

Sistema de Vigilância Inteligente para Estacionamento de Veículos

Lino Lopes de Barros Filho¹, Clésio Luis Tozzi¹

¹Faculdade de Engenharia Elétrica e de Computação (FEEC)
Universidade Estadual de Campinas (UNICAMP)
Caixa Postal 6101, 13083-970 – Campinas, SP, Brasil

{linobarr, clesio}@dca.fee.unicamp.br

Abstract. *This paper presents a semi-supervised intelligent surveillance system that detects anomalies in a parking lot. The proposed methodology extracts speed characteristics and change of direction from the trajectories to represent and classify the behavior of objects. The training phase aims to create a model of the objects normal behavior from a set of manually labeled trajectories. Then, the classification of the behavior is performed through similarity measure between the extracted features and the model. The results showed that the proposed system presented fast response and good accuracy in the anomaly detection.*

1. Introdução

Compreensão do comportamento humano (reconhecimento de atividades humanas e descoberta de padrões) é uma área com aplicações em interfaces homem-computador, multimídia, vigilância em vídeo e outras. Dentro da área de vigilância em vídeo, a vigilância inteligente tem sido abordada em diversas pesquisas devido à crescente necessidade de segurança e do controle eficaz de ambientes, tais como aeroportos, estações de trem, shoppings, arenas esportivas, instalações militares e outros. Além disso, o uso de operadores humanos para monitorar os sistemas tradicionais de vigilância em vídeo é dispendioso, limitado, gera resultados subjetivos e está sujeito a falhas humanas [Popoola and Wang 2012].

A finalidade de um sistema de vigilância inteligente (SVI) em vídeo é a detecção de anomalias, ou seja, detectar comportamentos que não satisfazem uma especificação de comportamento normal da cena. Para realizar a detecção de anomalias é necessário extrair, representar e interpretar informações visuais para aprendizado e classificação do comportamento [Morris and Trivedi 2008, Suriani et al. 2013].

Diversos trabalhos têm sido publicados na área de vigilância inteligente buscando responder algumas questões, tais como determinar o nível de supervisão do sistema, extrair características que descrevem o comportamento das diferentes classes de objetos da cena e gerar uma representação compacta das características extraídas. A seguir serão apresentados os trabalhos mais relevantes em termos de detecção de anomalias a partir de trajetórias utilizando uma única câmera estática.

Lin e Siang [Lin and Siang 2012b] apresentam um SVI semi-supervisionado que detecta anomalias em estacionamento de veículos. O sistema proposto apresentou uma boa taxa de precisão na detecção de eventos simples. Porém, são necessários dados rotulados para a fase de treinamento e existem certas restrições na cena. Morris e Trivedi

[Morris and Trivedi 2011] propõe um SVI não-supervisionado que utiliza um Modelo Oculto de Markov (Hidden Markov Model - HMM) para classificar os comportamentos. Li *et al* [Li et al. 2012] no qual é desenvolvido um SVI não-supervisionado que combina características da trajetória com informações da imagem para detectar anomalias em cenas desertas. O SVI proposto apresentou bons resultados devido as altas taxas de acerto do modelo de subtração de fundo e o rastreamento de objetos. Lin e Siang [Lin and Siang 2012a] propõe um SVI não-supervisionado que separa os eventos em classes através da clusterização de trajetórias, conseguindo alta taxa de acurácia na detecção de anomalias simples.

Em geral, os trabalhos encontrados na literatura tratam a detecção de anomalias a partir de trajetórias completas (trajetórias que contém informações desde o momento que o objeto foi detectado pela primeira vez até o seu desaparecimento na cena). Nestes casos, é necessário realizar uma normalização para que todas as trajetórias tenham o mesmo tamanho. Além disso, geralmente são utilizadas técnicas de redução de dimensionalidade no conjunto de características extraídas a partir das trajetórias com intuito de obter uma representação compacta. A principal consequência da utilização de trajetórias completas é a classificação do comportamento após à saída do objeto da cena.

Neste contexto, é proposto um SVI para auxiliar operadores nas tarefas de monitoramento e tomada de decisão em um estacionamento de veículos. No estacionamento de veículos busca-se a obtenção de informações referentes aos veículos e pessoas para um controle mais eficaz, pois existem situações que caracterizam risco e necessitam de atenção, como, por exemplo, uma pessoa que corre pela extensão do estacionamento e que pode caracterizar a fuga de um assalto. Essas situações de risco podem ocorrer em qualquer lugar e instante. Por isso, deseja-se que o SVI seja capaz de detectar quando e onde ocorrem as situações de risco, além gerar o mínimo possível de alarmes falsos. O sistema proposto extrai características de velocidade e mudança de direção dinamicamente a partir das trajetórias dos objetos para classificar o seu comportamento ao longo do tempo.

Este artigo se encontra organizado da seguinte forma: a Seção 2 descreve as etapas da metodologia proposta; a Seção 3 apresenta e discute os resultados obtidos; a Seção 4 apresenta as conclusões e os trabalhos futuros.

2. Metodologia

A arquitetura do sistema proposto é exibida na Figura 1 e apresenta os seguintes componentes principais: módulo de detecção de movimento implementado pelo o modelo de subtração de fundo Mistura de Gaussianas (MoG - *Mixture of Gaussians*); módulo de rastreamento de objetos utiliza o filtro de Kalman em conjunto com descritores de forma e textura para determinar a posição dos objetos de foreground; o módulo de classificação de objetos utiliza descritores de forma e movimento em conjunto com a área em *pixel* do objeto para determinar à qual classe das três classes predefinidas o objeto pertence; o módulo de representação do comportamento extrai características a partir das trajetórias para representar o comportamento dos objetos; o módulo de modelo de comportamento constrói o modelo do comportamento normal na fase de treinamento; e, o módulo de identificação de evento determina a similaridade dada uma amostra na fase de teste.

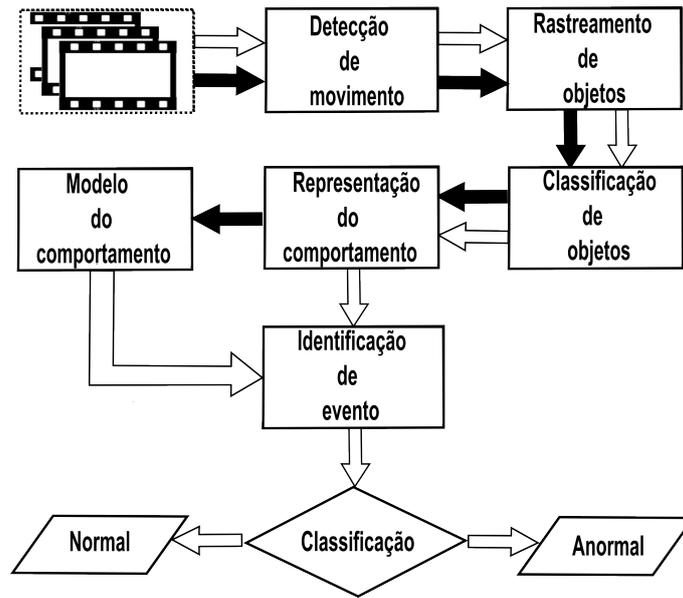


Figure 1. Arquitetura do sistema. Figura adaptada de [Lin and Siang 2012a].

2.1. Detecção de movimento

A etapa de detecção de movimento consiste em separar os objetos de interesse, ou seja, identificar os objetos de *foreground*. Para realizar essa etapa foi utilizado o modelo de subtração de fundo proposto em [Stauffer and Grimson 1999], denominado Mistura de Gaussianas. Neste método, a história recente de cada *pixel* $\{X_1, \dots, X_t\}$ é modelado por uma mistura de K Gaussianas, com a probabilidade de se observar a intensidade corrente do *pixel* dada por

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} * \eta(X_t, \mu_{i,t}, \Sigma_{i,t}), \quad (1)$$

tal que K é o número de distribuições, $\omega_{i,t}$ é o peso da i -ésima Gaussianas da mistura no instante t , $\mu_{i,t}$ é a média da i -ésima Gaussianas da mistura no instante t , $\Sigma_{i,t}$ é a matriz de covariância da i -ésima Gaussianas da mistura no instante t , e η é uma função densidade de probabilidade Gaussianas.

Em seguida, as K distribuições Gaussianas são ordenadas pelo valor de $\frac{\omega}{\sigma}$. As primeiras B distribuições são escolhidas para compor o modelo de *background*, onde

$$B = \operatorname{argmin}_b \left(\sum_{k=1}^b \omega_k > T \right), \quad (2)$$

e T é um *threshold* predefinido.

Cada novo valor assumido pelo *pixel* é testado nas K distribuições Gaussianas existentes, com intuito de verificar à qual distribuição esse *pixel* corresponde. Uma correspondência entre um *pixel* e uma distribuição Gaussianas ocorre se o seu valor distar da média até 2.5 vezes o desvio padrão de uma distribuição. Caso não exista nenhuma correspondência a distribuição Gaussianas com menor valor $\frac{\omega}{\sigma}$ é substituída por uma nova distribuição.

2.2. Rastreamento de objetos

O rastreamento consiste em determinar a trajetória de cada objeto ao longo do vídeo através da sua localização em cada quadro. Para isso, foram utilizadas métricas de posição, área e textura [Lei and Xu 2006, Freitas and Tozzi 2010]. O objeto detectado é descrito por um vetor de características $F_i = \{C_i, A_i, T_i\}$, tal que C_i é a posição do centróide do objeto i na imagem, A_i é a área do objeto i em *pixels* e T_i é a textura do objeto i obtida através do histograma calculado a partir do *Local Binary Patterns* [Ojala et al. 1994, Ojala et al. 1996]. A posição do centróide C_i é usada por um Filtro de Kalman [Kalman 1960] para estimar a posição do objeto i no quadro seguinte. A_i e T_i são empregados em situações em que apenas a posição estimada não é confiável, como em falhas ocasionadas por oclusões (parciais e totais), desaparecimento de objetos da cena e ambiguidade, uma vez que a posição estimada pode estar associada às posições de vários objetos.

2.3. Classificação de objetos

A classificação de objetos consiste em determinar a qual das três classes predefinidas (pessoas, carros e motocicletas) os objetos de *foreground* detectados pelo modelo de subtração de fundo pertencem. Para isso, foram utilizados descritores de forma e movimento em conjunto com a área em *pixel* do objeto. O objeto detectado é descrito por um vetor de características $F_i = \{R_i, A_i, V_i\}$, tal que R_i é a relação entre a largura e altura do *Bounding Box* do objeto i , A_i é a área do objeto i em *pixels* e V_i é a velocidade média do objeto i . Na etapa de treinamento, são rotulados manualmente no vídeo os objetos que pertencem às classes pessoas, carros e motocicletas. Em seguida, são obtidos os valores de R_i , A_i e V_i dos objetos rotulados. Para cada classe de objetos, os valores obtidos de R_i , A_i e V_i são modelados separadamente por uma distribuição Gaussiana, com intuito de encontrar *thresholds* que separem as três classes. Na etapa de teste, o objeto é classificado analisando as características R_i , A_i e V_i nos *thresholds* obtidos.

2.4. Representação do comportamento

Geralmente, os objetos que possuem comportamento normal realizam movimentos com o mínimo possível de mudanças na direção e velocidade média aproximadamente constante.

A trajetória do i -ésimo objeto pode ser representada por

$$T_i = \{(x_{i_t}, y_{i_t}), t = 1, \dots, N\}, \quad (3)$$

tal que x_{i_t} e y_{i_t} correspondem as coordenadas dos centróides do objeto i na imagem a cada instante de tempo t .

A velocidade de cada objeto i em um instante de tempo t é obtida definindo o vetor \vec{v}_{i_t} entre dois pontos consecutivos de sua trajetória, com intuito de calcular a norma euclidiana $\|\vec{v}_{i_t}\|$. A Figura 2 mostra a construção dos vetores \vec{v}_{i_t} .

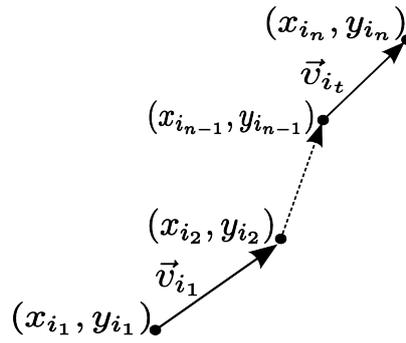


Figure 2. Construção dos vetores \vec{v}_{it} entre pontos consecutivos da trajetória. Figura adaptada de [Lin and Siang 2012a].

Seja o ponto da trajetória (x_{i_1}, y_{i_1}) definido como ponto de referência, como mostra a Figura 3(a). Seja \vec{v}_{i_1} o vetor definido entre os pontos (x_{i_1}, y_{i_1}) e (x_{i_2}, y_{i_2}) , e θ_{i_1} o ângulo determinado entre o vetor \vec{v}_{i_1} e o eixo horizontal (Figura 3(a)). Seja $\vec{v}_{i_{j-1}}$ o vetor definido entre os pontos (x_{i_1}, y_{i_1}) e (x_{i_j}, y_{i_j}) , e $\theta_{i_{j-1}}$ o ângulo determinado entre o vetor $\vec{v}_{i_{j-1}}$ e o eixo horizontal, tal que $j > 2$. Um objeto i muda de direção em um instante de tempo t quando o resultado da subtração entre os ângulos θ_{i_1} e $\theta_{i_{j-1}}$ é maior do que um *threshold* predefinido (Figura 3(b)). A cada instante que é identificada uma mudança de direção, o ponto de referência é alterado para onde ocorreu a mudança e o processo continua, como mostra a Figura 3(c). A Figura 3 ilustra a identificação de mudança de direção.

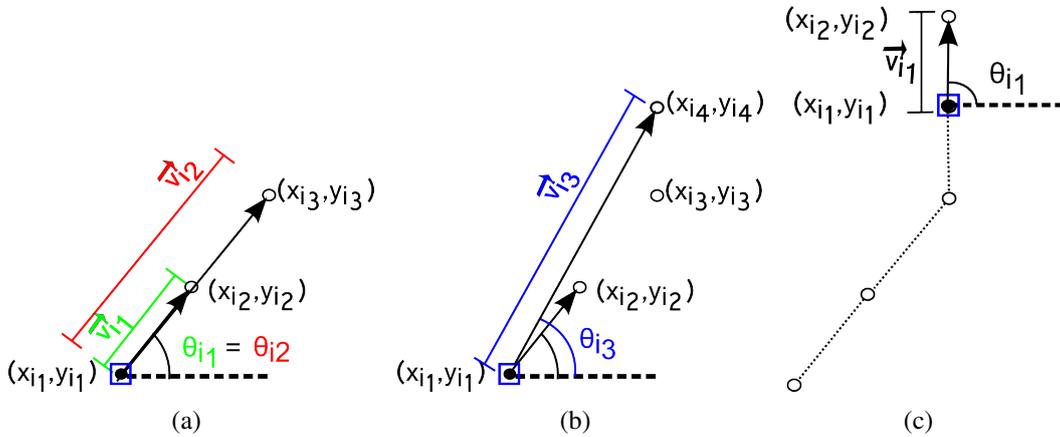


Figure 3. Detecção de mudança de direção (ponto de referência (x_{i_1}, y_{i_1}) delimitado pelo quadrado azul) : (a) não ocorre mudança de direção, pois $\theta_{i_2} - \theta_{i_1} \leq threshold$; (b) ocorre mudança de direção, pois $\theta_{i_3} - \theta_{i_1} > threshold$; (c) ilustra a mudança do ponto de referência (x_{i_1}, y_{i_1}) ocasionada pela mudança de direção. Figura adaptada de [Lin and Siang 2012a].

2.5. Modelo do comportamento

Nesta etapa, para cada classe de objetos (pessoas, carros e motocicletas) um conjunto de trajetórias manualmente rotuladas é utilizado para extrair padrões do comportamento normal. Para cada trajetória é extraída a velocidade a cada instante de tempo t e a quantidade de vezes que ocorreu uma mudança de direção. Os valores de $\|\vec{v}_{it}\|$ e a quantidade de

mudanças de direção são modelados separadamente por distribuições Gaussianas, com o objetivo de estabelecer *thresholds* para classificar o comportamento.

2.6. Classificação do comportamento

A cada instante de tempo t que o objeto está presente na cena é calculada a velocidade e a ocorrência de mudança de direção. O comportamento do objeto (pessoa, carro ou motocicleta) é classificado como normal se o valores de $\|v_{i_t}^{\vec{}}\|$ e o número de mudanças de direção distarem da média até 2.5 vezes o desvio padrão da sua respectiva Gaussiana.

2.7. Descrição das imagens utilizadas no teste do SVI

A sequência de vídeos utilizada para testar o SVI proposto foi obtida a partir de uma câmera de vídeo instalada no prédio ao lado do estacionamento monitorado. A Figura 4 mostra o local monitorado pela câmera de vídeo.



Figure 4. Imagem do estacionamento monitorado.

Os vídeos foram gravados em uma taxa de 15 quadros por segundo, onde cada quadro possui tamanho 704×480 . A gravação ocorreu em diferentes dias da semana, geralmente nos horários das 11:30 às 13:00, devido ao maior fluxo de pessoas e veículos no local. O local monitorado é um cenário complexo, pois se trata de um ambiente externo, não controlado e apresenta dificuldades particulares para algumas etapas do SVI:

1. Modelo de subtração de fundo: mudanças de iluminação (gradual e repentina), árvores que balançam e variações na geometria do *background*;
2. Rastreamento de objetos: alto número de interações entre os objetos da cena;
3. Classificação de objetos: variações no tamanho dos objetos no vídeo devido ao problema de perspectiva;
4. Classificação de comportamento: variedade de movimentos realizados pelos objetos no vídeo, por exemplo, objetos que realizam movimentos com baixa ou alta velocidade, bem como movimentos com trajetórias circulares.

3. Experimentos e resultados

A fim de analisar o desempenho do SVI proposto, esse sistema foi testado na sequência de vídeos descrita na Seção 2.7. Para avaliar a classificação do comportamento, as trajetórias foram classificadas manualmente e comparadas com a saída do SVI. A Tabela 1 apresenta para alguns dos casos estudados os instantes de ocorrência e detecção de anomalias identificadas pelo SVI. As Tabelas 2, 3 e 4 mostram a quantidade de comportamentos classificados corretamente pelo SVI.

Classe objeto	Tempo entrada	Tempo saída	Ocorrência anomalia	Detecção anomalia	Tipo anomalia
Pessoas	00:12:58	00:13:04	00:12:59	00:13:01	Alta velocidade
Pessoas	00:18:43	00:19:23	00:18:49	00:18:52	Trajatória circular
Carros	01:19:31	01:19:37	01:19:34	01:19:36	Baixa velocidade

Table 1. Informações de tempo de ocorrência e detecção de anomalias classificadas pelo SVI.

A Tabela 2 exibe em sua diagonal principal a quantidade de comportamentos classificados corretamente da classe pessoas, ao todo foram classificados corretamente 11 de 13. A Tabela 2 mostra o bom nível do SVI na detecção de anomalias, detectando corretamente 7 de 8 anomalias. Considerando que os objetos da cena realizam movimentos que apresentam uma grande variedade de velocidade e forma da trajetória, este resultado mostra que as características de velocidade e mudança de direção descreveram eficientemente o comportamento.

Classific. desejada	Classificação obtida	
	Normal	Anormal
Normal	4	1
Anormal	1	7

Table 2. Classificação de comportamento da classe pessoas.

A Tabela 3 apresenta em sua diagonal principal a quantidade de comportamentos classificados corretamente da classe carros, ao todo foram classificados corretamente 29 de 29. A Tabela 3 mostra que o SVI foi capaz de detectar todas as anomalias ocorridas. O melhor resultado era esperado dado que os movimentos realizados pelos objetos dessa classe apresentam menor dificuldade, com o padrão de velocidade e mudança de direção bem definidos.

Classific. desejada	Classificação obtida	
	Normal	Anormal
Normal	23	0
Anormal	0	6

Table 3. Classificação de comportamento da classe carros.

A Tabela 4 exibe os resultados da classificação de comportamento dos objetos da classe motocicletas. Para essa classe de objetos não foi incluída nenhuma anomalia nos dados de teste. A Tabela 4 mostra que o SVI reconheceu bem o padrão do comportamento, não ocasionando falsos alertas para os operadores.

A Figura 5 mostra o resultado da classificação do comportamento de um objeto pertencente à classe pessoas. O objeto entrou na cena no instante 01:33:56 e o SVI reconheceu o seu comportamento anormal no instante 01:33:59. Inicialmente, o sistema detectou o objeto na cena e o seguiu através do rastreamento de objetos, como mostram as Figuras 5(a)-5(b). Após 30 quadros o objeto seguido foi classificado como pertencente

Classific. desejada	Classificação obtida	
	Normal	Anormal
	Normal	4
Anormal	0	0

Table 4. Classificação de comportamento da classe motocicletas.

à classe pessoas (Figura 5(c)). Em seguida, o comportamento do objeto foi classificado como Anormal, como mostra a Figura 5(d). O comportamento foi classificado como Anormal pois tanto a forma da trajetória quanto a velocidade do objeto foge ao padrão esperado, como mostra a Figura 6. A Figura 6(a) exibe a forma circular da trajetória do objeto. Neste caso, o SVI detectou duas mudanças na direção, fugindo ao padrão que é apenas uma mudança. Na Figura 6(b) é apresentado o perfil de velocidade fora do padrão do objeto, haja vista que normalmente os objetos pertencentes à classe pessoas transitam no estacionamento com velocidade média de 40 *pixels* por segundo. Pode ser observado que o objeto apresenta uma velocidade próxima de 120 *pixels* por segundo até o 3º segundo. Em seguida, a velocidade do objeto decresce, pois o mesmo está realizando uma curva. Após o objeto realizar a curva a sua velocidade volta a crescer (a partir do 5º segundo) ultrapassando a velocidade apresentada anteriormente até o 3º segundo.

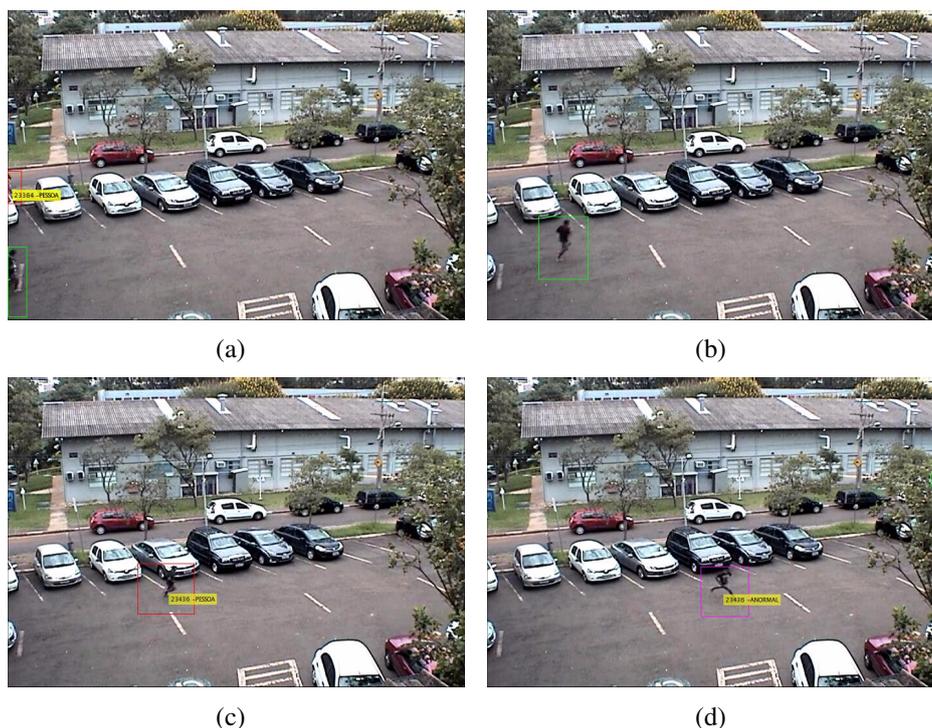


Figure 5. Resultado da classificação do Comportamento: (a) objeto aparece no estacionamento, quadro 84536; (b) objeto seguido pelo rastreamento, quadro 84550; (c) objeto é classificado como pessoa, quadro 8467; (d) classificação do comportamento como anormal, quadro 84581.

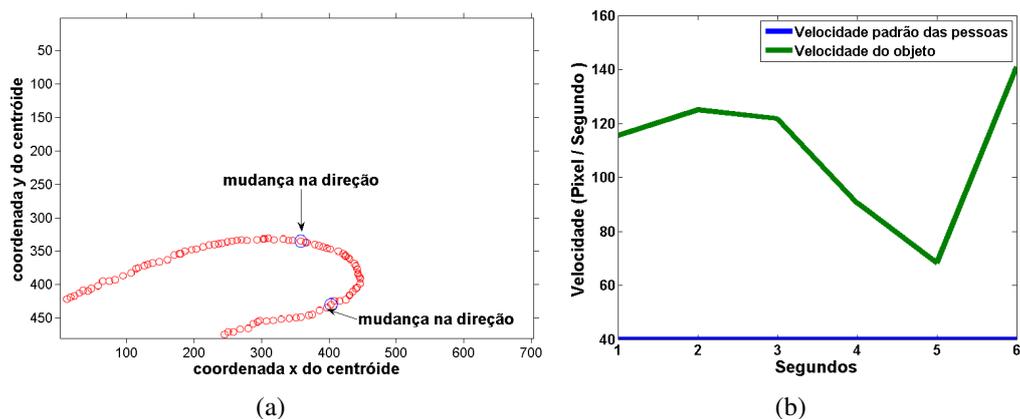


Figure 6. Descrição do comportamento do objeto: (a) forma da trajetória do objeto; (b) perfil da velocidade do objeto.

4. Conclusões

Um sistema de vigilância inteligente que detecta anomalias em um estacionamento de veículos foi apresentado neste artigo. A metodologia proposta tem como objetivo classificar o comportamento a cada instante através da análise das características de velocidade e mudança de direção extraídas a partir das trajetórias dos objetos. Como efeito da perspectiva da imagem as características dos objetos que se moviam na cena eram diferentes em cada região da imagem. Portanto, a cena foi dividida em duas regiões (região superior e inferior), onde a classificação foi realizada separadamente para cada região. Os experimentos apresentaram bom nível de precisão na detecção de comportamentos anormais. As características de velocidade e mudança de direção se mostraram eficientes na descrição do comportamentos dos objetos. O sistema apresentou tempo satisfatório de resposta na detecção de anomalias, dois segundos em média entre a ocorrência e a detecção. O estudo e implementação de outras técnicas podem ser realizadas visando a melhoria do SVI proposto: remoção de sombras com intuito de obter precisão maior na identificação de *foreground*; tratamento de oclusões para lidar com erros encontrados na construção das trajetórias dos objetos; descritores de aparência (cor e textura) visando aumentar o espaço de características dos objetos na etapa de classificação de objetos; e, abordagem de classificação de comportamento não-supervisionada para diminuir o nível de interação do usuário com o SVI. Além disso, é necessário testar o SVI em sequências de vídeos que ocorrem em outros tipos de cenários.

References

- Freitas, G. and Tozzi, C. (2010). Object tracking by multiple state management and eigen-background segmentation. *International Journal of Natural Computing Research*, 1(4):29–36.
- Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME. Journal of Basic Engineering*, 82:35–45.
- Lei, B. and Xu, L.-Q. (2006). Real-time outdoor video surveillance with robust foreground extraction and object tracking via multi-state transition management. *Pattern Recognition Letters*, 27(15):1816–1825.

- Li, H., Achim, A., and D.Bull (2012). Unsupervised video anomaly detection using feature clustering. *Signal Processing, IET*, 6(5):521–533.
- Lin, N. and Siang, C. (2012a). Trajectory learning for event recognition in video surveillance. In *Proceedings of the IEEE Image Electronics and Visual Computing Workshop*.
- Lin, N. and Siang, C. (2012b). Vision-based activities recognition by trajectory analysis for parking lot surveillance. In *Circuits and Systems (ICCAS), 2012 IEEE International Conference on*, pages 137–142.
- Morris, B. and Trivedi, M. (2008). A Survey of Vision-Based Trajectory Learning and Analysis for Surveillance. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(8):1114–1127.
- Morris, T. and Trivedi, M. (2011). Trajectory learning for activity understanding: Unsupervised, multilevel, and long-term adaptive approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(11):2287–2301.
- Ojala, T., Pietikainen, M., and Harwood, D. (1994). Performance evaluation of texture measures with classification based on kullback discrimination of distributions. *Pattern Recognition*, 1:582–585.
- Ojala, T., Pietikainen, M., and Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51–59.
- Popoola, O. and Wang, K. (2012). Video-Based Abnormal Human Behavior Recognition - A Review. *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, 46(6):865–878.
- Stauffer, C. and Grimson, W. (1999). Adaptive background mixture models for real-time tracking. *Computer Vision and Pattern Recognition*, 2:246–252.
- Suriani, N., Hussain, A., and Zulkifley, M. (2013). Sudden Event Recognition: A Survey. *Sensors - Open Access Journal*, 13(8):9966–9998.