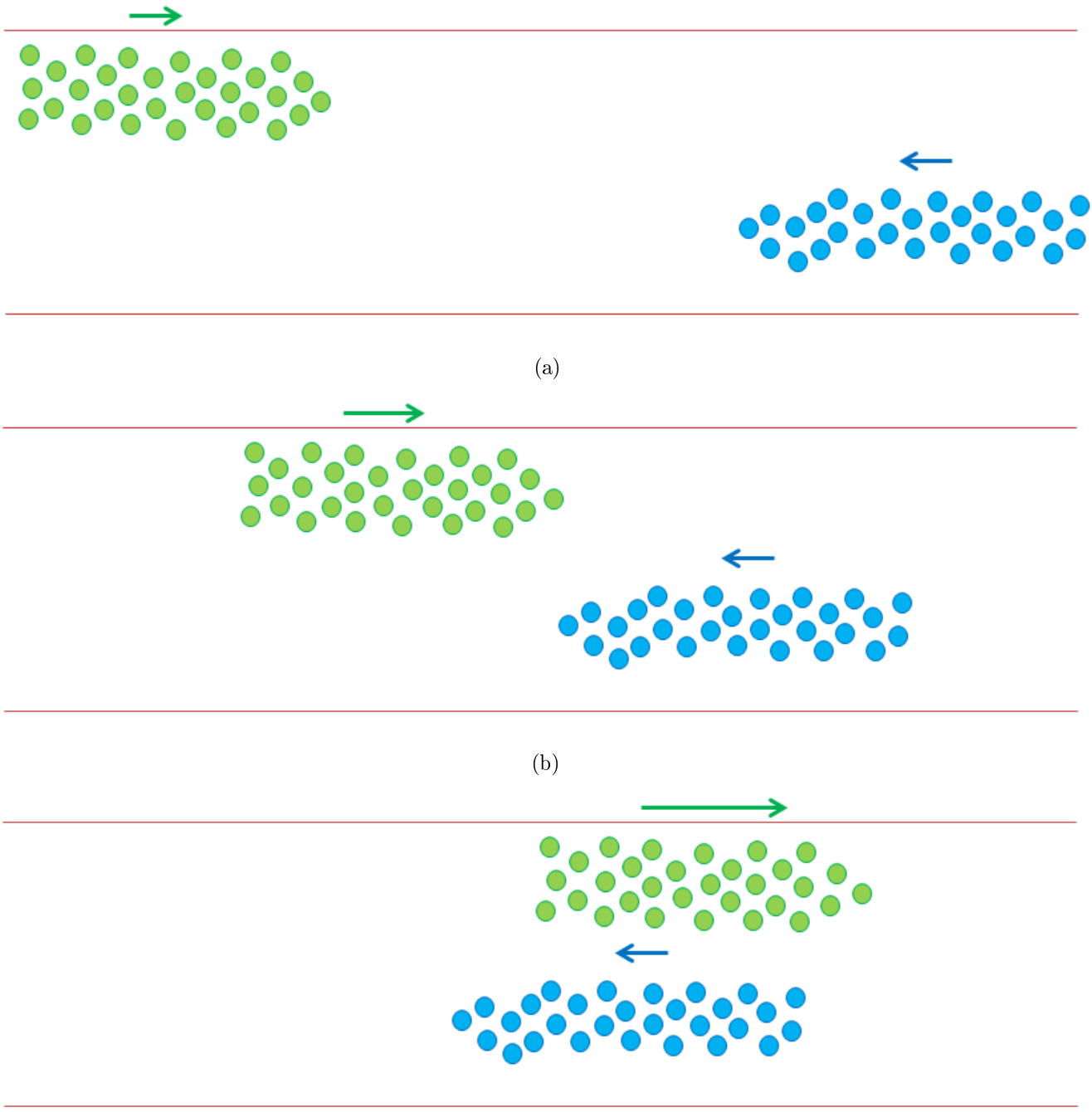
Fonte: O próprio autor.

dois grupos começam se movendo no mesmo sentido, porém com velocidades diferentes, e um deles altera a orientação, como mostrado na Fig. 4.8; e, por último, iii) dois grupos se movendo em sentidos opostos, se cruzam no meio do caminho e após um dos grupos se dispersa em várias direções correndo, como ilustrado na Fig. 4.9.

Para cada cenário foram realizados 10 testes, em cada teste foram variados o número de agentes presentes na cena, e as informações características de cada um. Para simular uma multidão real, as alterações de velocidade no primeiro cenário e velocidade e orientação no terceiro cenário foram realizadas gradualmente, assim como no cenário dois, onde os agentes diminuem a velocidade rapidamente e alteram a sua orientação, voltando a correr, porém na direção contrária. O principal objetivo destes testes é verificar se a abordagem de encontrar grupos na cena e determinar onde ocorre a mudança é válida, ignorando a forma de extração das informações dos agentes.

O *ground truth* utilizado para o primeiro cenário foi o instante em que os agentes aumentam a velocidade para iniciar a corrida, enquanto que no cenário dois foi utilizado o instante em os agentes diminuem a velocidade para iniciarem a corrida na direção oposta, e no último cenário o *ground truth* utilizado foi o instante de alteração de orientação dos

Figura 4.7: Cenário em que há dois grupos em direções opostas, e o grupo verde altera o comportamento.



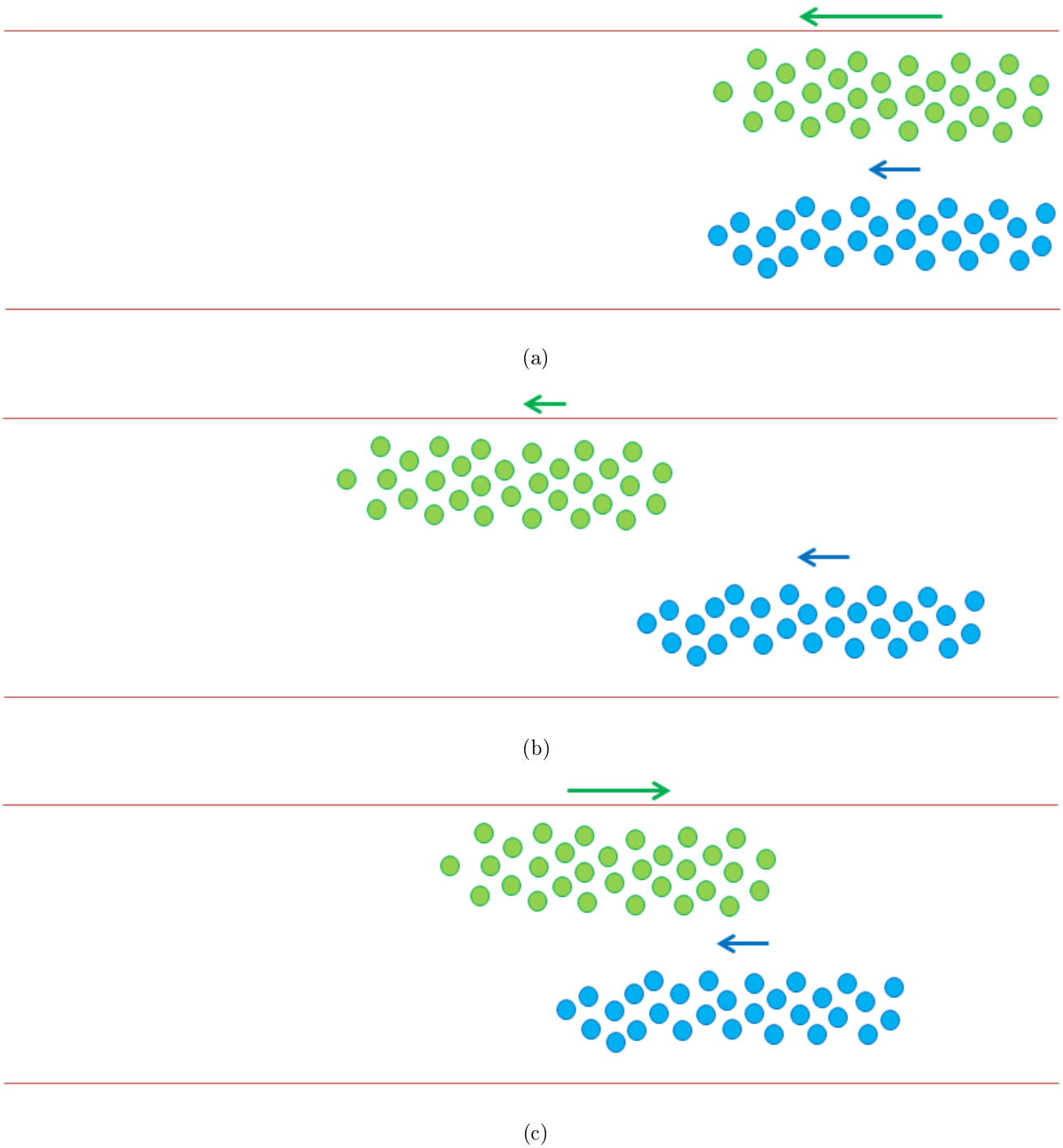
(a)

(b)

(c)

Fonte: O próprio autor.

Figura 4.8: Cenário em que há dois grupos na mesma direção, e o grupo verde altera a direção de deslocamento do grupo.



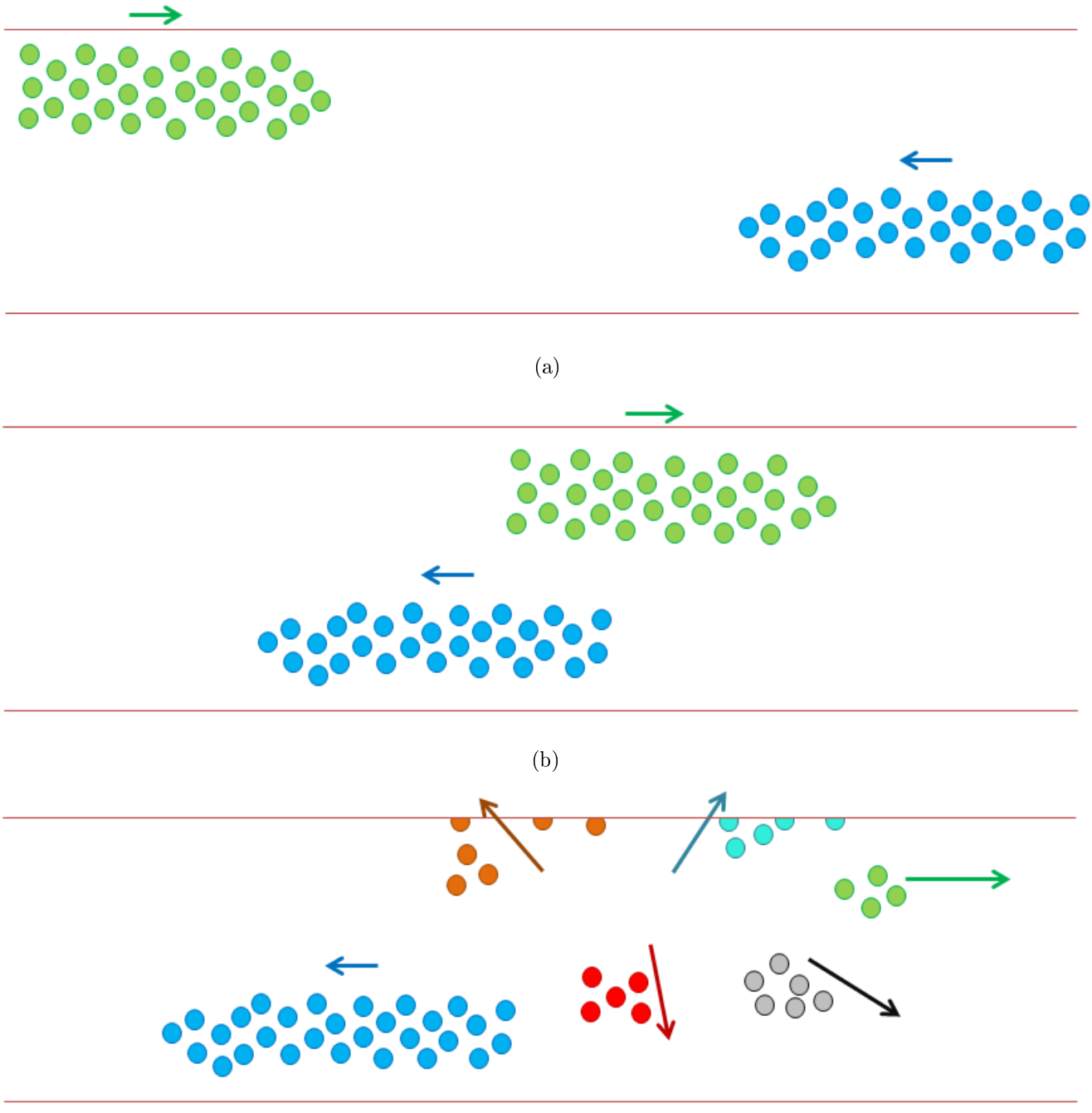
(a)

(b)

(c)

Fonte: O próprio autor.

Figura 4.9: Cenário em que há dois grupos que se cruzam no meio do caminho, e os membros do grupo verde se dispersam correndo.



(a)

(b)

(c)

Fonte: O próprio autor.

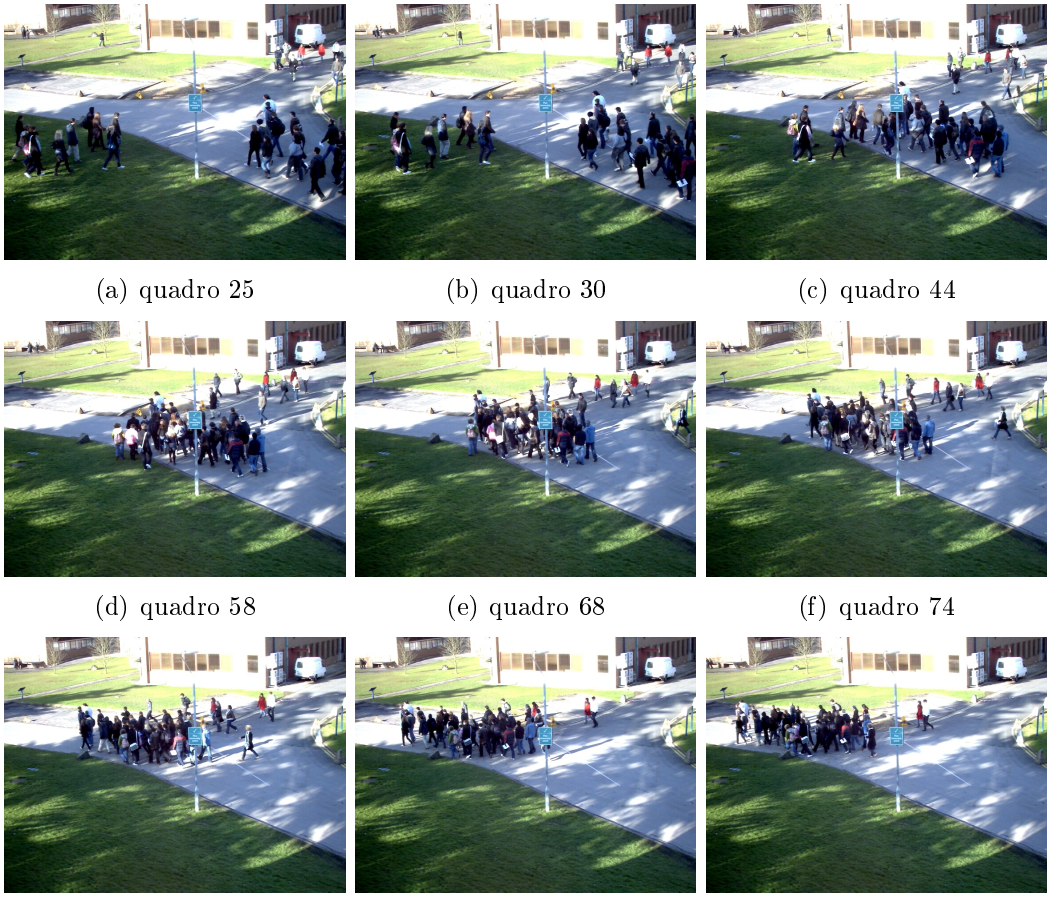
agentes presentes na cena. Em todos os testes realizados nos dois primeiros cenários a alteração no grupo é detectada menos de um segundo após o instante determinado pelo *ground truth*. Já os resultados dos testes no cenário três tiveram alterações de acordo com os dados de entrada, pois em nenhum dos casos de teste neste cenário foi possível determinar a mudança no grupo, pois o mesmo dividia-se em vários outros. Quando o grupo que ocorria a alteração de comportamento era maior que o grupo que permanecia com o comportamento durante toda a cena ou a diferença entre ambos era pequena, a mudança foi detectada utilizando a abordagem global, porém quando o grupo onde ocorre a mudança é menor que o outro grupo, a mudança de comportamento não se torna perceptível no histograma global.

Para determinar a mudança de comportamento do cenário três, podemos analisar eventos de união e divisão de grupos, como o método apresentado na Subseção 3.2.2. Para validar a detecção de eventos entre grupos, foram utilizados conjuntos de dados PETS2009 S3, onde há múltiplos grupos. O objetivo é detectar quando há a união ou divisão de grupos, tanto para determinar o ocorrido como um evento quanto para poder utilizar esta informação para a análise individual em cada grupo. O valor para $L = 100$, como explicado anteriormente, onde L é o limiar que define no pós-processamento se dois *clusters* são de um mesmo grupo real, e para γ foi determinado $\gamma = L$, e γ é o limiar que define se a alteração no número de grupos é resultado de divisão/união dos grupos da cena ou se há um grupo que entrou/saiu da cena.

Na formação dos grupos utilizamos as informações de posição (x, y) no plano da imagem e os valores do fluxo ótico adquiridos para a cena analisada. E estes são as informações dos centroides dos *clusters* que utilizamos para determinar a distância entre os *clusters*, de forma a avaliar se o surgimento de um *cluster* no quadro t corresponde a uma divisão de um *cluster* no quadro $t - 1$.

A primeira sequência da base de dados analisada é demonstrada na Fig. 4.10. Nesta cena são apresentados inicialmente três grupos se deslocando para o centro da cena, e quando os dois grupos externos se unem ao grupo central, eles adquirem a mesma orientação do grupo do centro. O primeiro grupo a se unir ao grupo central é o grupo da esquerda da imagem, e ocorre no quadro 58, enquanto que a união do grupo superior a direita ao grupo central é realizada no frame 97

Figura 4.10: Podemos observar nesta sequência, os três grupos apresentados se unindo ao final da sequência. Em (d) ocorre a primeira união, entre o grupo a esquerda e o central e em (h) a união do grupo superior a direita com o grupo principal.



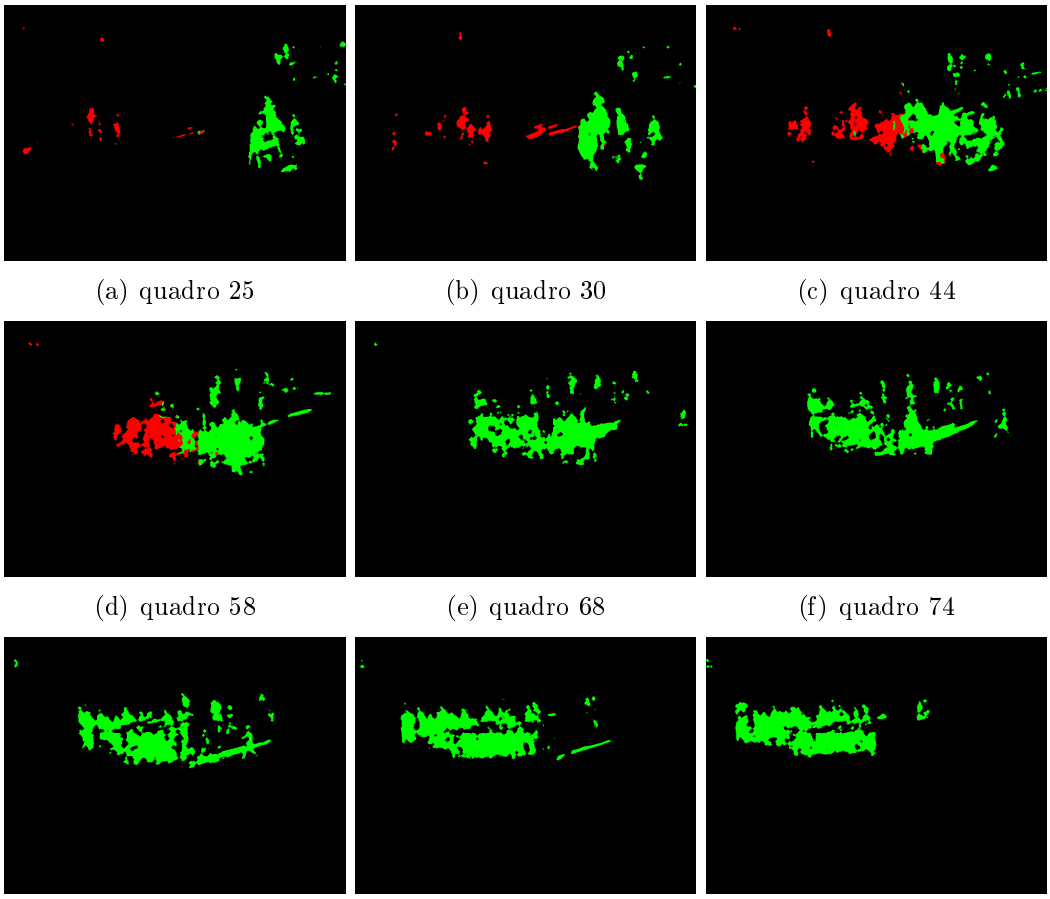
(a) quadro 25 (b) quadro 30 (c) quadro 44

(d) quadro 58 (e) quadro 68 (f) quadro 74

(g) quadro 86 (h) quadro 97 (i) quadro 105

Fonte: Ferryman e Ellis (2010).

Figura 4.11: O resultado encontrado pela nossa abordagem, percebe-se que não foi possível encontrar os três grupos, como visto em (a), unindo os dois grupos da direita em apenas um. Já a união dos dois grupos encontrados foi identificada em (e), apenas 7 quadros após o ocorrido em (d).



(a) quadro 25 (b) quadro 30 (c) quadro 44

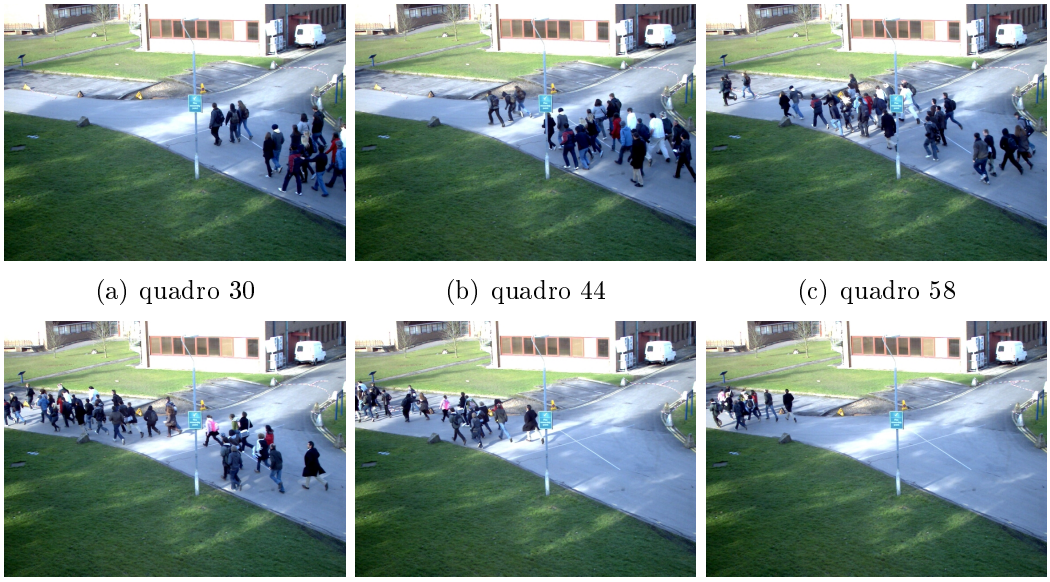
(d) quadro 58 (e) quadro 68 (f) quadro 74

(g) quadro 86 (h) quadro 97 (i) quadro 105

Fonte: O próprio autor.

Os resultados encontrados pela nossa abordagem são apresentados na Fig. 4.11, onde podemos observar que são determinados apenas dois grupos na cena, o grupo da esquerda, e o que se desloca a direita da imagem, logo assumindo que ambos os grupos que saem da direita são apenas um. Isto ocorre por que a distância entre os grupos, no plano da imagem, é pequena, e ainda o fluxo ótico dos pixels destes grupos é semelhante. Como pode ser observado na Fig. 4.11, a nossa abordagem detecta a união dos dois grupos no quadro 68, ou seja, apenas 1.43 segundo após a união ter ocorrido. A similaridade do grupo do quadro 67 que desapareceu da cena no quadro 68 com o único grupo do quadro 68 é igual a 31.7, configurando uma união de grupos.

Figura 4.12: Sequência de vídeos onde há apenas um grupo durante a cena.



(a) quadro 30 (b) quadro 44 (c) quadro 58

(d) quadro 72 (e) quadro 86 (f) quadro 100

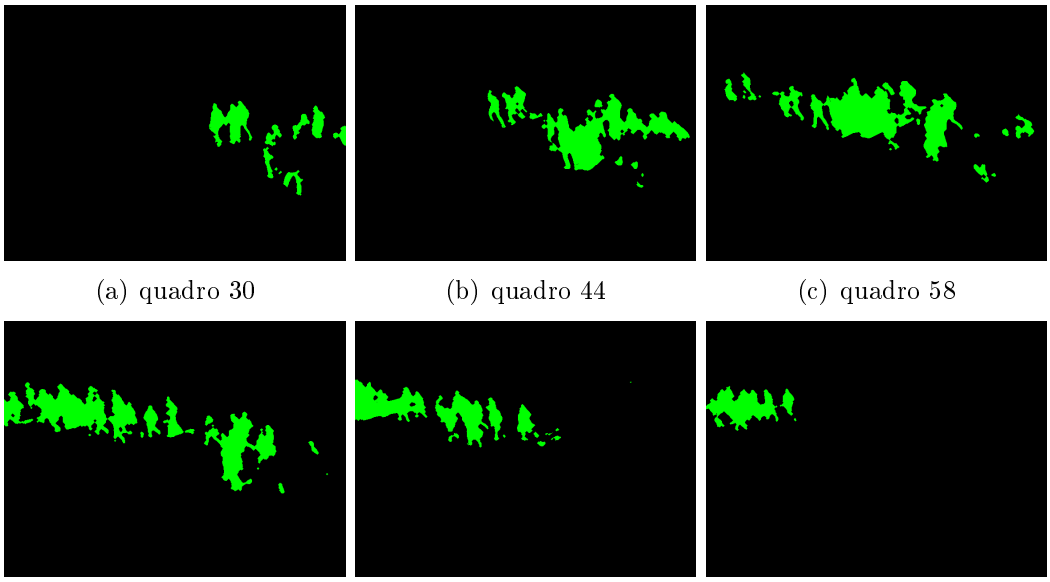
Fonte: Ferryman e Ellis (2010).

Outra sequência analisada é apresentada na Fig. 4.12, onde há apenas um grupo do início ao fim da cena, e há uma alteração no comportamento deste grupo, já explorada na seção anterior. O objetivo neste teste é observar se a nossa abordagem de grupos consegue identificar que há apenas um grupo nesta cena. O resultado pode ser observado na Fig. 4.13, onde nota-se que a abordagem utilizada para determinar os grupos, não está criando grupos desnecessários.

Outro tipo de sequência analisada é demonstrado na Fig. 4.14. Nesta sequência há apenas um grupo inicialmente, e este se divide em três grupos. A primeira divisão ocorre no quadro 56, com um grupo se separando do grupo principal em direção ao canto inferior esquerdo da cena. Enquanto que a segunda divisão ocorreu no quadro 85, com um grupo de pessoas alterando a sua orientação e se deslocando para a parte superior do cenário.

Os resultados que o nosso método encontraram para esta sequência são demonstrados na Fig. 4.15. Nossa abordagem não conseguiu detectar a primeira divisão que ocorre na cena, permanecendo após o quadro 56, onde de acordo com o *ground truth* ocorre a divisão, identificando apenas um grupo no vídeo, já a segunda divisão é detectada no quadro 83, 1 segundo após a divisão ter ocorrido, estes dois grupos não são detectados como um

Figura 4.13: Resultados encontrados para a segunda sequência de vídeo, onde podemos observar que foi encontrado, corretamente, apenas um grupo na cena.



(a) quadro 30 (b) quadro 44 (c) quadro 58

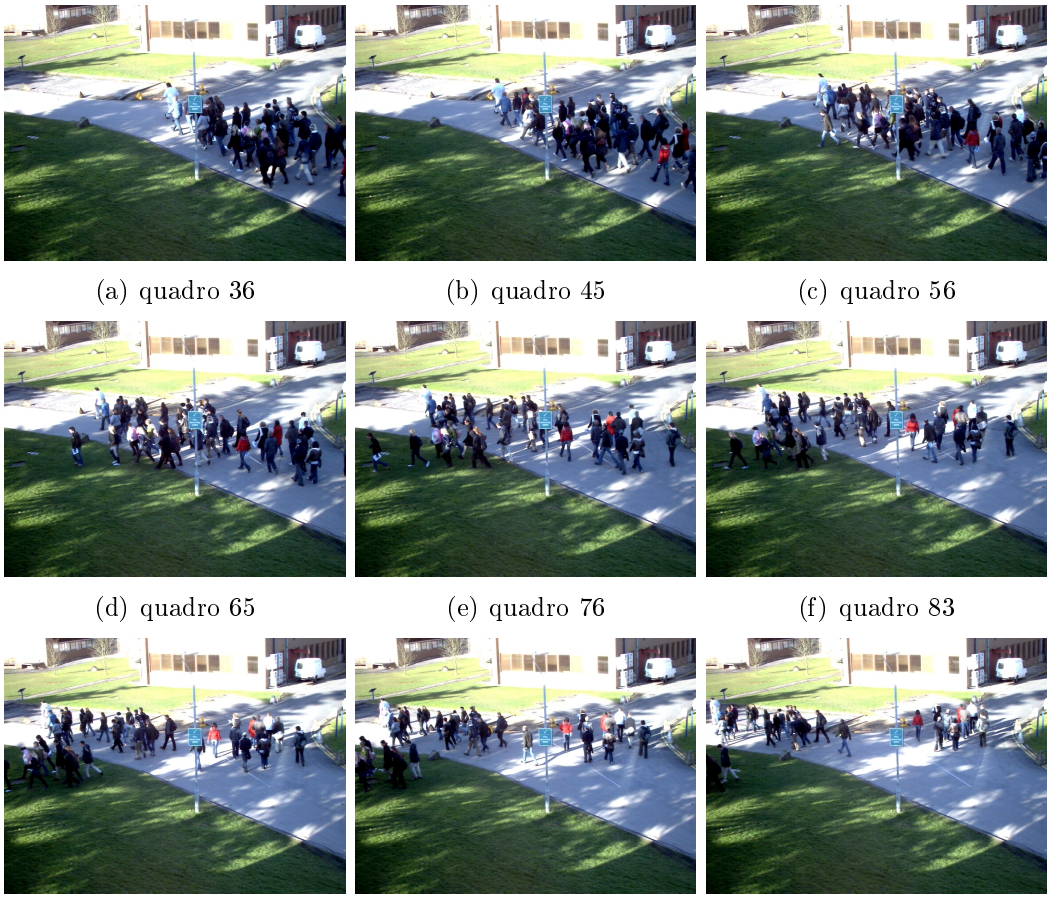
(d) quadro 72 (e) quadro 86 (f) quadro 100

Fonte: O próprio autor.

mesmo grupo no pós processamento da clusterização, pois a distância encontrada entre eles no quadro 83 é igual a 145.9. Ainda é verificado se este grupo é uma ramificação do grupo já presente anteriormente na cena ou se está entrando na cena, para isso é calculada a similaridade entre o novo grupo e o grupo do quadro 82. O valor encontrado para esta similaridade foi 19.08, menor que o nosso γ, logo é determinado que o grupo é uma ramificação do grupo anterior.

Todos os testes foram realizados na plataforma Windows, através de scripts MatLab v7.11.0. O tempo de execução total do método, incluindo a etapa de cálculo do fluxo ótico e subtração de fundo, foi de $57,38s/quadro$, sendo que a maior porção deste tempo é utilizado no cálculo do fluxo ótico, em torno de $44,96s/quadro$ em média. Apesar de não ser realizado em tempo real, é possível obter resultados mais rápidos, utilizando algoritmos em linguagem C ou C++, por exemplo.

Figura 4.14: Sequência de vídeos onde há um grupo no início da cena e este grupo se divide em três, a primeira ocorrendo no (c) quadro 56 e a segunda no (e) quadro 76



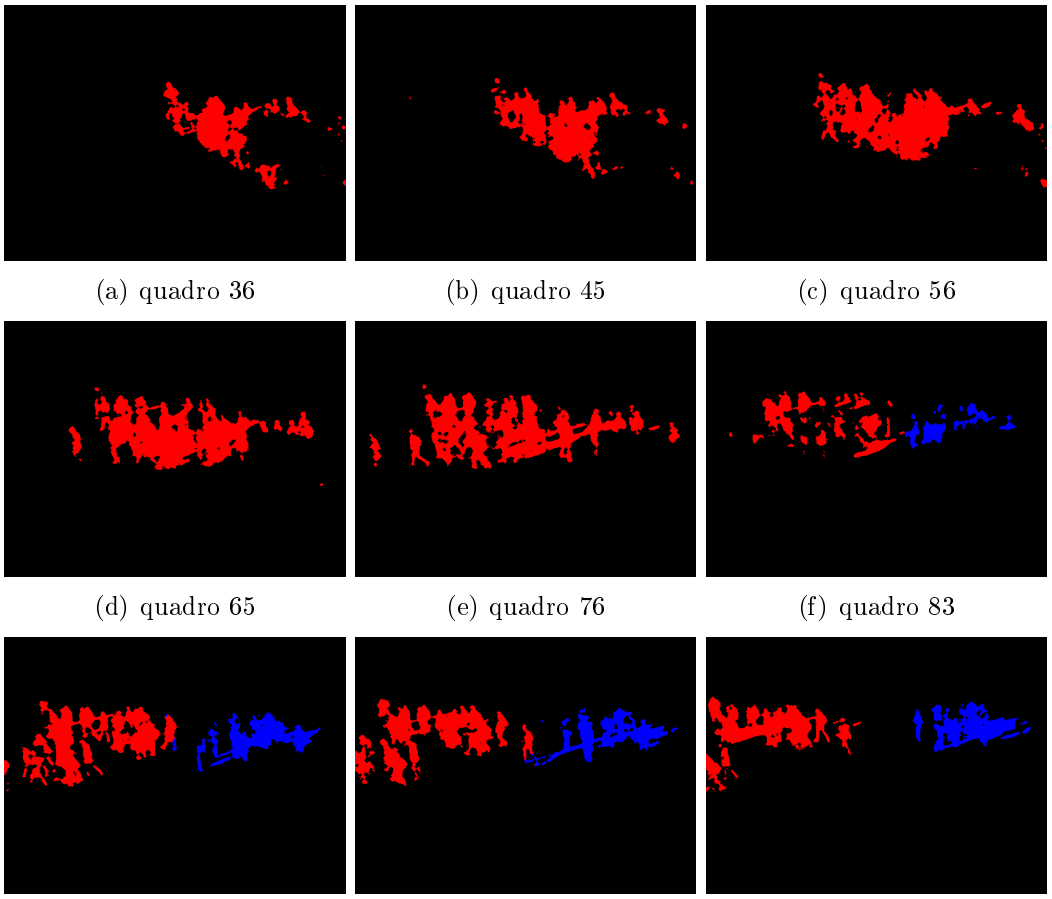
(a) quadro 36 (b) quadro 45 (c) quadro 56

(d) quadro 65 (e) quadro 76 (f) quadro 83

(g) quadro 89 (h) quadro 95 (i) quadro 105

Fonte: Ferryman e Ellis (2010).

Figura 4.15: Nesta sequência são identificados inicialmente um grupo, e detecta a divisão dos grupos no (f) quadro 83, 7 quadros após a divisão dos grupos de fato ocorrerem.



(a) quadro 36 (b) quadro 45 (c) quadro 56

(d) quadro 65 (e) quadro 76 (f) quadro 83

(g) quadro 89 (h) quadro 95 (i) quadro 105

Fonte: O próprio autor.

Capítulo 5

Conclusão

Neste trabalho apresentamos uma abordagem para detectar e localizar mudanças de comportamento em cenas de multidões. O método proposto é baseado na extração de do fundo para identificar os membros da multidão, e fluxo óptico para obter o campo vetorial de deslocamento. Este campo vetorial é mapeada para coordenadas globais, e histogramas 2D de velocidade e orientações são calculados para a cena. A similaridade de histogramas de movimento em vários quadros é o usado para detectar mudanças no comportamento das multidões, e classificá-los como de curto prazo ou de longo prazo. A abordagem utilizada para localização dos eventos é realizada separando a multidão em grupos utilizando algoritmos de *clustering*, e aplicando o método de detecção em cada grupo separadamente. Os resultados do *clustering* ainda são utilizados para identificar interações entre os grupos da cena.

Os resultados experimentais para a detecção de alterações no comportamento da multidão com conjuntos de dados públicos disponíveis indicam que a abordagem proposta apresenta potencial para detectar mudanças de comportamento, apresentando uma precisão equivalente (ou melhor) às abordagens existentes. Embora a necessidade de uma câmera calibrada possa ser uma desvantagem da abordagem proposta, é importante notar que existem algoritmos de autocalibração para a obtenção da homografia (BOSE; GRIMSON, 2003), e métodos semiautomáticos estão se tornando populares (ZHANG et al., 2013). Em qualquer caso, o mesmo procedimento utilizado, com o campo vetorial de deslocamento de coordenadas do mundo, pode ser usado com o campo vetorial de deslocamento de coordenadas da imagem (com a definição manual dos tamanhos dos bins de

velocidade).

Os resultados encontrados nos experimentos para a identificação de interação entre grupos (união/divisão de grupos) demonstram que a abordagem adotada consegue identificar os eventos da cena quando há diferença no movimento dos grupos perceptíveis no plano da imagem. Quando a diferença no movimento dos grupos é visível em coordenadas de mundo, porém pouco clara no plano da imagem, a abordagem adotada não identifica a divisão entre os grupos efetivamente, devido aos efeitos de perspectiva da câmera.

Um problema encontrado na fase de experimentos do método é a pouca quantidade de vídeos públicos para a detecção de eventos em multidões. Principalmente quando buscamos detectar eventos em diferentes grupos de pessoas, pois os poucos conjuntos de dados disponíveis apresentam alterações em todas as pessoas da cena, e não apenas em um grupo de pessoas presentes na cena.

Como continuidade deste trabalho, há a possibilidade de explorar interações entre os grupos utilizando informações em coordenadas de mundo, visando possibilitar a detecção dos eventos mais precisamente. Em vista da pequena quantidade de vídeos públicos para a detecção de eventos em multidões, uma continuidade possível deste trabalho, é a utilização de multidões sintéticas, como (COURTY et al., 2014), para a validação dos métodos e a possibilidade de maior controle da cena, possibilitando a criação de um número maior de cenas. É possível também explorar outros eventos anormais em multidões que não foram abordados neste trabalho, como efeito gargalo e efeito ondas de choque.

Bibliografia

ANDERSSON, M. et al. Recognition of anomalous motion patterns in urban surveillance. *Selected Topics in Signal Processing, IEEE Journal of*, v. 7, n. 1, p. 102–110, Feb 2013. ISSN 1932-4553.

ANDRADE, E.; BLUNSDEN, S.; FISHER, R. Modelling crowd scenes for event detection. In: *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. [S.l.: s.n.], 2006. v. 1, p. 175–178. ISSN 1051-4651.

BERKHIN, P. A survey of clustering data mining techniques. In: KOGAN, J.; NICHOLAS, C.; TEBOULLE, M. (Ed.). *Grouping Multidimensional Data*. Springer Berlin Heidelberg, 2006. p. 25–71. ISBN 978-3-540-28348-5. Disponível em: <http://dx.doi.org/10.1007/3-540-28349-8_2>.

BOSE, B.; GRIMSON, E. Ground plane rectification by tracking moving objects. In: *IEEE Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*. [S.l.: s.n.], 2003. p. 94–101.

BRIASSOULI, A.; KOMPATSIARIS, I. Spatiotemporally localized new event detection in crowds. In: *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. [S.l.: s.n.], 2011. p. 928–933.

BROSTOW, G. J.; CIPOLLA, R. Unsupervised bayesian detection of independent motion in crowds. In: *IEEE Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2006. p. 594–601.

BROX, T.; MALIK, J. Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 33, n. 3, p. 500–513, 2011.

CHEN, D.-Y.; HUANG, P.-C. Motion-based unusual event detection in human crowds. *Journal of Visual Communication and Image Representation*, v. 22, n. 2, p. 178–186, 2011.

CHENG, K.-W.; CHEN, Y.-T.; FANG, W.-H. Abnormal crowd behavior detection and localization using maximum sub-sequence search. In: *Proceedings of the 4th ACM/IEEE International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Stream*. [S.l.]: ACM, 2013. (ARTEMIS '13), p. 49–58.

COURTY, N. et al. Using the agoraset dataset: Assessing for the quality of crowd video analysis methods. *Pattern Recognition Letters*, v. 44, n. 0, p. 161 – 170,

2014. ISSN 0167-8655. Pattern Recognition and Crowd Analysis. Disponível em: <http://www.sciencedirect.com/science/article/pii/S016786551400018X>.

DATASET, U. of M. C. A. *http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi*. Disponível em: <http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi>.

DEE, H. M.; CAPLIER, A. Crowd behaviour analysis using histograms of motion direction. In: *Image Processing (ICIP), 2010 17th IEEE International Conference on*. [S.l.: s.n.], 2010. p. 1545–1548.

FERRYMAN, J.; ELLIS, A. Pets2010: Dataset and challenge. *Advanced Video and Signal Based Surveillance, IEEE Conference on*, IEEE Computer Society, Los Alamitos, CA, USA, v. 0, p. 143–150, 2010.

GREEN, M. W.; LABS. ALBUQUERQUE, N. S. N. Book, Microform, Online. *The Appropriate and Effective Use of Security Technologies in U.S. Schools. A Guide for Schools and Law Enforcement Agencies [microform] / Mary W. Green.* Distributed by ERIC Clearinghouse [Washington, D.C.], 1999. 282 p. p. Disponível em: <http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED436943>.

GREENEWALD, K.; HERO, A. Detection of anomalous crowd behavior using spatio-temporal multiresolution model and kronecker sum decompositions. *ArXiv e-prints*, jan. 2014.

HAQUE, M.; MURSHED, M. M. Panic-driven event detection from surveillance video stream without track and motion features. In: *IEEE International Conference on Multimedia and Expo*. [S.l.: s.n.], 2010. p. 173–178.

HU, Y.; ZHANG, Y.; DAVIS, L. Unsupervised abnormal crowd activity detection using semiparametric scan statistic. In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*. [S.l.: s.n.], 2013. p. 767–774.

HWANG, J.; LAY, S.; LIPPMAN, A. Nonparametric multivariate density estimation: a comparative study. *IEEE Transactions on Signal Processing*, v. 42, n. 10, p. 2795–2810, October 1994.

JUNG, C. R. Efficient background subtraction and shadow removal for monochromatic video sequences. *IEEE Transactions on Multimedia*, v. 30, n. 8, June 2009.

KUHN, H. W. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, Wiley Subscription Services, Inc., A Wiley Company, v. 2, n. 1-2, p. 83–97, 1955. ISSN 1931-9193. Disponível em: <http://dx.doi.org/10.1002/nav.3800020109>.

LAFFERTY, J. D.; MCCALLUM, A.; PEREIRA, F. C. N. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: *Proceedings of the Eighteenth International Conference on Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001. (ICML '01), p. 282–289. ISBN 1-55860-778-1.

LEE, D.-G.; SUK, H.-I.; LEE, S.-W. *Crowd Behavior Representation Using Motion Influence Matrix for Anomaly Detection*. 2013. 110-114 p.

LI, W.; MAHADEVAN, V.; VASCONCELOS, N. Anomaly detection and localization in crowded scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE Computer Society, Los Alamitos, CA, USA, v. 36, n. 1, p. 18–32, 2014. ISSN 0162-8828.

MEHRAN, R.; OYAMA, A.; SHAH, M. Abnormal crowd behavior detection using social force model. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2009. p. 935–942.

OLFATI-SABER, R.; MURRAY, R. Consensus problems in networks of agents with switching topology and time-delays. *Automatic Control, IEEE Transactions on*, v. 49, n. 9, p. 1520–1533, Sept 2004. ISSN 0018-9286.

PATHAN, S.; AL-HAMADI, A.; MICHAELIS, B. Crowd behavior detection by statistical modeling of motion patterns. In: *Soft Computing and Pattern Recognition (SoCPaR), 2010 International Conference of*. [S.l.: s.n.], 2010. p. 81–86.

PELLEG, D.; MOORE, A. X-means: Extending k-means with efficient estimation of the number of clusters. In: *In Proceedings of the 17th International Conf. on Machine Learning*. [S.l.]: Morgan Kaufmann, 2000. p. 727–734.

ROTSTEIN, A. et al. Preferred transition speed between walking dand running: Effects of training status. *Medicine and science in sports and exercise*, n. 37, p. 1864–1870, 2005.

SHI, J.; TOMASI, C. Good features to track. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 1994. p. 593–600.

SOLMAZ, B.; MOORE, B. E.; SHAH, M. Identifying behaviors in crowd scenes using stability analysis for dynamical systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 34, n. 10, p. 2064 –2070, oct. 2012.

ULLAH, H.; ULLAH, M.; CONCI, N. Real-time anomaly detection in dense crowded scenes. *Proc. SPIE 9026, Video Surveillance and Transportation Imaging Application*, v. 9026, p. 902608–902608–7, 2014.

WU, S.; MOORE, B. E.; SHAH, M. Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes. In: *IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2010. p. 2054–2060.

XIANG, T.; GONG, S. Video behaviour profiling and abnormality detection without manual labelling. In: *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*. [S.l.: s.n.], 2005. v. 2, p. 1238–1245 Vol. 2. ISSN 1550-5499.

XU, R.; WUNSCH D., I. Survey of clustering algorithms. *Neural Networks, IEEE Transactions on*, v. 16, n. 3, p. 645–678, May 2005. ISSN 1045-9227.

ZHANG, Z. et al. Practical camera calibration from moving objects for traffic scene surveillance. *Circuits and Systems for Video Technology, IEEE Transactions on*, v. 23, n. 3, p. 518–533, March 2013. ISSN 1051-8215.