

3 Vigilância para veículos

Este capítulo apresenta alguns algoritmos para resolver questões da vigilância eletrônica para veículos. Essas tarefas são apresentadas cada uma em uma seção separada, e são apresentadas na seguinte ordem: modelagem de veículos, detecção e remoção de sombras, detecção e remoção de oclusões, rastreamento de veículos, classificação e, por fim, a determinação de velocidade.

A partir do vídeo de entrada é realizada a subtração de fundo, apresentada no capítulo anterior, para criar um modelo de fundo responsável por gerar uma primeira estimativa dos veículos em movimento; a seguir, é realizada a detecção de silhueta e do *bounding-box* que servem de entrada juntamente com as informações da pista para a modelagem utilizada para representar geometricamente o veículo. A seguir são realizadas as etapas da remoção de sombras e de oclusão para que a classificação, a determinação de velocidade e do tracking sejam realizadas. A Figura 9 ilustra o fluxograma de relacionamento dessas etapas no trabalho.

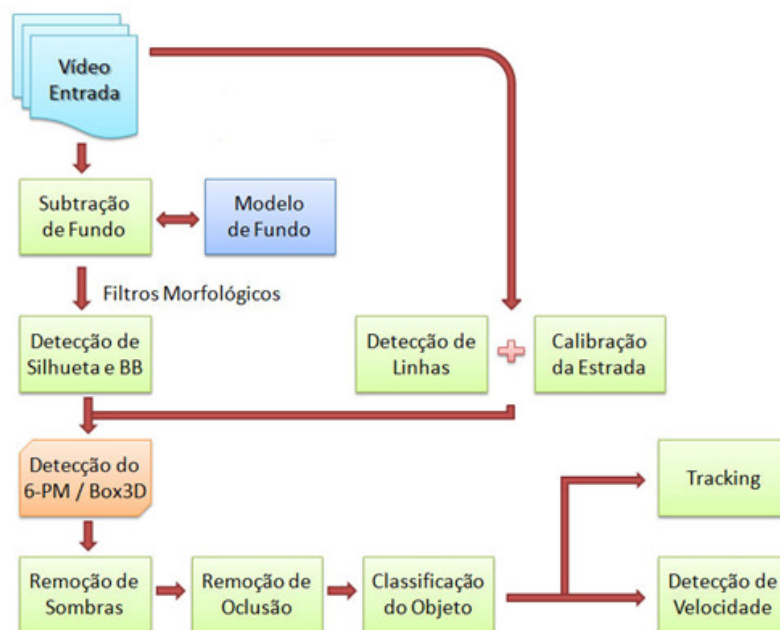


Figura 9: Fluxograma do modelo proposto

3.1

Modelagem de veículos

A modelagem dos objetos em movimento é de fundamental importância para melhor caracterizar a morfologia destes e disponibilizar mais informações para a aplicação em que atuam.

A maioria dos trabalhos da literatura utiliza a silhueta para representar os objetos em movimento, como por exemplo, os trabalhos de Haritaoglu et al. [1] e Cucchiara et al. [6]. Um dos maiores problemas de se utilizar a silhueta para esta representação é a carência de informações que ele possui sobre o objeto e a deformação que ela sofre devido à perspectiva da cena e que se agrava na presença de sombras e oclusões.

Uma modelagem mais eficaz é de fundamental importância para melhor caracterizar a morfologia dos veículos e disponibilizar mais informações ao sistema como a noção de perspectiva e volume. Trabalhos surgiram buscando a solução desse problema, como Ma e Grimson [15], que utilizam características das arestas dos veículos para auxiliar em seu rastreamento (seu modelo é ilustrado na Figura 10 (a)). Leotta e Mundy [16] desenvolveram um modelo veicular representado por uma malha tridimensional (Figura 10 (b)). Rother e Nagel [17] propuseram um modelo chamado “*Plate-Model*”, ilustrado na Figura 10 (c). Wai-Sing [18] propôs um modelo baseado na simplificação de um cubóide 3D para um modelo 2D, exemplificado pela Figura 10 (d).

Neste trabalho, é utilizada uma modelagem para o veículo sugerida por Yoneyama et al. [4] que é bastante similar à de Wai-Sing da Figura 10 (d). A forma de encontrar esta modelagem foi desenvolvida originalmente para este trabalho. Yoneyama propõe uma representação bidimensional do veículo utilizando apenas seis pontos. Este formato foi selecionado primeiramente devido à sua grande semelhança com a morfologia do veículo, pois os seis pontos interligados formam um cubóide projetado no plano segundo a Figura 11 (a). Além disso, a simplicidade matemática unida a bons resultados foi decisiva para a escolha desta modelagem chamada neste trabalho de “Representação Tridimensional Veicular”, ou simplesmente RTV.

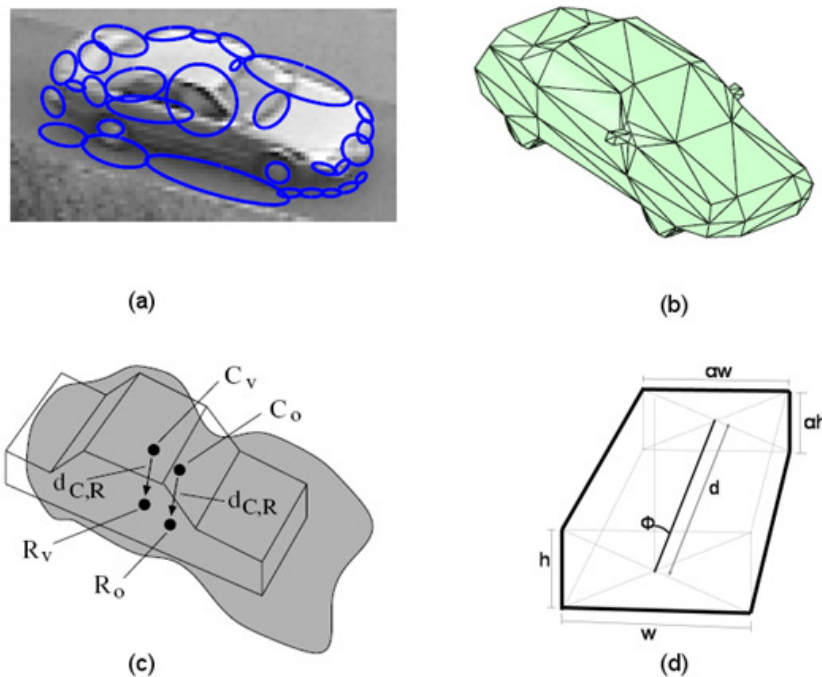


Figura 10: Modelagem para veículos de (a) Ma et al. [15], (b) Leotta et al. [16], (c) Rother et al. [17] e (d) Wai-Sing et al. [18]

Para manter a corretude desta modelagem foram feitas algumas considerações:

1. A rodovia de interesse deve ser aproximadamente plana e ocupar pelo menos um terço da imagem capturada pelo vídeo.
2. Os limites paralelos de cada faixa devem ser visíveis e facilmente extraídos da imagem.
3. O ângulo α formado entre a reta do eixo central da rodovia e a reta paralela ao eixo x deve estar entre 30 e 150 graus.
4. Os veículos das pistas selecionadas devem se movimentar no mesmo sentido, o que facilita o rastreamento na etapa seguinte.
5. As faixas não necessariamente deverão existir no vídeo, desde que sejam marcadas virtualmente.
6. É desejável que a câmera capture a rodovia com maior alcance possível da vista, de tal maneira que se possa extrair com facilidade

o comportamento dos veículos, principalmente em relação à mudança de faixas.

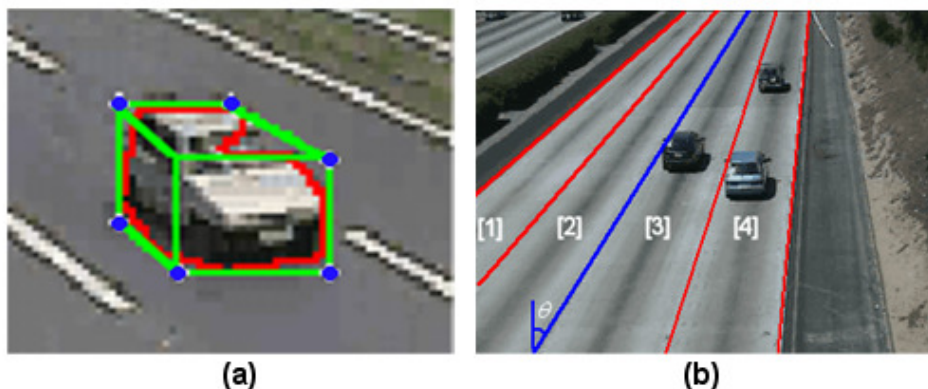


Figura 11: (a) RTV de um veículo, (b) Pistas segmentadas de uma rodovia.

A RTV fornece ao sistema as dimensões, o volume e a posição em relação à pista do veículo. Para encontrar o RTV é preciso primeiramente coletar algumas informações geométricas da rodovia e do veículo. No caso da rodovia é preciso informar por quantas pistas a rodovia é composta juntamente com duas equações de reta que a definem, gerando o resultado conforme ilustra a Figura 11 (b). A partir daí é possível identificar de qual pista cada veículo faz parte, utilizando algoritmo de ponto em polígono a partir da silhueta e cada plano formado pelas pistas.

Com estas informações é possível encontrar o modelo de seis pontos de articulação (MSPA). Esta representação é um cubóide simplificado no espaço tridimensional, o qual é composto de altura (h), largura (w) e comprimento (l) transformado no plano da imagem 2D.

A montagem do MSPA é realizada da seguinte forma: P_1 tem mesma coordenada de B_1 e P_4 de B_4 , sendo B_1 , B_2 , B_3 e B_4 os pontos que formam a *bounding-box* da silhueta conforme ilustra a Figura 12.

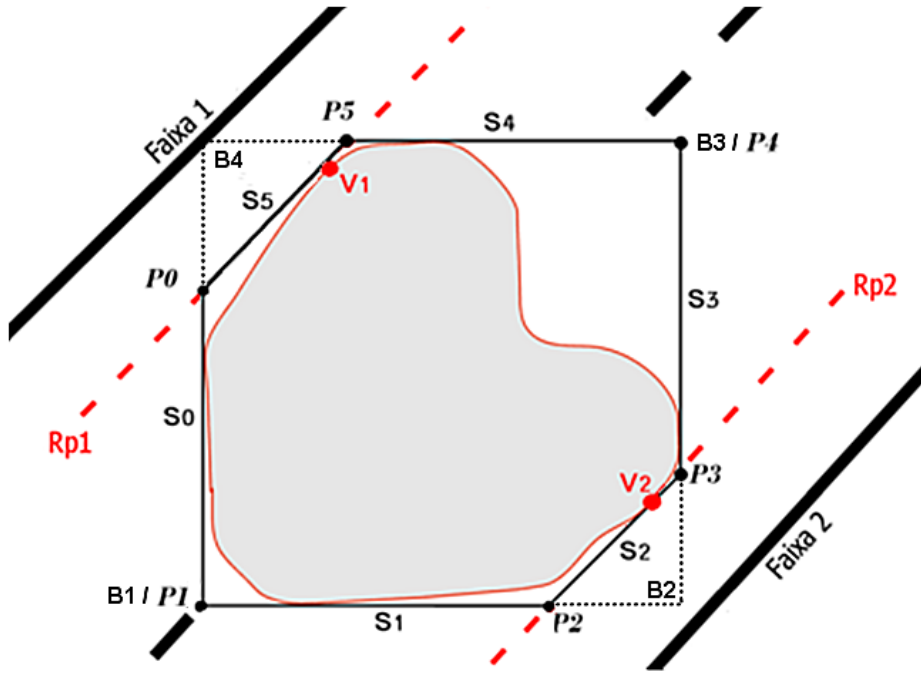


Figura 12: Representação do MSPA.

V_2 é o ponto pertencente à silhueta que é mais próximo da faixa 2 da pista do veículo. Rp_2 é a reta paralela a da faixa 2 que contém esse ponto V_2 e Rp_1 é a reta que contém V_1 paralela a faixa 1. P_0 é obtido pela interseção de Rp_1 com $\overline{B_1B_4}$.

Assim, P_2 é obtido através da interseção da reta $\overline{B_1B_2}$ do *bounding-box* com a esta reta Rp_2 . P_3 é obtido através da interseção de $\overline{B_2B_3}$ com Rp_2 e P_5 de $\overline{B_3B_4}$ com Rp_2 .

Após o cálculo desses pontos a modelagem está pronta e o seu resultado é bastante dependente da qualidade da silhueta encontrada pela segmentação de fundo, pois quanto mais perfeita ela for, mais próxima da realidade o MSPA será. Além disso, ela ainda pode incluir sombras e oclusões que são removidas através dos algoritmos descritos nas seções seguintes.

3.2

Detecção e remoção de sombras

A sombra é uma região de escuridão gerada pela presença de um objeto opaco localizado entre uma fonte de luz e esta região. Ela em geral é associada ao objeto de interesse, podendo causar distorções em sua forma, união com outros objetos distintos e até mesmo a perda destes, fazendo com que a sua remoção se torne uma etapa fundamental para a corretude da detecção dos veículos em movimento.

Muitos pesquisadores apostaram nas propriedades bem particulares de cor da sombra para detectá-las e removê-las. Cucchiara et al. [6] investigaram as propriedades da sombra no espaço de cor HSV e Salvador et al. [19] propuseram o espaço de cor RGB normalizado para esta tarefa. Porém, ambos os métodos são extremamente dependentes de seleção de limiares e, por realizarem uma filtragem por pixel, acabam gerando resultados bastante ruidosos na segmentação.

3.2.1

Abordagem

Inicialmente este trabalho procurou resolver o problema utilizando uma abordagem de cor para remover as sombras, principalmente pelas características tão distintas das componentes dos espaço IHSV. Nessa primeira tentativa, utilizou-se a abordagem proposta no trabalho de Cucchiara et al. [6], na qual se utiliza a Equação 10, determina se um pixel é sombra se:

$$SP_i(x, y) = \begin{cases} 1, & \text{se } \alpha \leq \frac{I_i(x, y)V}{B_i(x, y)V} \leq \beta \text{ e} \\ & |I_i(x, y)H - B_i(x, y)H| \leq \tau_H \text{ e} \\ & |I_i(x, y)S - B_i(x, y)S| \leq \tau_S \text{ e} \\ 0, & \text{caso contrário} \end{cases} \quad (10)$$

Onde α é um limiar de valores de 0.75 a 0.85; β é outro limiar com valores de 0.9 a 0.97, τ_H e τ_S também são limiares com o valor de 0.15; $I_i(x, y)H$ é o valor observado de matiz; $I_i(x, y)S$ da saturação e $I_i(x, y)V$, de luminância;

$B_i(x,y)H$ é o valor da média da matiz ao longo do treinamento; $B_i(x,y)S$, da saturação; e $B_i(x,y)V$, da luminância.

Caso o resultado de SP nesta equação seja 1, o pixel será classificado como sombra; caso contrário como movimento. A Figura 13 ilustra o resultado encontrado da remoção de sombra por cinco diferentes algoritmos, sendo que no grupo das primeiras quatro imagens a sombra é representada pela cor azul e o objeto em movimento pela cor vermelha. A imagem representada pela sigla “SP” (*statistical parametric*) é uma abordagem adotada pelo algoritmo de Mikic et al. [20], já a imagem representada pela sigla “SNP” (*statistical nonparametric*) tem como exemplo o algoritmo de Horprasert et al. [21] e por fim as imagens representadas pelas siglas “DNM1” e “DNM2” (*deterministic non-model*) têm como exemplo o algoritmo de Cucchiara et al. [6] e Stauder [22].

Na última imagem da Figura 13, encontra-se o resultado obtido pela implementação neste trabalho da equação do trabalho de Cucchiara et al. [6], onde a sombra é representada pela cor cinza, o fundo pela cor preta e os objetos em movimento pela cor branca.

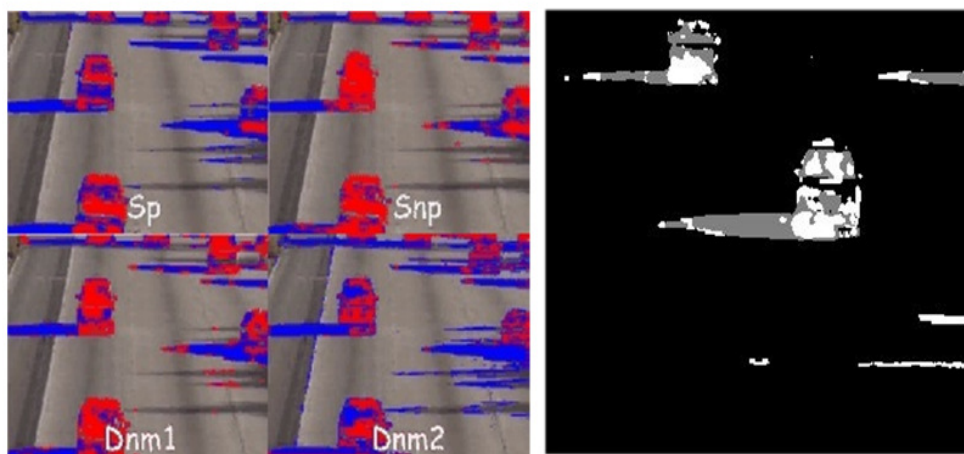


Figura 13: Resultado da segmentação da sombra utilizando abordagem de cor.

Como se pode observar em todas essas imagens, os resultados obtidos pelos algoritmos apresentam muitas regiões de falhas e imperfeições que podem ser irrelevantes, dependendo da aplicação a que essa detecção de movimento se

destina. No caso desta dissertação, como se tem tarefas do sistema que demandam precisão nas silhuetas, como por exemplo a classificação do tipo de cada veículo, esses resultados se mostraram insatisfatórios. Além disso, muitas vezes as regiões de sombra projetadas não foram removidas por completo, levando a uma detecção de novos objetos errônea.

Pensando nesses problemas, buscou-se uma solução fora dessa idéia de classificar cada pixel da imagem como sombra ou não. Analisando o trabalho de Yoneyama et al. [4], percebeu-se que resultados melhores poderiam ser obtidos com uma abordagem que também utiliza informações geométricas do veículo e da pista.

Conforme afirmado na seção anterior, o problema de vigilância para veículos fornece informações geométricas que são extremamente relevantes para auxiliar na remoção de sombras. Utilizando esta abordagem geométrica, Yoneyama et al. [4] modelam a sombra no plano 2D segundo a Figura 14 (a). Nesta figura, encontra-se um objeto no formato de um poste representado pela cor vermelha e sua respectiva sombra representada pela cor cinza, e ela é decomposta em três vetores: sh , sl e sw para representar a “altura do objeto”, o “comprimento da sombra” e a “largura da sombra”, respectivamente.

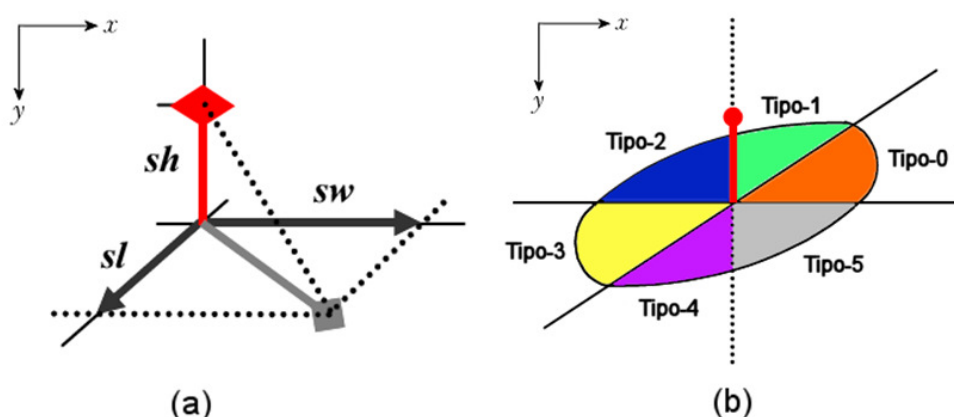


Figura 14: (a) Modelo da sombra; e (b) os seis tipos de sombras projetadas para cada direção de iluminação. Retirado de Yoneyama et al. [4].

Outra particularidade do problema é que somente são consideradas cenas de ambientes diurnos. Portanto, a fonte global de iluminação é a solar. Assim, a primeira etapa do algoritmo é buscar em qual dos seis lados do MSPA a luz incide. Esta aresta é chamada de S_f . A partir daí, sabe-se que no caso de existirem sombras na cena, elas serão geradas respeitando a projeção do lado de incidência. Yoneyama et al. [4] sugerem a projeção da sombra no plano 2D pode adquirir seis tipos distintos, conforme a Figura 15. Esta figura ilustra o relacionamento de M' representado pela cor azul que é o MSPA de um veículo e de M'_s , representado pela cor laranja, que é o MSPA deste veículo unido a sua sombra.

Então, para determinar qual aresta do modelo a luz incide, Yoneyama et al. [4] afirmam que, ao examinar cada um dos seis lados de M' , aquele que tiver a maior luminância média será o S_f , conforme ilustrado na Equação 11, onde N é o número de pixels em $S_{j,n}$ e lum_k representa o valor da luminância na posição (x,y) da imagem capturada.

$$L_n = \frac{1}{N} \sum_{k=0}^{N-1} lum_k, \quad \{k | (x,y) \in S_{j,n}\},$$

$$L_{max} = \max(L_n : S_{j,n} \in M'_j), \quad (11)$$

A última etapa do algoritmo é desenvolvida a partir do tipo de sombra identificado, os parâmetros do modelo da sombra S_l , S_h e S_w podem ser determinados. Yoneyama et al. [4] afirmam que é possível obter estes parâmetros através do comprimento da sombra projetada. Para isso, analisa-se para cada aresta a curva de luminância pela distância. Assim onde ocorrer o decrescimento mais acentuado desta curva será o ponto de quebra do objeto e determinará a região onde inicia a sombra. Neste trabalho, a abordagem para detectar e remover as sombras foi idêntica à utilizada por Yoneyama et al. [4].

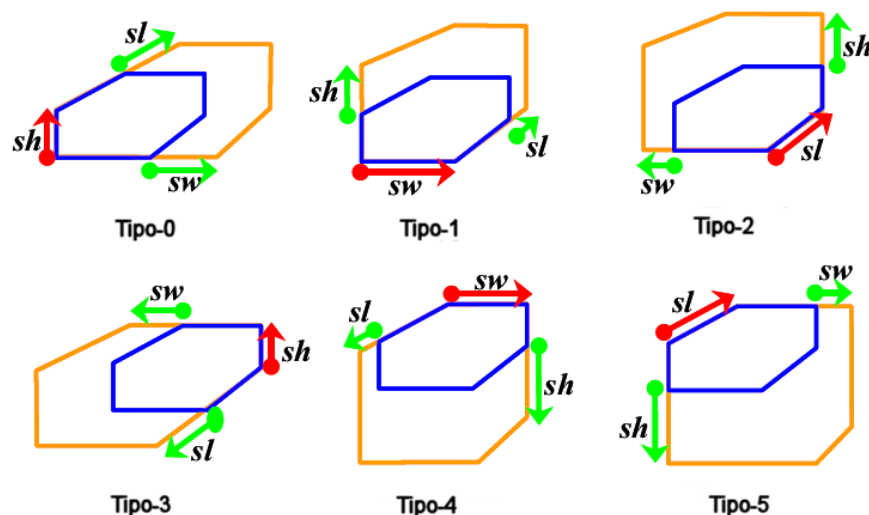


Figura 15: Seis tipos possíveis da sombra em um MSPA (Yoneyama et al. [4]).

3.2.2 Resultados parciais

Foram realizados testes baseados em um “*ground-truth*”, obtidos manualmente, similar ao da subtração de fundo para 3 vídeos para testar a corretude do algoritmo proposto. A Figura 16 ilustra o resultado para o vídeo “Highway 2” onde a caixa representada pela cor verde representa o MSPA original do veículo e a caixa branca a sua respectiva representação com a remoção da sombra.



Figura 16: Resultado da segmentação das sombras utilizando a abordagem geométrica

O algoritmo de detecção e remoção de sombras foi realizado em duas etapas. A primeira foi para verificar a corretude da detecção da posição da luz em relação ao MSPA do veículo. Para isso, analisou-se a média dos de acerto de cada veículo em que na maior parte de sua presença no vídeo a aresta correta tenha sido identificada. O resultado encontrado para essa etapa foi de 80% de acerto médio para três diferentes vídeos. Esse resultado mostra que esta abordagem é eficiente para se detectar a posição da luz em relação a cada veículo.

A segunda etapa da fase de treinamento foi para verificar a corretude da segmentação das regiões em sombra. E para isso foram utilizadas as mesmas métricas de Karaman et al. [14] utilizadas na subtração de fundo só que em relação ao *ground-truth* da sombra. Avaliando o quesito F o resultado médio obtido dos mesmo três vídeos foi de 77%.

3.2.3

Conclusões parciais

O resultado desse algoritmo mostrou-se satisfatório na maioria dos casos. Devido à sua dependência geométrica ao MSPA, esta região sofreu algumas oscilações conforme a movimentação do carro ao longo da rodovia gerando oscilações na região em sombra. A tabela completa dos resultados se encontra no apêndice B.

3.3 Transformação projetiva

Um grande problema a ser solucionado nas próximas etapas é a deformação da projeção que a imagem capturada pela câmera sofre. Ao passo que se analisa a projeção dos veículos no vídeo, observa-se que há divergência na dimensão destes. Por exemplo, um ônibus pode ser confundido com um carro próximo se ele estiver longe da câmera. Assim, o objetivo é encontrar uma transformação projetiva que levem pontos da imagem da rodovia do mundo tridimensional para o plano 2D da imagem.

3.3.1 Abordagem

Neste trabalho, a transformação projetiva é calculada utilizando-se um algoritmo que tem como entrada quatro pontos da pista, o comprimento, a largura e o número de faixas contidas na cena. A homografia é definida por uma matriz H que possui 9 parâmetros e é convenientemente descrita através da Equação 12.

$$H \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} u_s \\ v_s \\ t_s \end{bmatrix} \quad (12)$$

Supondo-se que o plano da pista é dado por $z=0$, em que $P = (x,y,0,w)$ é um ponto do mundo da cena e $P = (u,v,t)$ um ponto no plano da imagem. Dado que, neste trabalho, a escala geral não é relevante, o último elemento da matriz é 1, levando a oito parâmetros necessários para encontrar a homografia H . Cada ponto de calibração leva a duas equações, logo faz-se necessário quatro pontos não-colineares para se encontrar a solução exata dos oito elementos desconhecidos de H .

O mapeamento entre as coordenadas da imagem (P') e do plano da pista (P) será denominado conforme a Equação 13.

$$\begin{aligned} P' &= HP \\ P &= H^{-1}P' \end{aligned} \tag{13}$$

Assim, cada ponto do plano da imagem de entrada pode ser convertido para o plano da pista selecionado como por exemplo a região delimitada pelas linhas vermelhas da Figura 17.

3.3.2 Resultados parciais

O resultado da transformação projetiva do vídeo “Highway 2” é ilustrado pela Figura 17 (a) e do vídeo “Velocidade” pela Figura 17 (b).

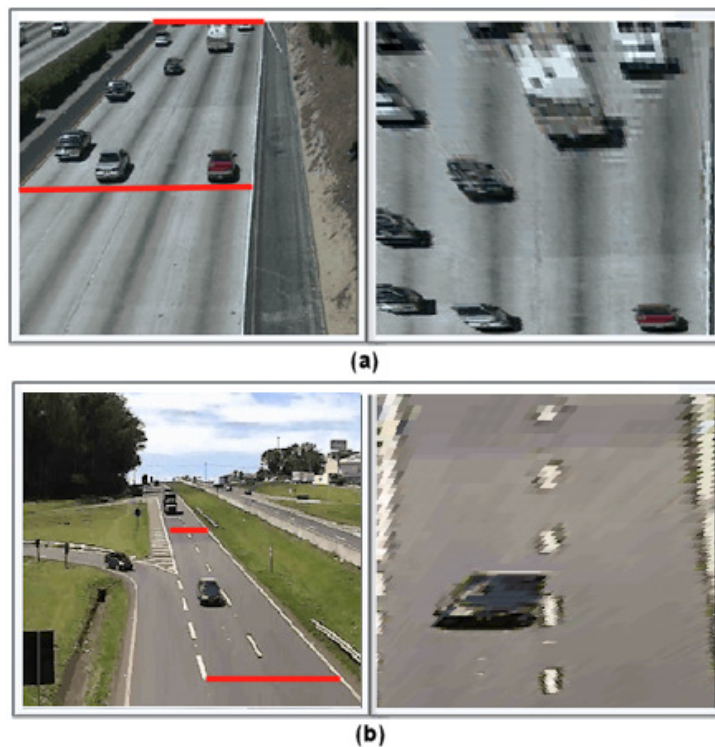


Figura 17: Imagem da cena real e da sua respectiva projeção no plano para os vídeos: (a) “Highway 2” e (b) “Velocidade”

Observando o resultado pela figura anterior nota-se um pouco de deformação na região mais distantes no vídeo, o que naturalmente é causado pela falta de informação em relação a regiões mais próximas. A transformação projetiva foi testada para quatro vídeos diferentes através da distância euclidiana entre as coordenadas originais da imagem passada para o cálculo da homografia e destes calculados pela transformação projetiva.

3.4

Detecção e remoção de oclusão

A oclusão entre veículos gera distorção em suas formas originais e a união destes levando a erradas detecções a respeito de suas trajetórias, classificação e identificação. Pensando neste problema, Koller et al. [23] utilizam a interseção das regiões de profundidade ordenadas associadas aos objetos. Yoneyama et al. [4] utilizam informações geométricas da pista e dos veículos, a partir da relação das imagens geradas por múltiplas câmeras sincronizadas.

3.4.1

Abordagem

Neste trabalho foi adotado o algoritmo de Kanhere [24], em que a idéia central é a detecção da base frontal de cada veículo, ou “*Vehicle Base Front*” (VBF). Esta decisão foi tomada porque a base frontal ou traseira de um veículo dificilmente é ocluída por completo ao longo da cena, o que é bastante vantajoso em relação aos outros métodos que buscam esta mesma segmentação, analisando todo o contorno do veículo.

Kanhere [24] propõe que para cada imagem de entrada de resultado da subtração de fundo, como na Figura 18 (b), é realizado o operador de diferença na direção vertical para gerar uma nova imagem binária, segundo a Equação 14, onde $V(x,y)$ é um pixel da nova imagem gerada, $F(x,y)$ é o pixel do quadro atual da subtração de fundo e $F(x,y+1)$ é o pixel do próximo quadro somente na direção vertical.

$$\begin{aligned}
 V(x,y) &= 1, & \text{se } F(x,y) - F(x,y+1) > 0 \\
 V(x,y) &= 0, & \text{Caso contrário}
 \end{aligned}
 \tag{14}$$

Esse operador é utilizado somente no caso dos veículos estarem se movimentando em direção à câmera, ou seja, que a componente y esteja sempre diminuindo ao passo que o veículo se movimenta e o seu resultado será a base frontal de cada veículo. No caso do seu movimento ser de afastamento da câmera, busca-se a base traseira e o operador de diferença é calculado segundo a Equação 15.

$$\begin{aligned} V(x, y) &= 1, \quad \text{se } F(x, y) - F(x, y - 1) > 0 \\ V(x, y) &= 0, \quad \text{Caso contrário} \end{aligned} \quad (15)$$

O resultado desta operação é ilustrado na Figura 18 (c), que por sua vez é projetada no plano da pista utilizando a homografia encontrada. Com objetivo de realçar o resultado no eixo x da imagem, é aplicado um filtro morfológico com o kernel K da Equação 16.

$$K = \begin{bmatrix} -1 & -1 & -1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad (16)$$

3.4.2 Resultados parciais

O resultado desta operação é ilustrado pela Figura 18 (e). Assim, a próxima etapa deste processo é realizar uma análise dos componentes conexos da imagem e identificar regiões próximas através de um limiar. A Figura 18 (f) ilustra esse resultado onde cada grupo foi identificado por uma cor diferente, onde supostamente cada cor desta seria um veículo diferente. Com esta informação cada MSPA será limitado ou criado a partir da reta encontrada.

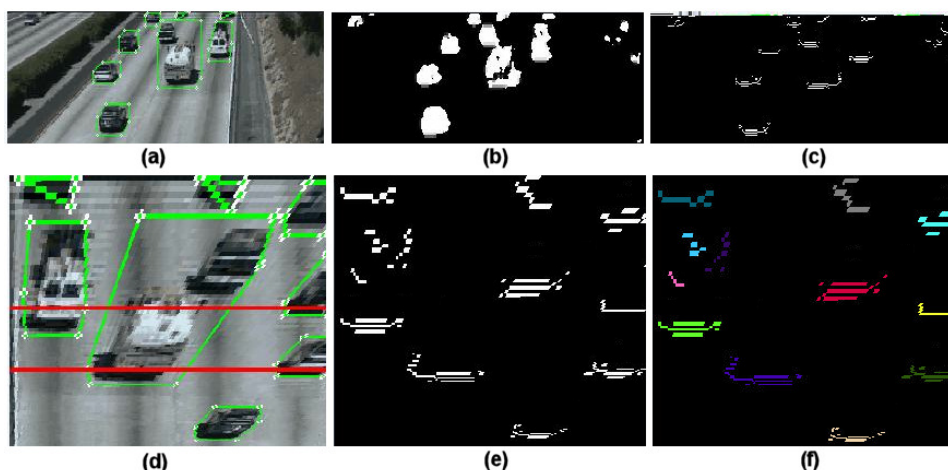


Figura 18: Etapas da remoção de oclusão.

Foram realizados testes para essa etapa em diferentes tipos de vídeos, a porcentagem de acerto médio do algoritmo foi de 85% para o total de 24 veículos. O resultado completo para essa etapa encontra-se no Apêndice D.

3.4.3 Conclusões parciais

O algoritmo falha quando o ângulo da câmera com o chão é pequeno, gerando oclusões em que o carro de trás fica “escondido”. Neste caso, como a base frontal detectada é nenhuma ou muito pequena, fazendo com que seja normalmente descartada nos filtros morfológicos. Apesar disso, os resultados obtidos foram satisfatórios, tanto para a detecção da oclusão quanto para a sua remoção.

3.5

Rastreamento de Veículos

O processo de rastreamento em vídeo ou “*video tracking*” tem como objetivo identificar objetos em movimento e manter a coerência dessa identificação ao longo de toda a cena. Além disso, o rastreamento permite a análise da trajetória dos veículos, podendo assim detectar colisões na estrada, avaliar o fluxo do tráfego e o estudo do comportamento do motorista.

Na literatura são apresentadas diversas formas de realizar o rastreamento, como o rastreamento de modelos tridimensionais por Baker e Sullivan [25], o de regiões por Kilger [26], o de contornos ativos (*snakes*) por Koller et al. [23] e o de características por Kanhere [24] e Yoneyama et al. [4].

3.5.1

Abordagem

Neste trabalho, a abordagem escolhida para realizar o rastreamento foi a por característica. Este tipo de rastreamento busca particularidades do objeto de interesse como pontos e linhas. A decisão de utilizar esta abordagem se deu pelo fato de já se possuir dados geométricos do veículo, além da simplicidade deste método.

O algoritmo proposto particulariza bastante o problema, pois os veículos sempre mantêm uma trajetória unidirecional e as interseções são bastante incomuns. Por isso, as possíveis trajetórias dos veículos em uma rodovia são extremamente previsíveis e fáceis de rastrear levando a uma simplificação do problema. Pensando nisso, nenhuma técnica complexa de *tracking* foi utilizada; o algoritmo busca apenas a relação dos objetos do quadro atual com o anterior a partir da distância euclidiana dos centros dos MSPAs.

Considerando que V_c é este centro de um veículo, ele pode ser encontrado segundo a Equação 17, onde V_p é o número de pixels pertencentes ao MSPA do veículo, e V_{xi} representa a coordenada x do i-ésimo pixel, enquanto V_{yi} representa a coordenada y.

$$V_c = \left[\sum_{i=1}^{N_p} \frac{V_{xi}}{N_p}, \sum_{i=1}^{N_p} \frac{V_{yi}}{N_p} \right] \quad (17)$$

Portanto, considerando V_{CM} e V_{CN} o centro do veículo no quadro atual e no seu consecutivo, respectivamente, a distância é dada pela Equação 18.

$$E_d(V_{CM} - V_{CN}) = \sqrt{(V_{CM1} - V_{CN1})^2 + (V_{CM2} - V_{CN2})^2} \quad (18)$$

Assim, algumas regras foram estipuladas para manter a corretude do processo, como mostra a Figura 19. São elas: veículos somente serão rastreados se estiverem na região de interesse da rodovia; um veículo ao entrar na cena pela primeira vez, adquire o estado “novo”; caso o rastreo do veículo com estado de “novo” permaneça com sucesso por mais de dez quadros consecutivos, ele receberá um novo status de “rastreado”; se o rastreo se perder por mais de dez quadros consecutivos a partir de um estado novo, ou por mais de 30 quadros consecutivos a partir de um estado “rastreado”, o veículo será classificado como “eliminado” e retirado da lista de rastreios; a partir do estado de “rastreado”, o veículo recebe um identificador único em relação a todos os outros, ao longo da cena; admite-se que antes da realização do rastreamento toda e qualquer oclusão já foi removida.

3.5.2 Resultados parciais

O teste principal desta etapa foi em relação à contagem de quantos veículos passaram durante todo o vídeo. Esse teste foi realizado para quatro vídeos com os resultados exibidos no Apêndice E. Os vídeos utilizados foram: “Highway”, com 49 carros e o total de acertos do algoritmo foi 71%; “Velocidade” com 222 carros e 74% de acerto; “Passarela 1”, com 2 carros e 50% de acerto; e por fim, “Passarela 2”, com 29 carros e 86% de acerto.

Esse resultado mostrou que o algoritmo falha quando ruídos ou falhas na remoção de sombra e oclusão ocorrem, pois a maioria dos erros desta contagem foi justamente reconhecer mais carros do que se deveria, isso fica bem claro no caso do vídeo “Velocidade”, onde o algoritmo reconheceu 280 veículos, quando na realidade só passaram 222.

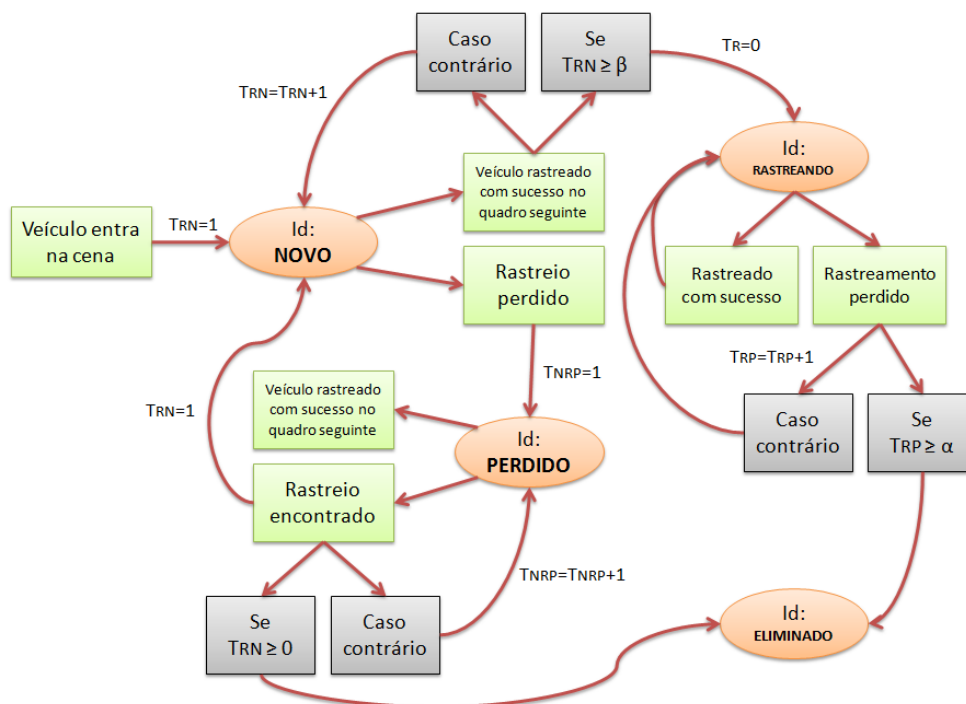


Figura 19: Fluxograma do algoritmo de rastreamento.

3.5.3 Conclusões parciais

De uma maneira geral o rastreamento proposto, apesar de sua simplicidade, foi bastante eficiente, eliminando a necessidade de outros algoritmos complexos como, o filtro de Kalman.

Ele mostrou também uma tendência de realizar mais detecções em alguns vídeos e menos detecções em outros.

Essa característica depende muito do vídeo em análise. Por exemplo, no vídeo “velocidade”, a iluminação varia muito ao longo do tempo, gerando erros na subtração de fundo, ocasionando uma segmentação da silhueta dos veículos levando assim mais detecções do que o correto. No caso do vídeo “highway2” já ocorre o oposto, são realizada menos detecções do que o correto, pois neste caso a posição da câmera fez com que do início da pista ao final, os veículos seguirem tão próximos que não foi possível realizar uma segmentação destes.

Outro quesito importante para o rastreamento é a trajetória. Porém, neste trabalho isso não foi analisado, pois as trajetórias dos veículos são bem definidas e em uma única direção e sentido.

3.6

Classificação de veículos

Utilizando todas as informações obtidas até aqui é possível avaliar as características do veículo e classificá-lo de acordo com o seu tipo. Essa tarefa permite que seja determinado qual tipo de veículo é predominante em uma rodovia, auxiliando em estudos como: avaliação da infra-estrutura e da segurança, engenharia de tráfego, entre outros.

Poucos trabalhos na literatura discutem o tema de classificação de veículos. Dentre esses poucos, pode-se citar Kanhere [24] que classifica os veículos como carro ou caminhão utilizando informações de altura e largura da caixa envolvente destes e Huang e Liao [5] que identifica sete diferentes tipos de veículos através de uma “classificação hierárquica”, onde eles são ordenados a partir de características das silhuetas e de limiares.

3.6.1

Abordagem

Neste trabalho, os veículos são classificados em três tipos: moto, carro ou de grande porte (como por exemplo, ônibus e caminhão). Através do fato que as características geométricas desses veículos são tão distintas em adição à posse da homografia, é possível que se realize esta etapa de forma bem simples.

A principal idéia do algoritmo desenvolvido neste trabalho é estimar a área de cada veículo em pixels e classificá-los de acordo com as informações reais da pista e dos veículos.

A transformação projetiva apresentada na seção anterior permitiu que os veículos ficassem aproximadamente com tamanho proporcional ao da pista, sendo removida a deformação que a perspectiva original gera no formato destes.

Ao se ter o conhecimento da largura e do comprimento da pista é possível estimar a área em pixel que uma moto, um carro e um caminhão ocupam na pista. Para isso foi considerado que o comprimento e a largura reais possuem em média

2.0m e 0.8m respectivamente no caso da moto, 4m e 1.6m para o carro e 6.3m e 2.2m ou maior para um veículo de grande porte.

Assim, a classificação pode ser escrita segundo a Equação 19, onde α , β e γ são limiares calculados de acordo com o erro da homografia, $A(v)$ é a área em pixel do veículo v que se deseja classificar, $A_m(v)$ é a área refencial da moto em pixels e $A_c(v)$ é a área do carro e $A_{gp}(v)$ de veículos de grande porte.

$$C(v) = \begin{cases} \text{Moto}, & \text{Se } A(v) \leq \alpha A_m(v) \\ \text{Carro}, & \text{Se } A(v) \leq \beta A_c(v) \\ \text{Grande porte}, & \text{Se } A(v) \geq \gamma A_{gp}(v) \end{cases} \quad (19)$$

3.6.2 Resultados parciais

Apesar da simplicidade do método, os resultados obtidos nesta etapa foram bastante satisfatórios. Os testes foram realizados para três diferentes cenários e a média de acerto da classificação foi de 92%. A maioria das vezes em que o resultado errou a sua classificação foi quando ocorreram eventos mal detectados de sombras, oclusões ou uma má subtração de fundo.

3.6.3 Conclusões parciais

O resultado mostrou que somente a informação da área dos veículos foi suficiente para diferenciá-los, dado que principalmente as etapas da subtração de fundo e a transformação projetiva obtêm bons resultados.

3.7

Determinação da velocidade

A determinação de velocidade é comumente obtida nas estradas por ondas sonoras disparadas por dispositivos eletrônicos. A presente seção demonstra que esta mesma tarefa pode ser realizada de forma mais barata e simples, além de ter uma instalação e manutenção mais facilitadas.

Grammatikopoulos et al. [27] utilizam a geometria da pista e pontos de fuga. O VISATRAM de Zhu et al. [28] utilizam a vista panorâmica e um plano epipolar para cada pista da rodovia. Bauer et al. [29] apresentam algoritmos para determinar a velocidade através de um sensor assíncrono temporal de contraste. Zhongzhen et al. [30] utilizam conhecidas marcações da pista, chamadas de “*virtual loops*”, que são as linhas pontilhadas que dividem as faixas da rodovia e de suas distâncias.

3.7.1

Abordagem

O algoritmo desenvolvido neste trabalho para determinar a velocidade dos veículos é relativamente simples e utiliza a tradicional equação de distância sobre o tempo.

Para determinar a distância são necessários, da mesma forma que para a classificação, as informações das medidas reais da pista convertidas para valores em pixel. Assim, uma vez determinado o trecho de interesse, ilustrado por exemplo pela região entre as duas linhas vermelhas da Figura 20, e sabendo a sua distância real em metros da mesma, é possível calcular essa equivalência em pixels.

A partir do primeiro quadro que o veículo seja detectado com sucesso dentro da região de interesse, carro de seleção amarela na Figura 20, a sua posição e o número do quadro atual do vídeo são anotados e o mesmo é feito após sua saída da região. Essa distância percorrida em pixels na projeção do plano é convertida para metros através da homografia. E o tempo em número de quadros é

multiplicado pela razão de $1/N$ para obter-se o valor em segundos, em que N é o número de quadros por segundo no qual o vídeo foi gravado.

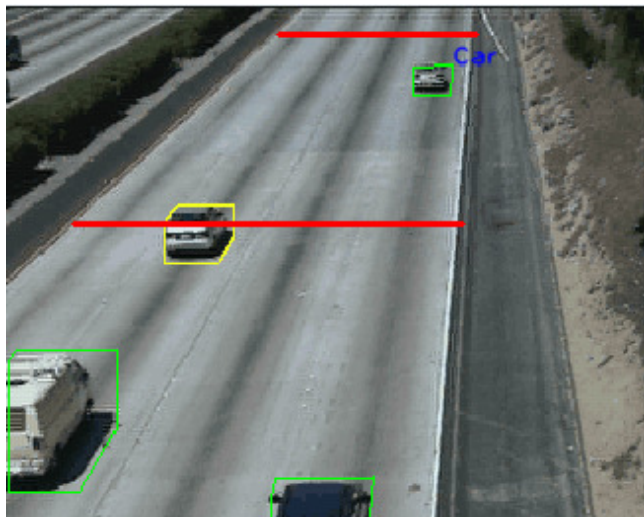


Figura 20: Área de seleção para medição da velocidade.

3.7.2 Resultados parciais

Os testes do algoritmo foram realizados em um vídeo de 18 minutos gravados na rodovia SP-310 (via Washington Luiz), próximo ao km 235. Foi utilizado um radar com precisão de ± 10 km para adquirir os valores reais para comparação com o algoritmo. O Apêndice G exibe a tabela dos resultados das velocidades comparativas de cada veículo. A média do erro da velocidade encontrada pelo algoritmo em relação aos dados reais foi de ± 11.90 km/h, o que forneceu o valor de acerto em 87% e um erro máximo de 56 km/h. A Figura 21 ilustra o gráfico da porcentagem de acerto da velocidade ao longo do vídeo em relação aos dados obtidos pelo radar.

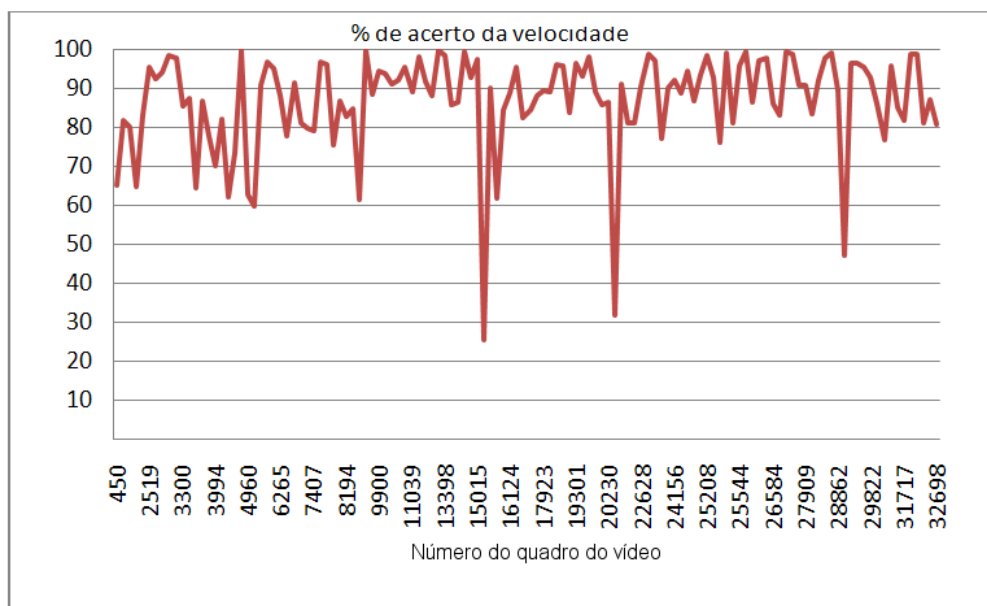


Figura 21: Gráfico da porcentagem de acerto da velocidade ao longo do vídeo em relação aos dados obtidos pelo radar

3.7.3 Conclusões parciais

O algoritmo falha quando a detecção do veículo demora para ocorrer dentro da área de seleção da pista, levando a amostras de tempo e espaço muito baixos, consecutivamente aumentando o erro da velocidade. Porém, analisando esses resultados, conclui-se que o algoritmo desenvolvido mostrou-se bastante preciso em relação aos dados coletados pelo radar, estando na maior parte do tempo dentro da faixa de precisão do aparelho.