

# **MÉTODO PARA CONTAGEM AUTOMÁTICA DE PESSOAS EM MULTIDÕES UTILIZANDO DRONES**

Nayara Cielly Freire do Ramo



CENTRO DE INFORMÁTICA  
UNIVERSIDADE FEDERAL DA PARAÍBA

João Pessoa, 2017



Nayara Cielly Freire do Ramo

# **MÉTODO PARA CONTAGEM AUTOMÁTICA DE PESSOAS EM MULTIDÕES UTILIZANDO DRONES**

Monografia apresentada ao curso Engenharia de Computação do Centro de Informática, da Universidade Federal da Paraíba, como requisito para a obtenção do grau de Bacharel em Engenharia de Computação.

Orientador: Alisson Brito

Junho de 2017

Ficha Catalográfica elaborada por  
Rogério Ferreira Marques CRB15/690

R175m Ramo, Nayara Cielly Freire do.  
Método para contagem automática de pessoas em multidões  
utilizando drones / Nayara Cielly Freire do Ramo. – João Pessoa, 2017.  
52p. : il. -

Monografia (Bacharelado em Engenharia de Computação) –  
Universidade Federal da Paraíba - UFPB.  
Orientador: Prof. Dr. Alisson Vasconcelos de Brito.

1. Informática. 2. Contagem automática de pessoas. 3. Drone. 4.  
Aeronave não tripulada. I. Título.

UFPB/BSCI

CDU: 004 (043.2)



CENTRO DE INFORMÁTICA  
UNIVERSIDADE FEDERAL DA PARAÍBA

Trabalho de Conclusão de Curso de Engenharia de Computação intitulado **Método para contagem automática de pessoas em multidões utilizando drones** de autoria de Nayara Cielly Freire do Ramo, aprovada pela banca examinadora constituída pelos seguintes professores:

---

Prof. Dr. Alisson Vasconcelos de Brito  
Universidade Federal da Paraíba

---

Prof. Dr. José Antônio Gomes de Lima  
Universidade Federal da Paraíba

---

Prof. Dr. Tiago Pereira do Nascimento  
Universidade Federal da Paraíba

---

Coordenador(a) do Curso  
Thaís Gaudêncio do Rêgo  
CI/UFPB

João Pessoa, 13 de junho de 2017



## **AGRADECIMENTOS**

Agradeço primeiramente a meus pais por todo apoio, carinho e incentivo que recebi até hoje, e também a todos da minha família que sempre torceram por mim. Agradeço ao Professor Alisson Brito pela orientação no desenvolvimento deste trabalho e também ao aluno Josué da Silva. Agradeço a todos os professores que contribuíram para a minha formação acadêmica e aos meus colegas de curso pelo companheirismo e apoio durante todo o curso. Agradeço também a Polícia Militar da Paraíba que cedeu imagens úteis para a realização deste trabalho.

## RESUMO

Este trabalho apresenta um método para contagem automática de pessoas em multidões através de fotos capturadas por um drone, com o objetivo de substituir a contagem manual. A contagem é feita em solo e o método proposto detecta cabeças utilizando um classificador Adaboost que utiliza características chamadas Haar Features. Para captura de imagens de até 30 metros os resultados foram satisfatórios, para alturas acima disso, os resultados precisam melhorar.

**Palavras-chave:** contagem automática de pessoas, contagem de multidões drones.

## **ABSTRACT**

This work presents a method for crowd counting that combines information from multiple photos captured by a drone, in order to replace manual counting. The counting is done in soil and the proposed method detects heads using an Adaboost classifier that uses features called Haar Features. For capturing images below 30 meters the results were satisfactory, for heights above that, the results need to improve.

**Key-words:** people automatic counting, crowd counting, drones.

## LISTA DE FIGURAS

Figura 2.1: Quatro tipos de características retangulares Haar.....	24
Figura 2.2: Oclusões dificultam o cálculo do número de pessoas.....	25
Figura 2.3: (a) Exemplo de oclusão estática.....	26
Figura 2.3: (b) Exemplo de oclusão dinâmica.....	26
Figura 3.1 Processo Metodológico.....	30
Figura 3.2 Exemplos de Imagens Positivas.....	32
Figura 3.3: Características Haar.....	33
Figura 4.1(a): Resultados para uma foto tirada a aproximadamente 15 metros de altura.....	40
Figura 4.1(b): Resultados para uma foto tirada a aproximadamente 15 metros de altura.....	41

Figura 4.2: Resultados para uma foto tirada a aproximadamente  
25 metros de altura.....42

Figura 4.3: Simulação de uma foto tirada a aproximadamente  
35 metros de altura.....43

Figura 4.4: Resultados para uma foto tirada a aproximadamente  
35 metros de altura.....44

Figura 4.5 Resultados para uma Imagem  
enviada pela PM da Paraíba.....45

## **LISTA DE TABELAS**

Tabela 1: Resultados.....	45
---------------------------	----

## **LISTA DE ABREVIATURAS**

PM	–	Polícia Militar
VANT	–	Veículo aéreo não tripulado
CLP	–	Controladores Lógicos Programáveis
OpenCV	–	Open Source Computer Vision Library
LBP	–	Local binary patterns



# SUMÁRIO

<b>1.</b>	<b>INTRODUÇÃO</b>	<b>17</b>
1.1.	MOTIVAÇÃO	17
1.2.	PROBLEMA	19
1.1.1.	Objetivo geral	19
1.1.2.	Objetivos específicos	20
1.3.	ESTRUTURA DA MONOGRAFIA	20
<b>2.</b>	<b>CONCEITOS GERAIS E REVISÃO DA LITERATURA</b>	<b>21</b>
2.1	REFERENCIAL TEÓRICO	21
2.2	ESTADO DA ARTE	26
<b>3.</b>	<b>METODOLOGIA</b>	<b>29</b>
3.1	BASE DE IMAGENS	31
3.2	HAAR CASCADE	32
3.3	O ALGORITMO DE VIOLA E JONES NO OPENCV	35
3.4	JUNÇÃO DE IMAGENS	37
<b>4.</b>	<b>APRESENTAÇÃO E ANÁLISE DOS RESULTADOS</b>	<b>39</b>
<b>5.</b>	<b>CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS</b>	<b>46</b>
	<b>REFERÊNCIAS</b>	<b>48</b>



## **1. INTRODUÇÃO**

A utilização de drones para realizar as mais diversas tarefas é uma tecnologia na qual se investe cada vez mais. As aplicações que utilizam drones vão desde monitoramento de plantações até operações militares, passando por vigilância de estradas e combate a incêndios florestais. Encontraram também um lugar na indústria do entretenimento, porque vários deles têm câmeras embutidas e lhes permitem filmar em lugares difíceis. Os drones têm sido muito adotados por fotógrafos e cinegrafistas como suporte para câmeras com o objetivo de fazer imagens aéreas de casamentos, atividades esportivas e outras festividades. Gravações para o cinema, tv e publicidade ganharam um incremento com o uso de drones, que possibilitam a filmagem em ângulos dificilmente alcançados pelas câmeras [1].

No Brasil, as aeronaves não tripuladas são utilizadas por órgãos como a Polícia Militar para aplicações como o auxílio no combate ao crime, como a perseguição de fugitivos, monitoramento de reservas ambientais, monitoramento de rebeliões em unidades carcerárias, monitoramento de áreas de tráfico de drogas e para a vigilância de eventos com o objetivo de aumentar a segurança do ambiente[2]. Policiais já fazem curso sobre o uso de drone na segurança pública[3]. Além do monitoramento, os drones em uso pela polícia militar podem servir a outros propósitos, como o de estimar a quantidade de pessoas presentes nestes eventos. Essa estimativa além de dados estatísticos, possibilitam uma melhor distribuição de policiamento nos eventos, podendo prevenir brigas, rixas e depredações do patrimônio público. Através da Visão Computacional, este trabalho propõe a criação de um sistema automático que realiza essas tarefas para auxiliar o trabalho da Polícia Militar da Paraíba, possibilitando a substituição da contagem manual pela contagem automática.

### **1.1. Motivação**

A contagem de multidões é uma técnica utilizada para contar ou estimar o número de pessoas em uma multidão, em geral em espaço aberto como parques e

ruas. Em eventos com ingresso, catracas são muitas vezes utilizadas para contar com precisão o número de pessoas que entram em um local. Em eventos sem entrada controlada, especialmente eventos que acontecem nas ruas ou em um parque, em vez de um local fechado, a contagem da multidão é mais difícil e menos precisa. Para muitos eventos, especialmente comícios e protestos, o número de pessoas em uma multidão carrega um significado político, já que a quantidade de participantes é uma forma de medir a relevância política do evento [4].

Hoje, o método utilizado pela polícia para estimar a quantidade de pessoas em uma multidão é feito com base na separação da área e averiguação de quantas pessoas existem nela, depois é realizado o somatório de todas as áreas. A área delimitada pela projeção horizontal coberta pela multidão é denominada mancha. Tudo se resume a uma relação que envolve a área da mancha e a taxa de ocupação na mesma. Caso a mancha seja composta por áreas com densidades distintas, é necessário subdividir a mancha em áreas menores, cada uma correspondendo a uma taxa de ocupação. Uma vez construída a escala é possível estimar a densidade[5].

Há diversos outros métodos para se medir uma multidão. O mais conhecido foi criado por um professor universitário de jornalismo, Herbert Jacobs, nos anos 60 – e por isso é conhecido como "método de Jacobs". A grosso modo é simples: calcula-se a área do local, estima-se o número de pessoas por  $m^2$ , e multiplicam-se os dois números. Se as pessoas se concentram de forma desigual, leva-se isso em conta [4].

Apesar do uso de drones e do acesso a fotos tiradas por eles, a contagem de multidões ainda é feita de forma manual, podendo levar muito tempo para ser concluída. Uma contagem automática pouparia tempo e esforço humano. Além do que, é comum haver uma grande discrepância nas estimativas sobre participação de público divulgadas por autoridades, imprensa e organizadores dos eventos. Isso ocorre basicamente por dois motivos: é muito difícil fazer a contagem precisa de uma grande aglomeração de pessoas em uma área aberta; e há sempre fortes interesses envolvidos nas estimativas, o que faz com que sejam superestimadas ou subestimadas.

## 1.2. Problema

Ao tratar cenários com grande fluxo de seres humanos, os algoritmos de detecção enfrentam problemas muito maiores do que contar pessoas individualmente, já que os indivíduos estão em multidão (ou seja, um grande número de pessoas). Essa dificuldade deve-se ao fato de ocorrer um grande número de oclusões dinâmicas e estáticas. Além disso, a maioria dos métodos existentes que tratam do problema de contagem de pessoas em multidão realiza a contagem de pessoas somente de cenários mais restritos, como ambientes fechados que possuem área de interesse menor, por exemplo, shoppings centers e metrô. Por isso, este trabalho visa tratar também o problema de contagem de pessoas em multidões em um cenário mais irrestrito devido a utilização do drone.

Em se tratando de detecção existem muitas outras variáveis que podem prejudicar a detecção, como:

- **Orientação de cabeças:** Há diversas orientações, de maneira frontal, perfil, inclinada e de costas.
- **Iluminação:** A variação de iluminação pode causar sombras na imagem, dificultando a detecção.
- **Escala:** A variação dos tamanhos e formas que servem de representação das pessoas é muito grande.

### 1.1.1. Objetivo geral

Este trabalho tem como objetivo geral, propor um método para contagem de pessoas em uma multidão de forma automática, utilizando imagens capturadas por drone.

### **1.1.2. Objetivos específicos**

Podem ser listados como objetivos específicos:

- Desenvolver um software que realize a contagem de pessoas, através de imagens;
- Estimar a escala das imagens;
- Utilizar um sistema criador de mosaicos de imagens, para junção das fotos.

### **1.3. Estrutura da monografia**

O trabalho está organizado da seguinte maneira. No capítulo 2, são apresentados os Conceitos Gerais e a Revisão da Literatura. Este capítulo apresenta as técnicas de contagem de pessoas e as explica detalhadamente. Além disso, contém outras seções que apresentam artigos com o tema contagem de multidões, também explica detalhes sobre as bases de fotos que foram usadas. O capítulo 3 traz a metodologia utilizada para desenvolver o software a partir da abordagem apresentada na revisão. O capítulo 4 possui os resultados obtidos das implementações feitas. E, finalmente, o capítulo 5 apresenta uma conclusão sobre os métodos propostos e a perspectiva para trabalhos futuros.

## **2. CONCEITOS GERAIS E REVISÃO DA LITERATURA**

Na última década, estudos sobre veículos aéreos não tripulados e estudos nas áreas de Visão Computacional, estimularam o uso destes em grande escala, e essas tecnologias estão sendo utilizadas em diversas aplicações incluindo monitoramento, rastreamento e contagem de multidões, como [6, 7, 8]. O reconhecimento de imagens requer bastante conhecimento prévio sobre o mundo e bastante capacidade de julgamento para se fazer conclusões sobre os dados da imagem e, então, tomar uma ação adequada às necessidades do agente. A disseminação e o interesse por drones resultou no aumento de suas capacidades, diminuição dos custos e aumento de produtividade em alguns setores.

Este capítulo apresentará as definições de métodos e elementos para uma melhor compreensão dos termos utilizados nos próximos capítulos. Em seguida, será apresentada uma análise sobre até onde vai a pesquisa atual no campo de estudos abordado neste trabalho e estabelecerá as bases sobre as quais ele será desenvolvido.

### **2.1 Referencial Teórico**

Nesta seção são definidos os conceitos teóricos dos métodos e elementos utilizados e abordados neste trabalho, são eles: Uso de Drones, método Haar Features, OpenCV e Oclusão.

#### **Uso de Drones**

Veículo aéreo não tripulado (VANT) ou drone, é todo e qualquer tipo de aeronave que não necessita de pilotos embarcados para ser guiada e são controladas à distância por meios eletrônicos e computacionais, sob a supervisão de humanos, ou mesmo sem a sua intervenção, por meio de Controladores Lógicos Programáveis (CLP). São muitas vezes, equipados com câmeras de alta qualidade [9].

Os drones foram idealizados para fins militares, foram concebidos, projetados e construídos para serem usadas em missões tradicionalmente de elevado risco para

humanos, nas áreas de inteligência militar, apoio e controle de tiro de artilharia, apoio aéreo a tropas de infantaria e cavalaria no campo de batalha, controle de mísseis de cruzeiro, atividades de patrulhamento urbano, costeiro, ambiental e de fronteiras, atividades de busca e resgate, entre outras[10].

Além do uso militar, os drones estão sendo utilizados por civis, como por exemplo por fotógrafos e cinegrafistas em festas de aniversários, casamentos ou eventos em geral. Um drone consegue captar melhores ângulos para fotos e filmagens mantendo a câmera estável por mais tempo facilitando também por conseguinte a produção de vídeos. Estas valências técnicas faz com que também sejam usados por emissoras de TV e pela polícia[11].

## **OpenCV**

OpenCV (Open Source Computer Vision Library). Originalmente, desenvolvida pela Intel, em 2000, é uma biblioteca multiplataforma, totalmente livre ao uso acadêmico e comercial, para o desenvolvimento de aplicativos na área de Visão Computacional. O OpenCV possui módulos de Processamento de Imagens e Vídeo I/O, Estrutura de dados, Álgebra Linear, GUI (Interface Gráfica do Usuário) Básica com sistema de janelas independentes, Controle de mouse e teclado, além de mais de 350 algoritmos de Visão Computacional como: Filtros de imagem, calibração de câmera, reconhecimento de objetos, análise estrutural e outros. O seu processamento é em tempo real de imagens. Esta biblioteca foi desenvolvida nas linguagens de programação C/C++. Também, dá suporte a programadores que utilizam Java, Python e Visual Basic e desejam incorporar a biblioteca a seus aplicativos [12].

## **Haar Features**

O método de segmentação proposto por Viola e Jones[13] e discutido em [14] é uma técnica de detecção baseada na aparência do objeto. Métodos desse tipo aprendem características a partir de conjuntos de imagens de treinamento que capturam a variedade da classe do objeto. Como sugerido em [15], esse método foi desenvolvido *a priori* para detecção de características faciais humanas, mas pode ser

usado para detectar características de outros objetos. É feita uma etapa inicial de treinamento, onde ocorre a seleção de dois conjuntos de imagens, um positivo e outro negativo. O conjunto de imagens positivas contém recortes onde está contido apenas o objeto a ser detectado; já o conjunto de imagens negativas são imagens que não contêm o objeto, geralmente paisagens onde o objeto pode ser encontrado. O processo de extração de características é baseado em máscaras convolucionais inspiradas em funções de Haar [16] (Figura 2.1). Tal característica é localizada em uma sub-região de uma sub-janela e varia em forma, tamanho e localização dentro da sub-janela. Para uma sub-janela de tamanho 20x20, pode haver dezenas de milhares de características, para diferentes formas, tamanhos e posições. Essas características são calculadas pela convolução na imagem do objeto baseadas numa decisão binária a partir de um threshold. Na Figura 2.1 duas características retangulares estão representadas em (A) e (B), as quais capturam transições de nível de cinza verticais e horizontais. Em (C) e (D) são mostradas máscaras que capturam características diagonais. Na etapa de treinamento o método utiliza o algoritmo AdaBoost [17] para construir o classificador. O objetivo desse algoritmo é construir um classificador eficiente a partir de uma série de classificadores fracos baseados em decisões binárias a partir da convolução dessas máscaras na imagem do objeto. O uso de características, baseado em máscaras convolucionais, reduz a variedade de dados na classe e aumenta a variedade de dados fora da classe em comparação à dados de entrada originais. Além disso, as características geralmente codificam conhecimento sobre o domínio, o que é difícil aprender a partir de um conjunto cru e finito de dados. Em uma imagem de 64 por 64 pixels temos um conjunto de mais de 120 mil características possíveis.

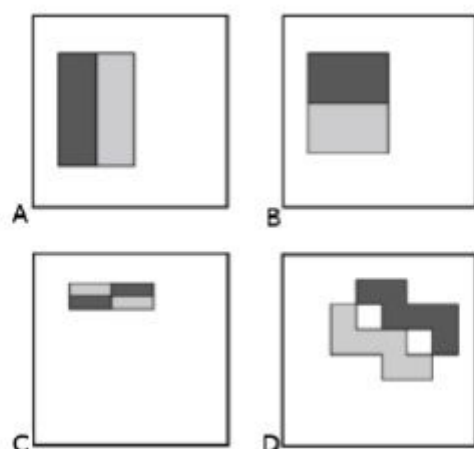


Figura 2.1: Quatro tipos de características retangulares Haar. Um característica é um escalar calculado pela soma dos pixels de uma região branca, onde é subtraída a soma da região escura

Dentre as vantagens do uso das características Haar pode-se citar a eficiência computacional, a independência de escala e cor, a baixa variedade intra-classes, a alta variação inter-classes e o fato de ser orientada a diferença de intensidades locais. As relações mútuas de intensidade de várias regiões pouco varia à medida que a intensidade absoluta de diferentes regiões da imagem muda drasticamente [18]. Tendo em vista o grande custo computacional de varrer uma imagem aplicando vários tipos de características Haar em diversas escalas, Viola e Jones [13] criaram uma representação intermediária da imagem, chamada de Imagem Integral, que pode ser computada através de poucas operações por pixel e, uma vez calculada, pode ser obtida em qualquer escala em tempo constante.

### **Oclusão**

Multidões, como mostra a Figura 3.1 aumentam a complexidade do problema de contagem com a ocorrência de oclusões.

Figura 2.2: Oclusões dificultam o cálculo do número de pessoas.



Fonte: The Heureka[19]

Existem dois tipos de oclusão:

- Oclusão estática: quando o indivíduo locomove-se para trás de um objeto da cena, como mostra a Figura 2.3 (a), recortada da Figura 2.2.
- Oclusão dinâmica: quando uma pessoa encontra-se na frente da visão que a câmera possui de outro indivíduo, como mostra a Figura 2.3 (b), também recortada da Figura 2.2;



(a) (b)

Figura 2.3: (a) Exemplo de oclusão estática; (b) Exemplo de oclusão dinâmica.

Em um cenário em que ocorrem diversas oclusões, a chance de uma cabeça ser oclusa é menor do que a de características ao longo do corpo, por possuir uma área menor.

## 2.2 Estado da Arte

A contagem de multidões e a estimativa da densidade ainda são importante tarefa em videovigilância e monitoramento de eventos. Muitos trabalhos introduziram a Aprendizagem Profunda (Deep Learning) em várias aplicações de fiscalização, como a reidentificação de pessoas [20], detecção de pedestres [21, 22, 23], rastreamento [24], análise de comportamento da multidão [25] e segmentação de multidão [26]. Seu sucesso se beneficia do poder discriminativo dos Modelos Profundos (Deep Models). [27] mostrou que as características extraídas de Modelos Profundos são mais eficazes do que recursos manuais para muitas aplicações. Em [28] foi proposto uma Rede Neural Convolutiva Profunda (CNN) para a contagem de multidões, treinada alternativamente com dois objetivos de aprendizado relacionados, densidade de multidão e contagem de multidão. Em [29] foi proposto um sistema para contar pessoas em uma cena lotada usando câmeras de vigilância detectando a região da cabeça. O detector de cabeça é baseado no classificador Adaboost combinado com um Soft Cascade, um derivado de classificadores cascade, mas apenas usando um único estágio com muitos classificadores fracos para obter detecções mais precisas.

Muitos estudos sobre contagem de pessoas através de detecção de cabeça foram feitos. [30] introduziu um método de detecção de cabeça usando o contorno de

cabeça parcial, usando o modelo de eclipse para ajustar o contorno. [31] utilizou a cor do cabelo e o contorno da cabeça para detectar cabeças humanas. Sabe-se que as cabeças humanas têm uma forma circular aproximada, pelo que o modelo de contorno é comumente utilizado na detecção de cabeças. Mas o método baseado em modelo de contorno é muitas vezes influenciado por outros objetos semelhantes a cabeças humanas em cenas. Em [32] propuseram um método efetivo de contagem de passageiros, que primeiro usa a transformação de círculo de Hough [16] para detectar cabeças humanas e, em seguida, executa o seguimento da cabeça usando o método de fluxo óptico [33]. No entanto, o método de rastreamento baseado no fluxo óptico requer uma grande quantidade de cálculo. No trabalho realizado por [34] é proposto um método de contagem de pessoas baseado na detecção e rastreamento de cabeças para avaliar o número de pessoas que se movem sob uma câmera. As partes principais no método proposto são: extração do primeiro plano, detecção e rastreamento de cabeças. O método proposto utiliza primeiro, um eficaz método de extração do primeiro plano para obter as regiões deste de pessoas em movimento e algumas operações morfológicas são utilizadas para otimiza-las. Em seguida, ele explora um recurso LBP baseado no classificador Adaboost para detecção de cabeça nas regiões de primeiro plano otimizado. O LBP cascade é uma versão mais rápida de treinar, porém menos precisa a Haar. Após a detecção ser realizada, as cabeças detectadas são rastreadas por um método local de rastreamento de cabeças baseado no algoritmo Meanshift, um método robusto de encontrar um local de densidade extrema em um conjunto de dados. É um processo fácil para distribuições contínuas, onde o contexto é essencialmente picos de densidade aplicados a um histograma de densidade de dados [35].

Métodos diferentes do classificador Adaboost e Haar Features foram usados no trabalho de [36], no qual é explorado o contexto para a detecção humana em multidões densas na forma de uma priori de escala localmente consistente, que capta a similaridade de escala nos arredores locais e sua variação suave sobre a imagem. Utilizando a escala e a confiança das detecções obtidas de um detector humano subjacente, são inferidas escala e priores de confiança usando Markov Random Field. Em um mecanismo iterativo, as confianças das hipóteses de detecção são modificadas para refletir consistência com os priores inferidos, e os priores são atualizados com

base nas novas detecções. O conjunto final de detecções obtidas é então fundamentado para oclusão usando *Binary Integer Programming*, onde sobreposições e relações entre partes de indivíduos são codificadas como restrições lineares. Tanto a detecção humana quanto o raciocínio de oclusão na abordagem proposta são resolvidos com restrições vizinhas locais dependentes, respeitando assim a interdependência entre características individuais e a densa análise das multidões. Além disso, propõe um mecanismo para detectar diferentes combinações de partes do corpo sem exigir anotações para combinações individuais.

Apesar destes trabalhos comentados anteriormente abordarem detecções de pessoas, nenhum é indicado para uma quantidade elevada de pessoas e nem para grandes alturas. Ou funcionam apenas para câmeras de vigilância em um determinado ambiente e não para imagens, ou foram desenvolvidos apenas para a detecção de corpos e rostos humanos, sendo portanto inviável para ser usado em uma detecção de pessoas cuja foto foi tirada de um ângulo aéreo, onde apenas o topo da cabeça aparece. Portanto, é proposto neste trabalho o uso de drones que capturem fotos de ângulos que eliminem ao máximo as oclusões e detectores que funcionem para alturas consideráveis .

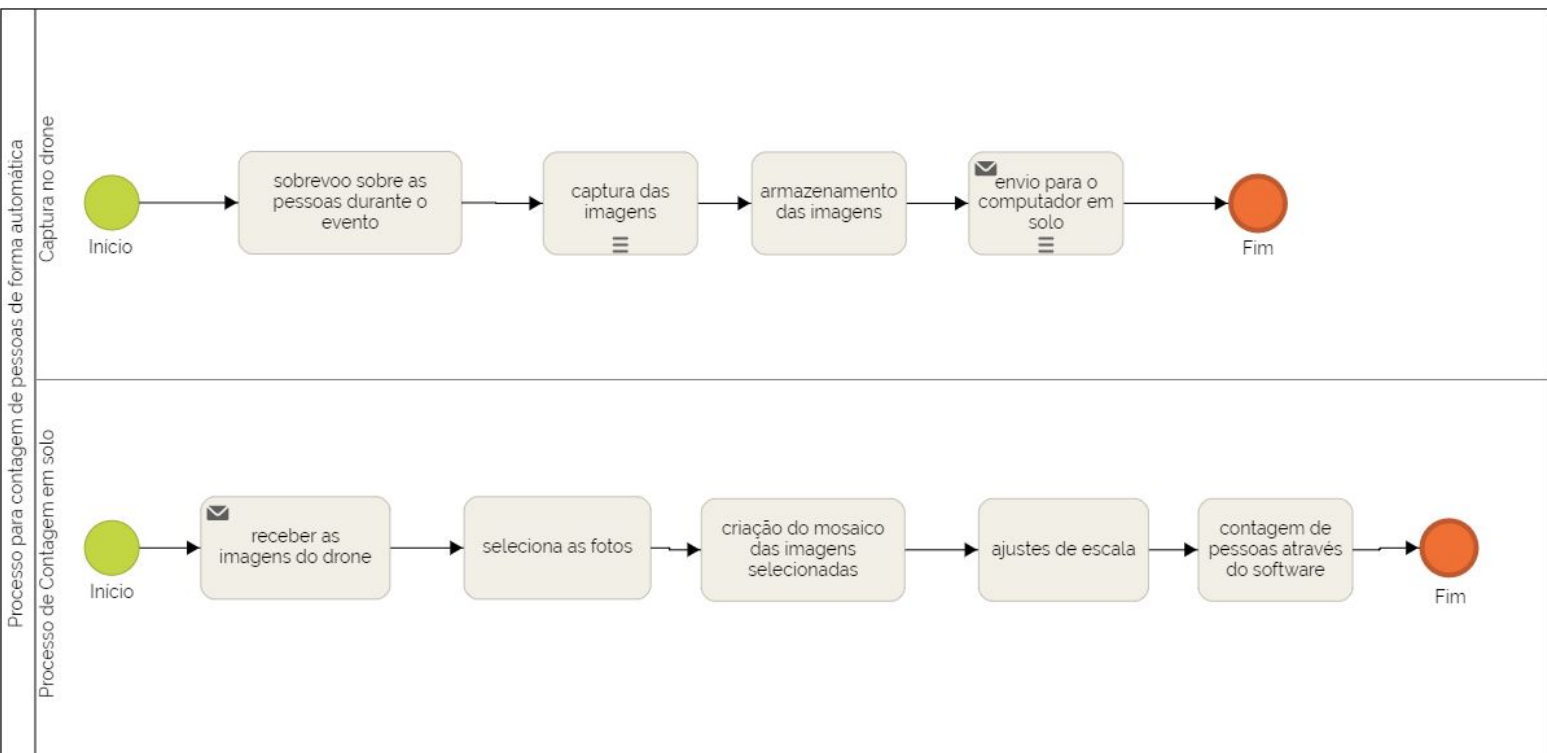
### 3. METODOLOGIA

Este capítulo encontra-se dividido em quatro tópicos que descrevem os passos realizados para o desenvolvimento deste trabalho. Primeiramente, serão justificadas as escolhas de ângulo e altura para a criação da base de dados utilizada no treinamento dos classificadores. No segundo tópico serão descritos os passos utilizados para a criação de um identificador Haar Cascade. No terceiro tópico será descrito o desenvolvimento do software e os testes realizados. No quarto e último tópico uma explicação a respeito da importância da junção de imagens para esse trabalho.

A figura 3.1 é um diagrama que descreve o processo metodológico e seus agentes de execução. O processo inicia no estado de início para os dois agentes:

1. Captura no Drone
  - Sobrevoou sobre as pessoas durante o evento
  - Captura das imagens. Os pequenos traços em baixo desta tarefa significa que esta se repete várias vezes de forma sequencial.
  - Armazenamento das imagens.
  - Envio para o computador em solo. Esta tarefa também se repete várias vezes, finalizando assim o processo com a ação do drone.
2. Processo de Contagem em Solo
  - Esta são as ações realizadas no computador, iniciando pelo processo de receber as imagens capturadas pelo drone.
  - Seleção de fotos
  - Criação do mosaico das imagens selecionadas.
  - Ajuste de escala.
  - Por fim, a contagem de pessoas através do software.

Figura 3.1 Processo Metodológico



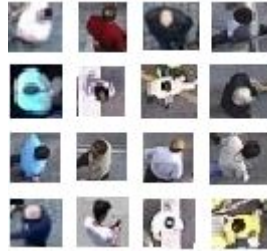
Fonte: Autoria própria

### 3.1 Base de Imagens

Para facilitar o problema da oclusão, foi determinado que seria usada somente uma visão aérea das cenas, chamada *top-down*, onde apenas o topo da cabeça de cada pessoa é visto. Para o treinamento seriam necessários então imagens positivas de cabeças de pessoas em um ângulo *top-down*. Uma das primeiras dificuldades encontradas no desenvolvimento de um sistema detector de pessoas, é a obtenção de imagens que são necessárias para o treinamento de um classificador e para a verificação da acurácia do detector. Muitas bases de imagens humanas foram pesquisadas e baixadas, entre elas, as mais conhecidas bases para treinamentos visando a detecção de pessoas, como *INRIA Person Dataset*, *Caltech Pedestrian Detection Benchmark*, *MIT Pedestrian Dataset*, *PETA dataset*. Bases com termos mais específicos, como “Human Head Database”, foram pesquisadas, e bases de imagem como a *Head Pose Image Database - PRIMA* e as disponíveis em [37] e [38] foram encontradas, porém, só continham imagens de cabeça tiradas de frente e/ou de muito perto. Não foram encontradas disponíveis para baixar nenhuma base de imagens que contivessem pessoas vistas de cima e da altura necessária. Devido à esta falta de base de imagens o primeiro passo foi recolher na internet algumas imagens capturadas através de drones onde aparecem pessoas vistas do ângulo e altura desejados.

Outras bases de imagens que não possuíssem pessoas foram baixadas para compor a base de imagens negativas. As imagens foram redimensionadas, cortadas e reunidas para inicializar o treinamento do classificador. Para que ao fim deste trabalho uma análise e uma comparação entre alturas possam ser feitas, foram criados 3 classificadores, um para cada uma das alturas das fotos que foram utilizadas nos testes. As alturas escolhidas foram 15m, 25m e 35m. Na figura 3.2 podem ser vistos alguns exemplos de imagens positivas utilizadas na base de imagens.

Figura 3.2 Exemplos de Imagens Positivas



Fonte: Dreams Time[39]

Para os testes com o classificador, além de imagens coletadas do Google, foram cedidas pela PM da Paraíba imagens para testes, capturadas pelo drone utilizado por ela.

### 3.2 Haar Cascade

As características Haar (Haar Features) são a base da contagem de pessoas utilizada neste trabalho. Ela consiste na diferença de intensidade entre regiões retangulares de uma imagem. As relações entre as diferentes regiões são expressas como restrições sobre os valores dos coeficientes. Essas características foram baseadas no trabalho do matemático húngaro Alfred Haar de 1909, a Transformada de Haar, que pode ser utilizada na representação de uma função.

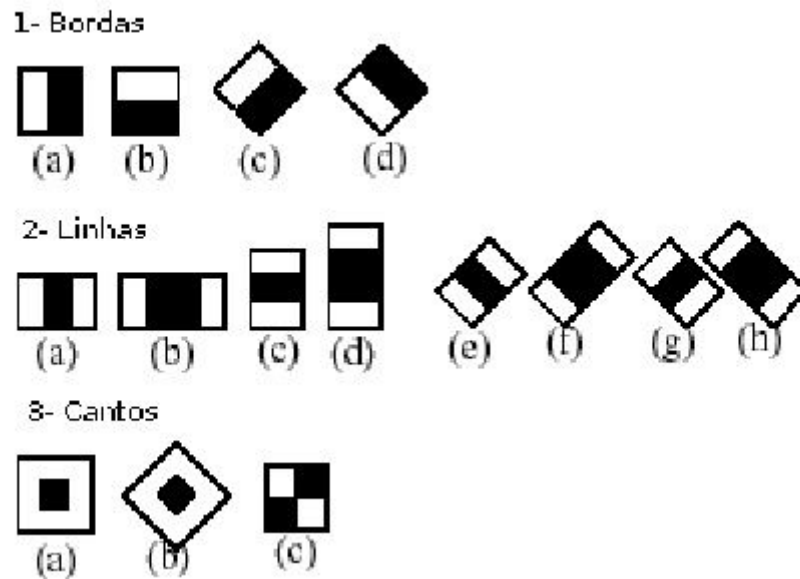


Figura 3.3: Características Haar [13]

O treinamento foi feito através de uma adaptação do algoritmo de Aprendizado de Máquina AdaBoost proposto por [13] utilizando o utilitário opencv-training e tendo como entradas para as alturas de 15, 25 e 35 metros, 230 imagens positivas e 1414 imagens negativas. Para iniciar o treinamento foi utilizado o repositório disponível em [40], ele inclui alguns utilitários e alguns diretórios vazios que serão preenchidos ao longo do treinamento. As imagens positivas foram colocadas no diretório ./positive\_images e as negativas no diretório ./negative\_images. Em seguida, os seguintes comandos foram utilizados para que fossem criadas listas em .txt com o nome de todas as imagens.

```
>find ./positive_images -iname "*.jpg" > positives.txt
>find ./negative_images -iname "*.jpg" > negatives.txt
```

Depois de criadas as listas, é necessário criar amostras positivas com o script bin/createsamples.pl e salvá-las na pasta ./samples:

```
>perl bin/createsamples.pl positives.txt negatives.txt samples 1500\  
"opencv_createsamples -bgcolor 0 -bgthresh 0 -maxxangle 1.1\  
-maxyangle 1.1 maxzangle 0.5 -maxidev 40 -w 80 -h 40"
```

Usando as ferramentas /mergevec.py as amostras em ./samples foram mescladas em um arquivo:

```
>python ./tools/mergevec.py -v samples/ -o samples.vec
```

Para começar a treinar o classificador com opencv\_traincascade, que veio com o OpenCV, e salvar os resultados em ./classifier usou-se o seguinte comando, onde numStages é a quantidade de estágios que serão treinados:

```
>opencv_traincascade -data classifier -vec samples.vec -bg negatives.txt\  
-numStages 20 -minHitRate 0.999 -maxFalseAlarmRate 0.5 -numPos 1000\  
-numNeg 600 -w 80 -h 40 -mode ALL -precalcValBufSize 1024\  
-precalcIdxBufSize 1024
```

Depois de iniciado, o programa de treinamento imprime seus parâmetros e, em seguida, inicia o treinamento. Cada etapa imprimirá algumas análises à medida que forem treinadas:

```
==== TRAINING 0-stage ====  
<BEGIN  
POS count : consumed 1000 : 1000  
NEG count : acceptanceRatio 600 : 1  
Precalculation time: 11  
+-----+-----+-----+  
| N | HR | FA |  
+-----+-----+-----+  
| 1 | 1 | 1 |  
+-----+-----+-----+  
| 2 | 1 | 1 |  
+-----+-----+-----+  
| 3 | 1 | 1 |  
+-----+-----+-----+  
| 4 | 1 | 1 |
```

```
+-----+-----+-----+
|  5|      1|      1|
+-----+-----+-----+
|  6|      1|      1|
+-----+-----+-----+
|  7|      1| 0.711667|
+-----+-----+-----+
|  8|      1|    0.54|
+-----+-----+-----+
|  9|      1|    0.305|
+-----+-----+-----+
END>
Training until now has taken 0 days 3 hours 19 minutes 16 seconds.
```

Vinte estágios foram treinados e todos foram longos, tendo cada um levado em média 12 horas para ser concluído. Depois de concluídos todos os estágios, um arquivo cascade.xml é gerado.

### 3.3 O Algoritmo de Viola e Jones no OpenCV

O software objetivo deste trabalho foi criado, um programa em python nomeado detect.py, com ele foi possível testar os arquivos .xml utilizando imagens de boa resolução coletadas da internet. A biblioteca OpenCV já possui uma implementação do algoritmo de Viola e Jones. Esse método já implementado, pode ser utilizado por meio da função: **detectMultiScale**. Essa função possui 5 parâmetros de entrada e como resultado de saída retorna uma sequência de retângulos, e cada um desses retângulos representa uma pessoa. Os parâmetros de entrada e saída são:

- image: A imagem onde será feita a procura.
- objects: Sequência de retângulos com a localização das imagens procuradas.

- cascade: O classificador do tipo HaarCascade (na forma de arquivo .xml) lido pelo OpenCV.
- storage: Espaço alocado na memória para armazenar a sequência dos resultados da detecção do objeto(resultados em forma de retângulos).
- scaleFactor: O fator pelo qual a janela de procura da função será aumentada para percorrer novamente a imagem
- minNeighbors: faz o agrupamento de retângulos vizinhos, sendo esse parâmetro utilizado para mesclar retângulos similares.
- flags: Faz um processamento específico disponibilizado pela biblioteca OpenCV.
- minSize: Valor em pixels da menor janela de procura, largura e altura da janela.

Os valores dos parâmetros precisaram ser modificados ao longo dos testes devido ao variado tamanho em *pixels* das imagens coletadas na internet para realizá-los. A princípio os valores dos parâmetros scaleFactor, minNeighbors, minSize foram definidos respectivamente como 1.3, 10 e (32, 96).

O software varre a imagem procurando por características Haar em janelas inicialmente de tamanho 32x96 pixels e que variam sua escala em 10 por cento até encontrar objetos de interesse, caso eles existam. Ao final do processo, as pessoas presentes na imagem são marcadas com um retângulo colorido e o número aproximado de pessoas na foto é exibido. Uma vez que as características Haar são invariantes à cor, todas as imagens foram transformadas para tons de cinza para efeitos de treinamento. Nos testes de detecção as imagens de entrada são coloridas, contudo, posteriormente elas são convertidas para escala de cinza para facilitar a detecção e, uma vez detectadas as pessoas, estas são marcadas na imagem colorida correspondente, que é a imagem de saída.

Os testes começaram a ser feitos com os arquivos .xml gerados a partir do estágio 15. 3 fotos foram testadas 20 vezes cada uma, sendo variados a cada teste o cascade.xml correspondente a cada estágio, e os valores dos parâmetros scaleFactor, minNeighbors, minSize. Para as fotos tiradas a aproximadamente 15 metros de altura os estágios de 15 a 18 não obtiveram resultados muito satisfatórios, porém o

classificador correspondente ao estágio 20 obteve bons resultados. Os testes seguintes foram feitos em imagens tiradas de aproximadamente 25 metros de altura. Todas as etapas realizadas anteriormente para as fotos de 15 metros foram repetidas em mais 2 fotos. O estágio que apresentou melhor resultados foi o de número 15.

Por fim o valor do parâmetro `minNeighbors` foi alterado para 8, e `minSize` foi alterado para (100,100) visto que as imagens selecionadas para os testes possuíam larguras que iam de 4200 a 6400 pixels, e alturas que iam de 2800 a 4200 pixels. Esses valores de parâmetros combinados com os respectivos arquivos .xml selecionados para cada altura apresentaram os melhores resultados.

### **3.4 Junção de Imagens**

Em um projeto de mapeamento aéreo com Drones são capturadas diversas imagens. Utilizando um software de processamento de dados, estas imagens são unificadas em uma única imagem cobrindo toda área de interesse. Esta imagem é georreferenciada, ou seja, cada pixel da imagem possui suas coordenadas conhecidas possibilitando a realização de medidas nesta imagem.

Para este trabalho, a contagem tem uma maior garantia se houver uma junção de uma série de imagens, pois com ela evita-se contar as mesmas pessoas várias vezes, além de conseguir toda a área de interesse. Após o processo de treinamento prévio com imagens numa perspectiva top-down utilizando o haar cascade, o processo de captura das imagens com o drone é iniciado, em um sobrevoo sobre toda a área de interesse de contagem e faz o processo de junção das imagens. Depois disso é iniciado o processo de contagem, passando a imagem geral resultado da junção para o sistema de contagens automática que utiliza o treinamento prévio já descrito.

Para a junção de imagens é necessário que haja também um controle de escala pois a construção dessas junções, denominadas mosaicos controlados, é feita a partir de fotografias retificadas na mesma escala. Para isso, utilizam-se pontos de controle, os quais são impressos em uma folha base sobre a qual são ajustadas as fotografias retificadas. Essas fotografias retificadas são imagens que consistem em modificar os

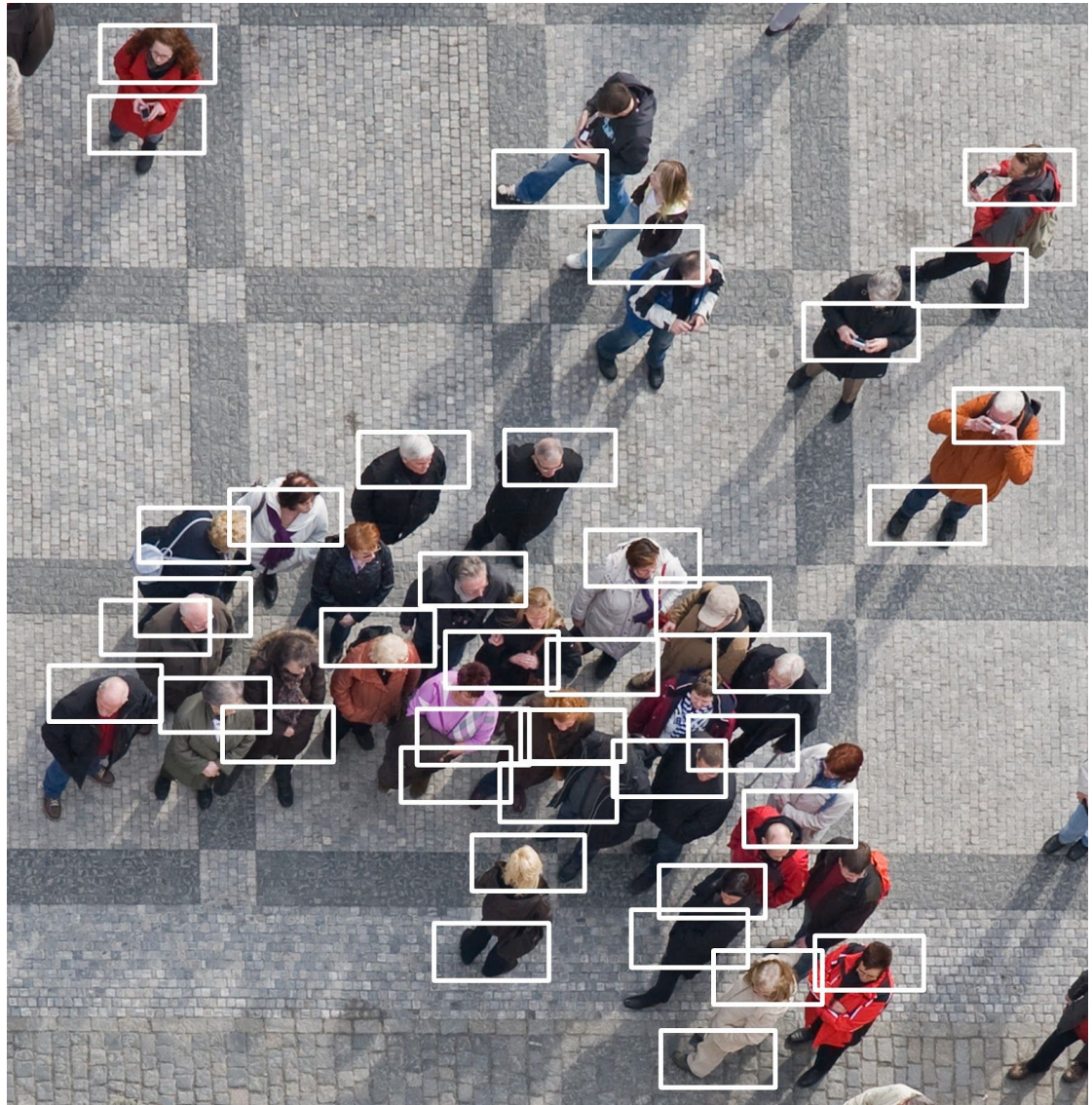
ângulos referentes à atitude da câmara, bem como a distância focal, resultando em projetá-la, para um plano horizontal, segundo seu feixe perspectivo, conforme [41]. Na ortorretificação são eliminadas as distorções causadas pelo relevo da área fotografada, pois há uma mudança da perspectiva cônica para a perspectiva ortogonal, segundo[42].

#### **4. APRESENTAÇÃO E ANÁLISE DOS RESULTADOS**

Duas medidas são importantes para avaliar a qualidade do algoritmo: A quantidade de objetos que foram incorretamente identificados como pessoas(falso positivo) e quantidade de pessoas que não foram identificadas(falso negativo). Idealmente, o algoritmo teria ambos os valores nulos. Apesar dos resultados mostrados a seguir apresentarem números bem aproximados, vários falsos positivos e negativos podem ser encontrados, porém o valor de ambos é proporcional, equilibrando assim o resultado final.

O primeiro teste foi feito em imagens tiradas a 15 metros de altura: Na figura 4.1(a), um existe um total de 34 pessoas e foram detectadas na imagem 39. Vinte e quatro positivos, 15 falsos positivos e 7 falsos negativos. Apresentando uma taxa de 71% de acerto.

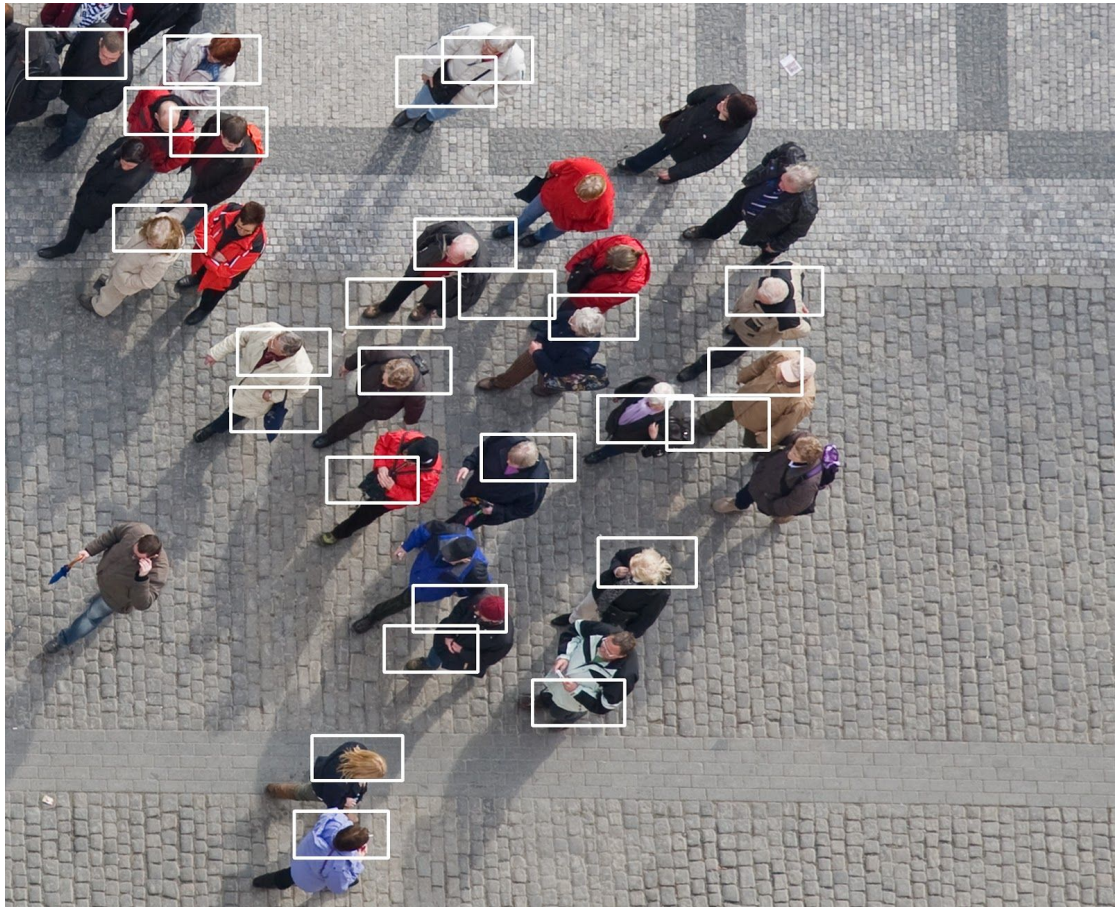
Figura 4.1(a): Resultados para uma foto tirada a aproximadamente 15 metros de altura



Fonte: JLL Attends Davos 2016: WEF Annual Meeting.[43]

Na figura 4.1(b), existe um total de 29 pessoas e foram detectadas na imagem 26. Dezenove positivos, 7 falsos positivos, 9 falsos negativos. Apresentando uma taxa de 66% de acerto.

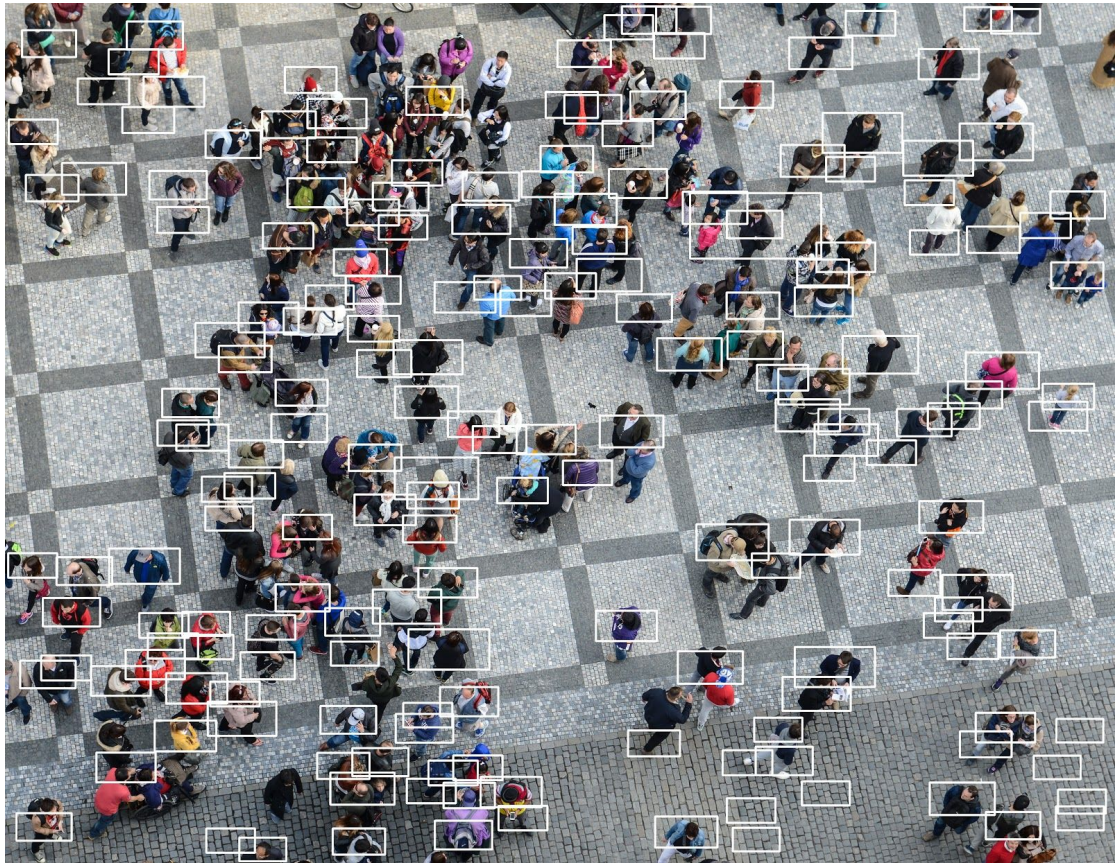
Figura 4.1(a): Resultados para uma foto tirada a aproximadamente 15 metros de altura



Fonte: JLL Attends Davos 2016: WEF Annual Meeting.[43]

Na figura 4.2, existe um total de 194 pessoas e foram detectadas na imagem 191. 126 positivos verdadeiros, 65 falsos positivos e 68 falsos negativos, apresentando uma taxa de 66% de acerto.

Figura 4.2: Resultados para uma foto tirada a aproximadamente 25 metros de altura



Fonte: Deposit Photos [44].

Devido a dificuldade de encontrar disponíveis na internet, figuras adequadas para os testes, foi usada a figura 4.3, que simula uma foto tirada a aproximadamente 35 metros de altura. Os resultados, encontrados na figura 4.4, foram pouco conclusivos, o número total de pessoas encontradas corresponde a apenas 50% do número de pessoas na imagem, e devido a proximidade de cada uma, é difícil dizer a quantidade exata de falsos negativos e positivos.

Figura 4.3: Simulação de uma foto tirada a aproximadamente 35 metros de altura



Fonte: Bigstock Photo[45]

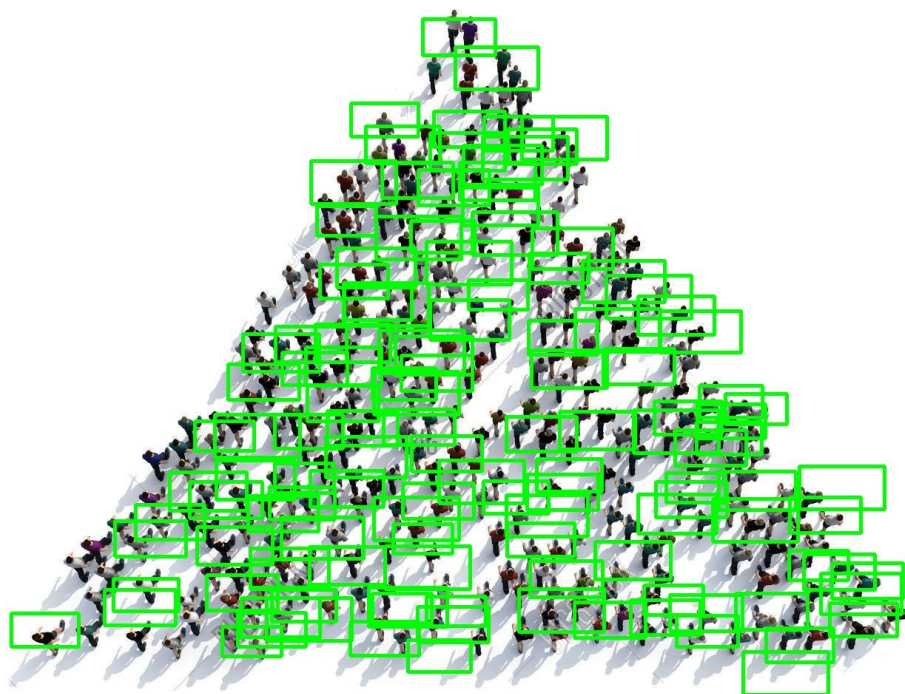


Figura 4.4: Resultados para uma foto tirada a aproximadamente 35 metros de altura

Posteriormente, as fotos enviadas pela PM da Paraíba, que foram capturas a 70 metros de altura, foram testadas, porém devido a diferença entre as alturas, os resultados não foram bons. Uma nova base de dados com imagens positivas foi criada através destas fotos. Um novo treinamento foi iniciado, porém ao chegar no estágio 11 observou-se que os resultados, mostrados na figura 4.6, ainda não correspondiam ao esperado, o treinamento foi então cancelado. Mesmo manipulando os parâmetros de busca, cortando a imagem em tamanhos menores e realizando novos treinamentos, o viola-jones não apresentou bons resultados, por não conseguir verificar características que pudessem definir os objetos buscados.

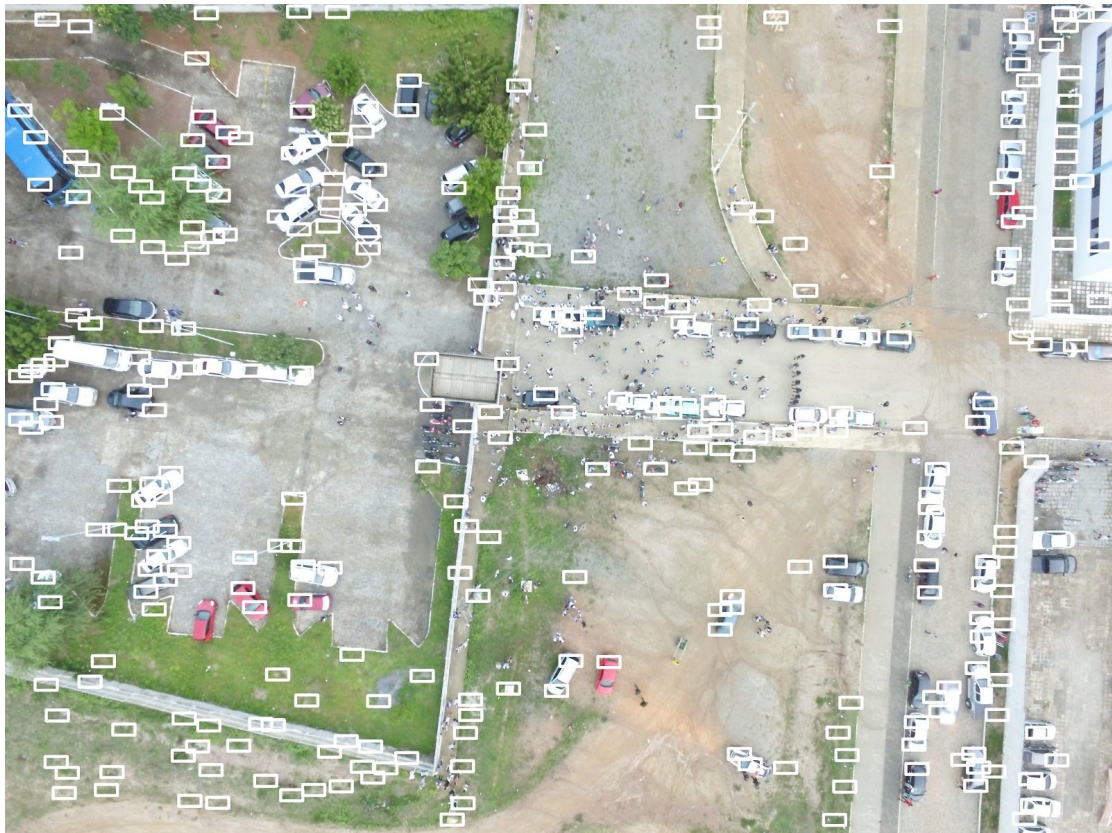


Figura 4.5 Resultados para uma Imagem enviada pela PM da Paraíba

Altitude	Taxa de Acerto
15m	entre 66% e 71%
25m	66%
35m	abaixo de 50%
70m	resultados inconclusivos

Tabela 1: Resultados

Com a criação do contador de multidões utilizando o método haar cascade , o objetivo geral deste trabalho foi concluído. Para alturas até 30 metros o método apresentou uma boa acurácia. A escala de imagens a partir de fotos foi determinada, porém fotos foram capturadas de grandes altitudes o método apresentado não foi bem sucedido. O objetivo de utilizar um sistema criador de mosaicos de imagens, para junção das fotos não pode ser concluído pela falta de um conjunto de imagens ,onde fosse encontrada uma multidão, que pudesse ser testado.

## 5. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Este trabalho é um ponto de partida para um estudo mais profundo envolvendo técnicas mais eficientes. A realização deste projeto possibilitou não somente o aprendizado da técnica aplicada, Haar cascade, mas também pôde-se observar a importância do tema principal, que é a contagem de multidões. Um sistema que realize contagens de multidão pode auxiliar a PM da Paraíba em vários aspectos.

A maior dificuldade neste projeto foi a etapa de treinamento do classificador. Para que o classificador tenha uma precisão ótima é necessário o treinamento com mais imagens, o que ocasiona um aumento no tempo do processamento para a criação do classificador. Como consequência à essa redução, obteve-se um classificador com a acurácia não tão elevada. Um dos pontos que neste trabalho tornou-se negativo para a obtenção de um classificador eficiente, foi com relação ao tamanho das imagens positivas utilizadas na fase de treinamento. Outro motivo para o qual foi obtido esses resultados, refere-se tanto à qualidade das imagens utilizadas, quanto ao tamanho empregado na etapa de treinamento. As imagens positivas, que já não eram tão nítidas, foram redimensionadas para tamanhos entre 20x20 e 30x30. Para processar imagens de tamanhos mais adequados, é necessário um computador com maior poder de processamento. Assim, o classificador poderá retirar mais informações das imagens, melhorando então o seu desempenho.

Ao analisar os resultados dos experimentos com o método de detecção de cabeça conclui-se que ainda há muito a ser melhorado para que ele obtenha resultados excelentes. Esses resultados seriam taxa de acerto médio acima dos 85% e uma baixa taxa de falsos positivos. A oclusão é um obstáculo difícil de ser resolvido e está presente em todos os métodos. Tendo em vista este problema o ângulo escolhido minimizou as oclusões, possibilitando uma melhor contagem. Para alturas acima de

30 metros não foram obtidos resultados satisfatórios, mas para alturas de até 30 metros pode-se concluir que o método utilizado possui boa acurácia.

Como trabalho futuro, existem alguns pontos que devem ser modificados para melhorar os resultados do método proposto. Primeiramente, os classificadores testados ainda possuem falhas ao criar falsos positivos e deixar de detectar algumas cabeças não oclusas. Um número maior de imagens positivas e a mudança de algumas imagens negativas ajudaria bastante no resultado final do método de detecção.

## REFERÊNCIAS

- [1] TECHTUDO. O que é drone e para que serve? Tecnologia invade o espaço aéreo. Disponível em: <<http://www.techtudo.com.br/noticias/noticia/2013/10/o-que-sao-e-para-que-serve-m-os-drones-tecnologia-invade-o-espaco-aereo.html/>>. Acesso em: Maio de 2016.
- [2] BLITZ DIGITAL. Drones Policiais. Disponível em: <<https://blitzdigital.com.br/tec-menu/apoio-aereo-de-policiamento/>>. Acesso em: Junho de 2016.
- [3] AGÊNCIA DE NOTÍCIAS DO PARANÁ. Policiais fazem curso sobre uso de drone na segurança pública. Disponível em: <<http://www.aen.pr.gov.br/modules/noticias/article.php?storyid=83331/>>. Acesso em: Junho de 2016.
- [4] WIKIPEDIA. Contagem de multidões. Disponível em: <[https://pt.wikipedia.org/wiki/Contagem\\_de\\_multid%C3%B5es/](https://pt.wikipedia.org/wiki/Contagem_de_multid%C3%B5es/)>. Acesso em: Maio de 2016.
- [5] ENGERT, B. Engert. Estimativa do Quantitativo de Indivíduos em Multidões. Centro de Estudos e Pesquisa de Desastres. 1ª Edição, 2006.
- [6] GOOD DRONE LAB. Disponível em: <<http://gooddronelab.com/index.html/>>. Acesso em: Maio de 2016.
- [7] MÜLLER, T.; MÜLLER, M. "Vision-based drone flight control and crowd or riot analysis with efficient color histogram based tracking", *Proc. SPIE 8020, Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications VIII*, 80200R. 2011.
- [8] IMAMURA, Y.; OKAMOTO, S.; HOON, J. Human Tracking by a Multi-rotor Drone Using HOG Features and Linear SVM on Images Captured by a Monocular Camera. Proceedings of the International MultiConference of Engineers and Computer Scientists 2016. v.1, IMECS, 2016.
- [9] EVENTIALS. Drones, modelos, usos e aplicações. Disponível em <<https://dlq8vi77lxj74.cloudfront.net/media/e55e91b2cc22ba117ba8d1546537f7a4c037cf67/d675ef0f9d8a788f6d5a3803769697a930d4d5d0/1411838094/dronemodelosuseaplicacoes.pdf/>>. Acesso em: Maio de 2016.

[10] CANALTECH. Robôs voadores: Conheça um pouco sobre a tecnologia dos drones militares. Disponível em:<<https://canaltech.com.br/analise/seguranca/Robos-voadores-conheca-um-pouco-sobre-a-tecnologia-dos-drones-militares/>>. Acesso em: Junho de 2016.

[11] INSTITUTO ALVARO VALLE. O uso civil dos drones. Disponível em:<[http://www.institutoalvarovalle.org.br/artigo\\_0082\\_2013.html/](http://www.institutoalvarovalle.org.br/artigo_0082_2013.html/)>. Acesso em: Junho de 2016.

[12] WIKIPEDIA. OpenCV. Disponível em:<<https://pt.wikipedia.org/wiki/OpenCV/>>. Acesso em: Maio de 2016.

[13] VIOLA, P.; JONES, M. “Rapid Object Detection using a Boosted Cascade of Simple Features”, Conference on Computer Vision and Pattern Recognition, 2001.

[14] LIENHART, R.; MAYDT, J. “An Extended Set of Haar-like Features for Rapid Object Detection”, International Conference on Image Processing, 2002.

[15] VIOLA, P.; JONES, M.; SNOW, D. Detecting pedestrians using patterns of motion and appearance. In IEEE Conference on Computer Vision, p. 734–741, Nice, France, 2003. IEEE Press.

[16] GONZALEZ, R. C.; WOODS, R. E. Digital Image Processing (3rd Edition). [S.l.]:Prentice Hall, 2007.

- [17] FREUND, Y.; SCHAPIRE, R. "A Short Introduction to Boosting", Journal of Japanese Society for Artificial Intelligence, v. 14, p. 771-780, 1999.
- [18] OREN, M. et al. Pedestrian Detection Using Wavelet Templates. Computer Vision and Pattern Recognition 1997, 1997.
- [19] THE HEUREKA. Disponível em: <<http://theheureka.com/berlins-factory-opening-day>> Acesso em: Maio de 2016.
- [20] LI, W.; ZHAO, R.; XIAO, T.; WANG, X. DeepReID: Deep filter pairing neural network for person re-identification. CVPR, 2014.
- [21] ZENG, X.; OUYANG, W.; WANG, M.; WANG, X. Deep learning of scene-specific classifier for pedestrian detection. ECCV. 2014.
- [22] ZENG, X.; OUYANG, W.; WANG, X. Multi-stage contextual deep learning for pedestrian detection. ICCV, 2013.
- [23] OUYANG, W.; WANG, X. Joint deep learning for pedestrian detection. ICCV, 2013.
- [24] WANG, N.; YEUNG, D. Y. Learning a deep compact image representation for visual tracking. NIPS, 2013.
- [25] JING, S.; KAI, K.; CHEN, L.; XIAOGANG, W. Deeply learned attributes for crowded scene understanding". in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. CVPR, 2015.
- [26] KAI, K.; XIAOGANG, W. Fully convolutional neural networks for crowd segmentation. 2014.
- [27] SERMANET, P.; EIGEN, D.; ZHANG, X.; MATHIEU, M.; FERGUS, R. LECUN Y. Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv, 2013.
- [28] ZHANG, C.; LI, H.; XIAOGANG, W.; XIAOKANG, Y. Cross-scene Crowd Counting via Deep Convolutional Neural Networks. University of Electronic Science and Technology of China, 2015.

[29] SUBBURAMAN, V.; DESCAMPS, A.; CARINCOTTE, C. Counting people in the crowd using a generic head detector. Image Department, 2012.

[30] CHAROENPONG, T. Human head detection by using partial head contour. Proceedings of the Third International Conference on Knowledge and Smart Technologies, p. 29-32, 2011.

[31] ZHAO, M.; SUN, D.; FAN, W. Hair-color model and adaptive contour templates based head detection. Proceedings of the 8th World Congress on Intelligent Control and Automation, p. 6104-6108, 2010.

[32] MUKHERJEE, S.; SAHA, B.; JAMAL, I.; LECLERC, R.; RAY N. A novel framework for automatic passenger counting, IEEE International Conference on Image Processing, p. 2969-2972, 2011.

[33] HORN, B. K. P.; SCHUNCK, B. G. Determining optical flow. Artif. Intell., v. 17, p. 185-203, 1981.

[34] LI, B.; ZHANG, J.; ZHANG, Z.; XU, Y. A People Counting Method Based on Head Detection and Tracking. Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, 2014.

[35] BRADSKI, G.; KAEHLER, A. Learning OpenCV: Computer vision with the OpenCV library. O'Reilly Media, Incorporated. 2008.

[36] IDRESS, H.; SOOMRO, K.; SHAH, M. Detecting Humans in Dense Crowds using Locally-Consistent Scale Prior and Global Occlusion Reasoning. Pattern Analysis and Machine Intelligence (PAMI), IEEE Transactions, 2015.

[37] ICS. Head pose estimation dataset. Disponível em: <<http://www.ics.forth.gr/cvrl/headpose/>> Acesso em: Março de 2016.

[38] RESEARCH UNIVERSITY PARIS. Context-aware CNNs for person head detection. Disponível em: <<http://www.di.ens.fr/willow/research/headdetection/>> Acesso em: Março de 2016.

[39] DREAMSTIME. Disponível em:< <https://dreamstime.com>>Acesso em: Abril de 2016.

[40] OPEN HAAR CLASSIFIER TRAINING. Learn how to train your own OpenCV Haar classifier. Disponível em:<<https://github.com/mrnugget/opencv-haar-classifier-training>>. Acesso em: Abril de 2016.

[41] ANDRADE, J. B. de. Fotogrametria. Curitiba: SBEE, 1998.

[42] MENESES, L. F. de. Ortorretificação de Fotografias Aéreas de Pequeno Formato Obtidas com Câmara Digital Convencional. João Pessoa, 2007. 76 f. Monografia (Tecnólogo em Geoprocessamento). Centro Federal de Educação Tecnológica da Paraíba. OLIVEIRA, C. de. Dicionário Cartográfico. 4. ed. Rio de Janeiro: IBGE, 1993.

[43] HOUSTONBLOG. JLL Attends Davos 2016: WEF Annual Meeting. Disponível em:<<http://houstonblog.jll.com/jll-attends-davos-2016-wef-annual-meeting/>> Acesso em: Maio de 2016.

[44] DEPOSITPHOTOS. Disponível em:<<https://cn.depositphotos.com/76652961/stock-photo-crowd-of-people-top-view.html/>> Acesso em: Abril de 2016.

[45] BIGSTOCKPHOTO. Disponível em:<[www.bigstockphoto.com](http://www.bigstockphoto.com)> Acesso em: Abril de 2016.