JAIME DALLA VALLE JUNIOR

CONTAGEM AUTOMÁTICA DE PESSOAS EM CENAS DE VÍDEO USANDO VISÃO COMPUTACIONAL

Dissertação apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Informática.

Curitiba
2007

JAIME DALLA VALLE JUNIOR

CONTAGEM AUTOMÁTICA DE PESSOAS EM CENAS DE VÍDEO USANDO VISÃO COMPUTACIONAL

Dissertação apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Informática.

Área de Concentração: Ciência da Computação

Orientador: Alceu de Souza Britto Junior
Co-orientador: Alessandro Lameiras Koerich

Curitiba
2007

Dalla Valle Jr., Jaime
CONTAGEM AUTOMÁTICA DE PESSOAS EM CENAS DE VÍDEO USANDO VISÃO COMPUTACIONAL. Curitiba, 2007.

Dissertação - Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Informática.

1. Visão Computacional 2. Contagem de Pessoas 3. Câmeras de Circuitos Fechados de TV I.Pontifícia Universidade Católica do Paraná. Centro de Ciências Exatas e Tecnologia. Programa de Pós-Graduação em Informática II - t

Dedico este trabalho aos meus pais e à minha namorada pela paciência, apoio e carinho.

Agradecimentos

Ao orientador Alceu de Souza Britto Junior e ao co-orientador Alessandro Lameiras Koerich pela paciência, estímulo e confiança dispensados para a realização deste trabalho.

Aos meus pais, Jaime e Salete, à minha irmã, Alice, e à minha namorada linda, Marilisa, pelo eterno incentivo e carinho.

A todos os amigos que de uma forma ou de outra me estimularam e apoiaram.

Sumário

Agradecimentos ii
Sumário iii
Lista de Figuras v
Lista de Tabelas vii
Lista de Algoritmos viii
Resumo ix
Abstract x

Capítulo 1

Introdução 1
1.1 Motivação 1
1.2 Definição do Problema 2
1.3 Objetivos 4
1.4 Método Proposto 4
1.5 Contribuições 5
1.6 Estrutura do Documento 6

Capítulo 2

Estado da Arte 7
2.1 Segmentação do Movimento 8
2.2 Rastreamento 10
2.3 Contagem de Pessoas 12

Capítulo 3

Método para Contagem de Pessoas 16
3.1 Captura e Pré-processamento 16
3.2 Segmentação 18

3.2.1 Tratamento de Ruídos 20
3.2.2 Tratando Objetos Desconectados 22
3.3 Rastreamento dos Objetos 22
3.4 Contagem de Pessoas 25
3.4.1 Abordagem Baseada em Limiares 26
3.4.2 Análise da Região Superior do Objeto 27
3.4.2.1 Características 28
3.4.2.2 Classificador 28

Capítulo 4

Experimentos e Resultados 33
4.1 Base de Dados 33
4.2 Calibração do Ambiente 35
4.3 Experimentos Realizados 36
4.3.1 Abordagem Baseada em Limiares Sobre Grupos de Pessoas 37
4.3.2 Análise da Região Superior do Objeto Sobre Grupos de Pessoas . . 38
4.4 Discussão 40

Capítulo 5

Conclusão 41
Referências Bibliográficas 43

Lista de Figuras

Figura 1.1 Sequência de imagens exemplificando a oclusão 3
Figura 1.2 Imagem exemplificando o problema da proximidade entre pessoas . 3
Figura 1.3 Exemplos da região superior de objetos 4
Figura 1.4 Visão geral do método proposto 5

Figura 2.1 a)Imagem Original em níveis de cinza. b)Resultado do método da subtração do fundo. c)Resultado do método da diferença temporal 9
Figura 2.2 Imagem exemplificando o posicionamento vertical de uma câmera (SNIDARO; MICHELONI; CHIAVEDALE, 2005) 13

Figura 3.1 Estrutura básica do método proposto 17
Figura 3.2 a)quadro de entrada. b)subtração do fundo usando o filtro da mediana. c) subtração do fundo sem o filtro da mediana. 18
Figura 3.3 a)quadro de entrada em níveis de cinza. b)resultado da subtração do fundo 20
Figura 3.4 a)quadro de entrada em níveis de cinza. b)resultado da subtração do fundo. c)resultado do filtro da abertura 21
Figura 3.5 a)quadro de entrada em níveis de cinza. b)resultado da subtração do fundo. c)resultado do filtro da abertura. d)resultado da Dilatação + Erosão 23
Figura 3.6 Exemplos de regiões de contagem 26
Figura 3.7 Região superior de um objeto. w indica a largura da *bounding box* do objeto. 27

Figura 3.8 Método baseado em limiares. a)Objetos rastreados em movimento. b)Quando o objeto identificado como 3 entra na região de contagem sua largura e área são verificadas e o contador é atualizado. c)O mesmo acontece para o objeto identificado como 2. d)Oclusão entre os dois objetos, rastreador trata o problema prevendo o movimento do objeto 3. e)Novo objeto em cena. f)Objeto identificado como 4, entra na região de contagem, com isso sua largura e área são verificadas e o contador atualizado. g)Objetos saindo da região de contagem. h)Objetos saem de cena. Resultado da contagem: 5 pessoas. 30
Figura 3.9 Geração do vetor de característica para n igual a 10. Exemplo da região superior de um objeto contendo duas pessoas 31
Figura 3.10 Vetor de Características 31
Figura 3.11 Método da Análise da região superior do objeto. a)Objetos rastreados em movimento. b)Quando o objeto identificado como 3 entra na região de contagem, a cada quadro do vídeo um vetor de característica é extraído dele e rotulado através do *k-NN*. c)o mesmo é feito pra o objeto identificado como 2. d)Oclusão entre os dois objetos, rastreador trata o problema prevendo o movimento do objeto 3. e)Novo objeto em cena. f) Quando o objeto identificado como 4 entra na região de contagem o processo descrito na letra a) é aplicado a ele. Objeto 2 sai da região de contagem, procedimento de extração e rotulação de vetores é interrompido para ele. g)Objeto 2 sai de cena, o voto da maioria é aplicado sobre seu vetores rotulados atualizando o contador. h)O mesmo acontece com os objetos 3 e 4. Resultado da contagem: 5 pessoas. 32

Figura 4.1 Vídeos da base de dados. a)quadro de um vídeo da CAVIAR Mall. b)quadro de um vídeo da Hall CCET 34
Figura 4.2 Exemplo de objeto onde a abordagem baseada em limiares falha contando duas pessoas ao invés de uma (limiar L=29). A abordagem da Análise da Região Superior do Objeto não falharia neste caso. 38
Figura 4.3 Exemplo de objeto onde a abordagem da Análise da Região Superior do Objeto falha contando uma pessoa ao invés de duas. 39
Figura 4.4 Exmplo de falha na etapa de segmentação. A cabeça da pessoa foi perdida devido ao baixo contraste com o fundo. 40

Lista de Tabelas

Tabela 4.1 Distribuição das classes 34
Tabela 4.2 Quantidade total de pessoas 34
Tabela 4.3 Calibração do ambiente fase segmentação 35
Tabela 4.4 Calibração do ambiente fase rastreamento 36
Tabela 4.5 Calibração do ambiente fase contagem 36
Tabela 4.6 Resultados das contagens automáticas 37
Tabela 4.7 Matriz de Confusão CAVIAR Mall - (Abordagem Baseada em Limiares) 37
Tabela 4.8 Matriz de Confusão CCET Hall - (Abordagem Baseada em Limiares) 38
Tabela 4.9 Matriz de Confusão CAVIAR Mall - (Análise da Região Superior do Objeto) 38
Tabela 4.10 Matriz de Confusão CCET Hall - (Análise da Região Superior do Objeto) 39
Tabela 4.11 Taxa de Acerto CAVIAR Mall (Análise da Região Superior do Objeto) 39
Tabela 4.12 Taxa de Acerto CCET Hall (Análise da Região Superior do Objeto) 39

Lista de Algoritmos

1 Filtro da mediana 19
2 Segmentação 20

Resumo

Este trabalho propõe um novo método para a contagem automática de pessoas utilizando técnicas de visão computacional a partir de vídeos capturados por uma câmera de circuitos fechados de TV (CFTV). O método proposto consiste em segmentar e rastrear os objetos do primeiro plano em cenas de vídeo para posteriormente realizar a contagem. Neste contexto, um dos principais problemas está na estimação do número de pessoas em grupos. Para isso, duas abordagens foram propostas. Para cada objeto rastreado que entra em uma região de contagem virtual do ambiente, uma das abordagens pode ser aplicada. A primeira delas é baseada em dois limiares. Um representa a largura média dos objetos que contém apenas uma pessoa, o outro representa a área média da região superior desses objetos, região esta que normalmente engloba as cabeças das pessoas contidas em um objeto. Comparando a largura e a área de cada novo objeto com estes limiares, decide-se se nele contém uma, duas ou três pessoas; A segunda abordagem utiliza um classificador previamente treinado para, dado um objeto, decidir se ele contém uma, duas, ou três pessoas. Para isso, é utilizado um esquema de zoneamento do objeto e características como área e largura são extraídas de sua região superior. O método proposto foi avaliado em duas bases de vídeos e as abordagens apresentaram resultados encorajadores.

Palavras-chave: Visão Computacional, Contagem de Pessoas, Câmeras de Circuitos Fechados de TV.

Abstract

This work presents a novel method for automatic people counting using computer vision techniques in videos captured through a conventional closed-circuit television camera (CCTV). The proposed method consists in segmenting and tracking foreground objects in video scenes to further make the counting. The main problem consistis in estimating the number of persons when such persons walk in groups, one very close to each other. To tackle this problem, two approaches are proposed. From each tracked object that reaches virtual counting zone, one of the proposed approaches can be applied. The first one is based on two thresholds which are related to the average width and to the average area of a blob top zone, which represents a person head. By matching the width and the head region area of a current blob against these thresholds it is possible to estimate if the blob encloses one, two or three persons; The second one, employs a classifier to decide to which class a object belongs (one, two or three persons). For such an aim, a zoning scheme is applied to each object and features such as area and width are extracted from the top region of the object. The method was evaluated in two video databases and the approachs have shown encouraging results.

Keywords: Computer Vision, People Counting, Closed-Circuit Television Camera.

Capítulo 1

Introdução

Atualmente existe um interesse crescente por sistemas de monitoramento de ambientes baseados em vídeo. Isto se deve, principalmente, a grande preocupação com a segurança em razão da crescente violência mundial, sobretudo pelos recentes incidentes terroristas, provocando assim, constantes discussões sobre segurança e, conseqüentemente, a busca de soluções para o problema. Muitas destas soluções estão voltadas ao desenvolvimento de métodos para o controle automático do número de pessoas em um determinado ambiente.

1.1 Motivação

Os sistemas de segurança baseados em vídeo disponíveis no mercado, popularmente conhecidos como circuitos fechados de TV (CFTV), em sua maioria estão limitados a captura, armazenamento e visualização dos vídeos feitos por uma ou mais câmeras. Alguns sistemas CFTV mais modernos contam com algoritmos de detecção de movimento, os quais são utilizados geralmente para ativar algumas funcionalidades do sistema, como a gravação de vídeo. Há também sistemas mais sofisticados que são capazes de classificar o comportamento de pessoas, podendo emitir alertas quando movimentações não convencionais ocorrem.

Neste contexto, a contagem automática do fluxo de pessoas em cenas de vídeo aparece como um novo tema de estudo em visão computacional. O crescente interesse por este tema não se deve somente a questões de segurança. Inúmeras aplicações podem ser associadas à contagem automática de pessoas. Uma aplicação que pode ser considerada, por exemplo, é o controle automático das funcionalidades de ambientes. Um edifício pode controlar a intensidade do ar-condicionado de cada andar, como em um sistema de aquecimento, ventilação e condicionamento de ar (HVAC), baseando-se no número

de pessoas presentes. Controlar a iluminação, somente acendendo ou apagando as luzes necessárias, de acordo com a quantidade e localização das pessoas em seu interior. O departamento de marketing de uma loja pode avaliar o impacto de uma campanha de marketing, contida em uma seção qualquer da loja, através do número de pessoas que ela atrai. Já em bancos e supermercados, para que não haja clientes insatisfeitos com a demora do atendimento, pode-se controlar o correto número de caixas necessários de acordo com o crescimento das filas, ou associado à quantidade de pessoas que entraram nesses estabelecimentos. Outra possível aplicação é projetar mudanças de *lay-out* em um estabelecimento baseando-se na trajetória do fluxo de seus clientes, e assim melhorar a circulação das pessoas. Também se pode estimar o fluxo de turistas em um determinado local e através desta estimativa alocar o número certo de funcionários para cada horário. Enfim, são várias as aplicações onde o monitoramento visual inteligente, com destaque para a área de contagem automática do fluxo de pessoas, pode ser empregado.

1.2 Definição do Problema

Mesmo com uma ampla aplicabilidade para sistemas automáticos de contagem de pessoas, grande parte do controle e medição do fluxo de pessoas em ambientes ainda é feita de maneira manual. Além disto, estes sistemas automático disponíveis no mercado, como os baseados em sensores infravermelho, não são capazes de contar o número de pessoas contidas em grupos, nos quais as pessoas encontram-se muito próximas umas das outras. Assim, em muitos casos grupos de pessoas são contados como sendo apenas uma única pessoa, prejudicando o resultado da contagem. Em outras palavras, tais sistemas não portam alguma inteligência para realizar tal tarefa. Já a solução manual depende da disposição do indivíduo que está coletando os dados, podendo facilmente se confundir em situações de muito movimento.

Considerando um ambiente real não controlado, com imagens obtidas através de uma câmera fixada no canto superior do ambiente, posicionada obliquamente a região a ser monitorada, como câmeras de segurança de um circuito interno de TV, um sistema completo de visão aplicado à contagem automática de pessoas deve ser capaz de identificar e perseguir as pessoas que se movimentam em cena e contar cada indivíduo apenas uma vez. Além disso, deve também ter a capacidade de estimar a quantidade de pessoas existentes em um grupo a fim de aperfeiçoar sua precisão. Para tal, o método precisa tratar situações de oclusão e de proximidade entre pessoas em um grupo.

A oclusão caracteriza-se pelo encobrimento, total ou parcial, do objeto rastreado por outro durante sua trajetória em cena (Figura 1.1). Já a proximidade entre pessoas

é observada quando duas ou mais pessoas andam em grupos (Figura 1.2). Devido ao posicionamento oblíquo da câmera estas duas situações acontecem naturalmente num ambiente onde pessoas entram e saem de cena a todo instante. Por exemplo, quando uma pessoa passa por trás de outra ou quando duas ou mais pessoas andam muito próximas umas das outras, os dois problemas são identificados. É possível evitá-los quase que na totalidade posicionando a câmera verticalmente à região supervisionada, porém o intuito é não realizar mudanças físicas no ambiente monitorado, ou seja, reutilizar as câmeras de segurança já instaladas.

Figura 1.1: Sequência de imagens exemplificando a oclusão

Figura 1.2: Imagem exemplificando o problema da proximidade entre pessoas

Desta forma, a presença de grupos normalmente acarreta em erros na contagem de pessoas. Assim, para o sucesso de um método de contagem automático, deve-se estimar com relativa precisão a quantidade de pessoas contidas em um grupo. Após análise minuciosa de vários vídeos observou-se que a região onde há maior discrepância entre grupos de pessoas é a região onde estão contidas as cabeças das pessoas que fazem parte destes grupos (Figura 1.3). Assim, a seguinte hipótese foi elaborada: a quantidade de in-

divíduos em um grupo de pessoas é estimável analisando a região que contém as cabeças dos membros deste grupo.

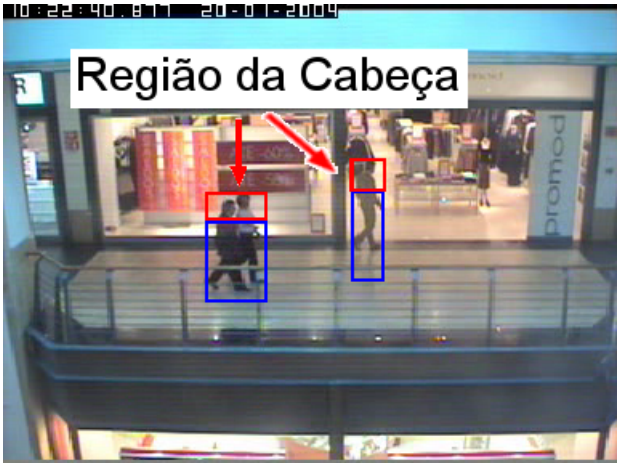

Figura 1.3: Exemplos da região superior de objetos

1.3 Objetivos

O objetivo principal deste trabalho é desenvolver um método para a contagem automática de pessoas a partir de vídeos capturados por uma camera de CFTV, posicionada de maneira oblíqua, utilizando técnicas de visão computacional minimizando o problema de oclusão e de grupos de pessoas. Assumindo uma área de contagem virtual definida no campo de visão da câmera de CFTV, o sistema determina o número de pessoas que circularam na área de contagem.

1.4 Método Proposto

O método proposto consiste em um sistema de visão completo e envolve quatro etapas básicas (Figura 1.4), a saber:

1. Captura e pré-processamento de vídeo.
2. Segmentação de objetos de interesse (separação fundo/primeiro plano).
3. Rastreamento de objetos de interesse (pessoas).
4. Contagem de pessoas dentro de uma região delimitada (resolução de oclusões e problema de pessoas em grupos).

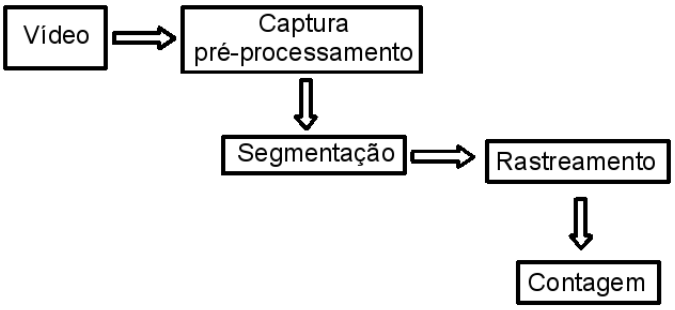

Figura 1.4: Visão geral do método proposto

A captura da cena é feita a partir de uma única câmera que deve estar fixada sobre o ambiente e disposta obliquamente à área monitorada, conforme a Figura 1.1. Cada quadro da seqüência capturada é pré-processado com intuito de reduzir pequenos ruídos decorrentes da variação na iluminação. O próximo passo consiste em separar o fundo do primeiro plano, onde os objetos de interesse (em movimento) depois de detectados são segmentados. Esta etapa baseia-se no método da subtração do fundo, que se mostra eficaz quando utilizado em ambientes fechados. Detectados os objetos de interesse, desenvolveu-se um rastreador capaz de perseguir um objeto quadro a quadro mesmo na presença de oclusão, tanto parcial quanto total. Para isto, são exploradas informações capazes de prever a posição do objeto ocluso. É importante destacar que o fluxo de pessoas no ambiente deve ser moderado, do contrário, em ambientes lotados, o método poderá incorrer em muitos erros. Finalmente é realizada a contagem de pessoas propriamente dita. Primeiramente é definido dentro do campo de visão da câmera uma região de contagem. Quando um objeto é rastreado para dentro desta região as abordagens que estimam o número de pessoas que este possui são aplicadas. Foram investigados dois procedimentos distintos: o primeiro é baseado em dois limiares: um representa a largura média de objetos que contêm apenas uma pessoa e o outro representa a área média da região superior desses objetos. Comparando a largura e área de cada novo objeto com estes limiares, define-se se este contém uma, duas ou três pessoas. O segundo procedimento considera a utilização de um classificador, o qual é usado para definir a qual classe o objeto pertence (uma, duas ou três pessoas). Para este fim, um esquema de zoneamento e extração de características da região da cabeça dos objetos é utilizado para a criação dos descritores das classes.

1.5 Contribuições

Dentre as principais contribuições deste trabalho destacam-se:

- No campo científico e tecnológico.
  - Apresentação de um novo método automático para contagem do fluxo de pessoas que poderá ser utilizado em câmeras de vigilância de circuitos internos de TV, ou seja, sem necessidade de instalar novas câmeras dedicadas a isto.
  - Avaliação de diferentes abordagens para tratamento do problema da contagem de pessoas quando agrupadas.
  - Disponibilização de uma base de vídeos rotulada que pode ser utilizada para avaliar métodos de contagem de pessoas.
- Do ponto de vista econômico e ambiental.
  - Adequação das condições de um ambiente em função do número de pessoas nele presente, evitando, por exemplo, o desperdício de energia.
  - Adequação do número de funcionários em função do número de pessoas presente em um estabelecimento comercial, evitando clientes insatisfeitos com a demora no atendimento e favorecendo o controle mais preciso do número de funcionários que o estabelecimento necessita.

1.6 Estrutura do Documento

O texto está organizado em cinco capítulos. O Capítulo 2 apresenta o estado da arte referente ao método proposto. Os trabalhos foram agrupados de acordo com as etapas do método. Em seguida, o Capítulo 3 descreve o método proposto, ou seja, como os procedimentos foram implementados. O Capítulo 4 mostra todos os resultados experimentais obtidos durante a avaliação do método. Por fim, o Capítulo 5 relata a conclusão do trabalho, suas contribuições, vantagens, desvantagens e trabalhos futuros.

Capítulo 2

Estado da Arte

O monitoramento visual é uma área amplamente pesquisada atualmente, e muitos trabalhos têm sido publicados, como pode ser observado nas revisões (HU et al., 2004) (KASTRINAKI; ZERVAKIS; KALAITZAKIS, 2003). Os trabalhos de pesquisa nesta área, direcionados ao processo de contagem e estimação do número de pessoas, geralmente utilizam algoritmos para a segmentação de objetos em movimento do restante da cena junto com algoritmos de rastreamento de objetos em movimento ao longo de quadros sucessivos do vídeo para realizarem assim a contagem dos objetos. A utilização destes algoritmos em sistemas de monitoramento visual é a opção para sistemas que objetivam a classificação dos objetos, bem como suas atividades e comportamentos, através da análise dos seus formatos e suas trajetórias. Estes sistemas possuem em comum a característica da separação dos objetos que se movimentam (primeiro plano) do fundo, pois as análises são realizadas sobre este plano (SNIDARO; MICHELONI; CHIAVEDALE, 2005) (MASOUD; PAPANIKOLOPOULOS, 2001) (KETTNAKER; ZABIH, 1999) (HARITAOGLU; FLICKNER, 2002) (HARITAOGLU; HARWOOD; DAVIS, 2000).

Neste capítulo serão apresentados os principais trabalhos relacionados à contagem de pessoas, utilizando técnicas de visão computacional em ambientes monitorados. Será apresentada também uma revisão bibliográfica das principais técnicas que serão necessárias para o desenvolvimento do trabalho. Os trabalhos analisados foram agrupados de acordo com as etapas do método proposto, sendo que em alguns casos as etapas se confundem, por exemplo, o final da etapa de rastreamento pode ser considerado o final da etapa de contagem, onde cada objeto localizado e rastreado é considerado uma pessoa.

2.1 Segmentação do Movimento

A segmentação do movimento contido em uma seqüência de imagens, consiste em detectar e separar os objetos que se moveram com o passar do tempo, como, por exemplo, veículos e pessoas.

Um dos mais populares métodos é o da subtração do fundo. Esse método é simples, rápido e separa o primeiro plano do fundo fazendo a diferença, pixel a pixel, de cada novo quadro do vídeo com a imagem de referência do fundo (SNIDARO; MICHELONI; CHIAVEDALE, 2005) (HARITAOGLU; FLICKNER, 2002) (HARITAOGLU; HARWOOD; DAVIS, 2000) (BJöRGVINSSON, 2006). É dependente de um fundo estático, sem muitas mudanças, pois é extremamente sensível a elas.

Em (HARITAOGLU; HARWOOD; DAVIS, 2000), é utilizado a subtração seguida de limiarização para determinar quais os pixels fazem parte do fundo em um ambiente externo. Porém, por ser um método sensível a mudanças de iluminação, a imagem de fundo usada como referência para a realização da subtração é construída durante um período de treinamento estatístico e é atualizada de tempos em tempos através de dois métodos: atualização baseada no pixel e atualização baseada no objeto. O primeiro adapta-se às mudanças de iluminação e o último adapta-se a mudanças físicas, por exemplo, um carro que estacionou e não se moveu por um longo período se tornará parte do fundo. Isto é necessário, pois num ambiente externo não se espera ter o mesmo fundo por longos períodos de tempo. No trabalho (KIM K.-S. CHOI; KO, 2002) também foi utilizado um método para atualização da imagem de fundo e, assim, evitar a segmentação de falsos objetos decorrentes de pequenas variações da iluminação.

Masoud e Papanikolopoulos (2001) fazem a segmentação através da subtração do fundo seguida de uma limiarização, onde o limiar é definido em uma etapa de treinamento sobre a imagem referência do fundo. Além disso, esta imagem é atualizada periodicamente através de uma função recursiva que captura mudanças lentas do fundo, como a mudança na iluminação devido à passagem de uma nuvem. Em seguida, um filtro de fechamento é aplicado sobre a imagem para a remoção de pequenos grupos.

Haritaoglu e Flickner (2002) detectam objetos através de um processo estatístico da subtração do fundo. A imagem de fundo é modelada durante um período de treinamento representando cada pixel com três valores de intensidade: mínima, máxima e a diferença máxima entre consecutivos quadros. Para reduzir os erros, causados pela movimentação da câmera, um pixel será considerado do primeiro plano se, e somente se, for detectado como tal nos últimos cinco quadros consecutivos. Assim, a silhueta dos objetos é obtida combinado o resultado de consecutivas detecções do primeiro plano.

Já em (LIPTON; FUJIYOSHI; PATIL, 1998), é utilizado a técnica da diferença temporal para segmentar os objetos em movimento. O método da diferença temporal, calcula, em uma seqüência de imagens, a diferença entre dois ou três quadros consecutivos para a segmentação do movimento. Como ela não segmenta objetos fixos (pessoas que ficam paradas no mesmo lugar são consideradas objetos fixos pelo método), houve a necessidade de utilizar um classificador que definisse se o objeto rastreado é um humano ou um veículo, pois se um veículo estacionar, ele se torna parte do fundo, ao contrário de um ser-humano, aprimorando o método. Na Figura 2.1 é possível comparar o resultado do método da subtração do fundo com o da diferença temporal.

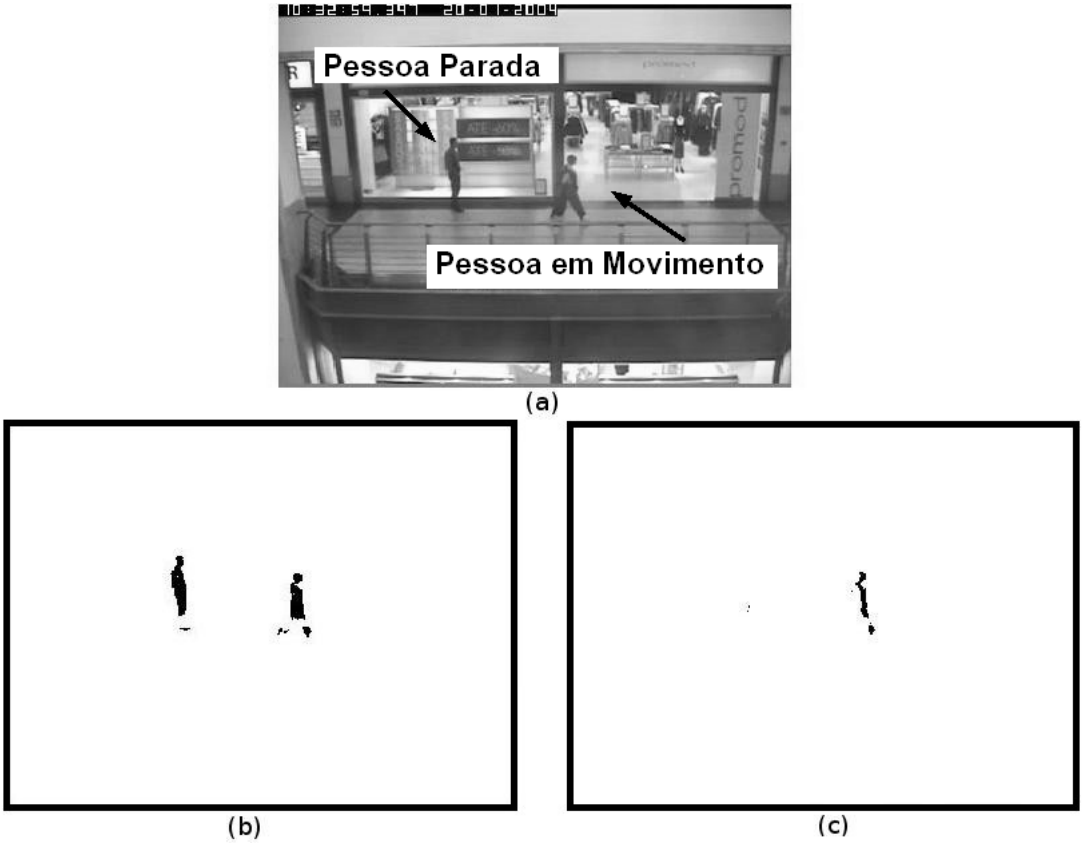

Figura 2.1: a)Imagem Original em níveis de cinza. b)Resultado do método da subtração do fundo. c)Resultado do método da diferença temporal

O trabalho descrito em (DIAS, 2005) também utiliza a subtração do fundo para realizar a tarefa de segmentação. No entanto junto com este método é realizado o tratamento de sombras, onde as componentes de cromaticidade são analisadas para eliminá-las (PRATI et al., 2001), isto é possível pois o sistema de cores usado tem o brilho expresso em apenas uma componente de cor. Para deixar a segmentação mais precisa, o fundo é estimado a cada novo quadro do vídeo para tratar variações de iluminação no ambiente mesmo na presença de pessoas.

Stauffer e Grimson (2000) realizam a segmentação fazendo a modelagem de cada pixel como sendo uma mistura Gaussiana. Baseando-se na persistência e na variância de cada Gaussiana da mistura, são determinadas quais Gaussianas correspondem ao fundo. Os valores dos pixels que não se encaixam nas distribuições do fundo são considerados partes do primeiro plano até haver uma Gaussiana que os inclua numa nova mistura do fundo. Este método é capaz de absorver movimentos contínuos. Isso quer dizer que após um curto período de tempo, um objeto que executa movimentos periódicos, como pequenas ondas em uma lagoa, é absorvido por uma das curvas da Guassiana e considerado como pertencente ao fundo, aprimorando o método.

Cada método possui suas vantagens e desvantagens que são diretamente dependentes da aplicação onde serão empregados, podendo, até mesmo, serem utilizados em conjunto. Os métodos da subtração do fundo junto com o método da diferença temporal são os mais utilizados nos trabalhos estudados. Isto se deve ao fato de eles serem simples de implementar e por serem muito rápidos, ao contrário da abordagem das Gaussianas, proposta por Stauffer e Grimson, que é computacionalmente mais custoso. A subtração do fundo é capaz de segmentar inclusive objetos que param em cena, porém é o método mais sensível a variações na iluminação. Já a diferença temporal e o método das Gaussianas são mais tolerantes a estas variações, porém não são capazes de segmentarem objetos que param no meio da cena e, além disso, quando os objetos se movimentam lentamente, somente as suas bordas são segmentadas. O que deve ser notado é que mesmo realizando a atualização da imagem de fundo periodicamente e a aplicação de filtros sobre os quadros do vídeo, em seqüências de imagens do mundo real objetos indesejados podem ser segmentados como parte do primeiro plano.

A seguir, alguns métodos de rastreamento de objetos são apresentados. Estes métodos são aplicados sobre os objetos do primeiro plano, ou seja, são aplicados após a segmentação.

2.2 Rastreamento

Rastrear um objeto significa perseguir, quadro a quadro, sua trajetória durante o tempo em que está em cena. O rastreamento é, na maioria das vezes, realizado sobre o resultado da segmentação obtido na etapa anterior.

Existem inúmeras aplicações para o rastreamento de pessoas e cada uma possui necessidades particulares, como mostra a revisão (YILMAZ; JAVED; SHAH, 2006). Há aquelas que precisam de um rastreamento que mostre detalhes das partes do corpo humano (WREN et al., 1997) (POLAT; YEASIN; SHARMA, 2001), como, por exemplo, o acompanhamento da

postura de um atleta visando a melhora de sua performance ou a tradução automática das libras, linguagem de sinais utilizados por surdos para comunicação, para a linguagem oral, pois são diretamente dependentes deste detalhamento, e existem as aplicações que não exigem um nível tão preciso de detalhamento como o controle de tráfego de veículos (KASTRINAKI; ZERVAKIS; KALAITZAKIS, 2003) e a detecção de eventos não convencionais.

Em (SNIDARO; MICHELONI; CHIAVEDALE, 2005) foram utilizados *bounding boxes* em torno de cada objeto segmentado para delimitar a área de extração de características (densidade, coordenadas da *bounding box*, coordenadas do centróide, média dos valores de cor, histograma, entre outras) para alimentarem a entrada do método conhecido como filtros de kalman (CUEVAS; ZALDIVAR; ROJAS, 2005) que é capaz de prever o novo estado do sistema se baseando no passado do objeto (características extraídas) e, assim, rastreá-lo.

No trabalho de (KIM K.-S. CHOI; KO, 2002) o rastreamento é realizado extraíndo características da *bounding box* em torno dos objetos, como área e ponto do centro. Para extrair características mais precisas, o fecho convexo do objeto dentro da *boundig box* é calculado. Para encontrar o objeto no quadro corrente, comparam-se as características do passado com os objetos mais próximos do quadro atual. Caso o objeto se uniu com outro no quadro atual, sendo assim segmentados como sendo apenas um objeto, seus movimentos são previstos analisando suas velocidades e direções até o momento, tratando desta maneira a oclusão de objetos.

O trabalho descrito em (BJöRGVINSSON, 2006) localiza os segmentos do quadro anterior no atual através da distância Euclidiana dos pontos. Os centros dos objetos indicam suas posições no quadro. Assim, os segmentos que possuem a menor distância são considerados como sendo a mesma pessoa.

Já Haritaoglu, Harwood e Davis (2000) constroem um modelo de movimento para cada pessoa, baseando-se em sua velocidade e aceleração, para estimar sua localização nos quadros subseqüentes. A posição da pessoa no quadro atual é calculada comparando a sua posição estimada e seu tamanho com as posições e tamanhos dos objetos ainda não identificados do quadro corrente, onde os mais próximos são associados.

Haritaoglu e Flickner (2002) realizam o rastreamento baseado na aparência. Para isso, é criado um modelo de aparência dinâmica para cada pessoa detectada em cena. O rastreamento é feito comparando-se o modelo construído com os novos segmentos. Este modelo é atualizado durante o rastreamento utilizando a informação de dois componentes: um componente textural que representa a aparência em níveis de cinza e um componente de forma que representa a informação do formato esperado da silhueta do objeto. Basicamente o método computa a correspondência entre as silhuetas segmentadas do quadro

atual com as silhuetas rastreadas do quadro anterior usando a correlação dos modelos de aparência dinâmicos criados. Ramanan e Forsyth (2003) também baseiam-se na aparência para realizar o rastreamento. Há uma etapa de aprendizagem onde os quadro do vídeo são varridos em busca de prováveis segmentos do corpo de um indivíduo, que são os segmentos que se mantêm unidos e similares com o passar do tempo. Em seguida esses segmentos são agrupados de maneira que representem o corpo de uma pessoa. Na etapa de busca, onde se esperam encontrar os modelos aprendidos nos quadros subseqüentes, uma rede Bayesiana é usada.

Em (TSAI; SHIH; HUANG, 2006) é proposto um método para rastrear pessoas em cenas lotadas. Cada objeto em cena é representado por um modelo de cor de duas regiões do corpo, a do torso (roupas) e a inferior (calças). Isto é feito usado o modelo da mistura Gaussiana. Para estimar a posição do objeto segmentado no próximo quadro a técnica do fluxo óptico é usada. Quando ocorre oclusão entre dois objetos, o fluxo óptico irá falhar, pois eles estão sobrepostos. Para esta situação, o modelo de cor do objeto é usado para estimar sua posição enquanto ocluso. O método é capaz de rastrear múltiplos objetos em cenas lotadas, onde oclusões ocorrem frequentemente.

Latecki e Miezianko (2006) localizam os objetos do quadro anterior através de uma função de custo baseada na distância, a qual utiliza informações como posição, tamanho da *bounding box* e direção nesta comparação. Se um objeto do quadro anterior não for encontrado no quadro atual, sua posição é estimada através de seu vetor de movimento. Se não reaparecer em um tempo determinado, ele é incorporado ao objeto mais próximo. Tratando desta maneira a oclusão entre objetos.

Os rastreadores geralmente localizam os segmentos correspondentes através de uma função baseada em distância, comparando características do passado dos objetos rastreados com as do presente. Características do passado também são usadas para tratar o problema da oclusão, prevendo os movimentos dos objetos oclusos. A seguir são apresentados alguns métodos de contagem de pessoas que em sua maioria utilizam do rastreamento para realizar ou ajudar na contagem.

2.3 Contagem de Pessoas

Para a contagem de pessoas muitas técnicas foram propostas. Em sua maioria utilizam um algoritmo de rastreamento como base e abordam o problema da proximidade entre pessoas, porém existem aqueles que consideram cada segmento do primeiro plano como sendo uma pessoa (KHAN et al., 2001) (HARITAOGLU; FLICKNER, 2002).

Em (SNIDARO; MICHELONI; CHIAVEDALE, 2005), é estimada a quantidade de pes-

soas dentro de um objeto com base em sua área. Devido ao posicionamento da câmera sobre o ambiente (Figura 2.2), que soluciona a maioria dos problemas gerados pela oclusão, o tamanho do objeto que representa uma única pessoa é de certo modo constante. Assim, quando um objeto cruza uma linha de contagem, estima-se a quantidade de pessoas contidas nele, comparando sua área com a área ocupada pelo objeto que representa uma única pessoa. Kim et al. (2002) e Kim et al. (2003) também utilizaram este posicionamento para a câmera. Porém sua contagem é realizada somente com o rastreador, ou seja, cada objeto é contado como uma pessoa quando entra no campo de captura da câmera e é rastreado até uma das linhas de contagem.

Figura 2.2: Imagem exemplificando o posicionamento vertical de uma câmera (SNIDARO; MICHELONI; CHIAVEDALE, 2005)

No trabalho (HARITAOGLU; HARWOOD; DAVIS, 2000) a contagem do número de pessoas é feito identificando suas cabeças através da análise de características globais e locais. Para isso, são combinados dois métodos baseados na geometria do formato do objeto. Um analisa o contorno superior do objeto, verificando se o formato dos pixels desta região é similar ao formato da curvatura de uma cabeça. O outro se baseia na projeção do histograma vertical do objeto, os picos significantes são usados para filtrar os resultados da análise local do formato feito pelo primeiro método. Formas parecidas com cabeças são guardadas somente se há uma projeção significativa dos picos em sua vizinhança. Em outras palavras, picos que não atingem um limiar (altura) necessário, que foram selecionados pelo método anterior, são eliminados. Nos picos que restaram é analisado se o eixo do torso, linha traçada do pico até o chão, passa por dentro da silhueta segmentada. Se sim, é considerada uma pessoa para a contagem.

Em outra oportunidade Haritaoglu e Flickner (2002) também analizaram a silhueta do objeto a procura de regiões parecidas com cabeças para a localização das pessoas em cena. O objetivo do trabalho consiste em contar o tempo que cada indivíduo passa observando uma vitrine e em descobrir o seu gênero (masculino ou feminino). Para isso, primeiro é necessário identificar cada pessoa em cena. O sistema para localização de pessoa

foi complementado, pois devido à posição da câmera, a cabeça de uma determinada pessoa poderia estar no interior da silhueta de uma outra pessoa, impossibilitando sua localização através da silhueta. Assim, um esquema de regiões de coerência temporal foi aplicado, onde a detecção é baseada em padrões de movimento do passado. Para identificar quem esta olhando para os avisos, as pupilas das pessoas são detectadas por meio de iluminação infravermelha. Para descobrir o gênero das pessoas, um classificador SVM é utilizado.

Em (KETTNAKER; ZABIH, 1999) foi proposto um método diferente para se contar pessoas. Um método que utiliza múltiplas câmeras que não se sobrepõem, ou seja, cada câmera está monitorando um lugar diferente do mesmo ambiente. A idéia principal deste método são as restrições que o ambiente monitorado pode fornecer. Em outras palavras, uma pessoa não pode, por exemplo, estar no campo de captura da câmera três sem antes ter passado pelas câmeras um e dois. Assim, a contagem de pessoas é realizada com base nas restrições topológicas do ambiente junto com um modelo estatístico da aparência construída para cada pessoa, para que elas possam ser identificadas nas câmeras subseqüêntes.

No trabalho (DIAS, 2005) a localização e contagem de pessoas é feita baseada na técnica da coerência de movimento, uma adequação ao método descrito por Shapiro e Stockman (2001), que parte de três hipóteses principais.

- Segmentos pertencentes a uma só pessoa se movimentam praticamente da mesma maneira entre quadros consecutivos.
- O deslocamento de um segmento entre quadros consecutivos é moderado.
- Pixels pertencentes a uma só pessoa formam um único componente conectado.

O método localiza para cada segmento do quadro atual seu correspondente do quadro anterior através de uma medida de similaridade baseada no cálculo do módulo da diferença de cores entre os pares de pixels correspondentes. Em seguida, vetores que definem o movimento de cada uma das regiões encontradas é computado. Por fim, o segmento que não encontrou sua região correspondente é agregado ao grupo de segmentos adjacentes, com movimentos coerentes de um quadro ao outro, em formação. Assim, cada grupo formado corresponde a um conjunto de segmentos coerentes e adjacentes ou apenas adjacentes, que são considerados pertencentes a uma única pessoa.

Björgvinsson (2006) utiliza câmeras posicionadas verticalmente sobre o ambiente, saídas e entradas de lojas, para contar quantas pessoas saíram e entraram no estabelecimento. Quando um objeto passa pela área de fluxo de contagem ele é rastreado até sair de cena. Se desaparecer através da área de entrada, o contador de entrada é incrementado.

Se o contrário ocorrer, ou seja, se desaparecer através da área de saída, o contador de saída é incrementado. Um objeto pode ser dividido em dois, quando sua largura ultrapassa a largura máxima que ele possa ter. Assim, a contagem se torna mais precisa.

Em (SIDLA et al., 2006), foi proposto uma maneira de contar pedestres em ambientes superlotados. Os pedestres são diferenciados a partir de características extraídas de sua região ombro/cabeça, desta maneira são rastreados quadro a quadro a partir destas características, onde cada uma destas regiões representa uma pessoa. Para a localização destas regiões, em cada quadro do vídeo é aplicando o algoritmo detector de contornos conhecido como *Canny*. Em seguida, as regiões são encontradas comparando o conjunto de pontos resultante com um conjunto de pontos modelo extraídos manualmente.

Os trabalhos sobre contagem de pessoas sempre utilizam, de alguma forma, um algoritmo de rastreamento para que uma mesma pessoa em cena não seja contada mais de uma vez. As soluções com câmeras posicionadas verticalmente ao ambiente evitam muitos problemas decorrentes da oclusão, pois oclusões totais são impossíveis de ocorrer com este posicionamento. Porém, necessitam de câmeras especialmente instaladas para este fim. No entanto, para os outros posicionamentos de câmera o problema da oclusão acontece normalmente no ambiente monitorado, trazendo a necessidade de tratar tal problema, além de haver a necessidade de tratar o problema de grupos de pessoas para que a contagem seja a mais próxima da real. Com base nos trabalhos apresentados, o próximo capítulo mostra o método proposto para realizar a contagem automática de pessoas.

Capítulo 3

Método para Contagem de Pessoas

No capítulo anterior revisamos os principais trabalhos que compõem um sistema de contagem de pessoas. Vimos que a grande parte deles utiliza um algoritmo de rastreamento, para tratar o problema da oclusão, e analisam os objetos do primeiro plano, para tratar o problema da proximidade entre pessoas, para enfim realizar a contagem. Visando resolver alguns destes problemas, neste capítulo foram propostos duas abordagens para contagem de pessoas: uma baseada em limiares; e outra baseada em um classificador e em características extraídas da região da cabeça dos objetos. Eles, principalmente propõem tratar uma das principais causas de erros em soluções deste tipo, que é estimar de maneira mais precisa o número de pessoas em um grupo. Deste modo, a solução proposta para atender os objetivos deste trabalho consiste nas seguintes etapas (Figura 3.1):

- Captura de vídeo e pré-processamento dos quadros.
- Segmentação dos objetos em movimento (separação entre fundo e objetos do primeiro plano).
- Rastreamento dos objetos segmentados ao longo dos quadros do vídeo.
- Contagem de pessoas em regiões de interesse.

Nas seções seguintes cada uma das etapas do método proposto são explicadas detalhadamente.

3.1 Captura e Pré-processamento

Para que o método proposto funcione corretamente é necessário que a aquisição dos vídeos seja realizada por uma câmera posicionada obliquamente ao ambiente a ser

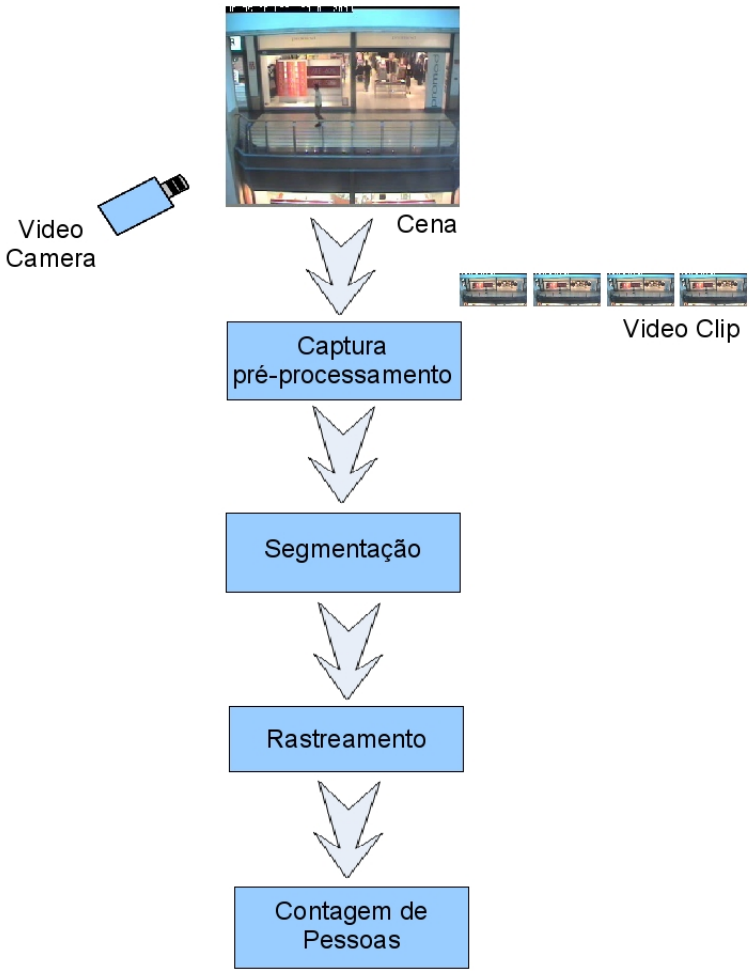

Figura 3.1: Estrutura básica do método proposto

monitorado. Esta restrição é imposta, pois a hipótese proposta levou em conta o posicionamento que as câmeras de CFTV geralmente têm. A Figura 1.3 mostra dois exemplos de como a câmera deve ser posicionada em relação as pessoas que transitam em seu campo de visão.

Quando uma câmera é colocada sobre o ponto de contagem, minimiza-se o problema de oclusão e o de agrupamento de pessoas, fornecendo o campo de visão ideal para contagem. O presente método propõe não utilizar uma câmera dedicada e sim reaproveitar as câmeras de CFTV já instaladas e logo, está mais susceptível a erros decorrentes dos dois problemas citados.

Em relação ao pré-processamento dos vídeos capturados, é aplicado um filtro da mediana 3x3 bi-dimensional (Algoritmo 1) sobre cada um dos canais de cores (RGB) para a redução de ruídos dos quadros do vídeo (DIAS, 2005). Esta etapa é importante para corrigir certas imperfeições durante a aquisição do vídeo, principalmente por conta da iluminação (Figura 3.2). Por fim, o vídeo capturado é transformado para níveis de cinza.

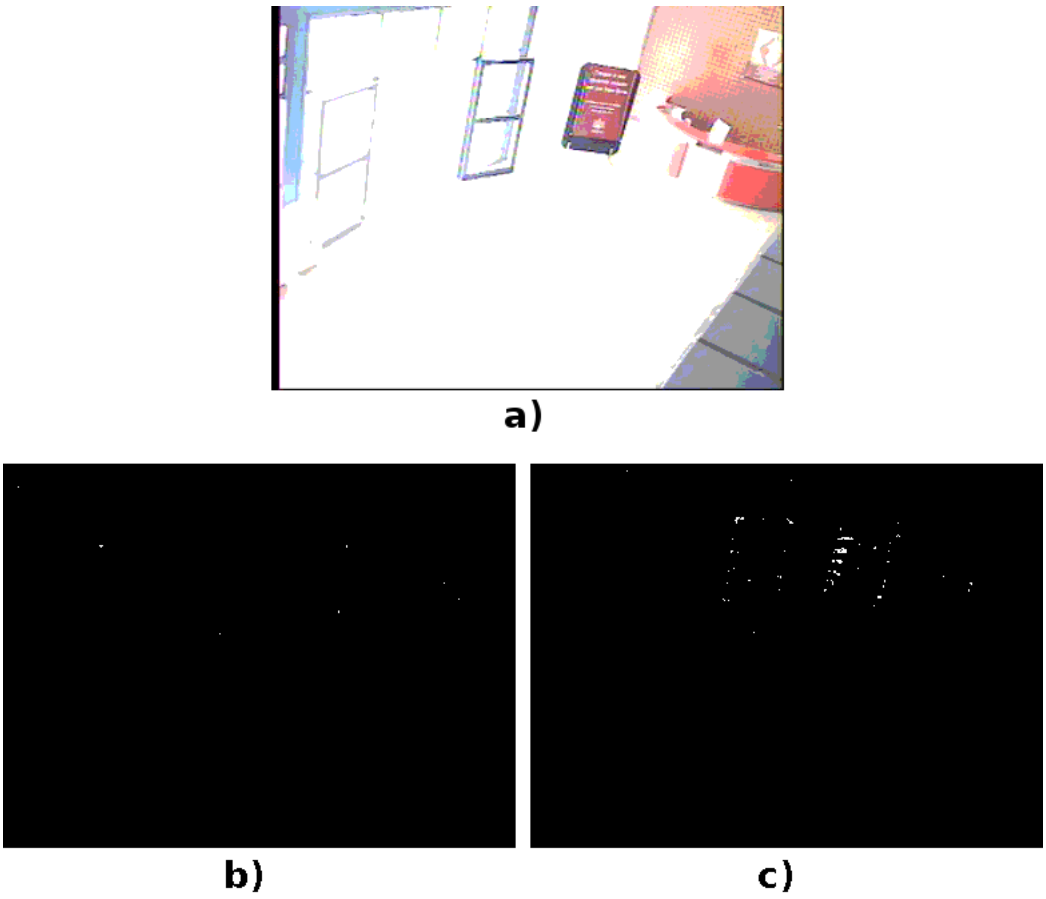

Figura 3.2: a)quadro de entrada. b)subtração do fundo usando o filtro da mediana. c) subtração do fundo sem o filtro da mediana.

3.2 Segmentação

A etapa de segmentação objetiva distinguir nos quadros do vídeo os pixels que representam o fundo dos que representam o primeiro plano. O resultado desta etapa é uma imagem binária onde o fundo é representado pela cor preta e os objetos do primeiro plano pela branca.

O método aplicado para chegar a esta imagem binária é o da subtração do fundo. Um método muito utilizado em sistemas de visão, mesmo este sendo muito sensível a variações na iluminação. A partir de uma imagem de referência do fundo, indicado por F_{fixo} o método consiste em subtrair cada nova imagem F^n, onde n é o quadro do vídeo que a imagem representa, da imagem F_{fixo}, pixel a pixel. A idéia deste método é identificar as regiões que se movimentaram em relação a esta imagem de referência. A imagem de referência do fundo é uma imagem em que não há objetos no primeiro plano, mas somente o ambiente. Geralmente o primeiro quadro do ambiente, quando não há ninguém em cena, é guardado como imagem de referência do fundo. A equação 3.1 mostra a operação da

Algoritmo 1: Filtro da mediana

Entrada: uma imagem no sitema de cores RGB de tamanho $x \times y$: Img
Saída: imagem filtrada: $filtrada$

```
1  para cada canal ← 1 até 3 faça
2      para i ← 2 até y − 1 faça
3          para j ← 2 até x − 1 faça
4              k ← 1;
5              para i_janela ← i − 1 até i + 1 faça
6                  para j_janela ← j − 1 até j + 1 faça
7                      janela[k] ← Img[canal, i_janela, j_janela];
8                      k ← k + 1;
9                  fim
10             fim
               /* Sort ordena o vetor de modo crescente                 */
11             Sort(janela);
12             filtrada[canal, i, j] ← janela[5];
13         fim
14     fim
15 fim
16 retorna filtrada
```

subtração do fundo,

$$Sub^n = |F_{fixo} - F^n| \tag{3.1}$$

onde Sub^n indica a n-ésima imagem resultante do processo de subtração pixel a pixel da imagem de x linhas e y colunas do fundo da cena, indicado por F_{fixo}, com o n-ésimo quadro do vídeo, indicado por F^n.

Em seguida o resultado da subtração é então limiarizado utilizando um limiar global e fixo para todos os quadros do vídeo de um ambiente. Baseando-se na imagem Sub^n resultante da subtração e de em limiar fixo l, o valor deste limiar é definido empiricamente através da análise de desempenho de cada valor testado de maneira visual sobre os vídeos. A limiarização é realizada, pixel a pixel, da seguinte maneira (Figura 3.3):

$$L^n = \begin{cases} 0, & \text{Se } Sub^n < l \\ 1, & \text{caso contrário.} \end{cases} \tag{3.2}$$

onde L^n é a imagem, de y linhas e x colunas, limiarizada resultante. O 0 representa a cor preta e 1 a cor branca.

Desta maneira se obtém a imagem binária com o fundo e o primeiro plano, aparentemente, distintos (Algoritmo 2). Aparentemente, pois existem muitos fatores que fazem

Figura 3.3: a)quadro de entrada em níveis de cinza. b)resultado da subtração do fundo

com que aconteça a segmentação de falsas regiões do primeiro plano. O principal deles é a variação de iluminação na cena, pois esta gera muitos ruídos e sombras, dois dos principais problemas indesejáveis nesta etapa.

Algoritmo 2: Segmentação

Entrada: F^n e F_{fixo} de tamanho $x \times y$
Saída: L^n

```
1  para i ← 1 até y faça
2      para j ← 1 até x faça
3          Sub^n[i,j] ← |F_fixo[i,j] − F^n[i,j]|;
4          se Sub^n[i,j] < l então
5              L^n[i,j] ← 0;
6          senão
7              L^n[i,j] ← 1;
8          fim
9      fim
10 fim
11 retorna L^n
```

Como todos os experimentos são realizados sobre vídeos de ambientes fechados não se observa problemas com sombras. Como há presença de ruídos, é necessária a aplicação de um método específico para tratar deste problema. O método proposto será descrito a seguir.

3.2.1 Tratamento de Ruídos

Devido à persistência de ruídos nos quadros limiarizados por conta da variação de iluminação, amenizada pela aplicação do filtro no vídeo de entrada, faz-se novamente

necessário à aplicação de um filtro nesta fase.

Para a eliminação destas ocorrências é aplicado uma operação morfológica de "abertura binária"no quadro limiarizado L^n. A operação de abertura consiste em erodir e depois dilatar o resultado da erosão (Figura 3.4), como é mostrado na equação 3.3,

$$L'^n = L^n \circ B \tag{3.3}$$

onde ($\circ$) indica a operação de abertura e B o elemento estruturante usado. Nos experimentos descritos no Capítulo 4 utilizou-se como elemento estruturante um quadrado 3x3.

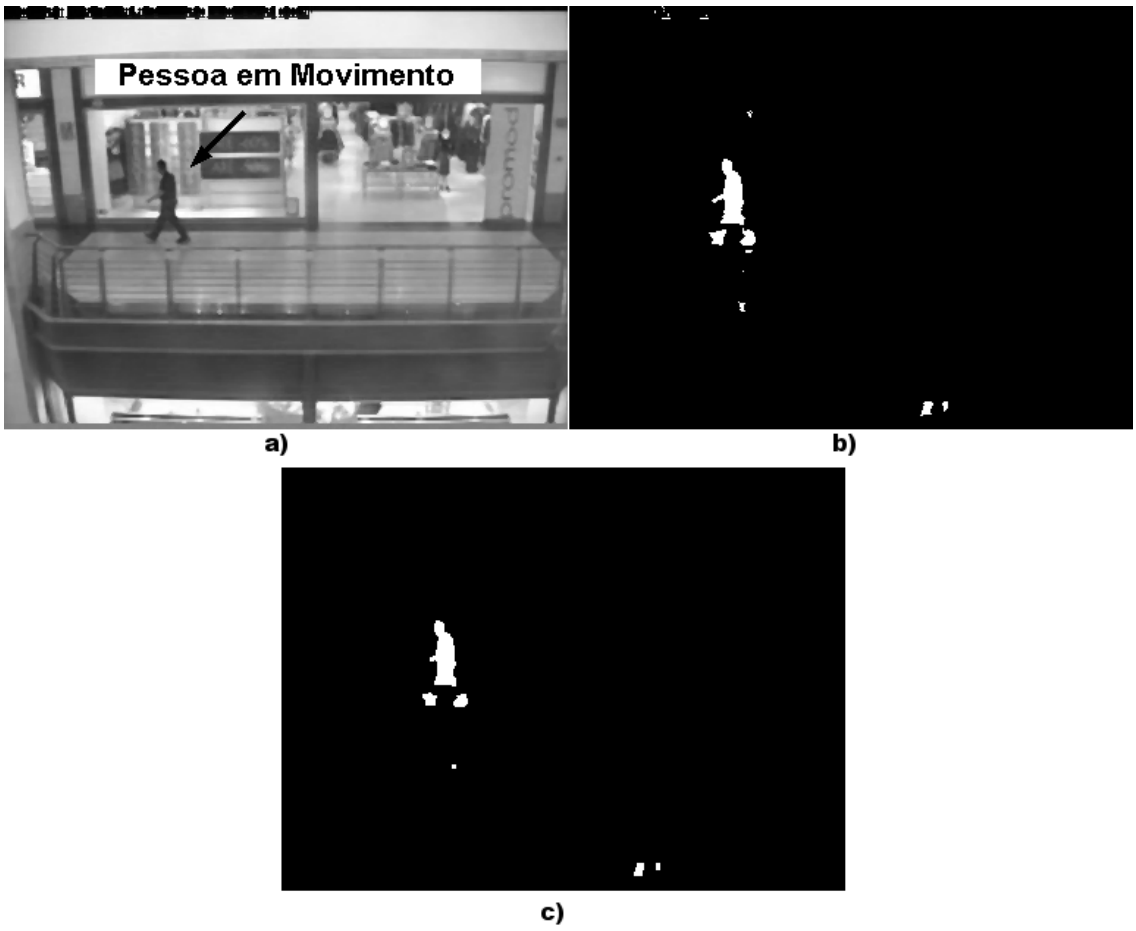

Figura 3.4: a)quadro de entrada em níveis de cinza. b)resultado da subtração do fundo. c)resultado do filtro da abertura

Com esta nova filtragem o resultado que se obtém é o conjunto dos pixels dos objetos em movimento segmentados. Todavia, neste resultado há a possibilidade de existirem regiões de um mesmo objeto desconectadas. Trazendo a necessidade de conectá-las para que elas não sejam identificadas como objetos distintos e sim como um mesmo objeto na

etapa de rastreamento. A seguir é descrito o método proposto para tratar este problema.

3.2.2 Tratando Objetos Desconectados

O baixo contraste entre o fundo e o objeto em movimento pode levar a não segmentação de algumas partes dele, fazendo com que um único objeto tenha duas ou mais partes desconectadas. Estes segmentos devem ser agrupados a fim de representar cada objeto por apenas um único conjunto de pixels conectados.

Assim, com o objetivo de conectar estas partes para se obter um único objeto, são utilizadas duas operações morfológicas. Primeiro é aplicada a operação de dilatação para conectar os segmentos mais próximos. Em seguida é aplicada a erosão para que as características de forma e tamanho dos objetos sejam mantidas. As equações referentes a estas operações são descritas abaixo:

$$L^{''n} = L^{'n} \oplus B \tag{3.4}$$

na qual ($\oplus$) é a operação de dilatação aplicada a imagem $L^{'n}$ usando elemento estruturante B (quadrado 3x3).

$$L^{'''n} = L^{''n} \ominus B \tag{3.5}$$

na qual ($\ominus$) é a operação de erosão aplicada a imagem $L^{''n}$ usando elemento estruturante B (quadrado 3x3).

Com isso as regiões desconectadas de um objeto são conectadas tomando-se o cuidado para não modificar o seu tamanho (Figura 3.5). Ao final desta etapa temos todos os objetos em movimento segmentados, prontos para serem utilizados no próximo passo, o de rastreamento.

3.3 Rastreamento dos Objetos

O rastreamento de um objeto em movimento consiste em localizá-lo na imagem e seguir sua trajetória nos quadros subseqüentes enquanto ele permanece em cena. Um rastreador de objetos deve ser capaz de identificar um objeto do quadro anterior no quadro atual até mesmo quando ocorrer oclusão. Ou seja, mesmo quando o objeto desaparece por alguns quadros, não se perde seu rastreamento. Pensando nisto, o rastreador escolhido baseia-se na abordagem proposta por Latecki e Miezianko 2006 que trata o proble da oclusão fazendo a predição do movimento do objeto ocluso. Este método é explicado a

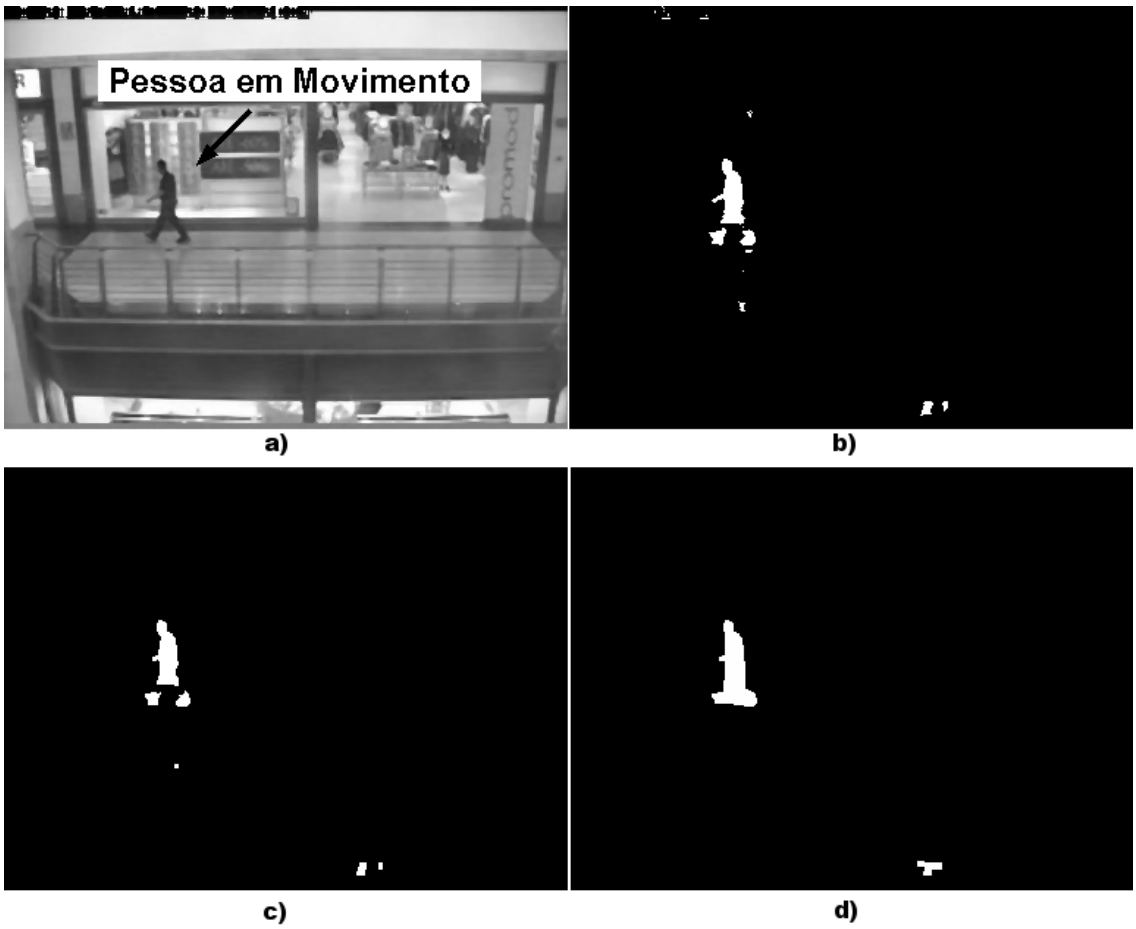

Figura 3.5: a)quadro de entrada em níveis de cinza. b)resultado da subtração do fundo. c)resultado do filtro da abertura. d)resultado da Dilatação + Erosão

seguir.

De posse dos segmentos do primeiro plano, um filtro de tamanho é aplicado sobre eles com o intuito de descartar aqueles objetos que não possuem altura ou largura suficientes para serem considerados uma pessoa. A idéia de utilizar filtros para eliminar regiões não desejáveis foi proposta por Lei e Xu 2005.

Assumindo que F^n representa o quadro resultante do processo de erosão do passo anterior, ou seja, $F^n = L^{'''n}$ e que F^{n+1} igualmente representa um quadro resultante do mesmo processo, a partir dos objetos que permaneceram após a filtragem de tamanho aplicada, suponha um objeto O^i no quadro F^n, onde O^i denota um objeto rastreado. No próximo quadro F^{n+1}, dadas j regiões de movimento, R^j, o objetivo é saber qual R^j representa o objeto O^i do quadro anterior. A equação 3.6 representa a função de custo usada para este fim

$$Custo = (w_P * d_P) + (w_S * d_S) + (w_D * d_D) + d_T \tag{3.6}$$

no qual w_P ,w_S e w_D são pesos cuja soma é igual a 1. d_P é a distância Euclidiana entre os pixels que representam os centros dos objetos, d_S é a diferença de tamanho entre as *bounding boxes* das regiões de movimento, d_D é a diferença da direção entre a posição do objeto estimada pelo algoritmo do fluxo óptico proposto por de Lucas-Kanade (LUCAS; KANADE, 1981) no quadro atual e a diferença entre o centro da região de movimento e o centro do objeto, e d_T é a diferença do tempo de vida (*TTL-time to live*) do objeto. Estes parâmetros são melhores descritos nas equações a seguir.

A distância Euclidiana representa o custo da distância entre dois segmentos e é estimada pela equação 3.7

$$d_P = |R_c^j - O_c^i| \tag{3.7}$$

no qual R_c^j é ponto central da região de movimento no quadro atual e O_c^i é o ponto central do objeto no quadro anterior. O valor de d_P não deve ser maior que um limiar de proximidade P_{max} medido em pixels.

O custo da diferença de tamanho entre uma região de movimento e um objeto é calculada através da equação 3.8

$$d_S = \frac{|R_r^j - O_r^i|}{(R_r^j + O_r^i)} \tag{3.8}$$

no qual R_r^j e O_r^i denotam respectivamente o tamanho da região de movimento e o tamanho do objeto.

O custo da diferença da direção entre dois segmentos é dado pela equação 3.9

$$d_D = |arctan(O_s^i - O_c^i) - arctan(R_c^j - O_c^i)| \tag{3.9}$$

no qual O_s^i é o vetor de movimento (ponto central) do objeto i predito pela equação 3.11.

A persistência de um objeto é calculada pela equação 3.10

$$d_T = (TTL_{MAX} - O_{TTL}^i)/TTL_{MAX} \tag{3.10}$$

no qual TTL_{MAX} é a persistência máxima em quadros que um objeto pode ter e O_{TTL}^i é a persistência do objeto. Se o objeto é encontrado no quadro atual, o valor de O_{TTL}^i é igual a TTL_{MAX}. Se o objeto não foi encontrado, mas foi associado a uma região de movimento pelo algoritmo de predição, que será explicado logo a seguir, seu O_{TTL}^i é decrementado de um. Por último, se o objeto não for encontrado e nem associado a uma região de movimento seu O_{TTL}^i é decrementado de três até tornar-se igual a zero, onde o objeto deve ser eliminado do rastreamento. O valor de TTL_{MAX} foi configurado em três vezes a

taxa de quadros por segundos do vídeo, ou seja, três segundos.

Cada objeto do quadro anterior deve ser absorvido por um região de movimento no quadro atual guiado pelo menor custo. Os valores do objeto são atualizados com os valores da região de movimento associada a ele. Se uma região não foi associada a nenhum objeto, então um novo objeto é criado com os valores desta região. Se há um objeto que não foi associado a nenhuma região de movimento, este objeto pode estar ocluso e o algoritmo do fluxo óptico irá falhar na predição do movimento. Neste caso, o movimento destes objetos é predito seguindo a equação 3.11

$$O_s^i = S * O_s^i + (1 - S) * (R_c^j - O_c^i) \tag{3.11}$$

no qual S é um valor fixo de velocidade. A região de movimento R_c^j deve ser a mais próxima região do objeto, respeitando o limiar de proximidade P_{max}. Encontrado a região a que o objeto ocluso pertence, uma nova posição do objeto e de sua *bounding box* são previstas e computadas de acordo com as equações 3.12 e 3.13.

$$O_c^i = O_c^i + O_s^i \tag{3.12}$$

$$O_r^i = O_r^i + O_s^i \tag{3.13}$$

Desta forma o problema da oclusão é tratado pelo algoritmo de rastreamento. Com isso temos como resultado desta etapa os objetos rastreados quadro a quadro. Na próxima seção é apresentado a maneira como as pessoas são contadas, bem como os métodos propostos para a análise de grupos. Uma etapa muito importante do sistema que é aplicada sobre os objetos rastreados.

3.4 Contagem de Pessoas

Uma vez os objetos do primeiro plano segmentados, localizados e rastreados, a contagem de pessoas pode ser realizada. Como o monitoramento do ambiente é permanente, uma pessoa começa a ser rastreada assim que entra no campo de visão da câmera, porém ela somente será contada se caminhar para dentro de uma região de contagem virtual (LIU et al., 2005). A região de contagem é uma área, definida dentro do ambiente monitorado, onde os objetos são analisados para que seja efetuada a contagem das pessoas que passam por ali (Figura 3.6).

Nesta etapa, o principal ponto a ser levado em consideração é o tratamento da

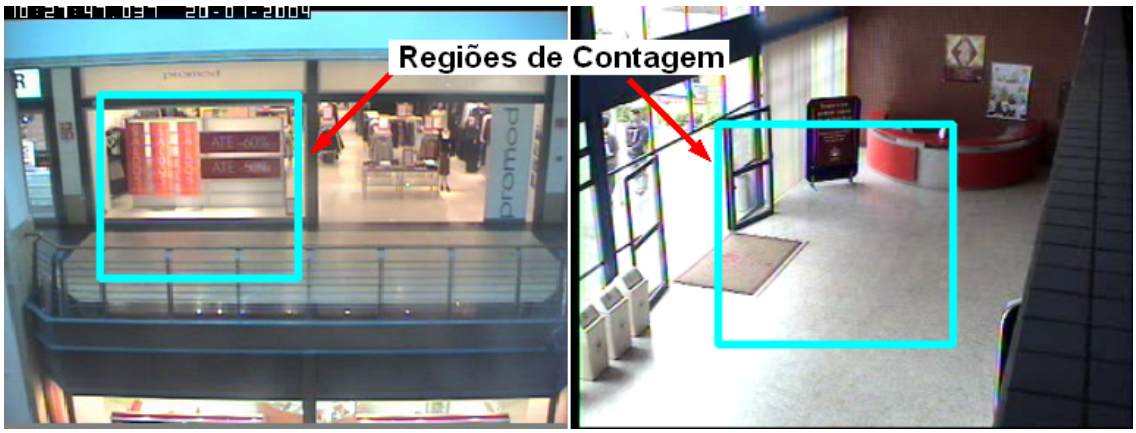

Figura 3.6: Exemplos de regiões de contagem

proximidade entre pessoas. Pois, um único objeto pode conter mais de uma pessoa, implicando em erros caso o problema não seja levado em consideração. Com base nisto, foram propostas duas abordagens para o tratamento deste problema. Uma baseada somente em limiares, e outra que faz o tratamento através de um classificador que analisa a região superior dos objetos, região da cabeça (Figura 1.3). A seguir as abordagens propostas para a análise dos objetos dentro da região de contagem são explicados detalhadamente.

3.4.1 Abordagem Baseada em Limiares

Esta abordagem é uma proposta simples para o problema da proximidade entre pessoas. Ela é aplicada logo que um objeto entra na região de contagem. A presente abordagem consiste em comparar a largura w do objeto analisado (Figura 3.7), obtida a partir de sua *bounding box*, com o limiar L que representa a largura que um objeto contendo uma única pessoa geralmente tem, empiricamente definido. Caso seu valor seja menor ou igual, incrementa-se uma pessoa ao contador, caso contrário, a área a da região superior deste objeto é computada e dividida com o limiar A que representa a área da região superior que um objeto contendo uma única pessoa geralmente tem, também empiricamente definido. Com isso, incrementa-se três ao contador se este valor for maior que 2 e dois caso contrário. Como mostra a equação 3.14. A região superior de um objeto é encontrada dividindo a altura de sua *boundig box* em quatro retângulos idênticos. O primeiro retângulo da extremidade superior da *bounding box* é a região que contém a cabeça do objeto (Figura 3.7).

Figura 3.7: Região superior de um objeto. w indica a largura da *bounding box* do objeto.

$$Count = \begin{cases} Count + 1, & \text{se } w \leq L \\ Count + 2, & \text{se } w > L \text{ e } a/A \leq 2 \\ Count + 3, & \text{se } w > L \text{ e } a/A > 2. \end{cases} \tag{3.14}$$

No qual $Count$ é a variável que armazena a contagem do sistema.

A Figura 3.8 exemplifica a situação de quando dois objetos se sobrepõem, ocorrendo assim a oclusão. O rastreador identificou os dois objetos tratando o problema da oclusão. O método baseado em limiares é aplicado em cada um dos objetos assim que invadam a região de contagem apenas uma vez. Depois de somados a contagem, eles não são mais verificados. Mesmo se sairem e entrarem novamente na região de contagem.

Embora seja uma abordagem simples e de rápida implementação, a necessidade de limiares fixos o torna não confiável, pois tende a cometer mais erros em situações triviais como a de diferenciar uma pessoa de um grupo. A seguir é apresentado uma nova abordagem no intuito de evitar este tipo de erro.

3.4.2 Análise da Região Superior do Objeto

Esta abordagem propõem a classificação dos objetos em três possíveis classes (“1”, “2”ou “3”pessoas) com o objetivo de diminuir os erros decorrentes do problema da proximidade entre pessoas e melhorar a confiabilidade do sistema em relação a abordagem baseada em limiares apresentada anteriormente.

O procedimento de contagem para este método se inicia assim que um objeto caminha para dentro de uma região de contagem. Em seguida, um conjunto de vetores de características $\mathcal{V}$ começa a ser extraído deste objeto, um vetor para cada quadro do vídeo. O número de vetores contidos no conjunto $\mathcal{V}$ varia de acordo com o tamanho da janela temporal. O processo de classificação é composto por dois estágios: primeiro, cada $V_i \in \mathcal{V}$ é classificado e rotulado com uma das classes; em seguida, baseando-se no rotulamento

realizado, a regra de voto da maioria é aplicado no conjunto $\mathcal{V}$ para decidir a qual classe o objeto pertence assim que ele sair de cena. Detalhes sobre as características usadas se encontram na seção 3.4.2.1 e sobre o classificador utilizado na seção 3.4.2.2.

3.4.2.1 Características

As características são extraídas da região superior do objeto, a qual contém a cabeça das pessoas, para construir os vetores de características usados na etapa de classificação. Esta região foi escolhida com base no posicionamento obliquo da câmera, onde acredita-se que informações nela contidas são suficientes para o tratamento de grupos de pessoas.

Os vetores de características são gerados a partir de um esquema de zoneamento e extração de características da região superior dos objetos. O esquema de zoneamento consiste em dividir verticalmente esta região em n sub-regiões iguais. Com isto, a área de cada sub-região é calculada e normalizada dividindo cada valor obtido pela área total da região superior do objeto. A área de uma região é a quantidade de pixel do primeiro plano que ela contém. Adicionalmente é atribuído mais uma característica ao vetor, o valor da largura da *boundin box* (Figura 3.10). Desta maneira são construídos os vetores de características com dimensão $n + 1$ para o sistema (Figura 3.9). Cada vetor de característica gerado para a base de treinamento é rotulado de acordo com a classe que representa ("uma", "duas"ou "3"pessoas), são extraídos do vídeo Z vetores representando todas as possíveis classes.

3.4.2.2 Classificador

O método proposto baseia-se em um classificador não-paramétrico de aprendizagem baseada em instâncias, mais especificamente o algoritmo dos k vizinhos mais próximos (k-*NN*) (AHA; KIBLER; ALBERT, 1991). Este classificador foi escolhido devido a simplicidade e a baixa dimensionalidade dos vetores de características. As instâncias são vetores de características de referência que compõem a base de treinamento Z.

Para o algoritmo do k-*NN*, a distância euclidiana entre os vetores de características do conjunto $\mathcal{V}$ e os Z vetores de referência é calculada. A distância Euclidiana entre um vetor de características V_i D-dimensional e um vetor de características de referência V_z é definida:

$$d(V_i, V_z) = \sqrt{\sum_{d=1}^{D} (V_{id} - V_{zd})^2} \tag{3.15}$$

Os k mais próximos vetores de características de referência irão rotular cada vetor do conjunto $\mathcal{V}$ com uma das possíveis classes. Depois de todos os vetores de $\mathcal{V}$ serem classificados, a decisão final da quantidade de pessoas contidas em um objeto é dada pelo voto de cada membro do conjunto $\mathcal{V}$, assim a classificação em "um", "dois"ou "três"é associada ao objeto de acordo com o voto da maioria. Por exemplo, se há oito vetores de características em $\mathcal{V}$ classificados pelo k-NN como "uma pessoa", dois como "duas pessoas"e um classificado como "três pessoas", o rótulo "uma pessoa"é associado ao objeto. Quando este objeto sair de cena, o número de pessoas que o rótulo associado a ele representa é somado a contagem.

A Figura 3.11 exemplifica a situação de quando ocorre oclusão dentro da região de contagem. Os objetos identificados como 2 e 3 entram na região e se sobrepõem. Observe que o rastreador foi capaz de tratar a oclusão não trocando os seus identificadores. Inicia-se a aplicação desta abordagem sobre os objetos que entram na região de contagem e só termina quando eles saem de cena.

Esta abordagem se mostrou mais confiável em relação a abordagem anterior, quase sempre acertando na diferenciação entre uma pessoa de um grupo de pessoas, sem precisar definir limiares. As duas abordagens são dependentes de uma boa performance na fase de segmentação, pois esta performance reflete diretamente na fase de contagem de pessoas. No próximo capítulo serão descritos todos os experimentos realizados, bem como os resultados obtidos para a avaliação das abordagens propostas.

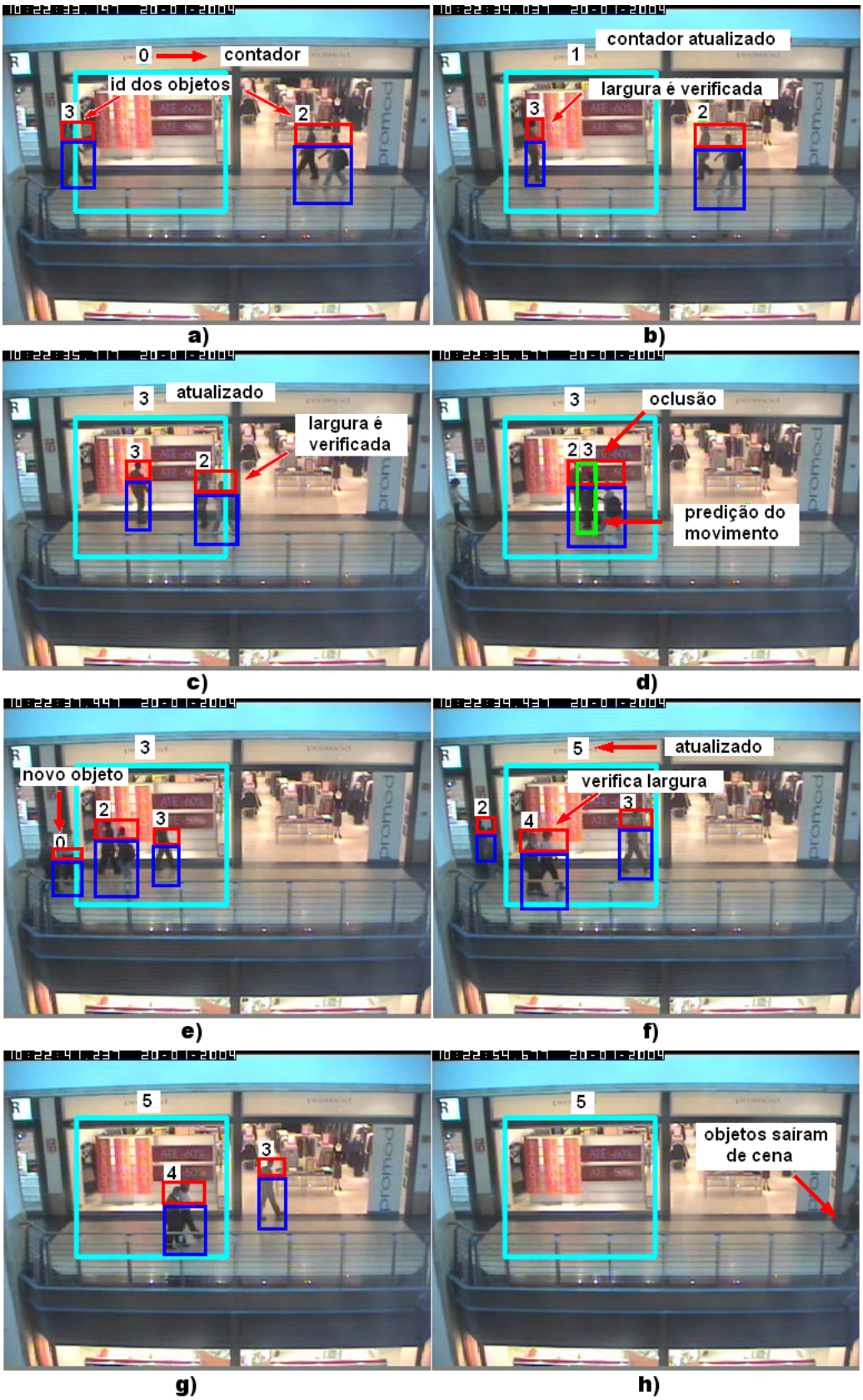

Figura 3.8: Método baseado em limiares. a)Objetos rastreados em movimento. b)Quando o objeto identificado como 3 entra na região de contagem sua largura e área são verificadas e o contador é atualizado. c)O mesmo acontece para o objeto identificado como 2. d)Oclusão entre os dois objetos, rastreador trata o problema prevendo o movimento do objeto 3. e)Novo objeto em cena. f)Objeto identificado como 4, entra na região de contagem, com isso sua largura e área são verificadas e o contador atualizado. g)Objetos saindo da região de contagem. h)Objetos saem de cena. Resultado da contagem: 5 pessoas.

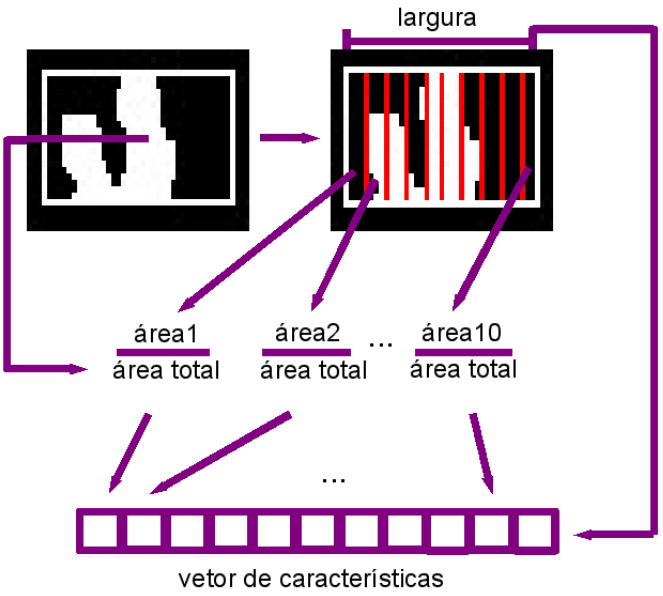

Figura 3.9: Geração do vetor de característica para n igual a 10. Exemplo da região superior de um objeto contendo duas pessoas

área 1 / área total	área 2 / área total	. . .	área n / área total	largura do *Blob*

Figura 3.10: Vetor de Características

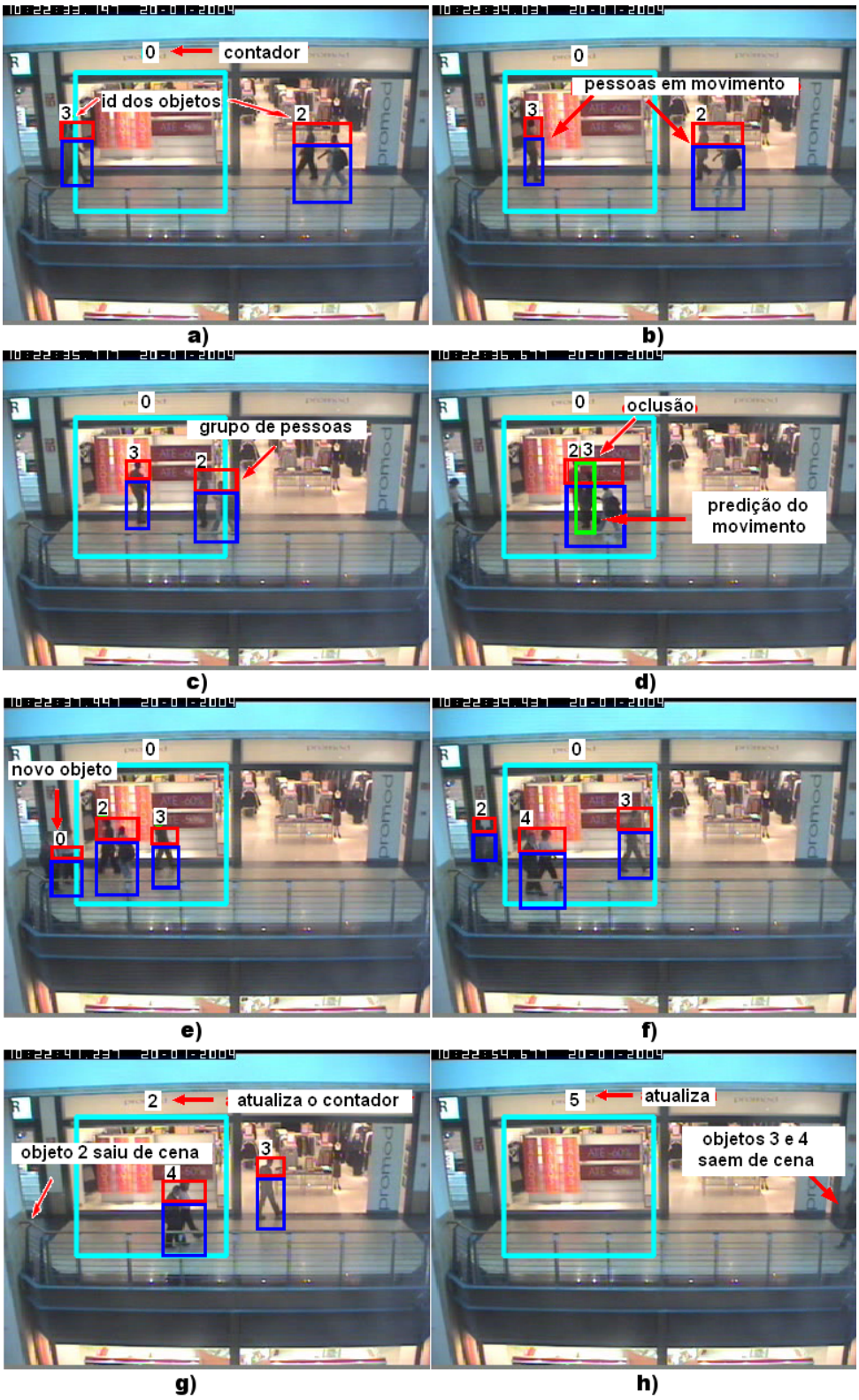

Figura 3.11: Método da Análise da região superior do objeto. a)Objetos rastreados em movimento. b)Quando o objeto identificado como 3 entra na região de contagem, a cada quadro do vídeo um vetor de característica é extraído dele e rotulado através do *k-NN*. c)o mesmo é feito pra o objeto identificado como 2. d)Oclusão entre os dois objetos, rastreador trata o problema prevendo o movimento do objeto 3. e)Novo objeto em cena. f) Quando o objeto identificado como 4 entra na região de contagem o processo descrito na letra a) é aplicado a ele. Objeto 2 sai da região de contagem, procedimento de extração e rotulação de vetores é interrompido para ele. g)Objeto 2 sai de cena, o voto da maioria é aplicado sobre seu vetores rotulados atualizando o contador. h)O mesmo acontece com os objetos 3 e 4. Resultado da contagem: 5 pessoas.

Capítulo 4

Experimentos e Resultados

Este capítulo apresenta os resultados dos experimentos realizados sobre os métodos propostos, descrevendo o plano experimental elaborado e discutindo os resultados obtidos. Na primeira parte descrevemos a base de dados utilizada para os testes. A segunda parte mostra a calibração do ambiente, ajustes de parâmetros. Já a terceira parte apresenta os resultados obtidos, tanto da contagem em geral quanto dos métodos que tratam a proximidade entre pessoas. Por fim, é apresentado uma discussão a respeito dos resultados obtidos.

4.1 Base de Dados

Os métodos propostos foram desenvolvidos e testados sobre dois conjuntos de videos digitais, sendo estes capturados em diferentes ambientes. O primeiro conjunto de vídeos fazem parte da base de dados CAVIAR[1] e foram filmados em um shopping center (Figura 4.1 a). Um total de 1337 segundos de vídeo simulados em um ambiente fechado com iluminação artificial. Possuem meia-resolução padrão PAL (384 x 288 pixels, 25 quadros por segundo).

O segundo conjunto de vídeos foram capturados no hall de entrada de um bloco acadêmico de uma universidade (Figura 4.1 b). Um total de 500 segundos de vídeo reais de um ambiente fechado com iluminação artificial e natural. Possuem meia-resolução padrão PAL (320 x 240 pixels, 25 quadros por segundo). Para cada ambiente, aproximadamente um minuto e meio de vídeo foi usado para treinamento e calibração das variáveis e o restante para teste.

Ambos conjuntos de vídeos possuem exemplos de proximidade entre pessoas, com grupos de duas, três e até mais de três pessoas caminhando juntas. A Tabela 4.1 apresenta

[1] http://homepages.inf.ed.ac.uk/rbf/CAVIAR/

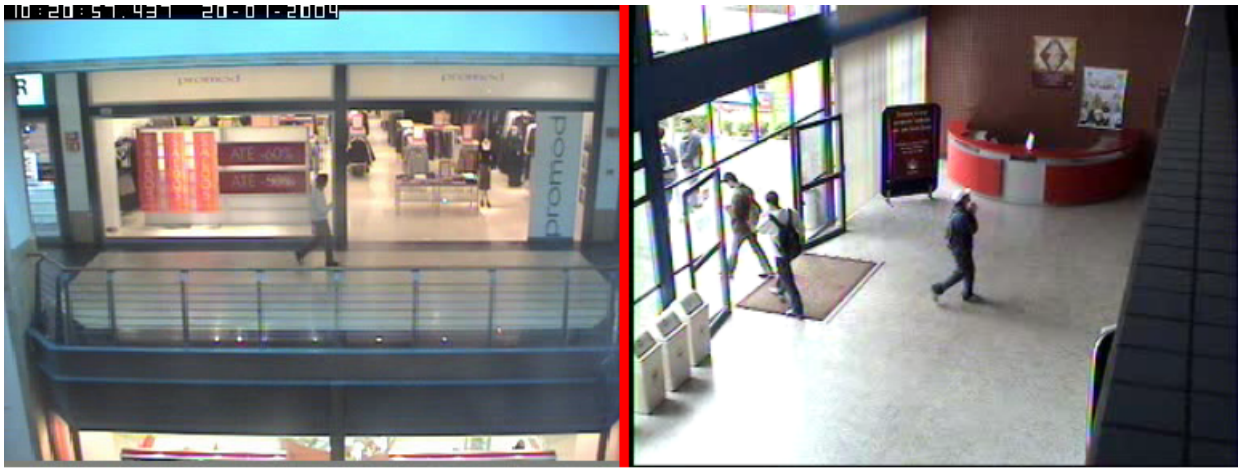

Figura 4.1: Vídeos da base de dados. a)quadro de um vídeo da CAVIAR Mall. b)quadro de um vídeo da Hall CCET

a distribuição das classes (uma, duas e três pessoas) nos dois ambientes, enquanto que a Tabela 4.2 mostra a quantidade total de pessoas que circularam dentro da região de contagem virtual de cada ambiente.

Tabela 4.1: Distribuição das classes

Classe	CAVIAR Mall n^o de amostras	CCET Hall n^o de amostras
1	49	45
2	14	20
3	4	11

Tabela 4.2: Quantidade total de pessoas

	N^o de Pessoas
CAVIAR Mall	92
CCET Hall	128

Como o processo de rotulação dos vídeos foi realizado manualmente, em alguns casos o resultado pode ser subjetivo pois depende da interpretação humana e isso pode causar ruído à base rotulada. Assim, a rotulação se deu assistindo os vídeos quadro a quadro e contando as pessoas que entravam em suas regiões de contagem, sendo que cada pessoa em cena foi contada somente uma vez. Em outras palavras, se uma pessoa entra na região de interesse, em seguida sai e entra de novo, ela não é contada duas vezes. Porém, se esta mesma pessoa sai de cena, depois aparece novamente e caminha para dentro de

uma região de contagem, ela é contada outra vez. Para a contagem das amostras de cada classe não foi diferente, quando um grupo de pessoas entrava na região de contagem era verificado quantas pessoas ele continha e assim somado a classe que pertencia. Sendo que grupos com mais de três pessoas foram considerados como parte da classe 3, devido a pouca quantidade de exemplos contidos nos vídeos da base. Deste modo o rotulamento dos dois ambientes foi criado para ser utilizado nos experimentos que são apresentados na próxima seção.

4.2 Calibração do Ambiente

Para realizar a avaliação do método proposto, primeiro foi necessário ajustar os parâmetros do sistema de acordo com o ambiente. Os valores foram ajustados de modo empírico sobre 90 segundos de vídeo de cada ambiente, onde havia exemplos de pessoas caminhando em grupos e sozinhas. Nas etapas de segmentação e rastreamento os mesmos valores foram usados nos dois conjuntos de vídeo, isto devido ao tamanho das pessoas que foram gravadas nos dois ambientes serem próximos. Alguns valores da etapa de rastreamento, como os pesos w_P, w_S e w_D, o limiar de proximidade P_{max} e a velocidade S foram ajustados de acordo com (LATECKI; MIEZIANKO, 2006). Os valores ajustados são apresentados por etapas nas tabelas seguintes, onde a Tabela 4.3 é referente a etapa de segmentação e a Tabela 4.4 a etapa de rastreamento.

Tabela 4.3: Calibração do ambiente fase segmentação

	CAVIAR Mall/ CCET Hall
Limiar l	25 (nível de cinza)
Dilatação $\oplus$	5 iterações
Erosão $\ominus$	5 iterações

Para a etapa de contagem, onde as abordagens propostas para tratar a proximidade entre pessoas são aplicadas, alguns valores a serem ajustados são distintos, ou seja, cada ambiente possui o seu próprio valor. A Tabela 4.5 apresenta o limiar utilizado pela abordagem Baseada em Limiares, o valor de n que representa o número de zonas em que a região superior do objeto foi dividida para o esquema de zoneamento e a quantidade de vetores extraídos dos vídeos para cada classe para compor o conjunto Z da abordagem Análise da Região da Cabeça do Objeto, onde cada conjunto de vídeo possui seus próprios vetores.

Como cada ambiente possui suas próprias características os ajustes dos parâmetros

Tabela 4.4: Calibração do ambiente fase rastreamento

	CAVIAR Mall/ CCET Hall
Altura mínima	30 pixels
Largura mínima	10 pixels
limiar P_{max}	40 pixels
peso w_P	0,4
peso w_S	0,1
peso w_D	0,5
velocidade S	0,9

Tabela 4.5: Calibração do ambiente fase contagem

	CAVIAR Mall	**CCET Hall**
limiar L	29 pixels	40 pixels
n	10	10
classe 1	30 vetores	30 vetores
classe 2	30 vetores	30 vetores
classe 3	30 vetores	30 vetores

apresentados é dependente do ambiente a ser monitorado. Por exemplo, a distância da câmera em relação ao ambiente monitorado influência o filtro de tamanho, ou seja, as variáveis referentes a largura e altura mínima mudam. A altura e a inclinação da câmera influenciam os vetores de características de referência. Já a iluminação influência diretamente a fase de segmentação. Por isso a importância da etapa de calibração.

4.3 Experimentos Realizados

Os experimentos buscaram avaliar os métodos propostos para contar pessoas, focando no problema da proximidade entre elas. Os primeiros experimentos consistiram em avaliar o resultado final da contagem, ou seja, comparar o resultado da contagem manual feita assistindo os vídeos com o da contagem automática resultante da aplicação das abordagens propostas. Os resultados obtidos são apresentados na Tabela 4.6, onde é mostrado a quantidade total contada por cada abordagem nos dois ambientes.

No intuito de ilustrar a necessidade de tratar o problema da proximidade entre pessoas também foi gerado resultados somente utilizando o algoritmo de rastreamento para realizar a contagem, pois ele não trata tal problema. Conforme se observa, o total de acerto de sua contagem foi de 80,43% em relação a contagem manual para a base CAVIAR Mall e de 73,74% para a base CCET Hall. Quando aplicado um tratamento

Tabela 4.6: Resultados das contagens automáticas

	Contagem Manual	Somente Rastreamento	Baseada em Limiares	Análise da Região Superior (11-nn)	(5-nn)	(1-nn)
CAVIAR	**92**	74	81	92	93	91
CCET	**128**	94	155	142	149	157

para o problema, neste caso a abordagem baseada em limiares, obteve-se 88,04% e 82,58% de acerto na contagem para o primeiro e segundo ambiente, respectivamente. Empregando a abordagem da análise da região superior do objeto para o tratamento de grupos a taxa de acerto para a base CAVIAR Mall e CCET Hall, para k igual a 11, foi de 100% e 90,14% respectivamente. Para k igual a 5, foi de 98,92% e 85,91%. Já para k igual a 1, foi de 98,91% e 81,53%. Devido ao fato do fluxo de pessoas da base CCET Hall ser maior que a CAVIAR Mall o algoritmo de rastreamento tende, em alguns momentos de muito movimento, a se perder com facilidade por conta da quantidade de oclusões ocorrendo ao mesmo tempo, errando a conta para mais.

Para avaliar o comportamento das abordagens acerca do problema da proximidade entre pessoas, outros experimentos se fizeram necessários. O principal objetivo destes experimentos foi de avaliar a capacidade de cada abordagem em estimar a quantidade de pessoas em um único objeto. Nas seções seguintes são apresentados os resultados destes experimentos.

4.3.1 Abordagem Baseada em Limiares Sobre Grupos de Pessoas

Os resultados obtidos aplicando-se a abordagem baseada em limiares em relação ao problema da proximidade entre pessoas podem ser observados na Tabela 4.7 e na Tabela 4.8. Elas mostram as matrizes de confusão, bem como a taxa de reconhecimento de cada classe para cada um dos ambientes testados. A Figura 4.2 mostra um exemplo de objeto em que a abordagem erra.

Tabela 4.7: Matriz de Confusão CAVIAR Mall - (Abordagem Baseada em Limiares)

Classe	1	2	3	Taxa Rec.
1	**44**	5	0	89,79%
2	8	**6**	0	42,85%
3	0	2	**2**	50%
			TOTAL	77,61%

Como pode ser observado há uma diferença entre o resultado total dos dois ambientes de aproximadamente 15 pontos percentuais. Isto se deve ao fato de que as pessoas

Tabela 4.8: Matriz de Confusão CCET Hall - (Abordagem Baseada em Limiares)

Classe	1	2	3	Taxa Rec.
1	**33**	7	5	73.33%
2	5	**9**	6	45%
3	0	5	**6**	54.54%
			TOTAL	63.15%

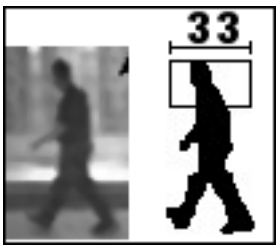

Figura 4.2: Exemplo de objeto onde a abordagem baseada em limiares falha contando duas pessoas ao invés de uma (limiar L=29). A abordagem da Análise da Região Superior do Objeto não falharia neste caso.

da base CCET Hall serem, em sua maioria, estudantes e, por conta disso, alguns andam com mochilas de costas, aumentando a largura do objeto que o representa e ficando maior que o limiar de uma pessoa.

4.3.2 Análise da Região Superior do Objeto Sobre Grupos de Pessoas

A Tabela 4.9 e a Tabela 4.10 apresentam os resultados obtidos através da aplicação da abordagem da análise da região superior do objeto para k igual a 11. Nelas são mostradas as matrizes de confusão para cada conjunto de vídeo testado e a taxa de reconhecimento de cada classe.

Tabela 4.9: Matriz de Confusão CAVIAR Mall - (Análise da Região Superior do Objeto)

Classe	1	2	3	Taxa Rec.
1	**49**	0	0	100%
2	3	**9**	2	64,28%
3	0	2	**2**	50%
			TOTAL	89,55%

Já na Tabela 4.11 e na Tabela 4.12 são apresentados as taxas de acerto da abordagem proposta em relação ao número de pessoas contidas em cada classe. Com base na contagem real (manual) e no resultado obtido pela abordagem aplicada, com o valor de k igual a 11, é mostrado a taxa de acerto total, bem como a taxa de acerto para cada

Tabela 4.10: Matriz de Confusão CCET Hall - (Análise da Região Superior do Objeto)

Classe	1	2	3	Taxa Rec.
1	**43**	2	0	95,55%
2	5	**14**	1	70%
3	0	7	**4**	36,36%
			TOTAL	80,26%

uma das classes separadamente. A Figura 4.3 mostra um exemplo de objeto em que a abordagem erra.

Tabela 4.11: Taxa de Acerto CAVIAR Mall (Análise da Região Superior do Objeto)

CAVIAR Mall Classe	Contagem Real	Contagem	Taxa de Acerto
1	49	49	100%
2	28	27	96,43%
3	12	10	83,33%
TOTAL	**89**	**86**	**96,37%**

Tabela 4.12: Taxa de Acerto CCET Hall (Análise da Região Superior do Objeto)

CCET Hall Classe	Contagem Real	Contagem	Taxa de Acerto
1	45	47	95,74%
2	40	36	90%
3	33	26	78,79%
TOTAL	**118**	**109**	**92,37%**

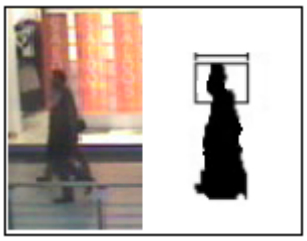

Figura 4.3: Exemplo de objeto onde a abordagem da Análise da Região Superior do Objeto falha contando uma pessoa ao invés de duas.

Como pode ser observado os resultados desta abordagem em relação as duas bases testadas são, de certa forma, próximos. Isto mostra que soluções baseada em limiares não são tão confiáveis quanto a baseada em classificadores.

4.4 Discussão

A partir dos experimentos descritos na seção anterior, observa-se a necessidade de se utilizar uma abordagem que trate do problema da proximidade entre pessoas. Afinal, sem qualquer tratamento o sistema comete muitos erros, pois não é capaz de diferenciar uma pessoa de grupos de pessoas e, principalmente, não é capaz de estimar a quantidade de pessoas nos grupos. Desta maneira, aplicando-se as abordagens propostas houve uma considerável melhora nos resultados obtidos (13 pontos percentuais, em média).

Na contagem geral, aquela que não observa o meio como foi obtido o resultado da contagem, onde somente interessa o resultado final, as duas abordagens propostas conseguiram resultados encorajadores. Porém, quando o meio é detalhado, observou-se que a abordagem da análise da região superior do objeto é mais confiável. Pois não necessita de limiares fixos, passível de muitos erros, melhorando a confiabilidade do sistema.

A maioria dos erros acontecem devido a falhas na etapa de segmentação. O baixo contraste entre as roupas, cor da pele, das pessoas com o fundo fazem com que elas sejam segmentadas deformadas Figura 4.4. Isto faz com que uma pessoa seja representada por dois ou mais segmentos, prejudicando a contagem. Uma falha na etapa de segmentação afeta todo o restante do sistema, principalmente na etapa de classificação, pois extrair características de um objeto deformado prejudica sua classificação.

Figura 4.4: Exmplo de falha na etapa de segmentação. A cabeça da pessoa foi perdida devido ao baixo contraste com o fundo.

Apesar dos resultados encorajadores, o método criado proposto precisa ser testado em outros ambientes que possuam mais amostras de pessoas caminhando em grupo, até mesmo para a geração dos vetores de características de referência, principalmente vetores que representam três pessoas. Pois se observou que grupos de três ou mais pessoas não são comuns de se encontrar, pelo menos nos ambientes testados.

Capítulo 5

Conclusão

Neste trabalho foi proposto um novo método para a contagem automática do fluxo de pessoas em ambientes através de câmeras de vídeos. Tal método visa oferecer um sistema que não necessite da instalação de câmeras dedicadas a ele. Em outras palavras, o sistema reutiliza as câmeras do circuito interno de TV (câmera de CFTV) devido ao posicionamento oblíquo destas em relação ao ambiente monitorado, o qual permite diferenciar uma pessoa de grupos de pessoas. Além disso, as abordagens propostas são capazes de contar e estimar a quantidade de pessoas em grupos, caminhando muito próximas umas das outras, na situação em que são segmentadas em um único objeto.

As abordagens foram testadas em dois ambientes, um simulado e o outro real. Elas apresentaram resultados encorajadores que podem ser analisados no capítulo 4. Além disso, devido ao pequeno conjunto de características e pelo método ser empregado sobre imagens em níveis de cinza a contagem acontece em tempo-real. Algumas contribuições a se destacar deste trabalho são: a possibilidade sobre economia de energia elétrica, o controle real das filas em estabelecimentos comerciais, a não necessidade de câmeras dedicadas a um sistema automático de contagem e a criação de duas abordagens que tratam o problema da proximidade entre pessoas.

Os resultados obtidos pelas abordagens propostas de contagem automática de pessoas estão entre 80% e 100% de taxa de acerto. SNIDARO; MICHELONI; CHIAVEDALE (2005) e KIM K.-S. CHOI; KO (2002) obtiveram em seus trabalhos resultados entre 90% e 100%, dependendo das condições dos testes. Levando-se em conta que em seus trabalhos foi utilizado o campo de visão ideal para a contagem de pessoas, ou seja, a câmera foi posicionada verticalmente ao ambiente monitorado, os resultados obtidos pelo método proposto estão próximos aos de trabalhos que possuem condições mais favoráveis para se realizar a contagem automática de pessoas.

Uma das desvantagens do modelo implementado, é que ele pode cometer muitos

erros em ambientes com muito movimento. Isto se deve ao método escolhido para a segmentação do movimento (separação fundo/primeiro plano), onde pode acontecer de todo ambiente ser segmentado como um único objeto por conta de sua superlotação. O rastreador usado também tende a cometer muitos erros em ambientes muito movimentado, onde oclusões ocorram seguidamente. Outra desvantagem que pode ser apontada é a necessidade da calibração manual do ambiente, melhor seria usar, se fosse possível, os mesmos valores para qualquer ambiente ou o ideal seria a calibração automática do ambiente.

Dentro dos aspectos que podem ser aprimorados para que aumente a confiabilidade do sistema destacam-se: estudar um método para o tratamento de objetos desconectados, por exemplo, utilizar a coerência dos movimentos para a realização desta tarefa; avaliar outras características que possam diferenciar ainda mais uma classe da outra, características de forma parece ser o caminho; por ser um dos grandes causadores de erros, pesquisar outra maneira de realizar a tarefa da segmentação; encontrar maneiras de lidar com ambientes muito movimentado/superlotados.

Referências Bibliográficas

AHA, D. W.; KIBLER, D.; ALBERT, M. K. Instance-based learning algorithms. *Machine Learning*, v. 6, n. 1, p. 37–66, 1991.

BJöRGVINSSON, T. *Peocounter-People Counting Software*. Dissertação (Mestrado) — Chalmers University of Technology, Gothenburg, 2006.

CUEVAS, E.; ZALDIVAR, D.; ROJAS, R. Kalman filter for vision tracking. In: *Technical Report B , Freie Universität Berlin, Fachbereich Mathematik und Informatik.* [S.l.: s.n.], 2005. p. 05–12.

DIAS, P. M. *Sistema de Contagem de Pessoas*. Dissertação (Mestrado) — Pontifícia Universidade Católica do Rio de Janeiro, Brasil, 2005.

HARITAOGLU, I.; FLICKNER, M. Attentive billboards: Towards to video based customer behavior. In: *WACV '02: Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision*. Washington, DC, USA: [s.n.], 2002. p. 127. ISBN 0-7695-1858-3.

HARITAOGLU, I.; HARWOOD, D.; DAVIS, L. S. W4: Real-time surveillance of people and their activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Los Alamitos, CA, USA, v. 22, n. 8, p. 809–830, 2000. ISSN 0162-8828.

HU, W. et al. A survey on visual surveillance of object motion and behaviors. *IEEE Transactions. Systems, Man, Cybernetics, Part C*, v. 34, n. 3, p. 334–352, 2004.

KASTRINAKI, V.; ZERVAKIS, M. E.; KALAITZAKIS, K. A survey of video processing techniques for traffic applications. *Image Vision Comput.*, v. 21, n. 4, p. 359–381, 2003.

KETTNAKER, V.; ZABIH, R. Counting people from multiple cameras. In: *ICMCS '99: Proceedings of the IEEE International Conference on Multimedia Computing and Systems Volume II-Volume 2*. Washington, DC, USA: IEEE Computer Society, 1999. p. 267–271. ISBN 0-7695-0253-9.

KHAN, S. et al. Human tracking in multiple cameras. In: *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. [S.l.: s.n.], 2001. v. 1, p. 331–336.

KIM, J.-W. et al. Real-time system for counting the number of passing people using a single camera. In: MICHAELIS, B.; KRELL, G. (Ed.). *DAGM-Symposium*. [S.l.]: Springer, 2003. (Lecture Notes in Computer Science, v. 2781), p. 466–473. ISBN 3-540-40861-4.

KIM K.-S. CHOI, B.-D. C. J.-W.; KO, S.-J. Real-time vision-based people counting system for the security door. *Proc. of 2002 International Technical Conference On CircuitsSystems Computers and Communications*, Coréia, 2002.

LATECKI, L. J.; MIEZIANKO, R. Object tracking with dynamic template update and occlusion detection. In: *18th Intl Conf on Pattern Recognition*. Washington, USA: [s.n.], 2006. p. 556–560. ISBN 0-7695-2521-0.

LEI, B.; XU, L. Q. From pixels to objects and trajectories: A generic real-time outdoor video surveillance system. In: *IEE Intl Symp Imaging for Crime Detection and Prevention*. London, UK: [s.n.], 2005. p. 117–122.

LIPTON, A. J.; FUJIYOSHI, H.; PATIL, R. S. Moving target classification and tracking from real-time video. In: *WACV '98: Proceedings of the 4th IEEE Workshop on Applications of Computer Vision (WACV'98)*. Washington, DC, USA: IEEE Computer Society, 1998. p. 8. ISBN 0-8186-8606-5.

LIU, X. et al. Detecting and counting people in surveillance applications. *avss*, IEEE Computer Society, Los Alamitos, CA, USA, v. 0, p. 306–311, 2005.

LUCAS, B. D.; KANADE, T. An iterative image registration technique with an application to stereo vision. In: *7th International Joint Conference on Artificial Intelligence*. Vancouver, Canada: [s.n.], 1981. p. 674–679.

MASOUD, O.; PAPANIKOLOPOULOS, N. A novel method for tracking and counting pedestrians in real-time using a single camera. *IEEE Transactions on Vehicular Technology*, p. 1267–1278, 2001.

POLAT, E.; YEASIN, M.; SHARMA, R. Tracking body parts of multiple people: A new approach. In: *WOMOT '01: Proceedings of the IEEE Workshop on Multi-Object Tracking (WOMOT'01)*. Washington, DC, USA: IEEE Computer Society, 2001. p. 35.

PRATI, A. et al. Shadow detection algorithms for traffic flow analysis: A comparative study. In: *Proceedings of the 4th IEEE International Conference on Intelligent Transportation Systems*. Oakland, CA: [s.n.], 2001. p. 340–345.

SHAPIRO, L. G.; STOCKMAN, G. C. *Computer Vision*. [S.l.]: Prentice Hall, 2001. 275-285 p.

SIDLA, O. et al. Pedestrian detection and tracking for counting applications in crowded situations. In: *AVSS '06: Proceedings of the IEEE International Conference on Video and Signal Based Surveillance*. Washington, DC, USA: IEEE Computer Society, 2006. p. 70. ISBN 0-7695-2688-8.

SNIDARO, L.; MICHELONI, C.; CHIAVEDALE, C. Video security for ambient intelligence. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, v. 35, n. 1, p. 133–144, 2005.

STAUFFER, C.; GRIMSON, W. E. L. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 8, p. 747–757, 2000.

TSAI, Y.-T.; SHIH, H.-C.; HUANG, C.-L. Multiple human objects tracking in crowded scenes. In: *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*. [S.l.: s.n.], 2006. p. 51–54. ISBN 0-7695-2521-0.

WREN, C. R. et al. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 19, n. 7, p. 780–785, 1997.

YILMAZ, A.; JAVED, O.; SHAH, M. Object tracking: A survey. *ACM Comput. Surv.*, ACM, New York, NY, USA, v. 38, n. 4, p. 13, 2006. ISSN 0360-0300.